CuFun: Revolutionizing Temporal Point Process Modeling with Cumulative Distribution Functions

Maolin Wang
City University of Hong Kong
Hong Kong SAR, China
morin.wang@my.cityu.edu.hk

Ruocheng Guo
Independent Researcher
Hong Kong SAR, China
rguo.asu@gmailãĂĆcom

Xiangyu Zhao
City University of Hong Kong
Hong Kong SAR, China
xy.zhao@cityu.edu.hk

Abstract

Temporal Point Processes (TPPs) are crucial for modeling event sequences in diverse domains such as social networking and e-commerce. These processes utilize interactions and transactions to identify patterns and predict future trends. Despite their utility, the complexity of the data often hinders the accuracy of these predictions. Recent advancements have merged Neural Networks with TPPs, creating sophisticated models capable of handling complex, nonlinear temporal data. However, challenges persist regarding the flexibility in modeling intensity functions, the computational overhead of integral calculations, and the capture of long-range dependencies. We introduce the CuFun model, a novel TPP approach employing a monotonic neural network to represent the Cumulative Distribution Function (CDF). This model enhances adaptability and accuracy by incorporating past events as scaling factors, simplifies log-likelihood calculations, and extends the model's applicability beyond conventional density function constraints. CuFun excels in capturing long-range temporal patterns, making it particularly effective for precise forecasting in real-world applications involving extended temporal sequences. Our proposed CDF-based TPP model integrates past event data to improve future predictions.

Keywords

Multimodal Sequential Recommendation, Information Flow Control, Cross-Modal Alignment, Entropy-Aware Fusion

ACM Reference Format:
Maolin Wang, Ruocheng Guo, and Xiangyu Zhao. 2025. CuFun: Revolutionizing Temporal Point Process Modeling with Cumulative Distribution Functions. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1 (KDD '25), August 3–7, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/xxxxx.xxxxxxxxx

1 Introduction

Event sequences play a crucial role in various domains, notably in social networking [5, 17, 25] and e-commerce [6, 27]. They are instrumental in advancing recommendation systems [5, 6] and information retrieval strategies [27]. Each user action on a social network or e-commerce transaction is an event that provides insights into behavioral patterns. Analyzing these events helps predict future patterns, enhancing the predictive capabilities of recommendation systems. A key challenge in this field is accurately forecasting future occurrences based on these event patterns [21, 24, 26].

Temporal Point Processes (TPPs) [1, 3], with their robust theoretical foundation, have become increasingly popular for modeling these temporal event sequences. TPPs treat time intervals between events as random variables, providing a more nuanced representation than traditional discrete-time models. The *(conditional) intensity function*, a core concept in TPPs, has been parameterized in various forms, including Stationary and Non-stationary Poisson processes[11], Hawkes processes [8, 13], and other processes [1, 10, 18]. Despite their theoretical robustness, specific TPP modeling methods sometimes struggle to align with the complexities of real-world phenomena [9, 14, 15]. This misalignment is evident in scenarios where existing models, such as the Hawkes process, which typically assumes a positive excitation effect from past events, do not accurately represent real-world dynamics, such as inhibitory effects in dietary choices [9].

Neural Networks integrated with Temporal Point Processes (TPPs) have advanced modeling of temporal data. However, key challenges remain:

- **Computational Instability:** Integral calculations of intensity functions face numerical challenges, particularly in maintaining normalized density functions [15, 19].
- **Limited Extreme-frequency Handling:** Traditional intensity function approaches struggle with abrupt changes and extreme-frequency events [4].
- **Long-Range Dependencies:** Current models inadequately capture extended temporal patterns crucial for real-world applications [20, 22, 23].

In addressing these limitations, we present the **Cu**mulative Distribution **Fun**ction-based Temporal Point Process (**CuFun**), a pioneering model that innovatively focuses on directly modeling the CDF within TPPs. To our knowledge, CuFun is the first to employ a monotonic neural network for CDF representation, utilizing past events as a scaling factor to enhance the prediction of future events [12, 16].

2 Methodology

This section delineates the methodologies employed in our study, focusing on the intricate relationships among key functions in Temporal Point Processes (TPPs) and their implications in modeling event timings. We introduce novel approaches to parameterize the Cumulative Distribution Function (CDF) using Recurrent Neural Networks (RNN) and Monotonic Neural Networks (MNN), thereby enhancing the accuracy and numerical stability of our TPP models.

KDD '25, Toronto, ON, Canada

ACM ISBN xxxxxxxxxx
https://doi.org/xxxxx.xxxxxxxxx

Relationships among Functions in TPP.

To begin, we first explore the fundamental relationships among key functions in Temporal Point Processes (TPPs), emphasizing their interdependencies and pivotal roles in accurately modeling the timing of events. In the realm of temporal point processes, relationships among various functions are pivotal. Given the survivor function $S^{(\tau)}$ and the density function $p^{(\tau)}$, the intensity function $\lambda^{(t)}$ is articulated as per the equation [3]:

$$\lambda^{(t)} = \frac{p^{(\tau)}}{S^{(\tau)}}. \tag{1}$$

A key function associated with the survivor function is the cumulative distribution function (CDF). To elucidate, the interrelationship among the CDF $F^{(\tau)}$, survivor function $S^{(\tau)}$, and density function $p^{(\tau)}$ is delineated as follows:

$$\begin{cases} S^*(\tau) = 1 - F^*(\tau), \\ p^*(\tau) = \frac{\partial F^*(\tau)}{\partial \tau}. \end{cases} \tag{2}$$

The primary objective here is to model the CDF, accounting for historical event data. This task bifurcates into: firstly, devising a method to encapsulate past events and their integration into the main network; and secondly, ensuring the network's output corresponds to a valid CDF.

The Recurrent Neural Network (RNN) is frequently the go-to for sequential data to distill underlying patterns. In line with prevalent literature [4, 15, 19], we employ an RNN to model past event influences. Here, an input vector x_i serves as the RNN's input. A straightforward representation of x_i could be the inter-event interval, defined as $x_i = t_i - t_{i-1}$, or its logarithmic form $x_i = \log(t_i - t_{i-1})$. The RNN's hidden state $\boldsymbol{h_i}$ updates as follows:

$$\boldsymbol{h}_i = \max\left\{ f(W^h \boldsymbol{h_{i-1}} + W^x x_i + b^h), 0 \right\}, \tag{3}$$

where W^h, W^x, b^h, and f represent the recurrent weight matrix, input weight matrix, bias term, and an activation function, respectively. The hidden state $\boldsymbol{h}$ of the RNN is then conceptualized as a condensed vector representation of historical events. The subsequent step involves leveraging this historical data and structuring the model to produce a valid CDF.

Cumulative Distribution Function-based TPP

CDF modeling, devoid of integral computations and approximations, ensures numerical stability and improved effectiveness, particularly in capturing the intricate dynamics [7, 12] of rapid event sequences.

Inspired by the application of monotonic neural networks in probabilistic function modeling [2], we utilize the inter-event interval as the input to a feedforward neural network. Notably, the hidden state $\boldsymbol{h_{i-1}}$ receives $x_{i-1} = t_{i-1} - t_{i-2}$ as input, which is obtained independently of the monotonic neural network. In essence, $\boldsymbol{h_{i-1}}$ represents the **covariate** variable, while τ_i corresponds to the **response** variable. As shown in Figure **??**, the cumulative distribution function is represented through a monotonic neural network (MNN$(\cdot)$) as follows:

$$F^*(\tau = \tau_i | \boldsymbol{h}_{i-1}) = \text{MNN}\tau = \tau_i(\tau, \boldsymbol{h} = \boldsymbol{h_{i-1}}). \tag{4}$$

However, a generic neural network does not inherently ensure a valid cumulative distribution function output. To certify a function as a valid cumulative distribution function, it must adhere to constraints of monotonicity and boundedness at both positive and negative infinities, as delineated by the following three conditions:

$$\begin{cases} ① : \lim_{\tau \to -\infty} F(\tau | \boldsymbol{h}) = 0, \\ ② : \lim_{\tau \to +\infty} F(\tau | \boldsymbol{h}) = 1, \\ ③ : \frac{\partial F(\tau | \boldsymbol{h})}{\partial \tau} >= 0. \end{cases} \tag{5}$$

We now elucidate the entire network structure and explain how these constraints are fulfilled. The neural network consists of two primary components. Initially, as previously mentioned, past event information is encoded into a compact vector $\boldsymbol{h_{i-1}}$ via an RNN. Once $\boldsymbol{h_{i-1}}$ is obtained, it, along with τ_i, serves as the input to the monotonic neural network. The crux of our model lies in the second component, the monotonic neural network, highlighted in a red dotted box. Here, $\boldsymbol{h}$ and τ are processed through two distinct single-layer networks, ensuring congruent output sizes. Generally, in monotonic neural networks, these outputs are summed. However, significant magnitude disparities between the two can lead to a dominance of one value over the other. Our experimental results uncover significant variations in output magnitudes, as elaborated in the experimental section of our report. These findings suggest that the influence of past events might often be undervalued when predicting future occurrences. This underestimation could lead to substantial discrepancies in the accuracy and reliability of our predictive models, indicating a need for more nuanced consideration of historical data within our forecasting methodologies.

To address this, we replace the addition operation with an element-wise product, interpreting it as a scaling effect of historical information on future event predictions. This product then serves as the input to another feedforward neural network. In this network, the initial hidden layer is the result of the element-wise product, with each unit in the median and last layers applying activation functions *tanh* and *sigmoid* respectively, to produce the output. The distinct activation functions are denoted by varying colors.

The compliance with the three specified constraints is examined as follows: The positive weights from τ to the final output, in conjunction with the positive attributes of $\boldsymbol{h}$ as outlined in Eq. 3, inherently satisfy constraint ③. This ensures that the output function increases monotonically with τ. Moreover, employing a *sigmoid* activation function in the final unit secures adherence to constraints ① and ②, as inferred from ③. As a result, the network's output can approximate the cumulative distribution function. **Such a modeling approach circumvents the need for integral computations, relying instead on numerically stable differentiation operations.**

3 Conclusion

This paper presents CuFun, a novel approach to temporal point process modeling that addresses several critical challenges in the field. Through direct CDF modeling and the incorporation of a historical information scaling mechanism, our method demonstrates significant improvements over traditional approaches. The integration of monotonic neural networks for CDF representation marks a departure from conventional intensity function-based methods, offering both theoretical elegance and practical advantages.

References

[1] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. 2008. *Survival and event history analysis: a process point of view.* Springer Science & Business Media.

[2] Pawel Chilinski and Ricardo Silva. 2020. Neural likelihoods via cumulative distribution functions. In *Conference on Uncertainty in Artificial Intelligence*. PMLR, 420–429.

[3] Daryl J Daley and David Vere-Jones. 2003. An introduction to the theory of point processes, volume 1: Elementary theory and methods. *Verlag New York Berlin Heidelberg: Springer* (2003).

[4] 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 *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1555–1564.

[5] Bin Guo, Yasan Ding, Lina Yao, Yunji Liang, and Zhiwen Yu. 2020. The future of false information detection on social media: New perspectives and trends. *ACM Computing Surveys (CSUR)* 53, 4 (2020), 1–36.

[6] Shashank Gupta and Manish Bansal. 2023. A Neural Bag-of-Words Point Process Model for User Return Time Prediction in E-commerce. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 177–181.

[7] Osman Hasan and Sofiene Tahar. 2007. Verification of probabilistic properties in HOL using the cumulative distribution function. In *Integrated Formal Methods: 6th International Conference, IFM 2007, Oxford, UK, July 2-5, 2007. Proceedings 6*. Springer, 333–352.

[8] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.

[9] Kiyoshi Kanazawa and Didier Sornette. 2023. Asymptotic solutions to nonlinear Hawkes processes: A systematic classification of the steady-state solutions. *Physical Review Research* 5, 1 (2023), 013067.

[10] John Frank Charles Kingman. 1992. *Poisson processes*. Vol. 3. Clarendon Press.

[11] John Frank Charles Kingman. 2005. P oisson processes. *Encyclopedia of biostatistics* 6 (2005).

[12] Qi Li and Jeffrey S Racine. 2008. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics* 26, 4 (2008), 423–434.

[13] Thomas Josef Liniger. 2009. *Multivariate hawkes processes.* Ph. D. Dissertation. ETH Zurich.

[14] Maya Okawa, Tomoharu Iwata, Takeshi Kurashima, Yusuke Tanaka, Hiroyuki Toda, and Naonori Ueda. 2019. Deep mixture point processes: Spatio-temporal event prediction with rich contextual information. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 373–383.

[15] Takahiro Omi, Kazuyuki Aihara, et al. 2019. Fully neural network based model for general temporal point processes. In *Advances in Neural Information Processing Systems*. 2122–2132.

[16] J Orbach. 1962. Principles of neurodynamics. Perceptrons and the theory of brain mechanisms. *Archives of General Psychiatry* 7, 3 (1962), 218–219.

[17] Brian A Primack, Ariel Shensa, Jaime E Sidani, César G Escobar-Viera, and Michael J Fine. 2021. Temporal associations between social media use and depression. *American journal of preventive medicine* 60, 2 (2021), 179–188.

[18] Patricia Reynaud-Bouret and Sophie Schbath. 2010. Adaptive estimation for Hawkes processes; application to genome analysis. (2010).

[19] Oleksandr Shchur, Marin Biloš, and Stephan Günnemann. 2019. Intensity-Free Learning of Temporal Point Processes. *arXiv preprint arXiv:1909.12127* (2019).

[20] Oleksandr Shchur, Ali Caner Türkmen, Tim Januschowski, and Stephan Günnemann. 2021. Neural temporal point processes: A review. *arXiv preprint arXiv:2104.03528* (2021).

[21] Yan Wang, Zhixuan Chu, Tao Zhou, Caigao Jiang, Hongyan Hao, Minjie Zhu, Xindong Cai, Qing Cui, Longfei Li, James Y Zhang, et al. 2023. Enhancing event sequence modeling with contrastive relational inference. *arXiv preprint arXiv:2309.02868* (2023).

[22] Shuai Xiao, Mehrdad Farajtabar, Xiaojing Ye, Junchi Yan, Le Song, and Hongyuan Zha. 2017. Wasserstein learning of deep generative point process models. In *Advances in Neural Information Processing Systems*. 3247–3257.

[23] Shuai Xiao, Junchi Yan, Mehrdad Farajtabar, Le Song, Xiaokang Yang, and Hongyuan Zha. 2019. Learning time series associated event sequences with recurrent point process networks. *IEEE transactions on neural networks and learning systems* 30, 10 (2019), 3124–3136.

[24] Siqiao Xue, Yan Wang, Zhixuan Chu, Xiaoming Shi, Caigao Jiang, Hongyan Hao, Gangwei Jiang, Xiaoyun Feng, James Y Zhang, and Jun Zhou. 2023. Prompt-augmented temporal point process for streaming event sequence. *arXiv preprint arXiv:2310.04993* (2023).

[25] Yizhou Zhang, Karishma Sharma, and Yan Liu. 2021. Vigdet: Knowledge informed neural temporal point process for coordination detection on social media. *Advances in Neural Information Processing Systems* 34 (2021), 3218–3231.

[26] Xiangyu Zhao, Maolin Wang, Xinjian Zhao, Jiansheng Li, Shucheng Zhou, Dawei Yin, Qing Li, Jiliang Tang, and Ruocheng Guo. 2023. Embedding in Recommender Systems: A Survey. *arXiv preprint arXiv:2310.18608* (2023).

[27] Yuchen Zhou, Yanan Cao, Yanmin Shang, Chuan Zhou, Shirui Pan, Zheng Lin, and Qian Li. 2023. Explainable hyperbolic temporal point process for user-item interaction sequence generation. *ACM TOIS* 41, 4 (2023), 1–26.