Qualitative Reasoning

29th International Workshop on Qualitative Reasoning

July 11th, 2016, New York, USA

IJCAI 2016

In cooperation with

University of Amsterdam

INRA

Palo Alto Research Center

Preface

This volume contains the papers presented at the 29th International Workshop on Qualitative Reasoning held on July 11, 2016 in New York.

The Qualitative Reasoning (QR) community develops qualitative representations to understand the world from incomplete, imprecise, or uncertain data. Our qualitative models span natural systems (e.g., physics, biology, ecology, geology), social systems (e.g., economics, cultural decision-making), cognitive systems (e.g., conceptual learning, spatial reasoning, intelligent tutors, robotics), and more.

The QR community includes researchers in Artificial Intelligence, Engineering, Cognitive Science, Applied Mathematics, and Natural Sciences, commonly seeking to understand, develop, and exploit the ability to reason qualitatively. This broadly includes:

- Developing new formalisms and algorithms for qualitative reasoning.
- Building and evaluating predictive, prescriptive, diagnostic, or explanatory qualitative models in novel domains.
- Characterizing how humans learn and reason qualitatively about the (physical) world with incomplete knowledge.
- Developing novel, formal representations to describe central aspects of our world: time, space, change, uncertainty, causality, and continuity.

The International Workshop on QR provides a forum for researchers from multiple perspectives to share research progress toward these goals. Topics of interest include:

- Qualitative modeling in physical, biological and social sciences, and in engineering.
- Representations and techniques for QR.
- Methods that integrate QR with other forms of knowledge representation, including quantitative methods, machine learning and other formalisms.
- Using QR for diagnosis, design, and monitoring of physical systems.
- Applications of QR, including education, science, and engineering.
- Cognitive models of QR, including the use of existing QR formalisms for cognitive modeling and results from other areas of cognitive science for qualitative reasoning.
- Using QR in understanding language, decision-making, sketches, images, and other kinds of signals and data sources.
- Formalization, axiomatization, and mathematical foundations of QR.

July 2, 2016
Amsterdam

Bert Bredeweg
Kamal Kansou
Matthew Klenk

Program Chairs

Bert Bredeweg	University of Amsterdam, The Netherlands
Kamal Kansou	Centre de recherche Angers-Nantes, France
Matthew Klenk	Palo Alto Research Center, USA

Program Committee

Nuria Agell	ESADE, Ramon Llull University, Spain
Elena Andonova	New Bulgarian University, Bulgaria
Mehul Bhatt	University of Bremen, Germany
Stefano Borgo	Laboratory for Applied Ontology, ISTC-CNR, Italy
Ivan Bratko	University of Ljubljana, Slovenia
Bert Bredeweg	University of Amsterdam, The Netherlands
George Coghill	University of Aberdeen, Scotland, UK
Johan de Kleer	Palo Alto Research Center, USA
Zoe Falomir	University of Bremen, Germany
Ken Forbus	Northwestern University, USA
Christian Freksa	University of Bremen, Germany
Scott Friedman	SIFT, LLC, USA
Tomoya Horiguchi	Kobe University, Japan
Liliana Ironi	IMATI-CNR, Italy
Kamal Kansou	INRA, France
Matthew Klenk	Palo Alto Research Center, USA
Andrew Lovett	Northwestern University, USA
Wei Pang	University of Aberdeen, Scotland, UK
Chris Price	Aberystwyth University, Wales, UK
Paulo Salles	University of Brasilia, Brazil
Qiang Shen	Aberystwyth University, Wales, UK
Peter Struss	Technical University Munich, Germany
Thora Tenbrink	Bangor University, Wales, UK
Stefania Tentoni	IMATI-CNR, Italy
Louise Travé-Massuyès	LAAS-CNRS, France
Jure Zabkar	University of Ljubljana, Slovenia

Additional Reviewers

Dylla, Frank
Grigoleit, Florian

Table of Contents

Assessing learner-constructed conceptual models and simulations of dynamic systems*

Bert Bredeweg[1] and Jochem Liem[1] and Christiana Nicolaou[2]

[1] Informatics Institute, University of Amsterdam, The Netherlands
B.Bredeweg@uva.nl, Jochem.Liem@gmail.com
[2] Department of Educational Studies, University of Cyprus, Cyprus
chr.nic@ucy.ac.cy

Abstract. Learning by conceptual modeling is seeing uptake in secondary and higher education. However, assessment of conceptual models is underdeveloped. This paper proposes an assessment method for conceptual models. The method is based on a metric that includes 36 types of issues that diminish model features. The approach was applied by educators and positively evaluated. It was considered useful and the derived grades corresponded with their intuitions about the models quality.

Keywords: Assessment, Conceptual modeling and simulation, Dynamic systems, Systems thinking

1 Introduction

Acquiring knowledge by constructing and using models is seeing uptake in secondary and higher education [5]. Recently, the approach is applied in a novel way using *conceptual models* and accompanying tools, which allow modelers to develop and simulate conceptual representations of dynamic systems [9, 2].

To implement modeling in classroom practice, formative and summative assessment techniques [7] for evaluating learner-constructed models are indispensable [19]. Assessment is one of the four vital parameters for science education, together with curriculum, instruction and professional development [17]. However, the assessment of conceptual models is underdeveloped, hampering its usage [4]. This means that there is a lack of criteria of what constitutes a good conceptual model. Consequently, it is difficult to give feedback to learners regarding the models they create. The problem is even more pressing when learning is self-regulated, and (groups of) learners develop their own unique models with different viewpoints, conceptualisations, and levels of abstraction. Comparison between learner-constructed models, and even comparison with a norm model, becomes impractical and inadequate for assessment.

This paper focusses on how assessment of *conceptual* models can be performed. The central idea is that learner-constructed conceptual models are rich

* Research co-funded by EU FP7, Project no. 231526, http://www.dynalearn.eu.

representations, and as such provide evidence of learning. Particularly, the number of correctly modeled ingredients compared to the total number of model ingredients (determined through a catalogue of modeling suboptimalities) can function as a measure of the modeling competence of the learner. This evidence can be identified, enumerated and scored by an assessment method and as such provide the basis for feedback, both formative and summative, and for learners and teachers. Hence, the question guiding the presented research is: What are the main components of an assessment method which can successfully evaluate diverse and different learner-constructed conceptual models?

2 Educational context and relevance

A scientific model is a construct that represents a system, and that consists of a set of objects and their properties, and a number of law statements indicating the behaviors of these objects in terms of their properties [3]. A *conceptual model* is a special kind of model that represents the referents in the domain through particular concepts as distinguished by the modeling language. For instance, it represents an explicit conceptual account of the physical structural and the behavioral features of the system under study, as well as the network of causal relationships underlying the behavior of the systems [10]. *Modeling competence* refers to the ability to construct and improve models [6].

Computer modeling is widely advocated as a way to offer students a deeper understanding of (complex) systems [14]. Consequently, the need for learners to master modeling competencies, e.g. being able to perform proper cause-effect reasoning. However, acquiring this competence is not so easily accomplished. Modeling is complex and both teachers and learners need to be well supported in order to successfully engage in modeling activities [18].

Learning by modeling is a process of engaging learners in (co)constructing models to gain understanding of systems. It is intrinsically related to the constructivist approach to teaching and learning, which is based on the idea that learners, through the use of the appropriate tools, construct their knowledge through building artifacts, here conceptual models. These artifacts encompass evidence of knowledge and skills on behalf of the learner and as such are rich sources of information of their modeling competence.

A model assessment instrument could thus provide valuable support for all stakeholders. However, well-suited methods for assessing conceptual models are sparse [13]. Some of the existing approaches use norm models [12, 18]. That is, the learner-constructed model is compared to a norm model and then scored. However, such approaches do not provide tools that systematically address deviations in learner-constructed models. Deviations that sometimes are erroneous but often also valuable variations on the norm. Moreover, in the context of self-directed learning activities, learners vary on topics, levels of granularity, perspectives and assumptions taken, etc., leading to a significant yet natural variation in the models constructed, which makes the a priory construction of norm models impractical (if not impossible).

Some other approaches use open-ended techniques addressing the model as a whole and evaluate paper-based models (drawings) [1]. The open-ended methods score the models on very general features, such as comparison and abstraction, while drawings models are not dynamic by nature. Both are limited evaluation procedures by design. Important details may be overlooked by the assessor and the scoring may end up being based on irrelevant or incorrect evidence.

In summary, assessment of learner-constructed models is important, yet usable methods are sparse. This hampers the use of modeling as an educational instrument. The work presented in this paper addresses the problem, particularly concerning the assessment of conceptual models.

3 Conceptual models and assessment needs

Our research addresses science education and particularly the challenge of making *learning by modeling* common practice in secondary education. We focus on conceptual models (as opposed to numerical) because they allow learners to directly interact with vocabulary that is necessary for the conceptual understanding they need to acquire. As a modeling tool we use DynaLearn [2], which has been used successfully in different educational settings as a workbench for learners to develop their understanding of how systems work (cf. [15]). The full workbench provides a sequence of workspaces with increasing complexity that facilitates a stepwise approach toward developing conceptual modeling expertise (for details see [11], Ch. 3 & 4).

3.1 Learner-constructed models - Identifying suboptimalities

Consider the learner-constructed model shown in Fig. 1. It was created during a course on conceptual modeling, within an environmental science bachelor, in which learners worked through a series of modeling assignments using DynaLearn (Learning Space 4, LS4). For the final, inquiry-based assignment, learners were asked to choose a system based on their interest, pose a question about that system and develop a model that answers this question. There were no norm models. The only constraint was that at least two processes causing change in the system were modeled. The learners worked in pairs. Model assessment in the context of such a self-regulated learning activity is quite a challenge.

Let us start by interpreting the domain details shown in the diagram. The model represents a field of quinoa being irrigated using salt water. The amount of water absorbed by the quinoa is determined by the concentration of salt in the roots of the quinoa and the salinity of the earth near the roots of the quinoa. As water is absorbed, the quinoa grows and the yield increases.

There are no major issues with the representation of the physical structure of the system, although *Seeds* (and *Saponin*) can be considered superfluous. Quantities, on the other hand, can be improved. *Volume* of *Salt water* is positively influencing *Soil saturation*. However, causal dependencies of type *I-* or *I+* are used for processes, while in the model the dependency seems to be a proportionality, that is *P-* or *P+*. Hence, this can be considered an incorrect causal relation (issue #20[3]) in the model. However, the model makes more sense if the volume quantity is interpreted as the irrigation process. Therefore, this issue is considered to affect the correctness of the model.

Quantity *Root zone salinity* refers to a mixture of notions including an entity and a quantity. As a result, it can be conceptually decomposed (issue #9). The simplest solution is to rename the quantity *Salinity*. Similarly, *Root salt concentration* can be conceptually decomposed (issue #9). The details in the model representing the physical structure of the system can be augmented by explicitly modeling the roots of the quinoa and indicating that these roots contain salt. This salt entity should have a quantity concentration.

The quantity spaces of *Root salt concentration* and *Root zone salinity* can be improved. There is no clear distinct behavior associated with reaching the landmark *Boundary* (issue #14). Consequently, this value and the value *Higher* can be removed. Secondly, the value *Higher* is vague (issue #15). That is, it is context dependent (higher compared to what?). Renaming this value to whatever happens above the value *Boundary*, or removing the value, would resolve it.

Causality has 2 issues. First, quantity *Root zone salinity* is affected by both a positive influence (from *Water uptake*) and a positive proportionality (from *Soil saturation*). Mixing causality types is incorrect (issue #23). Either a quantity is affected by a process directly, or change propagates to this quantity. In this case the proportionality should be removed. Second, when there is no more water in the soil, there can be no more water uptake (which is modeled using a value correspondence between the magnitudes *Zero* of *Water update* and *Zero* of *Soil saturation*). However, for this to occur, *Water uptake* should decrease as *Soil saturation* decreases. This can be modeled using a positive proportionality from *Soil saturation* to *Water uptake*. This is missing in the model (issue #21).

There are 4 issues with inequalities and correspondences, all resulting in inconsistencies (issue #24) when simulating: value correspondence from *Volume* of *Salt water* to *Soil saturation*, from *Volume* of *Salt water* (derivative) to *Soil saturation* (derivative), and the two correspondences from *Water uptake* to *Growth*.

Finally, simulation has 2 issues (Fig. 2). First, quantity *Soil saturation* has no value (issue #32). Second, quantity *Root salt concentration* has the value *Plus* and is decreasing in state 3, but never reaches *Zero*. This is a so-called *dead-end* (issue #34), caused by an inconsistency.

4 Instrument for assessing conceptual models

Within the conceptual modeling community, there is the belief that *"(...) a conceptual model can only be evaluated against people's (tacit) needs, desires and expectations. Thus the evaluation of conceptual models is by nature a social rather than a technical process, which is inherently subjective and difficult to*

[3] Our method identifies 36 issue types, each with a unique number (see Section 4).

formalise" [16]. We argue that it is possible to elevate model assessment from being a social process to one that is largely standardized and objective.

Our approach is based on the notions of *verification* and *validation*. Verification involves determining whether a product satisfies the conditions defined before development [21]. For a software program, knowledge base, or scientific model, such conditions typically include adhering to the syntactical and semantic requirements of the formalism used to develop the product. By contrast, validation determines whether the product performs adequately for its intended purpose and is satisfactory for the end user. As such, verification can be considered the assessment of internal (or internalized) *quality characteristics*, while validation tests external (purpose-oriented) quality characteristics [16].

Appraising internal quality characteristics (verification) should be an *objective* task. For example, conceptual models that allow for inferences (e.g., simulation) have an internal logic that imposes constraints that can be checked automatically. By contrast, validation is more subjective as a result of being domain and goal dependent. For example, different experts may disagree on whether a model is a *correct* domain representation [20] and can cite different resources to support their case. Here, we focus particularly on verification.

We propose *model features* that attest to the quality of a model (Table 1). These features are categorized into two verification categories. First, *formalism features* apply only to conceptual models developed in formalisms that allow for inferences, such as DynaLearn [2]. These features can be assessed using the internal logic of the formalism (e.g., consistency). The second category, *domain features* apply to conceptual models generally, and rely on the human interpretation of the model to be assessed. For example, the model feature *conformance to ontological commitments* requires that a referent in the domain is represented using the correct model ingredient in the formalism (e.g., biomass should be represented as a quantity). We claim both features can be checked objectively. Algorithms can be created to automatically detect them and suggest corrections.

Next step is to determine which model characteristics can be used to actually measure the quality of a conceptual model in terms of formalism and domain features. Correctness, completeness, and parsimony have been proposed as such quality characteristics (e.g. [20]). *Correctness* indicates that a model is free from errors. *Completeness* means that everything of relevance is included in the model. *Parsimony* implies that the model does not include redundancies. The following sections identify model features that attest to these quality characteristics.

4.1 Formalism-based features

Consistency is a prerequisite for the *correctness* of a conceptual model, and requires that ingredients in the model do not contradict each other (in terms of the possible inferences). For example, a quantity cannot be increasing and decreasing at the same time. *No unassigned variables* is a model feature that is important for the *completeness* of a model. An unassigned variable after reasoning (e.g., simulation) is an indicator that information in the model is missing to allow a particular reasoning step to succeed. *Reasoning relevance* means that

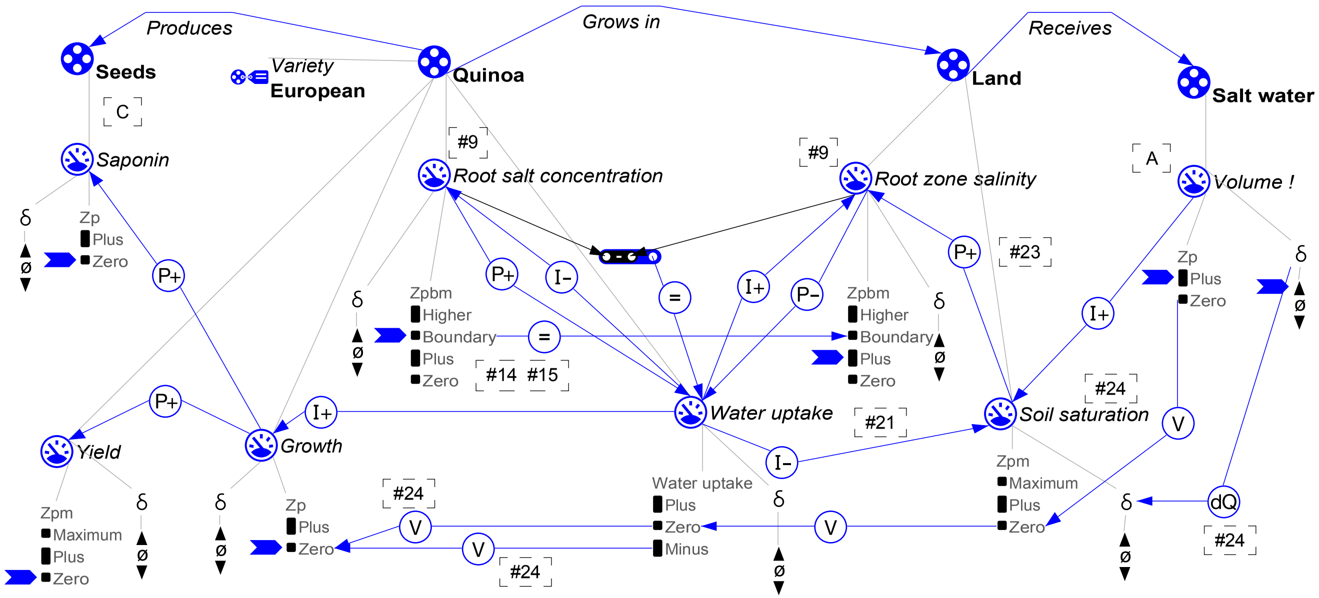


Fig. 1. Learner-constructed DynaLearn [2] model, using learning space 4, modeling the effects of watering quinoa using salt water. The amount of water absorbed (*Water uptake*, WU) by the quinoa is determined by the concentration of salt in the roots (*Root salt concentration*, RSC) of the quinoa and the salinity (*Root zone salinity*, RZS) of the earth near the roots of the quinoa ($RSC - RZS = WU$). As water is absorbed, the quinoa grows (*Water uptake* $I+$ *Growth*) and the yield increases (*Growth* $P+$ *Yield*). Particularly well-modeled is the so-called *equilibrium seeking mechanism* that determines the uptake of water, which consists of two negative feedback loops. The water uptake (if *Water uptake* = *Plus*) decreases the salt concentration in the roots of the quinoa ($I-$), and increases the salinity of the soil surrounding the roots ($I+$). The water uptake decreases as the salt concentration in the root decreases ($P+$), and the water uptake also decreases as the salinity of the soil surrounding the roots increases ($P-$). Note, the layout has been changed by the authors. The model issue numbers (verification) and the validation issues (A: correctness, C: parsimony) are indicated in dashed boxes.

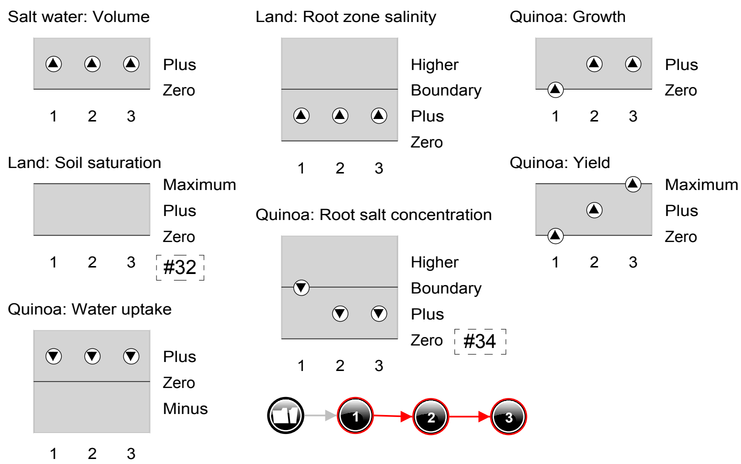


Fig. 2. The state-graph (4 connected circles) and value history (7 squares) of the quinoa model (Fig. 1). The model issues (#32 and #34) are indicated in dashed boxes.

Table 1. Model features that attest to quality characteristics of verification categories.

Verification Category	*Quality Characteristic*	*Model Feature*
Formalism	Correctness	Consistency
	Completeness	No unassigned variables
	Parsimony	Reasoning relevance
Domain representation	Correctness	Conformance to ontological commitments
		Falsifiability
	Completeness	Conceptual decomposition
		No missing representations
	Parsimony	No repetition
		No synonyms

each of the elements in the representation should have a function in terms of the reasoning. If not, the ingredient is superfluous and the model not *parsimonious*. For example, including a quantity without relating it to other quantities.

4.2 Domain representation-based features

Two domain features contribute to the *correctness*. *Conformance to ontological commitments* indicates that referents in the domain are represented using the correct model ingredients. For example, biomass being represented as an entity is an example of a type error. *Falsifiability* is the property of a claim, hypothesis or theory, to be proven false if the 'outcome' cannot be observed in reality. A conceptual model is falsifiable if its simulation results can be shown to be false through comparison with observations. Using vague values, such as 'small' or 'large', is an example of what makes a model unfalsifiable, as it becomes unclear what observations would conflict with the model's simulation results.

Two domain features contribute to *completeness*. *Conceptual decomposition*, which can be called the *'single concept per model ingredient rule'*, states that model ingredients that represent aggregated concepts should be broken down into multiple ingredients. For example, the use of quantities *Water temperature* and *Air temperature* can be an indicator that *Temperature* is a missing independent model ingredient that should have its own representation. As a guideline, a model ingredient can be considered conceptually decomposed when the represented concept can be found in an encyclopaedia, dictionary or glossary. *No missing representations* means that referents that are important in the domain are represented.[4] For example, given that *Mortality* and *Population size* are represented, there has to be a causal relation between these quantities.

Two domain features contribute to *parsimony*. *No synonyms* means that a domain concept, such as natality, should only be represented once, and consequently identified using a unique term. Hence, a model in which both *Natality* and *Birth rate* occur breaks this rule. Thesauri can be helpful in determining whether two terms are synonyms. *No repetition* indicates that there are no reoccurring arrangements of related ingredients. Such arrangements should be represented once and reused throughout the model (only at learning space 6 [2]).

4.3 Assessment metric

We have developed a best practice for conceptual modeling in the form of a catalogue of 36 modeling issues, checks to detect them, and modeling actions to ameliorate them (available via ([11] Ch. 5 (p. 99) and App. A.1 (p. 201), section 3.1 gives examples). Each of the issues affects one or more of the model features and thus the overall model quality. The issues are categorized based on whether they affect particular model ingredients, namely (*i*) Structure, (*ii*) Quantities, (*iii*) Quantity spaces, (*iv*) Causality, (*v*) Inequalities and correspondences, (*vi*) Model fragments (only at learning space 6), and (*vii*) Simulation results. For instance, issues #14 en #15 (see Section 3.1) both affect Quantity spaces. Next, we have established a metric that reflects a model's overall quality, based on the best practice (Table 2[5]). The quality metric results in a score between 0 and 100, which, when interpreted as a percentage, can be converted to grades.

How particular assessment categories are weighted is subjective. To minimize the potential for contention about the overall quality metric, we take the position that 50% (or more) of the overall quality measure should be based on objective criteria (hence verification). The other half of the weight is meant for model validation and is equally distributed between how well the model functions as

[4] May contribute to internal and external characteristics. Here the focus is on internal.
[5] Validation is not addressed in this paper. It is assessed using a rubric, see [11].

Table 2. Model assessment categories and weights.

Assessment categories	*Subcategories*	*Weight*
Verification: Model issues (50%)	Structure	10.00%
	Quantities	5.00%
	Quantity spaces	5.00%
	Causality	10.00%
	Inequalities and correspondences	5.00%
	Model fragments	5.00%
	Simulations	10.00%
Validation: adequate domain representation for goal (25%)	Correctness	10.00%
	Completeness	10.00%
	Parsimonious	5.00%
Validation: Communication (25%)	Layout of the model	5.00%
	Documentation	20.00%

a domain representation, and how well the model suites communication. Of course, when deemed appropriate users can change the distribution emphasizing different aspects of conceptual modeling for a particular assignment.

Given the proposed weights, the metric should reflect both those things that have been done correctly and the errors that have been made, as learners need to learn both from their errors, and be motivated by those aspects of modeling that they have done correctly. This results in the following calculation (shown for the Structure ingredients):

$$\text{Structure score} = 100 \times \frac{\#\text{entity + configuration definitions} - \#\text{structure issues}}{\#\text{entity + configuration definitions}}$$

When applying the metric, something is counted as an issue if it requires a single correction. As such, repeated reuse of an entity that is not conceptually decomposed counts as a single issue. However, repeated issues of the same type are counted as individual issues. Also, mistakes in smaller models are penalized more heavily, compared to mistakes in larger models. This is done by basing the scores on the ratio between the correctly modeled part and the whole model.

When all the steps needed to grade a model have been taken, the final score can be calculated. For the model in Fig. 1, the results are as follows:

Structure	$= 100 \times \frac{4+3-0}{4+3}$	$= 100.0$ (11%)	
Quantities	$= 100 \times \frac{8-2}{8}$	$= 75.0$ (6%)	
Quantity spaces	$= 100 \times \frac{4-2}{4}$	$= 50.0$ (6%)	Correctness $= 60$ (10%)
Causality	$= 100 \times \frac{10+1-2}{10+1}$	$= 81.8$ (11%)	Completeness $= 100$ (10%)
Ineq. and corresp.	$= 100 \times \frac{5+2-4}{5+2}$	$= 42.9$ (5%)	Parsimony $= 80$ (5%) Layout $= 80$ (5%)
Simulation	$= 100 \times \frac{3-2}{3}$	$= 33.3$ (11%)	Documentation $= 80$ (20%)

The weight for model fragments (not available in LS4) is distributed over the other verification subcategories (except inequalities and correspondences), hence 11 instead 10%, 6 instead of 5%, etc. The causality score is adjusted because a causal relation was missing (issue #21, Section 3.1). Consequently, a causal relation is added to the total number of Causality (10+1). Similarly, mistakes as subtracted: Quantities 8-2 (2x issue #9), Quantity spaces 4-2 (issue #14 & #15), Causality 11-2 (2x issue #23), Ineq. and corresp. 7-4 (4x issue #24), and Simulation (3-2) (issue #32 & #34).

Validation is not discussed here. However, as mentioned before, correctness, completeness, parsimony, layout and documentation are graded using a rubric. The results are shown above, RHS. The final score is 73.3.

5 Evaluating the assessment method

A pilot study was conducted with four evaluators who used the instrument to grade 34 models submitted by the student pairs in the course (two evaluators graded 9 models). The pilot focussed on whether the grades derived using the assessment method are comparable to grades that evaluators proclaim a model deserves. To this end, before having graded any models, the evaluators were asked to *intuitively* grade one set of models assigned to another evaluator[6]. The instruction was to analyse each model for 5 minutes, write down the grade, and proceed to the next model.

The *agreement* between the intuitive and actual grades was calculated. For this the different evaluators are assumed equal, and therefore all assessment method grades are considered of one evaluator (34 grades), and all intuitive grades of another (34 + 10 = 44 grades) (data available via [11] Ch. 5, p. 140). Typical statistical methods for inter-rater agreement (Cohen's kappa and Fleiss' kappa) cannot be used as they require a fixed number of mutually exclusive categories. IntraClass Correlation (ICC) and the Concordance Correlation Coefficient (CCC) can be used. Both were calculated, and both indicate strong agreement of about 0.89 ($r^{ICC} = 0.887$, 99%-confidence interval: $0.765 < r^{ICC}$

[6] One evaluator coincidently graded 2 sets.

< 0.947, $r^{CCC} = 0.885$, 99%-confidence interval: $0.767 < r^{CCC} < 0.945$). Suggesting the method's grades are acceptable.

Evaluators were able to detect model issues easily and only had difficulty in understanding one issue (#9. Ambiguous process rate *quantities*). This suggests that the assessment method is understandable and usable for evaluators. The evaluators required about 45 minutes per model to derive grades. As the model contributed 40% of the final grade, 45 minutes was considered reasonable.

6 Conclusion and discussion

Assessment of learner-constructed models is of great importance for effective development of the modeling competence on behalf of learners, and enabling learning by modeling as common practice in classrooms. Yet, ready to use assessment methods are sparse. We propose an assessment instrument based on a set of model features that attest to the quality of *conceptual* models. The model features address verification, and are categorized as formalism and domain features. The former apply only to conceptual models that allow for inferences, while the latter apply generally. The model features are further categorized as attesting to the quality characteristics correctness, completeness and parsimony.

A pilot study using the assessment method suggests that the derived grades correspond to evaluators' intuition of what a model is worth. The assessment method proved understandable, and the time required to apply it is considered reasonable. A listing of all the issues in a model serves as both an argument why a particular grade was given and as valuable feedback for learners.

As ongoing research we are investigating how the presented approach can be used as a real-time operating instrument, particularly for formative assessment, which requires automated detection of modeling issues. When issues are detected automatically, feedback may also be automated, but can also be left to the teacher. Another interesting future challenge would be to extend the current approach to the assessment of models created by domain experts, such as [8].

References

1. Bamberger, Y. M., Davis, E. A. (2013). Middle-School Science Students Scientific Modelling Performances across Content Areas and within a Learning Progression. International Journal of Science Education, 35(2), 213-238.
2. Bredeweg, B., Liem, J., Beek, W., Linnebank, F., Gracia, J., Lozano, E., Wißner, M., Bühling, R., Salles, P., Noble, R., Zitek, A., Borisova, P., Mioduser, D. (2013). DynaLearn - An Intelligent Learning Environment for Learning Conceptual Knowledge. AI Magazine, 34(4), 46-65.
3. Bunge, M. (1983). Treatise on Basic Philosophy: Volume 5: Epistemology & Methodology I: Exploring the World. Springer. Dordrecht: Reidel.
4. Eurydice: The information network on education in Europe. (2006). Science teaching in schools in Europe. DG for Education and Culture, Brussels.

5. Goberta, J.D., O'Dwyer, L., Horwitz, P., Buckley, B.C., Levy, S.T., Wilensky, U. (2011). Examining the Relationship Between Students' Understanding of the Nature of Models and Conceptual Learning in Biology, Physics, and Chemistry. International Journal of Science Education 33(5), 653-684.
6. Halloun, I. (2007). Mediated Modeling in Science Education. Science and Education, 16, 653-697.
7. Harlen, W., James, M. (1997). Assessment and Learning: differences and relationships between formative and summative assessment. Assessment in Education: Principles, Policy & Practice, 4(3), 365-379.
8. Kansou, K., Nuttle, T., Farnsworth, K., Bredeweg, B. (2013). How plants changed the world: Using qualitative reasoning to explain plant macroevolution's effect on the long-term carbon cycle, Ecological Informatics, 17, 117-142.
9. Leelawong, K., Biswas, G. (2008). Designing Learning by Teaching Agents: The Betty's Brain System. Int. J. of Artificial Intelligence in Education, 18(3), 181-208.
10. Leiba, M., Zuzovsky, R., Mioduser, D., Benayahu, Y., Nachmias, R. (2012). Learning about Ecological Systems by Constructing Qualitative Models with DynaLearn. Interdisciplinary Journal of E-Learning and Learning Objects, 8(1), 165-178.
11. Liem, J.: Supporting Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning. University of Amsterdam (2013). *https://jochemliem.files.wordpress.com/2014/01/liem2013-thesisdigital.pdf*
12. Löhner, S., van Joolingen, W. R., Savelsbergh, E. R., van Hout-Wolters, B. (2005). Students Reasoning during Modeling in an Inquiry Learning Environment. Computers in Human Behavior, 21(3), 441.
13. Louca, L., Zacharia, Z., Michael, M., Constantinou, C. (2011). Objects, Entities, Behaviors, and Interactions: A Typology of Student-Constructed Computer-based Models of Physical Phenomena. Journal of Educational Computing Research, 44(2), 173-201.
14. Louca, L., Zacharia, Z. (2012). Modeling-based Learning in Science Education: Cognitive, Metacognitive, Social, Material and Epistemological Contributions. Educational Review, 64(4), 471-492.
15. Mioduser, D. et. al. (2012). Final report on DynaLearn evaluation studies, DynaLearn, EC FP7 STREP project 231526, D7.4.
16. Moody, D.L. (2005). Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. Data & Knowledge Engineering, 55, 243-276.
17. National Research Council. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press.
18. Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R. (2005). The Difficult Process of Scientific Modelling: An Analysis of Novices Reasoning during Computer-Based Modelling. International Journal of Science Education, 27(14), 1695-1721.
19. Songer, N.B., Ruiz-Primo, M.A. (2012). Assessment and science education: Our essential new priority? Journal of Research in Science Teaching, 49(6), 683-690.
20. Teeuw,W., van den Berg, H. (1997). On the quality of conceptual models. In Proc.16th Int. Cont. on Conceptual Modeling (ER97), Los Angeles, California, USA.
21. IEEE standard for software verification and validation. (2004). TR 1012-2004, IEEE.

How Much Qualitative Reasoning is Required in Elementary School Science Test Questions?

Maxwell Crouse & Kenneth D. Forbus
Qualitative Reasoning Group, Northwestern University
2133 Sheridan Road, Evanston, IL, 60208, USA
MaxwellCrouse2020@u.northwestern.edu; forbus@northwestern.edu

Abstract

Understanding how to build cognitive systems with commonsense is a difficult problem. Since one goal of qualitative reasoning is to explain human mental models of the continuous world, hopefully qualitative representations and reasoning have a role to play. But how much of a role? Standardized tests used in education provide a potentially useful way to measure both how much qualitative knowledge is used in commonsense science, and to assess progress in qualitative representation and reasoning. This paper analyzes a small corpus of science tests from US classrooms and shows that QR techniques are central in answering 13% of them, and play a role in at least an additional 16%. We found that today's QR techniques suffice for standard QR questions, but integrating QR with broader knowledge about the world and automatically understanding the questions as expressed in language and pictures provide new research challenges.

Introduction

When children are learning about science, their initial education is qualitative in nature. It ties scientific concepts to everyday experiences, teaching them how to think about the world around them in terms of more fundamental ideas, including processes (e.g. evaporation, predation) and patterns (e.g. life cycles, food webs). Since these concepts are used in education, there are teaching materials that are accessible to children (and easier for natural language understanding systems to learn from) and standardized tests that measure knowledge in human-normed ways. For example, the New York State Board of Regents makes their exams publically available after they have been given, providing a corpus that supports research. Thus *commonsense science*, as it is sometimes called, provides an excellent frontier for research on qualitative reasoning, since it involves broad-ranging knowledge and multiple kinds of reasoning.

This is not a novel observation. Project Aristo (Clark et al. 2016) identified elementary school science as a productive research area for studying learning by reading and commonsense reasoning. The Science Learning and Teaching working group (which Forbus is part of) adopted such tests as the first phase in a longer research trajectory, with the long-term (2050) goal of AI systems that can help any person learn any area of science, at whatever level they are interested in. This effort is one of multiple efforts that, collectively, are being designed as a replacement for the Turing Test (Forbus, 2016).

That such tests require deeper knowledge can be seen from the recent Allen Institute Science Challenge on Kaggle[1], which used 8th grade science tests. The tag line was "Is your model smarter than an 8th grader?" The answer, for the 738 teams competing, was clearly no. The questions were limited to multiple-choice tests, without diagrams. The rules of the competition were such that no licensed data or software could be used, i.e. no resources from the Linguistic Data Consortium, nothing from Cyc, Watson, or any other system or data that could not be completely open-licensed. Thus the only techniques applied were off-the-shelf machine learning components (including deep learning) and statistical NLP. The best scores achieved on this challenge – which is only a subset of the types of questions on real exams – topped out at 60%[2]. This suggests that deeper knowledge is indeed needed to achieve 8th grade science literacy. Our analysis below argues further that QR is needed as part of that deeper knowledge.

This paper examines how useful qualitative reasoning might be in elementary school science tests. We focus on 4th grade examinations, since that is what Project Aristo has been examining. A prior study of such exams (Clark et al. 2013) provided a useful decomposition of question types,

[1] https://www.kaggle.com/c/the-allen-ai-science-challenge/
[2] Public presentations, Oren Etzioni, Peter Clark, AAAI 2016.

but did not take into account a qualitative reasoning perspective. Hence the questions we ask here are (1) what fraction of exam questions use qualitative representations? (2) How well do today's QR approaches handle the reasoning needed for such questions? After examining the contents of six Regents 4th grade exams, the answers so far are (1) qualitative knowledge is needed for at least 29% percent of exam questions and (2) the standard QR-related questions are naturally handled by existing qualitative reasoning techniques.

An Analysis of Science Tests

Much QR research has focused on specific scientific and engineering domains. By contrast, commonsense science is remarkable for its breadth – such tests cover physics, biology, chemistry, and other areas. Instead of a small vocabulary of structural elements (e.g. circuit components), the entire range of everyday objects is fair game. After all, the purpose of learning science in elementary school and middle school (grades 1-6 and grades 7-8th respectively, in the US) is to ground scientific ideas in a child's experience.

Some questions, such as Figure 1, look exactly like traditional QR scenarios. We call these *standard QR* questions. By viewing the flame as the source, the wire as the destination, and the contact surface with the flame as the path, any reasonable model of heat flow will predict that the temperature of the wire will rise. But translating that insight into heat travelling through the wire involves thinking of the wire itself as a kind of path, which makes the decoding of the language more subtle.

Some problems set up scenarios that are used in multiple questions. Here is an example:

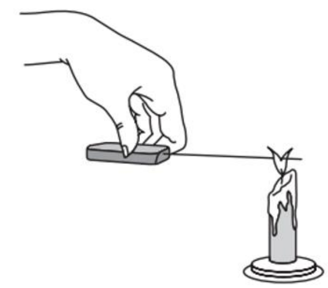
> *One hot, summer day it rained very heavily. After the rain, a plastic pan on a picnic table had 2 cm of rainwater in it. Four hours later, all of the rainwater in the pan was gone.*

One question asked about this scenario was which process caused the disappearance, given condensation, evaporation, precipitation, and erosion as choices. Examining the conditions and influences of these processes enables honing in quickly on the answer. Another question was, if the day was cool instead of hot, would the rainwater have disappeared slower, faster, or in the same amount of time? This is a classic comparative analysis question (Weld, 1986), and again well within the scope of today's QR systems.

Other types of questions require QR, but involve deeper visual reasoning, e.g. comparing which of two inclined planes it would be harder to push a weight up, or choosing among visual configurations as answers to a question posed. Prior research suggests that such problems can be handled via QR, but with additional complexities of visual reasoning, case-based reasoning, or both (e.g. Klenk et al. 2011; Chang et al. 2014). Hence we argue that, to fully capture human capabilities in commonsense science, we should expand our notion of domain theories to include both specific examples and knowledge of patterns of behavior. We call questions that make use of such knowledge *extended QR* questions, because answering them with off-the-shelf purely first-principles QR techniques might be doable, but would be a stretch.

Closely related are questions about patterns found in nature, e.g. food webs, the water cycle, and life cycles of different sorts of living creatures. Such questions are often accompanied by diagrams, showing for example the participants in a food chain or the stages in a life cycle. We refer to these as *pattern questions.* Once a pattern is introduced, some follow-on questions end up being standard QR questions. For example, questions about food webs often require performing comparative analysis on population size changes, taking into account predation. But other pattern questions simply involve placing states in a correct sequence, e.g. the phases of an animal's life cycle.

The drawing below shows a copper wire with a wooden handle being held in a flame.



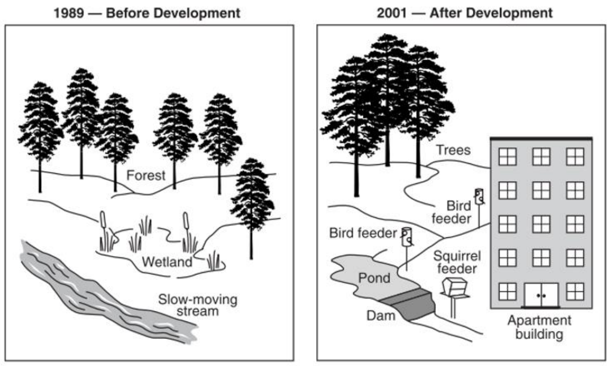
After a few minutes, what will most likely happen?

A. The light will change to electricity.
B. The heat will travel through the wire.
C. The flame will get brighter.
D. The flame will go out.

Figure 1: An example test question

While the picture in Figure 1 may help a child understand the problem better, the caption provides, in some sense, all that is needed to solve the problem. But in some problems a deeper understanding of diagrams is necessary to answer the question. Questions often involve decoding information from graphs, tables, and/or drawings of measurement instruments. For example, each exam typically has at least one question about graphs, which requires reading the graph and answering qualitative or quantitative questions about it (e.g. given a population graph, "How many times was there a decrease in the deer population from one year to the next […]?") Problems with pictures often involve recognition, e.g. the different animals in a food web, the different stages in a life cycle, weather icons on a map. Sometimes these pictures have labels, when recognition would be too demanding, as in Figure 2. We call such questions *visual questions*. This problem is especially interesting because it requires integrating the scenario across two modalities, language and vision, and generating an answer, rather than selecting from multiple-choice answers.

A company bought land in 1989 to build apartments. The diagram labeled 1989 shows the land before the company built the apartments. The diagram labeled 2001 shows the same land after the company built the apartments.



Describe one positive way and one negative way that the organisms living in the area have been affected by the changes shown in the diagrams.

Figure 2: A multimodal scenario problem

Any question that does not fit in one of the above categories we classify as a *world knowledge* question. This is a grab-bag category, involving many different kinds of knowledge. For example, some kinds of questions involve properties of objects, e.g. which object from a list (wax crayon, plastic spoon, rubber eraser, iron nail) is the best conductor of electricity? These involve QR, in that conductivity can be thought of as a parameter – while binary in this case, a harder question would involve iron, tap water, and salt water. But many others involve knowledge about non-continuous aspects of the world. For example, "which characteristic can a human offspring inherit?", where the answers include facial scars, long hair, broken leg, and blue eyes. Another sub-category of questions concern function, e.g. "The functions of a plant's roots are to support the plant and", with "make food", "produce fruit", "take in water and nutrients", and "aid in germination" as alternatives.

While more fine-grained analyses of commonsense science questions are possible, this set of categories suffices to address the first of our two questions. To identify the degree to which QR is needed in commonsense science, we analyzed a corpus of six 4th grade science exams[3]. The results are shown in Table 1.

This analysis suggests that QR knowledge about continuous causality is a necessary part of doing well on the exam: the highest score a student could get would be 71% otherwise. On the other hand, QR is not sufficient to do well on the exam, as indicated by 71% of the questions not involving QR.

Type	# Problems	%
Standard QR	31	13%
Extended QR	38	16%
Patterns	36	15%
Visual	55	22%
World	85	35%

Table 1: Analysis of question types on science exams

Solving QR-based Problems

Now let us turn to the second question: Can current QR techniques solve the QR problems that arise in such science tests? To examine this question, we selected the set of 31 standard QR questions from the corpus of New York Regents exams. To factor out issues in natural language understanding, we hand-coded queries corresponding to each question. We used knowledge base contents from ResearchCyc[4], with our own extensions for qualitative, visual, and analogical reasoning and learning.

While our KB already had a substantial portion of the knowledge needed, some extensions were required. We used qualitative process theory (Forbus, 1984) to express the new domain knowledge. Specifically, we encoded 8 additional physical processes (precipitation, evaporation, erosion, freezing, melting, birth, death, growth) and 5 other model fragments (buoyancy, organism populations, standard gravity, predator/prey, friction, and magnetism), along with 6 new types of quantities (fluid level in a container, fluid displaced, heat produced, friction force applied against an object, magnetic force attracting an object, and roughness) and one ordinal relationship (smooth objects are less rough than rough objects). The rest of the QP domain theory came from previously existing knowledge. It consisted of 2 types of processes (boiling and heat flow) and 7 types of quantities (population size, mass, weight, volume, temperature, size, amount of a substance, and distance). Extending the domain theory required approximately two months of work.

[3] Specifically, the New York Regents science exams for 2004, 2005, 2006, 2009, 2010, and 2011. These and exams for other years and grade levels are available on their web site.

[4] http://www.cyc.com/platform/researchcyc/

```
(isa FreezingProcess QPProcessType)
(mfTypeParticipant FreezingProcess ?thing-freezing LiquidTangibleThing
                   focusOf)
(mfTypeParticipant FreezingProcess
  ?sub ChemicalCompoundTypeByChemicalSpecies substanceOf)
(mfTypeParticipantConstraint FreezingProcess
                              (substanceOfType ?thing-freezing ?sub))
(mfTypeParticipantConstraint FreezingProcess
                 (relationAllInstance freezingPoint ?sub ?f-temp))
(mfTypeCondition FreezingProcess
                (qLessThan (TemperatureFn ?thing-freezing) ?f-temp))
(mfTypeBiconditionalConsequence FreezingProcess
  (hasQuantity ?self (SolidGenerationRateFn ?self)))
(mfTypeConsequence FreezingProcess
  (qGreaterThan (SolidGenerationRateFn ?self) 0))
(mfTypeConsequence FreezingProcess
  (qprop- (SolidGenerationRateFn ?self)
          ((QPQuantityFn Temperature) ?thing-freezing)))
(mfTypeConsequence FreezingProcess
  (i+ (AmountOfFn ?sub Solid-StateOfMatter ?thing-freezing)
      (SolidGenerationRateFn ?self)))
(mfTypeConsequence FreezingProcess
  (i- (AmountOfFn ?sub Liquid-StateOfMatter ?thing-freezing)
      (SolidGenerationRateFn ?self)))
```

Figure 3: Representation of the process of freezing

Figure 3 shows the description of the axioms for the process of freezing (`FreezingProcess`) as an example. That it is a type of process specified by QP theory is indicated by the isa statement placing it as a member of the collection `QPProcessType`, whose instances are members of QPProcess, e.g. a particular instance of freezing. Type-level predicates are used to define model fragment types. The participants are specified by `mfTypeParticipant`, e.g. here `FreezingProcess` has three participants, whose types are the third argument (e.g. `LiquidTangibleThing`, a pre-existing concept in Cyc), whose template variables are the second argument (e.g.`?thing-freezing`), and whose fourth argument is a role relation that is used to refer to this participant in axioms about instances (e.g. `solidOf`). `mfTypeCondition` expresses the conditions that must hold for an instance to be active. These are interpreted as conjunctions, although this process has only one, i.e. that the temperature of the thing freezing is less than its freezing point. The consequences are expressed via `mfTypeConsequence` and `mfTypeBiconditionalConsequence`, the latter for statements that can only be true when an instance is active. An example of such a constitutive relationship is the existence of a rate at which the process occurs, which does not make sense outside the process acting. The usual causal qualitative mathematics of QP theory appear in the consequences, e.g. `i+` and `i-` for direct influences (i.e. partial specifications of derivatives) and `qprop` and `qprop-`, for indirect influences (i.e. partial specifications of functional dependencies). Wherever possible, we link these descriptions into the Cyc ontology, e.g. `LiquidTangibleThing` comes from the Cyc ontology, so that axioms about them already in the knowledge base can provide leverage. Sometimes the Cyc ontology takes a slightly different perspective on the world. For example, the Cyc concept of Temperature concerns specific values for temperatures, e.g. `Hot` or `(DegreeCelsius 25)`. In QP theory, quantities are fluents, in that they are not values but conceptual entities whose value changes over time. We link the two notions via the logical function `QPQuantityFn`, a second-order function whose domain is Cyc quantities and whose range are functions denoting fluents, here `((QPQuantityFn Temperature) ?thing-freezing)` denotes the fluent representing the temperature of `?thing-freezing`.

Figure 4 provides an example of a model fragment, a description of an object floating in a fluid (`ObjectFloatingInFluid`). It is an instance of `ConceptualModelFragmentType`, that is, instances of this type of model fragment are conceptual knowledge about the situation. (Some types of model fragments indicate the existence of something, such as a contained fluid or population, those are instances of `PhysicalModelFragmentType`.) Note the multiple condition statements, which are interpreted conjunctively. `activeMF` is true when the model fragment instance which is its argument is active.

The queries to solve these problems were relatively straightforward applications of qualitative reasoning. For example, some problems describe a situation and ask what kind of process is involved in the change that is occurring in it. Performing model formulation on the situation and in-

```
(isa ObjectFloatingInFluid ConceptualModelFragmentType)
(mfTypeParticipant ObjectFloatingInFluid ?csolid SolidTangibleThing
                   solidOf)
(mfTypeParticipant ObjectFloatingInFluid ?cfluid FluidTangibleThing
                   fluidOf)
(mfTypeParticipant ObjectFloatingInFluid ?b-mf FluidDisplacement
                   displacementOf)
(mfTypeParticipantConstraint ObjectFloatingInFluid
                             (fluidOf ?b-mf ?cfluid))
(mfTypeParticipantConstraint ObjectFloatingInFluid
                             (contains-Underspecified ?cfluid ?csolid))
(mfTypeCondition ObjectFloatingInFluid (activeMF ?b-mf))
(mfTypeCondition ObjectFloatingInFluid
  (qLessThanOrEqualTo
    ((QPQuantityFn Weight) ?csolid)
    ((QPQuantityFn Weight) (FluidDisplacedFn ?b-mf))))
(mfTypeConsequence ObjectFloatingInFluid
  (qprop (FluidDisplacedFn ?b-mf) ((QPQuantityFn Weight) ?csolid)))
```

Figure 4: Representation of the model fragment describing a floating object

specting the instantiated processes provides a straightforward way to answer such questions. Sometimes chaining is needed, that is, searching through dependencies among model fragments. For example, the question "Which form of energy is needed to change water from a liquid to a gas?" with answers being "heat", "mechanical", "chemical", and "sound" requires a breadth-first search through model fragments, beginning with model fragments whose consequences involve direct influences on amounts of substances of different phases, a negative influence on the liquid version and a positive influence on that substance in the gas phase, and expanding on model fragments that are mentioned as conditions, in this case, heat flow. When questions involve comparisons, differential qualitative analysis (Weld, 1986, 1990) is used to determine the changes to quantities of interest that have occurred.

To provide a sense of how these problems are solved, let us return to the scenario presented earlier:

One hot, summer day it rained very heavily. After the rain, a plastic pan on a picnic table had 2 cm of rainwater in it. Four hours later, all of the rainwater in the pan was gone.

(Q7) Which process caused the rainwater in the pan to disappear as it sat outside in the hot air?

(Q8) If the day were cool instead of hot, the rainwater in the pan would have disappeared ______

Q7 is answered by constructing a qualitative model for the state of the scenario in which water was sitting in a pan and examining the influences on it (see Figure 5), to see which process is responsible for decreasing the amount of water. As Figure 5 illustrates, the model fragments are tied to the Cyc ontology, including the use of Cyc's `ScalarInterval` system for underspecified values (e.g. `Hot`), but which have ordinal relationships tied to other underspecified values in the same dimension (e.g. `Cool`). For Q8, an additional qualitative state is created to represent the cooler day, with everything the same except for that the temperatures of the air and rainwater are `Cool` instead of `Hot`. We use analogy to perform comparative analysis: the analogical mapping[5] provides information about how the two states correspond, in both their values and their causal structure. Figure 6 illustrates the correspondences computed between these two states. The system checks first to see if there is enough

```
(isa LiquidTangibleThing)
(substanceOfType rainwater014 Water)
(isa air435 GaseousTangibleThing)
;;From "the rainwater ... sat outside in the hot air"
(touches-Directly rainwater014 air435)
;; QPQuantityFn converts Cyc's value notion to QP's fluents
(qEqualTo ((QPQuantityFn Temperature)
air435) Hot)
(qEqualTo ((QPQuantityFn Temperature)
rainwater014) Hot)
```

Figure 5: Partial representation of Q7 evaporation scenario

information about the goal quantity to directly determine if it is different. (For example, if in a different question the system were asked about the temperature of the desert during the day (`Hot`) and during the night (`Cool`), the ordinal difference between these values would be sufficient to answer the question.) Otherwise, it looks for causal structure that specifies the goal quantity in terms of others, and recursively seeks their comparative values. Here, the aligned causal influences (`qprop` relations) linking the evaporation rate and temperature of each scenario enable the system to

[5] Mappings are computed using SME, the Structure-Mapping Engine (Falkenhainer et al. 1989; Forbus et al. in press).

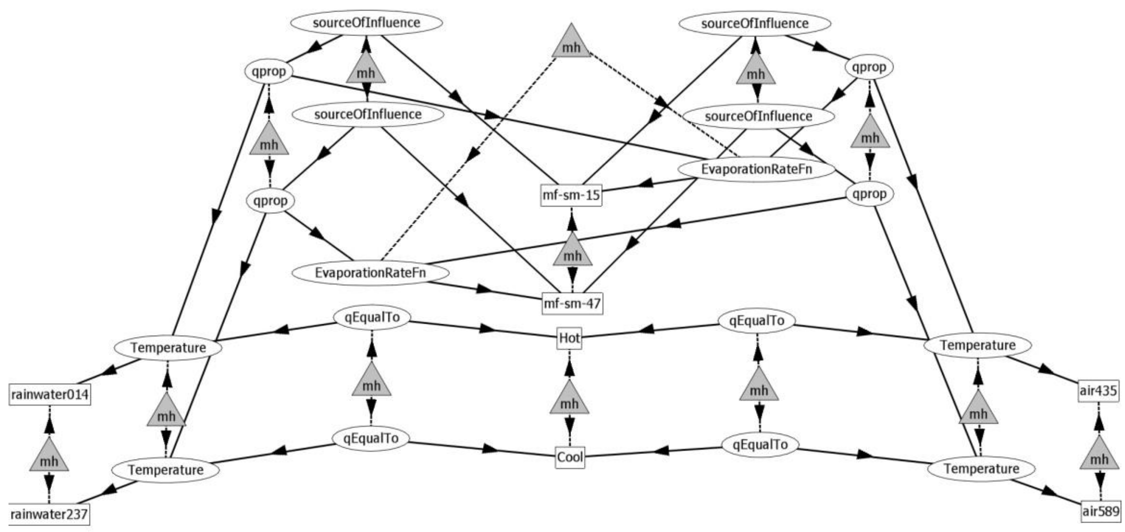


Figure 6: Partial depiction of the analogical mapping between two evaporation scenarios for differential qualitative analysis. Correspondences are indicated match hypothesis ("mh") triangles.

infer that, since the temperature is reduced, and the rate of evaporation depends on temperature, then the rate of water disappearance would be slower in the new scenario.

Using existing qualitative reasoning techniques, the system was able to solve all 31 standard QR problems. Less success was achieved on the extended QR problems. Of the few we tackled, none were solvable with strictly first-principles QR techniques. They require more knowledge of the everyday world. Consider again the problem shown in Figure 2. This problem requires inferring that there are fewer trees after construction than before construction. This is indicated schematically by there being fewer trees on the right, but also by the associated labels, e.g. "forest" versus "trees". Students must know that trees provide habitats for birds and squirrels, which are part of what helps determine the size of their populations, and hence that fewer trees means less habitat and hence a negative effect on population. On the other hand, adding feeders improves their food supply. (Whether or not this benefit outweighs the loss of food supply from habitat loss seems dubious, but nevertheless it is a positive influence, even if dominated by another factor.) Other examples involve richer interactions between dynamics and spatial knowledge (e.g. knowing that liquids take the shape of their container). Again, some forays into representing these ideas have been done before in QR, e.g. Kim's bounded stuff ontology (Kim 1993), but domain theories which tightly integrate qualitative dynamics and spatial representations are few and limited in coverage currently. Accumulating examples to reason from (e.g. Klenk et al 2011), plus more flexible multimodal interaction (e.g. Chang & Forbus, 2015) should be helpful for teaching systems the knowledge that they need to tackle problems like these.

Related Work

The most successful system thus far in answering elementary science exam questions is AI2's Aristo (Clark et al. 2016) which combines techniques from information retrieval, statistical NLP, and rule-based systems. The success of Aristo relied on both the ensemble of techniques and its ability to estimate which technique's answer is most likely to be correct. With its diverse set of techniques, Aristo achieved a score of 71.3% on a corpus of 129 Regents 4th grade non-diagram multiple-choice-only questions. In the analysis of its performance, five types of questions were identified as being challenging for Aristo to solve. The question types were comparison questions, simple arithmetic reasoning, complex inference, structured questions, and story questions. In our application of QR techniques to Regents exam questions, we found that a large proportion of the 31 questions solvable by our techniques were comparison and story questions, indicating that the addition of QR to Aristo may boost its performance.

We further note that most attempts to solve problems such as these focus on information retrieval techniques over text (e.g. Sachan et al. 2016), or lightweight knowledge representation schemes where the tokens in semi-structured representations are still words or phrases (e.g. Khashabi et al. 2016). By contrast, we are using deductive reasoning over conceptual representations. While we agree that there are roles for maintaining linguistic information in extracting knowledge from text, we also believe that the refinement of such knowledge into conceptual knowledge is a crucial, but underexplored, component of learning by reading. Efforts to date at such refinements include Semantic Construction Grammar (Schneider & Witbrock, 2015) and Companion-based learning by reading (Barbella & Forbus, 2015).

One of the foundational works for qualitative reasoning was Hayes' (1979) naïve physics manifesto, which encouraged the field to look at commonsense physical reasoning. Some research has focused on broad, axiomatic accounts of phenomena, e.g. liquids (Hayes 1985), matter (Davis, 2010), and containers (Davis et al. 2013), but none of these efforts were tied into a large, overarching ontology. We believe that integrating such efforts into the Cyc ontology (which can be used freely, by staying with OpenCyc) would radically improve the ability to create the kind of larger-scale, integrated accounts needed to broadly cover commonsense science. In 4^{th} grade science, qualitative simulation seems unnecessary, but that is unlikely to be true at higher grades, at which point qualitative simulators like Garp3 (Bredeweg et al. 2009) may prove valuable.

Discussion and Future Work

We agree with AI2 that commonsense science is a useful approach to studying the nature of commonsense reasoning more broadly. We are encouraged that over a quarter of the exam questions involve qualitative representations and reasoning, and that standard QR techniques perform well on this portion of 4^{th} grade exams. Prior research by Bruce Sherin[6] indicates that the content of middle-school science remains focused on qualitative knowledge, to provide a firm foundation for integrating with algebra and calculus later on. An analysis of 8^{th} grade exams, in progress, looks likely to provide additional evidence for the centrality of qualitative representations and reasoning for commonsense science.

We note that, like in prior projects, the broad contents of the ResearchCyc knowledge base provide significant leverage for this kind of research. Being able to draw on a wide-ranging ontology is useful to reduce tailorability, but more importantly, it provides leverage on its own (e.g. `ScalarInterval` as a simple form of qualitative value well suited for capturing the ambiguities inherent in natural language). Even when there are design choices that are not optimal from a particular perspective (e.g. formalizing some quantities as values instead of fluents), simple coercions typically suffice to put the knowledge in a more useful form.

Much future work remains, of course. First, we plan to extend the Companion natural language facilities to automatically interpret exam questions to generate the kinds of queries that here were created by hand. Second, we plan to extend our learning by reading work (e.g. Lockwood & Forbus, 2009; Barbella & Forbus, 2015) to provide the broad-scale knowledge needed to handle these kinds of questions. Third, we plan to use a combination of computer vision techniques and sketch understanding (Forbus et al. 2011) to automatically process the visual aspects of questions. Finally, we plan on exploring interactive training of Companions on commonsense science, by posing scenarios and asking questions, including follow-up questions aimed at exposing misconceptions gleaned from learning by reading.

Acknowledgements

This research was supported by the Intelligent and Autonomous Systems Program of the Office of Naval Research.

References

Barbella, D. and Forbus, K. (2015). Exploiting Connectivity for Case Construction in Learning by Reading. *Proceedings of the Third Annual Conference on Advances in Cognitive Systems*.

Bredeweg, B., Linnebank, F., Bouwer, A., and Liem, J., 2009. Garp3 - Workbench for Qualitative Modelling and Simulation. *Ecological Informatics* 4(5-6), 263-281.

Clark, P. (2015) Elementary School Science and Math Tests as a Driver for AI: Take the Aristo Challenge! *Proceedings of IAAI-2015*, pp. 4019-4021.

Clark, P., Harrison, P., & Balasubramanian, N., (2013). A study of the knowledge base requirements for passing an elementary school science test. *Proceedings of the 2013 Workshop on Automated Knowledge Base Construction*, pp 37-42.

Clark, P., Etzioni, O., Khashabi, D., Khot, T., Sabharwal, A., Tafjord, O., & Turney, P. (2016) Combining Retrieval, Statistics, and Inference to Answer Elementary Science Questions. *Proceedings of AAAI 2016.*

Davis, E. (2010) Ontologies and Representations of Matter, *Proceedings of AAAI-2010.*

Davis, E., Marcus, G., & Chen, A. (2013) Reasoning from Radically Incomplete Information: The Case of Containers. *Proceedings of the 2nd Annual Conference on Advances in Cognitive Systems.* Baltimore, MD.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, 41(1), 1-63.

Forbus, K. (1984) Qualitative Process Theory. *Artificial Intelligence*, 24, pp. 85-168.

Forbus, K. (2016) Software Social Organisms: Implications for measuring AI Progress. *AI Magazine*, 37(1), pp 85-90.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: Sketch understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science*. 3(4), pp 648-666.

[6] Personal communication

Forbus, K., Ferguson, R., Lovett, A., & Gentner, D. (in press). Extending SME to handle large-scale cognitive modeling. *Cognitive Science*.

Hayes, P. (1979). The naive physics manifesto. In D. Michie (Ed.), *Expert Systems in the Micro-Electronic Age*. Edinburgh University Press.

Hayes, P. (1985). Naive Physics I: Ontology for liquids. In Hobbs, R., & Moore, R. (Eds.), *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex Publishing Corporation.

Khashabi, D., Khot, T., Sabharwal, A., Clark, P., Etzioni, O., & Roth, D. (2016) Question Answering via Integer Programming over Semi-Structured Knowledge. *Proceedings of IJCAI 2016.*

Kim, H. (1993). Qualitative Reasoning about Fluids and Mechanics. Ph.D. Dissertation and ILS Technical Report, Northwestern University.

Klenk, M., Forbus, K., Tomai, E., & Kim, H. (2011) Using analogical model formulation with sketches to solve Bennett Mechanical Comprehension Test problems. *Journal of Experimental and Theoretical Artificial Intelligence*, 23(3): 299-327.

Lockwood, K. and Forbus, K. (2009). Multimodal knowledge capture from text and diagrams. *Proceedings of KCAP-2009*.

Sachan, M, Dubey, A., & Xing, E. (2016) Science Question Answering using Instructional Materials. Arxiv:1602.04375v5, April 5th.

Schneider, D. & Witbrock, M. (2015) Semantic Construction Grammar: Bridging the NL/Logic Divide. *Proceedings of WWW-2015.*

Weld, D. (1986) Comparative Analysis. *Artificial Intelligence*, 36, pp. 333-373.

Weld, D. (1990) *Theories of Comparative Analysis*. MIT Press. Cambridge, MA.

Towards A Qualitative Descriptor for Paper Folding Reasoning*

Zoe Falomir
Bremen Spatial Cognition Centre
Universität Bremen, Germany
zfalomir@informatik.uni-bremen.de

Abstract

Paper folding tests are used to measure spatial abilities in humans. Artificial agents with strong intelligence must have reasoning mechanisms to solve spatial problems cognitively. In this paper, a qualitative descriptor for paper folding (QD-PF) is presented. QD-PF defines the folding actions which appear in the problem and the areas in the paper where a hole can be punched. Reasoning tables for inferring location equivalences after paper folding are created. An approach towards solving paper folding questions logically when a hole is punched after one-to-three foldings of the paper is also provided.

1 Introduction

Qualitative Spatial and Temporal Representations and Reasoning (QSTR) [Cohn and Renz, 2007; Ligozat, 2011] models and reasons about *time* (i.e. coincidence, order, concurrency, overlap, granularity) and also about properties of *space* (i.e. topology, location, direction, proximity, geometry, intersection, etc.) and their evolution in time between continuous neighboring situations. Maintaining the consistency and constraints in space and time are the basics in qualitative reasoning when solving spatial problems (i.e. path finding, orientation, relative position, etc.) and temporal problems (i.e. constraint satisfaction, schedule optimisation, precedence) [Guesgen and Bhatt, 2009]. As a result, well-defined qualitative models and reasoning techniques have appeared in the literature which can deal with imprecise and incomplete knowledge on a symbolic level. Spatio-temporal reasoning has been proved to be successful in many areas and applications such as: robotics [Kunze *et al.*, 2014b; Falomir *et al.*, 2013], computer vision [Kunze *et al.*, 2014a; Falomir *et al.*, 2011; Cohn *et al.*, 2006], ambient intelligence [Bhatt and Dylla, 2009; Bhatt *et al.*, 2011, 2013; Falomir and Olteţeanu, 2015], architecture and design [Bhatt and Freksa, 2015], spatial query solving in geographic information systems [Fogliaroni, 2013; Al-Salman, 2014], classification of volunteered geographic information [Ali *et al.*, 2016], etc.

Furthermore, qualitative representations are thought to be closer to the cognitive domain, as shown in cognitive models of sketch recognition [Lovett *et al.*, 2006], spatial problem solving tasks (i.e. visual oddity tasks) [Lovett and Forbus, 2011] and in mental rotation tasks [Lovett and Schultheis, 2014]. Therefore, novel models which combine QSTR, cognitive spatial thinking and common sense are a challenge which envision further advances in Artificial Intelligence and its applications.

Moreover, spatial cognition studies have shown that there is a strong link between spatial abilities and success in Science, Technology, Engineering and Math (STEM) disciplines [Newcombe, 2010; Wai *et al.*, 2009]. For example, children at 4 years old have already informal awareness of spatial relations such as parallel relations for two dimensional shape identification and description before they are properly taught about them [Sinclair *et al.*, 2013]. For this reason, researchers in US and Canada study the actualities and possibilities of spatial reasoning in contemporary school mathematics [Sinclair and Bruce, 2014], also because spatial learning and reasoning can be taught easily using visual and kinetic interactions offered by new digital technologies [Highfield and Mulligan, 2007].

In cognitive psychology, games like *Upside Down World* are used to evaluate spatial skills of students who are challenged to recreate buildings composed of multilink cubes in their upright orientation and use spatial language to describe the composition of the buildings to their colleagues to build accordingly [Sinclair and Bruce, 2014]. Also, a test of the German Academic Foundation to find out children with gifted brains consists in finding out the consistent view/projection for a 3D object usually corresponding to a technological drawing[1]. In previous works by the same author, a qualitative model for 3D object description was developed and promising results are obtained [Falomir, 2015].

Regarding paper folding, in the literature, there are research works which dealt with the problem of modeling origami computationally [Ida *et al.*, 2015]. However, there is no previous works, as far as we are concerned, that modeled paper folding and punching actions qualitatively.

This paper presents a Qualitative Descriptor for Paper Folding (QD-PF) which is motivated by the fact that paper folding tasks (see Figure 1 as an example) have been extensively used in psychological cognitive tests to measure people spatial abilities as a form of intelligence [Ekstrom *et al.*, 1976]. Note that paper folding involves a coordinated sequence of spatial transformations.

The Dental Admission Test (DAT) by the American Dental Association[2] (ADA) includes a Perceptual Ability Test (PAT) which asks questions about paper folding. An example of a question in this test is that provided in Figure 1. The instructions provided are the following: *A flat square is folded one or more times. The broken lines indicate the original position of the paper. The solid lines indicate the position of the folded paper.*

*Correspondence to: Dr.-Ing. Zoe Falomir, Bremen Spatial Cognition Centre (BSCC), FB3 - Informatik und Mathematik, Universität Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany. E-mail: zfalomir@informatik.uni-bremen.de

[1] *Test der Studienstiftung: Gehirnjogging für Hochbegabte*, see *Spiegel* Online: `http://www.spiegel.de/quiztool/quiztool-49771.html`

[2] American Dental Association: `http://www.ada.org`

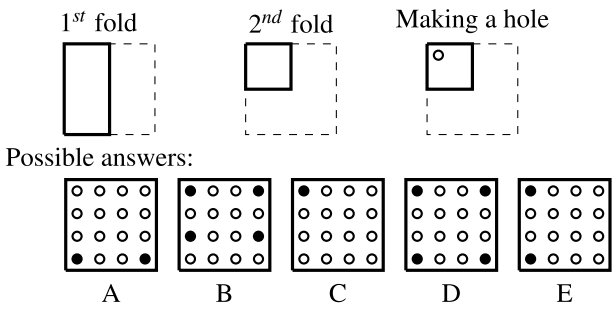


Figure 1: Example of a Paper Folding question in the Perceptual Ability Test part in the DAT.

The paper is never turned or twisted. The folded paper always remains within the edges of the original square. There may be from one to three folds in each item. After the last fold a hole is punched in the paper. Your task is to mentally unfold the paper and determine the position of the holes in the original square. Choose the pattern of black circles that indicates the position of the holes on the original square.

This paper explores the challenge of defining a model which can:

- help people to understand how to solve paper folding tests, so that they can improve their spatial skills and therefore they can enhance their success in STEM; and also,
- be used by artificial intelligent agents to solve paper folding problems so that they can learn the spatial transformations happening when folding and then developing a new framework in spatial reasoning.

The rest of the paper is organized as follows. Section 2 presents the Qualitative Descriptor for Paper Folding (QD-PF). Section 3 shows how paper folding actions bring paper areas into correspondence and present the reasoning tables obtained. Section 4 describes how several consequent paper folding actions are related to hole punching. Section 5 presents our approach for solving paper folding questions. Section 6 discusses the results and then future work is presented.

2 A Qualitative Descriptor for Paper Folding

This section presents a model for describing a spatial reasoning task named as *Paper Folding* and the logical reasoning process to be followed by intelligent agents to describe qualitatively the actions and results involved in this task.

The Qualitative Descriptor for Paper Folding (QD-PF) describes qualitative areas on a paper, the possible locations of holes in these areas and the actions that can be done on that paper:

QD-PF= {PaperAreas$_{RS}$, HoleLocation$_{RS}$, FoldingActions$_{RS}$}

The QD-PF model is defined by three reference systems (RS): PaperAreas$_{RS}$, HoleLocation$_{RS}$ and FoldingActions$_{RS}$, which are described next.

PaperAreas$_{RS}$= {w, h, PaperAreas$_{LB}$, PaperAreas$_{INT}$}

where w and h are the width and height dimensions of the paper; PaperAreas$_{LB}$ presents the set of names/labels for the possible areas in which a paper can be divided into (as illustrated in Figure 2); and PaperAreas$_{INT}$ defines the corresponding dimensions of these areas using the Cartesian product of intervals (see Table 1).

PaperAreas$_{LB}$= {left(l), right(r), up(u), down(d), middle(m), centre(c), corner(o)}

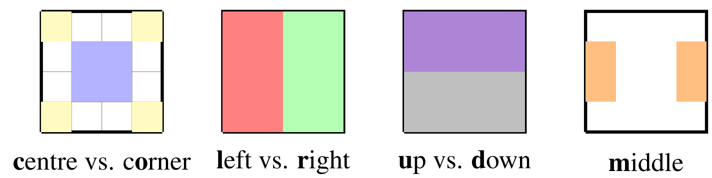


Figure 2: Illustration of paper areas: left, right, up, down, centre, middle and corners.

PaperAreas$_{LB}$	**PaperAreas$_{INT}$**
left	$[0,w/2] \times [0,h]$
right	$[w/2,w] \times [0,h]$
up	$[0,w] \times [h/2,h]$
down	$[0,w] \times [0,h/2]$
centre	$[w/4,3/4w] \times [h/4,3/4h]$
corner	$[0,w/4] \times [0,h/4]$ and $[0,w/4] \times [3/4h,h]$ and $[3/4w,w] \times [0,h/4]$ and $[3/4w,w] \times [3/4h,h]$
middle	$[0,w] \times [h/4,3/4h]$ and $[3/4w,w] \times [h/4,3/4h]$

Table 1: Defining PaperAreas$_{INT}$ using the Cartesian product of intervals based on w and h parameters.

The reference system for the location of holes (HoleLocation$_{RS}$) is defined next as:

HoleLocation$_{RS}$= {w, h, HoleLocation$_{LB}$, HoleLocation$_{INT}$}

where w and h are the width and height dimensions of the paper; HoleLocation$_{LB}$ refers to the names/labels given to the possible location of punched holes, as illustrated in Figure 3 (see Table 2 for a correspondence of those locations with the intersection of paper areas); and HoleLocation$_{INT}$ defines the possible locations of punched holes using Cartesian product of intervals (see Table 2).

HoleLocation$_{LB}$= {luo, lu, ru, ruo, lum, luc, ruc, rum, ldm, ldc, rdc, rdm, ldo, ld, rd, rdo / l,r,u,d,m,c,o ∈ PaperAreas$_{LB}$ }

luo	lu	ru	ruo
lum	luc	ruc	rum
ldm	ldc	rdc	rdm
ldo	ld	rd	rdo

Figure 3: Approximate location of holes in the paper.

Hole Location$_{LB}$	Paper Areas$_{LB}$	Hole Location$_{INT}$
luo	left up corner	$[0,w/4]\times[3/4h,h]$
lu	left up	$[w/4,w/2]\times[3/4h,h]$
ru	right up	$[w/2,3/4w]\times[3/4h,h]$
ruo	right up corner	$[3/4w,w]\times[3/4h,h]$
lum	left up middle	$[0,w/4]\times[h/2,3/4h]$
luc	left up centre	$[w/4,w/2]\times[h/2,3/4h]$
ruc	right up centre	$[w/2,3/4w]\times[h/2,3/4h]$
rum	right up middle	$[3/4w,w]\times[h/2,3/4h]$
ldm	left down middle	$[0,w/4]\times[h/4,h/2]$
ldc	left down centre	$[w/4,w/2]\times[h/4,h/2]$
rdc	right down centre	$[w/2,3/4w]\times[h/4,h/2]$
rdm	right down middle	$[3/4w,w]\times[h/4,h/2]$
ldo	left down corner	$[0,w/4]\times[0,h/4]$
ld	left down	$[w/4,w/2]\times[0,h/4]$
rd	right down	$[w/2,3/4w]\times[0,h/4]$
rdo	right down corner	$[3/4w,w]\times[0,h/4]$

Table 2: Notation of hole locations.

The reference system for the folding actions to carry out on the paper (FoldingActions$_{RS}$) is defined next as:

FoldingActions$_{RS}$=$\{w$,h,FoldingActions$_{LB}$, FoldingActions$_{PP}\}$

where FoldingActions$_{LB}$ correspond to the names/labels given to the actions showed in the Perceptual Ability Test (PAT) and FoldingActions$_{PP}$ correspond to the points which define the lines appearing in the folded paper when those actions are carried out. Figure 4 shows the possible folding actions (see further illustrations in Section 3) and Table 3 provides the corresponding points for their mathematical definition.

FoldingActions$_{LB}$= {btw-up-middle, in-middle, btw-down-middle, in-left, in-centre, in-right, left-up-corner, right-up-corner, left-down-corner, right-down-corner, diagonal-left, diagonal-right}

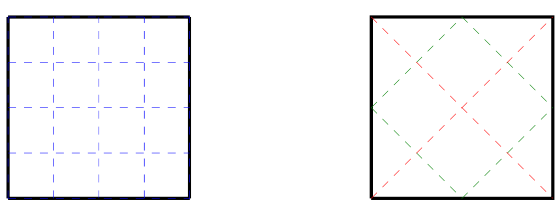


Figure 4: Possible folding actions: (i) first paper, in horizontal: *btw-up-middle, in-middle, btw-down-middle*; and in vertical: *in-left, in-centre, in-right*; (ii) second paper, in green: *left-up-corner, right-up-corner, left-down-corner, right-down-corner*; and in red: *diagonal-upward, diagonal-downward*.

Each folding action is intuitively defined by 2 points which are connected by a line when the paper is folded. Computationally, this folding line can be defined by the mathematical equation defined by two points, that is, given 2 points $p_1=(a_1,b_1)$ and $p2=(a_2,b_2)$, the folding line defined by them is $f(x)=b_1+(b_2-b_1)/(a_2-a_1)(x-a_1)$. Table 3 provides the points that define the folding actions (FoldingActions$_{PP}$) in our reference system.

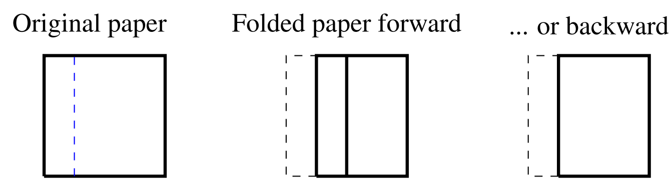
FoldingAction$_{LB}$	FoldingAction$_{PP}$
btw-up-middle	$(0,3/4h),(w,3/4h)$
in-middle	$(0,h/2),(w,h/2)$
btw-down-middle	$(0,h/4),(w,h/4)$
in-left	$(w/4,0),(w/4,h)$
in-centre	$(w/2,0),(w/2,h)$
in-right	$(3/4w,0),(3/4w,h)$
left-up-corner	$(0,h/2),(w/2,h)$
right-up-corner	$(w,h/2),(w/2,h)$
left-down-corner	$(0,h/2),(w/2,0)$
right-down-corner	$(w/2,0),(w,h/2)$
diagonal-upward	$(0,0),(w,h)$
diagonal-downward	$(0,h),(w,0)$

Table 3: Definition of folding actions depending on the lines they created on the paper by connecting 2 points.

3 Reasoning about Folding Actions and Hole Locations

The QD-PF model can relate folding actions with hole locations using the following inference tables.

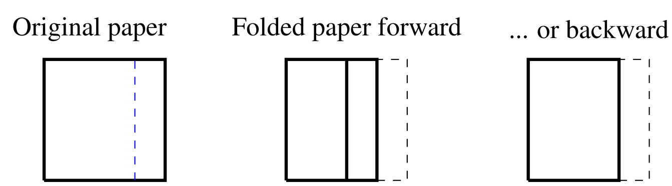
The folding action/operation *in-left* represents a paper folded in the *left* side (according to Figure 2). The inference table in Figure 5 shows this action and how the location left-up-corner (luo) is overlapping left-up (lu). Then a hole made in lu after folding *in-left* is also transmitted to luo. Similarly, left-up-middle (lum) location is equivalent to left-up-centre (luc), left-down-middle (ldm) to left-down-centre (ldc) and left-down-corner (ldo) to left-down (ld); and viceversa. Note that this relations do not depend on how the paper is folded: forward or backward.



Folding	lu	luc	ldc	ld
in left	luo	lum	ldm	ldo

Figure 5: Folding Action: in-left.

The folding action/operation *in-right* represents a paper folded in the *right* side (according to Figure 2). The inference table in Figure 6 shows that, when a paper is folded in the right side, the hole location right-up-corner (ruo) is equivalent to right-up (ru) and viceversa. Similarly, right-up-middle (rum) location is equivalent to right-up-centre (ruc), right-down-middle



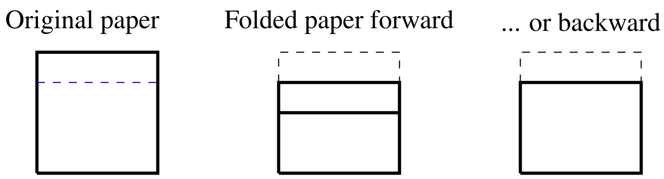
Folding	ru	ruc	rdc	rd
in right	ruo	rum	rdm	rdo

Figure 6: Folding Action: in-right.

(rdm) to right-down-centre (rdc) and right-down-corner (rdo) to right-down (rd); and viceversa. Note that this relations do not depend on how the paper is folded: forward or backward.

It is also important to note that the folding actions *in-right* and *in-left* are symmetrical operations, since the inferences are the same and the only parameter changed is both inference tables is right (r) for left (l).

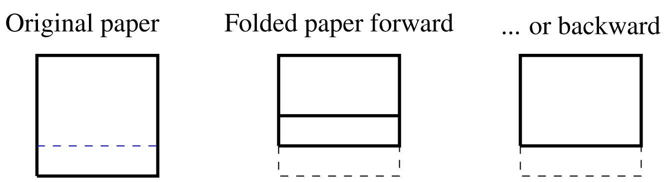
The folding action/operation *between up-middle* represents a paper folded between *up* and *middle* areas (according to Figure 2). Figure 7 shows that in this situation, the hole location left-up-corner (luo) is equivalent to left-up-middle (lum), and viceversa. Similarly, left-up (lu) location is equivalent to left-up-centre (luc), right-up (ru) to right-up-centre (ruc) and right-up-corner (ruo) to right-up-middle (rum); and viceversa.



Folding	lum	luc	ruc	rum
between up-middle	luo	lu	ru	ruo

Figure 7: Folding Action: between-up-middle.

The folding action/operation *between down-middle* represents a paper folded between *down* and *middle* areas (according to Figure 2). The inference table in Figure 8 shows that, in this situation, the hole location left-down-corner (ldo) is equivalent to left-down-middle (ldm), and viceversa. Similarly, left-down (ld) location is equivalent to left-down-centre (ldc), right-down (rd) to right-down-centre (rdc) and right-down-corner (rdo) to right-down-middle (rdm); and viceversa.



Folding	ldm	ldc	rdc	rdm
between down-middle	ldo	ld	rd	rdo

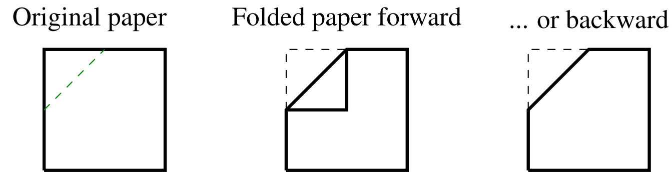
Figure 8: Folding Action: between-down-middle.

Note that the folding actions *between-up-middle* and *between-down-middle* are symmetrical operations, since the inferences are the same and the only parameter changed is both inference tables is up (u) for down (d).

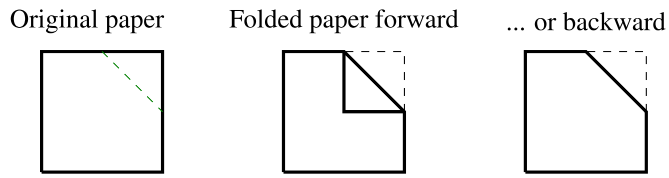
The folding action/operation on corners are shown in Figure 9. The *left-up corner* folding action represents a paper folded on that corner. In this situation, if a hole is made in the location left-up-centre (luc), then it appears in left-up-corner (luo) too.

Similarly, the folding action/operation *right-up corner* represents a paper folded on the right-up corner. In this situation, a hole made in the right-up-centre (ruc) location, also affects the location right-up-corner (ruo).

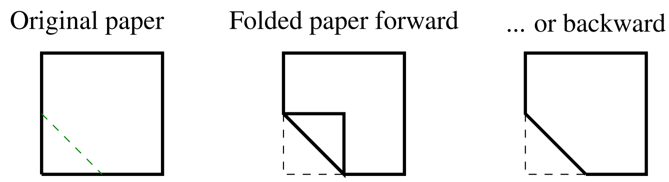
On the down side, the folding action/operation *left-down corner* represents a paper folded on that corner. In this situation, a hole made in the left-down-centre (ldc) location, also affects the left-down-corner (ldo) location.



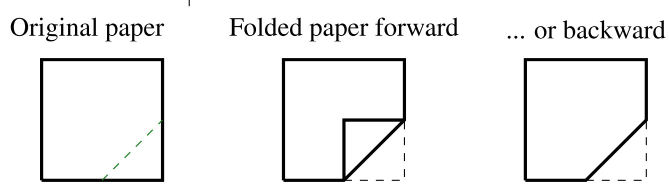
Folding	luc
left-up corner	luo



Folding	ruc
right-up corner	ruo



Folding	ldc
left-down corner	ldo



Folding	rdc
right-down corner	rdo

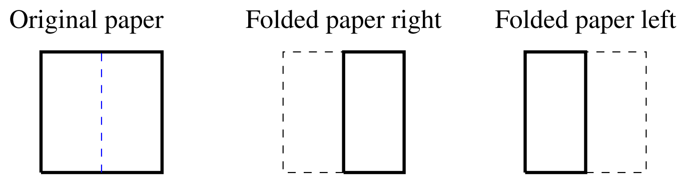
Figure 9: Folding Action on corners: left-up-corner, right-up-corner, left-down-corner, and right-down-corner.

The folding action/operation *right-down corner* represents a paper folded on the right-down corner. In this situation, a hole made on the right-down-centre (rdc) location also affects the right-down-corner (rdo) location.

It is interesting to note that:

- the folding actions *left-up corner* and *right-up corner* are symmetrical operations, since the inferences only differ in a parameter: right (r) for left (l) or viceversa.
- the folding actions *left-down corner* and *right-down corner* are also symmetrical operations, since the inferences only differ in a parameter: right (r) for left (l) or viceversa.
- the folding actions *left-up corner* and *left-down corner* are symmetrical operations, since the inferences only differ in a parameter: up (u) for down (d) or viceversa.
- the folding actions *right-up corner* and *right-down corner* are symmetrical operations, since the inferences only differ in a parameter: up (u) for down (d) or viceversa.

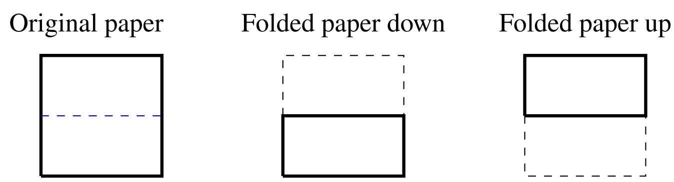
The folding action/operation *in-centre* represents a paper folded on the vertical middle. Figure 10 shows that, in this situation, the paper can be folded towards the left or towards the right side. If a hole is made on the left side of the paper, it is transmitted to the right side of the paper, and viceversa. In the inference table, note that all the locations on the left area of the



Folding	luo	lum	ldm	ldo	lu	luc	ldc	ld
in-centre	ruo	rum	rdm	rdo	ru	ruc	rdc	rd

Figure 10: Folding Action: in-centre.

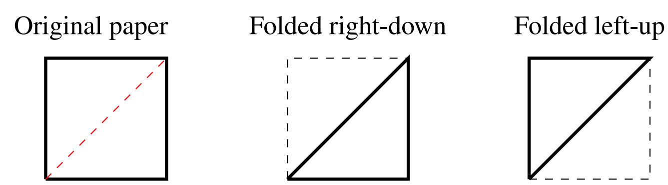
paper (luo, lum, ldm, ldo, lu, luc, ldc, ld), are equivalent to those locations on the right area of the paper (ruo, rum, rdm, rdo, ru, ruc, rd, rd). That is, all the locations are symmetrical since they only differ in one parameter: left (l) or right (r).



Folding	luo	lu	ru	ruo	lum	luc	ruc	rum
in middle	ldo	ld	rd	rdo	ldm	ldc	rdc	rdm

Figure 11: Folding Action: in middle.

The folding action/operation *in middle* represents a paper folded on half horizontally. Figure 11 shows that, in this situation, the paper can be folded towards *down* or towards *up*. If a hole is made on the up area of the paper, it is transmitted to the down area of the paper, and viceversa. In the inference table, note that all the locations on the *up* area of the paper (luo, lu, ru, ruo, lum, luc, ruc, rum), are equivalent to those locations on the *down* area of the paper (ldo, ld, rd, rdo, ldm, ldc, rdc, rdm). That is, all the locations are symmetrical since they only differ in one parameter: up (u) for down (d).

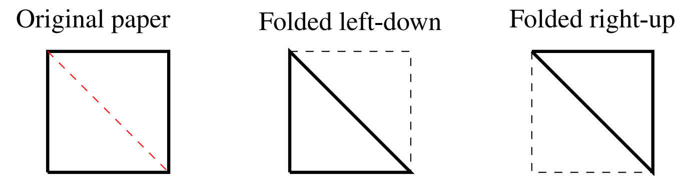


Folding	ldm	lum	luo	luc	lu	ru
diagonal upward	ld	rd	rdo	rdc	rdm	rum

Figure 12: Folding Action: diagonal-upward.

The folding action/operation *diagonal-upward* represents a paper folded on the diagonal that goes up from *left* to *right*. Figure 12 shows that, in this situation, the paper can be folded towards right-down or towards left-up. And in both situations, opposite locations are related (i.e., left-up-corner –luo– and right-down-corner –rdo–), as the inference table shows.

Similarly, the folding action/operation *diagonal-downward* represents a paper folded on the diagonal that goes down from *left* to *right*. Figure 13 shows that, in this situation, the paper can be folded towards left-down or towards right-up. And in both situations, opposite locations are related (i.e., right-up-corner –ruo– and left-down-corner –ldo–), as the inference table shows.



Folding	lu	ru	ruc	rum	rdm	ruo
diagonal downward	lum	ldm	ldc	ld	rd	ldo

Figure 13: Folding Action: diagonal-downward.

4 Connecting Several Foldings

Let us consider the following example situation in which two folding actions are done one after the other:

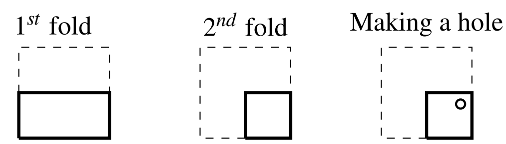


Figure 14: Example of how several foldings are related to each other.

Note that, when folding the paper *in-middle*, all the HoleLocation in the PaperArea *up (u)* are connected to those in the PaperArea *down (d)*, as it is shown by the corresponding inference table:

Folding	luo	lu	ru	ruo	lum	luc	ruc	rum
in middle	ldo	ld	rd	rdo	ldm	ldc	rdc	rdm

Then note also that, when folding the paper *in-centre*, all the HoleLocations in the PaperArea *right (r)* are connected to those in the PaperArea *left (l)*, as it is shown by the corresponding inference table:

Folding	luo	lum	ldm	ldo	lu	luc	ldc	ld
in-centre	ruo	rum	rdm	rdo	ru	ruc	rdc	rd

Therefore, when making a hole at the HoleLocation$_{LB}$ *rdm* (or right-down-middle), which other areas were punched? Note that *rdm* was connected to *rum* by folding *in-middle*, then it is inferred that two holes are punched located respectively at *rdm* and *rum*. Moreover, as these *rum* and *rdm* were connected to *lum* and *ldm* by the folding action *in-centre*, finally four holes are obtained in total: *rdm*, *rum*, *lum* and *ldm*.

In the next Section more details are given regarding the reasoning implementation process.

5 Solving Paper Folding Reasoning Questions

In order to test the QD-PF, we selected Prolog programming language [Sterling and Shapiro, 1994], which is based on Horn clause logic [Lloyd, 1987]. SwiProlog[3] was the testing platform [Wielemaker *et al.*, 2012].

As a testing dataset, we selected the *Paper Folding Test - VZ-2* from the *Manual for Kit of Factor-Referenced Cognitive Tests* by Ekstrom *et al.* [1976]. In this Educational Training Service (ETS) [Ekstrom *et al.*, 1976] factor kit Visualization (Vz) was defined as "the ability to manipulate or transform the image of

[3]SWI-Prolog: `http://www.swi-prolog.org/`

spatial patterns into other visual arrangements". The Vz-2 Paper Folding Test (suggested by Thurstone's Punched Holes [Carroll, 1993]) is described by as: *For each item successive drawings illustrate two or three folds made in a square sheet of paper. A drawing of the folded paper shows where a hole is punched in it. The participant selects one of 5 drawings to show how the sheet would appear when fully opened.*

The QD-PF inferences explained in Section 3 have been written using Prolog facts, such as:

```
action(FoldingAction, Hole1, Hole2).

Ex: action(right-down-corner,rdc,rdo).
```

And the QD-PF questions included in paper folding tests are written using Prolog facts as:

```
question(QuestionNumber, FoldingActions, Hole).
```

For example, the question showed in Figure 14 where the paper is first folded *in-middle* then *in-centre* and finally punched in *rdm*, is written as:

```
question(question2,[in-middle,in-centre],rdm).
```

When the problem only involves one folding action, it is easy to formulate and solve. See, for example, the question in Figure 15, which can be formulated as:

```
question(question1, [left-down-corner],ldc).
```

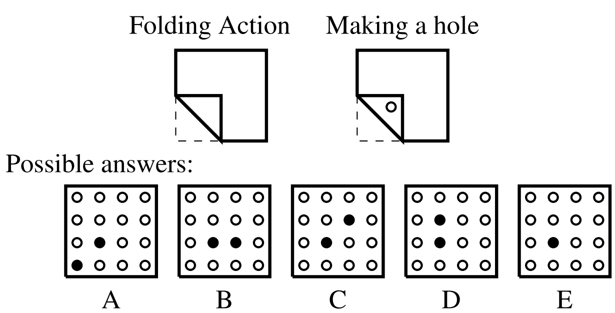


Figure 15: Question 1 in the *Paper Folding Test - VZ-2* by Ekstrom *et al.* [1976].

And the solution is easy to gather by applying the following logical inferences:

$$paper\_folding(Question, [Hole, H2]) \rightarrow question(Question1, Folding\_Action, Hole) \wedge action(Folding\_Action, Hole, H2). \quad (1)$$

In Prolog, we obtain:

```
?-paper_folding(question1, HoleList).
Inference:  action(left_down_corner, ldc, ldo).
HoleList: [ldc,ldo]
```

Note that the solution to Question 1 is option A.

When the problem involves two folding actions, we can have two situations: (i) the intermediate inferred hole location must be propagated to the next folding action; and (ii) the original hole location must be propagated to the next folding action.

Figure 16 shows an example of situation (i), which can be solved logically as:

$$paper\_folding([Folding\_A, Folding\_B], Hole, [Hole, H2, H3, H4]) \rightarrow action(Folding\_B, Hole, H2), action(Folding\_A, H2, H3), action(Folding\_B, H3, H4). \quad (2)$$

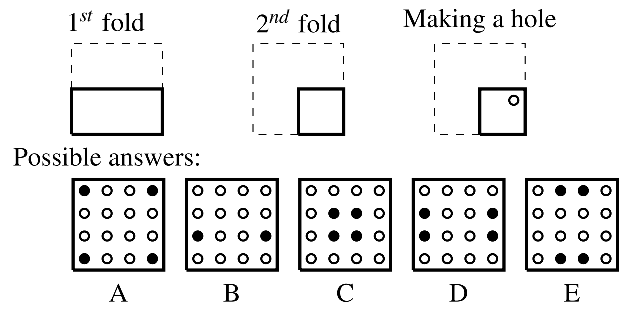


Figure 16: Question 2 in the *Paper Folding Test - VZ-2* by Ekstrom *et al.* [1976].

In Prolog, we obtain:

```
?-paper_folding(question2, HoleList):-
Folding actions:   [in-middle,in-centre]
Hole:   rdm
Inferences:
  action(in-centre, rdm, ldm),
  action(in-middle, ldm, lum),
  action(in-centre, lum, rum).
HoleList: [rdm, ldm, lum, rum]
```

Note that the solution to Question 2 is option D.

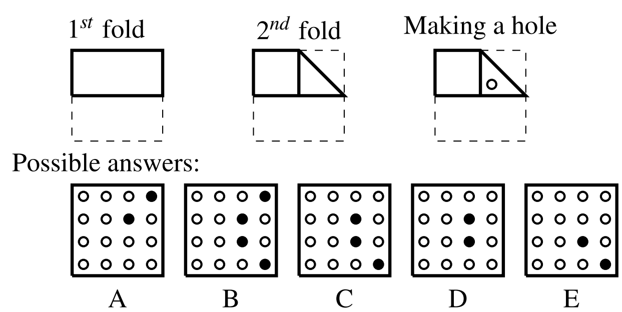


Figure 17: Question 5 in the *Paper Folding Test - VZ-2* by Ekstrom *et al.* [1976].

Figure 17 shows an example of situation (ii), which can be solved logically as:

$$paper\_folding([Folding\_A, Folding\_B], Hole, [Hole, H2, H3, H4]) \rightarrow action(Folding\_B, Hole, H2), action(Folding\_A, H2, H3), action(Folding\_A, Hole, H4). \quad (3)$$

In Prolog we obtain:

```
?-paper_folding(question5, HoleList):-
Folding actions:  [in-middle,right-up-corner]
Hole:  ruc
Inferences:
  action(right-up-corner, ruc, ruo),
  action(in-middle, ruo, rdo),
  action(ridht-up-corner, rdo, ruo), %repeated hole
  action(in-middle, ruc, rdc),
HoleList: [ruc,ruo,rdo,rdc]
```

Note that the solution to Question 5 is option B.

In general, in those problems involving 3 folding actions or more, multiple combinations of inferences are possible, so this propagation must be done automatically using a list of all hole inferred locations applied to the corresponding sequence of folding actions:

$$\begin{aligned}
&paper_folding(QuestionNum, HoleList) \rightarrow \\
&\quad question(QuestionNum, FoldingActions, HoleLocation), \\
&\quad reverse(FoldingActions, UnFoldingActions), \\
&\quad hole_propagate(UnFoldingActions, HoleLocation, HoleListA), \\
&\quad inferences_in_between(UnFoldingActions, [HoleLocation], \\
&\quad HoleListB), \\
&\quad mix(HoleListA, HoleListB, HoleListC), \\
&\quad remove_duplicates(HoleListC, HoleList).
\end{aligned} \tag{4}$$

Let us exemplify this with the question 8 showed in Figure 18:

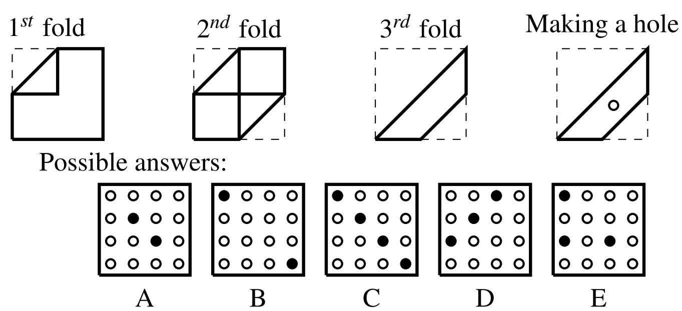


Figure 18: Question 8 in the *Paper Folding Test - VZ-2* by Ekstrom *et al.* [1976].

In Prolog, we obtained:

```
?-paper_folding(question8, HoleList)
FoldingActions:
 [left-up-corner,right-down-corner,diagonal-upward]
Hole:  rdc
Inferences:
 diagonal-upward (rdc,luc)
 right-down-corner (rdc,rdo)
 left-up-corner (rdc,rdc)
 diagonal-upward ([rdc],luc)
 right-down-corner ([luc,rdc],luc)
 left-up-corner ([luc,luc,rdc],luo)
HoleList: [luo,luc,rdo,rdc]
```

Note that the solution to Question 8 is option C.

6 Discussion about Future Work

A way to measure and train spatial intelligence in humans is using spatial tests, and paper folding is one of them. Cognitive intelligent agents must have common-sense reasoning mechanisms to be able to solve daily-living spatial problems. Manipulating paper for carrying out daily activities at home such as: wrapping a present, opening/closing the post, making signs/envelops/notes, and so on, it is a kind of spatial intelligence which home-robotics must learn.

This paper presented a qualitative definition for describing paper folding actions and for reasoning about their overlapping areas (QD-PF). Reasoning tables have been presented for inferring where a hole will appear after being punched in a paper area which resulted from a sequence of paper foldings. An approach towards solving paper folding questions logically when a hole is punched after one-to-three paper foldings is provided.

As future work, we intend: (i) to further develop the QD-PF by including algorithms for consistency checking to ensure that the system realizes in which locations a hole is not possible to be punched because –after carrying out certain folding actions– there is no paper in that area; (ii) to implement a paper-folding "smart" game for training people's spatial skills by providing them with feedback to improve their spatial intelligence; and (iii) to relate the folding actions here described with robot sensory-motor actions and contingencies to find out how a robot would be able to learn how to carry out actions with paper, such as wrap a gift, play origami, etc.

Acknowledgments

The funding and support by the Bremen Spatial Cognition Center[4] (BSCC), by the Universität Bremen under project *Cognitive Qualitative Descriptions and Applications* (CogQDA) and by the Deutscher Akademischer Austausch Dienst (DAAD) is gratefully acknowledged. I also thank the reviewers for their comments, which helped to improve this paper.

References

R. Al-Salman. *Qualitative Spatial Query Processing: Towards Cognitive Geographic Information Systems*. PhD thesis, University of Bremen, 2014. Supervised by Prof. Christian Freksa (University of Bremen) and Prof. Christian Jensen (Aalborg University).

A. Loai Ali, Z. Falomir, F. Schmid, and C. Freksa. Rule-guided human classification of volunteered geographic information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016. to appear.

M. Bhatt and F. Dylla. A qualitative model of dynamic scene analysis and interpretation in ambient intelligence systems. *I. J. Robotics and Automation*, 24(3), 2009.

M. Bhatt and C. Freksa. Spatial computing for design an artificial intelligence perspective. In John S. Gero, editor, *Studying Visual and Spatial Reasoning for Design Creativity*, pages 109–127. 2015.

M. Bhatt, H. Guesgen, and J.C. Augusto, editors. *Proc. of the Workshop on Space, Time and Ambient Intelligence (STAMI 2011)*, Pasadena, California, USA, 2011. International Joint Conference on Artificial Intelligence (IJCAI 2011).

M. Bhatt, H. Guesgen, and H. Cook, editors. *Proc. of the Workshop: Space, Time, and Ambient Intelligence: Spatio-Temporal Aspects of Human-Activity Interpretation (STAMI 2013)*, Bellevue, Washington USA, july 2013. 27th AAAI Conference (AAAI-13).

J.B. Carroll. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press, 1993.

A. G. Cohn and J. Renz. *Qualitative Spatial Reasoning, Handbook of Knowledge Representation*. Elsevier, Wiley-ISTE, London, 2007.

A. Cohn, D. Hogg, B. Bennett, V. Devin, A. Galata, D. Magee, C. Needham, and P. Santos. Cognitive vision: Integrating symbolic qualitative representations with computer vision. In H. Christensen and H-H. Nagel, editors, *Cognitive Vision Systems*, volume 3948 of *Lecture Notes in Computer Science*, pages 221–246. Springer Berlin/Heidelberg, 2006.

R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen. *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ, 1976.

Z. Falomir and A.-M. Olteţeanu. Logics based on qualitative descriptors for scene understanding. *Neurocomputing*, 161:3–16, 2015.

[4] http://bscc.spatial-cognition.de/

Z. Falomir, E. Jiménez-Ruiz, M. T. Escrig, and L. Museros. Describing images using qualitative models and description logics. *Spatial Cognition and Computation*, 11(1):45–74, 2011.

Z. Falomir, L. Museros, V. Castelló, and L. Gonzalez-Abril. Qualitative distances and qualitative image descriptions for representing indoor scenes in robotics. *Pattern Recognition Letters*, 38:731–743, 2013.

Zoe Falomir. A qualitative model for reasoning about 3d objects using depth and different perspectives. In T. Lechowski, P. Waga, and M. Zawidzki, editors, *LQMR 2015 Workshop*, volume 7 of *Annals of Computer Science and Information Systems*, pages 3–11. PTI, 2015.

P. Fogliaroni. *Qualitative Spatial Configuration Queries. Towards Next Generation Access Methods for GIS*, volume 9 of *Dissertations in Geographic Information Science*. IOS Press, 2013. ISBN 978-1614992486.

H. Guesgen and M. Bhatt, editors. *Proc. of the Workshop on Spatial and Temporal Reasoning*, Pasadena, California, USA, 2009. International Joint Conference on Artificial Intelligence (IJCAI).

K. Highfield and J. Mulligan. The role of dynamic interactive technological tools in preschoolers' mathematical patterning. In J. Watson and K. Beswick, editors, *Proc. of the 30th annual conference of the Mathematics Education Research Group of Australasia*, volume 1, pages 372–381. MERGA, 2007.

Tetsuo Ida, Fadoua Ghourabi, and Kazuko Takahashi. Formalizing polygonal knot origami. *Journal of Symbolic Computation*, 69:93 – 108, 2015. Symbolic Computation in Software Science.

L. Kunze, C. Burbridge, M. Alberti, A. Tippur, J. Folkesson, P. Jensfelt, and N. Hawes. Combining top-down spatial reasoning and bottom-up object class recognition for scene understanding. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'14)*, Chicago, Illinois, US, 2014.

L. Kunze, C. Burbridge, and N. Hawes. Bootstrapping probabilistic models of qualitative spatial relations for active visual object search. In *Qualitative Representations for Robots, Proc. AAAI Spring Symposium, Technical Report SS-14-06*, pages 81–80, 2014. ISBN 978-1-57735-646-2.

G. Ligozat. *Qualitative Spatial and Temporal Reasoning*. MIT Press, Wiley-ISTE, London, 2011.

John Wylie Lloyd. *Foundations of logic programming. Symbolic computation: Artificial intelligence*. Springer-Verlag, 2nd, extended edition edition, 1987.

A. Lovett and K. Forbus. Cultural commonalities and differences in spatial problem-solving: A computational analysis. *Cognition*, 121(2):281 – 287, 2011.

A. Lovett and H. Schultheis. Modeling spatial abstraction during mental rotation. In *Proc. on the Cognitive Science Conference*, 2014. in press.

A. Lovett, M. Dehghani, and K. Forbus. Learning of qualitative descriptions for sketch recognition. In *Proc. 20th Int. Workshop on Qualitative Reasoning (QR), Hanover, USA, July*, 2006.

N. Newcombe. Picture this: Increasing math and science learning by improving spatial thinking. *American Educator*, 34(2):29–35, 2010.

N. Sinclair and C. D. Bruce. Spatial reasoning for young learners, Research Forum. In *Proc. of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education (PME 38 / PME-NA 36)*, pages 173–205, Vancouver, Canada, July 2014.

N. Sinclair, E. de Freitas, and F. Ferrara. Virtual encounters: the murky and furtive world of mathematical inventiveness. *ZDM. The International Journal on Mathematics Education*, 45(2):239–252, 2013.

L. Sterling and E. Shapiro. *The Art of Prolog (2nd Ed.): Advanced Programming Techniques*. MIT Press, Cambridge, MA, USA, 1994.

J. Wai, D. Lubinksi, and C. P. Benbow. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4):817–835, 2009.

Jan Wielemaker, Tom Schrijvers, Markus Triska, and Torbjörn Lager. SWI-Prolog. *TPLP*, 12(1-2):67–96, 2012.

Reconciling Function and Structure in Scientific Models

Scott Friedman, Mark Burstein, David McDonald
SIFT, Minneapolis, MN, USA
{friedman, burstein, dmcdonald}@sift.net

James Pustejovsky, Peter Anick
Brandeis University, Waltham, MA, USA.
{jamesp, panick}@cs.brandeis.edu

Rusty Bobrow
Bobrow Computational Intelligence, LLC
rjbobrow@gmail.com

Brent Cochran
Tufts University School of Medicine, Boston, MA, USA
brent.cochran@tufts.edu

Abstract

Despite our increasing understanding of the structure and dynamics of scientific domains, functional knowledge and functional language— such as referring to a central purpose or function of a molecule— permeate scientific articles. Cognitive systems that collaborate with scientists must therefore represent functional knowledge to support machine reading and explanation. This paper describes our progress on automatically inferring and representing functional knowledge in R3 (*Reading, Reasoning, and Reporting*). R3 automatically reads biology articles from PubMed Central, using a massive domain model from Pathway Commons (www.pathwaycommons.org/) as background knowledge. R3 now relates functional language to its background structural model and explains functional knowledge, which is the central contribution of this paper. We motivate the representation of functional knowledge in the biology domain— which many existing ontologies omit— using examples from PubMed articles. We then describe how R3 automatically adds functional knowledge to its model by parsing textual summaries of biological processes and extracting semantics. We then describe how R3 builds event structures and compositional models with functional knowledge, and we illustrate how R3 uses its functional knowledge to diagram protein activity from the information it learned from reading.

Introduction

The concepts and factors we use to model scientific domains for our intelligent systems are often incommensurable with the concepts and factors we use to communicate scientific findings to our human peers. This is for a good reason, since intelligent systems and humans often serve complementary roles in the scientific process: machines engage in parallel search and discovery over vast structural models and networks of entities, while people frequently learn and communicate salient forms of entities with functional or intentional language. For instance, biologists often describe proteins and other natural kinds with functional contextual descriptors such as "active" and "inactive," and they compactly refer to the "activity" of an entity as its central function within a complex system.

Biologists often use artifactual mental models— such as *molecular switches*— to describe and reason about proteins. The molecular switch metaphor explicitly describes natural kinds (i.e., proteins) as *artifacts* (i.e., on/off switches), rather than just describing the behavior or capability of the natural kinds. For instance, this sentence from Akinleye et al. (2013) describes the proteins of the Ras family as molecular switches that are *inactive* (i.e., functionally *off*) when bound to GDP and *active* (i.e., *on*) when bound to GTP:

> "H-Ras, K-Ras, and N-Ras function as molecular switches when an inactive Ras-GDP is converted into an active Ras-GTP."

This relates a structural change (i.e., GDP/GTP binding) to a contextual function: when bound to GTP, Ras is able to perform its agreed-upon function (as opposed to many other reactions that Ras engages in) within a specific cell signaling pathway.

Intelligent systems that learn by reading must bridge this structural-functional gap: given only the *structural* knowledge from an ontology describing a complex system, a system can not resolve references to "active" or "inactive" entities that collectively "contribute to" some macro behavior, nor can it resolve references to the "activity" or functional capabilities of an entity. Our *Reading, Reasoning, and Reporting* (R3) system, developed as part of DARPA's Big Mechanism program (Cohen, 2015), reads articles in molecular biology to extend and revise its structural and functional models of biological mechanisms (Friedman et al., submitted, McDonald et al., 2016).

Our recent extensions to R3 aim to *automatically* bridge the structural-functional gap. This involves extending traditional compositional modeling semantics (Falkenhainer and Forbus, 1991) to support event structure (Pustejovsky, 1991b) and *telic qualia* (i.e., functional descriptions) from Generative Lexicon (GL) theory (Pustejovsky, 1991a).

Extending the modeling semantics gives R3 the representational capabilities, but it also needs the *content* to construct and populate these models. Fortunately, the model we are extending is annotated with English summaries of the individual reactions involved, written by human experts. R3 is thus able to extend the domain model to include critical functional information by automatically reading these textual summaries embedded within the model. R3 extracts the functional semantics from the summaries and automatically extends its model by adding functional characterizations of the reactions. It automatically identifies (1) events that comprise the entity's function (2) structural preconditions for

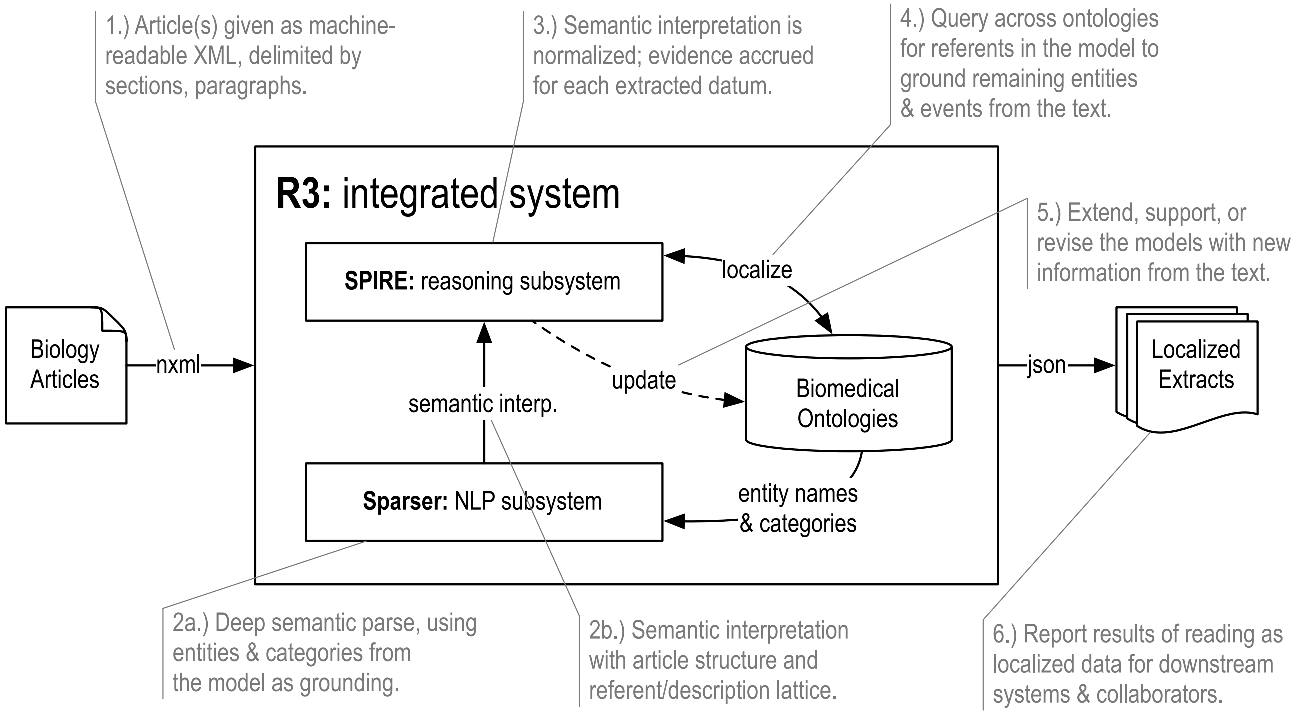


Figure 1: The R3 architecture, and the flow of information by which R3 reads articles, updates its mechanism models, and publishes extracted knowledge for human and machine collaborators.

functional status, and (3) precursor events that enable functional status.

R3's resulting models relate the structure of entities to the functional status of entities, as well as the events that comprise their function. This allows R3 to marshal functional knowledge into context when reading about structure, and also, to read-between-the-lines and ground functional references.

We begin by describing how R3 extracts and recognizes biological events and interactions from text. We then discuss R3's use of traditional qualitative modeling semantics with GL semantics in order to describe constituent entities, event structure, and functional knowledge. We discuss R3's approach to automatically extending its domain model with functional knowledge, and we present R3's automatically-generated results, including event structures and compositional models that accurately describe the well-known functions of the corresponding entities. We review related knowledge representation work for expressing the function of systems and devices, and distinguish our linguistically-motivated knowledge representation. We close with a discussion of the implications and future work for R3.

Deep Parsing in Molecular Biology

Biomedical research articles are written to be read by other professional biologists who are presumed to have the requisite technical background. Building a system such as R3 that can read-with-a-model in the biomedical domain poses key research challenges for the general task and the specific domain:

- Texts frequently use the same word to mean different types of objects in the model. *"RAS"* can refer to a protein, a gene, or a larger multi-protein complex, within a single article.
- Texts may describe things at different levels of abstraction than the model. For example, authors frequently talk about the the *function* of events while a purely mechanistic model may only describe the biochemical reactions taking place.
- One process or event may be part of many other processes or events in the domain model.

To address these challenges, R3 integrates deep semantic parsing, ontology mapping, and reasoning about structure, function, and mechanism-level causality, as shown in Figure 1.

Deep parsing allows R3 to extract precise semantics and determine entity types from local lexical context. R3 uses the SPARSER (McDonald, 1996) rule-based, type-driven semantic parser to read the texts. SPARSER'S rules succeed only if the types of the constituents to be composed satisfy the type constraints (value restrictions) specified by the rule. SPARSER compiles a semantic grammar from a semantic model of the information to be analyzed and a specification of all the ways each of the concepts can be realized in the language of the genre, such as biomedical research articles. This ensures that everything SPARSER is able to parse it can

Datum	Model Fragment	Generative Lexicon
Category of the resulting instance.	Type	Formal role
Instances & types of constituents/parts.	Participants	Argument structure
Existence criteria over participants.	Constraints	Constitutive role
Applicability of the model to the class of problem.	Modeling assumptions	N/A
Criteria for asserting behavior.	Conditions	Telic role(s) LHS
Behavior of the object/relation/event.	Consequence	Telic role(s) RHS
Events and state changes over explicit time.	Encapsulated histories*	Event structure
Function or purpose of the object/relation.	N/A	Telic role(s)

Figure 2: High-level mapping of qualitative model fragment semantics against Generative Lexicon semantics. [*] Encapsulated histories (Forbus, 1984) are defined *outside* of model fragments, but like GL event structure, encapsulated histories make explicit reference to time and state.

model, and that every rule in the compiled grammar has an interpretation.

R3 performs ontology-mapping with the SPIRE reasoner to map from SPARSER'S ontology into the BioPAX ontology of the domain model. This allows R3 to perform structure-mapping between semantic parses and the domain model in order to transfer knowledge from the parse into the model. This is crucial for learning by reading, as we discuss below.

Modeling with Generative Lexicon

Generative Lexicon (GL) theory assumes that word meaning is structured according to four generative *qualia roles* that describe how people understand entities and events and relations in the world. In GL, *lexemes* are words, word roots or phrases and their variants. They refer to entities, events, or relations, together with their associated semantics represented by *argument structure* and *qualia*, and organized by *habitats*.

There are four qualia roles of any lexeme:

- *Formal*: the basic ontological category.
- *Constitutive*: the relationships among constituent parts.
- *Telic*: its purpose or function.
- *Agentive*: how it came into being.

In addition to qualia roles, GL *argument structure* describes the constituents (i.e., the primitive entities, other lexemes, or sets thereof) that jointly participate in the lexeme, and their role within the lexeme. The GL *habitat* of an entity is a partial minimal model that enhances its qualia structure (Pustejovsky, 2013), and describes the *event structure* (i.e., events and sub-events) with reference to time. We provide example lexemes below to illustrate all of these GL concepts in the biology domain.

GL theory semantics supports the basic model fragment semantics for compositional modeling (Falkenhainer and Forbus, 1991, Rickel and Porter, 1997) as shown in Figure 2. The Figure 2 mapping has gaps: GL does not have an analog for meta-level modeling assumptions, so it cannot natively specify problem-level information such as applicable levels of granularity. We overload the GL telic role to describe both behavior (i.e., model fragment conditions and consequences) and function (i.e., the teleology or purpose of the lexeme). These gaps exist for good reason, since the two representations aim to solve different problems, but as shown in Figure 2, composition is a central capability of both representations.

Our lexeme representation violates the *no-function-in-structure* principle (de Kleer and Brown, 1981), which states that the rules for specifying the behavior of any constituent part of a system can in no way refer— even implicitly— to how the overall system functions. These lexeme-based models violate the no-function-in-structure principle by explicitly representing the function of proteins (through telic roles and habitats) within the larger cell signaling pathway.

GL theory distinguishes between *artifacts* and *natural kinds* via the telic role: artifacts have a telic role to express their function, whereas natural kinds have no inherent function, and therefore have no telic role. Biology articles do not adhere to this distinction: biologists frequently refer to a protein's *"function"* or *"activity,"* which effectively ascribes a purpose to a natural kind. We therefore model proteins as *artifactual* types with telic roles.

Artifactual models of natural kinds allow us to represent the widely-used artifactual "molecular switch" model of proteins within a larger pathway. Two such molecular switch lexemes induced by R3 are shown below, one for RAS, and one for MAPK. Both are summarized from the original R3 output for simplicity.

The RAS and MAPK molecular switches are defaulted to *off*, but as described in their telic roles, when events such as GTP-binding or phosphorylation occur, the molecules enter an *on* state. In the RAS lexeme, the *on* state enables RAS to function as a reactant in the activation of RAF, as a catalyst in the activations of MAP2K and MAPK, and as a catalyst in its own deactivation.

$$\begin{bmatrix} \textbf{H-RAS(x)} & \\ \text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \textbf{Protein(x):}id = P01111 \\ \text{TELIC} = \mathcal{C} \rightarrow [\textbf{bind}(x, \textbf{GTP})]On(x) \end{bmatrix} \\ \text{HABITAT} = & \begin{bmatrix} \text{ON(X)} = reactant(x, act(\textbf{RAF})) \\ \text{ON(X)} = reg_{+}(x, act(\textbf{MAP2K})) \\ \text{ON(X)} = reg_{+}(x, act(\textbf{MAPK})) \\ \text{ON(X)} = reg_{+}(x, deact(x)) \end{bmatrix} \end{bmatrix}$$

In the MAPK lexeme, activation occurs when the MAPK1 dimer is phosphorylated at Threonine and Tyrosine sites (not represented below), which enables its function of (1) positively regulating the deactivations of RAF and MAP2K and (2) translocating from the cytosol to the nucleoplasm.

$$\begin{bmatrix} \textbf{MAPK1\_dimer\_cytosol(x)} & \\ \text{ARGSTR} = & \begin{bmatrix} \text{SELF} = x\textbf{:Complex} \\ \text{COMPONENT} = y\textbf{:Protein:}id = P28482 \\ \text{COMPONENT} = z\textbf{:Protein:}id = P28482 \end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \textbf{Complex(x)} \\ \text{CONST} = [comp(x, y), comp(x, z)] \\ \text{TELIC} = \mathcal{C} \rightarrow [\textbf{phos}(\{y, z\})]On(x) \end{bmatrix} \\ \text{HABITAT} = & \begin{bmatrix} \text{OFF(X)} = loc(x, \textbf{cytosol}) \\ \text{ON(X)} = reg_{+}(x, deact(\textbf{MAP2K})) \\ \text{ON(X)} = reg_{+}(x, deact(\textbf{RAF})) \\ \text{ON(X)} = reactant(x, deact(x)) \\ \text{ON(X)} = [\textbf{move}(x)]loc(x, \textbf{nucleoplasm}) \end{bmatrix} \end{bmatrix}$$

These lexemes bridge the gap between structural status of entities (e.g., the phosphorylation status, molecular subcomponents, and molecule bindings) and the salient functional capabilities of entities within a larger system (e.g., their ability to translocate and activate other entities). With these lexemes available to its reading operations, R3 can utilize textual references to *"active RAS"* and *"RAS function"* to marshal important background knowledge about the structure of RAS when in the functionally active state, and the events that constitute RAS' function when active, respectively.

The events described in the habitats of the above auto-generated lexemes comprise small fractions of the RAS and MAPK events in R3's initial domain model; these are the events that R3 ascribed only to the active forms (i.e., *on* state) of these proteins.

As we describe below, R3 has no *a priori* knowledge of what constitutes the "active" functional form(s) (i.e., *on* state) of any single molecule, since its initial BioPAX model contains no information about "active" forms. Furthermore, activation is associated with different molecular configurations for different proteins, so active states cannot generally be inferred directly from chemical makeup. Consequently, R3 reads texts to extend its domain model with this functional knowledge, using human characterizations of activity when specifically attributed to different proteins in the model. We next describe how R3 learns by reading, and we present results of R3 using its learned functional knowledge to generate diagrams of protein function.

Learning-by-Reading Experiment

Here we describe R3's approach to automatically mining a large, existing biology model to generate lexical knowledge and event models that can be used to represent and reason about protein function. R3 parses textual data embedded within the model, including experts' summaries and reaction/molecule display names, into semantic interpretations. R3 extends the model with novel knowledge from these interpretations about the *function* of the described proteins and complexes, since the model is initially describes only structures, locations, and reactions. R3 propagates and analyzes these functional labels— such as whether a protein is active or inactive— in order to characterize the protein's function with the following dimensions:

- Structural conditions (e.g., phosphorylation status), location conditions (i.e., where within the cell), and binding conditions (e.g., in complex with another molecule) for protein activity.
- Event precursors that enable or disable the above conditions. These are the protein's activation and deactivation events, respectively.
- Events that depend on the active form of the protein and *not* the inactive form, such as reactions where the active protein is a catalyst or a scaffold. These events— which often activate or deactivate other proteins— comprise the protein's function within the signaling pathway.

We continue with a description of the dataset and a summary of R3's reading operations, whereby R3 extends its model with functional knowledge to describe active and inactive forms of proteins. We then present the results of its functional knowledge mining and lexical KB population using these labels.

Dataset

For this experiment, we used the entire "Signaling by EGFR" subset of the open-source, peer-curated Reactome pathway database.[1] Reactome pathway models describe reactions, reactants (i.e., complexes, proteins, and other molecules), catalysis and regulation relations, and protein modifications (e.g., phosphorylation, ubiquitination). These are downloadable as BioPAX (Demir et al., 2010), an RDF/OWL-based standard for describing pathways and the molecules and reactions that comprise them. The "Signaling by EGFR" Reactome subset contains 128 biochemical reactions and 911 molecules (i.e., proteins, complexes, small molecules, and other physical objects).

At the time of writing, BioPAX (Level 3) lacks categories and relations to describe the functional activity of a given protein with respect to a pathway, e.g., when it is active or inactive *within the context of the pathway*. We extended BioPAX to add categories and relations to represent these functional forms.

Machine Reading

R3 parsed summaries and display-names (i.e., descriptive labels) of 17 reactions that refer to molecules as "active," "inactive," "stimulated," or "activated," or that refer to the "activation" of a protein or describe how a protein "activates" another. Consider this example:

[1]The BioPAX OWL files are downloadable via the pathway browser: http://www.reactome.org/PathwayBrowser/

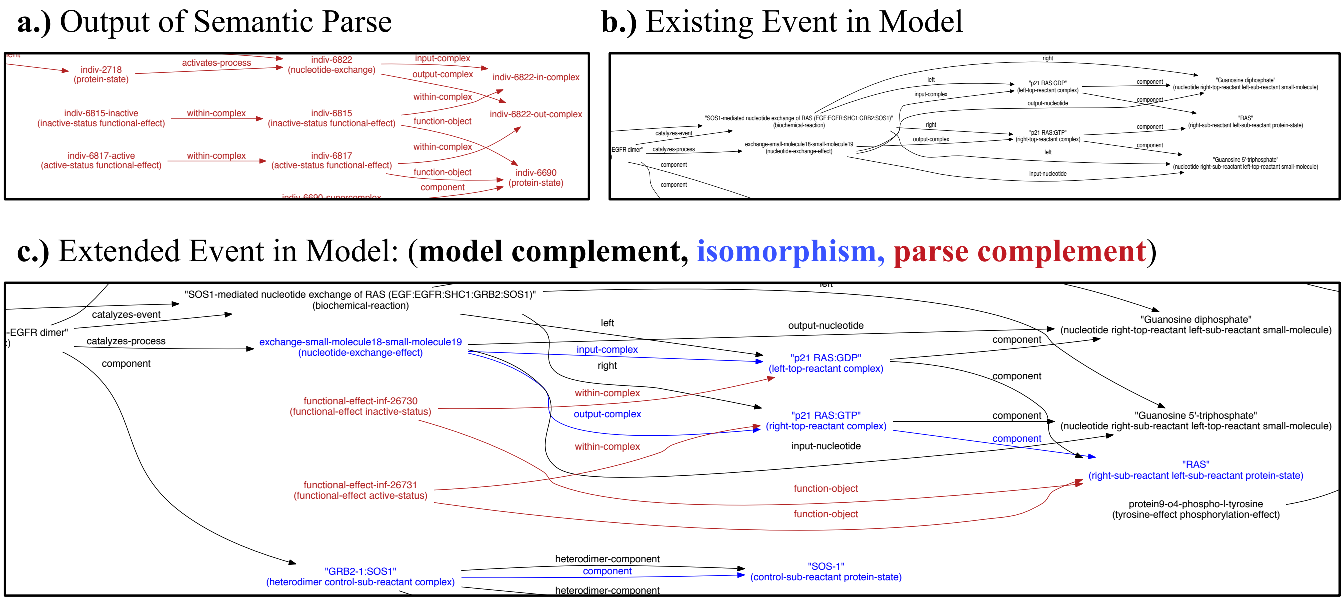


Figure 3: R3 extends its model with functional knowledge by reading: (a) its semantic parser interprets an expert's summary into a semantic graph; (b) it builds the semantic graph of the corresponding reaction; and finally (c) it computes the maximum common subgraph between the two, projecting the complement of the interpretation semantics into the model.

"SOS1 is the guanine nucleotide exchange factor (GEF) for RAS. SOS1 activates RAS nucleotide exchange from the inactive form (bound to GDP) to an active form (bound to GTP)."

R3 parses this summary to produce a semantic interpretation graph, a portion of which is displayed in Figure 3(a). It then uses the BioPAX semantic graph of the reaction, shown in Figure 3(b), to match against the interpretation graph, using a structure-mapping algorithm (Falkenhainer, Forbus, and Gentner, 1989) that supports additional constraints for identicality-matching. This produces the mapping shown in Figure 3(c). Here we illustrate that some entities and relations of the biology model (shown in blue) are isomorphic to the entities and relations of the semantic interpretation, despite using different symbols to describe entities. Provided the isomorphic subgraph, R3 projects a portion of the interpretation into the model: when the protein RAS is bound to GTP (to form the complex "p21 RAS:GTP"), it has an *active-status*, and when it is bound to GDP (to form the complex "p21 RAS:GDP") it has an *inactive-status*. R3 propagates this inference to all relevant super-complexes and reactions that contain these forms of RAS. R3 performs these steps for every summary and display-name it reads.

Populating Lexemes & Event Models

After identifying all of the active and inactive forms of proteins referenced in the summaries of the "Signaling by EGFR" subset of Reactome, R3 analyzes each active protein to generate lexemes.[2] This involves identifying structural preconditions, location preconditions, and molecular binding preconditions for *"active"* status, as shown in the RAS and MAPK lexemes listed above. In total, R3 generated 15 lexemes to describe active variants of RAS (3 lexemes), RAF (4 lexemes), MAP2K (3 lexemes), and MAPK (4 lexemes). There are multiple lexemes of each protein, since within the EGFR signaling subset of Reactome, *"active RAS"* refers to GTP-bound HRAS, KRAS, and NRAS of the RAS family. Similarly, *"active MAP2K"* can refer to a phosphorylated homodimer of MAP2K1, a phosphorylated homodimer of MAP2K2, or a phosphorylated MAP2K1/MAP2K2 heterodimer.

After identifying active components and building lexemes, R3 uses the functional knowledge in its lexemes to build an event graph of protein activation and protein function, as shown in Figure 4.

The event structure describes the active and inactive forms of molecules across cellular locations, as well as the biochemical-reactions ("R" nodes) that activate, deactivate, and translocate them. The triangular arrowheads indicate input and output reactants to the reactions, and the circular arrowheads indicate direct regulatory relationships— such as catalysis— between entities and reactions.

This automatically-generated event structure comprises a very small subset of R3's original BioPAX model, and it closely resembles the well-known RAS-RAF-MAP2K-MAPK activation cascade.

[2]Some proteins, such as SOS-1, are not referenced as being "active" in any textual Reactome summary, despite having an "active" form in the wider literature, so R3 does not have functional knowledge about these proteins.

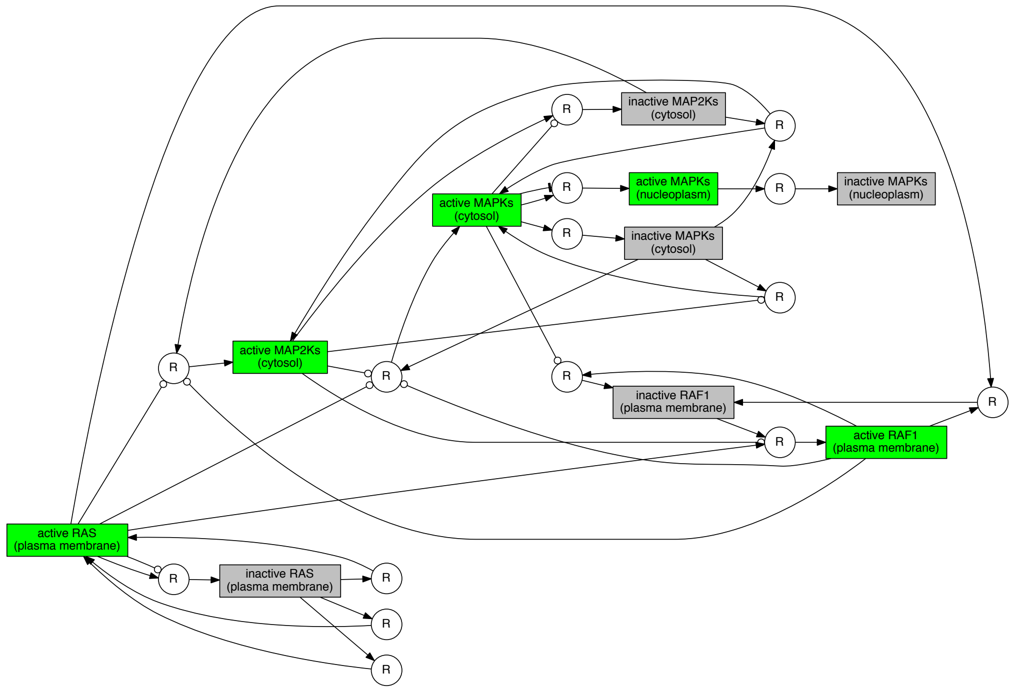


Figure 4: R3 automatically infers and displays the interrelated event structures of RAS, RAF1, MAP2K (MEK), and MAPK (ERK), including activations, deactivations, and translocations of the entities.

Related Work

Other work in philosphy and biology has discussed biological function, dysfunction, and malfunction (Krohs and Kroes, 2009) and identified different notions of function, including activity, role, biological advantage, and effect (Wouters, 2003). This work on R3 evokes the notion of function as biological activity, which we achieve by modeling natural kinds as artifacts.

Other research has used compositional modeling methods in biological domains (e.g., Mallavarapu et al., 2009, Rickel and Porter, 1997) to represent structure and behavior. These compositional modeling methods are useful for reasoning with the output of R3, which automatically produces scientific model components from text.

In engineering, functional knowledge has been used in simulation, diagnosis, design, and other tasks for decades (e.g., Chandrasekaran, 1994, Freeman and Newell, 1971, Goel, 2013). Knowledge-based systems use *functional representation* (FR) languages for describing the function of systems or components and the structural and causal processes that achieve the function. Previous approaches generally focus on encoding the *intent* of devices— in a top-down FR that complements traditional qualitative modeling— and then annotating causal transitions to explicate how the device achieves the intent (Chandrasekaran, 1994). Conceptual design systems also utilize functional knowledge— such as structure-behavior-function (SBF) models— by taking FR specifications and producing a structural specification that achieves the desired function (Goel, Rugaber, and Vattam, 2009, Goel, 2013).

Like other FRs, R3's lexeme-based approach explicitly describes the relation of structure to function; however, since R3 is a learning-by-reading system, its functional representation uses linguistically— and cognitively— motivated representations (Pustejovsky, 1991a,b). This will allow R3 to retrieve lexemes at parse-time and marshal background knowledge, e.g., about structure and function, into its semantic interpretation.

Conclusion & Future Work

This paper outlined our approach to reconciling structural and functional knowledge in the biology domain. We described extensions to traditional qualitative modeling semantics by inforporating telic roles and habitats from GL theory to encode functional knowledge. This enabled us to model the popular "molecular switch" artifactual perspective of natural kinds in our R3 system. We presented results of R3 learning functional knowledge by reading in order to to populate functional models, and then we demonstrated that R3 can use its functional knowledge to display protein activity (i.e., the function of a protein when the molecular switch is *"on"*).

This paper does not demonstrate R3's using its learned functional knowledge *while it reads*; this is a primary focus of our present work on R3. For instance, if R3 encounters a mention of "active Ras-GTP" in an article (e.g., Akinleye et al., 2013), it will use the lexemes induced in this work to marshal relevant functions of active Ras from the lexeme's habitat, including the phosphorylation— and activation— of Raf. When it reads the next sentence of the same article, stating that “Ras recruits and activates Raf kinases,” the corroborating background knowledge about the event and processes is already present in R3’s reasoning context. These lexemes thereby support R3’s overall objective of reading-with-a-model.

Acknowledgments

This work was supported in part by the DARPA “Big Mechanism” program, BAA 14-14, under ARO contract W911NF-14-C-0109. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the position or policy of the Government.

References

Akinleye, A.; Furqan, M.; Mukhi, N.; Ravella, P.; and Liu, D. 2013. Mek and the inhibitors: from bench to bedside. *J Hematol Oncol* 6(1):27.

Chandrasekaran, B. 1994. Functional representation: a brief historical perspective. *Applied Artificial Intelligence an International Journal* 8(2):173–197.

Cohen, P. R. 2015. Darpa’s big mechanism program. *Physical Biology* 12.

de Kleer, J., and Brown, J. S. 1981. Mental models of physical mechanisms and their acquisition. *Cognitive skills and their acquisition* 285–309.

Demir, E.; Cary, M. P.; Paley, S.; Fukuda, K.; Lemer, C.; Vastrik, I.; Wu, G.; D’Eustachio, P.; Schaefer, C.; Luciano, J.; et al. 2010. The biopax community standard for pathway data sharing. *Nature biotechnology* 28(9):935–942.

Falkenhainer, B., and Forbus, K. D. 1991. Compositional modeling: finding the right model for the job. *Artificial intelligence* 51(1):95–143.

Falkenhainer, B.; Forbus, K. D.; and Gentner, D. 1989. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence* 41(1):1 – 63.

Forbus, K. D. 1984. Qualitative process theory. *Artificial intelligence* 24(1):85–168.

Freeman, P., and Newell, A. 1971. A model for functional reasoning in design. In *Proc. 2nd Int. Joint Conf. Artificial Intelligence*, 621–633.

Friedman, S.; Burstein, M.; McDonald, D.; Paullada, A.; Plotnick, A.; Bobrow, R.; Cochran, B.; Pustejovsky, J.; and Anick, P. submitted. Learning by reading: Extending & localizing against a model. In *Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems*.

Goel, A. K.; Rugaber, S.; and Vattam, S. 2009. Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 23(01):23–35.

Goel, A. K. 2013. A 30-year case study and 15 principles: implications of an artificial intelligence methodology for functional modeling. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 27(03):203–215.

Krohs, U., and Kroes, P. 2009. *Functions in biological and artificial worlds: comparative philosophical perspectives*. Mit Press.

Mallavarapu, A.; Thomson, M.; Ullian, B.; and Gunawardena, J. 2009. Programming with models: modularity and abstraction provide powerful capabilities for systems biology. *Journal of The Royal Society Interface* 6(32):257–270.

McDonald, D.; Friedman, S.; Paullada, A.; Bobrow, R.; and Burstein, M. 2016. Extending biology models with deep nlp over scientific articles. In *Proceedings of the AAAI-16 Workshop on Knowledge Extraction from Text*.

McDonald, D. D. 1996. The interplay of syntactic and semantic node labels in partial parsing. In Bunt, H., and Tomita, M., eds., *Recent Advances in Parsing Technology*. Kluwer Academic Publishers. 295323.

Pustejovsky, J. 1991a. The generative lexicon. *Computational linguistics* 17(4):409–441.

Pustejovsky, J. 1991b. The syntax of event structure. *Cognition* 1(41):47–81.

Pustejovsky, J. 2013. Dynamic event structure and habitat theory. *Proceedings of GL2013* 1–20.

Rickel, J., and Porter, B. 1997. Automated modeling of complex systems to answer prediction questions. *Artificial Intelligence* 93(1):201–260.

Wouters, A. G. 2003. Four notions of biological function. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 34(4):633–668.

QSRlib: a software library for online acquisition of Qualitative Spatial Relations from Video

Y Gatsoulis[1], **M Alomari**[1], **C Burbridge**[3], **C Dondrup**[2], **P Duckworth**[1], **P Lightbody**[2], **M Hanheide**[2], **N Hawes**[3], **D C Hogg**[1], **A G Cohn**[1]
[1] University of Leeds, UK; [2] University of Lincoln, UK; [3] University of Birmingham, UK.

Abstract

There is increasing interest in using Qualitative Spatial Relations as a formalism to abstract from noisy and large amounts of video data in order to form high level conceptualisations, e.g. of activities present in video. We present a library to support such work. It is compatible with the Robot Operating System (ROS) but can also be used stand alone. A number of QSRs are built in; others can be easily added.

Introduction

Humans can effectively make sense of their surroundings and easily recognise the activities that take place using their perceptual abilities. Although these cognitive abilities are highly complex and not yet fully understood with many theories being suggested (Johnson-Laird 2008), there is general consensus that spatio-temporal relations and reasoning play an important role (Ragni, Fangmeier, and Bruessow 2010). For this, there are dedicated areas in our brains (Amorapanth, Widick, and Chatterjee 2010), which are able to form efficient and categorical qualitative representations of spatial relations resulting in powerful levels of flexible abstractions for inference and reasoning while avoiding information overflow and computational bottlenecks (Laeng 2013).

Within the field of artificial intelligence there has been considerable interest in developing techniques to make intelligent systems with similar spatial reasoning abilities. To this end, many spatial qualitative representations and calculi have been developed (Cohn and Renz 2008; Chen, Cohn, and Liu 2015), and have been used in a number of real-world application domains such as geographical systems (Van de Weghe et al. 2005), video analysis (Sridhar, Cohn, and Hogg 2011a), human activity recognition (Behera, Cohn, and Hogg 2012; Tayyub et al. 2014), etc. Despite their successes, there is no formal methodology to help decide which QSR is best suited for a given task, and often this is determined using domain knowledge, data analysis and empirical experimentation (Sridhar, Cohn, and Hogg 2011b).

As a result of all theoretical work that has taken place in developing new calculi for QSR, there have been a number of implemented systems with the aim of offering generic QSR reasoning services. For example SNARK (Stickel, Waldinger, and Chaudhri 2000) is an automated theorem proving system which builds in the mereotopological calculus, RCC, as does the commense knowledge system, CYC (Grenon 2003). SparQ [1] (Dylla et al. 2006) provides implementations of a wide variety of QSR calculi, as does CLP(QS). Another system providing qualitative spatial reasoning services is the Qualitative Algebra Toolkit (QAT) [2] (Condotta, Ligozat, and Saade 2006) which is a Java based toolkit implementing a constraint based approach for reasoning and has XML definitions of several qualitative algebras. GQR[3] (Westphal, Wölfl, and Gantner 2009) is another constraint based solver which supports binary constraint calculi for qualitative spatial/temporal reasoning.

However, the focus of all these systems is on *symbolic reasoning* – i.e. they generally assume that knoweldge is expressed already in symbolic form as a knowledge base of assertions involving qualitative relations over spatial entities and the reasoning services provided are primarily consistency checking and compositional reasoning; the issue of abstracting from metric data to form the qualitative knowledge base is not addressed at all (CLP(QS) does in fact allow for polygons to be directly input, but the focus is stil on symbolic reasoning). QSRlib is complementary to these systems in that it does not provide symbolic reasoning, but concentrates on qualitative abstraction from metric data. Moreover, these systems are not purposely designed for use on ROS-based robotic platforms, primarily owing to their choice of programming languages.

On the other hand there is one form of symbolic reasoning concerning *conceptual neighbourhoods* which these systems do not provide; a *conceptual neighbourhood* diagram (CND) – see Figure 3 – specifies which relations are conceptual neighbours – i.e. relations $R1$ and $R2$ are conceptual neighbours if $R1(a, b)$ holds and subsequently $R2(a, b)$ holds owing to the entities involved continuously deforming/translating, and there is no intermediate relation $R3(a, b)$ which holds. However, when data is acquired from sensors, and objects may be moving fast relative to the frame speed and relationship granularity, or objects may be occluded, then it is possible that a relation in the sequence

[1] http://www.sfbtr8.uni-bremen.de/project/r3/sparq/

[2] http://www.cril.univ-artois.fr/~saade/QAT/

[3] http://www.sfbtr8.spatial-cognition.de/de/projekte/reasoning/r4-logospace/research-tools/gqr/

through a CND may not be observed. In this case it may be useful to perform reasoning to detect this discrepancy and take corrective action (which might either be to insert the missing relations or to discard the observation, as in (Fernyhough, Cohn, and Hogg 1998)). This service is however provided in QSRlib.

As such, researchers often re-invent the wheel by repeating the implementations of QSRs. In an attempt to resolve these issues and speed up background development allowing better use of research time, we have developed a modern library, named QSRlib[4,5], with the aims of:

- Providing a number of QSRs that are well known, and in common use in scientific community.
- Exposing these QSRs via a standard IO interface that allows quick and easy re-usability, including a ROS interface to allow use in cognitive robotic systems.
- Providing a flexible and easy to use infrastructure to allow rapid development of new QSRs that extend the library.
- Delivering abstracted QSRs over time in an aggregated representation that facilitates further inference.

A typical usage of QSRlib would be an intelligent system, such as a robot for example, which acquires visual data via an RGB-D camera, such as a Kinect, and via object recognition and skeleton tracking is able to perceive the individual entities in the world. The system can then make calls to QSRlib in order to abstract this input data and form a qualitative representation of the perceived world scene. This could then be used to recognise activities in natural scenes such as the one shown in Figure 1, using already learnt models expressed using QSRs in the QSRlib library.

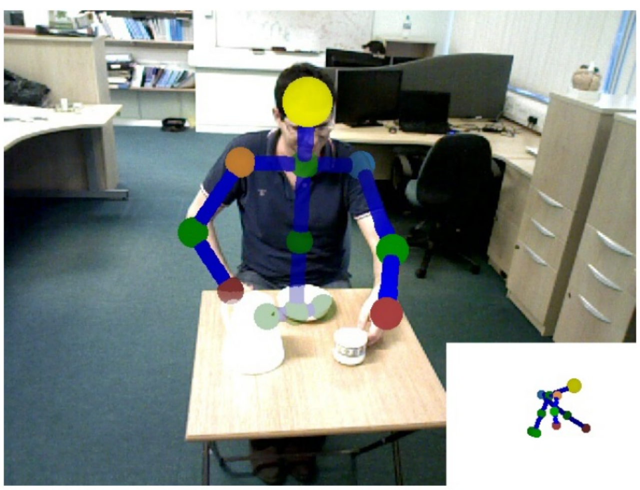


Figure 1: Activity recognition in a table top setting. Dyadic QSR relations between detected objects/skeleton points can be computed (bottom right inset).

[4]http://github.com/strands-project/strands_qsr_lib

[5]http://qsrlib.readthedocs.io

Description

QSRlib is based on a client-server architecture implemented in python 2.7 although measures for compatibility with 3.x have been adopted. Furthermore, the library can also be used with the Robot Operating System[6] (ROS), allowing its use in complex intelligent systems, such as robots. Figure 2 presents a flowchart with the main step processes for computing QSRs via the library. Qualitative spatial relations typically operate on object data such as their Cartesian positions, rotations, edges or bounding boxes describing their shape, velocities, etc., retrieved from single or multiple frames using 2D/3D trackers that are separate from QSRlib, hence allowing the use of state of the art developments in tracking. The raw data needs to be firstly converted into the common input data format of QSRlib, which represents a timeseries of the states of the perceived objects. Utility functions are provided that allow easy conversion of the raw data to this standard input data structure. This input data structure, the names of the requested QSRs to be computed and other options that control their behaviours are used to create a request message to the QSRlib server, which upon computation returns a response message that includes the computed QSRs as an output data structure similar to the input one, i.e. a timeseries of the QSRs between the objects, as well as other information. The QSRs computation is independent from each other, however, multiple QSRs can be requested and be computed in any frame and returned in both a standard data structure (equivalent to list of ground atomic formulas) as well as in a special graph structure, called a QSTAG (see below) which integrates them all into a single structure over a period of time. In our robot systems in real world deployment tests QSRlib was working from input of a 2D/3D vision system operating at over 20 frames per second.

QSRlib currently consists of directional, distance-based, motion-based, topology-based, and combined direction-topology-based QSRs (see also Table 1). Each of these consists of a set of *Jointly Exhaustive and Pairwise Disjoint* (JEPD) relations between the involved objects (typically two objects); i.e. exactly one relation should hold between any tuple of involved objects. We now briefly describe each of the currently implemented QSRs, and give citations to the literature where the formal definitions and semantics of these relations can be found.

Distance-based: A *Qualitative Distance Calculus* (QDC) (Clementini, Felice, and Hernandez 1997) provides qualitative relations based on a set of parameterised labels and distance boundaries, e.g. 'touch': $< .5m$, 'near': $> 5m$ and 'far': $> 5m$. In QSRlib we extend this and also provide a *Probabilistic Qualitative Distance Calculus* (PQDC) in which there are overlaps between the boundaries, and a probabilistic decision mechanism based on a Gaussian model.

Direction-based: *Cardinal Direction* (Frank 1990; 1996) relations specify compass-like directional relations between two objects with respect to their origin. The *Ternary Point Configuration Calculus* (Moratz, Nebel, and Freksa 2003;

[6]www.ros.org

Table 1: Description of supported qualitative spatial relation families

qualitative spatial relation families	type	num of relations / variations	kind of entities
Qualitative Distance Calculus	distance	user specified	2D points
Probabilistic Qualitative Distance Calculus	distance	user specified	2D points
Cardinal Directions	direction	9	2D rectangles
Moving or Stationary	motion	2	2D points
Qualitative Trajectory Calculus	motion	B11: 9, C21: 81	2D points
Rectangle/Block Algebra	topology & direction	169/2197	2D/3D rectangles
Region Connection Calculus	topology	2, 4, 5, 8	2D rectangles
Ternary Point Configuration Calculus	direction	25	2D points

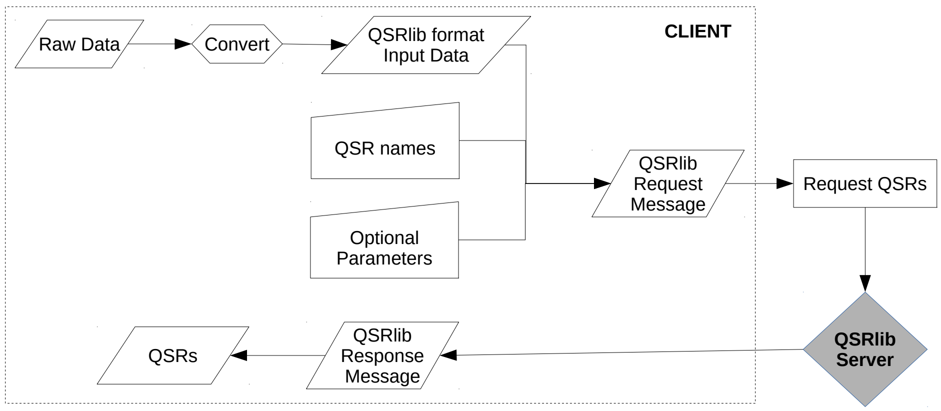


Figure 2: System Architecture.

Dylla and Moratz 2004) is more flexible as it allows the origin to be specified as a parameter, which is an advantage if dealing with multiple frames of reference, and it also further integrates distance-based relations, computed with respect to the horizon defined by the (variable) origin and the relative point of interest.

Topology-based: The *Region Connection Calculus* is a well established calculus for representing and reasoning about the mereotopology of regions. There are different sets of relations depending on the granularity desired, with the most common being RCC-8, which defines eight relations between two regions. These can be seen in Figure 3. The coarser calculus RCC-5 is often more useful in computer vision applications since the tangency distinctions made in RCC-8 but not RCC-5 are unreliable to compute due to noise in low level vision computations. In fact the user is able to supply a *quantisation factor* which allows control of how far apart two regions have to be before they become disconnected. QSRlib implements different variations of RCC, including those shown in Figure 3.

Combined Direction and Topology based: Allen's Interval algebra (Allen 1983) originally put forward as a calculus for qualitative temporal reasoning has also been used for reasoning about space, particularly in its 2D form, the *Rectangle Algebra* wherein objects are projected to the x and y axes and a relation consists of a pair of Allen relations. There are 13 Allen relations (see Figure 4, and hence 13×13=169 Rectangle Algebra relations. There also exists a 3D version of the interval calculus, called the *Block Algebra* (Balbiani, Condotta, and del Cerro 2002) having 13^3 relations which is also implemented in QSRlib.

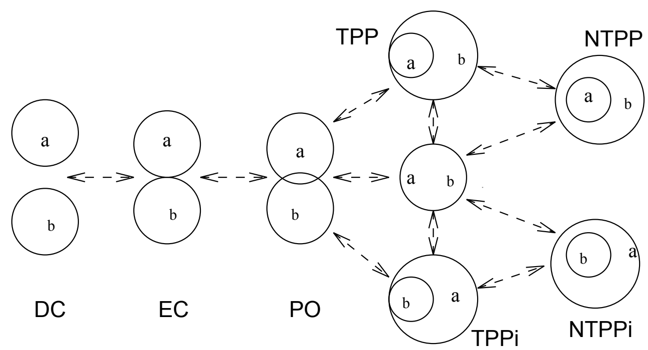


Figure 3: A 2D depiction of RCC-8 relations; RCC-5 is a coarser calculus which collapses DC,EC to DR, TPP, MTPP to PP and TPPi, NTPPi to PPi. The arrows depict the continuous transitions between relations and the entire diagram represents the *conceptual neighbourhood* diagram for RCC8.

Motion-based: The *Qualitative Trajectory Calculus* (QTC) (Van de Weghe et al. 2005; Delafontaine, Cohn, and Van de Weghe 2011) is a calculus for representing relations and reasoning about moving point-objects. There are several variations of QTC which define different types of motion-based relations. For example, QTC-B variants rep-

relation	symbol	inverse	meaning
preceeds	p	pi	
meets	m	mi	
overlaps	o	oi	
starts	s	si	
during	d	di	
finishes	f	fi	
equals	e	e	

Figure 4: The 13 jointly exhaustive and pairwise disjoint Allen interval calculus relations.

resent whether an object is approaching towards or departing from another object, and QTC-C variants compute the relative direction of movement of one object with respect to the other. The full specification of these QTC calculi also allows the specification of whether one object is moving faster, slower, or the same speed as the other. This functionality (present in QTC_{B12} and QTC_{B22}) is currently omitted from QSRlib. The particular versions currently implemented from (Delafontaine, Cohn, and Van de Weghe 2011) in QSRlib are: (1) QTC_{B11} which records the instantaneous motion of two point-objects x and y towards/away from each other: + meaning away from, – means towards and 0 means stationary; since x could be moving towards y but y moving away from x, each relation consists of a pair of these symbols: $\langle\alpha,\beta\rangle$ where $\alpha,\beta \in \{+,0,-\}$; (2) QTC_{C21} which extends QTC_{B21} by adding two further components forming a quadruple of relations$\langle\alpha,\beta,\gamma,\delta\rangle$ where δ represents whether x is moving to the left (–) or right (+) of the vector $\vec{xy}$ and γ whether y is moving to the left (–) or right (+) of the vector $\vec{yx}$.

Although QTC relations indirectly imply motion, this might not always be the case. For example, QTC_{B21} might compute that one object remains unchanged with respect to another ($\langle 0,0\rangle$, but this might be the outcome of both objects moving in parallel to each other. For this reason, QSRlib also includes a simple QSR, with just two relations, called *Moving or Stationary*, which as the name implies, determines whether an object is moving or is stationary.

Spatio-temporal relations over timeseries of QSRs For activity recognition we are usually interested in representing the temporal aspects that exist in a timeseries of QSRs. These can be represented in a standard knowledge base of QSR facts and temporal relations between the intervals involved, e.g. (Dubba et al. 2015). An alternative method, which has many advantages for data mining and learning is to use relational graphs, where the edges encode spatial/temporal relations. One particular such representation, known as *Qualitative Spatio-Temporal Activity Graphs* (QSTAG) (Sridhar, Cohn, and Hogg 2010), which provides a compact and efficient graph structure to represent both qualitative spatial and temporal information about objects of interest. In a QSTAG, the temporal relationship between a number of qualitative spatial timepoints is abstracted using Allen's Interval Algebra (Figure 4). A QSTAG has the further advantage of allowing the use of standard graph based methods, such as (approximate) matching. An example QSTAG is shown in Figure 5. QSRlib provides the ability to output QSTAGs abstracted from the input data over a period of time. As can be seen in Figure 5, the layer two spatial nodes can incorporate relations from more than one QSR calculus, in this case QDC (with relations touch/near/...) and QTC_{B21} (with relations $+/0/-$). Thus, QSRlib provides the ability to compute relations from multiple calculi simultaneously and output them in one integrated spatio-temporal representation.

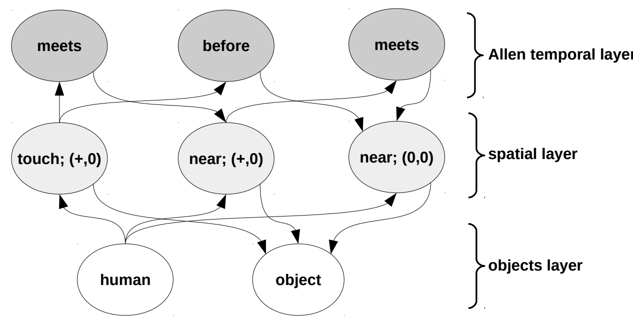


Figure 5: Example of a Qualitative Spatio-Temporal Activity Graph (QSTAG) between a human and an object; each spatial layer node encodes QSRs from two calculi: a QDC relation (touch/near) and a QTC_{B21} one ((+,0)/(0,0)).

Example Usage

QSRlib has been used in various research and teaching projects. We briefly describe some of this usage here to illustrate its possible future uses.

In (Duckworth et al. 2016a) QSRlib was used to rapidly experiment with multiple different types of qualitative representations in order to identify the most suitable one for learning human motion behaviours as perceived by a mobile robot that was deployed for a duration of six weeks in an office environment. QSRlib was used to quickly experiment with suitable representations for classifying scenes and environments from visual data (Thippur et al. 2015; Kunze et al. 2014). The library was also used to compute qualitative relations between a robot and humans moving in order to plan and execute safe path navigation taking into consideration their movement patterns (Dondrup et al. 2015). In (Duckworth et al. 2016b), multiple QSRs were used to represent a detected person's skeleton positions within a semantic map. The QSTAG framework was used to recover latent, semantically meaningful, patterns of how humans move within a scene in an unsupervised setting.

The library has also been used in a number of teaching projects (e.g. recognizing gestures for controlling a device, recognizing someone having breakfast, etc.), allowing the students to concentrate on the more interesting, high level, parts of their projects rather than spending a good portion of their project time in developing the low level tools they need (and which are the same from project to project).

Summary

This paper has presented a software library that allows easy and fast computation of qualitative spatial and temporal relations from objects tracked in video data. A number of implemented systems cited here have employed this library to understand and make predictions about the behaviour of physical systems, even in the presence of noisy quantitative information. Despite the large number of qualitative spatial calculi in the literature, any system aiming to use them to understand video data has had to implement an *ad hoc* solution to abstract the video data to qualitative spatio-temporal relations. QSRlib provides a library to do just this and has implementations of a number of commonly used qualitative calculi.

The standardized I/O data structures of QSRlib allow users to compute and process any set of desired QSRs included in the library with no additional effort. This means that researchers can focus their work on experimenting and analysing with the different types of QSRs, rather than spend time in implementing them and changing the format of their inputs. We have also provided the tools that allow contributors to extend the set of provided QSRs easily, quickly and with flexibility as needed. For example, the online documentation shows a minimal working example of developing and integrating a new QSR to the library; the process contains less than ten lines of key code. QSRlib is open source, well-documented and at a stable state with an active group of developers.

QSRlib is not a system for *reasoning* about qualitative relations, but rather a system for *acquiring* them. Having created a knowledge base of qualitative facts extracted via QSRlib it would be possible then to use the existing complementary QSR reasoning systems such as CLP(QS) or SparQ to perform conventional composition-based reasoning on the extracted qualitative facts. In fact, if data is abstracted from a single source, then it is likely that the extracted representation will be consistent since the world from which it is abstracted will be consistent; of course errors in low level computations and if data is obtained from multiple sources may mean this is no longer the case. Nevertheless, for the case of forming abstracted representations and building models of activity (e.g. learning event models) the principal use of QSRs in the literature has been as a *representation* language, rather than as a *reasoning* mechanism: qualitative spatio-temporal languages are able to abstract away from small metric variations in performance, and from noise in low level visual processing so that activity representations become (more) similar when represented qualitative, which greatly facilitates learning and interpretation of activities in video. QSRlib facilitates the implementation of systems taking this approach to activity.

Acknowledgements

The financial support of the EC under FP7 project STRANDS, 600623, is gratefully acknowledged.

References

Allen, J. F. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(11):832 – 843.

Amorapanth, P. X.; Widick, P.; and Chatterjee, A. 2010. The neural basis for spatial relations. *Journal of cognitive neuroscience* 22(8):1739 – 53.

Balbiani, P.; Condotta, J.; and del Cerro, L. F. 2002. Tractability results in the block algebra. *J. Log. Comput.* 12(5):885–909.

Behera, A.; Cohn, A. G.; and Hogg, D. C. 2012. Workflow Activity Monitoring using the Dynamics of Pair-wise Qualitative Spatial Relations. In *International Conference on MultiMedia Modeling (MMM), Klagenfurt, Austria*.

Chen, J.; Cohn, A.; and Liu, D. 2015. A survey of qualitative spatial representations. *The Knowledge Engineering Review* 30(1):106136.

Clementini, E.; Felice, P. D.; and Hernandez, D. 1997. Qualitative representation of positional information. *Artificial Intelligence* 95(2):317–356.

Cohn, A., and Renz, J. 2008. *Qualitative Spatial Representation and Reasoning*. Elsevier B.V. chapter 13, 551 – 596.

Condotta, J. F.; Ligozat, G.; and Saade, M. 2006. The QAT: A Qualitative Algebra Toolkit. In *2006 2nd International Conference on Information Communication Technologies*, volume 2, 3433–3438.

Delafontaine, M.; Cohn, A. G.; and Van de Weghe, N. 2011. Implementing a qualitative calculus to analyse moving point objects. *Expert Systems with Applications* 38(5):5187–5196.

Dondrup, C.; Bellotto, N.; Hanheide, M.; Eder, K.; and Leonards, U. 2015. A Computational Model of Human-Robot Spatial Interactions Based on a Qualitative Trajectory Calculus. *Robotics* 4(1):63 – 102.

Dubba, K. S. R.; Cohn, A. G.; Hogg, D. C.; Bhatt, M.; and Dylla, F. 2015. Learning relational event models from video. *Journal of Artificial Intelligence Research* 53:41–115.

Duckworth, P.; Gatsoulis, Y.; Jovan, F.; Hawes, N.; Hogg, D. C.; and Cohn, A. G. 2016a. Unsupervised Learning of Qualitative Motion Behaviours by a Mobile Robot. In *Proc. of the Intl. Conf. on Autonomous Agents and Multiagent Systems (AAMAS'16)*.

Duckworth, P.; Alomari, M.; Gatsoulis, Y.; Hogg, D.; and Cohn, A. 2016b. Unsupervised activity recognition using latent semantic analysis on a mobile robot. In *22nd European Conference on Artificial Intelligence (ECAI) (to appear)*.

Dylla, F., and Moratz, R. 2004. Empirical complexity issues of practical qualitative spatial reasoning about relative position. In *Workshop on Spatial and Temporal Reasoning, ECAI'04*.

Dylla, F.; Frommberger, L.; Wallgrün, J. O.; and Wolter, D. 2006. SparQ: A toolbox for qualitative spatial representation and reasoning. In *Workshop on Qualitative Constraint Calculi: Application and Integration at KI 2006*, 79 – 90.

Fernyhough, J. H.; Cohn, A. G.; and Hogg, D. C. 1998. Building qualitative event models automatically from visual input. In *Proceedings of the International Conference on Computer Vision ICCV98*, 350–355. IEEE.

Frank, A. U. 1990. Qualitative Spatial Reasoning about Cardinal Directions. In Mark, M., and White, D., eds., *Autocarto 10*. Baltimore: ACSM/ASPRS.

Frank, A. U. 1996. Qualitative Spatial Reasoning: Cardinal Directions as an Example. *Geographical Information Systems* 10(3):269–290.

Grenon, P. 2003. Tucking RCC in Cyc's ontological bed. In *IJCAI*, 894–899.

Johnson-Laird, P. 2008. *How We Reason*. Oxford University Press (OUP).

Kunze, L.; Burbridge, C.; Alberti, M.; Thippur, A.; Folkesson, J.; Jensfelt, P.; and Hawes, N. 2014. Combining Top-down Spatial Reasoning and Bottom-up Object Class Recognition for Scene Understanding. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'14)*.

Laeng, B. 2013. *Representation of Spatial Relations*. Oxford University Press.

Moratz, R.; Nebel, B.; and Freksa, C. 2003. Qualitative spatial reasoning about relative position: The tradeoff between strong formal properties and successful reasoning about route graphs. In Freksa, C.; Brauer, W.; Habel, C.; and Wender, K. F., eds., *Lecture Notes in Artificial Intelligence 2685: Spatial Cognition III*. Berlin, Heidelberg: Springer Verlag. 385–400.

Ragni, M.; Fangmeier, T.; and Bruessow, S. 2010. Deductive spatial reasoning: From neurological evidence to a cognitive model. In *Proceedings of the 10th International Conference on Cognitive Modeling*, 193 – 198.

Sridhar, M.; Cohn, A. G.; and Hogg, D. C. 2010. Unsupervised learning of event classes from video. In *Association for the Advancement of Artificial Intelligence (AAAI)*.

Sridhar, M.; Cohn, A.; and Hogg, D. 2011a. From Video to RCC8: Exploiting a Distance Based Semantics to Stabilise the Interpretation of Mereotopological Relations. In *Spatial Information Theory - 10th International Conference (COSIT 2011)*, 110 – 125.

Sridhar, M.; Cohn, A. G.; and Hogg, D. C. 2011b. Benchmarking Qualitative Spatial Calculi for Video Activity Analysis. In *Proc. IJCAI Workshop Benchmarks and Applications of Spatial Reasoning*, 15–20.

Stickel, M. E.; Waldinger, R. J.; and Chaudhri, V. K. 2000. A guide to SNARK. Technical report, AI Center, SRI International, Menlo Park, USA.

Tayyub, J.; Tavanai, A.; Gatsoulis, Y.; Cohn, A.; and Hogg, D. 2014. Qualitative and Quantitative Spatio-Temporal Relations in Daily Living Activity Recognition. In *Proc. of the 12th Asian Conference on Computer Vision (ACCV'14)*. Springer.

Thippur, A.; Burbridge, C.; Kunze, L.; Alberti, M.; Folkesson, J.; Jensfelt, P.; and Hawes, N. 2015. A Comparison of Qualitative and Metric Spatial Relation Models for Scene Understanding. In *29th Conference on Artificial Intelligence (AAAI'15)*.

Van de Weghe, N.; Cohn, A.; De Tré, B.; and De Maeyer, P. 2005. A Qualitative Trajectory Calculus as a basis for representing moving objects in Geographical Information Systems. *Control and Cybernetics* 35(1):97–120.

Westphal, M.; Wölfl, S.; and Gantner, Z. 2009. GQR: A fast solver for binary qualitative constraint networks. In *Benchmarking of Qualitative Spatial and Temporal Reasoning Systems, Papers from the 2009 AAAI Spring Symposium, Technical Report SS-09-02, Stanford, California, USA, March 23-25, 2009*, 51–52.

Hole in One: Using Qualitative Reasoning for Solving Hard Physical Puzzle Problems

Xiaoyu Ge
Research School of Computer Science
Australian National University

Jae Hee Lee
QCIS, FEIT
University of Technology Sydney

Jochen Renz and **Peng Zhang**
Research School of Computer Science
Australian National University

Abstract

The capability of determining the right physical action to achieve a given task is essential for AI that interacts with the physical world. The great difficulty in developing this capability has two main causes: (1) the world is continuous and therefore the action space is infinite, (2) due to noisy perception, we do not know the exact physical properties of our environment and therefore cannot precisely simulate the consequences of a physical action.

In this paper we define a realistic physical action selection problem that has many features common to these kind of problems, the minigolf hole-in-one problem: given a two-dimensional minigolf-like obstacle course, a ball and a hole, determine a single shot that hits the ball into the hole. We assume gravity as well as noisy perception of the environment. We present a method that solves this problem similar to how humans are approaching these problems, by using qualitative reasoning and mental simulation, combined with sampling of actions in the real environment and adjusting the internal knowledge based on observing the actual outcome of sampled actions. We evaluate our method using difficult minigolf levels that require the ball to bounce at several objects in order to hit the hole and compare with existing methods.

1 Introduction

One of the grand visions of Artificial Intelligence is to build robots with similar everyday capabilities as humans, who can live among us and assist us with many of our daily tasks. There is a multitude of applications such as caring for the sick, the young or the elderly or household robots that can relief us from many of our daily chores.

In order to progress towards more capable and more human-like robots, we need to develop methods and technology that allow robots to successfully interact with their environment. It requires AI or robots to perceive their environment using their available sensors and to select and to perform physical actions or a sequence of physical actions that achieves a given task. Dealing with physical actions is very hard for a number of reasons:

1. Since the available information about the environment is based on perception, it will most likely be incomplete and imprecise
2. Since the world is continuous, there are typically infinitely many different actions available, each of which could have a different outcome. For example, the exact angle or force that is used to interact with another object determines its behavior.
3. The outcome of a physical action might be unknown before executing it or before accurately simulating it.

Accurately predicting the outcome of physical actions is essential for selecting the right action, but there are potentially infinitely many possible physical actions to consider.

When we humans are faced with a "*physical action selection problem*", i.e., a problem that requires selecting a physical action [1] that achieves the desired goal (out of an infinite number of possible actions), we are very good at coming up with a qualitative solution and with a qualitative prediction of the consequences of an action. A qualitative solution means that we can describe the physical action in words as well as what we expect will happen as a consequence of executing the physical action. Based on these predictions we can describe a physical action or action sequence that could potentially achieve the goal. Whether it does achieve the goal or not, we can only find out once we execute the action.

Physical action selection problems can come in many variants and it is not possible to formalize all of them as a single meaningful problem that covers all cases. We have therefore selected one particular physical action selection problem that is an actual real-world problem and that covers many common aspects of physical action selection problems. We call our problem the "*Hole-in-One*" problem in reference to the problem in mini golf of identifying and executing a shot that sinks the ball with this single shot: Given a static environment of physical objects C (the mini golf "obstacle course"), an object B (the "ball") at start location S, and a target location H (the "hole"), all of which we define more precisely in section 3. Identify the force vector P (the "putt") that, when applied to B at location S, results in B reaching the target H. The idea is that in order to achieve this, the ball needs to bounce at several objects that are part of C in order to reach H with only one shot. But which objects have to be hit and in which order needs to be determined.

[1] An action has several parameters e.g. the exact angle and force of a putt. Actions are different if their parameters have different values

(a) (b)

Figure 1: (a) A real-world mini golf course (b) Illustration of the problem domain in this paper. The goal region H is the green region. The trajectory of B given by an identified solution is shown as the sequence of red dots.

This is a major difference to papers in the literature who study physical action selection of robots, some of which even study minigolf (Zickler and Veloso 2009). Variants of the hole-in-one problem occur frequently, not just in mini golf, in Pachinko, in pool billiard, curling or in a multitude of physics-based video games such as Angry Birds, but also in many everyday situations. These variants can be in 2D or in 3D environments and involve gravity or not.

What these problems have in common is that there are infinitely many possible force vectors. Once a force vector is given and the physical setting, that is the physical properties of all objects and the environment is exactly known, it is possible to compute the exact trajectory of the ball and to see if that force vector solves the problem. However, the task we need to solve is the inverse problem and therefore much harder. We have to identify a force vector out of infinitely many possibilities that solves the problem. While a geometrical or analytical solution of these problems is typically not possible if the obstacle "course" is non-trivial, humans are very successful in solving these kind of problems. We also very much enjoy solving these problems, which is demonstrated by the fact that they often occur in a game-like setting. These problems become even harder to solve when we consider that we usually do not know the exact physical setting. We often only know what we can see and our perception is thus the limiting factor in what we know about the physical setting. Because of the uncertainty about the physical environment, every potential solution to the problem needs to be executed in the actual environment before we can be sure that it is a solution. If it is no solution, we need to find ways of adjusting it so that it will eventually lead to a solution.

In this paper we propose to solve this problem similar to how humans are believed to be doing it: by a combination of qualitative reasoning and mental simulation as well as through a repeated process of limited sampling in the actual environment, observation of the consequences and adjusting our mental simulation. The method can solve even very complicate instances of the hole-in-one problem, which we demonstrate in Section 5.

2 Background

There are two key research streams in reasoning about physical systems, namely qualitative physical reasoning (Davis 2008) and simulation-based reasoning (Battaglia *et al.* 2013). One goal of qualitative physics is to formalise the commonsense knowledge (Kuipers 1989) about real-world physics and solve physical reasoning problem within the framework. The main advantage of qualitative physical reasoning is that it can rapidly draw useful conclusions from incomplete information (Davis *et al.* 2013). However, these approaches are often specific to a narrow domain and there have been very few implementations of these theories. The most relevant work in this field is (Forbus 1981) which proposed a framework for reasoning about the motion of a 2D ball by qualitative simulation. The rules used for state transitions are based on qualitative process theory (Forbus 1984). While most of the previous work focuses on representing physical systems and describing (or predicting) consequences of actions, our method is solving a considerably harder problem as it has to find applicable actions from the infinite action space.

Simulation-based reasoning was inspired by findings in cognitive psychology that mental simulation may play a pivot role in intuitive physical reasoning (Craik 1967). Mental simulation is viewed as a probabilistic simulation in which inferences can be drawn over the principles of classical mechanics. The prediction power drops off drastically when the model is inaccurate or the observation is radically incomplete. (Davis and Marcus 2015) discussed the limitations of the simulation-based reasoning methods. (Smith *et al.* 2013) analysed how humans make physical predictions about the destination of a moving object in the simulated environment where a ball is moving on a bumper table.

In robotics, there has been extensive research on motion planning (Kumar and Chakravorty 2012) or manipulation planning (Dogar 2013) For example, (Westphal *et al.* 2011) uses qualitative reasoning for generating manipulation plans. It models the spatial layout of objects using a spatial constraint network. The plan is found when there is a consistent constraint network under the goal constraints. (Kunze and Beetz 2015) combines a qualitative reasoning method and physics simulations to envision possible effects of actions when making a pancake. The actions and plans of making a pancake seem hard-coded. Our problem is different from most manipulation or motion planning problems where robots can actively adjust their motion path while executing it. In our case, we need to identify a single impulse that solves the given task and no further adjustments are possible. Our task is not easier compared with those complex robotic tasks. For example, a recent paper (Wolter and Kirsch 2015) developed a framework aiming at combining learning and planing and employing qualitative reasoning and linear temporal logic. They used their framework to solve a "ball throwing" problem which is a simplified version of our problem. Their problem is to throw an object to hit a fixed goal location. Their approach does not plan the path, instead, it only looks at the final result (e.g., is the ball left or right of the goal) and adjusts the initial action accordingly. There has been some work on teaching robots to play mini golf. (Khansari-Zadeh *et al.* 2012) proposed mathematical models for learning control parameters of hitting motions while they do not focus on solving the planning problem. (Zickler and Veloso 2009) proposed a framework for physics-based planning that integrates high level behavior planning and lower level motion planning. The method uses random-based sampling to find solutions. In the evaluation, we will show a comparison between our method and a random-based sampling method.

3 Modeling the Physical Environment

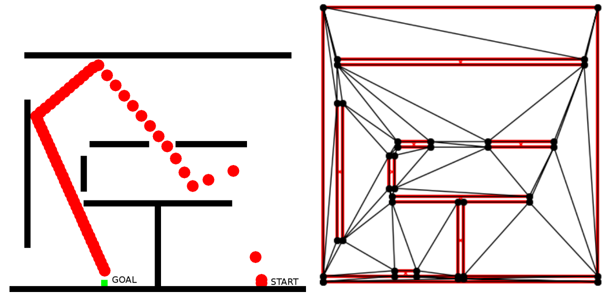
In this paper, we choose the following idealisation of the physical environment, which is often used in physics puzzle games such as mini golf:

- The environment is a restricted 2D plane.
- Objects are 2D rigid bodies with polygonal or circular shape.
- There may be a uniform downward gravitational force.
- Object movements and interactions obey Newtonian physics.
- Physics parameters of objects and the environment remain constant.

We call this environment PHYS2D. An instance of PHYS2D, or a *scenario* is a tuple $\langle E, \mathbf{O}, \mathbf{P}, \mathbf{T}\rangle$, where E is the restricted plane where the objects are located, $\mathbf{O}$ a finite set of static rigid objects O, each of which has a shape, a location, an angle and a type, $\mathbf{P}$ is a set of physics parameters that hold in the environment, such as gravity, and $\mathbf{T}$ is a set of object types and their respective physics properties such as mass and friction, or whether the object can move after being hit or remains static. We assume that all objects are initially static and stable under gravity.

Given such a scenario we can now apply physical actions to it and define physical reasoning problems and tasks. A *physical action* (short: *action*) can be applied to an object O by exerting an impulse at a certain position p of the exterior boundary of O (denoted as ∂O). An *impulse* is defined as a pair (θ, μ) where $\theta \in [0, 2\pi)$ is the direction and $\mu \in [0, \mu_{\max}]$ the magnitude of the impact. $\mu_{\max}$ is the maximal magnitude allowed in the environment, both θ and μ are continuous, therefore the number of possible actions is infinite. An action, i.e., a triple $\langle p, \theta, \mu\rangle$ applied to O can bring O into motion and, as an effect, can cause a chain of new actions on other objects. We call the initial action a *direct* action and the resulting new actions *indirect* actions.

While there are many possible problems that can be defined in this environment, the problem we want to solve is to identify a single action applied to a specified *start object*, such that the action results in a specified *goal region* to be hit by at least one of the objects. This problem is quite general in the sense that it can be applied to various physical games, such as computational pool (Archibald *et al.* 2010), Angry Birds (Renz 2015) and digital curling (Yamamoto *et al.* 2015). These games can have several objects that could be used as start objects or have several goal regions (e.g. holes in pool or pigs in Angry Birds) that the agent has to reach or destroy.In this paper we assume there is only *one* start object and *one* goal region. Despite this restriction, the problem is very challenging as there are many ways to use the chain of indirect actions on intermediate objects to reach the goal region after acting upon the start objects (see Example 2). Furthermore there are infinitely many possible actions, each of which might have a different outcome that needs to be determined in advance. We call this physical action selection problem the Hole-in-One problem.



(a) (b)

Figure 2: (a) Scenario MINI7 (b) Triangulation of Mini7

Definition 1. (Hole-in-One)
Instance: An instance of the physical action selection problem Hole-in-One is a tuple $\langle E, \mathbf{O}, \mathbf{P}, \mathbf{T}, B, H\rangle$, where we use a scenario of PHYS2D and determine a ball $B \in \mathbf{O}$ as the start object and H as the target hole, a polygon in E with a given location.
Solution: A *solution* is an action or *putt* $P = \langle p, \theta, \mu\rangle \in E \times [0, 2\pi) \times [0, \mu_{\max}]$ applied to object B such that B is delivered to the hole H as a consequence of the putt. To simplify the problem, we fix p to be the centroid of B. Formally, the problem can be described as finding a putt P such that the forward model f w.r.t. scenario $\langle E, \mathbf{O}, \mathbf{P}, \mathbf{T}\rangle$ induces a continuous function $R : [0, 1] \to E$, called the route R of ball B given P, such that $R(1) \in H$.

As the forward model f is not explicitly given and the search space is infinite, it is difficult to devise a systematic method for Hole-in-One and to analyze its complexity.

Example 2. In Figure 2a the start object B is the red object. The goal region H is the green region. Here, the solution is to make the start object bounce several times to reach the region. Since the action space to hit B is infinite, intelligent search is required to solve the problem.

What we have described is the actual physical environment we are dealing with. Solving physical action selection problems in this continuous environment is hard, but the problems we are facing are even harder as we do not have complete information about this environment. We only know what we can perceive and perception is typically noisy. The method we develop in this paper aims at solving these physical action selection problems under noisy perception, which requires an extended approach compared to having perfect information.

4 A Method for Identifying Physical Actions

The Hole-in-One problem distinguishes itself from common AI planning problems in that its search space is *infinite* and in particular the action space is *continuous*: Infinitely many different actions can allow an object to take infinitely many different paths. We propose the following method to solve this problem:

- As input scenario, we take the information about the physical environment that we obtained through potentially noisy perception.
- We first partition the free space of the given scenario into finitely many triangular zones (Figure. 2b).
- We defined qualitative rules that describe the physics that govern the transition of moving objects between the triangular zones. We use these rules to generate sequences of qualitative transitions between zones that coincide with potential real paths a moving object can take to achieve the goal. We call such a sequence a *qualitative path*.
- Once all qualitative paths are determined, we rank them by their likelihood of being realized, before we try to realize them (see Section 4).
- We now use a physics simulator based on our perceived input scenario (that is the simulator does not know the real environment, denoted as Sim_I) to search for actions that realize the qualitative paths in their ranked order, i.e., actions that allow objects to follow the qualitative paths.
- The solutions we obtain here are not necessarily solutions in the real environment. Therefore, whenever we obtain a solution in Sim_I, we immediately apply the solution to the real environment. If it does not lead to a real solution, we adjust the object information in Sim_I according to the observation made in the real environment before we continue with the previous step. We will not adjust the triangulation or the qualitative paths when we adjust objects in Sim_I.

We now describe these points in more detail.

Qualitative Rules for State Transitions

Given a instance, we triangulate the free zone (i.e., the space not occupied with objects) of the scene using constrained Delaunay triangulation (Shewchuk 1996) where the object boundaries and the boundaries of E are part of the triangulation. We then generate *qualitative paths* based on the triangulation, that are sequences of qualitative states Q that represent how objects can move through the triangulation.

Definition 3. The *qualitative state* $\langle \triangle, e, \Theta, \mathtt{mt} \rangle$ of the ball B is defined by

- $\triangle$: the triangular zone where the ball is located;
- e: the edge of $\triangle$ the ball crossed to enter $\triangle$;
- Θ: possible direction range when entering $\triangle$ via e;
- mt: the motion type of the ball. We distinguish three types of motions: FLY, BOUNCE, and SLIDE.

We obtain a qualitative path by expanding a qualitative state Q_i subsequently to a set of its next possible states $\mathbf{Q_i}$. Qualitative paths form a tree of states with branching factor at most 6, two outgoing edges with three possible motion types.

The procedure for expanding a qualitative state $Q_1 = \langle \triangle_1, e_1, \Theta_1, \mathtt{mt}_1 \rangle$ to a state $Q_2 = \langle \triangle_2, e_2, \Theta_2, \mathtt{mt}_2 \rangle \in \mathbf{Q_1}$ is as follows:

1. For each edge e_2 of triangle $\triangle_1$ with $e_2 \neq e_1$ we determine whether the current direction range Θ_1 allows the ball to move from e_1 to e_2.
2. If possible, we choose a motion type $\mathtt{mt}_2$ for the next state according to a set of physical rules (see below).
3. We also set the zone of he next state to $\triangle_2$ and set e_2 and determine Θ_2 according to our rules.

In the following, we describe the rules that govern state transitions between different motion types. Note that we always write "ball" to denote the moving object, as that is the moving object we use in the Hole-in-One problem. But the rules equally apply to other moving objects, not just to balls.

Rule 1 (FLY $\to$ FLY)**:** This rule is triggered when the current motion type is FLY and e_2 is an edge between $\triangle_1$ and $\triangle_2$, as the ball can keep flying after it entered $\triangle_2$ through e_2. Since the ball has entered $\triangle_1$ through e_1, the range of directions that the ball can fly from e_1 to e_2 is limited. Let v be the vertex of triangle $\triangle_1$ that is shared by e_1 and e_2, and let v_1 and v_2 be the remaining other vertices of e_1 and e_2, respectively. To compute Θ_2, we set first Θ_{Ran} as the range between the direction from v_1 to v and the direction from v to v_2. Then we can ensure that only free flying point-like objects with direction $\theta \in \Theta_{\text{Ran}}$ can fly from e_1 to e_2. To compute Θ_2 we also take the effect of gravity into consideration and apply gravity to the current direction Θ_1, obtaining a new range $\Theta_{\mathtt{Gra}}$. Then the method computes Θ_2 by intersecting $\Theta_{\mathtt{Gra}}$ with Θ_{Ran}. The next state $\langle \triangle_2, e_2, \Theta_2, \mathtt{FLY} \rangle$ will be created if $\Theta_2 = \Theta_{\mathtt{Gra}} \cap \Theta_{\text{Ran}} \neq \emptyset$. •

The following two rules are triggered when $\mathtt{mt}_1 = \mathtt{FLY}$ and e_2 is a surface of an object. In this case the ball flying to e_2 can bounce. We assume it will be bounced back in the reflection direction range with respect to the outward normal of the surface. To this end, we determine Θ_2 in the same way as described in the (FLY $\to$ FLY) rule.

Rule 2 (FLY $\to$ BOUNCE $\to$ FLY)**:** This rule will always be applied when $\Theta_2 \neq \emptyset$. The method computes the range $\Theta_\uparrow$ of bouncing directions using Θ_2 and the normal vector of the surface. and adds $\langle \triangle_1, e_2, \Theta_\uparrow, \mathtt{FLY} \rangle$ to $\mathbf{Q_1}$. •

Rule 3 (FLY $\to$ BOUNCE $\to$ SLIDE)**:** In addition, there is the possibility that the ball *slides* on e_2 when the gravity force is towards the e_2, which assures that the ball can receive support from the surface, which we call the SLIDE condition. If the condition hold, we add $\langle \triangle_2, e_2, \Theta_2, \mathtt{SLIDE} \rangle$ to $\mathbf{Q_1}$. •

The following three rules are triggered when $\mathtt{mt}_1 = \mathtt{SLIDE}$. We assume that once the ball enters into the SLIDE motion, it will keep sliding until it hits a "wall" or until the surface cannot support it anymore. Let e_1 be the surface on which the ball is sliding, and e_2 be the surface connected to e_1 through their common vertex.

Rule 4 (SLIDE $\to$ SLIDE)**:** We add state $\langle \triangle_2, e_2, \Theta_2, \mathtt{SLIDE} \rangle$ to $\mathbf{Q_1}$, where Θ_2 is the direction from the common vertex to the other vertex of e_2. $\triangle_2$ is the triangular zone to which e_2 belongs. •

Rule 5 (SLIDE $\to$ BOUNCE $\to$ SLIDE)**:** This rule applies if e_2 forms a "wall" in front of the current direction. Specifically, when Θ_2 as defined in Rule 3 is between the surface e_1 and the surface normal, then the ball will bounce back and slide on e_1 in the opposite direction. •

Rule 6 ($\texttt{SLIDE} \rightarrow \texttt{FLY}$)**:** If e_2 can neither give any support to the ball nor allow the bounce, the ball will start to fly from the end of e_1. Hence, we modify the current state by changing the motion type from SLIDE to FLY and apply the $\texttt{FLY} \rightarrow \texttt{FLY}$ rule to infer the next states. •

So far we have only considered changes to the state of a moving ball, but whenever a moving ball bounces off another object, the other object can start moving too (provided it is a movable object). While this is not allowed in the Hole-in-One problem where all objects remain static, we still add this possibility for the sake of generality. This is covered by the following rules.

Rule 7 (Hitting Movable Objects)**:** For each edge e that belongs to a movable object we generate a state $\langle \triangle, e, \Theta, \texttt{FLY} \rangle$ where $\triangle$ is the triangle in the free space that also has e as its edge. This triangle is uniquely determined, as edge e represents a surface of an object that is shared by only one triangle in the free space. Θ will be the same direction range as the direction range Θ_1 of the object that hit the movable object. Should the SLIDE condition (see Rule. 3) apply, we will also add a state with motion type SLIDE. We then continue expanding each of the generated qualitative states using the other applicable rules. •

Note that once an object starts moving, we would have to adjust the triangulation as the free space changes. However, in this paper we do not explicitly handle more complicated cases where moving objects repeatedly interact. Instead, we assume here that these cases are covered by the simulator when trying to realize paths and leave more detailed rules capturing this to future work should it be necessary.

Generating Qualitative Paths

To derive qualitative paths from the ball B to hole H, we starts with all possible initial states. An initial state $\langle \triangle_{\texttt{ini}}, e_{\texttt{ini}}, \Theta_{\texttt{ini}}, \texttt{mt}_{\texttt{ini}} \rangle$ of B is given as follows:

- $\triangle_{\texttt{ini}}$: the triangular zone containing the centroid of B;
- $e_{\texttt{ini}}$: the surface on which B is located;
- $\Theta_{\texttt{ini}}$: possible direction range to reach the next edge.
- $\texttt{mt}_{\texttt{ini}}$: there will be at most four different initial states as B can SLIDE on $e_{\texttt{ini}}$ in two different directions or FLY to each of the other edges of $\triangle_{\texttt{ini}}$.

We expand each initial state by successively applying all applicable rules to determine possible successor states. We stop expanding a state when it reaches the goal state $\langle \triangle_H, e_H, \Theta_H, \texttt{mt}_H \rangle$ where e_H is a surface of $\triangle_H$ that contains hole H. A qualitative path is a sequence of states from an initial state to a goal state where the successor state of each state is obtained by using our qualitative rules. We record which rule is applied at which state in order to rank qualitative paths.

Any qualitative path that does not lead to the goal state will be deleted. We use the following rules to ensure that,

Rule 8 (Do not move through a smaller edge)**:** Whenever an object O is required to pass through an edge e that is smaller than O itself, we remove any qualitative path that does not include a bounce transition at e. •

Rule 9 (Limit the number of states)**:** If a state is exactly the same as a previous state (including the same direction range Θ), we delete that qualitative path as it may generate infinite cycles. If a qualitative path reaches a preset maximum of states without reaching the goal state, we also delete it. •

Ranking Qualitative Paths

Before trying to physically realize the different qualitative paths, we will rank them according to the likelihood of leading to a solution. The idea is to take into account the magnitude of an impulse required to realize a qualitative path. If a qualitative path is too long or involves many bounces, which reduces the speed of the moving object, then the path will be less likely to be realized. Therefore, we assume that two factors play a role in ranking qualitative paths: the actual length of a path and the number of bounces along the path.

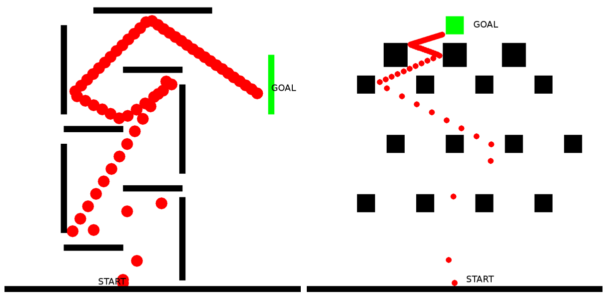
Let $be_0, be_1, \dots be_n$ be the sequence of edges, where be_0 is the initial edge from which B is launched, $be_1, \dots, be_{n-1}$ are all surfaces where bouncing takes place, and be_n is the entering edge of the goal state. Then the cost of a qualitative path is given by

$$\sum_{i=0}^{n-1} d(be_i, be_{i+1}) \cdot (1+\gamma)^i, \quad (1)$$

where $d(be_i, be_{i+1})$ is the Euclidean distance between two edges and $(1+\gamma)$ with $\gamma \in (0,1)$ is a penalty term. The penalty term penalizes the part of a qualitative path that happens after each bouncing. Therefore, given two paths of similar lengths, a path with less bounces will be preferred to a path with more bounces. γ can be set to a smaller value when the ball does not lose much kinetic energy after each bouncing, and vice versa.

Realizing Qualitative Paths

In this section we describe how to use these qualitative solutions to reduce the search space of finding a real solution. Recall that a solution is a physical action, i.e., an impulse $\texttt{imp} := \langle \theta, \mu \rangle$, on the centroid of ball B that delivers it to hole H. The idea is as follows: for each qualitative path, we sample possible actions using Sim_I within a range $\Theta_{\texttt{imp}}$ of directions that can potentially realize the path. We cluster qualitative paths that share similar direction ranges $\Theta_{\texttt{imp}}$ and sample only within these shared ranges. (We mainly focus on the direction parameter θ, as it has a larger effect on the solution. For the magnitude μ of the impulse, we always sample within its full range.) We also rank clusters by their average costs based on the cost function (1). If we do not find any action that can realize any qualitative path after going through all the clusters, we discard less promising clusters and restart sampling. A cluster is identified as less promising when none of the sampled actions has achieved the different bounces required to follow a qualitative path.



(a) (b)

Figure 3: Mini golf scenarios usually require several bounces. (a) Scenario MINI6 and (b) Scenario MINI2 with identified solutions

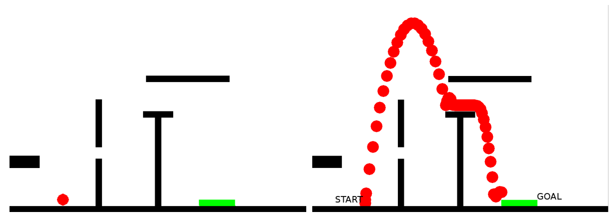
Testing Potential Solutions in the Real Environment

The visual input to the internal simulation can be noisy with imperfect perception. Therefore, applying an action in the real environment may have a different outcome from that predicted by Sim_I. To deal with this, we progressively adjust the internal simulation whenever a proposed action does not lead to a solution in the real environment. This is done to minimize the difference between the outcomes. The outcome of an action is represented by the sequence of triangular zones of the qualitative path generated by that action. This qualitative representation is less sensitive to visual noise compared with using trajectory points which are highly affected by noise. We use the Levenshtein distance (Levenshtein 1966) to quantify the similarities between two sequences. Once our method found an action that solves the problem in the internal simulation, it will execute the action in the real environment and observes the trajectory of the moving object. It then obtains the sequence of triangular zones from the observed trajectory and compares the sequence with its counterpart in the internal simulation. If the edit distance between them is greater than a threshold ϵ, the method will adjust the spatial configuration of relevant objects to minimize the distance. The relevant objects are the objects with which the moving object collided in the both internal simulation and real environment.

5 Evaluation

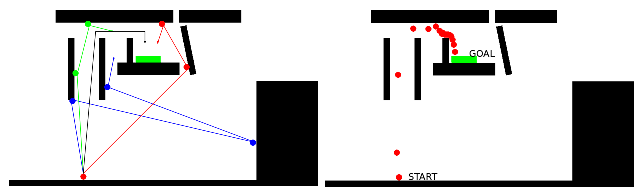
We simulated the real environment using the Box2D (www.box2d.org) physics engine. The method also uses Box2D for its internal simulator Sim_I with an incomplete knowledge of the real environment. We generated scenarios that allow us to evaluate different aspects of our proposed method. A scenario contains a set of movable and immovable objects. Objects have three physical properties, namely, density, friction, restitution. The goal region H is specified as rectangular region that is initially away from any movable object. An action is performed by exerting an impulse $(\mu, \theta), \mu \in I_\mu, \theta \in I_\theta$ on the centroid of B with $I_\mu = (0, 5000]$ and $I_\theta = [0, 2\pi)$. A problem is solved when H is contacted by B.

We first tested if our method can find different qualitative paths. In Figure. 4a, one possible solution is to let B hit



(a) (b)

Figure 4: (a) Scenario S_1 (b) An identified solution to S_1



(a) (b)

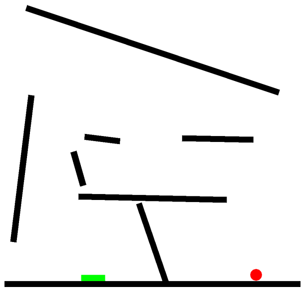
Figure 5: (a) Groups of qualitative paths detected in S_2 (b) An identified solution to S_2

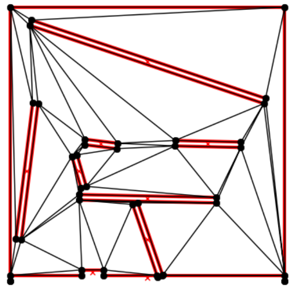
the platform and fall to the green goal region. To realize this path, one has to know how an object files under gravity. To test the effectiveness of our clustering and ranking strategy, we designed some scenarios that have many qualitative paths that may lead to solve the problem. For example, in S2 (Figure 5a), our method detected 595 qualitative paths and divided them into 14 groups by their $\Theta_{\texttt{imp}}$. The figure illustrates four interesting groups of qualitative paths. Each colored arrow represents a group of qualitative paths, and shows the rough direction B takes. The four groups of paths were ranked in descending order of average costs as: black, green, red, blue. The blue group is ranked last because it takes several long distance bounces. The black group is ranked first because it suggests to hit the goal region via a direct trajectory. However, the black group was identified as an un-realizable group after a few sampled actions in the real environment. We further created several mini-golf scenarios. The scenario designs are inspired by the game levels of a popular video game of mini-golf[2]. Unlike S_1 and S_2, there is no gravity in these mini-golf scenarios. The scenarios usually require more than 5 bounces to solve (e.g. see Figure. 3).

Dealing with Noisy Information To test the effectiveness of our method with imperfect perception, we perturb the visual input of a scenario by rotating each object at an angle θ in radians. The angle is sampled from a zero mean Gaussian with a variance $\sigma \in \{0.1, 0.2, 0.3\}$. The method uses the perturbed visual input for its internal simulation and keeps adjusting the angle of objects using the method described in Section 4. Figure 7 shows a perturbed mini-golf scenario in Figure 2a with $\sigma = 0.3$. It is clear to see that even a small rotation will substantially distort the scenario.

We use each designed scenario as a template to automat-

[2]http://www.eivaagames.com/games/mini-golf-game-3d/





(a) (b)

Figure 6: Perturbed scenario MINI7 with $\sigma = 0.3$ and its corresponding triangulation

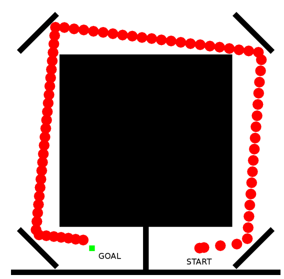
ically generate scenarios for testing. Given a template scenario, we randomly vary the spatial configuration of the objects. In the end, we obtained 72 levels. The proposed method uses penalty term $\gamma = 0.7$. At each round, the method samples 200 impulses with $\theta \in \Theta_{\texttt{imp}}, \mu \in [0, 5000]$ for each group of the identified qualitative paths.

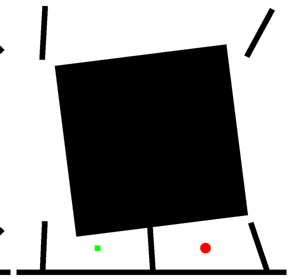
We compare our method M_1 with a solver M_2 which uses a goal-oriented sampling strategy. The sampling strategy of M_2 is similar to the one used in (Wolter and Kirsch 2015) that adjusts actions according to the distance between the final position of the ball and the target destination. Specifically, M_2 evaluates an action using the minimum distance D between the trajectory of any movable object and the goal region. The goal is to find an action with $D = 0$, which solves the problem. M_2 obtained trajectory points directly from the actual environment (simulator), which are noise-free. Given a problem, M_2 performs several rounds of searching. At each round, M_2 makes several random trial actions in the actual environment and selects the action that yields the minimum D_1. It then performs a local search around the action, and picks the sample with the minimum $D_2 < D_1$ and repeat the procedure until it finds a solution or reaches a cut-off threshold. The evaluation result is summarized in Table 1. To give an indication of how much our method can reduce the sampling space, we show the proportion (AR) of the direction range of all qualitative paths to the entire range $[0, 2\pi)$. The actual solution space can be even smaller.

Summary of the Evaluation Our method outperforms M_2 in all the scenarios. M_2 is less efficient because there could be many local optima in a problem. By contrast, our method tries to realize each group of qualitative paths, which helps to avoid these local optima. Consequently, it can detect more different types of solutions (if there are any). It can still effectively find solutions when the solutions are far apart and potentially disconnected in the solution space. Qualitative reasoning and triangulation can be achieved efficiently; it takes on average 4 seconds to generate qualitative paths based on a triangulation with around 60 zones. The reason why no solution was found for all the ten MINI5T levels is that the solution space of MINI5 is very small (see Figure 7a); Varying positions of any object (especially the middle black square) will eliminate these solutions.

Table 1: Results on the generated scenarios with imperfect perception. SN: The number of actions made in the actual environment until the first solution is found. *MINI1T: the scenarios created based on MINI1. The average of AR and respective SN are shown for those entries.

Scenarios	AR	SN(M1) $\sigma = 0.1$	SN(M1) $\sigma = 0.2$	SN(M1) $\sigma = 0.3$	SN(M2)
S1	0.05	421	621	829	4229
S2	0.05	375	837	837	>10000
MINI1	0.07	320	223	407	8529
MINI1T	0.05	371	332	319	426
MINI2	0.03	216	230	288	1736
MINI2T	0.04	32	23	36	280
MINI3	0.03	7	34	34	370
MINI3T	0.02	3	12	7	223
MINI4	0.02	133	509	967	987
MINI4T	0.04	42	93	172	254
MINI5	0.04	28	56	31	537
MINI5T	N/A	>10000	>10000	>10000	>10000
MINI6	0.04	68	199	208	2932
MINI6T	0.04	32	239	757	529
MINI7	0.05	41	77	236	3208
MINI7T	0.03	75	236	852	706





(a) (b)

Figure 7: (a) An identified solution to MINI5 (b) A randomly generated scenario based on MINI5 using $\sigma = 0.3$

As the noise level increases, our method can still detect and realize qualitative paths that lead to the goal. Such qualitative paths have similar bounce sequences as their counterparts derived from perfect triangulation. However, it takes on average longer to detect a solution than with accurate perception. Because as the noise increases, accuracy of the triangulation is getting worse and we generate more unrealisable qualitative paths. There will be triangular zones where there is supposed to be a surface while actually not or vice versa. These information can only be potentially adjusted after making several trial shots in the actual environment. Inaccurate perception also affects the ranking of qualitative paths.

6 Discussion

The Hole-in-One problem is just one example of a physical action selection problem, where a physical action has to be determined that achieves a given goal in a physical environment. The difficulty of these problems lies in the fact that the action space is infinite Another source of difficulty is that the sequence of interactions with other objects as well as the required number of interactions is unknown. The problem becomes even harder when the exact physical properties of objects and their locations are not exactly known.

Despite having considered only one example problem, the method we developed to solve the Hole-in-One problem for 2D environments, under gravity and noisy perception is more general. Other physical action problems in 2D environments with different optimization criteria can be solved in a similar way, by triangulating the free space and by determining qualitative paths between the triangles that obey the general physics rules we defined. Our approach of sampling proposed solutions in the real environment and adjusting our internal knowledge through observations can be used for other physical action problems.

While some of the techniques we use are not new, clearly we have not invented triangulation and also sampling based adjustment has been done before, what is novel in our work is that we can solve arbitrary instances of a complex physical action selection problem without hardcoding or predefining actions or action sequences. We only define standard physics rules that determine what changes can happen to an object when it interacts with other objects or moves through empty space. We can do this under gravity and under noisy perception, and we do it in a similar way to how humans supposedly solve these kind of problems (Trial and error). One possible extension of our work is to consider 3D environments, where we could partition the free space into zones similar to how we do it in 2D. Another possible extension is to lift the restriction that all objects other than the ball are static.

7 Conclusion

We studied a realistic problem that contains some of the essential components AI needs to successfully interact with the real world: being able to predict the consequences of physical actions and to select a physical action out of an infinite number of actions that achieves a specified goal. This problem becomes even harder with noisy perception. The proposed method involves a combination of qualitative reasoning and internal simulation together with testing proposed actions in the real environment and, if necessary, adjusting our internal knowledge based on the new observations. While it is not our intention to build a robot that becomes the new minigolf world champion, we have seen in our experiments that our approach is able to identify some remarkable shots involving several bounces before achieving a hole in one. We are rather interested in coming closer to being able to solve physical action selection problems in general, particularly in noisy environments. A solution to this problem will have a major impact on AI and we believe that our approach forms a good starting point to achieving that. As a side note, we just read about golf playing robot LDRIC that managed to score a hole in one. But of course (still somehow surprisingly) a hole in one in golf seems to be easier to achieve than a hole in one in a difficult minigolf level that involves several physical interactions with other objects.

References

Christopher Archibald, Alon Altman, Michael Greenspan, and Yoav Shoham. Computational Pool: A New Challenge for Game Theory Pragmatics. *AI Magazine*, 31(4):33–41, December 2010.

Peter W Battaglia, Jessica B Hamrick, and Joshua B Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332, 2013.

Kenneth James Williams Craik. *The nature of explanation*. CUP Archive, 1967.

Ernest Davis and Gary Marcus. The scope and limits of simulation in cognitive models. *arXiv preprint arXiv:1506.04956*, 2015.

Ernest Davis, Gary Marcus, and Angelica Chen. Reasoning from radically incomplete information: The case of containers. In *Proceedings of the Second Annual Conference on Advances in Cognitive Systems ACS*, volume 273, page 288, 2013.

Ernest Davis. Physical reasoning. *Handbook of Knowledge Representation*, page 597, 2008.

Mehmet R Dogar. *Physics-Based Manipulation Planning in Cluttered Human Environments*. PhD thesis, Massachusetts Institute of Technology, 2013.

Kenneth D. Forbus. A study of qualitative and geometric knowledge in reasoning about motion. Technical report, Cambridge, MA, USA, 1981.

Kenneth D Forbus. Qualitative process theory. *Artificial intelligence*, 24(1):85–168, 1984.

Seyed Mohammad Khansari-Zadeh, Klas Kronander, and Aude Billard. Learning to play minigolf: A dynamical system-based approach. *Advanced Robotics*, 26(17):1967–1993, 2012.

Benjamin Kuipers. Qualitative reasoning: modeling and simulation with incomplete knowledge. *Automatica*, 25(4):571–585, 1989.

Sandip Kumar and Suman Chakravorty. Adaptive sampling for generalized probabilistic roadmaps. *Journal of Control Theory and Applications*, 10(1):1–10, 2012.

Lars Kunze and Michael Beetz. Envisioning the qualitative effects of robot manipulation actions using simulation-based projections. *Artificial Intelligence*, 2015.

VI Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet Physics Doklady*, volume 10, pages 707–710, 1966.

Jochen Renz. AIBIRDS: The Angry Birds Artificial Intelligence Competition. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, March 2015.

Jonathan Richard Shewchuk. Triangle: Engineering a 2d quality mesh generator and delaunay triangulator. In *Applied computational geometry towards geometric engineering*, pages 203–222. Springer, 1996.

K Smith, Eyal Dechter, J Tenenbaum, and Edward Vul. Physical predictions over time. In *Proceedings of the 35th annual meeting of the cognitive science society*, pages 1–6, 2013.

Matthias Westphal, Christian Dornhege, Stefan Wölfl, Marc Gissler, and Bernhard Nebel. Guiding the generation of manipulation plans by qualitative spatial reasoning. *Spatial Cognition & Computation*, 11(1):75–102, 2011.

Diedrich Wolter and Alexandra Kirsch. Leveraging qualitative reasoning to learning manipulation tasks. *Robotics*, 4(3):253–283, 2015.

M. Yamamoto, S. Kato, and H. Iizuka. Digital curling strategy based on game tree search. In *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 474–480, August 2015.

Stefan Zickler and Manuela Veloso. Efficient physics-based planning: sampling search via non-deterministic tactics and skills. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 27–33. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

Preparing MILA−S for College

Ashok K. Goel
Design & Intelligence Laboratory
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30308
goel@cc.gatech.edu

David A. Joyner
Design & Intelligence Laboratory
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30308
david.joyner@gatech.edu

Taylor J. Hartman
Design & Intelligence Laboratory
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30308
tjhartman@gatech.edu

ABSTRACT

Scientists use both conceptual and simulation models to make sense of the world. MILA−S is an interactive system for authoring conceptual models of ecological phenomena and spawning agent-based simulations of the ecological systems directly from models. We have used MILA−S in middle school science to foster learning about both ecological systems and scientific modeling. We now seek to use MILA−S to promote learning about ecological systems and scientific modeling in college-level introductory biology classes. Compared to middle school students, college-level students typically study more complex ecological systems. In this paper, we present extensions and enhancements to MILA−S in preparation for deployment in college.

INTRODUCTION

Much cognitive systems research on science education seeks to introduce authentic practices of real scientists into science classrooms (Edelson, Gordon & Pea 1999). Scientists in general make sense of the world through cycles of model construction, use, evaluation and revision (Clement 2008; Darden 1989; Halloun 2000; Nersessian 1989, 2008; Schwarz et al. 2009). Further, scientists use many kinds of models to generate, specify, share, test and critique their ideas [Carruthers et al. 2002]. Two of the techniques scientists commonly use are construction of conceptual models and execution of simulation models of the phenomenon or system of interest (Clement 2008; Nersessian 2008). Conceptual models are abstract representations of the components, relations, and processes of the phenomenon (Clement 2008; Darden 1989; Nersessian 1989, 2008; Novak 2000; White & Fredriksen 1990). A conceptual model specifies a scientist's current understanding of a phenomenon and evidence for the understanding, allowing externalization, sharing and critiquing of that understanding, as well as use of the model to guide further investigation Like conceptual models, simulation models too specify the scientists' current understanding of the system and guide further investigation. Simulation models are executable with specific values for the system's input variables, enabling determination of the temporal evolution of the values of the system's output variables (Clement 2008; de Jong & van Joolingen 1998; Jackson, Krajcik & Soloway 2000; Nersessian 2008; White & Fredriksen 1990).

MILA−S is an interactive technology for authoring conceptual models of ecological phenomena and generating simulations based on the conceptual models, preserving the capacity for rapid revision and knowledge sharing allowed by the conceptual models while extending them to provide the repeated testing and feedback of more precise simulations (Goel & Joyner 2015; Joyner & Goel 2015; Joyner, Goel & Papin 2014). MILA−S uses agent-based simulations (Bonabeau 2002) because the paradigm of agent-based simulation is especially well suited for ecological systems (Grimm et al. 2006). MILA−S uses Component-Mechanism-Phenomena models (Joyner, Majerich & Goel 2013; Joyner, Goel, & Papin 2014) for authoring the conceptual models of ecological phenomena and the NetLogo simulation engine (http://ccl.northwestern.edu/netlogo) for agent-based simulations of ecological systems [Wilensky & Resnick 1999]. MILA−S implements a translator that directly compiles the conceptual models into agent-based simulations.

A pilot study entailed deployment of MILA−S in middle school science classrooms in metro Atlanta, and its use by about 50 students for modeling a local aquatic ecosystem. Preliminary results from the study indicated that the students used MILA–S to engage in the desired cycle of mode; construction, use, evaluation and revision (Joyner, Goel, & Papin 2014). Similar studies have also shown that using the MILA family of tools leads to an improvement in both the quality of the conceptual models of ecological phenomena and understanding of the process of scientific modeling −of ecological systems (Goel & Joyner 2015, Joyner & Goel 2015).

Much cognitive systems research has explored interactive tools for qualitative modeling and qualitative simulation and their use for promoting science education (Bredeweg & Forbus 2003). MILA−S parallels Bredeweg et al.'s (2009) Garp-3 system that allows the user to create first qualitative models of ecological phenomena and then qualitatively simulate them. In contrast to Garp-3, MILA−S uses Component-Mechanism-Phenomena modeling for authoring conceptual models, the off-the-shelf NetLogo engine for running agent-based simulations, and a translator between the two for directly spawning the simulations from the conceptual models.

Figure 1: A conceptual model constructed by a team of 7th grade students using MILA–S.

Given the success of MILA−S for fostering learning of ecological systems as well as scientific modeling in middle school science, we now seek to use MILA−S to promote learning about ecological systems and scientific modeling in college-level introductory biology classes. However, college-level students are cognitively more developed than middle school students, the ecological systems they study are more complex, and they have more prior knowledge of ecological systems and scientific modeling. This raises the question of how to extend and enhance MILA−S to match complexity of systems that they study? In this paper, we summarize MILA−S and its use for learning about ecological systems and scientific modeling, and then describe extensions and enhancements to MILA−S in preparation for its deployment in college.

DESIGN OF MILA−S

MILA (**M**odeling & **I**nquiry **L**earning **A**pplication)_ is a family of interactive tools for supporting student learning about scientific discovery. The core MILA tool enables middle school students to investigate and construct models of complex ecological phenomena. MILA–S also allows students to simulate their conceptual models (Goel & Joyner 2015; Joyner, Goel & Papin 2014).

MILA builds on a line of exploratory learning environments including the Aquarium Construction Toolkit (ACT; Vattam et al. 2011) and the Ecological Modeling Toolkit (EMT; Joyner et al. 2011). ACT and EMT were shown to facilitate significant improvement in students' deep, expert-like understanding of complex ecological systems. For conceptual modeling, ACT used Structure-Behavior-Functions models that were initially developed in AI research on conceptual design of technical systems (Goel 2013; Goel, Rugaber & Vattam 2009). In contrast, EMT used Component-Mechanism-Phenomenon (or CMP) conceptual models that are variants of Structure-Behavior-Function models adapted for modeling natural systems (Joyner et al 2011). Both ACT and EMT used NetLogo simulations as the simulation models (Wilensky & Reisman 2006; Wilensky & Resnick 1999). Like most interactive tools for supporting modeling in science education (vanLehn 2013), both ACT and EMT provided one set of tools for constructing and revising conceptual models and another tool set for generating and using simulations.

Conceptual Models

Components in CMP modeling can be either biotic or abiotic. Each component has a set of variables associated with it, four for biotic components, and one for abiotic components. Biotic components are defined by their population quantity, lifespan, energy level, and likelihood to breed; abiotic components are defined only by their quantity. Figure 1 illustrates a causal model constructed by a team of 7th grade life science students early in their interaction with MILA–S. In this model, there are three components: Sunlight, Oxygen, and "Fishies". The Sunlight and Oxygen are abiotic components, and they have only Amount as a variable that is designated on the node for the component. "Fishies" is a biotic component, and thus has Population, Age, Birth Rate, and Energy as variables; Population is designated on the "Fishies" node itself, while the notations for the other three variables extend downward from the main node.

CMP modeling draws causal relations between the variables associated with the different components. For example, the presence of a chemical like ammonia in the ecosystem that is poisonous to fish may decrease the lifespan of the fish, or it may directly decrease the population of the fish (additional information on the difference between the two is provided later in this paper). MILA−S provides the user with a set of prototypes that describe causal relationships among system variables. The choice among the available prototypes is determined by the variables on either end of the relation and the type or direction of the relation. For example, a relation from the Population of a biotic component to the Amount of an abiotic component, such as that from Fish Population to Oxygen Amount, 'consumes', 'produces', or 'becomes upon death,' etc. Similarly, a relation from an abiotic Amount to a biotic Population could be 'destroys' or 'feeds'. Similar relationship prototypes are available for links between two biotic and two abiotic components. In the model shown in Figure 1, the prototypes chosen are 'consumes' for the relationship between Fish and Oxygen, and 'produces' for the relationship between Sunlight and Oxygen. The direction of the arrow between the variables of two components indicates the direction of causal influence. For example, the arrow from Fish to Oxygen in

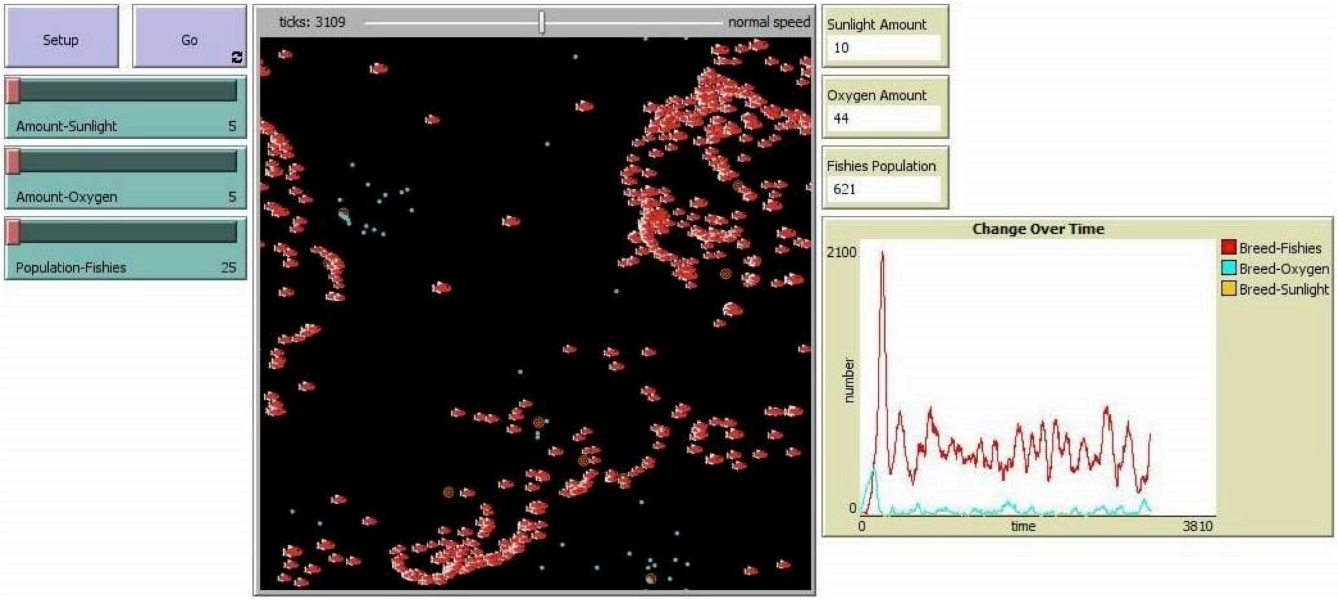


Figure 2: The results of NetLogo simulation of the conceptual model illustrated in Figure 1.

Figure 1 indicates that the Population of Fish influences the Amount of Oxygen.

A Mechanism in CMP modeling is a causal chain of component variables connected by causal relations. For example, Figure 1 illustrates a mechanism hypothesized by a team of students according to which the Amount of Sunlight (an abiotic component) influences the Amount of Oxygen (another abiotic component) and the Population of Fish (a biotic component) also influences the Amount of Oxygen.

A Phenomenon in CMP is an observation about the system of interest. For example, the phenomenon for the mechanism illustrated in Figure 1 is a change in the Amount of Oxygen in an aquatic ecosystem.

A user starts the process of CMP causal modeling using MILA−S with the goal of constructing a causal explanation for explaining a given phenomenon. She then specifies a mechanism as the explanation for the phenomenon, incrementally composing the mechanism from the components of the system, their variables, and the relations between the variables. As Figure 1 illustrates, a CMP model in MILA−S is an external visual representation with textual annotations.

NetLogo Simulations

Figure 2 illustrates the result of the NetLogo simulation for the conceptual model of Figure 1. Note that all three components of the causal model (Figure 1) are represented in the simulation (Figure 2): the Fish are in red, Sunlight hits the water at the location of the brown dots, and the Oxygen produced by that interaction appears as blue dots.

As Figure 2 illustrates, NetLogo depicts the agents in a window showing their actions and behaviors. Also as Figure 2 illustrates, NetLogo provides graphs and counters for illustrating the temporal evolution of various variables of the simulation. Before running a simulation, the user sets the simulation's start condition. The input variables are set through the sliders and toggles on the left side of the simulation window illustrated in Figure 2. The user then clicks the Setup button to apply those changes to a new simulation. The user next clicks the Go button to start the time steps of the simulation.

NetLogo simulations are typically designed with its own dedicated programming language, which allows for enormous flexibility, However, this flexibility of designing simulations makes rapid evaluation and revision of models difficult. First, it requires at least a rudimentary background in programming. Secondly, even if the simulation designer is relatively experienced in NetLogo, it can still take significant time to make non-trivial changes to the way in which the simulation operates: these changes can involve writing all-new methods, creating new variables, or defining new agents. Clearly, it would be useful if the cost of generating NetLogo simulations could be controlled.

MILA–S provides one technique for controlling the cost of generating NetLogo simulations: it automatically generates the simulations from user’s casual model. Note also that the generation of the CMP causal model illustrated in Figure 1 does not require any knowledge of programming. Instead, MILA−S provides a visual syntax for CMP modeling.

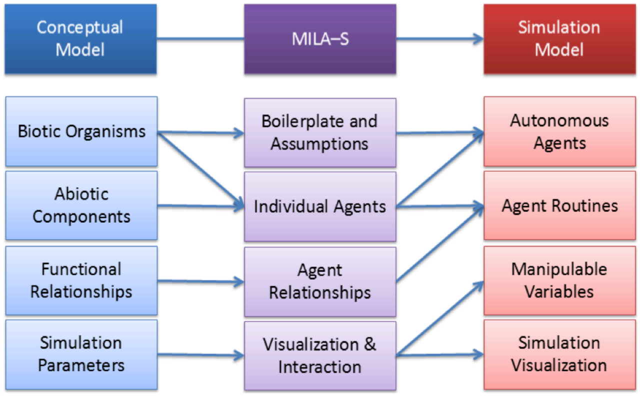


Figure 3: Scheme of translation of CMP conceptual models into NetLogo agent-based simulations.

Translating Conceptual Models into Simulations

After constructing a CMP conceptual model, a student first uses a template to set values of the input variables to the system, and then clicks a 'Run Sim' button for simulation generation. MILA–S iterates through some initial boilerplate settings, then gathers together all the components for initialization along with their individual parameters. After this, MILA–S writes the functions based on the relations specified in the CMP model. A key part of this is a set of assumptions that MILA–S makes about the nature of ecological systems. For example, MILA–S assumes that if a biotic component consumes a certain other component, then it must need that other component to survive. A model with 'Fish' that contains 'consumes' connections to both 'Plankton' and 'Oxygen' would infer that fish need both Plankton and Oxygen to survive. MILA–S also assumes that species will continue to reproduce to fulfill their carrying capacity rather than hitting other arbitrary limitations. These assumptions do limit the range of simulations that MILA–S can generate, but they also facilitate the higher-level rapid model revision process that is the learning objective of this project. Figure 3 illustrates the general scheme for translating the semantics of CMP conceptual models into the semantics of the Netlogo agent-based simulations; Joyner, Goel & Papin (2014) provide a more detailed account of the translation scheme and process.

USE OF MILA–S IN MIDDLE SCHOOL SCIENCE

Prior to engagement with MILA–S, the 50 students in our pilot study received a two-week curriculum on modeling and inquiry, featuring five days of interaction with CMP conceptual modeling in MILA. In the first part of the study using MILA, students also used pre-programmed NetLogo simulations that did not respond to students' models, but nonetheless provided students experience with the NetLogo interface and toolkit. Thus, when given MILA–S, students already had significant experience with CMP conceptual modeling, NetLogo simulations, and the interface of MILA.

Constructed Models

During engagement with MILA, students produced models that can be described as retrospective and explanatory. Students started from an observable phenomenon, the aforementioned fish kill, and recounted a series of events that led to that result. Causal relationships were captured throughout the model, but only those that lay directly in the causal path leading to the observed phenomenon, and only in the specific way in which the chain occurred in the phenomenon. For example, one team modeled multiple feedback cycles to explain the phenomenon. In their model, a heat spike caused algae populations to grow out of control, then die off due to a lack of carbon dioxide to breathe and a lack of sunlight to produce energy (due to the thick algae clouding the lake). This led to a spike in algae-decomposing bacteria which suddenly had an ample food supply, as well as a drop in the population of oxygen-producing algae. These bacteria, then, consumed an enormous quantity of oxygen, causing the fish population to suffocate. This led to more dead matter in the lake, thus encouraging more bacteria reproduction, exacerbating the fish kill further.

This model presented a complete explanation for why and how the fish kill occurred in the lake; however, the model only captured a retrospective view of the series of events applicable in this situation. Although students could use mental simulation to imagine the results, these models do not explicitly capture dynamic relationships in the system, and thus are of limited use describing what would have happened under different circumstances. For example, had the temperature changed more slowly and allowed the algae to grow steadily rather than exploding and plummeting in quick succession, could the lake have sustained the increased algae population? Would the increased algae population have produced sufficient oxygen to allow the fish population to grow and thrive as well? Thus, models constructed with MILA were explanatory and retrospective.

With MILA–S, students constructed fundamentally different kinds of models that aimed not to capture the series of events that occurred, but rather to capture the dynamic relationships that gave rise to that series of events. Thus, the models constructed in MILA–S invoked a layer of abstraction and generalization away from the models constructed in MILA. For example, one team constructed an initial model that captured the three relationships they considered most pertinent in the system. These students already believed that the fish kill was caused by a sudden drop in oxygen, suffocating the fish. Thus, their first relationship was that fish consume oxygen. They similarly knew that oxygen is produced from sunlight, and thus included the relationship between sunlight and oxygen. These connections differed fundamentally from those modeled in MILA, such as accounting for trends in multiple directions (i.e. oxygen

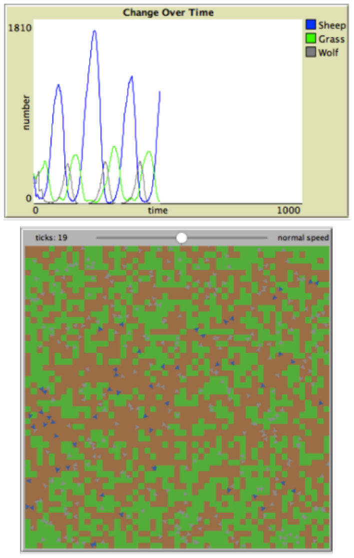


Figure 4: Long-durable stability of agent-based simulations.

production varies directly, up or down, with sunlight presence). The model was not constructed to directly explain the phenomenon, but rather to provide the relationships necessary so that under the right conditions, the phenomenon may arise on its own.

Model Construction Process

During prior engagement with MILA, we observed students engage in the model construction cycle. Model construction occurred as students constructed their initial hypotheses, typically connecting only a cause to a phenomenon with no mechanism in between. This was then used to guide investigation into other sources of information such as observed data or other theories to look for corroborating observations or similar phenomena. The conceptual model was then evaluated according to how well it matched the findings; in some cases, the findings directly contradicted the model, while in other cases, the findings lent evidence or mechanism to the model. Finally, the conceptual models were revised in light of this new information (or dismissed in favor of stronger hypotheses, reflecting revision at a higher level) and the process began again.

During engagement with MILA–S, however, we observed a profound variation on the model construction process. The four phases of model construction were still present, but the nature of model use and evaluation changed. Students started by constructing a small number of relationships they believe to be relevant in the system, the model construction phase. After some initial debugging and testing to become familiar with the way in which conceptual models and simulations fit together, students generated simulations and used them to test the implications of their conceptual models. After running the simulation a few times, students then evaluated how well the results of the simulation matched the observations from the phenomenon. This evaluation had two levels: first, did the simulation accurately predict the ultimate phenomenon (in this case, the fish kill)? Once this basic evaluation was met, an advanced evaluation followed: did other variables, trends, and relationships in the simulation match other observations from the phenomenon? For example, one team successfully modeled a fish kill by causing the quantity of food available to the fish to drop, but evaluated this as a poor model nonetheless because nothing in the system indicated a disturbance to the fish's food supply. Finally, equipped with the results of this evaluation, students revised their models to more closely approximate the actual system.

Thus, students still constructed and revised conceptual models, but through the simulation generation framework of MILA–S, the model use and evaluation stages took on the practical rigor, repeatable testing, and numeric analysis facilitated by simulations. Rather than speculating on the extent to which their model could explain a phenomenon, students were able to directly test its predictive power. Then, when models were shown to lack the ability to explain the full spectrum of the phenomenon, students were able to quickly return and revise their conceptual models and iterate through the process again.

Challenges

MILA–S provided an effective framework at simulating the interactions between a small number of components and their variables. However, some of the systems that students were examining involved several more components than these, along with multiple relationships between their variables. Upon reaching a level of complexity slightly higher than shown in Figure 2, the NetLogo simulations generated by MILA–S stopped providing meaningful feedback to students. The number of agents would explode based on the multiple consumption and production relationships at play, slowing the simulation down and rendering the visualization elements indistinguishable. Repeated runs of the same simulation with the same initial parameters sometimes generated wildly varied responses as the number of agents and methods exacerbated the influence of random chance on the simulation's outcomes.

It is likely that with proper parameters and relationships, MILA–S could still have generated usable simulations that gave meaningful feedback. The challenge was that most executions of the simulations gave limited or no feedback as to the changes that needed to be made to more closely replicate the phenomenon. The simulations contained too much noise to facilitate the process of model evaluation and revision.

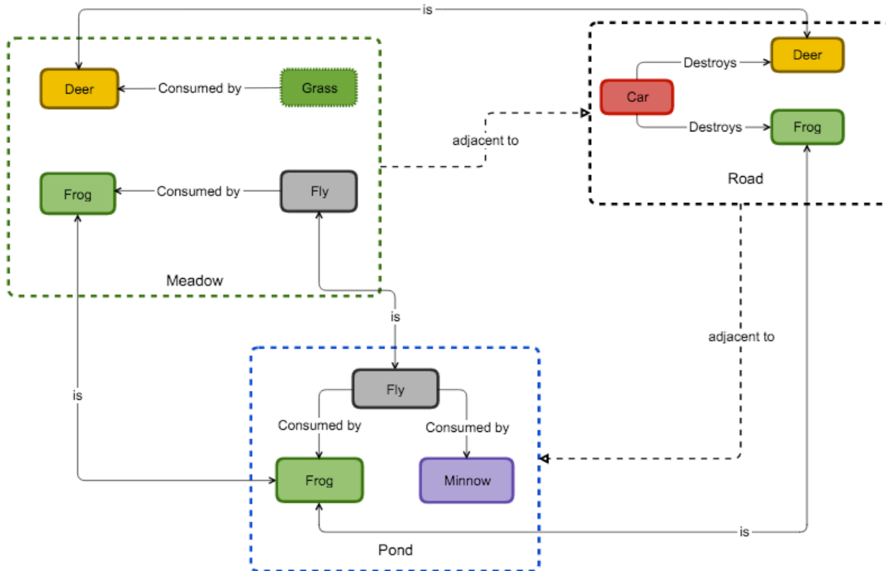


Figure 5: Enhancement of the CMP conceptual models by adding spatial relations.

FROM SCHOOL TO COLLEGE

In preparing MILA−S for use in college-level introductory biology classes, three factors are especially noteworthy. First, compared to middle school students in US in the 11-14 years range, college level students typically are 18-22 years old and therefore are cognitively more developed. Second, college-level students typically have more prior knowledge both about the systems of interest and the process of scientific modeling. Third, compared to middle-school science, college-level biology classes typically entail modeling of more complex ecological systems, with larger numbers and variety of species and larger number and range of interactions among them. Thus, to deploy MILA−S in college, we need to extend and enhance its capability in several ways.

Long-Duration Stable simulations

While ecological phenomena do not always sum up to a neat mathematical equation, there are emergent behaviors in an ecosystem that one comes to expect. For example, when a simple food chain ecosystem is modeled, one expects the resultant simulations to show the fluctuating predator-prey population cycles that can be mathematically modeled by the Lotka-Volterra equations. At the time of initial experimentation, it was difficult to get MILA−S to produce this expected behavior when simulating a food chain consisting of all biotic populations. In order to correct this, the concept of a "base population" was added to the conceptual model. This base population was implemented in NetLogo as patches instead of turtles like every other component. We found that in order to produce the cyclic behavior of predator-prey relationships the organism present at the bottom of the food chain needed to have the ability to repopulate and keep its population in tact without relying on it interacting with other members of the population. Essentially, once the base population could produce agents without interacting with other members of its species the simulations immediately stabilized and could be created much faster and with more success than experimenting with the organism's parameters. Figure 4 illustrates the stable results of this implementation.

Spatial Simulation

In addition to simulating food chain ecology and simple relationships between biotic and abiotic organisms, we are integrating spatially explicit relationships into the simulation. Integrating a spatial dimension allows users to model where organisms are allowed to exist and how they interact or are affected by their habitat. These simulations could be used to explore phenomena such as boundary effects, migration patterns, and urbanization effects. Figure 5 illustrates an initial expansion of the CMP language to include spatial regions such as meadow and pond, and spatial relations such as adjacency.

More Powerful Agent-Based Simulation Engines

As we noted above in the discussion on deploying MILA−S into middle school classroom, as the number of species and the variety of interactions among the species in the conceptual model increased, the NetLogo simulations became too slow to be useful. This means that for college-level ecological systems we may need more powerful agent-based simulation engines. Thus, we are integrating another off-the shelf agent-based simulation engine called Repast Simphony (http://repast.sourceforge.net/) into MILA−S. We chose Repast Simphony because it is an open-source agent-based simulation engine compatible with MILA−S, because it is similar to NetLogo in many respects but more powerful, and because it supports modeling of complex ecological systems. In the current version of MILA−S, we have partially integrated Repast Simphony into MILA−S; we are now testing the MILA−S' compiler for translating CMP conceptual models into the simulator's constructs.

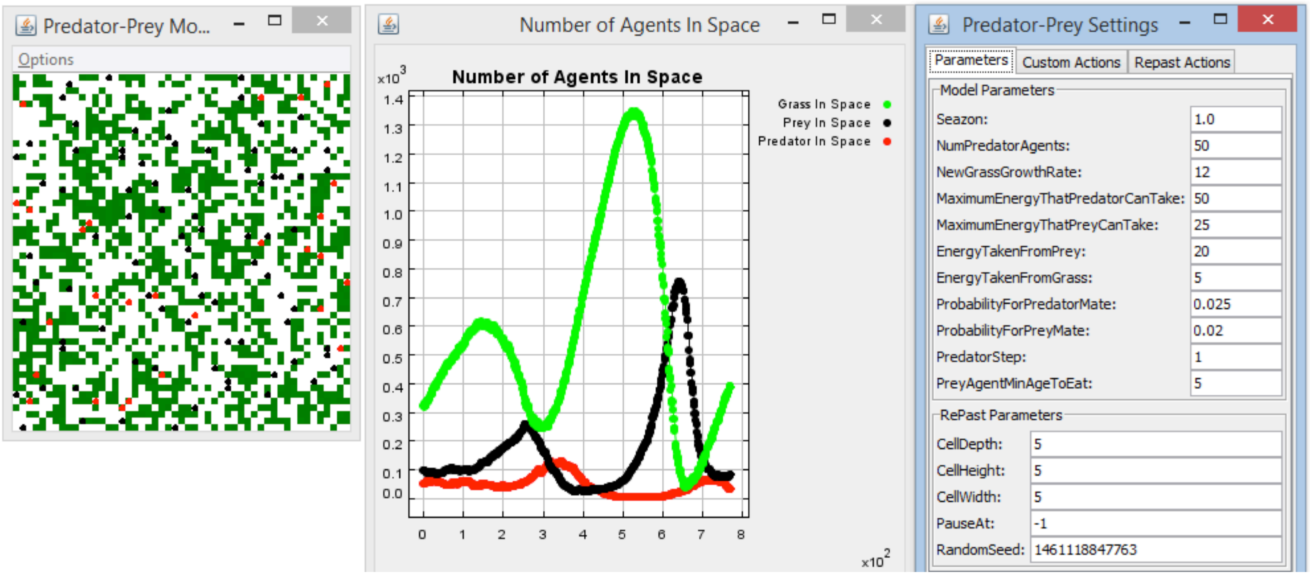


Figure 6: The output of Repast Simphony's agent-based simulation engine.

CONCLUSIONS

Cognitive systems research on qualitative reasoning typically focuses on qualitative modeling and qualitative simulation. Thus, in a parallel project on evaluating conceptual designs early in the design process, we have developed a technique for qualitative simulation of functional models of design concepts (Wiltgen & Goel 2016). In contrast, agent-based simulations are especially appropriate for modeling ecological systems. The question then becomes how can we use agent-based simulations in conjunction with qualitative modeling?

This paper has described the design of an interactive system called MILA–S for generating agent-based simulations from qualitative conceptual models of ecological systems. MILA–S not only enables construction of causal models of components and mechanisms in an ecosystem, but it also takes as input the causal model and autonomously generates an agent-based simulation that shows the temporal evolution of the system according to the causal model. The user needs to simply use a visual syntax for generating causal models and the interactive tool automatically generates the corresponding simulation. Further, because the simulation directly corresponds to the causal model, the results of the simulation directly evaluate the model and point to the revisions needed to the model.

Initial results from a pilot study with 50 students in a middle school provided preliminary evidence in favor of the hypothesis. Firstly, students approached the modeling process from a different perspective from the outset, striving to capture dynamic relationships among the components of the ecological system. These dynamic relationships then promoted a more abstract and general perspective on the system. Secondly, the process of model construction, use, evaluation, and revision presented itself naturally during this intervention, with the simulations playing a key role in supporting the cyclical process of constructing conceptual models.

Compared to middle school students, college-level students typically study more complex ecological systems. In this paper, we present extensions and enhancements to MILA–S in preparation for deployment in college. In particular, we described three extensions to MILA–S. (1) The ability to generate long-duration stable simulations. (2) The ability to take spatial relationships into account in both the conceptual and simulation models. (3) The ability to generate simulations that can capture a range of interactions in a variety of species. The next step will be to introduce MILA–S into college-level biology classes.

ACKNOWLEDGEMENTS

We thank Dr. Christine Spencer for many discussions about interactive technology for ecological modeling in introductory college-level biology courses, and Divya Vijayaraghavan for her work on integrating the Repast Simphony platform for agent-based simulations into MILA–S.

REFERENCES

1. Bonabeau, E. (2002) Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. In *Procs. National Academy of Sciences*, 99(3).
2. Bredeweg, B., & Forbus, K. (2003) Qualitative Modeling in Education. *AI Magazine*, 24(4): 35-45.
3. Bredeweg, B., Linnebank, F., Bouwer, A., Liem, J. (2009) Garp3 – Workbench for Qualitative Modeling and Simulation. *Ecological Informatics*, 4: 263-281.

4. Carruthers, P., Stitch, S., & Siegal, M. (editors, 2002) *The Cognitive Basis of Science,* Cambridge University Press.
5. Clement, J. (2008). *Creative Model Construction in Scientists and Students: The Role of Imagery, Analogy, and Mental Simulation.* Dordrecht: Springer.
6. Darden, L. (1998). Anomaly-driven theory redesign: computational philosophy of science experiments. In T. Bynum, & J. Moor (Eds.), *The Digital Phoenix: How Computers are Changing Philosophy*, (pp. 62–78). Oxford: Blackwell.
7. de Jong, T., & van Joolingen, W. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179-201.
8. Edelson, D., Gordin, D., & Pea, R. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences*, 8(3-4), 391-450.
9. Goel, A. (2013) One Thirty Year Long Case Study; Fifteen Principles: Implications of an AI Methodology for Functional Modeling. AIEDAM 27(3): 203-215, 2013.
10. Goel, A., & Joyner, D. (2015) Impact of a Creativity Support Tool on Student Learning About Scientific Discovery Processes. In *Procs. Sixth International Conference on Computational Creativity*. Provo, Utah.
11. Goel, A., Rugaber, S., & Vattam, S. (2009). Structure, Behavior & Function of Complex Systems: The SBF Modeling Language. *AIEDAM* 23: 23-35.
12. Grimm, V. et al (2005) Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science,* 310 (5750).
13. Halloun, I. (2007). Mediated Modeling in Science Education. *Science & Education*, 16(7), 653-697.
14. Jackson, S., Krajcik, J., & Soloway, E. (2000) Model-It: A Design Retrospective. In M. Jacobson & R. Kozma (editors), *Innovations in Science and Mathematics Education: Advanced Designs for Technologies of Learning* (pp. 77-115). Lawrence Erlbaum.
15. Joyner, D. & Goel, A. (2015). Improving Inquiry-Driven Modeling in Science Education through Interaction with Intelligent Tutoring Agents. In *Procs. 20th International Conference on Intelligent User Interfaces*. Atlanta, Georgia.
16. Joyner, D., Majerich, D. & Goel, A. (2013) Facilitating Authentic Reasoning about Complex Systems in Middle School Science Education. In *Procs. Fourth Conference on Systems Engineering Research*, Atlanta, March 2013; pp. 1043-1052.
17. Joyner, D., Goel, A., & Papin, N. (2014). MILA–S: Generation of Agent-Based Simulations from Conceptual Models of Complex Systems. In *Proc. 19th International Conference on Intelligent User Interfaces*, Haifa, Israel.
18. Joyner, D., Goel, A., Rugaber, S., Hmelo-Silver, C., & Jordan, R. (2011). Evolution of an Integrated Technology for Supporting Learning about Complex Systems. In *Proc. 11th IEEE International Conference on Advanced Learning Technologies*, Athens, GA.
19. Joyner, D., Goel, A., & Papin, N. (2014). MILA–S: Generation of Agent-Based Simulations from Conceptual Models of Complex Systems. In *Proc. 19th International Conference on Intelligent User Interfaces*, Haifa, Israel.
20. Nersessian, N. (1998) Conceptual Change in Science and Science Education. *Synthese* 80: 163-183, 1989.
21. Nersessian, N. (2008). *Creating Scientific Concepts*. Cambridge, MA: MIT Press.
22. North, M., et al. (2013) Complex Adaptive Systems Modeling with Repast Simphony, Complex Adaptive Systems Modeling, Springer, Heidelberg, FRG.
23. Novak, J. (2010) *Learning, Creating and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*. New York: Routledge.
24. Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632-654.
25. VanLehn, K. (2013). Model construction as a learning activity: a design space and review. *Interactive Learning Environments*, 21(4), 371-413.
26. Vattam, S., Goel, A., Rugaber, S., Hmelo-Silver, C., Jordan, R., Gray, S, & Sinha, S. (2011) Understanding Complex Natural Systems by Articulating Structure-Behavior-Function Models. *Journal of Educational Technology & Society*, 14(1): 66-81.
27. White, B, & Frederiksen, J. (1990) Causal Model Progressions as a Foundation of Intelligence Learning Environments. *Artificial Intelligence*, 42(1): 99-157.
28. Wilensky, U., & Reisman, K. (2006). Thinking Like a Wolf, a Sheep, or a Firefly: Learning Biology Through Constructing and Testing Computational Theories-An Embodied Modeling Approach. *Cognition and Instruction*, 24(2), 171-209.
29. Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology,*8:3-19.
30. Wiltgen, B., & Goel, A. (2016) Functional Model Simulation for Evaluating Design Concepts. *Advances in Cognitive Systems*, 4.

Testing scientific models using a QR model: Application to cellulose biodegradation

[1]Kamal Kansou and [2]Bert Bredeweg
[1]INRA, Biopolymères Interactions and Assemblages, BP 71267, 44316 Nantes, France
[2]Informatics Institute, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
Kamal.Kansou@nantes.inra.fr, B.Bredeweg@uva.nl

Abstract

The rapidly growing set of scientific publications makes it difficult for researchers to keep track of the progress towards adequate mechanistic explanations of phenomena. However, high-level representations can support integrating seemingly different results and ideas presented in the literature. This paper reports on our effort to deploy the qualitative reasoning framework as an instrument towards this end.

1 Introduction

The accumulation of scientific information is enormous. Keeping up to date in some fields of natural science is getting more and more difficult for the domain specialists (Fraser and Dunstan, 2010). For example, searching for "cellulose and hydrolysis and enzyme" in the Web Of Science yields more than 3000 scientific publications since 1995. Even experts find it difficult to keep integrating new mechanistic information about ligno-cellulose hydrolysis and envision the consequences on the system dynamics.

An emerging question is whether the (new) pieces of knowledge found in publications about a topic provide a way forward to a better (possibly complete) understanding of the underlying mechanism.

Higher-level representations can support literature integration by reviewing and assembling information provided in scientific papers in a computable model. Higher-level (conceptual) modelling formalisms can integrate scattered qualitative information about a mechanism and provide a valuable envisioning of the system dynamics. Our objective is to explore solutions for representing and manipulating mechanistic explanations from publications using a computational model. We focus on the analysis of cause-effect relations to identify/test putative explanations for a set of evidences.

1.1 Domain – Cellulose hydrolysis limitation

We are interested in explanations for processes limiting the cellulose hydrolysis. Cellulose is the main component of plant cell wall, and an abundant and accessible renewable source of carbon. As such, cellulose is of central interest for the many natural and industrial processes, including the production of biofuel. Hydrolysis of solid cellulosic substrate into soluble cellodextrins by a cocktail of cellulases is characterised by progress-curves determined by the amount of the carbohydrates released in a solution. The curve shows a saturation-shape that reflects the catalytic activity. It is known that the efficiency of the depolymerisation of solid cellulose chains gradually declines with time. This means that the cellulases activity gets less efficient as the reaction proceeds (Lynd et al., 2002; Zhang and Lynd, 2004).
Numerous observations pertaining cellulose hydrolysis can be found in the literature. However, establishing a mechanistic explanation of the declining rate is still an important and unsolved issue. This missing insight hampers the global conversion efficiency of cellulose into ethanol (Lynd et al., 2002; Zhang and Lynd, 2004).

1.2 Potential of QR as an instrument

Quantitative approaches (e.g. ODE) need precise data, and are very dependent on experimental conditions. Even if the model structure can be applied in a variety of experimental situations, the need to get sufficient data to perform precise parametrization is a limitation factor. Furthermore, quantitative models cannot readily represent an informal description of a mechanistic explanation in an easy manner, as for instance text or diagrams can. Finally, mathematic formulation of a physical process is not directly interpretable in terms of cause-effect.
In the work presented in this paper, we use the Qualitative Reasoning (QR) framework, which does provide representations of cause-effect and is also able to generate simulations of the system dynamics. QR modelling is complementary to quantitative approaches in the sense that it allows for formulating distinct paradigms and for providing a first assessment to a range of evidences without requiring precise measurements of parameters or specific experimental conditions.

1.3 The challenge

We present an approach to stepwise construct a mechanistic explanation from selected papers about cellulose hydrolysis rate slowing-down using the QR framework. Many studies have investigated the cause of the phenomenon; both enzyme and substrate-related factors can be held responsible for the decline of hydrolysis rate. However, an integrated or unified explanation is not available.

We have developed three QR models. Two models are derived from published mechanistic models. The third model is derived from experimental observations from the literature and analysis of the simulations of the two other models. Our paper also reflects on methodological issues relevant to creating and assessing such models exploiting observations from publications. Our primary objective is to demonstrate how the QR framework can be used for this.

2 QR for mechanism modeling

QR strives for inferring behaviour from physical system structure in a symbolic, human-like manner. We use Garp3 (Bredeweg et al., 2009), a workbench for constructing and simulating QR models. To illustrate the use of QR, consider the basic enzymatic reaction: $E+S \rightleftharpoons ES \rightarrow P$, with E (enzyme), S (substrate), ES (complex enzyme-substrate), and P (product). The ODE system representing this phenomenon computes the derivatives of the E, S, ES and P. concentrations. These simulations are well known. Fig. 1 shows the kinetic curves (coloured lines), produced with dummy values for the kinetic constants.

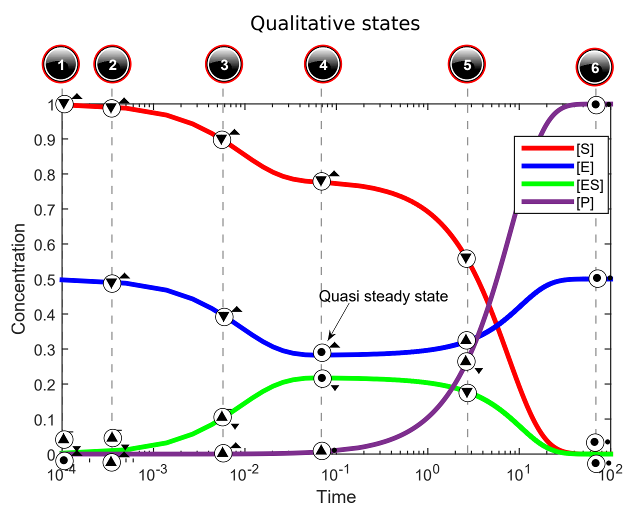


Figure 1. Simulation results for an enzymatic reaction with logarithmic time. The top row shows corresponding qualitative states, produced by simulating a QR model. Value histories of the quantities are placed on top of simulation curves. Key states are: initial state 1 (substrate starts being complexed with enzyme), state 4 (quasi-steady state), and end-state 6 (substrate conversion complete).

The Garp3 model implements a process-centric view, which emphasizes rates. Thus a Garp3 model of Equation 1 includes four entities (E, S, ES, P) each with a quantity *Concentration,* but also the rates *Ratein* and *Rateout* for respectively formation rates (for ES and P) and disappearing rate (for ES). In Garp3, quantities are characterized by: <*Magnitude, Derivative*>. The domain of allowable magnitudes associated with each quantity is called the Quantity Space (QS). *Concentration [in E]* and *Concentration [in S]* are assigned QS: {*Zero*, *Plus*, *Max*}, the other quantities have QS: {*Zero*, *Plus*}. All derivatives have QS {▼,●,▲} representing decreasing, steady, and increasing.

Garp3 provides two primitives for capturing causal dependencies between quantities, direct influence (I+ and I-) to model a *rate* influencing a concentration, and qualitative proportionality (P+; P-) to model the propagation of changes from one quantity to the next (cf. Forbus, 2008). P* is special kind of proportionality that captures the relation between the terms of a product and the result of this product.

Simulation results for the enzymatic reaction model, starting from maximum magnitudes for *Concentration [in E]* and *Concentration [in S]* includes a state-graph of 9 states. A Behaviour Path (BP) is a possible behaviour defined as a succession of qualitative states along a complete timeline. In Fig. 1 the BP [1→2→3→4→5→6] and value histories corresponding to the simulation curves are provided.

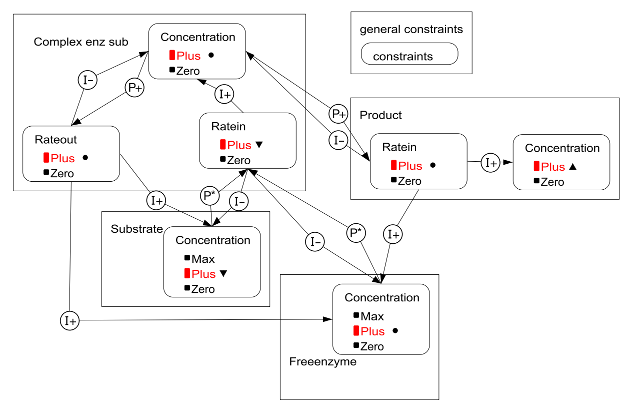


Figure 2. Causal dependencies compiled by Garp3 for state 4, providing a causal account for what is depicted by the value history graphs.

This shows that this particular BP matches the numerical simulation given Fig. 1. Key qualitative states of the process are identified this way, thus state 4 of the BP represents the quasi-steady state. The assembly of the causal chain active in state 4 is shown in Fig. 2. From this graph one can identify interacting feedbacks. For instance: two positive feedbacks, one productive including *Ratein [in P],* one unproductive including *Rateout [in ES]*, determine the reaction overall efficiency.

3 Explanatory model based on scientific publications

3.1 Behaviour path (BP)

In Garp3, a qualitative simulation of system behaviour uses a set of quantities $x_i \in X, i = \{1, \dots n\}$ linked by causal dependencies, and constrained by inequalities. A Qualitative State (QS) describes the system at time t such as: $\mathrm{QS} = \{\langle t, <\ x_i, \text{magnitude} = \alpha, \text{derivative} = \beta >\rangle, \forall x_i \in X\}$ with α, some value of the QS assigned to x_i, β a value of the QS assigned for derivatives (in Garp3: $\{\blacktriangledown, \bullet, \blacktriangle\}$). A Behaviour Path (BP) is a finite sequence of m qualitative states that represents a possible qualitative behaviour over time. All QSs of a BP but the last one have a transition relation towards a possible and qualitatively distinct successor such as: BP = $QS_0 \rightarrow \cdots \rightarrow QS_m$

Each BP is associated to a discrete timeline, T, composed of m time periods such as $T := \langle t_0, \dots t_m \rangle$. Depending on the nature of the state, a period of time can be an instant or an interval. Note that, if two similar QSs are met at different times, as for a periodic behaviour, Garp3 refers to the same QS. The BP is then a loop.

3.2 Target behaviour (TB)

A Target Behaviour (TB) is a qualitative abstraction of one or more observations of actual behaviours exhibited by a real (target) system, whose structure is unknown and investigated by domain scientists. A TB captures distinctive features as Target States (TS) ordered in time for which the model needs to provide an explanation. A TS describes the target system for a given time period, t, through a set V of nt quantities with known magnitudes and/or derivatives: $\mathrm{TS} = \{\langle t, <\ x_i, \text{magnitude} = \alpha', \text{derivative} = \beta' >\rangle, \forall x_i \in V\}$.

A model must include the variables of the TS to have a chance to satisfy it, therefore $V \subseteq X$. Similarly α' and β' belong to QSs also included in the corresponding qualitative model. For a TS α' and β' can be subsets of quantity spaces excluding the empty set. The full QS is noted "?".

In agreement with the QR formalism, a TB represents the change of the magnitude and the derivative of some quantities at distinct time intervals. Contrary to a BP, a TB does not need to cover a complete timeline, that is, from an initial state to an end state. A TB is defined as a finite sequence of mt target states, strictly ordered in time such as: TB = $TS_0 \rightarrow \cdots \rightarrow TS_{mt}$

The successor relation indicates simply that the next TS occurs some time later. Here again two successive TSs must be distinct. A TB applies to BPs produced by a simulation model to classify the possible behaviours of the system. It imposes that the TSs of a TB are satisfied in the right order by the QSs, therefore $mt \leq m$. It is often desirable that a TB covers a continuous time-period to rule out false positives.

Suppose that the curves in Fig.1 are observations (not the result of simulation). We can select states such as (*i*) initial-state of the reaction, (*ii*) intermediate state where [ES] is at a peak, and (*iii*) end-state. Those are likely to be characteristic states of the system under investigation. Then a possible TB could describe magnitudes and derivatives for the ES and P concentrations at three moments ($t_0 < t_1 < t_2$), as shown in Table 1.

Table 1. TB capturing qualitative features of Fig. 1 curves

Time index	Concentration [ES]	Concentration [P]
t_0 (initial state)	<Zero, ? >	<Zero, ?>
t_1 (intermediate state)	<Plus, ●>	<Plus, ▲>
t_2 (end state)	<Zero, ●>	<Plus, ●>

Note, '?' can be one of {▼,●,▲}.

Trying to explain the TB using a QR model of the enzymatic reaction produces the BP: [1→2→3→4→5→6] (Fig.1) consistent with Table 1: state 1 matches the initial-state, state 4 matches the intermediate (quasi-steady) state, and state 6 the end-state. All BPs containing these 3 states in the right order are consistent with Table 1. Therefore a QR model of the enzymatic reaction would provide a sufficient explanation for the TB.

3.23 Assessing QR models versus literature information

Using QR, it is possible to capture the causal information described in publications into qualitative cause-effect models, simulate these models, and thereby envision the information in terms of system behaviour. However, capturing causal links indistinctively from a set of papers will quickly make the qualitative simulation intractable and inappropriate for conveying a meaningful explanation to domain experts. Instead, we adopt an incremental model-building approach driven by a Target Behaviour (TB).

Establishing the TB is the first step in the modelling process, as it determines the modelling goal and orients the choices of entities, quantities, and QSs relevant for simulating the observed behaviours (Kansou and Bredeweg, 2014). In the ideal case, the TB is a mapping of existing time-series data. However, in natural sciences building a TB from a dataset obtained in specific experimental conditions can be insufficient to discriminate between concurrent explanations. Qualitative abstraction smoothens the peculiarities of experimental conditions reported in papers. This enables integration of observations from different sources into a composite TB, albeit with loss of some precision. A TB is built primarily on source papers and/or experimental results. This phase involves domain experts as main beneficiaries of the work for guidance about the literature and/or conducted experiments.

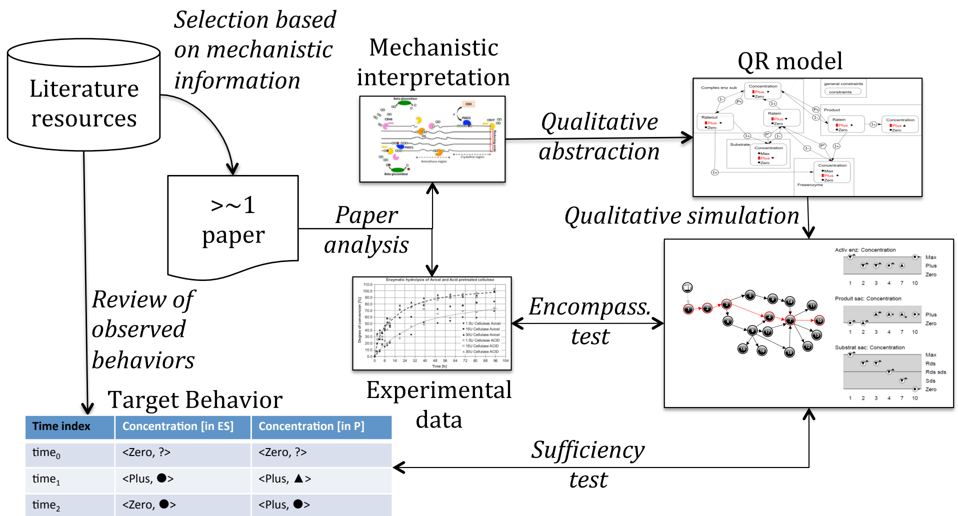


Figure 3. Using QR as an instrument to integrate scientific information from literature.

Our modelling methodology is depicted in Fig. 3. Selected papers introduce observations or simulation results related to the TB and provide useful mechanistic interpretations. Papers describing a quantitative model, usually with ODEs, are especially interesting as they propose a formal representation of a mechanism. For each version of the QR model its legitimacy as a faithful representation of the discovered knowledge is assessed using data and observations provided in the source papers, using the *encompassment* and the *sufficiency* test. The tests are defined as follows:

Encompassment: The QR model is a consistent representation of the interpretations given in the source papers. The model generates behaviours that match the observed data, numerical simulations or qualitative observations supplied in these papers.

Sufficiency: The QR model implements a plausible explanation for the target behaviour. The model generates a behaviour from which a plausible explanation for the target behaviour can be derived.

4 Testing cellulose hydrolysis paradigms

4.1 Defining target behaviour

To compose the TB, a short review of publications pertaining to the cellulose hydrolysis rate decline over time was performed. We strived for selecting publications addressing the most basic conditions, involving common cellulosic substrates with common hydrolytic enzyme, cellulase. The most important cellulase in this system has a processive action (enzyme complexed on a cellulose strand chops it up step-by-step as small sugars of similar size). The goal was to extract observations caused by basic processes that will take place regardless of the substrate nature (cellulosic or ligno-cellulosic) or the enzymatic cocktail complexity.

Hydrolysis rate decline: decline rate is related to the absolute quantity of bound enzymes as well as the specific rate per adsorbed enzyme (Lynd et al., 2002). The phenomenon extends over different time-scales. The hydrolysis decreases exponentially, immediately after an initial burst of catalytic activity and then at a much slower pace (Praestgaard et al., 2011), up to few days (Gan et al., 2003).

Restart experiments: Amongst the experiments in the domain, typical "perturbations" of the system include the addition of fresh enzyme in the course of the reaction, so-called "restart experiment". This type of experiment provides information about the system state, in particular about the state of the enzymatic component (Lynd et al., 2002; Eriksson et al., 2002). It has been observed that the addition of fresh enzymes, shortly after the reaction initialization, causes a clear restart of the hydrolysis (Cruys-Bagger et al., 2012). After longer time, it will cause a weak restart unless the cellulose surface is cleaned up beforehand (Yang et al., 2006).

We propose three TBs, (TB1, TB2 and TB2'), to capture prominent aspects of the experimental observations reported above (Table 2-4). TB2 (Table 3) depicts the restart phenomenon as the conversion of free enzyme in solution into catalytic active enzyme so that, the recruitment of active enzyme increases as long as the free enzyme quantity is increasing. In TB2' (Table 4) there is some process(es) limiting and eventually interrupting the restart phenomenon, so that increase in free enzyme might not result in an increase of catalytic rate.

Table 2. Hydrolysis rate is declining following an initial burst of hydrolytic activity (TB1).

Time index	Free enzyme	Catalytic rate
t_0	<Max, ?>	<0, ▲>
t_1	<{Zero, Plus}, ?>	<Plus, ▲>
t_2	<{Zero, Plus}, ?>	<Plus, ●>
t_3	<{Zero, Plus}, ?>	<Plus, ▼>

Table 3. Second dose of enzyme brings about a hydrolysis restart (TB2).

Time index	Free enzyme	Catalytic rate
t_0	<Plus, ▲>	<Plus, ●>
t_1	<Plus, ▲>	<Plus, ▲>
t_2	<Plus, ●>	<Plus, {●,▲}>

Table 4. Restart, but distinct from TB2 in that it represents a limited restart due to extra processes (TB2').

Time index	Free enzyme	Catalytic rate
t_0	<Plus, ▲>	<Plus, ●>
t_1	<Plus, ▲>	<Plus, {●, ▲}>
t_2	<Plus, ▲>	<Plus, ●>
t_3	<Plus, ●>	<Plus, ?>

4.2 Establishing models

We developed three QR models to test paradigms about cellulose hydrolysis proposed in the domain literature. To present the models structure we adopted a diagrammatic representation describing the causal linkages between quantities (Figs. 4-5) where rectangular box represents *concentration* or *amount_of* something, ellipse represents rate and causal linkages are labelled "P +/-" or "I +/-". "P+" and "P-" (proportionality relations) can connect two boxes together, a box to an ellipse or two ellipses together. "I+" and "I-" are direct influences. In the graph they can relate only an ellipse to a box. Algebraic relations can be implemented in Garp3 through qualitative algebra. Operators are represented by the symbols ⊕, ⊖ and ⊗.

The first QR model (M1) implements the surface-coverage limitation explanation based on modified Langmuir-Michaelis-Menten equations (proposed by Maurer et al., 2012). The system accounts for three processes: (*i*) reversible adsorption on the surface, (*ii*) reversible formation of surface enzyme-substrate complex, and (*iii*) hydrolysis of substrate generating a product without release of the active enzyme. The principle of the model is:

$$Ef + AS \leftrightarrow Ea + S \leftrightarrow ESa \rightarrow ESa + P.$$

The corresponding mass balance relates the accessible surface concentration (AS) and the free enzyme concentration (Ef) to the production rate (dP/dt) via the surface concentration of adsorbed cellulase in an uncomplexed form (Ea) and in a complexed and catalytic active form (ESa). S stands for the substrate concentration surface cellulose chain, assumed constant in the model.

The model is depicted in Fig. 4. *Free enzyme* first adsorbs on *Accessible surface*, to form *Adsorbed enzyme*. *Adsorbed enzyme* can form *Active enzyme* that degrades the cellulose at *Catalytic rate*, or get back to the *Adsorbed enzyme* form. The *Covered surface*, populated by *Adsorbed enzyme* and *Active enzyme* reduces the *Accessible surface*.

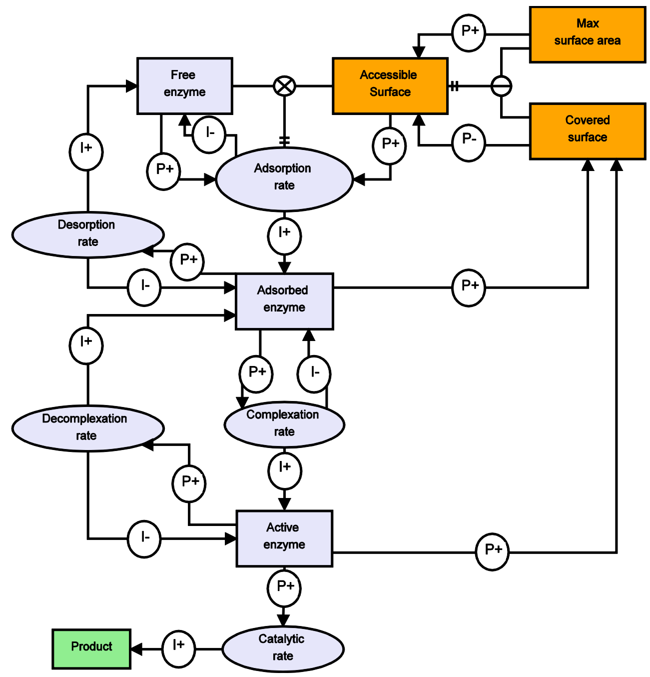


Figure 4. Model M1

The second QR model (M2, Fig. 5) implements an explanation related to putative presence of obstacles at the cellulose surface limiting the processive action of cellulase (presented in Jalak and Valjamae (2010), also implemented as kinetic model in Prastegaard et al., (2011) and in Cruys-Bagger et al., (2012)). The kinetic model implements the stalling of the processive enzyme when it reaches a surface obstacle during the catalytic process. The QR model has a global *Adsorption/complexation rate* of *Free enzyme* with cellulose, to form *Active enzyme*. *Active enzyme* degrades the cellulose strands processively at *Catalytic rate*. Next, it can either desorb (*Desorption rate1*) or get stalled if it meets an obstacle at *Stalling rate* and becomes *Stalled enzyme*. The *Desorption rates* (*Desorption rate1* + *Desorption rate2*) refill the amount of *Free enzyme* fuelling the turn-over. In our model, hydrolysis is a single step process performed by all *Active enzyme*, and not a summation of hydrolytic acts occurring along the cellulose strands as in the original model (Cruys-bagger et al., 2012). At the qualitative level, it would make the system and the ensuing explanation needlessly complicated. The relation between the *Catalytic rate* and the *Stalling rate* is modelled using a proportional dependency (P+).

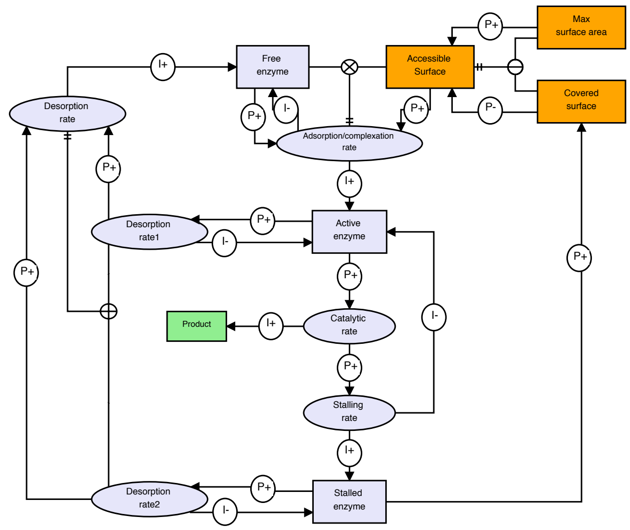


Figure 5. Model M2 and M3. M3 includes the surface limitation model fragment represented with orange boxes.

The third model (M3, Fig. 5) is an extension of M2 including the surface limitation from M1. It accounts for surface contamination by enzyme, which can hinder the hydrolytic activity. The process by which surface enzyme hinders the hydrolytic process is not clarified. Limitation of the adsorption due to *Covered surface* is similar to M1 (Fig. 4); this model also includes the case where *Covered surface* affects the complexation process. By extrapolating the impact of *Stalled enzyme* at the surface, we assume a proportional dependence (P+) between the *Stalled enzyme* and the *Covered surface*. In doing so, we test a new mechanism by which *Stalled enzyme* hinders the Adsorption and/or the complexation rate. Naturally other linkages of this kind could be tested as well. A more complete screening of the possible model structures is envisaged in future work.

4.3 Results

4.3.1 Simulation scenarios

The simulations used for testing the decline of the hydrolysis rate (TB1 and TB2') start from a scenario with only substrate and free enzyme (no product nor enzyme other than free in solution). Simulations of the restart phenomenon (TB2) are produced from a perturbation scenario reproducing the addition of a second dose of enzyme. Starting from a system in a state of equilibrium, with all the rates of the model (e.g. *Catalytic rate*, *Stalling rate*) being positive and stable, the addition of new *Free enzyme* is modelled through a feeding rate, exogenous to the system. The feeding rate is imposed to decrease over time. This accounts for the enzyme diffusion in the solution and limits the perturbation in time.

4.3.2 Description of model simulations

Models M1 on one hand and M2 and M3 on the other, exhibit very distinct behaviours. The M1 state-graph has seven states ordered linearly, with one stable end-state (state 5). The simulation envisions a conversion of *Free enzyme* into *Adsorbed enzyme* and then into *Active enzyme*. Models M2 and M3 produce state-graphs of 27 and 41 states each having a characteristic water lily leaf shape with a unique end-state at the centre (state 4). After a common starting branch (states 1→2) the system either: (*i*) goes directly to state 4 via the BP [1→2→3→4], (*ii*) initiates oscillations before reaching state 4 (*e.g.* Fig. 6), (*iii*) oscillates without reaching the end state. In the present situation, the system can be interpreted as a damped oscillator moving towards a steady state. State 4 is the equilibrium state with all quantities of the system steady, except for the concentration of *Product* that increases at a constant rate. The equilibrium state is characterized by the following equalities between the rates:

Ads/Comp rate = *Des rate* = *Des rate1* + *Des rate2*
⇒ δ(*Free enzyme*) = 0
Ads/Comp rate = *Stalling rate* + *Des rate1*
⇒ δ(*Active enzyme*) = 0
Stalling rate = *Des rate2* ⇒ δ(*Stalled enzyme)* = 0
⇒ δ(*Accessible surface*) = 0

M2 and M3 both envision the accumulation of *Stalled enzyme* governed by the balance between the *Stalling rate* and *Desorption rate2*. Inclusion of *Accessible surface* in M3 implements a negative feedback from *Stalled enzyme* concentration to *Adsorption/Complexation rate*. This leads to more complicated oscillations than for model M2. This may reflect a longer establishment of the equilibrium state.

4.3.3 Testing the Encompassment of sources

To investigate the encompassment of M1 for the interpretation by Maurer et al., (2012) simulation curves have been produced of the published ODE model (not shown here). The longest BP (seven states) produced by M1 maps exactly onto the quantitative simulation. It depicts the burst and then the decline of *Adsorbed enzyme*, while *Active enzyme* increases up to maximum level from which it stabilizes. Limitation of *Active enzyme* can be traced back to decline of *Accessible surface* and *Free enzyme*. Even if *Accessible surface* can regulate the *Adsorption rate*, (Fig. 4) deleting this model fragment does not change the system behaviour. The encompassment for M2 regarding Cruys-Bagger et al., (2012) is depicted in Fig. 6. The BP [1→2→3→5→6→4] matches the simulation curves provided in that publication. A fraction of enzyme being stalled at the cellulose surface, it is easy to infer from Fig. 5: a low *Desorption rate2* will create a bottleneck effect impacting the turnover between free and active enzymes. M2 conveys successfully the idea that obstacles at the cellulose surface would slow-down the hydrolytic activity. Interestingly, the shortest BP of the state-graph ([1→2→3→4]) also matches one of the experimental curves of Cruys-Bagger et al., (2012) (not shown here) obtained with the lowest substrate concentration. Here, the hydrolysis rate levels out close to its maximum value so that the burst is barely noticeable. In this situation the *Adsorption/Complexation rate* is certainly limitative compared to the other rates of the system.

M3 also fulfils the encompassment test for the Cruys-Bagger et al., (2012) results. It produces the same BP [1→2→3→5→6→4] as shown in Fig. 6. *Accessible surface* inclusion in M3 is not a representation of existing theory. As such, it does not encompass specific papers.

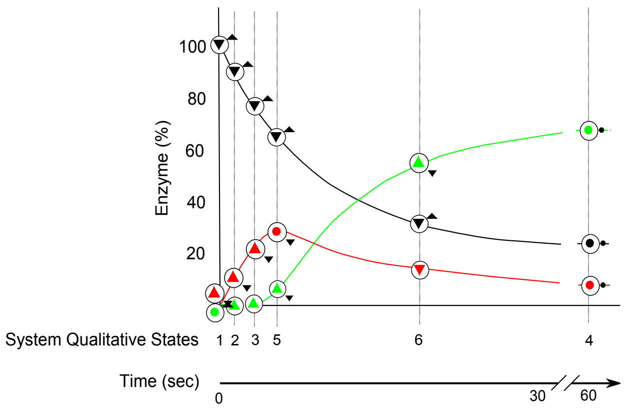


Figure 6. Enzyme evolution in M2 and M3 for the BP [1→2→3→5→6→4] placed on top on simulation curves from Cruys-Bagger et al., (2012). Red is Active, black is Free and green is Stalled enzymes.

4.3.4 Testing the sufficiency of the explanation

Results of the insufficiency test are shown in Table 5. M1 and M2 provide incomplete explanation for one of the three TBs. Particularly, M1 produces no BP with a decline of the hydrolysis rate. Indeed, following *Active enzyme* evolution, the *Catalytic rate* increases then stabilizes, which does not satisfy TB1, t_3 (Table 2). M1 produces BPs in line with TB2: addition of *Free enzyme* generates a restart of the hydrolysis process. It can also produce BPs satisfying TB2', as the reduction of *Accessible surface* due to the accumulation of *Adsorbed* and *Active enzyme*s can counteract the restart due to a second dose of *Free enzyme*. Compliance to TB2' is detailed below for M2 and M3.

Table 5. Results of the sufficiency test

Model	TB1	TB2	TB2'
M1	-	x	x
M2	x	x	-
M3	x	x	x

Fig. 6 shows that M2 provides an explanation for the decline of the *Catalytic rate* (directly proportional to the concentration of *Active enzyme*) in agreement with TB1. TB2 is assessed through a scenario that mimics the addition of *Free enzyme* in a system at the equilibrium, with a forced increase of *Free enzyme* while *Active enzyme* and *Stalled enzyme* are stable. First steps of this simulation are reported in Fig. 7a,b.

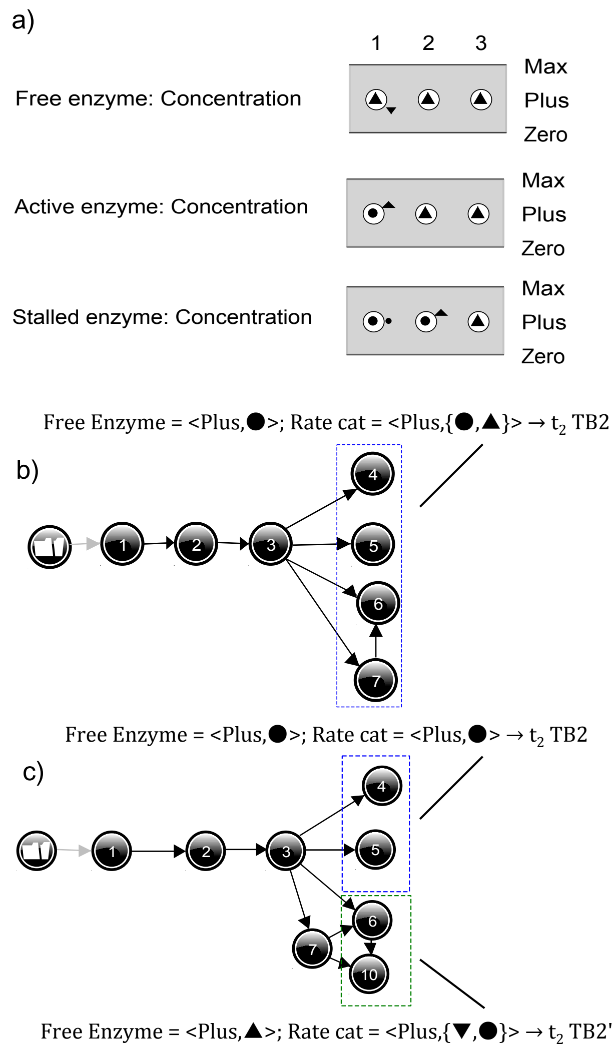


Figure 7. Partial simulation results of Restart scenario for model M2 and M3. a) Value history of the 3 first states for M2 and M3, b) first steps of M2 simulation in agreement with TB2, c) first steps of M3 simulation in agreement with TB2 and TB2'. Rate cat stands for Catalytic rate.

Addition of *Free enzyme* increases the adsorption of enzyme on the cellulose and, necessarily, brings about the increase of *Active enzyme* ([1→2→3] (Fig. 7a,b). This is consistent with TB2 (Table 3). From state 3 onwards, all possible BPs encompass the *Free enzyme* stabilization (*Free enzyme* = <Plus,●>, in states: 4, 5, 6, 7) with *Catalytic rate* = <Plus,●>, in states 4, 5 or *Catalytic rate* = <Plus, ▲> in states 6, 7. Both comply with t_2 of TB2 (Table 3) (Fig. 7b). This model implies that the second dose of

Free enzyme is completely transformed into *Active enzyme*, and causes a burst of hydrolysis anew. This behaviour was observed in concrete experiments as reported in Praestgaard et al., (2011) and in Cruys-Bagger et al., (2012). Regarding TB2' (Table 4), as shown in Fig. 7b, all the BPs produced with the restart scenario envision a stabilization of *Free enzyme* concentration prior to the stabilization of the catalytic rate, which does not match the t_2 stage. Hence, M2 does not provide an explanation for a weak restart.

M3 extends M2. M3 also meets TB1 and TB2. Regarding TB2', first steps of the simulation are given in Fig. 7c. It shows a restart of the hydrolysis in the path [1→2→3], Fig. 7a. For the next steps, some BPs satisfy TB2'. One of them starts with [1→2→3→6]. For this path *Free enzyme* = <Plus, ▲> so the addition of new enzyme is still ongoing. However, in state 6 the *Adsorption/Complexation rate* and the *Catalytic rate* stabilize (<Plus,●>), in agreement with the stage (t_2) of TB2'. The t_3 of TB2' is met in the following steps (not shown here). Given the model structure (Fig. 5), it can be inferred from *Adsorption/Complexation rate* = <Plus,●> and *Free enzyme* =<Plus, ▲> that *Accessible surface* = <Plus, ▼>. Therefore, the reduction of the *Accessible surface* limits the adsorption of *Active enzyme*, canceling out the restart phenomenon. Including *Accessible surface* in M3 does provide an explanation for a weaker restart effect.

5 Discussion

The presented work prepares the ground for a structured approach of literature integration using QR, using TB as a cornerstone. In addition, M3, shows that it is relatively simple to move from known paradigms to new ones.

Despite the fact that several mechanistic models of cellulose hydrolysis have been proposed in the literature and match experimental data, a scientist of the domain must still feel unsure about which one explains the observed the kinetics best, not to mention the parametrization or data collection techniques. This illustrates the difficult problem of verification and validation of numerical models in natural sciences (Oreskes et al., 1994). QR techniques can help overcome some of these difficulties as they focus on reproducing more abstract and generic, and accounting for diverse observations of the phenomenon.

Mechanistic interpretations, as well as observations, are available in the literature. Extraction qualitative information from selected papers has been performed manually for the work presented in this paper. Automatic composition of QR model structure is expected from treatment of natural language in the future (McFate et al., 2014).

A key question to be addressed concerns the assessment of the models, especially the genericity of the explanation they convey. Our approach used TB as reference for assessing explanations. Hence, it is the properties of the TB that determines the property of an explanation model.

References

Bredeweg, B., Linnebank, F., Bouwer, A. & Liem, J. 2009. Garp3 – Workbench for qualitative modelling and simulation. Ecological informatics, 4(5-6), 263–281.

Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K. & Westh, P. 2012. Pre-steady-state Kinetics for Hydrolysis of Insoluble Cellulose by Cellobiohydrolase Cel7A. *J. Biol. Chem*., 287(22), 18451-18458.

Eriksson, T., Karlsson, J. & Tjerneld, F. 2002. A model explaining declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglucanase I (Cel7B) of Trichoderma reesei. *Applied Biochemistry and Biotechnolog*y, 101(1), 41-60.

Forbus, K. 2008. Qualitative modeling. In F. Harmelen*,* V. Lifschitz*, &* B. Porter eds*. Handbook of knowledge representation (pp.* 361–394*).* New York: Elsevier

Fraser, A.G. & Dunstan, F.D. 2010. On the impossibility of being expert. *BMJ*, 341.

Gan, Q., Allen, S.J. & Taylor, G. 2003. Kinetic dynamics in heterogeneous enzymatic hydrolysis of cellulose: An overview, an experimental study and mathematical modeling. *Proc Biochem* 38, 1003–1018.

Jalak, J. & Väljamäe, P. 2010. Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnology and Bioengineering*, 106(6), 871-883.

Kansou, K. & Bredeweg, B. 2014. Hypothesis assessment with qualitative reasoning: Modelling the Fontestorbes fountain. *Ecological Informatics*, 19, 71-89.

Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol Mol Biol Rev*. 66, 506–577.

Maurer, S.A., Bedbrook, C.N. & Radke, C. J. 2012. Cellulase Adsorption and Reactivity on a Cellulose Surface from Flow Ellipsometry. *Ind. Eng. Chem*. Res., 51, 1389−11400.

McFate, C.J., Forbus, K. & Hinrichs, T. 2014. Using Narra-tive Function to Extract Qualitative Information from Natural Language Texts. *Proceedings of AAAI 2014*, Québec, Canada

Oreskes, N., Shrader-Frechette, K. & Belitz, K. 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263(5147), 641-646.

Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K.C., Borch, K. & Westh, P. 2011. A kinetic model for the burst phase of processive cellulases. *FEBS J.*, 278(278), 1547–1560.

Yang, B., Willies, D.M., & Wyman, C.E. 2006. Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion. *Biotechnology and Bioengineering*, 94(6), 1122-1128.

Zhang, Y.-H. P. & Lynd, L. R. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering*, 88(7), 797-824.

Challenges in Formulating Explanatory Models for Co-morbidities

Matthew Klenk and **Marzieh Nabi** and **Adam Arvay**
Johan de Kleer and **Daniel G. Bobrow**
Palo Alto Research Center
3333 Coyote Hill Rd
Palo Alto, CA, 94304
klenk,marziehn,arvay,dekleer,bobrow@parc.com

Abstract

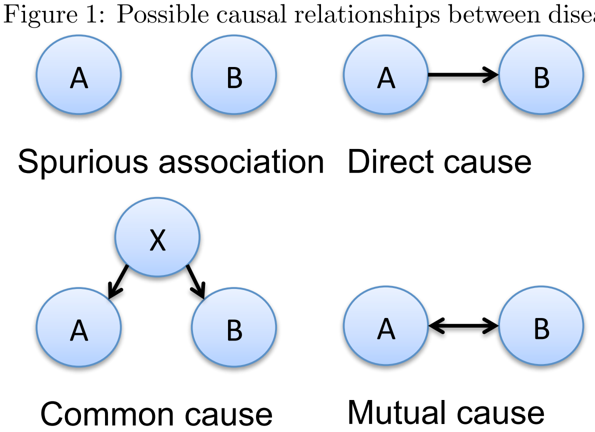
Patients with multiple health conditions pose significant challenges for modern healthcare. Understanding if and how these conditions are linked is essential to providing effective treatment. Physicians and researchers create explanatory models to develop hypotheses for these connections. In this paper, we discuss the breadth of domains these explanations draw upon as well as the diversity of applications of these models. Throughout the paper, we use example explanatory models from published literature and discuss the state-of-the-art of knowledge representation to support clinicians.

1 Introduction

Patients with multiple health conditions, or co-morbidities, pose significant challenges for modern healthcare. These patients make up 71% of total healthcare spending in America and 93% of Medicare spending.[1] To effectively treat patients with multiple health conditions, healthcare providers must consider if there is a causal relationship between the conditions and, if so, what mechanism underlies this relationship. Consider the hypotheses that obesity causes type 2 diabetes. Some treatments for type 2 diabetes cause weight gain (e.g., thiazolidinediones) potentially leading a vicious cycle of increasing insulin resistance (Kenkre, Tan, and Bloom 2013). To support healthcare providers, comorbidity researchers use electronic medical records to identify statistical relationships between diseases. Given a statistical relationship, it is necessary to understand if it occurs by either chance/sampling bias or through a causal mechanism. Figure 1 shows some of the different ways in which diseases may be causally connected.

Simply establishing a causal connection is insufficient. Effective treatment requires models of the underlying conditions and their interactions. Consider the hypothesized relationship between obesity and diabetes shown in Figure 2 (Liebman 2010). Under this model, increased levels of cortisol are the result

Figure 1: Possible causal relationships between diseases



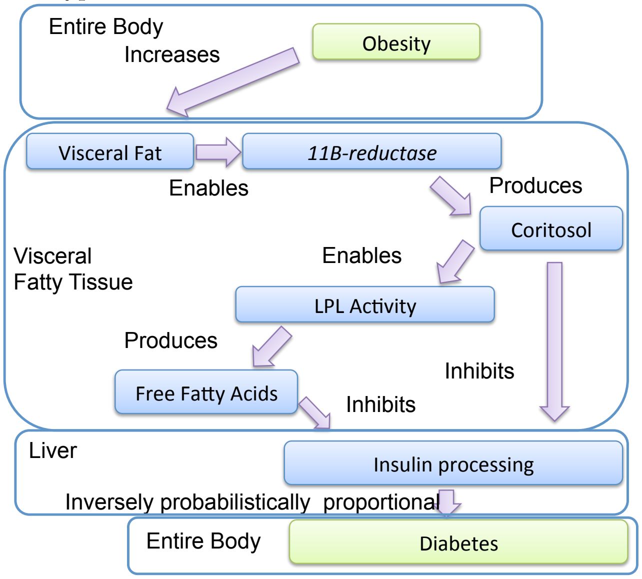
of increased 11B-reductase activity in visceral fatty tissue. The increased cortisol is transported from the fatty tissue to the liver through the portal vein. In the liver, cortisol promotes insulin resistance both directly and through the production of additional free fatty acids. Liver insulin resistance may spread throughout the body resulting in the patient having type 2 diabetes. Under this model, surgical treatments that remove visceral fat and drugs that regulate 11B-reductase activity or the level of cortisol will be effective treatments to break the link, but treatments that regulate the free fatty acid level in the liver will be insufficient.

While this explanatory model is an example of direct causation in Figure 1, when combined with the model of diabetes medicine causing weigh gain, the relationship is one of mutual causation. Simply identifying relationships between diseases with arrows and perhaps weights misses significant opportunities to support healthcare providers. We argue that representations of the explanatory models used by healthcare workers would enable new tools that could improve health outcomes.

Automatically constructing models to support these inferences is an exciting problem. The qualitative reasoning (QR) community's focus on understanding

[1] `http://www.hhs.gov/ash/initiatives/mcc/final-whcoa-mcc-slides-remediated.pdf`

Figure 2: Possible causal mechanism between obesity and type 2 diabetes.



the modeling process places it in a unique position to address this problem. QR researchers avoid ad hoc modeling in favor of reusable compositional models (Falkenhainer and Forbus 1991). The explicit representation of views enables component models referring to different levels of abstraction to be included in a single model. Furthermore, QR has developed rich models of causal reasoning that capture rules of thumb, probabilistic associations and well-understood mathematical relationships (Forbus 1984)(Weld and de Kleer 1989)(Kuipers 1994). While these features are essential for formulating the explanatory models identified in this paper, automatically creating them will require broadening established qualitative reasoning theories. The results of such an endeavor would not only transform the model-based reasoning community but also have significant impacts on medical research and practice.

This paper analyzes explanatory models found in literature to identify their properties and articulate the challenges. We make no claim as to the validity of these models, but include them to understand how practioners reason and communicate about comorbidities.

2 Example Comorbidity Explanations

In addition to the obesity and diabetes relationship described in Figure 2, we present three more explanatory models of comorbidities that will be referenced throughout the rest of this paper.

2.1 Autism and Asthma

Autism is frequently diagnosed before asthma. This could lead one to consider a direct causal relationship from autism to asthma. On the contrary, one proposed causal mechanism indicates that there might be common cause relationship between the conditions (Gidaya et al. 2016). In particular, Beta-2-adrenergic receptor inhalers (B2AR) are treatments for asthma. The use of B2AR during pregnancy has been associated with increased risk of autism developing in the child. Furthermore, asthma has a hereditary component resulting in an increased likelihood of children born to mothers using B2AR during pregnancy developing asthma and autism.

2.2 Diabetes and Lower Leg Amputation

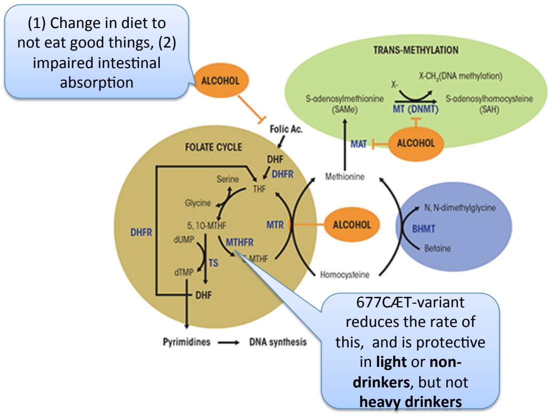
Given correlation between diabetes and lower leg amputation, Mayfield *et al.* (Mayfield et al. 1998) explore the likelihood of different mechanisms and how they affect treatment decisions. Here, we discuss a subset of the potential explanations involving the altered biomechanics of the patient. The following alterations in biomechanics can lead to ulcers and other lower leg trauma for which amputation is a treatment:

- Diabetics have increased body mass putting additional strain on the lower extremities.
- Lower leg trauma caused by limited joint mobility resulting from bone deformities and soft tissue damage.
  - Diabetes leads to bone deformities via motor neuropathy, the failure of neurons to communicate with certain muscles.
  - Diabetics have changes in their skin due to glycosylation, a metabolic process affected by diabetes, in soft tissue cells. These changes result in less pliable skin that is more prone to breaking.
- Neuropathy, a common symptom of diabetes, may prevent people from changing their gait as damage accumulates. Damage may be identified by temperature increase that would be perceptible to a clinician.
  - Peripheral vascular disease, which is frequently associated with diabetes, may lower foot temperature.
  - Neuropathy, a symptom of diabetes, may raise foot temperature.

2.3 Alcoholism and Cancer

Boffeta and Hashibe discuss the causal associations between alcohol consumption and different kinds of cancers (Boffetta and Hashibe 2006). While the mechanism is not well understood, the authors present two possible mechanisms acting through different metabolic pathways: ethanol and folate. Ethanol metabolism occurs in two steps. First, ethanol is transformed into acetaldehyde at a rate governed by the ADH and CYP2E1 gene families. Next, acetaldehyde is transformed into acetate at a rate governed by the ALDH gene. Genetic variation in ALDH gene

Figure 3: The possible effects of alcohol on folate metabolism.



family affect the rate of ethanol reactions by up to 90x. Alcoholism leads to increased alcohol consumption which in conjunction with genetic variation can lead to increased concentrations of acetaldehyde which is a known carcinogen. For example, the ALDH2 gene significantly slows the production of acetate allowing acetaldehyde to remain in the body at higher concentrations. Furthermore, this gene has been associated with increased risk for oral and throat cancers likely due to high concentrations of alcohol at those locations.

Figure 3 illustrates the possible effects of alcohol on folate metabolism. An important function of the folate cycle is DNA synthesis. Damaging this process increases the risk of developing cancer. First, alcohol reduces the amount of folate that enters the body. This is either through the poor diet of heavy drinkers or through alcohol affecting the intestinal absorption of folate. Within the folate cycle, Alcohol inhibits the expression of the MTR gene (Platek et al. 2009). The C677T variant of the MTHFR gene reduces the rate of 5-methylenetetrahydrofolate and appears to mediate the risk of colon cancer for light and moderate drinkers, but not for heavy drinkers.

3 Domains of Knowledge

From these example explanatory models, we identified six different domains of knowledge represented: genetic and metabolic pathways, physiology, mechanical, spatial, disease, and drug. For each domain of the knowledge, we present snippets from our examples and identify existing applicable biomedical knowledge bases.

3.1 Genetic and Metabolic Pathways

Genetic and metabolic pathways describe the chemical reactions that underlie biological phenomena. 11B-reductase from the obesity causing diabetes mechanism is a metabolic reaction. Proteins are the participants of metabolic reactions and genetic variation affects the rate which these proteins are transformed. Two different metabolic pathways, collections of reactions, have been identified as the possible mechanism concerning how alcoholism causes cancer.

Biologists are encoding the knowledge necessary to represent and reason about these pathways in wikipathways[2] and reactome (Joshi-Tope et al. 2005). Given the ontological structure of these models involving processes, rates, and concentrations, it is not surprising that members of the qualitative reasoning community have built systems to reason with this knowledge (Bredeweg et al. 2012). There is also work that links diseases to pathways that bioinformatics researchers have used to explain comorbidities through associations with the same pathways (Li and Agarwal 2009).

3.2 Physiological Models

Physiological models refer to the physical structure of the body and their functions. For example, physiological models are used multiple times in the associations between alcohol and cancer. First, when discussing the ethanol metabolic pathway, the explanation focuses on changes in oral and throat cancer rates due to the their roles in ingesting alcohol. Second, when analyzing reasons for decreased folate intake, the explanation discusses decreased intestinal function.

The majority of physiological modeling efforts have focused on linking genes and proteins to physiological functions (e.g., GO_MF (Ashburner et al. 2000) and Chemical Entities of Biological Interest (Degtyarenko et al. 2008)). There are ongoing efforts to link these ontologies to tissue-level descriptions (De Bono et al. 2015).

3.3 Mechanical Models

Mechanical models produce inferences from the physical connections of the body. That is, how the body moves and how different parts respond to forces applied to them. For example, the causal model concerning limited joint mobility in diabetics contains both static and dynamic models. Dynamic models include the fact that bone deformities restrict the range of movement of joints. Static models include the fact that changes in skin due to glycosylation increase the likelihood of breaking.

At this point, we are not aware of any reusable mechanical models of the human body that are used for healthcare. Standard practice appears to be to create a mechanical model for a specific purpose (Fung 2013). For reusable models, the most closely related efforts come from safety engineering (e.g., simulation of vehicle crashes) (Vezin and Verriest 2005).

[2] www.wikipathways.org

3.4 Spatial Models

Spatial models localize phenomena and interactions. While exact spatial locations may not be necessary, the representation of containers and connections is important. In diabetes model, the connection between visceral fatty tissue and liver through the portal vein is a central part of the explanation. Furthermore, the concentrations of different proteins must be understood with respect to a container.

The Open Biology Ontologies (OBO) include relationships for containment and adjacency for all of their ontologies (Smith et al. 2007). The Biological Spatial Ontology offers extensions to define precise, relative positions within an organism (Dahdul et al. 2014).

3.5 Disease Models

These models define diseases in terms of their signs, symptoms, and transmission. Disease symptoms can then be used to create patient specific models to identify how diseases may relate to one another. For example, in the obesity causes diabetes model, it is necessary to consider that obese people have more visceral fat. Then, the model is completed with the fact that increased insulin resistance is the defining signature of type 2 diabetes. Other examples of symptoms used in our examples include the fact that alcoholism has a symptom of increased alcohol consumption and reduces the amount folate in the diet. Representing disease transmission is necessary to form the causal link between the asthmatic mother and risk of asthma in their child.

The Disease Ontology is an ontology for describing the classification of human diseases organized by etiology, or causation (Kibbe et al. 2015). Alternatively, International Statistical Classification of Diseases and Related Health Problems (ICD-10) is used by many electronic medical record systems and contains codes for diseases, signs and symptoms, and abnormalities (Organization, Organization, and others 1992).

3.6 Drug models

Drug models describe how drugs interact with the body in multiple ways. Drug models that describe how a drug affects metabolic pathways may be used to design treatments. For example, a drug regulates 11B-reductase activity can be used to mitigate the risk of diabetes resulting from obesity. At the level of medical conditions and symptoms, drug models describe what diseases or symptoms drugs are used for and what their side effects are. For example, B2AR is a drug that treats asthma and that this drug has a side effect in pregnant women that increases the risk of autism in their children.

The National Institute of Health keeps records of drug interactions and side effects and makes this data available through APIs.[3] The OBO ontologies include multiple efforts to represent drugs and their effects. The DRON ontology supports comparative effectiveness researchers studying claims data.[4] The DINTO ontology categorizes drug-drug interactions (DDIs). This includes a representation of the possible mechanisms that can lead to them (including both pharmacodynamic and pharmacokinetic DDI mechanisms) (Herrero-Zazo et al. 2015).

In this section, we identified five different domains of knowledge that appear in our example explanatory models. In the next section, we highlight how this knowledge enables different types of inferences for healthcare providers.

4 Purposes of Explanatory Models

Explanatory models of comorbidities are important because they guide treatment decisions. Simple models of the form shown in Figure 1 in which nodes represent conditions and arrows represent causal relationships are insufficient for planning treatments. To illustrate, consider the abstract case of disease A causing disease B. If a patient has both diseases, will simply treating disease A be sufficient? Perhaps, but without an explanatory model to guide treatment this causal connection is not useful. In the rest of this section, we describe how explanatory models support model validation, patient observation, and treatment using examples.

4.1 Support for Model Validation

Medical researchers begin with statistical relationships between conditions. With the introduction of electronic medical records, bioinformatics researchers have developed new tools to identify orders of magnitude more potential relationships between diseases (Li and Agarwal 2009). In the diabetes lower leg amputation example, to validate the causal link between the increased body mass from diabetes and lower leg amputation, researchers conducted experiments to measure the peak plantar pressure in diabetic people. They found that body weight only accounts for less than 14% of the variance thus weakening the importance of this causal connection. This model validation step is essential in determining treatment decisions.

4.2 Support for Clinical Decisions

Healthcare providers use explanatory models to determine patient treatment and guide the monitoring of a single condition to ensure that it does not cause other conditions. In the asthma-autism connection, this involves changing the asthma medication for potential mothers. In the diabetes-amputation connection, the explanatory model explores if lower-leg stress can be identified through changes in temperature. The purpose of this statement is to support clinicians who are monitoring diabetic patients to identify those that

[3] `https://wwwcf2.nlm.nih.gov/nlm_eresources/eresources/search_database.cfm`

[4] `http://www.obofoundry.org/ontology/dron.html`

are at risk of developing ulcers that would lead to lower leg amputation.

Given a comorbid patient, explanatory models guide the treatment process. Instead of treating each condition in isolation, it is important to identify potential interactions between them. In the obesity-diabetes case, it is important that the treatment of diabetes does not lead to weight gain as that will counteract whatever treatment is being given to obesity.

5 So What? Advanced Tools for Clinical Support

In this work, we have identified a diverse set of knowledge domains necessary to create useful explanatory models of comorbidities. In Section 3, we illustrated some of the current efforts for creating reusable model libraries in each domain. From this landscape, we see two exciting research questions:

1. What kinds of inferences are possible from these explanatory models?
2. How can explanatory models be automatically constructed?

In the previous sections, we have already discussed aspects of the first question. A subset of the inferences that can be drawn from these models includes determining intervention decisions (e.g., not recommending B2AR inhalers for pregnant mothers), guiding future experiment design and data collection (e.g., determining the strength of the causal relationship between increased body mass and lower leg amputation), and directing healthcare monitoring (e.g., importance of watching weight for people taking diabetes drugs). Further research must explore the context of these decisions and other clinical decisions made by providers.

The second question concerns model formulation (Falkenhainer and Rajamoney 1988)(Rickel and Porter 1997). Given a question, this process typically involves (1) generating a model from a domain theory and experience, (2) evaluating its utility, and (3) revising the model based on its evaluation. Steps 2 and 3 continue until the modeler is satisfied with the results. One area where current work falls short for our application concerns the representation of the evaluation criteria and how these can change during the model revision process. Instead, current approaches typically address prediction questions (e.g., "What will happen to a quantity in a particular scenario?"). The explanatory models described in this work are often exploratory in nature (e.g., "How are these two conditions related?"). Their construction is important for communication between scientists, providers, their patients, and the public.

Understanding the context and inferences that are important to healthcare providers coupled with new techniques of model formulation and revision could enable new classes of clinical support tools.

6 Discussion

In this work, we argue that the simple causal network models that are shown in Figure 1 are insufficient for clinical support. Scientists and healthcare providers create explanatory models that expand the thin arrows in Figure 1 into mechanistic explanation consisting of entities and relationships.

The qualitative reasoning (QR) community's focus on understanding the modeling process places it in a unique position to bridge the gap between the causal network models that are derivable from data and the explanatory models used by clinicians. Compositional modeling's emphasis on reusable components, or model fragments, (e.g., 11B-reductase reaction is studied in isolation from diabetes patients) and explicit representation of assumptions are essential components of explanatory models. QR has developed rich models of causal reasoning that capture world knowledge (e.g., treatments of a pregnant mother affect the fetus), probabilistic associations (e.g., increasing insulin resistance increases the likelihood of a patient exhibiting diabetes) and understood mathematical relationships (e.g., systems biology models of chemical reactions, such as 11B-reductase, use differential equations).

While these features are essential for formulating the explanatory models identified in this paper, we also illustrate additional challenges that will require extending current QR theories and research. In particular, the extension of automated model formulation and revision from prediction and system identification tasks to the open-ended problem of comorbidity explanation. Such advances could enable a radical transformation of clinical support tools signficantly improving healthcare outcomes.

Acknowledgments

This work is supported by the Xerox Innovation Group.

References

[Ashburner et al. 2000] Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. 2000. Gene ontology: tool for the unification of biology. *Nature genetics* 25(1):25–29.

[Boffetta and Hashibe 2006] Boffetta, P., and Hashibe, M. 2006. Alcohol and cancer. *The Lancet Oncology* 7:149–156.

[Bredeweg et al. 2012] Bredeweg, B.; Linnebank, F.; Beek, W.; and Liem, J. 2012. Systems thinking in biology–the case for qualitative representations and reasoning1. In *Proceedings of the QR12 workshop on deep knowledge representation challenge*. Citeseer.

[Dahdul et al. 2014] Dahdul, W. M.; Cui, H.; Mabee, P. M.; Mungall, C. J.; Osumi-Sutherland, D.; Walls, R. L.; and Haendel, M. A. 2014. Nose to tail, roots to shoots: spatial descriptors for phenotypic diversity in the biological spatial ontology. *Journal of biomedical semantics* 5(1):1.

[De Bono et al. 2015] De Bono, B.; Safaei, S.; Grenon, P.; Nickerson, D. P.; Alexander, S.; Helvensteijn, M.; Kok, J. N.; Kokash, N.; Wu, A.; Yu, T.; et al. 2015. The open physiology workflow: modeling processes over physiology circuitboards of interoperable tissue units. *Frontiers in physiology* 6.

[Degtyarenko et al. 2008] Degtyarenko, K.; De Matos, P.; Ennis, M.; Hastings, J.; Zbinden, M.; McNaught, A.; Alcántara, R.; Darsow, M.; Guedj, M.; and Ashburner, M. 2008. Chebi: a database and ontology for chemical entities of biological interest. *Nucleic acids research* 36(suppl 1):D344–D350.

[Falkenhainer and Forbus 1991] Falkenhainer, B., and Forbus, K. D. 1991. Cmpositional modeling: finding the right model for the job. *Artificial intelligence* 51(1):95–143.

[Falkenhainer and Rajamoney 1988] Falkenhainer, B., and Rajamoney, S. 1988. The interdependencies of theory formation, revision, and experimentation. In *Proceedings of the 5th International Conference on Machine Learning*, 353–366.

[Forbus 1984] Forbus, K. D. 1984. Qualitative process theory. *Artificial intelligence* 24(1):85–168.

[Fung 2013] Fung, Y.-c. 2013. *Biomechanics: mechanical properties of living tissues.* Springer Science & Business Media.

[Gidaya et al. 2016] Gidaya, N. B.; Lee, B. K.; Burstyn, I.; Michael, Y.; Newschaffer, C. J.; and Mortensen, E. L. 2016. In utero exposure to -2-adrenergic receptor agonist drugs and risk for autism spectrum disorders. *Pediatrics*.

[Herrero-Zazo et al. 2015] Herrero-Zazo, M.; Segura-Bedmar, I.; Hastings, J.; and Martínez, P. 2015. Dinto: using owl ontologies and swrl rules to infer drug–drug interactions and their mechanisms. *Journal of chemical information and modeling* 55(8):1698–1707.

[Joshi-Tope et al. 2005] Joshi-Tope, G.; Gillespie, M.; Vastrik, I.; D'Eustachio, P.; Schmidt, E.; de Bono, B.; Jassal, B.; Gopinath, G.; Wu, G.; Matthews, L.; et al. 2005. Reactome: a knowledgebase of biological pathways. *Nucleic acids research* 33(suppl 1):D428–D432.

[Kenkre, Tan, and Bloom 2013] Kenkre, J.; Tan, T.; and Bloom, S. 2013. Treating the obese diabetic. *Expert review of clinical pharmacology* 6(2):171–183.

[Kibbe et al. 2015] Kibbe, W. A.; Arze, C.; Felix, V.; Mitraka, E.; Bolton, E.; Fu, G.; Mungall, C. J.; Binder, J. X.; Malone, J.; Vasant, D.; et al. 2015. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic acids research* 43(D1):D1071–D1078.

[Klenk et al. 2012] Klenk, M.; de Kleer, J.; Bobrow, D. G.; Yoon, S.; Hanley, J.; and Janssen, B. 2012. Guiding and verifying early design using qualitative simulation. In *Proceedings of the ASME 2012 IDETC and CIE.*

[Kuipers 1994] Kuipers, B. 1994. *Qualitative reasoning: modeling and simulation with incomplete knowledge.* Cambridge, MA, USA: MIT Press.

[Li and Agarwal 2009] Li, Y., and Agarwal, P. 2009. A pathway-based view of human diseases and disease relationships. *PloS one* 4(2):e4346.

[Liebman 2010] Liebman, T. 2010. The role of cortisol and abdominal obesity in the epidemic of type 2 diabetes. *Undergraduate Research Journal for the Human Sciences* 9.

[Mayfield et al. 1998] Mayfield, J. A.; Reiber, G. E.; Sanders, L. J.; Janisse, D.; and Pogach, L. M. 1998. Preventive foot care in people with diabetes. *Diabetes Care* 21(12).

[Organization, Organization, and others 1992] Organization, W. H.; Organization, W. H.; et al. 1992. Icd-10. *International statistical classification of diseases and related health problems (10th edition). Geneva, SR: World Health Organization (WHO).*

[Platek et al. 2009] Platek, M. E.; Shields, P. G.; Marian, C.; McCann, S. E.; Bonner, M. R.; Nie, J.; Ambrosone, C. B.; Millen, A. E.; Ochs-Balcom, H. M.; Quick, S. K.; et al. 2009. Alcohol consumption and genetic variation in mthfr and mtr in relation to breast cancer risk. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 18(9):2453.

[Rickel and Porter 1997] Rickel, J., and Porter, B. 1997. Automated modeling of complex systems to answer prediction questions. *Artificial Intelligence* 93(1):201–260.

[Smith et al. 2007] Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L. J.; Eilbeck, K.; Ireland, A.; Mungall, C. J.; et al. 2007. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* 25(11):1251–1255.

[Vezin and Verriest 2005] Vezin, P., and Verriest, J. P. 2005. Development of a set of numerical human models for safety. In *19th International Technical Conference on the Enhanced Safety of Vehicles, Washington DC*, 6–9.

[Weld and de Kleer 1989] Weld, D., and de Kleer, J. 1989. *Readings in Qualitative Reasoning about Physical Systems.* Morgan Kaufmann.

From Qualitative Absolute Order-of-Magnitude to the Extended Set of Hesitant Fuzzy Linguistic Term Sets

Jordi Montserrat-Adell* and **Mónica Sánchez** and **Francisco Javier Ruiz**
UPC-BarcelonaTech, Barcelona, Spain
{jordi.montserrat-adell,monica.sanchez,francisco.javier.ruiz}@upc.edu

Núria Agell
Esade - Universitat Ramon Llull, Barcelona, Spain
nuria.agell@esade.edu

Abstract

Hesitant fuzzy linguistic term sets were introduced to grasp the uncertainty existing in human reasoning. In this paper, inspired by absolute order-of-magnitude qualitative reasoning techniques, an extension of the set of hesitant fuzzy linguistic term sets is presented to capture differences between non-compatible preferences. In addition, an order relation and two closed operations over this set are also introduced to provide a lattice structure to the extended set of hesitant fuzzy linguistic term sets. Based on this lattice structure a distance between hesitant fuzzy linguistic term sets is defined.

Keywords: Linguistic modeling, Group decision making, Uncertainty and Fuzzy Reasoning, Hesitant fuzzy linguistic term sets.

Introduction

Techniques based on order-of-magnitude qualitative reasoning have provided theoretical models to deal with non-numeric variables (Agell et al. 2012; Forbus 1996; Travé-Massuyès and Dague 2003; Travé-Massuyès et al. 2005). One of the advantages of qualitative reasoning is its capability to tackle problems in such a way that the principle of relevance is preserved; that is to say, each variable involved in a real problem must be valued at the precision level required. Order-of-magnitude models are among the essential theoretical tools available for qualitative reasoning about real systems. They aim to capture order-of-magnitude commonsense inferences, as used by human beings in the real world.

In addition, different approaches involving linguistic assessments have been introduced in the fuzzy sets literature to deal with the impreciseness and uncertainty connate with human reasoning (Espinilla, Liu, and Martínez 2011; Herrera, Herrera-Viedma, and Martínez 2008; Herrera-Viedma, Herrera, and Chiclana 2002; Parreiras et al. 2010; Tang and Zheng 2006). Additionally, different extensions of fuzzy sets have been presented to give more realistic assessments when uncertainty increases (Deschrijver and Kerres 2003; Greenfield and Chiclana 2013; Rodríguez, Martínez, and Herrera 2012). To describe human reasoning with different levels of precision similarly to absolute order-of-magnitude qualitative models, Hesitant Fuzzy Linguistic Term Sets (HFLTSs) were introduced in (Rodríguez, Martínez, and Herrera 2012) and a lattice structure is provided to the set of HFLTSs in (Montserrat-Adell et al.).

In this paper, inspired by previous woks over absolute order-of-magnitude qualitative models (Agell et al. 2012; Prats et al. 2014), we present an extension of the set of HFLTSs, $\overline{\mathcal{H}_\mathcal{S}}$, based on an equivalence relation on the usual set of HFLTSs. This enables us to establish differences between non-compatible HFLTSs. An order relation and two closed operation over this set are also introduced to define a new lattice structure in $\overline{\mathcal{H}_\mathcal{S}}$. A distance between HFLTSs is defined based on the lattice of $\overline{\mathcal{H}_\mathcal{S}}$.

This structures may be very useful in management situations such as marketing or human resources problems, where order-of-magnitude labels are used to assess. For instance, a common linguistic scale in the human resources field is: outstanding, exceeds expectations, meets expectations, below expectations and unsatisfactory.

The rest of this paper is organized as follows: first, Section 1 presents a brief review of HFLTSs and its lattice structure. The lattice of the extended set of HFLTSs is introduced in Section 2. In Section 3, the distances between HFLTSs are defined. Lastly, Section 4 contains the main conclusions and lines of future research.

1 The Lattice of Hesitant Fuzzy Linguistic Term Sets

In this section we present a brief review of some concepts about HFLTSs already presented in the literature that are used throughout this paper (Montserrat-Adell et al. ; Rodríguez, Martínez, and Herrera 2012).

From here on, let $\mathcal{S}$ denote a finite total ordered set of linguistic terms, $\mathcal{S} = \{a_1, \ldots, a_n\}$ with $a_1 < \cdots < a_n$.

Definition 1. (Rodríguez, Martínez, and Herrera 2012) A *hesitant fuzzy linguistic term set (HFLTS)* over $\mathcal{S}$ is a subset of consecutive linguistic terms of S, i.e. $\{x \in \mathcal{S} \mid a_i \leq x \leq a_j\}$, for some $i, j \in \{1, \ldots, n\}$ with $i \leq j$.

*Corresponding author

The HFLTS $\mathcal{S}$ is called the *full HFLTS*. Moreover, the empty set $\{\} = \emptyset$ is also considered as a HFLTS and it is called the *empty HFLTS*.

For the rest of this paper, the non-empty HFLTS, $H = \{x \in \mathcal{S} \mid a_i \leq x \leq a_j\}$, is denoted by $[a_i, a_j]$. Note that, if $j = i$, the HFLTS $[a_i, a_i]$ is expressed as the singleton $\{a_i\}$.

The set of all the possible HFLTSs over $\mathcal{S}$ is denoted by $\mathcal{H}_\mathcal{S}$, being $\mathcal{H}^*_\mathcal{S} = \mathcal{H}_\mathcal{S} - \{\emptyset\}$ the set of all the non-empty HFLTSs. This set is provided with a lattice structure in (Montserrat-Adell et al.) with the two following operations: on the one hand, the *connected union of two HFLTSs,* $\sqcup$, which is defined as the least element of $\mathcal{H}_\mathcal{S}$, based on the subset inclusion relation $\subseteq$, that contains both HFLTSs, and on the other hand, the intersection of HFLTSs, $\cap$, which is defined as the usual intersection of sets. Notice that the usual union of sets cannot be considered given that it may not result a HFLTS. In addition, the reason of including the empty HFLTS in $\mathcal{H}_\mathcal{S}$ is to make the intersection of HFLTSs a closed operation in $\mathcal{H}_\mathcal{S}$.

For the sake of comprehensiveness, let us introduce the following example that is used throughout all this paper to depict all the concepts defined.

Example 1. Given the common set of linguistic labels, used in performance appraisal processes for human resources, $\mathcal{S} = \{a_1, a_2, a_3, a_4, a_5\}$, being a_1 = *unsatisfactory*, a_2 = *below expectations*, a_3 = *meets expectations*, a_4 = *exceeds expectations*, a_5 = *outstanding*, possible linguistic assessments and their corresponding HFLTSs by means of $\mathcal{S}$ would be:

Assessments	*HFLTSs*
A = *"below or meets expectations"*	$H_A = [a_2, a_3]$
B = *"below expectations"*	$H_B = \{a_2\}$
C = *"above meets expectations"*	$H_C = [a_4, a_5]$
D = *"below meets expectations"*	$H_D = [a_1, a_2]$
E = *"not outstanding"*	$H_E = [a_1, a_4]$

2 The Extended Lattice of Hesitant Fuzzy Linguistic Term Sets

With the aim of describing differences between couples of HFLTSs with empty intersections, an extension of the intersection of HFLTSs is presented in this section, resulting their intersection if it is not empty or a new element that we will call *negative HFLTS* related to the rift, or gap, between them if their intersection is empty. In order to present said extension of the intersection between HFLTSs, we first need to introduce the mathematical structure that allows us to define it as a closed operation. To this end, we define the extended set of HFLTSs in an analogous way to how integer numbers are defined based on an equivalence relation on the natural numbers. To do so, we first present some needed concepts:

Definition 2. Given two non-empty HFLTSs, $H_1, H_2 \in \mathcal{H}^*_\mathcal{S}$, we define:

(a) The *gap between* H_1 *and* H_2 as:

$$gap(H_1, H_2) = (H_1 \sqcup H_2) \cap \overline{H_1} \cap \overline{H_2},$$

where $\overline{H}$ represents the complement of H.

(b) H_1 and H_2 are *consecutive* if and only if $H_1 \cap H_2 = \emptyset$ and $gap(H_1, H_2) = \emptyset$.

Proposition 1. Given two non-empty HFLTSs, $H_1, H_2 \in \mathcal{H}^*_\mathcal{S}$, the following properties are met:

1. $gap(H_1, H_2) = gap(H_2, H_1)$.
2. If $H_1 \subseteq H_2$, $gap(H_1, H_2) = \emptyset$.
3. If $H_1 \cap H_2 \neq \emptyset$, $gap(H_1, H_2) = \emptyset$.
4. If $H_1 \cap H_2 = \emptyset$, $gap(H_1, H_2) \neq \emptyset$ or H_1 and H_2 are consecutive.
5. If H_1 and H_2 are consecutive, there exist $j \in \{2, \dots, n-1\}$, $i \in \{1, \dots, j\}$ and $k \in \{j+1, \dots, n\}$, such that $H_1 = [a_i, a_j]$ and $H_2 = [a_{j+1}, a_k]$ or $H_2 = [a_i, a_j]$ and $H_2 = [a_{j+1}, a_k]$.

Proof. The proof is straightforward. □

Note that neither $[a_1, a_j]$ nor $[a_i, a_n]$ can ever be the result of the *gap* between two HFLTSs for any i and for any j.

Notation. Given two consecutive HFLTSs, $H_1 = [a_i, a_j]$ and $H_2 = [a_{j+1}, a_k]$, then $\{a_j\}$ and $\{a_{j+1}\}$ are named as the linguistic terms that provide the consecutiveness of H_1, H_2.

Example 2. Following Example 1, $gap(H_B, H_C) = \{a_3\}$, while the HFLTSs H_A and H_C are consecutive and their consecutiveness is given by $\{a_3\}$ and $\{a_4\}$.

Definition 3. Given two pairs of non-empty HFLTSs, (H_1, H_2) and (H_3, H_4), the *equivalence relation* $\sim$*,* is defined as:

$$(H_1, H_2) \sim (H_3, H_4)$$

$$\Updownarrow$$

$$\left\{ \begin{array}{c} H_1 \cap H_2 = H_3 \cap H_4 \neq \emptyset \\ \vee \\ gap(H_1, H_2) = gap(H_3, H_4) \neq \emptyset \\ \vee \\ \text{both pairs are consecutive and} \\ \text{their consecutiveness is provided} \\ \text{by the same linguistic terms} \end{array} \right.$$

It can be easily seen that $\sim$ relates couples of non-empty HFLTSs with the same intersection if they are compatible, with consecutiveness provided by the same linguistic terms if they are consecutive and with the same *gap* between them in the case that they are neither compatible nor consecutive.

Example 3. Following Example 1, the pairs of HFLTSs (H_A, H_B) and (H_A, H_D) are related according to $\sim$ given that they have the same intersection, $\{a_2\}$. Additionally, $(H_C, H_B) \sim (H_C, H_D)$ since they have the same *gap* between them, $\{a_3\}$.

Applying this equivalence relation over the set of all the pairs of non-empty HFLTSs, we get the quotient set $(\mathcal{H}^*_\mathcal{S})^2/\sim$, whose equivalence classes can be labeled as:

- $[a_i, a_j]$ for the class of all pairs of compatible non-empty HFLTSs with intersection $[a_i, a_j]$, for all $i, j = 1, \dots, n$ with $i \leq j$.
- $-[a_i, a_j]$ for the class of all pairs of incompatible non-empty HFLTSs whose *gap* is $[a_i, a_j]$, for all $i, j = 2, \dots, n-1$ with $i \leq j$.
- α_i for the class of all pairs of consecutive non-empty HFLTSs whose consecutiveness is provided by $\{a_i\}$ and $\{a_{i+1}\}$, for all $i = 1, \dots, n-1$.

For completeness and symmetry reasons, $(\mathcal{H}_{\mathcal{S}}^*)^2/\sim$ is represented as shown in Figure 1 and stated in the next definition.

Example 4. Subsequent to this labeling, and following Example 1, the pair (H_C, H_B) belongs to the class $-\{a_3\}$ and so does the pair (H_C, H_D). The pair (H_C, H_A) belongs to the class α_3 and the pair (H_C, H_E) belongs to the class $\{a_4\}$.

Definition 4. Given a set of ordered linguistic term sets $\mathcal{S} = \{a_1, \dots, a_n\}$, the *extended set of HFLTSs, $\overline{\mathcal{H}_{\mathcal{S}}}$,* is defined as:

$$\overline{\mathcal{H}_{\mathcal{S}}} = (-\mathcal{H}_{\mathcal{S}}^*) \cup \mathcal{A} \cup \mathcal{H}_{\mathcal{S}}^*,$$

where $-\mathcal{H}_{\mathcal{S}}^* = \{-H \mid H \in \mathcal{H}_{\mathcal{S}}^*\}$ and $\mathcal{A} = \{\alpha_0, \dots, \alpha_n\}$.

In addition, by analogy with real numbers $-\mathcal{H}_{\mathcal{S}}^*$ is called the *set of negative HFLTSs,* $\mathcal{A}$ is called the *set of zero HFLTSs,* and, from now on, $\mathcal{H}_{\mathcal{S}}^*$ is called the *set positive HFLTSs.*

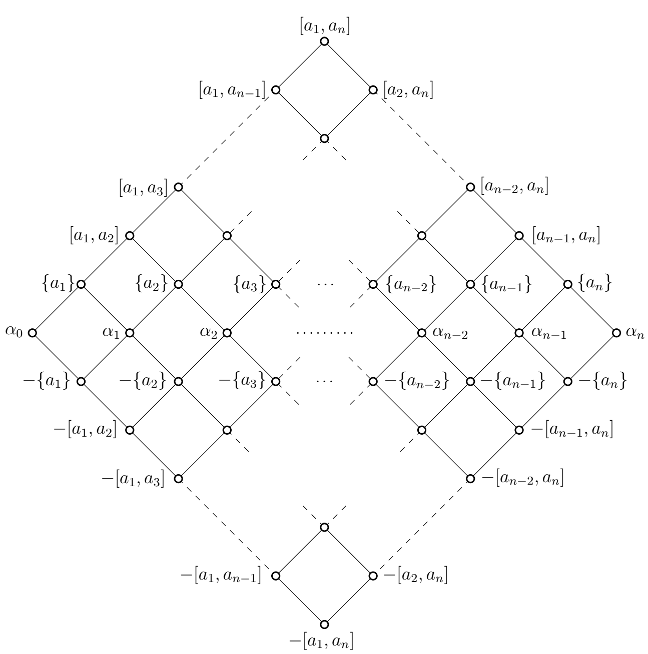


Figure 1: Graph of the extended set of HFLTSs.

Note that HFLTSs can be characterized by couples of zero HFLTSs. This leads us to introduce a new notation for HFLTSs:

Notation. Given a HFLTS, $H \in \overline{\mathcal{H}_{\mathcal{S}}}$, it can be expressed as $H = \langle \alpha_i, \alpha_j \rangle$, where the first zero HFLTS identifies the bottom left to top right diagonal and the second one identifies the top left to bottom right diagonal. Thus, $\langle \alpha_i, \alpha_j \rangle$ corresponds with $[a_{i+1}, a_j]$ if $i < j$, with $-[a_{i+1}, a_j]$ if $i > j$ and α_i if $i = j$.

This notation is used in the following definition that we present in order to latter introduce an order relation within $\overline{\mathcal{H}_{\mathcal{S}}}$.

Definition 5. Given $H \in \overline{\mathcal{H}_{\mathcal{S}}}$ described by $\langle \alpha_i, \alpha_j \rangle$ the *coverage of H* is defined as:

$$cov(H) = \{\langle \alpha_{i'}, \alpha_{j'} \rangle \in \overline{\mathcal{H}_{\mathcal{S}}} \mid i' \geq i \wedge j' \leq j\}.$$

Example 5. The coverage of H_A from Example 1 can be seen in Figure 2.

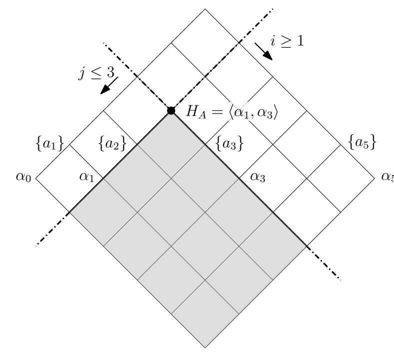


Figure 2: Coverage of H_A.

The concept of coverage of a HFLTS enables us to define the *extended inclusion relation* between elements of $\overline{\mathcal{H}_{\mathcal{S}}}$.

Definition 6. The *extended inclusion relation in $\overline{\mathcal{H}_{\mathcal{S}}}$, $\preccurlyeq$,* is defined as:

$$\forall H_1, H_2 \in \overline{\mathcal{H}_{\mathcal{S}}}, \quad H_1 \preccurlyeq H_2 \iff H_1 \in cov(H_2).$$

Note that, restricting to only the positive HFLTSs, the extended inclusion relation coincides with the usual subset inclusion relation. According to this relation in $\overline{\mathcal{H}_{\mathcal{S}}}$, we can define the *extended connected union* and the *extended intersection* as closed operations within the set $\overline{\mathcal{H}_{\mathcal{S}}}$ as follows:

Definition 7. Given $H_1, H_2 \in \overline{\mathcal{H}_{\mathcal{S}}}$, the *extended connected union of H_1 and H_2, $H_1 \sqcup H_2$,* is defined as the least element that contains H_1 and H_2, according to the extended inclusion relation.

Definition 8. Given $H_1, H_2 \in \overline{\mathcal{H}_{\mathcal{S}}}$, the *extended intersection of H_1 and H_2, $H_1 \sqcap H_2$,* is defined as the largest element being contained in H_1 and H_2, according to the extended inclusion relation.

It is straightforward to see that the extended connected union of two positive HFLTSs coincides with the connected union presented in (Montserrat-Adell et al.). This justifies the use of the same symbol. About the extended intersection of two positive HFLTSs, it results the usual intersection of sets if they overlap and the *gap* between them if they do not overlap. Notice that the empty HFLTS is not needed to make the extended intersection a closed operation in $\overline{\mathcal{H}_{\mathcal{S}}}$.

Proposition 2. Given two non-empty HFLTSs, $H_1, H_2 \in \mathcal{H}_{\mathcal{S}}^*$, if $H_1 \preccurlyeq H_2$, then $H_1 \sqcup H_2 = H_2$ and $H_1 \sqcap H_2 = H_1$.

Proof. The proof is straightforward. □

Example 6. Figure 3 provides an example with the extended connected union and the extended intersection of H_B and H_C and of H_A and H_E from Example 1: $H_B \sqcup H_C = [a_2, a_5]$, $H_B \sqcap H_C = -\{a_3\}$, $H_A \sqcup H_E = H_E$ and $H_A \sqcap H_E = H_A$. Note that $H_B \cup H_C = \{a_2, a_4, a_5\}$ is not a HFLTS.

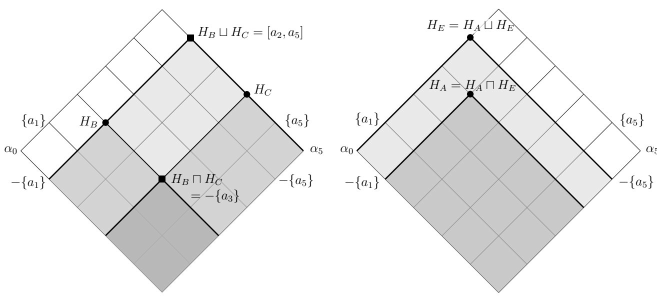


Figure 3: $\sqcup$ and $\sqcap$ of HFLTSs.

Proposition 3. $(\overline{\mathcal{H}_{\mathcal{S}}}, \sqcup, \sqcap)$ is a distributive lattice.

Proof. According to their respective definitions, both operations, $\sqcup$ and $\sqcap$, are trivially commutative and idempotent.

The associative property of $\sqcup$ is met since $(H_1 \sqcup H_2) \sqcup H_3 = H_1 \sqcup (H_2 \sqcup H_3)$ given that both parts equal the least element that contains H_1, H_2 and H_3. About the associativeness of $\sqcap$, $(H_1 \sqcap H_2) \sqcap H_3 = H_1 \sqcap (H_2 \sqcap H_3)$ given that in both cases it results the largest element contained in H_1, H_2 and H_3.

Finally, the absorption laws are satisfied given that: on the one hand $H_1 \sqcup (H_1 \sqcap H_2) = H_1$ given that $H_1 \sqcap H_2 \preccurlyeq H_1$ and on the other hand $H_1 \sqcap (H_1 \sqcup H_2) = H_1$ given that $H_1 \preccurlyeq H_1 \sqcup H_2$.

Furthermore, the lattice $(\overline{\mathcal{H}_{\mathcal{S}}}, \sqcup, \sqcap)$ is distributive given that none of its sublattices are isomorphic to the diamond lattice, M_3, or the pentagon lattice, N_5. □

3 A Distance between Hesitant Fuzzy Linguistic Term Sets

In order to define a distance between HFLTSs, we introduce a generalization of the concept of cardinal of a positive HFLTS to all the elements of the extended set of HFLTSs.

Definition 9. Given $H \in \overline{\mathcal{H}_{\mathcal{S}}}$, the *width of* H is defined as:

$$\mathcal{W}(H) = \begin{cases} card(H) & if \quad H \in \mathcal{H}_{\mathcal{S}}^*, \\ 0 & if \quad H \in \mathcal{A}, \\ -card(-H) & if \quad H \in (-\mathcal{H}_{\mathcal{S}}^*). \end{cases}$$

Note that the width of a HFLTS could be related as well with the height on the graph of $\overline{\mathcal{H}_{\mathcal{S}}}$, associating the zero HFLTSs with height 0, the positive HFLTSs with positive heights and the negative HFLTSs with negative values of heights as shown in Figure 4.

Proposition 4. $D(H_1, H_2) = \mathcal{W}(H_1 \sqcup H_2) - \mathcal{W}(H_1 \sqcap H_2)$ provides a distance in the lattice $(\overline{\mathcal{H}_{\mathcal{S}}}, \sqcup, \sqcap)$.

Proof. $D(H_1, H_2)$ defines a distance given that it is equivalent to the geodesic distance in the graph $\overline{\mathcal{H}_{\mathcal{S}}}$. The geodesic distance between H_1 and H_2 is the length of the shortest path to go from H_1 to H_2. Due to the fact that $H_1 \sqcap H_2 \preccurlyeq H_1 \sqcup H_2$, $\mathcal{W}(H_1 \sqcup H_2) - \mathcal{W}(H_1 \sqcap H_2)$ is the length of the minimum path between $H_1 \sqcup H_2$ and $H_1 \sqcap H_2$. Thus, we have to check that the length of the shortest path between $H_1 \sqcup H_2$ and $H_1 \sqcap H_2$ coincides with the length of the shortest path between H_1 and H_2.

If one of them belong to the coverage of the other one, let us suppose that $H_1 \preccurlyeq H_2$, then $H_1 \sqcup H_2 = H_2$ and $H_1 \sqcap H_2 = H_1$ and the foregoing assertion becomes obvious. If not, H_1, $H_1 \sqcup H_2$, H_2 and $H_1 \sqcap H_2$ define a parallelogram on the graph. Two consecutive sides of this parallelogram define the shortest path between $H_1 \sqcup H_2$ and $H_1 \sqcap H_2$ while two other consecutive sides of the same parallelogram define the shortest path between H_1 and H_2. Thus, the assertion becomes true as well. □

Proposition 5. Given two HFLTSs, $H_1, H_2 \in \overline{\mathcal{H}_{\mathcal{S}}}$, then $0 \leq D(H_1, H_2) \leq 2n$. If, in addition, $H_1, H_2 \in \mathcal{H}_{\mathcal{S}}^*$, then $0 \leq D(H_1, H_2) \leq 2n - 2$.

Proof. For the lower bound, notice that since $H_1 \sqcap H_2 \subseteq H_1 \sqcup H_2$, then $\mathcal{W}(H_1 \sqcap H_2) \leq \mathcal{W}(H_1 \sqcup H_2)$, and therefore $D(H_1, H_2) \geq 0$.

For the upper bound, if $H_1, H_2 \in \overline{\mathcal{H}_{\mathcal{S}}}$, then, the most distant pair is α_0 and α_n. Then,

$$\mathcal{W}(\alpha_0 \sqcup \alpha_n) - \mathcal{W}(\alpha_0 \sqcap \alpha_n) =$$

$$\mathcal{W}([a_1, a_n]) - \mathcal{W}(-[a_1, a_n]) =$$

$$n - (-n) = 2n.$$

If $H_1, H_2 \in \mathcal{H}_{\mathcal{S}}^*$, then, the most distant pair is $\{a_1\}$ and $\{a_n\}$. Then,

$$\mathcal{W}(\{a_1\} \sqcup \{a_n\}) - \mathcal{W}(\{a_1\} \sqcap \{a_n\}) =$$

$$\mathcal{W}([a_1, a_n]) - \mathcal{W}(-[a_2, a_{n-1}]) =$$

$$n - (-(n-2)) = 2n - 2.$$

□

Notice that for positive HFLTSs, $D(H_1, H_2)$ coincides with the distance $D_2(H_1, H_2)$ introduced in (Montserrat-Adell et al.). Additionally, in this case, the distance presented can also be calculated as $D([a_i, a_j], [a_{i'}, a_{j'}]) = |i - i'| + |j - j'|$.

Example 7. Figure 4 shows the width of the extended connected union and the extended intersection of H_B and H_C from Example 1. According to these results, $D(H_B, H_C) = \mathcal{W}(H_B \sqcup H_C) - \mathcal{W}(H_B \sqcap H_C) = 4 - (-1) = 5$.

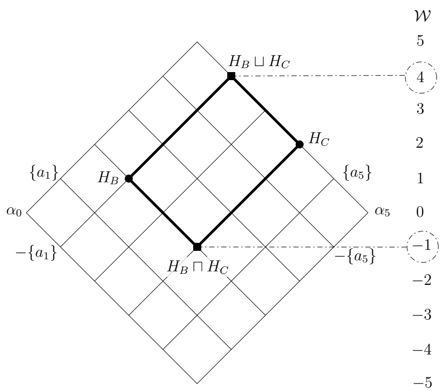


Figure 4: Distance between HFLTSs.

4 Conclusions and future research

This paper presents, inspired by previous works over absolute order-of-magnitude qualitative models, an extension of the set of Hesitant Fuzzy Linguistic Term Sets by introducing the concepts of negative and zero HFLTSs to capture differences between pairs of non-compatible HFLTSs. This extension enables the introduction of a new operation studying the intersection and the gap between HFLTSs at the same time. This operation is used to define a distance between HFLTSs that allows us to analyze differences between the assessments given by a group of decision makers.

There is, nowadays, a wide range of areas of application for distances between linguistic assessments, from managerial to medical or engineering. Future research is focused in two main directions. First, the study of the consensus level of the total group assessments to analyze the agreement or disagreement within the group. And secondly, a real case study will be performed in the marketing research area to examine consensus and heterogeneities in consumers' preferences.

References

Agell, N.; Sánchez, M.; Prats, F.; and Roselló, L. 2012. Ranking multi-attribute alternatives on the basis of linguistic labels in group decisions. *Information Sciences* 209:49–60.

Deschrijver, G., and Kerres, E. E. 2003. On the relationship between some extensions of fuzzy set theory. *Fuzzy Sets and Systems* 133:227–235.

Espinilla, M.; Liu, J.; and Martínez, L. 2011. An extended hierarchical linguistic model for decision-making problems. *Computational Intelligence* 27(3):489–512.

Forbus, K. D. 1996. *Qualitative Reasoning*. CRC Handbook of Computer Science and Engineering, CRC Press,.

Greenfield, S., and Chiclana, F. 2013. Accuracy and complexity evaluation of defuzzification strategies for the discretised interval type-2 fuzzy set. *International Journal of Approximate Reasoning* 54(8):1013–1033.

Herrera, F.; Herrera-Viedma, E.; and Martínez, L. 2008. A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems* 16(2):354–370.

Herrera-Viedma, E.; Herrera, F.; and Chiclana, F. 2002. A consensus model for multiperson decision making with different preference structures. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 32:394–402.

Montserrat-Adell, J.; Agell, N.; Sánchez, M.; Prats, F.; and Ruiz, F. J. Modeling group assessments by means of hesitant fuzzy linguistic term sets. (To appear in Journal of Applied Logic).

Parreiras, R. O.; Ekel, P. Y.; Martini, J. S. C.; and Palhares, R. M. 2010. A flexible consensus scheme for multicriteria group decision making under linguistic assessments. *Information Sciences* 180(7):1075–1089.

Prats, F.; Roselló, L.; Sánchez, M.; and Agell, N. 2014. Using l-fuzzy sets to introduce information theory into qualitative reasoning. *Fuzzy Sets and Systems* 236(1):73–90.

Rodríguez, R. M.; Martínez, L.; and Herrera, F. 2012. Hesitant fuzzy linguistic terms sets for decision making. *IEEE Transactions on Fuzzy Systems* 20(1):109–119.

Tang, Y., and Zheng, J. 2006. Linguistic modelling based on semantic similarity relation among linguistic labels. *Fuzzy Sets and Systems* 157(12):1662–1673.

Travé-Massuyès, L., and Dague, P. 2003. *Modèles et raisonnements qualitatifs*. Paris, France: Ed. Hermès.

Travé-Massuyès, L.; Prats, F.; Sanchez, M.; and Agell, N. 2005. Relative and absolute order-of-magnitude models unified. *Annals of mathematics and artificial intelligence* 45(1):323–341.

Collaborative Communication of Qualitative Spatial Perceptions for Multi-Robot Systems

Danilo H. Perico and **Paulo E. Santos** and
Reinaldo A. C. Bianchi
Centro Universitário FEI, SBC, Brazil
{dperico, psantos, rbianchi}@fei.edu.br

Ramon López de Mántaras
Artificial Intelligence Research Institute
CSIC, Campus UAB, Spain
mantaras@iiia.csic.es

Abstract

The goal of this work is to develop a collaborative communication system of spatial perceptions for vision-based multi-robot systems using qualitative spatial reasoning, where the representation of the domain is built upon the perspective of the Elevated Oriented Point Algebra ($\mathcal{EOPRA}$) and the reasoning itself is made by a combination between the Oriented Point Algebra ($\mathcal{OPRA}$) and a quantitative triangulation. The motivation of using qualitative information is to obtain a level of abstraction closer to the human categorisation of space and, also, to have a more effective way of interaction between robots and humans. Results allowed us to conclude that the method proposed is an effective way to address the high-level communication between only-robots agents or between robots and humans by using some spatial prepositions.

Introduction

Robots will soon achieve a level of electrical and mechanical development that would allow their insertion into the common (non-industrial) human environment. This fact brings atop the importance of developing robots that can interact with humans in a seamless way (Dylla, Kreutzmann, and Wolter 2014). To this end, the present paper describes our investigation on the development of a collaborative communication system of spatial perceptions for vision-based multi-robot systems using qualitative spatial information.

The use of qualitative representations is motivated by the fact that humans do not normally use numerical descriptions to talk about the commonsense space, so a seamless human-robot interaction implies a non-metrical representation of their common environment. Besides, there are cases where communicating qualitative relations are more effective than metrical information. For instance (Freksa 1991), imagine an aquarium full of fishes and two observers, one observer wants to point a particular fish to the other. Let's assume that there is only one red stone inside the aquarium. Pointing to this particular fish in terms of metric information (e.g. "the fish is 10 cm away from the aquarium's left wall, 5 cm from its bottom, 8 cm from the rear wall and 1 m away from you") is much harder to understand than pointing to it in a purely qualitative way (e.g. "the fish near the red stone"). In order to deal with qualitative representations, this paper assumes formalisms developed in the area of Qualitative Spatial Reasoning (QSR) (Ligozat 2013; Cohn and Renz 2007). QSR is a subfield of Knowledge Representation in Artificial Intelligence that develops formalisations of space by means of qualitative relations. The use of qualitative methods allow reasoning with incomplete knowledge (Renz and Nebel 2007) and facilitates meaningful abstractions of the physical world (Moratz 2006). From qualitative representations of space, high-level communication is favoured. This promotes the application of QSR to Multi-Robot Systems in human environments.

This work assumes the interaction of groups of robots from the RoboCup Soccer Humanoid League as a domain where the ideas developed here are evaluated. In the present paper the robots collaborate by sharing their individual visual observations of a scene with each other in order to enhance their knowledge about the environment their are immersed in. Two experiments were conducted: in the first, the group of robots had to answer spatial queries using the information perceived by each robot. This information was shared among the group members and inference over qualitative relations was used to combine the multiple pieces of data. In the second experiment, the qualitative calculus was used to communicate the observations of one robot about a target that was occluded with respect to another robot.

The collaborative communication system proposed in this paper uses the discretisation of the Elevated Oriented Point Algebra with granularity 6 ($\mathcal{EOPRA}_6$) (Moratz and Wallgrün 2012). The $\mathcal{EOPRA}_m$ notation is derived from $\mathcal{OPRA}_m$ (Moratz 2006) and allows a joint representation of qualitative direction and distance between points. The reasoning of this paper is a combination of $\mathcal{OPRA}_m$ and a quantitative triangulation.

Qualitative Spatial Reasoning

One of the main challenges of QSR is the development of formal systems to represent the spatial configuration of entities in purely qualitative terms, also permitting reasoning using this representation (Cohn and Renz 2007; Dylla 2009). These formal systems use a limited amount of qualitative categories to represent the possible spatial relations between entities (Renz and Nebel 2007). Applications of QSR vary from high-level computer vision, semantic of spatial propositions, reasoning about commonsense knowledge, geographical information systems, among others (Cohn and Renz 2007). In particular, the formalism named Oriented Point Algebra ($\mathcal{OPRA}_m$) (Moratz 2006; Mossakowski and Moratz 2012) has been very influential for representing and reasoning about objects with intrinsic fronts (Dylla et al. 2007), such as cars and boats (Dylla 2009), but also robots (Moratz 2006; Dylla, Kreutzmann, and Wolter 2014). This formalism is essential in the development of the present work and it is described as follows.

Oriented Point Algebra ($\mathcal{OPRA}_m$)

The Oriented Point Algebra with granularity m ($\mathcal{OPRA}_m$) is a qualitative calculus in which objects are represented as oriented points, that are represented by Cartesian coordinates, x and y, and an orientation, θ. Each point defines a relative reference frame of granularity m where $m \in \mathbb{N}$. This granularity is used in order to obtain the angular resolution, which is equal to $\frac{2\pi}{2m}$ (Mossakowski and Moratz 2012).

In $\mathcal{OPRA}_m$, if the Cartesian coordinates of two oriented points, A and B are different (cf. Figure 1), the relationship between the points is represented by $A_m\angle_i^j B$, which means: given the granularity m, the relative position of B with respect to A is described by i and the relative position of A with respect to B is j. For example, the relation between A and B, in Figure 1, is $A_4\angle_{11}^{1} B$, meaning that A is in the sector 1 of B; B is in the sector 11 of A, and 4 is the granularity of the relative frame. Such as in other methods of QSR, the $\mathcal{OPRA}_m$ reasoning is done through a composition table, where this table is constructed with the set of all relations between three-oriented points, for example, $A_m\angle_i^j B$, $B_m\angle_k^l C$ and $A_m\angle_s^t C$, where i, j, k, l, s, t are variables that describe the relations between the points (Moratz 2006).

For example, Figure 1 shows the composition of the relations $A_4\angle_{11}^{1} B$ and $B_4\angle_{13}^{9} C$ from which the relation between the points A and C can be inferred.

$\mathcal{OPRA}_m$ works only with orientation, however, in the real world, another important spatial information is distance. Distance can be defined qualitatively by using the idea of elevated point, described below.

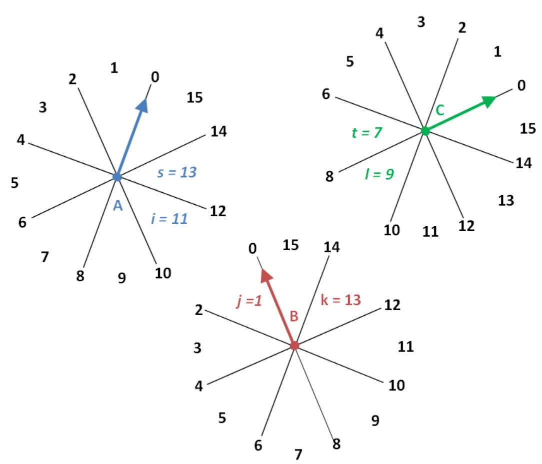
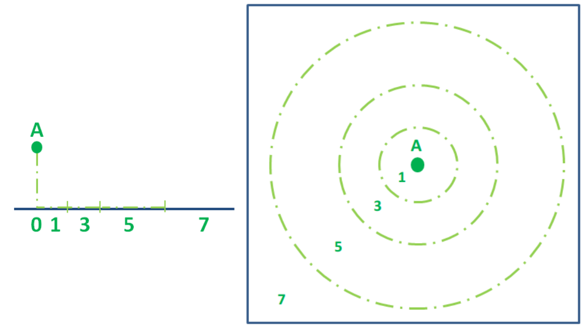


Figure 1: Composition of $A_4\angle_{11}^{1} B$ with $B_4\angle_{13}^{9} C$ can result in $A_4\angle_{13}^{7} C$, for example.



(a) Side view. (b) Bird's eye view.

Figure 2: Qualitative distances with $m = 4$: $\delta \times 0$, $\delta \times 1/2$, $\delta \times 1$ and $\delta \times 2$ (Moratz and Wallgrün 2012).

Elevated Point as Reference for Qualitative Distance

A definition of relative distances, based on local distance references (*elevations*), was proposed by (Moratz and Wallgrün 2012). Elevations are defined by the height of observers, whose projection in the 2D plane defines a circle around the observer's locations, that is used as a distance reference (Gibson 1986). The size of this projection is represented by δ, and all the distance ratios are calculated taking into consideration m and δ (Dorr and Moratz 2014). Granularity (also represented by m in the distance representation) also applies to elevations in order to provide the appropriate level of abstraction for distance relations. Distance relations between two points A and B are represented as $A\ {}_m\bigcirc_e^f\ B$, where e represents the relative distance of B with respect to A and f, the relative distance of A with respect to B.

The function $b_A(e)$, shown in the Equation 1, calculates the boundaries of qualitative distances around the elevated point A, where $0 \leqslant e \leqslant 2m$ and e must be an even number (Moratz and Wallgrün 2012). Figure 2 shows an example of a qualitative distance for $m = 4$.

$$b_A(e) = \begin{cases} \infty & \text{if } e = 2m, \\ \frac{e\delta_A}{m} & \text{if } e \leqslant m, \\ \frac{m\delta_A}{2m-e} & \text{otherwise.} \end{cases} \tag{1}$$

Elevated Oriented Point Algebra ($\mathcal{EOPRA}_m$)

$\mathcal{EOPRA}_m$ is an extension of $\mathcal{OPRA}_m$ that includes qualitative distances as elevated points. The $\mathcal{EOPRA}_m$ notation is derived from $\mathcal{OPRA}_m$, allowing a joint representation of qualitative direction and distance between two points as: $A_m\angle_{i\ e}^{j\ f} B$, where m is the common arbitrary granularity between distance and direction, i and j are orientation relations, and e and f are distance relations. Figure 3 represents the relation $A_4\angle_{11\ 3}^{1\ 5} B$ in $\mathcal{EOPRA}_m$ for two points A and B with distinct elevations.

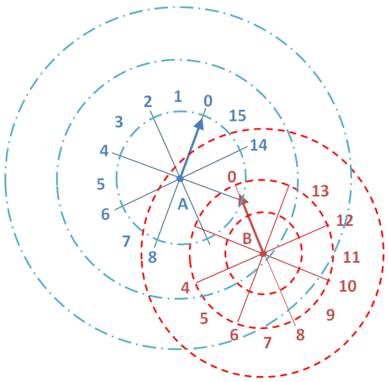


Figure 3: $\mathcal{EOPRA}_4$ relation $A \; {}_4\angle^{1}_{11} \, {}^{5}_{3} \; B$.

Collaborative Communication of Spatial Perceptions for Multi-Robot Systems

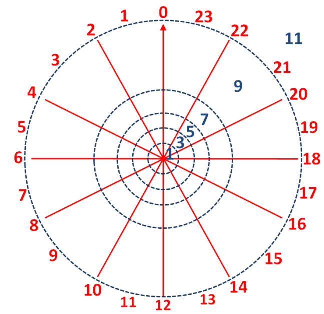
This section describes our proposal of a collaborative communication system of spatial perceptions for vision-based multi-robot systems, where the representation of the domain is built upon the perspective of the Elevated Oriented Point Algebra ($\mathcal{EOPRA}$) and the reasoning itself is made by a combination between the Oriented Point Algebra ($\mathcal{OPRA}$) and a quantitative triangulation. $\mathcal{EOPRA}_m$ discretisation is suitable for this purpose since it treats both direction and distance and allows for relative spatial perception communication, whereby a robot can locate itself "*through the eyes*" of the other robots in the domain.

In this work the granularity $m = 6$ was chosen. $\mathcal{EOPRA}_6$ discretisation is exemplified in Figure 4a. Also, the numerical regions defined were labelled by means of spatial prepositions, as shown in Figure 4b, where fr, l-fr, l, l-b, b, r-b, r and r-fr stand for $front$, $left$-$front$, $left$, $left$-$back$, $back$, $right$-$back$, $right$ and $right$-$front$, respectively. Likewise, at, vc, c, f, vf and ft stand for at, $very$-$close$, $close$, far, $very$-far and $farthest$.

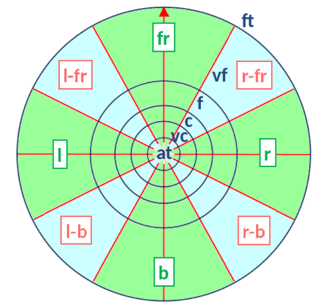
The multi-robot collaboration method proposed allows a robot to answer spatial queries even if the robot is not directly involved in the relation queried, or if it has incomplete knowledge of the domain. In this work, inference processes of directions and distance are made separately.

Due to the poverty conjecture (Forbus, Nielsen, and Faltings 1991), it is known that is, in fact, impossible to achieve a purely qualitative spatial reasoning mechanism (Cohn and Renz 2007). Thereby, distance inference is accomplished by quantitative triangulation using the law of cosines. This is possible because distances are quantitatively estimated before being discretised by means of elevations. In the same way, quantitative data are going to be used for restricting the number of possible relations during the direction inference.

Direction inference is based on $\mathcal{OPRA}_m$. However, $\mathcal{OPRA}_m$ algorithm (Mossakowski and Moratz 2012) only checks whether a composition made by the relations of the oriented points holds, i.e., it does not directly infer a composition. So, we introduce Algorithm 1, which allows a systematic way for inferring the set of possible orientations s, or s and t, in $\mathcal{OPRA}_m$ composition of relations $A_m\angle_i^j B$,



(a) $\mathcal{EOPRA}_6$



(b) Translating the numerical regions to spatial prepositions

Figure 4: $\mathcal{EOPRA}_6$: proposed discretisation.

$B_m\angle_k^l C$ and $A_m\angle_s^t C$. This algorithm checks which values s can assume, when t is given; or which values of s and t, when t is not given, and returns all compositions that hold.

Algorithm 1 may return a disjunction of relations as a result of a given composition. It is possible to reduce the number of possible relations in this disjunction by using triangulation (as represented in Algorithm 2). An example of obtaining this restriction is shown in Figure 5, where the blue robot should locate the green robot with respect to the red one. By using $\mathcal{OPRA}_m$ inference method, and assuming that t is not given, s could assume any of the following values: $0, 1, 2, 3, 17, 18, 19, 20, 21, 22, 23$, which means, fr, l-fr, r and r-fr. The blue robot can easily obtain the angle β, so it can calculate the angle α using the law of cosines. In the example shown in Figure 5, the quantitative angle α is equal to 46^o, however, as β is negative, α should also be negative in order to form a triangle. After obtaining α, this angle is discretised according to $\mathcal{OPRA}_6$ definitions, resulting in the relation $i_\alpha = -3$. Then, this new relation is added up to the relation i, where $i = 3$, leading to $i_{restr} = 0$. Now Algorithm 2 checks if i_{restr} is an even number. If so, this region is transformed into a set comprised of two odd regions ($[i_{restr} + 1, i_{restr} - 1]$). If not, i_{restr} is kept. If this odd region, or the items of the set of two regions, is contained in the set of relations inferred by $\mathcal{OPRA}_6$, then it becomes the output of the system; if not contained, the system returns failure (i.e. a contradiction has been found).

The next Section will show some preliminary experiments with our proposed $\mathcal{OPRA}_6$ combined with a quantitative triangulation that uses the $\mathcal{EOPRA}_6$ representation for performing collaborative communication of spatial perceptions for multi-robot systems.

Algorithm 1 Inferring the set of relations $\hat{s}$ or $\hat{s}$ and $\hat{t}$ of $\mathcal{OPRA}_m$ for non-coincident points.

Function $OpraInference(m, i, j, k, l, t)$
if $t = \emptyset$ **then**
  for $s_{test} = 0$ to $4m$ **do**
    for $t_{test} = 0$ to $4m$ **do**
      if $opra(m, i, j, k, l, s_{test}, t_{test})$ **then**
        add s_{test} to $\hat{s}$ and add t_{test} to $\hat{t}$
      end if
    end for
  end for
  return $\hat{s}, \hat{t}$
else
  for $s_{test} = 0$ to $4m$ **do**
    if $opra(m, i, j, k, l, s_{test}, t)$ **then**
      add s_{test} to $\hat{s}$
    end if
  end for
  return $\hat{s}$
end if
Function $opra(m, i, j, k, l, s, t)$
if $\exists\ 0 \leqslant u, v, w < 4m.\ turn_m(u, i, -s) \wedge turn_m(v, k, -j) \wedge turn_m(w, t, -l) \wedge triangle(u, v, w)$ **then**
  return $True$
end if
Function $turn_m(o, p, q)$
if $|(o + p + k + 2m) \bmod 4m) - 2m| \leqslant (o \bmod 2) \times (p \bmod 2)$ **then**
  return $True$
end if
Function $triangle_m(u, v, w)$
if $turn_m(u, v, w - 2m) \wedge (u, v, w) \neq (2m, 2m, 2m) \wedge sign_m(u) = sign_m(v) = sign_m(w)$ **then**
  return $True$
end if
Function $sign_m(q)$
if $(q \bmod 4m = 0) \vee (q \bmod 4m = 2m)$ **then**
  return 0
else if $(q \bmod 4m < 2m)$ **then**
  return 1
else
  return -1
end if

Experiments and Results

Experiments were made in two phases: first, the tests were performed in a simulated environment, using the RoboFEI-HT Soccer Simulator, in order to evaluate the proposed method with a considerably quantity of data points. Then, the method was validated in real humanoid robots.

Each phase was comprised of two experiments, used for evaluating the method proposed in this paper. The first experiment involves three robots, where each robot has to answer queries about the location of the other robots with respect to itself, or with respect to the other agents. In the second experiment, two robots have to locate a ball in a soccer field, according to their own positions. However, the target is only perceived by one robot, but not by the other (i.e., the ball may be occluded, out of the field of view, or the robot might have a faulty sensor). The inference method proposed in this work is used in this case in order to allow the latter robot to locate the ball, using the observation provided by the former and the relative locations of both robots with respect to each other.

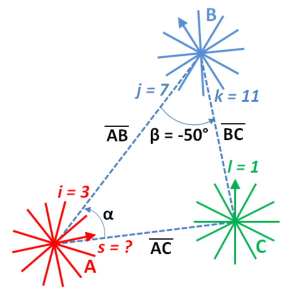


Figure 5: $\mathcal{OPRA}_6$ restricted by quantitative triangulation.

In all experiments conducted the robots were dressed with distinct colours, so that colour segmentation could be used to identify each agent. Orientation was obtained from the position of the motor in charge of the pan movement in robot's head. Distance was estimated by approximation functions, since all the sizes of the robots and other domain objects are known. The communication between the robots was conducted via broadcast using the User Datagram Protocol (UDP). The elevation δ, used for discretising the distances, was set to 1 meter, that is approximately twice the robot's

Algorithm 2 Restricting the set of $\hat{s}$ relations by the quantitative triangulation.

Function $RestrictingOpra(m, \hat{s}, \alpha)$
$i_{aux} = DiscretizeToOpra(\alpha)$
$i_{restr} = (i + i_{aux}) \bmod (4m)$
if i_{restr} = even number **then**
  $\hat{c} = [i_{restr} + 1, i_{restr} - 1]$
else
  $\hat{c} = [i_{restr}]$
end if
for $n = 0$ to $len(\hat{s})$ **do**
  for $x = 0$ to $len(\hat{c})$ **do**
    if $s_n = c_x$ **then**
      add s_n to $\hat{a}$
    end if
  end for
end for
if $\hat{a} = empty$ **then**
  return $fail$
else
  return $\hat{a}$
end if
Function $DiscretizeToOpra(\alpha)$
$i_\alpha = round(angle/(180/m) * 2)$
return i_α

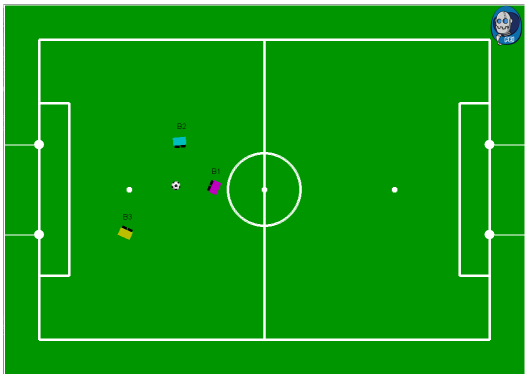


Figure 6: RoboFEI-HT Soccer Simulator.

height. This value provides the appropriate distance reference for the chosen domain.

Simulated Environment

The soccer simulator, shown in Figure 6, used for performing the first phase of the experiments was designed and developed by the authors, in order to simulate the control and the vision system of the real humanoid robots. One of the qualities of this simulator is that code made for it can be also used in our real robots.

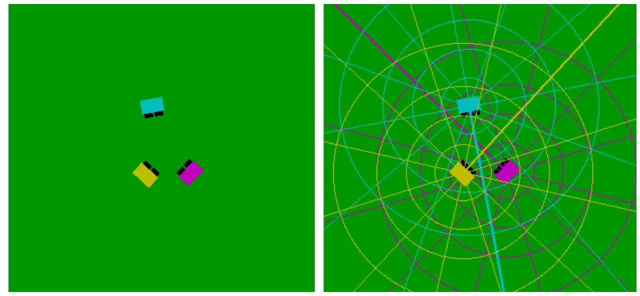
The simulator simulated Gaussian errors in the vision system for both, directions and distances, with standard deviation of 2^o and 10 cm respectively.

Experiment 1: Communication Effectiveness with Three Robots. This first experiment analyses the effectiveness of the inference process and check the behaviour of the overall system, including the vision system. The evaluation was conducted by verifying spatial queries such as "$l(x, red\_robot)$" or "$f(y, red\_robot)$", respectively, "which is the robot x that is on the left of the red_robot?" and "which is the robot y that is far from the red_robot?".

The queries were broadcast by a human agent to all robots in the experiment via UDP. Every robot had to answer every query, even if the agent is not a variable in a query. The inference process is then performed by the robots to allow the attainment of the relative relations of direction and distance.

Robots were randomly arranged 30 times for each question, and the queries were always made in relation to the red robot (R). The inferences were made by the blue robot (B) and by the yellow robot (Y). An inference consisted in finding the relation of distance and direction that holds. Figure 7 depicts the simulator with three robots positioned according to the $\mathcal{EOPRA}_6$ relation $R_6\angle_3^{21}Y$, $Y_6\angle_3^{23}B$, and $R_6\angle_{23}^{1}B$.

For example, considering only direction, the blue robot (B) can infer its own location with respect to the red robot (R) by using the following information: (i) $R_6\angle_i^j Y$, (ii) $Y_6\angle_k^l B$, and (iii) $R_6\angle_s^t B$, where (iii) is obtained by the composition of (i) and (ii). In these relations i is reported by R; j and k are reported by the yellow robot (Y); l and t are obtained from B's sensor data; and s is unknown. Considering the arrangement shown in Figure 7, we can say that the blue robot has received $i = 3$, $j = 21$, and $k = 3$, whereas it has perceived $l = 23$ and $t = 1$. It can then infer the



(a) Without visible $\mathcal{EOPRA}_6$ discretisation (b) With visible $\mathcal{EOPRA}_6$ discretisation

Figure 7: RoboFEI-HT Soccer Simulator: one of the robots' position during the first experiment.

relation $s = 23$ using the $\mathcal{OPRA}_m$ inference restricted by triangulation. This represents the location of B w.r.t. R.

From the set of answers obtained, precision, recall and accuracy were calculated. Table 1 shows the rates obtained for direction-only queries; Table 2 shows the values for distance-only queries; and, Table 3 shows the rates for combined queries. Even considering the noise purposely added in the vision system, and the inaccuracies found during $\mathcal{OPRA}_m$ inference, the results show a precision of above 80% in most cases, as well as the recall.

The lower precision results were found for the queries that involves the small relations of direction, i.e., r-b and l-fr. This happens because the frontiers of this relations are closer to each other, so it is easier for the inference process to return a wrong region.

Experiment 2: Communication of Spatial Perceptions to Handle with Occlusion. In the second experiment, the blue robot (B) uses the proposed reasoning for communicating the relative location of a ball (O) to the red robot (R), that cannot see the target due to occlusion (Figure 8). This information is then communicated to R.

The blue robot was able to see both the ball and the red robot, whereas the red robot could only see the blue robot. The set of relations obtained by the blue robot is: $\{R_6\angle_i^j B, B_6\angle_k^l O, R_6\angle_s^t O\}$. During this experiment both s and t are inferred, even if only s is necessary. As the ball does not have an intrinsic orientation, we assumed that it is oriented toward the blue robot, i.e. $l = 0$.

This experiment was conducted by randomly positioning the ball 30 times in different positions inside of each qualitative region of the robot R. Then, the blue robot (B) inferred the ball position, i.e. direction and distance, in relation to the red one (R), using the qualitative inference system restricted by triangulation. As the discretisation of direction is symmetric, only the regions fr, l-fr, l and l-b were chosen for being evaluated. So, the results of the ball's position inference, made by the blue robot (B) are presented in two confusion matrix: Table 4 for direction and Table 5 for distance. Each column of the tables represents the inference made while the rows represent the actual position of the ball w.r.t. the red robot (R). It is possible to notice that, as well as seen in the last experiment, the error is higher for the small

Table 1: Direction-only queries in simulator: evaluation of responses

Spatial Query	Precision	Recall	Accuracy
$fr(x, R)$	100.00%	84.62%	96.67%
$r(x, R)$	88.89%	94.12%	95.45%
$r\text{-}fr(x, R)$	90.00%	69.23%	92.42%
$r\text{-}b(x, R)$	66.67%	100.00%	98.48%
$l(x, R)$	94.12%	94.12%	96.97%
$l\text{-}fr(x, R)$	66.67 %	100.00%	95.45%
$l\text{-}b(x, R)$	81.82%	100.00%	96.97%

Table 2: Distance-only queries in simulator: evaluation of responses

Spatial Query	Precision	Recall	Accuracy
$at(x, R)$	90.91%	90.91%	96.97%
$vc(x, R)$	100.00%	92.86%	98.48%
$c(x, R)$	80.00%	88.89%	95.45%
$f(x, R)$	84.62%	64.71%	87.88%
$vf(x, R)$	80.00%	92.31%	94.03%

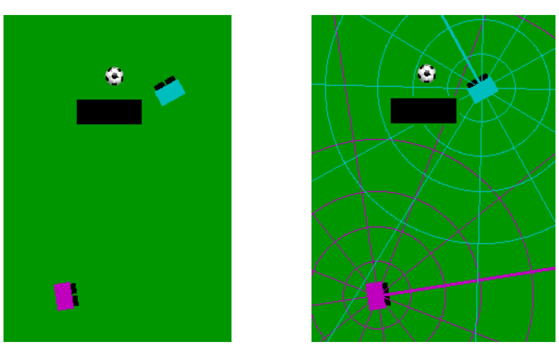
Table 3: Queries combining distance and direction in simulator: evaluation of responses

Spatial Query	Precision	Recall	Accuracy
$fr(x, R) \wedge c(x, R)$	100.00%	83.33%	98.48%
$l(x, R) \wedge vc(x, R)$	85.71%	75.00%	95.45%
$r\text{-}fr(x, R) \wedge at(x, R)$	100.00%	71.43%	96.97%
$r(x, R) \wedge f(x, R)$	100.00%	100.00%	100.00%

regions of the qualitative direction ($l\text{-}fr$ and $l\text{-}b$).

The simulated robots work with the Cross Architecture concept (Perico et al. 2014), as well as the real robots. So, it is possible to program both simulated and real robots, in several languages, such as C++ and Python. This feature allowed us to extend the simulated research to our real humanoid robots without many changes. The simulated experiments were performed in an Intel i5 with 8GB SDRAM running Ubuntu 14.04[1].

[1]The simulator used in this work, along with the source code of the proposal, are available at the URL `http://fei.edu.br/~rbianchi/software.html`



(a) Without visible discretisation (b) With visible discretisation

Figure 8: RoboFEI-HT Soccer Simulator: one of the ball's position during the second experiment. Ball is always occluded from the red robot (R).

Table 4: Confusion matrix for regions of direction: ball's positioning inference w.r.t. to the red robot (R).

regions	*r-fr*	*fr*	*l-fr*	*l*	*l-b*	*b*
fr	7.8%	76.4%	15.8%	0.0%	0.0%	0.0%
l-fr	0.0%	10.0%	72.5%	17.5%	0.0%	0.0%
l	0.0%	0.0%	18.4%	73.7%	7.9%	0.0%
l-b	0.0%	0.0%	0.0%	20.4%	61.2%	18.4%

Table 5: Confusion matrix for regions of distance: ball's positioning inference w.r.t. to the red robot (R).

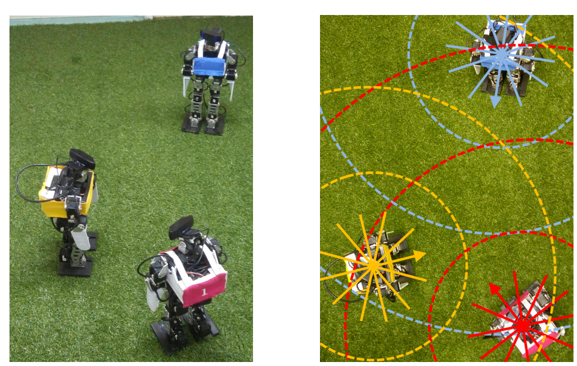
regions	*at*	*vc*	*c*	*f*	*vf*
at	86.6%	13.3%	0.0%	0.0%	0.0%
vc	3.3%	90.0%	6.6%	0.0%	0.0%
c	0.0%	0.0%	93.3%	6.6%	0.0%
f	0.0%	0.0%	6.6%	93.3%	0.0%
vf	0.0%	0.0%	0.0%	3.3%	96.7%

Real Robots

The experiments performed with real robots were conducted with humanoid robots inspired on the DARwIn-OP design (Ha et al. 2011). The robots have height of 490 mm; 20 degrees of freedom (6 for each leg, 3 for each arm and 2 in the neck) and a Full-HD camera located in the robot's head. On-board processing is made by an Intel NUC Core i5 with 8GB SDRAM. Figure 9a shows the robots in the first experiment.

Experiment 1: Communication Effectiveness with Three Robots. The effectiveness test with real robots considered the arrangement depicted in Figure 9, that is the reproduction of the arrangement shown in Figure 7. Such as executed in the simulator, the queries were made in relation to the red robot (R). The inferences were made by the blue robot (B) and by the yellow robot (Y). This test followed almost the same procedure adopted in the simulation, the difference is the number of times that each spatial query was repeated, that was 10 for the real robot.

Table 6 shows precision, recall and accuracy of the given answers, as well as Table 7 and 8. The vision system in real robots are lagged and noisy, so the errors can be even worse



(a) Humanoid robots (b) Bird's eye view

Figure 9: Robot's position for the first experiment.

Table 6: Direction-only queries in real robots: evaluation of responses

Spatial Query	Precision	Recall	Accuracy
$fr(x, R)$	76.00%	90.48%	80.00%
$r(x, R)$	–	–	100.00%
$r\text{-}fr$	–	–	87.50%
$r\text{-}b$	–	–	100.00%
$l(x, R)$	–	–	82.50%
$l\text{-}fr$	100.00%	75.00%	87.50%
$l\text{-}b$	–	–	100.00%

Table 7: Distance-only queries in real robots: evaluation of responses

Spatial Query	Precision	Recall	Accuracy
$at(x, R)$	–	–	97.50%
$vc(x, R)$	94.74%	100.00%	97.37%
$c(x, R)$	80.00%	70.59%	80.00%
$f(x, r)$	–	–	100.00%
$vf(x, R)$	–	–	100.00%

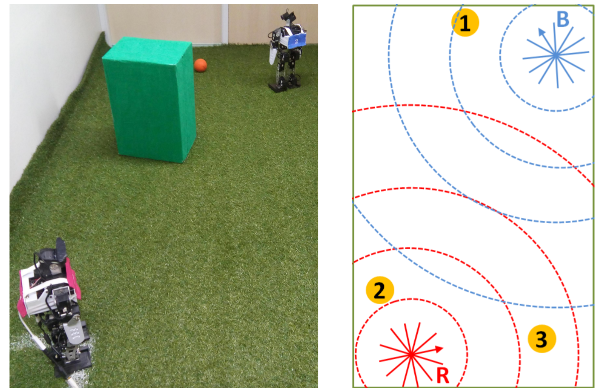
Table 8: Queries combining distance and direction in real robots: evaluation of responses

Spatial Query	Precision	Recall	Accuracy
$fr(x, R) \wedge c(x, R)$	87.50%	77.78%	85.00%
$l(x, R) \wedge vc(x, R)$	–	–	100.00%
$r\text{-}fr(x, R) \wedge at(x, R)$	–	–	100.00%
$r(x, R) \wedge f(x, R)$	–	–	100.00%

than those one considered in the simulation. Even taking these errors into account, the lower precision found is 75% while the lower recall was around 70%. It is also possible to see in Tables 6, 7 and 8 that some precisions and recalls were not calculated, due to the fact that there were not true positives in the involved regions.

Experiment 2: Communication of Spatial Perceptions to Handle Occlusion. Since we have already had, from the simulation, a quantitative evaluation for the inference of an occluded ball, the second experiment was made in order to validate the results obtained in the simulation in the real robots. Moreover, this experiment aims to provide a comparison between the purely qualitative inference system of $\mathcal{OPRA}_m$ and this same qualitative inference system restricted by triangulation. Each part of the experiment was conducted with an orange ball in three distinct positions, while the blue robot (B) and the red robot (R) remained in the same position during all the experiment. The position of the spatial entities is depicted in Figure 10.

The first column of Table 9 represent the ground truth for this experiment. The results obtained by the qualitative-only inference and the results obtained by the qualitative location restricted by triangulation are shown in the second and third columns respectively. The grey regions represent the position of the ball – real, in the first column, and inferred, in second and third columns. It is worth noting that, in some situations, the purely qualitative inference executed by $\mathcal{OPRA}_m$ is not possible, such as the situation where the ball is positioned in 3 (Figure 10b), represented in the bottom line of Table 9. On the other hand, due to the inclusion of triangulation as a constraint for $\mathcal{OPRA}_m$ inference, the system



(a) Hidden ball. (b) The 3 ball's positions.

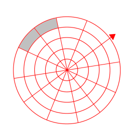
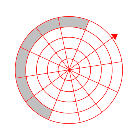
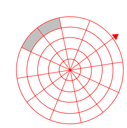
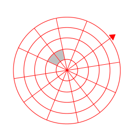
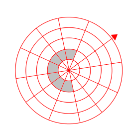
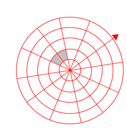
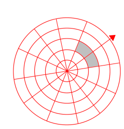
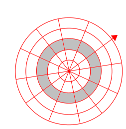
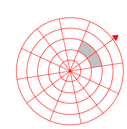
Figure 10: The arrangement of the second experiment: the ball had 3 distinct positions while the robots keep their initial position.

was capable to infer the correct position.

Related Work

$\mathcal{OPRA}_m$ has been used as a method for integrating local knowledge in a quadruped mobile robot (Moratz and Ragni 2008), where, during the experiments, the robot was able to distinguish between colours and simple objects using a monocular vision system. Despite using computer vision, the robot had no prior knowledge of the size of objects. Thus, distance estimation was not possible; the only information available was the local orientation of the robot in relation to the objects. The robot was able to complete the task "move to the yellow cube behind the red disk" using the $\mathcal{OPRA}_6$ as an engine of reasoning. The task was transmitted to the robot by human speech commands. In another work, $\mathcal{OPRA}_4$ was applied to formalise the Inter-

Table 9: Inference of the ball's position wrt the red robot.

Ball Pos.	Ground Truth	Blue Inference: Purely Qualitative	Blue Inference: Qualitative Restricted by Triangulation
1			
2			
3			

national Regulations for Preventing Collisions at Sea (Col-Regs) (Dylla et al. 2007). The authors focused their work on translating the navigation rules from natural language descriptions to a qualitative formalisation for agent control. $\mathcal{OPRA}$ calculus was chosen because direction information is extremely important in sea navigation. International Regulations for Preventing Collisions at Sea define, for each pair of vessels, which one has to give way and which is the privileged one, where different types of vessels require different rules. The qualitative representation was then obtained considering three vessels actions and the reasoning was made by constraint networks formed by the transition systems of the applicable rules. $\mathcal{OPRA}$ has also been used for navigation in street networks that are described via local observations (Lucke, Mossakowski, and Moratz 2011) and, when combined with the Region Connection Calculus (RCC-8) (Cohn et al. 1997), it has been used for defining the set of qualitative relations that represent social conventions (Dylla, Kreutzmann, and Wolter 2014).

The present work extends these qualitative methodologies applying them in real autonomous agents, where the distance concept was included. Also, by considering quantitative triangulation restricting $\mathcal{OPRA}_m$, the implementation described here shows better precision than its predecessors.

Conclusion

This work proposed a collaborative communication of spatial perceptions for multi-robot systems defined over $\mathcal{EOPRA}_m$ representation, where the reasoning is made by $\mathcal{OPRA}_m$ compositions combined with quantitative triangulations. This implementation was tested on a simulated environment and also on groups of real humanoid robots involved in collaborative tasks. Two experiments were considered: in the first, the group of robots had to answer spatial queries using the information perceived by each robot. This information was shared among the group members and the proposed inference was used to combine the multiple pieces of data, from where an answer to the query could be obtained by a simple predicate-unification process. In the second experiment, the proposed inference was used to communicate the observations of one robot about a target that was occluded with respect to another robot. The results obtained indicate that $\mathcal{EOPRA}_m$ representation is a suitable tool for representing (and sharing) qualitative spatial knowledge in groups of robots. Its qualitative nature allows for the definition of a small number of relations, that are closer to spatial predicates used in natural languages.

Acknowledgments

D. H. Perico acknowledges support from CAPES; P. E. Santos acknowledges support from FAPESP (2012/04089-3).

References

Cohn, A. G., and Renz, J. 2007. Qualitative spatial reasoning. In van Harmelen, F.; Lifschitz, V.; and Porter, B., eds., *Handbook of Knowledge Representation*. Elsevier.

Cohn, A. G.; Bennett, B.; Gooday, J. M.; and Gotts, N. 1997. RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica* 1:275–316.

Dorr, C. H., and Moratz, R. 2014. Qualitative shape representation based on the qualitative relative direction and distance calculus eopram. *CoRR* abs/1412.6649.

Dylla, F.; Frommberger, L.; Wallgrün, J. O.; Wolter, D.; Wölfl, S.; and Nebel, B. 2007. Sailaway: Formalizing navigation rules. In *Proc. AISB'07 Artif. and Ambient Intell. Symp. on Spatial Reasoning and Communication*.

Dylla, F.; Kreutzmann, A.; and Wolter, D. 2014. A qualitative representation of social conventions for application in robotics. In *2014 AAAI Spring Symposium - Qualitative Representations for Robots*, 34–41. AAAI Press.

Dylla, F. 2009. *Qualitative Spatial Reasoning for Navigating Agents: Behavior Formalization with Qualitative Representations*, volume 3 of *Ambient Intelligence and Smart Environments*. IOS Press. 98 – 128.

Forbus, K.; Nielsen, P.; and Faltings, B. 1991. *Qualitative Spatial Reasoning: The CLOCK Project*.

Freksa, C. 1991. Qualitative spatial reasoning. In Mark, D. M., and Frank, A. U., eds., *Cognitive and Linguistic Aspects of Geographic Space*. Dordrecht: Kluwer. 361–372.

Gibson, J. 1986. *The Ecological Approach to Visual Perception*. Resources for ecological psychology.

Ha, I.; Tamura, Y.; Asama, H.; Han, J.; and Hong, D. W. 2011. Development of Open Humanoid Platform DARwIn-OP. In *Proc. SICE Annual Conference*, 2178–2181. IEEE.

Ligozat, G. E. 2013. *Qualitative Spatial and Temporal Reasoning*. ISTE. Wiley.

Lucke, D.; Mossakowski, T.; and Moratz, R. 2011. Streets to the opra - finding your destination with imprecise knowledge. In *Workshop on Benchmarks and Applications of Spatial Reasoning (IJCAI)*, 25–32.

Moratz, R., and Ragni, M. 2008. Qualitative spatial reasoning about relative point position. *J. of Visual Languages & Computing* 19(1):75–98.

Moratz, R., and Wallgrün, J. O. 2012. Spatial reasoning with augmented points: Extending cardinal directions with local distances. *J. Spatial Information Science* 5(1):1–30.

Moratz, R. 2006. Representing relative direction as a binary relation of oriented points. In *ECAI 2006 Proc. 17th European Conf. on Artif. Intelligence*.

Mossakowski, T., and Moratz, R. 2012. Qualitative reasoning about relative direction of oriented points. *Artif. Intell.*

Perico, D.; Silva, I.; Vilao, C.; Homem, T.; Destro, R.; Tonidandel, F.; and Bianchi, R. 2014. Hardware and software aspects of the design and assembly of a new humanoid robot for robocup soccer. In *Robotics: SBR-LARS Robotics Symposium and Robocontrol*, 73–78.

Renz, J., and Nebel, B. 2007. Qualitative spatial reasoning using constraint calculi. In Aiello, M.; Pratt-Hartmann, I.; and Van Benthem, J., eds., *Handbook of Spatial Logics*. Springer Netherlands. 161–215.

Using qualitative reasoning to evaluate performance: An application in the retail sector

Xari Rovira and **Núria Agell** and **Josep M. Sayeras**
ESADE Business School, Universitat Ramon Llull, Av. Pedralbes 62, Barcelona, Spain
e-mail: {xari.rovira, nuria.agell, josep.m.sayeras}@esade.edu

Mónica Sánchez
Universitat Politècnica de Catalunya, UPC-BarcelonaTech, Jordi Girona, 1-3, Barcelona, Spain
e-mail: monica.sanchez@upc.edu

Abstract

This paper offers a new method inspired by classic importance-performance analysis (IPA) that provides a global index of importance versus performance for firms together with a new version of the IPA diagram. The index compares two rankings of the same set of features regarding importance and performance, taking into account under-performing features. The marginal contribution of each feature to the proposed global index defines a set of iso-curves that represents an improvement in the IPA diagram. The defined index, together with the new version of the diagram, enables, by means of qualitative reasoning techniques, the assessment of a firm's overall performance and therefore enhance decision making in the allocation of resources. The proposed method has been applied to a Taiwanese multi-format retailer and managerial perceptions of performance and importance are compared to assess the firm's overall performance.

Introduction

In Importance-performance analysis (IPA), firm features are ranked regarding either their importance and their performance. Differences between importance and performance rankings of features are considered when assessing a firm's resource allocation. Initial approaches in the late 70s were based on simple and intuitive graphic techniques (Martilla and James 1977). The traditional IPA methodology basically consists of representing ratings of importance and performance for several features on a two-dimensional chart. The resulting importance-performance grid is divided into four quadrants. To interpret the results, Martilla and James give a name to each quadrant to help managers determine the highest and lowest priorities for improvement, as shown in Figure 1 (Martilla and James 1977).

We present in this paper a new similarity index to compare the importance and performance rankings of the same set of features. The proposed index is based on induced ordered weighted averaging (IOWA) operators (Yager and Filev 1999), and importance and performance rankings are obtained by means of qualitative assessments of the different features considered in a firm evaluation. These assessments, given by a set of experts, are expressed using order-of-magnitude models, allowing the experts to use different levels of precision for each feature. Differences between the importance and performance ordered lists are considered to define the index of similarity. This index, when applied to a firm's features rankings for both importance and performance, enables a firm's global performance to be assessed. There are two main differences between our index and existing indexes such as Kendall's Tau and Spearman's Rho correlation coefficients. On the one hand, the asymmetry of the features treatment, i.e., it just takes into account under-performing features, and, on the second hand, the specific relation between the weights and the importance, i.e., the more important an under-performing feature, the greater its weight is considered in the similarity index.

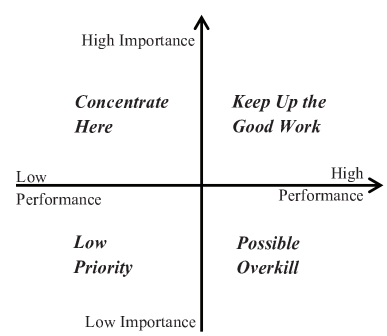


Figure 1: IPA diagram (Martilla and James 1977)

In addition, in this paper, a new IPA diagram, based on the proposed similarity index is presented to select features where resource allocation is necessary. The new IPA diagram is defined via the iso-curves obtained when considering the marginal contribution of the features to the proposed similarity index.

An application of the presented method to the retail sector has been conducted. The starting point of our application is a set of 44 features used in the retail sector that were selected by expert managers as the main performance variables. The similarity index is applied to compare the two rankings of this set of features. Whilst the proposed similarity index could have broader applications, the specific application in this paper throws light on company resource allocation (Deng 2007).

Theoretical framework

Several authors have conducted various analytical measures to compare the gap between performance and importance in the features that describe a firm. The index considered in this paper is based, on the one hand, on a ranking method that uses qualitative linguistic descriptions, and, on the other hand, on IOWA operators.

A ranking method using qualitative linguistic descriptions

In the proposed ranking method, each feature is characterized by the judgments of k evaluators, and each evaluator makes his/her judgements by means of qualitative labels belonging to an order-of-magnitude space $\mathbb{S}_{m_h}$ with granularity m_h for $h = 1, \dots, k$. The evaluations are then synthesized by means of the distance to a reference k-dimensional vector of labels. In this way, the process considered for ranking features assessed by k expert evaluators can be split in the following four steps:

Step 1. Feature representation as k-dimensional vectors of labels Features are represented by a k-dimensional vectors of labels belonging to the set $\mathbb{X}$, which is defined as:

$$\mathbb{X} = \mathbb{S}_{m_1} \times \cdots \times \mathbb{S}_{m_k} = \{\mathbf{X} = (X_1, \dots, X_k) \mid X_i \in \mathbb{S}_{m_h} \ \forall h = 1, \dots k\}.$$

For every component monotonicity is assumed, i.e., $X_h \leq X'_h$ indicates that the evaluation made by the evaluator h corresponding to the feature X' is better or equal to the one corresponding to X. The order relation defined in each $\mathbb{S}_{m_h}$ is extended to the Cartesian product $\mathbb{X}$:

$$\mathbf{X} = (X_1, \dots, X_k) \leq \mathbf{X}' = (X'_1, \dots, X'_k) \iff X_h \leq X'_h \ \forall h = 1, \dots, k.$$

This order relation in $\mathbb{X}$ is partial, since there are pairs of non-comparable k-dimensional vectors of labels. And $\mathbf{X} < \mathbf{X}'$, that is to say, $\mathbf{X} \leq \mathbf{X}'$ and $\mathbf{X} \neq \mathbf{X}'$, means that feature $\mathbf{X}$ is preferred to feature $\mathbf{X}'$ by all the evaluators.

Step 2. A distance between k-dimensional vectors of labels The method presented in (Agell et al 2012) via a codification of the labels in each $\mathbb{S}_{m_h}$ given by a location function is considered. The location function codifies each element $X_h = [B_i, B_j]$ in $\mathbb{S}_{m_h}$ by a pair of integers $(l_1(X_h), l_2(X_h))$, where $l_1(X_h)$ is the opposite of the number of basic elements in $\mathbb{S}_{m_h}$ that are "between" B_1 and B_i, that is, $l_1(X_h) = -(i-1)$, and $l_2(X_h)$ is the number of basic elements in $\mathbb{S}_{m_h}$ that are "between" B_j and B_{m_h}, i.e., $l_2(X_h) = m_h - j$.

The extension of the location function to the set $\mathbb{X}$ of k-dimensional vectors of labels is defined as:

$$L(\mathbf{X}) = L(X_1, \dots, X_k) = (l_1(X_1), l_2(X_1), \dots, l_1(X_k), l_2(X_k)).$$

A distance d between labels $\mathbf{X}, \mathbf{X}'$ in $\mathbb{X}$ is then defined via a weighted Euclidian distance in $\mathbb{R}^{2k}$ between their codifications:

$$d(\mathbf{X}, \mathbf{X}') = \sqrt{\sum_{h=1}^{k} w_h[((l_1(X_h) - l_1(X'_h))^2 + (l_2(X_h) - l_2(X'_h))^2]}.$$

where w_i are considered to be the weights assigned to the k evaluators and $\sum_{h=1}^{k} w_h = 1$. This function inherits all the properties of the weighted Euclidian distance in $\mathbb{R}^{2k}$.

Step 3. Building a reference k-dimensional vector of labels The reference k-dimensional vector of labels considered in this ranking method is the supreme with respect to the order relation $\leq$ of the set of feature representations.

Let $\{\mathbf{X}^1, \dots, \mathbf{X}^n\} \subset \mathbb{X}$ be the set of n features representations to be ranked, then the supreme of the set $\mathbf{X}^{\text{sup}}$, i.e., the minimum label in $\mathbb{X}$ which satisfies $\mathbf{X}^r \leq \mathbf{X}^{\text{sup}}, r = 1, \dots, n$, is computed as follows:

Given $\mathbf{X}^r = (X_1^r, \dots, X_k^r)$, with $X_h^r = [B_{i_h}^r, B_{j_h}^r]$ for all $h = 1, \dots, k$, and for all $r = 1, \dots, n$, then,

$$\mathbf{X}^{\text{sup}} = \sup\{\mathbf{X}^1, \dots, \mathbf{X}^n\} = (\tilde{X}_1, \dots, \tilde{X}_k),$$

where:

$$\tilde{X}_h = [\max\{B_{i_h}^1, \dots, B_{i_h}^n\}, \max\{B_{j_h}^1, \dots, B_{j_h}^n\}].$$

Step 4. Obtaining the features ranking from the values $d(\mathbf{X}, \mathbf{X}_{sup})$ Let d and $\mathbf{X}^{\text{sup}}$ be respectively the distance and the reference label defined in Steps 2 and 3. Then the following binary relation in $\mathbb{X}$:

$$\mathbf{X} \ll \mathbf{X}' \iff d(\mathbf{X}', \mathbf{X}^{\text{sup}}) \leq d(\mathbf{X}, \mathbf{X}^{\text{sup}})$$

is a pre-order, i.e., it is reflexive and transitive. This pre-order relation induces an equivalence relation $\equiv$ in $\mathbb{X}$ by means of:

$$\mathbf{X} \equiv \mathbf{X}' \iff [\mathbf{X} \ll \mathbf{X}' \ , \ \mathbf{X}' \ll \mathbf{X}] \iff d(\mathbf{X}', \mathbf{X}^{\text{sup}}) = d(\mathbf{X}, \mathbf{X}^{\text{sup}}).$$

In the quotient set $\mathbb{X}/\equiv$ the following relation between equivalence classes:

$$\text{class}\,(\mathbf{X}) \trianglelefteq \text{class}\,(\mathbf{X}') \iff \mathbf{X} \ll \mathbf{X}' \iff d(\mathbf{X}', \mathbf{X}^{\text{sup}}) \leq d(\mathbf{X}, \mathbf{X}^{\text{sup}})$$

is an order relation. It is trivially a total order.

In this way, a set of features $\mathbf{X}^1, \dots, \mathbf{X}^n$ can be ordered as a chain with respect to their proximity to the supreme: $\text{class}\,(\mathbf{X}^{i_1}) \trianglelefteq \cdots \trianglelefteq \text{class}\,(\mathbf{X}^{i_n})$.

If each class $(\mathbf{X}^{i_j}), j = 1, \dots n$, contains only a feature representation $\mathbf{X}^{i_j}$, the process is finished and we obtain the ranking $\mathbf{X}^{i_1} \trianglelefteq \cdots \trianglelefteq \mathbf{X}^{i_n}$. If there is some class $(\mathbf{X}^{i_j})$ with more than one feature representation, then the same ranking process is applied to the set of the feature representations belonging to class $(\mathbf{X}^{i_j})$, and continued until an iteration of the process gives the same ranking as the previous iteration. The final ranking $\mathbf{X}^{m_1} \trianglelefteq \cdots \trianglelefteq \mathbf{X}^{m_n}$ is then obtained.

IOWA operators

As rankings are generated for both importance and performance when measuring the same set of features, the definition of a suitable indicator of their differences is a relevant issue. A comparison of rankings may be undertaken with different techniques, but most techniques do not take into account the relative importance of the ranked items, and only consider their relative ranked position. The index considered in this paper, based on induced ordered weighted averaging (IOWA) operator's concept (Chiclana et al 2007; Yager and Filev 1999) enables importance and performance rankings to be compared more sensitively. IOWA operators were introduced in (Yager and Filev 1999) as an extension of ordered weighted averaging (OWA) operators (Yager 1988).

The OWA operators are a type of weighted mean that enables tuning the weights by means of the relative importance of the considered variable. To this end, values of the considered variable are ordered before being weighted.

Definition 1 *(Yager 1988) An OWA operator of dimension n is a mapping $f : \mathbb{R}^n \to \mathbb{R}$ such that:*

$$f(x_1, \dots, x_n) = \sum_{i=1}^{n} w_i x_{(i)},$$

where $x_{(i)}$ are the same values as x_i ordered from the largest to the smallest, and w_i are a set of weights such that $w_i \in [0, 1]$ and $\sum_{i=1}^{n} w_i = 1$.

On the other hand, IOWA operators consider two related variables: First, the order inducing variable, and second, the argument variable. The argument variable values are aggregated using a set of weights based on the order of the values of the first variable.

Definition 2 *(Yager and Filev 1999) An IOWA operator of dimension n is a mapping $\Phi : (\mathbb{R} \times \mathbb{R})^n \to \mathbb{R}$ such that:*

$$\Phi((u_1, x_1), \dots, (u_n, x_n)) = \sum_{i=1}^{n} w_i x_{\sigma(i)},$$

where $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ is a permutation such that $u_{\sigma(i)} \geq u_{\sigma(i+1)}$, $\forall i = 1, \dots, n-1$, and w_i are a set of weights such that $w_i \in [0, 1]$ and $\sum_{i=1}^{n} w_i = 1$.

Both OWA and IOWA operators have been deeply studied and applied in multi-criteria and group decision-making literature (Chiclana et al 2007). In addition, several extensions of the above-mentioned operators have been introduced in other studies to deal with situations where fuzzy or linguistic variables are considered in the decision-making process (Herrera and Herrera-Viedma 1997; Herrera-Viedma et al 2006).

An index for comparing importance and performance

The following definitions consider differences between performance and importance in features ordered from the most important to the least. The global index proposed in this paper is a convenient weighted mean of these differences, i.e., an IOWA operator, with importance as order inducing variable and these differences as argument variable.

Let n be the number of features considered to describe a firm and I_i and P_i be the importance and performance positions in the rankings of the ith feature respectively. I_i and P_i are numbers from 1 to n such that the feature corresponding to $I_i = 1$ is the most important and the feature corresponding to $P_j = 1$ is the best performed.

Note that from now on, the features are considered ordered with respect to their importance position in the ranking, i.e., the $(i)th$ feature is the feature with importance position in the ranking $I_{(i)} = i$, and so $I_{(1)} = 1 \dots, I_{(n)} = n$.

Definition 3 *The importance-performance vector of a firm F is the vector:*

$$IPR(F) = ((1, P_1), \dots, (n, P_n))$$

whose components are the pairs of ranking values of its considered features with respect to importance and performance, ordered with respect to their importance position in the ranking.

The n components of the $IPR(F)$ vector of a firm F can be represented as points in the IPA diagram, each point (x, y) corresponding to one of the n considered features. To include all these n points in the classical IPA diagram, the reverse positions in the ranking with respect to performance and importance, centered in $(\frac{n+1}{2}, \frac{n+1}{2})$, have to be computed, i.e., $x = \frac{n+1}{2} - P_i$ and $y = \frac{n+1}{2} - i$.

Note that the ranking values (i, P_i) of the considered features with respect to importance and performance can be obtained via any ranking method. Agell et al (2012) proposes a ranking method based on the absolute order-of-magnitude qualitative model.

From now on, let us denote by IPR^* the importance-performance vector of the ideal best performed firm, i.e., $IPR^* = ((1, 1), \dots, (i, i), \dots, (n, n))$ and IPR_* the importance-performance vector of a firm in the opposite situation, i.e., $IPR_* = ((1, n), \dots, (i, n-i+1), \dots, (n, 1))$.

To focus on the features in which resources must be allocated, and from the importance-performance vector of a firm $IPR(F) = ((1, P_1), \dots, (n, P_n))$, the next definition introduces a new vector that takes into account only underperforming features, i.e., those features where their performance position in the ranking is worse than their importance position in the ranking.

Definition 4 *Let $IPR(F) = ((1, P_1), \cdots, (n, P_n))$ be the importance-performance vector of a firm F. The non-negative performance-importance differences vector of the firm is the n-dimensional vector $DV(F) = (X_1, \dots, X_n)$, where $X_i = \max(P_i - i, 0)$, for all $i = 1, \dots, n$.*

Note that for any firm F, the components of $DV(F)$, are $X_i \geq 0$ for all $i = 1, \dots, n$ and nonzero components correspond to under-performing features.

In the two cases described above, corresponding to the ideal best performed firm and its opposite situation, the associated non-negative performance-importance differences vectors are respectively:

$DV^* = (0, \dots, 0)$ and $DV_* = (n-1, \dots, \max(n - 2i + 1, 0), \dots, 0)$.

Based on the usual partial order in $\mathbb{R}^n$, the next definition establishes a preference relation between differences vectors introduced in Definition 4, and therefore between the importance-performance status of firms.

Definition 5 *Let $DV(F^1) = (X_1^1, \dots, X_n^1)$ and $DV(F^2) = (X_1^2, \dots, X_n^2)$ be two differences vectors, then $DV(F^1)$ is preferred to $DV(F^2)$, $DV(F^1) \preceq DV(F^2)$,* *when $DV(F^1) \leq DV(F^2)$ with the usual order in $\mathbb{R}^n$, i.e., $X_i^1 \leq X_i^2$ for all $i = 1, \ldots, n$.*

In this way, $DV(F^1)$ is preferred to $DV(F^2)$ when F^1 performs better than F^2 for all under-performing features. Differences vectors introduced in Definition 4 enable us to define an index via an IOWA operator that preserves this preference relation:

Definition 6 *Let $DV(F) = (X_1, \ldots, X_n)$ be the differences vector of a firm, the Global Importance-Performance Index ($\mathcal{G}$) of the firm is:*

$$\mathcal{G}(X_1, \ldots, X_n) = \sum_{i=1}^{n} w_i X_i$$

where weights are computed using Borda-Kendall method (Kendall1962), i.e., $w_i = \dfrac{2(n-i+1)}{n(n+1)}$ for all $i = 1, \ldots, n$.

Note that $w_i \in [0,1]$ for all $i = 1, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$. These weights express the ratio between the reverse importance position in the ranking $n - I_i - 1 = n - i - 1$ of the ith feature and $\sum_{i=1}^{n} i$. Indeed, the weights decrease from $\dfrac{2n}{n(n+1)}$ to $\dfrac{2}{n(n+1)}$. In this way, features with greater importance have greater weights in the weighted mean defining the $\mathcal{G}(X_1, \ldots, X_n)$ of a given firm.

Note that $\mathcal{G}(X_1, \ldots, X_n)$ is an IOWA operator with importance as order inducing variable and the non-negative performance-importance differences as argument variable.

In the following proposition, some properties of $\mathcal{G}(X_1, \ldots, X_n)$ are provided.

Proposition 1 *$\mathcal{G}(X_1, \ldots, X_n)$ satisfies the following properties:*

1. *$\mathcal{G}(X_1, \ldots, X_n) \geq 0$.*
2. *$\mathcal{G}(X_1, \ldots, X_n) = 0$ if and only if $P_i = i$ for all $i = 1, \ldots, n$, i.e., $(X_1, \ldots, X_n) = (0, \ldots, 0) = DV^*$.*
3. *If n is even $\mathcal{G}(DV_*) = \dfrac{5n-2}{12}$, and if n is odd $\mathcal{G}(DV_*) = \dfrac{(n-1)(5n+3)}{12n}$.*
4. *$\mathcal{G}(X_1, \ldots, X_n)$ preserves the $\preceq$ relation.*

Proofs can be found in (Sayeras et al 2015)

The following proposition establishes an intuitive property for the $\mathcal{G}$ index, relating it with the partition of the IPA diagram in (Abalo et al 2007) (see Figure 2) and determining relevant importance-performance situations. Abalo et al (2007) use a partition that combines the quadrant and diagonal-based schemes, enlarging the top left quadrant as shown in Figure 2.

Proposition 2 *The features that contribute to the $\mathcal{G}$ index are all features above the principal diagonal of the IPA diagram, i.e., those classified as "Concentrate Here" in the partition of the IPA diagram in (Abalo et al 2007).*

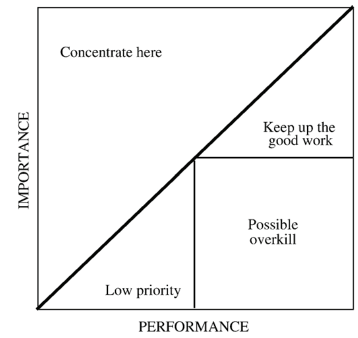


Figure 2: A partition of the IPA diagram, Abalo et al. (Abalo et al 2007)

PROOF. The proof is straightforward, because only features above the diagonal $I = P$ provide non-negative performance-importance differences.$\square$

The following proposition determines the level curves (iso-curves) of the marginal contribution of the features to the $\mathcal{G}$ index in the IPA diagram, giving decision makers a precise information about where to concentrate resources to improve performance.

Proposition 3 *The level curves of the marginal contribution of a feature to the $\mathcal{G}$ index in the IPA diagram are:*

$$\frac{n+1+2y}{n(n+1)}(y-x) = k,$$

for any $k \in \mathbb{R}^+$ (see Figure 3).

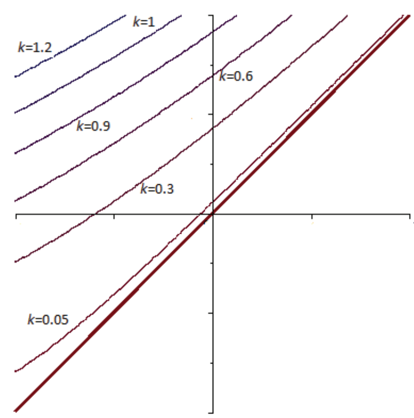


Figure 3: Level curves of the marginal contribution of the features to the $\mathcal{G}$ index

PROOF. Let us consider $x = \dfrac{n+1}{2} - P_i$ and $y = \dfrac{n+1}{2} - i$. From Definition 5, the level curves equations of the marginal contribution of the ith feature to the G index are:

$$\frac{2(n-i+1)}{n(n+1)}(P_i - i) = k,$$

for all features with non-negative performance-importance difference (otherwise the features do not contribute to the $\mathcal{G}$ index). By substituting P_i and i by their expressions in terms of x and y respectively, we obtain:

$$\frac{2(n-(\frac{n+1}{2}-y)+1)}{n(n+1)}((\frac{n+1}{2}-x)-(\frac{n+1}{2}-y))=k,$$

which is equivalent to:

$$\frac{2n-(n+1-2y)+2)}{n(n+1)}(y-x)=k,$$

Finally:

$$\frac{n+1+2y}{n(n+1)}(y-x)=k.\square$$

Figure 3 shows the level curves of the marginal contribution of the under-performing features to the $\mathcal{G}$ index over the IPA diagram partition in (Abalo et al 2007). Features in the same level curve are those with the same degree of under-performance, i.e., for each k the corresponding level curve contains features “with degree of under-performance k”. In Figure 3, level curves corresponding to $k = 0.05, 0.3, 0.6, 0.9, 1$ and 1.2 are represented.

This representation clearly improves the approach in (Abalo et al 2007) to determine the target features for resource allocation. The “Concentrate Here” zone of the diagram can be dynamically selected depending on the available resources and the admitted level of under-performance.

Two are the main differences between the $\mathcal{G}$ index and other well known correlation coefficients defined to compare rankings. On the one hand, the $\mathcal{G}$ index takes into account only under-performing features. On the other hand, since the $\mathcal{G}$ index is defined through an IOWA operator applied to the non-negative performance-importance differences of a firm, not all the features contribute to it in the same way. The more under-performing and the more important a feature is, the greater its contribution to the $\mathcal{G}$ index.

Let us highlight the advantages and disadvantages of our proposal in comparison with other existing IPA approaches. The IPA framework has been widely accepted due to its simplicity of calculations and intuitive graphical representation. From a computational point of view, the proposed method represents an improvement since the marginal contribution of each feature to the $\mathcal{G}$ index is determined. These marginal contributions provide information about how the current performance of a firm can be improved giving decision makers information about where to concentrate resources. From a graphical point of view, the innovative contribution of the proposed approach is that features can be drawn in a new diagram with the level curves of the marginal contribution of each feature to the $\mathcal{G}$ index, so managers can easily capture different levels of intensity regarding under-performed features.

As a possible drawback of the proposed method, we can note that $\mathcal{G}$ index compares attributes' importance and performance within a particular company. In a situation of limited information about competitors, it provides managers a framework to work with and to explore the strengths and weaknesses of the company. Nevertheless, the proposed method including the $\mathcal{G}$ index could be improved by adding measures of attributes' performance based on comparisons of products and services of either competing companies or the sector. In this direction, some extensions of IPA are reviewed in (Kim and Oh 2011). In particular, some approaches modify the original IPA by considering three or more dimensions, being competitors' performance one of them. These studies consider, instead of the four quadrants in the original IPA grid, either eight octants or even more different outcomes' areas. However, adding dimensions in the IPA grid implies loosing simplicity of attribute display and data interpretation.

In general in decision-making aid systems, one should note that there is no single method which outperforms all other methods in all aspects. However, the simplicity in user-interaction is, indeed, one of the main values that share most of the IPA methods, and it is closely related to the grid dimensionality.

A real-case application to the retail sector

In this section, an application of the proposed method to assess importance-performance in a Taiwan retail company is presented, after a brief introduction to the performance evaluation framework for the retail sector.

Evaluating performance in the retail sector

In recent years, the role of knowledge within strategic management has become the subject of substantial advances in research (Braz et al 2011; Chini 2004; Gherardi2006; Nonaka and Teece 2001; Teece 2000). Nevertheless, most of these studies relate to aspects of the transfer of knowledge rather than the application of knowledge in the evaluation of performance.

Despite the relative paucity of research in a retail context, the use of expert knowledge by managers is an important factor at a micro-level in the success of retailers and at the macro-level for sectorial re-structuring. Managers bring to bear their individual expert knowledge to solve problems at operational and strategic levels in the retail firm. The knowledge they hold and apply depends mainly on their perceptions of the levels of current performance and the levels of importance of specific features. An issue that arises, deriving from this view of the diversity of knowledge held by retail managers, is how to synthesize the individual perceptions of managers in ways that can be useful in strategic management. Thus, aggregating managerial opinions on the relative performance of some specific features and analyzing the contribution of these different features to the overall performance of the retailer are considered crucial.

In this research context, these individual and differing perceptions of the relevance of the various resources can be gathered through qualitative data collection. Given that managers will view differently the relative importance of the various features, a method to compare the opinions of managers and synthesize these qualitatively framed opinions would be useful.

In the next subsections, we conduct a full experiment that first includes the selection of relevant performance related

Table 1: The resource attributes used as variables in the evaluation procedures

Resource area	Resource concept	Number of features
Physical resource	Reach ability	2
Legal resource	Brand strength	2
Human resource	Human management	2
Organizational resources	Expansion ability	2
	Productivity	2
	General management	2
	Technology management	2
	Organizational management	2
	Inventory management	2
	Marketing management	2
	Financial management	2
	Product innovation	2
	Loan repay ability	3
	Diversification	1
Informational resources	Market segment risk	2
	Strategic vision	2
Relational resources	Stakeholder relations	3
External factors	Actions from outside stakeholders	3
	Political environmental	2
	Technological environmental	2
	Socio-culture environmental	2

variables. Secondly, we present a survey of senior managers that measures their perceptions of the importance and performance of the selected variables, based on an order-of-magnitude qualitative model. Thirdly, the ranking method detailed in (Agell et al 2012), is applied to obtain rankings of the selected variables, aggregating expert opinions with respect to importance and performance respectively. Finally, the global index $\mathcal{G}$, together with the iso-curves of the feature contribution to the index introduced in Section 3, is used to summarize the differences in these rankings and identify features to which resources should be allocated.

Design of the empirical study and data collection

A study involving senior managers as experts was undertaken in a major chain store organization. President Chain Store Corporation is a multinational retailer based in Taiwan that operates a multi-format strategy through a range of organizational structures. It is the largest retailer in Taiwan. Using literature surveys and 25 in-depth interviews with a cross-section of retailer stake-holders, 170 performance-related variables relevant to retailing were identified. From this list, after rationalization and classification in terms of the nature of the resource, 44 features or variables related to resources used in retailing were selected as the main performance variables. The selection was undertaken by reference to the views of interviewees and research literature on resource based theories of the firm. Seven resource areas were established within these 44 features, as shown in Table 1.

A survey was then undertaken with managers in the Taiwan head office. Data was collected from 84 senior managers across all the managerial functions. Managers were divided into five main groups depending on broad functional area: marketing (15); operations and store operations (17); accounting, finance and audit (24); R&D and information systems (14); and other (e.g. human resources, law) (14).

Managers were asked to use their expertise to assess each of the 44 variables in terms of their perceived importance to the performance of the firm. An ordinal scale of 1 to 4 was used as: (1) extremely important; (2) very important; (3) moderately important; (4) not very important; with (5) as ”don't know”. The managers were asked to repeat the exercise in terms of the perceived performance of the firm based on the same variables, with the scale being: (1) extremely good (or extremely strong); (2) very good (or very strong); (3) moderately good (or moderately strong); (4) not very good (or not very strong); with (5) again used as ”don't know”.

Data analysis and results

This subsection is devoted to analyzing and comparing the evaluations of importance and performance of the 44 features in Table 1. Using the ranking method described in (Agell et al 2012) the features were ranked with respect to their importance and with respect to their performance from the responses from all 84 experts.

In this case, the non-negative performance-importance differences vector of the firm is the 44-dimensional vector:
$DV(F) = (10, 12, 1, 4, 10, 13, 13, 0, 3, 16, 0, 0, 10, 27, 14, 0, 0, 12, 3, 0, 6, 11, 1, 0, 0, 10, 16, 0, 0, 1, 0, 7, 0, 3, 0, 0, 0, 0, 3, 0, 0, 0, 1, 0)$.

Then, weights are computed using Borda-Kendall law, obtaining $w_i = \dfrac{45-i}{990}$ for all $i = 1, \ldots, 44$. With these values, the $\mathcal{G}$ index introduced in Section 3 to compare rankings with respect to importance and performance is computed and produces a global importance-performance index $\mathcal{G}(DV(F)) = 6.329$. Taking into account that the ideal best performing firm has $\mathcal{G}(DV^*) = 0$ and the firm in the opposite situation has $\mathcal{G}(DV_*) = 18.167$, as proven in Proposition 1, there is therefore a significant divergence between the two considered rankings (corresponding to about one third of the range of variation, precisely a 34.8%). This fact shows that there is room for resource allocation improvement. Note that, similar conclusions can be obtained when we compute other well-known correlation coefficients, such as Kendall's Tau or Spearman's Rho, for the same pairs of importance-performance rankings. In these two cases we obtain 0.378 and 0.506 respectively.

The comparison of the two rankings given by our method and shown in Figure 4 also points out the directions for this improvement. The added value of our contribution to the comparison of both rankings is the combination of the $\mathcal{G}$ index and the level curves of the marginal contribution of the features to this index. In Figure 4 an example of the level curve corresponding to $k = 0.3$ is depicted (see Proposition 3).

As detailed in Proposition 2, among the 44 features selected, the 24 features that plot above the principal diagonal are those that contribute to the $\mathcal{G}$ value of the firm. These are

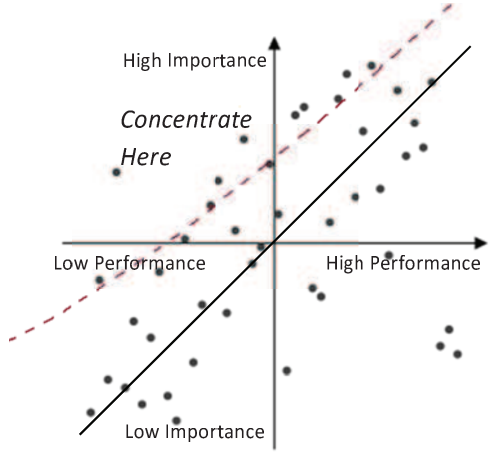


Figure 4: Representation of features with respect to managers' perceptions of importance and performance

aspects of the firm that are perceived by managers as under-performing and coincide with aspects in the "Concentrate Here" region defined in (Abalo et al 2007). Similarly, Figure 4 shows the region labeled as "Concentrate Here" in the Martilla's classical IPA diagram, which contains $seven$ features.

In addition, in this paper, as explained in Section 3, we propose a step forward in understanding which features may be improved. Beyond the IPA diagram, we suggest concentrating resources in those features that contribute most to the $\mathcal{G}$ value of the firm. In Figure 4, these features have been visualized over the dotted line for the case $k = 0.3$. This line is the iso-curve of the marginal contribution of the features to the $\mathcal{G}$ index in the IPA diagram corresponding to $k = 0.3$ (see Proposition 3). Visually, most of the contribution to the $\mathcal{G}$ index can be seen as focussing on a limited number of features. These 10 extreme values are listed in Table 2.

Table 2: Variations in the ranking of expert managers when importance is ranked much higher than performance

Features	Ranking of importance	Ranking of performance	Contribution to $\mathcal{G}$
Market positioning	1	11	0.444
Number of customer visits	2	14	0.521
Customer complaints management	5	15	0.404
Sales per store	6	19	0.512
Store opening strategy	7	20	0.499
Franchise system	10	26	0.566
Spending-per-visit rate	13	23	0.323
Staff training	14	41	0.845
Quality of data collection and process system	15	29	0.424
Innovation of new technology equipment	18	30	0.327

Most are directly or indirectly associated with firm growth. Six out of the ten relate directly to organizational resources, three relate to physical, human, and relational resources respectively, and the final one relates to external factors. Note that in this case, the value $k = 0.3$ has been used, however depending on the available resources, different values of k could be considered.

Discussion and managerial implications

Hansen and Bush pointed out that IPA is a simple and effective technique that can assist in identifying improvement priorities (Hansen and Bush 1999). IPA has been applied as an effective means of evaluating a firm's competitive position in the market, identifying improvement opportunities, and guiding strategic planning efforts. However, typically, managers must work with limited resources in competitive business environments. For this reason, the proposed method, able to decide how to best allocate scarce resources in order to maximize importance-performance, is very helpful.

The results of the empirical testing of the method show how to identify areas of perceived under-performance of the firm. In our real case, 44 features related to resources used in retailing were selected as main performance variables. Managers in the President Chain Store Corporation then evaluated the perceived importance and the perceived performance of the firm for these 44 features. From these evaluations, the features were ranked with respect to these two concepts. The proposed $\mathcal{G}$ index is computed, and the iso-curves of the marginal contribution of the features to the $\mathcal{G}$ index enabled recognition of the perceived under-performing features of the firm. The method used, by taking into account the qualitative perceptions held by managers, provides a useful tool for decision making for the retailer.

Considering the iso-curve of the marginal contribution to the $\mathcal{G}$ index as corresponding to a contribution of $k = 0.3$, ten features appeared as being under-performing in that firm, thus they can potentially be improved. This level of contribution ($k = 0.3$) corresponds, as a percentage, to 4.7% of the $\mathcal{G}$ index. As we can see in Table 2, the "staff training" feature, which belongs to the human resources area, is perceived as the most under-performing feature, contributing more than 13% (0.13351 = 0.845/6.329) to the $\mathcal{G}$ index. There are seven features whose contribution to the $\mathcal{G}$ index varies between 6.4% and 9%, with two features contributing about 5.1% each. The remaining under-performing features, below the considered iso-curve, contribute less than 4.7% each to the $\mathcal{G}$.

As stated, when modifying the value of k, a different number of features for focus would be obtained. The strength of the method proposed is its adaptable nature, which helps managers to improve the efficiency of the firm. Therefore, the $\mathcal{G}$ index could be considered as a valuable decision-support tool to better allocate resources within the firm.

Conclusions and future research

This paper contributes to improving importance-performance analysis by providing a new measure that captures the overall relationship between importance and performance. This measure is obtained by considering the relevant features that describe a firm and so enable a firm's managers to improve decision-making in resource allocation. The developed method, together with a new version of the classical IPA diagram, enables managers to assess a firm's overall performance and detect features where resources should be allocated. The presented global importance-performance index ($\mathcal{G}$), inspired by OWA operators, is a weighted sum of the non-negative performance-importance differences, where weights depend on the importance of the feature.

Moreover, the $\mathcal{G}$ index also leads to an enhancement of the IPA diagonal-based scheme with a new representation: Contribution-to-$\mathcal{G}$ iso-curves. These level curves show a more accurate picture of the most-needed-investment features, and determine a new "Concentrate Here" zone in the classical IPA diagram. A real-case application in the retail sector has been used to show that the presented method can lead to a more accurate importance-performance analysis of a firm's situation. The real-case application gives us an example of how $\mathcal{G}$ could benefit managerial decision-making processes in resource allocation.

As future work, a marginal sensitivity analysis of the $\mathcal{G}$ index incorporating changes in resource allocation would be a major future contribution for decision-making processes. It could be of interest in a more advanced study of $\mathcal{G}$ properties to determine the upper-boundary of the index for relative comparisons of company performances. Additional analysis that separately considers the functional area of managers could be performed to infer how the area of expertise influences perceptions and modifies the $\mathcal{G}$ index.

References

Abalo, J., Varela, J. and Manzano, V. (2007). Importance values for importance-performance analysis: A formula for spreading out values derived from preference rankings. *Journal of Business Research* 60, 115-121.

Agell, N., Sánchez, M., Prats, F. and Roselló, L. (2012), Ranking multi-attribute alternatives on the basis of linguistic labels in group decisions. *Information Sciences* 209: 49-60.

Braz, R., Scavarda, L. and Martins, R. (2011). Reviewing and improving performance measurement systems: An action research. *International Journal of Production Economics* 133(2): 751-760.

Chiclana, F., Herrera-Viedma, E., Herrera,F., Alonso, S.(2007). Some Induced Ordered Weighted Averaging Operators and Their Use for Solving Group Decision-Making Problems based on Fuzzy Preference Relations. *European Journal of Operational Research* 182(1): 383-399.

Chini, T. (2004). *Effective Knowledge Transfer in Multinational Companies*. Palgrave Macmillan, Basingstoke, U.K.

Deng, W. (2007). Using a revised importanceï¿$\frac{1}{2}$performance analysis approach: The case of Taiwanese hot springs tourism. *Tourism Management* 28: 1274-1284.

Gherardi, S. (2006). *Organizational Knowledge*. Blackwell, Oxford, U.K.

Hansen, E. and Bush, R.J. (1999). Understanding customer quality requirements. Model and Application. *Industrial Marketing Management* 2: 119-130.

Herrera, F., Herrera-Viedma, E. (1997). Aggregation Operators for Linguistic Weighted Information. *IEEE Transactions on Systems, Man and Cybernetics* 27: 646-656.

Herrera-Viedma, E., Pasi, G., López-Herrera, A.G., Porcel, C. (2006) Evaluating the Information Quality of Web Sites: A Methodology Based on Fuzzy Computing with Words. *Journal of American Society for Information Science and Technology* 57(4): 538-549.

Kim, B-Y. and Oh, H. (2011). An Extended Application of Importance-Performance Analysis. *Journal of Construction Engineering and Management* 137(12): 1142-1152.

Kendall, M. (1962) *Rank Correlation Methods*, Hafner, New York, USA.

Martilla, J.A. and James, J.C. (1977). Importance-Performance Analysis. *Journal of Marketing* 41(1): 77-79.

Nonaka, I. and Teece, D. (2001). *Managing Industrial Knowledge*. Thousand Oaks: Sage.

Sayeras, J.M., Agell, N., Rovira, X., Sánchez, M., Dawson, J.A. A measure of perceived performance to assess resource allocation*Soft Computing*. DOI 10.1007/s00500-015-1696-3

Teece, D. (2000). *Managing Intellectual Capital: Organizational, Strategic and Policy Dimensions*. Oxford: Oxford University Press.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems Management, Cybernetics* 18: 183-190.

Yager, R. R., Filev, D.P.(1999). Induced ordered weighted averaging operators, *IEEE Transaction on Systems, Man and Cybernetics* 29: 141-150.

Eclipse in Occlusion
A perspectival mereotopological representation of celestial eclipses

Paulo E. Santos
AI and Automation lab.
Centro Universitário FEI,
S. Paulo, Brazil

Roberto Casati
Institut Jean Nicod,
ENS-EHESS-CNRS,
Paris, France

Hannah M. Dee,
Computer Science,
Aberystwyth University,
SY23 3DB, Wales, UK.

Carl Schultz, Mehul Bhatt,
University of Münster
University of Bremen
Germany

Abstract

This paper presents a formalisation and exploration of the concept of *eclipse* from the perspective of qualitative spatial reasoning. Building upon theories of spatial connectivity, occlusion and shadows, we show that eclipses can be described and reasoned about using mereotopological relations, and that these formalisms can be used to disprove some commonsense misconceptions about the nature of celestial phenomena.

Introduction

Qualitative Spatial Reasoning (QSR) (Ligozat 2011) is the subfield of Knowledge Representation in Artificial Intelligence that develops and applies formal representations of qualitative knowledge about spatial phenomena. QSR formalisms have found a range of applications in areas such as Geographical Information Systems, Architecture Design, Cognitive Vision and Robotics etc (Cohn and Renz 2008; Bhatt et al. 2011; Bhatt, Schultz, and Freksa 2013). However, to the best of our knowledge, qualitative theories about space have not been significantly applied in education, e.g., specifically as a tool for checking students' conceptions of the physical world, or in the modeling of astronomical events. In this paper we describe the initial steps towards generating a formalisation of a basic astronomical event (eclipses) that could be used in an autonomous tutorial system to both interact with the student's understanding of the concept and to address particular learning errors.

Several authors point out an apparent prevalence of alternative conceptions about the causes of natural events that contradict basic scientific knowledge, even after formal instruction on the subject (Libarkin and Kurdzie 2001). This has been linked to current educational theories which claim that people obtain new knowledge based on their existing beliefs (D.Bransford, L.Brown, and R.Cocking 2000) and, when existing beliefs clash with new knowledge, the former prevails over the latter. In naïve astronomy, there is a common misconception that the phases of the Moon are due to lunar eclipses (Bailey and Slater 2004). The formalisation presented below, although preliminary, provides a way to mitigate this common misconception.

The work presented here belongs to a family of QSR formalisms which make explicit the notion of viewpoints in their ontologies (as described in Section § **Related Work**). Here we apply one such method to model a few basic concepts (linking occlusion relations and visual appearance) of eclipses as described in Section §**Eclipses**. The formalism used in the paper is an occlusion calculus defined upon a mereotopology as introduced in Section §**Region Occlusion Calculus**. The main results of this paper are presented in Section §**A qualitative formalisation of eclipses**. The formalism thus defined is implemented within the Constraint Logic Programming system CLP(QS) in our penultimate section.

Related Work

Much work in Qualitative Spatial Reasoning does not explicitly model the observer's viewpoint. Without including a *point of view* or *observer location* as one of the variables in a formalisation, a theory is limited in its capacity to model perception. This limitation precludes the ability to reason about concepts and inferences involved in naïve astronomy.

There are, however, a few QSR formalisms that consider viewpoints when accessing whether a particular spatial relation holds or not. Most of these formalisms have spatial occlusion (or motion parallax) as a key aspect of their ontology. Spatial occlusion occurs when an object is located between another object and the observer's viewpoint; it is one of the primary cues used by the human perceptual system to construct a 3D interpretation of the visual world as it provides an estimate of relative distances (Randell, Witkowski, and Shanahan 2001). Perhaps the first qualitative formalisation of spatial occlusion was proposed in (Petrov and Kuzmin 1996) where a set of axioms is designed to constrain a point-based notion of occlusion. Assuming 2D convex objects, rather than points, (Galton 1994) proposes the *Lines-of-Sight* calculus that represents the relative positions between pairs of bodies as seen from a viewpoint. Based on this idea, the *Region Occlusion Calculus* (ROC) (Randell, Witkowski, and Shanahan 2001) defines occlusion and image parallax within a mereotopological theory. More recently, (Guha, Mukerjee, and Venkatesh 2011) proposed a set of 14 occlusion relations making explicit the distinctions of whether the observed objects are fragmented or not, and whether the occluder is a moving object or part of the background. (Tassoni et al. 2011) develops a 3D qualitative formalism about visibility that has occlusion at its kernel. (Bhatt, Lee, and Schultz 2011) propose a declarative spatial resoning system where mixed qualitative-quantitative spatial reasoning may be (declaratively) performed in a logic programming setting; the system can be used to formally reason about a range of spatial representations, including for instance, visibility and occlusion relations such as in (Tassoni et al. 2011).

Occlusion has also been recently defined within an interval algebra in (Ligozat, Santos, and Samghabadi 2015) providing the appropriate tools for operating consistency checking in qualitative constraint networks.

In previous work, we proposed a dynamic formalism about occlusion in which qualitative changes observed by a mobile robot are the building blocks of the system (Santos 2007). Later, (Souchanski and Santos 2008) refined this formalism within a *reasoning about actions and change* framework, in order to facilitate autonomous inferences about the behaviour of other agents (as observed from a viewpoint). *Region Occlusion Calculus* has recently been applied within a Bayesian filter in (Santos et al. 2016) to generate an efficient algorithm for qualitative self-localisation for a mobile robot.

Although there is some work on formalising the perception of shadows by means of a Qualitative Spatial Reasoning theory (Santos, Dee, and Fenelon 2009), there is currently no literature describing attempts to represent the observation of astronomical phenomena, such as eclipses, within a QSR framework. The representation and simulation of physical systems at a qualitative level has long been one of the goals of the Qualitative Reasoning (QR) field (Bredeweg and Struss 2004; Kleer 1990). However, to the best of our knowledge, celestial eclipses (or other astronomical occlusion-related events) have not yet been targeted by this community. These phenomena can be seen as a special case of shadow casting, and this paper is our attempt to provide the framework for describing and reasoning about them.

Eclipses

Since antiquity, eclipses have been well-understood astronomic phenomena. Periodical patterns of eclipses were known to the Babylonians and led in some cases to successful predictions of the occurrence and the type of a lunar eclipse (Neugebauer 1952). A full mathematical characterisation (and prediction of both lunar and solar eclipses and, in the latter case of the precise visibility area) was possible only after the work of Newton, however. One particular contingency makes solar eclipses geometrically interesting: the ratio Moon-diameter/Moon-distance from Earth is approximately the same as the ratio Sun-diameter/Sun-distance from Earth, which means that the Moon can almost perfectly occlude the Sun (in this case we say that it is an eclipse of *magnitude one*). Due to the slightly elliptical nature of orbits, these distances fluctuate. So at times the Moon may more than completely hide the Sun as it is located on its orbit at a distance smaller than the one required for the perfect occlusion ratio. The extent of the occlusion accounts for the length of the eclipse (this is the case of an eclipse of magnitude greater than one). In other cases, the Moon is more distant and thus cannot fully occlude the Sun, giving rise to annular eclipses (this is an eclipse of magnitude less than one). In an annular eclipse the Moon obscures the centre of the solar disk, but not the whole disk, giving a remarkable bright halo effect sometimes called the "*ring of fire*".

The various states of a solar eclipse are shown in Figure 1. A partial eclipse of the sun occurs when the observer is located at the penumbra region (depicted in blue in Figure 1), a total eclipse occurs when the observer is in the umbra region (coloured in brown in Figure 1) and an annular eclipse happens when the observer is in the antumbra region (depicted in the colour green). When the observed images of the Moon and the Sun are seen as in contact for the first time is an event known as *First Contact*; the case when the Moon and the Sun are in contact for the last time during an eclipse is called Fourth contact. Second and Third contacts occur when the borders of the Sun and Moon images (as seen from a viewpoint) meet for the second and third times (as shown in Figure 1). These events will be defined in terms of mereotopological concepts below.

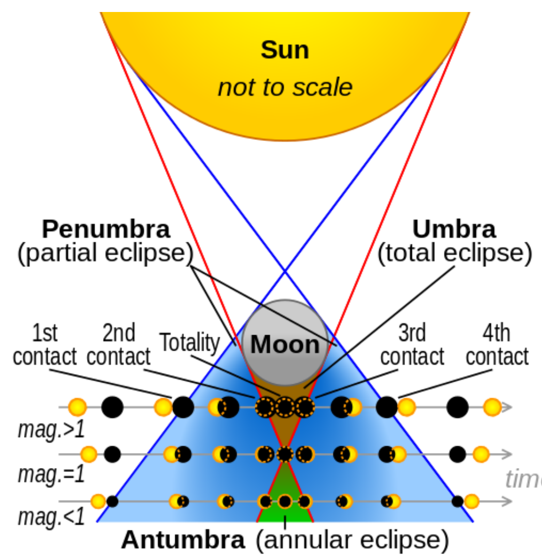


Figure 1: Visualisations of an eclipse. Figure adapted from `https://en.wikipedia.org/wiki/Eclipse`, accessed on Feb. 23, 2016.

Region Occlusion Calculus

The basic spatial theory used in this work is the Region Occlusion Calculus (ROC) (Randell, Witkowski, and Shanahan 2001), which is an extension of the Region Connection Calculus (RCC) (Randell, Cui, and Cohn 1992). RCC is a first-order axiomatisation of spatial relations based on a reflexive, symmetric and non-transitive dyadic primitive relation of *connectivity* ($C/2$) between two spatial regions. Informally, assuming two regions x and y, the relation $C(x, y)$, read as *"x is connected with y"*, is true if and only if the closures of x and y have at least one point in common.

Assuming the $C/2$ relation, and two spatial regions x and y, the following base relations can be defined: *disconnected from* (DC), *part of* (P), *equal to* (EQ), *overlaps* (O); *partially overlaps* (PO); *externally connected* (EC); *tangential proper part* (TPP); *non-tangential proper part* ($NTPP$). RCC also includes the inverse relations of P, TPP and $NTPP$, which are represented by a capital 'I' appended to the relative relation: PI, $TPPI$ and $NTPPI$.

The set constituted by the relations DC, EQ, PO, EC, TPP, $NTPP$, $TPPI$, and $NTPPI$ is the jointly exhaustive and pairwise disjoint set (JEPD) usually referred to as RCC8. The continuous transitions between the RCC8 relations, for two regions x and y, are shown as a *conceptual neighbourhood diagram* (CND) in Figure 2. By continuous transitions we mean that in between adjacent vertices of the graph there can be no other possible relation qualifying the state of the two regions. That is, assuming that the objects move continuously on the plane, these are the only transitions that are possible.

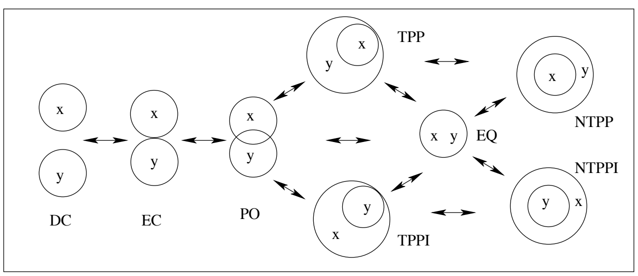


Figure 2: The RCC8 relations and their conceptual neighbourhood diagram (Randell, Cui, and Cohn 1992).

Let a and b be two physical (possibly non-convex) bodies, and ν an observer viewpoint. Using RCC8 relations, along with the primitive relation $TotallyOccludes(a, b, \nu)$ (which stands for "a totally occludes b with respect to the viewpoint ν"), the Region Occlusion Calculus (ROC) (Randell, Witkowski, and Shanahan 2001) defines the 20 base JEPD relations representing the various occlusion relations between two bodies. ROC distinguishes the occupancy regions of bodies and their images (or projections) from the viewpoint of an observer by assuming two functions: the function $region(a)$, which maps a body a to its 3D occupancy region, and the function $image(a, \nu)$ that maps a body a to the body's 2D projection, as seen from a viewpoint ν. The viewpoint in ROC is modelled as a pinhole camera whose parameters are not relevant here.

Figure 3 shows a graphical representation of the ROC relations between two bodies, represented as a white and a shaded region. In this figure, the shaded region corresponds to the first argument, and the white region to the second argument of ROC relations. For instance, the relation $PartiallyOccludesTPP(a, b)$ is depicted with the shaded object a occluding the white object b, while the 2D projection of the shaded object is a tangential proper part (TPP) of the 2D projection of the white object. It is worth noting that the relations on *mutual occlusion* only occur if and only if at least one of the objects is non-convex. ROC also defines a conceptual neighbourhood diagram (introduced in (Randell and Witkowski 2002)) that we do not present in this paper for brevity.

It is worth pointing out also that the "I" in the relations $TotallyOccludesTPPI(a, b, \nu)$ and $TotallyOccludesNTPPI(a, b, \nu)$ represents the inverse of TPP and $NTPP$, respectively; so, for instance, $TotallyOccludesTPPI(a, b, \nu)$, means that the body a totally occludes the body b, but $image(b)$ is the tangential proper part of $image(a)$ (i.e., $TPPI(image(a, \nu)$,-$image(b, \nu))$). The superscript "-1" in some ROC relations represents the inverse of the occlusion part of the relation.

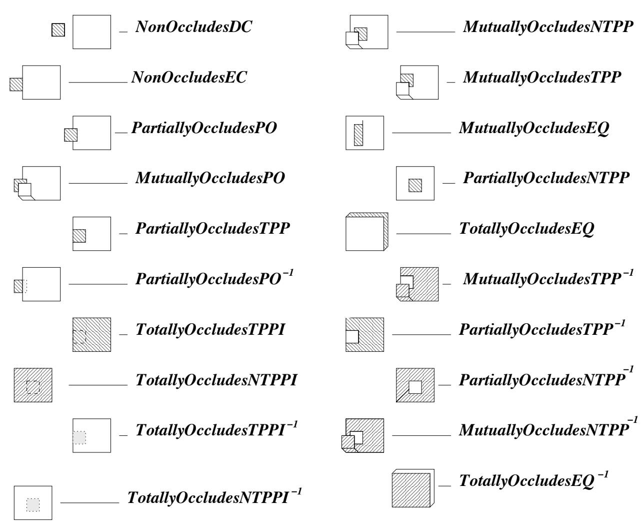


Figure 3: ROC relations between two objects (white and shaded regions).

As we are dealing with eclipses and therefore with celestial bodies, there are certain ROC relations which cannot hold in our situation. These relations are those which involve mutual occlusion.

Relative Positions

As well as the 20 ROC relations, this work assumes observer-relative positions of pairs of objects by means of the relations $Left$ and $Right$. Given two distinct bodies a and b and a viewpoint ν (where ν, in this case, is any viewpoint on Earth observing the eclipse) the relative positions between a and b with respect to their centroids are as follows:

- $Left(a, b, \nu)$, representing the fact that "a is on the left of b from the viewpoint ν";

Table 1: ROC relations representing eclipse states. The numbers in the left column are abbreviations used in the remainder of this paper.

1	$NonOccludesDC(M,S,v) \wedge Right(M,S,v)$.
2	$NonOccludesEC(M,S,v) \wedge Right(M,S,v)$.
3	$PartiallyOccludesPO(M,S,v) \wedge Right(M,S,v)$.
4	$PartiallyOccludesPO(M,S,v) \wedge Left(M,S,v)$.
5	$NonOccludesEC(M,S,v) \wedge Left(M,S,v)$.
6	$NonOccludesDC(M,S,v) \wedge Left(M,S,v)$.
7	$TotallyOccludesEC(M,S,v)$.
8	$PartiallyOccludesTPP(M,S,v) \wedge Right(M,S,v)$.
9	$PartiallyOccludesNTPP(M,S,v)$.
10	$PartiallyOccludesTPP(M,S,v) \wedge Left(M,S,v)$.
11	$TotallyOccludesTPPI(M,S,v) \wedge Right(M,S,v)$.
12	$TotallyOccludesNTPPI(M,S,v)$.
13	$TotallyOccludesTPPI(M,S,v) \wedge Left(M,S,v)$.

- $Right(a,b,\nu)$, representing the fact that "a is on the right of b from the viewpoint ν".

A qualitative formalisation of eclipses

Qualitative reasoning about eclipses proves to be a difficult task (Sorensen 1999). It may even be at the heart of some inconsistent commonsense reasoning, such as the common belief that the phases of the Moon are caused by the shadow cast by the Earth on the Moon (Barnett and Moran 2002; Bailey and Slater 2004). In this section we tackle qualitative reasoning about eclipses by means of the Region Occlusion Calculus described in the previous section.

We first assume two constants: M and S representing, respectively, the Moon and the Sun. With these constants, and the ROC relations described above, the qualitative states of a solar eclipse can be described by the conceptual neighbourhood diagram shown in Figure 4, where the dark object represents the Moon and the bright one, the Sun. The numbers assigned to each state in Figure 4 are abbreviations of conjunctions of ROC and Relative Position relations representing eclipse states, as shown in Table 1. In all of these diagrams and formulae the viewpoint v, if not explicitly stated, is assumed to be a location upon the surface of the Earth.

It is worth noting that the transition from 3 to 4 in Figure 4 represents a partial eclipse; the transitions from 3 to 7 and then from 7 to 4 represent an eclipse of magnitude equal to 1 (cf. Figure 1). Similarly, the transitions $3 \rightarrow 8 \rightarrow 9 \rightarrow 10 \rightarrow 4$ and $3 \rightarrow 11 \rightarrow 12 \rightarrow 13 \rightarrow 4$ represent, respectively, eclipses of magnitude greater than 1 ($mag. > 1$) and less than 1 ($mag. < 1$).

This formalisation includes all the states of an eclipse, as shown in Figure 1 and also represented in Figure 5 (the numbers in Figure 5 stand for the relations in Table 1).

Figure 5 also provides us an opportunity to define the astronomical terms in Figure 1 by means of ROC relations as follows:

- *Penumbra (or partial eclipse)* happens in the region where $PartiallyOccludesPO(M,S,v) \wedge Right(M,S,v)$ or $PartiallyOccludesPO(M,S,v) \wedge Left(M,S,v)$;

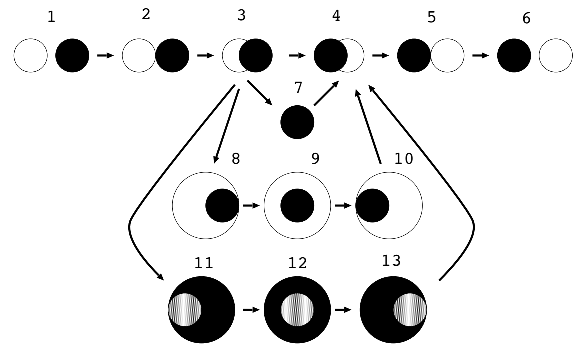


Figure 4: Solar eclipse as occlusion relations.

- *Umbra (or total eclipse)* happens in the region where $TotallyOccludes(M,S,v)$ holds (i.e. in 7, 11, 12 and 13);
- *Antumbra (or annular eclipse)* happens in the region where $PartiallyOccludesTPP(M,S,v) \wedge Right(M,S,v)$, or $PartiallyOccludesNTPP(M,S,v)$, or $PartiallyOccludesTPP(M,S,v) \wedge Left(M,S,v)$;
- *First Contact* can be defined as $NonOccludesEC(M,S,v) \wedge Right(M,S,v)$ (state 2 in Figure 5);
- *Second Contact* can be defined as $TotallyOccludesTPPI(M,S,v) \wedge Right(M,S,v)$ (state 11 in Figure 5);
- *Third Contact* can be defined as $TotallyOccludesTPPI(M,S,v) \wedge Left(M,S,v)$ (state 13 in Figure 5);
- *Fourth Contact* can be defined as $NonOccludesEC(M,S,v) \wedge Left(M,S,v)$ (state 5 in Figure 5).

Thus far we have shown that the formalisation here encompasses the phenomena of solar eclipses. With a minimal amendment, we can also accommodate lunar eclipses. The ROC relations in Table 1 are described from the perspective of an observer on the surface of the Earth.

To model a lunar eclipse we need to consider shadows more explicitly. In a lunar eclipse, the Moon is in the shadow of the Earth, and no (pointlike) light source can see the shadows cast by objects that intercept light rays emanating from it (as observed by Da Vinci). Thus we have an analogous situation (the Earth occludes the Sun with respect to the Moon).

As an instance of a qualitative derivation within the proposed calculus, with the formalisation introduced above it is straightforward to derive the negation of the common belief that the phases of the Moon are the result of eclipses. Assume, reasoning by *Reductio ad absurdum*, that the phases of the Moon are indeed due to lunar eclipses, then the Moon (M) and the Earth (E) must be in one of the occlusion relations $\{3,4,7,8,9,10,11,12,13\}$ with respect to

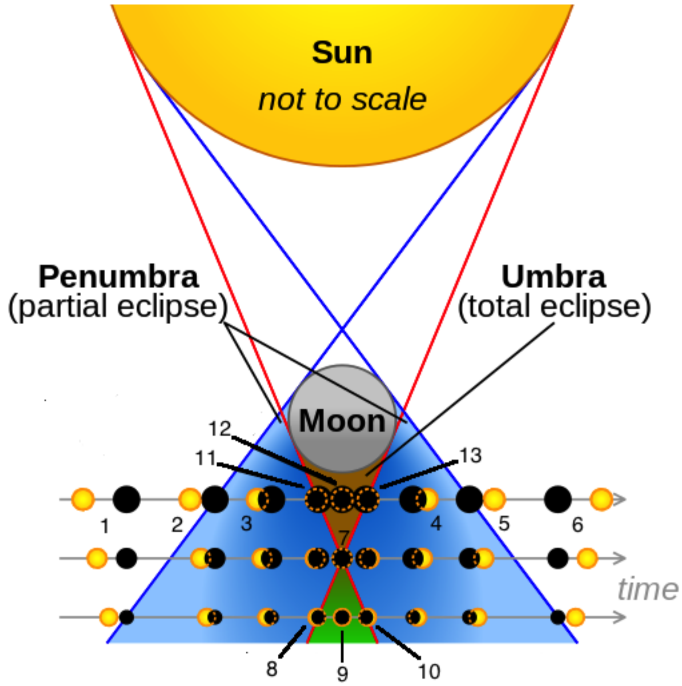


Figure 5: Solar Eclipse as occlusion relations: regions labelled with their ROC eclipse states from Table 1.

the illuminating surface of the Sun (viewpoint ν). Thus, from the definitions of the ROC relations we have that $C(image(M,\nu), image(E,\nu))$ (as presented in Section §*Region Occlusion Calculus*). However, in the case of a lunar phase it is a fact that $\neg C(image(M,\nu), image(E,\nu))$. The concluding step is a direct result of a qualitative model of lunar phases (not described in this paper).

A Declarative Implementation with CLP(QS)

In this section we present our implementation of our qualitative solar eclipse model in the Constraint Logic Programming system CLP(QS) (Bhatt, Lee, and Schultz 2011; Schultz and Bhatt 2012). Our implementation provides two key features:

- *Intelligent diagrams*: users can manipulate the objects in the diagram, and the system automatically updates other objects so that the qualitative spatial relations are maintained at all times;
- *Spatial Question / Answering*: users specify state constraints at both the domain level (solar eclipse states) and qualitative spatial level (topological and orientation relations between the Sun and the Moon) and CLP(QS) determines whether the constraints are consistent, and updates the intelligent diagram accordingly.

The interactive and dynamic aspects of intelligent diagrams (also referred to as *dynamic geometry*) makes them highly attractive for use in education domains such as teaching high-school level geometry (Winroth 1999). We implement our qualitative model in CLP(QS) and generate intelligent diagrams from two perspectives: (1) *top-down* and (2) *from-Earth*.

We define facts and rules for referring to solar objects from different perspectives, e.g. the following query specifies that the Moon and Sun in the top-down perspective are topologically *disconnected*:

```
object(type(moon), perspective(top_down), Moon),
object(type(sun),  perspective(top_down), Sun),
topology(rcc(dc), Moon, Sun).
```

For brevity, in the following we omit these type casting predicates when there is no ambiguity about perspective.

Top-down perspective. In the *top-down* perspective Earth is a *point*, the Moon is a *circle*, and the Sun is a *circle*. The Moon's orbit is a circle centred on Earth, such that the centre of the Moon is coincident to the orbital circle. Earth's orbit is a circle concentric with the Sun, such that Earth is coincident to the orbit.

```
point(Earth),
circle(MoonOrbit),
centre(MoonOrbit, Earth),

circle(Sun),
circle(SunOrbit),
incidence(concentric, SunOrbit, Sun),
size(larger, SunOrbit, Sun),

circle(Moon),
centre(Moon, MoonCentre),
incidence(coincident, MoonCentre, MoonOrbit),
incidence(coincident, Earth, SunOrbit),
size(smaller, Moon, Sun),
```

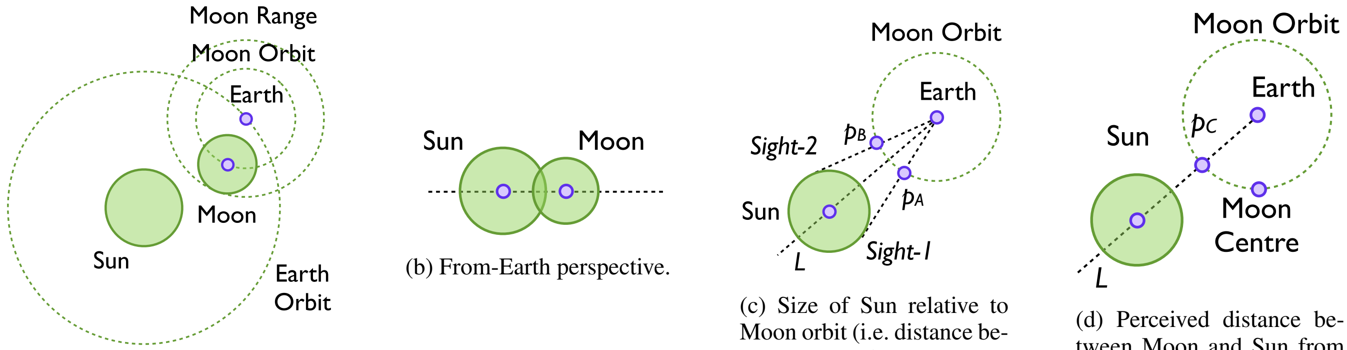
We add further constraints on the relative size and topology of our solar objects and their orbits. Firstly, the Earth is exterior to the Moon. Secondly, we want that the Moon and Sun never overlap. We thus define a Moon range circle, concentric with the Earth, such that the Moon is a tangential proper-part, and the Sun is disconnected (Fig. 6a).

```
incidence(exterior, Earth, Moon),
circle(MoonRange),
centre(MoonRange, Earth),
topology(rcc(tpp), Moon, MoonRange),
topology(rcc(dc), Sun, MoonRange),
```

From-Earth perspective. In the *from-Earth* perspective the Moon and Sun are *circles* such that the line between their centroids is horizontal (Fig. 6b).

```
circle(Moon), circle(Sun),
centre(Moon,MoonCentre),
centre(Sun,SunCentre),
orientation(horizontal,line(MoonCentre,SunCentre)),
```

Connecting perspectives. To qualitatively relate the top-down and from-Earth perspective we need to relate the size and relative positions of the solar objects. That is, when we manipulate the objects in one diagram we want the objects



(a) Top-down perspective.

(b) From-Earth perspective.

(c) Size of Sun relative to Moon orbit (i.e. distance between p_A, p_B).

(d) Perceived distance between Moon and Sun from Earth.

Figure 6: Implementing solar eclipse model in CLP(QS).

in the other diagram to also correctly change (e.g. if we resize and move the Sun in the top-down diagram then the Sun in the from-Earth diagram should also automatically change so that the diagrams remain consistent with each another).

Firstly, we set the Moon size in both perspectives to be equal.

```
object(type(moon), perspective(top_down), MoonTopDown),
object(type(moon), perspective(from_earth), MoonFromEarth),
size(equisized, MoonTopDown, MoonFromEarth).
```

Next we constrain the relative Sun size, that is, we need to relate the radii of the circles representing the Suns in the top-down and from-Earth perspectives. We can not simply constrain the Suns' radii to be equal, otherwise the Sun in the from-Earth perspective would be far too large relative to the Moon. Instead, we need to express the *perceived* diameter of the Sun relative to the Moon in the top-down perspective, and use this diameter to equal the diameter of the Sun in the from-Earth perspective.

To accomplish this, in the top-down diagram we add two lines-of-sight from Earth to either side of the Sun; these represent the left-most and right-most points of the Sun that are visible from Earth, respectively. Consider the points p_A, p_B where these lines-of-sight intersect the Moon's orbit (Fig. 6c). The length of the arc between p_A, p_B along the Moon orbit circle is the perceived diameter of the Sun relative to the size of the Moon. Thus, the diameter of the Sun in the from-Earth perspective is set to equal the length of the arc along the Moon orbit circle between points p_A, p_B.

Let L be the line between the Earth and the Sun's centre point in the top-down perspective. We define two sight lines (Sight-1, Sight-2) such that (a) they are tangent to the Sun, (b) one of the end-points of each sight line is coincident to Earth, (c) the other end-point is both coincident to the Sun, and either to the left or right of L, respectively for each sight line. We then define intersection points p_A, p_B between the Moon's orbit and each sight line. Let r be the radius of the Moon orbit circle. The length of the arc of radius r, centred on the Earth, from p_A to p_B is equal to the diameter of the Sun in the from-Earth perspective (Fig. 6c).

```
object(type(sun), perspective(top_down), SunTopDown),
object(type(sun), perspective(from_earth), SunFromEarth),
...
L = line(Earth, SunCentreTopDown),
Sight1 = line(Earth, SightPoint1),
Sight2 = line(Earth, SightPoint2),

orientation(tangent, Sight1, SunTopDown),
orientation(tangent, Sight2, SunTopDown),
incidence(coincident, SightPoint1, SunTopDown),
incidence(coincident, SightPoint2, SunTopDown),
orientation(left,  SightPoint1, L),
orientation(right, SightPoint2, L),

point(PA), point(PB),
incidence(coincident, PA, MoonOrbit),
incidence(coincident, PB, MoonOrbit),
incidence(collinear, PA, Sight1),
incidence(collinear, PB, Sight2),

size(equal, arc(MoonOrbit, PA,PB),
            diameter(SunFromEarth)).
```

Finally, we constrain the relative distance and orientation of the Earth and the Moon, that is, if the position of the Moon in the top-down diagram is changed then the relative position between the Sun and Moon in the from-Earth perspective must also change (and vice versa).

In the top-down diagram let p_C be the intersection point between L and the Moon's orbit, and again let r be the radius of the Moon's orbit. The length of the arc with radius r, centred on the Earth, from p_C to the Moon's centre is the perceived distance between the centre of the Sun and Moon in the from-Earth perspective.

In the top-down diagram the Moon's centre point is either left of, collinear to, or right of an "arrow" pointing from Earth to the centre of the Sun (i.e. L). This relative orientation relation between L and the centre of the Moon is constrained to be the same as the relative orientation of the centre of the Moon and a vertical "arrow" from the centre of the Sun, pointing upwards, in the from-Earth diagram (Fig. 6d).

```
incidence(coincident,PC,MoonOrbit),
incidence(collinear,PC,L),
centre(Moon, MoonCentre),
size(equal, arc(MoonOrbit,PC,MoonCentre),
     line(SunCentreFromEarth, MoonCentreFromEarth)),
orientation(Relation, MoonCentreTopDown, L),
orientation(Relation, MoonCentreFromEarth, L3),
```

Defining eclipse states. We implement all states including *penumbra*, *first contact*, etc. and magnitude relations. For example, the penumbra occurs when the Sun is externally connected (ec) or partially overlapping (po) the Moon in the from-Earth perspective. First-contact occurs during penumbra when the centre of the Moon is to the right of the centre of the Sun. The magnitudes define the perceived relative size of the Moon and Sun in the from-Earth perspective.

```
state(penumbra) :-
  object(type(moon), perspective(from_earth), Moon),
  object(type(sun),  perspective(from_earth), Sun),
  (topology(rcc(ec), Moon, Sun);
   topology(rcc(po), Moon, Sun)).
```

```
state(first_contact) :-
  state(penumbra),
  object(type(earth),perspective(top_down), Earth),
  object(type(moon), perspective(top_down), Moon),
  object(type(sun),  perspective(top_down), Sun),
  centre(Moon, MoonCentre),
  centre(Sun,  SunCentre),
  orientation(right, MoonCentre, line(Earth, SunCentre)).
```

```
magnitude('>1') :-
  object(type(moon), perspective(from_earth), Moon),
  object(type(sun),  perspective(from_earth), Sun),
  size(larger,Moon,Sun).
```

Spatial Q/A. Users can express Prolog queries about eclipse states and qualitative spatial relations in both perspectives seamlessly. For example, what is the topological relation between the Moon and Sun (from Earth's perspective) during the *umbra*?

```
?- state(umbra),
|  object(type(moon), perspective(from_earth), Moon),
|  object(type(sun),  perspective(from_earth), Sun),
|  topology(Relation, Sun, Moon).
Relation = rcc(tpp);
Relation = rcc(ntpp);
Relation = rcc(eq);
false.
```

What magnitude corresponds with the antumbra?

```
?- state(antumbra),
|  magnitude(Magnitude).
Magnitude = '<1';
false.
```

If the Moon is to the *left* of sight line Sight-1, can the Moon and Sun partially overlap in the from-Earth perspective?

```
?- object(type(moon),perspective(top_down), MoonTopDown),
|  object(type(sight_1), perspective(top_down), Sight1),
|  orientation(left, MoonTopDown, Sight1),
|
|  object(type(moon), perspective(from_earth),
|        MoonFromEarth),
|  object(type(sun),  perspective(from_earth),
|        SunFromEarth),
|  topology(rcc(po), MoonFromEarth, SunFromEarth).
false.
```

Intelligent Diagrams. As users manipulate diagrams of either perspective, the objects in both diagrams are automatically updated so that all qualitative spatial relations are maintained. Thus, users can explore different configurations of objects and observe the relationship between the perspectives.[1] Moreover, at any stage the user can query the diagram via a Prolog query as above, or modify the diagram by enforcing spatial constraints, specifying states, etc.

Conclusion and Future Research

In this paper we have sketched a route to a qualitative theory of eclipses, building upon previous work in occlusion and shadow reasoning. The incorporation of viewpoints into reasoning about perception and shadows provides a route to understanding celestial phenomena in a new way. Our model allows the characterisation of our visual experiences of eclipses in terms of occlusion relations, and the prediction of visual experiences given occlusion relations.

We have gone on to implement this qualitative model for eclipses in terms of the Constraint Logic Programming system CLP(QS). This implementation provides *intelligent diagrams*, with which users can interact with the qualitative model by manipulating its objects, and *Spatial Question/Answering* about the domain modeled. These features facilitate a seamless interaction between users and the domain. This could be used as a tool for a hypothesis testing procedure in an educational context (in a similar way as described in (Forbus et al. 2005)). In this paper we have demonstrated the intelligent diagram aspects and the question answering aspects, however the actual use and evaluation of this method in the classroom is a task for future research.

An interesting issue for a future work is to use a qualitative calculus in order to infer, given what is observed from Earth during an eclipse, what can be seen of the Earth-Moon system from any point on the surface of the Sun at the same time. For our purposes, the Sun can be modeled as an infinity of pointwise light sources, each of which can be considered a viewpoint. Each of these viewpoints sees the Moon fully occluding its own shadow cast on Earth, surrounded by a penumbral area which is the union of the shadows cast by all other points on the solar surface. No point in the Sun, thus, sees the umbra, and all points see a portion of the penumbra, which is partly occluded by the Moon itself. These viewpoints belong to different visibility classes, defined by the ROC relations holding on Earth.

[1] We use FreeCAD as a front end for the intelligent diagrams.

In this paper we assumed ROC relations as defined for single observation points. However, the Earth is sufficiently large to support salient parallax effects during the occurrence of an eclipse. Totality is visible in a certain area (the shadow's path of totality), the surrounding areas only experience partial eclipses, and the areas further away see tangential eclipses or no eclipses at all. Similarly, along the path of totality, totality sets in at different times, so that occlusion relations differ at different places. An opportunity thus arises to consider the combination of the information obtained by the multiple viewpoints into a single (qualitative) description of the phenomena. This is also an issue left for future research.

Acknowledgements

Roberto Casati's contribution was funded by grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*. Paulo Santos is partially funded by CNPQ and FAPESP.

References

Bailey, J., and Slater, T. 2004. A review of astronomy education research. *Astronomy Education Review* 2(2):20–45.

Barnett, M., and Moran, J. 2002. Addressing childrens alternative frameworks of the moons phases and eclipses. *International Journal of Science Education*.

Bhatt, M.; Guesgen, H. W.; Wölfl, S.; and Hazarika, S. M. 2011. Qualitative spatial and temporal reasoning: Emerging applications, trends, and directions. *Spatial Cognition & Computation* 11(1):1–14.

Bhatt, M.; Lee, J. H.; and Schultz, C. 2011. CLP(QS): A Declarative Spatial Reasoning Framework. In *COSIT 2011 - Spatial Information Theory*, 210–230. Berlin, Heidelberg: Springer-Verlag.

Bhatt, M.; Schultz, C.; and Freksa, C. 2013. The 'Space' in Spatial Assistance Systems: Conception, Formalisation and Computation. In *Representing space in cognition: Interrelations of behavior, language, and formal models. Series: Explorations in Language and Space*. 978-0-19-967991-1, Oxford University Press.

Bredeweg, B., and Struss, P. 2004. Current topics in qualitative reasoning. *AI Magazine* 24(4):13–16.

Cohn, A. G., and Renz, J. 2008. *Handbook of Knowledge Representation*. Elsevier. chapter Qualitative Spatial Representation and Reasoning, 551–596.

D.Bransford, J.; L.Brown, A.; and R.Cocking, R., eds. 2000. *How People Learn: Brain, Mind, Experience, and School*. NATIONAL ACADEMY PRESS.

Forbus, K. D.; Carney, K.; Sherin, B. L.; and II, L. C. U. 2005. VModel: A visual qualitative modeling environment for middle-school students. *Journal of Geoscience Education* 26(3):63–72.

Galton, A. 1994. Lines of sight. In *Proc. of the Seventh Annual Conference of AI and Cognitive Science*, 103–113.

Guha, P.; Mukerjee, A.; and Venkatesh, K. S. 2011. Ocs-14: You can get occluded in fourteen ways. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, IJCAI'11, 1665–1670. AAAI Press.

Kleer, D. S. W. d., ed. 1990. *Readings in Qualitative Reasoning About Physical Systems*. Morgan Kaufmann.

Libarkin, J. C., and Kurdzie, J. P. 2001. Research methodologies in science education assessing students alternative conceptions. *Journal of Geoscience Education* 49(4):378–383.

Ligozat, G.; Santos, P. E.; and Samghabadi, M. 2015. An occlusion calculus based on an interval algebra. In *Brazilian Conference on Intelligent Systems (BRACIS)*.

Ligozat, G. 2011. *Qualitative Spatial and Temporal Reasoning*. Wiley.

Neugebauer, O. A. 1952. *The Exact Sciences in Antiquity*. Princeton University Press.

Petrov, A. P., and Kuzmin, L. V. 1996. Visual space geometry derived from occlusion axioms. *Journal of Mathematical Imaging and Vision* 6(2-3):291–308.

Randell, D., and Witkowski, M. 2002. Building large composition tables via axiomatic theories. In *Proc. of KR*, 26–35.

Randell, D.; Cui, Z.; and Cohn, A. 1992. A spatial logic based on regions and connection. In *Proc. of KR*, 165–176.

Randell, D.; Witkowski, M.; and Shanahan, M. 2001. From images to bodies: Modeling and exploiting spatial occlusion and motion parallax. In *Proc. of IJCAI*, 57–63.

Santos, P. E.; Martins, M. F.; Fenelon, V.; Cozman, F. G.; and Dee, H. M. 2016. Probabilistic self-localisation on a qualitative map based on occlusions. *Journal of Experimental & Theoretical Artificial Intelligence* 1–19. In press, online version available.

Santos, P.; Dee, H.; and Fenelon, V. 2009. Qualitative robot localisation using information from cast shadows. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, 220–225.

Santos, P. 2007. Reasoning about depth and motion from an observer's viewpoint. *Spatial Cognition and Computation* 7(2):133–178.

Schultz, C. P. L., and Bhatt, M. 2012. Towards a declarative spatial reasoning system. In *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012*, 925–926.

Sorensen, R. 1999. Seeing intersecting eclipses. *The Journal of Philosophy* 96(1):25–49.

Souchanski, M., and Santos, P. 2008. Reasoning about dynamic depth profiles. In *Proc. of the 18th European Conference on Artificial Intelligence (ECAI)*, 30–34. Amsterdam, The Netherlands, The Netherlands: IOS Press.

Tassoni, S.; Fogliaroni, P.; Bhatt, M.; and De Felice, G. 2011. Toward a qualitative model of 3D visibility. In *25th International Workshop on Qualitative Reasoning*.

Winroth, H. 1999. *Dynamic projective geometry*. Tekniska högsk.

Structuring the Domain Knowledge for Model-based Decision Support to Water Management in a Peri-urban Region in India

P. Struss[1], F. Steinbruch[2], C. Woiwode[2]
[1]Tech. Univ. of Munich, [2]Indo-German Center for Sustainability, IIT Madras, Chennai
struss@in.tum.de, franziska.steinbruch@gmail.com, cwoiwode@iitm.ac.in

Abstract

The paper presents first results in designing and realizing a computer tool that is meant to support decision making in the context of peri-urban development with a focus on issues of water management. This is associated with research on the development in the Sriperumbudur region near Chennai that is conducted by the Indo-German Center for Sustainability at IIT Madras. We describe and illustrate the chosen approach, which is based on establishing a library of models of the relevant physical, technical, social, economic, and governance processes relevant to the problem domain and then using it for both predictive and explanatory tasks.

Introduction

Planning and decision making in urban and peri-urban development are facing significant challenges of different origin:

- They have to reflect complex interactions among a variety of factors, including natural aspects such as topographic, hydrological, climatic, and ecological ones, as well as social, economic, technological, and administrative impacts: migration, housing, water and energy supply, waste management, traffic, etc.
- Turning plans and decisions into reality often involves a large number of stakeholders with particular interests such as law-makers, administrative units, expert committees, lobbyists, migrant labor, real estate sector, and others.
- While several of the aspects are general and shared by different regions, there will always be significant differences, e.g. regarding specific natural, cultural, administrative, political, and economic conditions, which prohibits the development of universally applicable rules and patterns, but requires region-specific analyses and decisions.

Computer aid to conducting the necessary research and for decision making is desirable, but as of now limited to acquisition and storage of data (aerial and satellite data, geographical information systems, GIS) and to modeling and simulation systems (e.g. hydrological models) that have to be handcrafted and adapted to local conditions. Producing decision support systems for each region and problem would be costly if feasible at all.

In order to reflect these challenges, we propose a knowledge-based-systems approach, more precisely model-based decision support systems:

- i.e. systems capturing the knowledge about all known (or hypothesized) potentially relevant interdependencies and the decision processes and governance structures, stakeholders, their roles and guiding principles in models, i.e. a computer-based representation that can be used to support, for example, interpretations of observations, prediction of future evolution, and planning of interventions.
- This knowledge base has to be structured as a collection of elementary, independent model elements that can be combined to form models tailored to different individual scenarios for different regions, contextual conditions, or hypotheses. This also allows for an easy modification of individual model fragments and an extension of the entire model library.
- Based on this, computer tools can be developed that support both researchers and decision makers in their work. This could happen to evaluate and suggest certain policies of government and administrational institutions, but also to private companies, which, for instance, are obliged to invest in improvements if they are located close to water bodies.
- Such tools should be utilizable without requiring that users have detailed knowledge about the content, the kind of representation, and technical aspects of the models and the algorithms that exploit them.

This paper presents the first results of an analysis laying the foundations of a planned project along these lines which focuses on water management in the Sriperumbudur region near Chennai, India. It is linked to ongoing studies at the Indo-German Center for Sustainability (IGCS) at the Indian Institute of Technology Madras (IITM), Chennai, on problems of water, land use, energy and waste in the area.

The following section discusses peri-urban development in the Sriperumbudur region. Section 3 outlines the foundations of model-based decision support. Then, an approach to systematize and structure the model library is presented, and we introduce some sample models for the library.

Peri-urban Development in Sriperumbudur

The Area

Until quite recently, much of the research on climate change adaptation has focused either on cities and their adaptive capacity (service infrastructure, disaster preparedness, vulnerable groups) or on rural areas identified as specifically vulnerable to droughts or floods. Peri-urban areas bring up new issues that move across disciplinary but also governance and institutional boundaries, and as such define a complex and highly dynamic environment in their own right. The peri-urban, which can be treated as a transitional zone between the urban and rural of rapid changes, is now being recognized as an entirely separate socio-spatial configuration with distinct challenges of sustainability [Adell 99], [Simon 08].

In India, peri-urban areas are typically the sites of fastest growth outside the area of metropolitan corporations and the so-called Tier-II cities [Vishwanath 13]. In the face of climate change, it has become imperative to assess and steer the development process in these regions in order to ensure their sustainability and resilience. Though still relatively low, the level of urbanization in India has experienced a very rapid ever-increasing urban growth over the recent years.

As one of the six most urbanized states (with an urban population much higher than the national average of 31.1%), the state of Tamil Nadu is undergoing rapid urbanization. The City of Chennai (the former Madras) of The Greater Chennai Corporation metropolitan region is the most important location of the Indian and foreign automobile industry. Since Ford Motor Co., Hyundai Motor Co., Nissan Motor Co., Renault SA, Daimler AG and BMW AG are based in the Chennai region, some call it the 'Detroit of India' (The Wall Street Journal 2010).

Because of its geographically confining location in the northernmost part of Tamil Nadu, close to the border of Tamil Nadu's neighboring state Andhra Pradesh and the Bay of Bengal, the traditional industries in Chennai evolved mostly around the port and the vicinity of formerly abundant water resources in the northern outskirts of the old city, where industrialization reached stagnation due to unavailability of land. State driven modern industrialization started relatively late, after the occurrence of land reclassifications of the nature of waste land to industrial parks (including SEZs, i.e. special economic zones) or farm land to residential areas, which paved the way for the growth of industrial estates outside the city and triggered the evolution of a suburban/peri-urban region, which today may be seen as a positive situation.

The selected peri-urban study area of the IGCS project (see [Adelina et al. 15]) has been initially defined by administrative boundaries, constituting one of the ten taluks (the administrative divisions below the district) in Kancheepuram district of Tamil Nadu West of Chennai, which covers 371.94 sq.kms (of 4432 sq.kms of the district) and contains a population of 316,918 persons as per 2011 census (8% of the district). Within this district, Sriperumbudur Town Panchayat (i.e. the smallest unit of local administration) has been transformed from a village into an industrial hub in the past 20 years.

Since the peri-urban is characterized by the absence of clear boundaries, the region as a spatial unit is flexibly applied to the context of analysis and action. For instance, in relation to the water basin or the institutional setting and planning areas there are multiple overlaps, interfaces and levels which define peri-urban Sriperumbudur in various ways beyond the taluk boundaries.

In order to cope with the rapid urban growth, the boundaries of the Greater Chennai Corporation were gradually expanded in 2011. The process can be considered in two ways; as an attempt to formally integrate peri-urban areas with the metropolitan region or as swallowing of formerly independent areas to secure or satisfy natural resources needs for the city.

Impact on Water Management

Urbanization and industrialization change land use and use of water resources in a structural, qualitative, quantitative, and also in a political way. The impact on water and land resources has a number of different consequences:

- An increase of the consumption of water for domestic or industrial purposes due to piped network requirements and different consumer behavors (Minimum basic water need in Indian cities is 120 l per capita day and in rural areas 55 l per capita day) combined with an increase in waste water generation.
- A reduction of the available water resource quantities (lowering the ground water table, reducing reservoir capacities, eliminating traditional means of water storage and water harvesting).
- For coastal areas, salt water intrusion due to excessive groundwater abstraction.
- Generation of solid waste which accumulates in lower laying areas, usually water-logged areas or traditional water storages.
- A change in the kind of waste water (volumes and physico-chemical composition).

• A deterioration of the quality of resources for drinking water (ground water, surface water) due to absence or only partial treatment of waste water and absence of sanitary landfills.
• Increased amounts of storm water due to increased direct run-off of rain water due to lack of detentio/ retention areas and sealed surfaces.
• Reduced rainfall over urbanized areas due to heat island effect and changed albedo. Increased temperature in urbanized areas due to reduced moisture availability / reduced evaporation because soils are sealed and open water bodies are covered in solid waste or dried up.
• destruction of systems of interconnected water reservoirs and, thus, the ability to avoid or mitigate flooding, to store water for irrigation and drinking, or to allow the percolation of water to recharge aquifers.

In peri-urban development in the Sriperumbudur area, specific aspects add to this:
• Loss of traditional powers and knowledge to manage established rainwater collection and cascading water supply and storage systems [John 15].
• change in purpose which leads to the interruption of water infrastructure systems made for agricultural irrigation and used now for urban water supply with tankers. Continuous subsidies for the abstraction of groundwater by farmers (unlimited free-of-charge power consumption for pumping) creating a flourishing water market (former farmers sell water to newcomers in the peri-urban areas) and to city water supply.
• Upgrading or value-addition of former wetlands and man-made lakes to some other public infrastructure such as highways, sports stadiums, railway stations thus destroying traditional localized means for water supply from ponds and rain water harvesting.

Climate change, with its impact on the temporal distribution and amount of precipitation is projected to increase the number of extreme rainfall events per year and at the same time reduce the number of rainy days per year in Tamil Nadu [Chaturvedi et al. 10].

Model-based Decision Support

The Foundation

In our approach, we follow Qualitative Process Theory ([Forbus 84]) and our logical and computational reconstruction of it ([Heller 2001], [Heller-Struss 02], [Struss 11]). A process states that certain effects will be established whenever its preconditions are satisfied, i.e. an implication:

$$\text{StructuralConditions} \wedge \text{QuantityConditions} \Rightarrow \text{StructuralEffects} \wedge \text{QuantityEffects},$$

where StructuralConditions and StructuralEffects assign existence to objects and structural relations, and QuantityConditions and QuantityEffects contain assignments of values (or ranges) to quantities. In addition, QuantityEffects include also **influences**, which capture the contribution to the process to the dynamics of the systems (which may rival with counteracting influences of other processes). We assume that there exists a **process library** representing the core of the domain knowledge.

Constructing situations involves starting from the given observations (which may be considered as facts or default assumptions) and iteratively completing them in two directions:
• Forward completion: adding all implications of an intermediate result of the construction process and the process library, i.e. instantiating processes whose preconditions are satisfied and their effects. This establishes the causal impact of the intermediate result, but it does not address the main goal, namely finding a reason for what has been observed.
• Consistency check: if the resulting situation is consistent, it is possible answer to situation assessment. Otherwise, it is incomplete causally upstream and requires
• Backward completion: this looks for process candidates whose effects yield changes in quantities and/or existence of objects and relations that are unexplained. If there is no such process, the search is cut off here. Since there can be several candidates, the search may branch.

The somewhat surprising fact that pure consistency-based reasoning yields an abductive result is due to two axioms:
• **Influence resolution**: if its result yields an inconsistency, it can only be resolved by an additional process; a special case is a change in a quantity that is not influenced in the current model .
• **Existence default**: objects and relations do not exist unless they are given as observations, effects of active processes, or as introducibles, which are discussed below.

These closures are supported by closed-world assumptions that are associated with existence variables and quantities. In the Generalized Diagnosis Engine G+DE ([Heller 01], [Heller-Struss 02]) consistency-based diagnosis ([de Kleer-Williams 87], [Struss 08]) is performed to deliver (minimal) sets of assumptions that create an inconsistency. Resolving such an inconsistency may involve simply dropping some inappropriate default assumptions, while revising a closed-world assumption means searching for additional processes that provide an effect on the respective variable and, hence, performing backward completion in an informed and focused way.

The concept of **introducibles** is crucial for terminating the search: otherwise, repeated backward completion would usually ultimately result in an inconsistency, because some object remains unexplained. This reflects that each model library has a limited horizon of what can be explained. For instance, in the context of our project, we may require the system to discover the origin of a high arsenic concentration in drinking water (e.g. minerals containing arsenic in certain rock layers in the ground that are touched by wells), but not an explanation in terms of geological processes that created those minerals). All objects that do not occur in a StructuralEffect of any process in the library have to be marked as introducible. However, for a particular task, the model boundary, i.e. the scope of the explanation, may be tighter.

As a result, situation assessment, starting from initial assertions, includes causally downstream processes and their impact, but also causally upstream processes and their impact. Since backward search will commonly detect several potential explanations, it will usually deliver alternative results, and the user may have to pick the most plausible one.

Characterized in a more formal way, the **result of situation assessment** should be a **minimal** situation

• containing **all facts**
• otherwise only **introducibles** and
• **effects** of occurring processes
• being **closed** w.r.t. **effects**,
• in which **a maximal set of assumptions** holds.

Challenges of Water Management Addressed

The various issues mentioned in section 2 interact closely. These complex interdependencies make both the task of assessing the impact of previous or planned steps in peri-urban development on water and the development of adequate policies and the planning of specific interventions and constructions difficult.

In addition, any establishment of regulations and the realization of any constructive activities have to reflect the (formal and informal) governance structures, i.e. the interest, role, and competence of various stakeholders (central and local administrative institutions, communities, companies etc.).

These governance structures are complex and specific to a particular country, state and area (e.g. there is an overlap of responsibilities of the city of Chennai and the Sriperumbudur taluk, because of the delineated jurisdiction of the Chennai Metropolitan Development Authority CMDA).

Analysis and proposal of useful interventions combine qualitative and quantitative aspects. While, on the one hand, constructive actions need a numerical specification (e.g. regarding the diameter of pipes or throughput through a treatment plant), on the other hand, scenarios to be analyzed will often be characterized in qualitative or symbolic terms (“monsoon rainfall”, “Toxic waste water”), and information about real situations tends to be approximate and

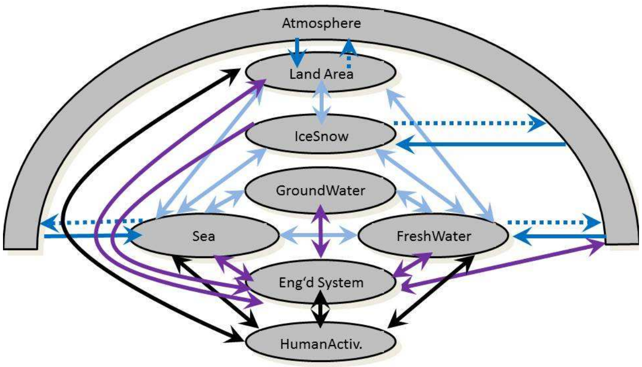


Figure 1 High-level Domain Concepts

stating tendencies (amount of required water supply”, “reduction of agricultural land use”).

Structuring the Model Library

In our application domain, the model library has to comprise model fragments that represent a fairly large number of interdependencies with a high diversity:

• Basic **physical processes**, such as downhill flow of water, salinization of ground water, and evaporation
• **Chemical and biological processes** like reactions due to pollution, its impact on living organisms, or algal bloom
• Various kinds of **human activities**: transporting and polluting water, constructing dams and dwellings, etc.
• **Economic developments** (changes in cost of land and accommodation, financial investments)
• **Political and legal regulations** and processes such as: administrative procedure, lobbying activities, formal and informal decision processes, water allocation priorities, water allocation tools, bulk water provisions, and water use charges (use and pollution) .
• The impacts of different kinds of **land use** (agriculture/horticulture, grazing, mining, …)

and many more. Filling such a diverse library is a genuine challenge and will not succeed without initial work on categorizing and decomposing the relevant body of knowledge and, based on this, systematizing and structuring the model library in an appropriate way.

In this section, we discuss the results of a first analysis. The high-level structure is illustrated in Fig. 1. The top-level distinction is between the interdependencies that are not controlled by humans but by natural “laws”, and human activities, artifacts and interactions among themselves and the natural objects (the bottom part of Fig. 1). We discuss these two “hemispheres” in the following two subsections.

The Physical Hemisphere

A central class of objects comprises land areas, more precisely: their surfaces that are usually not covered by water and possibly exploited by human activities, e.g. in agriculture, industrial production, or dwelling. Such land areas will be categorized by different land uses. They are further characterized by attributes, in our context (with a focus on water-related issues) by topographic information, such as altitude and slope (influencing flow of surface water), type of vegetation (affecting transpiration and water run-off), geological features (determining run-off, infiltration, aquifer formation).

Five classes are representing areas of water flow and storage:

• **Fresh water** bodies on the surface (lakes, streams, wetlands)
• **Groundwater**,
• **Oceans**, storing and moving salt water
• **Ice/snow** and glaciers, and
• **Atmosphere**

The latter will be the entry point for impacts of climate change, which becomes an essential factor in analyzing and planning sustainable development, e.g. through the amount and distribution of moisture, cloud coverage and permeability for sunlight, or heat exchange. From the perspective of water management, it directly interacts (in both directions) with land areas and the water bodies except for ground water, thus defining different basic processes in the model library: The solid dark blue arrows in Fig. 1 represent precipitation (snowing, frost, rain, dew), while the dotted dark blue arrows indicate transportation of water back to the atmosphere by evaporation (from water bodies and temporary surface water in land areas), transpiration (by vegetation on land areas), and sublimation (of solid water).

Already at this level of abstraction, it is evident that for different areas and/or different climatic conditions, the relevance of different elements of the model changes or vanishes. Decision support on water management in Sriperumbudur can safely ignore snow and glaciers, as well as direct influence by ocean water, while in Ladakh, snow and glaciers are essential, and in Chennai, salt water from the ocean infiltrates ground water. Other basic processes will occur in each model, which emphasizes the idea of configuring models from reusable model elements.

Like land areas, water bodies (except for groundwater) may need an association with different economic uses, such as fishing areas and fish farms or recreation (water sports, skiing) and, hence, linked to the concepts in the next subsection.

As a challenge to modeling, the spatial extension of the instances of the classes is not static: land, oceans, and surface water bodies may be seasonally covered by ice and snow, glaciers shrink, and land areas may be flooded.

Land areas and water bodies potentially exchange water among themselves, as indicated by the light blue arrows in Fig. 1. Besides the exchange between fresh water and the ocean (usually in one direction, unless affected by tides or storms), water from fresh water bodies or the ocean may inundate land areas and, in turn, runs off from land areas and melting snow and ice to the water bodies. Through infiltration, ground water is fed by fresh water, but also by oceans, if the water table is low, and discharges back into them. A high water table or an artesian aquifer may also generate flooding of land, unless the model opts for treating the result as a (temporary) fresh water body.

Actually, the arrows in the diagram discussed, so far, more or less correspond to the basic processes to be represented without requiring much refinement, and those ones that represent water flow are simply governed by gravity, i.e. directed towards the area with a lower elevation. This is what is mainly covered by numerical hydrological models.

From the perspective of management of drinking and waste water, not only the amount of moved or stored water is relevant, but also what is transported and captured in the water, such as salt, pathogens, toxic substances, and organic material. The distribution and transportation of such elements and possibly their transformation has to be captured by model fragments, as well.

What has been discussed, so far, captures basic natural processes of hydrological, biological etc. systems, chemical reactions etc. (although they may be triggered or influenced by humans (for instance, through water pollution). While this is a necessary ingredient to model-based decision support in the domain, water management has to consider human activities and, even stronger, is targeted at human interventions that, in interaction with the natural processes, achieve the fulfillment of certain goals. Therefore, we need to include a realm of human-related, social, administrative processes in the model library, which have to have a clear interface with respect to the natural processes.

The Human Hemisphere

Human interference with the processes described above often happens through constructed systems. They are very diverse and comprise water reservoirs for drinking water, wells, dams and water gates to prevent or control flooding, pumping and irrigation systems for agriculture, water treatment and desalinization plants, cisterns, urban retention/detention areas (on roof tops), . The level of scale and granularity of the model may vary: transfer of water may be related to individual water reservoirs or between different catchments. Even international trade with bottled water may be included as a mechanism for drinking water supply. We summarize this in Fig. 1 as a class of **engineered systems**. They **interact** (shown by purple arrows in Fig. 1) with

• **Ground water** (wells for extraction of ground water and injection wells for recharging it)

• **Freshwater** (canals and water gates for controlling flow and storage, extraction for drinking water treatment and disposal of water from waste water treatment or cooling systems of power plants, etc.)

• **Sea water** (dikes, desalinization plants, industrial run-off)

• **Ice and snow** (perhaps mainly one-directional: for instance, ice blocking water ways or the operation of facilities)

• **Land areas** (irrigation and draining systems, water distribution systems for industrial and domestic use)

• The **atmosphere** (systems for rain water harvesting, evaporation from stored water)

Finally, there is the most diverse field: **human activities**. They are linked to the above entities and processes (black arrows in Fig. 1)

• With **water bodies indirectly** through construction, operation, and use of the engineered systems discussed above (or suffering from their malfunction); ground water is exclusively accessible through them and

• **directly** by dumping waste water into sea water and fresh water, using water bodies for transportation, recreation, and economic exploitation (fishing and fish farms, salt production)

• With **land areas** by involvement in different land uses and their modifications (farming, grazing, acquiring or selling land, constructing “unauthorized colonies” (slums, gated communities that is))

• With the **atmosphere** by polluting it (and in turn suffering from it), modifying local climate conditions, e.g. increased temperatures in cities)

As we stated before, including **governance structures** in the model is essential for effective decision support, because the mere proposal of attempts to impose certain changes in the physical system is unlikely to be brought to a realization.

A more detailed and extended view on the lower part is depicted in Fig. 2. We introduce activities for

• **Building** the **engineered systems** (dams, pipelines, treatment plants, …) and for

• **Operating** them (distribution of drinking water, opening water gates, etc.).

Indeed, many decisions to be taken related to water management will aim at such interventions. In the modeling formalism, the first kind of process has the effect of bringing new objects into existence, while the second one changes states of such objects. In our application domain, the latter may trigger natural processes: for instance, opening a water gate, say, separating a lake from a riverbed, results in a modification of the relative order of water levels in the lake, the water gate, and the river bed and will trigger a flow from one water body to the other one via the water gate. In a treatment plant, opening a certain valve may result in a flow of an oxidant into a water tank, which is a precondition for process of oxidizing dissolved iron contained in the water.

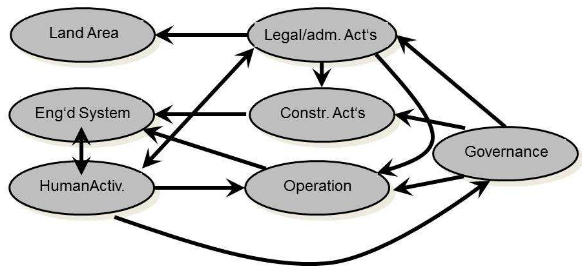


Figure 2 Human Activities and Governance Procedures

As a third type of activities in this context, we consider

• **Administration**, legislation, and financial acts, such as imposing limits on water consumption, controlling or subsidizing certain kinds of land use, changing ownership of land areas, providing compensation for damage or income loss.

Except for direct effects on land ownership, such activities will usually influence the physical world only via other human activities, social, economic, etc. ones or by setting the context for building or operation processes. For instance, financial incentives for growing certain crops influence decisions of farmers, which modifies consumption of irrigation water, hence the amount of water stored in reservoirs, etc. An impact on human behavior could also be exerted by information distribution, educational campaigns etc., and one might include them in the above class or create an additional one.

Finally, we need to explicitly describe

• **governance** structures and rules.

While the issue will occur in one way or another in other parts of the world, in the Sriperumbudur case, as in India in general, there are extremely complex structures. There are easily half a dozen institutions, legal bodies, advisory boards etc. at different levels (community, district, state, federal) involved in a single decision. Collecting and representing information about the various stakeholders, their role and interests, is an important part of the ongoing project at IGCS. A crucial issue is that many stakeholders may not operate in a coordinated manner; and, in addition, decision are also made ‘off-the-rule-book’. i.e. by informal power plays.

In the decision support system we are aiming at, these governance procedures will need to be included as preconditions for certain interventions, i.e. it should not only recommend building a canal that connects two reservoirs, but

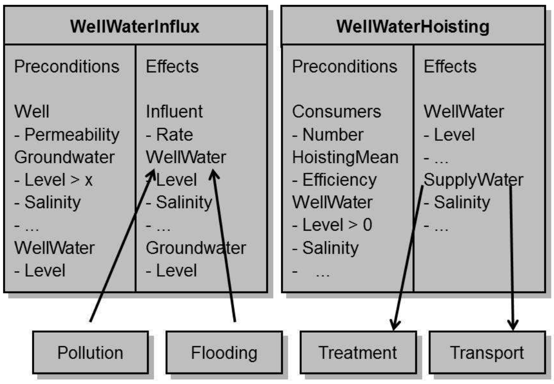


Figure 3 Engineered System "Well" and Human Activity

also which bodies will have to be involved in which role in order to plan, decide upon, finance, and construct it.

The governance models are **not** expected to actually **predict** the outcome of the respective planning or decision steps based on a detailed representation of the criteria, legal restrictions etc. The goal is rather enabling the system to **infer the preconditions** for the realization of certain proposed interventions. From the technical perspective, this means including their prerequisites in the backward completion of the model, thus determining, for instance, that the approval of a certain state government board is required for building a canal (but not predicting whether it will actually be approved). Furthermore, dependent on the location of the canal, certain local administrative bodies have to be involved in the planning, and the resulting plan needs financial contributions from the state government, the district, and, perhaps, companies in the area.

In order to achieve this, we need to represent the various institutions and legal entities that participate in certain procedures, the role they fulfill (officially or unofficially), preconditions for their activity etc. Such roles could be categorized as "planning", "deciding", "approval", "advisory" etc.

Sample Processes

We illustrate what has been introduced above using the example of a well. We reduce the process descriptions to the essential aspects, omitting, in particular, the detailed constraints and influences among the various quantities. The process WellWaterInflux (Fig. 3) is an instance of the interaction between an engineered system (the well) and a natural water body (groundwater) and follows straightforward physical principles. WellWaterHoisting (Fig. 3) captures the human activity related to the engineered system, namely extracting water from the well (which may then be subject to a treatment or transportation process). The figure also indicates how some other processes, such as pollution or flooding may affect the well water.

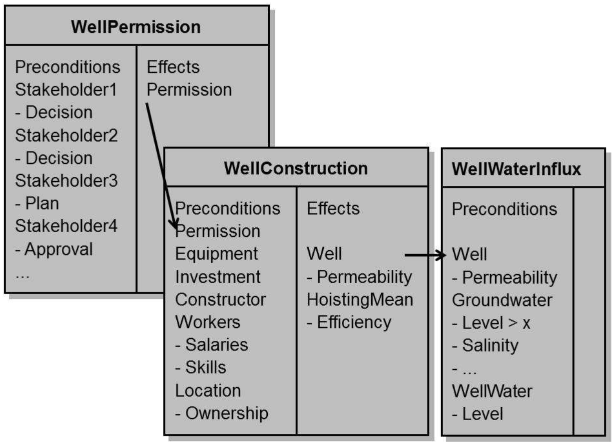


Figure 4 Constructing a Well requires a Permission

Fig. 4 shows that the well comes into existence through a construction process (see Figure 2). The precondition "permission" is subject to a respective process as part of the governance model that lists the stakeholders and their roles (Fig. 4). This can be further related to more detailed activities if needed and possible.

Discussion

We developed foundations for applying a generic approach to model-based decision support to the domain of peri-urban development with a focus on water management. Since the enterprise of building an appropriate model library has to be done in a distributed way by several (groups of) experts, it is essential to develop a systematic approach and a common ontology for integrating the various sections of the model library. Based on our current analysis, we can identify certain challenges and draw some preliminary conclusions.

• **Uniform Representation and Reasoning**: The analysis above suggests that we can actually use a single modeling formalism that allows integration of processes of quite diverse scope: natural physical phenomena, building and running man-made plants that interact with the natural environment, legal and administrative procedures that are relevant to proper planning of interventions etc.

• **Qualitative vs. quantitative** representation and reasoning: most of the interdependencies that have to be considered are of qualitative nature and lack precise numerical parametrization, which makes qualitative modeling and reasoning appropriate. However, some interventions need to be specified by numbers (e.g. determining the size and depth of a well, the required capacity of a water reservoir, the amount of water to be released through a water gate). This will require links to a numerical level, for instance, more precise hydrological models, calculations of water demand and supply etc. Also certain thresholds may have to be represented, esp. when effects are accumulated: the impact of a single well on the ground water table may be negligible, but too many of them located in a region may cause a significant reduction.

• Qualitative **spatial representation**: we expect that representing various land use areas and natural water bodies as compartments with certain topological relations (rather than detailed and quantified geometric features) in 2D will mainly suffice. Obvious exceptions are groundwater and ice or snow covering ground. This representation has to be linked to an existing geographical information system (GIS), which captures detailed information about the Sriperumbudur district.

• Multi-level **temporal representation:** A problem to be solved is the co-existence of processes that have some intrinsic natural temporal units at quite different levels. This may include daily peak periods of water demand, weekday-weekend patterns, monthly variations in requirements on agricultural irrigation, seasonal weather patterns over the year, long-term trends in changes in land use, climate etc.

• **Non-determinism**: many processes in this domain, especially related to human activities, but also with regard to biological and ecological systems and weather, cannot be described comprehensively as a causal interdependency, because there will be a large set of influencing factors that will be impossible to be included in the model. This applies, in particular, to the governance procedures that we want to include. Whether a certain decision will be taken or not, may depend on personal preferences, the date of the next elections, corruption, power relationships, etc. As discussed with respect to governance processes, this is not a fatal problem in the abductive direction, i.e. when generating possible explanations or establishing prerequisites. But the predictive power of the model suffers. Rather than following the common suggestion of using probabilities (which are not available, at least not comprehensively, and which make everything "somewhat likely"), a solution within the proposed system is the introduction of "unknown preconditions", whose fulfillment cannot be predicted by the model, but which can be subject to assumptions and explored in their consequences.

• Representing **intentions** and goals of stakeholders and human activities is a related issue. In reality, they have a crucial impact on triggering certain processes; but it is hard to impossible to derive them as a result of other processes.

The next step in the project will be consolidating and refining the conceptual structure and ontology by considering sample scenarios and building sections of the knowledge-base. We will need to focus on a feasible scope within the water domain. But we need to be open to later including links into the closely interacting aspect of energy, health, and waste, which are also studied at IGCS.

References

[Adelina et al. 15] C. Adelina1, C. Hill, R. John, T. Kuttler, S. C. Rajan, A. Roul, C. Woiwode: *Peri-urban Dynamics and Sustainability in Chennai: The Case of Sriperumbudur*, N-AERUS XVI, Dortmund, 2015

[Adell 99] Adell, G. *Theories and Models of the Peri-Urban Interface: A Changing Conceptual Landscape*. Working Paper, London: Development Planning Unit, University College London.

[Chaturvedi et al. 10] , Chaturvedi, R.K., Gopalakrishnan, R., Jayaraman, M., Bala, G., Joshi, N.V., Sukumar, R., and Ravindranath, N.H,, *Impact of climate change on Indian forests: a dynamic vegetation modeling approach*. Mitig Adapt Strateg Glob Change, 2010

[de Kleer-Williams 87] J. de Kleer and B. Williams: *Diagnosing Multiple Faults*. Artificial Intelligence, 32(1), 1987

[Forbus 84] K. Forbus: Qualitative Process Theory. Artificial Intelligence 24(1-3), 1984.

[Heller 2001] U. Heller: *Process-oriented Consistency-based Diagnosis-Theory, Implementation and Applications*. Dissertation, Fakultät für Informatik, Technische Universität München, 2001

[Heller-Struss 02] U.Heller, P.Struss: *Consistency-Based Problem Solving for Environmental Decision Support*. In: Computer-Aided Civil and Infrastructure Engineering 17 (2002) 79-92

[John 15] John, R. *Governance of peri-urban water resources: A case study of the discursive practises around sustainable water management in greater Chennai region, India*. Master Thesis, University of Freiburg, 2015

[Simon 08] Simon, D. *Urban environments: issues on the peri-urban fringe*. Annual Review of Environment and Resources, 33, 167-185.

[Struss 08] P. Struss, *Model-based Problem Solving* In: van F. Harmelen, V. Lifschitz, and B. Porter (eds.), Handbook of Knowledge Representation, Elsevier, 2008, 395-465

[Struss 11] Struss, P. : *A Conceptualization and General Architecture of Intelligent Decision Support Systems*. In: Chan, F., Marinova, D. and Anderssen, R.S. (eds) MODSIM2011, 19th International Congress on Modelling and Simulation. pp. 2282-2288

[Vishwanath 13] Vishwanath, T., Lall, S. V., Dowall, D., Lozano-Gracia, N., Sharma, S., & Wang, H. G. *Urbanization beyond Municipal Boundaries: Nurturing Metropolitan Economies and Connecting Peri-Urban Areas in India* (No. 75734) World Bank: Washington, DC. doi:10.1596/978-0-8213-9840

On the use of qualitative deviation models for diagnosis

Franz Wotawa
Graz University of Technology, Institute for Software Technology,
Inffeldgasse 16b/2, 8010 Graz, Austria
e-mail: wotawa@ist.tugraz.at

Abstract

Detecting and even locating faults in systems is an important but also very much resource consuming task, which is especially true for finding and fixing bugs in programs. In literature someone finds different approaches for supporting the fault localization task for programs including statistical methods like spectrum-based fault localization, methods based on control and data dependences like slicing, and even model-based diagnosis relying on a logical or constraint representation of a program for computing diagnosis candidates. One issue that hampers the use of model-based diagnosis for debugging is its computational requirements especially when relying on a more or less one-to-one representation of the underlying source code. In order to decrease computational requirements abstract models have to be used. In this paper, we discuss the use of deviation models and provide a framework for comparing such models making use of an abstraction function. First experimental results indicate that some abstract models behave similar to concrete models for diagnosis but come with a much lower computational footprint enabling their use in practice even for larger programs.

Introduction

When a program exhibits an unexpected behavior, the identification of its corresponding root cause can be a very laborious and time consuming task. This is due to several reasons including: (1) The interactions and data communicated with the program leading to the unexpected also contains a lot of information that is not needed for bug localization. (2) The chain of computations from a root cause to its effect, i.e., the failure, might be very long. (3) And it might also be very difficult to state an expected value. The latter occurs for example in cases of complicated computations where we know that the value should be within a range but nothing more.

Let us illustrate this third case using an example from the Spreadsheet domain taken from (Hofer et al. 2015). The spreadsheet given in Figure 1 computes the cardiac index of a person using the diastolic and systolic volume, the heart rate, and the body surface area as inputs. For illustrative purposes we added the cell's formulae directly beside the spreadsheet where we introduced a fault in cell C6, which should be B2-B3. Because of this bug the resulting cardiac index is 72 instead of 2,160. Someone experienced in estimating the cardiac index may easily detect this far too low value but may not be able to specify the real expected output value. Hence, means for specifying deviations from expected values in a qualitative way would be very valuable for automated debugging.

	A	B	C	D
1	Cardiogenic Shock Estimator			
2	End Diastolic Volume	120 mL		
3	End Systolic Volume	60 mL		
4	Heart Rate	72 bpm		
5	Body Surface Area	2 m2	Formulas	
6	Stroke Volume	2 mL	=B2/B3	Fault: "/" instead of "-"
7	Cardiac Output	144 mL/min	=B6*B4	
8	Cardiac Index	72 mL/min/m2	=B7/B5	
9				
10				

Figure 1: Spreadsheet "Cardiogenic Shock Estimator"

In this paper, we follow this idea of using qualitative representations for fault localization instead of real values. Someone should also bear in mind that using values from domains like integer or even reals in models for diagnosis might not be feasible. For example, (Hofer et al. 2015) reported that computing single faults took 25.1 seconds even for smaller spreadsheets having up to 70 non-empty cells. Hence, for larger spreadsheets representations used for diagnosis may hardly use quantitative models. Instead qualitative models that are able to handle deviations, i.e., differences between the expected and the observed value should be used providing that such models come with a smaller computational footprint.

In the following, we define diagnosis based on constraint solving and introduce different models including a value-based variant considering integer values, a dependency model capturing information about the correctness (or incorrectness) of certain values, and a model where we are able to state whether a value is smaller, equivalent, or larger than expected. The latter model we refer as comparison model. In addition to these models we discuss a framework where we are able to compare models with respect to diagnosis accuracy, which we define as the ability of a model for reducing the diagnosis search space. Moreover, we introduce a definition of abstraction that allows for compar-

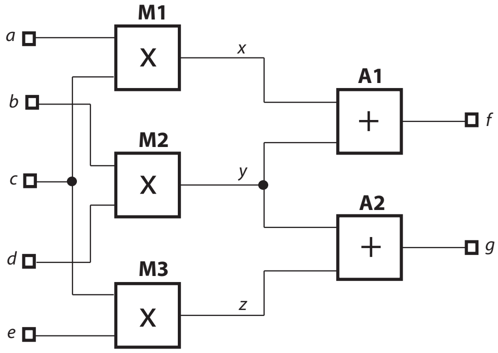


Figure 2: The d74 circuit

ing models. Afterwards, we present the first experimental results when using the different models for diagnosis. The results indicate that the comparison model, which we later plan to use for debugging spreadsheets, has a good running time performance and good diagnosis capabilities.

Basic definitions

In the following we introduce the basic definitions where we rely on the classical definitions of model-based diagnosis (Reiter 1987; de Kleer and Williams 1987) but adapt them to fit to the underlying constraint-based representation of models. For illustration purposes we make use of the famous d74 circuit example depicted in Figure 2.

We first start defining constraints systems and their corresponding constraint satisfaction problem. We define a constraint system as a tuple $(VARS, DOM, CONS)$ where $VARS$ is a finite set of variables, DOM is a function mapping each variable to its domain comprising at least one element, and $CONS$ a finite set of constraints. Without restricting generality we define a constraint c as a pair $((v_1, \ldots, v_k), tl)$ where $(v_1, \ldots, v_k)$ is a tuple of variables from $VARS$, and tl a set of tuples $(x_1, \ldots, x_k)$ of values where for each $i \in \{1, \ldots, k\}$: $x_i \in DOM(v_i)$. The set of tuples tl represents allowed variable value combination. For simplicity, we assume a function $scope(c)$ for a constraint c returning the tuple $(v_1, \ldots, v_k)$, and a similar function $tl(c)$ returning the set of tuples tl of c.

For example, the constraint representation M_{VB} of the d74 circuit to be used for diagnosis purposes has the following variables:

$$VARS = \left\{ \begin{array}{c} a, b, c, d, e, f, g, x, y, z, \\ ab\_m1, ab\_m2, ab\_m3, ab\_a1, ab\_a2 \end{array} \right\}$$

In this set the variables ab_X represent the fault status of a component X, i.e., ab_X is true if component X is said to be abnormal and false, otherwise. The domain for the variables representing connection are integers: $\forall w \in \{a, b, c, d, e, f, g, x, y, z\} : DOM(w) = \mathbb{W}$ and for the fault status we use Boolean values, i.e.: $\forall w \in \{ab\_m1, ab\_m2, ab\_m3, ab\_a1, ab\_a2\} : DOM(w) = \{T, F\}$.

For each component, we have to introduce a constraint. For the multiplication components M1, M2, M3, we have the following constraints:

$((ab\_M_{1|2|3}, a, c, x), \{(F, u, v, u \cdot b) | u, v, w \in \mathbb{W}\} \cup \{(T, u, v, w) | u, v, w \in \mathbb{W}\})$

For the adders A1 and A2 we have similar constraints:

$((ab\_A_{1|2}, a, c, x), \{(F, u, v, u + b) | u, v, w \in \mathbb{W}\} \cup \{(T, u, v, w) | u, v, w \in \mathbb{W}\})$

Note that both types of constraints indicate that in case of a fault all possible value combinations may be observed whereas for the correct behavior the respective restricting relationship among the connections have to be fulfilled.

In order to define a constraint satisfaction problem, we first introduce the concept of value assignments for variables. Given a constraint system $(VARS, DOM, CONS)$, and variable $v \in VARS$, then $v = x$ with $x \in DOM(v)$ is a single assignment of a value x to the variable v. We further say that a set of single assignments where there at the maximum one single assignment for a variable as value assignment. A constraint c with scope $(v_1, \ldots, v_k)$ fulfills a value assignment $\{\ldots, v_1 = x_1, \ldots, v_k = x_k, \ldots\}$, if there exists a tuple $(x_1, \ldots, x_k)$ in $tl(c)$. Otherwise, we say that such a value assignment contradicts the constraint.

A constraint satisfaction problem for a given constraint system is the question whether a value assignment exists that fulfills all given constraints. If there is such a value assignment, then the constraint satisfaction problem is said to be itself fulfilled.

For example, the value assignment $a = 2, b = 2, c = 3, d = 3, e = 2, x = 6, y = 6, z = 6, f = 12, g = 12, ab\_M1 = F, ab\_M2 = F, ab\_M3 = F, ab\_A1 = F, ab\_A1 = F$ fulfills all constraints of the d74 circuit constraint system, whereas $a = 2, b = 2, c = 3, d = 3, e = 2, x = 6, y = 6, z = 6, f = 10, g = 12, ab\_M1 = F, ab\_M2 = F, ab\_M3 = F, ab\_A1 = F, ab\_A1 = F$ does not. Hence, the d74 constraint satisfaction system can be fulfilled.

Solving a constraint satisfaction problem is basically a search procedure for a value assignment that fulfills all constraints. This search for constraint systems having only constraints with finite tuple lists is well known to be exponential and its corresponding problem is well known to be NP-complete. For details about algorithms and heuristics we refer to interested reader to (Dechter 2003).

In the following we discuss the diagnosis problem and show how constraint solving can be used to solve the classical diagnosis problem. According to (Reiter 1987) a diagnosis problem is a tuple $(COMP, SD, OBS)$ where $COMP$ is a set of components, SD a logical sentence describing the behavior of the system, i.e., the system description, and OBS a set of observations. In out constraint based representation of the diagnosis problem, we assume a constraint representation of the system and additional constraints specifying the observations. The constraint representation of a diagnosis problem (or the diagnosis problem for short) is a tuple $(VARS, DOM, CONS \cup COBS)$ where $(VARS, DOM, CONS)$ is a constraint representation of a system comprising variables ab_C for every component C of the system, and $COBS$ is the constraint representation of all observations OBS.

For our d74 circuit, the constraint representation M_{VB} together with the constraint representation $COBS = \{(a,b,c,d,e,f,g), \{(2,2,3,3,2,10,12)\}\}$ specifying observations forms a diagnosis problem.

The results of a diagnosis problem, i.e., the diagnoses, are subsets of the set of components $COMP$. We obtain these subsets from the solutions of the corresponding constraint problem via taking one value assignment that is a solution, and putting all components C for which the corresponding variable ab_C is set to T into a set, i.e., if s is a solution of the constraint representation of a diagnosis problem, then its corresponding diagnosis is $\Delta_s = \{C | ab\_C = T \in s\}$. When computing all solutions from the constraint representation, we obtain all possible diagnosis. As usual, we define a diagnosis to be minimal if there exists no subset, which is itself a diagnosis. Of course, we are always interested in only computing minimal diagnosis in the most efficient way.

In the following we discuss briefly a diagnosis algorithm that computes minimal diagnosis of increasing size. This can be achieved via restricting the number of ab_C variables to be set to true using constraints. In this way we are able to compute diagnoses up to a pre-specified size. The necessary additional constraints are added during diagnosis computation in diagnosis algorithms like ConDiag. (Nica and Wotawa 2012) introduced the ConDiag algorithm that computes minimal diagnoses up to a predefined size using a constraint representation of the diagnosis problem. (Nica et al. 2013) compared ConDiag with other diagnosis algorithms showing a good overall runtime. In order to be self-contained we briefly discuss the ConDiag algorithm, which is given in Algorithm 1.

Algorithm 1 ConDiag$((VARS, DOM, CONS \cup COBS), COMP, n)$

Input: A constraint model $(VARS, DOM, CONS \cup COBS)$ of a system having components $COMP$ and the desired diagnosis cardinality n

Output: All minimal diagnoses up to the predefined cardinality n

1: Let DS be $\{\}$
2: Let M be $CONS \cup COBS$
3: **for** $i = 0$ **to** n **do**
4: $\quad CM = M \cup \{|\{ab_C | C \in COMP \wedge ab_C = T\}| = i\}$
5: $\quad S = \mathcal{P}\left(\textbf{CSolver}(VARS, DOM, CM)\right)$
6: $\quad$ **if** i is 0 **and** S is $\{\{\}\}$ **then**
7: $\qquad$ **return** S
8: $\quad$ **end if**
9: $\quad$ Let DS be $DS \cup S$.
10: $\quad M = M \cup \{\neg(\mathcal{C}(S))\}$
11: **end for**
12: **return** DS

The ConDiag algorithm computes diagnoses starting with cardinality 0 to the predefined size n that has to be provided as parameter. In each step, we are searching for solutions that have exactly a size of i (Step 4). All these solutions are added to the set of solutions in Step 9. In order to prevent the computation of non-minimal diagnoses additional constraints saying that we are not interested in superset diagnoses are added (see Step 10). ConDiag returns all minimal diagnoses up to size n and the empty diagnosis if the system works as expected.

When using ConDiag on the M_{VB} model of the d74 circuit, we obtain two single fault diagnoses $\{M1\}$ and $\{A1\}$ and two double fault diagnoses $\{M2, A2\}$ and $\{M2, M3\}$.

Qualitative models for diagnosis

In the previous section we illustrated the basic definitions using the quantitative model of the d74 circuit M_{VB} based on constraints over integer values. In order to speed up the diagnosis computation especially for large systems comprising thousands of components, we have to use appropriate abstractions. In software debugging data and control dependencies can be used for this purpose like in program slicing (Weiser 1982). Based on static slices (Friedrich, Stumptner, and Wotawa 1999) developed a model that could be easily integrated into model-based diagnosis, and which was later proved to be equivalent to program slicing (Wotawa 2002).

All these abstractions are not abstractions in the sense of homomorphic functions applied to a quantitative space in order to obtain a qualitative representation. Instead these abstractions introduce the idea of classifying variables or values of connection between components to be either correct or incorrect in a particular diagnosis problem. Hence, instead of using particular values, e.g., from the integer domain, the dependency-based models use classifications, which are based on deviations between the actual and the expected behavior. For a very detailed analysis of such deviation models in the context of diagnosis we recommend to consult (Struss 2004).

In the following, we discuss two dependency-based models, and show how they can be represented using constraints. We start with the dependency-based model of (Friedrich, Stumptner, and Wotawa 1999) we call M_D^{orig}. There the authors introduce a model of a component C having m inputs and one output. This models states that the output can only be correct, if the component is correct and all inputs have a correct value, i.e.: $\neg ab_C \rightarrow (\bigwedge_i^m in_i = ok \rightarrow out = ok)$. In the case the component is correct, but one input is not, the output may be correct or not correct. In the case of a faulty component, there is no way of determining the correctness status of the output from any correctness information of the input. When taking this thoughts into consideration, then we obtain the following table constraint for a component C with $n = 2$ inputs.

ab_C	in_1	in_2	out
F	ok	ok	ok
F	$\neg ok$	ok	ok
F	ok	$\neg ok$	ok
F	$\neg ok$	$\neg ok$	ok
F	$\neg ok$	ok	$\neg ok$
F	ok	$\neg ok$	$\neg ok$
F	$\neg ok$	$\neg ok$	$\neg ok$
T	.	.	.

In this table and also the following ones a $'.'$ stands for any possible value. Hence a row with $'.'$ represents multiple rows when changing the placeholder $'.'$ with possible values.

If we use this component model for all components of the d74 circuit from Figure 2 and further set the observations as follows $COBS = \{(a,b,c,d,e,f,g), \{(ok, ok, ok, ok, ok, \neg ok, ok\}\}$, which represent the observations used for diagnosing the d74 in the previous section, then we obtain the following three minimal diagnoses: $\{M1\}$, $\{M2\}$, and $\{A1\}$. When comparing this result with the previous one we see that when considering integers, we have two diagnoses which are supersets of $\{M2\}$. Hence, when considering dependencies only we lose some information, which leads to a larger search space of potential diagnoses including all their supersets.

The underlying reason for this decrease in precision of diagnosis is that the model does not consider the case where a faulty value does not propagate through the whole system. For example, if we consider a logical and gate and we know that one input is false, then the other input does no longer determine the value of the output. Hence, any faulty value occurring will never be visible on sider of the output. This behavior is named coincidental correctness in software debugging and always influences the fault localization capabilities.

In order to handle coincidental correctness using a dependency-based model, we have to distinguish two cases: (1)There are component where coincidental correctness may occur, e.g., for logical gates. (2) There are other cases, where coincidental correctness is at least very unlikely, e.g., when considering a function for adding two integers. In the latter case, we can state that a correct output value for a working component also implies that all inputs are working, i.e., $\neg ab_C \rightarrow (\bigwedge_i^m in_i = ok \leftrightarrow out = ok)$. (Hofer and Wotawa 2014) introduced this improved model for debugging Spreadsheet programs handling coincidental correctness we call M_D^{CC}. The constraint representation of a two inputs component takes care of the bi-implication used in the component model, which is only allowed to be used if no coincidental correctness may occur.

ab_C	in_1	in_2	out
F	ok	ok	ok
F	$\neg ok$	ok	$\neg ok$
F	ok	$\neg ok$	$\neg ok$
F	$\neg ok$	$\neg ok$	$\neg ok$
T	.	.	.

When using the model M_D^{CC} for diagnosing the d74 circuit and the previously used set of observations, we obtain again two single fault diagnoses $\{M1\}$, $\{M2\}$ and also one double fault diagnosis $\{M1, A2\}$. The other double fault diagnosis $\{M2, M3\}$ is missing. The reason here is that this model is not able to handle the case that two faulty inputs may lead to a correct output, which might happen even for operations on integer domains. Adding the tuple $(F, \neg ok, \neg ok, ok)$ to the table solves this issue. In the following we refer to the model M_D^{CC} extended with the tuple as M_D.

The dependency-based models discussed can be seen as the most abstract form of deviation model only considering values to be either correct (i.e., ok) or incorrect (i.e. $\neg ok$). A less abstract model may allows to distinguish cases where a value is smaller, equivalent, or larger than an expected value. In the following we discuss such a model and introduce abstraction formally in order to allow comparing such models with others.

When dealing with comparisons like smaller $<$, equivalent $=$, or larger $>$, we have to introduce tabular constraints for the different operators. In case of multiplication and addition, the constraints are the same but for others like subtraction adaptation for capturing the different semantics are necessary. In the following table we summarize the constraints handling the behavior of addition and multiplication components. There we state that in case of equivalent inputs we also obtain an output value with no deviation. In case one input is smaller (or larger) and the other is equivalent, the output also is expected to be smaller (or larger respectively). In case we have one smaller and one larger input value, we cannot say anything about the output. Hence, in such a case all output values may occur. If the operator (or component) is said to be faulty, all combinations of values are possible.

ab_C	in_1	in_2	out
F	$=$	$=$	$=$
F	$<$	$=$	$<$
F	$=$	$<$	$<$
F	$<$	$<$	$<$
F	$>$	$=$	$>$
F	$=$	$>$	$>$
F	$>$	$>$	$>$
F	$<$	$>$	$=$
F	$<$	$>$	$<$
F	$<$	$>$	$>$
F	$>$	$<$	$=$
F	$>$	$<$	$<$
F	$>$	$<$	$>$
T	.	.	.

We call the comparison model based on such a table M_C.

Obviously, the deviation model based on comparison gives additional information for the diagnosis process. However, the question is whether there is an improvement of the accuracy of the obtained diagnosis.

Let us start with continuing the d74 example. From the value-based model we obtain the following observations:

a	b	c	d	e	f	g
$=$	$=$	$=$	$=$	$=$	$<$	$=$

This together with the model for the components $M1, M2, M3, A1, A2$ allows for computing again 2 single fault diagnoses $\{M1\}$ and $\{A1\}$, and the 2 double fault diagnoses $\{M2, M3\}$ and $\{M2, A2\}$. Hence, there is no improvement in accuracy for this example.

If we change the observations, i.e., assuming $g = 10$ and $f = 14$ the situation changes. The value-based model allows for computing no single fault but 8 double fault diagnoses: $\{M1, M2\}$, $\{M1, M3\}$, $\{M1, A2\}$, $\{M2, M3\}$,

$\{M2, A1\}$, $\{M2, A2\}$, $\{M3, A1\}$, and $\{A1, A2\}$. The same diagnoses can be obtained when using the more accurate deviation model and the observations:

a	b	c	d	e	f	g
$=$	$=$	$=$	$=$	$=$	$<$	$>$

In case of the both the original and the improved dependency-based model we obtain one single fault diagnosis $\{M2\}$, and 4 double fault diagnoses $\{M1, M3\}$, $\{M1, A2\}$, $\{M3, A1\}$, and $\{A1, A2\}$. Hence, we see that the more abstract deviation models lead to the computation of less accurate diagnoses in some cases. We depict the diagnosis search space that includes the minimal diagnoses as well as all of their supersets in Figure 3.

In the next section we introduce abstraction and diagnosis accuracy formally, and further more discuss their relationship in detail.

Domain abstraction

We start with defining abstraction formally. Given two constraint models for diagnosis $M_1 = (VARS, DOM_1, CONS_1 \cup COBS_1)$ and $M_2 = (VARS, DOM_2, CONS_2 \cup COBS_2)$. We say that M_1 is more abstract than M_2, i.e., $M_1 \prec M_2$, if there exists a function $h : DOM_2 \mapsto DOM_1$ that makes the constraints equivalent, i.e., $\forall c_1 \in CONS_1 \cup COBS_1$ and $c_2 \in CONS_2 \cup COBS_2$ having the same scope, $h(c_2) = c_1$.

For this definition of abstraction, we use the following definition of the application of a function f on a constraint $((v_1, \ldots, v_k), tl)$: $f(((v_1, \ldots, v_k), tl)) = ((v_1, \ldots, v_k), f(tl))$ where f is defined on tuple list as follows: $f(tl) = \{(f(x_1), \ldots, f(x_k)) | (x_1, \ldots, x_k) \in tl\}$.

From our running d74 example using the function h defined as $h(=) = ok$, $h(<) = \neg ok$, and $h(>) = \neg ok$ we can easily check that the improved dependency-based model considering coincidental correctness M_D is more abstract than the comparison-based model M_C, i.e., $M_D \prec M_C$. It is worth noting that there is no such function h for the original dependency-based model and the original model handling coincidental correctness.

From the definition of $\prec$ the definition of model equivalence follows immediately. Let M_1 and M_2 are constraint models for diagnosis. M_1 and M_2 are equivalent, i.e., $M_1 \equiv M_2$, if and only if $M_1 \prec M_2$ and $M_2 \prec M_1$. Obviously, if the same function h can be used to show that $M_1 \prec M_2$ and $M_2 \prec M_1$, then h has to be a bijective function.

In the following we define diagnosis accuracy. For this purpose, we bear in mind that in case of pure consistency-based diagnosis, all supersets of minimal diagnoses are also diagnoses. This is ensured in all cases where we only be aware of the behavior of a correct component but do not know a component's incorrect behavior. If we use models of the faulty behavior minimal diagnoses are not characterizing all possible diagnoses anymore. See (de Kleer, Mackworth, and Reiter 1992) for a detailed discussion on this topic. In this paper we assume models that capture the correct behavior only. Hence, we know that all possible diagnosis can be characterized as follows. Let Δ-MIN be the set of all minimal diagnoses obtained from a constraint model $M = (VARS, DOM, CONS \cup COBS)$. The set of all diagnoses comprises the minimal diagnoses and all of their supersets, i.e., $\Delta\text{-SET}^M = \{\Delta | \exists \Delta' \in \Delta\text{-MIN} : \Delta \supseteq \Delta'\}$. Thus $\Delta\text{-SET}^M$ spans the whole search space of diagnosis. Using this definition we are able to define accuracy as the ability of a model M to come up with the smallest possible set $\Delta\text{-SET}^M$. To compare two constrain models used for diagnosis, we only need to compare their search spaces.

Given two constraint models for diagnosis $M_1 = (VARS, DOM_1, CONS_1 \cup COBS_1)$ and $M_2 = (VARS, DOM_2, CONS_2 \cup COBS_2)$. We say that M_1 is less accurate than M_2, i.e., $M_1 \prec_A M_2$, iff $\Delta\text{-SET}^{M_1} \supset \Delta\text{-SET}^{M_2}$. M_1 is as accurate as M_2, i.e., $M_1 \doteqdot M_2$, iff $\Delta\text{-SET}^{M_1} = \Delta\text{-SET}^{M_2}$. M_1 is less or equal accurate M_2, i.e., $M_1 \preceq_A M_2$, iff $M_1 \doteqdot M_2$ or $M_1 \prec_A M_2$.

When considering the search spaces for diagnosing the d74 circuit given in Figure 3 we see that the model M_D is less accurate than M_C, i.e., $M_D \prec_A M_C$.

In the following theorem we manifest the relationship between abstraction and diagnosis accuracy.

Theorem 1. *Given two constraint models for diagnosis $M_1 = (VARS, DOM_1, CONS_1 \cup COBS_1)$ and $M_2 = (VARS, DOM_2, CONS_2 \cup COBS_2)$. If M_1 is more abstract than M_2, then M_1 is less or equal accurate than M_2, i.e., $M_1 \prec M_2 \rightarrow M_1 \preceq_A M_2$.*

Proof. To prove the theorem we first assume that we have two models M_1 and M_2 where $M_1 \prec M_2$. From this follows that there exists a function h, which maps the elements of DOM_2 to elements of DOM_1 such that the tuple sets for each component becomes equivalent using only elements from DOM_1. What we have to proof is that $\Delta\text{-SET}^{M_1} \supseteq \Delta\text{-SET}^{M_2}$, i.e., for all $\Delta \in \Delta\text{-SET}^{M_2}$ it follows that $\Delta \in \Delta\text{-SET}^{M_1}$. We prove this by contradiction. Let Δ be in$\Delta\text{-SET}^{M_2}$ but $\Delta \notin \Delta\text{-SET}^{M_1}$. Because $\Delta \in \Delta\text{-SET}^{M_2}$ we know that the constraint model $(VARS, DOM_2, CONS_2 \cup COBS_2 \cup ((\Delta), \{(T, \ldots, T)\}))$ is satisfiable. Because of the definition of $\prec$ we would get a corresponding constraint model $(VARS, DOM_1, CONS_1 \cup COBS_1 \cup ((\Delta), \{(T, \ldots, T)\}))$ when applying h to the constraints. Note that h has no effect on the constraint $((\Delta), \{(T, \ldots, T)\})$. But this constraint system has to be also satisfiable because of the construction of h. Hence, Δ is also element of $\Delta\text{-SET}^{M_1}$ contradiction our assumption, and the theorem hold. □

Note that a more abstract model does not cause less accurate diagnoses in all cases. For the d74 example, we saw that depending on the observations we obtain the same or a less accurate diagnosis for the more abstract model M_D when compared to M_C. It is also worth noting that the definition of more or less accurate can also be used independently from abstraction.

From Theorem 1 we obtain the following lemma, which states that equivalent models also have an equivalent diagnosis accuracy.

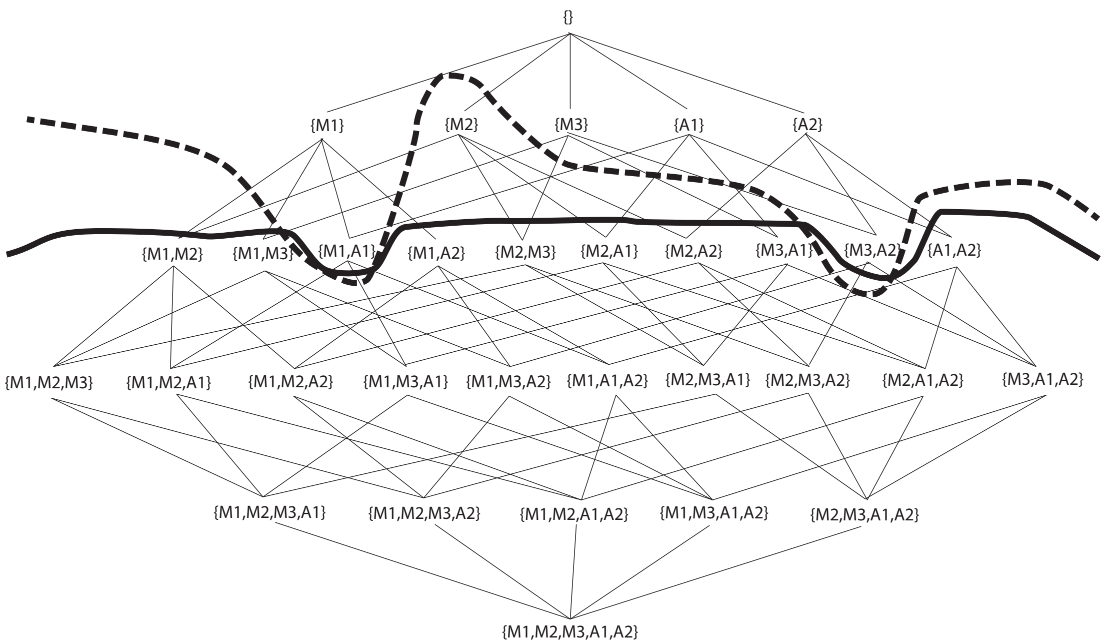


Figure 3: The diagnosis search space for the d74 circuit and assuming both outputs to behave incorrectly, i.e., $f = 10$ and $g = 14$. The dotted line shows the search space for the improved dependency-based model handling coincidental correctness whereas the solid line indicates the search space for both the value-based and the comparison-based model.

Lemma 1. *Let M_1 and M_2 be equivalent models, i.e., $M_1 \equiv M_2$, then M_1 and M_2 have the same diagnosis accuracy, i.e., $M_1 \equiv M_2 \rightarrow M_1 \doteq M_2$.*

Proof. The lemma follows directly from Theorem 1 and the definition of equal diagnosis accuracy. □

Obviously, there might be cases where two models have the same accuracy but there is no mapping between model elements. Hence, we are not allowed to conclude model equivalence from equal diagnosis accuracy.

Experimental results

In order to motivate the use of qualitative models for diagnosis we carried out some experiments based on a parametrizable circuit comprising components for adding and multiplying integers. Our underlying research questions are: (1) whether the discussed qualitative models decrease the running time of diagnosis compared to a model based on integer values, and (2) whether the accuracy of diagnosis does not decrease substantially.

For generating a parametrizable circuit we implemented a circuit generator having 2 parameters: (1) the number of components directly connected to the inputs, and (2) the number of outputs. The generator constructs the circuit level by level, where in each level the number of components is reduced by 1. We stop at a level where the number of components is equivalent to the wanted number of outputs. We further assume that each component is either a component for adding two integers or multiplying two integers. The functionality of each component changes at every level. Components from level i are only connected to components from level $i + 1$ where two components of $i + 1$ share one output of a component of level i. For example, in Fig-

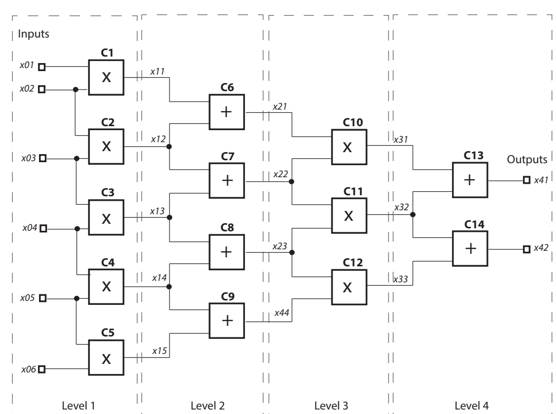


Figure 4: A generated circuit having 5 components directly connected to the inputs and 2 outputs.

ure 4 we depict the circuit, which can be generated using 5 components in level 1 and 2 outputs. Obviously, the number of inputs is always $n + 1$ if n is the number of components in the first level, and the number of wanted outputs k has to fulfill equation $n \geq k \geq 1$. For the experiment, we use the values 2 and 3 in an alternating way and computed the expected outputs when constructing the circuit. The implementation of the circuit generator return a value-based, an improved functional dependency model, and a comparison model of the circuit using the parameters. All the models can be executed using the Minion (Gent, Jefferson, and Miguel 2006) constraint solver. For the experiments we used the latest Minion Version 1.8 (available at `http://constraintmodelling.org/`).

In the experiment we generated 6 smaller circuits with $2 \ldots 7$ components in the first level and exactly two outputs. We named the circuit c22, c32, ..., c72. The purpose of this experiment was to compare the diagnosis results and the running time for computing all single fault diagnosis using our 3 models, i.e.: the value-based model M_{VB}, the improved functional dependency based one M_D, and the comparison model M_C. Besides the predefined inputs, which are either 2 or 3 for M_{VB} and ok for the other models, we assumed every output except the last one at the bottom of the circuit to be correct. For the last output we set its value to 0 (for M_{VB}), to $\neg ok$ for M_D, and to $<$ for M_C. In Table 1 we summarize the obtained results when running the search for single fault diagnoses on a MacBook Pro, 2,8 GHz Intel Core i7, 16 GB memory, and OS X version 10.11.3.

Note that for the value based model we used a integer domain ranging from -300 to 300 in order to compute the diagnoses. From Table 1 we see that even for small circuit there is a substantially larger running time when using M_{VB} compared to the other models. It is worth noting that with the given integer domain range the circuits c62 and c72 cannot be solved using M_{VB}. Moreover, the functional dependency model M_D produces many more diagnoses than both other models. The reason here is the introduction of a tuple in the constraint table that allows for masking faults in case of two wrong input values and the highly interconnected structure of the circuits. The best model in terms of running time and diagnosis accuracy is M_C. It is also interesting to see that for the slightly larger circuits computing diagnoses within 10 seconds was not possible when using M_{VB}. In order to complete the first experiments we further studied the influence of the used integer domain to the running time of diagnosis when using the value-based model. See Figure 5 for the results. When considering the logarithmic scale we see an exponential increase of running time when doubling the space of integers. Hence, for larger numbers diagnoses using M_{VB} becomes infeasible and that even for very small systems.

The results show that qualitative models for expressing the propagation of deviations from expected values are very valuable for diagnosis purposes. The introduced comparison model M_C provides a good diagnosis running time and accuracy especially when compared with the functional dependency model M_D.

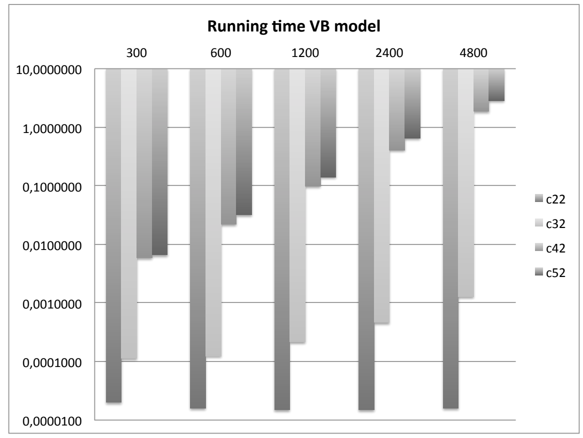


Figure 5: Minimum running time in seconds as a function of the size of the used integer domain.

Related research

The idea of using abstraction for diagnosis and in particular model-based diagnosis (Reiter 1987; de Kleer and Williams 1987) is not new. Initial work including (Mozetič 1991) and later (Autio and Reiter 1998) discussed the concept of structural abstraction, where sets of interconnected components are mapped to one component. The behavior of such a component is given using the sets of interconnected components. When using such an abstraction, we obtain a hierarchical model, where a component model in one level is given using the structure and behavior of the corresponding interconnected components. (Autio and Reiter 1998) discussed the resulting properties of such an approach in detail.

(Struss 1992) discussed modeling including abstraction and refinement in very much detail. (Sachenbacher and Struss 2003; 2005) introduced a different abstraction approach where quantitative domains are mapped to qualitative ones considering value boundaries influencing the behavior of the system. Such boundaries depend on the given diagnosis problem. Therefore, the authors suggested to use an automated abstraction approach for solving this issue.

In contrast, to these previous papers, we focus on deviation models for diagnosis and state a theory allowing to compare them using the introduced definition of abstraction, which is close to (Struss 1992). Although, we used example from classical hardware diagnosis to illustrate the concepts, we are driven by the idea of coming up with automated debuggers for programs. There qualitative representations seems to be very useful and appropriate.

It is also worth mentioning other work related to abstraction of programs. (Cousot and Cousot 1977) introduced the concept of program abstraction providing a theoretical framework. The focus of (Cousot and Cousot 1977) was on the execution part and there the consequences of introducing abstraction. In our work, the focus is on fault localization and deviation models.

Circuit		M_{VB}				M_D				M_C			
name	comps	singl. f.	min. T	avg. T	max T	singl. f.	min. T	avg. T	max T	singl. f.	min. T	avg. T	max T
c22	2	1	0.000020	0.000026	0.000035	1	0.000018	0.000024	0.000031	1	0.000018	0.000024	0.000033
c32	5	2	0.000113	0.000172	0.000258	2	0.000051	0.000060	0.000075	2	0.000051	0.000066	0.000081
c42	9	3	0.005931	0.006842	0.009888	4	0.000093	0.000121	0.000144	3	0.000084	0.000119	0.000147
c52	14	4	0.006591	0.006940	0.007364	7	0.000163	0.000200	0.000241	4	0.000169	0.000194	0.000213
c62	20	-	-	-	-	16	0.000313	0.000417	0.000479	5	0.000272	0.000306	0.000344
c72	27	-	-	-	-	42	0.000679	0.000882	0.000966	6	0.000411	0.000462	0.000561

Table 1: Empirical diagnosis results obtained for the different models. Besides the number of diagnoses, the minimum, average, and maximum running time in seconds for every model is given.

Conclusions

In this paper we formalized diagnosis as constraint satisfaction problem and introduced deviation models for fault localization. In addition, we discussed a framework that allows for comparing different models and to state whether one model is an abstraction of another model. Moreover, we present first empirical results showing that a deviation model based on the qualitative values *smaller*, *equivalent*, or *larger* behaves similar to a representation based on concrete values. The obtained running time for computing single fault diagnosis is also very much promising and may raise its usability for fault localization in programs.

The empirical evaluation is of course limited and has to be extended in the future. Moreover, it is planned to use the comparison model M_C in the domain of fault localization of spreadsheets, where a fast response time is required even in cases of large spreadsheets comprising hundreds of non-empty cells. There we expect that the use of qualitative deviation models improves fault localization substantially.

Acknowledgments

The work described in this paper has been been funded by the Austrian Science Fund (FWF) project DEbugging Of Spreadsheet programs (DEOS) under contract number I2144 and the Deutsche Forschungsgemeinschaft (DFG) under contract number JA 2095/4-1.

References

Autio, K., and Reiter, R. 1998. Structural abstraction in model-based diagnosis. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*.

Cousot, P., and Cousot, R. 1977. Abstract interpreation: A unified lattice model for static analysis of programs by construction of approximation of fixpoints. In *in Proc. POPL'77*, 238–252. Los Angeles: ACM.

de Kleer, J., and Williams, B. C. 1987. Diagnosing multiple faults. *Artificial Intelligence* 32(1):97–130.

de Kleer, J.; Mackworth, A. K.; and Reiter, R. 1992. Characterizing diagnosis and systems. *Artificial Intelligence* 56.

Dechter, R. 2003. *Constraint Processing*. Morgan Kaufmann.

Friedrich, G.; Stumptner, M.; and Wotawa, F. 1999. Model-based diagnosis of hardware designs. *Artificial Intelligence* 111(2):3–39.

Gent, I. P.; Jefferson, C.; and Miguel, I. 2006. Minion: A fast, scalable, constraint solver. In *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 2006)*.

Hofer, B., and Wotawa, F. 2014. Why does my spreadsheet compute wrong values? In *Proceedings of the International Symposium on Software Reliability Engineering (ISSRE)*, volume 25, 112 –121.

Hofer, B.; Wotawa, F.; Abreu, R.; and Außerlechner, S. 2015. Testing for distinguishing repair candidates in spreadsheets - the mussco approach. In *27th International Conference on Testing Software and Systems (ICTSS)*, 124–140.

Mozetič, I. 1991. Hierarchical model-based diagnosis. *International Journal of Man-Machine Studies* 35:329–362.

Nica, I., and Wotawa, F. 2012. Condiag – computing minimal diagnoses using a constraint solver. In *Proceedings of the International Workshop on Principles of Diagnosis (DX)*, 185–191.

Nica, I.; Pill, I.; Quaritsch, T.; and Wotawa, F. 2013. A route to success – a performance comparison of diagnosis algorithms. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial Intelligence* 32(1):57–95.

Sachenbacher, M., and Struss, P. 2003. Automated qualitative domain abstraction. In Gottlob, G., and Walsh, T., eds., *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, 382–387. Morgan Kaufmann.

Sachenbacher, M., and Struss, P. 2005. Task-dependent qualitative domain abstraction. *Artif. Intell.* 162(1-2):121–143.

Struss, P. 1992. What's in sd? towards a theory of modeling for diagnosis. *Readings in model-based diagnosis* 419–449.

Struss, P. 2004. Deviation models revisited. In *Working Papers of the 15th International Workshop on Principles of Diagnosis (DX-04)*.

Weiser, M. 1982. Programmers use slices when debugging. *Communications of the ACM* 25(7):446–452.

Wotawa, F. 2002. On the Relationship between Model-Based Debugging and Program Slicing. *Artificial Intelligence* 135(1–2):124–143.

Unified approach to qualitative motion planning in dynamic environments

Domen Šoberl and **Ivan Bratko**
Faculty of Computer and Information Science
Večna pot 113, Ljubljana, Slovenia
{domen.soberl, ivan.bratko}@fri.uni-lj.si

Abstract

Traditional motion planning methods rely on precise kinematic models to either compute the goal trajectory off-line, or to make on-line decisions based on current observations from a dynamic environment. With the increasing use of qualitative modeling in cognitive robotics, different planning approaches are needed to handle the lack of numerical data in manually constructed or autonomously learned qualitative domain theories. We propose a new motion planning algorithm that makes on-line decisions based on given qualitative domain description to reach a goal state. Decisions are stated in the form of simple qualitative actions that can easily be interpreted by robot's controller and transformed to a numerical output. We demonstrate its use on three classical problems: pursuing, obstacle avoidance and object pushing.

Introduction

Motion planning techniques for robotic systems in closed and controlled environment have become very efficient in recent years. Methods such as *Probabilistic RoadMap* (PRM) (Kavraki et al. 1996) or *Rapidly-exploring random tree* (RRT) (LaValle 2006) can produce a detailed motion plan, if precise mathematical model of the system's dynamics is provided. Such off-line planning is often unsuitable in real-world scenarios where complete and robust theories are rare, while environment is unpredictably changing, which is especially the case when human interaction is present.

To deal with challenges of motion planning in dynamic environments, reactive planners were proposed even in the early beginnings of motion planning, with emphasis on path construction and obstacle avoidance using a car-like vehicle (Fraichard, Hassoun, and Laugier 1991). Later works employ adaptations of certain off-line motion planning algorithms for dynamic plan modification. In (Leven and Hutchinson 2002) the PRM algorithm is used in two stages. A roadmap that corresponds to an obstacle-free environment is first built off-line and later dynamically updated with obstacle information by an on-line planner. A partial replanning with RRT is possible during the execution of the plan by recomputing only the necessary tree branches (Ferguson, Kalra, and Stentz 2006).

For a planner to predict exact future configurations that follow certain actions, a precise kinematic model of the system is needed. With the increasing use of qualitative modeling in cognitive robotics during the last few years, new challenges in automated planning emerged due to the lack of numerical information in qualitative domain theories. One of the main reasons to prefer qualitative over traditional numerical modeling, especially in the area of autonomous concept discovery, is its tendency to capture more general relations and express meaningful concepts which can significantly simplify the agent's theory (Bratko 2011). It has been demonstrated (Troha and Bratko 2011) that a wheeled robot can learn qualitative physics of pushing a rectangular box by experimentation. We later showed how such models can be used by a robot to plan the pushing of arbitrary convex polygonal objects (Šoberl, Žabkar, and Bratko 2015). However, our planner was specialized for planning actions of pushing and possibilities for a more general solution still needed to be addressed.

In this paper we propose a general motion planning algorithm together with a new domain description language which allows domain relations to be stated in the form of *monotonic qualitative constraints* (e.g. (Bratko and Šuc 2003)). We demonstrate the algorithm on three classical problems: pursuing, obstacle avoidance and object pushing. Our work differs from *symbolic qualitative planning* (Wiley, Sammut, and Bratko 2014), where qualitative plans are elaborate and used as a basis for further numerical learning. Our planner produces more basic qualitative actions, interpreted directly by the robot's controlling mechanism, and is therefore suitable for more straightforward tasks.

The rest of the paper is structured as follows. In the following section we give a general description of our planning mechanism and define the notion of robotic domain and qualitative action as used by our planner. Next we introduce our domain description language and describe its individual elements. We continue with in-depth analysis on how the planner interprets the given domain description to produce appropriate actions. We then describe three different experiments that we conducted in a simulated environment and present the results. Finally, we conclude and discuss our future work.

Our planning approach

We presume that the configuration space of the robot is continuous and connected. Let $x_1, \ldots x_n$ denote domain attributes where $x_1, \ldots, x_k$ are directly controllable, meaning that the control mechanism is able to increase or decrease their value, and so they represent output signals. All other attributes are controllable indirectly, using relations defined by the given qualitative model. A relation $x_3 = M^{+-}(x_1, x_2)$ states that x_3 monotonically increases in x_1 and decreases in x_2. If both x_1 and x_2 increase or decrease, we deem the direction of x_3 inconclusive. We define qualitative action as a mapping of controllable attributes $x_1, \dots, x_k$ to $\{+, -, 0\}$ and write $a = (x_1^+, x_2^-, x_3^0)$ to denote an action a that maps $x_1 \mapsto +$, $x_2 \mapsto -$ and $x_3 \mapsto 0$, meaning that the controller should increase the value of x_1, decrease the value of x_2 and keep the current value of x_3. It is then up to the controller to choose the numerical step. In the past we achieved satisfying results with simulated and real robots using the trivial mapping: $x_i^+ \mapsto \max(x_i)$, $x_i^- \mapsto \min(x_i)$ and $x_i^0 \mapsto x_i$. Say such a signal represents the output power to a motor. Setting it to the highest value takes some time for the motor to actually reach the highest speed. Providing a sufficiently high refresh frequency, observed attributes rarely reach their extremes, as the planner tends to guide them to a certain value. This differs from the classical PID control principle in the fact that with PID controllers the target value of the output signal is known in advance, whereas in our case it is the task of the planner to make such decisions.

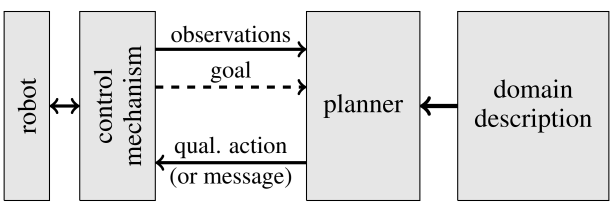


Figure 1: Reactive qualitative planning mechanism. Goal can be stated dynamically or set statically as a part of domain description.

Communication between the planner and the robot's control mechanism is shown in Fig. 1. The controller starts a new communication cycle by updating the values obtained from the robot's sensing system. The planner uses values from the previous cycle to record velocity (the signed speed) of each attribute. Those velocities are needed internally by the planner, but being regarded as attributes they can be used as a part of domain description by adding the apostrophe character `'` after an identifier (e.g. `x'` to denote the velocity $\dot{x}$). The goal state can be specified statically as a part of domain description, or given dynamically, together with input attribute values. A Goal condition is given as a set of simple equations $\{x_i = g_i\}$, where g_i is the goal value of attribute x_i and can be a constant or a numerical expression. If the robot should pursue an object, the goal can be specified as equality of their respective coordinates. If more than one goal is given (dynamically, statically or mixed), the planner chooses the most promising one for that cycle. The planner can implicitly set additional goals to satisfy given numerical constraints or avoid their violation. If the robot should avoid an obstacle, a constraint $D > 0$, where D is the distance between the robot and the object, should imply actions that lead away from the obstacle, especially if no other goal is given.

As soon as input values are set, the planner responds with the next action to be performed, or with one of the following messages:

- *No goal.* No goal to reach, neither explicit or implicit. This happens if no goal has been specified or when expressions that are part of the goal definition failed to be evaluated. This is usually the result of a poor domain definition or insufficient input data.
- *No solution.* Constraints have been violated. Depending on the design of the problem, this can represent an unwanted situation (e.g. a configuration where a reset is needed) or a part of the planning process. We demonstrate the latter case in our third experiment - pushing an object - where a constraint on action score fires *no solution*, when no available action is good enough. We then reposition the robot to a more favorable pushing position.
- *Goal reached.* According to the current speed of the system, goal attributes are close enough to their goal configurations. This means that for each goal attribute the distance between its current and goal value is smaller than the distance it can make between two communication cycles. However, the controller always has the liberty to impose its own conditions and end the process on its own terms.

Returning an answer the communication cycle ends. It is not required to invoke cycles with a constant frequency, but it is helpful for the planner to properly evaluate the tolerance of the goal state.

Domain description language

To describe the problem domain, qualitative model alone is usually insufficient. Besides defining actions and specifying goal conditions, additional numerical constraints often need to be set. Those define undesirable system configurations and make the configuration space non-convex. We propose a new domain description language, suitable to our planning approach.

A domain is described by a set of statements $\{S_1, S_2, \dots, S_n\}$, where each statement can evaluate as *true*, *false* or *inconclusive*. Statements can be interdependent, and since the language is declarative, it is the task of the planner to resolve the dependency of their evaluation. If one of the statements fails (is evaluated as *false*), the *no solution* answer is triggered. A statement is evaluated as *inconclusive* if insufficient data is provided, in which case it is ignored. This may also lead to inconclusive goal states and the *no goal* answer in the case no goal state is conclusive. There are seven types of statements: boolean expression, attribute range, numerical equation or inequation, qualitative relation, conditional statement, definition of action and specification of a goal state.

Boolean expressions

Statements can be combined into boolean expressions using the standard logical operators. The use of primitives *true* and *false* is also permitted. If one or more statements in the expression evaluate as *inconclusive*, the expression evaluates to *inconclusive* only if the result cannot be derived from other values. For instance, boolean expression *true* $\vee$ *inconclusive* evaluates as *true*, while *true* $\wedge$ *inconclusive* evaluates as *inconclusive*.

Attributes and their ranges

Attributes are used as variables in classical programming languages, but their nature is significantly different. All statements must agree on a single value of an attribute for the current cycle. Statements that propose different values to the same attribute, fail. If the value of an attribute is not set, statements that use it are evaluated as *inconclusive*. Besides the computed (or observed) numerical value, a predicted qualitative[1] value can also be assigned to an attribute. However, this can only be done implicitly by the planner.

An attribute can be bound to a range. There are three distinct ranges:

- *An interval*. Open, closed or half-closed intervals can be stated. Statement `x in I[a, b)` binds the attribute x to half-closed interval $[a, b)$ where a and b are arbitrary numerical expressions. The statement succeeds if no statement evaluates x to a value outside the given boundaries. This also sets an implicit goal for the planner to avoid both extremes. Infinity keywords `-inf` and `inf` can also be used instead of a or b respectively, making the planner avoid only one extreme.
- *A sphere* S^1. Statement `x in S[a, b)` makes the attribute x circular and normalizes its value to the given range. This statement always succeeds, sets no implicit goal, but allows the planner to choose between two directions to reach a desired state. A typical use of circular attributes is to specify rotations, e.g. orientation of the robot within $[0°, 360°)$.
- *A qualitative range*. Allows the user to limit the predicted value. Statement `x in Q[0+]` results in elimination of all actions for which the planner deduces a negative predicted value.

Numerical equations and inequations

There is no clear distinction between comparison and assignment. If all attributes contained in a statement evaluate to some value, the statement either succeeds or fails. Simple equations (a single attribute on either the left or the right side) try to set the value so that the statement succeeds. This can happen only if all other statements agree on the same value. Simple inequations set an implicit goal to the single attribute to avoid the violation of the rule. Complex comparisons remain inconclusive if one or more attributes cannot be evaluated. Note that a statement of the form `x = x + 1`, so frequent in classical programming languages, definitely fails here.

[1]Currently, our planner can only predict qualitatively. Incorporating some form of numerical machine learning should enable the planner to also make numerical predictions.

Qualitative relations

Qualitative relations between attributes are the basis for the planner to predict the outcome of an individual action. A domain is well defined when one can track relations from controllable to goal attributes. Relations are stated in the form of equations, with the M-notation on one side, and a single attribute on the other. As in the case of numerical equations, this can be interpreted as both, an assignment and comparison.

Consider the relation $\dot{D} = M^{-+}(v, \omega)$ from our second example, which states that (under certain conditions) the speed of distancing the robot from the obstacle increases by monotonically decreasing its forward speed v and increasing its rotational speed ω. Recall that every attribute holds two values, one numerical and one qualitative (its predicted future dynamic), of which either one can be set or unset. If the planner can derive qualitative values of all three attributes, $\dot{D}$, v and ω, and those values agree on the stated relation, the statement will succeed. If the qualitative value of $\dot{D}$ cannot be derived from other stated relations, the statement will set it so that the statement succeeds, while its numerical value will remain intact. If two or more independent attributes contradict, say we have v^+ and ω^+, the above statement is inconclusive. It sets no values and has no impact on the final outcome.

Conditional statements

A conditional statement is an implication of the form $\{G_1, \ldots, G_m\} \Rightarrow \{S_1, \ldots, S_n\}$, where statements G_i represent the guard. Only if all guarding statements G_i succeed, the implied statements S_j are considered a part of domain definition. An inconclusive guarding statement makes the guard and thus the whole statement inconclusive. By the rule of implication, a failed guard evaluates the whole statement as *true*, while omitting the implied statements S_i from the rest of definition.

If a conditional statement succeeds, its guarding statements G_i also becomes a part of domain definition. Consider a statement of the form $\{x = 1\} \Rightarrow \{S_0, \ldots, S_n\}$ whose guard succeeds by setting the attribute x to 1. The equality $x = 1$ is always considered by statements $S_0, \ldots, S_n$, but becomes a globally valid assertion only if the whole implication succeeds.

Conditional statements can be used to describe qualitative models as sets of qualitative relations that hold under given conditions, e.g. the implication $\{\varphi > 90°\} \Rightarrow \{\dot{D} = M^{-+}(v, \omega)\}$ states that the right-hand relation holds only when $\varphi > 90°$. Models obtained by programs such as QUIN (Bratko and Šuc 2003) or Padé (Žabkar, Bratko, and Demšar 2007) can easily be rewritten using such a form.

Actions

Classes of legal actions are defined by statements of the form `action(`$x_1, x_2, \ldots, x_n$`)`, giving the planner the freedom to choose among 3^n possible actions. If more than one such statement is used, the planner has the option to choose among different types of actions. Specifying no actions is considered a poor domain description and results in the *no* *solution* message, unless a goal state has been reached initially.

Each action is evaluated and assigned a score. A higher value indicates an action that will have a more desirable effect in favor of reaching the goal state. A positively scored action is predicted to advance the configuration closer to the goal state while negatively scored action should result in moving away from the goal. A zero value could mean an action without an effect or leading to a state equally distant to the goal state. The planner evaluates an action by applying it to the current attribute state and deducing its qualitative effect through stated attribute relations, tracked down to a goal statement. If qualitative deduction reaches a valid goal statement, the score is computed and assigned to the action, otherwise the action is discarded. In the case of multiple goals, the closest goal is selected according to Manhattan distance over all attributes, normalized by their speeds:

$$\text{distance}(G) = \sum_{i=1}^{n} \frac{|g_i - x_i|}{|\dot{x_i}|} \tag{1}$$

where $G = \{g_1, \ldots, g_n\}$ are explicit goal values of attributes $x_1, \ldots, x_n$. When all actions are evaluated, the planner returns the one with the highest value, or triggers *no solution* if all actions were discarded.

It is possible to state additional constraints on action scores. Statement `s = action(`$x_1, x_2, \ldots, x_n$`)` will compare / assign the highest score to attribute s, which can further be used in other numerical statements. In our third experiment we used this feature to trigger *no solution* when no positively evaluated action was found, which was achieved by numerical constraint $s > 0$.

Goals

A goal condition is defined as a set of simple equations $\{x_i = g_i\}$, where x_i is an attribute and g_i a numerical constant or expression. To specify an explicit goal we use statements of the form `goal(`$x_1 \to a, x_2 \to b, \ldots$`)`. The *arrow* symbol makes a clearer distinction between the attribute and its goal value and also denotes an operation a bit different from the usual numerical comparison. The goal statement not only compares numerical values, but also assesses its distance and predicts future dynamics based on currently set qualitative values of attributes. For attributes with unset predictions, the planner will try to make predictions based on the direction of attribute's speed. To each explicitly defined goal, implicit goals may be added by the planner internally, to avoid possible constraint violations. An implicit goal is stated as a set of pairs $\{(x_i, I_i)\}$, where I_i is the interval to which the attribute x_i is bound. Attributes that are used in explicit goals are not used in implicit goals.

The goal statement always succeeds unless some of the values cannot be deduced, in which case it remains inconclusive. When successful, it outputs the score which is assigned to the action that is currently being evaluated. We discuss further details on score evaluation in the following section.

The planning algorithm

During each cycle, the planner follows the following algorithm:

1. Check if input values conflict with any of the statements given in domain description. If so, return *no solution*.
2. Locate valid *action* statements and generate the list of possible actions.
3. For each action repeat:

3.1. Set predicted values of attributes as specified by the action (e.g. action (v^+, w^-) sets qualitative part of v and w to $+$ and $-$, respectively). Check if set values conflict with any of the statements. If so, discard the action and return to step 3.

3.2. Following valid qualitative relations, deduce predictions for all possible attributes.

3.3. Locate valid *goal* statements. If no goal is found, discard the action and return to step 3.

3.4. Compute the score of each goal and assign the highest value to the action. If conditions of some goal are met, return *goal reached*.

3.5. If the assigned score conflicts with numerical statements, discard the action.

4. Return the action with the highest score. If no action is left, return *no solution*.

We say *valid* actions, goals and relations to emphasize the fact that any statement can be conditioned using conditional statements. Putting an action or a goal under a guard makes it possible to divide the planning problem into separate tasks or phases.

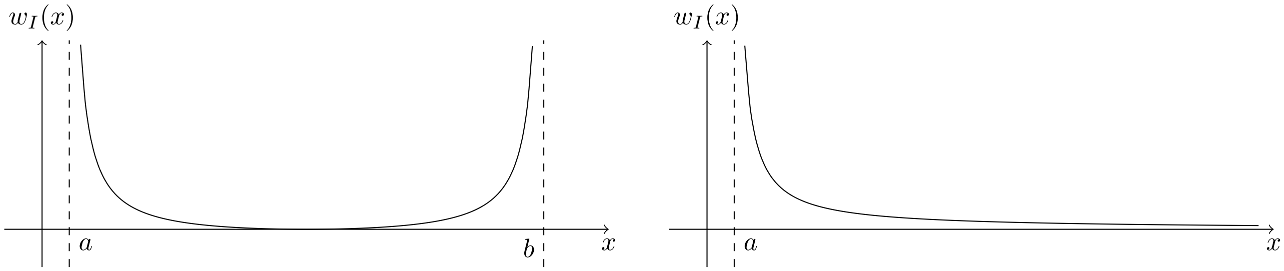
Consider an action a that is being evaluated under a goal $(x_1 \to g_1, \ldots, x_k \to g_k \mid (x_{k+1}, I_{k+1}), \ldots, (x_n, I_n))$, where g_i are explicitly defined goal values and I_i intervals, assigned to non-goal attributes $x_{i>k}$, and so comprise an implicit goal. Assume that all numerical values of x_i and g_i are set, and that all qualitative predictions of x_i were deduced under action a, making the goal statement successful. The score of action a is set using the following function:

$$\begin{aligned}\text{score}(a) = \sum_{i=1}^{k} p(x_i) \cdot \frac{g_i - x_i}{|\dot{x_i}|} \; + \\ + \sum_{i=k+1}^{n} p(x_i) \cdot w_I(x_i) \cdot |\dot{x_i}|\end{aligned} \tag{2}$$

For explicitly defined goals, each attribute contributes a weight proportional to the distance from its goal value and normalized by its speed. The greater the distance, the more important the attribute. Function $p(x_i)$ maps the qualitative value $+$, $-$ or 0 of x_i to its respective numerical representation $+1$, -1 or 0. This results in contributing a negative weight if the attribute is predicted to move away from its goal value.

Attributes that are part of an implicitly defined goal, thus bound to some interval, contribute their weights according to the w function, defined as

$$w_I(x) = \cos\left(\frac{\pi}{b-a} \cdot \left(x - \frac{a+b}{2}\right)\right)^{-1} - 1 \tag{3}$$



(a) Function w in the case $x \in I = (a, b)$.

(b) Function w in the case $x \in I = (a, \infty)$.

Figure 2: The weight of a bounded attribute is based on the w function.

for bounded intervals $I = (a, b)$, and

$$w_I(x) = \begin{cases} (x-a)^{-1}, & I = (a, \infty) \\ (b-x)^{-1}, & I = (-\infty, b) \end{cases} \tag{4}$$

for half-bounded intervals. The same function is used for closed and half-closed intervals. When an attribute is fully bounded, the planner will tend to keep its value close to the midpoint between both extremes. The contributed weight will stay relatively low for the major part of the interval, but will start to rise very rapidly when the attribute gets close to its extreme, as shown in Fig. 2. Weight is then multiplied by the speed of the attribute (see equation (2)), so faster attributes contribute more to the final score, as they are in higher danger to hit their forbidden zone quicker.

Experiments

To asses the performance of our planner we conducted three different experiments using a simple two-wheeled robot vehicle. Experiments were done in a simulator, assuming an overhead camera and object recognition system as sensory input. The sensory system recorded absolute location and orientation of single objects. All data were passed directly to the planner without any preprocessing. The refresh rate was 25 Hz and the planner was invoked whenever a change in configuration of objects was observed. Otherwise the values decided by the last cycle were held on the output.

The controller was able to interpret qualitative actions in the form (v^{+-0}, ω^{+-0}), where v is translational and ω angular velocity of the robot, being positive in the CCW direction. Using an independent sensory feedback, controller was able to adjust and maintain given speeds, but when instructed to increase / decrease one of them, it aimed for the maximum / minimum achievable value. Having a system with limited output capabilities, some action might not always be executable. Scenarios such as moving at a full speed forward while receiving instruction to increase ω and keeping v unchanged need a special consideration. In such cases our robot first lowered the forward speed by half, maintaining the same ratio between the left and the right wheel, and then proceeded with the intended action execution.

Pursuing objects

The goal of this task is to follow and eventually catch an animate object by choosing the appropriate translational v and angular ω speed of the robot at every step of the process. For this experiment we conducted no learning phase and designed the model manually. This way we demonstrated what we believe is an advantage of qualitative modeling. We were able to describe the domain intuitively, as we understood it, in a concise non-algorithmic way, and the robot behaved as we intended it.

Our reasoning was the following. We understand pursuing as the process of decreasing the distance to the target while orienting towards it. We believe the most efficient way to describe this domain is to use an egocentric approach (the robot sees itself as the center of the world). Let D denote the distance to the target and φ its angular offset, as depicted in Fig. 3. These need to be derived from the absolute robot position (x_0, y_0), its orientation θ, and position of the target (x, y). This can be done using the following equations:

$$\begin{aligned} D &= \sqrt{(x-x_0)^2 + (y-y_0)^2} \\ \varphi &= \operatorname{atan2}(y-y_0, x-x_0) - \theta \end{aligned} \tag{5}$$

We understand that increasing or decreasing the speed of the robot does not directly affect the distance D, but rather its speed $\dot{D}$. In relation to that, we recognize two distinct qualitative states: *the target is in front of the robot* ($|\varphi| \leq 90$), and *the target is behind the robot* ($|\varphi| > 90$). We also understand that the speed of φ decreases by increasing angular speed and vice versa. The goal should be reached by orientating towards the target ($\varphi \to 0$) and decrease the distance ($D \to 0$).

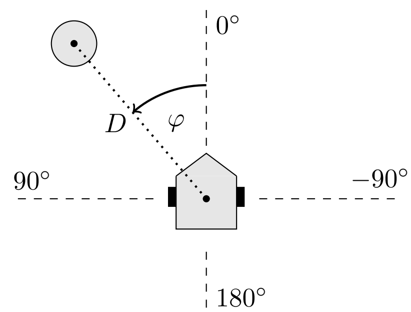


Figure 3: Relation of the robot to a target.

Listing 1 shows our domain description. Note that our implementation of the language demands statements to be terminated by a semicolon, and that we use the # symbol for comments. Single-statement guards and implications need not be enclosed in curly braces.

Listing 1 Pursuing an object

```
# Input values:
# x0, y0, theta - robot configuration
# x, y          - target position

phi in S[-180, 180); # phi is circular

# Egocentric values
D = sqrt((x - x0)^2 + (y - y0)^2);
phi = atan2(y - y0, x - x0) - theta;

# Qualitative model
abs(phi) <= 90; => D' = M-(v);
abs(phi) > 90;  => D' = M+(v);
phi' = M-(w);

action(v, w);
goal(D -> 0, phi -> 0);
```

A trajectory made by our robot during one of the trials is shown in Fig. 4. We initially positioned the robot facing backwards to its target. The pursuing began as soon as the target started moving from the left to the right with a constant speed. The robot first made a backward turn by 90° and so made a transition from qualitative state $|\varphi| > 90°$ to $|\varphi| \leq 90°$. We observed that by choosing action (v^-, ω^-), the planner was able to simultaneously utilize rules $\dot{D} = M^+(v)$ and $\dot{\varphi} = M^-(\omega)$, and therefore satisfy both goal directions, $D \to 0$ and $\varphi \to 0$. We find such a maneuver visually very intuitive from a human perspective. Being slightly faster than the target, the robot managed to stay in the $|\varphi| \leq 90°$ qualitative state until the end of the task, alternating between actions (v^+, ω^+) and (v^+, ω^-) to simultaneously shorten the distance and regulate its orientation towards the target.

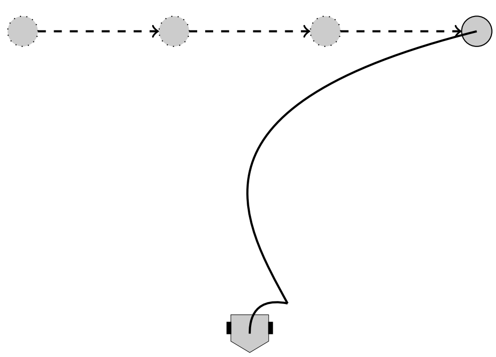


Figure 4: Trajectory made by the robot while pursuing a moving target.

It is possible to introduce more than one target and let the planner choose the closest one. We set up an additional scenario with one stationary and one moving target that we were able to move interactively. The behavior of the robot was as expected. Initially, the robot went for the stationary target, which was being positioned closer to it. Before hitting it, we moved the secondary target closer and distracted the robot away from its primary target. As soon as we moved our target out of reach, the robot headed back to its primary goal. To achieve such behavior, we had to separately describe dynamics of both targets, as shown in Listing 2, although both descriptions are identical.

Listing 2 Pursuing two objects

```
# Input values:
# x0, y0, theta - robot configuration
# x1, y1        - target 1 position
# x2, y2        - target 2 position

phi1 in S[-180, 180); # phi1 is circular
phi2 in S[-180, 180); # phi2 is circular

# Egocentric values of target 1
D1 = sqrt((x1 - x0)^2 + (y1 - y0)^2);
phi1 = atan2(y1 - y0, x1 - x0) - theta;

# Egocentric values of target 2
D2 = sqrt((x2 - x0)^2 + (y2 - y0)^2);
phi2 = atan2(y2 - y0, x2 - x0) - theta;

# Qualitative model 1
abs(phi1) <= 90; => D1' = M-(v);
abs(phi1) > 90;  => D1' = M+(v);
phi1' = M-(w);

# Qualitative model 2
abs(phi2) <= 90; => D2' = M-(v);
abs(phi2) > 90;  => D2' = M+(v);
phi2' = M-(w);

action(v, w);
goal(D1 -> 0, phi1 -> 0);
goal(D2 -> 0, phi2 -> 0);
```

Avoiding obstacles

The general idea to implementing obstacle avoidance is to introduce additional constraint to the pursuing scenario, making the configuration space non-convex. The tendency to avoid an obstacle is therefore the tendency to avoid violating constraint $D > 0$, where D is the distance from the border of the obstacle. This represents an implicit goal and we found it very efficient to construct the qualitative model of avoidance as shown in Fig. 5. We identify three qualitative states:

- *The obstacle is front left* ($0 \leq \varphi < 90$). The robot can increase $\dot{D}$ by decreasing v and increasing angular speed towards CW.
- *The obstacle is front right* ($-90 < \varphi < 0$). The robot can increase $\dot{D}$ by decreasing v and increasing angular speed towards CCW.
- *The obstacle is behind* ($\varphi \geq 90 \vee \varphi \leq -90$). The robot can increase $\dot{D}$ by increasing v.

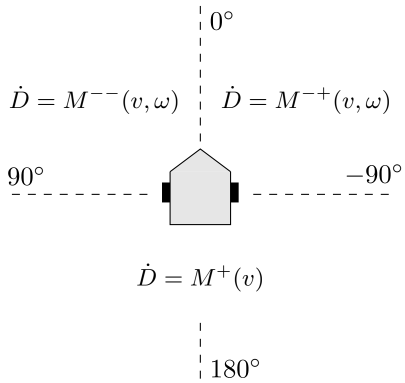


Figure 5: The qualitative model of avoidance with three qualitative states.

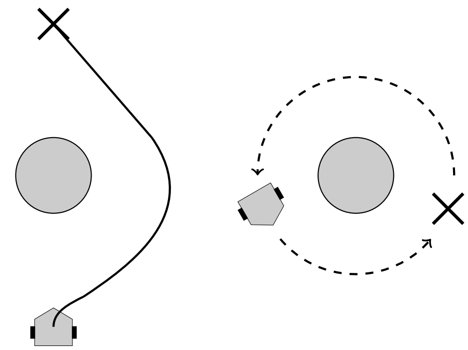


Figure 6: Trajectory made by the robot avoiding an obstacle while pursuing a stationary target (left) and chasing a moving target around the obstacle (right).

Listing 3 Avoiding an obstacle

```
# Input values:
# x0, y0, theta - robot configuration
# x1, y1        - goal position
# x2, y2        - obstacle position
# r             - obstacle radius

phi1 in S[-180, 180); # phi1 is circular
phi2 in S[-180, 180); # phi2 is circular
D2 > 0; # obstacle constraint

# Egocentric values of goal
D1 = sqrt((x1 - x0)^2 + (y1 - y0)^2);
phi1 = atan2(y1 - y0, x1 - x0) - theta;

# Egocentric values of obstacle
D2 = sqrt((x2 - x0)^2 + (y2 - y0)^2) - r;
phi2 = atan2(y2 - y0, x2 - x0) - theta;

# Qualitative model to pursue
abs(phi1) <= 90; => D1' = M-(v);
abs(phi1) > 90;  => D1' = M+(v);
phi1' = M+-(w);

# Qualitative model to avoid
phi2 in I[0, 90);      => D2' = M--(v, w);
phi2 in I(-90, 0);     => D2' = M-+(v, w);
phi2 in I[90, 180) or
phi2 in I[-180, -90]; => D2' = M+(v);
phi2' = M-(w);

action(v, w);
goal(D1 -> 0, phi1 -> 0);
```

Domain description shown in Listing 3 is very similar to description of the two-target pursuing domain, replacing the second qualitative model of pursuing with the model of avoidance, and the second explicit goal with constraint `D2 > 0`. Trajectories made by two different scenarios are shown in Fig. 6. The first setting involved a stationary target, placed straight ahead of the robot but behind an obstacle. Because $\varphi = 0°$ falls into the first qualitative state, the robot chose to avoid the obstacle by its right side. The second setting involved a moving target, circling with approximately the same speed as the robot, about one third of the circle in front. The robot made a circular trajectory trying to catch the target.

Pushing objects

Using our new planner and domain description language we were able to reproduce experiments described in (Šoberl, Žabkar, and Bratko 2015). This way we showed that this planner is at least as powerful, but more universal than our previous planning methods. We used the same qualitative model of pushing, which was learned by autonomous robotic experimentation. Domain attributes (depicted in Fig. 7) are the following: position of the object (x, y), orientation of the object β, goal position (x_g, y_g), goal orientation γ, orientation of the robot θ, the point of contact $\tau \in [-1, 1]$ and the angle of pushing $\varphi \in [-30°, 30°]$. The egocentric approach is used, making the above values relative to the robot's position and orientation.

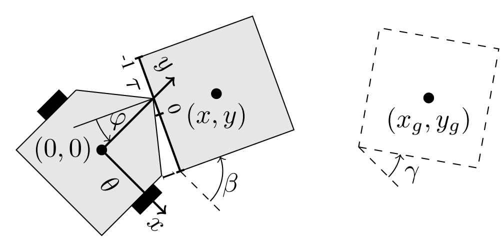


Figure 7: Attributes of the pushing domain.

Qualitative relations derived in the original work are the following:

$$
\begin{aligned}
\dot{y} &= M^{+}(v) & \dot{\varphi} &= M^{+}(\omega) \\
\dot{x} &= M^{--}(\tau, \omega) & \dot{\tau} &= M^{--}(\omega, \varphi) \\
\dot{\beta} &= M^{+--}(\tau, \varphi, \omega) & \dot{\theta} &= M^{+}(\omega)
\end{aligned}
\tag{6}
$$

An example trajectory is shown in Fig. 8. A rectangular box is placed 1×1 meter from its goal location and rotated by $180°$. The goal is to match the position and orientation of the box with that of the goal. It can be seen from the form of the trajectory that the robot is trying to satisfy both goal conditions simultaneously, until none of the possible actions works in favor of the goal directions. We make the planner discard *non-positive* actions by adding the constraint:

```
score = action(v, w);
score > 0;
```

We can see the trajectory being composed of 4 smooth curves. Their joints are the points where no action was scored above 0, and therefore the *no solution* message was received, meaning that the robot had to reposition in order to continue towards the goal. We used a separate solution to reposition the robot to an exact initial position, however, our planner did make the decision about which initial position is best. Whenever such a decision had to be made, we computed attribute values for all possible initial states and sent each initial state to the planner. For each setting we then obtained the best possible action together with its score. The state permitting the highest evaluated action was then selected as the next initial state.

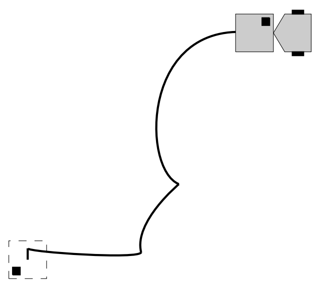


Figure 8: Trajectory made by a rectangular box being pushed to a goal configuration.

Conclusion

We have shown that qualitative models in the form of qualitative monotonic constraints contain enough information to allow simple motion planning without the need for additional numerical learning. We introduced a new qualitative planning method that can handle basic motion planning problems and proposed a new domain description language that allows concise non-algorithmic description of robotic domains using qualitative relations and additional numerical constraints. We demonstrated the intuitiveness of qualitative modeling in robotic planning and its ability to produce desired results without the need of doing any precise numerical measurements or modeling to describe the domain. We believe this way the robot exhibits similar behavior to a living being making fast instinctive decisions as it moves through an unfamiliar terrain in pursuit of some goal. However, at this point of research, our planning method is still somewhat shortsighted and unable to learn from its past mistakes or successes. In a few occasions we managed to bring the robot to a dead loop, alternating between two qualitative states as it was trying to reach two equally distant goals, oblivious of its past states and decisions. Only the fact of slight randomness due to certain sensory noise and communication delays eventually brought the robot out of a self-made trap. However, such situations were rare and we had to make extra effort to invoke them.

The final form of the language is still under our consideration and we shortly plan to add some extra elements. In our first experiment we had to duplicate the model of pursuing to introduce the second target. This problem could be tackled by introducing vector-like structures, combining attributes of the same type that belong to different objects, e.g. `D = [D1, D2, ...]` to combine distances to multiple objects. A single qualitative or numerical constraint would then hold for all attributes within that vector, e.g. stating only `D = M-(v)` instead of a separate statement for each object.

References

Bratko, I., and Šuc, D. 2003. Learning qualitative models. *AI magazine* 24(4):107–119.

Bratko, I. 2011. Autonomous discovery of abstract concepts by a robot. In *Proc. ICANNGA*, volume 1, 1–11. Springer Lecture Notes.

Ferguson, D.; Kalra, N.; and Stentz, A. 2006. Replanning with RRTs. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, 1243–1248.

Fraichard, T.; Hassoun, M.; and Laugier, C. 1991. Reactive motion planning in a dynamic world. In *Advanced Robotics, 1991. 'Robots in Unstructured Environments', 91 ICAR., Fifth International Conference on*, 1029–1032 vol.2.

Kavraki, L.; Svestka, P.; Latombe, J.-C.; and Overmars, M. 1996. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *Robotics and Automation, IEEE Transactions on* 12(4):566–580.

LaValle, S. M. 2006. *Planning Algorithms*. Cambridge: Cambridge University Press.

Leven, P., and Hutchinson, S. 2002. A Framework for Real-time Path Planning in Changing Environments. *The International Journal of Robotics Research* 21(12):999–1030.

Šoberl, D.; Žabkar, J.; and Bratko, I. 2015. Qualitative planning of object pushing by a robot. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9384, 410–419. Springer Verlag.

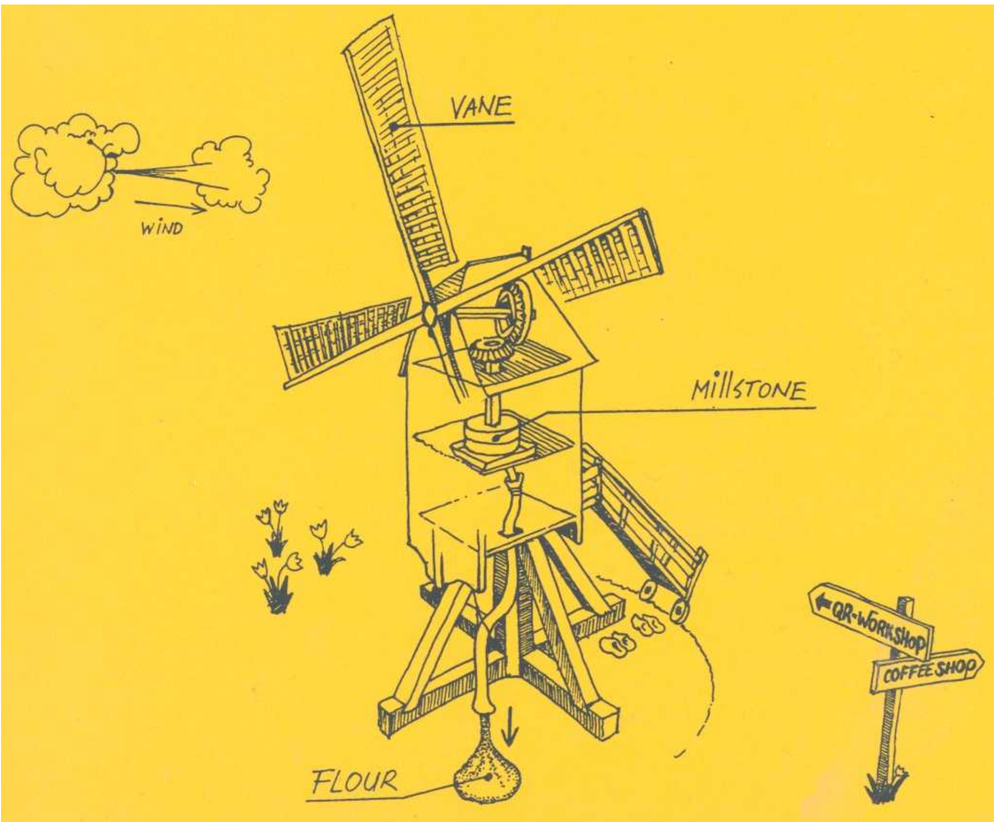
Troha, M., and Bratko, I. 2011. Qualitative Learning of Object Pushing by a Robot. In *25th International Workshop on Qualitative Reasoning*, 175–180.

Wiley, T.; Sammut, C.; and Bratko, I. 2014. Qualitative Planning with Quantitative Constraints for Online Learning of Robotic Behaviours. In *28th AAAI Conference on Artificial Intelligence*, 2578–2584.

Žabkar, J.; Bratko, I.; and Demšar, J. 2007. Learning qualitative models through partial derivatives by Padé. In *Proceedings of the 21st Annual Workshop on Qualitative Reasoning*, 193–202.

Author Index

Keyword Index


VANE
WIND
MILLSTONE
FLOUR
QR-WORKSHOP
COFFEE SHOP