Recommendation Method for Improving Customer Lifetime Value

Tomoharu Iwata, Kazumi Saito, and Takeshi Yamada, *Member*, *IEEE*

Abstract—It is important for online stores to improve customer lifetime value (LTV) if they are to increase their profits. Conventional recommendation methods suggest items that best coincide with user's interests to maximize the purchase probability, and this does not necessarily help improve LTV. We present a novel recommendation method that maximizes the probability of the LTV being improved, which can apply to both measured and subscription services. Our method finds frequent purchase patterns among high-LTV users and recommends items for a new user that simulate the found patterns. Using survival analysis techniques, we efficiently find the patterns from log data. Furthermore, we infer a user's interests from the purchase history based on maximum entropy models and use the interests to improve recommendation. Since a higher LTV is the result of greater user satisfaction, our method benefits users as well as online stores. We evaluate our method using two sets of real log data for measured and subscription services.

Index Terms—Personalization, marketing, recommender system, collaborative filtering, survival analysis.

1 Introduction

Recommender systems are widely used in online stores [1] because they can improve both user convenience and store profits. Conventional recommendation methods recommend items that best coincide with a user's interests to maximize the purchase probability [2], [3], [4], [5], [6]. Although these methods can increase short-term sales, they do not necessarily maximize long-term profits. For example, if an online store recommends an electronic product that has a lot of peripheral devices, the user is likely to revisit the store to purchase peripheral devices in the future. The recommendation of a DVD that is the first of a series can lead to the purchase of other DVDs in the series. Long-term profit is related to customer lifetime value (LTV) [7], which is defined as the total profit that a customer generates over his/her entire purchase history. Since acquiring new customers is not easy, it is important for stores to increase the LTV of existing customers.

In this paper, we propose a novel recommendation method that maximizes the probability of the LTV being improved. Our method finds frequent purchase patterns among high-LTV users and recommends items for a new user that simulate the found patterns. Since the possibility of purchasing the recommended item depends on the user's interests, we take the interests into consideration in order to generate effective recommendations. To find the patterns, we use survival analysis techniques [8]. To estimate user's interests, we use maximum entropy models. Then, we combine the found patterns and the estimated user's interests in a probabilistically principled framework. Since a higher LTV is the result of higher user satisfaction, our method benefits both users and online stores. Therefore, our method can be seen as a tool for customer relationship management (CRM) [9]. CRM is important in terms of improving relationships between online stores and their users.

The charges that online store users pay relate to the type of service they use and these can be categorized as either measured or subscription services. With a measured service, users pay for purchased items individually. With a subscription service, users pay for the periodic (e.g., monthly or yearly) use of, for example, magazines, music, movies, software, and cell-phone services. To increase LTV, online stores providing measured services must encourage users to purchase many items. On the other hand, online stores providing subscription services need to encourage users to extend their subscription periods. Therefore, online stores should recommend items using different strategies depending on measured or subscription services, and we propose different recommendation methods for each of them.

The remainder of this paper is organized as follows: In Section 2, we briefly review related work. In Section 3, we describe our method for improving LTV for measured services, and in Section 4, we show its validity using the log data of an online music store. In Section 5, we modify our method for subscription services, and in Section 6, we apply it to the log data of an online cartoon distribution service. Finally, we offer concluding remarks and a discussion of future work in Section 7.

- *T. Iwata and T. Yamada are with NTT Communication Science Laboratories, 2-4 Hikaridai, Seika-cho, Keihanna Science City, Kyoto 619-0237, Japan. E-mail: {iwata, yamada}@cslab.kecl.ntt.co.jp.*
- *K. Saito is with the School of Administration and Informatics, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan. E-mail: k-saito@u-shizuoka-ken.ac.jp.*

Manuscript received 12 Apr. 2007; revised 4 Dec. 2007; accepted 4 Mar. 2008; published online 14 Mar. 2008.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDE-2007-04-0158.
Digital Object Identifier no. 10.1109/TKDE.2008.55.

2 Related Work

A number of recommendation methods have been proposed, such as collaborative filtering (CF) [4], [6], content filtering [3], and their hybrids [2], [5]. CF is a method for predicting a user's interests using other users' interests. Content filtering predicts interests using item information. These approaches recommend items that best coincide with

1041-4347/08/$25.00 © 2008 IEEE Published by the IEEE Computer Society

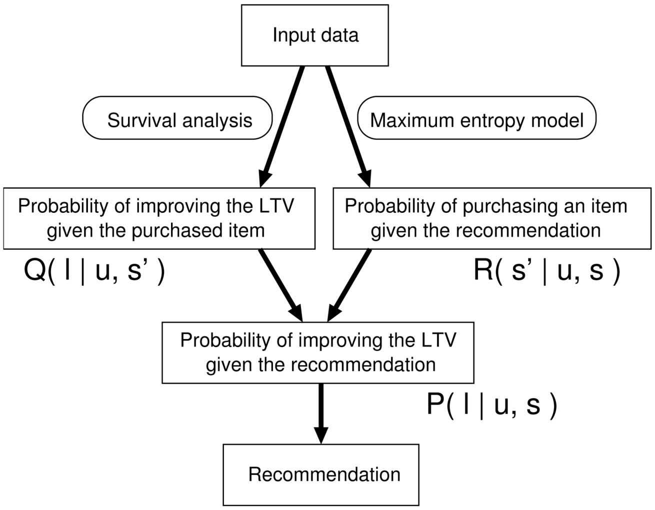

Fig. 1. Framework of our method.

a user's interests to maximize the purchase probability. Our aim, which is to improve LTV, is different from theirs.

The use of LTV estimation to identify loyal customers has been studied by various researchers using survival analysis and data mining techniques [7], [10], [11], [12], [13], [14]. However, they are not used for recommendation. Piatetsky-Shapiro and Masand [15] and Rosset et al. [16] proposed models for estimating the effects of marketing activity on LTV. These models are related to our method because a recommendation can be considered as a marketing activity. However, since our method focuses on the recommendation of items, our method can automatically perform marketing for each user based on the log data. We proposed a recommendation method for subscription services in [17]. In this paper, we extend the method to make it applicable to both measured and subscription services.

3 Proposed Method

3.1 Recommendation for Improving Customer Lifetime Value

Our method recommends item $\hat{s}$ that maximizes $P(l|u,s)$, which is the probability of improving the LTV of user u when item s is recommended, as follows:

$$\hat{s} = \arg\max_{s \in \boldsymbol{S}} P(l|u,s), \tag{1}$$

where S represents a set of items, l represents the improvement of the LTV, and s represents the recommended item. We can also recommend m items with the highest $P(l|u,s)$ values. In real applications, candidates for recommendation would be a subset of item set S, for example, only items not yet purchased by the user.

In general, $P(l|u,s)$ cannot be directly estimated because it is not possible to observe whether the LTV is improved given a recommendation. Therefore, we decompose $P(l|u,s)$ into two components with a number of assumptions so that $P(l|u,s)$ can be estimated from data that can be easily obtained by online stores. Let s' be the item purchased by user u when item s is recommended. If the recommendation does not influence the user purchase behavior, it is natural to think that the recommendation does not influence the LTV, either. Therefore, we assume that improving LTV l and recommendation s are independent conditioned on purchased item s' and user u. Based on this assumption, $P(l|u,s)$ can be decomposed into two components: $Q(l|u,s')$, which is the probability of the LTV being improved when user u purchases item s', and $R(s'|u,s)$, which is the probability of item s' being purchased when item s is recommended to user u, as follows:

$$\begin{aligned} P(l|u,s) &= \sum_{s' \in \boldsymbol{S}} P(l,s'|u,s) \\ &= \sum_{s' \in \boldsymbol{S}} Q(l|u,s') R(s'|u,s). \end{aligned} \tag{2}$$

We can estimate $Q(l|u,s')$ and $R(s'|u,s)$ from the log data with survival analysis techniques and maximum entropy models, respectively. We summarize the framework of our method in Fig. 1.

TABLE 1
An Example of a Purchase Log

user	item	purchase time
1	3	2004/8/16 12:06:28
1	1	2004/8/16 13:01:21
2	2	2004/8/16 18:51:43
1	6	2004/8/16 21:35:06
3	2	2004/8/17 16:42:11
⋮	⋮	⋮
N	10	2005/10/28 23:15:14

3.2 Frequency Model

Assuming the profit generated by an item is constant for all items, the LTV is proportional to the purchase frequency in measured services. We derive the probability of improving LTV given the purchased item $Q(l|u,s')$ using purchase frequency models.

Let e_n^k be the status of the kth purchase of user u_n as follows:

$$e_n^k = \begin{cases} 0 & \text{if the } k\text{th purchase of user } u_n \text{ is the last purchase,} \\ 1 & \text{otherwise.} \end{cases} \tag{3}$$

Let t_n^k be the interpurchase time of the kth purchase of user u_n as follows:

$$t_n^k = \begin{cases} d_{end} - d_n^k & \text{if } e_n^k = 0, \\ d_n^{k+1} - d_n^k & \text{if } e_n^k = 1, \end{cases} \tag{4}$$

where d_n^k is the time of the kth purchase of user u_n, and d_{end} is the last time that the given log data were modified. We assume that interpurchase time t is a discrete variable. The status e and the interpurchase time t are obtained from the purchase log. The purchase log lists the user, the item, and the time of each purchase. Table 1 shows an example of a purchase log. Fig. 2 shows relationships among interpurchase time t, purchase time d, last modification time d_{end}, and status e. As input data for modeling purchase frequencies, we use a set of users, interpurchase times, purchase histories, and statuses.

We model the purchase frequency, or interpurchase time, using frailty models [18], which are used in survival analysis for modeling repeated events. The purchases are

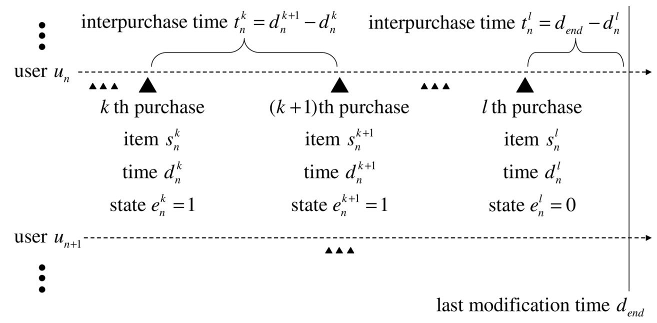

Fig. 2. Relationships among interpurchase time t, purchase time d, last modification time d_{end}, and status e.

repeated events in the sense that a user purchases items repeatedly, as shown in Fig. 2. The purchase frequencies can differ among users, and some heavy users purchase many items and some users purchase only a few items. Frailty models can account for such heterogeneity across users by incorporating a user specific effect into models. Let $h(t|\boldsymbol{x}_n, u_n)$ be a hazard function that represents the instantaneous rate of purchase in interpurchase time t of user u_n with purchase history $\boldsymbol{x}_n$. $\boldsymbol{x}_n = (x_{nb})_{b\in \boldsymbol{B}}$ is a column vector of features for the purchase history of user u_n, where $\boldsymbol{B}$ is a set of feature indices. Examples of features include whether the user has purchased item s_i, and whether the user has purchased item s_i directly after item s_j. In frailty models, the hazard function $h(t|\boldsymbol{x}_n, u_n)$ can be represented as follows:

$$h(t|\boldsymbol{x}_n, u_n) = \lambda_0(t)\lambda_{u_n} \exp(\boldsymbol{\lambda}^\top \boldsymbol{x}_n), \tag{5}$$

where $\lambda_0(t)$ is the baseline hazard function, λ_{u_n} is the frailty effect of user u_n for handling heterogeneity, $\boldsymbol{\lambda} = (\lambda_b)_{b\in\boldsymbol{B}}$ is an unknown parameter vector, and $\boldsymbol{\lambda}^\top$ represents the transpose of $\boldsymbol{\lambda}$. Under the frailty models, the global optimum of the estimation is guaranteed, and $Q(l|u, s')$ can be written in a closed form as described below.

We can estimate unknown parameters $\boldsymbol{\lambda}_u = \{\lambda_{u_n}\}$ and $\boldsymbol{\lambda}$ by maximizing the log partial likelihood using optimization methods such as quasi-Newton methods [19]. The log partial likelihood with the Breslow approximation [8] is defined as follows:

$$\begin{aligned} PL(\boldsymbol{\lambda}_u, \boldsymbol{\lambda}) =&\ \log \prod_t \frac{\prod_{(n,k)\in D(t)} h\big(t|\boldsymbol{x}_n^k, u_n\big)}{\Big(\sum_{(m,j)\in E(t)} h\Big(t|\boldsymbol{x}_m^j, u_m\Big)\Big)^{|D(t)|}} \\ =&\ \sum_t \sum_{(n,k)\in D(t)} \big(\log \lambda_{u_n} + \boldsymbol{\lambda}^\top \boldsymbol{x}_n^k\big) \\ &- \sum_t |D(t)| \log \sum_{(m,j)\in E(t)} \lambda_{u_m} \exp\big(\boldsymbol{\lambda}^\top \boldsymbol{x}_m^j\big), \end{aligned} \tag{6}$$

where $D(t) = \{(n,k)|t_n^k = t \wedge e_n^k = 1\}$ is the set of purchases for which the interpurchase time is equal to t and the status is 1, and $|D(t)|$ is its size, $E(t) = \{(n,k)|t_n^k \geq t\}$ is the set of purchases for which the interpurchase time is more than or equal to t, and $\boldsymbol{x}_n^k$ is the feature vector of the purchase history of user u_n at the kth purchase. Note that we do not need to estimate the baseline hazard function $\lambda_0(t)$ in the estimation of unknown parameters $\boldsymbol{\lambda}_u$ and $\boldsymbol{\lambda}$.

In frailty models, features that have high λ_b (> 0) characterize purchase patterns with a short interpurchase time, and features that have low λ_b (< 0) characterize patterns with a long interpurchase time. These patterns are informative for the online store. For example, they enable the store to understand the relationship between purchase histories and purchase frequencies, or to determine new items to be distributed to increase purchase frequency.

3.3 Probability of Increasing Purchase Frequency Given the Purchased Item

With measured services, if the interpurchase time is shortened, the purchase frequency or LTV increases. Therefore, we assume that $Q(l|u, s')$ is the probability of shortening the interpurchase time when user u purchases item s'. We derive $Q(l|u, s')$ from hazard function $h(t|\boldsymbol{x}, u)$.

Let $\boldsymbol{x}$ be the purchase history of user u, and let $\boldsymbol{x}_{+s'}$ be the updated purchase history when item s' is purchased. For simplicity, we refer to the user who purchases item s' as $u_{+s'}$. We assume that either u or $u_{+s'}$ purchases an item at time t, while neither user purchases the next item before time t. At t, the hazard functions of u and $u_{+s'}$ are $h(t|\boldsymbol{x}, u)$ and $h(t|\boldsymbol{x}_{+s'}, u_{+s'})$, respectively, where we assume that the frailty effect does not change by the purchase $\lambda_u = \lambda_{u_{+s'}}$. The probability that user $u_{+s'}$ purchases the item at t is equal to the probability of shortening the interpurchase time when user u purchases item s' as follows:

$$\begin{aligned} Q(l|u, s') &= \frac{h(t|\boldsymbol{x}_{+s'}, u_{+s'})}{h(t|\boldsymbol{x}, u) + h(t|\boldsymbol{x}_{+s'}, u_{+s'})} \\ &= \frac{1}{1 + \exp\big(-\boldsymbol{\lambda}^\top(\boldsymbol{x}_{+s'} - \boldsymbol{x})\big)}, \end{aligned} \tag{7}$$

which is a sigmoid function. Note that the probability does not depend on frailty effect λ_u. While we can recommend an item that maximizes $Q(l|u, s')$, the user may not purchase the recommended item when the user is not interested in the recommended item. In this case, the recommendation is useless with respect to improving LTV. Therefore, we need to consider whether the recommended item is purchased by the user taking the user's interests into consideration.

3.4 Probability of Purchasing an Item Given the Recommendation

We explain the estimation of $R(s'|u, s)$, which is the probability that user u purchases item s' when item s is recommended. Let $R(s'|u)$ be the probability that user u purchases item s' without recommendation, where $\sum_{s'\in\boldsymbol{S}} R(s'|u) = 1$. The recommendation of item s will increase the probability of the item being purchased. We assume that the probability increases γ times as follows:

$$R(s'|u, s) = \begin{cases} \frac{\gamma}{Z(u,s)} R(s'|u) & s = s', \\ \frac{1}{Z(u,s)} R(s'|u) & \text{otherwise}, \end{cases} \tag{8}$$

where $Z(u, s) = 1 + (\gamma - 1)R(s|u)$ is the normalization term, and $\gamma \geq 1$. γ represents the effect of the recommendation on purchase behavior and depends on the way that the recommendation is presented in the online store, including considerations such as display size and position.

If an item matches the user's interests, the probability of the user purchasing the item becomes high, and if it does not match, the probability is low. Therefore, $R(s'|u)$

TABLE 2
Number of Users, Transactions, and Features in the Log Data for Frequency Model Evaluation

	2005/08/31	2005/09/30	2005/10/31
number of users	10,923	13,612	17,123
number of transactions	55,416	74,582	102,165
number of features	4,234	5,662	8,283

represents the degree of agreement between the interests of user u and item s'. Since conventional recommendation methods suggest items that coincide with a user's interests, we can use conventional methods to obtain $R(s'|u)$. We employ maximum entropy models [2], [4], [20], which estimate a probabilistic distribution that maximizes entropy under the constraints of the given data. Maximum entropy models can integrate arbitrary features such as purchase histories and user attributes, and they are used for a variety of natural language tasks [21], [22] as well as recommendation. In maximum entropy models, the probability that user u purchases item s' is represented as follows:

$$R(s'|u) = \frac{1}{Z(u)} \exp\left(\sum_c \alpha_c y_c(u, s') \right), \tag{9}$$

where $Z(u) = \sum_{s \in \mathcal{S}} \exp\big(\sum_c \alpha_c y_c(u, s)\big)$ is the normalization term, y_c is a feature of the purchase history, α_c is an unknown parameter to be estimated, and c is an index of each feature.

The unknown parameters $\boldsymbol{\alpha} = \{\alpha_c\}$ can be estimated by maximizing the following log likelihood using optimization techniques such as quasi-Newton methods:

$$\begin{aligned} L(\boldsymbol{\alpha}) &= \sum_n \sum_k \log R\big(s_k^n | u_k^n\big) \\ &= \sum_n \sum_k \sum_c \alpha_c y_c\big(u_k^n, s_k^n\big) \\ &\quad - \sum_n \sum_k \log \sum_s \exp\left(\sum_c \alpha_c y_c\big(u_k^n, s\big) \right), \end{aligned} \tag{10}$$

where s_k^n is the kth item in the purchase sequence of user u_n, and u_k^n is the purchase history of user u_n before the kth item is purchased. In maximum entropy models, we can obtain a global optimum solution. By using a Gaussian prior with a zero mean on unknown parameter α, overfitting can be reduced [23]. We use a Gaussian prior in our experiments.

4 Experimental Results for a Measured Service

We evaluated our method by employing the log data of an online music download service in Japan from 1 April 2005. We set the unit time as one day.

4.1 Evaluation of Frequency Models

We model purchase frequencies with frailty models, in which we use purchase histories as their covariates. To evaluate our purchase frequency model, we compared the frailty models $h(t|\boldsymbol{x}_n, u_n) = \lambda_0(t)\lambda_{u_n} \exp(\boldsymbol{\lambda}^\top \boldsymbol{x}_n)$ that use both the purchase history and the user heterogeneity information described in Section 3.2 with the Cox proportional hazards models [24] $h(t|\boldsymbol{x}_n) = \lambda_0(t)\exp(\boldsymbol{\lambda}^\top \boldsymbol{x}_n)$ that use the purchase history but not the user heterogeneity information, and models that do not use the purchase history information $h(t)$. We used the following features for the frailty models and Cox proportional hazards models:

$$x_{n,i} = \begin{cases} 1 & \text{if user } u_n \text{ has purchased item } s_i, \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

where we omitted features that appeared fewer than 10 times in the training data.

We used three data sets consisting of the log data up to 31 August 2005, 30 September 2005, and 31 October 2005. We used the last interpurchase time of each user as test data and used others as training data, in which the number of users, transactions, and features were as listed in Table 2. For the evaluation measurement, we used the average log partial likelihood for test data, which is the log partial likelihood divided by the number of transactions. A higher average log partial likelihood indicates a higher predictive performance of the model. Table 3 shows the results. The average log partial likelihoods of the frailty models were higher than those of the Cox proportional hazards models and the models that do not use purchase histories. This result shows that frailty models with both purchase history and user heterogeneity information can predict purchase frequencies more precisely than models without them.

TABLE 3
Average Log Partial Likelihoods of Frequency Models

	2005/08/31	2005/09/30	2005/10/31
models without purchase histories	-8.594	-8.807	-9.031
Cox models	-8.282	-8.504	-8.732
frailty models	**-8.270**	**-8.493**	**-8.721**

4.2 Evaluation of Purchase Models

We evaluated purchase models based on the maximum entropy models described in Section 3.4, which is the probability that user u purchases item s', $R(s'|u)$. We used first-order Markov transitions as features, since we considered the last purchased item revealed the user's interests:

$$y_{s_a,s_b}(u, s_j) = \begin{cases} 1 & \text{if item } s_a \text{ is the last purchased} \\ & \text{item of user } u \text{ and } s_b = s_j, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

We used three data sets consisting of the log data up to 31 August 2005, 30 September 2005, and 31 October 2005, from which we omitted transitions to the same item, items that appeared fewer than 10 times, and users that purchased fewer than five items. We divided each set of data into training and test data. The number of users, transactions, and items were as shown in Table 4. We compared maximum entropy models with uniform distributions, multinomial distributions, item-based CF [25], and probabilistic latent semantic analysis (PLSA) [26]. Uniform distributions do not use the information in the log data at all. Multinomial distributions use the information about the number of each item purchased by all users but do not consider individual interests. The unknown parameters of the multinomial distribution were estimated by the maximum likelihood method. In item-based CF, we used the cosine similarity that is defined as follows:

TABLE 4
Number of Users, Transactions, and Items in the Log Data of a Measured Service for Purchase Model Evaluation

	2005/08/31	2005/09/30	2005/10/31
number of users	5,129	6,857	9,214
number of transactions	35,301	49,916	72,156
number of items	1,091	1,405	1,811

$$sim(s_i, s_j) = \cos(\boldsymbol{w}_i, \boldsymbol{w}_j) = \frac{\boldsymbol{w}_i^\top \boldsymbol{w}_j}{\|\boldsymbol{w}_i\| \, \|\boldsymbol{w}_j\|}, \tag{13}$$

where $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{iN})^\top$ is a column vector, in which $w_{in} = 1$ if user u_n has purchased item s_i, $w_{in} = 0$ otherwise, and $\|\cdot\|$ is the euclidean norm. The probability of purchasing item s' is proportional to the summation of the cosine similarities of items in the purchase history u as follows:

$$R(s'|u) \propto \sum_{s \in u} sim(s', s), \tag{14}$$

which implies that items similar to purchased items are likely to be purchased. In PLSA, the probability that user u purchases item s' is represented as follows:

$$R(s'|u) = \sum_{z=1}^{Z} P(s'|z) P(z|u), \tag{15}$$

where z is a latent class, Z is the number of latent classes, $P(s'|z)$ is the probability that a user in class z purchases item s', and $P(z|u)$ is the probability that user u belongs to class z. The probabilities $P(s'|z)$ and $P(z|u)$ can be estimated by maximizing the likelihood using the EM algorithm. We used $Z = 10$.

For the evaluation measurements, we used the average log likelihood and the accuracy of the next purchase predictions. Tables 5 and 6 show the results. The predictive performance of the maximum entropy models was the highest. Even though item-based CF and PLSA are widely used for unknown rating prediction, this result indicates that they are inadequate for the prediction of the next purchase items since they do not consider purchase orders. Maximum entropy models with purchase histories can predict user purchase behavior and interest more precisely than those without them.

4.3 Purchase Frequencies and Purchase Probabilities

Conventional recommendation methods recommend items that have a high probability of being purchased. If high-LTV users tend to purchase high-purchase-probability items, conventional methods are sufficient to improve LTV. We investigated the relationship between LTV or purchase frequencies and purchase probabilities using the frailty model estimated using the log data up to 31 October 2005.

The hazard function in frailty models is multiplied by $\exp(\lambda_i)$ with the existence of feature x_i. We expressed the effect on improving LTV of the purchase of an item s_i by $\exp(\lambda_i)$, and we expressed the purchase probability by the multinomial distribution parameter estimated by the maximum likelihood. Fig. 3 shows a scatter plot of the improving LTV effect of the purchase, $\exp(\lambda_i)$, and the purchase probability, $R(s_i)$. The correlation coefficient was -0.052, and the improving LTV effect and the purchase probability were negatively correlated. This result implies that the recommendation that suggest items that have a high probability of being purchased does not necessarily improve LTV.

TABLE 5
Average Log Likelihoods of Purchase Models for a Measured Service

	2005/08/31	2005/09/30	2005/10/31
uniform distribution	-6.995	-7.248	-7.502
multinomial distribution	-5.944	-6.190	-6.338
item-based CF	-6.066	-6.227	-6.326
PLSA	-5.897	-6.047	-6.198
maximum entropy model	**-5.558**	**-5.725**	**-5.834**

TABLE 6
Accuracies (Percent) of Purchase Models for a Measured Service

	2005/08/31	2005/09/30	2005/10/31
uniform distribution	0.09	0.07	0.06
multinomial distribution	4.90	3.63	2.56
item-based CF	8.16	6.76	6.46
PLSA	6.56	4.99	5.05
maximum entropy model	**8.71**	**8.39**	**9.01**

4.4 Simulation

In Section 4.1, we showed that frailty models could predict purchase frequencies, and in Section 4.2, we revealed that maximum entropy models could predict user purchase behavior. Here, we examine the effectiveness of our method by simulation. We simulate user behavior using the frailty model and the maximum entropy model that are estimated using the log data from 1 April 2005 to 31 October 2005.

The function of Algorithm 1 is to generate a purchase history, where t is the time, u is the purchase history, u_{+s} is the updated history when item s is purchased, ϕ is an empty history, MaxTime is the time period for the simulation, $\mathrm{Multinomial}(\psi)$ is the multinomial distribution of one event with j's success probability ψ_j, and $\mathrm{Exponential}(\lambda)$ is the exponential distribution with parameter λ. The first item that the user purchases is determined according to $R(s)$,

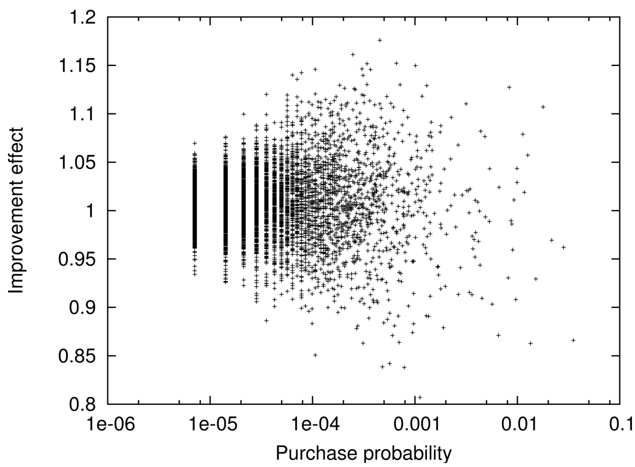

Fig. 3. Purchase probability versus LTV improving effect.

which is the probability of purchasing item s first (line 4). If the user has purchased items, we perform a recommendation using our method (line 6), and the item that the user purchases is determined according to $R(s|u, \hat{s})$ (line 7). The interpurchase time is sampled from the exponential distribution (line 10), and the time is updated (line 11). We estimated unknown parameters $R(s)$ and λ_{u_n} using the log data by the maximum likelihood method.

Algorithm 1 Simulation algorithm of a user behavior in a measured service.

1: Set $t \leftarrow 0$, $u \leftarrow \phi$
2: **while** $t \leq \mathrm{MaxTime}$ **do**
3: **if** $u = \phi$ **then**
4: Sample $s \sim \mathrm{Multinomial}(R(s))$
5: **else**
6: $\hat{s} \leftarrow \arg\max_s P(l|u, s)$
7: Sample $s \sim \mathrm{Multinomial}(R(s|u, \hat{s}))$
8: **end if**
9: Set $u \leftarrow u_{+s}$
10: Sample $\tau \sim \mathrm{Exponential}\big(\lambda_0 \lambda_{u_n} \exp(\boldsymbol{\lambda}^\top \boldsymbol{x})\big)$
11: Set $t \leftarrow t + \tau$
12: **end while**
13: Output u

We compared our method with the following recommendation methods:

- **Q Recommend** recommends an item that is most likely to increase the purchase frequency when the user purchases the item. Line 6 in Algorithm 1 is changed as follows:

$$\hat{s} \leftarrow \arg\max_s Q(l|u, s). \tag{16}$$

  This recommendation does not take the user's interests into consideration.
- **R Recommend** recommends an item that best coincides with the user's interests. Line 6 is changed as follows:

$$\hat{s} \leftarrow \arg\max_s R(s|u). \tag{17}$$

  This recommendation is the same strategy as that of conventional methods.
- **No Recommend** does not recommend any items. The item that the user purchases is determined solely according to the user's interests. Line 6 is omitted, and line 7 is changed as follows:

$$\text{Sample } s \sim \mathrm{Multinomial}(R(s|u)). \tag{18}$$

  This recommendation can also be achieved by using $\gamma = 1$ with Algorithm 1, which means that the recommendation has no effect on purchase behavior.

We generated 171,230 user histories with recommendations by each method where $1 \leq \gamma \leq 10$, in which we used each estimated λ_u 10 times. We set the time period for the simulation at 365 days. Fig. 4 shows the average number of purchased items with different γ. Our method was more successful than the others in increasing the number of purchased items. The number of purchased items increases with an increase in γ. This result indicates that if recommendations can influence user behavior, or $\gamma > 1$, our method can increase the purchase frequency. Moreover, the purchase frequency can be increased further by improving the effect of the recommendations. Q Recommend also increases the number of purchased items, although the effect was smaller than that of our method. This is because Q Recommend may recommend items that have low probabilities of being purchased by the user. On the other hand, our method recommends items taking user's interests into account in order to improve the recommendations. R Recommend reduces the number of purchased items because the purchase frequency is negatively correlated with the purchase probability as shown in Fig. 3.

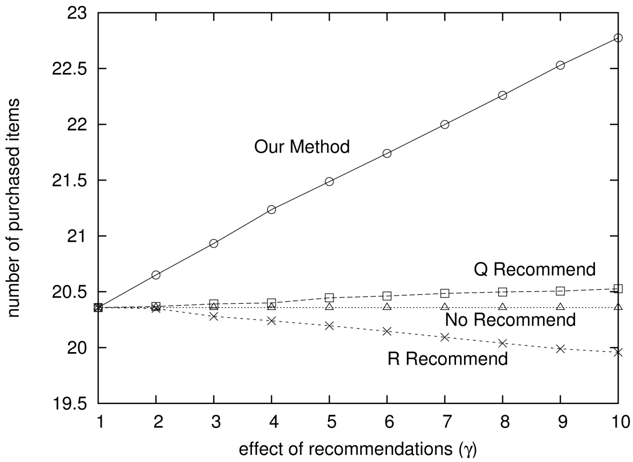

Fig. 4. Average number of purchased items in simulations with parameters estimated from log data.

5 Recommendation for Subscription Services

In this section, we describe a recommendation method designed to improve LTV for subscription services, which is obtained by modifying our method for measured services described in Section 3. With subscription services, the LTV is proportional to the subscription period and does not depend on the purchase frequency. Therefore, we modify the probability of improving the LTV given the purchased item to the probability of extending the subscription period given the purchased item.

5.1 Subscription Period Model

We model the subscription period using Cox proportional hazards models. Let $h(t|\boldsymbol{x})$ be the hazard function, which represents the instantaneous rate of unsubscription at period t of users with purchase history $\boldsymbol{x}$. In Cox proportional hazards models, the hazard function $h(t|\boldsymbol{x})$ is defined as follows:

$$h(t|\boldsymbol{x}) = \beta_0(t) \exp(\boldsymbol{\beta}^\top \boldsymbol{x}), \tag{19}$$

where $\boldsymbol{\beta}$ is an unknown parameter vector, and $\beta_0(t)$ is the baseline hazard function.

Let e_n be the status of user u_n as follows:

$$e_n = \begin{cases} 0 & \text{if user } u_n \text{ is still subscribing,} \\ 1 & \text{if user } u_n \text{ has already unsubscribed.} \end{cases} \tag{20}$$

TABLE 7
An Example of a Subscription Log

user	status	subscribed time	unsubscribed time
1	1	2004/8/16 11:50:30	2005/01/08 20:14:11
2	0	2004/8/16 18:01:28	
3	1	2004/8/17 16:10:51	2004/08/25 13:01:06
4	1	2004/8/17 21:39:29	2004/08/29 07:21:51
5	0	2004/8/18 01:44:17	
⋮	⋮	⋮	⋮
N	0	2005/10/28 23:10:03	

The subscription period t_n of user u_n is obtained as follows:

$$t_n = \begin{cases} d_{end} - d_n^{start} & \text{if } e_n = 0, \\ d_n^{end} - d_n^{start} & \text{if } e_n = 1, \end{cases} \tag{21}$$

where d_n^{start} is the subscribed time of user u_n, d_n^{end} is the unsubscribed time of user u_n, and d_{end} is the last time the log was modified. We assume that subscription period t is a discrete variable. The status e_n and the subscription period t_n are obtained from the subscription log. The subscription log consists of the subscribed time, the status (still subscribing or already unsubscribed), and, where relevant, the unsubscribed time of each user. Table 7 shows an example of a subscription log. Fig. 5 shows the relationships among subscription period t, subscribed time d^{start}, unsubscribed time d^{end}, last modification time d_{end}, and status e. Note that unsubscription is not a repeated event in the sense that one user can only unsubscribe once.

We can estimate unknown parameters $\boldsymbol{\beta}$ by maximizing the log partial likelihood as follows:

$$\begin{aligned} PL(\boldsymbol{\beta}) &= \log \prod_t \frac{\prod_{n \in D(t)} h(t|\boldsymbol{x}_n(t))}{\left(\sum_{m \in R(t)} h(t|\boldsymbol{x}_m(t))\right)^{|D(t)|}} \\ &= \sum_t \sum_{n \in D(t)} \boldsymbol{\beta}^\top \boldsymbol{x}_n(t) \\ &\quad - \sum_t |D(t)| \log \sum_{m \in R(t)} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{x}_m(t)\right), \end{aligned} \tag{22}$$

where $D(t)$ is the set of users unsubscribed at t, $R(t)$ is the set of users subscribing at t, and $\boldsymbol{x}_n(t)$ is the feature vector of user u at t. Note that we need to treat purchase history $\boldsymbol{x}_n$ as time-dependent variables since purchase history $\boldsymbol{x}_n$ change when user u_n purchases an item. Features that have low $\beta_b\ (< 0)$ are characteristic purchase patterns for long-subscription users, and features that have high $\beta_b\ (> 0)$ are characteristic patterns for short-subscription users.

5.2 Probability of Extending Subscription Period Given Purchased Item

With subscription services, if the subscription period is long, the LTV increases. We assume that $Q(l|u, s')$ is the probability of extending the subscription period when user u purchases item s'. We estimate $Q(l|u, s')$ from hazard function $h(t|\boldsymbol{x})$ in a manner similar to that described in Section 3.3.

Let $\boldsymbol{x}$ be the purchase history of user u, and $\boldsymbol{x}_{+s'}$ be the updated purchase history when item s' is purchased. For simplicity, we refer to the user when item s' is purchased as $u_{+s'}$. We assume that either u or $u_{+s'}$ unsubscribed at t while the other is still subscribing. At t, the hazard functions of u

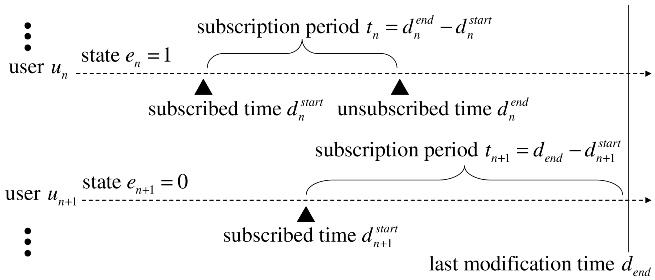

Fig. 5. Relationships among subscription period t, subscribed time d^{start}, unsubscribed time d^{end}, last modification time d_{end}, and status e.

and $u_{+s'}$ are $h(t|\boldsymbol{x})$ and $h(t|\boldsymbol{x}_{+s'})$, respectively. The probability that user u unsubscribed at t is equal to the probability of extending the subscription period when user u purchases item s' as follows:

$$\begin{aligned} Q(l|u, s') &= \frac{h(t|\boldsymbol{x})}{h(t|\boldsymbol{x}) + h(t|\boldsymbol{x}_{+s'})} \\ &= \frac{1}{1 + \exp\left(-\boldsymbol{\beta}^\top (\boldsymbol{x} - \boldsymbol{x}_{+s'})\right)}, \end{aligned} \tag{23}$$

which is a sigmoid function.

Our method for subscription services recommends item $\hat{s}$ that maximizes $P(l|u, s)$, which is the probability of extending the subscription period of user u when item s is recommended as follows:

$$\begin{aligned} \hat{s} &= \arg\max_{s \in \boldsymbol{S}} P(l|u, s), \\ &= \arg\max_{s \in \boldsymbol{S}} \sum_{s' \in \boldsymbol{S}} Q(l|u, s') R(s'|u, s), \end{aligned} \tag{24}$$

where we use (23) as $Q(l|u, s')$.

6 Experimental Results for a Subscription Service

We evaluated our method for extending subscription periods by using the log data of an online cartoon distribution service for cell phones in Japan. With this service, users pay monthly to read cartoons on their cell phones. Some cartoons have several volumes, and some users purchased an item more than once. We regarded a cartoon that had several volumes as one item, and the unit time was one day. This service began on 16 August 2004, and the last modification date of the log was 28 October 2005.

6.1 Evaluation of Subscription Period Models

With our method, we assume that we can estimate subscription periods more precisely using purchase histories. To evaluate this assumption, we compared the Cox proportional hazards models $h(t|\boldsymbol{x}) = \beta_0(t)\exp(\boldsymbol{\beta}^\top \boldsymbol{x})$ that use the purchase histories described in Section 5.1 and models that do not use purchase histories $h(t)$. We used the following three sets of features for the Cox proportional hazards models:

- **F1:** whether user u has purchased item s_i,

$$x_{n,i} = \begin{cases} 1 & \text{if user } u_n \text{ has purchased item } s_i, \\ 0 & \text{otherwise,} \end{cases} \tag{25}$$

TABLE 8
Number of Features of Subscription Period Models

	2005/06/30	2005/07/31	2005/08/31
Cox models (F1)	75	80	84
Cox models (F2)	2,671	3,159	3,485
Cox models (F3)	3,711	4,455	5,250

- **F2:** whether user u_n has purchased item s_i and item s_j,

$$x_{n,i,j} = \begin{cases} 1 & \text{if user } u_n \text{ has purchased} \\ & \text{item } s_i \text{ and item } s_j, \\ 0 & \text{otherwise,} \end{cases} \tag{26}$$

- **F3:** whether user u_n has purchased item s_j next to item s_i,

$$x_{n,i\to j} = \begin{cases} 1 & \text{if user } u \text{ has purchased} \\ & \text{item } s_j \text{ next to item } s_i, \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

where we omitted features that appeared fewer than 10 times in the purchase histories.

We used three sets of training and test data. The training data were the log data up to 30 June 2005, 31 July 2005, and 31 August 2005. The test data were the log data of subscribers on the end date of the training data, and the end date of the test data was 28 October 2005. The number of features was as shown in Table 8, and the number of subscribers and unsubscribers were as shown in Table 9.

For the evaluation measurement, we used the average log partial likelihood for test data. Table 10 shows the results. The average log partial likelihoods of the Cox proportional hazards models (F3) were higher than those for the model that does not use purchase histories. This result shows that Cox proportional hazards models can predict subscription periods more precisely by using purchase histories.

6.2 Evaluation of Purchase Models

We evaluated purchase models in a subscription service based on the maximum entropy models described in Section 3.4, which estimate the probability that user u purchases item s', $R(s'|u)$ as in Section 4.2. We used first-order Markov transitions as features. We used three sets of training and test data. The training data were the log data up to 30 June 2005, 31 July 2005, and 31 August 2005, from which we omitted transitions to the same item, items that appeared fewer than 10 times, and users that purchased fewer than five items. The test data were the log data from the end date of the training data to 28 October 2005, from which we omitted transitions to the same item and

TABLE 10
Average Log Partial Likelihoods of Subscription Period Models

	2005/06/30	2005/07/31	2005/08/31
models without purchase histories	-9.845	-9.465	-9.904
Cox models (F1)	-9.205	-9.445	-9.812
Cox models (F2)	-9.179	-9.422	-9.839
Cox models (F3)	**-9.129**	**-9.351**	**-9.798**

transitions that contained items that had not been distributed during the training data period. The number of users, transactions, and items were as shown in Table 11. We compared maximum entropy models with uniform distributions, multinomial distributions, item-based CF, and PLSA.

We used the average log likelihood and accuracy for the evaluation measurements. Tables 12 and 13 show the results. The maximum entropy models had higher average log likelihoods and accuracies than those of other methods except for the PLSA accuracy of 31 August 2005 data. These results show that we can predict next purchase items using the maximum entropy models in this subscription service as well as in the measured service.

6.3 Subscription Periods and Purchase Probabilities

We investigated the relationship between the effect on extending subscription periods and the purchase probabilities as in Section 4.3, where we used the Cox proportional hazards model (F3) and the maximum entropy model estimated using the log data up to 31 August 2005.

The expected subscription period given the purchase history in Cox proportional hazards models is multiplied by $\exp(-\beta_{i\to j})$ with the existence of feature $x_{i\to j}$. We expressed the effect on extending subscription periods of a transition by $\exp(-\beta_{i\to j})$. The probability of the transition was estimated using maximum entropy models. The features of the Cox proportional hazards model (F3) and the maximum entropy model are both first-order Markov transitions. Fig. 6 shows a scatter plot of the extending effects of transitions, $\exp(-\beta_{i\to j})$, and their transition probabilities, $R(s_j|s_i)$. The correlation coefficient was 0.159, and there was little correlation. This result implies that the recommendation of high-purchase probability items does not necessarily lead to extend the subscription period.

6.4 Simulation

We examined the effectiveness of our method for subscription services by simulation. We simulated user behavior using the Cox proportional hazards model and the maximum entropy model that we estimated using the log data from 16 August 2004 to 28 October 2005. The log data comprised 107 items.

TABLE 9
Number of Subscribers and Unsubscribers in the Log Data for Subscription Period Model Evaluation

	2005/06/30		2005/07/31		2005/08/31	
	training	test	training	test	training	test
number of subscribers	13,284	7,221	14,669	9,608	28,409	17,028
number of unsubscribers	4,988	6,063	8,802	5,061	9,765	11,381

TABLE 11
Number of Users, Transactions, and Items in the Log Data of a Subscription Service for Purchase Model Evaluation

	2005/06/30	2005/07/31	2005/08/31
number of users	6,088	7,474	10,180
number of transactions	53,866	67,524	89,932
number of items	74	80	84

TABLE 12
Average Test Log Likelihoods of Purchase Models for a Subscription Service

	2005/06/30	2005/07/31	2005/08/31
uniform distribution	-4.304	-4.382	-4.431
multinomial distribution	-4.344	-4.449	-4.316
item-based CF	-4.037	-4.152	-4.111
PLSA	-4.141	-4.199	-4.069
maximum entropy model	**-3.867**	**-3.965**	**-3.808**

The function of Algorithm 2 is to generate a subscription period t, where $\mathrm{Bernoulli}(\theta)$ is the Bernoulli distribution with success probability θ. First, from lines 3 to 4 in Algorithm 2, we decide whether the user unsubscribes or not in unit time using the unsubscription probability in unit time of a subscriber $h(t|\boldsymbol{x})$. Second, from lines 7 to 8, we decide whether the user purchases or not in unit time using the purchase probability in unit time, g. We assumed that g is constant over subscription period t. The first item that the user purchases is determined according to $R(s)$, where $R(s)$ is the probability of purchasing item s first (line 10). If the user has purchased some items, we perform a recommendation using our method (line 12), and the item that the user purchases is determined according to $R(s|u,\hat{s})$ (line 13). We estimated unknown parameters $\beta_0(t)$, g, and $R(s)$ using the log data by the maximum likelihood method.

Algorithm 2 Simulation algorithm of a user behavior in a subscription service.

1: Set $t \leftarrow 0$, $u \leftarrow \phi$
2: **while** $t \leq \mathrm{MaxTime}$ **do**
3: Sample $r_1 \sim \mathrm{Bernoulli}(h(t|\boldsymbol{x}))$
4: **if** r_1 is success **then**
5: break
6: **end if**
7: Sample $r_2 \sim \mathrm{Bernoulli}(g)$
8: **if** r_2 is success **then**
9: **if** $u = \phi$ **then**
10: Sample $s \sim \mathrm{Multinomial}(R(s))$
11: **else**
12: $\hat{s} \leftarrow \arg\max_s P(l|u,s)$
13: Sample $s \sim \mathrm{Multinomial}(R(s|u,\hat{s}))$
14: **end if**
15: Set $u \leftarrow u_{+s}$
16: **end if**
17: Set $t \leftarrow t+1$
18: **end while**
19: Output t

TABLE 13
Accuracies (Percent) of Purchase Models for a Subscription Service

	2005/06/30	2005/07/31	2005/08/31
uniform distribution	1.35	1.25	1.19
multinomial distribution	1.66	1.45	2.16
item-based CF	10.88	10.92	10.41
PLSA	11.48	11.48	**12.33**
maximum entropy model	**12.29**	**12.18**	11.95

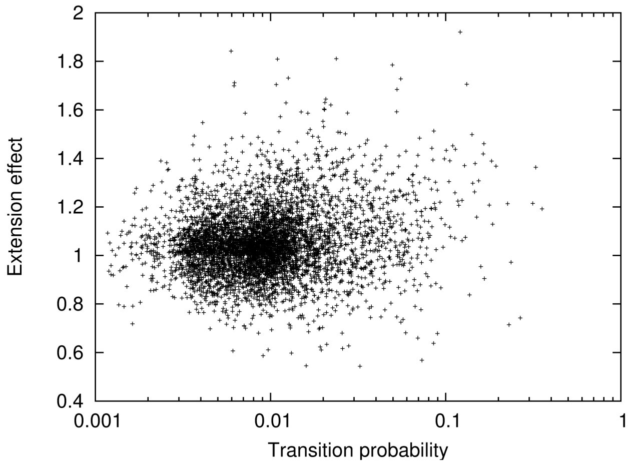

Fig. 6. Transition probability versus extension effect.

We compared our method with Q Recommend, R Recommend, and No Recommend. We generated 100,000 user subscription periods with recommendations by each method where $1 \leq \gamma \leq 10$. We set the maximum subscription period at 365 days. Fig. 7 shows the average subscription periods. Our method was more successful than the others in extending subscription periods. Since Q Recommend may recommend items that have no probability of being purchased by the user, the effect of Q Recommend is smaller than that of our method. R Recommend only slightly extended subscription periods because there was little correlation between the subscription periods and the purchase probability as in Fig. 6.

7 Conclusion

In this paper, we have proposed a novel recommendation method for improving LTV, which encourages users to purchase more items for measured services and to extend their subscription periods for subscription services. We used basic features in the experiments to make the novelty of our framework easy to understand. Our method can use other features such as high-order Markov transitions and

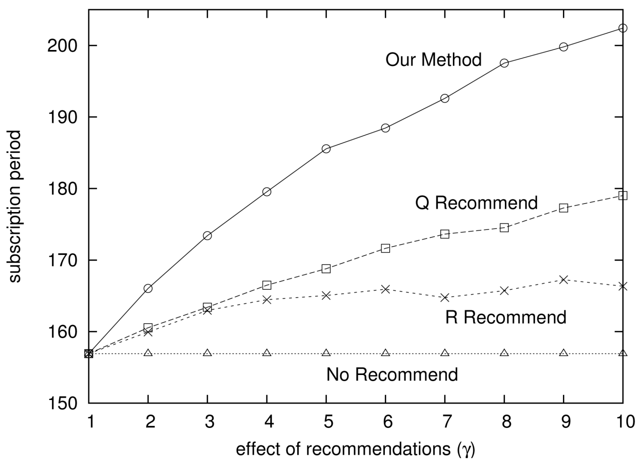

Fig. 7. Average subscription periods in simulations with parameters estimated from the log data.

user attributes. Since our method is divided into two modules, namely, the estimation of LTV and the estimation of user's interests, it can be further enhanced using survival analysis or CF techniques. For example, we can combine our approach with content filtering to estimate user's interests.

Although we have already obtained encouraging results, some directions remain in which we must extend our approach before it can become a useful tool for recommendation. First, it is important to improve the frailty model or Cox proportional hazards model and the feature selection in order to find informative purchase patterns among high-LTV users. Second, we need to estimate the effect of recommendations on purchase behavior from the log data automatically. This can be achieved by using the log data of purchase histories with and without recommendations. Finally, we want to apply our method to an online store and show how it can improve the LTV of real users.

References

[1] J.B. Schafer, J.A. Konstan, and J. Riedl, "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery,* vol. 5, pp. 115-153, 2001.

[2] X. Jin, Y. Zhou, and B. Mobasher, "A Maximum Entropy Web Recommendation System: Combining Collaborative and Content Features," *Proc. ACM SIGKDD,* 2005.

[3] R.J. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," *Proc. Fifth ACM Conf. Digital Libraries (ACM DL '00),* pp. 195-204, 2000.

[4] D.Y. Pavlov and D.M. Pennock, "A Maximum Entropy Approach to Collaborative Filtering in Dynamic, Sparse, High-Dimensional Domains," *Proc. Advances in Neural Information Processing Systems (NIPS '02),* pp. 1441-1448, 2002.

[5] A. Popescul, L. Ungar, D. Pennock, and S. Lawrence, "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments," *Proc. 17th Conf. Uncertainty in Artificial Intelligence (UAI '01),* pp. 437-444, 2001.

[6] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An Open Architecture for Collaborative Filtering of Netnews," *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW '94),* pp. 175-186, 1994.

[7] D.R. Mani, J. Drew, A. Betz, and P. Datta, "Statistics and Data Mining Techniques for Lifetime Value Modeling," *Proc. ACM SIGKDD '99,* pp. 94-103, 1999.

[8] M. Cleves, W.W. Gould, and R. Gutierrez, *An Introduction to Survival Analysis Using Stata,* revised ed. Stata Press, 2004.

[9] M.J.A. Berry and G.S. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management.* John Wiley, 2004.

[10] W.H. Au, K.C.C. Chan, and X. Yao, "A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction," *IEEE Trans. Evolutionary Computation,* vol. 7, no. 6, pp. 532-545, 2003.

[11] P.S. Fader, B.G.S. Hardie, and K.L. Lee, ""Counting Your Customers" The Easy Way: An Alternative to The Pareto/NBD Model," *Marketing Science,* vol. 24, no. 2, pp. 275-284, 2005.

[12] L. Jen, C.-H. Chou, and G.M. Allenby, "A Bayesian Approach to Modeling Purchase Frequency," *Marketing Letters,* vol. 14, no. 1, pp. 5-20, 2003.

[13] M.C. Mozer, R. Wolniewicz, D.B. Grimes, E. Johnson, and H. Kaushansky, "Predicting Subscriber Dissatisfaction and Improving Retention in The Wireless Telecommunications Industry," *IEEE Trans. Neural Networks,* vol. 11, no. 3, pp. 690-696, 2000.

[14] Y. Shono, Y. Takada, N. Komoda, H. Oiso, A. Hiramatsu, and K. Fukuda, "Customer Analysis of Monthly-Charged Mobile Content Aiming at Prolonging Subscription Period," *Proc. IEEE Int'l Conf. Computational Cybernetics (ICCC '04),* pp. 279-284, 2004.

[15] G. Piatetsky-Shapiro and B. Masand, "Estimating Campaign Benefits and Modeling Lift," *Proc. ACM SIGKDD '99,* pp. 185-193, 1999.

[16] S. Rosset, E. Neumann, U. Eick, and N. Vatnik, "Customer Lifetime Value Models for Decision Support," *Data Mining and Knowledge Discovery,* vol. 7, pp. 321-339, 2003.

[17] T. Iwata, K. Saito, and T. Yamada, "Recommendation Method for Extending Subscription Periods," *Proc. ACM SIGKDD '06,* pp. 574-579, 2006.

[18] J. Box-Steffensmeier and B. Jones, *Event History Modeling.* Cambridge Univ. Press, 2004.

[19] D.C. Liu and J. Nocedal, "On the Limited Memory BFGS Method for Large Scale Optimization," *Math. Programming,* vol. 45, no. 3, pp. 503-528, 1989.

[20] L. Zitnick and T. Kanade, "Maximum Entropy for Collaborative Filtering," *Proc. 20th Conf. Uncertainty in Artificial Intelligence (UAI '04),* pp. 636-643, 2004.

[21] K. Nigam, J. Lafferty, and A. McCallum, "Using Maximum Entropy for Text Classification," *Proc. Int'l Joint Conf. Artificial Intelligence Workshop Machine Learning for Information Filtering,* pp. 61-67, 1999.

[22] A. Ratnaparkhi, "A Maximum Entropy Model for Part-of-Speech Tagging," *Proc. First Conf. Empirical Methods in Natural Language Processing (EMNLP '96),* pp. 133-142, 1996.

[23] S.F. Chen and R. Rosenfeld, "A Gaussian Prior for Smoothing Maximum Entropy Models," CMUCS-99-108, technical report, 1999.

[24] D.R. Cox, "Regression Models and Life-Tables," *J. Royal Statistical Soc., Series B,* vol. 34, no. 2, pp. 187-220, 1972.

[25] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-Based Collaborative Filtering Recommendation Algorithms," *Proc. 10th Int'l World Wide Web Conf. (WWW '01),* pp. 285-295, 2001.

[26] T. Hofmann, "Probabilistic Latent Semantic Analysis," *Proc. 15th Conf. Uncertainty in Artificial Intelligence (UAI '99),* pp. 289-296, 1999.

Tomoharu Iwata received the BS degree in environmental information from Keio University, Kanagawa, Japan, in 2001 and the MS degree in arts and sciences from the University of Tokyo, Japan, in 2003. He is currently a researcher at Emergent Learning and Systems Research Group, NTT Communication Science Laboratories, Kyoto, Japan. His research interests include data mining, machine learning, information visualization, and recommender systems.

Kazumi Saito received the BS degree in mathematics from Keio University, Kanagawa, Japan, in 1985 and the PhD degree in engineering from the University of Tokyo, Japan, in 1998. In 1985, he joined the NTT Electrical Communication Laboratories, Kanagawa. In 2007, he joined the University of Shizuoka, Japan. He is currently a professor in the School of Administration and Informatics. His current research interests include machine learning and statistical analysis of complex networks.

Takeshi Yamada received the BS degree in mathematics from the University of Tokyo in 1988 and the PhD degree in informatics from Kyoto University, Japan, in 2003. He is currently a senior research scientist, supervisor, and leader of Emergent Learning and Systems Research Group, NTT Communication Science Laboratories. His research interests include data mining, statistical machine learning, graph visualization, metaheuristics, and combinatorial optimization. He is a member of the IEEE.