An Analysis of Differences between Preferences in Real and Supposed Contexts

Hideki Asoh
Intelligent Systems RI, AIST
AIST Tsukuba Central 2
1-1-1 Umezono, Tsukuba
Ibaraki 305 Japan
h.asoh@aist.go.jp

Yoichi Motomura
Center for Service Research, AIST
AIST Tokyo Waterfront
2-41-6 Aomi, Koto-ku
Tokyo 135-0064 Japan
y.motomura@aist.go.jp

Chihiro Ono
KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino
Saitama 365-8502 Japan
ono@kddilabs.jp

ABSTRACT
Modeling users' preferences is an important element of constructing recommender systems. In order to make statistical preference models, collecting large amount of data regarding users' preferences, such as ratings, through inquiries is normally done. In particular, to make the model context-aware, users' preference data should be collected under various contexts. Because putting subjects of inquiries into the real contexts and collecting answers is often difficult and/or costs much, collecting answers in supposed contexts, i.e. contexts where the subjects pretend or image that they are in the specific contexts, is often conducted. Although there may be differences between the preferences in real contexts and the preferences in supposed contexts, the differences have not been investigated. In our previous works, we collected users' preferences in both real and supposed contexts and shown that the differences of the whole rating distributions is statistically significant and not negligible. In this work, we analyzed the nature of the differences in more detail by comparing the rating distributions conditioned by contexts and items. As a result we found an interesting aspect of the differences, that is, the differences of rating distributions depending on contexts are much larger in the supposed contexts than in the real contexts. This result suggests that the subjects in the supposed contexts answered to the inquiries not by imaging to be in the contexts but by just using a simple knowledge about the general tendency of ratings in the specific contexts.

Categories and Subject Descriptors
G.3. [**Probability and Statistics**]: Contingency table analysis

General Terms
Experimentation, Human factors, Measurement

Keywords
Preference Modeling, Context-Aware Recommender Systems

CARS-2010, September 26, 2010, Barcelona, Spain.

1. INTRODUCTION
Context-awareness is an important research issue in the area of recommender systems [1, 2, 3]. In particular, it is indispensable for recommender systems on mobile phones which can be used in various user contexts. We have constructed several context-aware attribute-based recommender systems using Bayesian networks for modeling users' preferences [7, 8] In order to construct the statistical preference model, collecting large amount of data about users' preference, such as ratings of items, through inquiries is necessary. In particular, to make the model context-aware, users' preference data should be collected under various contexts. Because putting subjects of inquiries into real contexts and collecting answers from them is often difficult, collecting answers in supposed contexts, i.e. contexts where the subjects pretend or image that they are in the specific contexts, is often conducted. Although there may be difference between the preferences in real contexts and in supposed contexts, the differences have not so been investigated as far as we know.

In our previous works [9, 10], we collected users' preferences in both real and supposed contexts and shown that the differences of the whole distributions of preference ratings is statistically significant and not negligible. In this work, we analyzed the nature of the differences in more detail by comparing the rating distributions conditioned by contexts and items. As a result we found an interesting aspect of the differences, that is, the differences of rating distributions depending on contexts are much larger in the supposed contexts than in the real contexts. This result suggests that the subjects in the supposed contexts answered to the inquiries not by imaging to be in the contexts but by just using a simple knowledge about the general tendency of ratings in the specific contexts.

The rest of the paper is organized as follows. Section 2 is about the data we analyze in this work. Section 3 describes the result of analysis. Section 4 is for discussion, and section 5 is for conclusion and future work.

2. DATA SET
We designed internet questionnaire surveys in order to collect corresponding data, that is, we asked subjects the same question about preference both in real and supposed contexts and collect pairs of answers [9, 10]. The target contents were food menus provided in typical food court such as "chicken steak" or

Table 1: Example Records of Two Surveys

The 1st Survey

Real Context	Supposed Context	Menu	Rating
Full	×	Japanese Noodle	3
Full	Normal	Japanese Noodle	3
Full	Hungry	Japanese Noodle	1

The 2nd Survey

Real Context	Supposed Context	Manu	Rating
Hungry	×	Japanese Noodle	2
Hungry	Normal	Japanese Noodle	2
Hungry	Full	Japanese Noodle	2

Table 2: Example Records of Combined Corresponding Data Set

Menu	Context	Real Preference (Rating)	Supposed Preference (Rating)	Real Context for Supposed Preference
Japanese Noodle	Full	3	2	Hungry
Japanese Noodle	Hungry	2	1	Full
Beef Steak	Full	4	3	Normal
Beef Steak	Normal	3	1	Full
Chinese Fried Rice	Hungry	2	2	Full
Chinese Fried Rice	Full	5	4	Hungry

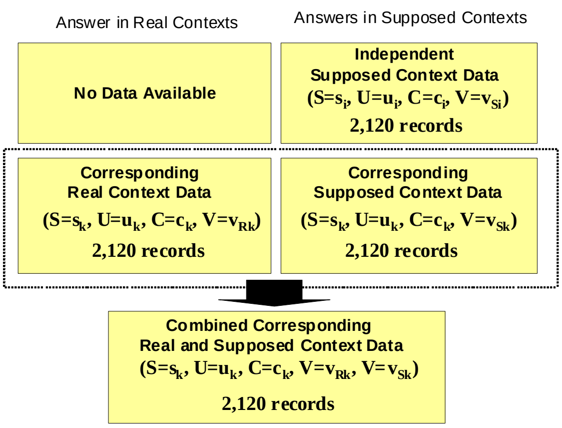

Figure 1: Acquisition of corresponding data.

"Japanese noodle" and the considered context is degree of hungriness of the subjects.

We choose the target domain and the context attribute because the difference of preferences depending contexts is expected to be large. Also the hungriness is easy to ask and is expected to distribute randomly when asking through internet questionnaires.

The survey was composed of two successive questionnaires. The first one was conducted from 16th to 17th in December 2008. The number of subjects was 746 and each subject evaluated 5 kinds of a la carte menus in 5 degree scale. The menus are randomly selected from 20 kinds of menus such as "chicken steak", "beef steak", "beef curry", "pasta with salted cod roe", "Chinese noodle in spicy source", etc. We choose typical food menus in a casual food court considering the variation of food stuff and food type.

At the same time the subjects answered their degree of hungriness in 3 levels (hungry, normal, full). After that, the subjects are asked to imagine they are in hungriness conditions different from one at that time and answered the preference for the same five menus. The latter answers become preferences in supposed contexts. In total, preferences for five menus in three different contexts (degree of hungriness) are collected. Among the three contexts, one is real and two are supposed.

Then, the second survey was conducted in other days from 22nd to 24th in December 2008. The all subjects who answered in the first survey were imposed the same questions as the first survey and we selected subjects who answered in different real hungriness from the first survey. After filtering out unreliable subjects, the number of subjects remained for analysis was 212.

By combining the results of the two surveys, we got corresponding preference for five food menus in two different degrees of hungriness per a subject. Hence the number of corresponding records was 2,120. Table 1 shows examples of answers in the two surveys, Table 2 shows examples of combined corresponding rating data, and Figure 1 shows how we got the corresponding data set.

3. DATA ANALYSES

In our previous works [9, 10], we showed that the difference of the whole distributions of preference ratings in the real and supposed contexts is statistically significant. In this work, we analyzed the differences of rating distributions conditioned by contexts and items.

Baltrunas and Ricci proposed the idea of "items splitting" for context-aware collaborative filtering [4, 5]. They divided rating data for each item into several subsets according to the contexts of the rating. Then the difference of distribution of the ratings in the subsets is evaluated by statistical measures such as information gain [11] and t-stet statistics [6]. If the difference between two distributions is larger than some threshold, they split the item depends on the contexts.

In order to clarify the detailed structure of difference between supposed context data and real context data, we applied their idea to our dataset and observed the differences of rating distributions in three contexts. We measured the difference between ratings in full and normal, normal and hungry, and full and hungry. We employed two measures, one is the information gain and another is the statistics for Kolmogorov-Smirnov test for comparing two probability distributions.

The information gain is popularly used as a criterion for splitting a set into two sets in the decision tree construction. When a set S of observed value is divided into two sets S_1 and S_2, the distribution of observed value also changes from P to P_1 and P_2. Then the information gain of the splitting is defined as

$$H(P) - (\frac{w_1}{w} H(P_1) + \frac{w_2}{w} H(P_2))$$

where $H(P)$ is the entropy of P, w_1 and w_2 are the number of elements of S_1 and S_2 respectively, and $w = w_1+w_2$ is the number of elements of S. In our case, we computed the information gain for the splitting rating data in two contexts (e.g. full and normal) into two sets according to the context. When the value of the information gain is larger, the distributions of two datasets are more different.

The Kolmogorov-Smirnov test is a test to determine if two datasets differ significantly. Contrary to the t-test which assume that the distribution of the target data is Gaussian, K-S test does not make assumptions about distribution of the target data and is suitable for applying to our data. When the statistics (Z-value) is larger, the distributions of two datasets are more different.

First we measured the differences depending on contexts. The results are shown in Table 3. The upper value in a low is the information gain, the lower value is the Z-score of the K-S test. * and ** mean that the difference between two datasets is significant in 0.05 and 0.01 levels, respectively. Then we computed the difference for each food menu item also. Table 4 shows the results. Each low corresponds to a food menu.

4. DISCUSSION

Table 3 clearly shows that the difference of the rating distributions between different degrees of hungriness is much larger in the supposed contexts than in the real contexts. We applied Kruskal-Wallis test for the difference of 3 groups also and confirmed that the difference is larger in the supposed contexts.

In addition, Table 4 shows that the difference depending on the hungriness varies according to the food menu. It is naturally expected that the difference between full condition and hungry condition is larger than the difference between both full-normal and normal-hungry conditions. As is shown by K-S test values in Table 4, in the supposed contexts data, these inequality relations are satisfied for 16 menus among 20 menus but for only 9 menus in the real context data.

These observations mean that the differences of preference ratings depending on the hungriness in supposed contexts distribute simpler than in the real contexts. The preferences in the real context behave in much more complex manner.

Table 3: The Difference between Distributions of Ratings in Different Hungriness for the Whole Data

	Full/Normal	Normal/Hungry	Full/Hungry
Real	0.023 1.756**	0.007 3.102**	0.041 3.630**
Supposed	0.775 6.397**	0.050 4.692**	0.129 8.242**

There could be several explanations of these observations. As an example, when the subjects are requested to put themselves in another context, they could tend to believe that the context makes a difference even if not. In other words, the subjects may tend to rate foods in the supposed contexts with logical inferences using simple rules from commonsense knowledge such as "When I am hungry I want to eat foods more than when I'm full." instead of imaging that he is hungry.

These results suggest that the ratings acquired in supposed contexts may have different nature from ratings in real contexts. Researcher should take care of the differences and it may be better to give some treatments to reduce the differences. For example, designing questionnaire to prevent from using the commonsense knowledge and to promote to remember the previous specific situation is thought to be effective.

5. CONCLUSION AND FUTURE WORK

We analyzed the corresponding data of food menu preferences acquired both in supposed contexts and real contexts, and found that preferences in the supposed contexts distribute much simpler than in the real contexts. These results suggest that the ratings acquired in supposed contexts may have different characteristics from the ratings in real contexts.

This result suggests that when asking about preference in supposed contexts, some methods preventing subjects from answering by using knowledge should be considered. As an example, using more concrete or detailed description of specific contexts may be effective to help subjects image the contexts easier. Confirming the effectiveness of such methods should be investigated in future. Explaining the observed differences in the light of a behavior model of the subjects is also an important research issue.

This work is very preliminary and the results may change depending on the form of questionnaires, the context attributes, and the domain of the target items/contexts. We want to execute other questionnaires to clarify the effect of those factors and analyze the data to find stable structures in user ratings.

We are also investigating model adaptation methods which can combine supposed contexts data and real contexts data in order to get more accurate preference predictions in real contexts [9, 10]. We would like to incorporate the findings in this work into the model adaptation procedures to get better recommendation with less real contexts data.

6. ACKNOWLEDGMENTS

We are grateful to Dr. Shigeyuki Akiba, President and CEO of KDDI R&D Laboratories, Inc. for his continuous support of this study. This work was supported in part by JSPS KAKENHI 20650030..

7. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734-749, 2005.

Table 4: Difference between the Distribution of Ratings in Different Hungriness for Real and Supposed Data

	Real Contexts			Supposed Contexts		
Menu	Full/Normal	Normal/Hungry	Full/Hungry	Full/Normal	Normal/Hungry	Full/Hungry
Chicken Steak Set	0.069 0.777	0.072 0.920	0.142 1.383*	0.062 1.130	0.182 1.551*	0.161 1.807**
Beef Yakiniku Set with Mashed Horseradish	0.084 0.547	0.094 1.276	0.128 1.169	0.184 1.770**	0.082 0.936	0.245 1.952**
Baked Pork and Cabbage with Chinese Miso Taste	0.041 0.713	0.061 0.677	0.144 1.497*	0.130 1.814**	0.156 1.355	0.233 2.076**
Mashed Tuna Bowl	0.026 0.601	0.039 0.619	0.003 0.458	0.086 0.807	0.070 0.773	0.175 1.514*
Pork Cutlet Bowl	0.062 0.416	0.072 0.634	0.103 1.002	0.267 2.080**	0.140 1.448*	0.359 2.560**
Beef Curry and Rice	0.049 0.595	0.047 1.521*	0.084 1.248	0.176 1.465*	0.196 1.104	0.182 1.932**
Chinese Noodle in Spicy Soup	0.028 0.733	0.026 0.989	0.052 0.784	0.067 1.664**	0.029 0.773	0.124 1.993**
Chinese Noodle in Soy Source Soup	0.025 0.404	0.037 0.915	0.050 1.289	0.082 1.621**	0.044 0.855	0.124 2.021**
Japanese Wheat Noodle in Curry Taste Soup	0.019 0.862	0.014 1.198	0.050 1.249	0.133 1.952**	0.100 1.225	0.160 2.302**
Beef Stew Set	0.038 0.818	0.037 0.922	0.047 1.172	0.139 2.087**	0.057 0.891	0.188 2.409**
Tempura Bowl	0.030 0.706	0.020 1.150	0.070 1.537*	0.202 2.634**	0.136 2.001**	0.341 3.061**
Pasta Bolognese with Eggplant	0.018 0.704	0.044 0.545	0.093 0.353	0.011 0.445	0.026 0.601	0.040 0.939
Pasta with Salted Cod Roe	0.026 0.733	0.014 0.529	0.051 0.945	0.173 2.292**	0.014 0.357	0.115 1.937**
Simmered Sole Set	0.069 1.039	0.074 0.600	0.033 0.900	0.205 1.925**	0.091 1.145	0.136 1.852**
Baked Salmon with Mashed Horseradish Set	0.046 0.684	0.025 1.693*	0.038 1.092	0.101 1.411*	0.095 1.554*	0.122 1.464*
Simmered Mackerel in Miso	0.056 0.923	0.008 0.725	0.049 0.838	0.087 1.493*	0.105 1.428*	0.143 1.874**
Chinese Boiled Tofu and Meat in Chili Source	0.045 0.507	0.007 0.553	0.033 0.551	0.089 1.555*	0.132 1.501*	0.187 2.086**
Chinese Fried Rice	0.034 1.112	0.038 1.064	0.085 1.635**	0.136 2.112**	0.070 1.440*	0.179 2.513**
Japanese Wheat Noodle with Beef	0.032 0.452	0.014 0.862	0.043 0.712	0.059 1.572*	0.059 1.181	0.140 2.304**
Japanese Soba Noodle with Mushroom	0.043 0.467	0.028 0.489	0.016 0.251	0.017 0.554	0.068 1.048	0.076 0.827
Average	0.042 0.690	0.039 0.894	0.066 1.003	0.120 1.619	0.093 1.158	0.172 1.971

[2] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 23(1):103-145, 2005.

[3] S. S. Anand and B. Mobasher. Contextual recommendation, In *Lecture Notes in Artificial Intelligence* 4737, 142-160, Springer-Verlag, 2007.

[4] L. Baltrunas and F. Ricci. Context-based splitting of item ratings in collaborative filtering. *Proceedings of the 3rd ACM Conference on Recommender Systems*, 245-248, 2009.

[5] L. Baltrunas and F. Ricci. Context-dependent recommendations with items splitting, *Proceedings of the 1st Italian Information Retrieval Workshop*, 2010. http:ims.dei.unipd.it/websites/iir10/index.html.

[6] L. Breiman, H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Statistics/Probabilities Series, Wadsworth Publisher Company, 1984.

[7] C. Ono, M. Kurokawa, Y. Motomura, and H. Asoh. A. Context-aware movie preference model using a Bayesian network for recommendation and promotion. In *11th International Conference, UM 2007, Corfu, Greece, July, 2007, Proceedings, Lecture Notes in Computer Science* 4511, 247-257, Springer-Verlag, 2007.

[8] C. Ono, M. Kurokawa, Y. Motomura,.and H. Asoh. Implementation and evaluation of a movie recommender system considering both users' personality and situation. *IPSJ Journal*, 49(1):130-140, 2008 (in Japanese).

[9] C. Ono, Y. Takishima, Y. Motomura, H. Asoh, Y. Shinagawa, M. Imai, and Y. Anzai. Context-aware users' preference models by integrating real and supposed situation data. *IEICE Transactions on Information and Systems*, E91-D(11):2552-2559, 2008.

[10] C. Ono, Y. Takishima, Y. Motomura, and H. Asoh. Context-aware preference model based on a study of difference between real and supposed situation data. *In User Modeling, Adaptation, and Personalization, 17th International Conference, UMAP2009 formerly UM and AH, Trento, Italy, June 2009, Proceedings, Lecture Notes in Computer Science* 5535, 102-113, Springer-Verlag, 2009.

[11] J. R. Quinlan, J. R. *C4.5: Programs for Machine Learning*. Morgan Kafuman Series in Machine Learning, Morgan Kaufmann, 1993