
Personalizing Marked Temporal Point Process Models

Alex Boyd Padhraic Smyth Robert Bamler Stephan Mandt

Department of Computer Science
University of California, Irvine
{alexjb,smyth,rbamler,mandt}@uci.edu

1 Introduction

Event sequences in continuous time occur across many real-world contexts, leading to a variety of associated data analysis applications such as forecasting consumer purchases, fraud detection in transaction data, and survival analysis in medicine. In this paper we focus on discrete events where each observed event has a discrete label, or *mark*, and a timestamp for when it occurred. Examples include medical events or diagnoses for a patient over time as well as products viewed or purchased by a consumer.

There has been significant amount of prior work in statistics for modeling such data, typically under the broad framework of marked temporal point process (MTPP) models [2]. These models characterize the instantaneous rate of occurrence, the *intensity function* for an MTPP, for events of various types, conditioned on the history of a sequence up until that point in time. Recent work in machine learning has sought to address these limitations via the use of deep recurrent neural networks (RNNs). These models use expressive representations for the intensity function, use event embeddings to avoid parameter explosion, and optimize the associated log-likelihood via stochastic gradient methods.

A common implicit assumption made in these techniques is that all of the sequences being modeled originate from the same source or user, which is often not the case in real-world datasets such as online consumer data or medical history records. Sufficiently powerful neural-based MTPPs can internally adjust for this heterogeneity after conditioning on a significant portion of a history; however, they will likely exhibit large predictive uncertainty at the beginning of sequences. It is important to develop techniques that *personalize* predictions to account for heterogeneity across users, as illustrated in Figure 1. To develop personalized MTPPs, we propose using variational autoencoders (VAEs) [3] in conjunction with previous neural-based MTPPs. We also employ a mixture-of-experts approach for VAEs in order to account for heterogeneity within the same user [6].

The main contributions of our paper are:

- A novel personalization scheme using VAEs for marked temporal point processes that is able to efficiently scale to many users as well as easily adapt to new users.
- Systematic experimental results on three large, real world datasets that clearly demonstrate improvements in predictive performance by using latent user preferences.
- New evaluation methodologies to assess the ability of MTPP models to generate future predictions and sampled trajectories conditioned on partial sequence histories.

2 Method and Experiments

Almost all of the neural-based MTPPs share common characteristics in their architectures. Namely, they utilize a recurrent unit that accepts the embedding associated with an event’s mark and somehow integrates the time of the event into the hidden state (either through direct input or some specific interpolation scheme). The hidden state is then used to evaluate various predicted intensity values.

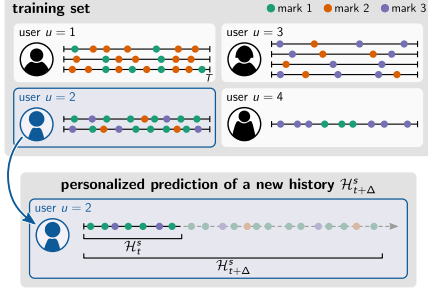


Figure 1: On top is a collection of users and sequences generated by them to be learned from. On bottom is an illustration of the personalized prediction task where a partial sequence from user $u = 2$ is conditioned on a partial history \mathcal{H} .

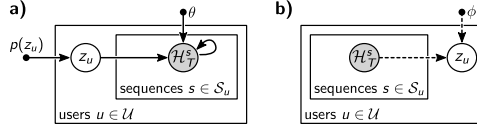


Figure 2: Graphical model of the proposed VAE for personalized point processes where \mathcal{H}_T^s is a sequence of timestamps and discrete marks from time 0 up to time T and z_u represents the user embedding for user u . a) generative model. b) inference model.

This work proposes using a bidirectional RNN to estimate a user’s latent distribution by encoding a reference sequence of events. Note that since we are under the mixture-of-experts paradigm, we allow for encoding sequences that are different from the ones we are decoding so long as they come from the same user. One interpretation of this setup is that it is no longer a variational autoencoder but rather a variational sequence to sequence model.

After drawing a latent state from the amortized distribution, our proposed framework incorporates this into most neural-based MTPPs by estimating the initial RNN hidden states from this latent state, as well as concatenating the latent state to the embedded representation of every mark that is inputted into the recurrent unit. This serves to contextualize the hidden state and personalize downstream intensity predictions. A graphical representation of our proposed framework can be seen in Figure 2.

To showcase the benefits of this framework, we fit two different models—the neural Hawkes process (NHP) and the recurrent marked temporal point process (RMTPP)—on three different event sequence datasets—Reddit comments where the subreddit posted to is the mark [1], Amazon reviews with the product category as the mark [5], and MemeTracker where a common phrase uttered on the internet is the event with the website being the mark [4]. The datasets had 1000, 737, and 5000 unique marks respectively and each yielded between 310K and 410K sequences with between 43K and 80K users.

After training both models in decoder-only and mixture-of-experts configurations on all three datasets, we find that the mixture-of-experts versions are superior when measured on test log-likelihood. Through a more in depth analysis, we find that the majority of the gains in performance are from better model certainty during the beginning portions of sequences.

We also proposed a new sampling task with two accompany novel metrics for evaluating a model’s sampling performance. The task is given a ground truth sequence, we truncate it at a certain percentage, and then sample a new trajectory for the sequence. The new forecasted future is then compared to the held out portion that was truncated via our new proposed metrics. The first of which measures the appropriateness of sampled marks with Jaccard distance. The second measures how well the sampled timestamps match the true timestamps via Wasserstein distance.

We found that in general, the mixture-of-experts neural-MTPPs had superior performance when compared to the decoder-only variants on all three datasets. Furthermore, the differences were more drastic when there was less events to condition on for sampling.

3 Conclusion

We address the problem of personalization in discrete-event continuous time sequences by proposing a new neural MTPP model based on a variational mixture-of-experts autoencoder approach. Experimental results on three large real-world event datasets demonstrate that the proposed approach systematically outperforms well-known alternative models, across a variety of performance metrics, by leveraging sequence heterogeneity in an effective manner.

References

- [1] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift reddit dataset. *arXiv preprint arXiv:2001.08435*, 2020.
- [2] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: Elementary Theory and Methods*. Springer Science & Business Media, 2007.
- [3] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [4] J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.
- [5] J. Ni, J. Li, and J. McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the EMNLP-IJCNLP Conference*, pages 188–197, 2019.
- [6] Y. Shi, N. Siddharth, B. Paige, and P. Torr. Variational mixture-of-experts autoencoders for multi-modal deep generative models. In *Advances in Neural Information Processing Systems*, pages 15692–15703, 2019.