ExPERT: Modeling Human Behavior under External Stimuli Aware Personalized MTPP

Ritvik Vij[*1], **Subhendu Khatuya**[*2], **Paramita Koley**[2], **Samik Datta**[2], **Niloy Ganguly**[2]

[1] Amazon [2] IIT Kharagpur, India
ritvikvi@amazon.com, {subha.cse143, paramita.2000, datta.samik}@gmail.com, niloy@cse.iitkgp.ac.in

Abstract

Marked Temporal Point Process (MTPP) – the de-facto sequence model for continuous-time event sequences – historically employed for modeling human-generated action sequences, lack awareness of external stimuli. In this study, we propose a novel framework developed over Transformer Hawkes Process (THP) to incorporate external stimuli in a domain-agnostic manner. Furthermore, we integrate personalization into our framework by employing language model-based representations of user and event descriptions, which is essential for modeling human-generated action sequences. Towards evaluating the efficacy, we put together a comprehensive benchmark comprising 5 datasets (2 novel additions, and 3 repurposed from existing open datasets) harvested from several domains, spanning education, e-commerce, online payment, and discussion forum. On average, we achieve 9.35% gain in type-prediction accuracy and 7.38% reduction in time-prediction RMSE across all datasets over SOTA MTPP baselines. We demonstrate the superior performance of our proposed model through extensive ablations and showcasing its ability to capture complex combinations of external stimuli in a synthetic set up.

Introduction

Continuous-time event sequences are ubiquitous in nature and society: e.g., opinions expressed in Online Social Network (De et al. 2016), buy and sell transactions in Stock Exchanges (Bacry, Mastromatteo, and Muzy 2015), earthquakes and tremors in geo-physics (Touati, Naylor, and Main 2014). In general, any system emitting a sequence of asynchronous events localized in time matches this description. In such sequences, each event is generally associated with a timestamp, a context, an actor, and an action information. For such sequences, MTPP (Daley and Vere-Jones 2007; Last and Brandt 1995) has emerged as a powerful framework and has been subsequently employed in a wide array of domains (De et al. 2016; Farajtabar et al. 2015; Rizoiu et al. 2017; Tabibian et al. 2019; Boyd et al. 2020; Gupta et al. 2022; Zhang et al. 2021; Qu et al. 2023; Xue et al. 2023; Sahebi et al. 2024).

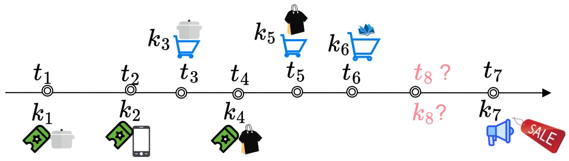


Figure 1: An illustrative example of the temporal prediction problem where the temporal sequence contains both the user-generated events (customer transactions) and a number of external stimuli (in form of various coupons across different categories). Our problem is to predict the time and type of the next user event (such as apparel shopping).

In particular, asynchronous event sequences representing human actions, – online payments (Manzoor and Akoglu 2017); sleep, eat, and exercise habit (Kurashima, Althoff, and Leskovec 2018); grocery runs (Bhagat et al. 2018; Sahebi et al. 2024); education (Yao et al. 2021b; Sahebi et al. 2024); human engagement in social media (Qu et al. 2023; Xue et al. 2023); user click behavior (Xue et al. 2023); mobile app usage behavior (Yang et al. 2023) – have been the subject of several recent studies, where MTPP served as the backbone sequence-model. While modeling the timestamp and the action information effectively, these attempts do not effectively model the context and the actor information.

Specifically, human actions are subject to external stimuli – e.g., incentives and reminders affecting payment behavior; coupons affecting e-commerce purchase behavior (we provide a detailed account in Figure 4b); assignments and examination deadlines affecting study behavior – whose presence and influence have been acknowledged (De, Bhattacharya, and Ganguly 2018; Zhang et al. 2021), but has not been incorporated into MTPP historically, except for (Manzoor and Akoglu 2017) and (Yao et al. 2021b). The approaches undertaken by them are domain-specific and labor-intensive, reminiscent of feature engineering, making them difficult to generalize across different domains. However, external stimuli are common in many important applications, highlighting the need for a general, domain-agnostic unified approach.

The response of an individual to such external stimuli is highly variable and is closely influenced by a range of personal factors. For instance, a student's reaction to a looming deadline or a customer's response to a specific product promotion is deeply contingent on their unique

[*] These authors contributed equally.

profile. This variation necessitates the integration of personalization within the modeling framework to accurately capture the effects of external stimuli on behavior. However, many existing MTPP (Marked Temporal Point Process) frameworks overlook this aspect. While models such as SSHP (Yao et al. 2021b) attempt to account for human factors by introducing numerous user-specific parameters, they face significant scalability challenges.

Finally, human-generated actions, such as customer transactions or student activities, as well as external stimuli like coupons or badges, are often accompanied by rich textual descriptions. When appropriately integrated into the model, these descriptions have the potential to significantly enhance the accuracy of predictions. Unfortunately, current MTPP models frequently disregard this valuable information.

In our research, we address all of the above mentioned limitations by developing a novel MTPP framework based on Transformer Hawkes process (Zuo et al. 2020) that incorporates external stimuli in a personalized, domain-agnostic manner, along with integrating the rich textual metadata available. This approach eliminates the need for excessive domain-specific customization while enhancing the model's applicability across diverse contexts. For the purpose of evaluation, we assemble a comprehensive array of 5 datasets, harvested from a wide variety of domains spanning online payment, online education (MOOC, as well as institute-wide online classes), and an online discussion forum on software engineering. Two of these datasets are carefully curated by us, whereas the other three have been repurposed (and reassembled) from existing open-domain datasets in order to feature both human action and external stimuli. We evaluate our approach on the next event prediction task, and benchmark it against several SOTA baselines, where our approach achieves salient performance gains on the next event prediction task. On average, we achieve **9.35%** gain in type-prediction accuracy and **7.38%** reduction in time-prediction RMSE across all datasets. Further ablation studies and extensive experiments, including those with synthetic data featuring complex combinations of external stimuli, demonstrate the effectiveness of our model.

Related Work and Preliminaries

MTPP. Traditionally, MTPP required an analytical form to specify the Conditional Intensity Function. This analytical form was customized to suit the nuances of the domain (Kurashima, Althoff, and Leskovec 2018), as well as to model factors such as external stimuli (Manzoor and Akoglu 2017; Yao et al. 2021a). Recently, the hand-written analytical form was replaced with a neural network, such as an RNN in (Du et al. 2016), and a Transformer in (Zhang et al. 2020; Zuo et al. 2020). Further customization, such as personalization, has also been built atop the neural network backbone (Boyd et al. 2020). However, neural MTPPs have not been extended to model external stimuli, to the best of our knowledge.

External Stimuli on Human Action. Effect of external stimuli, such as an incentive, has been quantitatively studied historically at a population level (Kusmierczyk and Gomez-Rodriguez 2018). Although acknowledged in recent literature on MTPP (De, Bhattacharya, and Ganguly 2018; Zhang et al. 2021), external stimuli have not yet been incorporated into the model, except for (Manzoor and Akoglu 2017; Yao et al. 2021b), which require domain-specific customization. In marketing, the effect of stimuli such as promotion has been extensively recorded (Elberg et al. 2019) but has not been distilled into personalized models.

Preliminaries

In this section, we recount the necessary background – first, the Marked Temporal Point Process (MTPP) framework, and next, Transformer Hawkes Process (THP).

Notation: We assume that a user, u, performs a sequence of actions, $\mathcal{H}_u$, influenced by a sequence of nudges and incentives, $\mathcal{C}_u$, from an agent. We model a user's actions, $\mathcal{H}_u$, and external stimuli, $\mathcal{C}_u$, as sequences $\{e_i := (t_i, k_i)_{i=1}^{L+L_C}\}$, where e_i has timestamp t_i and type $k_i \in [1, \ldots, K]$. For instance, in e-commerce, $\mathcal{H}_u$ includes purchase records and $\mathcal{C}_u$ includes coupons. Given a sequence $\mathcal{H}_u \bigcup \mathcal{C}_u$ of historical user actions and external stimuli, our problem is to predict the time and type of the next action taken by the user. Note that $\mathcal{H}_u$ is user-specific while $\mathcal{C}_u$ may be generic. For example, in e-commerce, mass promotional campaigns and special sale events are not personalized. [1] Figure 1 illustrates a concrete example of this problem formulation.

Marked Temporal Point Process: A Marked Temporal Point Process (MTPP) is a stochastic process where events are recorded as a sequence $\mathcal{H} := \big\{(t_1, k_1), \cdots, (t_n, k_n)\big\}$, with t_i denoting the time and k_i the type (mark) of each event. It is typically modeled using the Conditional Intensity Function (CIF), $\lambda(t)$, which embodies the probability that an event will occur within an infinitesimal temporal window, $(t, t+dt]$. The conditional expectation of the number of events in the interval $(t, t+dt]$, given past events, is $\lambda^*(t)dt$.

Transformer Hawkes Process: The Transformer Hawkes Process (THP) (Zuo et al. 2020) uses transformers (Vaswani et al. 2017) to model the conditional intensity $\lambda^*(t)$ of a marked temporal point process. It encodes sequences of timestamps and marks, processes these encodings through a transformer to obtain attention-based representations, and then applies a linear layer and softplus function to generate $\lambda^*(t)$. A separate linear layer and softmax are used to predict the mark distribution.

ExPERT: External Stimuli-aware Personalized MTPP

We introduce ExPERT, an external stimuli-aware personalized MTPP framework, built on the backbone of THP, having three main components: (1) personalization by encoding the context in the form of user and event metadata (2) external stimuli-aware query mechanism in attention layer, and (3) a causal mask to redirect external influence in the appropriate event type. Figure 2 depicts the overall architecture.

[1] To avoid cumbersome notation, when the context permits, we further drop the subscript u and denote $\mathcal{H}_u, \mathcal{C}_u$ simply with $\mathcal{H}, \mathcal{C}$.

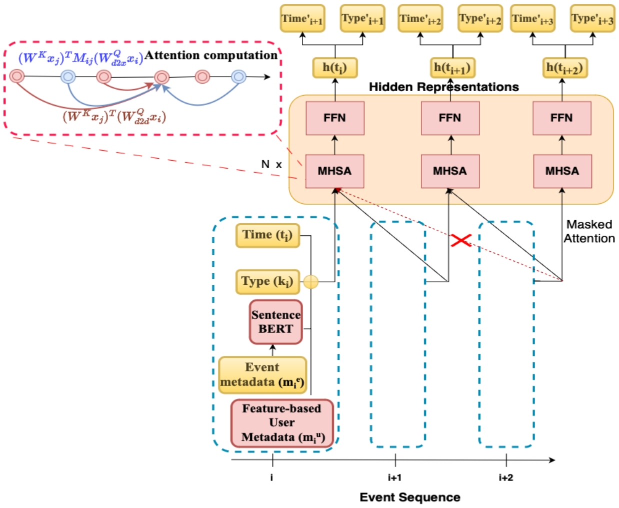
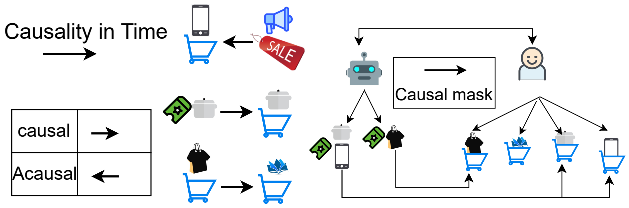


Figure 2: Overview of the model architecture of ExPERT. Time, type, and context of the sequence are initially encoded through an encoding layer. The encoding further passes through the attention layer to generate the hidden representation. The attention layer captures the effects of external stimuli and endogenous events through different networks (as shown in the red dotted block).



(a) Causality in time (b) Causality in type

Figure 3: Illustration of causality in ExPERT. Fig 3a shows that while external stimuli affect future user activity in general, they can also affect past user activity. Fig 3b shows that external stimuli (coupons in this case) often affect user events in specific categories only.

Encoding

The sequence of user-generated actions, $\mathcal{H}_u$ is encoded as follows. The timestamp, t_i, is embedded as $v_i \in \mathbb{R}^M$ with sinusoidal positional encoding (Vaswani et al. 2017). The type, k_i, is embedded with Uy_i, where $U \in \mathbb{R}^{M \times K}$ is a learnable projection matrix, and $y_i \in \{0,1\}^K$ is the one-hot vector representation of k_i. The embedding of the event, e_i, is expressed in an additive manner with $x_i := v_i + Uy_i$.

Personalization

We integrate personalization in our framework by incorporating user feature vectors and employing a vectorized representation of the textual event metadata in the model. Given sequence $\mathcal{H} \bigcup \mathcal{C}$, following (Wu et al. 2020), ExPERT attempts to encode the context as follows. Without loss of generality, we assume each sequence $\mathcal{H} \bigcup \mathcal{C}$ corresponds to a distinct user u with user-specific metadata $m^u \in \mathbb{R}^{M_u}$, where each element of m^u represents specific feature of the user-profile. m^u is passed through a linear layer to generate the user encoding $W_u m^u$, where $W_u \in \mathbb{R}^{M \times M_u}$ is the trainable parameter. Similarly, textual description, associated with particular event e_i (event metadata) are processed via Sentence-BERT (Reimers and Gurevych 2019) to generate vectorized representation $m_i^e \in \mathbb{R}^{M_e}$, which further passes through another linear layer $W_e \in \mathbb{R}^{M \times M_e}$ to generate event metadata encoding $W_e m_i^e$. The user and event metadata encoding are added with temporal and type encoding $(v_i + Uy_i)$ to generate the final event encoding $x_i = v_i + Uy_i + W_e m_i^e + W_u m^u$. Note that the same encoding layer is applied for both external stimuli and user actions for temporal, type, and contextual encoding.

External Stimuli Aware Attention

While computing the intermediate attention, the external stimuli and user events must be treated distinctly. Following (Yamada et al. 2020), ExPERT introduces external stimuli-aware query mechanism that maintains separate linear transformations for generating query embedding depending on the past influencing events. Towards this, ExPERT introduces two separate weight parameter matrices $W^Q_{d2x}, W^Q_{d2d} \in \mathbb{R}^{M \times M_K}$ for generating the query embeddings and $W^K \in \mathbb{R}^{M \times M_K}$, $W^V \in \mathbb{R}^{M \times M_V}$ for generating the key and value embedding respectively. Event encoding x_i employs W^Q_{d2d} when the past event x_j is user-generated event, and W^Q_{d2x} is employed when the past event x_j is external stimulus. Specifically, the embedding of individual events, x_i, are concatenated as $X = \{x_i\}_{i=1}^L$ and passed through a self-attention module, where the event encoding x_i generates s_i as

$$s_i = \sum_{j=1}^{L} \text{softmax}(\frac{w_{ij}}{\sqrt{M_K}}) W^V x_j$$

and the attention w_{ij} is computed as follows:

$$w_{ij} = \begin{cases} (W^K x_j)^T (W^Q_{d2d} x_i), & x_i, x_j \in \mathcal{H}, \\ (W^K x_j)^T (W^Q_{d2x} x_i), & x_i \in \mathcal{H}, x_j \in \mathcal{C} \end{cases} \tag{1}$$

Causal Mask

We notice that in certain domains, external stimuli affect only events with specific types. For example, in any customer transaction dataset, where coupons serve as external stimuli, they can only influence the transactions of their category. A coupon applicable to meat and dairy purchase can only influence the purchase of meat and dairy products and not other purchase categories like medicines. These causal and acausal relationships are also illustrated in figure 3. To utilize this domain knowledge, we propose using the following **causal mask**. Let $\mathbb{X}(k)$ denote the set of all internal event types that can be triggered by an external stimulus in type k. The attention computation is further modified as follows.

$$w^{\mathbb{M}}_{ij} = \begin{cases} w_{ij}, & x_i, x_j \in \mathcal{H}, \\ \mathbb{M}_{ij} w_{ij}, & x_i \in \mathcal{H}, x_j \in \mathcal{C} \end{cases} \tag{2}$$

where $w_{ij}^{\mathbb{M}}$ is the modified attention between events e_i and e_j. $\mathbb{M}_{ij} = 1$ if $k_i \in \mathbb{X}(k_j)$, otherwise 0. Here k_i and k_j are event types of the events e_i and e_j.

Moreover, while computing the attention, MTPP models generally do not consider the future events to ensure that events will not depend on future events. $w_{ij}^{\mathbb{M}}$ is masked if $i < j$. While this holds true for user actions in general, certain external stimuli occurring in the near future can influence past user actions. For example, student activity tends to surge as exams approach, assignments are typically submitted just before the deadline, and customers often redeem unused coupons just before expiry. To incorporate the impact of such impending external events, such as coupon expiry or assignment deadlines, on the past user actions, ExPERT unmasks attention from appropriate future external events, provided they are close enough in time.

Final Hidden Representation

ExPERT employs multiple attention layers whose outputs are aggregated and further processed to generate the final output. Specifically, n_{head} number of attention outputs are aggregated in final attention output as $S_{tot} = [S_1 || S_2 || \dots || S_{n_{\text{head}}}] W^O$ where $W^O \in \mathbb{R}^{HM_V \times M}$ is the aggregation matrix and $||$ is the concatenation operator. Finally S_{tot} passes through a feed-forward neural network to generate hidden representation $\mathcal{Z} = \text{ReLU}(S_{tot} W_1^{FC} + b_1) W_2^{FC} + b_2$, where $W_1^{FC} \in \mathbb{R}^{M \times M_H}, W_2^{FC} \in \mathbb{R}^{M_H \times M}, b_1 \in \mathbb{R}^{M_H}, b_2 \in \mathbb{R}^M$ are linear transformations. Each row $\boldsymbol{z}_i$ of $\mathcal{Z}$ represents the embedding of a particular event. In practice, the input is passed through attention modules sequentially through n_{layer} number of layers to capture high-level dependency.

Learning objective

The framework models the generating dynamics of human-generated events ($\mathcal{H}$) only. Events in $\mathcal{C}$ are generated independent of the internal dynamics ($\mathcal{C}$) and modeling of those events are out of scope of current work. Therefore, ExPERT excludes $\mathcal{C}$ during likelihood maximization, while exploiting it during the attention computation.

Let $\mathcal{H}_t$, and $\mathcal{C}_t$ be the human-generated event and external event sequence till time t (excluding t) respectively. $\lambda_k(t)$ be the conditional intensity of type k of the model, computed as $\lambda_k(t|\mathcal{H}_t \bigcup \mathcal{C}_t) = f_k\Big(\alpha_k \frac{t - t_j}{t_j} + \boldsymbol{w}_k^T \boldsymbol{z}_j + b_k\Big)$ where $t \in [t_j, t_{j+1})$ and $f_k(x) = \beta_k \log(1 + \exp(x/\beta_k))$ is the softplus function with softness parameter β_k. $\lambda(t)$ is the conditional intensity across all event types, computed as $\lambda(t|\mathcal{H}_t \bigcup \mathcal{C}_t) = \sum_{k=1}^{K} \lambda_k(t|\mathcal{H}_t \bigcup \mathcal{C}_t)$. The log-likelihood loss takes the following form.

$$\mathcal{L}_{\text{ll}}(\mathcal{H} \bigcup \mathcal{C}) = \sum_{e_j \in \mathcal{H}} \log \lambda_{k_j}(t_j | \mathcal{H}_{t_j} \bigcup \mathcal{C}_{t_j}) - \int_{t_1}^{t_L} \lambda(t|\mathcal{H}_t \bigcup \mathcal{C}_t) dt \tag{3}$$

Note that the presence of external stimuli in the history of λ, λ_{k_j} indicates the presence of external stimuli in attention computation. To learn the predictor parameters, cross-entropy loss for type predictions and squared loss for time predictions are employed. Let $\boldsymbol{k}_1, \boldsymbol{k}_2, \dots, \boldsymbol{k}_L$ be the ground-truth one-hot encodings for the event types. Then these two losses take the following form, where time and type prediction, $\hat{t}_j$ and $\hat{\boldsymbol{p}}_j$ follow THP.

$$\mathcal{L}_{\text{type}}(\mathcal{H} \bigcup \mathcal{C}) = \sum_{j > 1, e_j \in \mathcal{H}} -\boldsymbol{k}_j^T \log(\hat{\boldsymbol{p}}_j), \tag{4}$$

$$\mathcal{L}_{\text{time}}(\mathcal{H} \bigcup \mathcal{C}) = \sum_{j > 1, e_j \in \mathcal{H}} (t_j - \hat{t}_j)^2, \tag{5}$$

Note that ExPERT excludes external stimuli in the time and type prediction loss for the similar reason. Finally, for a given event sequence $\mathcal{H} \cup \mathcal{C}$, ExPERT optimizes the following weighted loss.

$$\mathcal{L}_{\text{total}}(\mathcal{H} \bigcup \mathcal{C}) = -\lambda_l \mathcal{L}_{\text{ll}}(\mathcal{H} \bigcup \mathcal{C}) + \lambda_k \mathcal{L}_{\text{type}}(\mathcal{H} \bigcup \mathcal{C}) + \lambda_t \mathcal{L}_{\text{time}}(\mathcal{H} \bigcup \mathcal{C}), \tag{6}$$

where λ_l, λ_k and λ_t are weights of the corresponding losses. For a input of N sequences $\mathcal{U} = \{\mathcal{H}_u \bigcup \mathcal{C}_u : u \in 1, ..., N\}$, the optimization problem becomes $\min_{\boldsymbol{\Theta}} \sum_{u=1}^{N} \mathcal{L}_{\text{total}}(\mathcal{H}_u \bigcup \mathcal{C}_u)$ where $\boldsymbol{\Theta}$ denotes all trainable weight matrices. Optimization is performed using stochastic gradient descent.

Experiments

In this section, we first outline the dataset, baselines, and evaluation setup. We then compare ExPERT against 7 TPP-based baselines across 5 real datasets, as summarized in Table 2. Next, we perform an ablation study to evaluate the impact of each component of ExPERT. We also highlight the diverse temporal patterns within the dataset and demonstrate ExPERT's effectiveness in capturing these patterns. Finally, through a comprehensive synthetic experiment, we showcase ExPERT's ability to respond to stimuli arising from a combination of multiple factors.

Dataset

The datasets selected for our benchmark should demonstrate three key characteristics: a) presence of human-generated activity sequence; b) presence of external stimuli; c) (optional) metadata describing the human actor and events. Towards this, we harvest 5 real-world datasets collected from a diverse set of domains – e-commerce, online education, online discussion forums, etc. Two of these datasets are curated by the authors, while the rest are reassembled from existing open-domain datasets. We summarize the data description and statistics in Table 1.

Dataset	#Seq	Avg Seq Length	Nudges/Seq (%)(Avg)	#Types (Nudge)	$\mu(\Delta t)$	$\sigma(\Delta t)$
Moodle-0.5M	1.1K	486	37.5%	15	0.037	0.19
Mooc-5M	21K	266	0.8%	2	0.05	0.51
SO-10M	32K	327	6.4%	1	0.30	9.43
Dunnhumby	2.5K	1079	3.8%	44	0.08	0.47
Transactions-6M	100K	67	25.4%	20	0.99	5.17

Table 1: Statistics of the datasets used in our experiments.

Moodle-0.5M (Curated by the authors) This dataset describes user activities collected across 10 courses taken during 2018-2021 from a university. The activities of professors and teaching assistants are marked as external stimuli, where student activities form user-generated events. Student activities include uploading submissions, checking grades, quiz attempts submitted, user profile viewed, etc., whereas external events include both targeted events, like grading a student or unlocking a student for late submission, and non-targetted events, like creating course module, etc.

Methods	RMSE					Type prediction accuracy				
	MOODLE	MOOC	SO	DUNNHUMBY	TRANSACTIONS	MOODLE	MOOC	SO	DUNNHUMBY	TRANSACTIONS
SSHP	0.3943	0.7556	2.9750	3.1650	4.0853	-	-	-	-	-
RMTPP	0.093	5.84	3.0568	3.4900	3.1205	36.64	63.08	58.45	63.49	50.97
Intensity RNN	0.098	6.31	3.0798	3.39	3.4010	55.26	63.55	57.73	63.72	52.63
SAHP	0.1770	3.2604	6.8131	1.574	4.8485	35.39	15.66	43.20	60.08	42.21
NHP	0.7096	1.297	5.000	0.7002	3.7910	28.99	29.37	19.81	60.04	45.73
AttNHP	0.7158	1.2937	4.2381	0.7196	3.7468	22.57	23.31	38.38	60.04	47.10
THP	0.1233	0.1006	2.2130	0.7070	3.3118	53.49	64.63	59.31	64.05	52.91
ExPERT	**0.0797**	**0.0956**	**2.0406**	**0.6990**	**2.5112**	**61.30**	**67.87**	**64.77**	**64.63**	**58.66**
RI(%)	+14.30%	+4.97%	+7.79%	+0.17%	+19.52%	+10.93%	+5.01%	+9.21%	+0.90%	+10.86%

Table 2: Comparative analysis of ExPERT against all baselines across all datasets in terms of both RMSE and type prediction accuracy. The best result is in **bold**, and the second-best is in underline. The last row RI(%) provides the relative improvement of our method over the best-performing baseline for each of the datasets. ExPERT consistently outperforms all the baselines.

Mooc-5M (repurposed from (Feng et al. 2019)). This dataset describes student activities in 247 online courses. Each sequence starts with the course-start event and ends with a course-completion event, both used as external stimuli. Human-generated activities include playing course videos, stating a discussion, posting a comment, etc (there are 22 such event types). Note that we only use a subset of the data corresponding to active students.

SO-10M (reassembled from (so2 2021), (Paranjape, Benson, and Leskovec 2017)) This dataset contains various user activities for a span of 2773 days in StackOverflow, the online question-and-answering website. There are three types of user events, namely 'answer-to-question', 'comment-to-question', and 'comment-to-answer', collected from (Paranjape, Benson, and Leskovec 2017). Moreover, the platform rewards each user with several badges at different times to promote user engagement. We combine the badges data (so2 2021) and user events (Paranjape, Benson, and Leskovec 2017) to create a unified user activity & badges dataset. The receipts of the badges are marked as external events here.

DUNNHUMBY (Gonen 2020) This dataset contains household-level transactions over two years from a group of 2500 households who are frequent shoppers at a retailer. Households were targeted with campaigns offering various coupons across diverse product categories. Here, coupons form external events, the customer transactions are marked as user events, and the item category serves as the event type. For multiple items from distinct categories in a shopping basket, distinct events are created with minimal perturbation in their times.

TRANSACTIONS-6M (Curated by the authors) This dataset contains user logs for a randomly sampled set of 100K customers over a period of 4 months on an online payment service in an emerging marketplace. Here, the user transactions include payment of phone bills, sending money to a business, shopping on e-commerce stores, etc. To drive user engagement, the platform rewards users with coupons that are tagged to a subset of the transaction categories and expire after a certain period. Here, customer transactions are marked as user events, whereas the coupons are marked as external events. Here, features as user age, engagement level, product purchase value, etc., are used as user metadata, and coupon features such as reward value, type of offer, popularity, are used as event metadata for external events.

For each of these datasets, the task involves predicting the timestamp and type of the next event by the corresponding user, which reduces to predicting next student activity for MOOC-5M and MOODLE-0.5M, customer transaction for TRANSACTIONS-6M and DUNNHUMBY, and user activity for SO-10M. In all courses, except TRANSACTIONS-6M, we consider user ID as user metadata.

Baselines

We compare ExPERT against following baselines: 1) SSHP (Yao et al. 2021b), 2) RMTPP (Du et al. 2016), 3) SAHP (Zhang et al. 2020)), 4) Intensity-RNN (Xiao et al. 2017), 5) NHP (Mei and Eisner 2017), 6) AttNHP (Yang, Mei, and Eisner 2021), and 8) THP (Zuo et al. 2020). Among them, SSHP trains a personalized Hawkes model with a domain-specific assumption of external stimuli, RMTPP and Intensity RNN rely on RNN-like structures, NHP employs continuous-time LSTM, and SAHP, THP, and AttNHP employ an attention-based mechanism for encoding past influences. (Refer to Supplementary for details.)

Evaluation Setup

For each dataset, we scale all event times to a range of $[0, 100]$. Next, we split our dataset into two random equally-sized sets of sequences A and B and construct the train and test data as follows - (1) Train : (Full sequences in A) + (first 70% events of all sequences in B) (2) Test : (last 30% events of all sequences in B). We keep the full sequences in train for 50% instances as some external events occur in the tail end of the sequence (e.g. course completion in the Mooc-5M dataset), and we need to train our model on capturing the influence of such external events. For each event in the test set, we predict the timestamp and type of the next event, using history of observations for that sequence till that event. We evaluate event type prediction by type prediction accuracy and event time prediction by root mean square error (RMSE).

Architectural Details. For training, we employ Adam optimizer with learning rate $1e^{-3}$ and batch size 64 for 30 epochs, where we select the model with least training error. The dimensions are set to 512, $M = M_K = M_V = 512$. Layer normalization and dropout of 0.1 are employed at the multihead attention and feed-forward layer. Number of attention heads and attention layers are set to 4. The feedforward network consists of $M_H = 1024$ hidden nodes with GeLU activation. We set softness parameter $\beta_k = 1$ and $\alpha_k = -0.1, \forall k$. Moreover we set the loss weights $(\lambda_l, \lambda_k, \lambda_t) = (1, 1, 0)$ while optimizing for type prediction and $(\lambda_l, \lambda_k, \lambda_t) = (1, 0, 1)$ while optimizing for time prediction, inspired from (Park et al. 2022). (Refer to Supplementary for detailed evaluation setup of the baselines.)

Comparative Analysis

Here, we compare ExPERT against the TPP-based baselines described above. Note that SSHP does not model marks; therefore, its type prediction accuracy is omitted. Table 2 summarizes the

Methods	RMSE					Type prediction accuracy				
	Moodle	Mooc	SO	Dunnhumby	Transactions	Moodle	Mooc	SO	Dunnhumby	Transactions
THP	0.1233	0.1006	2.213	0.7070	3.3118	53.49	64.63	59.31	64.05	52.91
ExPERT$_{\text{P(User)}}$	0.0942	0.0982	2.0900	0.7068	3.0811	51.78	64.91	60.75	64.05	53.57
ExPERT$_{\text{Ext}}$	0.1035	0.0971	2.0794	0.7004	2.8723	57.28	64.87	63.48	64.37	56.18
ExPERT$_{\text{Ext+P(User)}}$	0.0798	0.0970	2.0492	0.7012	2.7123	61.20	67.54	63.98	64.37	57.03
ExPERT$_{\text{Ext+P(User)+CM}}$	$*$	0.0962	$*$	**0.6984**	2.6504	$*$	67.79	$*$	64.52	58.32
ExPERT$_{\text{Ext+P(User+Event)+CM}}$	**0.0797**	**0.0956**	**2.0406**	0.6990	**2.5112**	**61.30**	**67.87**	**64.77**	**64.63**	**58.66**

Table 3: Ablation study. We compare ExPERT with its variants and report RMSE and average accuracy for the time and type predictions. We use $*$ when the ablation is not applicable to the dataset. The best results are highlighted in bold.

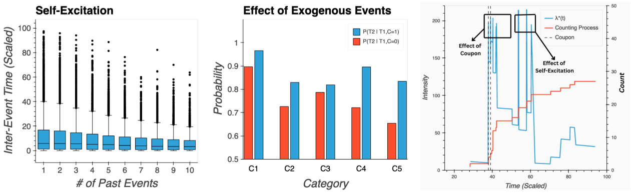
RMSE for the time predictions and the type prediction accuracies on the 5 real datasets. We observe that ExPERT achieves a substantial performance gain over THP, the next best-performing baseline across most datasets in terms of both time and type prediction. The relative improvements by ExPERT are particularly note-worthy for Moodle-0.5M, Transactions-6M, and SO-10M, significant for Mooc-5M and marginal for Dunnhumby. An observation is that sequences in the Moodle-0.5M, Transactions-6M, and SO-10M datasets are particularly rich in both external events and comprehensive textual descriptions, as outlined in Table 1, indicating that the impact of modeling external events and event metadata is more pronounced when the sequence contains a sufficient number of well-described external events. This hypothesis is further strengthened by the tiny improvement on the Dunnhumby dataset that is relatively sparse with external events in the sequences but also has the largest number of different types of external events.

Moreover, in these datasets, the nature of external events plays a significant role in shaping the future dynamics. For example, the external events available in the Moodle-0.5M data, such as assignment upload, grade upload, assignment deadline, etc. are very likely to trigger a major onset of events. Conversely, the Mooc-5M dataset is much sparser in external events, consisting of only the start and end date of the course, which has a relatively lower influence on events, potentially explaining the lower relative improvement.

Among the rest of the baselines, NHP and AttNHP mostly perform comparably with THP for time predictions, while performing much worse for type predictions. The rest of the baselines perform much worse than THP or ExPERT on almost all the datasets in both metrics. However, they perform comparatively better on Moodle-0.5M in terms of RMSE, which is considerably smaller in size, with small mean event inter-arrival times. Similarly, their type accuracy is considerably better on Dunnhumby, where some categories strongly dominate others. In summary, these methods fail to capture the dynamics of larger, complex datasets. SSHP provides unreliable performance in terms of time prediction error and does not provide type prediction. SSHP improves upon classic Hawkes by adding dynamic modules specifically designed for external stimuli in certain domains and fails to generalize to applications in other domains when the number and types of external stimuli vary. Moreover, because of scaling issues, SSHP is evaluated on clusters of sequences on larger datasets, reporting an average over the cluster, resulting in poor performances.

Ablation Study

We perform an ablation study to investigate the contribution of individual components of ExPERT. ExPERT comprises three main components, namely personalization (with user and event metadata), generation-aware attention for external events, and causal mask addition. Accordingly, we present results on ExPERT$_{\text{P(User)}}$, ExPERT$_{\text{Ext}}$, ExPERT$_{\text{Ext+P(User)}}$ (combining ExPERT$_{\text{P(User)}}$ and ExPERT$_{\text{Ext}}$), ExPERT$_{\text{Ext+P(User)+CM}}$ (combining ExPERT$_{\text{P(User)}}$, ExPERT$_{\text{Ext}}$ and Causal Mask) and ExPERT$_{\text{Ext+P(User+Event)+CM}}$ (which is essentially ExPERT).



(a) (b) (c)

Figure 4: Diverse temporal patterns in Transactions-6M dataset. Fig 4a shows self-excitation in data, as event inter-arrival time decreases with an increase in # past events. Fig 4b indicates the impact of external stimuli by showing an increment in the likelihood in presence of external stimuli. Fig 4c showcase the predicted intensity function portraying the effectiveness of ExPERT in capturing the effect of external stimuli and self-excitation.

Results. We compare the performance of the different components of ExPERT and the best-performing baseline THP and summarize the results in Table 3. Here, we observe all the versions, ExPERT$_{\text{P(User)}}$, ExPERT$_{\text{Ext}}$, and ExPERT$_{\text{Ext+P(User)}}$ improve over THP on all the datasets. Performance of ExPERT$_{\text{P(User)}}$ is generally better than THP, however inferior to external-influence aware variants, showing external-influence modeling to be more contributing factor than personalization. On the other hand, for almost all cases, ExPERT$_{\text{Ext+P(User)}}$ has nonnegative performance gain on ExPERT$_{\text{Ext}}$, validating the efficacy of personalization along with distinct attention-mechanism for external stimuli modeling. Also, we observe the addition of an additional weight matrix for generating query embedding (from THP to ExPERT$_{\text{Ext}}$) results in maximum performance gain for all the datasets, showing better modeling of the information about the external events contribute maximum to the performance gain of ExPERT over THP. Moreover, employing type-aware causal mask in the attention layer (ExPERT$_{\text{Ext+P(User)+CM}}$) enables significant performance gain for relevant datasets, like Mooc-5M, Dunnhumby, and Transactions-6M, where type-specific external events are present. Finally, adding event metadata further boosts the performance in all datasets, showing the impact of integrating textual event descriptions to further improve the representation of events.

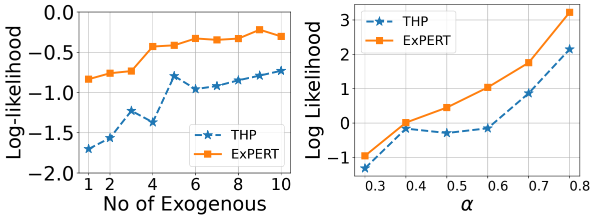
Data Analysis and Visualizations

Figure 4 (a), (b) illustrates some temporal patterns observed in Transactions-6M. (Refer to Supplementary for similar analysis on other datasets.) In Figure 4a, the presence of self-excitation in the data is emphasized through the distribution of next-event arrival times in an event category, varying the number of past events. Figure 4b depicts the probability of the next event's occurrence, with a blue bar representing the presence of a relevant external stimulus and a red bar indicating its absence. Across a randomly selected set of 5 event types, it shows that next event likelihood increases in the presence of external stimuli.

Figure 4c presents a visualization of predicted intensity function for a sequence from the Transactions-6M dataset, that captures self-excitation and external stimuli. Here, we observe huge peaks in the per-category intensity immediately after occurrence of the external stimulus (relevant coupon issuance), which validates that the model effectively captures the excitation due to the external stimuli (as observed in Figure 4b). After the coupon gets expired, the next sudden jumps in the predicted intensity is due to the self-excitation (as observed in Figure 4a) which ExPERT effectively captures.

Synthetic Experiments

Following the validation of our algorithm with real-world data across various domains, we extend our analysis by generating synthetic data in a controlled environment. This approach allows us to simulate complex combinations of external stimuli configurations that are control in real-world data. Here we systematically explore a range of configurations by controlling the type and count of external stimuli. Further, we introduce a hyperparameter α, for controlling the impact of the external stimuli. This approach provides a more granular understanding of our model's performance and robustness under different conditions.



(a) (b)

Figure 5: Performance comparison of ExPERT with THP in various synthetic settings. Figure 5a and 5b vary the count and impact of external stimuli respectively.

Vary the count of the external events. We gradually increase the number of external stimuli per sequence and observe how different methods perform against this variation. User behavior is generated using

$$\lambda(t) = \mu + \sum_{k=1}^{K} \mathbf{1}[\gamma_k t > t_k] b_k^{\frac{t-t_k}{s}} + \sum_{t_i<t} \alpha\beta \exp^{-\beta(t-t_i)} \quad (7)$$

We vary K, the number of external stimuli from 1 to 10. We keep the external stimuli parameters fixed at $\beta_k = 0.2, \gamma_k = 0.5$. t_k are chosen at fixed intervals from the sequence. Self-excitation parameters α, μ, β are fixed at $0.2, 0.4, 1$. Figure 5a presents the log-likelihood performance of ExPERT vs. THP. averaged across 500 sequences. We see that ExPERT outperforms THP across all configurations in terms of model fitting, with the margin being significantly large.

Vary self-excitation effect. We gradually increase the self-excitation effect with respect to external stimuli. Event intensities follows Eq. 7. More specifically, we vary α from 0.3 to 0.8, while setting $K = 2, (\beta_k, \gamma_k) = (0.2, 0.4)\forall k$. Other self-excitation parameters μ, β are fixed at 0.3, 1. Figure 5b summarizes the model fitting performance of ExPERT vs THP across 500 sequences. THP manages to almost close the gap between ExPERT and itself with increasing self-excitation, establishing that ExPERT is particularly useful over THP given there are abundant external influence in the event generation process.

External stimuli arising from complex effects. In this experiment, events are generated under the combination of the following influences, namely an exponentially decaying starting effect, periodic stimulations, self-excitation, and an approaching deadline. The intensity function takes the following form (Yao et al. 2021b).

$$\lambda(t) = \gamma^h (\sin(\frac{2\pi}{s}(t+p)) + c) + \gamma^o b^{\frac{t}{s}}$$

$$+\gamma^d (\frac{1}{\sqrt{2\pi v}(d - m - \frac{t}{s})} e^{-\frac{(\ln(d-m-\frac{t}{s}))^2}{v}}) + \sum_{x^\tau<t} \alpha\beta \exp^{-\beta(t-x^\tau)}$$

s is a scale parameter, d is the deadline, $d - m$ denotes the window in which the effect of an approaching deadline is prominent. v is standard deviation of log-normal distribution, α, β are self-excitation parameters. We sample each of the parameters from normal distributions with the following configuration. $A \sim \mathcal{N}(0.4, 0.1)$, $M \sim \mathcal{N}(0, 5)$, $\Gamma^d \sim \mathcal{N}(15, 3)$, $\Gamma^o \sim \mathcal{N}(5, 3)$, $\Gamma^h \sim \mathcal{N}(0.5, 0.1)$, $\boldsymbol{v} \sim \mathcal{N}(20, 10)$, $\boldsymbol{b} \sim \mathcal{N}(0.5, 0.3)$, $\boldsymbol{p} \sim \mathcal{N}(6, 4)$ and $\boldsymbol{c} \sim \mathcal{N}(1.2, 0.1)$. Events are simulated from corresponding intensity using the Ogata thinning algorithm (Ogata 1981). We keep the type of events 1.

We compare the time prediction performance of ExPERT with THP, the primary baseline, and find that ExPERT achieves an RMSE of 0.4821, representing a **22.65%** relative reduction compared to THP, which has an RMSE of 0.5845. The superiority of ExPERT demonstrates the ability of ExPERT to capture complex form of influences.

Conclusion

This work addresses personalized temporal modeling of human behavior under explicit external stimuli. Here, we have proposed a novel approach to explicitly include external stimuli in a THP-based framework in a personalized, domain-agnostic manner, along with a set of benchmark datasets with labeled external stimuli for this task. The ablation study establishes the fact that including external stimuli and personalization significantly boosts the performance, however, the impact of modeling external stimuli is more prominent. A thorough analysis of the dataset shows that ExPERT effectively models both self-excitation and responses to external stimuli. Additionally, we conducted an in-depth synthetic experiment to examine ExPERT's response to stimuli generated from a combination of multiple factors, revealing that it significantly outperforms its leading competitor. We believe this paper brings forward the ubiquity of external stimuli and takes a significant step at accurately representing that. This modeling success is expected to have a profound impact on various downstream tasks, such as designing recommendation systems, which will be pursued as part of our immediate future work.

References

2021. Stack Exchange Data Dump.

Bacry, E.; Mastromatteo, I.; and Muzy, J.-F. 2015. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01): 1550005.

Bhagat, R.; Muralidharan, S.; Lobzhanidze, A.; and Vishwanath, S. 2018. Buy it again: Modeling repeat purchase recommendations. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 62–70.

Boyd, A.; Bamler, R.; Mandt, S.; and Smyth, P. 2020. User-dependent neural sequence models for continuous-time event data. *Advances in Neural Information Processing Systems*, 33: 21488–21499.

Daley, D. J.; and Vere-Jones, D. 2007. *An introduction to the theory of point processes: volume II: general theory and structure.* Springer Science & Business Media.

De, A.; Bhattacharya, S.; and Ganguly, N. 2018. Demarcating endogenous and exogenous opinion diffusion process on social networks. In *Proceedings of the 2018 World Wide Web Conference*, 549–558.

De, A.; Valera, I.; Ganguly, N.; Bhattacharya, S.; and Gomez Rodriguez, M. 2016. Learning and forecasting opinion dynamics in social networks. *Advances in neural information processing systems*, 29.

Du, N.; Dai, H.; Trivedi, R.; Upadhyay, U.; Gomez-Rodriguez, M.; and Song, L. 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.

Elberg, A.; Gardete, P. M.; Macera, R.; and Noton, C. 2019. Dynamic effects of price promotions: Field evidence, consumer search, and supply-side implications. *Quantitative Marketing and Economics*, 17: 1–58.

Farajtabar, M.; Wang, Y.; Gomez Rodriguez, M.; Li, S.; Zha, H.; and Song, L. 2015. Coevolve: A joint point process model for information diffusion and network co-evolution. *Advances in Neural Information Processing Systems*, 28.

Feng, W.; Tang, J.; Liu, T. X.; Zhang, S.; and Guan, J. 2019. Understanding Dropouts in MOOCs. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence.*

Gonen, F. 2020. Dunnhumby - The complete journey.

Gupta, V.; Bedathur, S.; Bhattacharya, S.; and De, A. 2022. Modeling continuous time sequences with intermittent observations using marked temporal point processes. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(6): 1–26.

Kurashima, T.; Althoff, T.; and Leskovec, J. 2018. Modeling interdependent and periodic real-world action sequences. In *Proceedings of the 2018 world wide web conference*, 803–812.

Kusmierczyk, T.; and Gomez-Rodriguez, M. 2018. On the causal effect of badges. In *Proceedings of the 2018 world wide web conference*, 659–668.

Last, G.; and Brandt, A. 1995. Marked point processes on the real line. the dynamic approach, 1995.

Manzoor, E.; and Akoglu, L. 2017. Rush! targeted time-limited coupons via purchase forecasts. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1923–1931.

Mei, H.; and Eisner, J. M. 2017. The neural hawkes process: A neurally self-modulating multivariate point process. *Advances in neural information processing systems*, 30.

Ogata, Y. 1981. On Lewis' simulation method for point processes. *IEEE transactions on information theory*, 27(1): 23–31.

Paranjape, A.; Benson, A. R.; and Leskovec, J. 2017. Motifs in temporal networks. In *Proceedings of the tenth ACM international conference on web search and data mining*, 601–610.

Park, N.; Liu, F.; Mehta, P.; Cristofor, D.; Faloutsos, C.; and Dong, Y. 2022. Evokg: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, 794–803.

Qu, C.; Tan, X.; Xue, S.; Shi, X.; Zhang, J.; and Mei, H. 2023. Bellman meets hawkes: Model-based reinforcement learning via temporal point processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 9543–9551.

Reimers, N.; and Gurevych, I. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084.*

Rizoiu, M.-A.; Xie, L.; Sanner, S.; Cebrian, M.; Yu, H.; and Van Hentenryck, P. 2017. Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th international conference on world wide web*, 735–744.

Sahebi, S.; Yao, M.; Zhao, S.; and Feyzi Behnagh, R. 2024. MoMENt: Marked Point Processes with Memory-Enhanced Neural Networks for User Activity Modeling. *ACM Transactions on Knowledge Discovery from Data*, 18(6): 1–32.

Tabibian, B.; Upadhyay, U.; De, A.; Zarezade, A.; Schölkopf, B.; and Gomez-Rodriguez, M. 2019. Enhancing human learning via spaced repetition optimization. *Proceedings of the National Academy of Sciences*, 116(10): 3988–3993.

Touati, S.; Naylor, M.; and Main, I. G. 2014. Statistical modeling of the 1997–1998 Colfiorito earthquake sequence: Locating a stationary solution within parameter uncertainty. *Bulletin of the Seismological Society of America*, 104(2): 885–897.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wu, L.; Li, S.; Hsieh, C.-J.; and Sharpnack, J. 2020. SSE-PT: Sequential recommendation via personalized transformer. In *Proceedings of the 14th ACM Conference on Recommender Systems*, 328–337.

Xiao, S.; Yan, J.; Yang, X.; Zha, H.; and Chu, S. 2017. Modeling the intensity function of point process via recurrent neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Xue, S.; Wang, Y.; Chu, Z.; Shi, X.; Jiang, C.; Hao, H.; Jiang, G.; Feng, X.; Zhang, J.; and Zhou, J. 2023. Prompt-augmented temporal point process for streaming event sequence. *Advances in Neural Information Processing Systems*, 36: 18885–18905.

Yamada, I.; Asai, A.; Shindo, H.; Takeda, H.; and Matsumoto, Y. 2020. LUKE: deep contextualized entity representations with entity-aware self-attention. *arXiv preprint arXiv:2010.01057.*

Yang, C.; Mei, H.; and Eisner, J. 2021. Transformer embeddings of irregularly spaced events and their participants. *arXiv preprint arXiv:2201.00044.*

Yang, K.; Zhao, X.; Zou, J.; and Du, W. 2023. ATPP: A mobile app prediction system based on deep marked temporal point processes. *ACM Transactions on Sensor Networks*, 19(3): 1–24.

Yao, M.; Zhao, S.; Sahebi, S.; and Behnagh, R. F. 2021a. Relaxed clustered hawkes process for student procrastination modeling in moocs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4599–4607.

Yao, M.; Zhao, S.; Sahebi, S.; and Feyzi Behnagh, R. 2021b. Stimuli-sensitive Hawkes processes for personalized student procrastination modeling. In *Proceedings of the Web Conference 2021*, 1562–1573.

Zhang, P.; Iyer, R.; Tendulkar, A.; Aggarwal, G.; and De, A. 2021. Learning to Select Exogenous Events for Marked Temporal Point Process. *Advances in Neural Information Processing Systems*, 34: 347–361.

Zhang, Q.; Lipani, A.; Kirnap, O.; and Yilmaz, E. 2020. Self-attentive Hawkes process. In *International conference on machine learning*, 11183–11193. PMLR.

Zuo, S.; Jiang, H.; Li, Z.; Zhao, T.; and Zha, H. 2020. Transformer hawkes process. In *International conference on machine learning*, 11692–11702. PMLR.