Information cascade prediction based on T-DeepHawkes model

Shijie Wang, Lihua Zhou* and Bing Kong

School of Information, Yunnan University, 650500, Kunming, China

*E-mail: lhzhou@ynu.edu.cn

Abstract. It is an important research point of social network analysis to predict future propagation range of information based on its early propagation characteristics. DeepHawkes model combines the Hawkes model in the traditional methods with deep learning, which not only inherits the high interpretability of Hawkes model, but also contains the high prediction ability of deep learning. However, the DeepHawkes model ignores the effect of the text content of the information on the propagation. Therefore, on the basis of DeepHawkes model, this paper further considers the influence of the text content on the diffusion, and proposes a T-DeepHawkes model which merges a topic classification model into DeepHawkes model. The experimental result shows that T-DeepHawkes model has a better prediction accuracy.

1. Introduction
The emergence of Twitter, Facebook, WeChat and, Sina Weibo and other online social platforms has greatly promoted the generation and transmission of information, which profoundly changed the way of information transmission. Such a huge amount of information contains a wealth of knowledge, such as users’ preferences, retweeting behavior and the relationships between the users. It has extremely wide application value for mining of these knowledge in viral marketing, online advertising, information recommendation, rumor control and other aspects. So, it has very important research significance [1]. However, as social platforms are usually large-scale open systems, they will be affected by external factors [2], such as network topology, concern relationship, user interest, release time, privacy protection, etc. Due to the large scale of network users, the dynamic diffusion of information, the greater randomness of the transmission path, and the uneven "popularity" of different information, the accuracy of information cascade prediction is full of challenging [3].

In the prediction problem of information cascade, the current methods are mainly divided into feature-based methods [1-7] and generation methods [8-10]. The feature-based methods firstly extract various features of information based on existing experience, and then predict the popularity of future information through training regression/classification model. The generation approach dedicates to describe and model the process of an information notice, and make people easy to understand the underlying mechanisms governing the popularity dynamics of information [2]. However, the predict power performance of generation methods are not ideal, because they are not optimized by popularity prediction. So Cao et al. proposed DeepHawkes model by combining the high interpretability of Hawkes model with the high predictability of deep learning [2]. But DeepHawkes model aims to model the process of information transmission, ignoring the influence of the text content on message
transmission. In fact, the text content of the message is also an important factor in the prediction problem of message transmission.

So in this paper, we propose T-DeepHawkes (Topic-DeepHawkes) model which considers the effect of text content by a topic classification model. T-DeepHawkes model contains the topic of the messages and merges the influence of cascade and text content. The main work of this paper including: (1) The T-DeepHawkes model considering the topic is proposed, and the classification model is used to extract the topic of text content of every message, and the deep learning model is used to learn the representation vector of the topic. T-DeepHawkes model incorporates the topic of messages into DeepHawkes model, which models the information diffusion process more comprehensively and improves the accuracy of popularity prediction. (2) The sufficient experiments are carried out on the real data set, and comparing the prediction results of multiple algorithms, which verified the effectiveness of the proposed method.

2. Related work
This section mainly introduces the methods of information popularity prediction, including feature-based methods and generation methods.

Feature-based methods manually extract various features related to information popularity, such as time features [1], structural features [3] and content features [6], and regard popularity prediction task as a regression or classification problem. Zhu et al. [3] proposed the concept of propagation acceleration on information propagation. Pinto et al. [5] predicted the future popularity of news based on the historical information of popularity. Cheng et al. [6] found that in the study of information cascade prediction, temporal and structural characteristics are the key predictors of cascade size.

The generation method regards the information popularity accumulation as an incentive process of user retweeting behaviour, and then models the process. Based on the self-excitation point theory, Zhao et al. [7] designed a statistical model to predict the information popularity. Shen et al. [9] used the reinforcement Poisson process (RPP) to model three factors in social networks. DeepHawkes model [3] combines the high interpretability of feature-based methods with the high predictability of generation methods, bridging the gap between prediction and understanding of information cascade. However, this model only models the information cascade propagation and ignores the effect of information content on information propagation.

3. T-DeepHawkes model
The T-DeepHawkes model includes model component for cascade, model component for topics and the fusion of the above two components. The T-DeepHawkes model takes the information cascade and the text content of the message as input, and the output of the model as the predicted value of the cascade retweeting increment.

3.1. DeepHawkes component for cascading
The DeepHawkes component for cascade adopts the model proposed by Cao et al. [3]. The component takes the information cascade as input and converts the input cascade into a set of diffusion paths, and each diffusion path depicts the information retweeting process during the observation time. The component makes an analogy between the explicable factors of Hawkes process through the three components: user embedding, retweeting path encoding and time decay. For the details of user embedding, path encoding and time decay in the DeepHawkes components of cascade, please refer to literature [3].

3.2. Topic component for topics
The DeepHawkes component for topics takes the text content of the information as input, extracts the topic of the message through the LDA (Latent Dirichlet Allocation) topic model, and constructs the topic embedding matrix. The component builds a topic path for all the topics included by a message,
and passes to the cyclic neural network GRU for encoding, so as to simulate the impact of topics on message propagation.

1. Topic embedding

Different messages involve different topics. In this paper, topics in messages are extracted by LDA topic classification model, and topic information embedding matrix is constructed. The construction process is as follows:

LDA Topic Modelling. Input the document-word frequency matrix into the LDA topic classification model, and construct the topic-word frequency matrix and document-topic matrix. Topic-word frequency matrix stores the probability of each topic generating different words, and document-topic matrix stores the probability of each message corresponding to different topics.

Build Topic Embedded Matrix. Word2Vec method [11] is used to generate the word vector of each word, and then the average word vector of all keywords under each topic is taken as the representation vector of this topic. Each column of the topic embedding matrix $B \in \mathbb{R}^{L \times K}$ represents the representation vector of the topic, $L$ is the dimension of the topic representation vector, and $K$ is the number of topics.

2. Topic path coding

We suppose that a message contains $n$ topics $t_1, t_2, \ldots, t_n$, and these topics can build up a topic path $\{t_1, t_2, \ldots, t_n\}$. This topic path is represented as a one-hot vector $q^i \in \text{RK}$, where K is the total number of topics. All topics share an embedded matrix $B \in \mathbb{R}^{L \times K}$ where $L$ is the dimensions of the topic vector. Topic embedding matrix $B$ converts each topic to its representation vector:

$$y = Bq^i$$

(1)

Similar to user embedding matrix A, topic embedding matrix B is also studied under the supervision of future popularity in the training process. Therefore, the topic embedding matrix B learned is optimal for popularity prediction. The expression vector of the topic path is represented by the output $h_j^i$ of the last GRU unit on the topic path, and $j^i$ represents the number of topics that this message contains.

3.3. Fusion components

The pooling layer integrates the output of the last GRU of each path (cascade path and topic path) in the DeepHawkes components by a sum pooling, comprehensively reflecting the influence of cascade and topic on information diffusion. The cascading-topic $c_j^i$ representation of the message $m^i$ is defined as:

$$c_j^i = \sum_{j=1}^{K'} \lambda_{t_j^i \rightarrow c_j^i} h_j^i + h_j^i$$

(2)

where $K'$ is the number of topics covered by message $m^i$, and $h_j^i$ is the output of the topic path. The output of the sum pooling layer is passed to the full connection layer as its input. The output layer has only one output unit, and its output value is:

$$\Delta \hat{R}_j^i = \text{MLP}(c_j^i)$$

The minimization objective function is defined as:

$$\text{obj} = \frac{1}{M} \sum_{m^i} \Delta \hat{R}_j^i - \Delta R_j^i$$

where $\Delta \hat{R}_j^i$ is the predicted incremental popularity of the message $m^i$, $\Delta R_j^i$ is the true incremental popularity and $M$ is the total number of messages.

4. Experiment and results

4.1. Experiment preparation

Dataset. This experiment used a real dataset crawled from Sina Weibo. In this system, users' posting time has obvious daily and weekly distribution patterns, and the number of blog posts is shown as Weibull distribution. The retweeting and evaluation behaviors of blog posts have a strong correlation,
and the retweeting probability of blog posts is higher than the evaluation probability [12]. The crawl time of dataset is from 8:00 on June 20, 2018 to 16:00 on June 21, 2018. The crawled data contains the message ID, the message publisher ID, the retweeter ID, the time interval between the message retweeting time, the message publishing time (in seconds), and the content of the message. The dataset has 131888 messages, and the average \( R_t \) is 70.2, average path length is 1.29.

**Baseline model.** Three models such as feature-linear, DeepCas and DeepHawkes are used as Baseline models in this paper. Feature-Linear [6] is a cascade prediction model based on time features, structural features and time decay, etc. DeepCas [10] is a cascade prediction model based on representation learning, and DeepHawkes [3] is a cascade prediction model combining deep learning with Hawkes model.

**Evaluation Indicators.** Similar to literature [3], this paper uses the Mean Square log-transformed Error [3] and the median Square log-transformed Error [3] as evaluation indexes, defined as:

\[ MSLE = \frac{1}{M} \sum_{i=1}^{M} SLE_i^E, \]

\[ MLE = \frac{1}{M} \sum_{i=1}^{M} SLE_i^M, \]

where \( SLE_i^E = (\log \hat{R}_t^E - \log R_t^E)^2 \), \( SLE_i^M = (\log \hat{R}_t^M - \log R_t^M)^2 \) is the predicted and true increment of message popularity, respectively. \( mSLE \) is the median of \( SLE_i^E \) which can effectively reduce the impact of outliers, \( mSLE = median(SLE^E, ..., SLE^M) \).

### 4.2. Parameter settings

**Obvious Time.** The task of this paper is to predict the retweeting increment at a certain point in the future according to the retweeting situation observed during the observation time. This paper sets the observation time as 1 hour, 2 hours and 3 hours respectively.

**Training Set, Verification Set and Test Set.** Observation time of dataset starts from 13:00 on June 20, 2018. The first 70% of the data set is set as the training set, the middle 15% as the verification set, and the last 15% as the test set.

**Topics.** In this experiment, the topic number is set as \( K = 480 \) through the analysis of the dataset.

**Other Parameter.** LDA super-parameter is set to the default parameter \( \alpha = 50/K, \beta = 0.01 \). The dimension of user embedded matrix vector and topic embedded matrix vector is set as \( L = L^* = 50 \). The related parameters of the cascade component are consistent with those in DeepHawkes model [2].

### 4.3. Experiment results

Table 1 shows the performance and T-DeepHawkes and other algorithm on test dataset. T-DeepHawkes obtained the best MSLE and mSLE in all observation windows of the dataset, indicating that it is reasonable to introduce topic content features on the basis of the original model.

**Table 1. The performance of baseline model and T-DeepHawkes on dataset.**

| Dataset | Evaluation Metric | T 1h | T 2h | T 3h |
|---------|-------------------|------|------|------|
|         | MSE \( E \)       | mSLE | MSE \( E \) | mSLE | MSE \( E \) | mSLE | MSE \( E \) | mSLE |
| Feature-linear | 4.387 | 0.971 | 4.014 | 0.817 | 3.955 | 0.797 |
| DeepCas    | 3.744 | 1.066 | 3.351 | 0.876 | 3.258 | 0.907 |
| DeepHawkes | 2.596 | 0.801 | 2.387 | 0.699 | 2.359 | 0.723 |
| T-DeepHawkes | **2.582** | **0.776** | **2.353** | **0.654** | **2.297** | **0.712** |

### 5. Conclusion

On the basis of DeepHawkes model, this paper considers the influence of the text content of information on the diffusion, integrates LDA topic mode into DeepHawkes model, and proposes T-DeepHawkes model. T-DeepHawkes model not only considers the cascade factor, but also considers
the text content of the information, so as to model the information diffusion process more comprehensively. T-DeepHawkes further improves the accuracy of popularity prediction while inheriting the high interpretability power of DeepHawkes. The validity of T-DeepHawkes model is verified by experiments on the real data set. In this paper, neural networks are used to autonomously learn the embedding matrices of users and topics, which contain user behaviors, preferences, correlations between topics and other information. However, the accuracy of the model proposed in this paper is less, and how to further improve the prediction accuracy remains to be studied.

Acknowledgments
The authors thank the reviewers for their helpful comments and suggestions. This research was supported by the National Natural Science Foundation of China (61762090, 61262069, 61472346, 61662086), The Natural Science Foundation of Yunnan Province (2016FA026, 2015FB114), the Project of Innovative Research Team of Yunnan Province, Program for Innovation Research Team (in Science and Technology) in University of Yunnan Province (IRTSTYN).

References
[1] Ding Zhaoyun, Jia Yan, and Zhou Bin. Survey of Data Mining for Microblogs [J]. Journal of Computer Research and Development, 2014, 51(4): 691-706 (in Chinese)
[2] Cao Qi, Shen Huawei, Cen Keting, et al. DeepHawkes: Bridging the Gap between Prediction and Understanding of Information Cascades [C] //Proceedings of the 26th ACM International Conference on Information and Knowledge Management. Singapore: ACM, 2017: 1149-58
[3] Zhu Hailong, Yun Xiaochun, Han Zhishuai. Weibo Popularity Prediction Method Based on Propagation Acceleration [J]. Journal of Computer Research and Development, 2018, 55(6): 1282-93 (in Chinese)
[4] Li Wei, Wang lihong, Li Ruiguang. Topic-oriented measurement of microblogging network[J]. Journal on Communications, 2013, (11): 171-78 (in Chinese)
[5] Pinto H, Almeida J M, and Gonçalves M A. Using Early View Patterns to Predict the Popularity of Youtube Videos [C] //Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM ’13). Rome: ACM, 2013:365–74.
[6] Cheng Justion, Adamic Lada A., and Kleinberg J. Can Cascades be Predicted? [C] //Proceedings of the 23rd International Conference on World Wide Web (WWW ’14). Seoul: ACM, 2014: 925–36.
[7] Zhao Qingyuan, Ergodugu, Murar A.,et al. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity [C] //Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’15). Sydney: ACM, 2015: 1513–22
[8] Yang Shuanghong , Zha Hongyuan. Mixture of mutually exciting processes for viral diffusion [J]. Proceedings of the 30th International Conference on International Conference on Machine Learning. JMLR, 2013, 28(2): 1-9
[9] Shen Huawei, Wang Dashun, Song Chaoming, et al. Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes [C] //Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI’14). Canada: 2014: 291–97.
[10] Li Cheng, Ma Jiaqi, Guo Xiaoxiao. DeepCas: an End-to-end Predictor of Information Cascades [C] //Proceedings of the 26th International Conference on World Wide Web Companion (WWW ’17). Australia: ACM, 577–86.
[11] Zhang Zhifei, Miao Duqian, Gao Can. Short text classification using latent Dirichlet allocation [J]. Journal of Computer Applications, 2013: 33(6): 1587-1590(in Chinese)
[12] David M. Blei, Andrew Y. Ng, Michael I. Jordan. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022