A Graph Temporal Information Learning Framework for Popularity Prediction

Caipiao Yang
Beijing Jiaotong University
Beijing, China
19121729@bjtu.edu.cn

Peng Bao*
Beijing Jiaotong University
Beijing, China
baopeng@bjtu.edu.cn

Rong Yan
Beijing Jiaotong University
Beijing, China
20112202@bjtu.edu.cn

Jianian Li
Beijing Jiaotong University
Beijing, China
lijianian@bjtu.edu.cn

Xuanya Li
Baidu Inc.
Beijing, China
lixuanya@baidu.com

ABSTRACT

Effectively predicting the future popularity of online content has important implications in a wide range of areas, including online advertising, user recommendation, and fake news detection. Existing approaches mainly consider the popularity prediction task via path modeling or discrete graph modeling. However, most of them heavily exploit underlying diffusion structural and sequential information, while ignoring the temporal evolution information among different snapshots of cascades. In this paper, we propose a graph temporal information learning framework based on an improved graph convolutional network (GTGCN), which can capture both the temporal information governing the spread of information in a snapshot, and the inherent temporal dependencies among different snapshots. We validate the effectiveness of the GTGCN by applying it on a Sina Weibo dataset in the scenario of predicting retweet cascades. Experimental results demonstrate the superiority of our proposed method over the state-of-the-art approaches.

CCS CONCEPTS

• **Information systems** → *Information systems applications*; • **Human-centered computing** → *Social network analysis.*

KEYWORDS

popularity prediction, dynamic graph representation learning, graph convolutional network

ACM Reference Format:
Caipiao Yang, Peng Bao, Rong Yan, Jianian Li, and Xuanya Li. 2022. A Graph Temporal Information Learning Framework for Popularity Prediction. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3487553.3524231

*Corresponding Author.

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524231

1 INTRODUCTION

With the rapid development of online social network, there are abundant online contents and topics communication among massive users on some mainstream social media. Therefore, predicting the future dissemination trend of online content has important academic significance and broad application prospects, such as online advertising, user recommendation, and fake news detection.

Existing approaches for popularity prediction fall into three main categories: feature-driven approaches, generative approaches, and deep learning-based approaches. The first type of approaches mainly consider the content, user, structural, and temporal features, and further feed these extracted features into a regression or classification model to predict future popularity. Tsur et al.[8] proposed to determine whether the text contains specific features, such as hot topic hashtags and emojis. However, the performance of feature-driven approaches heavily depends on the quality of artificially designed features, which are often hand-crafted based on human's prior domain knowledge. In order to understand the underlying diffusion mechanisms, generative approaches regard the accumulation of popularity as the arrival process of attention, and focus on modeling the intensity function of each piece of information in the process of occurrence [1]. Shen et al.[7] proposed the RPP model, which uses a self-enhancing Poisson process to model the forwarding behavior. Mishra et al.[6] used the Hawkes self-excited point process to model the diffusion process, capturing that each forwarding would bring new excitation for itself to spread future news. However, generative approaches have the less desirable predictive ability since their parameters are not optimized under the supervision of future popularity.

Deep learning-based methods automatically learn valuable representations from data and extract abundant implicit features. Li et al.[5] firstly proposed an end-to-end deep learning model called DeepCas, which converts an information cascade graph into multiple node sequences based on a random walk algorithm. Cao et al.[3] proposed the DeepHawkes model to learn the three key factors of the Hawkes process. The model bridged the gap between prediction and understanding of information cascades. Chen et al.[4] proposed a recursive convolutional network model that employs a graph convolutional network to capture the differences in the structures of cascade graphs. However, these existing methods mostly utilize the order relationship between forwarding users to model

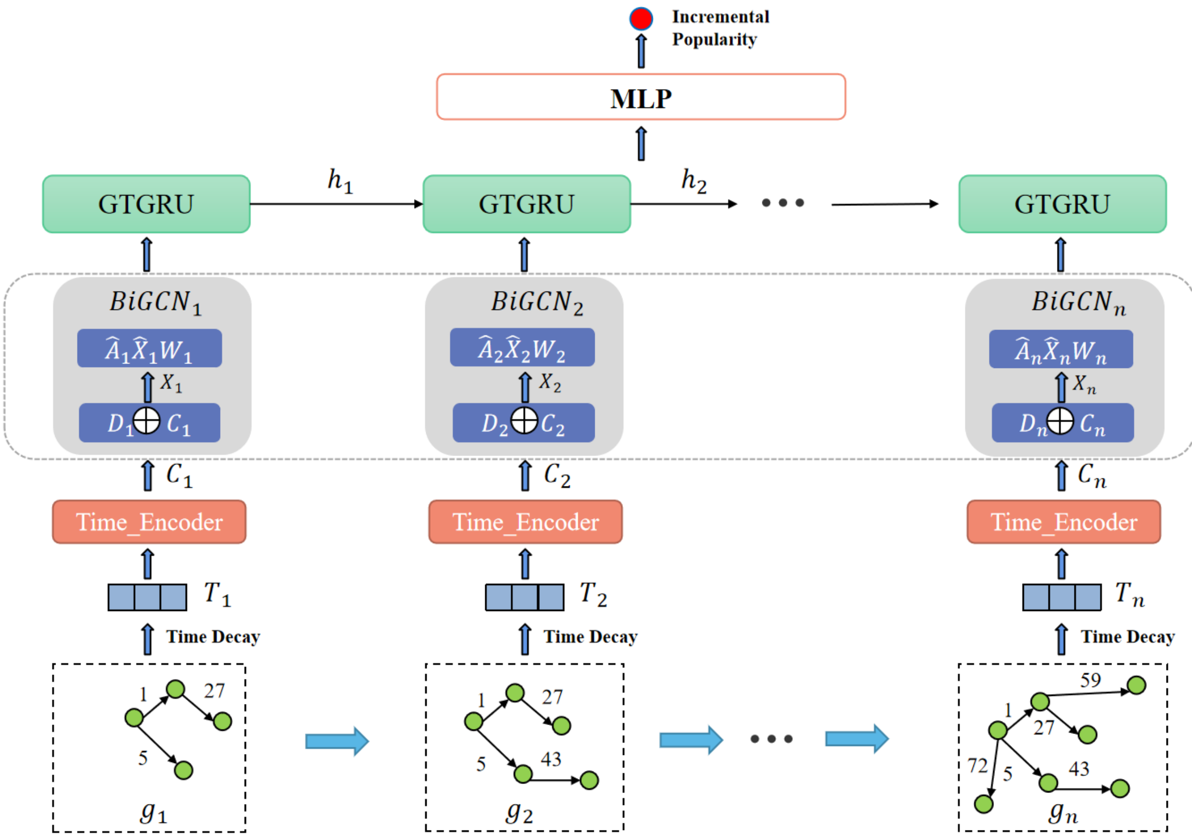

Figure 1: The framework of GTGCN

the sequential information, while largely ignoring the impact of temporal information on popularity prediction.

In this paper, we propose a novel graph temporal information learning framework based on an improved graph convolutional network called GTGCN. The framework first leverages a functional time encoding technique to deal with the timestamp information. Then, GTGCN improves the graph convolutional network to efficiently utilize underlying structural and temporal information in one snapshot by aggregating the time-encode features. Finally, we incorporate temporal encoding into a gated neural network to obtain temporal information among different snapshots. We verify the effectiveness of GTGCN by applying it to the Sina Weibo dataset in the scenario of predicting retweet cascades. Experimental results demonstrate that the proposed GTGCN significantly improves the ability of prediction and outperforms the state-of-the-art approaches for popularity prediction.

2 THE METHOD

The proposed GTGCN framework aims to capture the temporal evolution information in one snapshot and the dependencies among discrete graph sequences in the process of information diffusion. The framework is mainly divided into three components: temporal encoder, improved graph convolutional network, and temporal gated recurrent network. The overall framework is presented in Figure 1.

2.1 Temporal Encoder

The dissemination process of information cascade can form a complex diffusion network. Therefore, we define g to represent the diffusion graph structure at each time span. For simplicity, we divide the observation window [0, T) into n disjoint fine-grained time intervals. The whole cascade graph G can be described as:

$$G = \{g_1, g_2, ..., g_i, ..., g_n\} \tag{1}$$

The snapshot of cascade graph g_i that contains abundant diffusion information at time interval t_i can be represented as a shorthand $g_i = \{u, e, t\}$, where u, e and t respectively represent the users sets, edge sets, and the timestamps when the retweeting occur. Given the fixed prediction time window Δt, our task is to predict the increment size ΔS of users. The final predictive target is computed as $\Delta S = |u^{T+\Delta t}| - |u^T|$.

For each snapshot, existing works only consider modeling its structural features at the slice moment, neglecting its evolution process within the time interval. To capture the temporal information in each subgraph, we adopt the functional time encoding technique based on the classical Bochner's theorem from harmonic analysis referring to the TGAT[9] approach. When a cascade is retweeted gradually, the influence of each forwarding will appear as decay effect over time. Therefore, considering the time decay, the emergence timestamp t of each node can be optimized to $t^{'} = t_i - t$ where t_i is the split time point. We can use the equation below to encode the temporal information and acquire the representation matrix on all timestamps in a snapshot:

$$t2v(t) = \sqrt{\frac{1}{d}}\left[\cos\left(w_1 t\right)\sin\left(w_1 t\right), \ldots, \cos\left(w_d t\right)\sin\left(w_d t\right)\right] \quad (2)$$

where d controls the dimension of generated time vectors, and w_i is the hyperparameter.

In each snapshot g_i, we use T_i to represent the total timestamp matrix in g_i after the time decay processing. Then, we feed T_i into time encoder and obtain the feature representation matrix C_i:

$$C_i = t2v(T_i) \quad (3)$$

2.2 Improved Graph Convolutional Network

Graph convolutional network can effectively extract the structural information of each graph and merge these information in the representation matrix of the graph. In order to model the temporal information in the snapshot preferably, we propose an improved GCN model, which simultaneously models temporal dependencies and structural features by incorporating temporal representation vectors. Firstly, we design an initialization feature matrix $\hat{X}$, which adds the degree matrix D_i to the temporal feature matrix C_i.

The cascade graphs in social network are usually dynamic directed graphs. However, the classical GCN methods cannot be applied for cascades modeling since they focus on fixed and undirected graphs. To solve the problem, we employ the bi-directional graph convolutional network to deal with the directed graphs as Bian et al. proposed in [2]. The Bi-GCN layer is calculated as follows:

$$Bi\text{-}GCN\left(H_t^l, \widehat{A}\right) = \sigma\left(\hat{A}\hat{X}W^l\right), \quad (4)$$

where $\hat{A} = A + I$, which combines the adjacency matrix A and self-loops matrix I to perform graph convolution operation, $H_t^0 = \hat{X}$, and W_l represents the parameter matrix of the l-th layer.

Finally, a multi-layer Bi-GCN with the following layer-wise propagation rule:

$$H_t^{l+1} = Bi\text{-}GCN\left(H_t^l, \widehat{A}\right), \quad (5)$$

where H_t^l is the hidden feature representation of the l-th layer GCN operation.

To guarantee the algorithm performance, two layers model is adopted in our experiment to capture more implicit structural and temporal information.

2.3 Temporal Gated Recurrent Network

The RNNs is an optimal choice to model the temporal dependence of diffusion dynamics. Gate Recurrent Unit (GRU) is a stable and powerful variant of RNNs. More specifically, update gate z_t ignore some unimportant information and merge some information from input.

$$z_t = \sigma\left(W_z \mathrm{x}_t + U_z \mathrm{h}_{t-1}\right), \quad (6)$$

where x_t represents the convolution result of each snapshot H_t^{l+1}, $\sigma(\cdot)$ is the sigmoid activation function. W_z, U_z are GRU parameters learned during training.

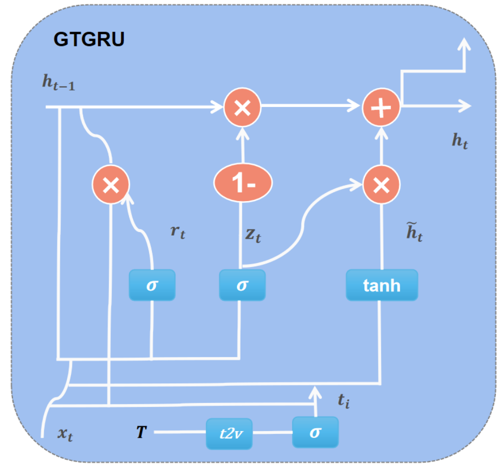

Figure 2: The architecture of GTGRU

Reset gate r_t remembers the current state and adds it to the hidden layer representation, which can be represent as follow:

$$r_t = \sigma\left(W_r \mathrm{x}_t + U_r \mathrm{h}_{t-1}\right). \quad (7)$$

Similarity, W_r, U_r are GRU parameters learned during training.

The actual activation of hidden state $\tilde{h}_t$ is then computed by

$$\tilde{h}_t = \tanh\left(W\mathrm{x}_t + U\left(r_t \odot \mathrm{h}_{t-1}\right)\right), \quad (8)$$

where

$$\mathrm{h}_t = \left(1 - z_t\right) \odot \mathrm{h}_{t-1} + z_t \odot \tilde{h}_t. \quad (9)$$

Note that $\odot$ represents element−wise product, W, U are learnable parameters.

The features and attributes of the cascade evolve over time among different snapshots. We propose a variant of the gated recurrent network to acquire the temporal dependencies in the evolution procedure and model the implicit dynamic information. First, we utilize the temporal encoder to acquire the representation vector on the time span T. Then, the feature vector is activated by sigmoid function and we combine the result with input x_t to form a time gate t_i.

$$t_i = \sigma\left(W_t x_t + \sigma\left(U_T * t2v(T)\right) + b\right), \quad (10)$$

where W_t and U_T are learnable parameters.

Finally, we incorporate the novel time gate t_i to the hidden state, which can be represent as follow:

$$\tilde{h}_t = \tanh\left(W x_t \odot t_i + U\left(r_t \odot h_{t-1}\right)\right) \quad (11)$$

By designing the temporal gate, it is efficient to model the temporal dependencies among the evolution information. The architecture of the temporal gated recurrent network is shown in Figure 2.

The node representation matrix in the final graph is obtained by concatenating the representation vector of its final hidden layer with the graph convolution matrix.

Afterwards, we use sum pooling to aggregate the node representation to graph representation. Finally, the incremental popularity of the cascade can be predicted by a multi-layer perceptron.

3 EXPERIMENTS

In this section, we compare the prediction performance of the proposed GTGCN framework with the state-of-the-art approaches.

3.1 Dataset

The dataset used in this paper is Sina Weibo provided in [3], which collects all original microblogs generated on June 1st in 2016, with all retweets of each post within the next 24 hours. The cascades with the publication time before 8 *am* and after 6 *pm* are filtered out and the dataset consists of 119,313 posts in total. We follow the similar experimental setup as DeepHawkes, i.e., the length T of observation time window is $T = 1$ hour, 2 hours, and 3 hours. We take 70% cascades as the training set, 15% cascades as the validation set, and the remaining 15% cascades as the test set. The statistics of the dataset are reported in Table 1.

Table 1: Statistics of the Sina Weibo dataset.

T	Data	1h	2h	3h
Number of Messages	train	29531	35403	38576
	val	6328	7586	8266
	test	6327	7586	8266
Average Path Length	train	1.24	1.26	1.27
	val	1.27	1.30	1.30
	test	1.25	1.27	1.28

3.2 Baselines

In order to quantitatively examine the effectiveness of GTGCN, we use the mean square log-transformed error (MSLE) as the evaluation metric. The smaller the MSLE is, the better the prediction approach performs. As we mentioned in Introduction, existing methods for popularity prediction are classified into three categories. We select state-of-the-art methods in each category as strong baselines. The generative methods have poor prediction performance so that we will ignore the type of methods. The selected five classical baselines are as follows.

- *Feature-Driven approaches*: We extract structural features (i.e. the number of leaf nodes and first layer nodes, average and max length of retweet) and temporal features (i.e. the mean time interval between each retweet, the cumulative popularity and incremental popularity every 10 minutes). We further feed them into a linear regression model with L2 regularization or a multi-layer perceptron, denoted as **Feature-linear** and **Feature-MLP**, respectively.

- *Deep Learning approaches*: **DeepCas** and **DeepHawkes** are representative methods that consider the cascade graph as sequence modeling.

- *Discrete approaches*: **CasCN** is the state-of-the-art method on popularity prediction by learning discrete graph sequences.

Table 2: Prediction performance of all methods

MSLE	1h	2h	3h
Feature-Linear	3.737	3.379	3.329
Feature-MLP	3.653	3.215	2.927
DeepCas	3.632	3.258	3.047
DeepHawkes	2.573	2.307	2.245
CasCN	2.316	2.254	2.131
GTGCN	**2.251**	**2.159**	**1.928**

3.3 Prediction Performance

Table 2 summarizes the performance comparison among GTGCN and baselines on three scenarios of the Sina Weibo dataset. In general, the proposed GTGCN framework outperforms traditional approaches, as well as superior to the state-of-the-art deep learning methods, with a statistically significant drop of MSLE.

Our model exhibits strong prediction performance, with an improvement of 39.76%, 38.38%, 38.02%, 12.51%, and 9.53% compared with Feature-Linear, Feature-MLP, DeepCas, DeepHawkes, and CasCN, respectively. The results reflect that the timestamp encoder and temporal modeling have significant effects for predicting the popularity of cascades.

4 CONCLUSIONS

In this paper, we propose a novel graph temporal information learning framework based on an improved graph convolutional network called GTGCN, which captures both the temporal information governing the spread of information in a snapshot and the inherent temporal dependencies among different snapshots. We apply it to the Sina Weibo dataset in the scenario of predicting retweet cascades. Experimental results demonstrate that our proposed framework can significantly outperform the state-of-the-art approaches. As for future work, we will dive deeper into the time-decaying effect modeling, as well as the model interpretability. Besides, codes will also be shared on MindSpore.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (61702031). Peng Bao is also sponsored by CAAI-Huawei MindSpore Open Fund and Baidu Program.

REFERENCES

[1] P. Bao and X. Zhang. 2017. Uncovering and predicting the dynamic process of collective attention with survival theory. *Scientific Reports* 7, 1 (2017), 2621.
[2] T. Bian, X. Xiao, T. Xu, and et al. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *Proc. of AAAI*. 549–556.
[3] Q. Cao, H. Shen, K. Cen, and et al. 2017. DeepHawkes: Bridging the Gap between Prediction and Understanding of Information Cascades. In *Proc. of CIKM*. 1149–1158.
[4] X. Chen, F. Zhou, K. Zhang, and et al. 2019. Information Diffusion Prediction via Recurrent Cascades Convolution. In *Proc. of ICDE*. 770–781.
[5] L. Cheng, J. Ma, X. Guo, and et al. 2017. DeepCas: An End-to-end Predictor of Information Cascades. In *Proc. of WWW*. 577–586.
[6] S. Mishra, M. Rizoiu, and L. Xie. 2016. Feature driven and point process approaches for popularity prediction. In *Proc. of CIKM*. 1069–1078.
[7] H. Shen, D. Wang, C. Song, and et al. 2014. Modeling and predicting popularity dynamics via reinforced poisson processes. In *Proc. of AAAI*, Vol. 28.
[8] O. Tsur and A. Rappoport. 2012. What's in a Hashtag? Content based Prediction of the Spread of Ideas in Microblogging Communities. In *Proc. of WSDM*. 643–652.
[9] D. Xu, C. Ruan, E. Korpeoglu, and et al. 2020. Inductive Representation Learning on Temporal Graphs. arXiv:2002.07962