

# Learning Neural Models for Continuous-Time Sequences

Vinayak Gupta  
IIT Delhi

vinayak.gupta@cse.iitd.ac.in

## ABSTRACT

The large volumes of data generated by human activities such as online purchases, health records, spatial mobility *etc.* is stored as a sequence of events over a continuous time. Learning deep learning methods over such sequences is a non-trivial task as it involves modeling the ever-increasing event timestamps, inter-event time gaps, event types, and the influences between events – within and across different sequences. This situation is further exacerbated by the constraints associated with data collection *e.g.* limited data, incomplete sequences, privacy restrictions *etc.* With the research direction described in this work, we aim to study the properties of continuous-time event sequences (CTES) and design robust yet scalable neural network-based models to overcome the aforementioned problems. In this work, we model the underlying generative distribution of events using marked temporal point processes (MTPP) to address a wide range of real-world problems. Moreover, we highlight the efficacy of the proposed approaches over the state-of-the-art baselines and later report the ongoing research problems.

### ACM Reference Format:

Vinayak Gupta. 2021. Learning Neural Models for Continuous-Time Sequences. In *Proceedings of the First International Conference on AI-ML Systems, October 21–23, 2021, Bangalore, India*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/XXXX.XXXX>

## 1 MOTIVATION

Continuous-time event sequences (CTES) have become pervasive across many applications ranging from healthcare [12], traffic [7, 19], social networks [3, 4], finance [8, 13], and sensor networks [10, 15]. Unlike images and text, the data quality of CTES is highly susceptible to the collection process *i.e.* a few missing events or shorter sequences can significantly hamper the data quality and consequently the performance of the neural models trained on such datasets. This situation is further exacerbated by privacy restrictions such as GDPR [2] and thus, overcoming the drawbacks in dataset quality is a non-trivial task unaddressed in the past literature.

## 2 PRELIMINARIES: MTPP

Marked temporal point processes (MTPP) are stochastic processes realized by a sequence of asynchronous events in continuous-time. We represent an MTPP  $\mathcal{S}_k = \{e_i = (m_i, t_i) | i \in [k], t_i < t_{i+1}\}$ ,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

AIMLSYS '21, October 21–23, 2021, Bangalore, India.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

ACM ISBN 978-1-4503-8446-9/21/11...\$15.00

<https://doi.org/10.1145/XXXX.XXXX>

where  $t_i \in \mathbb{R}^+$  is the occurrence time and  $m_i \in C$  is the discrete mark associated with the  $i$ -th event, with  $C$  as the set of all discrete marks. Here,  $\mathcal{S}_k$  denotes the sequence of first  $k$  events and we represent the inter-arrival times as,  $\Delta_{t,k} = t_k - t_{k-1}$ . If the sequences have a spatial component *i.e.*  $e_i = (m_i, t_i, d_i)$  then we represent the inter-event spatial differences as  $\Delta_{d,k} = d_k - d_{k-1}$ .

## 3 RESEARCH DIRECTIONS

In this section, we highlight the CTES data related problems addressed in our research plan and the proposed solutions. We classify them into the following categories:

### 3.1 Incomplete Sequences

Traditional models and inference methods for CTES assume a complete observation scenario *i.e.* the events in a sequence are completely observed with no missing events – an ideal setting and rarely applicable in real-world applications. Undeniably, any data collection procedure may not capture some events due to crawling and privacy restrictions by online platforms. As mentioned in Section 1, these missing events can drastically affect the quality of data and hence, the prediction performance of the learned model.

**Proposed Solution.** In Gupta et al. [7], we present an unsupervised model and inference method for learning neural MTPP over CTES with missing events. Specifically, we design a *coupled*-MTPP approach that first models the generative processes of both – observed events and missing events – where the missing events are represented as latent-random variables. Later, we jointly learn the distribution of all events via MTPP using variational inference.

The proposed model IMTPP (Intermittently-observed Marked Temporal Point Processes) models the generative distribution of observed events ( $e_k$ ) and missing events ( $e_r$ ) using MTPP denoted as  $p(\bullet)$  and  $q(\bullet)$  respectively. For generating future observed events, we follow a standard sampling procedure [13], however, the generation process of missing events is conditioned on the history as well as the *next* observed event. Specifically, we sample missing events between two observed events, till we reach the future observed event *i.e.* the MTPP are –  $p(e_k)$  and  $q(e_r | e_{k+1})$  respectively. We learn the parameter by maximizing a variational lower bound or evidence lower bound (ELBO) of the log-likelihood.

$$\mathbb{E}_q \sum_{k=0}^{K-1} \log p(e_{k+1}) - \sum_{k=0}^{K-1} \text{KL} \left[ q(e_r | e_{k+1}) || p_r(e_r) \right], \quad (1)$$

where  $p_r, \text{KL}$  denote prior MTPP and KL-divergence respectively. **Results.** We evaluated IMTPP across five datasets from different domains ranging from Amazon movies, Amazon toys, Twitter, Foursquare, and Stack-Overflow. To summarize the results, our observations were: (1) for predicting the mark and time of events in the test set, IMTPP outperformed the state-of-the-art

approaches [3, 13, 19], by up to 8% across all datasets, (2) for forecasting future events in a sequence, the gains IMTPP had over other baselines we consistent even for farther predictions, and (3) a scalability analysis over datasets with millions of events showed that other CTES models took up to 24 hours training times, whereas the times for IMTPP were under 5 hours.

**Ongoing Work and Challenges.** As an extension to IMTPP, we addressed the problem of imputing missing events in a sequence. Specifically, we sampled missing events using the posterior MTPP and evaluated over synthetically deleted events in a sequence. The results, including qualitative analysis, showed that the new approach outperforms other baselines and opens up new applications of neural models and CTES. This extension is under review in a reputed journal. A major challenge with imputations is to estimate the total number of missing events between each observed event.

### 3.2 Limited Training Data

The problem of limited training data is ubiquitous in all real-world applications ranging from recommender systems [16], vision [14], spatial models [17] *etc.* In contrast to incomplete data, defined in Section 3.1, we regard *limited* data as the problem of data-scarcity *i.e.* the available data is assumed to be complete but the volume is insufficient to effectively train a deep neural network.

**Proposed Solution.** Transfer learning (TL) has long been proposed as a feasible solution to overcome limited data problems [18]. Accordingly, in Gupta and Bedathur [6], we present a transfer approach for training neural MTPP on a data-rich dataset and fine-tune the model parameters on a data-scarce dataset. In detail, we consider the problem of mobility prediction, wherein we have spatial trajectories of users across different regions and regions with data-rich mobility sequence as a *source* region and the *target* region with scarce mobility data.

The proposed model REFORMD(**R**eusable **F**lows **f**or **M**obility **D**ata), learns the spatial and temporal distribution of a user trajectory using normalizing flows(NFs) [11]. We make the trained NFs *invariant* to a region, by restricting our model to learn the distribution of inter-event time intervals,  $\Delta_{t,\bullet}$ , and spatial distances,  $\Delta_{d,\bullet}$ . Since these features are independent of the underlying region, the trained NFs can be easily fine-tuned for any mobility sequence. Moreover, we use a *log-normal* flow to model both distributions *e.g.* for time, the generative distribution for future events is learned as:

$$p_t(\Delta_{t,k+1}|s_k) = \text{LogNormal}(\mu_t(s_k), \sigma_t^2(s_k)), \quad (2)$$

with  $[\mu_t(s_k), \sigma_t^2(s_k)] = [W_1 s_k + b_1, W_2 s_k + b_2]$  denote the *mean* and *variance* of the time distribution.  $s_k$  is the output of neural MTPP,  $W_\bullet$  and  $b_\bullet$  are trainable parameters. We sample the time of future events as  $\Delta_{t,k+1} \sim p_t$  and a similar procedure is followed for spatial flows. All model parameters are trained by maximizing the location recommendation accuracy and the likelihood of time and distance of events, Moreover, we use a standard procedure to fine-tune the model parameters on the sequences from the target region.

**Results.** We evaluated our model across eight mobility datasets from the US and Japan and our observations were: (1) for location recommendation and event-time prediction, REFORMD outperformed other MTPP models by up to 20% and 23% respectively (2) It

also demonstrated better and faster convergence on target dataset than other approaches, and (3) even in the absence of spatial flows, we outperform other methods by 3% for item recommendation and 14% for time prediction across four datasets from Amazon.

**Ongoing Work and Challenges.** A crucial drawback of our proposed model is its standard transfer learning procedure, whereas, modern neural models deploy a meta-learning [5] procedure to transfer model parameters. Therefore as future work, we plan to combine meta-learning with NF to design robust MTPP models.

### 3.3 Latent Features

The latent features that play a curcial role in CTES are inter-event and inter-sequence relationships. Farajtabar et al. [4] show that these relationships imitate an information diffusion process *e.g.* social preferences, community formations *etc.* and can be captured using MTPP models. Accordingly, we address the problem of community detection in the absence of the social network of a user. Specifically, we assign communities to users in a network based on their mobility preferences.

**Proposed Solution.** In our work [9] with Ankita, a former Ph.D. student, we devise a spatial-temporal point process method to learn the diffusion process to assign communities to users in the network. In detail, we train a self-exciting MTPP for each user with a common excitation matrix. Later, we assign a community to a user based on their personal MTPP and the shared matrix. We learn the latent community of users and our model parameters using stochastic variational inference.

**Results and Challenges.** The results across two spatial mobility datasets show that our model achieves improvements of up to 27% in location prediction and 8% for community prediction in comparison to other neural models. Though we use certain heuristics to estimate the user community, evaluating it in the absence of true labels is still an open problem.

## 4 ONGOING WORK

Here we highlight the ongoing works that are under review.

**Time Series as Graphs.** In this work, we represent a time series as a graph with weighted temporal edges. Later, we learn the dynamics of sequences using graph neural networks. Moreover, we overcome data scarcity using a meta-learning algorithm.

**Retrieving CTES.** Due to the disparate nature of CTES with events containing marks and if applicable, spatial features, the problem of sequences retrieval has been left undressed by the past literature. In this work, we propose a self-attention MTPP model for retrieving a similar sequence from a corpus, given a query sequence.

**Time Series and Databases.** In [1] with Garima, a former Ph.D. student, we learn the embedding of entities in temporally evolving relational databases.

## 5 CONCLUSION

In this paper, we present the research directions, possible data-related problems, and solutions for learning neural models on continuous-time sequences. In addition, we highlighted the ongoing works that are under review. As possible progress till October, we plan to get better results for meta-learning-based training of MTPP and notifications for the papers under review.

## REFERENCES

- [1] Siddhant Arora, Vinayak Gupta, Garima Gaur, and Srikanta Bedathur. 2021. BERT Meets Relational DB: Contextual Representations of Relational Databases. *arXiv preprint arXiv:2104.14914* (2021).
- [2] European Commission. 2018. 2018 Reform of EU Data Protection Rules. Available At: [https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes\\_en.pdf](https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf).
- [3] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *KDD*.
- [4] Mehrdad Farajtabar, Yichen Wang, Manuel Gomez Rodriguez, Shuang Li, Hongyuan Zha, and Le Song. 2015. COEVOLVE: A Joint Point Process Model for Information Diffusion and Network Co-evolution. In *NeurIPS*.
- [5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*.
- [6] Vinayak Gupta and Srikanta Bedathur. 2021. Region Invariant Normalizing Flows for Mobility Transfer. *arXiv preprint arXiv:2109.05738* (2021).
- [7] Vinayak Gupta, Srikanta Bedathur, Sourangshu Bhattacharya, and Abir De. 2021. Learning Temporal Point Processes with Intermittent Observations. In *AISTATS*.
- [8] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. In *Biometrika*.
- [9] Ankita Likhyan, Vinayak Gupta, PK Srijith, P Deepak, and Srikanta Bedathur. 2020. Modeling Implicit Communities from Geo-tagged Event Traces using Spatio-Temporal Point Processes. In *WISE*.
- [10] Sonam Maurya, Vinayak Gupta, and Vinod Kumar Jain. 2017. LBRR: Load Balanced Ring Routing Protocol for Heterogeneous Sensor Networks with Sink Mobility. In *WCNC*.
- [11] Danilo Rezende and Shakir Mohamed. 2015. Variational inference with normalizing flows. In *ICML*.
- [12] Marian-Andrei Rizoiu, Swapnil Mishra, Quyu Kong, Mark Carman, and Lexing Xie. 2018. SIR-Hawkes: on the Relationship Between Epidemic Models and Hawkes Point Processes. In *WWW*.
- [13] Oleksandr Shchur, Marin Bilos, and Stephan Günnemann. 2020. Intensity-Free Learning of Temporal Point Processes. In *ICLR*.
- [14] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. 2019. Meta-Transfer Learning for Few-Shot Learning. In *CVPR*.
- [15] Daniela Tulone and Samuel Madden. 2006. PAQ: Time series forecasting for approximate query answering in sensor networks. In *EWSN*.
- [16] Manasi Vartak, Arvind Thiagarajan, Conrado Miranda, Jeshua Bratman, and Hugo Larochelle. 2017. A Meta-Learning Perspective on Cold-Start Recommendations for Items. In *NeurIPS*.
- [17] Huaxiu Yao, Yiding Liu, Ying Wei, Xianfeng Tang, and Zhenhui Li. 2019. Learning from Multiple Cities: A Meta-Learning Approach for Spatial-Temporal Prediction. In *WWW*.
- [18] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How Transferable Are Features in Deep Neural Networks?. In *NeurIPS*.
- [19] Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. 2020. Transformer Hawkes Process. In *ICML*.