



Prediction of Social Image Popularity Dynamics

Alessandro Ortis^(✉), Giovanni Maria Farinella, and Sebastiano Battiato

Department of Mathematics and Computer Science, University of Catania,
Viale A. Doria, 6, 95125 Catania, Italy
{ortis,gfarinella,battiato}@dmi.unict.it

Abstract. This paper introduces the new challenge of forecasting the engagement score reached by social images over time. The task to be addressed is hence the estimation, in advance, of the engagement score dynamic over a period of time (e.g., 30 days) by exploiting visual and social features. To this aim, we propose a benchmark dataset that consists of $\sim 20K$ Flickr images labelled with their engagement scores (i.e., views, comments and favorites) in a period of 30 days from the upload in the social platform. For each image, the dataset also includes user's and photo's social features that have been proven to have an influence on the image popularity on Flickr. We also present a method to address the aforementioned problem. The engagement score dynamic is represented as the combination of two properties related to the dynamic and the magnitude of the engagement sequence, referred as shape and scale respectively. The proposed approach models the problem as the combination of two prediction tasks, which are addressed individually. Then, the two outputs are properly combined to obtain the prediction of the whole engagement sequence. Our approach is able to forecast the daily number of views reached by a photo posted on Flickr for a period of 30 days, by exploiting features extracted from the post. This means that the prediction can be performed before posting the photo.

Keywords: Image popularity prediction · Social media engagement

1 Introduction and Motivation

One of the most promising application of social media analysis is related to social marketing campaigns. In this context, several companies are interested in the automatic analysis of the level of engagement of potential customers with respect to the different posts related to their products. The user engagement is usually measured in terms of number of views, likes or shares. This information can be further combined with web search engines and companies website visits statistics to find correlations between social advertising campaigns and their aimed outcomes. Khosla et al. [7] proposed a log-normalized popularity score that has been then commonly used by the community to measure the level of

engagement of an image. Let c_i be a measure of the engagement related to a social media item i (e.g., number of likes, number of views, etc.). The popularity score of the item i is computed as:

$$score(i) = \log \left(\frac{c_i}{T_i} + 1 \right) \quad (1)$$

where T_i is the number of days since the item i has been uploaded on the social platform. In order to estimate the popularity with Eq. 1, the cumulative engagement obtained by an image post until a specific day is normalized with respect to the total number of days of the post. However, the formulation of the popularity score does not take into account the dynamic of the engagement of the image post. We observe that the relative increment of the engagements over time is not constant and tends to decrease over time. Therefore, the Eq. 1 tends to penalize social media contents published in the past with respect to more recent contents, especially when the difference between the dates of the upload of different posts is large. As claimed in previous works [12], the number of views increases in the first few days, and then remains stable over time. As consequence, the popularity score computed as in Eq. 1 decreases with the oldness of an image post, but this fact has not been considered in the popularity score formulation commonly used in the state of the art works. In other words, two pictures with similar engagement dynamic but very different lifetime, have to be ranked differently. The work presented in this paper proposes a new challenging task which finds very practical applications in recommendation systems and advertisement placement. A new benchmark dataset is released, consisting on a collection of $\sim 20K$ Flickr images, with related user and photo meta-data and three engagement scores tracked for 30 days. The presented dataset will allow the community to perform more accurate and sophisticated analysis on social media content engagement. Moreover, the proposed framework empowers the development of systems that support the publication and promotion of social contents, by providing a forecast of the engagement dynamics. This can suggest when old contents should be replaced by new ones before they become obsolete.

2 Related Works

Many researches tried to gain insights to understand which features make an image popular for large groups of users [3, 4, 7–9]. The authors of [7] analysed the importance of several cues related to the image or to the social profile of the user that lead to high or low values of popularity. The selected features were used to train a Support Vector Regressor (SVR) to predict the image popularity score defined in Eq. 1. Cappallo et al. [6] addressed the popularity prediction task as a ranking problem. They used a latent-SVM with an objective function defined to preserve the ranking of the popularity scores between pairs of images. Few works have considered the evolution of the image popularity over time. In [1] the authors considered the number of likes achieved within the first hour of the post lifecycle to forecast the popularity after one day, one week or one month. The

problem has been treated as a binary classification tasks (i.e., popular vs. non-popular) by thresholding the popularity scores. The paper [10] aims to predict popularity of a category (i.e., a brand) by introducing a category representation, based on user’s posts daily number of likes. The time-aware task proposed in the ACM Multimedia 2017 SMP Challenge¹ aims to predict the top-k popular posts among a group of new posts, given the data related to the history of the past posts in the social platform. Differently than the SMP Challenge, in this work we propose a dataset that reports the engagement scores (i.e., number of views, comments and favorites) of all the involved images recorded on a day-to-day basis for 30 consecutive days. This allows to compute the popularity score for an arbitrary day and to perform more specific analyses on the photo lifecycle. Time aware popularity prediction methods exploit time information to define proper image representations used to infer the image popularity score as in Eq. 1 at a precise time. Indeed, most of the existing works exploit features based on the popularity achieved in the first period. Bandari et al. [2] propose a method which aims to predict popularity of items prior to their posting, however it focuses on news article (i.e., tweets) and the popularity prediction results (i.e., regression) are not satisfactory, as claimed by the authors themselves. Indeed they achieve good results by quantizing the range of popularity values and performing a classification. On the contrary we predict a temporal sequence at time zero, without any temporal hints. Differently than previous works, our approach focuses on the prediction of the whole temporal sequence of image popularity scores (30 days) with a daily granularity, at time zero (i.e., using only information known before the post is published) and without any temporal hints. We designed a regression approach with the aim to estimate the actual values of popularity, instead of quantizing the ranges of possible popularity values or distinguish between popular and non-popular posts (i.e., binary classification). This makes the task very challenging, as there is not an upper bound for the popularity score values and regression errors are considered in the whole 30 days predicted sequences.

3 Proposed Dataset

Most of the datasets on image popularity prediction are based on Flickr. However, the platform only provides the cumulative values of the engagement scores. For this reason we built and released a new dataset including multiple daily values of the engagement scores over 30 days². More than 20K images have been downloaded and tracked considering a period of 30 days. In particular, the first day of crawling the procedure downloads a batch of the latest photos uploaded on Flickr. We run this procedure multiple times by varying the day of the week and time of download in order to avoid the introduction of bias in the dataset. All the social features related to the users, the photos and groups were collected, as well as a number of information useful to compute the engagement scores are downloaded for the following 30 days (i.e., views, comments and favorites). In

¹ <https://social-media-prediction.github.io/PredictionChallenge/>.

² The dataset is available at <http://iplab.dmi.unict.it/popularitydynamics>.

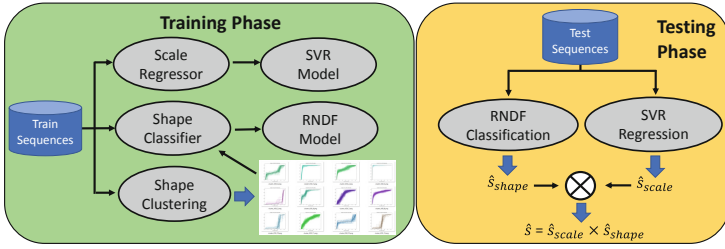


Fig. 1. The pipeline performs the estimation of two properties: the sequence engagement shape \hat{s}_{shape} and scale \hat{s}_{scale} . The shape is estimated by exploiting a Random Forest Classifier (RNDF), whereas a Support Vector Regressor (SVR) is employed to estimate the scale.

all the experiments described in this paper, the engagement score is computed using the number of views, as it is the most used score in the state of the art. However, the other two scores are available with the dataset for further studies. In order to obtain a fine-grained sampling, the first sample of the social data related to a post is taken within the first 2 h from the time of the image upload. Then, a daily procedure takes at least 2 samples per day of the social data for each image. Several social features related to users, photos and the groups have been retrieved. In particular, for each user we recorded: the number of the user’s contacts, if the user has a pro account, photo counts, mean views of the user’s photos, number of groups the user is enrolled in, the average number of members and photos of the user’s groups. For each picture we considered: the size of the original image, title, description, number of albums, groups the picture is shared in, the average number of members and photos of the groups in which the picture is shared in, as well as the tags associated by the user. Moreover, for each photo, the dataset includes the geographic coordinates, the date of the upload, the date of crawling, the date of when the picture has been taken, as well as the Flickr IDs of all the considered users, photos and groups³. We tracked a total of 21.035 photos for 30 days. Some photos have been removed by authors (or not longer accessible through the APIs) during the period of crawling. As result, there are photos that have been tracked for a longer period than others. In particular the dataset consists of: 19.213 photos tracked at least 10 days, 18.838 photos tracked at least 20 days and 17.832 photos tracked at least 30 days. In our experiments we considered the set of photos tracked for 30 days, as it represents the most challenging scenario.

4 Proposed Method

A given sequence s related to the number of views of a posted image over n days, can be formulated as the combination of two properties, namely the *sequence*

³ The IDs of all the entities allow further API requests.

shape and the *sequence scale*. We define these properties as in Eq. 2. In particular, s_{scale} is defined as the maximum value of s , whereas s_{shape} is obtained by dividing each value of s by s_{scale} :

$$\begin{aligned} s &= [v_0, v_1, \dots, v_n] \\ s_{scale} &= \max\{s\} = v_n \\ s_{shape} &= \left[\frac{v_0}{v_n}, \frac{v_1}{v_n}, \dots, \frac{v_n}{v_n} \right] \end{aligned} \quad (2)$$

Therefore, given s_{shape} and s_{scale} , we can obtain the sequence s as follows.

$$s = s_{shape} \times s_{scale} \quad (3)$$

The engagement sequence can be hence considered as a pair of shape and scale. The shape describes the general dynamic (i.e., trend) of the sequence in the monitored period, regardless its actual values. Moreover, sequences with similar shapes can be associated to a more general shape, named “shape prototype”. The sequence scale represents the degree of popularity reached by the photo in n days. We observed that two sequences with the same shape could have very different scales, and vice-versa. To motivate this assumption, we analysed the distributions of the s_{scale} values related to sequences grouped by the assigned shape prototype (i.e., s_{shape}^*). This distribution is shown in Fig. 2. There is a huge variability of the s_{scale} values within the sequences of the same shape. This demonstrates the independent relationship between the scale and the shape of a sequence. Based on this assumption, the proposed approach performs two separate estimations for the shape and the scale of the engagement related to a period of $n = 30$ days. Then, the two results are combined to perform the estimation of the final engagement sequence associated to the photo (see Fig. 1). The shape of the training sequences are grouped with a clustering procedure during the training phase. All the sequences in the same cluster are represented by a shape prototype denoted by s_{shape}^* . The obtained shape prototypes s_{shape}^* are used as labels to train a classifier in order to predict the shape prototype to which a new sequence has to be assigned, given the set input social features. The predicted shape prototype for a new sequence is denoted by \hat{s}_{shape} . A Support Vector Regressor (SVR) is also trained to infer the value of s_{scale} given the social features. The output of the regressor is denoted by \hat{s}_{scale} . The final engagement estimation of an image post for the period of n days is obtained as $\hat{s} = \hat{s}_{shape} \times \hat{s}_{scale}$. We evaluated the proposed approach by exploiting features extracted from the user information, the photo metadata or from the visual content. Although in previous works on popularity prediction the most effective results have been obtained by considering a combination of user information and photo meta-data, some experiments [7] suggest that the semantic content of the picture also influence the prediction. For this reason we evaluated 6 visual representation extracted from the pictures by exploiting three state of the art Convolutional Neural Networks (CNN). For each architecture we extracted the last two layers of activations before the softmax, referred here as $f1$ and $f2$.

Specifically we considered a CNN specialized to classify images into 1183 categories including 978 objects and 205 places (Hybridnet [13]); a CNN trained to assign an Adjective-Noun Pair to an input image among 4342 different ANPs (DeepSentiBank [5]); and a CNN for object classification trained on 1000 categories (GoogleNet [11]).

4.1 Shape Prototyping

Considering Eq. 2, all the s_{shape} sequences have values in the range $[0, 1]$. We consider that all the sequences with the “same” dynamics will have a very similar s_{shape} . Since groups of sequences with the same shape embody instances with a common engagement trend, in the first step of our approach we try to infer a number of popularity shape prototypes representing the different groups of shapes. To this aim, we perform a K-means clustering to group the training sequences. The obtained cluster centroids represent the dynamic models for the sequences within clusters (i.e., each centroid sequence is a shape prototype). In order to select a proper number of clusters for the training sequences (i.e., number of shape prototypes), we performed clustering by considering a wide range of values for K. Then we selected the optimal K considering the Within cluster Sum of Squares (WSS) and the Between cluster Sum of Squares (BSS) indices. The best results have been obtained with $K = 50$.

4.2 Shape Prediction

The result of the shape clustering is a set of shape prototypes. Given this “dictionary” of shapes, any sequence can be assigned to a cluster by comparing its shape with respect to the prototypes. By exploiting the set of shape prototypes we labelled all the sequences in the training and testing dataset by assigning each sequence to a prototype. Then, considering the training sequences, we built a classifier that takes only the social features associated to a post, and predicts the shape (i.e., the prototype) of the corresponding sequence. In order to find the best classifier, we evaluated a pool of algorithms common used for classification tasks, as well as several variations of the social features as input. In particular, we considered the following classifiers: Random Forest (RNDF), Decision Tree (DT), k-Nearest Neighbour (kNN), SVM with RBF or linear kernel and Multi-layer Perceptron (MLP). The best results have been achieved by using all the social features described in Sect. 3 as input of a RNDF classifier. Some of the considered methods required the selection of parameters (e.g., the number of neighbours K for kNN, the parameters C and γ for LSVM and RBFSVM, etc.) which we have established with a grid search method over the training data. In the proposed approach, given a new test image post, the RNDF classifier is hence used to assign a shape prototype based only on its social features.

4.3 Scale Estimation

In order to estimate the value s_{scale} , we trained a Support Vector Regressor (SVR). In particular, we consider the s_{scale} values to compute the popularity

score as in Eq. 1, then the SVR is trained to predict the popularity score. After the prediction, the estimation of the number of views is obtained by inverting Eq. 1. Let \hat{p} be the popularity score estimated by the SVR, the number of views is hence computed by the following equation:

$$\hat{s}_{scale} = (e^{\hat{p}} - 1) \times n \quad (4)$$

First, we evaluated the estimation performances to infer the scale by training the SVR with a single feature as input. For each experimental setting, we computed the Spearman’s correlation between the predicted score and the Ground Truth. This provided a measure of correlation between the features and the value of s_{scale} . In the second stage of experiments, we trained the SVR by considering the concatenation of social features as input. The groups of features have been selected in a greedy fashion considering the performance obtained individually. We have also evaluated two approaches that exploits as input the results obtained by the SVRs trained with the single features. As these methods perform a fusion of several outputs after the prediction, they are often referred as “late fusion” strategies. In the approach named *Late Fusion 1* the outputs of the SVRs are averaged, whereas in the approach *Late Fusion 2* the outputs of the SVRs are concatenated and used to train a new SVR. Scale estimation results, in terms of Spearman’s correlation, are reported in Tables 1 and 2.

5 Problem Analysis

As observed, previous works in popularity prediction simplify the task by quantizing the possible output values into two values [1, 2], or by just predicting a ranking between pictures [6] rather than the actual values of popularity. Furthermore, the majority of these works predict a single score, normalized at an arbitrary age of the post, rather than the dynamic of the post popularity. The few works that aim to predict dynamics lean on the early sequence values (e.g., [1]). This makes the addressed task very challenging. In order to understand the task difficulty, we performed an ablation analysis aimed to understand how the inference of the shape and the scale affect the estimation of the whole sequence. Each experiment exploits some Ground Truth knowledge about either the scale and the shape of the sequences. Therefore, the experiment error rates can be considered as lower bounds for the proposed approach. The considered experimental settings are the following:

- **Case A:** the inferred sequence is obtained by considering both the shape of the Ground Truth (s_{shape}^*) and the Ground Truth scale value (s_{scale}). This method achieves the minimum possible error as both values are taken from the Ground Truth. The measured error is due to the clustering approximation of the sequences.
- **Case B:** in this case the shape is predicted (\hat{s}_{shape}), whereas the scale value is taken from the Ground Truth (s_{scale}). The measured error is due to the clustering approximation of the sequences and the error of the classification to assign the shape prototype.

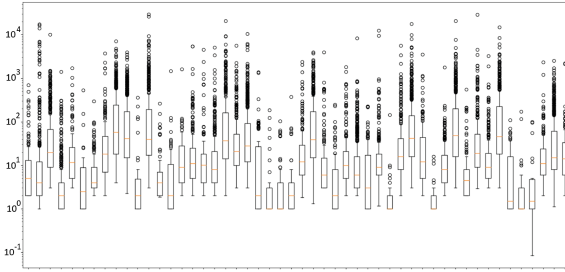


Fig. 2. Distributions of the s_{scale} values of sequences grouped by the s_{shape}^* prototypes assigned by the clustering procedure.

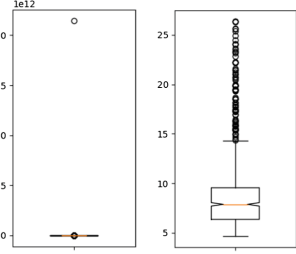


Fig. 3. Box-and-whisker plot of RMSE values (left) computed on a test set. Same distribution after removing the 25% highest/lowest values (right).

- **Case C:** this case combines the shape of the Ground Truth (s_{shape}^*) and the predicted scale value (\hat{s}_{scale}). The measured error is due to the clustering approximation of the sequences and the error in the estimation of the scale.

The mean RMSE errors, considering 10 random runs, of Case A and Case B are 4.97 and 7.51 respectively. These measures are not affected by the prediction of the scale and depend only on the clustering and shape prototyping steps. The results of Case C are affected by the scale estimation. As detailed above, the errors concerning the ablation experiments are related to either the shape and the scale estimations, depending on the definition of the prior Ground Truth knowledge. From this analysis we observed that the whole sequence estimation is more sensitive to errors in the scale prediction. Indeed, an error in the scale affects all the elements of the sequence and hence the RMSE value. Moreover, since we don't quantize the possible output values as done in previous works, there is not an upper bound for the estimated popularity scale and the predicted value could be very large with high magnitude. When the error is averaged, even a few large values could affect the final result. The left plot in Fig. 3 shows the box-and-whisker plot of the computed RMSE values for Case C. From this plot, it's clear that the mean of the RMSE errors is not a good method for the evaluation of the performance, as it is skewed by few large values. Figure 2 shows the presence of several s_{scale} values which are outliers with respect the data distribution. Indeed, the dataset includes few examples with very large scales (depicted as circles in the box-and-whisker plots). For instance, there are only 11 sequences with s_{scale} between $10k$ and $30k$ views. The presence of such few uncommon examples with very large magnitudes caused the skewness in the test error distribution observed in Fig. 3. For these reasons we considered two performance measures for the evaluation of either the proposed method and the Case C (i.e., the two methods that infer the scale of the sequences). Specifically, we considered the 25% truncated RMSE (tRMSE 0.25), also known as the interquartile mean, and the Median RMSE (RMSE MED). The interquartile mean discards an equal

Table 1. Results obtained by considering a single feature approach.

Feature source	Feature ID	Features	Spearman	Proposed method		Case C (shape)	
				tRMSE 0.25	RMSE MED	tRMSE 0.25	RMSE MED
User	0	Ispro	0,25	11,74	8,87	11,50	8,79
User	1	Contacts	0,52	10,48	8,96	10,22	8,69
User	2	PhotoCount	-0,06	11,00	9,65	10,70	9,31
User	3	MeanViews	0,73	9,68	8,19	9,44	7,95
User	4	GroupsCount	<u>0,53</u>	<u>10,42</u>	8,55	10,18	8,42
User	5	GroupsAvgMembers	0,44	11,19	9,37	10,91	9,11
User	6	GroupsAvgPictures	0,46	11,57	8,96	11,31	8,90
Photo	7	Size	0,05	11,03	9,45	10,72	9,18
Photo	8	Title	0,03	11,01	9,63	10,72	9,30
Photo	9	Description	0,06	10,96	9,58	10,66	9,21
Photo	10	NumSets	0,20	11,57	9,33	11,26	9,14
Photo	11	NumGroups	0,34	10,95	9,63	10,65	9,32
Photo	12	AvgGroupsMemb	0,34	10,88	9,54	10,59	9,24
Photo	13	AvgGroupPhotos	0,34	10,91	9,58	10,61	9,28
Photo	14	Tags	0,11	11,27	9,36	10,97	9,20
Visual	hybrid_f1	Hybridnet fc7	0,22	21,61	17,97	20,08	17,44
Visual	hybrid_f2	Hybridnet fc8a	0,26	13,37	11,76	12,95	11,38
Visual	senti_f1	DeepSentiBank fc7	0,25	21,16	18,27	20,60	17,82
Visual	senti_f2	DeepSentiBank fc8	0,30	16,52	14,32	15,99	13,77
Visual	google_f1	GoogleNet pool5/7 × 7_s1	0,26	13,85	12,15	13,43	11,74
Visual	google_f2	GoogleNet loss3/classifier	0,27	13,18	11,61	12,76	11,17

amount of either high and low tails of a distribution. This means that either the best and the worst 25% error values are removed from the mean computation (right plot in Fig. 3). After this process the distribution of the errors is more clear, yet it is still skewed by some outliers depicted as circles.

6 Popularity Dynamic Results

The experimental results have been obtained by averaging the output of 10 random train/test splits with a proportion of 1:9 between the number of test and training images⁴. Given an image represented by a set of social feature, the proposed approach exploits a Random Forest Classifier to predict the shape prototype \hat{s}_{shape} . Then an SVR is used to estimate the popularity score of the image after 30 days. This value is then transformed by using Eq. 4, in order to obtain the scale estimation \hat{s}_{scale} . Finally, the estimated shape \hat{s}_{shape} and the estimated scale \hat{s}_{scale} are combined to obtain the predicted sequence \hat{s} (Fig. 1).

Tables 1 and 2 show the results for Case C and for the proposed approach in terms of tRMSE and Median RMSE at varying of the input feature used by the scale regressor (i.e., the SVR). The results obtained by feeding the SVR with a single feature are detailed in Table 1. In particular, the Spearman’s correlation between the input feature and the Ground Truth popularity is reported in the fourth column. Considering the achieved Spearman’s values, one can observe that the features related to the user have higher correlation values with respect to

⁴ The same protocol has been applied during the ablation study described in Sect. 5.

Table 2. Evaluation results obtained by combining the features.

Feature ID	Feature	Spearman	Proposed method		Case C (shape)	
			tRMSE 0.25	RMSE MED	tRMSE 0.25	RMSE MED
all	concat(0–14)	0,63	9,59	7,43	9,38	7,26
user	concat(0–6)	0,66	9,86	7,03	9,67	6,92
photo	concat(7–14)	0,28	10,80	8,58	10,50	8,30
best_photo	concat(11, 12, 13)	0,34	10,63	9,09	10,35	8,84
user2	concat(3,4)	0,71	9,60	7,66	9,38	7,54
user3	concat(1,3,4)	0,71	9,48	7,51	9,28	7,37
user5	concat(1, 3, 4, 5, 6)	0,68	9,77	7,30	9,59	7,24
	concat(user2, best_photo)	0,72	9,49	7,55	9,27	7,43
	concat(user3, best_photo)	0,71	9,37	7,38	9,16	7,27
	concat(user5,best_photo)	0,69	<u>9,48</u>	<u>7,28</u>	9,28	7,20
	Late Fusion 1 (AVG)	0,59	10,70	9,34	10,43	9,06
	Late Fusion 2 (SVR)	0,46	11,91	8,91	11,66	8,86

the others. The other columns in Table 1 report the error rates on the estimation of the prediction of the whole sequence in terms of trimmed RMSE (tRMSE) and Median RMSE (RMSE MED) for the proposed method (columns 5 and 6) and the Case C (columns 7 and 8). The experiments pointed out that the visual features achieves higher error rates. Indeed, the popularity of a photo in terms of number of views is directly related to the capability of the user and the photo to reach as many users as possible in the social platform. Based on the results reported in Table 1, we further considered the combination of the most effective features for the estimation of the sequence scale. To this aim, we evaluated several early and late fusion strategies. The early fusion consists on creating a new input for the SVR, obtained by the concatenation of the selected features. In particular, we evaluated 10 different combinations of features. Each combination is assigned to an identifier for readability. The obtained results show that in the experiments which involve user related features, the achieved error rates are lower and similar. Indeed, the higher error rates are obtained by the combinations that do not consider user features. The best results in terms of tRMSE are obtained by combining the three best user’s features (i.e., identified by the ID “user3” in Table 2) and the best photo’s features (i.e., “best_photo”). Whereas the best results in terms of Median RMSE are obtained by using only the user related features (i.e., “user”).

7 Conclusions and Future Works

In this paper we introduced a new challenging task of estimating the popularity dynamics of social images. To benchmark the problem, a new publicly available dataset is proposed. We also describe a method to forecast the whole sequence of views over a period of 30 days of a photo shared on Flickr, without constrains on the estimated values, nor considering the early values of the sequences. Future works can be devoted to the extension of the dataset by taking into account other social platforms, relations among the followers, as well as features inspired

by human perception and sentiment analysis. Future experiments could consider the prediction of the popularity dynamics at different time scales. Also, more sophisticated approaches to define the shape prototypes could be evaluated.

References

1. Almgren, K., Lee, J., et al.: Predicting the future popularity of images on social networks. In: Proceedings of the 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016, p. 15. ACM (2016)
2. Bandari, R., Asur, S., Huberman, B.A.: The pulse of news in social media: forecasting popularity. In: ICWSM, vol. 12, pp. 26–33 (2012)
3. Battiato, S., et al.: Organizing videos streams for clustering and estimation of popular scenes. In: Battiato, S., Gallo, G., Schettini, R., Stanco, F. (eds.) ICIAP 2017. LNCS, vol. 10484, pp. 51–61. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68560-1_5
4. Battiato, S., et al.: The social picture. In: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, pp. 397–400. ACM (2016)
5. Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.F.: Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: Proceedings of the 21st ACM International Conference on Multimedia, pp. 223–232. ACM (2013)
6. Cappallo, S., Mensink, T., Snoek, C.G.: Latent factors of visual popularity prediction. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp. 195–202. ACM (2015)
7. Khosla, A., Das Sarma, A., Hamid, R.: What makes an image popular? In: Proceedings of the 23rd International Conference on World Wide Web, pp. 867–876. ACM (2014)
8. Ortis, A., Farinella, G.M., Torrisi, G., Battiato, S.: Visual sentiment analysis based on objective text description of images. In: 2018 International Conference on Content-Based Multimedia Indexing (CBMI), pp. 1–6. IEEE (2018)
9. Ortis, A., Farinella, G.M., D’Amico, V., Adesso, L., Torrisi, G., Battiato, S.: RECFusion: automatic video curation driven by visual content popularity. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 1179–1182. ACM (2015)
10. Overgoor, G., Mazloom, M., Worring, M., Rietveld, R., van Dolen, W.: A spatio-temporal category representation for brand popularity prediction. In: Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, pp. 233–241. ACM (2017)
11. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
12. Valafar, M., Rejaie, R., Willinger, W.: Beyond friendship graphs: a study of user interactions in Flickr. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, pp. 25–30. ACM (2009)
13. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems, pp. 487–495 (2014)