Modeling Individual Users' Responsiveness to Maximize Recommendation Impact

要　旨

【キーワード】
推薦システム、パーソナライゼーション、購買予測、最適化、機械学習

推薦システムは、クリック履歴や購買履歴などから推定されるユーザーの嗜好に応じて情報を提供する。このとき、好みの商品が推薦されればユーザーは反応するという前提が一般的であるが、実際の反応性はユーザーのタイプによって異なる。富士ゼロックスは、個々の反応性の違いを考慮した購買予測モデルを考案し、食品日用品に関し、購買の予測精度と推薦効果が改善することを確認した。推薦履歴が少ない場合、反応性は他の情報源から推測する必要がある。そのため、我々は、ユーザーや商品の特徴と、推薦に対する反応性との相関を調べた。相関のあった特徴をもとに推薦した場合も、十分な推薦履歴をもとに推薦した場合とほぼ同等の推薦効果が得られることがわかった。これは推薦履歴が少ない場合にも我々の手法が有効であることを意味している。本研究は、推薦に対する反応性の違いに基づく新たなパーソナライズの方向性を示唆している。

Abstract

【Keywords】
recommender systems, personalization, purchase prediction, optimization, machine learning

Recommender systems provide personalized information based on a user's preferences. Differences in preferences among users are estimated from past records such as click logs or purchase logs. Recommender systems typically assume that users will respond to recommendations, provided that their favorite items are correctly selected. However, the responsiveness to recommendations depends on the type of users; while some users might be easily persuaded to take action, others might be more hesitant. In this paper, we propose a purchase prediction model that incorporates the differences in the responsiveness. Improvement in purchase prediction and recommendation impact, which is defined as the increase in purchase probability through recommendations, was verified using a grocery shopping dataset. These results demonstrate the importance of modeling the responsiveness of individual users. In cases where recommendation logs are insufficient, the responsiveness needs to be estimated from other sources. Consequently, we investigated the correlation of the responsiveness with user attributes and item attributes. The recommendation impact of the model estimated from the correlated attributes was almost comparable to that of the model estimated from recommendation logs. These findings can help overcome the cold-start problem of inadequate recommendation logs. Our study presents a new direction in the field of personalization based on the responsiveness to recommendations.

Author

Masahiro Sato[*1]
Hidetaka Izumo[*2]
Takashi Sonoda[*1]

[*1] Communication Technology, Research & Technology Group
[*2] Fuji Xerox Information Systems Co., Ltd.

1. Introduction

Recommender systems are prevalent in many fields. Electronic commerce websites display items that users might like to buy, and social networking services find people whom users might know and want to connect with. Recommendation research has attracted the interest of both academics and practitioners.

Much effort has been dedicated to algorithms that estimate individual users' preferences. Preferences can be extracted from past records of explicit feedbacks such as five-point ratings and implicit feedbacks such as click logs and purchase logs. Traditionally, recommendation research has focused on the personal differences in item preferences. On the other hand, it has been indifferent to other personal differences.

Recently, however, new kinds of personal differences have drawn the attention of researchers. For example, the propensity to diversity depends on personality [23, 24]. Novelty-seeking behavior differs among users [12, 25]. Some users accept higher risks for higher returns from recommendations, while others avoid such risks [26]. Moreover, the change rate in preferences over time is unique to each user [19].

In addition to these personal differences, the responsiveness to recommendations, which is defined as the effect of recommendations to increase a user's rating of an item, might also depend on the user. However, responsiveness has been treated as independent of the users [3, 10, 22] and the individual differences have never been investigated, to the best of our knowledge. The responsiveness to recommendations is directly connected to the success of recommendation and requires further investigation in order to design better recommender systems.

Along with the dependence of the responsiveness on users, individual items might trigger different responses in users. This possibility is implied by in-situ experiments in real stores [14, 17]. Researchers have demonstrated that while certain items in some categories sell easily through recommendations, other items in other categories do not.

In this paper, we propose a recommender system that incorporates individual differences in responsiveness. We formulate the purchase probability as a sigmoid function of the sum of a rating and recommendation response. Recommendation responsiveness is decomposed into common responsiveness, user-specific responsiveness, and item-specific responsiveness. The responsiveness is inferred from a combination of purchase logs and recommendation logs, so as to maximize the likelihood of the model.

We evaluated the effectiveness of our model in terms of purchase prediction and impact maximization, using a grocery shopping dataset. The accuracy of the predictions made by our model was compared to the accuracy of a conventional model that assumes constant responsiveness. The recommendation impact, which we define as the increase in purchase probability as a result of recommendations, was also compared between the proposed model and the conventional model.

In order to clarify the characteristics of responsiveness and estimate responsiveness despite inadequate recommendation logs, we investigated the correlation between responsiveness and the other attributes of users and items. For the analysis, we used demographic information about users and features extracted from purchase records. The correlated features were then applied to predict user- and item-specific responsiveness. Furthermore, we evaluated the recommendation impacts of an individualized responsiveness model, estimated exclusively from the correlated features, without using recommendation logs.

The outline for this paper is as follows. In the next section, we review related work. In Section 3, we introduce our model, along with a conventional model and the dataset we used for our evaluation. Section 4 presents performance comparisons between our model and the conventional model in terms of prediction accuracy and recommendation impact. Section 5 describes the correlation of recommendation responsiveness to other attributes and the estimate of responsiveness from correlated attributes. Finally, we summarize and conclude this paper in Section 6.

2. Related Work

There are two branches of research that relate closely to this work; meta-personalization beyond item preference, and purchase prediction of recommended items.

2.1 Meta-Personalization

Accurately predicting item preferences does not in itself lead to user satisfaction [16]. Consequently, there is a discrepancy between online performance and offline performance [5]. New perspectives have thus been introduced to recommender systems. For instance, diversity and novelty are vogue topics in recommendation research [4].

As research into diversity and novelty progresses, it is becoming apparent that the desired degree of diversity and novelty differs among users. The propensity to diversity has been measured in terms of the entropy in item selection, and the diversity of recommendations for each user can be adjusted accordingly [6]. Indeed, the preference for diversity is correlated to personality [24], and in particular to "openness to experience" [23]. Recommender systems that adapt to the novelty-seeking traits of users have also been proposed [12, 25].

Such meta-personalization is not limited to diversity and novelty. Individual differences in risk tolerance have been introduced to recommender systems in order to adjust the allowable degree of the variance in rating estimates for each user [26]. Dynamics of preference, or fickleness, also differ among users, and this has been taken into account for recommendations [19].

We believe that our work in modeling individual users' responsiveness will shed new light on the field of meta-personalization.

2.2 Recommended Purchase Prediction

A conventional task of recommender systems with implicit feedback is to predict which items users will click or buy [9, 11, 18]. However, such predictions do not always consider the effect of recommendations.

Recommendation naturally increases the probability that an item will be clicked or purchased. Recently, the effect of recommendations on purchase predictions has been modeled in several ways. Shani et al. [22] assumed that the increase in purchase probability from recommendations is proportional to the purchase probability without recommendations. Jianga et al. [10] imposed the constraint that consumers buy an item only if the valuation is more than the price of the item. They assumed that recommendation increases the valuation of the item and that the increase is constant. Bodapati [3] decomposed purchase probability into awareness probability and satisfaction probability, and assumed that recommendations guaranteed awareness of the item.

Whereas the responsiveness to recommendations was considered to be independent of the user in the previous work, we introduce user-dependent responsiveness, in an effort to advance purchase prediction a step further.

3. Individualized Responsiveness

In this section, we first describe a base model for purchase prediction. We next explain the dataset we used in Subsection 3.2. Subsection 3.3 shows our preliminary experiment, which indicates individual differences in recommendation responsiveness. Finally, we introduce user- and item-specific responsiveness in Subsection 3.4.

3.1 Base Model for Purchase Prediction

The probability of binary implicit feedback, such as clicks or purchases, can be formalized in a sigmoid function of a rating of user u on item i (r_{ui}) [11]:

$$p = \sigma(r_{ui}) = \frac{1}{1+\exp(-r_{ui})}. \quad (1)$$

We used matrix factorization, which is known to perform well in rating prediction [13]. Matrix factorization decomposes a rating to the latent factors of the user and the item. Adding bias terms is a common technique, because ratings are not zero-centered. Hence rating r_{ui} is expressed as:

$$r_{ui} = b_c + b_u + b_i + \theta_u{}^{\mathrm{T}}\varphi_i, \quad (2)$$

where b_c is a bias common to all the users and items, b_u is a user-specific bias, and b_i is an item-specific bias. Further, θ_u and φ_i denote the latent factors of the user and the item, respectively. Equation (2) is expressed equivalently in matrix form as follows:

$$\begin{gathered} \mathbf{R} = \mathbf{B} + \Theta^{\mathrm{T}}\Phi, \\ \{\mathbf{R}\}_{ui} = r_{ui},\ \{\mathbf{B}\}_{ui} = b_c + b_u + b_i,\ \{\Theta\}_{u*} = \theta_u,\ \{\Phi\}_{i*} = \varphi_i. \end{gathered} \quad (3)$$

Recommending an item should increase the probability of purchasing the item. Adding recommendation responsiveness γ to rating r_{ui}, Equation (1) becomes

$$p = \frac{1}{1+\exp(-(b_c + b_u + b_i + \theta_u{}^{\mathrm{T}}\varphi_i + \delta_{\mathrm{rec}}\gamma))}, \quad (4)$$

where δ_{rec} is an indicator function of the recommendation. Here, $\delta_{\mathrm{rec}} = 1$ when item i is recommended to user u; otherwise, $\delta_{\mathrm{rec}} = 0$. Furthermore, γ can be constant or dependent on the user and the item, as discussed below in Subsection 3.4.

In our models, the parameters $\wp$ to be learned are:

$$\wp = \{b_c, b_u, b_i, \theta_u, \varphi_i, \gamma \mid u \in U, i \in I\}. \quad (5)$$

From purchase records and recommendation records, each term is determined such that it minimizes the negative log likelihood (NLL):

$$\begin{aligned} \mathrm{NLL} &= -\ln\Big(\prod_{\text{purchase}} \sigma(r_{ui} + \delta_{\mathrm{rec}}\gamma) \\ &\quad \times \prod_{\text{non-purchase}} (1 - \sigma(r_{ui} + \delta_{\mathrm{rec}}\gamma))\Big) \\ &= \sum_{\text{purchase}} \ln(1 + \exp(-(r_{ui} + \delta_{\mathrm{rec}}\gamma))) \\ &\quad + \sum_{\text{non-purchase}} \ln(1 + \exp(+(r_{ui} + \delta_{\mathrm{rec}}\gamma))). \end{aligned} \quad (6)$$

We define each term in the summation of purchase records and the summation of non-purchase records as l_{ui}^{purchase} and , $l_{ui}^{\text{non-purchase}}$ respectively.

$$l_{ui}^{\text{purchase}} \equiv \ln(1 + \exp(-(r_{ui} + \delta_{\mathrm{rec}}\gamma))), \quad (7)$$

$$l_{ui}^{\text{non-purchase}} \equiv \ln(1 + \exp(+(r_{ui} + \delta_{\mathrm{rec}}\gamma))). \quad (8)$$

We used a stochastic gradient descent (SGD) method for iterative learning. For each iteration, an SGD randomly picks a user-item pair and updates the parameters in the opposite direction of the gradient. The differentials of l_{ui}^{purchase} and $l_{ui}^{\text{non-purchase}}$ are:

$$\begin{aligned} &\frac{\partial}{\partial \wp} l_{ui}^{\text{purchase}} \\ &= -\left(\frac{1}{1+\exp(r_{ui} + \delta_{\mathrm{rec}}\gamma)}\right)\frac{\partial}{\partial \wp}(r_{ui} + \delta_{\mathrm{rec}}\gamma), \end{aligned} \quad (9)$$

$$\frac{\partial}{\partial \wp} l_{ui}^{\text{non-purchase}} = \left(\frac{1}{1+\exp(-(r_{ui}+\delta_{\text{rec}}\gamma))} \right) \frac{\partial}{\partial \wp}(r_{ui}+\delta_{\text{rec}}\gamma). \quad (10)$$

Parameters are updated as:

$$\wp \leftarrow \wp + \zeta_{\wp}\left(-\frac{\partial}{\partial \wp} l_{ui}^{\text{purchase}} - \lambda_{\wp}\wp \right), \quad (11)$$

$$\wp \leftarrow \wp + \zeta_{\wp}\left(-\frac{\partial}{\partial \wp} l_{ui}^{\text{non-purchase}} - \lambda_{\wp}\wp \right), \quad (12)$$

where $\zeta_{\wp}$ is the learning rate and $\lambda_{\wp}$ is the regularization coefficient of the parameter. Learning the parameters of the model with the SGD always converged in our experiments.

3.2 Dataset

We used proprietary data from a grocery shop. The dataset included purchase logs and recommendation logs. We could not find any publically available open data with recommendation logs, which are crucial for our experiments. Hence, we used only this dataset. The grocery shop mainly deals in foods like vegetables, meat, fish, and various processed foodstuffs. The club members of the shop received the catalogs of available items each week and made purchases by mail order. For each week, several "recommended items of the week" were selected by the shop owner. The recommended items were selected from diverse categories of foods in the shop. Flyers with one of the items printed were bundled with the catalog and posted for the club members over ten weeks. The members targeted for recommendation were chosen randomly each week. Members received at most one flyer per week and a flyer recommended only one item. Table 1 summarizes the dataset.

From the purchase records, we created non-purchase records, which are user-item pairs comprising users who use the shop on a certain week and the items that they do not purchase despite their availability in that week's catalog. The shop changes the merchandize assortment weekly. We generated 155,236,964 non-purchase records. Both purchase records and non-purchase records are necessary in order to evaluate the purchase probability of items.

We merged purchase records, non-purchase records, and recommendation records each week. We assumed that the influence of recommendation continued for a week, because the flyers showcased "recommended items of the week" and the merchandize assortment changed each week. Table 2 shows examples of the merged dataset. Recommended items differ depending on the week and the user. For example, Item 1 might be available on Week 1 but not on Week 2. Moreover, the same user can repeatedly buy the same item; in this example, User 1 buys Item 2 on both Week 1 and Week 2.

Table 1. Summary of the dataset

Type	#records	#users	#items	#weeks
Purchase	3,743,300	6,937	4,150	39
Recommend	30,174	6,897	36	10

Table 2. Sampling examples of the merged

User ID	Item ID	Week ID	Purchase?	Recommend?
1	1	1	True	True
1	2	1	True	False
2	1	1	True	False
2	2	1	False	True
1	2	2	True	True
1	3	2	False	False

3.3 Preliminary Experiment

We first learned the components of the rating, $b_c, b_u, b_i, \theta_u, \varphi_i$, from data without recommendations (Recommend? = False), so as to minimize the NLL. We reserved 10 % of the data for validation and tuned hyper-parameters such as the learning rate and regularization coefficient with the validation data. We adjusted the matrix dimensions from 10 to 1000, and the improvement of the NLL saturated at 300 dimensions. Hence, we set the matrix dimensions to 300.

We next investigated the relationship between the predicted purchase probability without recommendations and the observed purchase probability with recommendations. For all user-item pairs in the recommendation logs (Recommend? = True), we calculated the purchase probability without including the recommendation responsiveness γ. Then, we clustered user-item pairs according to the similarity of the probability. We averaged the estimated probability for each cluster. Finally, we calculated the observed purchase probability with recommendations for each cluster, which is defined as:

$$\frac{\textit{the number of purchases in a cluster}}{\textit{the size of the cluster}}. \quad (13)$$

Figure 1 shows the results. The x-axis and the y-axis represent the estimated purchase probability without recommendations and the observed probability with recommendations, respectively. If recommendations do not influence purchase probability, both the probabilities should be the same, i.e., y = x, as represented by the dotted line in Figure 1. The solid line represents the moving average. Here, the solid line is above the dotted line, meaning that recommendations boost the purchase probability. While the probability without recommendations is merely an estimate, we confirmed that the prediction is fairly accurate (the average NLL, defined later in Equation (15), was 0.032 for data without recommendations). In addition, we can assume that the prediction error is unbiased, and averaging within the clusters should decrease the error.

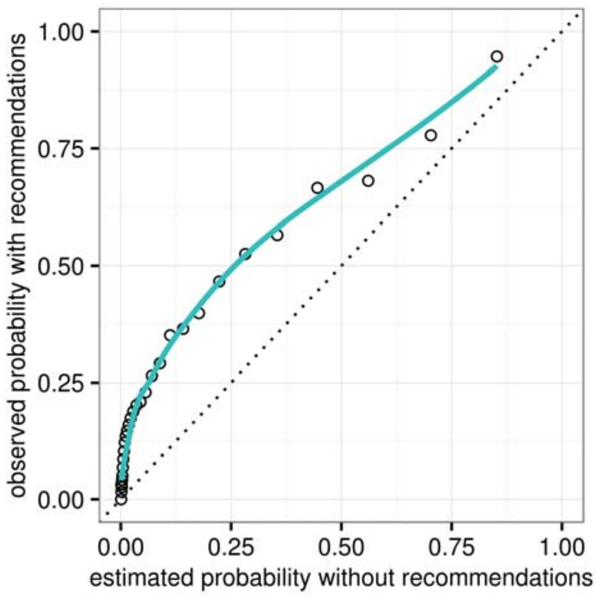

Figure 1. Purchase probabilities with and without recommendations

We compared the increase in purchase probability among users of different ages. While personality information regarding the users was unavailable to us, it is known that some personality traits are correlated with age. For example, age has positive correlations with agreeableness and conscientiousness, and negative correlations with neuroticism, extraversion, and openness [15]. Conscientious people might notice recommendations more often than others, and agreeable people might accept recommendations relatively easily. Hence, we expected that the effect of recommendations might depend on age. We split the cluster of user-item pairs according to age,

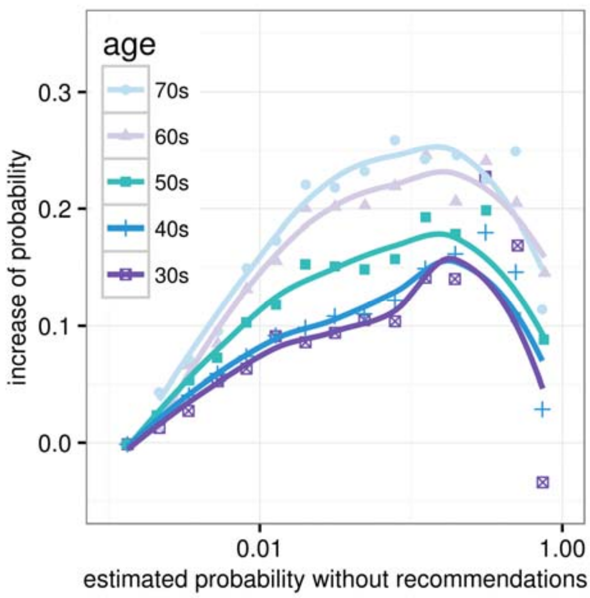

Figure 2. Increase in purchase probability from recommendations for various ages
（The x-axis is a log scale）

by grouping users in their 30s, 40s, 50s, 60s, and 70s. Figure 2 illustrates the difference in the probability increase. Indeed, the increase becomes more significant with advancing age. This result implies that the responsiveness to recommendations depends on the type of user. This supplies the motivation for personalizing responsiveness.

3.4 Individualized Responsiveness

From the observations in Subsection 3.3, we hypothesized that the responsiveness to recommendations differs for each user and each item. Whether a user accepts a recommendation might depend on his or her personality, e.g., the user's agreeableness. In addition, some items might induce impulse shopping, whereas others might entail more deliberation.

We split recommendation responsiveness γ into a common term γ_c, a user-specific term γ_u, and an item-specific term γ_i:

$$\gamma = \gamma_c + \gamma_u + \gamma_i \tag{14}$$

These terms can be obtained through SGD using the purchase logs and the non-purchase logs with recommendations (Recommend? = True in Table 2). We expect that this formulation will explain the observed differences in Subsection 3.3.

4. Comparative Evaluation

In this section, we evaluate the effect of individualizing recommendation responsiveness. We measured the accuracy of the purchase predictions and the impact of the recommendations. The effectiveness of the model was examined by comparing it with a conventional model, in which responsiveness is constant for all users and items.

4.1 Accuracy Comparison

We compared the accuracy of purchase prediction in terms of NLL and precision. We calculated the NLL for each user-item pair in the testing data and took the average of them.

$$l_{ui}^{\text{ave}} = \frac{\sum_{\text{purchase}} l_{ui}^{\text{purchase}} + \sum_{\text{non-purchase}} l_{ui}^{\text{non-purchase}}}{(\textit{the number of the test data})}. \tag{15}$$

The precision was calculated for user-item pairs of the top *n* % in purchase probability:

$$\text{Precision} = \frac{\textit{the number of purchase within top } n\%}{\textit{the number of } u-i \textit{ pairs within top } n\%}. \tag{16}$$

In the dataset, 27.1% of all the recommendations were purchased; the baseline for the precision obtained by random recommendation was thus 0.271. We set *n* = 27.1%, because precision and recall are the same at this threshold, and this facilitates the comparison.

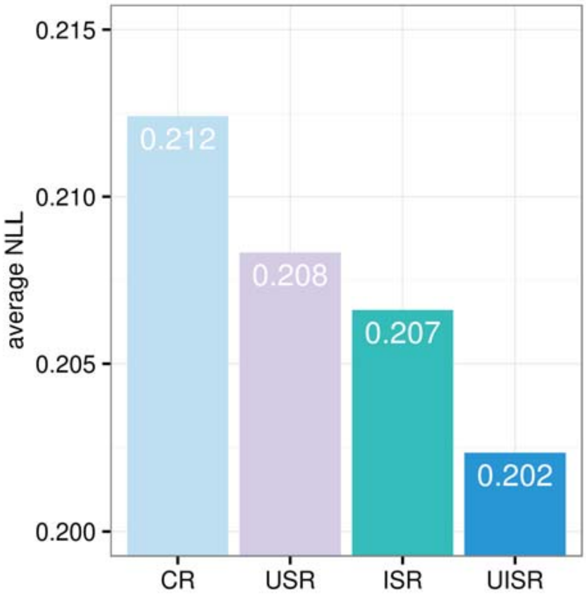

Figure 3. Comparison of the average NLL

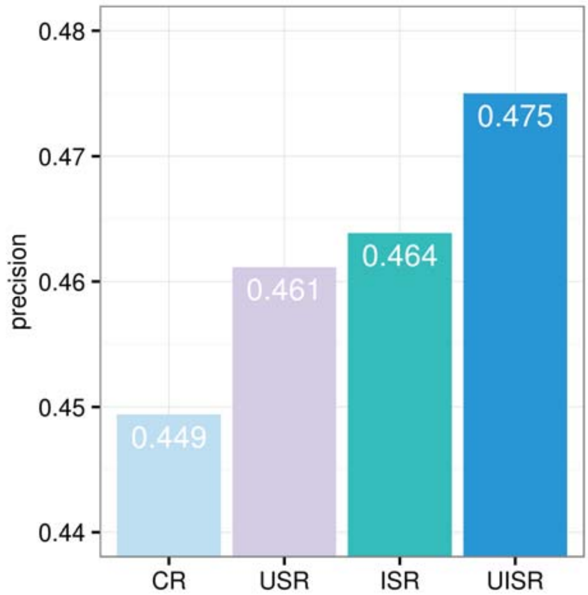

Figure 4. Comparison of precision

We compared four models: constant responsiveness ($\gamma = \gamma_c$, CR), user-specific responsiveness ($\gamma = \gamma_c + \gamma_u$, USR), item-specific responsiveness ($\gamma = \gamma_c + \gamma_i$, ISR), and user- and item-specific responsiveness ($\gamma = \gamma_c + \gamma_u + \gamma_i$, UISR). After pre-training of the components of the rating, $b_c, b_u, b_i, \theta_u, \varphi_i$, from the data without recommendations (Recommend? = False), γ for each model was trained using data with recommendations (Recommend? = True). We performed ten-fold cross validation on each model, and calculated the average of the results obtained.

Figure 3 shows a comparison of the mean NLL (l_{ui}^{ave}), and Figure 4 shows a comparison of the precision. UISR clearly outperformed CR with both metrics. Both user- and item-specific terms improved the accuracy and combining them further improved it.

We also confirmed the significance of the results. The Wilcoxon signed rank test was performed for CR vs. USR/ISR and USR/ISR vs. UISR in terms of both NLL and precision. All of the differences were significant; with p-value = 0.014 for CR vs. USR in precision and p-value < 0.007 for the other comparisons. These results demonstrate the effectiveness of modeling user- and item-specific responsiveness for accurate purchase predictions.

4.2 Impact Maximization

We next evaluated the recommendation impact, which is defined as the increase in purchase probability through recommendations. We believe this is an important evaluation metric, despite the fact that it is uncommon in the field of recommendation research.

Traditional recommender systems are designed to predict whether a user will purchase an item, regardless of whether it is recommended. They then recommend the item with the highest purchase probability. These systems adopt the tacit assumption that there is a positive correlation between the increase in purchase probability from recommendations and the purchase probability without recommendations:

$$(p(\delta_{\mathrm{rec}} = 1) - p(\delta_{\mathrm{rec}} = 0)) \propto p(\delta_{\mathrm{rec}} = 0) \quad (17)$$

However, this assumption is not necessarily true. Consider an extreme example where an item is recommended to a user who has already decided to buy the item in spirit; the purchase probability without a recommendation is almost 100% in this case, and there is no space for a recommendation to further increase this probability. This corresponds to $x \approx 1$ in Figure 1. On the other hand, recommending an item that a user has no intention of buying will not affect the purchase probability either. This corresponds to $x \approx 0$ in Figure 1. As can be seen in Figures 1 and 2, the increase in purchase probability from recommendations is a convex function of the purchase probability without recommendations. We argue that the convexity is universal in any recommendation domains based on the above observations, whereas peak positions might be domain-dependent. Recommending items that are most likely to be purchased without a recommendation is not an optimum strategy.

Recommender systems can be designed for various objectives [20]; End-users might want to maximize utility surplus, which is defined as item utility minus price [8], and maximizing profit is a major concern for retailers [2]. Maximizing recommendation impact can be seen as another form of maximizing the utility surplus or the profit. However, our definition of recommendation impact aims to evaluate the net influence of recommendations.

In order to calculate the recommendation impact, the purchase probability is needed both with and without recommendations. Although we cannot know their exact values, our model is capable of estimating them. Their difference yields the impact of each recommendation. Summing this impact is equivalent to the expected value of the increase in sales volume through recommendations. Hence, maximizing impact leads directly to profit maximization when commercial goods are recommended for purchase.

We compared the recommendation impacts obtained with two strategies: 1) the strategy used by traditional systems that recommend items that have the highest purchase probability without recommendations (HP); and 2) recommending items that will result in the largest increase in probability through recommendations (LI). For the latter strategy, we tested four models introduced in Subsection 4.2: the CR, USR, ISR, and UISR models (LI-CR, LI-USR, LI-ISR, and LI-UISR, respectively). We acquired recommendation logs for 6897 users and 36 items, and there are 248,292 possible user-item pairs. We selected the best m pairs (the highest probability for Strategy 1 and the largest increase for Strategy 2) from the possible combination, and calculated the average impact. We set $m = 3017$, which is the average number of recommendations per week in our dataset. We used the UISR model to estimate the impacts, because it is the most accurate.

The results are presented in Figure 5. LI-CR outperformed HP, proving that maximizing the increase in probability is a superior strategy. Furthermore, LI-UISR had more of an impact than LI-CR. Both LI-USR and LI-ISR surpassed LI-CR, meaning that both user- and item-specific responsiveness contribute for improvement. This result demonstrates the importance of individualized responsiveness for maximizing recommendation impact.

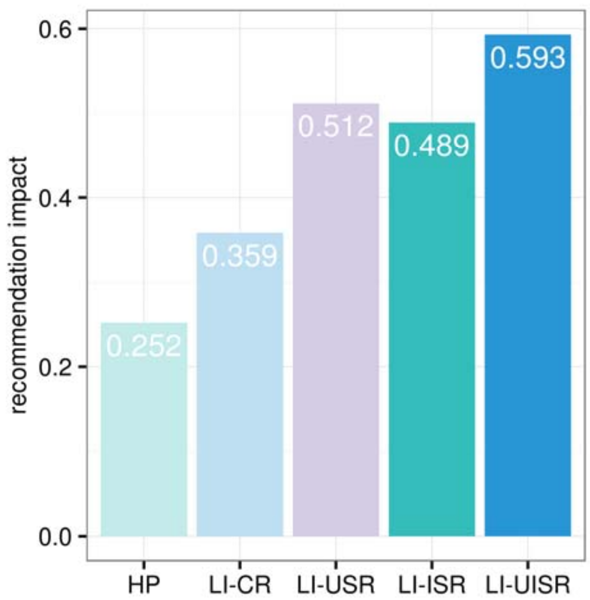

Figure 5. Comparison of recommendation impacts

5. Responsiveness Estimation

In order to clarify the characteristics of responsiveness and enable estimations of it despite insufficient recommendation logs, we investigated the correlation between responsiveness and other user and item attributes. This investigation is described in Subsection 5.1. We also evaluated the prediction accuracy of responsiveness from the correlated attributes, the results for which are described in Subsection 5.2.

5.1 Correlation Analysis

Understanding the correlation between recommendation responsiveness and user and item attributes can lead to reveal the origin of personal differences in recommendation responsiveness. We analyzed the demographic information of users and features derived from purchase records.

The demographic features available comprised age and family size. Among these features, only age was correlated significantly with user-specific recommendation responsiveness. The line shows a linear regression. A positive correlation was found, meaning that elderly people are more easily persuaded to buy an item. It is known that age is positively correlated with agreeableness and conscientiousness [15], and this result might originate from the positive correlations of user-specific responsiveness to agreeableness and conscientiousness.

Some users buy many items at once, while others buy a few at a time. We define the mean basket size of each user as the average number of items purchased at one time. The mean basket size was negatively correlated with recommendation responsiveness. This result suggests that bulk buyers tend to be indifferent to recommendations.

Regarding item-specific responsiveness, we examined the relationship between the number of weeks an item was displayed (η_{fam}) and the number of purchases per week (η_{pop}). We found that item-specific responsiveness increases the more time an item is displayed and decreases with the number of weekly purchases. The number of weeks an item is displayed is related to its familiarity to the user, and the number of weekly purchases tracks the popularity of the item. Hence these results suggest that familiar yet unpopular items are good candidates for recommendations.

5.2 Estimating Individual Responsiveness

Predicting user- and item-specific recommendation responsiveness is important when the recommendation logs are insufficient. Retail shops often keep purchase logs, but they rarely keep recommendation logs. Even when recommendation logs are properly recorded, we do not know the responsiveness when we first making recommendations to a user or when recommending a particular item for the first time.

The situation above resembles a situation, in which purchase logs of new users or new items are insufficient for extracting preferences. This problem is known as a cold-start problem in recommender systems [1, 21]. In our case, purchase logs are abundant, but recommendation logs are inadequate. This is a new form of the cold-start problem with our model.

In order to overcome the cold-start problem, we estimate the responsiveness from other sources. We built a linear regression model to predict individual responsiveness. Owing to the correlation analysis conducted in Subsection 5.1, effective predictors are already known. Thus, user-specific responsiveness can be predicted merely from the age and the mean basket size:

$$\gamma_u = a_1 \cdot \eta_{\text{age}} + a_2 \cdot \eta_{\text{bas}} + a_3 \tag{18}$$

and item-specific responsiveness can be predicted from the familiarity and the popularity:

$$\gamma_i = b_1 \cdot \eta_{\text{fam}} + b_2 \cdot \eta_{\text{pop}} + b_3 \tag{19}$$

The coefficients obtained were, $a_1 = 0.0052$, $a_2 = -0.0080$, $b_1 = 0.0047$, and $b_2 = -0.00055$. We confirmed that all of the coefficients are statistically significant ($p < 0.01$).

We evaluated the predictive performance with twelve-fold cross validation. We chose twelve-fold instead of ten-fold cross validation, because there are 36 items with item-specific responsiveness and 36 is divisible by 12. Figure 6 shows the root mean square errors (RMSEs) from predicting user-specific responsiveness ("User" in the figure) and item-specific responsiveness ("Item" in the figure). The accuracy of the linear regression model (LR) was compared to the accuracy of the mean estimate (Mean). The linear regression outperformed the mean estimate when predicting both user- and item-specific responsiveness. Indeed, item-specific responsiveness improved relatively more than user-specific responsiveness. However, both results were statistically significant (the p-value from the paired Wilcoxon signed rank tests was 2.6×10^{-6} for user-specific responsiveness and 0.047 for item-specific responsiveness). The individual responsiveness can be estimated at some level merely from the demographic information and purchase logs.

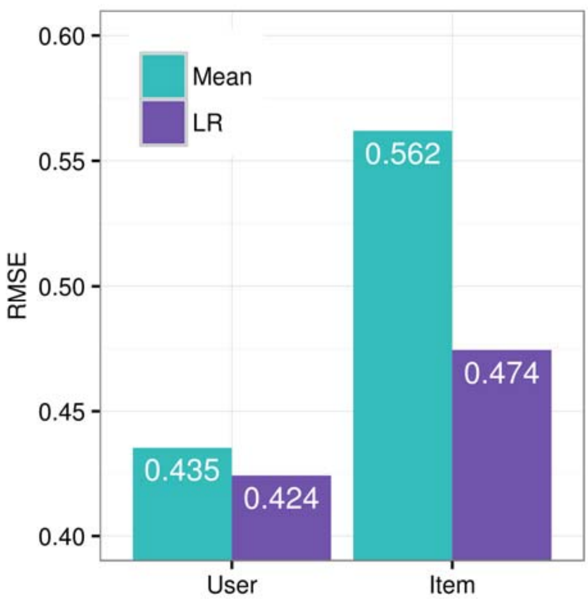

Figure 6. Predictive performance of user- and item-specific responsiveness: comparing mean estimates with linear regression estimates

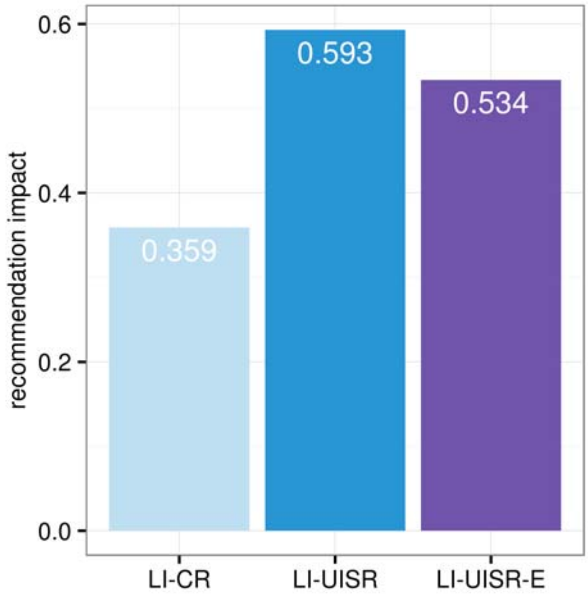

Figure 7. Comparison of recommendation impacts

Finally, we evaluated the recommendation impact obtained from the estimated responsiveness. We used the user- and item-specific responsiveness estimated from Equations (18) and (19) for the UISR model (UISR-E), and recommended an item with the largest increase in purchase probability for each user (Strategy 2 in Subsection 4.2, LI). Figure 7 shows the comparison of the impact among LI-CR, LI-UISR, and LI-UISR-E. Note that the results for the LI-CR and LI-UISR models are the same as the results in Figure 5. They are again provided in order to facilitate the comparison. LI-UISR-E exceeded LI-CR with the statistical significance ($p < 2.2 \times 10^{-16}$ by the Wilcoxon signed rank test). LI-UISR was superior to LI-UISR-E, and learning responsiveness directly from recommendation logs is desirable, where available. However, LI-UISR-E closely aligned with LI-UISR. This result shows the potential applicability of our model, despite inadequate recommendation logs.

6. Summary and Future Work

In this paper, we proposed a purchase prediction model that incorporates individual differences in recommendation responsiveness. Our model improved the accuracy of purchase prediction and the impact of recommendations. These results confirmed the importance of modeling individualized responsiveness. We found a correlation between user-specific responsiveness and both age and the mean basket size. We also found a correlation between item-specific responsiveness and both familiarity and popularity. The estimated responsiveness from the correlated attributes outperformed the mean estimates. We further confirmed that the recommendation impact of the user- and item-specific responsiveness model estimated from the correlated attributes exceeds the impact of the constant-responsiveness model. These findings demonstrate the applicability of our model, even when there are insufficient recommendation logs. This work offers a new research direction in personalizing recommender systems based on recommendation responsiveness.

In future work, we shall compare our impact-maximization approach with other approaches, such as diversity- and novelty-seeking approaches. This comparison would be helpful for uncovering the best recommendation tactics. Whereas we applied a sigmoid function to convert ratings into purchase probabilities, other methods are available, such as Poisson distribution [7]. We plan to evaluate these methods. This research was based on the analysis of purchases and recommendations in grocery shopping. Therefore, investigating the effectiveness of individualized responsiveness in other domains remains for future work. Finally, recommendation responsiveness might relate closely to personality, and investigating this relationship can lead to a better understanding of why users are affected by recommendation.

References

[1] Adomavicius, G. and Tuzhilin, A. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE T. Knowl. Data En.* 17, 6 (June 2005), 734-749. DOI=http://dx.doi.org/10.1109/TKDE.2005.99.

[2] Azaria, A., Hassidim, A., Kraus, S., Eshkol, A., Weintraub, O., and Netanely, I. 2013. Movie recommender system for profit maximization. In *Proceedings of the 7th ACM conference on Recommender systems* (Hong Kong, China, October 12 – 16, 2013). RecSys '13. ACM, New York, NY, USA, 121-128.

[3] Bodapati, A. V. 2008. Recommendation systems with purchase data. *J. Marketing Res.* 45(1) (February 2008), 77-93.

[4] Castells, P., Hurley, N. J., and Vargas, S. 2015. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*. Springer US, 881-918.

[5] Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., and Turrin, R. 2011. Looking for "good" recommendations: a comparative evaluation of recommender systems. In *Proceedings of the13th IFIP TC13 Conference on Human-Computer Interaction* (Lisbon, Portugal, September 05 – 09, 2011). INTERACT '11. Springer Berlin Heidelberg. 152-168.

[6] Di Noia, T., Ostuni, V. C., Rosati, J., Tomeo, P., and Di Sciascio, E. 2014. An analysis of users' propensity toward diversity in recommendations. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Silicon Valley, California, USA, October 06 - 10, 2014). RecSys '14. ACM, New York, NY, USA, 285-288.

[7] Gopalan, P., Hofman, J. M., and Blei, D. M. Scalable recommendation with hierarchical Poisson factorization. 2015. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence* (Amsterdam, Netherlands, July 12 – 16, 2015). UAI '15. AUAI Press.

[8] Gu, W., Dong, S., and Zeng, Z. 2014. Increasing recommended effectiveness with Markov chains and purchase intervals. *Neural Comput. Appl.* 25(5) (October 2014), 1153-1162.

[9] Hu, Y., Koren, Y., and Volinsky, C. 2008. Collaborative filtering for implicit feedback datasets. In *Proceedings of the 8th IEEE International Conference on Data Mining* (Pisa, Italy, December 15 - 19, 2008). ICDM '08. IEEE Computer Society, Washington, DC, USA, 263-272.

[10] Jiang, Y., Shang, J., Liu, Y., and May, J. 2015. Redesigning promotion strategy for e-commerce competitiveness through pricing and recommendation. *Int. J. Prod. Econ.* 167 (September 2015), 257-270.

[11] Johnson, C. C. 2014. Logistic matrix factorization for implicit feedback data. In *Proceedings of NIPS 2014 Workshop on Distributed Machine Learning and Matrix Computations*.

[12] Kapoor, K., Kumar, V., Terveen, L., Konstan, J. A., and Schrater, P. 2015. "I like to explore sometimes": adapting to dynamic user novelty preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria, September 16 – 20, 2015). RecSys '15. ACM, New York, NY, USA, 19-26.

[13] Koren, Y., Bell, R., and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (August 2009), 30-37.

[14] Lawrence, R. D., Almasi, G. S., Kotlyar, V., Viveros, M., and Duri, S. S. 2001. Personalization of supermarket product recommendations. *Data Min. Knowl. Discov*. 5, 1-2 (January 2001), 11-32.

[15] McCrae, R. R., Costa, P. T., de Lima, M. P., Simões, A., Ostendorf, F., Angleitner, A., Marušić, I., Bratko, D., Caprara, G. V., Barbaranelli, C., Chae, J. H., and Piedmont, R. L. 1999. Age differences in personality across the adult life span: parallels in five cultures. *Dev. Psychol*. 35(2) (March 1999) 466-477.

[16] McNee, S. M., Riedl, J., and Konstan, J. A. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montreal, Quebec, Canada, April 24 – 27, 2006). CHI EA '06. ACM, New York, NY, USA, 1097-1101.

[17] M Dias, M. B., Locher, D., Li, M., El-Deredy, W., and Lisboa, P. J. 2008. The value of personalised recommender systems to e-business: a case study. In *Proceedings of the 2nd ACM Conference on Recommender Systems* (Lausanne, Switzerland, October 23 - 25). RecSys '08. ACM, New York, NY, USA, 291-294.

[18] Pan, R., Zhou, Y., Cao, B., Liu, N. N., Lukose, R., Scholz, M., and Yang, Q. 2008. One-class collaborative filtering. In *Proceedings of the 8th IEEE International Conference on Data Mining* (Pisa, Italy, December 15 - 19, 2008). ICDM '08. IEEE Computer Society, Washington, DC, USA, 502-511.

[19] Rafailidis, D. and Nanopoulos, A. 2014. Modeling the dynamics of user preferences in coupled tensor factorization. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Silicon Valley, California, USA, October 06 - 10, 2014). RecSys '14. ACM, New York, NY, USA, 321-324.

[20] Said, A., Tikk, D., Shi, Y., Larson, M., Stumpf, K., and Cremonesi, P. 2012. Recommender systems evaluation: a 3D benchmark. In *Proceedings of RecSys 2012 Workshop on Recommendation Utility Evaluation: Beyond RMSE*. 21-23.

[21] Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M. 2002. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Tampere, Finland, August 11 - 15, 2002). SIGIR '02. ACM, New York, NY, USA, 253-260.

[22] Shani, G., Heckerman, D, and Brafman, R. I. 2005. An MDP-based recommender system. *J. Mach. Learn. Res*. 6 (December 2005), 1265-1295.

[23] Tintarev, N., Dennis, M., and Masthoff, J. 2013. Adapting recommendation diversity to openness to experience: a study of human behaviour. In *Proceedings of the 21th Conference on User Modeling, Adaptation, and*

Personalization (Rome, Italy, June 10 - 14, 2013). UMAP '13. Springer Berlin Heidelberg, 190-202.
[24] Wu, W., Chen, L., and He, L. 2013. Using personality to adjust diversity in recommender systems. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media* (Paris, France, May 1 – 3, 2013). HT '13. ACM, New York, NY, USA, 225-229.
[25] Zhang, F., Zheng, K., Yuan, N. J., Xie, X., Chen, E., and Zhou, X. 2015. A novelty-seeking based dining recommender system. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy, May 18 – 22, 2015). WWW '15. ACM, New York, NY, USA, 1362-1372.
[26] Zhang, W., Wang, J., Chen, B., and Zhao, X. 2013. To personalize or not: a risk management perspective. In *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong, China, October 12 – 16, 2013). RecSys '13. ACM, New York, NY, USA, 229-236.

Authors' Introductions

Masahiro Sato
Communication Technology, Research & Technology Group
Area of specialty: Recommender Systems, Collaborative Systems.

Hidetaka Izumo
Fuji Xerox Information Systems Co., Ltd.
Area of specialty: Recommender Systems, Customer Analysis.

Takashi Sonoda
Communication Technology, Research & Technology Group
Area of specialty: Physics (Ph.D.), Recommender Systems, Machine Learning.