HyFea: Winning Solution to Social Media Popularity Prediction for Multimedia Grand Challenge 2020

Xin Lai*
East China Normal University
Shanghai, China
51184501121@stu.ecnu.edu.cn

Yihong Zhang*
East China Normal University
Shanghai, China
51184501185@stu.ecnu.edu.cn

Wei Zhang†
East China Normal University
Shanghai, China
zhangwei.thu2011@gmail.com

ABSTRACT

Social Media Popularity (SMP) prediction focuses on predicting the social impact of a given post from a specific user in social media, which is crucial for online advertising, social recommendation, and demand prediction. In this paper, we present HyFea, our winning solution to the Social Media Prediction (SMP) Challenge for multimedia grand challenge of ACM Multimedia 2020. To address the multi-modality and personality issues of this challenge, HyFea carefully considers multiple feature types and adopts a tree-based ensembling method, i.e., CatBoost, which is shown to perform well in prediction. Specifically, HyFea involves the features related to Image, Category, Space-Time, User Profile, Tag, and Others. We conduct several experiments on the Social Media Prediction Dataset (SMPD), verifying the positive contributions of each type of features.

CCS CONCEPTS

• **Information systems → Content analysis and feature selection**; *Personalization.*

KEYWORDS

Social Media Popularity Prediction; Feature Construction; Ensemble Learning

ACM Reference Format:

Xin Lai, Yihong Zhang, and Wei Zhang. 2020. HyFea: Winning Solution to Social Media Popularity Prediction for Multimedia Grand Challenge 2020. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3394171.3416273

1 INTRODUCTION

With the proliferation of online social media, the amount of user generated content is surging, which causes a heavy information overload problem. As such, automatically selecting potential popular information is critical for improving the users' experience, by delivering popular information to them. Among different types of user generated content, social multimedia posts are ubiquitous across different networking platforms. If the social impact (e.g., total clicks) of a newly emerging post could be accurately predicted, platforms can quickly locate posts which will be popular, even without knowing the early interactions of posts with a large number of users.

Due to the significance of social multimedia posts, the Social Media Prediction (SMP) Challenge for multimedia grand challenge of ACM Multimedia 2020 is hosted. The task of the challenge setting [1] is formulated as that: given a photo (a.k.a. post) from a user, the target is to automatically predict the popularity score of the photo, e.g., view count for Flickr. A large-scale, multi-faced dataset, i.e., SMPD (Social Media Prediction Dataset), is provided accompanied by the competition [2–4]. SMPD contains 486K social multimedia posts generated by 70K users, and some detailed information about each post, such as image content, image tags, title, released time, etc. Moreover, some historical features of users, such as the total number of images previously published by the user, are also given.

The main issues to be addressed for the SMP challenge lie in the following two aspects: (1) Multi-modality. Each post itself belongs to multimedia content, containing both visual and textual features, such as the visual content and the descriptions of posts. In addition, each post has some other information, including the categories and tags of posts. The feature complexity revealed in the above information should be comprehensively considered to boost the performance. (2) Personality. The users (a.k.a. publishers) of posts should be investigated. This is a key factor to understand the unique characteristics of user generated content compared to traditional image/text modeling without considering personalization. In reality, the role of users heavily determines the popularity of social posts. It is intuitive that if a user has published some popular posts in the past, the published posts in the future might be popular as well. This phenomenon could also be explained by the fact that different users have different degrees of authority in various aspects. It is therefore necessary to consider the personality of social posts, hoping to capture the informative clues from users for popularity estimation.

To address the above two major issues encountered in popularity prediction, we develop a hybrid feature based approach, named as HyFea. It consists of two indispensable steps towards winning this competition. The first is about tailored feature engineering, covering different categories of information. To be specific, we consider posts' metadata and user related features, and fuse them

*Both authors contributed equally.
†Corresponding author.

MM '20, October 12–16, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-7988-5/20/10…$15.00
https://doi.org/10.1145/3394171.3416273

together as input to regression models. The second is to choose a well-performed model to utilize the crafted features. Inspired by the strong ability of modeling categorical features while retaining the benefits of tree-based ensemble learning, CatBoost [5] is leveraged. We perform cross-validation to tune the hyper-parameters of the adopted model for achieving better performance. The comprehensive experiments, including ablation study, are conducted on SMPD to verify the contributions of each type of feature. In the end, local tests provide a reference to submit the predictions. We release the key source code[1] for further research.

2 RELATED WORK

In recent years, the academic community has devoted many efforts to investigate the task of popularity prediction for social media posts.

Existing research mainly focuses on three aspects of feature selection: image content, text information, and user information. Some studies [6, 7] demonstrate that visual information can affect the trend of posts. Some other studies show that textual information could also be modeled to infer the popularity of social posts [6, 8–10]. The pre-trained models such as Word2Vec, GloVe, and BERT are usually employed to extract effective feature representations of text. The combination of image and text features is also verified to be useful for this task [11]. Moreover, 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 [6, 9, 10]. In addition, some researchers explore to utilize temporal information to help popularity prediction [2, 8].

Predicting social media popularity is a category of continuous value regression tasks. Some studies use the linear regression [7] model, which is a fast and easy-to-use basic model that combines all features in a linear fashion. Support vector regression minimizes the total deviation of all samples from the hyperplane, and is often used in this task [12]. Random forest [13] is a tree-based Ensemble method making more accurate predictions than individual models. The trees in random forests are performed in parallel and output the mean prediction of the individual trees for regression training. The gradient boosting decision tree (GBDT) [14] fits the residual in each iteration and performs well in the regression task, but the traditional implementation method has a large space overhead and takes a long time. The authors [6, 9, 10] have proved the advantages of the gradient boosting decision tree through experiments. The proposal of XGBoost [15] realizes a large-scale parallel implementation without degrading performance. Subsequently, the newly emerged LightGBM and CatBoost have other improvements compared to XGBoost. Among them, LightGBM effectively improves the calculation efficiency of GBDT [16]. The study [5] leverages CatBoost that is claimed to be good or even better than XGBoost and LightGBM in some cases. It combines the new Boosting algorithm to solve the problem of overfitting, while enriching the feature dimensions. In addition, [17] builds neural network models to predict popularity and also obtains good results.

[1]https://github.com/runnerxin/HyFea

3 METHOD

3.1 Overview

In this section, we introduce the approach of which the main components are shown in Fig. 1. It mainly consists of two components, data feature extraction and CatBoost-based regression model. We first extract the feature from six categories of features, i.e., Image, Category, Time, User Profile, Tag, and Others. In what follows, we combine all the features to train the CatBoost model and predict the social media posts popularity.

3.2 Feature Construction

For each social post, some features can be obtained directly, while others need to be calculated in some ways. The detailed features are as follows:

- **Image:** We can extract the basic features of an image, i.e., “ImgLength”, “ImgWidth”, “Pixel”, and “ImgModel”. Among them, "ImgModel" involves four modes: P, l, RGB and CMYK.
- **Category:** There is a three-level category hierarchy for social media posts. Specifically, the three levels are named as “Category” (with 11 classes), “Subcategory” (with 77 classes), and “Concept” (with 668 classes), respectively.
- **Space-time:** We first extract geographical features of social posts, such as the basic longitude and latitude information. The granularity level of locations is also considered as features. Then we construct temporal features of social posts, such as “HourInDay”, “HourInWeek”, “DayInWeek”, “DayInMonth”, and “WeekInYear”.
- **User Profile:** Three basic features, i.e., “Uid”, “UserPostCount”, “PhotoCount”and “Ispro”, from each user are first constructed, representing the user’s ID, the total number of published posts, the total number of published photos, and whether the user is a professional member, respectively. Then we can use the temporal information related to the first published image and the first taken image by the user, including their absolute timestamps and the temporal difference compared to the posting time of the current image.
- **Tag:** For this feature aspect, we consider the type of media file, which covers ‘photo’ and ‘video’. We also get the features "TitleLen", "TitleNumber", "AlltagsLen", and "AlltagsNumber" from the title and tags of the posts, which represent the number of words and characters for the title and tags, respectively. In addition, in order to get the statistical and semantic features in the title and tags, we obtain the Term Frequency-Inverse Document Frequency (TF-IDF) features, Since the original feature dimension is too large, we adopt SVD to reduce the dimension to 20. Besides, we use the average word embedding (pre-trained by Glove) to represent the semantic information of the title and tag words.
- **Others:** As complementary to the above feature aspects, we consider the Boolean feature "Ispublic", which indicates the post is authenticated with read permissions. To fully capture the user characteristics, we crawl additional user information from user page indicated by “pathalias” and build the following features: “FollowingCount”, “FollowerCount”, “TotalPhoto”, “TotalGroup”, “TotalFaves”, “TotalGeotagged”,

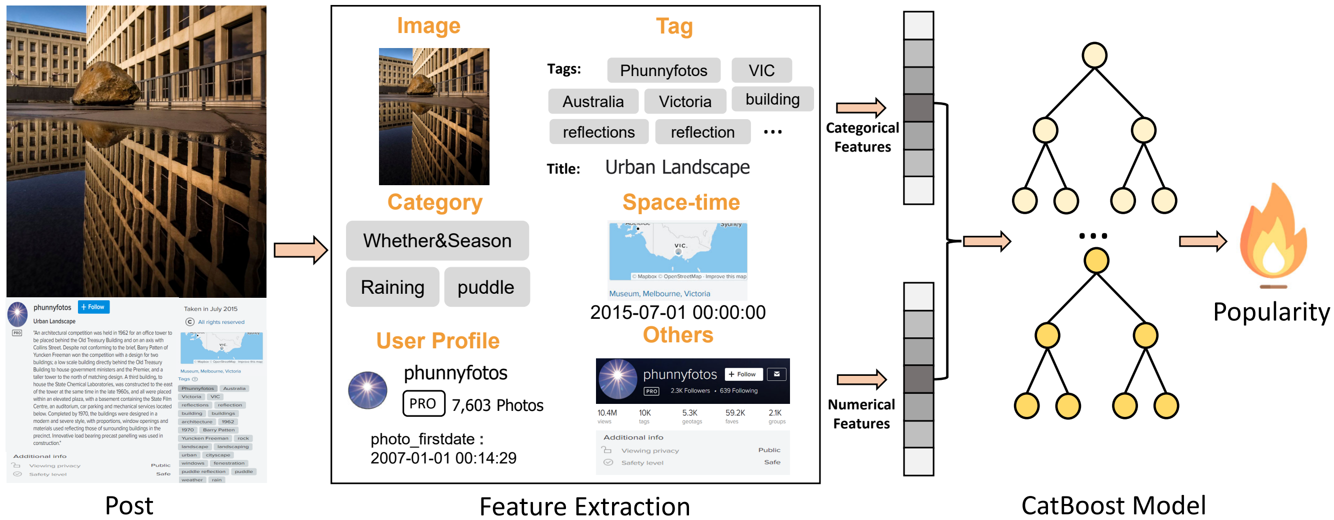

Figure 1: Illustration of the proposed method.

"TotalTags", and "TotalViews". Besides, we calculate some average features about users' images, e.g., "MeanFaves", "MeanTags", and "MeanView".

The above features are concatenated and fed into the selected model for training, which will be elaborated in the next section.

3.3 The CatBoost Model

We adopt CatBoost [5] as the regression model in this paper. It belongs to a tree-based gradient boosting algorithm. Compared to GBDT and XGBoost, CatBoost owns the intrinsic advantage of handling categorical features. There are some category features such as "Uid" in the constructed features. CatBoost first generates some statistics of category features, such as the occurrence frequency of a specific category feature, and then add parameters for generating new numerical features. More details about the mechanisms used in CatBoost can be referred to the paper [5].

4 EXPERIMENTS

4.1 Dataset

The Social Media Prediction Dataset (SMPD) is collected from Flickr, one of the largest photo-sharing platforms. This dataset contains 486k social multimedia posts that are generated by 70k users, and each post is associated with multiple types of information as mentioned before. We divide the data into training set and test set in a chronological order to gain model performance based on local testing.

4.2 Evaluation Metrics

To evaluate the performance of different methods, Spearman's Rho (SR) and mean absolute error (MAE) are adopted in our experiment, which are also leveraged to judge the performance of different teams online. SR measures the ranking correlation between the real popularity score y and the predicted popularity score $\widehat{y}$. Thus the larger the value is, the better the performance the corresponding method gains. SR is defined as follows:

$$SR = \frac{1}{n-1}\sum_{k=1}^{n}\left(\frac{y_k - \bar{y}}{\sigma_y}\right)\left(\frac{\widehat{y}_k - \bar{\widehat{y}}}{\sigma_{\widehat{y}}}\right), \tag{1}$$

where $\bar{y}$, $\bar{\widehat{y}}$ and σ_y, $\sigma_{\widehat{y}}$ denotes the mean and variance of corresponding real popularity score and the predicted popularity score, respectively. With respect to MAE, it measures the absolute difference between real and predicted popularity scores, which are given as follows:

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|\widehat{y}_k - y_k\right|. \tag{2}$$

In the above two equations, n denotes the number of testing instances.

4.3 Performance Comparison

To evaluate the performance of the proposed method, we choose some well-performed baselines for comparisons. The corresponding results are shown in Fig. 2. Among all the methods, HyFea and HyFea(K) are the CatBoost based methods and HyFea(K) additionally utilizes K-fold cross validation. By comparing all the methods, we first observe that the simple linear models, such as LR and Ridge, do not perform equally well compared to the ensemble baselines. Moreover, CatBoost achieves better performance than GBDT, XGBoost, and LightGBM, showing its advantages in the task setting. In particular, CatBoost with k-fold cross validation could further improve the performance slightly.

4.4 Importance of Features

In order to understand how different specific features contribute to the final performance, we calculate their importance weights based on CatBoost and show the top 20 most important features in Table 1. As we can see, "Uid" is the most important feature, which

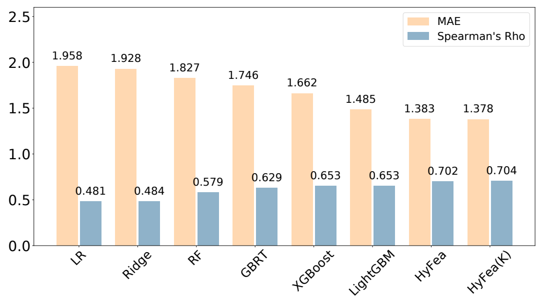

Figure 2: Performance of different models.

Table 1: Importance of different features.

Rank	Features	Imp	Rank	Features	Imp
(1)	Uid	14.411	(11)	TotalGroup	0.633
(2)	MeanView	7.849	(12)	TagsGloveVec282	0.488
(3)	Concept	4.163	(13)	PhotoCount	0.484
(4)	UserPostCount	3.951	(14)	Subcategory	0.471
(5)	AlltagsNumber	2.350	(15)	TotalTags	0.447
(6)	FollowerCount	1.986	(16)	TagsGloveVec224	0.443
(7)	MeanTags	1.421	(17)	TagsTfidfSvd16	0.439
(8)	AlltagsLen	1.405	(18)	HourInWeek	0.436
(9)	TotalPhoto	0.964	(19)	TotalViews	0.425
(10)	WeekInYear	0.646	(20)	HourInDay	0.418

is in line with the expectation because different users have different degrees of popularity, i.e., stars versus ordinary people. Moreover, we find that many user relevant features in the feature type "Others" play an important role, such as "Meanview", "FollowerCount", and "TotalPhoto", which aims at characterizing the authority of users as well. In total, the above phenomena demonstrate the significance of considering the personality of social media posts in popularity prediction.

4.5 Ablation Study

In this part, we further investigate the impact of different types of features through ablation study, wherein each time we remove one type of features from our method HyFea. The results are shown in Table 2, from which we have the following observations:

- Each type of feature has a positive contribution to the performance in most cases.
- "Category" and "Others" seem to have a larger influence on the performance.
- "Space-time" has a relatively small contribution. However, some features from this feature group occur in the top 20 most important features. This phenomenon reveals that there are correlations between some features. Thus combining features together might not achieve significantly better results.

Table 2: Ablation study w.r.t. the adopted features.

Methods	MAE	Spearman's Rho
HyFea	1.383	**0.702**
w/o Image	1.388	0.698
w/o Category	1.802	0.638
w/o Space-time	**1.381**	0.698
w/o User Profile	1.397	0.699
w/o Tag	1.411	0.685
w/o Others	1.610	0.620

Table 3: Performance of some alternatives.

Methods	MAE	Spearman's Rho
HyFea	**1.383**	**0.702**
w/o GloveVec	1.434	0.674
w/o TfidfSvd	1.384	0.700
w/ $\oplus$ ResNetImage	1.390	0.698
w/ $\oplus$ UserDes	1.402	0.697
w/ $\oplus$ LocationDes	1.387	0.699

4.6 Alternative Experiment

In this section, we verify some vectorized features in detail, including "GloveVec" (embeddings gotten by Glove), "TfidfSvd" (low-dimensional representations of Tfidf features), "ResNetImage" (image representations by ResNet), "UserDes", and "LocationDes" (feature vectors of the description of users and locations).

The results are shown in Table 3, from which we can observe that: 1) "GloveVec" is more effective than "TfidfSvd" for the task. 2) Using image feature vectors obtained by other deep learning models might not improve the performance. 3) The feature vectors about the description of users and locations could promote the improvement of results.

5 CONCLUSION AND FUTURE WORK

In this paper, we present our solution to the Social Media Prediction Challenge. Six types of features are constructed based on social media posts, aiming to capture the multi-modality and personality of them. CatBoost is leveraged due to its strong power of handling categorical features. We have conducted extensive experiments on the official SMPD2020 dataset and achieve the best performance, which verifies the effectiveness of the proposed approach.

In future work, we aim to improve the solution from the following two aspects. First, an end-to-end deep learning approach like UHAN [11] might be profitable to integrate the multiple types of features used throughout this paper, whereas our attempts on the SMPD dataset show they are computationally expensive. As such, devising an efficient approach is indispensable to the whole. Second, some additional advanced feature representation approaches, such as BERT [18], could be considered and fed into the adopted classification model.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (No. 61702190 and No. U1609220) and the foundation of Key Laboratory of Artificial Intelligence, Ministry of Education, P.R. China.

REFERENCES

[1] Bo Wu, Wen-Huang Cheng, Peiye Liu, Bei Liu, Zhaoyang Zeng, and Jiebo Luo. Smp challenge: An overview of social media prediction challenge 2019. In *ACM MM*, 2019.

[2] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, Huang Qiushi, Li Jintao, and Tao Mei. Sequential prediction of social media popularity with deep temporal context networks. In *IJCAI*, pages 3062–3068, 2017.

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

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

[5] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: Unbiased boosting with categorical features. In *NIPS*, page 6639–6649, 2018.

[6] Junhong Chen, Dayong Liang, Zhanmo Zhu, Xiaojing Zhou, Zihan Ye, and Xiuyun Mo. Social media popularity prediction based on visual-textual features with xgboost. In *ACM MM*, page 2692–2696, 2019.

[7] Jinna Lv, Wu Liu, Meng Zhang, He Gong, Bin Wu, and Huadong Ma. Multi-feature fusion for predicting social media popularity. In *ACM MM*, page 1883–1888, 2017.

[8] Wen Wang, Wei Zhang, Jun Wang, Junchi Yan, and Hongyuan Zha. Learning sequential correlation for user generated textual content popularity prediction. In *IJCAI*, pages 1625–1631, 2018.

[9] Ziliang He, Zijian He, Jiahong Wu, and Zhenguo Yang. Feature construction for posts and users combined with lightgbm for social media popularity prediction. In *ACM MM*, page 2672–2676, 2019.

[10] Peipei Kang, Zehang Lin, Shaohua Teng, Guipeng Zhang, Lingni Guo, and Wei Zhang. Catboost-based framework with additional user information for social media popularity prediction. In *ACM MM*, page 2677–2681, 2019.

[11] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. User-guided hierarchical attention network for multi-modal social image popularity prediction. In *WWW*, pages 1277–1286, 2018.

[12] Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn van Dolen. Multimodal popularity prediction of brand-related social media posts. In *ACM MM*, page 197–201, 2016.

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

[14] Wen Wang and Wei Zhang. Combining multiple features for image popularity prediction in social media. In *ACM MM*, page 1901–1905, 2017.

[15] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *SIGKDD*, page 785–794, 2016.

[16] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *NIPS*, pages 3146–3154. 2017.

[17] Keyan Ding, Ronggang Wang, and Shiqi Wang. Social media popularity prediction: A multiple feature fusion approach with deep neural networks. In *ACM MM*, page 2682–2686, 2019.

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL*, pages 4171–4186, 2019.