CasODE: Modeling Irregular Information Cascade via Neural Ordinary Differential Equations (Student Abstract)

Zhangtao Cheng[1], Xovee Xu[1], Ting Zhong[1]*, Fan Zhou[1], Goce Trajcevski [2]

[1]University of Electronic Science and Technology of China, 610054
[2]Iowa State University, Ames, Iowa, 50011
zhangtao980107@outlook.com, xovee@live.com, zhongting@uestc.edu.cn, fan.zhou@uestc.edu.cn, gocet25@iastate.edu

Abstract

Predicting information cascade popularity is a fundamental problem for understanding the nature of information propagation on social media. However, existing works fail to capture an essential aspect of information propagation: the temporal irregularity of cascade event – i.e., users' re-tweetings at random and non-periodic time instants. In this work, we present a novel framework CasODE for information cascade prediction with neural ordinary differential equations (ODEs). CasODE generalizes the discrete state transitions in RNNs to continuous-time dynamics for modeling the irregular-sampled events in information cascades. Experimental evaluations on real-world datasets demonstrate the advantages of the CasODE over baseline approaches.

Introduction

Online social platforms have become an integral element of personal daily life, enriching real-time communication among individuals and enabling information diffusion quickly. The dynamics of users' activities facilitate the fast propagation of information along social networks, forming an information cascade. Understanding information cascade has significant economic and societal impacts – and one of the typical tasks is information popularity prediction which forecasts the size of potentially affected users after a certain time (Zhou et al. 2021).

Challenges. Despite the success of existing neural network-based methods (Zhou et al. 2021), real-world information diffusion processes are often irregularly sampled (i.e., the time series of different user activities are non-uniform). Specifically, the following are the observations along these lines which motivate our work: *Irregularly-sampled cascade event.* Since people have personal preferences and requirements in real life, they may browse tweets/microblogs and retweet them at different times. This is also the case for other types of networks, such as paper citations and rumor propagation. The time intervals between adjacent events (e.g., retweets and citations) are irregular. RNNs are the dominant models for capturing the temporal patterns of the information diffusion, which, however, are initially designed for regularly-sampled sequences but cannot reflect the influence of irregular events due to the underlying inflexible iterative structure (Rubanova, Chen, and Duvenaud 2019).

Present Work. To address the mentioned challenge, we present a novel information **Cas**cade popularity prediction model based on neural **O**rdinary **D**ifferential **E**quations (Chen et al. 2018) (**CasODE**) for modeling the irregular-sampled events in information cascades. Specifically, CasODE generalizes discrete state transitions to continuous-time dynamics of the information cascade. It allows us to more appropriately model the real information propagation, obeying an ODE between successive observations to possess continuous hidden states. Once a new event occurs, the state will be updated by a gating mechanism, which jointly considers the new input and the temporal interval.

Methodology

In this section, we present the architecture of our proposed CasODE, which consists of three components: (1) Structural-equivalent feature extraction, (2) Irregular information diffusion modeling, and (3) Information cascade popularity prediction.

Structural-Equivalent Feature Extraction. Given an information cascade graph $\mathcal{G}_c(t_o)$ observed at time t_o, we have its weighted adjacency matrix $\mathbf{A}_c$ and diagonal degree matrix $\mathbf{D}_c$. Then an unnormalized graph Laplacian $\mathbf{L}_c = \mathbf{D}_c - \mathbf{A}_c = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ can be used to calculate the spectral graph wavelets Φ with heat kernel function on the spectrum (Donnat et al. 2018). Each column vector of $\Phi \in \mathbb{R}^{|\mathcal{V}_c|\times|\mathcal{V}_c|}$ is the wavelets for a node in cascade graph. In order to solve the graph mapping problem (i.e., solve the "isomorphism" problem between two nodes' neighbors), the wavelet coefficients are processed as a probability distribution. Then the empirical characteristic functions are utilized to obtain the final structural-equivalent node embeddings $\mathbf{E}$. The Chebyshev polynomials are used to calculate the wavelet coefficients. The overall complexity for cascade graph structure learning is linear in the number of edges in the graph.

Modeling Irregular Temporal Diffusion. First, given the sequence of user embeddings $\{\mathbf{E}_{u_i}|u_i \in \mathcal{V}_c\}$, we employ an LSTM cell before performing the ODE solver to avoid the vanishing or exploding of gradients. Subsequently, we use a numerical ODE solver – the Euler method (Chen et al. 2018)

*corresponding author.

– to evaluate the hidden states between successive observations and update the hidden states using a GRU cell at each observation. Besides, we propose a temporal gating mechanism (T-Gate) to merge the latent representation from the first two steps to learn the continuous-time dynamics. We use T-Gate to handle the irregular sampling problem caused by the time-varying from seconds to hours between successive cascade events that make the ODE solver challenging to evaluate continuous hidden dynamics.

Within the previous hidden state pair $(\mathbf{c}_{i-1}, \mathbf{h}_{i-1})$ at time t_{i-1} and user u_i's embedding $\mathbf{E}_{u_i}$, we first feed them into the LSTM cell to generate a new hidden state $(\mathbf{c}_i, \mathbf{h}'_i)$:

$$(\mathbf{c}_i, \mathbf{h}'_i) = \text{LSTMCell}\left(\theta_l, (\mathbf{c}_{i-1}, \mathbf{h}_{i-1}), \mathbf{E}_{u_i}\right), \qquad (1)$$

where $\mathbf{c}_{i-1}$ is the memory cell and $\mathbf{h}_{i-1}$ is the output state. θ_l denote the learnable parameters in LSTM Cell. Then, we feed $\mathbf{h}_{i-1}$ into the ODE solver based on the Euler method to obtain the ODE hidden state $\mathbf{z}_i$ at each step. This operation is to construct the continuous-time dynamics of the hidden states between irregular time intervals in all consecutive (t_{i-1}, t_i) pairs. To construct the true cascade representation at time t_i, we take two states of LSTM cell – output state $\mathbf{h}'_i$ and hidden state $\mathbf{z}_i$ – as the input for latent representation learning of u_i, which outputs $\mathbf{h}''_i$:

$$\mathbf{z}_i = \text{ODESolver}\left(f_\omega, \mathbf{h}_{i-1}, \mathbf{h}'_i, (t_{i-1}, t_i)\right), \qquad (2)$$

$$\mathbf{h}''_i = \text{GRUCell}\left(\theta_g, \mathbf{h}'_i, \mathbf{z}_i\right), \qquad (3)$$

where $\mathbf{z}_i \in \mathbb{R}^d$ is the solution at t_i to an ODE started from time t_{i-1}; $\mathbf{h}''_i$ is the updated hidden state; θ_g denotes the learnable parameters in GRU Cell. Given the latent states $\mathbf{h}'_i$ and $\mathbf{h}''_i$, we update $\mathbf{h}_i$ using the gating mechanism:

$$\mathbf{h}_i = \boldsymbol{\nu}_i \odot \mathbf{h}''_i + (\mathbf{1} - \boldsymbol{\nu}_i) \odot \mathbf{h}'_i, \qquad (4)$$

where the temporal gate $\boldsymbol{\nu}_i = e^{-(\Delta t_i)} \in \mathbb{R}^d$ helps the model determine how much of the state is solved by ODE that needs to be passed to the future. Finally, we compute the output states $\{\mathbf{o}_1 \dots \mathbf{o}_n\}$ via a fully-connected layer for downstream tasks, where $\mathbf{o} \in \mathbb{R}^d$ and n denotes the number of users $|\mathcal{V}_c|$ in the early evolution. Generally, we use the final output state $\mathbf{o}_n$ as the cascade latent representation $\mathbf{Z}$.

Prediction. We feed $\mathbf{Z}$ into multi-layer perceptrons (MLPs) to predict cascades' popularity. During training, we use the mean square logarithmic error (MSLE) as the objective to train CasODE.

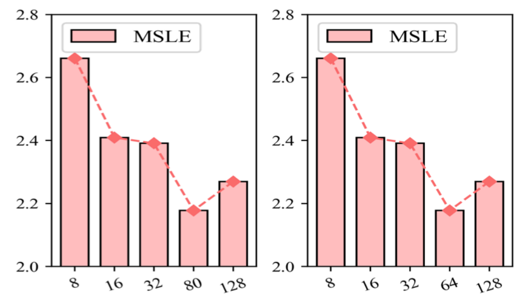
Experiments

Datasets and Baselines. We select two public datasets, i.e., Twitter and Weibo, and compare our model CasODE with five baselines: **DeepHawkes** (Cao et al. 2017), **CasCN** (Chen et al. 2019), **LatentODE** (Rubanova, Chen, and Duvenaud 2019) and **CasFlow** (Xu et al. 2021).

Performance Comparison. The performance of baselines and CasODE on two datasets are summarized in Table **??**. In particular, the proposed CasODE outperforms all baselines in terms of both MSLE and MAPE, demonstrating the benefits of exploiting continuous-time dynamics for modeling the irregular-sampled events in information cascades.

Model	Twitter		Weibo	
	MSLE	MAPE	MSLE	MAPE
DeepHawkes	7.216	0.587	2.796	0.282
CasCN	7.183	0.547	2.732	0.273
LatentODE	7.112	0.475	2.234	0.245
CasFlow	6.955	0.456	2.281	0.242
CasODE	**6.812**	**0.436**	**2.195**	**0.217**

Table 1: Performance comparisons on two datasets.



(a) Dimension of $\mathbf{E}$ (b) Dimension of $\mathbf{Z}$

Figure 1: Impact of hyper-parameters of CasODE on Weibo dataset, measured by MSLE.

Parameter Sensitivity. Figure 1(a) and (b) show that the best performance of CasODE is achieved when the dimensions of latent factor $\mathbf{Z}$ is 64 and the dimension of cascade embedding $\mathbf{E}$ is 80.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Grant No.62072077 and No.62176043), Natural Science Foundation of Sichuan Province (Grant No. 2022NSFSC0505), and National Science Foundation SWIFT (Grant No.2030249).

References

Cao, Q.; Shen, H.; Cen, K.; Ouyang, W.; and Cheng, X. 2017. DeepHawkes: Bridging the gap between prediction and understanding of information cascades. In *CIKM*.

Chen, R. T.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. 2018. Neural ordinary differential equations. In *NeurIPS*, 6572–6583.

Chen, X.; Zhou, F.; Zhang, K.; Trajcevski, G.; Zhong, T.; and Zhang, F. 2019. Information diffusion prediction via recurrent cascades convolution. In *ICDE*.

Donnat, C.; Zitnik, M.; Hallac, D.; and Leskovec, J. 2018. Learning structural node embeddings via diffusion wavelets. In *SIGKDD*, 1320–1329.

Rubanova, Y.; Chen, R. T.; and Duvenaud, D. 2019. Latent odes for irregularly-sampled time series. In *NeurIPS*.

Xu, X.; Zhou, F.; Zhang, K.; Liu, S.; and Trajcevski, G. 2021. CasFlow: Exploring Hierarchical Structures and Propagation Uncertainty for Cascade Prediction. *TKDE*, 1–14.

Zhou, F.; Xu, X.; Trajcevski, G.; and Zhang, K. 2021. A Survey of Information Cascade Analysis: Models, Predictions, and Recent advances. *ACM CSUR*, 54(2): 1–36.