International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:4, No:4, 2010

Context-aware Recommender Systems using Data Mining Techniques

Kyoung-jae Kim, Hyunchul Ahn, and Sangwon Jeong

Abstract**—**This study proposes a novel recommender system to provide the advertisements of context-aware services. Our proposed model is designed to apply a modified collaborative filtering (CF) algorithm with regard to the several dimensions for the personalization of mobile devices – location, time and the user's needs type. In particular, we employ a classification rule to understand user's needs type using a decision tree algorithm. In addition, we collect primary data from the mobile phone users and apply them to the proposed model to validate its effectiveness. Experimental results show that the proposed system makes more accurate and satisfactory advertisements than comparative systems.

Keywords**—**Location-based advertisement, Recommender system, Collaborative filtering, User needs type, Mobile user.

I. Introduction

RECOMMENDER systems are now regarded as an essential marketing tool in the e-commerce environment, because many customers who use e-commerce suffer serious information overload. Recommender systems may filter and provide essential and useful information to customers in e-commerce. As the use of e-commerce tends to extend their business area to mobile or ubiquitous environments, the importance of the recommender systems is increasing.

Of the many kinds of recommender systems, context-aware advertising has been a driver of third-generation mobile operators' marketing efforts over the past few years. Thus, many studies on location-based marketing or advertising have been proposed. However, these approaches have two common shortcomings. First, most of them just suggested theoretical architectures that are too abstract to apply to the real-world cases. Second, many of these approaches only consider the benefit to service providers rather than that to customers. To mitigate these limitations, this study proposes a new type of recommender systems that can be applied for recommending the advertisements of context-aware services to mobile users.

The proposed model in this study utilizes a collaborative filtering (CF) algorithm, but modifies the conventional CF algorithm using the multi-dimensional personalization model proposed by Schilke et al. [1] for enabling context-aware advertising to mobile users. Schilke et al. [1] proposed three dimensions, location, time, and interest, for personalization of mobile users. The location and time dimensions use information about the user's position and time from mobile devices. The interest dimension considers user preferences to match relevant products or services.

We can easily capture information about location and time, because mobile devices always communicate with the base station of the service provider. However, the third dimension, interest, is rarely captured, because interests of users differ and are diverse. In this study, we automate the process to capture user needs or interests using data mining techniques.

In addition, we also modify the user-item matrix to reflect three dimensions for the multi-dimensional personalization model. CF generally works by building the user-item matrix; this represents the satisfaction levels of users for items. However, CF in this study should reflect other additional information, such as location, time, and user's needs type in order to generate personalized recommendations for mobile users. Consequently, we propose a novel CF with a modified user-item matrix that makes recommendations for mobile users taking into account location, time and needs type. We obtain real-world data of mobile users and perform experiments to validate the usefulness of the proposed model.

The remainder of the paper is organized as follows. Section 2 reviews the basic process of conventional collaborative filtering and the concept of user's needs type. Section 3 proposes our recommendation model – Location-based Advertisement Recommender for Mobile Users (LARMU). Section 4 presents the explanation for the experimental design and results. In the final section, the conclusions of the study are presented.

II. Prior Research

A. Collaborative Filtering

A recommendation system can be described as an automated decision support system needed to provide personalized recommendations without going through a complicated search process. It utilizes user's requirements and preferences to help their selection process of available products or services. There are two popular types of recommender systems: content-based and collaborative filtering.

The recommendation system with the content-based (CB) approach analyzes the items, content, and creates a profile representing a user's interest in terms of items. Then, it reviews the content of items that are unknown to the user, and compares them against the user's profile. It finally recommends results

Kyoung-jae Kim is with Dongguk University, Seoul, South Korea (phone: +82-2-2260-3324; fax: +82-2-2260-3684; e-mail: kjkim@ dongguk.edu).
Hyunchul Ahn is with Kookmin University, Seoul, South Korea.
Sangwon Jeong is with Dongguk University, Seoul. South Korea.

with the new items that are likely to satisfy the user's preference [2]. CB has much strength but has an important limitation. CB may only be useful in a few domains, such as recommendation of news articles or web pages, because it is never easy to profile the items without human intelligence [3].

The CF method, however, mitigates the limitation of the CB approach by recommending items that other people with similar preference have liked [4]. First, a CF collects user ratings of items in question. Then, users are compared based on how similar their ratings are, and the CF system recommends the items favored by others who have similar preference.

The CF predicts the preference of an active user for a target item based on specific user preference. Generally, there are two types of CF approaches: memory-based and model-based. The former repeatedly searches the user-item matrix to choose peer groups for an active user. Then, a prediction result is computed by weighting the votes of users from the peer groups. The users in the peer groups are identified according to their similarity or nearness in preferences to the active user. The latter, the model-based CF approach, builds a user model from the database of rating history. Then, the predictions are performed utilizing the user model. This approach predicts the result in a shorter time compared to the former. However, it has a limitation in that it requires more time to search the dataset. In addition, it may be not suitable for situations in which user preference models must be updated rapidly or frequently [5]. Thus, we will focus on the memory-based CF algorithm in this study. There are two advantages. It does not require effort to characterize product items, and it can produce serendipitous recommendations, because it uses information on item content. Moreover, it is appropriate for environments that require recent recommendation results, because its database is updated continuously.

The recommendation process of the memory-based CF algorithm follows: The first step is the process of similarity calculation. Pearson correlation coefficient is used as a similarity measure between an active user and his/her neighbor.

The next step is neighbor selection. This step selects the n neighbors who have the highest similarity to the active user. The similarities calculated in the first step are used as a criterion for choosing nearest neighbors.

The thirst step predicts the active user's unanswered rating from a combination of the ratings of selected neighbors.

B. User's Needs Type

A need is something that is necessary for humans' healthy life [http://www.wikipedia.org]. In general, needs are defined as requirements for something essential or desirable that is lacking. That is, needs are the most basic factors and the starting point of the generating process of behavioral outcomes. Therefore, understanding a user's needs is quite important to satisfy the user. Prior research in the marketing context has identified numerous kinds of needs that influence the process of stimulating user behavior. However, we may classify them into two types: utilitarian and hedonic [6].

Utilitarian needs are explained as requirements for products that remove or avoid problems with life, while hedonic ones are requirements for products that provide social or aesthetic utility to users. For example, a user who participates in an online social network to obtain useful information for his/her life has utilitarian needs, but the user has hedonic needs when he/she uses it for a social relationship or amusement. Users are generally conscious of the needs stimulated by advertisements. Thus, advertisers can use utilitarian or hedonic appeals to stimulate users' utilitarian or hedonic needs.

III. Proposed Recommender System

In this study, we propose a new type of recommender system that can be applied to recommend the advertisements of location-based services to mobile users. For convenience, we call our recommender system LARMU (Location-based Advertisement Recommender for Mobile Users). LARMU is based on a collaborative filtering (CF) algorithm. In general, CF works by building a user-item matrix that represents the satisfaction levels of users for items. However, as explained in the previous section, CF in LARMU should reflect other additional information, location, time, and user's needs type to generate personalized recommendations for mobile users. Consequently, we develop a novel user-item matrix for our proposed system. Fig. 1 shows the schema of LARMU's user-item matrix.

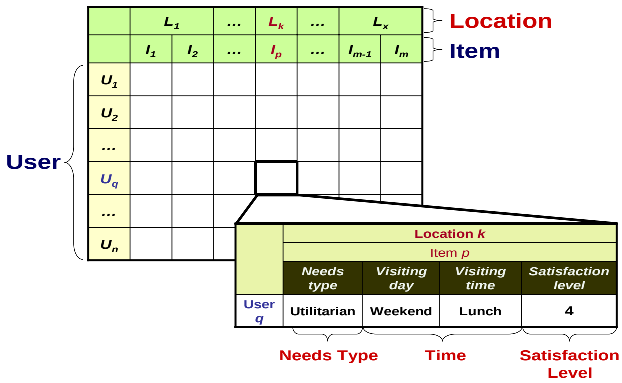

Fig. 1 The schema of LARMU's user-item matrix

A user-item matrix of conventional CF only contains information on user, item, and corresponding satisfaction level. However, our new scheme includes three additional dimensions – location, time and needs type as presented in Fig. 1. Here, we adopt 'location' as a higher dimension of items, because the items of our study are commercial spots located in a specific area (i.e. location). Thus, when implementing LARMU, the information on location may be used to filter items. In the case of 'time', we specify it into two factors – visiting day and visiting time. In general, the commercial spots that we visit may change according to the day of the visit (e.g. weekday or weekend) and the time frame of the visit (e.g. morning, lunch time, afternoon, or dinner time). Therefore, we design our recommender system to consider both the day and the time of the visit independently. In addition, we include

ISSN: 2517-9411
Vol:4, No:4, 2010

'needs type' as one of the additional information that affects user's satisfaction level. For needs type, we consider three cases – (1) the target user has hedonic needs, (2) the target user has utilitarian needs, and (3) the target user has both types of needs.

Fig. 2 represents the LARMU system architecture. The recommendation procedure of LARMU consists of three steps. The detailed explanation for each step is presented as follows:

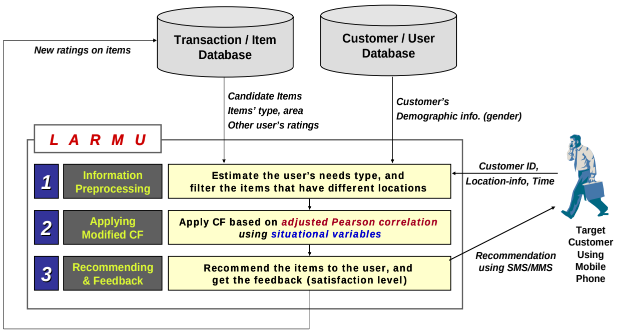

Fig. 2 LARMU system architecture

In the first step, the information about the target user is input to LARMU. This includes the identification code of the target user, his or her location (visiting area), and current date and time. All of these can be easily obtained automatically. The user's current need type is also required to generate recommendations using LARMU. However, it is very difficult and unrealistic to get the information on the user's current needs type by a direct inquiry of the user. Thus, LARMU is designed to estimate the user's needs type by using the information on the target user, and the background (location and time).

After collecting and estimating all the required information on the target user, LARMU searches for candidate items for recommendation. At this time, it filters items in the areas that are located in t areas far from the target user's current position. Thus, the system can reduce the search space dramatically; this results in reducing computation time.

After the first step, CF is performed to find similar neighbors to the target user, and to calculate the expected satisfaction level for the items that are outside the experience of the target user. In traditional CF, the similarity between the users is calculated using Pearson correlation of the satisfaction levels (ratings) between the users, as presented in Equation (1). Then, the expected satisfaction levels for the items of the target user are calculated by applying Equation (2). However, in LARMU, other additional information, such as day, time, and needs type should also be considered when calculating similarity and the expected satisfaction level. Thus, it adopted 'weighted Pearson correlation' when calculating similarity between users.

Therefore, the similarity between users can be improved by considering multi-dimensional factors, including day, time and needs type. Thus, more sophisticated recommendation results may be generated for mobile users.

In the last step, the optimal recommendation results produced by LARMU are provided to the target user. Then, the user will show the different types of responses. These are stored as a new source of data, and will be used to refine the next recommendation. The results can be transferred using SMS or MMS. If the system uses MMS, it may include more detailed information for the recommended places, such as a simple map, photos, bar codes for promotion, and long messages, although SMS can only send 80 alphabetical characters.

IV. Experiments

A. Experimental Design

We developed an experimental system that implements the core features of LARMU to validate its usefulness, applying it to real-world data. Although our proposed model is based on CF, we cannot use public datasets for CF, such as MovieLens or EachMovie, since we need additional information on location, time, and user's needs type to generate recommendations. Therefore, we built a Web-based data collection system to gather appropriate empirical data from users. This data collection system contained the places for shopping, eating, drinking, enjoyment, and learning in five major commercial locations of Seoul, South Korea. The system contained the information on 275 places.

The data collection system was designed to collect data including the visiting day, visiting time, user's needs type at the point of visit (the purpose of the visit), and the satisfaction level for these spots from mobile phone users. To simplify the input process, we discritize the candidate values of the input variables, as presented in Table 1. As shown in Table 1, we assign the numeric code in an interval scale to each candidate value of the most input variables (visiting time, needs type, and ratings). Consequently, it is possible to apply simple numeric operations for the inputted values.

TABLE 1
Description of the Input Variables

Dimension	Variable	Candidate values
Location	Commercial Zone (CZ)	Chongro Daehakro Shinchon / Ewha Univ. Kangnam Station Myungdong
Time	Visiting day (VD)	Weekday (Mon.-Fri.) Weekend (Sat./Sun.)
	Visiting time (VT)	Morning / AM08:00 – AM11:00 Lunch / AM11:00 – PM02:00 Afternoon / PM02:00 – PM05:00 Dinner / PM05:00 – PM08:00 Night / PM08:00 – PM11:00
Interest	Needs type (NT)	Hedonic (H) Utilitarian (U) Both (B)
Satisfaction level	Ratings (RT)	Very dissatisfied Dissatisfied Somewhat dissatisfied Neutral Somewhat satisfied Satisfied Very satisfied

International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:4, No:4, 2010

We collected the experimental from April to May 2006. In the two months, we collected 9980 ratings from 265 respondents in three universities in Korea. We eliminated some cases that were inappropriate, and finally selected 200 respondents and their ratings for 175 items for the experimental dataset.

In addition, as explained in the previous section, LARMU is designed to estimate the user's needs type by using the information of the target user, and the background (location and time). Thus, the algorithm to estimate the user's needs type should also be developed before applying the experimental dataset to our model. We adopted a decision tree technique for this. Decision trees are generally known to be good for handling discrete independent variables. In addition, they are easily applied, since the trees can be interpreted as a set of the simple IF-THEN rules. Thus, we applied the CART (classification and regression tree) algorithm, a famous decision tree technique. Five variables were used as independent variables – (1) visiting day, (2) visiting time, (3) location (area), (4) the user's gender, and (5) the type of the place (shopping mall, restaurant, etc.). The maximum depth of the tree was set to five. Fig. 3 presents the result of the tree. It has 14 leaf nodes that can be matched to 14 IF-THEN rules. 57% of the cases correctly predicted the needs type applying this prediction model.

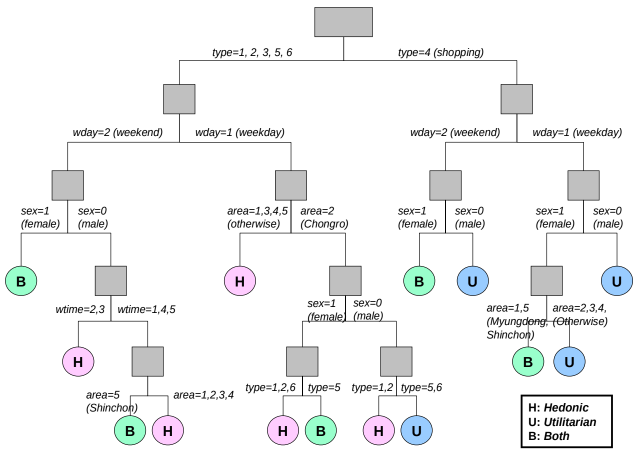

Fig. 3 The result of the CART for estimating the user's needs type

The experimental software to implement LARMU was developed using Microsoft Excel 2007 and its VBA (Visual Basic for Applications) feature. In addition, we used the 'Allbut1' method to predict the satisfaction level. This means that the test set for each test user contains a single randomly selected rating and the observed set contains the remaining ratings.

In addition, we also applied six comparison models to the same dataset to test the effectiveness of LARMU. The first model is the simple average (SimAVG) model. It is the simplest of the comparison models. This model is designed to predict a user's preference rating for a specific item to the average of all the ratings made by him or her [7].

The second model is the conventional CF approach (ConvCF). This model is designed to ignore all the information, except for the ratings. Thus, this model generates recommendation results only considering the pattern of ratings.

The third to sixth comparison models are the filtering models for each variable – visiting day (VD), visiting time (VT), actual needs type (ANT), and estimated needs type (ENT). In these models, the recommender only considers neighbors who have same value for each situational factor (i.e. visiting day, visiting time, or actual/estimated needs type) when calculating similarities between users. The efficiency of the model can be improved, since it does not consider other users that have different values for situational factors. However, this type of recommender model may suffer from the problem of information loss.

We also developed two different versions of the proposed model, LARMU. The first version termed 'LARMU-ANT' is designed to use 'actual needs type' that was directly inputted by the user whereas the second version 'LARMU-ENT' uses 'estimated needs type'. In the latter, the needs type is estimated using the decision tree in Fig. 4. Of these versions, the genuine proposed model is LARMU-ENT, since the needs type would be estimated rather than measured in practice.

B. *Experimental Results*

In this study, we set the average MAE (mean absolute error) as the criterion to evaluate the performance of the comparative models. MAE is frequently used in the CF literature, and represents the difference between the predicted and actual rating of users [8-10].

Table 2 shows overall results of the comparative models and our proposed model. Our proposed models, LARMU-ANT and LARMU-ENT, show the minimal average MAE among the comparative models. Thus, we may conclude that our model generates the most accurate prediction results in the recommendation for mobile users. We can also find that the comparative models that apply a filtering approach generally show unsatisfactory prediction accuracy. The reason of this phenomenon seems to be information loss due to filtering. We suspect that this may hinder the models from finding generalized patterns of various neighbors. Among the comparative models of VD, VT, ANT and ENT, the effect of actual needs type (ANT) is the most critical, and the one of visiting day (VD) is the least. From this, we can find that the consideration of the user's needs type is very important to recommend the appropriate location-based advertisements, although measuring or estimating it is not easy.

When comparing LARMU-ANT and LARMU-ENT, we can find that LARMU-ANT is better than LARMU-ENT from the perspective of 'prediction accuracy'. However, the gap between them is very small (about 0.0001). Thus, we may conclude that estimating needs type using decision trees has little impact on the effectiveness of the proposed recommender model.

TABLE II
THE RESULTS OF THE EXPERIMENTAL MODELS

NO	Model	Situational variables to be considered			Mean of MAE	S.D. of MAE
1	SimAVG	None			0.9488	0.2755
2	ConvCF	None			0.9161	0.2747
3	VD	Visiting day			0.9253	0.2730
4	VT	Visiting time			0.9604	0.2614
5	ANT	Actual needs type			0.9240	0.2741
6	ENT	Estimated needs type			0.9289	0.2754
7	LARMU-ANT	Visiting day	Visiting time	Actual needs type	0.9143	0.2747
8	LARMU-ENT	Visiting day	Visiting time	Estimated needs type	0.9144	0.2747

V. Conclusions

In this study, we proposed a novel recommender system for recommending the advertisements of the location-based services. LARMU, our proposed system, is designed to apply modified CF with consideration of the several dimensions for the personalization of mobile phone users – location, time and the user's needs type (interest). To validate the effectiveness of the proposed model, we collected primary data from the mobile phone users, and applied them to the proposed model. Experimental results showed that LARMU outperforms the conventional CF algorithm, as well as other comparative models. Moreover, we also found that estimating the user's needs type does not affect the prediction accuracy of LARMU significantly

Fig. 4 presents the sample scenario of the real-world application of LARMU. As presented in these figures, our recommender system can be applied by mobile service operators as a new business model. It delivers the target advertisements for location-based services to the mobile users who have given permission to the operator. As we all know, most advertisements of local places are thrown away. Most coupons that are distributed on the streets are also wasted, since they are not personalized. Nonetheless, the size of the local advertising market is immense, because many local stores need to promote themselves. We believe that our recommender model will be a good solution to mitigate this social inefficiency.

Fig. 4 LARMU sample usage scenario

Our study also has some limitations. First, the problem of data scarcity should be resolved. Although we collected data on the Web for about a month, the collected dataset had insufficient ratings. This may cause so-called 'sparsity problem', which means the problem of low-quality recommendations when the system has a few ratings of users, since the users' patterns for measuring the similarity between users become unclear. Thus, the average MAE of our experiment is quite high compared to other CF-related studies that use a denser data set. So, in the future, it is required to add some ideas to lessen the sparsity problem in LARMU.

Second, we currently set the relative importance (i.e. importance weights) of each situational variable (visiting day, visiting time, and needs type) to the same value (1/3).That is, we assumed that these components are equally important. However, this is not reasonable, as we indicated the levels of the effectiveness of each variable differ. Thus, how to determine the optimal weights of the situational variables can be a good research topic for future study.

Finally, the usefulness of LARMU should be validated in practice. The validation process in our study is quite restricted, because our model is not validated in the real-world mobile situation, although the experimental validation is performed using the data collected from real-world users. Thus, we hope to have a chance to implement and validate LARMU practically with a real-world mobile service provider in the future.

References

[1] Schilke, SW, Bleimann, U, Furnell, SM, Phippen, AD (2004) Multi-dimensional personalisation for location and interest-based recommendation. Internet Research 14(5):379-385
[2] Balabanovic, M, Shoham, Y (1997) Fab: Content-based, collaborative recommendation. Communications of ACM 40(3):66-72
[3] Billsus, D, Pazzani, MJ (1998) Learning collaborative information filters. In: Proceedings of the 15th ICML, Madison, WI, pp 46-54
[4] Lawrence, RD, Almasi, GS, Kotlyar, V, Viveros, MS, Duri, SS (2001) Personalization of supermarket product recommendations. Data Mining and Knowledge Discovery 5(1-2):11-32
[5] Schafer, JB, Konstan, J, Riedl, J (2001) Electronic commerce recommender applications. Data Mining and Knowledge Discovery 5(1-2):115-152
[6] Maclnnis, DJ, Jaworski, BJ (1989) Information processing from advertisements: Toward an integrative framework. Journal of Marketing 53(4):1-23

[7] Roh, TH, Oh, KJ, Han, I (2003) The collaborative filtering recommendation based on SOM cluster-indexing CBR. Expert Systems with Applications 25(3):413-423

[8] Breese, JS, Heckerman, D, Kadie, C (1998) Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence, San Francisco, CA, pp. 43-52

[9] Sarwar, BM, Konstan, JA, Borchers, A, Herlocker, J, Miller, B, Riedl, J (1998) Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. In: Proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW), Seattle, WA, pp. 345-354

[10] Goldberg, K, Roeder, T, Gupta, D, Perkins, C (2001) Eigentaste: a constant time collaborative filtering algorithm. Information Retrieval 4(2):133-151

[11] Green, SB, Salkind, NJ, Akey, TM (2000) Using SPSS for Windows. 2 Ed. Prentice Hall, NJ