

Modeling User Return Time using Inhomogeneous Poisson Process

Mohammas Akbari^{1,2}, Alberto Cetoli², Stefano Bragaglia², Andrew D.
O’Harney², Marc Sloan², and Jun Wang¹

¹ University College London

² Context Scout

Abstract. For Intelligent Assistants (IA), user activity is often used as a lag metric for user satisfaction or engagement. Conversely, predictive leading metrics for engagement can be helpful with decision making and evaluating changes in satisfaction caused by new features. In this paper, we propose User Return Time (URT), a fine grain metric for gauging user engagement. To compute URT, we model continuous inter-arrival times between users’ use of service via a log Gaussian Cox process (LGCP), a form of inhomogeneous Poisson process which captures the irregular variations in user usage rate and personal preferences typical of an IA. We show the effectiveness of the proposed approaches on predicting the return time of users on real-world data collected from an IA. Experimental results demonstrate that our model is able to predict user return times reasonably well and considerably better than strong baselines that make the prediction based on past utterance frequency.

Keywords: User Return Time Prediction · Intelligent Assistant

1 Introduction

Intelligent Assistants (IAs) are software agents that interact with users to complete a specific task. The success of an IA is directly linked to long-term user engagement which can be measured by observing when users return to reuse the IA. Indeed, the repeating usage of service is a typical characteristic for IAs, thus, monitoring a user’s usage pattern is a necessary measure of engagement.

Predicting when usage will next occur is therefore a useful indicator of whether a user will continue to be engaged and can serve as a reference to compare against when introducing new features, as well as a method for managing churn for business purposes. For example, predicted return times of customers can be utilized for clustering customers according to their activity and narrow their interest to investigate a specific groups of users with a short or long inter-arrival time for target marketing [12, 2]. Furthermore, it helps the service to prepare content for the customers in advance to better serve them in engagement utterance. For example, a target marketing and advertising program can be planned for the next engagement of the user.

Modeling the inter-arrival time of user engagement events is a challenging task due to the complex temporal patterns exhibited. Users typically engage with IA systems in an ad hoc fashion, starting tasks at different time points with irregular frequency. For example, Figure 1 shows the inter-arrival times (denoted by black crosses) of two users in one week. Notice the existence of regions of both high and low density of inter-arrival times over a one week interval. Users do not arrive in evenly spaced intervals but instead they usually arrive in times that are clustered due to completing several tasks in bursts. This may be attributed to a user’s personal preferences and their tasks’ priorities. Retrospective studies in modeling inter-arrival times between events treat events as independent and exponentially distributed over time with constant rate [1]. They hence fail to perform accurate predictions when there exists time-varying patterns between events [14].

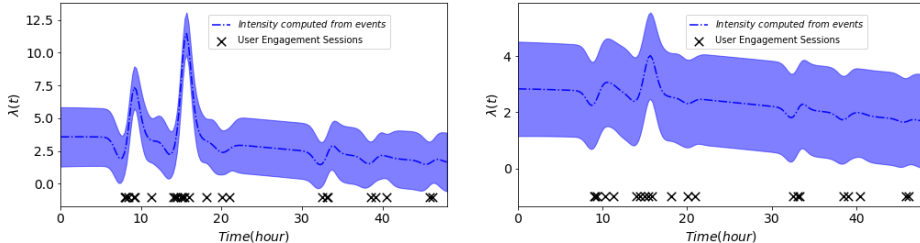


Fig. 1. Intensity functions (dotted lines) and corresponding predicted user inter-arrival times for different users (black crosses). Light regions depict the uncertainty of estimations.

To this purpose, we define User Return Time (URT) as the predicted inter-arrival time until next user activity, allowing us to predict user engagement to aid in creating an optimal IA system. To do so, we propose to model inter-arrival times between a user usage sessions with a doubly stochastic process. More specifically, we leverage the log-Gaussian Cox process (LGCP), an inhomogeneous Poisson process (IPP), to model the inter-arrival times between events. LGCP models return times that are generated by an intensity function which varies across time. It assumes a non-parametric form for the intensity function allowing the model complexity to depend on the data set. We evaluate the proposed model using a real-world dataset from an IA, and demonstrate that it provides good predictions for inter-arrival return times, improving upon the baselines. Even though the main application is return time estimation for an IA, one could apply the proposed approach to other events in e-commerce systems such as the arrival times of requests, return time of customers, etc.

The contribution of this paper can be summarized as,

- We propose User Return Time as a measure to predict user engagement with an IA.
- We leverage a doubly-stochastic process, i.e. Log-Gaussian Cox processes, to predict URT for an IA, and show that it effectively captures the time-varying patterns of usage between events.

- We verified the effectiveness of the proposed method on a real-world dataset and found that Gaussian and Periodic kernels can make accurate estimations.

2 Problem Statement

Let $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ denote a set of n different users who used the IA to assist with different tasks. Let $\mathcal{H}_i = \{t_{i,j}\}_{j=1}^{n_i}$ denote the task history of the i -th user, where $t_{i,j}$ represents the start time of the j -th task performed by the i -th user and n_i is the number of tasks performed by user i . Our aim is to estimate a user’s next start time from their past interactions with the IA, i.e., predict $t_{i,k+1}$ based on the start times of the previous sessions, i.e. $\{t_{i,j}\}_{j=1}^k$.

Based on the above discussion, we formally define the problem of **Predicting User Return Time** as: *Given a user history of session start times \mathcal{H}_i , predict the next times that the user will use the IA.*

3 Model

Poisson processes have been widely adopted for estimating the cross-interval times between different events such as failure of devices, social media events and purchase in e-commerce sites [11]. The Homogeneous Poisson process (HPP) is a class of point processes that assumes the events are generating with a constant intensity rate, i.e. λ , (with respect to the time and the product features). However, in an IA scenario, user engagement incidence often occurs with a varying-rate over time where there are several spikes of usage in a short period and a long-period of absence (when the user performs other tasks). Thus, we exploit the Inhomogeneous Poisson process (IPP) [6] that can model events happening at a variable rate by considering the intensity to be a function of time, i.e. $\lambda(t)$. For example, Figure 1 shows intensity functions learned from two different IPP models. Notice how the generated inter-event times vary according to the intensity function values.

To model inter-arrival times, we employed a log-Gaussian Cox process which models the intensity function of point processes as a stochastic function [7]. LGCP learns the intensity function, $\lambda(t)$, non-parametrically via a latent function sampled from Gaussian processes [9]. Here, to impose non-negativity to the intensity function (as an interval cannot be negative), we assume an exponential form for the intensity function, i.e., $\lambda(t) = \exp(f(t))$. We adopted a non-parametric approach to model the intensity function which utilizes Bayesian inference to train a model, where the complexity of the model is learned from the training data available. In the next section, we explain the details of the proposed model and how to learn model parameters from training data.

3.1 Modeling inter-arrival time

An inhomogeneous Poisson process (unlike HPP) uses a time varying intensity function and hence, the distribution of inter-arrival times is not independent and

identically distributed [10]. In IPP, the number of tasks y (signals of a returning user to the IA) occurring in an interval $[p, q]$ is Poisson distributed with rate $\int_p^q \lambda(t) dt$,

$$p(y|\lambda(t), [p, q]) = \text{Poisson} \left(y \mid \int_p^q \lambda(t) dt \right) = \frac{\left(\int_p^q \lambda(t) dt \right)^y \exp \left(- \int_p^q \lambda(t) dt \right)}{y!} \quad (1)$$

Assume that the k -th event occurred at time $t_k = p$ and we are interested in the inter-arrival time $\delta_k = t_{k+1} - t_k$ of the next event. The arrival time of next event t_{k+1} can be obtained as $t_{k+1} = t_k + \delta_k$. The cumulative distribution for δ_k , which provides the probability that an event occurs by time $p + q$ can be obtained as,

$$\begin{aligned} p(\delta_k \leq q) &= 1 - p(\delta_k > q | \lambda(t), t_k = p) \\ &= 1 - \exp \left(- \int_p^{p+q} \lambda(t) dt \right) = 1 - \exp \left(- \int_0^q \lambda(p+t) dt \right). \end{aligned} \quad (2)$$

The derivation is obtained by considering a Poisson probability for zero counts with rate parameter given by $\int_p^{p+q} \lambda(t) dt$ and applying integration by substitution to obtain Eq.(2). The probability density function of the random variable T_n is obtained by taking the derivative of Eq.(2) with respect to q ,

$$p(\delta_k = q) = \lambda(p+q) \exp \left(- \int_0^q \lambda(p+t) dt \right). \quad (3)$$

We associate a distinct intensity function $\lambda_i(t) = \exp(f_i(t))$ to each user u_i as they have different temporal preferences. The latent function $f_i(t)$ is modeled to come from a zero mean Gaussian process (GP)[9] prior. The Squared Exponential (SE) kernel is a common choice for GPs where the kernel is defined as,

$$k(t_i, t_j) = \sigma \exp \left(- \frac{(t_i - t_j)^2}{l} \right). \quad (4)$$

Eq(4) imposes smoothness over time on the intensity function. We also experiment with periodic kernels which allow the modelling of functions that repeat themselves exactly. Periodic kernels can model complex periodic structure relating to the working week by finding a proper periodicity hyperparameter in the kernel. The periodic kernel is defined as,

$$k_{\text{periodic}}(t_i, t_j) = \sigma^2 \exp \left(- \frac{2 \sin^2 \pi |t_i - t_j| / r}{l^2} \right), \quad (5)$$

where σ and l are the output variance and length-scale, respectively, and r is the periodicity hyperparameter.

3.2 Inference

We learn the model parameters by maximizing the marginal likelihood over all users in the dataset. The likelihood of the return times over the dataset is then obtained by taking the product of return times over all users.

4 Experiments

4.1 Dataset

We conducted an empirical study in which we used the access log of a commercial web IA to construct our dataset. The IA has been developed as an extension for the Chrome Web browser that assists recruiters to automatically find and segregate job candidate information such as skills and contact details, based on the information found on popular social networking platforms such as LinkedIn.

For the purpose of this experiment the tool was altered to log the interactions of users with the tool when they performed their tasks. Various interactions were logged with a time stamp, based on which we constructed our dataset. Table 1 depicts a sample log record for a specific user from our dataset. Inspired by [4, 3], we split action sequences into sessions based on a time gap of 15 minutes which is commonly used in information retrieval and web search to identify sessions. We collected all logs of a random set of users within a period of 3 months, from the beginning of April 2018 until the end of June 2018. The dataset consists of 2,999,593 interaction events committed by 133 distinct users.

Table 1. Example log of a user’s interaction sequence. The first three interactions occurred within a single task. The final interaction indicated the start of a new task.

User	Action	Time stamp
484	New web page	2018-07-01 17:35:25
484	Opened IA	2018-07-01 17:35:27
484	Clicked Contact Button	2018-07-01 17:35:51
484	New web page	2018-07-01 18:05:25

4.2 Baselines and Evaluation Metrics

Here the proposed model is compared against several methods to evaluate their effectiveness in predicting URT. We discuss the advantages, assumptions and limitations of each and provide empirical results on a real-world dataset. We examine the following distinct models: (1) **Linear Regression**: a linear regression model which is trained on a historical window of URT. We used the last 20 (computed empirically) inter-arrival times as features. (2) **HPP**: we also used a homogeneous Poisson process (HPP) [5] which models an exponentially distributed inter-arrival times with a fix rate λ . The rate parameter was learned based on the maximum likelihood approach. (3) **HP**: we compare against the Hawkes Process (HP) [13], a self exciting point process where an occurrence of an event increases the probability of the event arriving soon afterwards. We consider a univariate Hawkes process where the intensity function is modeled as $\lambda(t) = \mu + \sum_{t_i < t} k(t_i, t)$. We apply Ogata’s thinning algorithm for generating arrival times using Hawkes process [8]. (4) **GP**: we also exploited GP as a time-series baseline, where the inter-arrival time is modeled as a function of the time of occurrence of last tweet. We examine two commonly used kernels, SE and periodic, where kernel parameters are learned by maximizing the likelihood.

(5) **RNN**: the final baseline was selected from deep learning approaches, where we used an LSTM model with two Bi-LSTM layers and input length of 20 for predicting URT (we used the same window size as with linear regression).

In order to evaluate the model, we use mean absolute error (MAE) and root mean square error (RMSE) between the actual and the predicted times for each user in hours. Since the data varies in size for each user, we take the micro average of the errors to obtain the final result.

4.3 Results

Table 4.3 compares the predictive performance of LGCP against various baselines. We find that the standard kernel used in GP models, the SE kernel, performs poorly as expected due to the complex temporal patterns exhibited by users in their session start times. The SE kernel typically models smoothly varying functions and is not suitable to model this situation. The periodic kernel could model the periodicity in the data (for instance, users tend to be more active on weekends) and are found to perform better than SE in both GP and LGCP. The LGCP models with periodic kernel outperforms the baseline approaches such as HPP, linear regression, GP regression. RNN model can outperform all baselines except LGCP which shows that deep models can capture complex user behavior in using IA.

Table 2. Comparison of different approaches in terms of MAE and RMSE.

Method	Kernel	MAE	RMSE
Linear Regression		53.27(\pm 48.91)	86.10(\pm 79.13)
HPP		43.44(\pm 51.13)	63.70(\pm 53.60)
HP		28.12(\pm 28.47)	61.24(\pm 51.22)
RNN		19.11(\pm 28.01)	45.52(\pm 34.20)
GPR	SE	35.52(\pm 40.79)	42.51(\pm 54.22)
	Periodic	19.82(\pm 21.04)	38.30(\pm 48.71)
LGCP	SE	26.52(\pm 29.79)	54.12(\pm 58.37)
	Periodic	15.52(\pm 18.79)	32.54(\pm 38.15)

5 Conclusions

In this paper we proposed to use User Return Time as a predictive measure of engagement with an IA, for which the log-Gaussian Cox process was proposed as an appropriate prediction model. Through our experiments, we demonstrated that this model does indeed offer better predictive performance due to its ability to capture the complex temporal behaviour typical of IA users.

This approach can be generalized to model problems other than URT prediction for an IA, e.g. purchase time prediction, advertisement campaigns, and disaster management. The effectiveness of using an RNN in this research also shows that it is worth investigating the potential of deep sequence models for prediction in these scenarios. In future, we plan to examine user and context features to improve prediction performance.

References

1. Doerr, C., Blenn, N., Van Mieghem, P.: Lognormal infection times of online information spread. *PloS one* **8**(5), e64349 (2013)
2. Du, N., Wang, Y., He, N., Sun, J., Song, L.: Time-sensitive recommendation from recurrent user activities. In: *Advances in Neural Information Processing Systems*. pp. 3492–3500 (2015)
3. Halfaker, A., Keyes, O., Kluver, D., Thebault-Spieker, J., Nguyen, T., Shores, K., Uduwage, A., Warncke-Wang, M.: User session identification based on strong regularities in inter-activity time. In: *Proceedings of the 24th International Conference on World Wide Web*. pp. 410–418. International World Wide Web Conferences Steering Committee (2015)
4. He, D., Göker, A.: Detecting session boundaries from web user logs. In: *Proceedings of the BCS-IRSG 22nd annual colloquium on information retrieval research*. pp. 57–66 (2000)
5. Hosseini, S.A., Alizadeh, K., Khodadadi, A., Arabzadeh, A., Farajtabar, M., Zha, H., Rabiee, H.R.: Recurrent poisson factorization for temporal recommendation. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 847–855. ACM (2017)
6. Lee, S., Wilson, J.R., Crawford, M.M.: Modeling and simulation of a nonhomogeneous poisson process having cyclic behavior. *Communications in Statistics-Simulation and Computation* **20**(2-3), 777–809 (1991)
7. Møller, J., Syversveen, A.R., Waagepetersen, R.P.: Log gaussian cox processes. *Scandinavian journal of statistics* **25**(3), 451–482 (1998)
8. Ogata, Y.: On lewis’ simulation method for point processes. *IEEE Transactions on Information Theory* **27**(1), 23–31 (1981)
9. Rasmussen, C.E.: Gaussian processes in machine learning. In: *Advanced lectures on machine learning*, pp. 63–71. Springer (2004)
10. Ross, S.M.: *Introduction to probability models*. Academic press (2014)
11. Takeshi Sakaki, M.O., Matsuo, Y.: Earthquake shakes twitter users: Real-time event detection by social sensors (2010)
12. Yang, C., Shi, X., Jie, L., Han, J.: I know you’ll be back: Interpretable new user clustering and churn prediction on a mobile social application. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 914–922. ACM (2018)
13. Yang, S.H., Zha, H.: Mixture of mutually exciting processes for viral diffusion. In: *International Conference on Machine Learning*. pp. 1–9 (2013)
14. Zhang, R., Walder, C., Rizoïu, M.A., Xie, L.: Efficient non-parametric bayesian hawkes processes. In: *WWW* (2018)