Model Adaptation with Bayesian Hierarchical Modeling for Context-Aware Recommendation

Hideki Asoh
AIST
AIST Tsukuba Central 2
1-1-1 Umezono, Tsukuba
Ibaraki 305-8568 Japan
h.asoh@aist.go.jp

Yoichi Motomura
AIST
AIST Tokyo Waterfront
2-3-26 Aomi, Koutouku
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

Model adaptation is a process of modifying a model trained with a large amount of training data from the source domain to adapt a specific similar target domain by using a small amount of adaptation data regarding the target domain. Bayesian hierarchical modeling is well known as a general tool for model adaptation and multi-task learning, and widely used in various areas such as marketing, ecology, medicine, education, and so on in order to model the heterogeneity in the phenomena. In this work, we propose to apply the Bayesian hierarchical modeling to the problem of preference modeling, where a model trained with a large amount of supposed context data is adapted to the real context by using additional small amount of real context data. The effectiveness of the proposed method is evaluated by experiments using context-aware food preference data.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

General Terms

Experimentation, Human factors, Measurement

Keywords

Model Adaptation, Preference Modeling, Context Awareness

1. INTRODUCTION

Modeling users' preferences is an important element of recommender systems. We have constructed several context-aware attribute-based recommender systems. The systems use Bayesian networks for modeling users' preferences [2, 16]. In the course of the construction, collecting large amount of data about users' preference through inquiries is necessary. In particular, to make the model context-aware, users' preference data should be collected under various contexts. However, putting subjects of inquiries into various contexts and collecting answers from them is often difficult and costs much. Hence, 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 the real contexts and the supposed contexts, the differences are not taken seriously.

In our previous works, we collected users' preferences of various dishes in both real and supposed contexts and showed that the difference is statistically significant and not negligible [17]. We also analyzed the statistical nature of the differences and demonstrated that the structure of preferences in supposed contexts is simpler than that of the preferences in real contexts [3]. These studies suggested that it is dangerous to construct preference models using data collected only in supposed contexts.

In this work, we pursue the possibility to construct better preference models by combining data in the supposed contexts and the real contexts. Although there are differences between the preferences in the real contexts and in the supposed contexts, they are similar in some extent, and the cost to collect data in the supposed contexts is much cheaper than in the real contexts. Hence, if we can modify a model constructed by a large amount of supposed context data to adapt to the real contexts by using small amount of real context data, it helps much to realize better context-aware recommender systems with smaller cost.

This kind problems are known as "model adaptation", "learning to learn", "transfer learning", or "multi-task learning" in the area of the statistical machine learning, and studied actively in recent years [6, 13, 15, 20]. In the area, the methods to have good learning results (statistical models of data) by combining data in different but similar domains. Typical examples are acoustic model adaptation and language model adaptation in speech recognition systems [11, 9, 19]. The collaborative filtering can also be considered as a case of multi-task learning [24].

There has been proposed several methods for model adaptation. In this work, we will exploit the methods using Bayesian hierarchical modeling [7, 8] because the simple and natural nature of the method. We will construct a hierarchical model for preference model adaptation by combining real and supposed context data, and evaluate the model using the food preference data.

The rest of the paper is organized as follows. Section 2 briefly introduces the Bayesian hierarchical modeling and formulates our model for model adaptation in context-aware preference modeling. Section 3 describes experiments using food preference data, and Section 4 is for conclusion and future work.

CARS-2011, October 23, 2011, Chicago, Illinois, USA.

2. BAYESIAN HIERARCHICAL MODELING

Bayesian hierarchical modeling is an effective method for simultaneous estimation of several parameters over similar domains, and is used to capture heterogeneity of subjects in areas such as marketing and ecology [5, 12, 18].

We have already proposed to apply the following simple linear Gaussian hierarchical model to the problem of constructing context-aware preference model which can model and predict ratings r_{ucs} by users u for items c in contexts s [4].

$$
\begin{aligned}
r_{ucs} &\sim \mathrm{normal}(\mu_{ucs}, 1/\tau),\\
\mu_{ucs} &= \mu_0 + a_u + b_c + c_s,\\
\tau &\sim \mathrm{gamma}(\nu, \theta),\\
\mu_0 &\sim \mathrm{normal}(\mu, \sigma^2),\\
a_u &\sim \mathrm{normal}(0, 1/\tau_a),\\
b_c &\sim \mathrm{normal}(0, 1/\tau_b),\\
c_s &\sim \mathrm{normal}(0, 1/\tau_c),\\
\tau_a &\sim \mathrm{gamma}(\nu, \theta),\\
\tau_b &\sim \mathrm{gamma}(\nu, \theta),\\
\tau_c &\sim \mathrm{gamma}(\nu, \theta).
\end{aligned}
$$

Here, $\mathrm{normal}(\mu, 1/\tau)$ means Gaussian distribution with mean μ and variance $1/\tau$, and gamma means Gamma distribution.

In this paper, we will extend the above model for model adaptation by combining real and supposed context data as follows:

$$
\begin{aligned}
r^{(r)}_{ucs} &\sim \mathrm{normal}(\mu^{(r)}_{ucs}, 1/\tau),\\
r^{(s)}_{ucs} &\sim \mathrm{normal}(\mu^{(s)}_{ucs}, 1/\tau),\\
\mu^{(r)}_{ucs} &= \mu^{(r)}_0 + a^{(r)}_u + b^{(r)}_c + c^{(r)}_s,\\
\mu^{(s)}_{ucs} &= \nu^{(s)}_0 + a^{(s)}_u + b^{(s)}_c + c^{(s)}_s,\\
\tau &\sim \mathrm{gamma}(\nu, \theta),\\
\mu^{(r)}_0 &\sim \mathrm{normal}(\mu, \sigma^2),\\
\mu^{(s)}_0 &\sim \mathrm{normal}(\mu, \sigma^2),\\
a^{(r)}_u &\sim \mathrm{normal}(0, 1/\tau_a),\\
b^{(r)}_c &\sim \mathrm{normal}(0, 1/\tau_b),\\
c^{(r)}_s &\sim \mathrm{normal}(0, 1/\tau_c),\\
a^{(s)}_u &\sim \mathrm{normal}(0, 1/\tau_a),\\
b^{(s)}_c &\sim \mathrm{normal}(0, 1/\tau_b),\\
c^{(s)}_s &\sim \mathrm{normal}(0, 1/\tau_c),\\
\tau_a &\sim \mathrm{gamma}(\nu, \theta),\\
\tau_b &\sim \mathrm{gamma}(\nu, \theta).\\
\tau_c &\sim \mathrm{gamma}(\nu, \theta),
\end{aligned}
$$

Here, $r^{(r)}_{ucs}$ denotes a rating in a real context, and $r^{(s)}_{ucs}$ denotes corresponding rating in a supposed context. This model is composed of two hierarchical context-aware preference models, the generative model of the real context data and the supposed context data. They are connected through common hyper-hyper parameters $\tau, \tau_a, \tau_b, \tau_c$. Through the common hyper-hyper parameters, information in the ratings in supposed contexts can affect to the posterior probability distribution of predicted ratings in the real context model.

3. EXPERIMENTS

We applied the proposed model to our context-aware food preference data and evaluated the accuracy of predicted ratings in the real contexts for unknown cases.

3.1 Data acquisition and preparation

In our previous work [17], we designed an internet questionnaire survey in order to collect corresponding data, that is, we asked subjects the same question about food preference both in real and supposed contexts and collect pairs of answers. The target contents were typical dishes served in food courts.

The survey was composed of two questionnaire surveys. The first questionnaire survey was conducted from 16th to 17th in December 2008. The number of subjects was 746, each subject evaluated 5 kinds of a la carte dishes randomly selected from 20 kinds of dishes such as "chicken steak", "beef steak", "beef curry", "pasta with cod roe", "Japanese noodle", etc. using 5-grade rating scale from "I do not want to order the dish at all" to "I want to order the dish very much". At the same time the subjects answered the current degree of hunger in 3 levels (hungry, normal, full).

After that, the subjects are asked to imagine that they are in the different degree of hunger from the current, and answered the preference for the same 5 dishes. In total, preferences for 5 dishes in three different contexts (degree of hunger) are collected. Among the three contexts, one is real and two are supposed.

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 extracted subjects who answered different degree of hunger from the first survey. After filtering out unreliable subjects, the number of extracted subjects was 212.

By combining the result of two surveys, we got corresponding preference for 5 dishes in 2 different degree of hunger per a subject. Hence the number of total ratings was 2,120. Figure 1 shows the whole data set. Figure 2 shows examples of answerers in two surveys, and examples of combined corresponding data.

We divided the dataset into training data and test data. First, we randomly left one real context rating out of the 10 ratings of each subject for evaluation. The rest of the 9 rat-

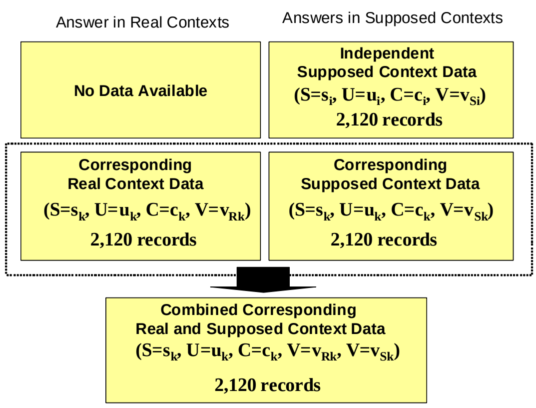

Figure 1: Structure of the whole data set [17]

1st Survey

Subject	Food	Real	Supposed	Preference
1	Noodle	Full	×	3
1	Noodle	Full	Normal	2
1	Noodle	Full	Hungry	2

2nd Survey

Subject	Food	Real	Supposed	Preference
1	Noodle	Hungry	×	1
1	Noodle	Hungry	Normal	1
1	Noodle	Hungry	Full	2

Subj.	Food Menu	Context	Real Preference	Supposed Preference
1	Noodle	Full	3	2
1	Noodle	Hungry	1	2
3	Steak	Full	2	3
3	Steak	Normal	2	2
1	Fried Rice	Hungry	1	2
1	Fried Rice	Full	2	3

Figure 2: Examples of ratings in two surveys and combined corresponding data [3]

ings per a subject in real contexts are used as training data. In order to evaluate the effect of the number of real context training data for constructing preference model, we change the number of real context ratings per a subject which are used for model construction from 0 (supposed only) to 9. For the supposed context data, all 10 ratings per a subject are used for model construction.

We repeated the experiment 10 times with different division of the real context data and evaluated the accuracy of the predicted ratings for the left out test data in real contexts. We evaluated average and standard deviation of the mean squared error (MSE) of the predictions. We also evaluate the prediction accuracy of the model constructed with only real context data.

Experiments were conducted with the open source statistical computing software R and software for Bayesian Monte Carlo simulation WinBUGS [14, 22]. For connecting R to WinBUGS, we used R package R2WinBUGS. We set $\mu = 2.0, \sigma = 10, \nu = 2.0, \theta = 1.0$, however, the results are robust with respect to the values of these parameters.

3.2 Result and discussion

Table 1 shows the average and standard deviation of MSE for various values of the number of real context data $L = 0, ..., 9$. Standard deviations are depicted in brackets. We also visualize the average of the MSE values in Figure 2.

This results demonstrate that as the number of real context data L increases, the MSE of the predicted rating in real context decreases monotonically. Hence, model adaptation by combining a small amount of real context data with a large amount of supposed context data is verified to be effective.

In particular, the performance for $L = 1$, that is, constructing model with 10 supposed context data + 1 real context data, is much better than the performance for $L = 0$, that is, constructing model only with supposed context data. This demonstrates the facts that

Table 1: Average of mean squared eerror for 10 experiments

L	Supposed Context Data + Real Context Data	Only Real Context Data
0	1.95 (0.10)	—
1	1.66 (0.17)	1.71 (0.20)
2	1.50 (0.14)	1.49 (0.14)
3	1.51 (0.11)	1.51 (0.12)
4	1.48 (0.12)	1.48 (0.12)
5	1.43 (0.13)	1.43 (0.12)
6	1.43 (0.13)	1.42 (0.11)
7	1.42 (0.11)	1.42 (0.11)
8	1.40 (0.12)	1.39 (0.13)
9	1.40 (0.12)	1.39 (0.13)

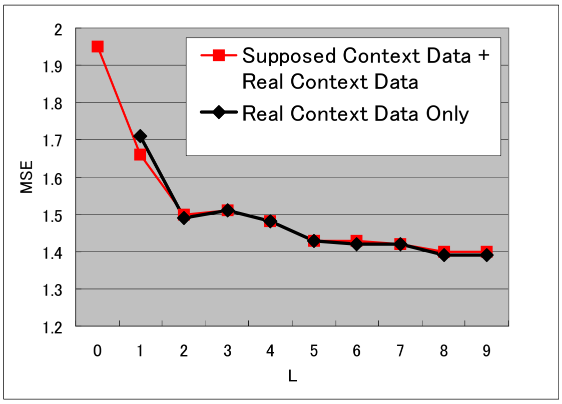

Figure 3: Effect of Model Adaptation

- Constructing preference model with only supposed context data is dangerous,
- Very small amount of real context data can improve the model.

However, the results showed that the models constructed with only small number of real context data perform rather well also. Even using only 2 real context data per a subject, the performance of the model is almost equal to the performance of the model constructed by combining supposed and real context data. This is because Bayesian hierarchical models is able to make robust prediction even though the number of training data is very small. This means that using the supposed context data is effective only for $L = 1$ (cold start) case.

4. CONCLUSION AND FUTURE WORK

In this paper, we propose to apply Bayesian hierarchical modeling to preference model adaptation by combining real and supposed context data. The results of the experiments with food preference data demonstrate that the model adaptation is effective in particular for the cases where very small amount of real context data is available. This means that the model adaptation provides a solution for the cold start problem in context-aware recommender systems. Note that Umyarov and Tuzhillin observed a very similar phenomena in different context. They showed a small amount of aggregated external rating data can significantly improve the performance of a Bayesian hierarchical preference model [21].

There are several future works. The first one is more intensive evaluation. In this paper we evaluated the method with small scale dataset. As the number of users, items and contexts increases, the more training data is necessary for constructing good preference models. Hence the importance of the model adaptation is expected to increase. Evaluating with data in domains other than food preference is also important.

The second one is to apply the method to different base models. The model adaptation technique with Bayesian hierarchical modeling is independent from the generative model of ratings. In this work, we used the simple linear Gaussian model of rating generation. More elaborated generative models of ratings such as probabilistic tensor factorization model [10, 23] can be used instead. Using generative models for ordered ratings may be effective also.

The third one is to investigate other model adaptation techniques. The proposed model adaptation technique is bi-directional. This means that the combined model is symmetric for source and target domains. Investigating more directional model adaptation techniques is an interesting future work.

Acknowledgments.

We thank Dr. Yasuyuki Nakajima, President and CEO of KDDI R&D Laboratories Inc., for his continuous support of this study. This work was supported in part by JSPS KAKANHI 20650030.

5. REFERENCES

[1] A. Ansari, S. Essegaier, and R. Kohli. Internet recommendation systems. Journal of Marketing Research, vol. 37, no. 3, 2000.

[2] H. Asoh, C. Ono and Y. Motomura. A movie recommendation method considering both users' personality and situation. In Proceedings of the ECAI2006 Workshop on Recommender Systems, pp. 45–48, 2006.

[3] H. Asoh, C. Ono and Y. Motomura. An analysis of differences between preferences in real and supposed contexts. In Proceedings of 2nd Workshop on Context-Aware Recommender Systems (CARS-2010), 2010.

[4] H. Asoh, C. Ono, and Y. Motomura. A Bayesian hierarchical preference model for context-aware recommendations, Adjunct Proceeding of UMAP 2010, 2010.

[5] P. Congdon. Bayesian Statistical Modelling, Second Edition. Wiley, 2006.

[6] H. Daume III, D. Marcu. Domain adaptation for sitasitical classifiers. Journal of Artificial Intelligence Research, vol. 26, pp. 101–126, 2006.

[7] H. Daume III. Bayesian multitask learning with latent hierarchy. In Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI), 2009.

[8] J. R. Finkel, C. D. Manning. Hierarchical Bayesian domain adaptation. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL, pp.602–610, 2009.

[9] D. Gildes and T. Hofmann. Topic-based language models using EM. In Proceedings of 6th European Conference on Speech Communication and Technology (Eurospeech 99), pp.2167–2170, 1999.

[10] A. Karatzoglou, X. Amatriain, N. Oliver, and L. Baltrunas. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In Proceedings of ACM Recommender Systems 2010, 2010.

[11] C. Leggetter, P. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Computer, Speech and Langage, vol. 9, no. 2, pp.171–185, 1995.

[12] M. A. McCarthy. Bayesian Methods for Ecology. Cambridge University Press, New York, 2007.

[13] NIPS2005 Workshop Inductive Transfer: 10 Years Later, http://iitrl.acadiau.ca/itws05/

[14] I. Ntzoufras. Bayesian Modeling Using WinBUGS. Wiley, 2009.

[15] S. J. Pan and Q. Yang. A survey of transfer learning. IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[16] 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, LNCS vol. 4511, pp. 247–257, Springer-Verlag, 2007.

[17] C. Ono, Y. Takishima, Y. Motomura and H. Asoh. Context-aware prefence model based on a study of difference between real and supposed context data. In User Modeling, Adaptation, and Personalization, 17th International Conference, UMAP2009, Proceedings, LNCS vol. 5535, pp. 102–113, 2009.

[18] P. E. Rossi, G. M. Allenby and R. McCulloch. Bayesian Statistics and Marketing. Wiley, 2005.

[19] T. Takiguchi. Statistical Acoustic Model Adaptation for Robust Speech Recognition in Noisy Reverberant Environments. Doctral Thesis, Nara Institute of Science and Technology, 1999.

[20] S. Thrun and L. Pratt (eds.). Learning to Learn, Kluwer Academic Publishers, 1998.

[21] A. Umyarov and A. Tuzhilin. Using External Aggregate Ratings for Improving Individual Recommendations, ACM Transactions on the Web, vol. 5, no. 1, article 3, 2011.

[22] http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.

[23] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. Carbonell. Temporal collaborative filtering with Bayesian probabilistic tensor factorization. In Proceedings of SIAM Data Mining 2010 (SDM 10), 2010.

[24] K. Yu and V. Tresp. Learning to learn and collaborative filtering. In NIPS 2005 Workshop "Indctive Transfer: 10 Years Later", 2005.