

# Early Detection of User Exits from Clickstream Data: A Markov Modulated Marked Point Process Model

**Conference Paper****Author(s):**

Hatt, Tobias; Feuerriegel, Stefan

**Publication date:**

2020-04

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000404486>

**Rights / license:**

[Creative Commons Attribution 4.0 International](#)

**Originally published in:**

<https://doi.org/10.1145/3366423.3380238>

# Early Detection of User Exits from Clickstream Data: A Markov Modulated Marked Point Process Model

Tobias Hatt  
ETH Zurich  
Zurich, Switzerland  
thatt@ethz.ch

Stefan Feuerriegel  
ETH Zurich  
Zurich, Switzerland  
sfeuerriegel@ethz.ch

## ABSTRACT

Most users leave e-commerce websites with no purchase. Hence, it is important for website owners to detect users at risk of exiting and intervene early (e. g., adapting website content or offering price promotions). Prior approaches make widespread use of clickstream data; however, state-of-the-art algorithms only model the sequence of web pages visited and not the time spent on them.

In this paper, we develop a novel Markov modulated marked point process (M3PP) model for detecting users at risk of exiting with no purchase from clickstream data. It accommodates clickstream data in a holistic manner: our proposed M3PP models both the sequence of pages visited and the temporal dynamics between them, i. e., the time spent on pages. This is achieved by a continuous-time marked point process. Different from previous Markovian clickstream models, our M3PP is the first model in which the continuous nature of time is considered. The marked point process is modulated by a continuous-time Markov process in order to account for different latent shopping phases. As a secondary contribution, we suggest a risk assessment framework. Rather than predicting future page visits, we compute a user's risk of exiting with no purchase. For this purpose, we build upon sequential hypothesis testing in order to suggest a risk score for user exits.

Our computational experiments draw upon real-world clickstream data provided by a large online retailer. Based on this, we find that state-of-the-art algorithms are consistently outperformed by our M3PP model in terms of both AUROC (+6.24 percentage points) and so-called time of early warning (+12.93 %). Accordingly, our M3PP model allows for timely detections of user exits and thus provides sufficient time for e-commerce website owners to trigger dynamic online interventions.

## CCS CONCEPTS

• **Mathematics of computing** → **Markov processes**; *Kalman filters and hidden Markov models*; • **Information systems** → **Web log analysis**; • **Computing methodologies** → **Modeling methodologies**; *Continuous models*; • **Applied computing** → *Online shopping*.

## KEYWORDS

Clickstream data, Marked point process, Continuous-time Markov process, Risk scoring, Online user behavior

### ACM Reference Format:

Tobias Hatt and Stefan Feuerriegel. 2020. Early Detection of User Exits from Clickstream Data: A Markov Modulated Marked Point Process Model. In *Proceedings of The Web Conference 2020 (WWW '20)*, April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3366423.3380238>

## 1 INTRODUCTION

In 2018, over 97 % of global e-commerce website users exited the website with no purchase [38]. Because of this, there exists considerable potential for online retailers to increase revenue by improving their conversion rate. Accordingly, online retailers demand solutions that help targeting users with online interventions in order to prevent them from exiting.

Detecting exiting users from clickstream data is a challenging undertaking. First, clickstreams are highly variable and, therefore, difficult to predict [36]. Second, the objective in practice is not only to detect users just before their exits but to detect users at risk of exiting at an early stage of their session. This is important in order to provide online retailer with sufficient time for interventions. Such interventions can have various forms, for instance, presenting marketing content, dynamically adapting website design, or offering price promotions [14, 24], with the goal of steering users that otherwise would have exited with no purchase towards conversion (e. g., turning a user into a buyer).

The effectiveness of interventions depends on the accuracy with which users at risk of exiting with no purchase are identified [11]. In order to make such predictions, different modeling approaches have been developed. Markov models are widely applied to predicting future pages in clickstream data [22, 34]. Hidden Markov models have become popular as they further accommodate latent shopping phases (e. g., "browsing-oriented" phase vs. "deliberation-oriented" phase), based on which interventions are suggested [11, 28]. However, the aforementioned approaches are discrete-time models: they assume that a clickstream process evolves in unit time-steps. As such, they are unable to model the time spent on a page (TSP), and thus an important source of heterogeneity in clickstream data is missed.

As our primary contribution, we develop a novel Markov modulated marked point process (M3PP) model with the objective of detecting users that exit with no purchase from clickstream data. Our model addresses the above mentioned research gap: it models both the sequence of pages visited *and* the TSP. Thereby, our model adheres to prior literature suggesting that TSP is highly relevant for

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '20, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7023-3/20/04.

<https://doi.org/10.1145/3366423.3380238>

predicting future user actions [9, 15, 18]. Accommodating the TSP that is continuous in time requires a continuous-time process. For this reason, our model builds upon a continuous-time marked point process: here the *marks* refer to the visited pages and the *points* to the TSP. The marked point process is modulated by a Markov jump process, which is a continuous-time extension of the discrete-time Markov chain. This allows the user's clickstream to be driven by the evolution of the user's latent shopping phases, which has been theoretically and empirically verified [26, 27, 29, 31].

As our secondary contribution, we suggest a new risk assessment framework: it quantifies a user's risk of exiting with no purchase based on sequential hypothesis testing [40]. This is in contrast to previous work which focuses on predicting the next  $m$  pages [36]. However, this has limitations: there does not exist a particular exit page, user exits can occur on any page. Hence, predicting the next  $m$  pages might miss user exits. In addition to common performance metrics, we further evaluate the model using the time of early warning, which focuses on how early a user at risk of exiting with no purchase is detected. Early detection is crucial in order to provide enough time for steering users appropriately.

**Results:** We conducted computational experiments based on a real-world clickstream dataset collected during 2019. The dataset was provided by *Digitel Galaxus*, the largest online retailer in Switzerland offering more than a million products. The results show that our M3PP model consistently outperforms state-of-the-art algorithms. Our M3PP model detects exiting users more accurately and earlier, that is, it improves the AUROC (+6.24 percentage points) and the so-called time of early warning (+12.93 %). On the basis of this, e-commerce website owners can target users at the risk of exiting with no purchase at an earlier stage of their sessions in order to convert them into buyers.

**Contributions:** In summary, our main contributions are the following:

- (1) *M3PP model:* We propose a new M3PP<sup>1</sup> model that models both the sequence of individual pages visited *and* the time spent on pages. Formally, this task requires a tailored model: instead of assuming that the clickstream is a discrete-time process with unit time-steps, our setting demands a continuous-time process. Accordingly, our model captures a user's complete clickstream. Hence, when detecting users at risk of exiting with no purchase, it achieves superior performance.
- (2) *Risk assessment framework:* We suggest a new risk assessment framework for detecting users at risk of exiting with no purchase. We provide a risk score that computes a user's probability of eventually exiting with no purchase in real-time. Based on our framework, the model performance is evaluated by standard metrics (ROC) and a metric called *time of early warning*, which focuses on early detection of users at risk. Early detection is crucial in order to provide sufficient time for triggering successful interventions. Note that our framework can be built on top of any clickstream model and thus ensures widespread applicability.

## 2 RELATED WORK

Several approaches have been proposed for modeling clickstream data [cf. 25, for an overview]. Particularly popular have been approaches designed for handling sequential data: deep learning and Markov models. The former builds upon, e.g., recurrent neural networks [e.g., 21, 35, 37, 39] and serves as one of our baselines. The latter, Markov models, present the focus of this paper and are reviewed.

**Markov models:** Most of the previous work on modeling clickstream data describes the sequence of pages visited using a discrete-time Markov chain. These models are applied to capture users' clickstreams where each state of the model corresponds to a page type (e.g., home, account, checkout) [4, 13, 23, 34, 36, 42]. Based on this, a user's next  $m$  visited pages are predicted. However, the family of discrete-time Markov chains assume that the process evolves in unit time-steps. Hence, the heterogeneity of the time spent on page is ignored.

Relational Markov models are an extension of Markov models where states can be of different types [3]. Higher-order Markov chains relax the memoryless property of Markov chains and can improve predictive performance on clickstream data [5, 10, 22]. Variable-length Markov chains extend higher-order Markov chains by allowing variable-length history [6]. Although higher-order Markov chains and variable-length Markov chains are usually more accurate for predicting users' clickstreams, they come at computational costs due to their exponentially large state space. Moreover, the increase in the number of states may even result in worse predictive performance and can substantially restrict their usefulness for applications requiring fast predictions, such as inferring a user's risk of exiting with no purchase in real time [10].

**Hidden Markov models:** Hidden Markov models (HMMs) describe the user's latent shopping phases using a discrete-time Markov chain with the pages visited being the emissions depended on the current shopping phase [11, 28]. HMMs build upon research according to which users' sessions undergo different latent shopping phases. These shopping phases represent a user's current goals or state of mind, e.g., "goal-directed" and "exploratory" search [20, 26], "flow" and "non-flow" phases [17], or "browsing-oriented" and "deliberation-oriented" phases [28]. The predictive performance of HMMs suggests that the clickstream may reflect a user's goals, which could be helpful to predict future movements on a website. However, they fail to holistically model both the sequence of pages visited and the TSP.

**Marked point processes:** In addition to Markov models, we review marked point processes, which have become popular due to its ability to capture both the timing of events and the type of the event. Marked point processes were successfully used to model check-in data [12, 32] and patients' clinical states in a medical context [1, 2, 19]. However, they have not been tailored to clickstream data yet.

**Research gap:** The family of discrete-time Markov chains cannot capture both the sequence of pages visited and the times spent on page, since it assume that the process evolves by unit time-steps. Hence, this requires a family of models that can deal with processes that evolve in continuous-time. In addition, previous work has not addressed early detection of user at risk of exiting with no purchase.

<sup>1</sup>Code is available at <https://github.com/tobhatt/M3PP>.

Usually, websites do not have pages dedicated to exiting; user exits can happen on any page. Hence, predicting a user's next  $m$  pages does not allow for predicting a user exit.

### 3 CLICKSTREAM DATA

In this section, we describe the typical structure of clickstream data. A clickstream dataset comprises a set of sessions; each session is a sequence of pages that have been visited by a user together with a timestamp when it was visited. Note that sessions can have variable lengths (i. e., number of pages visited).

We denote a clickstream dataset  $\mathcal{D}$  that comprises  $D$  sessions as  $\mathcal{D} = \{S^d\}_{d=1}^D$ , where  $S^d$  is the  $d$ -th user's session. The  $d$ -th user's session is defined as  $S^d = (\{t_m^d, p_m^d\}_{m=1}^{M_d}, T^d)$ , with  $p_m^d \in \mathcal{P}$  being the  $m$ -th page visited at time  $t_m^d$ , and  $\mathcal{P}$  is the set of available pages on the website. The total number of pages observed during this session is denoted by  $M_d$  and the duration of the session is denoted by  $T^d$ . The time between two successive page visits is the TSP, i. e.,  $t_m - t_{m-1}$ . The session starts when the first page is visited and is concluded by the outcome of the session. The outcome of the session is either a purchase or an exit with no purchase, which we aim to predict.

Usually, only the sequence of pages visited is considered, i. e.,  $\{p_m^d\}_{m=1}^{M_d}$ . In Fig. 1, we illustrate an example of a user's sequences of pages visited. We can see from the figure that any temporal information of when the user visited a page and the TSP is lost since the time when a page is visited is discarded. Most existing modeling approaches have not been tailored to adequately capture such information. To do so, we included the irregularly spaced time instances at which the pages are visited, i. e.,  $\{t_m^d\}_{m=1}^{M_d}$ . However, this requires to move from the perspective that time evolves in unit time-steps as in Fig. 1 to a perspective that time evolves continuously. In Fig. 2, we depict the continuous-time perspective of a user's session. This perspective enables us to include the irregularly spaced time instances at which pages are visited. We can see from the figure that the user spends more time on some pages than on others.

However, adopting a continuous-time perspective is not possible in previous work, since this requires a continuous-time model. As a remedy, we introduce our M3PP model in the following.

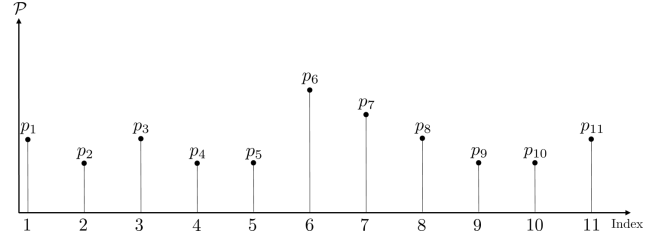
### 4 THE PROPOSED MARKOV-MODULATED MARKED POINT PROCESS MODEL

Now we present our Markov modulated marked point process, M3PP, for detecting users at risk of exiting with no purchase based on clickstream data  $\{S^d\}_{d=1}^D$ . The model has two components: (i) a Markov jump process modeling the latent shopping phases, and (ii) a marked point process modeling the clickstream, both of which are specified in the following.

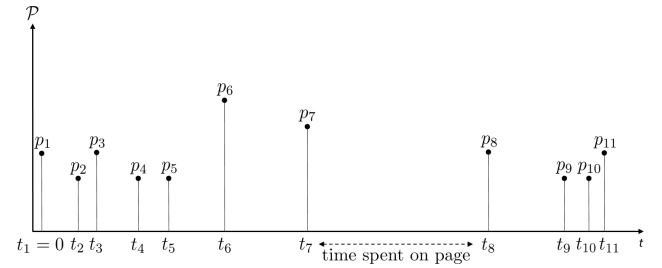
For ease of reading, we omit the superscript  $d$  in the subsequent sections.

#### 4.1 Modeling the Latent Shopping Phases

We model each user's session as being driven by an underlying latent process that represents different latent shopping phases over time. These latent shopping phases can not be observed directly;



**Figure 1: Example of a user's sequence of pages visited,  $\{p_m^d\}_{m=1}^{M_d}$ . The temporal information of how long a user spends on a page is discarded. Note that the vertical axis denotes categorical variables with no specific order.**



**Figure 2: Example of a user's sequence of pages visited including the irregularly spaced time instances at which pages are visited, i. e.,  $\{p_m^d, t_m^d\}_{m=1}^{M_d}$ .**

however, their presence has been validated empirically [26, 27, 29, 31] and, particularly, manifests in clickstream data [28].

We use a Markov jump process  $X(t)$  to model a user's latent shopping phases. A Markov jump process<sup>2</sup> is a continuous-time extensions of a discrete-time Markov chain. Its realization is a piecewise-constant function transitioning between  $N$  phases. Formally, a realization of a Markov jump process is given by

$$X(t) = \sum_{n=1}^K X_n \cdot \mathbf{1}_{\{\tau_n \leq t < \tau_{n+1}\}}, \quad (1)$$

for all  $t \in [0, T]$ , where every new phase  $X_n$  belongs to a finite space  $\mathcal{X} = \{1, 2, \dots, N\}$  and lasts from a jump time  $\tau_n$  until another jump time  $\tau_{n+1}$ . During a session, a total number of  $K$  phases are realized, which is a random variable itself. Note that  $t = 0$  corresponds to the beginning of a session, whereas  $T$  corresponds to the end of the session (i. e., the time when the user either purchases or exits with no purchase).

For the phase  $X_n$ , we define the waiting time  $\Delta\tau_n \triangleq \tau_{n+1} - \tau_n$ , with  $\tau_1 = 0$ . The waiting time  $\Delta\tau_n$  is drawn from an exponential distribution, i. e.,

$$\Delta\tau_n \mid (X_n = i) \sim \text{Exp}(\gamma_i), \quad (2)$$

whose rate parameter,  $\gamma_i$ , depends on the current phase  $X_n = i$  for  $i \in \mathcal{X}$ .

The transitions among phases are governed by a Markov transition matrix  $A = \{a_{ij}\}_{ij} \in \mathbb{R}^{N \times N}$ , i. e.,

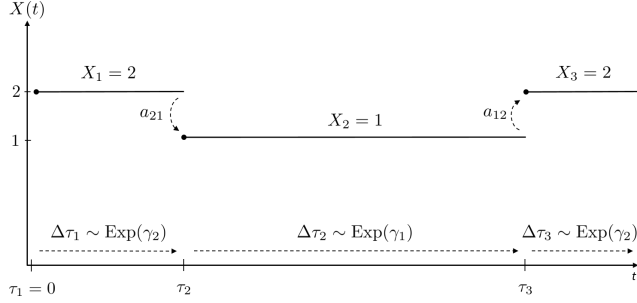
$$a_{ij} = P(X_{n+1} = j \mid X_n = i), \quad (3)$$

<sup>2</sup>For an in-depth discussion on continuous-time Markov processes, we refer to [30].

where self-transitions are eliminated, i. e.,  $a_{ii} = 0$ , for all  $i \in \mathcal{X}$ , since these are captured by the waiting times.

The initial phase distribution is given by  $\pi = \{\pi_i\}_{i=1}^N$ , where  $\pi_i = P(X(0) = i)$ , and describes the probability of starting a session in phase  $i \in \mathcal{X}$ .

Fig. 3 depicts an exemplary realization of a Markov jump process.



**Figure 3: Illustration of a Markov jump process  $X(t)$ . The process jumps at the jump times  $\tau_n$ . The probability of the next phase is determined by a Markov transition matrix  $A$ .**

A user's latent shopping phase  $X(t)$  manifests in a user's clickstream in two ways: (i) It modulates the distribution of the TSP, for instance, users in a "browsing-oriented" phase might spend more time on pages than in another phase. (ii) It modulates the distribution of the sequence of pages visited, for instance, users in a "deliberation-oriented" phase might be more likely to visit product pages than in another phase. We capture these two effects via a marked point process in the next section.

## 4.2 Modeling the Clickstream

We model a user's clickstream using a marked point process and allow the parameters of the marked point process to depend on the latent shopping phase  $X(t)$ .

A marked point process is a continuous-time process where the points are the time instances when pages are visited and the marks are the pages themselves, i. e.,  $\{p_m, t_m\}_{m=1}^M$ . For a clickstream  $\{p_m, t_m\}_{m=1}^M$ , the density  $f$  of a marked point process is given by

$$f(\{p_m, t_m\}_{m=1}^M) = \prod_m f(p_m, t_m | \mathcal{H}_{t_{m-1}}), \quad (4)$$

where the history  $\mathcal{H}_{t_{m-1}}$  denotes the clickstream up to time  $t_{m-1}$ . Depending on the application, one can design a variety of forms for  $f(p_m, t_m | \mathcal{H}_{t_{m-1}})$ . We choose the factorization  $f(p_m, t_m | \mathcal{H}_{t_{m-1}}) = f(t_m | \mathcal{H}_{t_{m-1}}) f(p_m | \mathcal{H}_{t_{m-1}})$  due to its simplicity; joint models can be adapted in a straightforward manner.

In the following sections, we derive parameterizations for (i)  $f(t_m | \mathcal{H}_{t_{m-1}})$ , i. e., the conditional density of time instances when pages are visited, and (ii)  $f(p_m | \mathcal{H}_{t_{m-1}})$ , i. e., the conditional density of the pages visited.

**4.2.1 Point Process for the Time Spent on Page.** We describe the conditional density for the time instances when pages are visited by a point process. The conditional density  $f(t_m | \mathcal{H}_{t_{m-1}})$  of a point

process is given by

$$f(t_m | \mathcal{H}_{t_{m-1}}) = \lambda^*(t_m) \exp\left(-\int_{t_{m-1}}^{t_m} \lambda^*(s) ds\right), \quad (5)$$

where  $\lambda^*(t) = \lambda(t | \mathcal{H}_{t-})$  is the intensity function conditioned on the history up to, but not including time  $t$ . It describes the probability of the occurrence of a new page within a small window  $[t, t + dt]$  [33]. Note that in Eq. 5, modeling the time when the next page is visited,  $t_m$ , also determines the time between two successive page visits,  $t_m - t_{m-1}$ , which is precisely the TSP. Formally, a point process is specified by the conditional intensity function  $\lambda^*(t)$ , which is given by

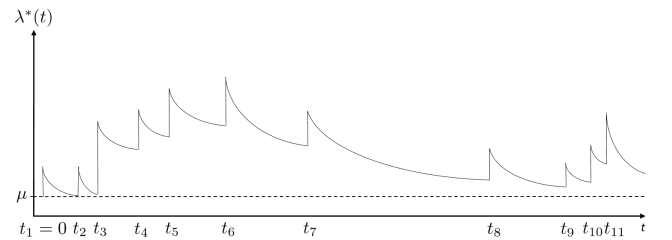
$$\lambda^*(t) dt = P(\text{page in } [t, t + dt] | \mathcal{H}_t). \quad (6)$$

The functional forms of the conditional intensity function  $\lambda^*(t)$  are often designed to describe the phenomena of interest [8]. In particular, we model  $\{t_m\}_{m=1}^M$  as a one-dimensional Hawkes process [16] whose conditional intensity function is modulated by the latent shopping phase  $X(t)$  and defined as

$$\lambda_i^*(t) \triangleq \lambda^*(t | X(t) = i) = \mu_i + \alpha_i \sum_{\tau \leq t_m < t} \phi_i(t, t_m), \quad (7)$$

for all  $i \in \mathcal{X}$ , where  $\mu_i \geq 0$  is the phase-dependent baseline intensity,  $\tau < t$  is the time of the most recent jump in the latent shopping phase process  $X(t)$ , and  $\phi_i(t, t_m) \geq 0$  is the kernel capturing temporal dynamics between pages. The sum over the kernel terms makes the conditional intensity dependent on the history from the most recent jump in the latent shopping phase process. We choose an exponential kernel  $\phi_i(t, t_m) = e^{-\beta_i(t-t_m)}$  such that the temporal dependence between pages decays over time. For  $\beta_i = \infty$  or  $\alpha_i = 0$ , we recover a modulated Poisson process as a special case [32].

A distinctive feature of the Hawkes process is that the occurrence of each page visit increases the conditional intensity by a certain amount and, hence, the probability of another page visit. In Fig. 4, we illustrate the conditional intensity function of a Hawkes process. We can see that the close succession of page visits increases the conditional intensity function and, therefore, the probability of another page visit.



**Figure 4: The conditional intensity function  $\lambda^*(t)$  of a Hawkes process. The dashed line represents the baseline intensity  $\mu$  while the solid line corresponds to the total intensity.**

We choose a Hawkes process due to this "exciting effect", since the display of information (e. g., advertisement) is likely to stimulate subsequent page visits [41]. In other words, prior exposure to a page can positively affect the probability of subsequent page visits. In the

light of this argument, the modulated Hawkes process appears to be a suitable model for  $\{t_m\}_{m=1}^M$  as it captures both the time-varying intensity and the "exciting effects" among page visits.

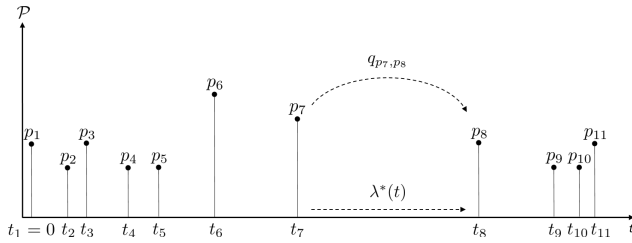
**4.2.2 Mark Process for the Sequence of Pages.** We capture the conditional density  $f(p_m | \mathcal{H}_{t_{m-1}})$  in the marked point process by a mark process, which describes the sequence of pages visited by a user. We build upon previous work on clickstream modeling that extensively studied suitable models for the sequence of pages [e.g., 36]. We describe the probability of choosing the next page depending on the current page, i.e.,  $f(p_m | \mathcal{H}_{t_{m-1}}) = f(p_m | p_{m-1})$ . In particular, the sequence of pages visited,  $\{p_m\}_{m=1}^M$ , is modeled using a discrete-time Markov chain given by

$$f(p_m | p_{m-1}, X(t) = i) = q_{p_{m-1}, p_m}^i, \quad (8)$$

where  $Q^i = \{q_{jk}^i\}_{jk} \in \mathbb{R}^{|\mathcal{P}| \times |\mathcal{P}|}$  describes the probability of visiting the  $k$ -th page conditioned on being on the  $j$ -th page and the latent shopping phase  $X(t) = i$ . In addition, the probability of the first page is given by the initial probabilities  $o^i = \{o_j^i\}_{j=1}^{|\mathcal{P}|}$ .

We call the above discrete-time Markov chain a *switching Markov chain* due to its dependence on the current latent shopping phase  $X(t) = i$  and in order to clearly distinguish it from the Markov jump process that describes the latent shopping phases.

Fig. 5 illustrates the marked point process consisting of the Hawkes process from Section 4.2.1 describing the TSP and the switching Markov chain describing the sequence of pages visited for a user's clickstream.

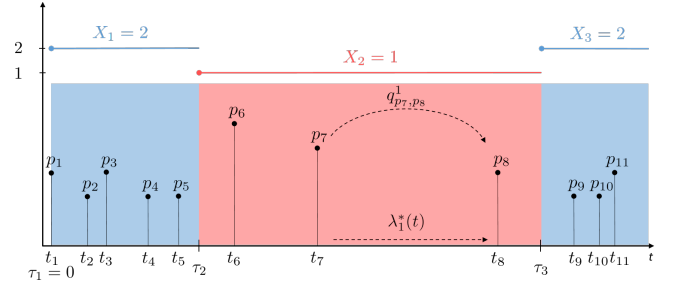


**Figure 5: The marked point process for a user's sequence of pages visited and the time spent on pages. The former is described by a switching Markov chain with conditional probabilities  $q_{p_{m-1}, p_m}$ , and the later is described by a Hawkes process with conditional intensity function  $\lambda^*(t)$ .**

Our M3PP model is then given by the combination of the point process and the mark process, both modulated by the latent shopping phase process  $X(t)$ . It describes a user's clickstream in a holistic manner. Fig. 6 presents the entire model for a user's session, where both the conditional intensity and the distribution of the sequence of pages visited depend on the latent shopping phase.

### 4.3 Estimating the Model Parameters

In this section, we describe the estimation procedure for the M3PP model parameters using the clickstream data  $\{\mathcal{S}^d\}_{d=1}^D$ . In particular, we state the likelihood of the complete M3PP model which can subsequently be used for any likelihood-based estimation procedure. We provide the code of our model on <https://github.com/tobhatt/M3PP>.



**Figure 6: A depiction of our proposed M3PP model that completely describes a user's session. The sequence of pages and the time spent on pages are drawn from a marked Hawkes process, whose parameters are modulated by a (latent) Markov jump process.**

For every latent phase  $i \in \mathcal{X}$ , we denote the Hawkes process parameters as  $\Lambda_i = (\mu_i, \alpha_i, \beta_i)$  and the switching Markov chain parameters as  $\mathcal{M}_i = (o^i, Q^i)$ . The set of model parameters is given by

$$\Omega = (A, (\pi_i, \gamma_i, \Lambda_i, \mathcal{M}_i)_{i \in \mathcal{X}}), \quad (9)$$

where  $A$  is the Markov transition matrix of the Markov jump process  $X(t)$  and  $\{\pi_i\}_{i=1}^N$  its initial phase distribution.

Estimating the model parameters is a daunting task due to the user's shopping phase  $X(t)$  being latent. Usually, the jump times among the latent phases,  $\{\tau_n^d\}_{n=1}^{K+1}$ , cannot be observed, since transitions can take place at any point in time.

However, in our specific setting, the user itself makes the decision when to visit another page. Hence, we assume that changes in the latent shopping phases are associated with the user taking direct action and visiting another page. Therefore, the times when the user visits another page, i.e.,  $\{t_m^d\}_{m=1}^{M_d}$ , is informative about changes in the latent shopping phase. Mathematically, this means that the set of jump times  $\{\tau_n^d\}_{n=1}^{K+1}$  is a subset of  $\{t_m^d\}_{m=1}^{M_d}$ . This simplifies the likelihood derivation, since it restricts the set of possible jump times to the finite set, i.e.,  $\{t_m^d\}_{m=1}^{M_d}$ , rather than the uncountable set,  $[0, T^d]$ .

Using the above, the likelihood of our model is given by

$$\mathcal{L}(\{p_m^1, t_m^1\}_{m=1}^{M_1}, \dots, \{p_m^d, t_m^d\}_{m=1}^{M_d} | \Omega) = \prod_{d=1}^D \sum_{i \in \mathcal{X}} \alpha_{t_m^d}^i(i), \quad (10)$$

where  $\alpha_t(i) \triangleq P(\{p_k, t_k \leq t\}, X(t) = i | \Omega)$  are the forward probabilities. These are recursively given for our M3PP model by

$$\begin{aligned} \alpha_{t_m^d}^i(i) &= \sum_{s=0}^m \sum_{j \in \mathcal{X} \setminus \{i\}} \alpha_{t_{m-s-1}^d}^j(j) A_{ji} \gamma_i e^{-\gamma_i(t_m^d - t_{m-s}^d)} \\ &\times \exp \left( - \int_{t_{m-s+1}^d}^{t_m^d} \lambda_i^*(s) ds \right) \prod_{k=m-s}^m \lambda_i^*(t_k^d) q_{p_{k-1}, p_k}^i, \end{aligned} \quad (11)$$

where  $\lambda_i^*(s)$  is the conditional intensity function of the Hawkes process in Eq. 7 depending on the latent shopping phase  $i \in \mathcal{X}$ . The derivation of the likelihood can be found in Section A.

## 5 RISK ASSESSMENT FRAMEWORK

We now describe our risk assessment framework that computes a user's risk of exiting with no purchase. For this, we introduce a real-time risk score based on sequential hypothesis testing [40]. This risk score can be used to detect users at risk whenever it exceeds a predefined threshold and, thereby, determines the so-called time of early warning.

### 5.1 Assessing a User's Exiting Risk

We denote the risk score at time  $t$  as  $R(t)$ , which is the probability that a user exits with no purchase given the observed clickstream up to time  $t$ .

More formally, the risk score is confronted with two hypotheses: (i) the null hypothesis  $\mathbf{H}_0$  which corresponds to the hypothesis that the user will ultimately purchase something, and (ii) the alternative hypothesis  $\mathbf{H}_1$  which corresponds to the hypothesis that the user will exit with no purchase. The risk score's aim to test the null hypothesis as more clickstream data from a user becomes available. Hence, the risk score is a sequential hypothesis test [40], where the true hypothesis is observed at the end of the sessions by the outcome of the session (i. e., purchase or exit). We denote the session outcome by the variable  $l^d \in \{0, 1\}$ , where  $l^d = 0$  if the null hypothesis holds true, and  $l^d = 1$  if the alternative hypothesis holds true, i. e.,

$$l^d = \begin{cases} 0, & \text{if } \mathbf{H}_0 \text{ (i. e., user purchases),} \\ 1, & \text{if } \mathbf{H}_1 \text{ (i. e., user exits with no purchase).} \end{cases} \quad (12)$$

Hence, following [40], we view the user's risk score as the test statistic of a sequential hypothesis test. That is, the user's risk score  $R(t)$  at time  $t$  is the posterior probability of hypothesis  $\mathbf{H}_1$  given the observations up to time  $t$ , i. e.,  $R(t) \triangleq P(\mathbf{H}_1 \mid \{p_m^d, t_m^d \leq t\})$ . Using Bayes' rule, the risk score can be rewritten as

$$R(t) = \frac{P(\{p_m^d, t_m^d \leq t\} \mid \mathbf{H}_1) P(\mathbf{H}_1)}{\sum_{l^d \in \{0,1\}} P(\{p_m^d, t_m^d \leq t\} \mid \mathbf{H}_{l^d}) P(\mathbf{H}_{l^d})}, \quad (13)$$

where  $P(\mathbf{H}_0)$  and  $P(\mathbf{H}_1)$  are the prior probabilities of a user purchasing and exiting with no purchase, respectively (i. e., the rate of purchasing and exiting sessions). This means that the risk score  $R(t)$  determines under which hypothesis the observed clickstream is more likely to occur.

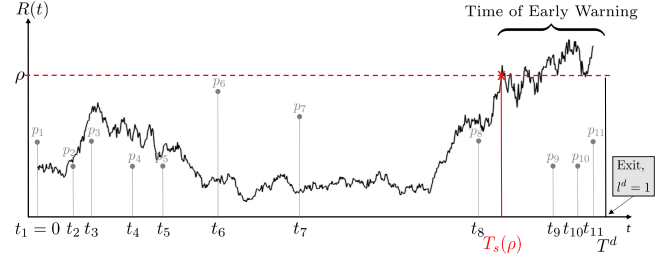
In practice, a threshold  $\rho$  must be set on the computed risk score  $R(t)$  in order to detect users at risk of exiting with no purchase. Whenever the risk score exceeds that threshold, a user is classified as at risk of exiting with no purchase (i. e., the hypothesis  $\mathbf{H}_1$  is declared). This is similar to the sequential hypothesis test, where the null hypothesis is rejected whenever the test statistic crosses a predefined threshold [40].

### 5.2 Time of Early Warning

We define the time at which the risk score exceeds the threshold  $\rho$  as the stopping time  $T_s(\rho)$ , i. e.,

$$T_s(\rho) \triangleq \inf\{t \in \mathbb{R}_+ : R(t) \geq \rho\}, \quad (14)$$

where we set  $\inf\{\} = \infty$ . Then the user's *time of early warning (TEW)* is the time between a user's stopping time,  $T_s(\rho)$ , and the end of the session,  $T^d$ . Fig. 7 depicts an example of the evolution of a computed risk score during a user's session.



**Figure 7: Example of a user's session with corresponding risk score  $R(t)$ . When the risk score exceeds the threshold  $\rho$  at stopping time  $T_s(\rho)$ , the user is classified as at risk of exiting with no purchase. The time between the stopping time,  $T_s(\rho)$ , and the end of the session,  $T^d$ , is the time of early warning.**

The threshold  $\rho$  can be varied in order to manipulate the time of early warning and control alarm fatigue. The smaller  $\rho$ , the earlier it is exceeded by the risk score. This yields an early stopping time,  $T_s(\rho)$ , and, therefore, a longer time period until the end of sessions,  $T^d$ . However, this also results in a sensitive risk assessment with many false alarms. On the contrary, the larger  $\rho$ , the later it is exceeded by the risk score, which yields a late stopping time,  $T_s(\rho)$ , and, therefore, a shorter time period until the end of sessions,  $T^d$ . This yields an insensitive risk assessment which misses many users that exit with no purchase. Note that users are classified as not at risk of exiting with no purchase (i. e., hypothesis  $\mathbf{H}_1$  is rejected), if  $\rho < R(t)$ , for all  $t \in [0, T^d]$ .

This risk assessment framework can be used for an arbitrary model for a given clickstream. In particular, we are using it for both our M3PP model and all baselines.

## 6 EXPERIMENTAL SETTING

In this section, we describe our clickstream dataset and the experimental settings. In particular, we introduce a variety of baselines and performance metrics, which we are going to use later for model comparison.

### 6.1 Data Description

We evaluate our M3PP on a real-world clickstream dataset provided by *Digitex Galaxus*, the largest online retailer in Switzerland offering more than a million products. Our partner company sells a large variety of products, ranging from electronic devices to fashion. Its website also contains content that provides information to prospective customers, for instance, product reviews.

Our clickstream dataset  $\mathcal{D} = \{\mathcal{S}^d\}_{d=1}^D$  displays the typical structure of clickstream data as described in Section 3. As a reminder, the  $d$ -th user's session is defined as  $\mathcal{S}^d = (\{p_m^d, t_m^d\}_{m=1}^{M_d}, T^d)$  with  $p_m^d$  being the  $m$ -th page visited at time  $t_m^d$ ,  $M_d$  is the total number of pages observed during this session and  $T^d$  is the duration of the session. Our clickstream dataset  $\mathcal{D}$  comprises  $D = 2'223$  user sessions, with a maximum number of pages in a session,  $M_d$ , of 50, and a maximum TSP of 20 minutes. The TSP is heterogeneous across all page visits in the dataset (mean: 48.70 sec., SD: 113.01). This supports our



premise that user sessions evolves in continuous-time rather than in unit time-steps. Each session is further associated with a variable  $l^d \in \{0, 1\}$  describing the session outcome (i. e., purchase, or exit). This variable is used in our risk assessment framework where it decodes whether the null hypothesis or the alternative hypothesis held true in Eq. 12 and is required to compute a user's risk score  $R(t)$  from Eq. 13 and to evaluate the predictive performance.

Our real-world clickstream dataset has the following two benefits. (i) Our dataset comprises user sessions collected during 2019 and is thus recent (different from many public datasets). (ii) We obtained the actual session outcomes,  $l^d$ . The session outcomes can not be extracted from the clickstreams themselves, as there is no page dedicated to a purchase or exit. Using a proxy by assuming that CHECKOUT or SHOPPING CART pages are associated with purchases yields noisy variables  $l^d$ .

**6.1.1 Data Preprocessing.** The following preprocessing was applied analogously to previous research on clickstream modeling [13, 22, 28]. (i) We extracted the pages that were actually visited (i. e., rendering of a page request in the user's browser window). (ii) We encoded page according to seven distinct page types: HOME, ACCOUNT, OVERVIEW, PRODUCT, MARKETING CONTENT, COMMUNITY, and CHECKOUT (incl. Shopping Cart). (iii) We excluded sessions from the dataset that originated from web crawlers as these were not conducted by humans and thus had different objectives. For this purpose, we made use of a built-in tool provided by our partner company. (iv) In addition, many sessions had only a few page visits. We excluded sessions with fewer than three pages visited. (v) We excluded sessions that contain more than 50 page visits. (vi) We considered a session has ended if the TSP was more than 20 minutes, since some users kept the website open for a long period of time and continued with the sessions later or another day. This results in the above dataset containing 2,223 sessions.

## 6.2 Baselines

We compare our M3PP model with the following state-of-the-art baselines analogous to the models in Section 2:

- (1) **Markov chain (MC):** We compare to Markov chains with varying higher orders from 1 to 3 [12, 22], denoted as **MC-1**, **MC-2**, and **MC-3**.
- (2) **Variable-length Markov chain (VLMC):** We describe the sequence of pages visited using a variable-length Markov chain (sometimes also called variable-order Markov chain), where the maximum context length is set to 5 [6].
- (3) **Hidden Markov model (HMM):** We fit a hidden Markov model with multinomial emission variables. Analogous to previous research, we use  $n = 2$  hidden states for describing different latent shopping phases [11, 28].
- (4) **Logistic regression (LR):** We fit a logistic regression using three features extracted from a given clickstream: the number of pages visited during the session, the average TSP, and total time spent on the website. These features are used to train the model together with the session outcome  $l^d$  as target variable.

- (5) **Long short-term memory network (LSTM):** Long short-term memory networks have been recently applied to clickstream data [e. g., 21, 37]. We train a LSTM, which is state-of-the-art in clickstream modeling, following the architecture of [37]. The LSTM has 3 layers and 60 neurons in each layer. The network is trained in 1,000 epochs using the *Adam* optimizer and the cross-entropy loss. The input consists of a user's clickstream; the label  $l^d \in \{0, 1\}$  represents the target variable.

We also compare to a marked temporal point process (MTPP) without latent shopping phases. This resembles the model from [12] but is based on a classic Hawkes process. The model form [32] is not applicable to clickstream data since it assumes continuous marks. For our M3PP model and the baselines (1)–(3), we use the risk score in Eq. 13. For the baselines (4)–(5), the expression  $P(\{p_m^d, t_m^d \leq t\} \mid \mathbf{H}_{1d})$  from Eq. 13 can not be computed. However, the way were trained these two models, we can directly compute the posterior probability of hypothesis  $\mathbf{H}_1$ , i. e.,  $P(\mathbf{H}_1 \mid \{p_m^d, t_m^d \leq t\})$ . This yields the same risk score, since  $l^d = 1$  for  $\mathbf{H}_1$ .

## 6.3 Performance Metrics

We compare our M3PP model to the baselines in Section 6.2 in two ways: (i) the accuracy of whether a user is detected correctly; and, if detected correctly, (ii) the time of early warning.

For (i), we use the receiver operating characteristics (ROC), which is commonly used in machine learning in order to evaluate the performance of classification problems. The ROC is the trade-off between the false positive rate (FPR) and the true positive rate (TPR). The FPR is the rate with which users are classified as at risk of exiting with no purchase although they are going to purchase, i. e.,

$$\text{FPR} = \frac{\# \text{ users with } l^d = 0 \text{ and } R(t) \geq \rho \text{ for some } t < T^d}{\# \text{ users with } l^d = 0}. \quad (15)$$

The TPR is the rate with which users are classified as at risk of exiting with no purchase and they are going to exit with no purchase, i. e.,

$$\text{TPR} = \frac{\# \text{ users with } l^d = 1 \text{ and } R(t) \geq \rho \text{ for some } t < T^d}{\# \text{ users with } l^d = 1}. \quad (16)$$

For (ii), we use the time of early warning as introduced in Section 5.2. In particular, we use the trade-off between the FPR and the TEW, which is tailored to early warnings but closely resembles the familiar ROC curves [7]. In particular, we use the expected TEW in order to account for all users in the dataset, i. e.,

$$\overline{\text{TEW}} = \mathbb{E}[T - T_s(\rho) \mid R(t) \text{ exceeds threshold, } l = 1]. \quad (17)$$

The threshold  $\rho$  can be varied in order to trade off between the FPR and the TPR, as well as to trade off between the FPR and the  $\overline{\text{TEW}}$ . The performance metrics above evaluate the model performance in terms of its accuracy, false alarm rate, and time of early warning in detecting users at risk of exiting with no purchase.

## 6.4 Estimation Details

In this section, we provide the detailed steps of how the model was estimated and, based on this, how the risk score was computed.

- (1) We split our clickstream dataset into a training set (75 %) and a test set (25 %).



Model	Performance metrics	
	AUROC	AUEW
LR	0.585	4.36
LSTM	0.597	4.36
MC-2	0.618	4.21
VLMC	0.631	4.23
M3PP	0.641	4.68
MC-3	0.642	4.41
HMM	0.651	4.51
MC-1	0.666	4.53
<b>Proposed M3PP Model</b>	<b>0.728</b>	<b>5.12</b>

**Table 1: Test set evaluation for various models. The proposed M3PP model outperforms all baselines along receiver operating characteristics and time of early warning.**

- (2) We further divide the training set into two set: one set on which  $\mathbf{H}_0$  holds true for all sessions and one set on which  $\mathbf{H}_1$  hold true for all sessions. For this, the session outcomes  $l^d$  are required.
- (3) Then, we estimate the model parameters  $\Omega$  on both sets from (2) using the likelihood from Section 4.3. We take a Bayesian approach to estimating the parameters via a maximum a posteriori estimation.<sup>3</sup> We place weakly informative priors on the unknowns  $A$ ,  $\pi$ ,  $\gamma$ ,  $\Lambda$ , and  $\mathcal{M}$ . In particular, we place normal priors on the rows of  $A$ , a Dirichlet prior on  $\pi$ , and gamma priors on  $\gamma$ . We further placed gamma priors on the Hawkes process parameters  $\Lambda$  and Dirichlet priors on the switching Markov chain parameters  $(\mathcal{M}_i)_{i \in \mathcal{X}}$ . This yields one model estimate for  $P(\{p_m^d, t_m^d \leq t\} | \mathbf{H}_0)$  and another model estimate for  $P(\{p_m^d, t_m^d \leq t\} | \mathbf{H}_1)$ .
- (4) Afterwards, we determine hyperparameters. Note that the number of latent shopping phases  $N$  is the only hyperparameter that has to be selected. Using the Bayesian information criterion, we selected  $N = 2$  shopping phases for each of the two model estimates.
- (5) Finally, we compute the risk score  $R(t)$  given by Eq. 13 on the test set using the two model estimates. The risk score is dynamically updated as more clickstream data from a user becomes available.

## 7 RESULTS

In this section, we present the predictive performance on the test sessions and report the parameter estimates of the M3PP model.

### 7.1 Predictive Performance

We compare our M3PP model with our baselines across different performance metrics. In practice, the threshold  $\rho$  has to be set in order to classify a user at risk of exiting with no purchase. We vary the threshold  $\rho$  and report the area under the curve (AUC) of the performance metrics. The AUC of the ROC curve is denoted by AUROC, and the AUC of the FPR vs.  $\overline{\text{TEW}}$  curve is denoted by AUEW.

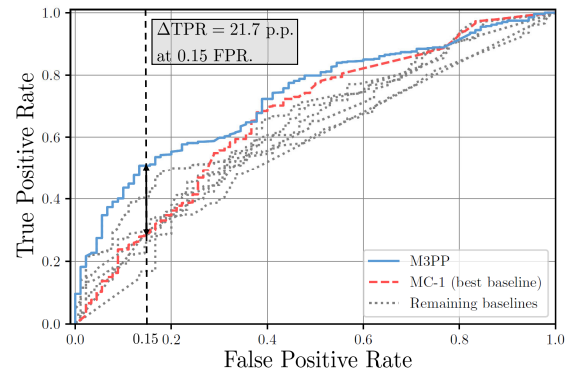
<sup>3</sup>We provide the implementation of the likelihood and the choice of priors on <https://github.com/tobhatt/M3PP>

We report the AUC values for all models under consideration in Table 1. We can see that our M3PP model outperforms the state-of-the-art baselines: it achieves an AUROC of 0.728 and an AUEW of 5.12. The best performing baseline is the first-order Markov chain (i. e., MC-1) with an AUROC of 0.666 and an AUEW of 4.53. Hence, our proposed model offers an improvement over the MC-1 of 6.2 percentage points (p.p.) in AUROC and 12.9 % in AUEW.

As mentioned in Section 2, the increase in the number of states results in inferior predictive performance for MC-2 and MC-3. The LSTM achieves an AUROC of 0.597 and an AUEW of 4.36. It is outperformed by our proposed M3PP model by 13.2 p.p. in AUROC and 17.4 % in AUEW.

For better illustration, we now make an example with a fixed threshold  $\rho$ . This differs from the above comparison (where we varied the value for  $\rho$ ). In practice, the threshold  $\rho$  should be set such that the resulting FPR is relatively low in order to avoid many false positive alarms and, therefore, falsely spent online interventions (e. g., price promotions). We set a different threshold  $\rho$  for each model that was chosen in a way such that the resulting FPR is 0.15 for each model. Note that different thresholds for each model are needed in order to achieve the same FPR.

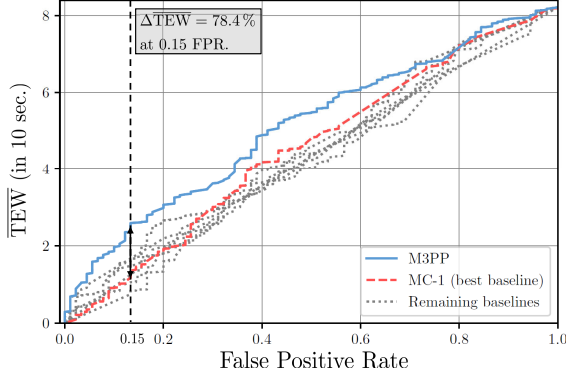
In Fig. 8, we depict the ROC curve, which shows the trade-off between the FPR and the TPR as a function of  $\rho$ . The dashed line in Fig. 8 denotes a FPR of 0.15. Our M3PP model detects users at risk of exiting with no purchase at a TPR of 0.51. The MC-1 baseline achieves at the same FPR only a TPR of 0.29. Hence, detections from the M3PP model are better by 21.7 p.p. better. Out of 100 users that exit with no purchase, the M3PP model detects 50 users compared to 30 users detected by the MC-1.



**Figure 8: ROC curve for our proposed M3PP model (blue), best baseline (red), and the remaining baselines (gray). Our M3PP model increases the true positive rate for any setting of the false positive rate below 0.75. For instance, at a false positive rate of 0.15, the M3PP model improves the true positive rate by  $\Delta\text{TPR} = 21.7$  percentage points compared to the best baseline.**

Fig. 9 depicts the FPR vs.  $\overline{\text{TEW}}$  curve, which shows the trade-off between the FPR and the  $\overline{\text{TEW}}$  as a function of  $\rho$ . Again, the dashed line in Fig. 9 denotes a FPR of 0.15. On average, our M3PP model triggers alarms 25.99 seconds before the end of the session. Compared to this, the MC-1 baseline triggers alarms 14.57 seconds

before the session ends. Hence, our proposed model detects on average user that will exit with no purchase 78.4 % earlier. If we measure the TEW in number of pages instead of seconds, then the M3PP detects a user at risk of exiting with no purchase 3 pages before the exit. Compared to this, the MC-1 detects a user at risk of exiting with no purchase 2 pages before the exit. This is considerable when comparing to the average session, which contains 10 pages.



**Figure 9: False positive rate vs. expected time to early warning curve for our M3PP model (blue), best baseline (red), and the remaining baselines (gray). Our M3PP model improves the expected time to early warning for any setting of the false positive rate below 0.7. For instance, at a false positive rate of 0.15, the M3PP model improves the expected time to early warning by  $\Delta\text{TEW} = 78.4\%$  compared to the best baseline.**

Note that for lower FPR, the performance difference tends to increase in favor of our M3PP model. In summary, our M3PP does not only detect more user at risk of exiting with no purchase correctly, but also substantially earlier than the best baseline.

## 7.2 Interpretation of Model Parameters

Now we present the estimated parameters for our M3PP model. In particular, we highlight that the marked point process parameters  $(\Lambda_i, M_i)$  for each latent shopping phase  $i \in \mathcal{X}$  differ substantially. In the following, we use these differences to attach an interpretation to the latent shopping phases using the terminology in [28].

As a result of the estimation in Section 6.4, we obtained two sets of estimated parameters: one set for the distribution conditioned on the null hypothesis,  $\mathbf{H}_0$ , denoted by  $\hat{\Omega}_0$ , and one set for the distribution conditioned on the alternative hypothesis,  $\mathbf{H}_1$ , denoted by  $\hat{\Omega}_1$ . We only report the estimated parameters  $\hat{\Omega}_1$  due to space constraints.<sup>4</sup> We remind of our model selection: using the Bayesian information criterion, we selected 2 latent shopping phases for the Markov jump process  $X(t)$ , i. e.,  $\mathcal{X} = \{1, 2\}$ .

**7.2.1 Estimated Point Process Parameters.** We modeled  $\{t_m^d\}_{m=1}^{M_d}$  using a Hawkes process in Eq. 7. The Hawkes process' conditional

intensity functions for  $\hat{\Omega}_1$  are given by the parameter estimates

$$\lambda_1^*(t) = \underbrace{0.68}_{\hat{\mu}_1} + \underbrace{0.01}_{\hat{\alpha}_1} \sum_{\tau \leq t_m < t} e^{-\underbrace{8.40}_{\hat{\beta}_1} (t-t_m)}, \quad (18)$$

$$\lambda_2^*(t) = \underbrace{12.04}_{\hat{\mu}_2} + \underbrace{0.02}_{\hat{\alpha}_2} \sum_{\tau \leq t_m < t} e^{-\underbrace{7.74}_{\hat{\beta}_2} (t-t_m)}, \quad (19)$$

for each of the two latent shopping phases, where  $\lambda_i^*(t)$  is measured in units of 10 seconds.

We interpret phase 1 as a "browsing-oriented" phase in which the user does not have a specific goal in mind. During this phase, the user tends to spend more time on an individual page, since the estimated baseline intensity,  $\hat{\mu}_1$ , is much lower than in phase 2 (i. e., 0.68 vs. 12.04). Phase 2 is interpreted as a "deliberation-oriented" phase in which the user has a specific goal in mind. Because of this, the user switches between pages much quicker and spends less time on an individual page, since the estimated baseline intensity,  $\hat{\mu}_2$ , is much higher than in phase 1. The temporal dynamics between pages visited differ from one phase to another: during a "deliberation-oriented" phase, past page visits influence the conditional intensity function more strongly than during a "browsing-oriented" phase (i. e., larger  $\alpha_2$  and smaller  $\beta_2$ ).

**7.2.2 Estimated Mark Process Parameters.** We modeled the sequence of pages visited using a switching Markov chain in Eq. 8. As above, the estimated parameters of the switching Markov chain substantially differ from one latent shopping phase to another. For each of the two latent shopping phases, Table 2 depicts the estimated parameters of the switching Markov chain, i. e.,  $Q^1$  and  $Q^2$ . We find that users in a "browsing-oriented" phase tend to visit OVERVIEW and PRODUCT pages more often, whereas users in a "deliberation-oriented" phase tend to visit ACCOUNT pages and MARKING CONTENT more often. This can be observed by summing over the columns in Table 2 for each phase. These findings coincide with previous findings on latent shopping phases [11, 28].

In summary, the above latent shopping phase-specific differences underline the importance of modulating the marked point processes: it allows us to integrate the user's clickstream together with her latent shopping phase while learning the user's clickstream behavior.

## 8 CONCLUSION

This paper proposes a novel Markov modulated marked point process model for detecting e-commerce website users at risk of exiting with no purchase using clickstream data. Our proposed model captures not only the sequence of individual page visits by a user but also the temporal dynamics between them, i. e., the time spent on pages. This requires extending previous discrete-time models to continuous-time.

As a secondary contribution, we suggested a risk assessment framework that assesses a user's risk of eventually exiting with no purchase using a real-time risk score. It computes a risk score based

<sup>4</sup>The latent shopping phases for the estimated parameters  $\hat{\Omega}_0$  have the same interpretation.

		Next page						
Page type		Home	Acct.	Overv.	Prod.	Mktg.	Comm.	Check.
Current page	Phase							
Home	1	<b>0.32</b>	0.15	0.14	0.27	0.07	0.01	0.05
	2	0.17	0.21	0.26	0.02	<b>0.34</b>	0.00	0.00
Account	1	0.05	<b>0.77</b>	0.02	0.11	0.00	0.00	0.04
	2	0.00	<b>0.95</b>	0.00	0.04	0.00	0.00	0.00
Overview	1	0.01	0.01	0.32	<b>0.64</b>	0.00	0.01	0.01
	2	0.00	0.00	<b>0.94</b>	0.00	0.06	0.00	0.00
Product	1	0.05	0.01	0.25	<b>0.60</b>	0.01	0.00	0.08
	2	0.00	0.03	0.01	<b>0.96</b>	0.00	0.00	0.01
Marketing	1	0.27	0.03	0.11	<b>0.45</b>	0.14	0.00	0.00
	2	0.03	0.00	0.00	0.00	<b>0.96</b>	0.00	0.00
Community	1	0.29	0.01	<b>0.67</b>	0.01	0.00	0.01	0.01
	2	0.03	0.05	0.06	0.07	0.04	<b>0.60</b>	0.14
Checkout	1	0.05	0.09	0.02	0.12	0.00	0.00	<b>0.72</b>
	2	0.09	0.20	0.01	<b>0.33</b>	<b>0.33</b>	0.03	0.01

**Table 2: Estimated switching Markov chain parameters,  $Q^1$  and  $Q^2$ , for  $\hat{\Omega}_1$ . For each page type, the first row shows the parameters for phase 1 and the second row for phase 2. For instance, being at a HOME page, the probability of the next page being MARKETING CONTENT is 0.07 in phase 1 and 0.34 in phase 2.**

on sequential hypothesis testing in order to detect users at risk of exiting with no purchase early during their sessions.

Our computational experiments use a real-world clickstream dataset provided by our partner company, the largest online retailer in Switzerland. The results demonstrate that our M3PP model consistently outperforms state-of-the-art models in terms of both AUROC (+6.24 percentage points) and AUTEW (+12.93 %). At the same level of the FPR, our model detects user exiting with no purchase not only more accurately, but also 78.4 % earlier, which provides more time for interventions.

For online retailers, the substantial predictive performance offered by our M3PP model promises a great improvement in the quality and timing of online interventions. By utilizing the proposed model combined with the risk assessment framework, e-commerce website owners can better focus their attention on users at risk of exiting with no purchase, and can intervene timely and effectively via online interventions (e.g., adapting website content or offering price promotions).

With a high exiting rate on e-commerce websites, deploying our M3PP model can help converting users into buyers; we are currently implementing the proposed M3PP model at our partner company.

## A DERIVATION OF THE LIKELIHOOD

We derive the likelihood of the M3PP model stated in Section 4.3. Eq. 10 follows directly by independence of the user sessions and the law of total probability. Hence, we need to derive the expression in Eq. 11 for the forward probabilities  $\alpha_t^d(i) = P(\{p_k^d, t_k^d \leq t\}, X(t) = i \mid \Omega)$ . In the following, we omit the dependence on the model parameters  $\Omega$  and the superscript  $d$ . We can rewrite the forward probabilities as,

$$\alpha_t(i) = \int_0^t P(\{p_k, t_k \leq t\}, X(t) = i, \Delta\tau = t - s) ds, \quad (20)$$

where  $\Delta\tau$  is the waiting time in the current latent shopping phase  $X(t) = i$ . Using the assumption in Section 4.3 that  $\{\tau_n\}_{n=1}^{K+1}$  is a subset of  $\{t_m\}_{m=1}^M$ , we see that  $\alpha_t(i)$  only takes values at  $\{t_m\}_{m=1}^M$ , and hence we can rewrite the forward probabilities as

$$\alpha_{t_m}(i) = \sum_{s=0}^m P(\{p_k, t_k\}_{k=1}^m, X(t_m) = i, \Delta\tau = t_{m+1} - t_{m-s}). \quad (21)$$

Splitting  $\{p_k, t_k\}_{k=1}^m$  into the observation during  $X(t_m) = i$  and before the phase  $i$  yields

$$\sum_{s=0}^m P(\{p_k, t_k\}_{k=1}^{m-s-1}, \{p_k, t_k\}_{k=m-s}^m, X(t_m) = i, \Delta\tau = t_{m+1} - t_{m-s}). \quad (22)$$

With Bayes' rule, and the transition probabilities of the Markov jump process  $X(t)$ , it follows that,

$$\begin{aligned} \sum_{s=0}^m \sum_{j \in \mathcal{X} \setminus \{i\}} P(\{p_k, t_k\}_{k=1}^{m-s-1}, X(t_{m-s-1}) = j) P(X(t_m) = i \mid X(t_{m-s-1}) = j) \\ \times P(\Delta\tau = t_{m+1} - t_{m-s} \mid X(t_m) = i) \\ \times P(\{p_k, t_k\}_{k=m-s}^m \mid X(t_m) = i), \end{aligned} \quad (23)$$

With the definition of the forward probabilities, it follows that

$$\begin{aligned} \sum_{s=0}^m \sum_{j \in \mathcal{X} \setminus \{i\}} \alpha_{t_{m-s-1}}(j) P(X(t_m) = i \mid X(t_{m-s-1}) = j) \\ \times P(\Delta\tau = t_{m+1} - t_{m-s} \mid X(t_m) = i) \\ \times P(\{p_k, t_k\}_{k=m-s}^m \mid X(t_m) = i), \end{aligned} \quad (24)$$

Using that the latent shopping phase process,  $X(t)$ , is Markov jump process with transition matrix  $A$  and rate parameter  $\{\gamma_i\}_{i=1}^N$ , the above can be rewritten as

$$\begin{aligned} \sum_{s=0}^m \sum_{j \in \mathcal{X} \setminus \{i\}} \alpha_{t_{m-s-1}}(j) A_{ji} \times \gamma_i e^{-\gamma_i(t_{m+1} - t_{m-s})} \\ \times P(\{p_k, t_k\}_{k=m-s}^m \mid X(t_m) = i). \end{aligned} \quad (25)$$

Since  $\{p_k, t_k\}_{k=m-s}^m$  follows a marked point process with independent points and marks as in Section 4.2, we can write

$$\begin{aligned} P(\{p_k, t_k\}_{k=m-s}^m \mid X(t_m) = i) &= \prod_{k=m-s}^m f(t_k, p_k \mid \mathcal{H}_{t_{k-1}}, X(t_m) = i) \\ &= \prod_{k=m-s}^m f(p_m \mid \mathcal{H}_{t_{m-1}}, X(t_m) = i) f(t_m \mid \mathcal{H}_{t_{m-1}}, X(t_m) = i). \end{aligned} \quad (26)$$

Together with the parameterization of the point process in Eq. 7 and the mark process in Eq. 8, this implies that

$$\begin{aligned} \alpha_{t_m}(i) &= \sum_{s=0}^m \sum_{j \in \mathcal{X} \setminus \{i\}} \alpha_{t_{m-s-1}}(j) A_{ji} \gamma_i e^{-\gamma_i(t_{m+1} - t_{m-s})} \\ &\times \exp\left(-\int_{t_{m-s}}^{t_m} \lambda_i^*(s) ds\right) \prod_{k=m-s}^m \lambda_i^*(t_k) q_{p_{k-1}, p_k}^i, \end{aligned} \quad (27)$$

where  $\lambda_i^*(s)$  is the conditional intensity function of the Hawkes process in Eq. 7.

## REFERENCES

- [1] Ahmed M. Alaa, Scott Hu, and Mihaela van der Schaar. 2017. Learning from clinical judgments: Semi-markov-modulated marked hawkes processes for risk prognosis. In *Proceedings of the 34th International Conference on Machine Learning*. JMLR.
- [2] Ahmed M. Alaa and Mihaela van der Schaar. 2018. A hidden absorbing semi-Markov model for informatively censored temporal data: Learning and inference. *Journal of Machine Learning Research* 19, 2015 (2018), 1–62.
- [3] Corin R Anderson, Pedro Domingos, and Daniel S Weld. 2002. Relational Markov models and their application to adaptive web navigation. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- [4] Fabricio Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgilio Almeida. 2009. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*.
- [5] José Borges and Mark Levene. 2000. Data mining of user navigation patterns. *Lecture Notes in Computer Science* 1836 (2000), 92–112.
- [6] José Borges and Mark Levene. 2007. Evaluating variable-length Markov chain models for analysis of user web navigation sessions. *IEEE Transactions on Knowledge and Data Engineering* 19, 4 (2007), 441–452.
- [7] Glen Wright Colopy, Stephen J. Roberts, and David A. Clifton. [n.d.]. Gaussian processes for personalized interpretable volatility metrics in the step-down ward. *IEEE Journal of Biomedical and Health Informatics* 3 ([n.d.]), 949–959.
- [8] David Roxbee Cox and Valerie Isham. 1980. *Point processes*. Vol. 12. CRC Press.
- [9] Kushal S Dave, Vishal Vaingankar, Sumanth Kolar, and Vasudeva Varma. 2013. Timespent based models for predicting user retention. In *Proceedings of the 22nd International Conference on World Wide Web*.
- [10] Mukund Deshpande and George Karypis. 2004. Selective Markov models for predicting web page accesses. *ACM Transactions on Internet Technology* 4, 2 (2004), 163–184.
- [11] Amy Wenxuan Ding, Shibo Li, and Patrali Chatterjee. 2015. Learning user real-time intent for optimal dynamic web page transformation. *Information Systems Research* 26, 2 (2015), 339–359.
- [12] 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*. ACM, 1555–1564.
- [13] Magdalini Eirinaki, Michalis Vazirgiannis, and Dimitris Kapogiannis. 2005. Web path recommendations based on page ranking and Markov models. In *Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management*.
- [14] Alex Gofman, Howard R. Moskowitz, and Tönis Mets. 2009. Integrating science into web design: Consumer-driven web site optimization. *Journal of Consumer Marketing* 26, 4 (2009), 286–298.
- [15] Şule Gündüz and M Tamer Özsu. 2003. A web page prediction model based on click-stream tree representation of user behavior. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 535–540.
- [16] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.
- [17] Donna L Hoffman and Thomas P Novak. 1996. Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing* 60, 3 (1996), 50–68.
- [18] Peter I Hofgesang. 2006. Relevance of time spent on web pages. In *Proceedings of KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- [19] Kazi T Islam, Christian R Shelton, Juan I Casse, and Randall Wetzel. 2017. Marked point process for severity of illness assessment. In *Machine Learning for Healthcare Conference*.
- [20] Chris Janiszewski. 1998. The influence of display characteristics on visual exploratory search behavior. *Journal of Consumer Research* 25, 3 (1998), 290–301.
- [21] Porter Jenkins. 2019. ClickGraph: Web page embedding using clickstream data for multitask learning. In *Companion Proceedings of The 2019 World Wide Web Conference*. ACM, 37–41.
- [22] Choudur Lakshminarayan, Ram Kosuru, and Meichun Hsu. 2016. Modeling complex clickstream data by stochastic models: Theory and methods. In *Proceedings of the 25th International Conference Companion on World Wide Web*. 879–884.
- [23] Mark Levene and George Loizou. 2003. Computing the entropy of user navigation in the web. *International Journal of Information Technology & Decision Making* 2, 3 (2003), 459–476.
- [24] William C McDowell, Rachel C Wilson, and Charles Owen Kile Jr. 2016. An examination of retail website design and conversion rate. *Journal of Business Research* 69, 11 (2016), 4837–4842.
- [25] Bamshad Mobasher. 2007. Data mining for web personalization. In *The Adaptive Web*. Springer, 90–135.
- [26] Wendy W Moe. 2003. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology* 13, 1-2 (2003), 29–39.
- [27] Wendy W Moe and Peter S Fader. 2004. Dynamic conversion behavior at e-commerce sites. *Management Science* 50, 3 (2004), 326–335.
- [28] Alan L Montgomery, Shibo Li, Kannan Srinivasan, and John C Liechty. 2004. Modeling online browsing and path analysis using clickstream data. *Marketing science* 23, 4 (2004), 579–595.
- [29] Sucheta Nadkarni and Reetika Gupta. 2007. A task-based model of perceived website complexity. *MIS Quarterly* (2007), 501–524.
- [30] James R Norris and James Robert Norris. 1998. *Markov Chains*. Number 2. Cambridge University Press.
- [31] Thomas P Novak, Donna L Hoffman, and Adam Duhachek. 2003. The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology* 13, 1-2 (2003), 3–16.
- [32] Jiangwei Pan, Vinayak Rao, Pankaj Agarwal, and Alan Gelfand. 2016. Markov-modulated marked poisson processes for check-in data. In *Proceedings of the 33rd International Conference on Machine Learning*. 2244–2253.
- [33] Jakob Gulddahl Rasmussen. 2011. Temporal point processes: The conditional intensity function. *Lecture Notes* (2011).
- [34] Narayanan Sadagopan and Jie Li. 2008. Characterizing typical and atypical user sessions in clickstreams. In *Proceedings of the 17th International Conference on World Wide Web*.
- [35] C Okan Sakar, S Olcay Polat, Mete Katircioglu, and Yomi Kastro. 2018. Real-time prediction of online shoppers’ purchasing intention using multilayer perceptron and LSTM recurrent neural networks. *Neural Computing and Applications* (2018), 1–16.
- [36] Ramesh R. Sarukkai. 2000. Link prediction and path analysis using Markov chains. *Computer Networks* 33, 1 (2000), 377–386.
- [37] Humphrey Sheil, Omer Rana, and Ronan Reilly. 2018. Understanding ecommerce clickstreams: A tale of two states. *KDD Deep Learning Day*. ACM (2018).
- [38] Statista. 2019. Global online shopper conversion rate 2017–2018.
- [39] Steven Tang, Joshua C Peterson, and Zachary A Pardos. 2016. Deep neural networks and how they apply to sequential education data. In *Proceedings of the 3rd ACM Conference on Learning@Scale*.
- [40] A. Wald. 1945. Sequential test of statistical Hypothesis. *Annals of Mathematical Statistics* 16, 2 (1945), 117–186.
- [41] Lizhen Xu, Jason A. Duan, and Andrew Whinston. 2014. Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science* 60, 6 (2014), 1392–1412.
- [42] Jianhan Zhu, Jun Hong, and John G Hughes. 2002. Using Markov chains for link prediction in adaptive web sites. *Building* (2002), 60–73.