IJST, Transactions of Electrical Engineering, Vol. 37, No. E1, pp 93-100
Printed in The Islamic Republic of Iran, 2013
© Shiraz University

A NOVEL COLLABORATIVE FILTERING MODEL BASED ON COMBINATION OF CORRELATION METHOD WITH MATRIX COMPLETION TECHNIQUE[*]

E. BOJNORDI [**] AND P. MORADI

Dept. of Computer Engineering, University of Kurdistan, Sanandaj, I. R. of Iran
Email: Bojnordi.ehsan@gmail.com

Abstract– One of the fundamental methods used in collaborative filtering systems is Correlation based on K-nearest neighborhood. These systems rely on historical rating data and preferences of users and items in order to propose appropriate recommendations for active users. These systems do not often have a complete matrix of input data. This challenge leads to a decrease in the accuracy level of recommendations for new users. The exact matrix completion technique tries to predict unknown values in data matrices. This study is to show how the exact matrix completion can be used as a preprocessing step to tackle the sparseness problem. Compared to application of the sparse data matrix, selection of neighborhood set for active user based on the completed data matrix leads to achieving more similar users. The main advantages of the proposed method are higher prediction accuracy and an explicit model representation. The experiments show significant improvement in prediction accuracy in comparison with other substantial methods.

Keywords– Recommender systems, collaborative filtering, correlation, matrix completion, convex optimization, nearest neighborhood

1. INTRODUCTION

The aim of a recommender system is to generate meaningful recommendations for the users based on historical rating data. Suggestions for books on Amazon, or movies on Netflix, are real-world examples of the recommender system. In most of these systems, users frequently provide ratings on a scale of 1 (disliked) to 5 (liked)[1]. Such a data source records the users' historical ratings.

Here the problem is that, the user ratings matrix is very big and sparse and the recommender system has to predict what rating a user would give to an item which has not been previously rated. Typically, the ratings are predicted for all items that have not been noticed by a user. Consequently, the items with the highest ratings will be presented as a recommendation. Personalized TV recommendations as a recent example of recommender systems was proposed by Pozrl et al [2].

Recommender systems are different in the ways that they analyze these data sources. Variations of this notion are discussed in [1, 3, 4] and they can be broadly categorized as:

- Collaborative Filtering: Systems that analyze historical interactions alone and recommend items based on all users' past ratings collectively.
- Content-based Filtering: Systems mostly based on profile attributes.
- Hybrid techniques attempting to combine the above mentioned designs.

Among these approaches, collaborative filtering systems have more potentiality [1]. These systems rely on intelligent agents such as historical ratings of users achieve better performance in most cases[5]. One of the traditional methods developed in this case is based on the nearest neighborhood collaborative

[*] Received by the editors November 11, 2012; Accepted April 7, 2013.

[**] Corresponding author

filtering that uses users' historical rating data and item preferences as input. This technique relies on the similarity between users and items. The collaborative filtering strategy in machine learning framework was proposed by Billsus & Pazzani [6], they combined different machine learning techniques. Zhang and Li introduced the different linear classifiers for the prediction of users' preferences, and compared this method with other memory-based methods [7]. Their experimental results show that, linear models are better than other models in this application. Ungar & Foster [8], review the two-sided clustering model for collaborative filtering; and describe how this model can be represented by Bayesian networks. Also, they expressed how to provide this model as a probabilistic relational model. Probabilistic latent semantic analysis (PLSA) is a new probabilistic graphical model to users' purchasing behavior model that was proposed by Hofmann [9].

Unified collaborative filtering model is based on the combination of latent features raised by Zhong & Li [5]. This unified model led to significantly more accurate predictions compared with other previous well-known methods. The remainder of the paper is organized as follows:

Section 2 introduces the fundamental concepts of the correlation method. A brief introduction to matrix completion via convex optimization is introduced in section 3. Section 4, is devoted to an introduction of the initiative collaborative filtering method and the empirical evaluations of the proposed approach. Finally the paper is concluded and the roadmap for future work is drawn in section 5.

2. CORRELATION BASED ON NEAREST NEIGHBORHOOD METHOD

The first step in correlation based on the nearest neighborhood method is finding a subset of users based on their similarity with active users. The second step in this method is generating predictions for active users based on a weighted combination of their existing ratings. A summarized algorithm of this method is given below [1]:

1. Assign weights to all the users based on their similarity with the active users.
2. Select k users with maximum similarity to the active users. (select neighborhoods)
3. Calculate the rating prediction based on the weighted combination of the neighbors' ratings.

In step 1, $\boldsymbol{W_{a,u}}$is the similarity between the user $\boldsymbol{u}$ and the active user$\boldsymbol{a}$. Pearson correlation coefficient as the common similarity metric is defined as follows[10]:

$$W_{a,u} = \frac{\sum_{i\in I}(r_{a,i}-\overline{r_a})(r_{u,i}-\overline{r_u})}{\sqrt{\sum_{i\in I}(r_{a,i}-\overline{r_a})^2\sum_{i\in I}(r_{u,i}-\overline{r_u})^2}} \tag{1}$$

where I, the set of items, is rated by both users. $r_{u,i}$ denotes the rating given to the item i by the user u. and $\overline{r_u}$ is the mean rating given by the user u. In step 3, predictions are generally calculated as the weighted average of deviation from the neighbors' mean, as in:

$$p_{a,i} = \overline{r_a} + \frac{\sum_{u\in K}(r_{u,i}-\overline{r_u})\times W_{a,u}}{\sum_{u\in K} W_{a,u}} \tag{2}$$

where, $p_{a,i}$ symbolizes prediction for the item i by the active user a and $W_{a,u}$ gives the similarity of the user u and the active user a. K, denotes the set of neighbors for the active user, a.

While determining the similarity between users, items that have been rated or not rated by all users have a semantic value less than those of other items[3]. Breese et al introduced a concept as the inverse user frequency, which is processed as$\boldsymbol{f_i = \log\left(\frac{n}{n_i}\right)}$,where $\boldsymbol{n}$ is the number of all users and $\boldsymbol{n_i}$ is the number of the users that have rated item $\boldsymbol{i}$ differently from the others. Therefore, in order to use this concept in (1) the rating of the item $\boldsymbol{i}$ must be multiplied by $\boldsymbol{f_i}$.

$$\mathbf{W}_{\mathbf{a},\mathbf{u}} = \frac{\sum_{\mathbf{i}\in\mathbf{I}} \mathbf{f}_\mathbf{i}\mathbf{r}_{\mathbf{a},\mathbf{i}}\mathbf{r}_{\mathbf{u},\mathbf{i}} - (\sum_{\mathbf{i}\in\mathbf{I}} \mathbf{f}_\mathbf{i}\mathbf{r}_{\mathbf{a},\mathbf{i}})(\sum_{\mathbf{i}\in\mathbf{I}} \mathbf{f}_\mathbf{i}\mathbf{r}_{\mathbf{u},\mathbf{i}})}{\sqrt{[\sum_{\mathbf{i}\in\mathbf{I}} \mathbf{f}_\mathbf{i}\mathbf{r}_{\mathbf{a},\mathbf{i}}^2 - (\sum_{\mathbf{i}\in\mathbf{I}} \mathbf{f}_\mathbf{i}\mathbf{r}_{\mathbf{a},\mathbf{i}})^2][\sum_{\mathbf{i}\in\mathbf{I}} \mathbf{f}_\mathbf{i}\mathbf{r}_{\mathbf{u},\mathbf{i}}^2 - (\sum_{\mathbf{i}\in\mathbf{I}} \mathbf{f}_\mathbf{i}\mathbf{r}_{\mathbf{u},\mathbf{i}})^2]}} \tag{3}$$

The hidden hypothesis in this approach is that items which are generally hated or liked by users have been rated more than the other ones.

3. MATRIX COMPLETION BASED ON CONVEX OPTIMIZATION

The fundamentals of the matrix completion deal with the answer to the question, “how it is possible to recover a low rank data matrix with only partial sampling of its entries?” Candes & Recht [11] proved that if the number of sampled entries or m obeys:

$$m \geq cn^{1.2} rlogn \tag{4}$$

For some positive numerical constant like c, with very high probability, most $n \times n$ matrices of the rank $r$can be perfectly recovered by solving a simple convex optimization problem [11]. This approach finds the appropriate entries in data matrix by minimizing the nuclear norm. The equation (4) is established with the assumption that the matrix rank is not too large.

In order to recover a square $\boldsymbol{n} \times \boldsymbol{n}$ matrix $\boldsymbol{M}$of the rank $\boldsymbol{r}$, this matrix could be displayed with $\boldsymbol{n}^2$numbers but in fact it only has$(\boldsymbol{2n} - \boldsymbol{r}) \times \boldsymbol{r}$ degree of freedom on its entries. This fact can be proved by counting parameters in the singular value decomposition (SVD) of the low rank matrix (the degree of freedom in the matrix is equal to the description of singular values and of the left and right singular vectors). So in low rank matrices the degree of freedom is considerably smaller than $\boldsymbol{n}^2$.

The utilization of the matrix completion technique in the Netflix problem is proposed by Candes & Recht [11] as a practical example of using this technique. In the Netflix problem, the data matrix consisting of all user ratings may be almost a low rank because users generally set their preferences and tastes based on a limited number of factors.

Matrix conditions

Candes & Recht exhibit a simple model for low rank matrix [11]. Suppose SVD of a matrix $\boldsymbol{M}$as:

$$M = \sum_{k=1}^{r} \sigma_k u_k \, v_k^* \tag{5}$$

Where u_k and v_k are left and right singular vectors and σ_k are singular values (the roots of eigen values of $M \times M$). Therefore $u_k$can be a generic low rank matrix as$\{u_k\}_{1\leq k\leq r}$, uniformly and randomly selected from a set of r orthonormal vectors, and the same with $\{v_k\}_{1\leq k\leq r}$. Both of these sets may or may not be independent from one another. Nevertheless, this model is called as the random orthogonal model.

Condition of sampling sets

It is obvious if we don’t have any samples in a row or column, then we certainly won’t be able to reconstruct any values of this row or column. Thus, there should be at least one observation (sample) in each row and each column. The observation set can be selected uniformly and randomly from matrix entries.

Algorithm of problem solving

Classic optimization problems can be divided into two general categories: The concave and convex optimization problems. Convex optimization problems try to find a minimum for optimization variables and there are many numerical methods for them. If a problem is not convex, it can be changed into a convex optimization problem to be solved more quickly. One of the most popular ways to convert a non-convex problem to a convex problem is the Lagrangian relaxation method[12].

On the subject of the matrix completion technique we aim to recover the data matrix by solving the following optimization problem:

$$\textit{Minimize rank }(X) \qquad \boldsymbol{Subject\ to\ X_{ij} = M_{ij}(i,j) \in \Omega} \tag{6}$$

Where, X is the decision variable, $rank(X)$is equal to the rank of the matrix $X$and Ω is the observed sampling set in matrix X. It is obvious that (6) is to fit the observed data. If the rank of the matrix is r,then this matrix will exactly have r non-zero singular values. Fazel expressed that (6) has little practical application because the minimization of optimization variables (or the number of no vanishing singular values) is not a convex optimization problem, and belongs to the set of NP-hard problems instead[13]; moreover, all known algorithms that provide accurate solutions require exponential time order to solve the problem[14]. Fazel proposed a solution to convert this NP-hard problem to a convex optimization problem, which requires the nuclear norm to be used in the optimization problem, instead of the number of singular values to be considered. This is an alternative strategy to minimize the sum of the singular values, which is called the nuclear norm:

$$||X||_* = \sum_{k=1}^{n} \sigma_k(X) \tag{7}$$

Where, $\sigma_k(x)$is equal to k^{th} largest singular value of X and the heuristic optimization is then given through:

$$\textit{Minimize } \|X\|* \quad Subject\ to\ X_{ij} = M_{ij}(i,j) \in \Omega \tag{8}$$

Accordingly in this approach, matrix entries are retrieved by the minimum nuclear norm. It is worth noticing that the nuclear norm is a convex function and can be effectively optimized through semi-definite programming. Further comments might be followed in [11-13].

To solve the convex optimization problem, there are several solvers. The CVX is one of the most popular solvers raised by Boyd[15, 16]. In this paper, the CVX solver has been used for completing the data matrix.

4. EXPERIMENTS

As previously mentioned, the correlation method relies on the principle that a subset of users is selected based on their similarity to the active user and consequently the weighted combination of their ratings is used to provide the prediction of rating for the active user. However, since the recommender systems have the matrix sparseness problem, in order to overcome this problem, the exact matrix completion technique via convex optimization was used to retrieve the missing values in the data matrix. In other words, the matrix completion technique was used as a pre-processing step in the correlation method. Since the lower number of the missing values in the data matrix leads to a higher level of accuracy in choosing more similar neighbors for the active user, it is obvious that this preprocessing stage improves the accuracy of the correlation method. The proposed approach was substantiated by the experiments in this study.

Data set

One of the popular benchmark data sets used in the collaborative filtering research is the EachMovie data set, which has been used in the present experiments. It is a very large data set. There are 1623 items (movies), 72916 users, within which there are 2.1 million ratings. These figures are too large to be used in the simulation and testing processes. Thus most of the previous studies incorporated only 0.03% to 8% of the original data set[17]. As for the current study, 2025 ratings, or about 0.09% of the original data set, was used in the experiments.

Evaluation metrics and experiment setup

There are two kinds of accuracy evaluation in recommender systems:

- Prediction accuracy evaluation when explicitly predicting active users' ratings on some unseen items.
- Ordering accuracy evaluation of a set of unseen items, in order to recommend top-ranked items to the active user.

The former focuses on the quality of the explicit prediction which is made on the active user's level of interest in some unseen items. Whereas the latter implies finding a tenuous order within a set of unseen items so that top-ranked items will be recommended to the active user.

The system goal in the former is to predict a user's rating on a special item. Evaluation metrics in this process are mean absolute error (MAE), root mean square error (RMSE) and 0/1loss error[3].

$$MAE = \frac{\sum_{i=1}^{N} |p_i - r_i|}{N} \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (p_i - r_i)^2}{N}} \tag{10}$$

$$\mathbf{0/1}\boldsymbol{Loss} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} |\boldsymbol{p_i} - \boldsymbol{r_i}| \tag{11}$$

$\mathbf{p_i}$: Prediction of rating for item **i**
$\mathbf{r_i}$: User rating to item **i**
N: Number of all items

It is noteworthy that the proposed approach in this paper concentrates on the former; therefore, in the first step a customized data set with items exceeding the threshold of 21 ratings is selected. In other words, a 50×50 data matrix of these items and users are selected to represent the fulfilled conditions of the matrix completion technique expressed in part 3. Secondly, ten votes in the new data matrix are randomly selected. To evaluate the obtained prediction accuracies, the values of the ten votes are replaced with Null values. At this stage, the data matrix would be the input to the matrix completion technique. Finally, the completed matrix which has been obtained through the convex optimization method is considered as the input data to be used in the correlation method based on the nearest neighborhood method which has been expressed in part 2.

A general scheme for the proposed approach can be expressed as follows:

- *First*: In order to follow the prerequisite of the data matrix, the rows and columns whose numbers of existing entries are less than the threshold of 21 will be eliminated.
- *Second*: At this stage, the input matrix is given to the CVX solver to be completed.
- *Third*: The completed matrix is used to select neighbors of the active user and then, as explained in part 2, the correlation method is applied on it.

Results of prediction procedure

Table 1 summarized the experimental results of the current study, as well as those of the Gaussian PLSA mixture method [9]and the unified method [5]. A comparison between the current results with those of the previous methods reveals that the users' ratings are more accurately predicted through the proposed approach in this study. Obviously, the novel proposed method including both correlation and matrix completion outperforms the memory-based method in terms of MAE and leads to a relative accuracy gain over the other methods.

Table 1. Prediction accuracy of various methods

Method	**Error**		
	MAE	***RMSE***	***0/1Loss***
Mixture pLSA (Hofmann , 2004)	0.848	1.170	63.4
Unified Method with no latent features (Zhong & Li , 2010)	0.906	1.185	66.8
Unified Method with no external features (Zhong & Li , 2010)	0.791	1.067	65.4
Unified Method with both features (Zhong & Li , 2010)	0.771	1.047	62.6
Correlation (Breese et al., 1998)	0.994	---	---
Bayesian Clustering (Breese et al., 1998)	1.103	---	---
Bayesian Networks (Breese et al., 1998)	1.066	---	---
Matrix Completion	1.074	7.263	10.7
Proposed Method	**0.449**	**2.716**	**4.4**

Table I indicates that using the matrix completion technique without a memory-based method is not enough to get an exact prediction. The results of this method are superior to those of the other methods and it is worth noticing that the proposed method which is derived from combining two weak methods leads to an approach that is more accurate than other significant methods. According to Table 1, in terms of MAE and 0/1 Loss, the combination of the matrix completion technique with a memory-based method such as correlation leads to a more accurate method.

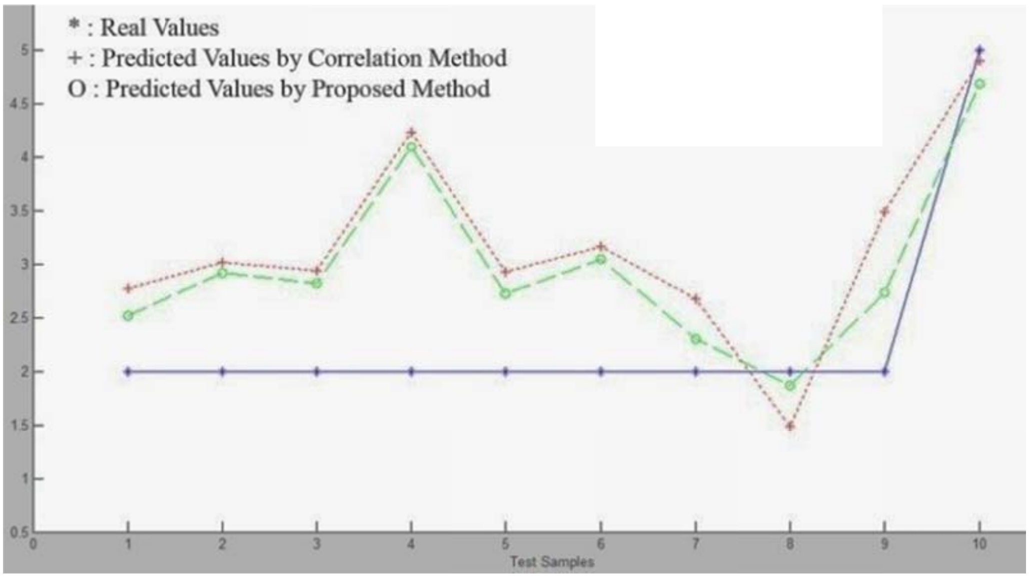

Fig. 1. Comparison of prediction of correleation method with proposed method

A comparison between the level of prediction in the correlation method and that of the proposed method is shown in Fig. 1. As it can be seen, the variety of the predictions with real values in the proposed method is smaller than that of the correlation method.

5. CONCLUSION AND FUTURE WORK

Among collaborative filtering methods, the correlation method belongs to the memory-based category while the matrix completion method belongs to the model-based one. These two methods were combined in this study and a new type of collaborative filtering method was proposed. The advantages of this novel hybrid method are a high level of accuracy in prediction, and overcoming the semantic comprehension problem in the web. This recent advantage is due to the facts that first, this method does not rely on syntax – as content based methods do – and second, it is based on an intelligent agent such as users' comments.

Further research includes the application of the matrix completion technique in other methods, such as the PLSA and hybrid methods. Moreover, as this approach is effective in increasing the first kind of accuracy (Prediction Accuracy Evaluation), it will probably increase the second kind of accuracy (Ordering Accuracy Evaluation). Besides, the operation of the currently proposed approach in hybrid personalized recommender systems might be investigated in the future.

REFERENCES

1. Melville, P. & Sindhwani, V. (2010). *Recommender System , Encyclopedia of Machine Learning*. C. Sammut and G.I. Webb, Editors, Springer US. pp. 829-838.
2. Pozrl, T., Kunaver, M., Pogacnik, M., Kosir, A. & Tasic, J. F. (2012). Improving human-computer interaction in personalized TV recommender, *Iranian Journal of Science & Technology, Transactions of Electrical Engineering*, Vol. **36**, No. E1, pp. 19-36.
3. Breese, J., Heckerman, S. D. & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*1998, Morgan Kaufmann Publishers Inc.: Madison, Wisconsin. pp. 43-52.
4. Deshpande, M. & Karypis, G. (2004). *Item-based top-N recommendation algorithms.* ACM Transactions on Information Systems, Vol. 22, pp. 143-177.
5. Zhong, J. & Li, X. (2010). Unified collaborative filtering model based on combination of latent features. Expert *Systems with Applications*, Vol. **37**, No. 8, pp. 5666-5672.
6. Billsus, D. & Pazzani, M. J. (1998). Learning collaborative information filters. *Proceedings of the Fifteenth International Conference on Machine Learning*1998, Morgan Kaufmann Publishers Inc. pp. 46-54.
7. Zhang, T. & Iyengar, V. S. (2002). *Recommender systems using linear classifiers. J. Mach. Learn. Res*., Vol. 2, pp. 313-334.
8. Ungar, L. & Foster, D. (1998). *Clustering methods for collaborative filtering*. Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence, Madison, Wisconsin, USA, AAAI Press. pp. 112-125.
9. Hofmann, T. (2004). Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, Vol. 22, pp. 89-115.
10. Adomavicius, G. & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans. on Knowl. and Data Eng*., Vol. 17, No. 6, pp. 734-749.
11. Candès, E. & Recht, B. (2009). *Exact* matrix completion via convex optimization. *Foundations of Computational Mathematics*, Vol. 9, No. 6, pp. 717-772.

12. Bonnans, J. F., et al. (2006). *Numerical optimization: Theoretical and practical aspects (Universitext)*. Springer-Verlag New York, Inc.
13. Fazel, M. (2002). *Matrix rank minimization with applications*. Stanford University.
14. Chistov, A. L. & Grigor'ev, D. Y. (1984). *Complexity of quantifier elimination in the theory of algebraically closed fields*, in *Mathematical Foundations of Computer Science 1984*. M.P. Chytil and V. Koubek, Editors, Springer Berlin Heidelberg, pp. 17-31.
15. *http://stanford.edu/~boyd/cvxbook.*
16. Grant, M. & Boyd, S. *www.cvxr.com/cvx*.
17. Im, I. & Hars, A. (2007). Does a one-size recommendation system fit all? the effectiveness of collaborative filtering based recommendation systems across different domains and search modes. *ACM Transactions on Information Systems*, Vol. 26, pp. 1-30.