Classification and Evaluation for Microblog Popularity Prediction

Gang Ji¹, Yuxuan Zhu¹*, Yukai Niu¹ and Kai Hu²

¹ Institute of Systems Engineering, Academy of Military Science, Beijing, 100083, China
² School of Microelectronics, Tianjin University, Tianjin, 300072, China
*Corresponding author’s e-mail: zhuyx16@tsinghua.org.cn

Abstract. In recent years, with the rapid development of the Internet, especially the mobile Internet, social networks have entered the stage of vigorous development and become one of the main sources of information. User-generated contents (UGC) on social platforms can spread information along social networks at an astonishing speed. Existing literature has proposed many prediction methods for the popularity prediction on social networks. This paper presents a classification and establishes a unified evaluation framework of popularity prediction methods for microblogs. More specifically, we divide these mainstream prediction methods into four types: feature based methods, time series methods, collaborative filtering methods and deep learning methods and conduct experiments on the real-world weibo data using these methods to predict. Finally, according to four indicators, including accuracy, efficiency, robustness and bias, we evaluate and compare the methods. Based on the prediction and evaluation results, this paper summarizes and draws the following research conclusions: (1) The deep learning method has the characteristics of high accuracy, high robustness and low bias. The DeepFM method, one of the deep learning methods, performs better than the other three prediction methods when using temporal data as its input. (2) The feature based methods only using temporal features are basically consistent with those using all available features, indicating that the temporal feature has strong prediction power. Therefore, the ‘peeking’ strategy that monitors the early response of users in the initial period after the items are posted is effective. Additionally, the predictive power of temporary features can be further amplified in time series methods and deep learning methods. (3) Due to the sparse user-item interaction in social networks, the accuracy and efficiency of collaborative filtering methods are low, which makes it impossible to predict the popularity of items in social networks well.

1. Introduction

For the past few years, the rapid development of social networks which relies on the Internet has greatly changed people's way of life and communication. We define the user-generated contents on social platforms as items. They can spread along social networks and have a certain impact. How do items gain popularity through the information dissemination has become a hot topic for researchers and marketers.

Researchers have done abundant work on prediction, since the popularity prediction for social network information based on big data has a lot of application scenarios. When the items are generated by official marketing accounts such as companies, bloggers and other marketers, the popularity prediction of items can help them find potential hot topics, create popular contents, improve marketing strategies, and achieve more effective marketing. In addition, predicting whether an item will become
popular and analyzing what contributes to an item popularity are useful as well for social network platforms such as Weibo. If they are able to predict which content will become popular in advance, then they can come up with hot topics. In the future, with the popularity of mobile Internet, the promotion of people’s online content consumption demand and the explosive growth of network information, it is expected that the use of social network will be more in-depth in people’s daily life, and the popularity prediction of social network information will play a greater role.

Due to the wide application prospect of popularity prediction and the availability of microblog and twitter data, many researchers have conducted research on it and proposed a series of prediction methods or models [13], and they are considering this problem either as regression [14] or classification [15] task. These methods perform well under their respective assumptions, but none of them provides whether their prediction methods can apply to other assumptions or under other circumstances. Aimed at exploring which model or method can make a more accurate, rapid and stable popularity prediction of social network information, we will sort out and summarize the mainstream and latest prediction methods. According to the prediction mechanism, they are divided into four categories: feature based methods, time series methods, collaborative filtering methods and deep learning methods. Through performing and testing the prediction methods under the unified framework, we can make a fair and objective comparison of the prediction methods and learn about the best prediction mechanism.

We uniformly evaluate 14 proposed prediction methods, including 6 feature based methods, 2 time series methods, 3 collaborative filtering methods and 3 deep learning methods. Our dataset is composed of a Weibo dataset with 0.3 million microblogs and 1.8 million users. We compare the methods in terms of accuracy, efficiency, robustness and bias. The results show that the deep learning method performs the best among the four methods for predicting the popularity of social items. It beats the other three prediction methods with higher accuracy, higher robustness and lower bias. In addition, the temporary feature shows strong prediction power and the power can be further amplified in time series methods. Collaborative filtering methods have a low prediction accuracy due to the sparsity problem.

Our main contributions can be summarized as follow:

- This paper constructs a systematic overview and classification of proposed popularity prediction methods on social networks, categorizing them into feature based methods, time series methods, collaborative filtering methods and deep learning methods. In this process, we analyze features of these methods and relationships between them.
- We compare deep learning with traditional methods for the first time, and the results show that deep learning methods have significant advantages in addition to efficiency.
- We build an evaluation framework which includes accuracy, efficiency, robustness and bias so that it can assess the prediction methods more comprehensively, compared with the traditional single standard evaluation.

2. Classification of popularity prediction methods

In this section, we will introduce proposed popularity prediction methods on social networks from four categories: feature based methods, time series methods, collaborative filtering methods and deep learning methods.

2.1. Feature Based Methods

There exists a series of factors in an item such as type of the content, the number of followers and followings of the poster, the personal information of the user and so on, which directly affect how far the content generated by the user can go [10]. These factors are intertwined with each other and influence users' retweeting behavior. Based on this idea, feature based methods try to extract and select useful features during the spread process and then train the traditional machine learning models to make the popularity prediction [11].

The extracted features usually fall into four categories: content-related features, user-related features, time-related features and network-related features. These features play a role during the whole
information diffusion process, including posting, reacting and retweeting. These features cover all components in social networks, including users, items and time series.

Feature based methods have some drawbacks. Firstly, these methods highly depend on feature engineering, causing great time and space consumption. Besides, it still needs further feature selection manually. Secondly, the feature selection requires much expert knowledge. If the selected features are not appropriate enough for the task, the model will perform badly. Thirdly, the transfer ability of the model is low, since different tasks require different features to characterize them.

2.2. Time Series Based Methods

Time series based methods mainly focus on the time information of the item in the social network. In time series based models, every cascade can be broken up into single retweet action that happen one by one, which can be described by a point process. The whole information diffusion process can be represented by a sequence of retweet times. Our goal is to utilize the existing time sequence to model the behavior dynamics in the social network and then make the prediction.

However, time series methods have some shortcomings. The models rely on strong hypotheses about the formation of popularity. If the hypotheses do not fit the dataset, it will cause poor performance in prediction. What is more, the models only consider time series information. For items with no retweet action observed, the time series models will not be able to predict them.

Zhao et al. [1] proposed a statistical model named SEISMIC based on the theory of self-exciting point process. SEISMIC model is an extension of Hawkes model, which assumes that the previous retweet actions will influence the future evolution of the process. It can successfully characterize the “rich-get-richer” phenomenon. Besides, the model takes human reaction times and post infectiousness into consideration.

Yu et al. [2] proposed NEtworked WEibull Regression model (NEWER) to model the behaviour dynamics in the social network and make the prediction of future popularity. The model is based on Weibull distribution, which can better preserve the characteristics of behavioral dynamics than Exponential and Rayleigh distributions. The model also applies strong correlations between the parameters of a node’s behavioral dynamics and its neighbor nodes behavioral features for parameter learning. In addition to the maximum likelihood estimation term, we also assume the parameters of a node can be regressed by the behavioral features of its neighbor nodes and thus impose networked regularizers to improve the interpretability and generality of the model.

2.3. Collaborative Filtering Methods

Different from feature based methods, this method mainly focuses on user interaction. Given user $u_i$ whose retweet actions have known, Matrix Factorization (MF) uses the similarity of two users to predict whether the user $u_j$ whose behaviors are unknown would retweet the item or not [12]. In this problem, there is only binary data between the users and items which records whether a user retweeted the item.

Salakhutdinov et al. [4] optimized the basic Matrix Factorization method and came up with the Probabilistic Matrix Factorization (PMF), which performs well on large, sparse, and very imbalanced Netflix dataset.

Jiang et al. [3] tended to solve the first defect by proposing a novel item clustering based matrix factorization model, which is Centroid-based Regularization Prediction Model (CRPM). It is likely that clustering information would improve the result of prediction since the users may be willing to retweet the items which are relative to their interests. To make similar items in one cluster close to each other in the latent feature space, they propose another item clustering based matrix factorization model, which is Individual-based Regularization Prediction Model (IRPM).

2.4. Deep Learning Methods

The deep learning method uses the deep neural network (DNN) as the core part for prediction. The deep learning method performs well in the application of the recommendation system. Since both popularity prediction and recommendation systems are microcosmic predictions of whether users will be interested
in items, the following methods are also included in this paper. According to the experimental results, the deep learning method used in this paper is better than the traditional method in popularity prediction.

In the recommendation system, the prediction of click-through rate (CTR) is a very important part. In CTR prediction, not only the single feature, but also combinations between features need to be considered. For feature combinations, there are two main types of mainstream practices: Factorization Machines (FM) series and the tree series.

Rendle [5] proposed the FM in order to solve the problem of how the features are combined in the case of data sparseness. The FM is defined as:

$$y_{FM} = \omega_0 + \sum_{i=1}^{n} \omega_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \omega_{ij} x_i x_j$$

where $x_i$ is the element of the feature vector $x$, $\omega_0$ is the global bias, $\omega_i$ denotes the weight of the $i$-th feature, and $\omega_{ij}$ denotes the weight of interaction between the $i$-th and $j$-th feature.

Guo et al. [6] proposed DeepFM to learn both low- and high-order feature interactions. DeepFM includes FM component and deep component, so its prediction is:

$$y_{DeepFM} = \text{sigmoid}(y_{FM} + y_{DNN})$$

where $y_{DeepFM}$ denotes the predicted of DeepFM, $y_{FM}$ denotes the output of FM component, and $y_{DNN}$ denotes the output of deep component. The FM component is a factorization machine. The deep component is a feed-forward neural network. DeepFM uses an embedding layer which is the input of both FM and deep component to compress the input vector to a low-dimensional, dense real-value vector.

Xiao et al. [7] presented the Attention Factorization Machine (AFM) which improves the representation ability and the interpretability of the FM model. They found that when making predictions, FM fixes a feature to a specific vector, and when the feature intersects with other features, the calculation would be done with the same vector. This is not reasonable, because different features of interactions have different levels of importance. Therefore, they propose the AFM model combining the attention component. AFM can be divided into five layers: sparse input layer, embedding layer, pair-wise interaction layer, attention-based pooling layer, and prediction layer.

He et al. [8] presented the Neural Factorization Machine (NFM) to prediction under sparse settings. The difference between NFM and FM is the third term. The third term in FM is the second-order feature interactions, whereas the third term in NFM is a multi-layered feed-forward neural network which models second-order and high-order feature interactions. NFM includes five layers: input feature vector layer, embedding layer, b-interaction layer, hidden layer, and prediction layer.

3. Experiments

3.1. Dataset Description

We use the Weibo dataset for evaluating the prediction ability of models, which is a very popular social platform. Weibo has the feature of asymmetry friendship, which leads that the items retweeted on Weibo are visible in the followers' feeds without the permission of publisher.

Zhang et al. [9] crawled the dataset from Weibo. At first, they chose 100 users randomly as seeds, and then collected their followees and followers. In the end, they got 1.7 million users and 0.3 billion following relationships in total, and each user has about 200 followees, 1,000 most recent microblogs including tweets and retweets. There are about 1 billion microblogs in total in the dataset. All the users' profiles contain name, gender, verification status, bi-following, followers, followees, and microblogs.

They selected 300,000 popular microblog diffusion episodes from the dataset to study users' retweet behaviors. The original item has been retweeted about 80 times and all of the retweets are included in each diffusion episode.

3.2. Experiment Setup

For feature based methods we choose Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), k-Nearest Neighbors (KNN), Naive Bayesian (NB) and Multi-Layer Perceptron (MLP) as the
classifier. For feature based methods and collaborative filtering methods, the training set and the testing set are from a random 80% and 20% split of the datasets. For time series methods, we use the timestamps of the retweet actions up to the observation time as input and finally predict the total number of retweets. Compared with feature based methods, the deep learning methods use the same feature input as feature based methods. The deep learning methods use the first 20 repost times and the sum of remaining reposts which belong to the observation time as input, when contrasting experiment with time series methods. Because the input of collaborative filter methods is matrix, the deep learning methods use the elements of the matrix as prediction objects and corresponding rows’ and columns’ features as input. For the classification methods in feature based methods and time series methods, we consider an item popular, if the count of reposts is more than 50. If the count of reposts is less than 10, the item will be discarded. We chose 50 as the threshold of the repost count for the preliminary screening of models.

4. Evaluation

4.1. Accuracy
We use F1 score and mean square error (MSE) as evaluation metrics. The accuracy of classification on the feature based methods is listed in Tables 1 and 2, where Table 1 shows the model with all features input and Table 2 shows the model with temporal features input. We tested data of 2, 4, 6 and 12 hours in experiments based on all features and only based on temporal features respectively. Additionally, data of 0 hours is considered in the experiments based on all features, which means all available features except temporal features are used. As shown in the two tables, we can see that the classification accuracy of the three types of input rank from high to low as the order: all feature input, temporal feature input, all feature input without temporal feature input, which also confirms the previous statement that temporal features have the greatest predictive power. Removing the temporal features would make the prediction worse, while only using the temporal features would certainly not reduce the classification accuracy of the model. The reason may be that only using the temporal feature eliminates some redundant features. For all-feature input, LR, NB and MLP have the highest classification accuracy, while that of KNN is the lowest. For time-feature input, LR and NB have the highest classification accuracy, while DT gets the lowest one.

| Time | Metrics | DT   | LR   | NB   | RF   | MLP  | KNN  |
|------|---------|------|------|------|------|------|------|
| 0h   | F1      | 0.5921 | 0.6614 | **0.6786** | 0.5944 | 0.6743 | 0.5576 |
|      | MSE     | 0.4599 | **0.3927** | 0.4342 | 0.4434 | 0.4832 | 0.4868 |
| 2h   | F1      | 0.6764 | 0.7152 | **0.7198** | 0.7003 | 0.7175 | 0.6011 |
|      | MSE     | 0.3648 | 0.4124 | 0.4131 | **0.3193** | 0.4267 | 0.4715 |
| 4h   | F1      | 0.6741 | 0.7176 | 0.7198 | 0.7005 | **0.7247** | 0.5994 |
|      | MSE     | 0.3669 | 0.4098 | 0.4131 | **0.3163** | 0.4325 | 0.4738 |
| 6h   | F1      | 0.6758 | 0.7184 | 0.721 | 0.7017 | **0.7216** | 0.5989 |
|      | MSE     | 0.3629 | 0.408 | 0.409 | **0.3147** | 0.4219 | 0.4731 |
| 12h  | F1      | 0.6792 | 0.7191 | 0.7198 | 0.702 | **0.7243** | 0.597 |
|      | MSE     | 0.3612 | 0.4091 | 0.4132 | **0.3164** | 0.4232 | 0.4768 |
Moreover, we make a contrast between time series based methods and deep learning methods. To ensure comparability, we use the same temporal data as their input. Table 3 shows their classification accuracy. In order to compare with feature based methods, we also select 2, 4, 6 and 12 hours as the time nodes. It can be seen from the table that all the models except for NEWER perform well, and they have much higher classification accuracy than the feature based methods on all times nodes of prediction. SEISMIC performs much better than NEWER, which can be attributed to their differences in input. NEWER only has one input that is the retweeting time of each retweet within the prescribed time, while SEISMIC has one more input, the number of followers of retweet users. According to the table, SEISMIC has a higher precision but a lower recall, while DeepFM and NFM are just the opposite, with higher recall and lower precision. The prediction accuracy of NFM with the data in 2, 4 and 6h is very high, but the prediction accuracy is almost not improved with the data increasing in the latter time nodes.

Moreover, we make a contrast between time series based methods and deep learning methods. To ensure comparability, we use the same temporal data as their input. Table 3 shows their classification accuracy. In order to compare with feature based methods, we also select 2, 4, 6 and 12 hours as the time nodes. It can be seen from the table that all the models except for NEWER perform well, and they have much higher classification accuracy than the feature based methods on all times nodes of prediction. SEISMIC performs much better than NEWER, which can be attributed to their differences in input. NEWER only has one input that is the retweeting time of each retweet within the prescribed time, while SEISMIC has one more input, the number of followers of retweet users. According to the table, SEISMIC has a higher precision but a lower recall, while DeepFM and NFM are just the opposite, with higher recall and lower precision. The prediction accuracy of NFM with the data in 2, 4 and 6h is very high, but the prediction accuracy is almost not improved with the data increasing in the latter time nodes.

| Time | Metrics | SEISMIC | NEWER | DeepFM | AFM | NFM |
|------|---------|---------|-------|--------|-----|-----|
| 2h   | F1      | 0.7769  | 0.6019 | **0.7919** | 0.7609 | 0.7884 |
|      | MSE     | **0.1554** | 0.252 | 0.2468 | 0.306 | 0.2613 |
| 4h   | F1      | 0.8104  | 0.6496 | 0.8173 | 0.7861 | **0.8454** |
|      | MSE     | **0.1348** | 0.2304 | 0.2209 | 0.2626 | 0.1902 |
| 6h   | F1      | 0.8249  | 0.6792 | 0.8419 | 0.7969 | **0.8627** |
|      | MSE     | **0.1237** | 0.2168 | 0.1881 | 0.2432 | 0.1672 |
| 12h  | F1      | 0.8695  | 0.7103 | **0.8958** | 0.7912 | 0.8914 |
|      | MSE     | **0.0936** | 0.205 | 0.1209 | 0.2611 | 0.136 |

Moreover, we make a contrast between time series based methods and deep learning methods. To ensure comparability, we use the same temporal data as their input. Table 3 shows their classification accuracy. In order to compare with feature based methods, we also select 2, 4, 6 and 12 hours as the time nodes. It can be seen from the table that all the models except for NEWER perform well, and they have much higher classification accuracy than the feature based methods on all times nodes of prediction. SEISMIC performs much better than NEWER, which can be attributed to their differences in input. NEWER only has one input that is the retweeting time of each retweet within the prescribed time, while SEISMIC has one more input, the number of followers of retweet users. According to the table, SEISMIC has a higher precision but a lower recall, while DeepFM and NFM are just the opposite, with higher recall and lower precision. The prediction accuracy of NFM with the data in 2, 4 and 6h is very high, but the prediction accuracy is almost not improved with the data increasing in the latter time nodes.

| Method | PMF | CRPM | IRPM | DeepFM | AFM | NFM |
|--------|-----|------|------|--------|-----|-----|
| F1     | 0.8209 | 0.8299 | 0.8596 | **0.8658** | 0.8397 | 0.8541 |
| MSE    | 0.0984 | 0.0899 | 0.0786 | **0.0710** | 0.0858 | 0.0768 |
The classification accuracy of the collaborative filtering models and the deep learning models are listed in Table 4. In the experiment of collaborative filtering model, in order to maximize the accuracy of collaborative filtering, we choose 100 epochs. The accuracy of the collaborative filtering models is relatively high and it is almost the same with the deep learning methods. DeepFM has the highest accuracy except recall rate, and its recall rate is only a little lower than IRPM.

4.2. Robustness

In this section we will discuss the stability of predictions. In the application, we tend to choose a prediction method that not only performs well on average but does not cause serious misjudgment.

![Figure 1. The robustness of methods (APE>150%)](image)

We use APE > 150% as the evaluation metric. Figure 1 shows that NEWER basically has no errors with APE greater than 150%, indicating that the prediction of NEWER is relatively stable though NEWER is less accurate than other methods. The robustness of the SEISMIC, DeepFM and AFM is getting better over time, among which the SEISMIC remains robust all the time. However, SEISMIC has some infinite predictions. DeepFM has a high error rate at the beginning but it declines rapidly over time. The robustness of DeepFM in 24h performs the best. The variation curve of robustness of NFM is similar to the accuracy of NFM. That is robustness improves in the first three time nodes and then fluctuates slightly.

4.3. Efficiency

We tested the program training time for the feature based and collaborative filtering methods as shown in Figure 2. Where the left figure of Figure 2 is the training time of feature based methods and corresponding deep learning methods. Among them, feature based methods are the time of one training, and deep learning is the time of one epoch in one training. It can be seen from the figure that the training time of deep learning methods is much longer than feature based methods. The right figure of Figure 2 is the training time of collaborative filtering methods and deep learning methods. The time listed is an epoch of collaborative filtering methods or deep learning methods. It can be seen from the figure that the epoch time of deep learning methods is much longer than collaborative filtering methods, but collaborative filtering methods uses 100 epochs in the experiment, and deep learning methods uses only 10 epochs to predict which have a good performance like collaborative filtering methods. Due to the difference in the number of epochs, the real training time difference will be reduced. The time series algorithm takes a long time, especially the SEISMIC training speed which is very slow. The training time of NEWER and deep learning are not very different. The time of prediction of all methods is generally small and the differences among them are not large.
4.4. Bias

By plotting the predicted size of the cascade (y-axis) versus the actual size (x-axis), as shown in Figure 3, we find that some prediction methods show bias. The accuracy of the time series methods has reached a high level by 6 hours and it has basically no improvements later. Therefore, we list the prediction scatter diagram of each algorithm for 6 hours, and randomly select 10,000 pairs of predicted value and true value for each algorithm. As shown in the figures, the predictions of NEWER are relatively pessimistic, and the prediction of AFM is rather pessimistic without any predicted value greater than 500. This shows that the AFM method fails and it is not suitable for using time series data as input. While SEISMIC, DeepFM, and NFM do not display any tendency. The variance of NFM is larger than that of SEISMIC and DeepFM, and the prediction results are poor when the prediction cascade is relatively high. After the actual size is greater than 2000, basically none of the prediction points of NFM are accurate. Whereas DeepFM with the smallest variance and SEISMIC can also perform well when the actual size is large.
5. Conclusion
In this paper, we compare methods for popularity prediction of single user-generated contents on social networks, through establishing a classification and evaluating the prediction results under a unified testing framework. For the classification, we divided the mainstream and latest prediction methods into four categories according to their prediction mechanism: feature based methods, time series methods, collaborative filtering methods and deep learning methods. Using these methods, we perform experiments on real world Weibo dataset under the same assumptions. Finally, we evaluate the experimental results from four perspectives: accuracy, robustness, efficiency and bias and analyze the prediction methods on social networks respectively.

The results show that DeepFM performs the best among the four types of methods for predicting the popularity of social items. It beats the other prediction methods by different degrees on the four main evaluation indexes. Since deep learning is able to build a deep neural network with a hierarchical structure, it can better fit the microblog data with both individual and time characteristics. Deep learning methods are specifically used to simulate human behaviors more intelligently, which decides in the face of high-dimensional data with complex structure and large samples, they have stronger ability to interpret and predict human social behaviors on social networks. In addition, it can be neatly combined with other algorithms such as FM to overcome some defects of traditional prediction methods and realize optimization based on existing models. Therefore, suitable deep learning methods are supposed to have a better performance in accuracy compared with the other three prediction methods. Experiments show that FM and DNN in parallel structure is more suitable for popularity prediction, while serial methods may not be suitable. In addition, because deep learning methods have more complex training and prediction mechanisms, it takes much longer in total time consumption than other prediction methods.

Through the analysis of the experimental results, it can be seen that the temporal features have the strongest prediction ability, which is consistent with previous literature. Using time series methods, temporal features are utilized to the maximum extent, so it has higher prediction accuracy. Whereas, in feature based methods, the temporal features of data are only taken as one of the input data features. The absence of important information leads to the fact that feature based methods could not perform as well as the time based methods. In the case of the absence of time data, it can also make predictions based on individual data when a microblog has not been adopted. Therefore, we recommend feature based methods when the temporal data are absent. Collaborative filtering methods are relatively inefficient to predict social filtering trends. It is mainly because when user-item interactions are sparse, collaborative filtering methods get too little information to analyze and predict. Unfortunately, real social networks operate in this case.

References
[1] Zhao, Q., Erdogdu, M. A., He, H. Y., Rajaraman, A., & Leskovec, J. (2015). Seismic: A self-exciting point process model for predicting tweet popularity. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1513-1522.
[2] Yu, L., Cui, P., Wang, F., Song, C., & Yang, S. (2015). From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics. In 2015 IEEE International Conference on Data Mining. pp. 559-568.
[3] Jiang, B., Liang, J., Sha, Y., & Wang, L. (2015). Message clustering based matrix factorization model for retweeting behavior prediction. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. pp. 1843-1846.
[4] Mnih, A., & Salakhutdinov, R. R. (2007). Probabilistic matrix factorization. Advances in neural information processing systems, 20: 1257-1264.
[5] Rendle, S. (2010). Factorization machines. In 2010 IEEE International Conference on Data Mining. pp. 995-1000.
[6] Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247.
[7] Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., & Chua, T. S. (2017). Attentional factorization machines: Learning the weight of feature interactions via attention networks. arXiv preprint arXiv:1708.04617.

[8] He, X., & Chua, T. S. (2017). Neural factorization machines for sparse predictive analytics. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval. pp. 355-364.

[9] Zhang, J., Liu, B., Tang, J., Chen, T., & Li, J. (2013). Social influence locality for modeling retweeting behaviors. In IJCAI. 13: 2761-2767.

[10] Matsubara, Y., Sakurai, Y., Prakash, B. A., Li, L., & Faloutsos, C. (2012). Rise and fall patterns of information diffusion: model and implications. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 6-14.

[11] Ma, Z., Sun, A., & Cong, G. (2012). Will this# hashtag be popular tomorrow?. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. pp. 1173-1174.

[12] Ma, H., Yang, H., Lyu, M. R., & King, I. (2008). Sorec: social recommendation using probabilistic matrix factorization. In Proceedings of the 17th ACM conference on Information and knowledge management. pp. 931-940.

[13] Gao, X., Cao, Z., Li, S., Yao, B., Chen, G., & Tang, S. (2019). Taxonomy and evaluation for microblog popularity prediction. ACM Transactions on Knowledge Discovery from Data (TKDD), 13(2): 1-40.

[14] Cao, Q., Shen, H., Gao, H., Gao, J., & Cheng, X. (2017). Predicting the popularity of online content with group-specific models. In Proceedings of the 26th International Conference on World Wide Web Companion. pp. 765-766.

[15] Liao, D., Xu, J., Li, G., Huang, W., Liu, W., & Li, J. (2019). Popularity prediction on online articles with deep fusion of temporal process and content features. In Proceedings of the AAAI Conference on Artificial Intelligence. 33: 200-207.