IJCSE International Journal of Computer Science and Engineering Open Access
Research Paper Volume-2, Issue-1 E-ISSN: 2347-2693

Enhancing Prediction in Collaborative Filtering-Based Recommender Systems

Majid Hatami[1*], Saeid Pashazadeh[2]

[1*] *Faculty of Electrical and Computer Engineering, University of Tabriz, Iran, hatami6889@gmail.com*
[2] *Faculty of Electrical and Computer Engineering, University of Tabriz, Iran, pashazadeh@tabrizu.ac.ir*

www.ijcaonline.org

Received: 03/12/2013 Revised: 08/01/2014 Accepted: 22/01/2014 Published: 31/01/2014

Abstract— Recommender systems (RS) are introduced to help users with finding the desired information. Collaborative filtering (CF) approach is one of the most widely used techniques in recommender systems. Prediction is the main part of all recommender systems. An enhanced prediction formula that could be employed in all CF-based methods is proposed in this paper. Resnick prediction formula that is the most well-known and employed formula in CF-based RS is used as basis in this paper. Not only the average of active user's ratings, but also the collective average of similar users' ratings and the average of all ratings given to the target item is used in the formula of this study. The results are promising and satisfactory. Results of enhanced prediction formula are compared with the results of unenhanced version to verify the effectiveness of proposed method.

Keywords/Index Term— Collaborative Filtering, Recommender Systems, Prediction Formula, Enhancement

I. INTRODUCTION

The volume of information continues to grow at an astonishing rate that brought some problems beside its benefits. Information overload is one of its most famous disadvantages, which refers to the hardness of finding desired information among the large amount of information. Recommender systems (RS) are introduced to help users for dealing with this problem. RS assist users to make choice when they have not enough experience of alternatives [1].

RS are generally divided into two main categories: (1) collaborative filtering (CF) and (2) content-based filtering (CB) [2]. Collaborative filtering refers to the behavior when people collaborate to help each other perform filtering by recording their reactions to the items they have encountered [3]. CF methods employ an information filtering technique based on the user's history of purchases or evaluation of items [2]. On the other hand, CB methods recommend items based on their descriptions and users' preferences [2].

Each technique normally has some problems when used alone. Hybrid RS is introduced to combine two or more recommendation techniques to obtain better qualities with fewer downsides [4]. Combining CF with CB is one of common hybrid models.

Prediction is one of most important stages in all recommender systems. The most famous and widely used formula for CF-based recommender systems has introduced by Resnick et al. [5].

In the following sections we are describing how we enhanced this formula and will compare the results of normal Resnick formula to the enhanced one. We have measured the prediction accuracy using Mean Absolute Error (MAE) metric and the recommendation quality with Precision and Recall metrics which are described by Bobadilla et al. [6] in detail. These metrics are the most common measuring methods for comparing recommender systems by researchers. We have carried out our experiments on MovieLens dataset.

II. RELATED WORKS

Resnick et al. [5] introduced a prediction formula that uses the similarity of users and their ratings for predicting rating of a user to an item. O'Donovan et al. [6] introduced an addition that is employable to some prediction formulas. They have proposed three way of combining trust value with prediction formula: Trust-Based Weighting, Trust-Based Filtering, and a combination of Trust-Based Weighting and Filtering.

III. IMPORTANCE OF THE STUDY

Since Resnick et al. [5] introduced their method of prediction, it has been used increasingly in research studies and real life projects. Making a small improvement on this method could cause significant improvement in recommender systems field.

Corresponding Author: *Majid Hatami*, hatami6889@gmail.com

Their formula uses the *K* most similar users to the target user for making a prediction. The main idea of this paper can be represented as a simple question. Why do we not use other information of these *K* most similar users? The next section explains how we benefit from the average of the K most similar users' ratings besides their similarities.

IV. Prediction Formula

The well-known Resnick prediction formula is as follows:

$$p_u^i = \bar{r}_u + \frac{\sum_{x \in K_u} Sim(u,x) \times (r_x^i - \bar{r}_x)}{\sum_{x \in K_u} Sim(u,x)} \quad (1)$$

Where P_u^i stands for the prediction about the rating of user *u* for the item *i*, and r_x^i represents the real rate which user *x* has given to the item *i*, and $\bar{r}_x$ indicates the average of all the ratings given by the user *x*. Function *sim*(*u*, *x*) calculates the similarity between user *u* and *x* which could be in the range of [-1, 1]. K_u denotes the top *K* most similar users to the user *u*.

In this paper, we have modified the Resnick's prediction formula by making use of two additions: (1) the average of all *K* similar users' ratings average, and (2) the average of all users' ratings to the target item (the overall opinion about it).

As a result, our proposed prediction formula is as follows:

$$p_u^i = \lambda + \frac{\sum_{x \in K_u} Sim(u,x) \times (r_x^i - \bar{r}_x)}{\sum_{x \in K_u} Sim(u,x)} \quad (2)$$

Where λ can be computed as follows:

$$\lambda = \alpha \times \bar{r}^i + (1-\alpha)\gamma \quad (3)$$

$$\gamma = \beta(\frac{1}{K}\sum_{x \in K_u} \bar{r}_x) + (1-\beta)\bar{r}_u \quad (4)$$

λ denotes a combination of $\bar{r}^i$ (the average of all the ratings given to the item *i* by all users), and γ (the collective average of the ratings given by the active user, *u*, and all his/her similar users). α and β are used to tune the impact of the average of all the ratings given to the item *i* and the average of all the similar users' average, respectively. In other words, α and β determine how much obtained averages should be mixed together and how much each one should affect the prediction. Both α and β are values between 0 and 1. We have set α=0.25 and β=0.5 as a default value, however these values could be different based on the nature of the RS and the taste of users.

V. Result & Discussion

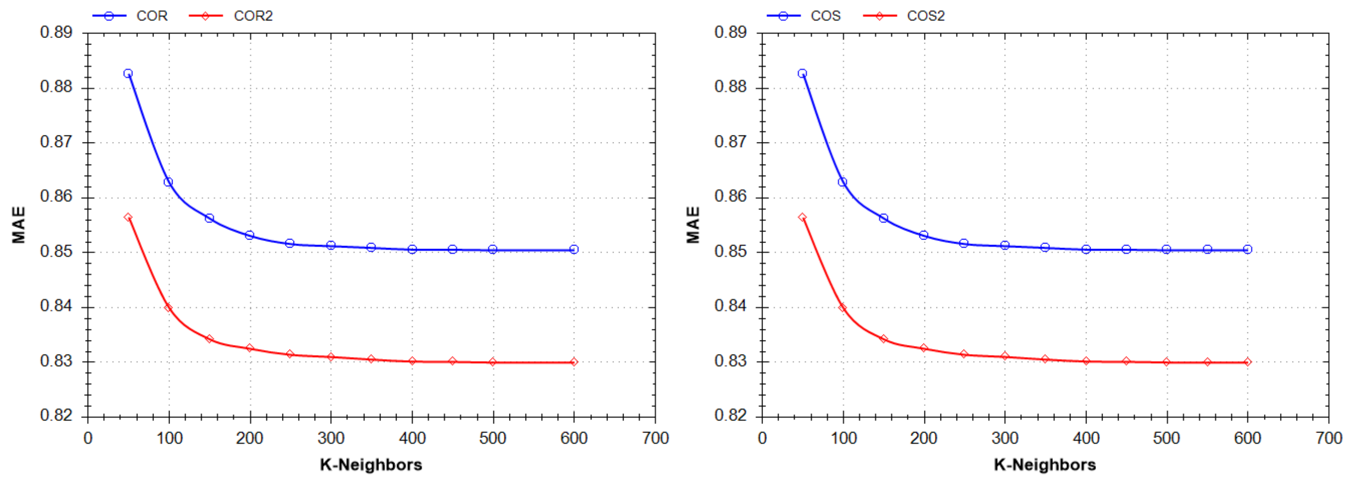
Bobadilla et al. [7, 8] provided different evaluation metrics to measure quality of recommender systems results. We have employed some of the most common used metrics such as MAE, precision and recall to compare our proposed method with the unenhanced method. A portion of MovieLens dataset that is one of the most popular and widely used datasets by researchers is used in this research. This portion contains 20,000 ratings given by 459 users to 1410 movies. Person Correlation and Cosine similarity measurements are used in our experiments.

Figures 1, 2, and 3 compares quality of predictions using two different similarity methods. In the part (a) of these figures, the blue line entitled 'COR' refers to the unenhanced version of Resnick formula, and the red line entitled 'COR2' refers to our proposed method while Pearson Correlation is used as the similarity method. The same goes for the part (b), 'COS' and 'COS2' lines except for their similarity method which is Cosine similarity.

Figure 1 shows the quality of prediction in terms of Mean Absolute Error (MAE). Figure 2 shows quality of predictions in terms of precision and Figure 3 shows quality of predictions in terms of recall. The top K most similar users in all experiments is varied from 50 to 600 in steps of 50. The number of recommendations for measuring precision and recall is varied from 2 to 20 in steps of 2.

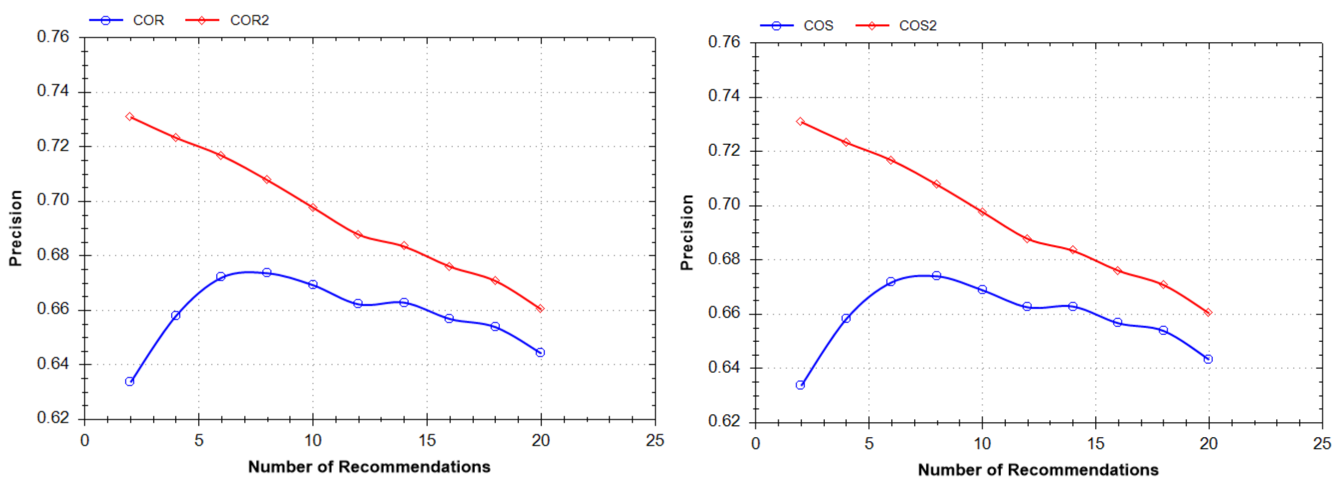
Experimental results in all quality metrics show that the result of using Pearson similarity and Cosine similarity as the similarity method, have the same results. Beyond that, results show that the proposed modification on prediction formula increases the quality of prediction in comparison with basic Resnick's prediction formula. Figures 2 and 3 show that the measure of improvement that our proposed method gains in comparison with basic Resnick's prediction formula is much more when we have lower number of recommendations, in other words, our experimental results indicate great improvement of proposed formula in the lower number of recommendations.

The trust approach which is introduced by O'Donovan et al. [6] needs extra and heavy calculations on ratings matrix, while our method is just using a simple mean on user's ratings. Moreover, our method is employable to their method as an extra enhancement, because they did not make use of similar users' ratings average in the way we have used.



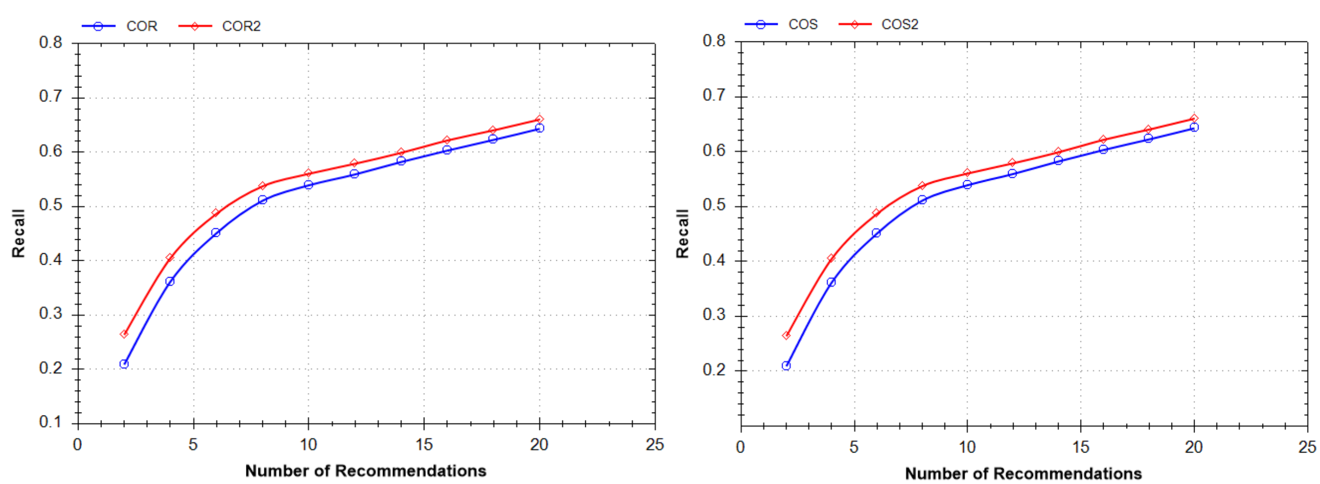
(a) MAE using Pearson correlation similarity (b) MAE using Cosine similarity

Figure 1. Experimental results of MAE with Pearson Correlation and Cosine similarity



(a) Precision using Pearson correlation similarity (b) Precision using Cosine similarity

Figure 2. Experimental results of precision using Pearson Correlation and Cosine similarity



(a) Recall using Pearson correlation similarity (b) Recall using Cosine similarity

Figure 3. Experimental results of recall using Pearson Correlation and Cosine similarity

VI. CONCLUSION

A modification to the Resnick prediction formula that is employable to all CF-based recommender systems is proposed in this paper. Experimental results show that proposed modifications increase quality of predictions in comparison with basic Resnick prediction formula. We have much better quality in precision quality metric when we make lower number of recommendations. The proposed modification can be employed to every prediction formula that makes its predictions based on the average of the active user's ratings. How much our proposed method will improve the qualities of a RS is depend on the nature of RS, its users' ratings, and their taste.

REFERENCES

[1] P. Resnick, and H. R. Varian, "Recommender systems," Commun. ACM, vol. 40, no. 3, pp. 56-58, 1997.
[2] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," Expert Systems with Applications, vol. 39, no. 11, pp. 10059-10072, 2012.
[3] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave information tapestry," Commun. ACM, vol. 35, no. 12, pp. 61-70, 1992.
[4] R. Burke, "Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331-370, 2002.
[5] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," In proceedings of the 1994 ACM Conference on Computer supported cooperative work, Sharing Information and Creating Meaning, pp. 175–186, 1994.
[6] J. O'Donovan, and B. Smyth. "Trust in recommender systems." In Proceedings of the 10th international conference on Intelligent user interfaces, pp. 167-174. ACM, 2005.
[7] J. Bobadilla, A. Hernando, F. Ortega, and J. Bernal, "A framework for collaborative filtering recommender systems," Expert Systems with Applications, vol. 38, no. 12, pp. 14609-14623, 2011.
[8] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutirrez, "Recommender systems survey", Knowledge-Based Systems, vol. 46, no. 0, pp. 109-132, 2013.

AUTHORS PROFILE

Majid Hatami is a M.Sc. student of Artificial Intelligence at University of Tabriz, Iran. His research interests include machine learning, data mining and pattern recognition.

Saeid Pashazadeh is Assistant Professor of Software Engineering and chair of Information Technology Department at Faculty of Electrical and Computer Engineering in University of Tabriz in Iran. He received his B.Sc. in Computer Engineering from Sharif Technical University of Iran in 1995. He obtained M.Sc. and Ph.D. in Computer Engineering from Iran University of Science and Technology in 1998 and 2010 respectively. He was Lecturer in Faculty of Electrical Engineering in Sahand University of Technology in Iran from 1999 until 2004. His main interests are modeling and formal verification of distributed systems, computer security, wireless sensor/actor networks, and applications of artificial neural networks. He is member of IEEE and senior member of IACSIT and member of editorial board of journal of electrical engineering at University of Tabriz in Iran.

