Towards SMP Challenge: Stacking of Diverse Models for Social Image Popularity Prediction

Xiaowen Huang[1] , Yuqi Gao[2] , Quan Fang[1] , Jitao Sang[1,2] , Changsheng Xu[1]
[1]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
[2]State Key Laboratory for Novel Software Technology, Software Institute, NJU, Nanjing 210093, China
{Xiaowen.huang77,gyqnju13,fangquanyi,sangjitao}@gmail.com,csxu@nlpr.ia.ac.cn

ABSTRACT

Popularity prediction on social media has attracted extensive attention nowadays due to its widespread applications, such as online marketing and economical trends. In this paper, we describe a solution of our team CASIA-NLPR-MMC for Social Media Prediction (SMP) challenge. This challenge is designed to predict the popularity of social media posts. We present a stacking framework by combining a diverse set of models to predict the popularity of images on Flickr using user-centered, image content and image context features. Several individual models are employed for scoring popularity of an image at earlier stage, and then a stacking model of Support Vector Regression (SVR) is utilized to train a meta model of different individual models trained beforehand. The Spearman's Rho of this Stacking model is 0.88 and the mean absolute error is about 0.75 on our test set. On the official final-released test set, the Spearman's Rho is 0.7927 and mean absolute error is about 1.1783. The results on provided dataset demonstrate the effectiveness of our proposed approach for image popularity prediction.

CCS CONCEPTS

• **Networks** → *Social media networks*; • **Human-centered computing** → *Social network analysis*;

KEYWORDS

popularity prediction, social media, image, Flickr

1 INTRODUCTION

Popularity prediction is becoming an ever important aspect of web 2.0 applications in an information overload age. In recent years, popularity prediction on social media has attracted extensive attention because of its widespread applications, such as presidential election[16], online marketing and economical trends [14]. Social media serves as essential reflections of the real world, e.g. the number of concerned Flickr images is closely related to the support rate of presidential candidate[6]. Successful popularity prediction of social media posts thus contributes to better understanding of real-word events and its evolution.

Tens of thousands of images are uploaded to the internet every day through various kinds of social networks and photo sharing platforms. While some images receive millions of views, others are completely ignored. Even those images are from the same publisher, different images receive different number of views. This makes us think about a question that can we predict the impact of sharing different posts for a publisher on social media? More specifically, can we predict the number of views the image will receive before the publisher upload it? Due to the importance of popularity prediction, many works have been proposed in bookmarking[9], video[4][13], and social[5] domains. Popularity prediction of social images on Flickr has also been addressed in recent years [11][10][7][15]. Most of these works exploited popularity by investigating into user and image information. Khosla et al. provided evidences that image content such as color, gradients, deep learning features, and social context such as number of friends that lead a high or low popularity of images[7]. Considering that many new images are annotated with no tags and initially, an image has no interaction data[10], McParlane et al. focused on image popularity prediction in a cold start scenario by considering an images context, visual appearance and user context. Wu et al. found that time information plays a crucial role on social media popularity, and presented a novel approach to factorize the popularity into user-item context and time-sensitive context for exploring the mechanism of dynamic popularity[15]. Inspired by these studies, in this work, we propose to address the social image popularity prediction of Social Media Prediction (SMP) challenge in MM 2017. The studies mentioned above demonstrate the effectiveness of heterogeneous features in social image popularity prediction, which lay foundations and motivates us to employ those kind of features into our proposed solution for the challenge.

The task of SMP challenge is designed to predict the impact of sharing different posts for a publisher on social media platform. Given a photo from a publisher, the goal is to automatically predict the popularity of the photo, e.g., view count for Flickr, pin count for Pinterest, etc. In this task, our goal is to predict the number of views of images which publishers upload it on Flickr. The SMP challenge dataset is collected from Flickr, one of the most popular photo sharing social media platforms. The dataset contains 135 users and

MM'17, October 23–27, 2017, Mountain View, CA, USA.
 ISBN 978-1-4503-4906-2/17/10. . . $15.00
DOI: https://doi.org/10.1145/3123266.3127899

Dataset	#image	#user	Temporal. Range(Years)	Avg. Title Length	Avg. Tag Count	Avg. Description Length	Avg. Views
Training set	389155	135	6.05	19.70	9.39	114.80	132.55
Test set	43235	135	5.42	23.73	11.24	106.66	125.46

Table 1: Summary statistics of the dataset used in the experiments

more than 400K posts. Popularity of an image on Flickr is represented by the number of view. In this task, the desired popularity score is log-scaled view count.

To address the prediciton task, we develop a stacking system containing diverse image popularity prediction models. A rich set of models are first utilized to measure the popularity by capturing different features from user-centered and image-centered information. We then combine the individual models to generate the final popularity score. The prediction result of each individual models will be used as input to the stacking model to learn the model parameters. User-centered, image content and image context features are utilized into our prediction models. Experimental results on the test set show competitive performance, which validate the effectiveness of our proposed system.

2 SYSTEM OVERVIEW

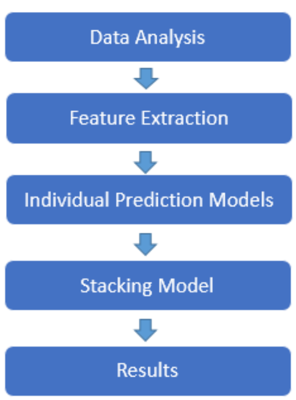

Figure 1: Our proposed prediction system

We first provide an overview of our proposed social image prediction system. As Figure 1 shows, our whole prediction system is divided into four sections. Inspired from the observations of data analysis, we design strategies for feature extraction, and then train some individual prediction models. A stacking model is finally developed to combine the individual models. In the data analysis section, we do some simple data analysis in order to make better decisions of feature and model selection. In feature extraction session, some hidden features are extracted for prediction model learning. We adopt several models to learn the prediction scores, and in the final section, a stacking model is employed to aggregate those individual prediction models to produce the final result.

3 DATA ANALYSIS

3.1 Dataset

The official-released SMP dataset contains 135 users and more than 400K posts, time span of all the posts up to six years. The data is split with time-order, resulting in 90% for training and 10% for testing. The training set is made up of user-centered information, image context specific information, a piece of image and view count corresponding to that image. The test set is available for the final evaluation. Table 1 provides the summary statistics of the dataset used in the experiments. Figure 2 shows some sample images from our prediction training dataset. The popularity of the image is sorted from more popular (left) to less popular (right).

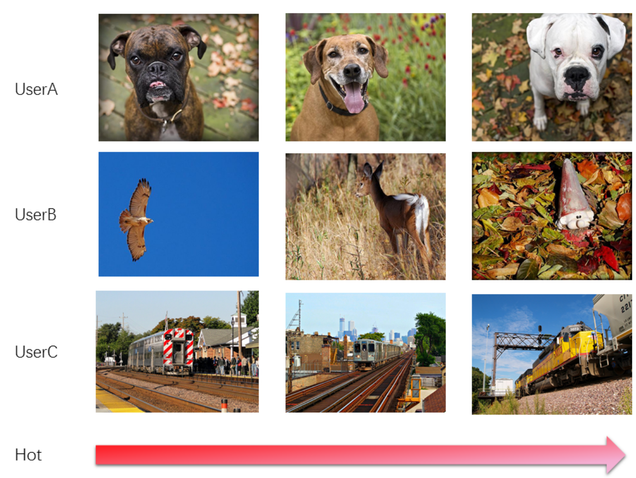

Figure 2: Some sample images from our prediction training dataset with hot annotation.

3.2 Data Analysis

The SMP dataset provides some original features, including unique picture id (pid) along with associate user id (uid). Also, metadata of the picture such as the post date (postdate), comment count (commentcount), whether has people tagged in the photo (haspeople), character length of the title and image caption (titlelen or deslen), number of tags in post. Meanwhile, user-centered information such as average view count, group count, and average member count are also included. Besides, a label set which contains the popularity score (log-scaled views of image) is provided. While image content is useful to predict image popularity to some extent, social cues play a significant role in the number of views an

Features	Avg. Views	Title Length	Description Length	Tag Count	Group Count	Avg. Membercount	Postdate
Pearson Correlation Coefficient	0.56	0.14	0.17	0.35	-0.16	0.11	0.09

Table 2: Pearson Correlation Coefficient of different features with popularity score

image will receive. A person with a larger number of contacts would naturally be expected to receive a higher number of average views. Similarly, we would expect that an image with more tags shows up in search results more often (assuming each tag is equally likely). Therefore, in this section, we do some simple data analysis to find which features are helpful for the task. Pearson Correlation Coefficient[1] is employed as the evaluating indicator of different features with popularity score. Pearson Correlation Coefficient of X and Y can be calculated as:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

where $cov(X,Y)$ is the covariance and σ is the standard deviation.

The results are showed in Table 2. From the Pearson Correlation Coefficient of different features with popularity score, we can observe that: (1) user's average view count performs extremely well in predicting the popularity of a new image with a Pearson correlation coefficient of 0.56. This is consistent with what we have described above. High average view count means that the publisher has more followers so that the greater exposure the image will receive after being released on the social media. Another noteworthy feature is tag count. We can see that the correlation coefficient is 0.35, this is again to be expected as having more tags contributes to increase the probability of retrieval due to browsers' different usage of keywords in the retrieval. In addition, title length, description length and average member count also play important roles on popularity.

One thing that is evident from these results is that the user-centered features and image content features are both necessary. Except for those features provided by the official-released dataset, there are many features that can be inferred and extracted from the data. In the following section, we will further detail the features used in our experiment.

4 FEATURE EXTRACTION

This section introduces both the explicit features and hidden features used by the prediction models.

4.1 Explicit features

According to the results of data analysis above, we remove some original features which are useless or would have side effects. Thus, we consider several user-centered features and image context specific features for model training. We refer to user-centered features as ones that are shared by all images of a single user. Social context specific features means the image specific features that refer to the context, which is the meta data of pictures as described above. They are listed below.

- uid: unique identification of users. Given a unique number for each publisher according to their average view count, the larger the average view count, the bigger the number
- Avg. view count: mean of number of normalized views of all public images of the given user
- Group count: number of groups the given user belongs to
- Avg. member count: average number of members in the groups a given user belongs to
- Tag count: number of tags of the image
- Title length: length of the image title
- Desc. length: length of the image description
- Post_date: specific time (time stamp) of the image being released

4.2 Hidden features

In addition to the explicit features described above, we find that there are some hidden features take positive effect to the popularity prediction.

We investigate some kind of features on image content that could be used to explain the popularity of images. The experimental result shows that there are three features extracted from images are useful for promoting the prediction score. The three image content features are listed below:

- Semantic feature: visual feature extracted from the FC7 layer of the pre-trained VGG-16 Net
- HSV feature: Hue, Saturation, Value features of an image. We only employ the Value feature in HSV space into training because the other two features have no positive effects
- Size feature: the size of image, a two-dimensional vector.

Deep learning algorithm such as convolutional networks (CNNs) have recently become popular for learning image representations [8]. In this paper, we extract the visual features from the final 1000 class classification layer. We choose top 3 categories represented by category index as the semantic features of the image. HSV features and size of image are the simplest image content features. The experiment result suggests that the content also plays an important role in predicting popularity.

In fact, time also effects crucial impact on the popularity but is often overlooked. According to [3], in order to get more attention, when content be shared on the social media is very essential. For example, early afternoon is the optimal time to post on Facebook. 1 p.m. will get you the most shares, while

[1] https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

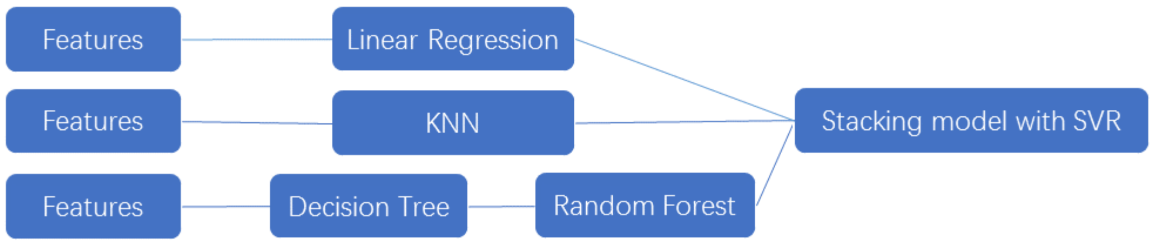

Figure 3: Stacking Framework.

3 p.m. will get you the most clicks. Therefore, for image context, we classify it based on the post time stamp. We introduce the following time-sensitive contextual features to represent an image:

- Post_day: which day the image is posted during a week, choosing from Monday to Sunday.
- Post_time: which time period the image is posted during a day, choosing from morning(06:00 to 11:59), afternoon(12:00 to 17:59), evening(18:00 to 23:59) or night(00:00 to 05:59)

5 INDIVIDUAL MODELS

In this section, we introduce several different approaches for scoring popularity of an image i, and let $p(i)$ denote the popularity score which is log-scaled view count.

5.1 Linear Regression Model

Making use of linear regression model to predict numeric values with numeric data is a simple but effective approach, especially the parameter *avgview* is instrumental during the process and endue the model with comprehensibility. In this way, the popularity score can be generated as:

$$s(x_i) = \mathbf{w}^T \mathbf{x}_i + b \tag{2}$$

where $\mathbf{x}_i$ denotes the features of image i, $\mathbf{w}$ is a weight vector. After the training procedure we can fix the value of $\mathbf{w}$ and b, in this way $p(i)$ is calculated as $p(i) = s(x_i)$. Because of the data continuity, we dropped the features which are based on *post_time* and semantic feature.

5.2 Neighbors-based Regression Model

The idea of this model is to make use of the continuous variables belonging to the features rather than discrete ones. Since most of the features are continuous rather than discrete, we adopt them without the feature *post_date* and semantic feature. Measuring the distance between pair of points is an important question during the process of nearest neighbors regression, and to avoid overlong processing time brought by naive brute-force approach, we utilize K-D Tree [1] to compute the nearest neighbors. To strengthen the influence of nearby points and make sure these points contribute more to the regression process than faraway points, weights were set inversely proportional to the distance from some point and in the way the local neighborhood contributes not equally to the classification. We will compare these two settings called *weighted* and *original* in the section below. As for the parameter, we empirically set the neighbors to generate is 5, that is $k = 5$.

5.3 Decision Tree Regression Model

5.3.1 Decision Tree Regression. Man-made parameters often brings artifacts and makes the model over-fitting, so a model with the ability to avoid those problem is what we need. Non-parametric decision trees is a appropriate choice in this case. Our features have many user-defined values like the make segmentation on time of day. Also, these values can be seen as categorical features. With a set of if-then-else decision rules, decision trees handle these situations well and even there are assumptions are violated which always happens when there are manual features. When it comes to regression task, decision tree using rules to fit a curve and in this way we can use the model to predict the popularity score with all features.

5.3.2 Random forest Regression. When the depth parameter in decision tree is too high, the decision tree would learn too many details. Under these circumstances, the ensemble method random forest [2] helps a lot. The random forest is a meta model fits the results from decision trees on different subsets of the original dataset. The averaging strategy random forest using can improve the accuracy and avoid over-fitting.

6 STACKING MODEL

As shown in Figure 3, we employ Support Vector Regression (SVR) in scikit-learn [12] to train a meta model of different individual models mentioned above. SVR make a assumption that there is ρ deviation between the output of model $s(x)$ and the actual output y, then the loss function can be written as:

$$\min_{w,b} \frac{1}{2}||w||_2{}^2 + k\sum_{i=1}^{n} g(w^T \mathbf{x}_i + b - y_i) \tag{3}$$

Models	Spearman's Rho	MSE	MAE
Linear Regression Model	0.78	3.7	1.41
Neighbors-based Regression Model	0.81	1.65	0.95
Decision Tree Regression	0.83	1.4	0.95
Random Forest Regression	0.87	1.15	0.85
Stacking Model	0.88	1.15	0.75

Table 3: Evaluation metric on different test set using different models

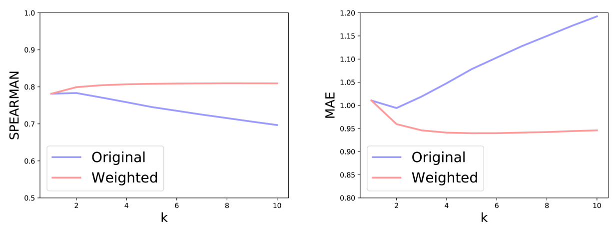

Figure 4: Comparation of different parameters on Spearman's Rho and MAE

where k is a constant parameter, n is the number of samples and $\mathbf{x}_i$ denotes the features of image i. In addition, the expression $g(m)$ equals 0 when $|m| \leq \rho$ otherwise $g(m) = |m| - \rho$. Besides, $\mathbf{x}$ represents the results generated by the individual models. With the solution w, b, we can conduct a meta model which integrates the individual models.

7 EXPERIMENTS AND RESULTS

In this section, we detail the model parameters setting and results of our experiment for image popularity prediction, respectively.

7.1 Model Parameters Setting

Because of the methods other than neighbors-based one don't have to set parameters manually, we only show the experiment of neighbors-based approach. As shown in Figure 4, the parameter *weighted* better than *original*, whether we focus on MAE or Spearman's Rho. As for the parameter k, there is no obvious difference when $k \geq 5$, we finally set k=5 through synthetical consideration.

7.2 Results and Analysis

To evaluate the prediction performance, we utilize the following three metrics:

- Spearman's rank correlation coefficient (Spearman's Rho)[2]:

$$SPEARMAN = \rho_{r_x, r_y} = \frac{cov(r_x, r_y)}{\sigma_{r_x} \sigma_{r_y}} \tag{4}$$

  where ρ_{r_x, r_y} denotes the Pearson correlation coefficient of rank variables r_x and r_y which are ranks converted from n raw scores x and y, $cov(r_x, r_y)$ denotes the covariance and σ_{r_x} and σ_{r_y} are the standard deviations.
- Mean Squared Error (MSE)[3]:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\widehat{X}_i - X_i)^2 \tag{5}$$

  where $\widehat{X}$ denotes a vector of n predictions, and X is the observed vector.
- Mean Absolute Error (MAE)[4]:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\widehat{X}_i - X_i| \tag{6}$$

  where the symbols share the same definition with the ones defined in (5).

Based on the dataset provided by SMP challenge, and features for model learning described in section 4, several models are employed to predict the popularity score. In our experiment, we randomly split the official-released training set into three parts, 6/9 of the data as training set for individual models, 2/9 as training set for stacking model, and the rest part is employed as our test set. The evaluation results of different models are listed in Table 3. For our random-split test set, we can observe that four individual models both achieve good performances. Even the simplest Linear Regression Model achieve 0.78 in Spearman' Rho, this indirectly reflects the features we extracted are effective. The decision tree regression model can better handle the categorical features, and the ensemble method random forest with averaging strategy can improve the accuracy and avoid overfitting. These approaches result in an efficient improvement compared with the first two individual models. The stacking model works best in the popularity prediction task due to its effective integration of the results generated by the individual models. The results demonstrate that our proposed popularity prediction system is feasible and effective.

On official final-released test set, the Spearman's rank correlation coefficient is 0.7927, mean squared error is 2.4973 and mean absolute error is about 1.1783. The evaluation results show a certain degree of decline compared with our random-split test set. Here we show some failure cases in Figure 5 and Figure 6 and make some discussions on them. These pictures in Figure 5 are both from a user with average log-scaled view of 3.67. We can make an obversion that

[2]https://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient

[3]https://en.wikipedia.org/wiki/Mean_squared_error

[4]https://en.wikipedia.org/wiki/Mean_absolute_error

Predict: 4.14 Truth: 0   Predict: 3.55 Truth: 1.6

Figure 5: Two example of failure predictions.

the method will fail when the actual view count of current sample is too low compared to the average. Also, the same obversion can be generated when the actual view count of current sample is too high compared to the average(The owner's average log-scale view is 7.16) as shown in Figure 6. As

Predict: 8.11 Truth: 10.2   Predict: 7.08 Truth: 10.4

Figure 6: Two example of failure predictions.

the Figures shows, our model fails when the current sample is too far away from the average value. That indicates our model is less variety to some extent and this is also the reason why our performance drops on the official final-released test set. Besides, when it comes to the reasons for specific images, we can see that our method does not achieved high discrimination on images with striking difference but analogous features, that means there are some features we did not pay enough attention to. These are the shortcomings of our approach and the aspects we need to improve in the future.

8 CONCLUSION AND DISCUSSION

In this paper, we introduce our solution for the social media prediction challenge in MM 2017. We present a stacking framework by combining a diverse set of models to predict the popularity of social images on Flickr. The results on the provided dataset demonstrate the effectiveness of our proposed approach for image popularity prediction. In the future, several lines of work can be conducted to further improve the results. (1) In this challenge, as the time zone information of the provided data is partially missing, we can only regard the Greenwich Mean Time as the local time of publishers who has no time zone information. This approximation may has some negative influence on our experiment result. (2) One important reason that popularity prediction in challenging is that we have no very litter knowledge about what feature can contribute to the prediction result. Now the extacted features for this challenge are manual-designed. It is critical to incorporate feature learning mechanism in the future solution to explore more powerful features.

9 ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (No. 61432019, 61225009, 61303176, 61272256, 61373122, 61332016).

REFERENCES

[1] Jon Louis Bentley. 1975. Multidimensional binary search trees used for associative searching. *Commun. ACM* 18, 9 (1975), 509–517.
[2] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
[3] Nathan Ellering. 2017. *What 20 Studies Say About The Best Times To Post On Social Media. in LATEX.* https://coschedule.com/blog/best-times-to-post-on-social-media/.
[4] Flavio Figueiredo. 2013. On the prediction of popularity of trends and hits for user generated videos. In *Proceedings of the sixth ACM international conference on Web search and data mining.* ACM, 741–746.
[5] Liangjie Hong, Ovidiu Dan, and Brian D Davison. 2011. Predicting popular messages in twitter. In *Proceedings of the 20th international conference companion on World wide web.* ACM, 57–58.
[6] Xin Jin, Andrew Gallagher, Liangliang Cao, Jiebo Luo, and Jiawei Han. 2010. The wisdom of social multimedia: using flickr for prediction and forecast. In *Proceedings of the 18th ACM international conference on Multimedia.* ACM, 1235–1244.
[7] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. 2014. What makes an image popular?. In *Proceedings of the 23rd international conference on World wide web.* ACM, 867–876.
[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems.* 1097–1105.
[9] Kristina Lerman and Tad Hogg. 2010. Using a model of social dynamics to predict popularity of news. In *Proceedings of the 19th international conference on World wide web.* ACM, 621–630.
[10] Philip J McParlane, Yashar Moshfeghi, and Joemon M Jose. 2014. Nobody comes here anymore, it's too crowded; predicting image popularity on flickr. In *Proceedings of International Conference on Multimedia Retrieval.* ACM, 385.
[11] Xiang Niu, Lusong Li, Tao Mei, Jialie Shen, and Ke Xu. 2012. Predicting image popularity in an incomplete social media community by a weighted bi-partite graph. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on.* IEEE, 735–740.
[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[13] Gabor Szabo and Bernardo A Huberman. 2010. Predicting the popularity of online content. *Commun. ACM* 53, 8 (2010), 80–88.
[14] Alexandru Tatar, Marcelo Dias de Amorim, Serge Fdida, and Panayotis Antoniadis. 2014. A survey on predicting the popularity of web content. *Journal of Internet Services and Applications* 5, 1 (2014), 8.
[15] Bo Wu, Tao Mei, Wen-Huang Cheng, Yongdong Zhang, et al. 2016. Unfolding Temporal Dynamics: Predicting Social Media Popularity Using Multi-scale Temporal Decomposition.. In *AAAI.* 272–278.
[16] Quanzeng You, Liangliang Cao, Yang Cong, Xianchao Zhang, and Jiebo Luo. 2015. A multifaceted approach to social multimedia-based prediction of elections. *IEEE Transactions on Multimedia* 17, 12 (2015), 2271–2280.