Recommendation Method for Extending Subscription Periods

Tomoharu Iwata, Kazumi Saito, Takeshi Yamada
NTT Communication Science Laboratories
2-4, Hikaridai, Seika-cho, Keihanna Science City
Kyoto 619-0237 Japan
{iwata,saito,yamada}@cslab.kecl.ntt.co.jp

ABSTRACT

Online stores providing subscription services need to extend user subscription periods as long as possible to increase their profits. Conventional recommendation methods recommend items that best coincide with user's interests to maximize the purchase probability, which does not necessarily contribute to extend subscription periods. We present a novel recommendation method for subscription services that maximizes the probability of the subscription period being extended. Our method finds frequent purchase patterns in the long subscription period users, and recommends items for a new user to simulate the found patterns. Using survival analysis techniques, we efficiently extract information from the log data for finding the patterns. Furthermore, we infer user's interests from purchase histories based on maximum entropy models, and use the interests to improve the recommendations. Since a longer subscription period is the result of greater user satisfaction, our method benefits users as well as online stores. We evaluate our method using the real log data of an online cartoon distribution service for cell-phone in Japan.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*; I.2.6 [**Artificial Intelligence**]: Learning; I.5.1 [**Pattern Recognition**]: Model—*Statistical*

General Terms

Algorithms

Keywords

Recommendation, User Purchase Pattern, Subscription Service, Maximum Entropy, Survival Analysis

1. INTRODUCTION

Recommender systems are widely used by online stores [13] because they can improve both user convenience and profits. 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 profits, online stores providing measured services need to 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, the desired user behavior for online stores are different between measured and subscription services.

Recommendation is one way to influence user behavior for online stores. Since the desired user behavior differs for measured and subscription services, online stores should recommend items using different strategies. Conventional recommendation methods recommend items that best coincide with user's interests to maximize the purchase probability [5, 8, 11]. With a measured service, these methods can increase profits. With a subscription service, however, those methods do not necessarily extend subscription periods, and therefore do not necessarily increase profits.

In this paper, we propose a novel recommendation method that maximizes the probability of the subscription period being extended. Our method finds frequent purchase patterns in long subscription period users, and recommends items for a new user to simulate the found patterns. For example, we may find a pattern in long subscription period users, in which they purchase items from a wide variety of genres. Since the difficulties involved in changing user purchase behavior differ depending on user's interests, we take user's interests into consideration so that we can make effective recommendations. To find the patterns, we use Cox proportional hazards models with survival analysis techniques. 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 longer subscription period 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) [2]. CRM is important in terms of improving relationships between online stores and the users.

2. RELATED WORK

A number of recommendation methods have been proposed, such as collaborative filtering [11], content filtering [8], and their hybrids [5]. Collaborative filtering is a method for predicting user's interests using other user's interests. Content filtering predicts interests using item information. These approaches recommend items that best coincide with user's interests to maximize the purchase probability. Our aim, which is to extend subscription periods, is different from theirs.

The prediction of subscription periods and the identification of users that unsubscribe within a period of time, which are known as churn analysis, have been studied by various researchers using survival analysis and data mining techniques [1, 9, 14]. Customer Lifetime Value (LTV) is related to the subscription period. LTV is usually defined as a combination of two components: tenure (or subscription period) and value over time [7]. In subscription services, the value does not change over time because of the fixed charge per given period, and LTV corresponds to the subscription period. Models of LTV have been proposed to facilitate CRM [7]. But they are not used for recommendations.

Piatetsky-Shapiro and Masand [10] and Rosset et al. [12] proposed models for estimating the effects of marketing activity on subscription periods or LTV. A recommendation can be considered a marketing activity. By focusing on the recommendation of items, our method becomes an automatic recommendation framework that can learn from the log data.

3. PROPOSED METHOD

3.1 Preliminaries

The log data usually possessed by online stores providing subscription services are the subscription and the purchase logs. The subscription log consists of the subscribed time, the status (still subscribing or already unsubscribed) and, where relevant, the unsubscribed time of each user. The purchase log consists of the user, the item, and the time of each purchase.

Consider a set of users $\boldsymbol{U} = \{u_n\}_{n=1}^{N}$ and a set of items $\boldsymbol{S} = \{s_i\}_{i=1}^{V}$. Each user u_n has subscription period t_n and status e_n ($e_n = 1$ if unsubscribed, $e_n = 0$ if subscribing). The subscription period t_n is obtained from the subscription log as follows:

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

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 at which the log was modified.

Each user is represented by the sequence of purchased items $u_n = (s_1^n, s_2^n, \ldots)$, where $s_k^n \in \boldsymbol{S}$ is the k-th item in the sequence of user u_n. The sequence is obtained from the purchase log. We have two sets of features of purchase history $x(u_n)$ and $y(u_n, s_j)$, where $x(u_n)$ is used for estimating the user subscription period, and $y(u_n, s_j)$ is used for estimating the user's interests. Examples of features include: whether the user has purchased item s_i, and whether the user has purchased item s_i next to item s_j. We rewrite a set of features $x(u_n)$ in the form of a column vector $\boldsymbol{x}_n = (x_1(u_n), x_2(u_n), \ldots)^T$ for convenient notation, where $x_j(u_n)$ is the j-th feature of $x(u_n)$. We use the subscription periods at which items are purchased for treating purchase histories as time-dependent variables when we estimate the subscription periods.

3.2 Recommendation for Extending Subscription Periods

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

$$\hat{i} = \arg\max_{i \in \boldsymbol{S}} P(l|u, r_i), \tag{2}$$

where l represents an event that the subscription period is extended, and r_i represents an event that item s_i is recommended. We can also recommend m highest $P(l|u, r_i)$ items, if we want to recommend m items. In real applications, candidates for recommendation would be a subset of item set $\boldsymbol{S}$, for example, only items not yet purchased by the user.

Let s_j be the item purchased by user u when item s_i is recommended. If the recommendation does not influence the user purchase behavior, it is natural to think that the recommendation does not influence the subscription period, either. Therefore, we assume that extending subscription period l and recommendation r_i are independent conditioned on purchased item s_j and user u. By using this assumption, $P(l|u, r_i)$ can be decomposed into two components: $Q(l|u, s_j)$, which is the probability of the subscription period being extended when user u purchases item s_j, and $R(s_j|u, r_i)$, which is the probability of item s_j being purchased when item s_i is recommended to user u, as follows:

$$\begin{aligned} P(l|u, r_i) &= \sum_{j \in \boldsymbol{S}} P(l, s_j|u, r_i) \\ &= \sum_{j \in \boldsymbol{S}} Q(l|u, s_j) R(s_j|u, r_i). \end{aligned} \tag{3}$$

We estimate $Q(l|u, s_j)$ and $R(s_j|u, r_i)$ using Cox proportional hazards models and maximum entropy models, respectively. We summarize the framework of our method in Figure 1.

3.3 Cox Proportional Hazards Models

It is generally considered that the subscription period depends on the purchase history. Let $\boldsymbol{x}$ be the feature vector of the purchase history of user u, and we refer to $\boldsymbol{x}$ as the purchase history for the simplicity. We estimate the extension probability given the purchased item $Q(l|u, s_j)$ from the hazard function $h(t|\boldsymbol{x})$, which is the rate that users with purchase history $\boldsymbol{x}$ subscribing until t unsubscribe at t. We estimate $h(t|\boldsymbol{x})$ using survival analysis techniques [4]. If users are still subscribing, we cannot know their true subscription period. These samples are called censored samples. Using survival analysis techniques, we can effectively estimate $h(t|\boldsymbol{x})$ using information found in the data including censored samples.

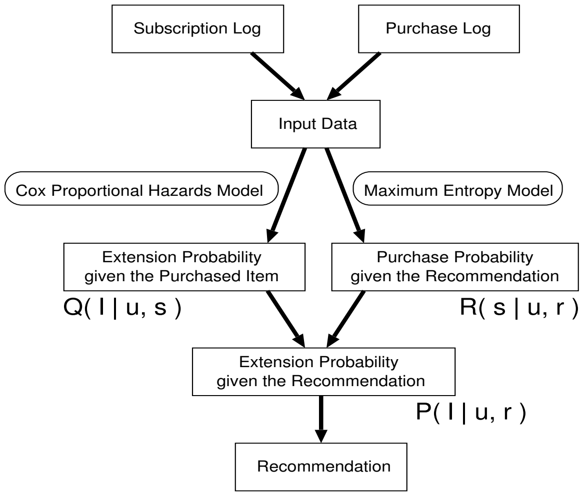


Figure 1: The framework of our method.

We assume Cox proportional hazards models for $h(t|\boldsymbol{x})$ as follows:

$$h(t|\boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{x}), \quad (4)$$

where $\boldsymbol{\beta}$ is the unknown parameter vector, $\boldsymbol{\beta}^T$ is the transpose of $\boldsymbol{\beta}$, and $\lambda_0(t)$ is the baseline hazard function, which represents the hazard function for users with $\boldsymbol{x} = \boldsymbol{0}$. We employ Cox proportional hazards models since the global optimum of the estimation is guaranteed and $Q(l|u, s_j)$ can be written in a closed form as described below.

The log partial likelihood with Breslow approximation [4] is defined as follows:

$$\begin{aligned} PL(\boldsymbol{\beta}) &= \log\prod_t \frac{\prod_{n\in D(t)} h(t|\boldsymbol{x}_n(t))}{(\sum_{m\in R(t)} h(t|\boldsymbol{x}_m(t)))^{|D(t)|}} \\ &= \sum_t \sum_{n\in D(t)} \boldsymbol{\beta}^T\boldsymbol{x}_n(t) \\ &- \sum_t |D(t)| \log \sum_{m\in R(t)} \exp(\boldsymbol{\beta}^T\boldsymbol{x}_m(t)), \quad (5) \end{aligned}$$

where $D(t)$ is the set of users unsubscribed at t and $|D(t)|$ is its size, $R(t)$ is the set of users subscribing at t, and $\boldsymbol{x}_n(t)$ is the feature vector of user u at t. We can estimate unknown parameters $\boldsymbol{\beta}$ by maximizing the log partial likelihood using optimization methods such as quasi-Newton methods [6].

In Cox proportional hazards models, features that have high $\exp(-\beta_b)$ (> 1) are characteristic purchase patterns in the long subscription users, and features that have low $\exp(-\beta_b)$ (< 1) are characteristic patterns in the short subscription users. These patterns are informative for the online store, for example, to understand the relationship between purchase histories and subscription periods, or to determine new items to be distributed for extending subscription periods.

3.4 Extension Probability given the Purchased Item

We explain the estimation of $Q(l|u, s_j)$, which is the probability of extending the subscription period when user u purchases item s_j, using hazard function $h(t|\boldsymbol{x})$. Let $\boldsymbol{x}$ be the purchase history of user u, and $\boldsymbol{x}_{+s_j}$ be the updated purchase history when item s_j is purchased. For the simplicity, we refer to the user when item s_j is purchased as u_{+s_j}. We assume that either one of u or u_{+s_j} unsubscribed at t while the other being still subscribing. At t, the hazard function of u and u_{+s_j} are $h(t|\boldsymbol{x})$ and $h(t|\boldsymbol{x}_{+s_j})$, respectively. The probability that u is the one who unsubscribed at t is equal to the probability of extending the subscription period when user u purchases item s_j as follows:

$$\begin{aligned} Q(l|u, s_j) &= \frac{h(t|\boldsymbol{x})}{h(t|\boldsymbol{x}) + h(t|\boldsymbol{x}_{+s_j})} \\ &= \frac{1}{1+\exp(-\boldsymbol{\beta}^T(\boldsymbol{x} - \boldsymbol{x}_{+s_j}))}, \quad (6) \end{aligned}$$

which is a sigmoid function. We can recommend an item that maximizes $Q(l|u, s_j)$, however, users may not purchase recommended items. In this case, the recommendation is useless as regards extending subscription periods. Therefore, we need to consider whether the recommended item is purchased by the user taking user's interests into consideration.

3.5 Purchase Probability given the Recommendation

We explain the estimation of $R(s_j|u, r_i)$, the probability that user u purchases item s_j when item s_i is recommended to user u . Let $R(s_j|u)$ be the probability that user u purchases item s_j without recommendations, where $\sum_{j=1}^{V} R(s_j|u) = 1$. The recommendation of item s_i increases the probability of item s_i being purchased. We assume that the probability increase γ times as follows:

$$R(s_j|u, r_i) = \begin{cases} \frac{\gamma}{Z(u,r_i)} R(s_i|u) & j = i, \\ \frac{1}{Z(u,r_i)} R(s_j|u) & j \neq i, \end{cases} \quad (7)$$

where $Z(u, r_i) = 1 + (\gamma - 1)R(s_i|u)$ is the normalization term, and $\gamma \geq 1$. γ represents the recommendation effect on user purchase behavior, and depends on the way the recommendation is presented in the online store, such as the size or position.

If an item matches the user's interests, the probability of the user purchasing the item is high, and if it does not match, the probability is low. Therefore, $R(s_j|u)$ represents the degree of coincidence between the interests of user u and item s_j. Since conventional recommendation methods suggest items that coincide with user's interests, we can use conventional methods to obtain $R(s_j|u)$. We employ maximum entropy models [5], 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. In maximum entropy models, the probability that user u purchases item s_j is:

$$R(s_j|u) = \frac{1}{Z(u)}\exp(\sum_c \alpha_c y_c(u, s_j)), \quad (8)$$

where $Z(u) = \sum_{j=1}^{V}\exp(\sum_c \alpha_c y_c(u, s_j))$ is the normalization term, y_c is a feature of the purchase history, α_c is an

Table 1: **The number of subscribers, unsubscribers and features.**

	2005/06/30		2005/07/31		2005/08/31	
	learning	test	learning	test	learning	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
number of features	3,711		4,455		5,250	

Table 2: The average log partial likelihoods of the model that does not use purchase histories, $h(t)$, and the Cox proportional hazards models, $h(t|\boldsymbol{x})$.

	2005/06/30		2005/07/31		2005/08/31	
	learning	test	learning	test	learning	test
model that does not use purchase histories $h(t)$	-8.865	-9.845	-9.165	-9.465	-9.513	-9.904
Cox proportional hazards model $h(t|\boldsymbol{x})$	-8.604	**-9.129**	-9.048	**-9.351**	-9.325	**-9.798**

unknown parameter, and c is an index of each feature.

The log likelihood of the maximum entropy model is:

$$\begin{aligned} L(\alpha) &= \sum_n \sum_k \log R(s_k^n | u_k^n) \\ &= \sum_n \sum_k \sum_c \alpha_c y_c(u_k^n, s_k^n) \\ &- \sum_n \sum_k \log \sum_j \exp(\sum_c \alpha_c y_c(u_k^n, s_j)), \end{aligned} \quad (9)$$

where s_k^n is the k-th item in the purchase sequence of user u_n, and u_k^n is the purchase history of user u_n before the k-th item is purchased. We can estimate unknown parameters α that maximize log likelihood $L(\alpha)$ by using optimization techniques such as quasi-Newton methods. In maximum entropy models, we can obtain the global optimum solution. By using a Gaussian prior with a zero mean on unknown parameters α, overfitting can be reduced [3]. We use a Gaussian prior in our experiments.

4. EXPERIMENTAL RESULTS

We evaluated our method with 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. Our method is applicable since the service employs a subscription log and a purchase log, and extension of the subscription period increases the profit of the online store. 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 a day. This service began on 16 August 2004, and the last modification date of the log was 28 October 2005.

4.1 Evaluation of Cox Proportional Hazards 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 $h(t|\boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x})$ that use purchase histories described in Section 3.3 and models that do not use purchase histories $h(t) = \lambda_0'(t)$. We used as features whether user u has purchased item s_j next to item s_i:

$$x_{s_i \to s_j}(u) = \begin{cases} 1 & \text{if user } u \text{ has purchased} \\ & \text{item } s_j \text{ next to item } s_i, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

where we omitted features that occurred fewer than ten times in the purchase histories.

We used three sets of learning and test samples. The learning samples were the log data up to 30 June 2005, 31 July 2005 and 31 August 2005. The test samples were the log data of subscribers on the end date of the learning samples, and the end date of the test samples was 28 October 2005. The number of subscribers, unsubscribers and features were as shown in Table 1.

For the evaluation measurements, we used the average log partial likelihood, which is the log partial likelihood divided by the number of unsubscribers. The higher average log partial likelihood for the test samples indicates the higher predictive performance of the model. Table 2 shows the results. The average log partial likelihoods for the test samples of the Cox proportional hazards models were higher than 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.

4.2 Evaluation of Maximum Entropy

We evaluated the maximum entropy models described in Section 3.5, which estimate the probability that user u purchases item s_j, $R(s_j|u)$. We used first-order Markov transitions as features, since we considered the last purchased item revealed the user's interests:

$$y_{s_a, s_b}(u, s_j) = \begin{cases} 1 & \text{if item } s_a \text{ is the last purchased} \\ & \text{item of user } u \text{ and } s_b = s_j \text{ ,} \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

We used three sets of learning and test samples. The learning samples 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. The test samples were the log data from the end date of the learning samples to 28 October

Table 3: The number of transitions and items.

	2005/06/30		2005/07/31		2005/08/31	
	learning	test	learning	test	learning	test
number of transitions	300,486	122,904	382,778	171,749	459,456	197,476
number of items	75		81		86	

Table 4: The average log likelihoods of the uniform distribution, the multinomial distribution and the maximum entropy model.

	2005/06/30		2005/07/31		2005/08/31	
	learning	test	learning	test	learning	test
uniform distribution	-4.317	-4.317	-4.394	-4.394	-4.454	-4.454
multinomial distribution	-3.875	-4.263	-3.938	-4.673	-3.975	-4.454
maximum entropy model	-3.554	**-3.551**	-3.581	**-3.732**	-3.605	**-3.762**

2005, from which we omitted transitions to the same item and transitions that contained items that had not been distributed during the learning sample period. The number of transitions and items were as shown in Table 3. We compared maximum entropy models with uniform distributions and multinomial distributions. Uniform distributions do not use the information in the log data at all. Multinomial distributions use the information of the purchased number of each item by all users, but do not consider the individual interests. The unknown parameters of the multinomial distribution were estimated by the maximum likelihood method.

We used the average log likelihood for the evaluation measurements. Table 4 shows the results. The average log likelihoods of the maximum entropy models for the test samples were higher than uniform and multinomial distributions. This means that maximum entropy models can predict user purchase behavior and interest more precisely by using purchase histories.

4.3 Simulation

In Section 4.1, we showed that Cox proportional hazards models can predict subscription periods, and in Section 4.2, we revealed that maximum entropy models can predict user purchase behavior. Here, we examined the effectiveness of our method by simulation. We simulated user behavior using the Cox proportional hazards model and the maximum entropy model that were estimated using the log data from 16 August 2004 to 28 October 2005. The log data comprised 107 items.

Algorithm 1 Simulate a user in a subscription service

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

The function of Algorithm 1 is to generate a subscription period t, where u is the sequence of purchased items, u_{+s_j} is the updated sequence when item s_j is purchased, ϕ is an empty sequence, $Bernoulli(\theta)$ is the Bernoulli distribution with success probability θ, and $Multinomial(\psi)$ is the multinomial distribution of one event with j's success probability ψ_j. First, from line 3 to 4 in Algorithm 1, 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 line 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 $R(s_j)$, where $R(s_j)$ is the probability of purchasing item s_j first (line 10). If the user has purchased some items, we make a recommendation using our method (line 12), and the item that the user purchases is determined according to $R(s_j|u, r_{\hat{i}})$ (line 13). We estimated unknown parameters $\lambda_0(t)$, g and $R(s_j)$ using the log data by the maximum likelihood method.

We compared our method with following recommendation methods:

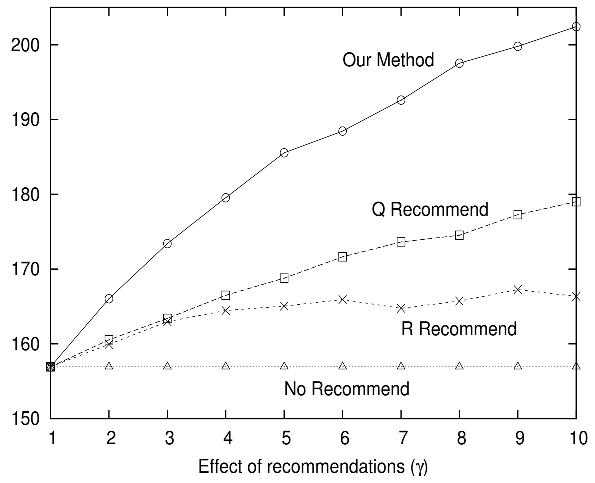
- **Q Recommend** recommends an item that is most likely to extend the subscription period when the user purchases it. Line 12 in Algorithm 1 is changed as follows:

$$\hat{i} \leftarrow \arg\max_i Q(l|u, s_i). \quad (12)$$

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 12 is changed as

Figure 2: The average subscription periods in simulations with parameters estimated from the log data.



follows:

$$\hat{i} \leftarrow \arg\max_i R(s_i|u). \quad (13)$$

This recommendation is the same 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 12 is omitted and line 13 is changed as follows:

$$\text{Sample } s_j \sim Multinomial(R(s_j|u)). \quad (14)$$

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

We generated $100,000$ user subscription periods with recommendations by each method in $1 \leq \gamma \leq 10$. We set the maximum subscription period as 365 days. Figure 2 shows the average subscription periods. Our method was more successful than the others in extending subscription periods. The subscription periods became longer with increases in γ. This result indicates that if recommendations can influence on user behavior, or $\gamma > 1$, our method can extend subscription periods, and subscription periods can be extended even further by improving the influence of the recommendations. Q Recommend extended subscription periods, although the effect was smaller than with our method. This is because Q Recommend may recommend items that have no probability of being purchased by the user. On the other hand, our method suggests items taking user's interests into account to improve the recommendations. R Recommend only slightly extended subscription periods because of little correlation between the subscription periods and the user's interests.

5. CONCLUSIONS

In this paper, we proposed a novel recommendation method that encourages users to extend their subscription periods for online stores providing subscription service. 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 long distance dependencies or user attributes. Since our method is divided into two modules: the estimation of subscription periods and the estimation of user's interests, it can be further enhanced using survival analysis or collaborative filtering techniques. For example, we can combine our approach with content filtering to estimate user's interests. Online stores providing measured services are eager to prevent their users from switching to other stores. Therefore, our method can also be used for recommendations in measured services by considering the subscription period as a period of continuous use of the online store.

Although we have obtained encouraging results to date, 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 Cox proportional hazards model and the feature selection for finding informative purchase patterns in long subscription period users. Second, we need to estimate the influence of recommendations on user purchase behavior from the log data automatically. This can be estimated 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 extend the subscription periods of real users.

6. REFERENCES

[1] 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*, 7(6):532–545, 2003.

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

[3] S. F. Chen and R. Rosenfeld. A Gaussian prior for smoothing maximum entropy models. Technical report, CMUCS–99–108, 1999.

[4] M. Cleves, W. W. Gould, and R. Gutierrez. *An Introduction to Survival Analysis Using Stata, Revised Edition*. Stata Press, 2004.

[5] X. Jin, Y. Zhou, and B. Mobasher. A maximum entropy web recommendation system: combining collaborative and content features. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2005.

[6] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Math. Programming*, 45(3):503–528, 1989.

[7] D. R. Mani, J. Drew, A. Betz, and P. Datta. Statistics and data mining techniques for lifetime value modeling. In *Proceedings of the Fifth ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 94–103, 1999.

[8] R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, pages 195–204, 2000.

[9] 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*, 11(3):690–696, 2000.

[10] G. Piatetsky-Shapiro and B. Masand. Estimating campaign benefits and modeling lift. In *Proceedings of the Fifth ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 185–193, 1999.

[11] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 175–186, 1994.

[12] S. Rosset, E. Neumann, U. Eick, and N. Vatnik. Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery*, 7:321–339, 2003.

[13] J. B. Schafer, J. A. Konstan, and J. Riedl. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5:115–153, 2001.

[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. In *Proceedings of IEEE Conference on Computational Cybernetics*, pages 279–284, 2004.