Image Popularity Prediction in Social Media Using Sentiment and Context Features

Francesco Gelli, Tiberio Uricchio,
Marco Bertini, Alberto Del Bimbo
MICC, Università degli Studi di Firenze
Viale Morgagni 65 - 50134 Firenze, Italy
{$name.surname$}@unifi.it

Shih-Fu Chang
Columbia University
500 West 120th Street, New York, NY, USA
sfchang@ee.columbia.edu

ABSTRACT

Images in social networks share different destinies: some are going to become popular while others are going to be completely unnoticed. In this paper we propose to use visual sentiment features together with three novel context features to predict a concise popularity score of social images. Experiments on large scale datasets show the benefits of proposed features on the performance of image popularity prediction. Exploiting state-of-the-art sentiment features, we report a qualitative analysis of which sentiments seem to be related to good or poor popularity. To the best of our knowledge, this is the first work understanding specific visual sentiments that positively or negatively influence the eventual popularity of images.

Categories and Subject Descriptors

H.3.1 [**Information Storage And Retrieval**]: Content Analysis and Indexing

General Terms

Algorithms, Experimentation

Keywords

Image popularity; social networks; visual sentiment; affective computing

1. INTRODUCTION

In the last decade users of social networks such as Flickr and Facebook have uploaded tens of billions of photos, often adding accompanying metadata by tagging and by providing a short description. Users interact with each other by forming groups of shared interests, following the status streams of each other, and by commenting the photos that have been shared. Inevitably, in the huge quantity of available media, some of these images are going to become very popular, while others are going to be totally unnoticed and end up in oblivion. Often, media may be popular because it conveys sentiments or it has a rich meaning in the social context it is put. In fact, sentiments have been known to affect popularity of visual media since the widespread watch of television programs [8]. Also, it was recently found to be related to popularity in tweets [1]. Being able to predict the popularity of a media may have a profound impact on several essential applications such as content retrieval and annotation, but also in other fields such as advertising and content distribution [9].

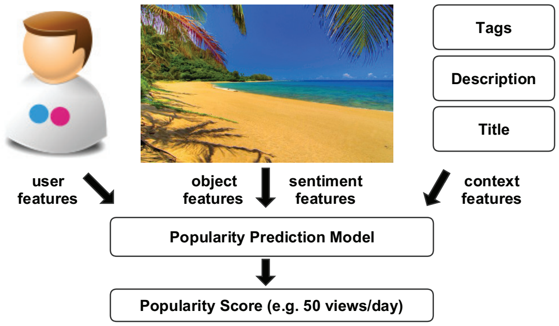


Figure 1: A schema of our approach to popularity prediction of images.

In this paper, we address the problem of predicting the popularity of an image posted in a social network, considering different scenarios that are typical of different situations. Despite the recent crop of literature that studies the question of what makes an image popular [12, 14, 16], none of these works addresses the question of how much the visual sentiment is influencing the popularity of media. As social context has been widely found important to predict media popularity [12], we show how to further improve popularity estimation by using a knowledge base to supplement the understanding of semantics in textual metadata.

The main contributions of this paper are:

- we propose to employ state-of-the-art visual sentiment features [2, 5] to perform image popularity prediction;
- we propose three new textual features based on a knowledge base, to better model the semantic description of an image, in addition to the social context features proposed in [12, 14];
- we show qualitative results of which sentiments seem to be related to a good or poor popularity.

To the best of our knowledge, this is the first work understanding specific visual concepts that positively or negatively influence the eventual popularity of images, beyond just numerical prediction of photo popularity.

Experiments performed on large scale datasets illustrate several benefits of the two types of proposed features, and

MM'15, October 26–30, 2015, Brisbane, Australia.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3459-4/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2733373.2806361 .

show how their combination impacts effectively on the performance of popularity prediction.

2. RELATED WORK

Popularity Prediction. Recently, a significant effort has been spent on investigating popularity of social media content. Regarding image popularity, the majority of works agree that social features have the greatest predictive power [12, 14, 16]. Visual content features are less powerful than social ones in terms of predictive power, but they are useful when no user metadata is present (e.g. no tags or description) or to address scenarios such as the case in which no social interactions have been recorded before posting the image (e.g. because the user has just joined the social network). Previous works vary in terms of popularity score definition (e.g. image views, reshares, mean views over a period) but they all share the same basic pipeline: they extract several content and context related features and successively employ a regressor to compute the popularity score.

In [12], Khosla *et al.* investigate both low-level features such as color, GIST, LBP, and content features such as the object predictions and network activations of a state-of-the-art CNN image classifier [13]. Together with user and image context features, they show promising results. McParlane *et al.* [14] propose to use image content, context features and user context to predict popularity. Their analysis is limited to a cold start scenario, i.e. where there exist no or little textual or interaction data. Totti *et al.* [16] investigate the use of aesthetics features such as blur, aspect ratio and color channel statistics together with the output of 85 object classifiers as content features.

Visual Sentiment. A few works have addressed the problem of multimedia sentiment analysis of social network images. Starting from the 24 basic emotions of Plutchik's Wheel of Emotions [15], Borth *et al.* [2] have recently presented a large-scale visual sentiment ontology termed SentiBank. They train 3,244 detectors on pairs of nouns and adjectives (ANPs) based on a combination of global and local features. Based on the recent breakthrough of convolutional networks for classification [13], Chen *et al.* [5] used a CNN to replace SVM in the approach of Borth *et al.* [2], obtaining an improved accuracy on ANPs.

The authors in [6] proposed an hierarchical system able to handle sentiment concept classification and localization on objects. They found individual concept detector of SentiBank [2] less reliable for object-based concepts.

Chen *et al.* [7] studied the correlation between the intended publisher sentiment and the actual induced in the viewer ('viewer affect concept'). They aim to recommend appropriate images for the publisher by predicting in advance the induced sentiment in the viewer.

3. THE PROPOSED METHOD

Our proposed method is based on two hypotheses: *i)* the popularity of an image can be fueled by the inherent visual sentiments conveyed; *ii)* semantic descriptions of an image is also important for its popularity, since it makes it easier to be found or looked at.

3.1 Measuring Popularity

It is difficult to precisely define a single score as measure of popularity, and several ways have been proposed to measure it. Khoshla *et al.* [12] used the number of views on Flickr as the principal metric. McParlane *et al.* [14] consider both the number of views and the number of comments for each image as they have been found correlated in video popularity [4]. However they only aim to predict two classes of popularity: high or low.

In this work we follow Khoshla *et al.* [12] and consider the number of views on Flickr as popularity metric. To cope with the large variation of views, we divide the popularity metric by the difference of time between the user upload and our retrieval, then we apply the log function.

3.2 Visual Sentiment Features

To discover which visual emotions are roused from the visualization of an image, a visual sentiment concept classification is performed based on the Visual Sentiment Ontology (VSO). The ontology, consisting in a collection of 3,244 Adjective-Noun-Pairs (ANPs), has been defined by Borth *et al.* [2]. In particular we used DeepSentiBank [5]: a convolutional neural network pre-trained from [13] has been fine-tuned to classify images in one of a subset of 2,096 ANPs. Similarly to its previous version [2], this tool provides a mid-level representation of an image.

For each image we extract two descriptors that we term respectively SentANPs and FeatANPs: the ANPs prediction layer of 2,096d and the rectified activations of the 7^{th} fully connected layer of 4,096d.

3.3 Object Features

Since image popularity is related also to the visual content of the image, we extract the convolutional neural networks features, initially proposed in [12]. A very deep CNN with 16 layers [3] was used to extract for each image the final output containing 1,000 objects from ILSVRC 2014 challenge (termed ObjOut) and the 4,096d representation of the 7^{th} rectified fully connected layer (termed ObjFC7).

3.4 Context Features

Image context information such as tags and description contains important cues that may reflect on the number of views that an image obtains. Entities like people, locations or tourist attractions can affect popularity as *i)* people may be more interested in photographs referring some particular subject; *ii)* the presence of tags and description, the submission of a photo to some groups, etc. make it easier to be found by other users. The extraction of entities from image context strongly depends on the nature of the text, i.e. tags and textual description; due to the different nature of these channels, two different approaches are proposed.

Entity Extraction from Tags. Starting from image tags, we define two new context features that we term TagType and TagDomain. They both rely on Freebase, a large collaborative ontology containing millions of interconnected topics. Given a tag, a search for a *Freebase topic* is performed: if the tag is related to some topics, the most popular one is picked, according to Freebase popularity ranking. Meaningless tags that do not have a match in Freebase topics are ignored, thus they do not act as a nuisance. When a Freebase topic is retrieved, another query is performed to extract its *Freebase types* with the "notable" property and its *Freebase domain.* While *types* are mostly specific (e.g. Person, Author) *domains* cover broader areas (e.g. Film, Music).

Due to the vast number of types in the ontology, a smaller specific type knowledge base is introduced. We first randomly sampled 10k tags from MIR-Flickr dataset vocabulary [11] and used them to extract Freebase types. We select the 100 most frequent types as our specific knowledge base.

The extraction of TagType feature for an image is then straightforward: each tag is used to query Freebase for a notable type. We count the matches to the 100 selected types and obtain a 100d histogram as final feature.

Regarding the TagDomain feature, we take the full list of 78 domains pre-defined by Freebase curators and count the tag matches, similarly as TagType. Thus, the eventual TagDomain feature result in a 78d histogram.

Entity Extraction from description. Differently from the concise tags, image descriptions allow users to comprehensively detail their images in natural language. We seek to recognize subjects and objects of this text to detail context. Hence, we adopt a well known CRF-based language model to perform Named Entity Recognition (NER) [10]. We used the pre-trained 7-class model for MUC that is able to recognize Time, Location, Organization, Person, Money, Percent, Date. We count the occurrences for each class and build a 7d feature that we term NER_7.

3.5 User Features

Previous works have found that the number of views that a photograph is going to obtain depends not only on the image itself and its context information, but also on the author data. In this work we used the same user features proposed by Khosla *et al.* [12]: among these features the most related one to popularity is the mean views of the images of the user, as it represents the popularity of the user himself.

3.6 Popularity prediction

In order to predict popularity as a concise score, we used an off-the-shelf Support Vector Machine. As we are working with large-scale dataset, we used a L2 regularized L2 loss Support Vector Regression (SVR) from LIBLINEAR package due to its scalability with large sparse data and huge number of instances compared to a kernelized version.

4. EXPERIMENTS

As different scenarios show different aspects of popularity, we structure our experimental setups similarly to those of Khosla *et al.* [12], using Flickr social network. Two datasets were used to represent two different scenarios:

- *One-Per-User (OPU)*: we randomly selected 250k images from the VSO Flickr Dataset [2]. This dataset represents the scenario of a Flickr search, where images belong to different users.
- *User Specific (US)*: 25 users from the VSO Flickr Dataset are selected at random to constitute 25 different trials. For each one, 10k images are randomly selected. This dataset represent the scenario of a user that wants to select which of his pictures should be uploaded to attract the attention of other users.

In each experiment, we extract and concatenate the selected features. We freely provide the extracted features on our website. Multidimensional features are L2 normalized, while scalar attributes are scaled in the [0, 1] range. We split every dataset in training and evaluation: half was randomly chosen as training set, while the remaining images were equally split in validation and testing set. The C of SVM was set in the range $[0.001 - 100]$.

After the prediction, testing images are ranked in descending popularity scores and compared to the correct ranking obtained by the ground truth scores. The correlation between these two lists r and s is computed using *Spearman's rank correlation* that ranges in $[-1, 1]$:

$$\rho = \frac{\sum_i (r_i - \overline{r})(s_i - \overline{s})}{\sqrt{\sum_i (r_i - \overline{r})^2} \sqrt{\sum_i (s_i - \overline{s})^2}} \qquad (1)$$

a score of 1 (or -1) corresponds to perfect (inverse) correlation, while 0 corresponds to random ranks.

4.1 Results

Experiments have been carried out for visual features, context ones and visual + context + user combination. We train a model with each single feature to show its predictive power. Then, we combine the features and compare a model with all of them against baselines implemented following the method of Khosla *et al.* [12] i.e. without our novel features. Results are reported in terms of *Spearman's rank correlation* and, for the User Specific dataset, the average scores between the 25 users are reported.

Visual Features. Visual content features include visual sentiment and object detections (Sec. 3.2, 3.3). The latter ones are used in this case as a baseline, including ObjOut and ObjFC7.

Dataset	SentANPs	FeatANPs	ObjOut	ObjFC7	Baseline	All
OPU	0.28	0.32	0.13	0.30	0.30	**0.36**
US	0.31	0.40	0.27	0.40	0.40	**0.43**

Table 1: Visual Features Results

Results are reported in Table 1: sentiment features are comparable with object features. As ANPs are learned starting from a similar network for classification, this suggests the existence of some correlation between them. Nevertheless, SentANPs is higher than ObjOut, suggesting that ANPs are better for popularity prediction than purely object classification. Our features are able to improve overall prediction in both scenarios.

Context Features. The performance of the proposed context features (Sec. 3.4) is compared with a baseline composed by the number of tags, the length of title and description (Table 2).

Dataset	TagType	TagDomain	NER_7	TagNum	TitleLen	DescLen	Baseline	All
OPU	0.42	0.36	0.50	0.55	0.22	0.48	0.61	**0.63**
US	0.44	0.37	0.13	0.23	0.17	0.20	0.33	**0.54**

Table 2: Context Features Results

Our features are comparable with other context-based ones in the OPU scenario. In the US scenario, all the features except TagType and TagDomain lose predictive power due to the limited context of a single user. This is because our features are able to better model semantically the single photos, regardless of the single user. When combined, our feature boost correlation to 0.54 from 0.33 of the baseline.

Visual + Context + User. In this experiment we combined visual, context and user features along with the total combination with and without our novel features. User features are added to resemble a state of the art pipeline. Each modality is singularly tested and finally combined together. Results are reported in Table 3. Note that User Features can't be used for the User Specific scenario as each model is trained for a single user.

User Features produce the highest correlation in the OPU scenario, confirming that popularity is highly related to the

Dataset	Method	Visual Content	Image Context	User Features	All
OPU	proposed	0.36	0.63	0.72	**0.76**
	baseline	0.30	0.61	0.72	0.74
US	proposed	0.43	0.54	n/a	**0.61**
	baseline	0.40	0.33	n/a	0.50

Table 3: Visual + Context + User Features Results

popularity of the author [12]. Despite this, the combination of the three modalities is helpful, boosting correlation from 0.72 to 0.74. Our features further improve upon this, bringing the value to 0.76. In the User Specific dataset, the improvement from the baseline is more pronounced, where a correlation of 0.61 vs 0.50 is obtained.

4.2 Qualitative Analysis

We investigate which specific ANP and semantic metadata correlated the most with the number of views of images. This analysis is performed for the One-Per-User scenario, as it aims to be as generic as possible. Fig. 2a shows the trained SVR weights for each of the 2089 ANPs, in descending order. According to the figure we split the visual sentiments in three categories.

A first group include those ANPs that have a positive impact on image popularity (e.g. *sexy legs*, *beautiful eyes*, *heavy rain*). The rapid drop evinces that a very short number of ANPs corresponds to strongly popular images in the training dataset. Then, we observe that some visual sentiments obtain very low weights, near zero: that ANPs are almost irrelevant to the number of views (e.g. *sunny trees*, *dry forest*). Finally a third group includes ANPs that are associated to a sufficiently negative score: the detection of those push an image towards unpopularity (e.g. *creepy eyes*, *silly clown*).

Extending our analysis to the 28 basic emotions of the Plutchick wheel, we found out that our model marked as unpopular those images that arouse emotions such as *annoyance* or *serenity*, while high scores are likely to be returned in case of sentiments as *amazement* or *ecstasy.* These last emotions derive from ANPs containing the adjective *sexy*, resulting in 10 occurrences in the top 35 visual emotions. A similar analysis on the 100 semantic entities is shown in Fig. 2b. This plot has a similar trend compared with that of visual sentiment, but for the extreme values: in this case the negatively weighted types (e.g. *religious practice* and *software genre*) have more prominent values than the positively weighted ones (e.g. *garment* and *film character*).

5. CONCLUSIONS

In this paper we proposed to employ state-of-the-art visual sentiment features and three new context features to address the problem of predicting whether an image posted on a social network may became popular. We are the first to show a qualitative analysis of which sentiments (as ANPs) are correlated to popularity. Our experiments suggest that some sentiments have a correlation with popularity, still smaller than user features. However, together with our novel context features, they have good prediction power, especially when user features are unavailable as in the User Specific scenario.

Acknowledgments. T. Uricchio is supported by Telecom Italia PhD grant funds and M. Bertini is partially supported by the "Social Museum and Smart Tourism" MIUR project (CTN01_00034_231545).

6. REFERENCES

[1] Y. Bae and H. Lee. Sentiment analysis of Twitter audiences: Measuring the positive or negative influence of popular twitterers. *JASIST*, 63(12):2521–2535, 2012.

[2] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proc. of ACM MM*, 2013.

[3] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *Proc. of BMVC*, 2014.

[4] G. Chatzopoulou, C. Sheng, and M. Faloutsos. A first step towards understanding popularity in YouTube. In *Proc. of INFOCOM*, 2010.

[5] T. Chen, D. Borth, T. Darrell, and S.-F. Chang. DeepSentiBank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv:1410.8586*, 2014.

[6] T. Chen, F. X. Yu, J. Chen, Y. Cui, Y.-Y. Chen, and S.-F. Chang. Object-based visual sentiment concept analysis and application. In *Proc. of ACM MM*, 2014.

[7] Y.-Y. Chen, T. Chen, W. H. Hsu, H.-Y. M. Liao, and S.-F. Chang. Predicting viewer affective comments based on image content in social media. In *Proc. of ICMR*, 2014.

[8] E. Diener and D. DeFour. Does television violence enhance program popularity? *JPSP*, 36(3):333, 1978.

[9] F. Figueiredo, J. M. Almeida, M. A. Gonçalves, and F. Benevenuto. On the dynamics of social media popularity: A YouTube case study. *TOIT*, 14(4):24, 2014.

[10] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proc. of ACL*, 2005.

[11] M. Huiskes, B. Thomee, and M. Lew. New trends and ideas in visual concept detection: the MIR Flickr retrieval evaluation initiative. In *Proc. of ACM MIR*, 2010.

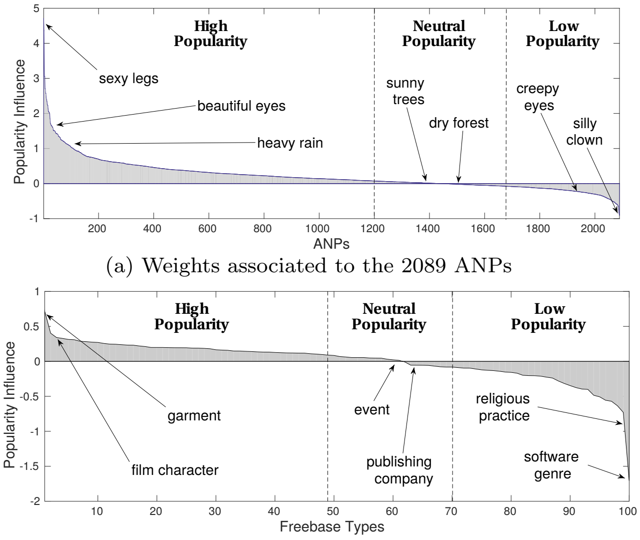
[12] A. Khosla, A. Das Sarma, and R. Hamid. What makes an image popular? In *Proc. of WWW*, 2014.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. of NIPS*, 2012.

[14] P. J. McParlane, Y. Moshfeghi, and J. M. Jose. Nobody comes here anymore, it's too crowded; predicting image popularity on Flickr. In *Proc. of ACM ICMR*, 2014.

[15] R. Plutchik. The nature of emotions. *American Scientist*, 89(4):344–350, 2001.

[16] L. C. Totti, F. A. Costa, S. Avila, E. Valle, W. Meira Jr, and V. Almeida. The impact of visual attributes on online image diffusion. In *Proc. of WebSci*, 2014.



(a) Weights associated to the 2089 ANPs

(b) Weights associated to the 100 Freebase Types

Figure 2: Influence of Multimedia Concepts on Popularity: weights of the 2089 ANP visual sentiment concepts *(top)*; weights of the 100 Freebase Types extracted from contextual image tags *(bottom)*.