Category Specific Post Popularity Prediction

Masoud Mazloom, Iliana Papi, and Marcel Worring

Informatics Institute, University of Amsterdam, Netherlands
m.mazloom@uva.nl,iliana.n.pappi@gmail.com,m.worring@uva.nl

Abstract. Social media have become dominant in everyday life during the last few years where users share their thoughts and experiences about their enjoyable events in posts. Most of these posts are related to different categories related to: *activities*, such as dancing, *landscapes*, such as beach, *people*, such as selfie, and *animals* such as pets. While some of these posts become popular and get more attention, others are completely ignored. In order to address the desire of users to make popular posts, several researches have studied post popularity prediction. Existing works focus on predicting the popularity without considering the category type of the post. In this paper we propose category specific post popularity prediction using visual and textual content for *action, scene, people* and *animal* categories. In this way we aim to answer the question *What makes a post belonging to specific action, scene, people or animal category popular?*. To answer to this question we perform several experiments on a collection of 65K posts crawled from Instagram related to four categories: *action, scene, people*, and *animal*.

1 Introduction

A huge amount of visual and textual information is posted everyday in social media such as Twitter, Instagram, Flickr, Facebook. It takes only a few seconds to share social activity in the form of an image, video, comments, and tags in any place at any time with a simple internet connection to a device such as a mobile or tablet. As a result, every day more and more users are generating and sharing multimodal contents in social media. However the destiny of the user generated posts in social media are completely different. While some posts receive a high number of likes and gain a lot of attention, others are more or less ignored.

Predicting how popular a post will be among other users in the user's network or in public, has become interesting for marketing and business [16], political and economic sciences [12] and decision-making strategies of campaigns targeting on social media crowds [11]. Moreover, predicting post popularity is important for the self-evolution of the social media [8]. Every user would like to know the best way to interact or get noticed in a social media platform, concerning both shared posts and quotes or comments. Defining what makes a user generated post become popular has been proven to be a challenging problem to solve [9].

Many approaches have been proposed to predict the popularity of post focusing on the effect of visual low- and high-level contents [4, 9, 14, 16, 6, 15, 13], textual contents such as tweets, user's tags and comments [18, 1, 7], and visual contents along with the

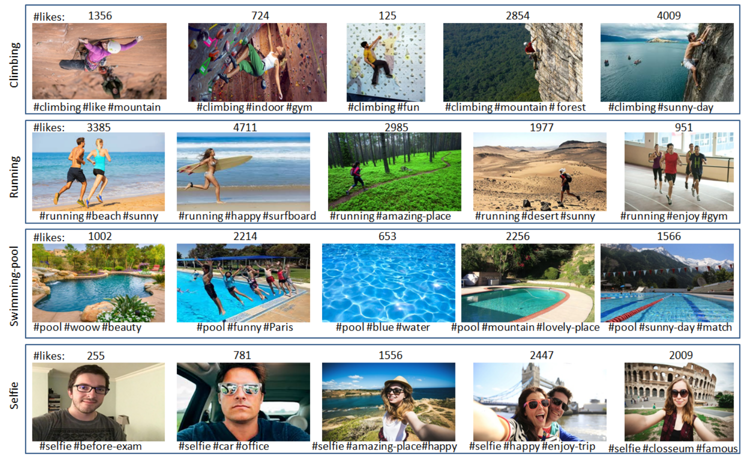

Fig. 1. Each row shows five posts related to the actions climbing and running, swimming-pool scene, and selfie respectively. We aim in this paper to investigate whether there are any visual semantic features affecting the popularity of posts belonging to a specific category.

textual contents of a user's post [9, 14, 16, 6, 15]. Inspired by the success of recent post popularity prediction in social media [4, 9, 14, 16, 6, 15, 18, 1, 7, 13] we continue the study by predicting the popularity of category-based posts. The above methods are successful using a general method which is the same for every category. We analyze the content of posts related to different categories such as *actions, scenes, people*, and *animal* for the study, following the analysis of brand-related contents for brand popularity prediction in [16]. We study how human action present in the image, scenery or background, the presence of people, or animals affect the popularity of a post.

For the purpose of this paper a new dataset was created from scratch, containing posts crawled from Instagram, a broadly used social network with emphasis on visual and textual contents. Among the different social media platforms, Instagram has a strong emphasis on self-expression by images with a description through captions and hashtags and is easy to crawl. In contrast, Twitter does not always have image content in a post, Facebook is complicated to crawl and has a lot of privacy rules and Flickr is less connected with social activity and more connected to photographers' communities. Since we aim to predict the popularity of post related to what people mostly enjoy, we crawled posts in terms of enjoyable activities, places, selfie, and pets. Figure 1 show some examples of our dataset.

In this paper we propose a multimodal framework for post popularity prediction especially when action, scene, people, and animal appear in the users' posts. We investigate which semantic features more affected the popularity of a post in social media. Especially we try to study the role of low and high-level visual features, along with textual features with specific characteristics related to e.g. action and scenery, in correlation with post popularity.

We make three main contributions in this paper: 1) We study the problem of post popularity prediction inside various categories in social media, 2) We investigate the correlation of semantic features with popularity prediction of posts for different categories which allows us to propose meaningful suggestion to a user, 3) We introduce a new dataset, for category-based post popularity prediction, obtained for free from Instagram by a simple crawling procedure.

We organize the remainder of this paper as follows. We start by considering related work in Section 2. Section 3 describes our proposal for predicting the popularity of posts. We introduce the experimental setup on our dataset in Section 4. Results and conclusions are presented in Section 5 and 6 respectively.

2 Related Work

Several studies in the literature addressing the post popularity prediction based on the textual or visual features, or a combination of them.

A study for predicting popularity in Twitter by Hong et al. [7], formulates the task as a classification problem, investigating a wide spectrum of features based on the content of the messages. Bae et al. [1] analyze Twitter posts and categorized the followers of a limited number of influential users to a positive and a negative audience. From there, they correlate the sentiment of the followers with the textual content of their posts and based on that defined a measure of influence. Szabo et al. in [18] investigate the popularity of videos in YouTube by analyzing the social cues, comments, and associated tags. All of these works use textual content for popularity prediction. However, visual content which also holds a lot of information, is not addressed in these methods.

Visual contents of posts are investigated for their correlation with popularity prediction in [4, 9, 14, 6, 15, 16]. Cappallo et al. in [4] developed a model for popularity prediction in social media based only on visual content. A latent ranking approach was proposed, which takes into account not only the distinctive visual cues in popular images, but also those in unpopular images. Khosla et al. in [9] report the importance of image cues such as color, gradients, low-level features and the set of objects present, as well as the importance of various social cues such as number of followers or number of photos uploaded by the user. Image popularity prediction in a "cold start" scenario, where there exists no or limited textual interaction data, by considering image context, visual appearance and user context was investigated by McParlane et al. [15]. The authors cast the problem as a classification task between highly popular and unpopular images. Mazloom et al. in [14] present an approach for identifying what aspects of posts determine their popularity. The proposed model was based on the hypothesis that brand-related posts may be popular due to several cues related to factual information, sentiment, vividness and brand engagement parameters. Gelli et al. in [6] investigate the effect of visual sentiment analysis and context features on image popularity in social media. Overgoor et al. in [16] investigate brand, as a category, popularity prediction in a spatio-temporal category representation framework. The results of this work confirm complementary of visual and textual features for predicting the popularity of a brand.

Different from [4, 15, 14, 9, 6] which perform popularity prediction of posts in a general setting, we aim to take into account the category type of a post in predicting it's popularity. Inspired by the success of brand popularity prediction in [16], we predict the popularity of posts related to different categories such as action, scene, people and animal.

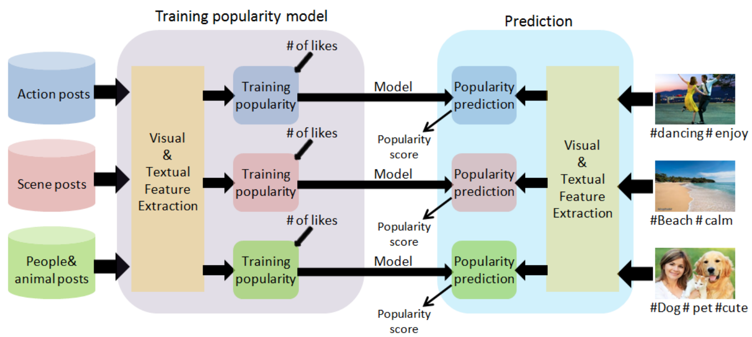

Fig. 2. Our proposal for computing popularity score for category specific posts.

3 Our Proposal

Our category-based popularity prediction framework consists of two main components, schematically illustrated in Figure 2. In the training phase, at first, we represent each category-based post using its visual and textual contents. We use the number of likes of each post received in social media as label and train a popularity model. Second, in the prediction phase, we represent an unseen post and compute a popularity score. Before introducing the proposed framework for popularity prediction, the notation and key concepts will be formally introduced.

3.1 Problem formalization

Given a specific category-based post, popularity prediction inside a category is the task of computing a score per post that shows how popular it will be in comparison to other posts in the same category. For consistency, we use P_{z_x} to indicate the given x^{th} post of category z. We aim to construct a real-valued function $f(P_{z_x})$ which produces a score for the popularity of P_x in category z. By sorting all posts of category z in the test set according to $f(:)$ in descending order, a list of most popular posts in category z will be obtained.

Suppose $C = \{C_1, C_2, ..., Cn\}$ is a dataset consist of posts related to all n categories, where each C_i is given by m_i posts, $C_i = \{(P_{i_1}, y_{i_1}), (P_{i_2}, y_{i_2})), ..., (P_{i_{m_i}}, y_{i_{m_i}})\}$ where P_{i_j} is j^{th} post of category C_i, carrying multimodal information, and y_{i_j} is its corresponding number of likes. We hypothesize that what makes a category-based post become popular on social media, number of likes a post will receive, depends on the content of generated post. It means the output of $f(P_{z_x})$ largely depends on the representation of P_x. In section 3.2 we show how to represent a post based on it's content.

3.2 Post representation

In this section we explain how a category-based post can be represented using its visual and textual contents. Suppose $P_{i_j} = (P^v_{i_j}, P^t_{i_j})$ consists of the two components visual and textual content. To find different representations of post P_{i_j} , we extract visual features, $F(P^v_{i_j})$ and textual features, $F(P^t_{i_j})$. The output of F is a representation of P_{i_j}. Similar to [14], we use both visual and textual contents in the representation of a post as they carry complementary information related to the popularity of a post. Depending on how to extract these contents, we present several representations of P_{i_j} here.

We present each P_{i_j} using it's visual channel, $P^v_{i_j}$, by three state-of-the-art features used in [14]:

- *Concepts Features*: To extract Concept features we utilized a deep neural network namely the GoogleNet Inception V3 [19]. The output of $F(P^v_{i_j})$ is a 1000-dimensional feature vector of the softmax output layer.
- *Low-level Features:* The output of $F(P^v_{i_j})$ is the 2048-dimensional low-level feature vector for each image from the Max Pooling of the Convolutional Pool 8x8 layer of the same network.
- *Visual Sentiment Features:* The sentiment in the visual content of each post was expressed through 1200-dimensional vectors, utilizing SentiBank detectors [5] of the Visual Sentiment Ontology (VSO) [2]. This bank consist of 1,200 Adjective-Noun pairs (ANP's) detectors such as "beautiful flowers". The output of $F(P^v_{i_j})$ in this case is a 1,200 dimensional vector with probabilities of ANP's being present in the image.

We also present each P_{i_j} using it's textual channel, $P^t_{i_j}$, by three state-of-the-art features used in [14]:

- *Word-to-Vec Features:* The Word-to-vec (W2V) model [17] leads to vector representations of words learned using word embeddings, where each word is represented by a 300-dimensional vector. The words for each post are represented by a W2V vector and the final representation of post, $F(P^t_{i_j})$, is obtained by average pooling over all word vectors.
- *Bag-of-Words Features:* Bag-of-Words (BOW) is a sparse representation of the counts of each word in the post, compared to a pre-constructed vocabulary, i.e., a sorted list of all the unique words in the dataset based on their frequency. We find a representation, $F(P^t_{i_j})$, with 1000-dimension as the best representation by cross validation over the different dimension sizes.
- *Textual Sentiment Features:* The output of $F(P^t_{i_j})$ is a 2-dimensional feature vector which represents the positive, ranging from 1 to 5, and negative, ranging from -1 to -5, sentiment score of the textual content in the post P_{i_j}. We use SentiStrength [20] for this task.

Next, we show in section 3.3 how to tackle category-based post popularity prediction using different representations.

3.3 Popularity prediction

After representing posts, we use it for the problem of post popularity prediction. Suppose C is a set of m posts, $m=\sum_{i=1}^{n} m_i$, where $C=\{(P_{i_1},y_{i_1}),...,(P_{i_{m_i}},y_{i_{m_i}})$ $,...,(P_{n_{m_n}},y_{n_{m_n}})\}$. We divide C into two parts, with $C = C_{tr} \cup C_{te}$ where C_{tr} is a training set consisting of k posts and C_{te} is a test set consisting of the other

Table 1. Statistics of our dataset used in experiments

Category							
Action		Scene		People		Animal	
Name	Post	Name	Post	Name	Post	Name	Post
#basketball	2055	#art-gallery	1499	selfie	4500	pet	3200
#climbing	2008	#bar	2039				
#cycling	2786	#beach	2264				
#dance	2473	#bedroom	2240				
#football	2549	#cafe	2568				
#horse-riding	2258	#canals	3382				
#hug	2290	#fields	2264				
#kiss	2909	#forest	3289				
#playing-music	1850	#home	2306				
#running	2139	#kitchen	2620				
#ski	2267	#street	2408				
#surfing	1622	#swimming-pool	2915				
	27206		29794		4500		3200

$m-k$ posts in C. By representing the posts in a training and test set using features explained in section 3.2, we define C_{tr} and C_{te} as two matrix representations of posts, $C_{tr}=[F(P_1),...,F(P_k)]$ and $C_{te}=[F(P_{k+1}),...,F(P_m)]$ where each row of C_{tr} and Cte represents a post.

We consider the popularity prediction of a post as a regression problem as considered in [9, 14, 16, 6]. Let $F(P_i)$ be a representation of post P_i from C_{te} and y_i the popularity of P_i in social media. The goal is to learn function $f(:)$ over C_{tr} to estimate the popularity of P_i, $\tilde{y}_i=f(F(P_i))=w^T F(P_i)$, where $|y_i-\tilde{y}_i|$ as an error is small. The idea is to optimize w, parameter of function $f_w()$, on C_{tr} to minimize the error. To solve the problem and find the optimal value of w we use different regressors such as L2 regularized L2 loss Support Vector Regression as used in [9, 14], Support Vector Regression using RBF kernel, Random Forest, and Multilayer Perceptron. After training the model and finding the optimum value of w on C_{tr}, we use it for prediction of post popularity on C_{te} and report the rank correlation between the predicted scores and ground truth.

4 Experimental Setup

4.1 Dataset

Since there is no existing dataset for predicting the popularity of a category-based posts, we created one by crawling Instagram. Our dataset consists of approximately 65k Instagram posts with visual and textual content as well as metadata related to *action, scene, people* and *animal* categories. All the posts in our dataset were generated between 1/1/2017 to 25/4/2017. The crawling of the posts was made according to hashtags of several enjoyable event related to the categories. The relevant hashtags used for the Instagram crawler are: for **actions**: we select those actions related to love, music, and sport such as *#playing-music, #running, #basketball, #surfing, #ski, #climbing, #cycling, #dance, #football, #horse-riding, #hug, #kiss*. For **places/scenes**: we consider those sceneries related to indoor, outdoor, nature, and hobby such as *#art-gallery, #bar, #beach, #bedroom, #cafe, #canals, #fields, #forest, #home, #kitchen,*

#street, #swimming-pool, #urban. For **people** we consider *#selfie*and *#pets*for **animal**. The statistics of our dataset are reported in Table 1.[1].

In order to explore the popularity of posts in all categories or in a specific category, we define two different settings: 1) **Category-mix** where we use the whole dataset, C, to build a general model for popularity prediction. 2) **Category-specific** where we used all data related to specific category C_i. We perform the training and evaluation independently for each category. In both settings we split the data randomly into training and test set. We train a model over 70% of the dataset as training set and report the popularity result over the other 30% of dataset as test set.

4.2 Implementation details

Popularity measurement Similar to [9, 14] we consider popularity prediction of a post as a ranking problem. We use the number of likes a post received in social media as the measure of its popularity. We find the majority of posts receive little likes and the minority of them receive a high number of likes. To deal with it we follow [9, 14] and consider the log number of likes.

We used different regressor methods for predicting the popularity of a post. We consider 5-fold cross-validation on training set for tuning all parameters of regressors. We find the optimal value of the regularization parameter $\lambda = 0.1$, in SVR, through $\lambda \in \{0.001, 0.1, 1, 10, 100, 1000\}$. We find Random Forrest Regressor (RFR) [3] with 100 tree estimators, as an optimal parameter giving the best results among the values $\{10, 100, 300, 1000\}$. We consider the default setting in results [10] for using Multi-layer Perceptron (MLP) as a regressor. The optimum value of $\alpha = 0.01$, the loss parameter, was tuned for $\alpha \in \{0.0001, 0.001, 0.01, 1, 10\}$.

Evaluation metric In this paper we evaluate the post popularity prediction model using Spearman's rank correlation coefficient, a statistical metric showing the monotonic relation between two vectors as used in [9, 14]. We compute the rank correlation between the prediction vector resulting from the model and ground truth vector which returns a value between [-1, 1]. A value close to 1 corresponds to perfect correlation.

4.3 Experiments

Experiment 1: Post popularity prediction in Category-mix dataset In this experiment we evaluate the effect of different visual and textual features explained in section 3.2 on predicting the popularity of a post in Category-mix datasets. To find an efficient popularity model we report the result of different regressors explained in 3.3, Linear-SVR (LSVR), RBF-SVR(RSVR), Random Forest Regressor(RFR), and MLP Regressor (MLPR). We use the best regressor for the other experiments. We also report the effect of combining the best visual and textual feature, based on their rank correlation results, by average pooling in a late fusion scenario.

Experiment 2: Post popularity prediction in Category-specific We report the result of post popularity prediction on Category-specific dataset in this

[1] http://isis-data.science.uva.nl/Masoud/MMM17Data.

Table 2. Experiment 1: Popularity Prediction on Category-mix dataset

	Visual features			
Model	Concepts	Low-Level	Visual sentiment	AvgPool
RSVR	0.221	0.201	0.152	0.234
LSVR	0.229	0.211	0.196	**0.253**
RFR	0.232	0.221	0.202	**0.260**
MLPR	0.183	0.237	0.192	0.230
	Textual features			
	Word2Vec	BoW	Textual sentiment	AvgPool
RSVR	0.328	0.415	0.102	0.374
LSVR	0.339	0.402	0.104	**0.390**
RFR	0.350	0.428	0.085	**0.395**
MLPR	0.409	0.320	0.099	0.355

experiment. We evaluate the effect of various visual and textual features on the popularity of category-based posts. We compare the result of applying a model which is trained on Category-specific data with the model trained on train set of Category-mix and report the result on the test set of each category. We report the result of fusing best visual and best textual features.

We also report the result of popularity prediction on each specific instance inside categories, such as popularity of dancing in action category. We evaluate the correlation of semantic visual features, Concepts and Visual sentiments, with popularity of post in different instances in categories. To do that we compute the weights of regressor models separately for each category and each semantic feature. We sort them for selecting the top semantic features which have high impact on popularity of post per category.

Experiment 3: Popularity prediction using specific visual concepts: In this experiment we evaluate the effect of specific visual Concepts on the popularity of a specific category. We use the *Concepts features* explained in 3.2 in this experiment. We manually labelled the 1000-Imagenet Concepts for *action, scene, people, animals*, and *general objects*. The outcome was a set of 50-dimensional feature vectors for action, 151-dimensional vectors for scene, 10-dimensional vectors for people, 404-dimensional vectors for animals, and 525-dimensional vectors for general objects within the Imagenet 1000 concepts. Many of the concepts had to be double-labelled, and as a result participating in two categories, as a strict division of concepts was not always possible in terms of semantic meaning. We report the popularity of posts using specific concept category for each specific category.

5 Results

5.1 Post popularity prediction in Category-mix dataset

We report the result of post popularity prediction in category-mix dataset in Table 2. Starting with the visual features and LSVR as a regressor method, the results show the 0.229, 0.211 and 0.196 rank correlation as a popularity of a post using Concepts, Low-level, and Visual sentiment features respectively. Using an average pooling for fusing these visual features, the result reaches 0.253. The results in Table 2 also show the importance of using RFR as a regressor, instead of using the other regressors, for predicting the popularity of a post using visual features where the result reach 0.232, 0.221, 0.202 and 0.260 rank correlation using Concepts, Low-level, Visual sentiment

Table 3. Experiment 2: Popularity on Category-specific dataset using visual features.

	Action			
training data	Concepts	Low-Level	Visual sentiment	AvgPool
Category-mix	0.186	0.241	0.144	0.285
Category-specific	0.286	0.317	0.211	0.345
	Scene			
	Concepts	Low-Level	Visual sentiment	AvgPool
Category-mix	0.151	0.201	0.112	0.221
Category-specific	0.221	0.227	0.153	0.250
	People			
	Concepts	Low-Level	Visual sentiment	AvgPool
Category-mix	0.168	0.191	0.152	0.224
Category-specific	0.187	0.223	0.198	0.244
	Animal			
	Concepts	Low-Level	Visual sentiment	AvgPool
Category-mix	0.221	0.201	0.162	0.234
Category-specific	0.165	0.216	0.224	0.247

and fusion of them respectively. The results show the advantage of using LSVR and RFR for training a popularity model over visual features in comparison with the other two regressors, RSVR and MLPR.

The result of popularity prediction using textual features in Table 2 also depict to superiority of using LSVR and RFR for training a popularity model. However the result of using RFR, over both visual and textual features, is slightly better than using LSVR, but it suffers from the time efficiency for training a model against LSVR. We keep LSVR regressor method and use it for the other experiments. By fusing the results of the best visual and textual features using LSVR model, which are Concepts and BoW respectively, with an average operator the rank correlation reaches 0.434 which shows the complementary of visual and textual features for popularity prediction.

5.2 Post popularity prediction in Category-specific dataset

We report the result of post popularity prediction using visual features in the category-specific dataset in Table 3. The result of predicting the popularity of post related to action category reaches 0.286, 0.317, 0.211, and 0.345 using Concepts, Low-level, Visual sentiment and fusion of them respectively where the model is trained on action specific data. However using a model trained on category-mix the result reaches to 0.186, 0.241, 0.144, and 0.285 by Concepts, Low-level, Visual sentiment and fusion of them. The result of fusing features shows 38% relative improvement in popularity of action posts where we use a model trained on specific action data against a model using all data. The result in Table 3 also depict to 13%, 9%, and 6% relative improvement in popularity of scene, people, and animal category respectively using specific data for training a popularity model against using all data.

We report the effect of training a popularity model over specific data, using different textual features, versus training a model using all data. It shows 17%, 29%, 8%, and 5% relative improvement in predicting the popularity of post related to action, scene, people, and animal category. The result of rank correlation by combining the best visual and textual features per category reaches to 0.488, 0.481, 0.245, and 0.248 for action, scene, people, and animal category respectively.

Table 4. Experiment 2: Popularity Prediction for each instance of categories.

		Visual features		
Category	instance	Concepts	Low-level	Visual sentiment
Action	Basketball	0.171	0.226	0.157
	Climbing	0.088	0.158	0.201
	Cycling	0.182	0.175	0.095
	Dancing	**0.400**	**0.321**	**0.255**
	Playing-football	0.120	0.074	0.095
	Horse-riding	0.100	0.106	0.083
	Hugging	0.135	0.200	0.076
	kissing	0.086	0.157	0.065
	Playing-music	0.087	0.100	0.108
	Running	0.057	0.124	0.045
	Skiing	0.119	0.114	0.107
	Surfing	0.070	0.067	0.035
Scene	Art-gallery	0.100	0.170	0.287
	Bar	0.102	0.112	0.114
	Beach	0.064	0.107	0.158
	Bedroom	0.100	0.137	0.149
	Cafe	0.093	0.087	0.035
	Canals	0.088	0.068	0.094
	Fields	0.081	0.103	0.087
	Forest	0.100	0.125	0.154
	Home	0.046	0.068	0.099
	Kitchen	0.069	0.032	0.055
	Street	0.035	0.1092	0.112
	Swimming-pool	**0.495**	**0.598**	**0.309**
People	Selfie	0.187	0.223	0.198
Animal	Pet	0.165	0.216	0.224

We show the result of popularity prediction per instance inside each category using visual features in Table 4. As we can see among all actions the best results are for the action *Dancing*, where the result reaches 0.400, 0.321, and 0.255 rank correlation using Concepts, Low-level and visual sentiment respectively. In the scene category, *Swimming-pool* has the highest rank correlation scores using all visual features among the others. The results are pretty good for these two instances showing the effect on popularity of pleasant actions and places for everyday life hidden in visual content.

We observe from Table 3, and 4 that the presence of visual sentiment for most of categories and instances has a positive effect on the predicted popularity. We highlight the important of sentiments with different impact on popularity for *Dancing, Swimming-pool, Selfie*, and *Pet* in Figure 3.

The results of experiment 2 confirm the effect of visual and textual features, also the combination of them, for predicting the popularity of post related to different categories. Moreover, in general, it emphasizes the accuracy of popularity prediction models trained on category-specific data. At the end it shows that there are some specific visual semantic features which make sense for popularity of posts in different categories.

5.3 Popularity prediction using specific visual concepts

We report the result of experiment 3 in Figure 4. It shows that for all categories, except for scene, all 1000 Concepts have more descriptive power of popularity. In this

Fig. 3. Experiment 2: Visual sentiments with high impact on post popularity for (a) Dancing, (b) Swimming-pool, (c) Selfie and (d) Pet. Font size correlates with the impact of visual sentiments with the popularity of instances.

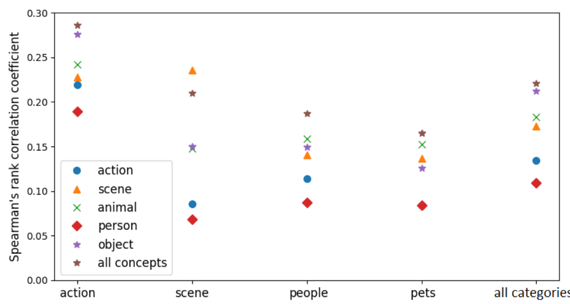

Fig. 4. Experiment 3: Correlation between specific concepts and categories

case, the scene category is observed to have the highest rank correlation with the specific concepts related to scene concepts, forming 151-dimensional concept feature vectors of 1000 concepts. The reported correlation is 0.255 versus all concepts with 0.221. The rank correlation result reaches 0.286, 0.271, 0.241, 0.232, 0.215, and 0.186 using all concepts, objects, animal, scene, action, and person concepts for action category. There are just 50 concepts out of 1000 concepts related to action. The results indicate that the low dimensional visual concepts are less effective for predicting the popularity of actions. Another reason is that the most of specific action concepts are not relevant with the actions in our dataset.

The result of experiment 3 confirm that higher-dimensional concept vector are contributing more to popularity prediction, since there are as many descriptors as possible for a subset category of posts.

6 Conclusion

We study the problem of popularity prediction of user generated post based on visual and textual content of post. Different from existing work, which investigate the effect of different visual and textual features on popularity of a post, we consider to predict the popularity of post inside different specific categories such as action, scene, people, and animal. We study if is there any visual semantic concepts in category which has positive effect on the popularity of category. By performing three experiments on a dataset of 65K posts related to different categories crawled from Instagram we find that: 1) Visual and textual contents have different impact on the popularity prediction of posts. Combining these contents improves the result. 2) In general, training a popularity model on specific-category data increases the accuracy of popularity per category. and 3) Concepts related to scene and different objects, have a descriptive power with the highest correlation with popularity prediction in all categories. Human faces and animals are also important for popularity prediction, as the adjective-noun pairs results show.

Acknowledgments. This research is supported by the Amsterdam Academic Alliance Data Science (AAA-DS) Program Award to the UvA and VU Universities. Thanks to Dr Efstratios Gavves for his collaboration in this project.

References

1. Younggue Bae and Hongchul Lee. Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *J. Am. Soc. Inf. Sci. Technol.*, 63(12):2521–2535, 2012.
2. Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *MM*, 2013.
3. Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
4. Spencer Cappallo, Thomas Mensink, and Cees G.M. Snoek. Latent factors of visual popularity prediction. In *ICMR*, 2015.
5. Tao Chen, Damian Borth, Trevor Darrell, and Shih-Fu Chang. Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *CoRR*, 2014.
6. Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. Image popularity prediction in social media using sentiment and context features. In *MM*, 2015.
7. Liangjie Hong, Ovidiu Dan, and Brian D. Davison. Predicting popular messages in twitter. In *WWW*, 2011.
8. O. Iordache. *Self-Evolvable Systems: Machine Learning in Social Media*. Understanding Complex Systems. Springer Berlin Heidelberg, 2012.
9. Aditya Khosla, Atish Das Sarma, and Raffay Hamid. What makes an image popular? In *WWW*, 2014.
10. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, 2014.
11. Cheng Li, Yue Lu, Qiaozhu Mei, Dong Wang, and Sandeep Pandey. Click-through prediction for advertising in twitter timeline. In *KDD*, 2015.
12. Michael B. MacKuen. Political drama, economic conditions, and the dynamics of presidential popularity. *American Journal of Political Science*, 27(2):165–192, 1983.
13. Masoud Mazloom, Iliana Pappi, and Marcel Worring. Multimodal context-aware recommender for post popularity prediction in social media. In *MM*, 2017.
14. Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn van Dolen. Multimodal popularity prediction of brand-related social media posts. In *MM*, 2016.
15. Philip J. McParlane, Yashar Moshfeghi, and Joemon M. Jose. "nobody comes here anymore, it's too crowded"; predicting image popularity on flickr. In *ICMR*, 2014.
16. Gijs Overgoor, Masoud Mazloom, Marcel Worring, Robert Rietveld, and Willemijn van Dolen. A spatio-temporal category representation for brand popularity prediction. In *ICMR*, 2017.
17. Xin Rong. word2vec parameter learning explained. *CoRR*, 2014.
18. Gabor Szabo and Bernardo A. Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, 2010.
19. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, 2015.
20. Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *J. Am. Soc. Inf. Sci. Technol.*, 61(12):2544–2558, 2010.