A Study of Weight-Based Techniques for Recommender System Application

Sumi Shin

S&T Information Center, Korea Institute of Science and Technology Information

---**ABSTRACT**--
This study proposed popularity and awareness-based weight techniques which can be used in collaborative filtering approach. In this concept, relative popularity change and features which change depending on users' awareness ratio are applied as they are. Therefore, it is recommendable according to trends. Furthermore, it can improve a probability of being satisfied when a user selects a product after its relative quality is considered for recommendation. This study proposes a method which increases product quality with weight. This applied weight to a 2-D recommendation space which is evaluated by the recommender system operated in conventional web environment for recommendation, making recommendation considering multidimensional elements possible. MAEs on five experimental data were compared. The NBCFA which adopted product quality as weight revealed high recommendation accuracy. When popularity weight is used, NBCFA and CMA reveal 1.4% and 4.3% respectively in performance improvement.
Keywords: Collaborative Filtering, Item Quality, Weight-based, Recommender System, Recommendation

I. INTRODUCTION

Recently a large number of users use the Internet to find various kinds of items such as music, video, books, and movies or to browse news articles, and scientific articles. In the Internet, 35% of Amazon's sales are result of its recommendation engine. Additively, there are many other usages of them such as recommending music, books, movies, or even articles7. As the number of these items increases, it becomes more and more difficult to find their preferred items and information. For such users' convenience, recommendation systems technology is required. The term "recommendation" is defined as a visualized form of the personalization of information or service for a specific group or person, and the system for recommendation is called recommender[1]. A recommender system analyzes users' needs and demand and finds and offers perfect products and services for them. For more accurate and effective analysis, there have been a lot of studies. A study on its application to the recommender system started with a collaborative filtering approach.

The basic idea of this recommendation technique is that a user may like the one preferred by other users who have similar preference with the user. Then, non-experienced products evaluated by users before are recommended. Therefore, this collaborative filtering approach needs a process to find other users (or neighbors) who have similar interest with the user through the details of user rating and filter products based on the similar users' rating on the products.

The conventional studies which have adopted this kind of collaborative filtering approach include Group Lens[2] and Ringo[3]. The matters considered by these techniques for recommendation are user propensity, characteristics of products, inter-user similarities and product rating by users. However, they failed to consider product quality. If the quality of products is provided as the standard for recommendation, it would be able to get closer to user needs and demand.

Therefore, this study proposes a method which increases product quality with weight. This applied weight to a 2-D recommendation space which is evaluated by the recommender system operated in conventional web environment for recommendation, making recommendation considering multidimensional elements possible.

II. RELATED WORKS

There are lots of items in the online shopping mall and most people can find what they want by searching the Web. One problem of the large number of items in the web such as online shopping mall is that users spend most of their time to find a correct item from search results. Thus, users start looking for a better system that can suggest relevant item instead of letting them go through all search results. We call such a system recommender system. Most of the recommendations rely on the rating given by the users[4]. A recommender system is often based on collaborative filtering (CF). The main purpose of collaborative filtering is to recommend items of prefer for a target user on the

basis of other users' preference. A variety of CF technique has been studied and their performance has been evaluated. This technique is based on a simple intuition: predictions for the target user should be based on the preference patterns of other users who have similar preferences. Therefore, the first step of these algorithms is to find similarities between user interests. As an example, e-commerce shopping mall has a database of user ratings for items, where users indicate their interests. In other words, users predict rating (scores) on a certain product based on others' rating. In this case, there is a doubt if the collaborative recommendation technique predicts product recommendation considering all users. If all users are considered for prediction, the accuracy would considerably decrease. With this kind of reason, the collaborative recommendation technique finds the users who have similar preference with certain users and predicts based on their rating. For example, let's assume that a man intends to watch a film. In this case, the recommender system analyzes the history of the films this user has watched and selects the users who have similar preference. Then, it predicts the films the films that users haven't watched yet based on other users' rating and recommends the most suitable films. Those that users watched before would be definitely excluded. Finding similar users and making a prediction based on similar users' rating are the basic process of the collaborative recommendation technique. This kind of collaborative recommendation technique has been applied to many systems such as GroupLens and Ringo. We have discussed a collaborative filtering technique based on users' similarities. This algorithm is called User-Based Collaborative Filtering. In a user-item matrix, calculation of users' similarity is equivalent to calculation of the similarity between rows. It is also possible to make predictions based on the similarity between columns, that is, the similarity between items. This technique is called item-based collaborative filtering. The main idea of item-based collaborative filtering is that the user will like an item that is similar to items other users liked earlier, and will dislike an item that is similar to items they disliked earlier.

The Content-based recommender system is an upgrade from the simple keyword comparison-based recommendation. It makes a recommendation through a user profile which includes user taste and preference. The user profile information can be derived from user explicitly or implicitly, and the profile consists of the keywords which represent user characteristics. Therefore, the content-based recommender system makes a recommendation by calculating weight on these keywords. The typical technique used to calculate these weights is Term Frequency/Inverse Document Frequency[5]. The Term Frequency/Inverse Document Frequency appears in the document and calculates the weights of the related keywords. Then, it makes a recommendation to users. The content-based recommender system also uses diverse machine-learning techniques such as Bayesian classifiers. However, because these methods in the content-based recommendation system calculate utility prediction not based on heuristic algorithm, they reveal differences with information search. Therefore, it can be said that they are based on a model which uses base data by taking advantage of statistical and machine learning technologies. However, the content-based recommendation technology has some problems.

First, it covers the features only, which are explicitly related with the recommending products. To have a set with these features, therefore, contents could be automatically interpreted by computer, or it would be possible to analyze product features directly. The information search technologies work well in extracting features from texts. However, they have a fundamental problem that features are automatically extracted in other domains. Therefore, it is difficult to apply the automatic feature extraction method to multimedia data.

In fact, many users have evolved and grown in recommendation technology along with the advancement of item search. In terms of the recommendation of items, a contents-based technique which recommends the items to users would show interest in the past and similar ones, using items profile and their contents. However, contents-based filtering is limited in reflecting items preference and level because it reflects the contents only. Therefore, this study proposes a personalized items recommendation system which supports meaning based on semantic relatedness as a way to improve contents-based filtering techniques.

The hybrid recommender is a recommendation system which combines the content-based recommender system and collaborative filtering-based recommendation system[6]. It performed collaborative filtering algorithm, considering relationship among users, groups, tag and resources after extracting similar items based on items. According to his test, when a hyper-graph model which combines both methods is used, it is more effective than the user-based collaborative filtering recommendation method. Hybrid CF algorithms, such as the content-boosted CF algorithm, are found helpful to address the sparsity problem, in which external content information can be used to produce predictions for new users or new items.

III. PROPOSED ALGORITHM

In this chapter, a proposal of the product popularity weight-reflected recommendation techniques and product quality-related concepts are described. Then a recommendation method which derives interest in users' preference and item quality is proposed.

In conventional recommendation techniques, a recommendation is made, using a set of product features as a rating element. Product quality is not included in this kind of evaluation elements because it is hard to get every feedback

on product quality from users. The fact that if product quality is high, the product would be used and perceived by a lot of users is conceptualized. Therefore, this study attempted to increase the efficiency by applying this kind of possibility to the recommender system as a weighted value.

In product quality, the given product is 'i,' and the probability that it would be loved by a user is 'Q(i).' In terms of the quality of popular products, they are mostly satisfied by users. Therefore, a proportional relation exists between product quality (Q) and popularity (P). If a product is wanted by more users, its quality is high. In other words, if product quality is high, a probability of being chosen increases. Under this mechanism, if a product is selected by more users, it is reasonable to say that its quality is high. In terms of product quality, users are satisfied with popular products. Therefore, if a product is wanted by more users, its quality is high. With popularity, it is able to measure quality based on the implicit feedback so-called 'product selection by users.' Furthermore, a proportional relation is found between product awareness (A) and quality.

Based on this relevance among product quality, popularity and awareness, product quality can be estimated. In other words, product quality that users generally feel can be calculated based on the followings: how broadly the product is known to people; how much it is preferred. It can be stated as equation 1 below:

$$(i)=P(i)*A(i) \tag{1}$$

The recommendation system creates the best candidates for users through a hybrid recommender system which adopted either collaborative or content-based filtering technique or contained the advantages of both techniques, and users select the one they want among the recommended candidates.

If the same question is asked by several users, the products 'a,' 'b' and 'c' are extracted as recommended candidates. If 'a' is most selected by users, followed by 'b' and 'c,' 'a' would be the most popular product. If selected more, the product is popular. Therefore, popularity can be defined as the ratio (P(i)) of the users who have chosen the product 'i.'

Users frequently find and select popular products. In fact, products can be popular because they are trendy or have good quality. After all, they are chosen because they meet many people's tastes. Users are encouraged to select and purchase high-quality products. Therefore, they are frequently searched by many users. Based on this fact, use awareness is defined as the ratio (A(i)) of the users who recognize the product 'i.'

Product popularity is calculated based on how much the product is selected by users. For example, if the product (i) is recommended 1 million times and selected by 100,000 users, its popularity is '0.1.'

Here, 'z(i)' refers to the number of the product (i) selected while 'y(i)' represents the number of the product (i) included in the total recommendation. The product quality proposed in this study means a probability of the product (i) being preferred among those who know the product.

For user awareness, it is needed to consider what would be used to estimate user awareness prior to the calculation. In a recommendation system, it is not easy to figure out how many users are aware of the product at a certain point in time. It is not possible to check all users if they know the product. Therefore, this study uses the history of product recommendations to estimate user awareness in the recommendation system. There is a higher possibility that the products recommended to users are perceived by users than those not recommended. In addition, the products with a high recommendation rank are higher than those with a low recommendation rank in terms of awareness probability. These data are used for awareness analysis.

Through the recommendation system, a user's similarity user set ('M') was recommended a total of 'x' times. If the number of recommendations including the product (i) is 'y,' and the number of the product (i) selected is 'z,' user awareness which means a probability that a user is aware of the product (i) is stated as equations 2.

$$A(i)=\frac{y(i)}{x(i)}*\frac{z(i)}{u(i)}*r(i) \tag{2}$$

In equation 2 above, 'y(i)/x(i)' refers to a probability of 'i' to be included while 'z(i)/y(i)' means a probability of 'i' to be selected. In case of 'r' among the rest factors needed for awareness calculation, for example, if the product (i) ranks 90th when a total of 100 products are recommended, 'r' is 0.1. The calculation of 'r' is stated in equation 3 below, Here, 't' refers to the average rank when 'i' is recommended.:

$$r(i)=1-\frac{\sum_{i=1}^{f}t(i)}{y(i)} \tag{3}$$

In addition, 'r' represents a relative rank when 'i' is recommended. Therefore, A(i) can be calculated through the probability of 'i' to be included, probability to be chosen and the relative rank of 'i.' As a result, user awareness on the product can be estimated depending on recommendation frequency and rank.

IV. PERFORMANCE ASSESSMENT

For performance assessment proposed in this study, user similarity was calculated against 6,040 users, using Pearson's Coefficient for the purpose of proving the improvement of recommendation accuracy. Statistical precision measurement technique as a way to assess the accuracy of the referral system (statistical accuracy metrics) measurement techniques and decision support (decision support metrics) are used. Prediction evaluation techniques and decision support measurement technique is recommended to measure how well support the user to select items of good quality by comparing the statistical accuracy of the measurement technique is that the user is predicted and evaluated values value to measure the accuracy of the prediction the valuation techniques. This paper uses a predictive evaluation metrics of MAE (Mean Absolute Error). MAE (Mean Absolute Error) are recommended to evaluate the performance of the average of the absolute difference between the predicted value of the user of the system like the actual evaluation values for the items by Statistical accuracy metrics techniques. This paper was compared to the case when the pure collaborative filtering algorithm 'NBCFA' and proposed weight were applied. With the experimental data, recommendation accuracy was compared between the NBCFA and the proposed product quality-applied NBCFA. In the NBCFA which used Pearson's Coefficient for similarity weight, MAE is stated as shown in TABLE 1 and Fig. 1 below:

Table. 1 Test Result

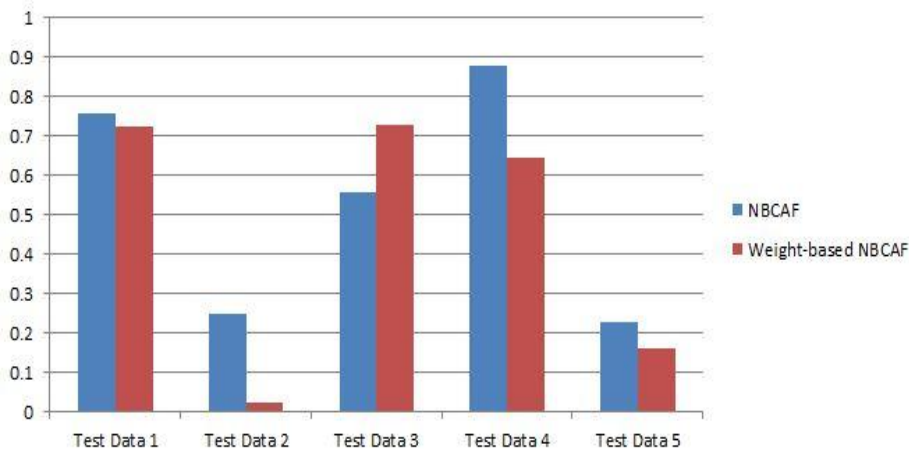| a)	Algorithm	b)	NBCAF	c)	Weight-based NBCAF
d)	Test Data 1	e)	0.7571	f)	0.7231
g)	Test Data 2	h)	0.2473	i)	0.0214
j)	Test Data 3	k)	0.5542	l)	0.7279
m)	Test Data 4	n)	0.8795	o)	0.6442
p)	Test Data 5	q)	0.2267	r)	0.1574

Figure 1: NBCAF vs. weighted-based NBCAF

As described in TABLE 1, MAEs on five experimental data were compared. Among them, the NBCFA which adopted product quality as weight revealed high recommendation accuracy. It shows that popularity weight is used, NBCFA reveal 4.5% respectively in performance improvement.

V. CONCLUSION

As computing environment advances, and the use of web technologies becomes more common recently, more products and services are now provided on the web. These web-based products and services have been chosen and provided in customized fashion. This kind of paradigm targeted to provide custom-made products and services is called 'Personalization.' It is 'recommendation' which has been studied as a field to provide personalized products and services. However, these recommendation techniques have a limitation of making a recommendation based on user ratings on the products and services only in computational space. Even so, it is hard to approach this problem caused by the limitation of a recommendation space as a field with continuous interest because processing costs can skyrocket with an addition of recommendation.

The recommendation system which predicts and recommends user-interested items automatically has long been studied in diverse fields such as information retrieval and expert system. This paper has examined recommendation system techniques and proposed a weighted-based personalized recommendation system which has applied item quality to collaborative filtering as a way to help a large number of items recommend in efficient and artificial fashion.

In the proposed system, the target preferred items were expanded using item quality, and other items viewed by the users who have already buy them were chosen as recommendation candidates. Then, utilization weight-based recommendation was added in addition to substantial similarity on the item was added. However, there should be further studies on the addition of an associated items filtering technique based on ranking on target recommendation user and highly matched item in terms of frequently finding and selecting popular item by the quality.

Therefore, this study proposes a technique which applies quality weights on the products as a way to improve recommendation performances without increase in recommendations and compares it with the conventional method to reveal performance improvement. This study proposed popularity and awareness-based weight techniques which can be used in collaborative filtering approach. In this concept, relative popularity change and features which change depending on users' awareness ratio are applied as they are. Therefore, it is recommendable according to trends. Furthermore, it can improve a probability of being satisfied when a user selects a product after its relative quality is considered for recommendation. When popularity weight is used, NBCFA reveal 4.5% respectively in performance improvement.

REFERENCES

[1]. Goldberg David, Nichols David, Oki Brian M. and Terry Douglas, Using collaborative filtering to weave an information tapestry , *Communications of the ACM, Vol. 35,* 1992, 61-70.
[2]. Konstan, Miller, and Herlocker, GroupLens: Applying Collaborative Filtering to Usenet News*, Comm. ACM, Vol.40,* 1997, 77-87.
[3]. U. Shardanand and P. Maes, Social Information Filtering:Algorithms for Automating 'Word of Mouth', *Proc. Conf. Human Factors in Computing Systems*, 1995.
[4]. K. Harinath, M. Kumaran. Scalable Recommendation Engine for Optimized Product Discovery. *Indian Journal of Science and Technology*, 8(29), 2015.
[5]. Salton, Automatic Text Processing (Addison-Wesley, 1989).
[6]. M. Pazzani, A Framework for Collaborative, Content-Based, and Demographic Filtering, *Artificial Intelligence Rev*.,1999. pp.393-408.

S.M Shin received the B.S. and M.S. degrees in Computer Engineering from Hongik University, Korea, in 1997 and 2005, respectively. She is a Ph.D. student in Dept. of Computer Engineering, Hongik University, Korea. And she is currently a senior researcher in the Department of Information Service, KISTI, Korea. She is interested in recommender system, semantic web, big data analysis, databases, data retrieval and Internet of Things.