Social Media Popularity Prediction: A Multiple Feature Fusion Approach with Deep Neural Networks

Keyan Ding
City University of Hong Kong
Hong Kong
keyanding2-c@my.cityu.edu.hk

Ronggang Wang
Shenzhen Graduate School, Peking University
Shenzhen, China
rgwang@pkusz.edu.cn

Shiqi Wang*
City University of Hong Kong
Hong Kong
shiqwang@cityu.edu.hk

ABSTRACT

Social media popularity prediction (SMPD) aims to predict the popularity of the post shared on online social media platforms. This task is crucial for content providers and consumers in a wide range of real-world applications, including multimedia advertising, recommendation system and trend analysis. In this paper, we propose to fuse features from multiple sources by deep neural networks (DNNs) for popularity prediction. Specifically, high-level image and text features are extracted by the advanced pretrained DNN, and numerical features are captured from the metadata of the posts. All of the features are concatenated and fed into a regressor with multiple dense layers. Experiments have demonstrated the effectiveness of the proposed model on the ACM Multimedia Challenge SMPD2019 dataset. We also verify the importance of each feature via univariate test and ablation study, and provide the insights of feature combination for social media popularity prediction.

CCS CONCEPTS

• **Information systems → Content analysis and feature selection**; • **Human-centered computing → Social media**.

KEYWORDS

Social media, image popularity, deep neural networks, features fusion

ACM Reference Format:
Keyan Ding, Ronggang Wang, and Shiqi Wang. 2019. Social Media Popularity Prediction: A Multiple Feature Fusion Approach with Deep Neural Networks. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3343031.3356062

1 INTRODUCTION

In the past decade, user-generated content (UGC) in online social networks is increased dramatically. For example, hundreds of thousands of photos are uploaded to the internet every minute through various social network and photo sharing platforms.These images tend to receive a quite diverse distribution of views, likes and comments, and this phenomenon motivates us to further analyze and model the popularity of social media. In particular, the task of popularity prediction tries to infer the degree of interaction between users and specific posts (e.g., predicting the number of received likes). Accurate and robust prediction of the social media popularity plays important roles in multiple application domains including content recommendation, advertisement, information retrieval. However, this task is non-trivial because many factors will affect the popularity, including image content, image caption, user information, upload time and location [5, 12, 20]. Moreover, different modalities of data may have different impacts, and it is difficult to model their relationships.

In this paper, we first try to find and analyze key factors that may affect the popularity of posts, and address the popularity prediction problem by considering features from four sources: visual content, text, user and temporal-spatial information. Their influences on the popularity are also analyzed. Furthermore, we propose to fuse features from these sources by the deep neural network (DNN) for popularity predictions. DNNs can intrinsically learn a high-level representation from the image and text, which avoids heavy work on feature engineering. Specifically, from the perspective of visual content, high-level image features are extracted by the pretrained ResNet [6]. We also use trained NIMA model [26] and IIPA model [4] to compute the image aesthetics score and the intrinsic image popularity score, respectively. For text content, deep text features are extracted by the pretrained BERT [3], and the statistical information, such as the count of tags and the length of captions, are also calculated. In addition, we capture other numerical features from the user and temporal-spatial information. By combining with the visual and text features, they are fed into multiple dense layers together for regression. Experiments demonstrate that the proposed model performs well on the SMPD2019 dataset [24]. Considering the fact that different features may have substantially different influence on the prediction task, we perform ablation study and univariate test to further analyze the importance of each feature.

2 RELATED WORK

In recent years, many efforts have been devoted to popularity prediction of social media contents. They share a common pipeline consisting of extracting various features and then using a regression model to compute the final popularity index.

Many studies aim to predict the popularity of images from the professional photo-sharing site Flickr. McParlane *et al.* [20] proposed an image popularity prediction method when limited textual

*Corresponding author.

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3356062

Figure 1: Example of posts in SMPD2019 dataset [24], which mainly contains user ID, image, caption, tags, category, upload time, geographic location and popularity label.

or interaction data are available. They focus on three aspects: image context, visual appearance, and user context to predict the number of comments and views. Khosla *et al.* [12] predicted the image popularity using the image content and the user context based on millions of images. They systematically analyzed the impact of low-level, middle-level and high-level features on the prediction accuracy. Wu *et al.* [29–31] incorporated multiple time-scale dynamics into sequential prediction of social image popularity. Zhang *et al.* [32] proposed a model with user-guided hierarchical attention mechanism for multi-modal content popularity prediction.

There are also several works based on other social network platforms. Mazloom *et al.* [19] examined several engagement parameters, such as sentiment, vividness and entertainment, to predict the popularity of brand-related posts on Instagram. Considering the preferences of different users to the items, they also presented a model to predict the popularity of posts related to a specific user [18]. Hessel *et al.* [7] compared multimodal content to social context for predicting relative popularity on Reddit.

Many researchers predicted social media popularity based on ACM Multimedia Challenge 2017 [28] and 2018 [23]. For example, Hidayati *et al.* [8] proposed an influence- and aesthetics-aware popularity predictor by an SVR-based and regression tree-based ensemble model. Li *et al.* [17] proposed a hybrid model which combines the convolutional neural network with XGBoost for predicting popularity. Hsu *et al.* [9, 10] presented an iterative refinement approach to predict popularity. Huang *et al.* [11] used the random forest algorithm to exploit post-related and user-related features for popularity prediction. They all achieved competitive performances.

3 DATA ANALYSIS

In this section, we aim to discover and analyze important factors that may affect the popularity of social multimedia. The data used in this paper are from ACM Multimedia Challenge SMPD2019 [24], a large social multimedia dataset collected from Flickr. Each of social media posts has rich contextual information (e.g. image, text, temporal-spatial information, user profile), and the popularity label is regarded as the log-scaled number of views. Fig. 1 shows an example of posts in the dataset. We will consider the popularity prediction from four perspectives: visual, text, user information, and temporal-spatial information.

From the perspective of visual clues, an image with higher quality and aesthetics tends to receive more likes and thus become more popular [8]. Therefore, assessing photo quality and aesthetics is important for image popularity prediction. In addition, the objects in the image will largely affect the popularity. For example, brilliant selfie often obtains more attention than ordinary photos without people on social media [1]. To understand the image content and extract the object-level features, the deep neural networks, which have been verified to be effective in many fields especially in computer vision and natural language processing (NLP) tasks, can be adopted in this scenario [15].

The text information including captions, tags and categories is also important for popularity prediction. For example, a hot tag contributes significantly to image popularity because of the extensive exposure to viewers beyond followers. Generally speaking, the more tags of a post, the greater chances of receiving more likes. To extract the text feature, NLP tools, such as Word2Vec [21], GloVe [25] and BERT [3] are beneficial.

The user information will largely affect the popularity of posts. Many studies have shown that there is a high correlation between image popularity and users [2, 11, 12]. The straightforward explanation is that different users may have different numbers of followers. In general, images posted by the user with more followers have a higher chance of receiving more views and likes. In the SMPD2019 dataset, some key information of users is missing, thus we crawl the extra information by the provided user alias, including the number of posts, followers and followings. For the unavailable user, these values are replaced with the average of other users.

In addition, the temporal and spatial information may also have an impact on popularity. The earlier the post is uploaded, the more views or likes the post receives. The differences in terms of uploading time in a day and geographical locations also affect the popularity. Considering the fact that the number of various features is not "the more the better" [17, 27], the contribution of each feature is of great interest, such that we can choose the useful features for the prediction task.

4 METHOD

Our method consists of two components: feature extraction and popularity regression. We first extract high-level feature representations of the image and text by pretrained DNNs, and useful statistical information from social metadata. Then we design a regressor with multiple dense layers to learn the popularity. Fig. 2 shows the framework of the proposed model for social media popularity prediction.

4.1 Feature Extraction

4.1.1 *Visual features.*

- **Deep image features.** Deep learning has become a very popular method for image representations. We utilize a recent deep learning model ResNet-101 [6] trained on ImageNet [14] for classification to extract features in the second to the last layer, resulting in a feature vector with 2048 dimensions.
- **Image aesthetics score.** We use NIMA model [26] to measure image aesthetics. NIMA is a CNN-based image aesthetics predictor trained on a large-scale AVA database [22], and achieved state-of-the-art performance on aesthetics evaluations. Besides, it can be used to compare the quality of images.

Instead of using hand-crafted features to measure aesthetics [8], we directly compute the image aesthetics score by the trained NIMA model as a visual feature.

- **Intrinsic image popularity score.** We use IIPA model [4] to measure the intrinsic image popularity. IIPA singles out the contribution of visual content to image popularity. It is trained on millions of images from Instagram by the deep ranking method, and achieves human-level performance. We compute the intrinsic image popularity score by the trained IIPA model as another image feature.

4.1.2 *Text features.*

- **Deep text features.** Each post is associated with the text information, including category, tag and caption. Instead of using simple one-hot encoding or shallow Word2Vec model [21], we adopt the recent BERT model [3] to extract the text features. BERT applies the bidirectional training of Transformer to learn a language model. It has achieved state-of-the-art results in a wide variety of NLP tasks, including text classification, machine translation and question answering. We obtain a 768 dimension text feature from the pretrained BERT model.
- **Tags count and caption length.** To further rich the text representation, we count the number of tags and the length of caption as the auxiliary text features.

4.1.3 *Numerical features.*

- **User features.** The log-scaled number of followers, followings and posts as used as the numerical features of users.
- **Temporal-spatial features.** We use the time of how long the post has been uploaded, and the geographical location (Latitude and Longitude) as the temporal and spatial features, respectively.
- **Features normalization.** Each numerical feature will be normalized with Z-score method[1] before training, including the additional numerical features of text (tags count and caption length) and image (aesthetics score and intrinsic popularity score).

4.2 Popularity regression

After extracting these features, we model the relationship between the features and popularity. There are many off-the-shelf regression models, such as Support Vector Machine (SVM) and Random Forest (RF), and they have demonstrated their effectiveness for popularity modeling [8, 10, 11].

In this paper, we adopt a DNN-based regressor for popularity prediction, as shown in Fig. 2. We divide the features to three categories: deep image feature (2048D), deep text feature (768D) and the set of numerical features (10D), and feed these features into the neural network. The network is composed of four dense layers, and each is followed by a batch normalization layer and a ReLU activation layer. The loss is the mean-squared error between the predicted score and the label score. To alleviate the potential overfitting, we add the L2-norm regularization term into the loss function.

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

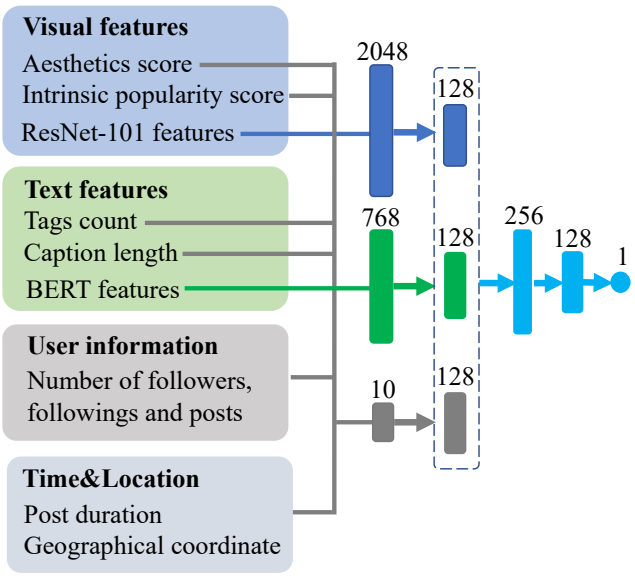

Figure 2: The proposed framework for social media popularity prediction. High-level feature representations of the image and text are extracted from the pretrained ResNet [6] and BERT [3], respectively. By combining with selected numerical features, they are fed into the dense layers together for regression. (Each arrow represent a dense layer followed by the batch normalization and the ReLU activation, and the number is the output dimension of the current layer).

5 EXPERIMENTS

In this section, we first describe the implementation details of the proposed method, including the dataset partition, training procedures and evaluation metrics. We then show the performance of our model and analyze the importance of each feature via univariate and ablation tests.

5.1 Implementation details

SMPD2019 dataset [24] contains 305,613 posts. We download 287,158 images via the provided URL and deleted the posts without images. In the experiments, we choose 80% posts for training and the remaining 20% for testing by three different partition manners. More specifically, the partition manners include 1) Set-I: partition by the upload time of the post; 2) Set-II: partition by users; and 3) Set-III: random partition. The evaluation criteria of the performance are Mean Absolute Error (MAE), Mean Squared Error (MSE) and Spearman Rank Correlation Coefficient (SRCC).

The training of network is implemented by PyTorch[2] framework with an Intel Xeon W-2123 3.60GHz CPU and an NVIDIA RTX2080Ti GPU. We use Adam [13] optimizer with an L2 penalty multiplier of 10^{-3} for training. The learning rate is set to 10^{-4}, and decays linearly by a factor of 0.9 after each epoch. When computing the deep feature of images, the input image is rescaled to

[2]https://pytorch.org

$224 \times 224 \times 3$. The training batch size is 128 and the total number of epochs is 10.

5.2 Model performance

Table 1 shows the performance of the proposed model on three different test sets. We can find that our model achieves the best results on the test Set-III (random partition) with a large margin, and the result of Set-I (partition by post time) is slightly superior to that of Set-II (partition by users). The main reason may be that there could be many repetitive users in the training and testing sets in Set-III, and little repetitive users in Set-I. Moreover, user information is a key factor for popularity prediction, which will be discussed in the next subsection. Therefore, the prediction of Set-II is more difficult than that of Set-I and Set-III. In the subsequent experiments, we focus on the Set-I for univariate and ablation studies.

Table 1: The performance of proposed model on three test sets split by different manners

Test set	MAE	MSE	SRCC
Set-I (time)	1.420	3.361	0.649
Set-II (user)	1.432	3.448	0.640
Set-III (random)	1.275	2.847	0.712

5.3 Univariate and ablation study

In this subsection, we evaluate the contribution and influence of each feature to the final performance by a univariate study. We also conduct an ablation study to examine the validity of each adopted feature by setting the unused features to zero.

Univariate study. Table 2 shows the results of the univariate study. The leftmost column indicates the only feature adopted. From Table 2, we can easily find that the user information (the number of followers, followings and posts) achieves the smallest MAE, MSE and the highest SRCC, implying that it plays the most important role in the prediction. This phenomenon agrees with many studies [12, 16, 27], and is in accordance with the fact that more followers often implies more popular. It is worth noting that since we replace the unavailable user information with the average in the dataset, there is still room to improve the performance if we can attain accurate information for all users. We also notice that visual and text features are conducive to popularity prediction, especially the deep text feature and the deep image feature. It shows that deep learning-based features have strong generalizability and are indeed useful for popularity prediction. In addition, the feature of tags count is meaningful, which is consistent with the phenomenon that more tags often leads to higher popularity. One key reason is tags can increase the visibility in the search results.

Ablation study. Since the user information, deep text feature, deep image feature and tags count play important roles in popularity prediction, in this experiment, we evaluate the model performance by discarding one of these four features. Table 3 shows the results of the ablation study, from which we find that the performance becomes worse obviously when the user feature is discarded. It means the user information contributes most to the final result, and is consistent with the one in the univariate study. When removing the deep image feature or deep text feature, the performance degrades slightly. However, discarding the feature of tags count has little influences on the final result, although it is a useful feature in the univariate study. The reason may be that the tags count is highly relevant with the deep text feature, such that the feature of tags count cannot provide the complementary information for prediction. It is worth noting that we did not consider the deep text feature in the final submission of SMPD challenge [24], which leads to a little performance degradation.

Table 2: Results of univariate study

Feature adopted	MAE	MSE	SRCC
User information	**1.521**	**4.140**	**0.558**
Visual features	1.842	5.312	0.320
Text features	1.775	4.923	0.395
Time&location	1.942	5.915	0.101
Deep image feature	1.872	5.478	0.302
Deep text feature	1.802	5.028	0.376
Tags count	1.860	5.358	0.338
Title length	1.954	5.958	0.128

Table 3: Results of ablation study

Feature discarded	MAE	MSE	SRCC
User information	**1.688**	**4.522**	**0.501**
Deep text feature	1.431	3.509	0.631
Deep image feature	1.423	3.525	0.640
Tags count	1.418	3.450	0.650

From the univariate and ablation study, we can find that among the various features to be considered, some of them may be useless for popularity prediction in our model. Deep features extracted from image and text has a powerful capability for popularity modeling. User-related features (e.g., the number of followers and following) are the most useful ones in the task of social multimedia popularity prediction.

6 CONCLUSIONS

In this paper, we have presented a social media popularity prediction strategy based on multiple features fusion with neural network for ACM Multimedia Challenge (SMPD2019). We consider four perspectives (visual, text, user, and temporal-spatial) to predict the popularity of posts, and train a DNN-based regression model to obtain the final popularity score. Both univariate and ablation studies provide useful insights regarding the importance of these features. In particular, deep image feature, text feature and user information are important clues to infer the social media popularity.

ACKNOWLEDGEMENTS

This work is supported by Hong Kong RGC Early Career Scheme 9048122 (CityU 21211018), City University of Hong Kong Applied Research Grant (ARG) 9667192 and by Shenzhen Research Projects of JCYJ20160506172227337 and GGFW2017041215130858.

REFERENCES

[1] Saeideh Bakhshi, David A Shamma, and Eric Gilbert. 2014. Faces engage us: Photos with faces attract more likes and comments on Instagram. In *SIGCHI Conference on Human Factors in Computing Systems*. 965–974.

[2] Ethem F Can, Hüseyin Oktay, and R Manmatha. 2013. Predicting retweet count using visual cues. In *ACM International Conference on Information & Knowledge Management*. 1481–1484.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805* (2018).

[4] Keyan Ding, Kede Ma, and Shiqi Wang. 2019. Intrinsic Image Popularity Assessment. *ACM International Conference on Multimedia*.

[5] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. 2015. Image popularity prediction in social media using sentiment and context features. In *ACM International Conference on Multimedia*. 907–910.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

[7] Jack Hessel, Lillian Lee, and David Mimno. 2017. Cats and captions vs. creators and the clock: Comparing multimodal content to context in predicting relative popularity. In *International Conference on World Wide Web*. 927–936.

[8] Shintami Chusnul Hidayati, Yi-Ling Chen, Chao-Lung Yang, and Kai-Lung Hua. 2017. Popularity meter: An influence-and aesthetics-aware social media popularity predictor. In *ACM International Conference on Multimedia*. 1918–1923.

[9] Chih-Chung Hsu, Chia-Yen Lee, Ting-Xuan Liao, Jun-Yi Lee, Tsai-Yne Hou, Ying-Chu Kuo, Jing-Wen Lin, Ching-Yi Hsueh, Zhong-Xuan Zhang, and Hsiang-Chin Chien. 2018. An iterative refinement approach for social media headline prediction. In *ACM International Conference on Multimedia*. 2008–2012.

[10] Chih-Chung Hsu, Ying-Chin Lee, Ping-En Lu, Shian-Shin Lu, Hsiao-Ting Lai, Chihg-Chu Huang, Chun Wang, Yang-Jiun Lin, and Weng-Tai Su. 2017. Social media prediction based on residual learning and random forest. In *ACM International Conference on Multimedia*. 1865–1870.

[11] Feitao Huang, Junhong Chen, Zehang Lin, Peipei Kang, and Zhenguo Yang. 2018. Random forest exploiting post-related and user-related features for social media popularity prediction. In *ACM International Conference on Multimedia*. 2013–2017.

[12] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. 2014. What makes an image popular?. In *International Conference on World Wide Web*. 867–876.

[13] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv:1412.6980* (2014).

[14] 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.

[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.

[16] Liuwu Li, Sihong Huang, Ziliang He, and Wenyin Liu. 2018. An effective text-based characterization combined with numerical features for social media headline prediction. In *ACM International Conference on Multimedia*. 2003–2007.

[17] Liuwu Li, Runwei Situ, Junyan Gao, Zhenguo Yang, and Wenyin Liu. 2017. A hybrid model combining convolutional neural network with xgboost for predicting social media popularity. In *ACM International Conference on Multimedia*. 1912–1917.

[18] Masoud Mazloom, Bouke Hendriks, and Marcel Worring. 2017. Multimodal context-aware recommender for post popularity prediction in social media. In *Thematic Workshops of ACM Multimedia*. 236–244.

[19] Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn Van Dolen. 2016. Multimodal popularity prediction of brand-related social media posts. In *ACM International Conference on Multimedia*. 197–201.

[20] Philip J McParlane, Yashar Moshfeghi, and Joemon M Jose. 2014. Nobody comes here anymore, it's too crowded; predicting image popularity on Flickr. In *ACM International Conference on Multimedia Retrieval*. 385–391.

[21] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[22] Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2408–2415.

[23] SMHP Challenge Organization. 2018. Social Media Headline Prediction. https://social-media-prediction.github.io/PredictionChallenge

[24] SMP Challenge Organization. 2019. Social Media Prediction Challenge. http://smp-challenge.com

[25] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[26] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment. *IEEE Transactions on Image Processing* 27, 8 (2018), 3998–4011.

[27] Wen Wang and Wei Zhang. 2017. Combining multiple features for image popularity prediction in social media. In *ACM International Conference on Multimedia*. 1901–1905.

[28] Bo Wu. 2017. TPIC: A social media dataset for temporal popularity prediction. https://social-media-prediction.github.io/TPIC2017

[29] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, and Tao Mei. 2016. Time matters: Multi-scale temporalization of social media popularity. In *ACM International Conference on Multimedia*. 1336–1344.

[30] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, Huang Qiushi, Li Jintao, and Tao Mei. 2017. Sequential prediction of social media popularity with deep temporal context networks. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

[31] Bo Wu, Tao Mei, Wen-Huang Cheng, and Yongdong Zhang. 2016. Unfolding temporal dynamics: Predicting social media popularity using multi-scale temporal decomposition.. In *AAAI Conference on Artificial Intelligence*. 272–278.

[32] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. 2018. User-guided hierarchical attention network for multi-modal social image popularity prediction. In *International Conference on World Wide Web*. 1277–1286.