Displaying User Profiles to Elicit User Awareness in Recommender Systems

Yoshinori Hijikata
Graduate School of
Engineering Science
Osaka University
Osaka, 560–8531, Japan
Email: hijikata@sys.es.osaka-u.ac.jp

Kazunori Okubo
Graduate School of
Engineering Science
Osaka University
Osaka, 560–8531, Japan

Shogo Nishida
Graduate School of
Engineering Science
Osaka University
Osaka, 560–8531, Japan

***Abstract*—When users understand their preferences and interests, they may find it easier to make a decision to buy an item new to them. Such understanding may also help users explore new categories because it enables them to judge differences from their current preferences or interests. In our research, we show users their user profiles created by a recommender system and ask them whether they learn any new knowledge about their preferences or interests. Because user profiles are usually implicitly created by machine learning techniques based on the users' usual activities such as browsing and shopping, they might include user preferences or interests of which the users are not explicitly aware.**

I. Introduction

While users have come to access information easily through the development of Web technology, they find some difficulty in obtaining information they need from the Web because of the existence of too much information. One promising solution is a recommender system. A recommender system is a system that shows users items (e.g., commercial goods, news articles, and online contents) that fit their preferences or interests.

Most studies of recommender systems have been driven by the goal of improving accuracy (how well the recommended items fit the users' preferences or interests). However, evaluation metrics that consider other aspects of usefulness such as novelty and serendipity have also been studied in recent years [1]. This is because it becomes difficult for users to acquire unknown items or items beyond their originally preferred categories when the system recommends items based only on accuracy metrics. In the future, people will expect that recommender systems not only output correct recommendations (fitting their preferences or interests) but also give a good user experience in the whole recommendation process.

In accordance with these new research directions that explore new user experiences with recommender systems, in this study we focus on users' awareness of their preferences or interests while using a recommender system. When users notice tendencies implicit in their preferences, they may come to make a decision to buy items with confidence, or they may explore new fields or categories that they have not tried before. For example, a user is interested in an item that appeals to her, but the item is in a category from which she has not bought before. In this case, she may be reluctant to buy this item because she is not confident that her selection is a good choice. When she notices that this item has a feature in common with items that she has bought before, she may decide with confidence to buy it, or she may start to search for new items having the above common feature and explore new categories of items.

In this study, we show a user her user profile with the aim of promoting user awareness. User profiles are basic information about users' preferences or interests and are used for recommending items to users in a recommender system. For example, they may be represented as a rating vector, in which each vector element corresponds to an item. In this paper, we regard a user profile as a discriminant model that classifies an item as a favorite one or a disliked one. A user profile is usually acquired implicitly by machine learning techniques from a browsing history, a purchase history, or a collection of item ratings (evaluation values the user has given to specific items) [2]. Therefore, it reflects not only the preferences or interests the user explicitly recognizes but also those the user has beneath her conscious awareness. Based on this concept, we show user profiles to the respective users in order to elicit their awareness of their preferences and interests.

Algorithms of recommender systems are categorized into those using content-based filtering and those using collaborative filtering [3]. A content-based filtering algorithm learns characteristics or features of items the user prefers and recommends items based on those. A collaborative filtering algorithm exploits other users' information, which does not include items' feature information, and recommends items that similar users prefer. We believe that items' characteristics are effective for eliciting users' awareness when the system shows them user profiles because they can understand the meanings of these characteristics. Therefore, we use several content-based filtering algorithms and display user profiles created by them.

In this study, we first identify the type of user profile format that makes it easy for users to understand their preferences or interests. We conduct an experiment in which user profiles created by several major recommendation algorithms are presented to users and the users are asked how easy it is to understand what the user profiles present. Then, we determine whether the user profile having the format found above leads to users' actual awareness of their preferences and interests by means of a user experiment. We ask users to input the content of their discoveries in this experiment, and we analyze their quantity and quality.

The following are the contributions of our study:

- We show that users can become aware of previously unknown preferences and interests by viewing their user profiles.
- We show that some of the users' discoveries are useful to them.
- We determine some typical types of content of the users' discoveries.
- We show whether users' discoveries help them to choose items to purchase.

The rest of the paper is organized as follows. Section II introduces related work. Section III explains the dataset used in our study. In Section IV, we address the type (format) of user profile to use when showing it to the user. Section V describes a method for recording the user's discoveries and evaluating them. Section VI explains the procedure of a user experiment using the above evaluation method, and Section VII shows its results. Finally, Section VIII gives a summary of the study and suggestions for future work.

II. Related work

A. Approach beyond the accuracy

A new research orientation is becoming popular in which researchers consider not only the accuracy of the recommendation results but other usefulness factors as well, especially regarding the user's knowledge acquisition. Herlocker et al. showed a notion of novelty and serendipity [1]; the former assesses whether the user does or does not know the recommended items, and the latter assesses how much the recommended items surprise the user. Ziegler et al. proposed an intra-list similarity metric, which measures the diversity of items in a recommendation list, and developed an algorithm that diversifies a recommendation list produced by collaborative filtering algorithms [4]. Adomavicius and Zhang showed that the stability of recommendation results is higher in memory-based algorithms than in model-based algorithms [5]. Hijikata et al. developed several algorithms that recommend items of high novelty based not only on users' own preference ratings but also on those of their acquaintances [6]. Vargas and Castells aimed to represent metrics for evaluating novelty and diversity in a generalized model [7].

Some researchers have conducted research on a human interaction model to improve the above evaluation metrics or user satisfaction. Herlocker et al. examined several types of explanation of recommendation result for a collaborative filtering system [8]. Castagnos et al. analyzed the movement of the user's point of gaze in the recommender system and ascertained that the diversification of the recommendation results are important to the user's decision making [9]. Bollen et al. examined the length of the recommendation list and found that it influences the attractiveness of the recommendation list [10]. Cremonesi et al. examined the relationship between the length of the user profiles (the number of ratings the user has input) and the user satisfaction [11]. Knijnenburg et al. examined the relationship between the inspectability and control (user intervention) based on structural equation modeling [12]. They reported that the user satisfaction increases when the user perceives that they can control the recommender system. Ekstrand et al. conducts a user experiment according to the relationship between novelty, diversity, accuracy, satisfaction and the degree of personalization [13]. Hijikata et al. developed a system in which users can intervene in the recommendation process and examined the relationship between the degree of user intervention and user satisfaction [14].

In this study, we consider user awareness as a new orientation toward research into recommender systems. The above studies considered the user's acquaintance with items, or user's experience or intervention in recommender systems. However, they did not examine the user's awareness on their preferences or interests.

B. Displaying user profiles

Some researchers have displayed a user profile constructed by a system. Ahn et al. showed a user keywords in which the user showed interest while using a recommender system of news articles [15]; they also allowed users to remove inappropriate keywords from the listed keywords or to add new keywords interesting to them. Hijikata et al. portrayed a user's preference model as a tree structure for a music recommender system [16]; they showed that the precision of the recommendation was improved when the system used the user profile revised by the user. Bostandjiev et al. represented user preferences regarding music through bar graphs that users can edit themselves [17].

When a system constructs a user profile via a machine learning technique, the user profile as learned sometimes includes errors. The objective of the above studies was to give the users a chance to resolve the errors. The objective of our study is to facilitate users' awareness of their preferences or interests by displaying user profiles.

III. Dataset

Many recommender systems utilize machine-learning techniques for building a user profile; we also take this approach. A learning data set is required in order to build the user profile. We need to select a domain for collecting the data. We selected two domains to use in both the profile format experiment and the user awareness experiment. One is of wallpaper figures for the desktop PC screen; the other is of crafted figurines for placing on a display shelf or work desk. Although movies, music, and news articles are popular items in recommender systems, we chose domains where it is difficult for users to present their preferences by themselves.

We downloaded 1000 pictures from Flickr [18], one of the popular photo-sharing sites, in the wallpaper domain and the crafted figurine domain, respectively. We used 200 pictures as the learning set and 800 pictures as the test set (for online test). When building a recommender system using a content-based filtering method, feature data are required for each item. We defined the set of features and their values for each domain, as shown in Tables I and II. All the features are categorical in nature. When we refer to a specific categorical value of a specific feature, we will call it a "feature type."

TABLE I. FEATURES AND THEIR CATEGORICAL VALUES FOR WALLPAPER.

Feature	Values
Weather	clear / sunny / cloudy / unknown
Season	spring / summer / fall / winter / unknown
Time of day	day / evening / night
Dominant color	white / black / red or yellow / blue / green / brown
Photographic subject	artifact / animal / plant / sea / mountain / field of grass / rock / sky / forest
Illuminant direction	ahead / cross direction / backward / unknown

TABLE II. FEATURES AND THEIR CATEGORICAL VALUES FOR CRAFTED FIGURINES.

Feature	Values
Motif (subject)	human / animal / industrial products / plant / (cartoon) character / food / nature / vehicle / abstract entity
Materials	wood / metal / glass / ceramic / fabric / plastic / plaster / stone or sand / leather
Dominant color	white / black / blue / red or yellow / green / brown / clear / multicolored / silver
Practical utility	true / false
Culture	Japanese / European / ethnic / Chinese / unknown
Era	modern / classic / future / old days or childhood / ancient

For assigning feature values to an item, three human evaluators viewed the item and selected one categorical value for each feature. When the evaluators' selections differed from each other, they discussed it and selected the most common or similar value that could achieve a resolution among them.

IV. EXAMINATION OF PROFILE FORMATS

In this section, we describe the determination of a profile format to be used for the user awareness experiment. With a user experiment, we identify the kind of profile format that allows users to easily understand the content of a user profile.

A. *Profile formats for comparison*

Among the many machine-learning algorithms available for recommender systems, for this examination we select several algorithms that allow users to understand the content of the learned discriminant model. We examine the ease of understanding the content of the learned user profile. In particular, we examine the following three user profile formats generated from their respective learning algorithms:

– Naive Bayes (NB)
– Decision tree (DT)
– Market basket analysis (MB)

Note that we will use the abbreviated names when referring to the profile formats.

1) Naive Bayes: Naive Bayes is an algorithm that learns the probability that the user likes an item having a specific feature type [19]. NB is a profile format in which the respective probabilities that the user prefers each feature type (these can be considered association rules) are shown in a list (see Figure 1).

If the probability that the user prefers a given feature type is higher than the probability that she dislikes it, the label "like" is assigned to the circle (shown to the right of the feature type); otherwise, "dislike" is assigned. The size of the circle representing the decision corresponds to the number of cases

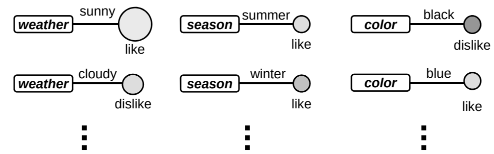


Fig. 1. Example of NB format, generated from the naive Bayes algorithm.

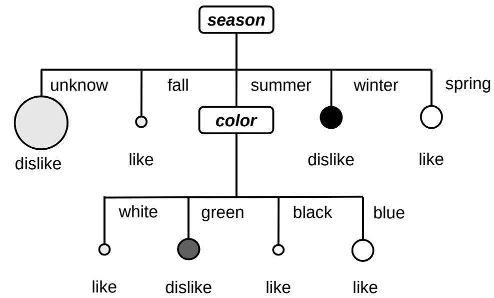


Fig. 2. Example of DT format, generated from the decision tree algorithm.

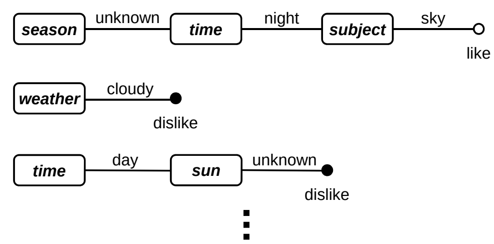


Fig. 3. Example of MB, generated using market basket analysis.

(data points) used in learning the rule, and the intensity of the color relates to the ratio of the number of cases with the label shown ("like" or "dislike") to the total number of cases that were used to learn the rule.

The size and color intensity of the circle have the same meanings in the other two formats, described in the sections that follow.

2) Decision tree: A decision tree is a learning algorithm in which the learned model is presented as a tree structure. A leaf shows a class ("like" or "dislike"), and the path to the leaf presents the feature type or the set of feature types that items in the class should have. We use C4.5 [20], one of the popular decision tree algorithms. DT presents the entire model learned by C4.5 (Figure 2).

3) Market basket analysis: The market basket analysis uses the Apriori algorithm [21] to learn the co-occurring feature types having the same label ("like" or "dislike"). MB presents the learned rules in a list (Figure 3).

B. *User experiment*

We conducted a user experiment to determine which format makes it easiest for users to understand the content of the user profile. We showed test subjects the three kinds of user profiles created from a specific user's learning data and asked them to evaluate the user profiles according to their ease in understanding the content of the user profile. We invited 15 graduate students as test subject.

TABLE III. EASE OF UNDERSTANDING FOR EACH PROFILE FORMAT.

Domain	NB format	**DT format**	MB format
Wallpaper	4.21	**4.93**	3.21
Crafted figurines	4.15	**4.77**	3.23

For preparation, we created user profiles using the three kinds of format in advance. We asked three graduate students (different ones from the above 15 students) to rate each of 100 items in each domain on a five-point scale, ranging from -2 (dislike) to 2 (like). We created a user profile in each of the three profile formats for each of the three users' rating sets; thus, we created nine user profiles in each domain. We used Weka 3.6 [22], one of the popular data mining tools, to create the user profiles.

Each of the 15 test subjects viewed and compared three user profiles at the same time; the three user profiles in each set were made from different users and in different formats. The test subjects were asked to guess the preference of the owner of each user profile. They reported their ease in guessing the owner's preference on a seven-point scale (1–7).

Table III shows the average of the scores for ease of understanding for each profile format. The bold value shows the highest value. We can see that DT allows the greatest ease in understanding the user's preference. We believe that the DT's characteristic that the user can see the whole model at a glance in a hierarchical manner contributes to the ease of understanding it.

V. EXAMINATION OF USERS' AWARENESS

This section explains the design of our user experiment. We aimed to examine the following four hypotheses related to profile presentation and users' awareness:

– Users discover their preferences or interests by seeing their displayed user profiles.

– Some of the users' discoveries are useful.

– Users discover knowledge other than that in the displayed information.

– Users' discoveries help them in making a decision.

In this section, we explain a method for obtaining the users' discoveries, evaluating the value of the users' discoveries, and determining whether the users' discoveries help them to make a decision.

A. Acquisition of users' discoveries

Test subjects give ratings to items in a five-point scale using the interface shown in Figure 4-(a) (The details are shown later). A decision tree is learned from the given ratings. The test subjects view the decision tree representing their own user profiles (Figure 4-(b)). If they notice something about their preference, they click on a button labeled "!" located at the upper right of the window, and a new window for inputting the content of the discovery is invoked (Figure 4-(c)). They input the content of their discovery, and the system stores it in the server. They repeat this until they do not notice anything more.

TABLE IV. NUMBER OF TAGS IN EACH DOMAIN.

Domain	AllTag	DBTag
Wallpaper	80,104	1,202
Crafted figurines	97,578	1,496

B. Evaluation of users' discoveries

We evaluate the users' discoveries from a quantitative viewpoint and a qualitative viewpoint. We also examine the types of content in the users' discoveries.

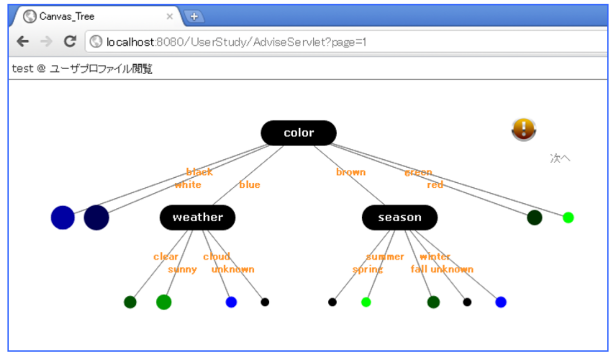
1) Quantity of discoveries: The simplest method for evaluating the quantity of the users' discoveries would be to count the number of inputs of discoveries. However, the detail or granularity of an input may differ between users because there are individual differences, making it difficult to measure the quantity under a uniform scale. In this study, we introduce a discovery checklist, created from a uniform perspective. After test subjects input their discoveries with a free-form description, they are asked to evaluate their discoveries using the checklist.

The checklist lists 100 tags with checkboxes (Figure 4-(d)). Tags are usually used for both calculating the recommendation scores (for making recommendation lists) [23]–[27] and explaining the recommendation results [28]–[30]. Thus, we think that these tags can be used for presenting items in the target domain (wallpaper or crafted figurines in this experiment) and the users' preferences or interests to them. These 100 tags are selected randomly in advance from a tag database that we created. We created this tag database by collecting tags that were given to the photos in the target domain in Flickr and selecting those that appear with high frequency (10 times or more). Table IV shows the number of tags collected from Flickr (AllTag) and the number of tags recorded in the tag database (DBTag) for each domain.

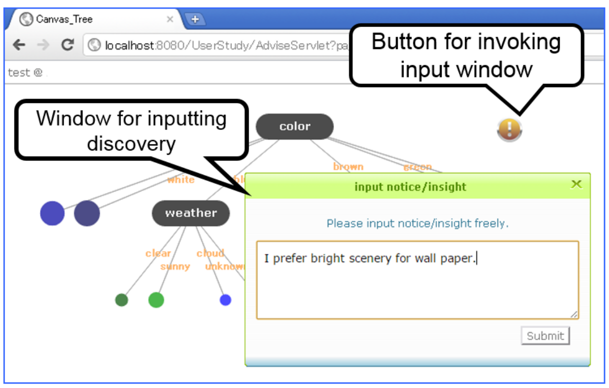
We believe that high-frequency tags are general words representing the features of items (although there still exists some variability in the granularity of description or the level of abstraction). Thus, we use the above tags as units of discovery. Furthermore, users can create a new tag and check it if they cannot find appropriate tags for their discoveries. We measure the quantity of users' discoveries by counting the number of descriptions and that of tags the user checked.

2) Quality of discoveries: We introduce the serendipity metric and the importance metric for evaluating the quality of the users' discoveries. The serendipity metric is a subjective measure representing how much the discovery surprises the user or how much the user was previously unaware of it. It is inspired by the serendipity metric of Herlocker et al. [1], which was applied to recommended items; here it is applied to the users' discoveries. The importance metric is a subjective measure representing how much the discovery affects the user's future selection (or purchase) of items. Each test subject scores both the serendipity and the importance of the description of the discovery and of each of its corresponding tags, using a five-point scale (1–5) (Figure 4-(d)). They input the serendipity score in the left (blue) dropdown list and the importance score in the right (orange) dropdown list. We define discoveries having both high serendipity and high importance as useful discoveries.

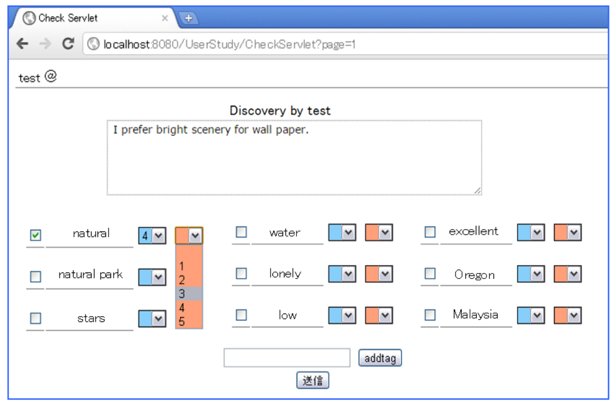
(a) Screen for rating a figure



(b) Screen of user profile presentation



(c) Window for inputting a discovery



(d) Screen for checking and evaluating the tags

Fig. 4. Examples of user profile presentation and other screens used in the experiment.

3) Content of discoveries: We analyze the content of the users' discoveries. In particular, we determine whether the users have discovered knowledge other than that in the presented information. We categorize the content of the discoveries into the following four types. (Note that expressions in parentheses are abbreviations for the categories.)

– A discovery that is the same as the information displayed in the user profile (SM).

– A discovery that is expressed as a word whose notion is a generalization of several pieces of information displayed in the user profile (GN).

– A discovery that is expressed as a more concrete word for information displayed in the user profile (CN).

– A totally new discovery that is not included in the user profile (NW).

C. Influence on decision making

To determine the influence of the users' discoveries on their decision making, we ask them to give ratings to items before and after the user profiles are presented. The first round (before the user profiles are presented) has been already conducted when learning the profile. The second round (after the user profiles are presented) will be conducted after the test subjects finish their browsing of the user profiles. We measure the time that the user takes to give the ratings and the variance of the ratings as objective evaluation metrics. We also ask the users about their degree of confidence in the correctness of their ratings given before and after the user profiles are presented. These will be used to calculate a subjective evaluation metric.

If the users do not understand their preferences well, they lack confidence in the ratings they give to items. For such users, we may expect the variances of the ratings to be small because they hesitate to give an extreme rating value such as -2 or 2 on the five-point scale (-2 to +2). We also expect that the time required to give a rating to an item may be longer because those users face difficulties in selecting an explicit rating value. We expect that users can make a decision with confidence once they make their discoveries and understand their preferences more deeply; this will lead to increasing the variance of the ratings and shortening the time needed to give ratings.

VI. Execution of Experiment

We conducted a user experiment using the experimental method described in the previous section. We invited 20 test subjects who were men and women with ages between 20 and 43.

A. System for experiment

We built a recommender system equipped with a user profile presentation function for the experiment. The server of the system was built using Java Servlet technology, and the client was built in JavaScript. The test subjects participated in the experiment by accessing the server from their own PCs using their preferred Web browsers. We used Weka 3.6 to create the user profiles. The decision tree algorithm for the recommendation was C4.5.

B. User task

The test subjects conduct the task in the following manner.

(1) Give ratings to items in the learning set: The test subjects give ratings to 100 items randomly selected from the items in the learning set on a five-point scale, from -2 (dislike) to 2 (like). (See Figure 4-(a))

(2) Input the degree of confidence (first round): They input the degree of confidence they have in the above item evaluations on a five-point scale (1 – Not confident at all, 2 – Not very confident, 3 – Somewhat confident, 4 – Quite confident, 5 – Very confident).

(3) Describe the tendency of their item preference: They describe the features or characteristics of their favorite (or unfavorite) items in a free-form description.

(4) Describe their discoveries after user profile browsing: They describe their discoveries in a free-form description (see Figure 4-(c)) after browsing their displayed user profile (see Figure 4-(b)).

(5) Check appropriate tags about their discoveries: They check tags in the checklist that match the content input in Step (4) (see Figure 4-(d)). They also evaluate each tag according to its serendipity and its importance, using a five-point scale for each.

(6) Give ratings to items in the test set: They give ratings to each of 100 items randomly selected from the items in the test set, on a five-point scale.

(7) Input the degree of confidence (second round): They input the degree of confidence they have in the above item evaluations on a five-point scale.

We prepared Step (3) so that the test subjects could easily evaluate the serendipity and the importance of their discoveries in Step (5).

1) Serendipity: They evaluated serendipity by choosing one of the following statements: 1 – I reconfirmed what I input in Step (3). 2 – I am aware of the tendency (content of the text description or tag) although I forgot to input in Step (3). 3 – I am vaguely aware of the tendency although I did not input it in Step (3). 4 – I am only slightly aware of the tendency. 5 – These are totally new discoveries for me.

2) Importance: When evaluating the importance of their discoveries, they are asked to consider the future purchase or selection of items. They evaluate it on a five-point scale (1 – Not helpful at all, 2 – Not very helpful, 3 – Somewhat helpful , 4 – Quite helpful, 5 – Very helpful).

VII. Experimental results

A. Quantity of discoveries

The experimental results for the quantity of discoveries are shown in Table V. In the table, #people denotes the number of test subjects who input at least one discovery when they saw their displayed user profiles, and #avg_notices and #avg_tags denote the average number of discoveries input via free-form description and the average number of tags input per test subject (considering only those test subjects who input at least one discovery), respectively.

TABLE V. Statistics of test subjects' discoveries.

Domain	#people	#avg_notices	#avg_tags
Wallpaper	17	2.9	15.1
Crafted figurines	18	2.0	9.4

TABLE VI. Number of test subjects who made useful discoveries.

		Free-form description			Tag		
Domain	All	s-h	i-h	s&i-h	s-h	i-h	s&i-h
Wallpaper	17	14	11	9	12	15	8
Crafted figurines	18	11	16	9	12	15	10

We can see that most of the test subjects do make some discoveries after seeing their user profiles. When they made discoveries, on average they input two or three free-form descriptions and checked 10–15 tags.

B. Usefulness of discoveries

We evaluated the quality of the discoveries from their serendipity and importance. Here, we regard discoveries whose serendipity is greater than three as discoveries of high serendipity and those whose importance is greater than three as discoveries of high importance. Further, we regard discoveries whose serendipity is greater than three and whose importance is greater than three as useful discoveries. Table VI shows the number of test subjects who made at least one discovery of high serendipity (s-h, for serendipity-high), the number of test subjects who made at least one discovery of high importance (i-h, for importance-high), and the number of test subjects who made at least one discovery of both high serendipity and high importance (useful discovery) (s&i-h, for serendipity&importance-high). In this table, "All" indicates the number of test subjects who made at least one discovery.

The numbers of test subjects differ between the free-form description and the tags for each of s-h, i-h and s&i-h. We infer that some people gave lower ratings to the tags because they lost the value of word combinations that the original free-form description had, and other people gave higher ratings to the tags because they reassessed the serendipity and the importance of their discoveries when they saw the tags.

C. Content of discoveries

We classified the discoveries into the four types defined in Section V. Table VII shows the number of discoveries of each type, along with some sample discoveries. Note that we would classify one discovery into several types when it included more than one type of content.

Most of the test subjects' discoveries were directly input from the description in the displayed user profile (SM). Meanwhile, we confirmed 25 discoveries of other types (GN, CN, and NW). This indicates that users sometimes notice content in addition to the presented content when seeing their user profiles.

TABLE VII. NUMBER OF DISCOVERIES FOR EACH DISCOVERY TYPE.

	Example	Wall-paper	Crafted figurines
SM	I love dogs. I reconfirmed that I love European objects. I dislike artificial materials.	43	27
GN	I prefer bright scenery for wallpaper. I dislike creatures. I love things that exist in nature.	5	7
CN	I figured out the reason that I like the combination of the color white and winter is that I love snow scenery. I figured out the reason that I like both European style and the color white is that I love elegant figures. I realized that the reason I like the color black and things unrelated to season is that I like photos with a small amount of light.	7	4
NW	I might dislike bright pictures because I keep away from things that are bad for my eyes in an unconscious way. I might not be satisfied with everyday scenery.	2	0

TABLE VIII. CHANGE IN VARIANCE OF ITEM RATINGS.

	Larger	No change	Smaller
Wallpaper	4	4	9
Crafted figurines	7	5	8

TABLE IX. CHANGE IN TIME NEEDED TO GIVE RATINGS TO ITEMS.

	Shorter	No change	Longer
Wallpaper	4	8	4
Crafted figurines	5	11	4

TABLE X. CHANGE IN DEGREE OF CONFIDENCE IN EVALUATING ITEMS.

	More confident	No change	Less confident
Wallpaper	4	11	2
Crafted figurines	3	15	2

D. Decision support

Table VIII shows the change in the variance of item rating values before and after browsing the user profile. The change is described as "became larger," "no change," or "became smaller." Table IX shows the change in the time required for evaluation. The change is described as "became shorter," "no change," or "became longer." If the rating value or required time changes by 10% or more, we regard the change as "became larger (smaller)" or "became longer (shorter)." From these figures, we cannot find any tendency for the variance of rating values to increase or for the time required to decrease.

Table X shows the change in the degree of confidence in evaluating items before and after browsing the user profile. The change is described as "became more confident," "no change," or "became less confident." We can see that the user's confidence hardly changes after browsing the user profile.

From the results of the experiments, we could not find any change in either of the objective evaluation metrics or in the subjective evaluation metric. This indicates that the users' discoveries do not lead to support for their decisions.

E. Discussion

From the results of the experiments, we confirmed the following points: (1) users are able to notice preferences of which they were previously unaware by seeing their user profiles, (2) some of the users' discoveries are useful, (3) there exist some typical types of content in the users' discoveries, and (4) users notice information other than that displayed. However, we did not find that displaying the user profile helps users in making their decisions on selecting items.

TABLE XI. RESULTS OF FOLLOW-UP QUESTIONNAIRE.

	Wallpaper	Crafted figurines
Came to think more deeply	4	6
No change	5	4
Came to think more simply	8	8

To ascertain why the useful discoveries did not help users in their decision making, we had an interview with six of the test subjects who participated in the experiment. We asked them whether there were any changes in their thinking process when evaluating items before and after browsing the user profile. The answers from the interview can be categorized into "came to think more deeply," "no change," and "came to think more simply." Thinking deeply means that the user takes many factors into consideration when evaluating the item. Thinking simply means that the user considers a limited set of features for the evaluation.

Based on the results of this interview, we sent a follow-up questionnaire to all of the test subjects who had made discoveries. In the questionnaire, we asked them to select from three options to describe how their thought process for evaluating items changed after browsing their user profile. The results are shown in Table XI.

From this table, we can see that the changes in the users' thinking process do not follow a consistent pattern. When thinking more deeply, it takes users more time to evaluate items. The variety of changes in thought process leads to the erratic results seen in the changes in the variance of rating values, time required for evaluating items, and degree of confidence.

VIII. CONCLUSION

We focused on users' discoveries as one factor enriching the user experience in recommender systems. We proposed displaying the user profile for eliciting users' discoveries of their preferences among items. To implement this function, we first examined different formats of user profile and found that the type of user profile created by the decision tree algorithm is the easiest for understanding the preferences of the user profile's owner.

We then conducted a user experiment to determine whether users can make discoveries when seeing their displayed user profiles in the decision tree format. From the experimental results, we found that some of the users make useful discoveries and that some of the users acquire knowledge that was not presented in the displayed user profile. However, we could not confirm that the discoveries made increase the rating variances or shorten the time for evaluation.

From the results of a follow-up questionnaire, we found that the change in users' thought process when evaluating items after browsing the user profiles differs between users. We plan to examine the relationship between the pattern of contemplation and the ease of decision making in a future study.

References

[1] J.L. Herlocker, et al, “Evaluating Collaborative Filtering Recommender Systems,” ACM Trans. on Information Systems, Vol.22, Issue 1, pp.5–53, 2004.

[2] Y. Hijikata, “User Profiling Technique for Information Recommendation and Information Filtering,” Journal of JSAI (The Japanese Society for Artificial Intelligence) , Vol.19, No.3, pp.365–372, 2004.

[3] D. Riecken, et al., “Personalized Views of Personalization,” Comm. of the ACM, Vol. 43, No. 8, pp. 26–158, 2000.

[4] C.N. Ziegler, et al., “Improving Recommendation Lists Through Topic Diversification”, Proc. of the 14th ACM International Conference on World Wide Web, ser. WWW’05, pp. 22–32, 2005.

[5] G. Adomavicius and J. Zhang, “On the Stability of Recommendation Algorithms,” Proc. of the fourth ACM Conference on Recommender Systems, ser. RecSys’10, 2010.

[6] Y. Hijikata, T. Shimizu and S. Nishida, “Discovery-oriented Collaborative Filtering for Improving User Satisfaction,” Proc. of the 14th ACM International Conference on Intelligent User Interfaces, ser. IUI’09, pp.67–76, 2009.

[7] S. Vargas and P. Castells, “Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems,” Proc. of the fifth ACM Conference on Recommender Systems, ser. RecSys’11, pp.109–116,2011.

[8] J.L. Herlocker, J.A.Konstan and J. Riedl, “Explaining Collaborative Filtering Recommendations,” Proc. of the 2000 ACM Conference on Computer Supported Cooperative Work, ser. CSCW’00, pp.241–250, 2000.

[9] S. Castagnos, N. Jones and P. Pu, “Eye-Tracking Product Recommenders’ Usage,” Proc. of the fourth ACM Conference on Recommender Systems, ser. RecSys’10, pp.29–36, 2010.

[10] D. Bollen, et al., “Understanding Choice Overload in Recommender Systems,” Proc. of the fourth ACM Conference on Recommender Systems, ser. RecSys’10, pp.63–70, 2010.

[11] P. Cremonesi, F. Garrzotto and R. Turrin, “User Effort vs. Accuracy in Rating-based Elicitation,” Proc. of the sixth ACM Conference on Recommender Systems, ser. RecSys’12, pp.27–34, 2012.

[12] B.P. Knijnenburg, et al.: Inspectability and Control in Social Recommenders, Proc. of the sixth ACM Conference on Recommender Systems, ser. RecSys’12, pp. 43–50, 2012.

[13] M. Ekstrand, et al., “User Perception of Differences in Movie Recommendation Algorithms,” Proc. of the 8th ACM Conference on Recommender Systems, ser. RecSys’14, pp. 161–168, 2014.

[14] Y. Hijikata, Y. Kai and S. Nishida, “The Relation between User Intervention and User Satisfaction for Information Recommendation,” Proc. of the 27th Annual ACM Symposium on Applied Computing, ser. SAC’12, pp. 2002–2007, 2012.

[15] Ahn, et al., “Open User Profiles for Adaptive News Systems: Help or Harm?,” Proc. of the 16th International Conference on World Wide Web, ser. WWW’07, pp.11–20, 2007.

[16] Y.Hijikata, K. Iwahama and S. Nishida, “Content-based Music Filtering System with Editable User Profile,” Proc. of the 2006 ACM symposium on Applied computing, ser. SAC’06, pp.1050–1057, 2006.

[17] S. Bostandjiev, J. O’Donovan, J., T. Hollerer, “Taste Weights: a Visual Interactive Hybrid Recommender System,” Proc. of the sixth ACM Conference on Recommender Systems, ser. RecSys’12, pp.35–42, 2012.

[18] http://www.flickr.com/.

[19] G.H. John and P. Langley, “Estimating Continuous Distributions in Bayesian Classifiers,” Proc. of the Eleventh Conference on Uncertainty in Artificial Intelligence, ser. UAI’95, pp. 338–345, 1995.

[20] J.R. Quinlan, C4.5, “Programs For Machine Learning,” Morgan Kaufmann Publishers, 1993.

[21] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules,” Proc. of the 20th International Conference on Very Large Data Bases, ser. VLDB’94, pp.487–499, 1994.

[22] http://www.cs.waikato.ac.nz/ml/weka/.

[23] J. Diederich and T. Iofciu, “Finding Communities of Practice from User Profiles based on Folksonomies,”. Proc. of Innovative Approaches for Learning and Knowledge Sharing, EC-TEL Workshop, pp. 288–297, 2006.

[24] E. Michlmayr and S. Cayzer, “Learning User Profiles from Tagging Data and Leveraging them for Personal(ized) Information Access, Proc. of the Workshop on Tagging and Metadata for Social Information Organization, Int. WWW Conf., 2007.

[25] A. Shepitsen, et al., “Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering,” Proc. of the 2008 ACM Conference on Recommender Systems, ser. RecSys’08, pp.259–266, 2008.

[26] M. Gemmis, et al., “Integrating Tags in a Semantic Content-based Recommender,” Proc. of the 2008 ACM Conference on Recommender Systems, ser. RecSys’08, pp.163–170, 2008.

[27] H. Liang, et al., “Connecting Users and Items with Weighted Tags for Personalized Item Recommendations,”, Proc. of the 21st ACM conference on Hypertext and hypermedia, ser. HT’10, pp.51–60, 2010.

[28] J. Vig, S. Sen and J. Riedl, “Tagsplanations: Explaining Recommendations using Tags,” Proc. the 14th international conference on Intelligent user interfaces, ser. IUI’09, pp. 47–56, 2009.

[29] S. Sen, J. Vig and J. Riedl, “Tagommenders: Connecting Users to Items through Tags,” Proc. of the 18th International Conference on World Wide Web, ser. WWW’09, pp. 671–680, 2009.

[30] F. Gedikli and D. Jannach, “Improving Recommendation Accuracy based on Item-specific Tag Preferences,” ACM Trans. on Intelligent Systems and Technology, Vol. 4, No. 1, Article 11, 2013.