

Linking Transformer to Hawkes Process for Information Cascade Prediction (Student Abstract)

Liu Yu¹, Xovee Xu¹, Ting Zhong¹, Gocet Trajcevski², Fan Zhou^{1*}

¹ University of Electronic Science and Technology of China

² Iowa State University

liu.yu@std.uestc.edu.cn, xovee@ieee.org, zhongting@uestc.edu.cn, gocet25@iastate.edu, fan.zhou@uestc.edu.cn

Abstract

Information cascade is typically formalized as a process of (simplified) discrete sequence of events, and recent approaches have tackled its prediction via variants of recurrent neural networks. However, the information diffusion process is essentially an evolving directed acyclic graph (DAG) in the continuous-time domain. In this paper, we propose a transformer enhanced Hawkes process (**Hawkesformer**), which links the hierarchical attention mechanism with Hawkes process to model the arrival stream of discrete events continuously. A two-level attention architecture is used to parameterize the intensity function of Hawkesformer, which captures the long-term dependencies between nodes in graph and better embeds the cascade evolution rate for modeling short-term outbreaks. Experimental results demonstrate the significant improvements of Hawkesformer over the state-of-the-art.

Introduction

Sharing content through social media platforms such as Twitter and Weibo has become a main channel to express individual opinions. The initial tweet, along with the subsequent retweets forms an *information cascade* (Zhou et al. 2021). Predicting the *size* of an information cascade after a certain time-period is one of the typical tasks. The core challenge is how to model the underlying diffusion process governing the popularity dynamics of information cascades.

Prior *Temporal Point Processes* (TPPs) based methods and *Deep Learning*-based approaches especially recurrent neural network (RNN) based sequential models exhibit three notable drawbacks: (i) they use simple intensity functions and make strong assumptions on the diffusion mechanism; (ii) the irregular time intervals between events and their order are important to describe the diffusion dynamics, however, they are difficult to capture in a way from discrete-time domain; (iii) they cannot fully exploit the diffusion process of DAGs and capture the long-term dependency due to the intrinsic limitations of recurrent models; and (iv) previous works (Tang et al. 2021) have shown that information cascades with short-term outbreaks are more likely to be popular in the future. They cannot capture such trends intrinsi-

cally and RNN-based models are prone to face error accumulation issue, especially when the cascade size is large.

Methodology

Let C denote an information cascade with a retweet history $\mathcal{H} = \{(t_j, u_j)\}_{j=1}^L$ with $t_j \in [0, t_o)$, where each pair (t_j, u_j) corresponds to infected user u_j occurs at time t_j . L denotes the number of retweets that have occurred up to the observation time t_o . Generally, we focus on the information cascade prediction as a complex and nonlinear regression task, predicting the popularity size $P(t_p|\mathcal{H})$ of a cascade over a period of time, where t_p is the time of prediction.

Hawkesformer provides an elegant mathematical tool for modeling event occurrence in continuous-time domain and takes the underlying DAG into consideration, which links a two-level attention (Vaswani et al. 2017) framework to parameterize the intensity function of Hawkes process (Hawkes 1971). We characterize the long-term dependence as $\mathbf{h}(t)$ and short-term outbreak as $\mathbf{\vartheta}(t)$. Given an information cascade C – e.g., a tweet, and its retweet history $\mathcal{H} = \{(t_j, u_j)\}_{j=1}^L$ – we model the continuous dynamics of temporal point processes, with the following conditional intensity function:

$$\lambda(t|\theta, \mathcal{H}) = f\left(\underbrace{\alpha \frac{t - t_j}{t_j}}_{\text{current}} + \underbrace{\mathbf{w}_1^\top \mathbf{h}(t_j)}_{\text{long-term}} + \underbrace{\mathbf{w}_2^\top \mathbf{\vartheta}(t_j)}_{\text{short-term}}\right) \quad (1)$$

where θ denotes the model parameters; retweet time t is defined on interval $[t_j, t_{j+1})$; and $f(\cdot)$ is a softplus function to constrain the intensity function to be positive. The first term *current* indicates the evolutionary process in the continuous-time domain; $\mathbf{w}_1, \mathbf{w}_2$ are the learned weights for long-term dependency and short-term outbreak at time t respectively; and $\lambda(t)$ denotes the arrival rate of a retweet at time t .

Specifically, at the first level, we design a long-term dependency module to dynamically capture the diffusion process, where we make a *primary/non-primary* path assumption to adaptively integrate the diffusion process on the underlying DAG, see Figure 1. This also enables each node attending the cascade at any position to update its current hidden states. At the second level, we design a short-term outbreak module which considers the evolution rate of a cascade and learns the local patterns in a time-slice window.

*Corresponding author

