

A Study of Neural Network Learning-Based Recommender System

Sumi Shin

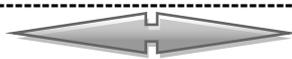
S&T Information Center, Korea Institute of Science and Technology Information, South Korea

ABSTRACT

A recommender system sorts and recommends the information which meets personal preferences among a huge amount of data provided by e-commerce. In particular, collaborative filtering (CF) is the most widely used technique in these recommendation systems. This method finds neighboring users who have similar preferences with particular users and recommends the items preferred by the former. This study proposes a neural network learning model as a new technique to find neighboring users using the collaborative filtering method. This kind of neural network learning model takes care of a sparseness problem during the analysis stage among those related with target users. The proposed method was tested with MovieLens data sets, and the results showed that precision improved by 6.7%.

Keywords: Collaborative Filtering, Recommender System, Recommendation, Neural Network Learning

Date of Submission: 20 January 2017



Date of Accepted: 05 February 2017

I. INTRODUCTION

There is lots of information in the Web and most people can find what they want by searching the Web. One problem of the large number of data in the Web is that we often spend most of our time to find a correct result from search results. Thus, people start looking for a better system that can suggest relevant information instead of letting users go through all search results: We call such systems recommender systems. 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 is a program designed to selectively provide the items which meet personal preferences or interest among a great number of items. Recommender systems are commonly used in e-commerce, Internet shopping malls and news websites.

Recommendation techniques include collaborative filtering, content-based filtering and demographic filtering. Collaborative filtering is a technique to recommend items after analyzing the preferences of the neighboring users associated with target users. In content-based filtering, users estimate preferences, using diverse item information. Demographic filtering is a method which estimates target users' preferences based on their personal information such as age, gender and occupation. The most popular recommendation technique is collaborative filtering. However, the conventional collaborative filtering methods weren't able to reflect importance or weighted value among items at estimation of inter-user similarity.

This paper proposes a neural network learning model which improves a method to find neighboring users, using collaborative filtering. The proposed method estimates target users' preferences based on the learning of correlations among users, using a neural network. This method enables the learning of the weighted values of items and can be extended to the technique which implements learning by adding the related information such as users' demographic data.

This paper is structured as follows: In chapter 2, the related works are described. In chapter 3, the proposed neural network learning-based recommender technique is explained. In chapter 4, the proposed method is compared to other techniques in terms of performances. In chapter 5, conclusion and future directions are stated.

II. RELATED WORKS

Collaborative filtering is an algorithm which recommends the items users are interested in by discovering and taking advantage of correlations among them. In other words, collaborative filtering starts from the following assumption: 'People's interest and taste are not randomly distributed. Instead, certain trend and pattern actually exist among a group of users or their preferences.' In fact, people would get recommendation from their acquaintances or other users on the Internet and make a decision based on such opinions. In GroupLens, correlations among users were calculated for the first time, using the k-nearest neighbor method. A certain user's k-nearest neighbor is selected, and the weighted values depending on the degree of proximity and item preferences are used in predicting target users' preferences.

The k-nearest neighbor method based on user preferences involves the following three steps: 1) Calculate similarities using user preferences, 2) choose neighbors of user A who wants recommendations (neighbors are a group of users who have similar preferences to user A) and 3) estimate the preference for a specific item based on the neighbors' ratings.

There are a number of studies on systems with a variety of collaborative filtering [2]. Also, there are studies on the clustering technique [3] based on the collaborative filtering for improving the collaborative filtering as well as studies on the improvement of the accuracy of recommendations based on users' roles [4] and matrix factorization [5]. Moreover, there are attempts to identify the user's preference in the recommendation process of the CF. But, these kinds of collaborative filtering methods do not reflect the importance of items in calculating correlations among users and can have a problem of sparsity. Sparsity refers to the distortion of inter-user correlations when there are few items which represent user preferences. Billsus, D and others propose a method which takes '0' as missing value for a neural network-based recommender technique. After reducing dimensions by applying SVD to the transformed learning data, user preferences by item are entered through a neural network. In terms of a learning-based technique, there also is a Bayesian classifier-based method [6]. This kind of a recommender system learns the prediction of like and dislike, using users' preference data.

III. NEURAL NETWORK LEARNING-BASED RECOMMENDATION

This study proposes a method to estimate preferences after analyzing correlations among the recommended items through neural network learning. Once the items to be recommended or not are chosen, target users' preferences are measured by entering other users' preferences on them. For example, when preference data such as Table 1 are available, neural network learning is given with the preferences on those other than target users.

Table 1. User-Item Matrix

	Item1	Item2	Item3	Item4
User1	3		2	
User3		1		5
User3		4	2	1

Based on the results, then, whether or not the preferences would be recommended as target items is decided. This kind of a learning model is advantageous in the online shopping malls having a great number of items. In collaborative filtering, if the number of the items in which users expressed their preferences is low, distortion can occur in correlations among users. To solve this kind of problem, in this study, the content information of the items is additionally used. Here, 'content information' refers to the characteristics of items.

Table 2. Preprocessed Matrix

	Item1	Item2	Item3	Item4
User1	0.6	0	0.2	0
User3	0	0.2	0	0.5
User3	0	0.8	0.4	0.2

For example, if an item represents a film, the items include a film genre and director. To include this kind of content information, neural network input nodes are added up to the number of the content information. In certain circumstances, input value should be codified or edited to '0,' '1' or other values for neural network-based learning. In this study, when data preferences aren't entered, preferences were set to '0' as stated in Table 2.

The methods to apply the matrix shown in Table 2 to learning, the neural network-based model is created by learning data of item preference of all users. And with this model, we can estimate the preference of the target item. In conventional recommender methods, there were difficulties in converging heterogeneous data. In the proposed algorithm, on the contrary, it is needed to add the related information as input target. In other words, the proposed technique has the advantage of easily coupling the items to be analyzed for recommendation by adding input nodes only. In this kind of technique, however, if there are more items, the number of input nodes should keep rising as well. This kind of limitation could be overcome by sorting those who are similar to target users in advance. This study does not consider the selection of these users as a top priority after choosing one genre only as additional information.

IV. EXPERIMENTS AND ASSESSMENTS

This study attempted to apply the recommender technique proposed based on the data collected from actual users and measure its performances. Among MovieLens data collected as a part of the project launched by the GroupLens research lab at the University of Minnesota, 1 million data including the same number of ratings were used. The MovieLens 1M dataset consists of 1,000,209 ratings on 3,952 films given by a total of 6,040 users. In this test, 600 people's data, 10 percent of total users, were used as test data.

Basically, neural network learning should continue until an error converges to zero. However, a recommender system has limitations in order for an error to converge to zero. Therefore, learning was terminated when the MSE converged to 0.05. In this study, 3 hidden nodes for a neural network were set. The proposed technique was evaluated with precision and recall values. With the assumption of a matrix such as Table 3, both precision and recall can be explained. In the following assumption: ‘a’ – the recommended item is actually bought; ‘d’ – non-recommended items aren’t bought; ‘c’ – recommended but not bought; ‘d’ – a non-recommended item is bought, precision refers to the percentage of group ‘a’ (actually bought) among the recommended groups (a + c) while recall represents the ratio of group ‘a’ (subject of recommendation) among the purchased groups (a + b).

Table 3. Recommendation-Purchase Matrix

Category	Recommendation	Non-recommendation
Purchase	a	b
Non-purchase	C	d

Table 4 compares the proposed technique with collaborative filtering. According to the results, precision and recall improved by 6.7% and 6.2% respectively.

Table 4. MAE Comparison

Category	Preference Learning	CF
Precision	38.1	35.7
Recall	20.6	19.4

Table 5 states the results of learning with an addition of a film genre to the proposed technique. According to the results, Compared to the learning with preferences only, precision and recall improved by 3.5% and 2.4%.

Table 5. Comparison with Genre Added Learning

Category	Proposed Technique	Genre Added Learning
Precision	38.1	35.7
Recall	20.6	19.4

V. CONCLUSION

This study proposed an algorithm which adds diverse item data to improve neural network learning-based recommender systems and neural network learning performances. This study proposes a method to estimate preferences after analyzing correlations among the recommended items through neural network learning. Once the items to be recommended or not are chosen, target users’ preferences are measured by entering other users’ preferences on them. This kind of neural network learning-based recommendation can be used regardless of data complexity. In addition, it has the advantage of utilizing diverse data in addition to items. This kind of a learning model is advantageous in the online shopping malls having a great number of items. In comparison to conventional collaborative filtering methods, furthermore, the excellence of the proposed algorithm was confirmed. It is planned to improve recommendation performances by adding data such as user profile. Moreover, there would a further study on how to express knowledge in a more diverse manner.

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*, pages 61-70, 1992.
- [2]. Robert Bell, Yehuda Koren, and Chris Volinsky. Modeling relationships at multiple scales to improve accuracy of large recommender systems. *In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 95–104, 2007.
- [3]. Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John Riedl. Itembased collaborative filtering recommendation algorithms. *In Proceedings of the 10th International World Wide Web Conference, WWW’01*, pages 285–295, 2001.
- [4]. Jason J. Jung. Attribute selection-based recommendation framework for shorthead user group: An empirical study by movielens and imdb. *Expert Systems with Applications*, 39(4): pages 4049–4054, 2012.
- [5]. Yehuda Koren, Robert M. Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8): pages 30–37, August 2009.
- [6]. Miyahara, K., Pazzani, M, J.. Collaborative Filtering with the Simple Bayesian Classifier, *Pacific Rim International Conference on Artificial Intelligence*, pages 679-689, 2000.

S.M Shin received the M.S. and Ph. D. degrees in Computer Engineering from Hongik University, Korea, in 2005 and 2017, respectively. 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.

