A Recommendation Method Based on Contents and User Feedback

So-Ryoung Kim
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
soryoung@cs.yonsei.ac.kr

Sang-Min Choi
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
jerassi@cs.yonsei.ac.kr

Lae-Hyun Kim
Bionics Center, Korea
Institute of Science and Technology,
Seoul, Republic of Korea
laehyunk@kist.re.kr

Yo-Sub Han
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
emmous@cs.yonsei.ac.kr

***Abstract*-Nowadays, user is provided with many contents, which the previous search engines failed to find, thanks to various recommendation systems. These recommendation algorithms are usually carried out using collaborating filtering algorithm, which predicts user's preference, or contents based algorithm, which calculates on the basis of the similarity between contents. In addition to the above algorithms, many algorithms using user's context have been recently developed. Based on the previous researches, this paper proposes a new system to categorize contents information into various factors and learn user's selection. First, we divide information of items into four types and make user preference pattern using each information type. The information types can express more various user preferences and user preference pattern can calmly deal with user preference. Then, we calculate the score for recommendation using user preference pattern. That is, our system is constructed on these three modules: item analyzing module, user pattern analyzing module and recommendation score module. Lastly, we provide entire system flow to show how they work.**

Keywords-Recommend method; Learning algorithm; User Preference; Recommendation system.

I. INTRODUCTION

With the help of recommendation systems, people can access a wider range of contents, which was not possible with the old search engines. For example, when we listen to music or buy contents like DVD titles, books, and laptops, the recommendation systems provide a rich selection of items to aid our decisions. With the growth of the Internet society, web systems are now capable of managing and processing a large amount of information. The recommendation systems provide more intelligent information considering the user's preferences. Google search engine and Amazon recommendation systems are well-known examples.

The recommended items' properties, the user's properties, and the user's current context are the fundamentals of the recommendation systems [1]. In the case of movie recommendations, the movie properties include genre, title, and cast. These properties are used to calculate the similarity between the movies. The user properties include the user's favorite genre, actors, and/or directors. For example, if a user likes action movies, the recommendation system will include more action movies in the recommended set. The recommendation system sets the basic direction based on items' properties and user's properties of information [2]. The user context information represents the environment the user's in, which highly influences the user's decisions [3, 4]. For instance, people prefer to watch scary movies in the summer than in the winter. The recommendation system can provide better recommendations taking the user's context information into consideration.

Many different algorithms and effective recommendation systems have used items' properties, user's properties and user's current context of information. Collaborative filtering is the most popular recommendation algorithm used in many recommendation systems [5]. This algorithm predicts the user's interests based on the user's preferences from the past. Many researchers tried to improve the algorithm using different approaches. The modified algorithms take the user's temporal context or genre preferences into consideration. This paper proposes an analysis method to categorize the information types and a learning algorithm to predict the user's item preferences based on the three information types mentioned above.

In Section II, we check the previous researches on recommendation systems using properties and learning algorithms. In Section III, we propose a new method of analyzing contents and present our recommended architecture. Then, in Section IV, we show how our system works using the system flowchart. Finally, the conclusion is given in Section V

II. RELATED WORK

A recommendation system compares a user profile with some reference characteristics and seeks to predict the rating or preference that a user would give to an item that they had not yet considered. There are lots of recommendation algorithms. One of them is the recommendation system based on the contents properties.

 ISBN: 978-1-61208-177-9

This system calculates a similarity between items and suggests contents to user using this calculated similarity. Another approach is learning algorithm. The learning algorithm analyzes the user pattern when user selects contents and applies these analyzed results to the recommendation.

Our system is based on contents properties and learning algorithm for user preference.

A. A recommendation system based on contents properties

A recommendation system based on contents properties consists of three steps. The first step is to calculate the correlation coefficient using contents preferences. The second step is to choose neighbors of content A selected by user. Neighbors mean a group of contents that have similar rating with content A. The third step is to estimate the preference for contents based on the neighbors ratings. The detailed explanation is given in the followings.

- *Calculating the correlation coefficient*

To calculate the correlation coefficient of the contents selected by user, we use Equation (1) Pearson correlation coefficient [6, 7, 8]:

$$\boldsymbol{\rho_{xy}} = \frac{\boldsymbol{\sigma_X \sigma_Y}}{\mathbf{cos}(\boldsymbol{X},\boldsymbol{Y})} = \frac{\sum(\boldsymbol{X_i}-\overline{\boldsymbol{X}})(\boldsymbol{Y_i}-\overline{\boldsymbol{Y}})}{\sqrt{\sum(\boldsymbol{X_i}-\overline{\boldsymbol{X}})^2} \times \sqrt{\sum(\boldsymbol{Y_i}-\overline{\boldsymbol{Y}})^2}} \quad \textbf{, (1)}$$

where

- X is a content.
- $\overline{X}$ is the average rating of X.
- X_i denotes the rating of the i^{th} user for X.
- Y is all the other contents
- $\overline{Y}$ is the average rating of Y
- Y_i denotes the rating of the i^{th} user for Y.
- ρ_{xy} is the Pearson correlation coefficient between X and the other contents Y.

- *Selecting neighbors*

In this step, neighbors are chosen using the results of Equation (1). A certain correlation coefficient value close to 1 is first selected as a threshold and contents with a correlation coefficient value greater than this threshold are selected as neighbors.

- *Predicting preferences*

The final step is to predict preferences based on the ratings of neighbors. This step uses Equation (2):

$$\mathbf{P} = \overline{\boldsymbol{X}} + \frac{\sum_{y \in raters}(Y_i - Y)\rho_{xy}}{\sum_{y \in raters}|\rho_{xy}|} \quad \textbf{, (2)}$$

where

- $\overline{X}$ is the average rating of content X.
- Y_i is the rating of the i^{th} item by other users.
- Y is the average the rating of the given content by the neighbors of X.
- ρ_{xy} is the Pearson correlation coefficient between X and the other contents Y.
- *raters* are a set of contents which have ratings.
- P is the predicted value of contents for a specific user.

B. Learning algorithm

Machine learning is concerned with the development of algorithms allowing the machine to learn via inductive inference [9]. This inductive inference is based on observation data that represents incomplete information about statistical phenomenon. One of machine learning systems attempts to eliminate the need for human intuition in data analysis while others adopt a collaborative approach between human and machine. In our approach, we use association rules for the learning method.

- *Association rule learning*

Association rule learning is a popular and well known method for discovering interesting relations between variables in a large database [10]. Many algorithms for generating association rules were presented over time. Some well-known algorithms are Apriori, Eclat and FP-Growth, but they only do half the job, since they are algorithms for mining frequent item-sets. Another step needs to be done after to generate rules from frequent item-sets found in a database [11]. Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time.

Association rule learning divides into two separate steps. In the first step, minimum support is applied to find all frequent item sets in a database. In the second step, these frequent item sets and the minimum confidence constraint are used to form rules. Finding all frequent item sets in a database is difficult because it involves searching all possible item set.

III. OUR APPROACH

Users, generally, selects contents on the basis of a variety of information. For example, if a user selects movie Scream, we assume that user likes horror genre or series. Thus, analyzing the user selections can help to predict user preference.

According to Korean Film Council (KOFIC) [12] survey, user considers genre, rating and casts when he/she chooses a movie. Table I shows the values that user considers when he/she selects movie.

We suggest a new classification method of the factors of items. We divide information of items into four types, basic information, create information, property information, related information. Then, we make user preference pattern

 ISBN: 978-1-61208-177-9

using each information type and calculate the score for recommendation using user preference pattern. Therefore, our system is composed of these three modules: item analyzing module, user pattern analyzing module and recommendation score module.

A. *Item Analyzing Module*

In item analyzing module, we first analyze properties of items. Then, we classify those analyzed factors by information types.

- **Basic information** is associated with basic information for items. Basic information includes three factors of item such as year, cast, country.
- **Create information** is associated with information of the production. Create information includes three factors of item such as creator, director, producer.
- **Property information** is associated with properties of items. Property information includes three factors of item genre, keyword, episodes.
- **Related information** is associated with information related with items. Related information includes three factors of item awards, top250, soundtrack.

B. *User Pattern Aanalyzing Module*

This module produces user preference pattern using the result of item analyzing module. User preference pattern is the configuration of individual weights on the given information type. Each information types have two patterns, short-term preference pattern and long-term preference pattern.

Short-term Preference Pattern (SPP) is the most basic pattern to define relation between items and user preferences. Preferences for items are defined by common factors between criterion item and compared item. SPP is calculated by each information type. Thus, the pattern has a total of four types, basic information pattern, create information pattern, property information pattern and related information pattern. These four patterns are used to calculate long-term pattern (LPP). Equation (3) shows how to draw the SPP.

$$\mathbf{SPP} = \frac{|A \cap B|_i}{\sum |A \cap B|_i} \quad \mathbf{, (3)}$$

where

- SPP is Short-term preference pattern.
- A is criterion item, B is compared item.
- *i* is the selection number of compared item.

Long-term Preference Pattern (LPP) represents learned preferences of user. It is configured about each information type and is determined by the average of the weights used in the past and short-term preference pattern of a selected item. Equation (4) shows how to draw the LPP.

$$\mathbf{LPP}_i = \frac{SPP_{n-1} \times LPP_{n-1}}{2} \quad \mathbf{, (4)}$$

where

- LPP is long-term preference pattern
- *i* is the selection number of compared item.

C. *Recommendation Score Module*

The recommendation score module calculates the prediction score of items. It is value to determine recommendation item. Equation (5) shows how to compute the recommendation score.

$$\mathbf{RS}_i = \sum LPP_i \times Rating_i \quad \mathbf{, (5)}$$

where

- RS is recommendation score.
- *i* is item number.
- Rating_i is rating on item *i*.

IV. SYSTEM FLOW

Our whole system uses values gained from SPP, LPP, and RS in Section III. Its system flow has two steps, shown in Figure 1.

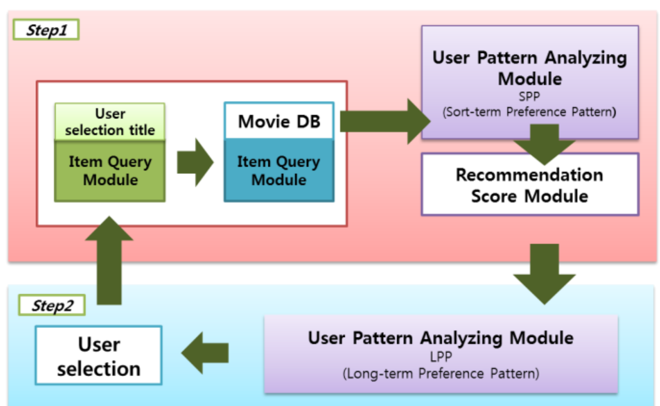

Figure 1. System flow

In Step 1, when user puts the title of item into the system, item analyzing module analyzes it based on information type. Then, analyzed data are compared with other data. The system has three small steps. First, the system analyzes the properties of compared item. Secondly, it makes short-term preference pattern. Thirdly, the system compares between the compared item and input data to calculate recommendation score with the weights put on each information type. Lastly, the system provides a recommendation list for user. For example, if user inputs the title *Harry Potter*, our item query module classifies its information, such as the genre *Fantasy* and cast *Emma Watson*. After classifying information of the item, user

 ISBN: 978-1-61208-177-9

pattern analyzing module compares common information factors between *Harry Potter* and other movies and makes user preference pattern for recommendation. Through all these process, the system makes recommendation score and suggests a recommendation list. If the score is high in Property information type, then it recommends other *Fantasy* genre movies like *The Chronicles of Narnia*. Or if it is high on Basic information, it suggests *Emma Watson*'s other movies.

In the Step 2, user pattern analyzing module operates right after user selects an item. First, it makes long term preference pattern comparing the weights of selected item and the weights in user profile. Here, the selected item becomes user input data. Ending this process, the system restarts step 1. In the *Harry Potter* case, if user selects *The Chronicles of Narnia*, then user's property information weight increases and a next recommendation list has more contents related to the genre *Fantasy* or key word *wizard*.

V. CONCUSION AND FUTHURE WORKS

For effective recommendation, we have suggested a new classification and learning method. Using the new classification method, we have divided information of item into four types. Analyzed information types can express more various user preferences. We also have made user preference pattern. It is learned to follow up preference of user whenever he/she selects an item. Thus, our system has possibility to recommend items to user more precisely than other systems because we use more information in items and history of user pattern.

In future works, we aim to test our system using large size of database such as IMDB and to verify our algorithm. And we will also expand our algorithm using user context. For example, we can check user preference variation over time or user's feeling.

ACKNOWLEDGEMENTS

This research was supported by the IT R&D program of MKE/IITA 2008-S-024-01 and the Basic Science Research Program through NRF funded by MEST (2010-0009168).

REFERENCES

[1] C. Ono, Y. Takeshima, Y. Motomura, and H. Asoh, "Context-Aware Users' Preference Models by Integrating Real and Supposed Situation Data", IEICE Trans. Inf. & vol. E91-D, no.11 November 2008

[2] H. Asoh, C. Ono, Y. Motomura, "A Movie Recommendation Method Considering Both User's Personality and Situation", ECAI 2006, August

[3] C. Ono, M. Kurakawa, Y. Motomura, and H. Asoh, "Implementation and evaluation of a movie recommender system considering both users' personality and situation," IPSJ J., vol. 49, no. 1, pp. 130-140, 2008

[4] I. Zekerman and D.W. Alberecht,"Predictive statistical models for user modeling," User Modeling and User-Adapted Interaction, vol.11, no.1-2, pp.5-18, 2001.

[5] M. Balabanovic and Y. Shoham. Feb: content-based collaborative recommendation Communications of the ACM, Vol 40, No 3, pp.66-72, 1997

[6] D. Billsus, M. J. Pazzani, in: The 15th International Conference on Machine Learning, pp. 46-54.

[7] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, in: ACM Conference on Electronic Commerce, pp. 158-167.

[8] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, in: Proceedings of the 10th International World Wide Web Conference, pp. 285-295.

[9] Bauer, M., Pohl, W. and Webb, G. (eds.): 1997, UM97 Workshop: Machine Learning for User Modeling. Online proceedings: http://www.dfki.uni-sb.de/_bauer/um- ws/

[10] Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, in G. Piatetsky-Shapiro & W. J. Frawley, eds, 'Knowledge Discovery in Databases', AAAI/MIT Press, Cambridge, MA.

[11]Rakesh Agrawal and Ramakrishnan Srikant.Fast algorithms for mining association rules in large databases. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

[12] Korea Research Team, Movie audience preference analysis,2009.

TABLE I. CONSIDERED ELEMENTS FOR MOVIE CHOICE

	Sex		**Age**					
	Man	Woman	19-24age	25-29age	30-34age	35-39age	40-44age	45-49age
Movie scenario	86.7	93.1	86.3	93.3	88.7	95.0	88.6	87.5
Other people's rating	75.6	79.9	79.4	78.0	75.0	80.9	77.1	76.0
Genre	74.5	79.0	74.0	73.7	79.9	77.9	77.6	77.5
Cast	70.0	78.4	70.6	77.5	72.5	77.9	75.6	71.0
Ranking	71.5	68.6	68.1	70.3	64.2	70.4	73.6	74.0
Expert rating	40.8	45.7	39.7	37.8	35.3	43.2	51.2	52.5
Director	41.9	38.4	33.8	43.5	35.8	36.7	49.8	41.5
Country	32.2	33.2	31.4	35.9	27.5	32.7	35.8	33.0
Awards	28.6	29.9	25.0	20.6	23.0	29.6	39.8	38.0

Production cost	24.9	14.2	15.2	19.1	15.7	22.6	25.4	19.5
publisher	11.1	8.7	7.4	6.2	7.8	9.5	19.9	9.0

 ISBN: 978-1-61208-177-9