Bridging Models for Popularity Prediction on Social Media

Swapnil Mishra

The Australian National University & Data61 CSIRO

Canberra, Australia

swapnil.mishra@anu.edu.au

ABSTRACT

Understanding and predicting the popularity of online items is an important open problem in social media analysis. Most of the recent work on popularity prediction is either based on learning a variety of features from full network data or using generative processes to model the event time data. We identify two gaps in the current state of the art prediction models. The first is the unexplored connection and comparison between the two aforementioned approaches. In our work, we bridge gap between feature-driven and generative models by modelling social cascade with a marked Hawkes self-exciting point process. We then learn a predictive layer on top for popularity prediction using a collection of cascade history. Secondly, the existing methods typically focus on a single source of external influence, whereas for many types of online content such as YouTube videos or news articles, attention is driven by multiple heterogeneous sources simultaneously - e.g. microblogs or traditional media coverage. We propose a recurrent neural network based model for asynchronous streams that connects multiple streams of different granularity via joint inference. We further design two new measures, one to explain the viral potential of videos, the other to uncover latent influences including seasonal trends. This work provides accurate and explainable popularity predictions, as well as computational tools for content producers and marketers to allocate resources for promotion campaigns.

KEYWORDS

Information diffusion; Hawkes Processes; Recurrent Neural Networks

1 INTRODUCTION AND BACKGROUND

Popularity refers to the total attention that a digital item receives. Understanding popularity has been an important problem since the early days of social media research. Important open questions include explaining observed popularity, predicting popularity in the future, and being able to reason about the effects of external interventions. Understanding popularity dynamics would provide

us with more insights of collective behavior by compiling seemingly incoherent individual responses to a single piece of information.

When presented with the task of predicting popularity, a clear gap appears between the two main classes of approaches: the feature-driven approaches and the generative approaches. *Feature-driven approaches* [1] rely on collecting large-scale datasets and constructing informative features. *Generative methods* [2, 6] leverage fine-grained timing information in the event series. Hence first research question we address in our work is: **how to unify feature-driven and generative models?**

Furthermore, recent models can be categorized into two distinct classes based on type of information being modelled. The first describes individual user actions, or discrete events in continuous time (e.g. tweets) [8]. The second is based on aggregate metrics of user actions [1] or aggregated event volumes (e.g. the number of daily views) [2, 7]. Each of these models specialise in a distinct data type, but it is common to observe data of different types for the same online item, for example in YouTube videos or news articles, attention is driven by multiple heterogeneous sources simultaneously – e.g. microblogs or traditional media coverage. It is desirable to develop a model that accounts for multiple heterogeneous series. This leads us to our second research question: **can we model multiple heterogeneous asynchronous streams of data driving content popularity?**

Moreover, state-of-the-art popularity models provide black-box predictions [3, 5, 8]. In practice one often demands simulations on various *what-if* scenarios, such as to quantify the effect of a unit amount of promotions, to capture seasonality or the response to outliers, to name a few. Lastly, the influence users have on popularity has been subject to constant debate in this research area. It is desirable to have one model on which the future effect of different users can be comparably studied. This motivates us to **design metrics for explainability and interpretability of state-of-the-art prediction models.**

2 CONTRIBUTIONS

Feature Driven and Generative Approaches In our work [3], we bridge the gap between feature driven and generative models for popularity prediction in context of predicting final size of retweet cascades. We build a generative model with Hawkes Process, which has a predictive layer on top, using model parameters as features, to make final size prediction. The properties modeled into the generative model are: inherent quality of the tweet, social influence of users and the length of "social memory". We develop a state of the art feature driven approach, where features can be computed on data containing solely the message content and basic user profiles. Hence these features can be computed very easily on any public social dataset. We further combine both models to create a new hybrid model. This helps us to take into account other

cascades and mitigate limitations of model assumptions and parameter estimation. We use all three models to do both regression (actual number of retweets) and classification (whether a cascade will double it's current size) task. This has enabled us to create common benchmarking for feature-driven and generative models.

Generalized Point Process Models Our generative model in previous work doesn't account for two important practical aspects of a social cascade – i) limited population size that a cascade can grow over and ii) unknown underlying distribution responsible for generating our data. We propose two new models HawkesN [6], accounting for limited size and Recurrent Point Process (RPP) [4], to tackle the restrictive expressive power of parametric generative processes. In our HawkesN model apart from modelling the self-exciting part with an additive kernel, we also modulate the intensity of arrival of events based on the fraction of population available for cascade to grow. This model captures the intuition that as information is spreading at first it's more viral but once it has diffused enough in network the self-excitation rate goes down as not enough nodes are available to be influenced. In RPP model we use a recurrent neural network to learn the generative process of events based on their history. The intuition behind this work is that the dependency of past events on future events is captured by the hidden state of a RNN model and hence we can learn any non-linear dependency on events rather than being restricted to a parametric form.

Modelling Popularity in Multiple Heterogeneous Streams We tackle the challenges related to handling heterogeneous asynchronous streams of data in our work [4], where we predict the daily views of a YouTube video in presence of a daily share series along with an event series of tweets about the same video. We develop a new *RNN-MAS* model (Recurrent Neural Networks for Multiple Asynchronous Streams), a flexible class of models learnable from social cascades that describe heterogeneous information streams, explain predictions, and compare user effects for both individuals and groups using recurrent neural networks. We link multiple recurrent neural networks by allowing them to exchange information across different asynchronous streams. The core idea of the work is to formulate the generative model of intensity process for event series as well the volume series with help of Recurrent Neural Networks. Once we have models for both series individually we combine them with help of the intensity process and learn parameters with joint inference.

Interpetability and Explainability *RNN-MAS* describes popularity under influence of heterogeneous streams of shares and tweets. We propose two metrics based on *RNN-MAS* to quantify average response to unit promotion, and the relative influence among users of different *fame*. The aforementioned metrics helps us to understand many *what-if* scenarios that can be answered with the help of our predictive models without necessitating a need to understand individual parameters of the network. We also utilize the model to estimate a response series to unseen influence for a video. Response to unseen influence allows us to consider effects of multiple sources of external influence, not captured explicitly by our model.

We obtain metrics by using learnt model for a video. For example, we can obtain the unseen influence by simulating the view series under zero promotion (zero shares and tweets). Now to obtain response to a unit promotion we can first simulate the model with a promotion of unit p and then subtract it from unseen response and normalize it with p. Similarly for promotion under a series of tweets we can first simulate the resulting tweet series, as a tweet excites other tweet, and then plug this simulated tweet series as promotion to our learn model. Hence we can use this to estimate the influence of Twitter users in YouTube. We define this influence by a user as new metric loudness level, for details refer [4].

3 SUMMARY AND DISCUSSION

We have introduced our work on bridging different models for popularity prediction in social media. We have successfully modelled future popularity with generative models in conjunction with feature-driven models. This has helped us and the community at large to gather more insights on trade-offs between different classes of models. Our results suggest we should use both classes of models together to get best of both world, high predictive power of featured-driven models and greater explainability of point process models. We have also been able to evolve the space of popularity modelling by presenting a novel model *RNN-MAS* that can handle multiple asynchronous streams. This new model has given us a way to use current progress in deep-learning to build superior predictive models. The main contribution in building the models is not just our superior performance but our formulation of new measures that provide us a tool to interpret and simulate popularity at scale. Especially, to the best our knowledge, we have developed the first tool and measure to estimate the influence of users across network boundaries.

Admittedly there are some gaps we would like to bridge further in this work. Specifically we would like our models to explicitly capture the dependencies between different content competing for attention. This will enable us to capture both temporal dynamic drifts and effect of limited attention in our models. Another important question we would like to address with our models is the ability to make predictions in cold start setting. Another area for future work is related to studying effects on performance of models on longitudinal data and see how performance evolves time horizon of the forecast and data used.

REFERENCES

[1] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can cascades be predicted?. In *WWW '14*.
[2] Riley Crane and Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS '08* 105 (2008).
[3] Swapnil Mishra, Marian-Andrei Rizoiu, and Lexing Xie. 2016. Feature Driven and Point Process Approaches for Popularity Prediction. In *CIKM '16*.
[4] Swapnil Mishra, Marian-Andrei Rizoiu, and Lexing Xie. 2018. Modeling Popularity in Asynchronous Social Media Streams with Recurrent Neural Networks. In *ICWSM '18*.
[5] Marian-Andrei Rizoiu, Young Lee, Swapnil Mishra, and Lexing Xie. 2018. Frontiers of Multimedia Research. ACM and M&C, Chapter Hawkes Processes for Events in Social Media.
[6] Marian-Andrei Rizoiu, Swapnil Mishra, Quyu Kong, Mark Carman, and Lexing Xie. 2018. SIR-Hawkes: Linking Epidemic Models and Hawkes Processes to Model Diffusions in Finite Populations. In *WWW '18*.
[7] Marian-Andrei Rizoiu, Lexing Xie, Scott Sanner, Manuel Cebrian, Honglin Yu, and Pascal Van Hentenryck. 2017. Expecting to be HIP: Hawkes Intensity Processes for Social Media Popularity. In *WWW'17*.
[8] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. 2015. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In *KDD '15*.