PAPER *Special Section on Data Engineering and Information Management*

Analyzing Temporal Dynamics of Consumer's Behavior Based on Hierarchical Time-Rescaling

Hideaki KIM[†a)], **Noriko TAKAYA**[†b)], ***Nonmembers*, *and* Hiroshi SAWADA**[††c)], ***Member***

SUMMARY Improvements in information technology have made it easier for industry to communicate with their customers, raising hopes for a scheme that can estimate when customers will want to make purchases. Although a number of models have been developed to estimate the time-varying purchase probability, they are based on very restrictive assumptions such as preceding purchase-event dependence and discrete-time effect of covariates. Our preliminary analysis of real-world data finds that these assumptions are invalid: self-exciting behavior, as well as marketing stimulus and preceding purchase dependence, should be examined as possible factors influencing purchase probability. In this paper, by employing the novel idea of hierarchical time rescaling, we propose a tractable but highly flexible model that can meld various types of intrinsic history dependency and marketing stimuli in a continuous-time setting. By employing the proposed model, which incorporates the three factors, we analyze actual data, and show that our model has the ability to precisely track the temporal dynamics of purchase probability at the level of individuals. It enables us to take effective marketing actions such as advertising and recommendations on timely and individual bases, leading to the construction of a profitable relationship with each customer.

key words: *self-exciting process, time-rescaling theorem, e-commerce, purchase behavior*

1. Introduction

With the explosive growth of e-commerce systems, industrial practitioners now have the potential to communicate with their customers wherever and whenever they want. This situation puts strong pressure on the practitioners to find the right time to communicate with each of their customers, that is, the time when she is most likely to make a purchase [1]–[3]. The purpose of this paper is to propose a feasible method for tracking the fluctuation of the underlying purchase probability of a customer, based on her transaction data.

The negative binomial distribution (NBD) model [4], which has been applied extensively in marketing studies, is a standard model for estimating the purchase probability of a customer. The model, however, assumes that the purchase probability is stationary over time, and is independent of all non-stationary marketing variables and the customer's purchase history. Thus the NBD model says nothing about when to communicate with a customer. Some extensions of the NBD model have been proposed, although to maintain model tractability, they are based on unrealistic assumptions such that the purchase behavior only depends on the last purchase decision, and that the effect of marketing variables is piecewise constant over time [5]–[12]. Our analysis finds that these assumptions are violated in actual data. To achieve a more realistic model, we have to overcome the trade-off between model flexibility and tractability, which is not possible with the conventional approach.

In this paper, by employing the novel idea of hierarchical time rescaling, we achieve a tractable but highly flexible model for estimating customers' purchase probability over time*. Due to its flexibility, the model is able to incorporate the effects of time-varying marketing variables and various types of history dependency in a continuous-time setting. Because of its tractability, the model offers the real-time tracking of each customer's purchase probability, as well as the efficient optimization of the model parameters. We call the proposed model, the *Hierarchical Time-Rescaling* model (HTRm).

We assess the potential of HTRm for capturing customers' purchase behaviors by applying it to real-world data consisting of four categories of transaction data. HTRm is introduced in general form, and we propose a specific form that is suitable for analyzing transaction data. It incorporates three possible rate fluctuation factors: seasonal sales, self-excitation, and the preceding purchase-events. We compare the predictive performance of the proposed model against the results achieved by the NBD model and other models with fewer factors, and show that HTRm's inclusion of various factors is essential for precisely estimating customer purchase dynamics.

The rest of the paper is organized as follows. In Sect. 2, we describe the problem of estimating a customer's purchase behavior via a point process, and outlines related work. In Sect. 3, we construct HTRm in a form specifically designed to analyze purchase data. In Sect. 4, we develop a feasible algorithm, based on the Monte Carlo EM algorithm, for estimating the model parameters. Section 5 applies HTRm to real-world data, and shows that the model has the ability to precisely estimate the underlying dynamics of customer purchase dynamics. Finally, Sect. 6 provides

Manuscript received June 21, 2016.
Manuscript publicized October 13, 2016.

[†]The authors are with NTT Service Evolution Laboratories, NTT Corporation, Yokosuka-shi, 239–0847 Japan.
[††]The author is with NTT Communication Science Laboratories, NTT Corporation, Kyoto-fu, 619–0237 Japan.
a) E-mail: kin.hideaki@lab.ntt.co.jp
b) E-mail: takaya.noriko@lab.ntt.co.jp
c) E-mail: sawada.hiroshi@lab.ntt.co.jp

DOI: 10.1587/transinf.2016DAP0010

*Earlier versions of this work were presented at the 23rd ACM International Conference on Information and Knowledge Management (CIKM '14) [13] as a conference paper.

our conclusions.

2. Preliminary

2.1 Estimating Purchase Behavior via Point Process

We first introduce a basic method for analyzing purchase events occurring in time. The theory of *point process* [14], which has been introduced in not only marketing science [4] but also such diverse disciplines as neuroscience [15] and seismology [16], provides a powerful tool for modeling and analyzing the stochastic structure of point events occurring in continuous time. In the point process framework, the purchase behavior of a customer is characterized by the *purchase rate*, that is, the instantaneous probability for a customer to make a purchase decision at each point in time [4], [7], [17]. Suppose that customer u $(1 \le u \le U)$ makes a sequence of purchase decisions, $\{t^u_j\}^n_{j=0} \equiv (t^u_0, t^u_1, \cdots, t^u_n)$, in the half-open observation period $(t^u_0, T]$. We can estimate u's purchase rate, $\lambda^u(t)$, by evaluating the probability density of the observation occurring, given by

$$
\begin{aligned}
&p(\{t^u_j\}^n_{j=0}|\lambda^u(t)) \\
&= \Big[\prod_{j=1}^{n} \lambda^u(t^u_j)\Big] \exp\Big(-\sum_{j=0}^{n-1}\int_{t^u_j}^{t^u_{j+1}} \lambda^u(t)dt - \int_{t^u_n}^{T} \lambda^u(t)dt\Big) \\
&= \Big[\prod_{j=1}^{n} \lambda^u(t^u_j)\Big] \exp\Big(-\int_{t^u_0}^{T} \lambda^u(t)dt\Big), \qquad (1)
\end{aligned}
$$

where the exponential on the right-hand side of Eq. (1), called the *survivor function* [14], represents the probability of no purchase events occurring in the open intervals (t^u_j, t^u_{j+1}) for $0 \le j \le n-1$ and the half-open interval $(t^u_n, T]$. If the purchase rate is estimated as $\hat{\lambda}^u(t)$ based on Eq. (1), we can predict future purchase times, $\{t^u_j\}_{j>n}$, again based on Eq. (1), as $p(\{t^u_j\}_{j>n}|\hat{\lambda}^u(t))$. For illustration, see Fig. 1. The notation is summarized in Table 1. Note that we take the initial purchase point t^u_0 as the start of the observation, because industrial practitioners usually start monitoring customer's transactions from the first occurrence. In that case, the initial purchase t^u_0 is not regarded as being generated from $\lambda^u(t)$, and thus the probability density (1) does not have factor $\lambda^u(t^u_0)$.

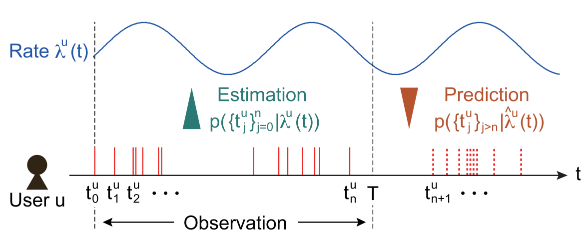

Fig. 1 A schema for the estimation and prediction procedure via point process. Solid bars represent observed purchase decisions, and dashed bars represent a most likely purchase sequence in the future.

Table 1 Symbols and definitions for data.

Symbol	Definition
U	# of customers
u	uth user, $1 \le u \le U$
t^u_j	jth purchase point of user u, $j \ge 0$
T	end point of observation
t^w_i	start point of ith winter sale
t^s_i	start point of ith summer sale

The difficulty in performing the estimation and the prediction procedures comes from the costly integral, $\int \lambda^u(t)dt$, in the survivor function. Therefore, to make the point process model tractable, a simple purchase rate $\lambda^u(t)$ should be designed so that the survivor function is obtained analytically. To pursue the dynamical purchase behaviors of customers, which may modulate intricately depending on the history of purchase decisions and/or time-varying marketing variables, we need to make $\lambda^u(t)$ as flexible as possible. To address the trade-off between model tractability and its flexibility, several purchase rate models have been proposed, as briefly outlined in the following. For simplicity, we omit the user index of u when discussing a model on an individual basis.

2.2 Previous Work

The most standard model for analyzing customers' purchase behavior is the NBD model [4]. In it, purchase rate $\lambda(t)$ is assumed to be stationary over time, $\lambda(t) = \lambda$, leading to the analytical survivor function, $\exp(-\lambda(T - t_0))$. This model is based on a time-homogeneous Poisson process [14], and thus its inter-event intervals, $t_j - t_{j-1}$, should obey an exponential distribution, and there should be no correlation between successive inter-event intervals.

Inclusion of History Dependency: To relax the limitations, the NBD model was extended by the renewal point process, which allows the purchase rate to be dependent on the last purchase decision [5]–[8], [11]. Thus the purchase rate $\lambda(t)$ can be expressed by using a cumulative distribution of the inter-event intervals, $F(t_j - t_{j-1})$, as $\lambda(t) = -\frac{d}{dt}\log[1 - F(t - t_j)]$, for $t_j < t < t_{j+1}$, leading to the analytical survivor function, $(1 - F(T - t_n))\prod_{j=1}^{n}(1 - F(t_j - t_{j-1}))$. The inter-event intervals, $t_j - t_{j-1}$, may obey a non-exponential distribution, $dF(t_j - t_{j-1})/dt$, but there should be no correlation between successive inter-event intervals.

Inclusion of Covariate: It was also proposed to incorporate time-varying covariates such as seasonality and price promotion into the renewal model, in which the survivor function can be obtained analytically, as long as the piecewise constancy of covariates is assumed [7]–[9], [12].

These NBD-based models marginally improve the explanatory power of the original NBD model, although they are inadequate for reproducing real purchase behavior. The assumption of piecewise constancy in covariates is violated when their modulation frequency is higher than the purchase rate, and the history dependency of purchase behavior is not necessarily very simple, both of which are found to be the case in the transaction data we analyzed (see Sect. 5).

2.3 Time-Rescaling Theorem

In preparation for our model construction, we here provide a description of the time-rescaling theorem. The theorem has been applied in such fields as neuroscience [15] and seismology [16], mainly for evaluating the goodness-of-fit of point process models to data. In this paper, we utilize the idea of time-rescaling for constructing our flexible and tractable model (Sect. 3). The time-rescaling theorem states that any point process may be transformed into a Poisson process with a unit rate, which can be easily demonstrated as follows. With rescaled time $\Lambda(t)$, defined by

$$\Lambda(t) \equiv \int_0^t \lambda(t')dt', \quad \Lambda_j \equiv \Lambda(t_j), \tag{2}$$

the probability density (1) at real time t is transformed into that at rescaled time $\Lambda(t)$ as,

$$\begin{aligned} p(\{\Lambda_j\}_{j=0}^n) &= p(\{t_j\}_{j=1}^n|\lambda(t)) \prod_{j=0}^n \Big[\frac{d}{dt}\Lambda(t_j)\Big]^{-1} \\ &= \Big[\prod_{j=1}^n 1\Big] \exp\Big(-1 \cdot \Lambda(T)\Big), \end{aligned} \tag{3}$$

which is identical to the probability density that the customer makes a sequence of purchase decisions, $\{\Lambda_j\}_{j=0}^n$, in the period $[0, \Lambda(T)]$, at the unit purchase rate. With this theorem, we can assess goodness-of-fit of the model by evaluating how well the statistics of the rescaled time sequence $\{\Lambda_j\}_{j=0}^n$ agree with those of a Poisson process with a unit rate.

The time-rescaling theorem implies that the difficulty of calculating a survivor function is equivalent to that of performing a time-rescaling procedure (2), which is a key point of our model construction.

3. Model

In this section, we construct a striking model that overcomes the limitations of the conventional NBD-based models, namely, the renewal property and the piecewise constancy, at the same time. We achieve complicated time rescaling by performing simple time-rescaling operations hierarchically. In each operation, time is rescaled by each of the factors that induce the purchase rate to fluctuate, such as marketing stimulus and intrinsic history dependency one by one, and finally the incorporation of various rate fluctuation factors is achieved. To the best of our knowledge, this is the first proposal of the concept and the first formalization of hierarchical time rescaling for capturing customers' purchase behavior.

3.1 Hierarchical Time-Rescaling Model

To allow the time-rescaling (2) to be performed analytically, we construct the purchase rate $\lambda(t)$ in the following two steps: (i) First, instead of converting t into $\Lambda(t)$ at one time, we consider converting t in a hierarchical manner as,

$$\Lambda^0(t) \equiv t, \quad \Lambda^{m+1}(t) \equiv \int_0^{\Lambda^m(t)} \lambda^m(\Lambda^m)d\Lambda^m, \tag{4}$$

for $0 \leq m \leq M-1$, where $\lambda^m(\Lambda^m(t))$ is a factor that rescales the mth rescaled time $\Lambda^m(t)$, and M denotes the highest level of the hierarchy. Here, we assume that each factor $\lambda^m(\Lambda^m)$ can be integrated analytically. (ii) Next, we construct purchase rate $\lambda(t)$ as a unit in the Mth rescaled time, $\Lambda^M(t)$, leading to the factorized rate,

$$\lambda(t) = \prod_{m=0}^{M-1} \lambda^m(\Lambda^m(t)). \tag{5}$$

For calculation details, see Appendix A.

Figure 2 illustrates how the proposed model performs time-rescaling hierarchically (5). In each time-rescaling step, a new time Λ^{m+1} is defined by rescaling the current time Λ^m with the time-inhomogeneous factor $\lambda^m(\Lambda^m)$, according to which the strongly fluctuating rate (the blue curve in the top) drops its fluctuation factors, λ^m, one by one in descending order, before eventually being converted into a unit (the blue curve at the bottom). The scales of the rescaled times, represented by the lower black bars, show that the time with a higher rate is elongated, and vice versa (dashed line), thus compensating for the fluctuation of the rate. It is worth noting that the sequence of purchase decisions generated from $\lambda(t)$ (red bars) seems to be uncorrelated or be aligned in a Poisson manner in rescaled time $\Lambda^M(t)$. We call the proposed model the *Hierarchical Time-Rescaling* model (HTRm).

The key advantage of HTRm is that under the weak assumption that each factor $\lambda^m(\Lambda^m)$ has an analytical integral, we can compute the rescaled time $\Lambda^M(t)$, and thus obtain the probability density (1) analytically. We place no further restriction on the functional form of the individual factors $\lambda^m(\Lambda^m)$. Thus, by the incorporation of multiple rate fluctuation factors, HTRm is able to reproduce a rich variety of purchase behaviors, while maintaining excellent tractability.

3.2 Realization of HTRm for Analyzing Purchase Data

In Sect. 5, our analysis of real-world transaction data produced a finding that contradicted the time-homogeneous Poisson assumption (see Table 3): one is the non-exponential distribution of inter-event intervals (IEIs), the other is the positive correlation between successive IEIs. Both could be explained by *self-excitation* and/or renewal behavior of a customer, although a time-varying marketing stimulus might cause the purchase rate to fluctuate, leading the customer to behave in a non-Poisson manner.

Thus, we provide HTRm with three possible rate fluctuation factors, namely, sale, self-excitation, and preceding purchase dependency: The sale factor is modeled by the sequence of alpha-functions, in which each of the sale events triggers a rapid excitation of the purchase rate followed by a slow decay; The self-excitation factor is expressed by the Hawkes process with exponential decay [18]–[20], where a

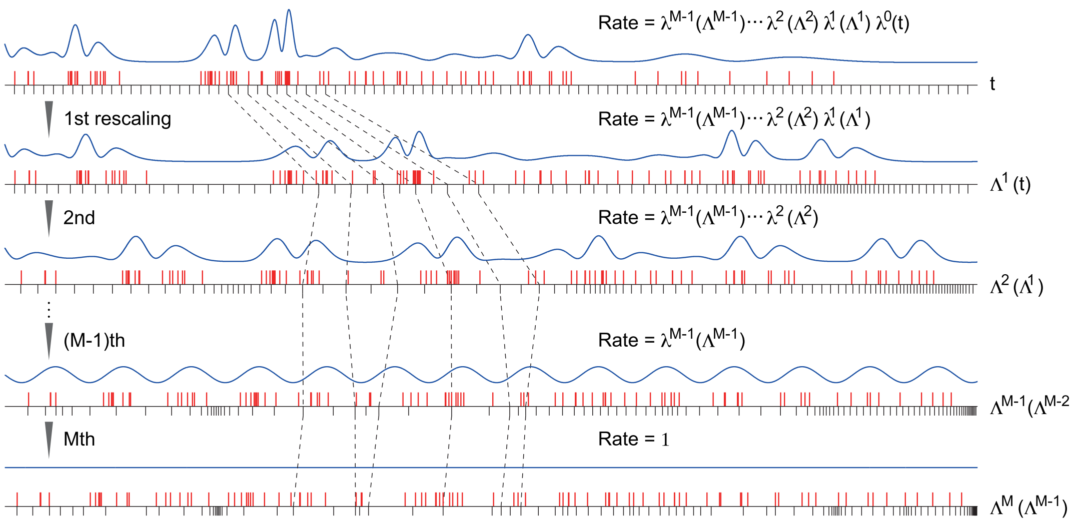

Fig. 2 Hierarchical time-rescaling model with M layers. For an explanation, see Sect. 3.1.

purchase event gives an additional rise in the purchase rate, which provokes further purchase events; The factor of the preceding purchase dependency can be modeled by using the Weibull hazard function [8]. Denoting the factors as λ^{sale}, λ^{excite}, and λ^{pre}, respectively, we propose the following purchase model for a customer:

$$\lambda(t) = \lambda^{\text{sale}}(t) \cdot \lambda^{\text{excite}}(\Lambda(t)) \cdot \lambda^{\text{pre}}(\Omega(t)), \tag{6}$$

$$p(\{t_j\}_{j=0}^{n}|\lambda(t)) = \Big[\prod_{j=1}^{n} \lambda(t_j)\Big] \exp(-\Psi(T)), \tag{7}$$

where the corresponding rescaled times are denoted by

$$\begin{aligned} \Lambda(t) &\equiv \int_{t_0}^{t} \lambda^{\text{sale}}(t')dt', \\ \Omega(t) &\equiv \int_{0}^{\Lambda(t)} \lambda^{\text{excite}}(\Lambda')d\Lambda', \\ \Psi(t) &\equiv \int_{0}^{\Omega(t)} \lambda^{\text{pre}}(\Omega')d\Omega', \end{aligned} \tag{8}$$

respectively.

Let N $(t > t_0)$ and $N_{w(s)}$ $(t > t_0)$ be the right-continuous counting process of the purchase events and the winter (summer) sale events, respectively: the sample paths of the counting processes jump 1 immediately following the associated event points, and are constant otherwise [14]. Thus first we model the sale factor $\lambda^{\text{sale}}(t)$ by using alpha functions as

$$\begin{aligned} \lambda^{\text{sale}}(t) = r_0\Big(1 &+ a_s \sum_{i=1}^{N_s(t)} (t - t_i^s) \exp[-b_s(t - t_i^s)] \\ &+ a_w \sum_{i=1}^{N_w(t)} (t - t_i^w) \exp[-b_w(t - t_i^w)]\Big), \end{aligned} \tag{9}$$

where r_0 represents the base rate, and $t_i^{w(s)}$, $a_{w(s)}$ and $b_{w(s)}^{-1}$ represent the start point, the impact, and the timescale of the ith winter (summer) sale, respectively. Second, by employing the Hawkes process with exponential decay [18]–[20], we model the self-exciting factor $\lambda^{\text{excite}}(t)$ as

$$\lambda^{\text{excite}}(\Lambda(t)) = 1 + \sum_{j=1}^{N(t)} a_h \exp[-b_h(\Lambda(t) - \Lambda(t_j))], \tag{10}$$

where a_h and b_h^{-1} are the impact and the timescale of a purchase event, respectively. Finally, we model the factor of the preceding purchase dependency, $\lambda^{\text{pre}}(t)$, by using the Weibull hazard function,

$$\lambda^{\text{pre}}(\Omega(t)) = \frac{f_\kappa(\Omega(t) - \Omega(t_{N(t)}))}{1 - F_\kappa(\Omega(t) - \Omega(t_{N(t)}))}, \tag{11}$$

where $f_\kappa(x) \equiv \kappa x^{\kappa-1} \exp(-x^\kappa)$ and $F_\kappa(x) \equiv 1 - \exp(-x^\kappa)$ are the Weibull distribution and its cumulative distribution with shape parameter κ, respectively. Substituting Eqs. (9)–(11) into Eq. (8), we obtain the corresponding rescaled times in the following analytical forms:

$$\begin{aligned} \Lambda(t) &= r_0(t - t_0) \\ &+ \frac{r_0 a_s}{b_s} \sum_{i=1}^{N_s(t)} \Big\{ e^{-b_s(0 \vee (t_0 - t_i^s))}[0 \vee (t_0 - t_i^s) + b_s^{-1}] \\ &\qquad - e^{-b_s(t - t_i^s)}[(t - t_i^s) + b_s^{-1}] \Big\} \\ &+ \frac{r_0 a_w}{b_w} \sum_{i=1}^{N_w(t)} \Big\{ e^{-b_w(0 \vee (t_0 - t_i^w))}[0 \vee (t_0 - t_i^w) + b_w^{-1}] \\ &\qquad - e^{-b_w(t - t_i^w)}[(t - t_i^w) + b_w^{-1}] \Big\}, \\ \Omega(t) &= \Lambda(t) + \sum_{j=1}^{N(t)} \frac{a_h}{b_h}\Big(1 - \exp[-b_h(\Lambda(t) - \Lambda(t_j))]\Big), \\ \Psi(t) &= \sum_{j=1}^{N(t)-1} (\Omega(t_{j+1}) - \Omega(t_j))^\kappa + (\Omega(t) - \Omega(t_{N(t)}))^\kappa, \end{aligned} \tag{12}$$

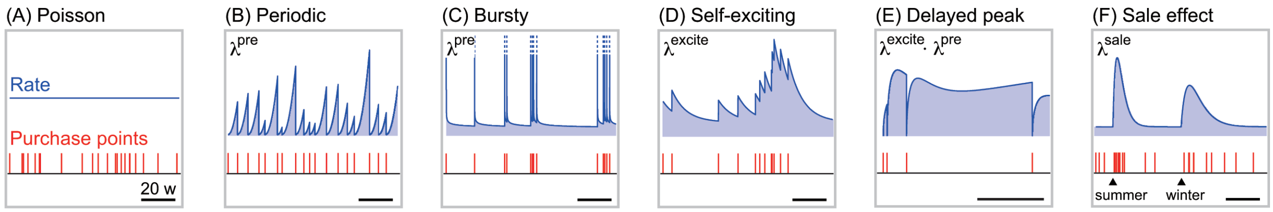

Fig. 3 Representative purchase patterns reproduced by the HTR model. The blue curve represents purchase rate, and the red bars represent sample purchase points generated from the rate. (A)–(E) The sets of parameters (a_h, b_h, κ) used for reproducing the history-dependent purchase patterns are $(0, *, 1)$ for A, $(0, *, 3)$ for B, $(0, *, 0.7)$ for C, $(1.5, 2, 1)$ for D, and $(1.8, 4, 1.5)$ for E. Pattern E is displayed on a larger scale for better visualization. In each case, the effect of sales is ignored ($a_s = a_w = 0$). (F) A purchase pattern with sales effect, $(a_s, a_w, b_s, b_w) = (10, 3, 0.4, 0.2)$. Each inverted triangle represents the start point of a sale. In (A)–(F), the base rate r_0 is adjusted so that the mean purchase rate is around 0.2 [1/week].

where $x \vee y = \max\{x, y\}$.

Let a set of individual parameters be denoted as $\theta \equiv (a_s, a_w, b_s, b_w, r_0, a_h, b_h, \kappa)$. Then, for the proposed model (6)–(12), we can evaluate the logarithm of the probability density of the data (7) efficiently by using $\lambda^{\text{sale}}(t_j)$, $\Lambda(t_j)$ and $\Lambda(T)$ as follows:

$$\begin{aligned}
&\log p(\{t_j\}_{j=0}^{n}|\theta) \\
&= \sum_{j=1}^{n} \log \lambda^{\text{sale}}(t_j) + \sum_{j=1}^{n} \log \lambda^{\text{excite}}(\Lambda(t_j)) \\
&+ \sum_{j=1}^{n-1} \log f_\kappa(\Omega(t_{j+1}) - \Omega(t_j)) + \log[1 - F_\kappa(\Omega(t_1))] \\
&+ \log[1 - F_\kappa(\Omega(T) - \Omega(t_n))],
\end{aligned} \tag{13}$$

where the first term is calculated from Eq. (9), the second term is calculated as

$$\begin{aligned}
&\sum_{j=1}^{n} \log \lambda^{\text{excite}}(\Lambda(t_j)) = \sum_{j=1}^{n} \log(1 + a_h A_j), \\
&A_1 = 0,\ A_{j+1} = (1 + A_j) e^{-b_h[\Lambda(t_{j+1}) - \Lambda(t_j)]},
\end{aligned} \tag{14}$$

the third term is calculated as

$$\begin{aligned}
&\sum_{j=1}^{n-1} \log f_\kappa(\Omega(t_{j+1}) - \Omega(t_j)) \\
&= (n-1)\log\kappa + (\kappa - 1)\sum_{j=1}^{n-1} \log(\Delta\Omega_j) - \sum_{j=1}^{n-1} (\Delta\Omega_j)^\kappa, \\
&\Delta\Omega_j = \Lambda(t_{j+1}) - \Lambda(t_j) + \frac{a_h}{b_h}(1 + A_j - A_{j+1}),
\end{aligned} \tag{15}$$

and the remaining terms (the logarithms of the left- and right-censored survivor functions) are given by

$$-(\Omega(t_1))^\kappa - (\Omega(T) - \Omega(t_n))^\kappa, \tag{16}$$

where

$$\begin{aligned}
&\Omega(t_1) = \Lambda(t_1), \\
&\Omega(T) - \Omega(t_n) = \Lambda(T) - \Lambda(t_n) \\
&\qquad + \frac{a_h}{b_h}(1 + A_n)\left(1 - e^{-b_h[\Lambda(T) - \Lambda(t_n)]}\right).
\end{aligned} \tag{17}$$

The proposed model (6)–(12) has the ability to mimic a variety of history-dependent purchase behaviors depending on the set of three parameters, (a_h, b_h, κ), which is summarized in Figs. 3 A–E. In practice, the effect of sales, which is shown in Fig. 3 F, is multiplied into each of the behaviors. In Appendix B, we provide an efficient procedure for generating sample purchase points from the proposed model.

3.3 Heterogeneity Across Customers

We developed a purchase model (6)–(12) on an individual basis, in which the eight parameters, $a_s, a_w, b_s, b_w, r_0, a_h, b_h$, and κ, differ among customers. Letting the set of individual parameters for customer u ($1 \le u \le U$) be denoted by

$$\theta^u \equiv (a_s^u, a_w^u, b_s^u, b_w^u, r_0^u, a_h^u, b_h^u, \kappa^u), \tag{18}$$

we assume that each of the parameters is generated from a gamma prior distribution:

$$p(\theta_l^u|\mu_{\theta_l}, \nu_{\theta_l}) = \frac{1}{\Gamma(\nu_{\theta_l})} \frac{\nu_{\theta_l}}{\mu_{\theta_l}} \left(\frac{\nu_{\theta_l}\theta_l^u}{\mu_{\theta_l}}\right)^{\nu_{\theta_l}-1} \exp\left(-\frac{\nu_{\theta_l}\theta_l^u}{\mu_{\theta_l}}\right), \quad \text{for } 1 \le l \le 8, \tag{19}$$

where μ_{θ_l} and ν_{θ_l} are the scale and shape parameter of the gamma distribution, respectively. The scale parameter μ_{θ_l} represents the mean of θ_l across all customers, and the shape parameter ν_{θ_l}, which determines the coefficient of variation of the distribution as $\nu_{\theta_l}^{-1/2}$, is an index of the homogeneity in θ_l across all customers. Then, the prior distribution of θ^u is expressed as $p(\theta^u|\mu, \nu) = \prod_{l=1}^{8} p(\theta_l^u|\mu_{\theta_l}, \nu_{\theta_l})$, where

$$\begin{aligned}
&\mu \equiv (\mu_{a_s}, \mu_{a_w}, \mu_{b_s}, \mu_{b_w}, \mu_{r_0}, \mu_{a_h}, \mu_{b_h}, \mu_\kappa), \\
&\nu \equiv (\nu_{a_s}, \nu_{a_w}, \nu_{b_s}, \nu_{b_w}, \nu_{r_0}, \nu_{a_h}, \nu_{b_h}, \nu_\kappa).
\end{aligned} \tag{20}$$

The notation is summarized in Table 2. The NBD model [4]

Table 2 Symbols and definitions for model.

Symbol	Definition
r_0^u	base rate of customer u
a_h^u	impact of a purchase decision on customer u's purchase rate
b_h^u	inverse of timescale of a purchase decision effect on customer u's purchase rate
κ^u	shape parameter of Weibull hazard function of customer u
$a_{w(s)}^u$	impact of a winter (summer) sale on customer u's purchase rate
$b_{w(s)}^u$	inverse of timescale of a winter (summer) sale effect on customer u's purchase rate
μ_{θ_l}	mean of individual parameter θ_l across all customers, $\theta_l \in a_{w(s)}, b_{w(s)}, r_0, a_h, b_h, \kappa$
ν_{θ_l}	index of homogeneity in individual parameter θ_l across customers

is a special case of the proposed model (6)–(12), (18)–(19) for $\mu_{a_h} = \mu_{a_{s(w)}} = 0$, $\mu_\kappa = 1$, and $\nu_{a_h} = \nu_{a_{s(w)}} = \nu_\kappa = \infty$.

4. Estimation Procedure

4.1 Determining Aggregate Parameters

Based on the empirical Bayes method, the sets of aggregate parameters, μ and ν, can be determined by maximizing the marginal likelihood,

$$p(D|\mu,\nu) = \int \prod_{u=1}^{U} p(\{t_j^u\}|\theta^u) p(\theta^u|\mu,\nu) d\theta^u, \tag{21}$$

where $\{t_j^u\}$ and D denote customer u's purchase data and all of the customers' data, respectively. Given a set of data D, the maximization is performed based on the Monte Carlo expectation maximization algorithm [21]: aggregate parameters are determined by iteratively maximizing the expected value of the complete-data log-likelihood, the Q function,

$$\begin{aligned} &Q(\mu,\nu|\mu^{(p)},\nu^{(p)}) \\ &= E\Big[\sum_{u=1}^{U} \log\big(p(\{t_j^u\}|\theta^u) p(\theta^u|\mu,\nu)\big)\Big| D, \mu^{(p)}, \nu^{(p)}\Big], \end{aligned} \tag{22}$$

where $\mu^{(p)}$ and $\nu^{(p)}$ are the aggregate parameters of the pth iteration, and $E[\,\cdot\,|\{t_j^u\},\mu^{(p)},\nu^{(p)}]$ represents the expectation with respect to the posterior distribution of θ^u under the pth estimate of the aggregate parameters,

$$E[\,\cdot\,|\{t_j^u\},\mu^{(p)},\nu^{(p)}] \equiv \int \cdot\; p(\theta^u|\{t_j^u\},\mu^{(p)},\nu^{(p)}) d\theta^u. \tag{23}$$

To perform the expectation, we employ the following Monte Carlo sum,

$$E[g(\theta^u)|\{t_j^u\},\mu^{(p)},\nu^{(p)}] \approx \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} g(\theta_i^u), \tag{24}$$

where $g(\theta^u)$ is a function of θ^u, N_{MC} is the Monte Carlo sample size, and θ_i^u is the ith sample of θ^u generated from its posterior distribution. The sampling is performed by a variant of the Markov Chain Monte Carlo (MCMC) method, or the Metropolis–Hastings algorithm [22], in which we chose a lognormal proposal distribution for θ^u. The $(p+1)$th estimate of (μ,ν) is then determined by the conditions for $dQ/d\mu = dQ/d\nu = 0$, leading to the following update rule:

$$\mu^{(p+1)} = \frac{1}{U} \sum_{u=1}^{U} E[\theta^u|\{t_j^u\},\mu^{(p)},\nu^{(p)}], \tag{25}$$

$$\begin{aligned} &\log(\nu^{(p+1)}) - \psi(\nu^{(p+1)}) \\ &\quad = \log(\mu^{(p+1)}) - \frac{1}{U}\sum_{u=1}^{U} E[\log(\theta^u)|\{t_j^u\},\mu^{(p)},\nu^{(p)}], \end{aligned} \tag{26}$$

where $\psi(\nu)$ is the digamma function. Because $\log(\nu) - \psi(\nu)$ is a monotonically decreasing function, we can obtain the root of Eq. (26) with the bisection method [23].

4.2 Estimating Individual Parameters

Given the aggregate parameters to be determined as $(\hat{\mu},\hat{\nu})$, we can obtain the posterior mean estimate of the individual parameters θ^u as

$$\hat{\theta}^u = E[\theta^u|\{t_j^u\},\hat{\mu},\hat{\nu}], \qquad 1 \le u \le U. \tag{27}$$

The expectation is performed by the Monte Carlo sum (24). The scalability for the estimation algorithm is discussed in Appendix C.

5. Experiments

We examine the potential efficiency of the HTR model to capture customers' purchase behaviors by applying it to e-commerce and panel data sets.

5.1 About Data Used

The e-commerce data set contains transactions about fashion items conducted over 230 weeks (October 1, 2009 ∼ February 27, 2014) as performed at a commercial website, where the first 200 weeks of data are used for model fitting, and the remaining 30 weeks serve to evaluate the model's predictive performance. We denote the data set by Fashion. We also consider its subset consisting of the loyal customers, those who made at least 20 transactions during first 200 weeks. We denote the subset by $\mathrm{Fashion_L}$. The timestamp of the winter and summer sales is given as: $t_1^s = 37$, $t_2^s = 88$, $t_3^s = 141$, $t_3^s = 193$, $t_1^w = 62$, $t_2^w = 114$, $t_3^w = 167$, $t_3^w = 218$ [week].

The panel data consists of three categories of transaction data sets, namely, milk, coffee drink and water†. We denote the data sets by Milk, Coffee and Water, respectively. The both contain transactions gathered over 50 weeks

†Individual consumer panel research data (SCI) collected by INTAGE Inc. (Tokyo, Japan).

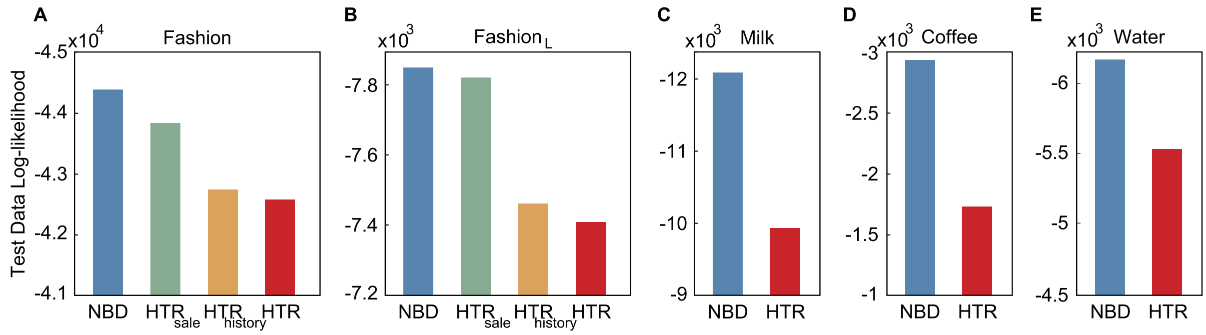

Fig. 4 Comparison of HTR with other models as regards predictive performance. (A–B) Log-likelihoods of test data achieved by NBD, HTR_{sale} ($\lambda = \lambda^{\text{sale}}$), $\text{HTR}_{\text{history}}$ ($\lambda = \lambda^{\text{excite}} \cdot \lambda^{\text{pre}}$), and HTR ($\lambda = \lambda^{\text{sale}} \cdot \lambda^{\text{excite}} \cdot \lambda^{\text{pre}}$). (C–E) Log-likelihoods of test data achieved by NBD and HTR ($\lambda = \lambda^{\text{excite}} \cdot \lambda^{\text{pre}}$).

Table 3 Data statistics.

Data set	Fashion	$\text{Fashion}_{\text{L}}$	Milk	Coffee	Water
# of customers	7,129	563	3,131	2,240	1,689
# of trans.	83,129	20,928	131,961	111,235	56,845
period [week]	230	230	50	50	50
mean IEI [week]	11.3	4.5	1.0	0.8	1.5
cv of IEI	1.37	1.39	1.50	2.26	1.71
correlation	**0.25	**0.23	**0.40	**0.25	**0.28

IEI: inter-event interval, cv: coefficient of variation

(January 1, 2013 ~ December 17, 2013), where the first 45 weeks of data are used for model fitting, and the remaining 5 weeks for model evaluation. We have no information about sales. For all five date sets, we omit customers who made less than 5 transactions during the fitting period. The purchase time point is measured in hours. The data statistics are summarized in Table 3.

As mentioned in Sect. 3.2, Table 3 shows that the Poisson behavior, assumed in the NBD model, is violated in all data sets: (i) the coefficient of variation for IEIs largely deviates from unity, indicating a non-exponential distribution of IEIs [24]; (ii) there is a significant ($p < 0.01$) correlation between successive IEIs. The non-exponential distribution and the positive correlation indicate that the customers makes purchase decisions depending on the last or more previous decisions, but the seasonal sale could also cause the customers' non-Poissonian behaviors. Using our HTRm, we elucidate the details of the history dependency and the degree of contribution of the seasonal sale on such non-Poissonian behaviors.

5.2 Predictive Performance

Based on the e-commerce data, we compare HTRm's predictive performance against the results achieved by the NBD model (NBD: $a_s = a_h = 0, \kappa = 1$), HTRm with only the sale effect (HTR_{sale}: $a_h = 0, \kappa = 1$), and HTRm with only the history dependency ($\text{HTR}_{\text{history}}$: $a_w = a_s = 0$). We denote the HTRm with both the sale effect and the history dependency by HTR. Based on the panel data, on the other hand, we compare the predictive performance of HTRm against that of the NBD model. Because we have no information about sales for the panel data, we denote HTRm developed from only the history dependency by HTR.

In accordance with the procedure described in Sect. 4, we estimated the aggregate and individual parameters of each model based on the training data. Using the estimated individual parameters $\hat{\theta}^u$, we evaluated the predictive performance of each model based on the log-likelihood for the test data,

$$LL \equiv \sum_{u=1}^{U} \log p(\{t_j^u \in D^{\text{test}}\}|\hat{\theta}^u), \tag{28}$$

where D^{test} represents the purchase timestamps in the test data. Figure 4 shows that HTR performed better than the other models for all four data sets. The comparison between Fig. 4 A and Fig. 4 B found that the effect of the seasonal sales is smaller in $\text{Fashion}_{\text{L}}$ than in Fashion, which is consistent with the fact that loyal customers are less responsive to the price promotion. The results indicate that the inclusion of various factors is essential for precisely estimating customers' purchase dynamics.

Using the time-rescaling theorem as a basis, we also evaluated the predictive performance by checking whether or not the purchase points rescaled by the estimated individual parameters, $\Psi(t_j)$, follow a Poisson process. Here, we checked the following two points: (i) whether the rescaled IEIs, $\Delta\Psi_j \equiv \Psi(t_{j+1}) - \Psi(t_j)$, follow an exponential distribution with mean of 1; (ii) whether the successive rescaled IEIs, $\Delta\Psi_j$ and $\Delta\Psi_{j+1}$, have no correlation between them. Figure 5 shows the semi-log plot of the empirical distribution of rescaled intervals $\Delta\Psi_j$ (upper figures) and the scatter diagram in the $(\log(\Delta\Psi_j), \log(\Delta\Psi_{j+1}))$ plane (lower figures). The significant correlation between successive intervals, found in the NBD model, was removed in HTRm perfectly for the Fashion and $\text{Fashion}_{\text{L}}$ data sets (Figs. 5 A and B), and partially for the Milk, Coffee and Water data sets (Figs. 5 C–E). The distribution of the rescaled intervals deviates greatly from the exponential with mean of 1 (dashed line) in the NBD model (Figs. 5 A–D), which is substantially improved in HTRm for all of the data sets except the Fashion data set. In the Fashion data set (Fig. 5 A), over half of the customers (63%) have less than 10 transactions, thus it is harder to estimate individual parameters accurately

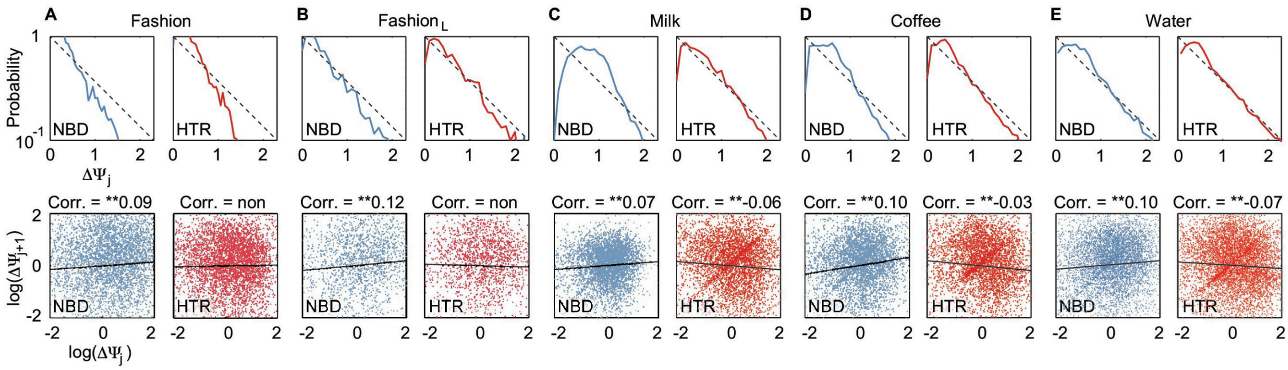

Fig. 5 (Upper figures) Empirical distributions of rescaled inter-event intervals $\Delta\Psi_j$. The empirical distributions were obtained by histogram density estimation with a bin size of 0.1. The diagonal dashed line represents an exponential distribution with unit mean. (Lower figures) Scatter diagrams of the log interval $\log(\Delta\Psi_j)$ against the successive one $\log(\Delta\Psi_{j+1})$. For better visualization, the abscissa and ordinate are the deviations of $\log(\Delta\Psi_j)$ and $\log(\Delta\Psi_{j+1})$ from their means. The coefficient of correlation (Spearman) between the successive intervals is depicted at the top of each figure, where ** and * represent the correlation significances, $p < 0.01$ and $p < 0.05$, respectively. Black lines represent the regression lines.

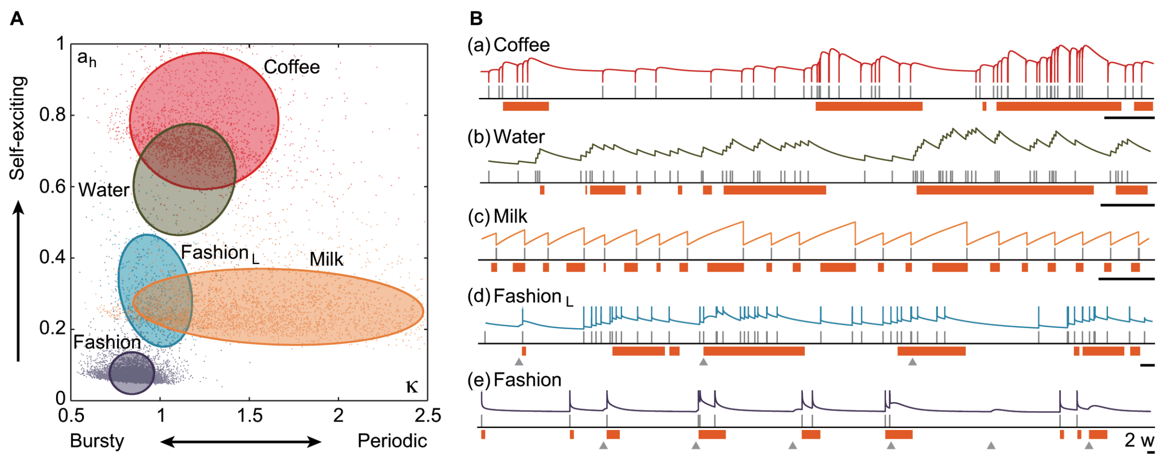

Fig. 6 (A) Distributions of estimated individual parameters for all the data sets. Ellipses represent 75% quantiles of two-dimensional Gaussian distribution fitted to the individual data sets. (B) Representative purchase rates estimated from the data sets. Shaded rectangles in each figure represent the intervals during which purchase rate is estimated online to be relatively high. Filled triangles in (d) and (e) represent seasonal sale points.

compared to in the other data sets. It should be emphasized that HTRm achieved higher predictive performance than the simpler NBD model (Fig. 4 A) even for this very sparse data set.

5.3 Estimated Dynamics of Purchase Decisions

The estimated individual parameters are summarized in Fig. 6. Figure 6 A shows that the history dependency of purchase decisions is widely distributed among item categories. The purchase rate in Fashion rises briefly after a purchase decision is made ((e) in Fig. 6 B), while due to its self-exciting dynamics, the rate in Fashion_L sometimes keeps its value high for several months ((d) in Fig. 6 B). The difference in the purchase dynamics might imply that because the prices of fashion items are relatively high, the psychological impact of a spending decision would be substantial for usual customers, and thus cancel out the intrinsic self-exciting effect seen in loyal customers. On the other hand, the purchase rate in Milk drops to zero immediately after a purchase decision occurred and the value remains low for a while, resulting in the periodic purchase pattern ((c) in Fig. 6 B). The purchase rate in Coffee also has a brief ‘refractory period’ after each purchase decision, which prevents quick successive purchases. Sometimes, however, it shows transient activation due to its strong self-exciting effect ((a) in Fig. 6 B).

Figure 6 A also shows that the temporal dynamics of purchase decisions differ largely across customers even in the same item category. Especially in the Milk category, the periodicity of purchase behavior is estimated to have strong heterogeneity. This might be caused by the fact that some customers occasionally purchase milk only when needed, while others, who drink milk every day, purchase new fresh milk before the expiration date, which regularizes the inter-purchase intervals. Regardless, the broad distributions in Fig. 6 A suggest that the temporal dynamics of customers' purchase behavior should be examined individually.

5.4 Real-Time Tracking of Purchase Rate

Given the individual parameter $\hat{\theta}^u$, the current state of the purchase rate, $\lambda^u(t_{\text{now}})$, can be computed so quickly (of the order of milliseconds) based on the individual observation $\{t_j^u\}$, because HTRm guarantees that $\lambda^u(t_{\text{now}}|\{t_j^u\})$ is expressed by analytical functions (see Sect. 3). Figure 6 B displays examples of real-time estimates of the purchase rate. In it, purchase decisions are observed frequently in the intervals during which the estimated purchase rate was high, indicated by the shaded rectangles. When the prediction of future state λ^u $(t > t_{\text{now}}|\{t_j^u\})$ is needed, we can generate sample paths of λ^u $(t > t_{\text{now}}|\{t_j^u\})$ efficiently by using the time rescaling theorem (see Appendix B). Based on the current and future state of the purchase rate, we can make the go/no-go decisions on marketing actions such as advertising and recommendation.

6. Conclusion and Future Work

In this paper, we constructed an innovative model that can track the temporal dynamics of a customer's purchasing decisions. Our model offers us a fast way of estimating the current and future state of the purchase rate, which is intricately influenced by various intrinsic and external factors. We incorporated the factors that are thought to dominate the modulation of the purchase rate, namely, seasonal sale, self-excitation, and preceding purchase-events, into the model, and confirmed that our proposed model achieved high predictive performance when challenged with real-world data in the categories of fashion, milk, and coffee beverages. This result indicates that our proposed model has the ability to discover the hidden dynamics of purchase behavior at the level of individuals, suggesting that the model will enable us to take effective marketing actions such as advertising and recommendations on timely and individual basis.

A natural extension of this study is to extend our model to incorporate the interaction among customers [12], [25] or among items [2], which is expected to improve its predictive performance. HTRm is not limited to purchase event data, but is widely applicable to any sequences of event points. Thus this model may help unveil the underlying dynamics of social interaction [26], financial markets [27], and so on.

References

[1] J. Wang, B. Sarwar, and N. Sundaresan, "Utilizing related products for post-purchase recommendation in e-commerce," RecSys '11, pp.329–332, ACM, 2011.
[2] J. Wang and Y. Zhang, "Opportunity model for e-commerce recommendation: Right product; Right time," SIGIR '13, pp.303–312, ACM, 2013.
[3] G. Zhao, M.L. Lee, W. Hsu, and W. Chen, "Increasing temporal diversity with purchase intervals," SIGIR '12, pp.165–174, ACM, 2012.
[4] A.S.C. Ehrenberg, "The pattern of consumer purchases," Applied Statistics, vol.8, no.1, pp.26–41, 1959.
[5] C. Chatfield and G.J. Goodhardt, "A consumer purchasing model with Erlang inter-purchase times," Journal of the American Statistical Association, vol.68, no.344, pp.828–835, 1973.
[6] D.G. Morrison and D.C. Schmittlein, "Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort?," Journal of Business & Economic Statistics, vol.6, no.2, pp.145–159, 1988.
[7] S. Gupta, "Stochastic models of interpurchase time with time-dependent covariates," Journal of Marketing Research, vol.28, no.1, pp.1–15, 1991.
[8] G.M. Allenby, R.P. Leone, and L. Jen, "A dynamic model of purchase timing with application to direct marketing," Journal of the American Statistical Association, vol.94, no.446, pp.365–374, 1999.
[9] C.-Y. Huang and C.-S. Lin, "Modeling the audience's banner ad exposure for internet advertising planning," Journal of Advertising, vol.35, no.2, pp.123–136, 2006.
[10] K. Kopperschmidt and W. Stute, "Purchase timing models in marketing: A review," AStA Advances in Statistical Analysis, vol.93, no.2, pp.123–149, 2009.
[11] G. Trinh, C. Rungie, M. Wright, C. Driesener, and J. Dawes, "Predicting future purchases with the Poisson log-normal model," Marketing Letters, vol.25, no.2, pp.219–234, 2014.
[12] R. Takahashi, H. Mizuta, N. Abe, R.L. Kennedy, V.J. Jeffs, R. Shah, and R.H. Crites, "Collective response spike prediction for mutually interacting consumers," ICDM, pp.727–736, IEEE, 2013.
[13] H. Kim, N. Takaya, and H. Sawada, "Tracking temporal dynamics of purchase decisions via hierarchical time-rescaling model," CIKM '14, pp.1389–1398, ACM, 2014.
[14] D.L. Snyder, Random Point Processes, Wiley, New York, 1975.
[15] E.N. Brown, R. Barbieri, V. Ventura, R.E. Kass, and L.M. Frank, "The time-rescaling theorem and its application to neural spike train data analysis," Neural Computation, vol.14, no.2, pp.325–346, 2002.
[16] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," Journal of the American Statistical Association, vol.83, no.401, pp.9–27, 1988.
[17] D.C. Schmittlein, D.G. Morrison, and R. Colombo, "Counting your customers: Who are they and what will they do next?," Management Science, vol.33, no.1, pp.1–24, 1987.
[18] C. Blundell, K.A. Heller, and J.M. Beck, "Modelling reciprocating relationships with hawkes processes," NIPS, pp.2609–2608, 2012.
[19] A.G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," Biometrika, vol.58, no.1, pp.83–90, 1971.
[20] Y. Matsubara, Y. Sakurai, B.A. Prakash, L. Li, and C. Faloutsos, "Rise and fall patterns of information diffusion: Model and implications," KDD '12, pp.6–14, ACM, 2012.
[21] G.C.G. Wei and M.A. Tanner, "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," Journal of the American Statistical Association, vol.85, no.411, pp.699–704, 1990.
[22] C. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.
[23] W.H. Press, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, 1992.
[24] S. Grün and R. Stefan, Analysis of Parallel Spike Trains, Springer, 2010.
[25] T. Iwata, A. Shah, and Z. Ghahramani, "Discovering latent influence in online social activities via shared cascade Poisson processes," KDD '13, pp.266–274, ACM, 2013.
[26] R. Crane and D. Sornette, "Robust dynamic classes revealed by measuring the response function of a social system," PNAS, vol.105, no.41, pp.15649–15653, 2008.
[27] C.G. Bowsher, "Modelling security market events in continuous

time: Intensity based, multivariate point process models," Journal of Econometrics, vol.141, no.2, pp.876–912, 2007.

Appendix A: Calculation Details in HTRm

From Eqs. (4) and (5), the differential of $\Lambda^M(t)$ with respect to real time t is given by

$$\begin{aligned}\frac{d}{dt}\Lambda^M(t) &= \lambda^{M-1}(\Lambda^{M-1}(t))\frac{d}{dt}\Lambda^{M-1}(t)\\ &= \lambda^{M-1}(\Lambda^{M-1}(t))\lambda^{M-2}(\Lambda^{M-2}(t))\frac{d}{dt}\Lambda^{M-2}(t)\\ &= \prod_{m=0}^{M-1}\lambda^m(\Lambda^m(t))\cdot\frac{d}{dt}\Lambda^0(t) = \lambda(t). \qquad (\text{A}\cdot 1)\end{aligned}$$

Thus, substituting Eq. (A· 1) into Eq. (1) yields the probability density under the Mth rescaled time as,

$$\begin{aligned}p(\{\Lambda_j^M\}_{j=0}^n) &= p(\{t_j\}_{j=0}^n|\lambda(t))\prod_{j=0}^n\Big[\frac{d}{dt}\Lambda^M(t_j)\Big]^{-1}\\ &= p(\{t_j\}_{j=0}^n|\lambda(t))\prod_{j=0}^n\lambda^{-1}(t_j)\\ &= \Big[\prod_{j=0}^n 1\Big]\exp\Big(-\int_0^T\frac{d}{dt'}\Lambda(t')dt\Big)\\ &= \Big[\prod_{j=0}^n 1\Big]\exp\Big(-1\cdot\Lambda^M(T)\Big), \qquad (\text{A}\cdot 2)\end{aligned}$$

where $\Lambda_j^M \equiv \Lambda^M(t_j)$. Equation (A· 2) indicates that the purchase rate at time $\Lambda^M(t)$ has a unit value.

Appendix B: Sample Generation Using Time-Rescaling Theorem

In HTRm, a sequence of purchase decisions $(t_1, t_2, \cdots)$, generated from the purchase rate $\lambda(t)$, is transformed into that in the rescaled time hierarchically as

$$t_j \to \Lambda_j \equiv \Lambda(t_j) \to \Omega_j \equiv \Omega(\Lambda_j) \to \Psi_j \equiv \Psi(\Omega_j),$$

where $(\Psi_1, \Psi_2, \cdots)$ is regarded as being realized from a Poisson process with a unit rate (see Sect. 3). This property suggests that we can generate a sequence of purchase points $(t_1, t_2, \cdots)$ by inversely transforming a sequence of time points $(\Psi_1, \Psi_2, \cdots)$, generated from a Poisson process with a unit rate, into that in real time as

$$\Psi_j \to \Omega_j \equiv \Psi^{-1}(\Psi_j) \to \Lambda_j \equiv \Omega^{-1}(\Omega_j) \to t_j \equiv \Lambda^{-1}(\Lambda_j).$$

The simulation algorithm proceeds as follows:

0. Assume that the last purchase occurred at t_0, and that $\Psi_0 \equiv \Psi(t_0)$, $\Omega_0 \equiv \Omega(t_0)$ and $\Lambda_0 \equiv \Lambda(t_0)$. Set $j = 1$.
1. Draw a random variable $\Delta\Psi_j \equiv \Psi_j - \Psi_{j-1}$ from the exponential distribution with mean of 1.
2. Find Ω_j as the root of $\Delta\Psi_j = \Psi(\Omega_j) - \Psi(\Omega_{j-1})$, leading to the solution, $\Delta\Omega_j \equiv \Omega_j - \Omega_{j-1} = (\Delta\Psi_j)^{1/\kappa}$.
3. Find Λ_j as the root of $\Delta\Omega_j = \Omega(\Lambda_j) - \Omega(\Lambda_{j-1})$, leading to the solution,

$$\begin{aligned}\Delta\Lambda_j &\equiv \Lambda_j - \Lambda_{j-1}\\ &= \Delta\Omega_j - B_j + \frac{1}{b_h}W\Big(b_h B_j e^{-b_h(\Delta\Omega_j - B_j)}\Big),\end{aligned}$$

where B_j is defined as

$$B_1 = \frac{a_h}{b_h}, \qquad B_j = \frac{a_h}{b_h}\Big(1 + B_{j-1}e^{-b_h\Delta\Lambda_{j-1}}\Big),$$

and $W(z)$ is the Lambert W function defined as the inverse function of $W(z)\exp(W(z)) = z$.
4. Find t_j as the root of $\Delta\Lambda_j = \Lambda(t_j) - \Lambda(t_{j-1})$. Because it cannot be solved analytically, we obtain its solution by employing the bisection method [23].
5. $j \to j + 1$, and go back to 1.

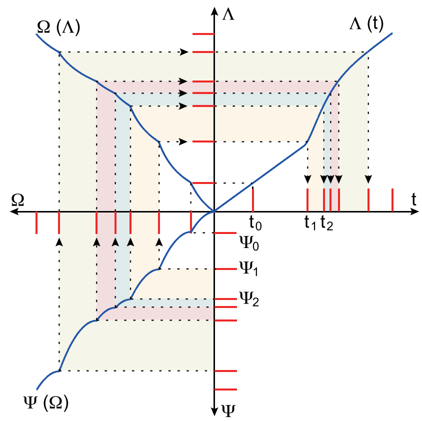

Fig. A· 1 The procedure for generating a sample sequence of purchase decisions from the HTR model. A sequence of purchase time points, $(t_1, t_2, \cdots)$, can be simulated by inversely transforming a Poisson sequence with a unit rate, $(\Psi_1, \Psi_2, \cdots)$, based on the rescaling functions $\Psi(\Omega)$, $\Omega(\Lambda)$, and $\Lambda(t)$.

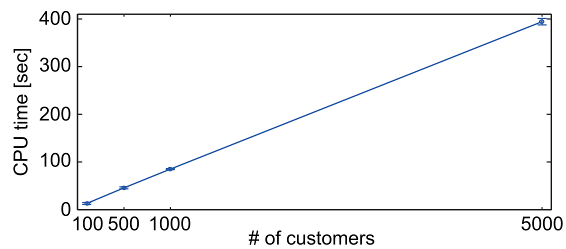

Fig. A· 2 The CPU time for estimating individual parameters against the number of customers in the Fashion data set. The Monte Carlo sample size and the burn-in size is 10,000 and 5,000, respectively. A 30-core CPU (2.6 GHz) computer was used to perform the estimation.

Figure A· 1 shows the simulation procedure described above. Note that the simulation procedure is a hierarchical extension of the original algorithm based on the time-rescaling theorem [15].

Appendix C: Scalability

Given the aggregate parameter (μ, ν), we can estimate $\hat{\theta}^u$ efficiently using parallel computation because each customer's parameter $\hat{\theta}^u$ is estimated based on his/her own data $\{t_j^u\}$. Figure A· 2 shows that the CPU time is proportional to the number of customers. When the EM iteration number and the CPU time for estimating $\hat{\theta}^u$ are denoted by $N_{\rm EM}$ and Δt respectively, the CPU time for determining the aggregate parameter (μ, ν) is represented by $N_{\rm EM}\Delta t$. The aggregate parameter can be determined efficiently based on a randomly selected subset of the whole customer data, which makes the CPU time $N_{\rm EM}\Delta t$ less than several hours. It should be emphasized here that we do not need to update the values of $\hat{\theta}^u$ and (μ, ν) frequently, but should do so after a certain amount of data is observed.

Hideaki Kim received the B.S. degree in Physics from Tohoku University, Sendai, Japan, in 2007, and the M.S. and Ph.D. degrees in Physics from Kyoto University, Kyoto, Japan, in 2009 and 2013, respectively. He joined NTT Corporation in 2013. His research interests include machine learning, latent variable model, point process, user modeling, and computational neuroscience. He is a member of Japan Institute of Marketing Science.

Noriko Takaya received the B.A. degree in English and American Literature from the Gakushuin University in 1987. In 1995, she joined NTT Advertising, Inc. Since 2010, she has been in NTT Service Evolution Laboratories. She is a member of Information Processing Society of Japan, Japan Marketing Academy and Japan Institute of Marketing Science.

Hiroshi Sawada received the B.E., M.E. and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993 and 2001, respectively. He joined NTT Corporation in 1993. He is now an executive manager at the NTT Communication Science Laboratories, Kyoto, Japan. His research interests include statistical signal processing, audio source separation, array signal processing, machine learning, latent variable model, graph-based data structure, and computer architecture. From 2006 to 2009, he served as an associate editor of the IEEE Transactions on Audio, Speech & Language Processing. He is an associate member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE SP Society. He received the Best Paper Award of the IEEE Circuit and System Society in 2000, the SPIE ICA Unsupervised Learning Pioneer Award in 2013, and the Best Paper Award of the IEEE Signal Processing Society in 2014. He is a member of the IEEE, the IEICE and the ASJ.