Modeling Item-Specific Temporal Dynamics of Repeat Consumption for Recommender Systems

Chenyang Wang
BNRist, DCST, Tsinghua University
Beijing, China
wangcy18@mails.tsinghua.edu.cn

Min Zhang*
BNRist, DCST, Tsinghua University
Beijing, China
z-m@tsinghua.edu.cn

Weizhi Ma
BNRist, DCST, Tsinghua University
Beijing, China
mawz14@mails.tsinghua.edu.cn

Yiqun Liu
BNRist, DCST, Tsinghua University
Beijing, China
yiqunliu@tsinghua.edu.cn

Shaoping Ma
BNRist, DCST, Tsinghua University
Beijing, China
msp@tsinghua.edu.cn

ABSTRACT

Repeat consumption is a common scenario in daily life, such as repurchasing items and revisiting websites, and is a critical factor to be taken into consideration for recommender systems. Temporal dynamics play important roles in modeling repeat consumption. It is noteworthy that for items with distinct lifetimes, consuming tendency for the next one fluctuates differently with time. For example, users may repurchase milk weekly, but it is possible to repurchase mobile phone after a long period of time. Therefore, how to adaptively incorporate various temporal patterns of repeat consumption into a holistic recommendation model has been a new and important problem.

In this paper, we propose a novel unified model with introducing Hawkes Process into Collaborative Filtering (CF). Different from most previous work which ignores various time-varying patterns of repeat consumption, the model explicitly addresses two item-specific temporal dynamics: (1) *short-term effect* and (2) *life-time effect*, which is named as **S**hort-Term and **L**ife-Time **R**epeat **C**onsumption (SLRC) model. SLRC learns importance of the two factors for each item dynamically by interpretable parameters. According to extensive experiments on four datasets in diverse scenarios, including two public collections, SLRC is superior to previous approaches for repeat consumption modeling. Moreover, due to the high flexibility of SLRC, various existing recommendation algorithms are shown to be easily leveraged in this model to achieve significant improvements. In addition, SLRC is good at balancing recommendation for novel items and consumed items (exploration and exploitation). We also find that the learned parameters is highly interpretable, and hence the model is able to be leveraged to discover items' lifetimes, and to distinguish different types of items such as durable and fast-moving consumer goods.

KEYWORDS

Recommender system, Repeat consumption, Temporal dynamics, Collaborative filtering, Hawkes process

*Contact author

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313594

ACM Reference Format:
Chenyang Wang, Min Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. 2019. Modeling Item-Specific Temporal Dynamics of Repeat Consumption for Recommender Systems. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3308558.3313594

1 INTRODUCTION

Recommender system plays a crucial role in users' engagements with web-services, which makes it easier for users to explore endless items that they have not consumed (novel items) [19, 23, 36]. Meanwhile, repeat consumptions come to account for a large proportion of user-item interactions and become increasingly important in recent years. Therefore, properly recommending items user consumed before becomes a vital part of recommender systems nowadays. For example, an online shopping website can recommend the milk powder user consumed before when it is time to be used up. It is widely agreed that providing consumed items at the right time can greatly improve both users' satisfaction and sellers' profit [3, 11, 12, 40].

Temporal dynamics is an important factor to accurately model repeat consumption. Previous study about user consumption sequence presents a holistic model of sequential consumption [4]. The model takes time factor, distance factor, quality factor into consideration and the time factor turns out to be the most important one. Du et al. [16] utilize temporal point process to model recurrent user activities. That is to say, previous consumptions are likely to trigger repeat consumptions for those items (self-excitation) and such excitation decays with time.

However, existing studies about repeat consumption (e.g. [3–5, 8, 16, 32]) do not focus on modeling how the tendency of repeat consumption fluctuates with time, especially the difference between items. As a result, they fail to capture various temporal dynamics across distinct items. Actually, it is noteworthy that different items demonstrate diverse temporal patterns of repeat consumption. In particular, there are both (1) *short-term* and (2) *life-time* effects. To illustrate these two factors, Figure 1 gives an example of two items' purchase tendencies (baby diaper and feeding bottle) drifting with time. For a new mother, both baby diaper and feeding bottle are needed, so the base purchase tendencies are similar at first. After she buys both items simultaneously, their purchase tendencies increase immediately and diminish with time (*short-term effect*), which means the mother probably consumes these two items again in a short term for some reasons such as future usage. On the

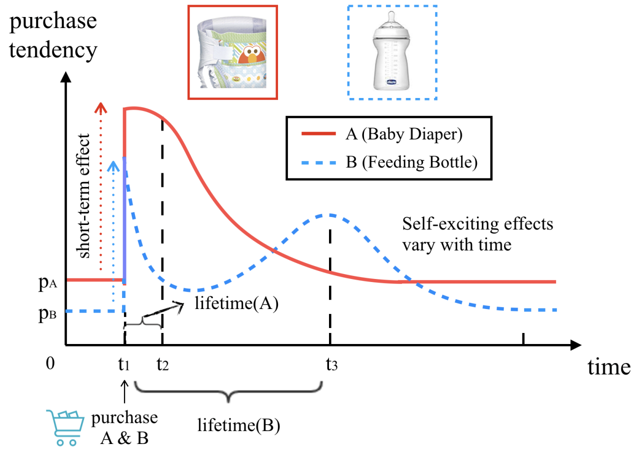

Figure 1: Illustration of two items' purchase tendencies drifting with time. There existing *short-term effect* for both baby diaper and feeding bottle after purchase, and such effect decays with time. On the other hand, *life-time effect* peaks after different periods of time, which is short for baby diaper (superposed with short-term effect) but long for feeding bottle. Both items' purchase tendencies finally decay to their basic values as time goes by.

other hand, users also tend to purchase the item again near the end of item's lifetime [20, 30]. In Figure 1, the purchase tendency of feeding bottle has another spike around t_3 (*life-time effect*), which is probably the time to change a new one. With regard to baby diaper, lifetimes of such fast-moving consumer goods are so short that the corresponding effect is superposed with short-term effect, resulting in higher tendency shortly after purchasing. As time goes by, the self-exciting effect diminishes and both items' purchase tendencies return to their basic values. Such difference between items also holds for many other cases such as snacks for A and mobiles for B.

Current methods modeling repeat consumption do not jointly model both two characteristics for different items. However, the lack of variety of temporal factors can lead to lower predictive ability. For example, when life-time effect is not considered, the dominated short-term effect will result in continuous recommendation for mobiles shortly after purchasing one, which would be annoying to users. And mobiles could not be recommended when it is the time to change to a new one.

In this paper, firstly we conduct a detailed analysis of repeat consumption and temporal dynamics on four real-world datasets spanning over 30 months. Two key factors are revealed to model the complex time-varying patterns of repeat consumption: (1) *short-term effect* and (2) *life-time effect*. We find there are many repeat consumption shortly after previous purchase even for durable goods like baby walker (short-term effect). Besides, a consumption is likely to trigger next consumptions for the same item when item's lifetime runs out (life-time effect). Moreover, the importance of short-term and life-time effect can vary wildly between distinct items.

Then, based on these findings, we propose a holistic model that can recommends both novel and consumed items simultaneously, called SLRC referring to **S**hort-Term and **L**ife-Time **R**epeat **C**onsumption model. SLRC innovatively combines collaborative filtering (CF) [24] and Hawkes Process [17], with the two temporal characteristics of repeat consumption modeled by the kernel function in Hawkes Process. The CF part of the model projects users and items to the same space, in which similar representations stand for similar properties. As a result, the CF part helps to better capture users' preferences and explore novel items. While Hawkes process utilizes patterns of repeat consumption to properly rank consumed items higher at specific time (when users need items previously consumed again), leading to a holistic model with a good balance of exploration and exploitation. Moreover, the item-specific modeling of repeat consumption makes it possible to reveal fine-grained temporal patterns and automatically discover distinct items' lifetimes, which often need to be predefined in previous work [27]. The main contributions of this work can be summarized as follows:

- Two temporal dynamics of repeat consumption: (1) short-term effect and (2) life-time effect are revealed through empirical study, which vary a lot for distinct types of items.
- A unified model, **S**hort-Term and **L**ife-Time **R**epeat **C**onsumption model (SLRC), is proposed with innovative combination of two techniques: Collaborative Filtering (CF) and Hawkes Process. The model adaptively incorporates two item-specific temporal dynamics of repeat consumption into kernel function of Hawkes Process. And the integration of CF helps SLRC to better capture users' intrinsic preference, achieving a good balance of exploration and exploitation.
- Comparative experimental results on four datasets indicate SLRC is superior to previous models of repeat consumption. SLRC is also flexible to leverage different state-of-the-art recommendation algorithms and further enhance the performance and interpretability of originals.
- With high interpretability, SLRC is able to discover distinct items' lifetimes, and hence distinguish different kinds of items such as durable goods and fast-moving consumer goods.

2 RELATED WORK

2.1 Repeat Consumption

Repeat consumption has been studied in various domains, including web revisitation [2, 9, 29, 45], repeated web search queries [40, 41], music listening [20], time-limited coupons [32], and predicting consumption rate [28]. All the work demonstrates that repeat consumption is common and important in various fields.

Early explorations of repeat consumption utilize marginal utility in economics to make prediction [42]. The *Law Of Diminishing Marginal Utility* [7] states that an item's marginal utility decreases as its quantity grows. However, this law may not always hold especially in item purchasing scenario. For example, users may buy fast-moving consumer goods again and again regularly although the accumulated quantities are growing.

There is a series of work focusing on some simplified tasks related to repeat consumption to better model intrinsic patterns. Some work predicts whether a purchase will be a repeat consumption according to item features in current time window [11]. And some other work predicts which consumed item user will prefer given the fact that current consumption is a repeat consumption [3, 5, 12]. Especially, Anderson et al. find *recency* is a key factor with a cache-based method, indicating that users tend to purchase items recently consumed [3]. Different from these studies, we focus to build a unified and holistic model that can recommend both consumed items and novel items simultaneously, where consumed items are only properly selected at the right time.

Further, Benson et al. [4] presents a holistic model of sequential repeat consumption, which introduces time factor, distance factor (*recency*) and quality factor. Most recently, Cai et al. [8] and Du et al. [16] utilize revamped Hawkes Process to model sequential online interactive behaviors and recurrent user activities, respectively. Although these methods can make holistic recommendation, they either estimate temporal patterns over all items or move one more step to model item-specific effect of previously consumed items (self-excitation or mutual-excitation), but the degree of excitation simply decays globally with time. All these studies are lack in taking various temporal dynamics of repeat consumption into consideration. Moreover, without integrating collaborative filtering methods, they may fail to accurately capture users' personalized preference.

2.2 Collaborative Filtering and Point Process

There are other two lines of research related to our work: *Collaborative Filtering* and *Point Process.*

Collaborative Filtering analyzes relationships between users and interdependencies among products to identify new user-item associations, which is domain free and more accurate than content-based techniques [25]. Latent factor model is one of the primary methods of collaborative filtering, in which users and items are characterized on some specific number of factors inferred from the interaction patterns. A series of latent factor models have been proposed to improve performance of item recommendation [6, 10, 18, 19, 23, 35, 36]. The utilization of mutual relationships between users and items enables these methods to better capture users' preference and discover novel items users may be interested in.

Point Process is good at modeling sequential events localized in time [14]. There have been many applications of point process, including modeling user influence in social network [39, 46], predicting earthquakes [33], paper citation count [44] and user return times [38]. As a variant of temporal point process, Hawkes Process [17] explicitly models the effect of previous events (self-excitation). The triggering kernel of intensity function in Hawkes Process controls how excitation varies with time. There have been several studies leveraging recurrent neural network (RNN) to derive intensity or replace triggering kernel [15, 34, 43], but it is hard for RNN to capture very long dependencies of repeat consumption when spanning over a long period of time. The lack of interpretability is also the main weakness. Recent work utilizes temporal point process to predict high-level human actions such as eating and biking, which accounts for three time-sensitive characteristics [26]. Compared to only 8-10 actions in the work, there are millions of items in personalized recommendation task, in which case their model contains intractably numerous parameters and will fail to explore endless novel items without integrating collaborative filtering.

We further build a holistic model with combination of Collaborative Filtering and Hawkes Process to take full advantage of both techniques. The CF part of the model captures users' inherent preference and helps to explore novel items, while two temporal dynamics of repeat consumption modeled by Hawkes Process makes it possible to properly exploit consumed items, leading to a good balance of exploration and exploitation.

Table 1: Statistics of Datasets.

Dataset	#user ($\lvert\mathcal{U}\rvert$)	#item ($\lvert\mathcal{I}\rvert$)	#entry ($\sum_u \lvert S_u\rvert$)	repeat ratio	time span
Baby	25.4k	7.6k	238.7k	17.8%	2012.09 - 2014.05
Order	34.6k	86.2k	820.4k	14.0%	2016.12 - 2018.03
Recsys2017	113.8k	44.7k	1240.4k	21.2%	2016.11 - 2017.02
BrightKite	29.7k	105.2k	3591.9k	90.0%	2008.03 - 2010.10

3 EMPIRICAL STUDY ON REPEAT CONSUMPTION

In this section, we first describe the task we are tackling. Then we make several observations about repeat consumption on four real-world datasets, which serve as the foundation of our model.

3.1 Task Definition

Given a target time $\hat{t}$ and a user's consumption history before $\hat{t}$, our task is to predict top-k items the user may be interested in at target time. Note that the recommendation list here can contain both novel items and consumed items. Formally, let $\mathcal{U}$ be the set of all users. Each user $u \in \mathcal{U}$ has a consumption sequence $S_u = \{(i_1, t_1), (i_2, t_2), \cdots, (i_{N_u}, t_{N_u})\} \in \mathcal{S}$ with N_u interactions. Each interaction consists of an item $i_n \in \mathcal{I}$ and corresponding time $t_n \in \mathbb{R}^+$ $(0 \leq t_n \leq T)$, which means user u consumed item i_n at time t_n. T is the total time span observed. $\mathcal{S}$ represents all the consumption sequence in data. The sequence is sorted by time in an ascending order, i.e., $t_n \leq t_{n'}$ for any $n < n'$. For simplicity, we denote a user's consumption sequence up to time t as $S_t^u = \{(i', t') | (i', t') \in S_u \wedge t' < t\}$.

Then, we define *repeat consumption* as below:

Definition 3.1. The interaction (i, t) conducted by user u is a *repeat consumption* if and only if $\exists (i', t') \in S_t^u : \ i' = i \wedge t' < t$

3.2 Dataset Description

To illustrate critical temporal patterns of repeat consumption, we use four datasets in distinct scenarios. The datasets are described below and the consumption statistics after filtering are summarized in Table 1.

Baby. This dataset contains online purchasing information within baby category from an e-commerce retailer.

Order. This dataset contains mobile payment records in supermarkets and convenient stores. We have provided the anonymized dataset online[1].

Recsys2017. Recsys Challenge 2017[2] is a competition for job recommendation, where the item is job posting and the consumption is clicking. Then a repeat consumption is that a user clicks the same job posting again after previous click. We use click interactions in open training data (offline phase).

[1]https://github.com/THUwangcy/SLRC/tree/master/data
[2]https://www.recsyschallenge.com/2017/

BrightKite. BrightKite was a location-based social networking website, where users could check in to physical locations. The data is publicly available[3] [13].

Each record in these datasets contains user-item pair and corresponding timestamp. Users and items with less than 5 associated interactions are filtered out. Besides, we remove two types of users: (1) interactions take place within one day (likely to be attracted by promotions); (2) consumes too frequently (more likely a merchant but a normal user). These four datasets cover consumptions in different scenarios, including both online and offline consumptions, as well as different scene of shopping, job seeking and check-in. Besides the ratio of repeat consumption varies across datasets. For *BrightKite*, 90% consumptions are repeat consumption, while the repeat ratio is moderate in other datasets.

3.3 Temporal Dynamics of Repeat Consumption

To reveal the temporal dynamics of repeat consumption, we try to figure out how previous consumption triggers next purchasing for the same item and how repurchasing tendency changes with time. However, it is hard to observe actual self-exciting pattern, which is latent and we can only observe consumption behavior as a result of the latent factor. Here, we resort to study the inter-consumption gap, which means the time interval between a user's two adjacent consumptions for the same item. Inter-consumption gap is a good indicator of how the tendency for consuming an item again varies with time. For example, if there is a lot of people purchasing milk one day after previous consumption, it indicates the tendency of repurchasing milk is high after one day.

Through investigating overall and item-level distributions of inter-consumption gap, we make two important empirical observations about repeat consumption. First, consumptions for most items are likely to trigger next consumption for the same item in the short term, which we call *short-term effect.* Second, some items tend to be consumed again centrally after a period of time, which we call *life-time effect.* These two types of effects serve as the foundation of the proposed SLRC model in the next section.

3.3.1 Short-term effect. Figure 2 shows the overall distribution of inter-consumption gap in the datasets. Except for *Recsys2017*, all the distributions monotonically decrease with a heavy tail. The distribution in *Recsys2017* also diminishes quickly in the beginning, which means generally there are a lot of repeat consumptions happening shortly after previous consumption. Such short-term effect is very common in daily life. Figure 3 shows different items' inter-consumption gap distributions in *Baby* and *Order* datasets. Take milk powder in Figure 3(b) for example, one may occasionally buy a kind of milk powder online and appreciates it very much. As a result, she purchases the milk powder again recently. Even for items that are not fast-moving consumer goods, like baby care cream in Figure 3(a), it is possible to repurchase them in a short term just as a gift to others or for other reasons. As a result, the time-decaying short-term effect can be leveraged to basically model the tendency to repurchase.

3.3.2 Life-time effect. Apart from short-term effect, there is also life-time effect for repeatedly consuming items. The distribution in

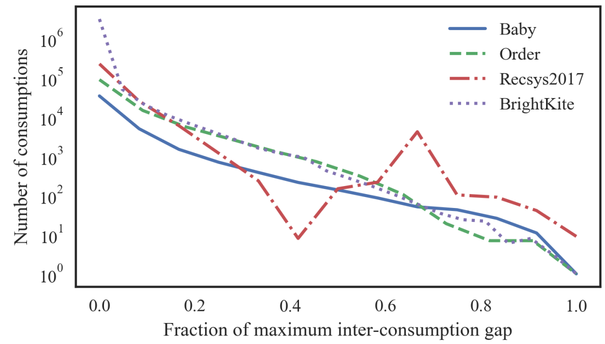

Figure 2: Distributions of inter-consumption gap. On the whole, most repeat consumptions happen shortly after the previous, while life-time effect is obvious in *Recsys2017* dataset.

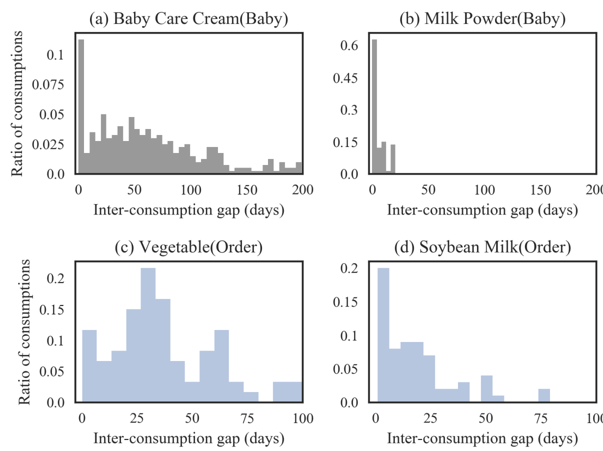

Figure 3: Inter-consumption gap distributions of different items. Various types of items inherently have distinct temporal dynamics of repeat consumption.

Recsys2017 dataset presents another spike after a long-period time, which means users may forget the information of previously clicked posts and click them again. Almost all items have their inherent lifetimes and demonstrates different time-varying patterns. For other datasets, the monotonically decreasing form on the whole can be a superimposed result of distinct distributions for various items. In Figure 3, we can see time-varying patterns differ from each other greatly. For example, in Figure 3(a)(c), baby care cream runs out in about 50 days and some families may go to supermarket monthly to resupply vegetables. When item's lifetime is going to run out, users will be more likely to purchase it again. Therefore, when taking life-time effect into consideration, we can make more accurate prediction at the right time (e.g. recommending durable goods consumed before after a long period when user tends to repurchase them).

[3]https://snap.stanford.edu/data/loc-brightkite.html

In summary, two temporal patterns of repeat consumption: (1) *short-term effect* and (2) *life-time effect* are uncovered through investigation into inter-consumption gap distribution. More importantly, the importance of these two factors varies wildly across different items. Therefore, addressing item-specific temporal dynamics of repeat consumption will help to make more relevant recommendation.

4 SLRC MODEL

Based on the empirical observations above, we propose a holistic model with combination of collaborative filtering and Hawkes Process, which is named as **S**hort-Term and **L**ife-Time **R**epeat **C**onsumption (SLRC) model. Some preliminaries about temporal point process are introduced first. Then we detailedly describe model definition and parameter learning.

4.1 Preliminaries About Temporal Point Process

Formally, a temporal point process is a random process of which the realization consists of a list of discrete events localized in time, $\{t_n\}_{n\in\mathbb{N}}$ with the time $t_n \in \mathbb{R}^+$. In the scene of item consumption, $\{t_n\}_{n\in\mathbb{N}}$ represents a series of time that a user purchases some specific item. Given the history time of past events S_t, temporal point process introduces conditional intensity function $\lambda(t|S_t)$ representing a stochastic model for the time of the next event given all the times of previous events. For simplicity, we omit conditional sign and denote $\lambda(t)$ as conditional intensity function. Then, the probability for the occurrence of a new event given the history time S_t within a small time window $[t, t + dt)$ can be expressed as:

$$\lambda(t)dt = \mathbb{P}\{\text{event in } [t, t + dt) \mid S_t\} \ . \tag{1}$$

Besides, the conditional probability that an event does not happen since last observed time t_n is

$$U(t) = \exp\left(-\int_{t_n}^{t} \lambda(\tau)d\tau\right) \ . \tag{2}$$

Combining these two components, we have the conditional density function $f(t|S_t) = \lambda(t)U(t)$ that an event happens exactly at time t given all previous consumptions S_t [1].

The conditional intensity function $\lambda(t)$ takes various functional forms. A constant intensity leads to homogeneous Poisson Process, while non-homogeneous Poisson Process has time-varying intensity function but the events are still independent. Hawkes Process models excitations between events whose intensity function takes the form of:

$$\lambda(t) = \lambda_0 + \alpha \sum_{t_j<t} \gamma(t - t_j) \ , \tag{3}$$

where λ_0 represents the base intensity and every history event has an addictive self-exciting effect to current intensity, which varies with time and temporal characteristics are controlled by triggering kernel γ. α represents the degree of excitation.

4.2 Model Definition

Inspired by the intensity function in Hawkes Process (Eq. 3), we integrate collaborative filtering in base intensity and address both short-term and life-time effects in kernel function, leading to an intensity function in the following form:

$$\lambda^{u,i}(t) = \overbrace{\lambda_0^{u,i}}^{base} + \overbrace{\alpha_i \sum_{(t',i')\in S_t^u} I(i' = i)\gamma_i(t - t')}^{self-excitation} \ , \tag{4}$$

where the total intensity can be seen as a summation of base intensity $\lambda_0^{u,i}$ and the self-excitation[4] of all previous consumptions for item i.

4.2.1 Base intensity. Base intensity $\lambda_0^{u,i}$ models inherent personalized preference for different items. To accurately model the basic preference, we leverage collaborative filtering methods to derive the base intensity, such as Bayesian Personalized Ranking (BPRMF) and Neural Collaborative Filtering (NCF). Collaborative filtering methods model user's preference based on what she has consumed in the past and assume similar users like similar items, which helps to explore novel items. In the case of BPRMF, there is a K-dimensional latent factor for each user and item, p_u and q_i respectively. Then base intensity

$$\lambda_0^{u,i} = p_u^T q_i + b \ . \tag{5}$$

In the case of NCF, the base intensity is derived by a multi-layer neural network, which can be formulated as

$$\lambda_0^{u,i} = \phi_{out}(\phi_X(\cdots\phi_2(\phi_1(p_u^T, q_i^T))\cdots)) \ , \tag{6}$$

where ϕ_{out} and ϕ_x respectively denote the mapping function for the output layer and x-th neural collaborative filtering layer, and there are X neural CF layers in total.

We will show it is easy for SLRC model to leverage many state-of-the-art algorithms with joint parameters learning.

4.2.2 Self-excitation. In self-excitation part, $I(c)$ is an indicator function and returns 1 when c is true, otherwise 0. The triggering kernel $\gamma_i(\cdot)$ controls the temporal characteristics of self-excitation and varies with inter-consumption time interval $\Delta t = t - t'$. To explicitly address the two factors, we resolve $\gamma_i(\cdot)$ as a mixture distribution modeling short-term effect with Exponential distribution and life-time effect with Gaussian mixture distribution:

$$\gamma_i(\Delta t) = \overbrace{\pi_0^i E(\Delta t|\ 1/\beta^i)}^{short-term} + \overbrace{\sum_{z\in[1,Z]} \pi_z^i N(\Delta t|\ \mu_z^i, \sigma_z^i)}^{life-time} \ . \tag{7}$$

$E(x|\lambda)$ represents Exponential distribution with parameter λ, while $N(x|\mu, \sigma)$ is a Gaussian distribution with mean μ and standard deviation σ. The coefficient π_z^i satisfy the limit that $\sum_{z\in[0,Z]} \pi_z^i = 1$. From Eq. 7, we can see there are two factors in self-excitation:

Short-term effect. As shown in Section 3.3.1, short-term effect is really common and important for modeling repeat consumption. From Figure 2, the short-term effect diminishes quickly as time goes by, in which case an item-specific Exponential distribution is suitable for modeling it.

Life-time effect. Life-time effect also greatly influences user's repurchasing tendency, which often causes a raise of tendency after specific time intervals (e.g. 50 days after purchasing baby care cream when it is time to run out). As a result, it's natural to choose Gaussian mixture distribution to model life-time effect, for sometimes item's lifetime can be multimodal. But in our experiments,

[4] Here we only consider self-excitation (influence by the same item) because we mainly focus on repeat consumption. We leave the mutual-excitation caused by other items as future work.

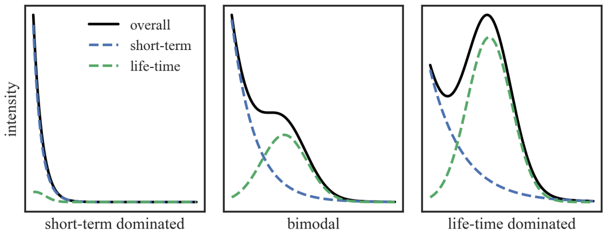

Figure 4: Illustration on the joint impact of various temporal dynamics on time-varying trends of overall intensity, charactered by the triggering kernel.

we find a single Gaussian distribution is strong enough, with its mean standing for item's expected lifetime. Therefore we choose to set $Z = 1$ without loss of generality. If items' lifetimes demonstrate more complex patterns in datasets, larger Z can be chosen through cross-validation. When $Z = 1$, the triggering kernel $\gamma_i(\cdot)$ can be simplified as:

$$\gamma_i(\Delta t) = (1 - \pi^i)E(\Delta t|\ 1/\beta^i) + \pi^i N(\Delta t|\ \mu^i, \sigma^i),\ \pi^i \in [0, 1]\ . \quad (8)$$

All the parameters related to self-excitation $\Theta_s = \{\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\}$ are item-specific. We make the degree of excitation α^i a summation of global degree and item bias: $\alpha^i = \alpha + \alpha_b^i$, in which case global α can capture average degree of excitation. Figure 4 conceptually visualize how we separately model two temporal dynamics in self-excitation part, which shows different time-varying trends of overall intensity (black) can be represented as the summation of short-term effect (blue) and life-time effect (green).

As intensity $\lambda^{u,i}(t)$ indicates the tendency of consumption, the recommendation list can be derived by the followings: we calculate the intensity value of each candidate item (including novel items and consumed items) for user u at time t firstly. Then, all candidate items are ranked according to their intensity value. The top-k items in the ranked list will be recommended to users.

Different from previous work [5, 8, 31, 32, 43], we focus to build a holistic model that can recommend both novel items and consumed items. The integration of collaborative filtering methods helps to better balance recommendation of the two kinds of items. Besides, we address two temporal dynamics of repeat consumption with specifically designed item-specific triggering kernel, which is concise and powerful with high interpretability. There may be other methods to address these two factors, but what we want to do is preliminarily exploring how to incorporate item-specific temporal dynamics of repeat consumption into recommender systems. Therefore we leave other forms of intensity function as future work.

4.3 Parameter Learning

As all candidate items are ranked according to intensity value $\lambda^{u,i}(t)$, we can adopt pairwise ranking loss for optimization. For each consumption (u, i_n, t_n), a negative item i_n^- is randomly sampled from items user haven't consumed, i.e. $\mathcal{I} \backslash I_u$ (I_u is a set of unique items consumed by user u). The pairwise ranking loss is defined as follows:

$$\mathcal{L}(\Theta|\mathcal{S}) = -\sum_{n\in\mathcal{U}}\sum_{n=1}^{N_u}\log\sigma\left(\lambda^{u,i_n}(t_n) - \lambda^{u,i_n^-}(t_n)\right) + \Omega(\Theta_b)\ . \quad (9)$$

Remind that the base intensity λ_0^{u,i_n^u} can be derived by different methods. We denote Θ_b as parameters related to calculating base intensity (e.g. $\Theta_b = \{\boldsymbol{p}, \boldsymbol{q}, b\}$ for BPRMF). Θ_b may have various forms when our model leverages different algorithms. We add a $\mathcal{L}_2$ regularizer $\Omega(\Theta_b)$ to prevent the base intensity overfitting data. On the other hand, parameters related to self-excitation can be denoted as Θ_s ($\Theta = \Theta_b \cup \Theta_s$). Θ_b and Θ_s are learned together with the unified loss (Eq. 9).

Due to the success of Adam algorithm [22] in many recommendation models' parameter learning procedures, we use Adam as the learning algorithm to minimize the pairwise ranking loss. Note that such optimization method makes it easy and efficient for joint learning of parameters when leveraging different algorithms to derive base intensity.

5 EXPERIMENTS

5.1 Experimental Settings

5.1.1 Datasets and evaluation protocols. We use the same datasets described in Section 3.2, including consumptions in various scenarios. For each consumption sequence $S_u \in \mathcal{S}$, we leave the interactions at the latest time as test dataset and the interactions at the secondly latest time as validation dataset. All the remaining data are used for training. The repeat consumption ratio of each test dataset is similar with overall repeat ratio (as shown in Table 1). And for each test or validation case, we recommend top-k ($k = 5, 10$) items from a given candidate set containing ground-truth items. Considering it is too time-consuming to rank all items for some methods when dataset is large, we randomly sample a certain number of negative items besides ground-truth items to construct candidate set. The negative items are randomly sampled from all the items except for ground truth, which means the candidate set may contain both items that user has never consumed and previously consumed items (but not bought this time, which are also negative examples for this consumption). The candidate sets stay the same for all models' training and testing. As for the size of candidate set, we make it relevant to the number of unique items in the dataset (500, 5000, 2500, 5000 for *Baby*, *Order*, *Recsys2017*, *BrightKite* respectively).

To evaluate the quality of recommendation, we use Recall and *Normalized Discounted Cumulative Gain* (NDCG) as evaluation metrics. Recall measures how many ground-truth items are recommended, while NDCG concerns about whether ground-truth items are ranked higher than others by accounting for the position of hit. We calculate both metrics for each user in test or validation dataset and report the average score.

5.1.2 Baseline Methods. We compare SLRC to nine baselines in different aspects. The first three baselines are collaborative filtering models and do not take repeat consumption into consideration:

- **BPRMF**. This method apply a pairwise ranking loss to optimize Matrix Factorization (MF) model, which is competitive for item recommendation [36].

Table 2: Experimental results (higher is better). Bold face indicates the best result of each column in a particular metric. Best baseline in each metric is underlined. The relative improvement SLRC gains compared to originals is listed within parentheses. Value** is significantly better than the strongest baseline ($p < 0.01$).

Method	*Baby*				*Order*				*Recsys2017*				*BrightKite*			
	topk=5		topk=10		topk=5		topk=10		topk=5		topk=10		topk=5		topk=10	
	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG
BPRMF	0.3686	0.2786	0.4496	0.3058	0.2307	0.2168	0.2796	0.2371	0.5754	0.5191	0.5879	0.5231	0.7609	0.7143	0.7872	0.7229
Tensor	0.4587	0.3543	0.5708	0.3923	0.2214	0.1971	0.2874	0.2240	0.5561	0.4655	0.5758	0.4719	0.7816	0.7280	0.8106	0.7375
NCF	0.3600	0.2717	0.4450	0.3002	0.2241	0.1987	0.2834	0.2231	0.5926	0.5322	0.6080	0.5372	0.7431	0.6771	0.7663	0.6847
FPMC	0.3835	0.3034	0.4655	0.3311	0.2346	0.2188	0.2897	0.2414	0.6180	0.5838	0.6324	0.5885	0.7369	0.6927	0.7651	0.7018
UtilitySVD	0.3431	0.2677	0.4197	0.2933	0.2282	0.2214	0.2734	0.2401	0.5730	0.5053	0.5850	0.5092	0.7453	0.7151	0.7610	0.7202
HCM	0.3703	0.2736	0.4824	0.3118	0.1445	0.1345	0.1863	0.1518	0.5633	0.4277	0.6326	0.4505	0.8019	0.7540	0.8276	0.7590
TSR-Hawkes	0.3797	0.2998	0.4576	0.3260	0.2565	0.2544	0.2907	0.2689	0.6629	0.5888	0.6658	0.5898	0.7611	0.7447	0.7622	0.7451
Neu-Hawkes	0.4203	0.3298	0.5088	0.3550	0.2645	0.2577	0.3087	0.2804	0.7078	0.6356	0.7280	0.6405	0.7801	0.7511	0.8006	0.7520
LS-Hawkes	0.4304	0.3457	0.5356	0.3786	0.2604	0.2554	0.2978	0.2723	0.7103	0.6389	0.7292	0.6409	0.7803	0.7513	0.8098	0.7587
$\text{SLRC}_{\text{BPRMF}}$	0.4014 (+8.9%)	0.3217 (+15.5%)	0.4833 (+7.5%)	0.3492 (+14.2%)	**0.2790**** (+20.9%)	**0.2723**** (+25.6%)	**0.3301**** (+18.0%)	**0.2935**** (+23.8%)	0.7058 (+22.7%)	0.6337 (+22.1%)	0.7250 (+23.3%)	0.6399 (+22.3%)	**0.8083**** (+6.2%)	**0.7778**** (+8.9%)	**0.8286**** (+5.3%)	**0.7844**** (+8.5%)
$\text{SLRC}_{\text{Tensor}}$	**0.4894**** (+6.7%)	**0.3919**** (+10.6%)	**0.5991**** (+5.0%)	**0.4290**** (+9.3%)	0.2622 (+18.4%)	0.2572 (+30.5%)	0.3057 (+6.4%)	0.2752 (+22.9%)	0.6650 (+19.6%)	0.6034 (+29.6%)	0.6724 (+16.8%)	0.6058 (+28.4%)	0.7923 (+1.4%)	0.7658 (+5.2%)	0.8119 (+0.2%)	0.7722 (+4.7%)
SLRC_{NCF}	0.3924 (+9.0%)	0.3124 (+15.0%)	0.4709 (+5.8%)	0.3387 (+12.8%)	0.2751 (+22.8%)	0.2683 (+35.0%)	0.3219 (+13.6%)	0.2876 (+28.9%)	**0.7207**** (+21.6%)	**0.6450**** (+21.2%)	**0.7364**** (+21.1%)	**0.6501**** (+21.0%)	0.8044 (+8.3%)	0.7768 (+14.7%)	0.8219 (+7.3%)	0.7824 (+14.3%)

- **Tensor**. Tensor factorization is widely used in context-aware recommendation [21]. Here we split the continuous time into bins and handle a three dimensional tensor (user-item-time).
- **NCF**. This is a state-of-the-art collaborative filtering method that utilizes neural networks to capture complex relationship between users and items in latent space [18].

Another three baselines incorporate repeat consumption information and can make holistic item recommendation:

- **FPMC**. This method combines MF and factorized Markov Chains with a user-specific transition matrix, which utilizes the sequence information of each user [37].
- **UtilitySVD**. This method utilizes marginal utility to model repeat consumption, where users' preference for items will change according to consumed quantity [42].
- **HCM**. It is a holistic generative model for sequential repeated consumption, which takes time, quality and distance factor into consideration. We use the last two parts of the model to conduct top-k recommendation [4].

The last three are related to Hawkes Process with similar task:

- **TSR-Hawkes**. This model aims at Time-Sensitive Recommendation with low-rank Hawkes Process, in which core parameters are assumed to have low-rank structure. The triggering kernel decays globally with time [16].
- **Neu-Hawkes**. Mei et al. [34] propose a neural Hawkes process model that uses recurrent layer to derive intensity according to events history. A novel continuous-time LSTM architecture is utilized where each memory cell exponentially decays toward some steady-state value.
- **LS-Hawkes**. The model captures long- and short-term dependency between actions with two different Hawkes Process, while both of the triggering kernels simply decay in a global way [8].

The first three baselines will further be leveraged to derive the base intensity $\lambda_0^{\boldsymbol{u},\boldsymbol{i}}$ in our SLRC model. We denote the unified model as $\text{SLRC}_{\text{method}}$, such as $\text{SLRC}_{\text{BPRMF}}$.

As for other work about repeat consumption [3, 5, 11, 12] referred to in Section 2, they are used either to classify repeat consumption or to predict only items consumed before. Neither of them focuses to recommend both consumed items and novel items simultaneously. So they cannot be taken as baseline methods.

5.1.3 Implement details. We implement SLRC model in *TensorFlow* and the code has been open source[5]. We tune hyperparameters based on NDCG@10 in the validation dataset and report the results in the test dataset. All models are trained with Adam optimizer until converge with a maximum of 200 epochs. Each experiment is repeated 5 times with different random seed and the average result is reported.

For fair comparison, the batch size is fixed to 128 and the latent dimension for all models is 100. As for parameter initialization, global excitation α and parameters in mixture distribution $\boldsymbol{\beta}, \boldsymbol{\sigma}$ are initialized with 1; item bias of excitation degree α_b^i is all initialized with 0. $\boldsymbol{\pi}$ is normally initialized with 0.5 mean; $\boldsymbol{\mu}$ is also normally initialized and the mean is global average inter-consumption time interval. Other parameters are normally initialized with 0 mean. All normal initializers have 0.01 standard deviation.

5.2 Performance

Table 2 shows the performance of all baselines and the improved methods when incorporating SLRC model.

First, the performances of different baselines vary across datasets. Tensor, Neu-Hawkes, LS-Hawkes, HCM performs the best among all baselines in *Baby*, *Order*, *Recsys2017*, *BrightKite*, respectively, which shows capacities of different models in various scenarios. Tensor performs quite well in *Baby*, which may be the result of long time span and time sensitivity in this dataset. As state-of-the-art models based on Hawkes Process, Neu-Hawkes and LS-Hawkes performs robustly in all the datasets, and become the most competitive baselines in *Order* and *Recsys2017*. As for HCM, it achieves the best performance among baselines in *Brightkite* because of its high repeat consumption ratio (90%). Explicitly predicting whether current consumption will be repeat consumption in HCM greatly narrow the range of candidate items to consider. But in other datasets with

[5]https://github.com/THUwangcy/SLRC

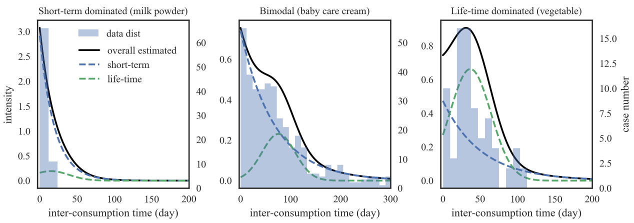

Figure 5: Some examples to validate the parametric modeling assumptions. For different types of items, overall intensity (black) closely approximates real distributions of inter-consumption gap (bar). Moreover, Exponential (blue) and Gaussian (green) distribution fit short-term and life-time effects well, respectively.

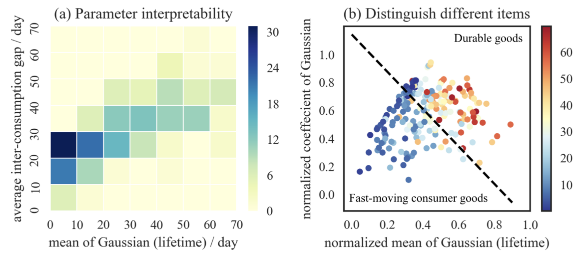

Figure 6: Correlation between item's inter-consumption gap and the learned lifetime parameter μ^i in Gaussian distribution (left), where color represents the number of items. And joint distribution of parameter μ^i and π^i (right), where red and blue stand for longer and shorter inter-consumption gap in average, respectively.

lower repeat consumption ratio (e.g. *Order*), due to the lack of collaborative filtering information, mainly modeling repeat consumption results in bad performance.

Second, although collaborative filtering and Hawkes Process methods gain encouraging results in different datasets, SLRC performs consistently better than other baselines in all datasets (the best combination with collaborative filtering method may be different in distinct datasets). This shows the proposed model is capable of utilizing both the strengths of collaborative filtering and temporal dynamics of repeat consumption in Hawkes Process, leading to great overall performance improvement consequently. Aside from the performance improvement over models about repeat consumption (i.e. FPMC, UtilitySVD, HCM), SLRC outperforms methods based on Hawkes Process without concern about item-specific temporal dynamics of repeat consumption (i.e. TSR-Hawkes, Neu-Hawkes and LS-Hawkes), which indicates the importance of modeling both short-term and life-time effects in kernel function. The integration of collaborative filtering methods also plays an important role, which estimates users' inherent preference for items more accurately and helps to better recommend novel items that meet users' interests.

Third, the experimental results demonstrate that the incorporation of SLRC leads to a big gap of improvement (up to 35.0%) compared to the original collaborative filtering baselines. This shows the information of repeat consumption and its item-specific temporal dynamics are critical to recommender systems. Notice that the improvement for traditional method BPRMF is extremely large in most datasets. SLRC introduces history sequence information and temporal factors into the model, so BPRMF may attentively model inherent preference more accurately. For other baselines that take more information into consideration themselves, SLRC also brings consistent performance gain on all datasets. Besides, the encouraging results on all the datasets in different scenarios show the scalability and flexibility of our model.

5.3 Parametric Assumptions Validating

Here we want to validate whether the learned parameters have interpretable meaning as we assume when designing the model.

First, we study how well the learned intensity function fits real data. Figure 5 gives some examples from different datasets. We draw both the distribution of inter-consumption gap in real data and how estimated intensity drifting with time[6]. The results show that the overall intensity (black) closely approximates real distribution of inter-consumption gap. Moreover, the *short-term effect* (blue) and *life-time effect* (green) capture the temporal characteristics of repeat consumption well, respectively. Take vegetable as an example (the right figure), we find there are numerous repurchasing behavior for vegetables around 30 days after previous consumption in Section 3.3. As expected, the Gaussian distribution for vegetable learned in SLRC shows around 30 days mean ($\mu^i = 36$ days) and the importance of *life-time effect* is high ($\pi^i = 0.65$), indicating *life-time effect* is in dominated position.

Next, remind that the mean (μ^i) and the coefficient (π^i) of life-time effect in Eq. 7 are item-specific. μ^i can be seen as item's expected lifetime and π^i is the importance of life-time effect. Figure 6 (a)[7] shows the correlation between item's inter-consumption gap in average and corresponding parameter μ^i. Parameters are trained in *Baby* dataset with SLRC$_{\text{BPRMF}}$ model. Other combinations in other datasets yield similar results. Note that the information of inter-consumption gap is not introduced in SLRC explicitly, but μ^i is reasonably in direct proportion to average inter-consumption gap (Pearson Correlation Coefficient: 0.7468), showing that items that are often consumed again after a long period of time (long lifetime) get larger μ^i just as assumed.

Further, we show the joint distribution of μ^i and π^i in Figure 6 (b), which demonstrates that items with higher average inter-consumption gap (red in color) almost have larger μ^i and π^i. The dashed line in figure obviously divides items into two kinds: fast-moving consumer goods (below) and durable goods (above). Some representative items above the dashed line are toys, baby walkers, beds and so on, while fast- moving consumer goods like paper diaper and milk powder are mostly under the dashed line. Another convincing and interesting observation is that most feeding bottles appear above the dashed line while nipples only appear below the dashed line, for the reason that users often change nipples but seldom change feeding bottles.

[6]Here we assume previous consumption happens at the origin of coordinates and show the temporal trend of self-excitation.

[7]We discard items that are never repeatedly consumed. And items with extremely high or low number of average inter-consumption gap are also removed.

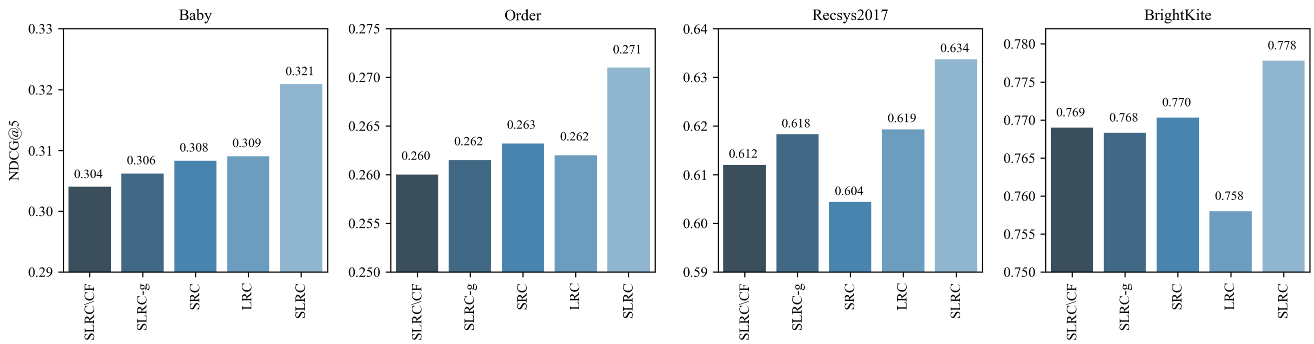

Figure 7: Performance comparison between SLRC and methods with no consideration for CF or variety of temporal dynamics. SLRC performs the best again various variations. All improvements are statically significant ($p < 0.05$).

Table 3: Difference comparison between variants of SLRC.

Variants	CF	Item-specific	Life-time	Short-term
SLRC\CF		✓	✓	✓
SLRC-g	✓		✓	✓
SRC	✓	✓		✓
LRC	✓	✓	✓	
SLRC	✓	✓	✓	✓

To sum up, SLRC gains remarkable performance improvement and the estimated parameters can further explain various temporal patterns of repeat consumption across different items perfectly.

6 FURTHER ANALYSIS

6.1 Ablation Study

To verify the effect of CF part and the two item-specific temporal dynamics (short-term and life-time effects) addressed in the model, we compare SLRC to a series of baselines without consideration for variety of temporal characteristics of repeat consumption.

- **SLRC\CF**. This model removes the CF part of SLRC and uses an item-specific parameter to estimate the base intensity.
- **SLRC-g**. This model does not concern the variety across different items. Parameters related to self-excitation are not item-specific (parameters in $\Theta_s = \{\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\}$ are global values).
- **SRC**. This model only consider short-term effect without life-time effect, where the triggering kernel in intensity function is an item-specific Exponential distribution.
- **LRC**. Similar with SRC, LRC only takes life-time effect into consideration, where an item-specific Gaussian distribution acts as the triggering kernel in intensity function.

Table 3 intuitively shows the difference between SLRC and its variants. Combinations with different baselines yield similar results, so here we only show results when combined with BPRMF (i.e. base intensity is calculated by BPRMF for SRC, LRC, SLRC-g). The results shown in Figure 7 indicate that although variants of SLRC achieve consistently better performance than BPRMF, none of them can outperform the full SLRC model.

Without the integration of CF, SLRC\CF leads to loss of performance, which demonstrates that the CF part of SLRC plays an important role for holistic recommendation task. As for SLRC-g, it neglects inherent properties (like lifetime) of distinct items. The short-term and life-time effects estimated over all items can be inaccurate when evaluating some specific items. Therefore the difference in temporal dynamics across items are critical to model repeat consumption behavior.

On the other hand, SRC and LRC do not take both temporal factors into consideration and hence get worse results, which demonstrates that the lack of variety of temporal factors can lead to lower predictive ability (e.g. continuous recommendation for mobiles shortly after purchasing one and mobiles could not be recommended when it is the time to change a new one). It is noteworthy that the performances of SRC and LRC vary across datasets. The results are similar in *Baby* and *Order* for the reason that both short-term and life-time effects are common in item purchasing scenario. But LRC is better than SRC in *Recsys2017*, which may be the result of obvious life-time effect in this dataset as shown in Section 3.3. The short-term effect is weaker compared with other datasets because users seldom click a job posting again in a short time. While in *BrightKite*, SRC outperforms LRC greatly. It is also reasonable because users often revisit hotels they appreciate recently but seldom go to the one visited long time ago. The short-term effect is much stronger than life-time effect in such scenario.

Further, dynamically modeling short-term and life-time effects for different items, our SLRC model is able to utilize strengths of CF and capture complex patterns of repeat consumption, achieving the best performance in all the datasets.

6.2 Discussion on Exploration and Exploitation

Besides the great performance improvement, we want to figure out how well our model balances the recommendation of novel items and consumed items. The self-excitation addressed in the model should not result in recommending more consumed items under all circumstances. Therefore, we construct a series of subsets (the same size) from the test dataset with different ratios of repeat consumption to simulate exploration dominated and exploitation dominated scenarios. This can be done by randomly sampling a certain ratio (namely, repeat ratio r) of cases from repeat consumption set and filling the rest with cases randomly sampled from the other part

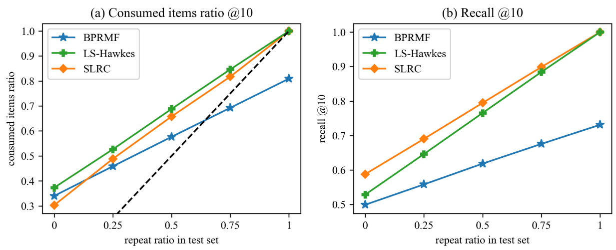

Figure 8: Results on test subsets with different ratios of repeat consumption case. In the left figure, the dashed line shows expected trend of consumed items ratio. SLRC shows a better balance of exploration and exploitation, and gains better result all the time.

containing no repeat consumptions. When repeat ratio r is small, exploration cases are in dominant position (if $r = 0$, all cases in the subset are consumptions for novel items), otherwise exploitation cases dominate the subset.

Figure 8 shows consumed items ratio and Recall@10 on subsets with different repeat ratio r. Consumed items ratio means the proportion of test cases whose corresponding recommendation list contains consumed items. Ideally, model should recommend less consumed items when repeat ratio r is small (i.e. users tend to seek novel items) but to retrieve more consumed items when r is large (i.e. users want to purchase items consumed before). In the left figure, the dashed line shows expected trend of consumed items ratio. We conduct top-10 recommendation in *Recsys2017* dataset and use BPRMF as the base method. Several observations from these results are worth highlighting:

- All the models recommend more consumed items when repeat ratio gets larger (i.e. users express more needs for consumed items) and yield better performance.
- SLRC recommends less items when repeat ratio is small, while recommends more consumed items otherwise, leading to a better balance of exploration and exploitation compared with BPRMF and LS-Hawkes.
- SLRC achieves better results under all circumstances. The improvement gap compared to BPRMF gets larger as repeat ratio increases. And only when all test cases are repeat consumption, LS-Hawkes can achieve similar performance with SLRC.

The main conclusion is that addressing both short-term and life-time effects helps SLRC to well balance exploration and exploitation. SLRC will not recommend extra consumed items when users do not need. While in the cases when users express demands for consumed items, our model properly recommends more consumed items. Notice that the performance of LS-Hawkes gets closer to SLRC's when repeat ratio increases, the reason may be that LS-Hawkes wrongly gives higher scores to some consumed items without considering various temporal dynamics and CF information, so novel items cannot be recommended for some exploration cases. As a result, only when there are all repeat consumption cases, LS-Hawkes can achieve similar performance with SLRC.

In summary, the results demonstrate that SLRC clearly captures users' need for consumed items, achieving a good balance of exploration and exploitation. Furthermore, in real-world scenarios where repeat consumption takes a moderate part, SLRC model performs significantly better than other methods.

7 CONCLUSION AND FUTURE WORK

Repeat consumption comes to be increasingly important recently. It is necessary to model repeat consumption behavior so that recommender system can recommend items user consumed before at proper time. In this work, we focus to build a holistic model that can recommend novel items and consumed items simultaneously. Two factors are revealed for modeling various time-varying patterns of repeat consumption: *short-term effect* and *life-time effect*. To address this variety, we propose a novel model, **S**hort-Term and **L**ife-Time **R**epeat **C**onsumption (SLRC), which explicitly takes the two factors into consideration with combination of two techniques: Collaborative Filtering (CF) and Hawkes Process. According to extensive experiments, significant performance improvement is achieved in four real-world datasets compared to state-of-the-art baselines. Besides, with high interpretability, SLRC is able to discover distinct items' lifetimes. And different types of items such as fast-moving consumer goods and durable goods can even be distinguished by the learned parameters, showing the model can perfectly captures various item-specific dynamics of repeat consumption. Generally, our model is able to be leveraged as a framework that utilize repeat consumption information to enhance existing recommender systems based on collaborative filtering and hence achieve better recommendation results in real-work applications.

In the future, we will turn to category level and consider mutual-excitation of similar items. Intuitively, a previous consumption not only has influence on the item itself, but also influences purchase tendency of similar items. The item-specific modeling method in SLRC may suffer from data sparsity greatly. Further, simultaneous consumption has some inherent relationship with repeat consumption. How to combine these two together is an interesting but seldom explored problem.

ACKNOWLEDGMENTS

This work is supported by Natural Science Foundation of China (Grant No. 61672311, 61532011) and The National Key Research and Development Program of China (2018YFC0831900). We would like to thank Maarten de Rijke and Diane Kelly for their valuable comments of the work.

REFERENCES

[1] Odd O. Aalen, ÃŸrnulf Borgan, and HÃěkon K. Gjessing. 2008. *Survival and Event History Analysis*. Springer New York. 457–491 pages.
[2] Eytan Adar, Jaime Teevan, and Susan T. Dumais. 2008. Large scale analysis of web revisitation patterns. In *Sigchi Conference on Human Factors in Computing Systems*. 1197–1206.
[3] Ashton Anderson, Ravi Kumar, Andrew Tomkins, and Sergei Vassilvitskii. 2014. The dynamics of repeat consumption. In *Proceedings of the 23rd international conference on World wide web*. ACM, 419–430.
[4] Austin R. Benson, Ravi Kumar, and Andrew Tomkins. 2016. Modeling User Consumption Sequences. In *International Conference on World Wide Web*. 519–529.
[5] Rahul Bhagat, Srevatsan Muralidharan, Alex Lobzhanidze, and Shankar Vishwanath. 2018. Buy It Again: Modeling Repeat Purchase Recommendations. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 62–70.
[6] Preeti Bhargava, Thomas Phan, Jiayu Zhou, and Juhan Lee. 2015. Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor

factorization on sparse user-generated data. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 130–140.

[7] Stanley L Brue, Campbell R Mcconnell, and McGrawHill. 2011. Economics: principles, problems and policies. *Economics Principles Problems and Policies* (2011).

[8] Renqin Cai, Xueying Bai, Zhenrui Wang, Yuling Shi, Parikshit Sondhi, and Hongning Wang. 2018. Modeling Sequential Online Interactive Behaviors with Temporal Point Process. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 873–882.

[9] Lara D Catledge and James E Pitkow. 1995. Characterizing Browsing Strategies in the World-Wide Web. In *International World Wide Web Conference*. 1065âĂŞ1073.

[10] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural Attentional Rating Regression with Review-level Explanations. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. 1583–1592.

[11] Jun Chen, Chaokun Wang, and Jianmin Wang. 2015. Will You" Reconsume" the Near Past? Fast Prediction on Short-Term Reconsumption Behaviors.. In *AAAI*. 23–29.

[12] Jun Chen, Chaokun Wang, Jianmin Wang, and S Yu Philip. 2016. Recommendation for repeat consumption from user implicit feedback. *IEEE Transactions on Knowledge and Data Engineering* 28, 11 (2016), 3083–3097.

[13] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility:user movement in location-based social networks. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Ca, Usa, August*. 1082–1090.

[14] David Roxbee Cox and Valerie Isham. 1980. Point processes. *Monographs on Statistics and Applied Probability* 65, 432 (1980), 47–98.

[15] 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.

[16] Nan Du, Yichen Wang, Niao He, and Le Song. 2015. Time-sensitive recommendation from recurrent user activities. In *International Conference on Neural Information Processing Systems*. 3492–3500.

[17] J. Durbin and G. S. Watson. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.

[18] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 173–182.

[19] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. Ieee, 263–272.

[20] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul Schrater. 2015. Just in time recommendations: Modeling the dynamics of boredom in activity streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 233–242.

[21] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. 2010. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 79–86.

[22] Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *Computer Science* (2014).

[23] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 426–434.

[24] Yehuda Koren and Robert Bell. 2015. Advances in collaborative filtering. In *Recommender systems handbook*. Springer, 77–118.

[25] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.

[26] Takeshi Kurashima, Tim Althoff, and Jure Leskovec. 2018. Modeling Interdependent and Periodic Real-World Action Sequences. *arXiv preprint arXiv:1802.09148* (2018).

[27] Guokun Lai, Wei Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. (2018).

[28] Moshe Lichman and Padhraic Smyth. 2018. Prediction of Sparse User-Item Consumption Rates with Zero-Inflated Poisson Regression. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 719–728.

[29] Jie Liu, Chun Yu, Wenchang Xu, and Yuanchun Shi. 2012. Clustering web pages to facilitate revisitation on mobile devices. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. 249–252.

[30] Xinyue Liu, Yuanfang Song, Charu Aggarwal, Yao Zhang, and Xiangnan Kong. 2017. BiCycle: Item Recommendation with Life Cycles. In *IEEE International Conference on Data Mining*. 297–306.

[31] Dixin Luo, Hongteng Xu, Yi Zhen, Xia Ning, Hongyuan Zha, Xiaokang Yang, and Wenjun Zhang. 2015. Multi-task multi-dimensional hawkes processes for modeling event sequences. In *International Conference on Artificial Intelligence*. 3685–3691.

[32] Emaad Manzoor and Leman Akoglu. 2017. RUSH!: Targeted Time-limited Coupons via Purchase Forecasts. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1923–1931.

[33] D Marsan and O LenglinÃĺ. 2008. Extending earthquakes' reach through cascading. *Science* 319, 5866 (2008), 1076–9.

[34] Hongyuan Mei and Jason M Eisner. 2017. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*. 6754–6764.

[35] Andriy Mnih and Ruslan R Salakhutdinov. 2008. Probabilistic matrix factorization. In *Advances in neural information processing systems*. 1257–1264.

[36] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 452–461.

[37] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web*. ACM, 811–820.

[38] Mingxuan Sun, Mingxuan Sun, Tao Ye, and Tao Ye. 2014. A hazard based approach to user return time prediction. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1719–1728.

[39] Yusuke Tanaka, Takeshi Kurashima, Yasuhiro Fujiwara, Tomoharu Iwata, and Hiroshi Sawada. 2016. Inferring Latent Triggers of Purchases with Consideration of Social Effects and Media Advertisements. In *ACM International Conference on Web Search and Data Mining*. 543–552.

[40] Jaime Teevan, Eytan Adar, Rosie Jones, and Michael Potts. 2006. History repeats itself: repeat queries in Yahoo's logs. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 703–704.

[41] Sarah K. Tyler and Jaime Teevan. 2009. Large scale query log analysis of re-finding. In *ACM International Conference on Web Search and Data Mining*. 191–200.

[42] Jian Wang and Yi Zhang. 2011. Utilizing marginal net utility for recommendation in e-commerce. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 1003–1012.

[43] Shuai Xiao, Junchi Yan, Stephen M. Chu, Xiaokang Yang, and Hongyuan Zha. 2017. Modeling The Intensity Function Of Point Process Via Recurrent Neural Networks. (2017).

[44] Shuai Xiao, Junchi Yan, Changsheng Li, Bo Jin, Xiangfeng Wang, Xiaokang Yang, Stephen M Chu, and Hongyuan Zha. 2016. On Modeling and Predicting Individual Paper Citation Count over Time.. In *IJCAI*. 2676–2682.

[45] Haimo Zhang and Shengdong Zhao. 2011. Measuring web page revisitation in tabbed browsing. In *Sigchi Conference on Human Factors in Computing Systems*. 1831–1834.

[46] Ke Zhou, Hongyuan Zha, and Le Song. 2013. Learning triggering kernels for multi-dimensional hawkes processes. In *International Conference on International Conference on Machine Learning*. III–1301.