Multi-dimensional Aggregation Recommendation Algorithm Based on Average Prediction

Jiaxin Han, Wenjuan Wei, Haiyang Xia

School of Computer Science, Xi'an ShiYou University Shaanxi, China 710065

Abstract. Traditional collaborative filtering recommendation algorithm uses single dimensional data to calculate the similarity between users or items, ignoring the user's preference, thus affect the recommendation accuracy. To this end, an averaging forecasting based multi-dimensional aggregation recommendation algorithm was proposed in this paper, which constructs the relationship aggregation function by user's total score and dimension scores firstly, then apply the aggregation function to the initial multi-dimensional score that calculated by the modified averaging forecasting algorithm. The experiment result shows that compared with the previous collaborative filtering based recommendation algorithm, it has higher recommendation accuracy.

1 Introduction

The collaborative filtering algorithm is a kind of most commonly used algorithms in the recommender system. It mainly include user-based collaborative filtering algorithm and item-based collaborative filtering algorithm [1-4], they searching "Neighbor" for achieving rating forecast [5]. The collaborative filtering algorithm was applied in many fields. However, the similarity measure is based on the assumption that users have only one score for each item [6, 7], which ignores user's preference thus affects the accuracy of recommendation.

To fill this research gap, this paper proposes a MDAA (Multi-Dimension Aggregation recommendation based on Average forecasting) algorithm. Firstly, MDAA algorithm uses the multi-dimensional score to reduce the impact of data sparsity on recommendation. Secondly, it adds the average measure to the traditional prediction algorithm, which reduces the impact of user preference on the recommendation. Finally, the algorithm achieves multi-dimensional score aggregation. After experimental verification, the algorithm can effectively improve the accuracy of the recommendation.

2 Collaborative filtering recommendation algorithms

Collaborative filtering recommendation algorithm is mainly divided into three steps: get user-item rating matrix, find nearest neighbor, and calculate unknown scoring [8-10].

The input is a scoring matrix as shown in Table 1, m is the number of users, n is the number of items, and the element in the matrix Rui represents the rating of user u on item i.

Table1. User-item Rating Matrix

	i_1	i_2	i_3	…	i_n
u_1	5	5	4	3	?
u_2	5	5	4	3	5
…	…	…	…	…	…
u_m	4	5	4	4	4

The collaborative filtering recommendation algorithm uses each row in the rating matrix as the user scoring eigenvector, and gets the top-N user sets which is closest to the target user (the nearest neighbor sets) by cosine similarity calculation. The calculation of similarity between users as follows:

$$s_{u,\mathrm{m}} = \frac{u \cdot m}{\|u\|\|m\|} = \frac{\sum R_{u,i} R_{m,i}}{\sqrt{\sum R_{u,i}^2 \sum R_{m,i}^2}} \tag{1}$$

Where $s_{u,\mathrm{m}}$ the similarity between user u and user m is, $R_{u,i}$ denotes the rating of user u on item i, $R_{m,i}$ is the rating of user m on item i.

According to the top-N neighbor sets, the item ratings of target user are calculated by:

$$R(u,i) = \overline{R(u)} + \frac{\sum_{m \in N(u)} s_{u,m} \cdot \left(R(m,i) - \overline{R(m)}\right)}{\sum_{m \in N(u)} \left|s_{u,m}\right|} \tag{2}$$

Where $N(u)$ represents the set of users that is similar to user u, the set is called neighbour group, and $sim(u,m)$ represents the similarity between user u and user m, $\overline{R(u)}$ refers to the average score of users in all items.

* Corresponding author: wenjuan2018@163.com

3 MDAA recommend algorithm

Traditional collaborative filtering recommendation algorithm uses single dimensional data to calculate the similarity between users or items, on the contrast, the MDAA recommendation algorithm is add additional rating information for original users [11-14] and consider the user's rating preferences at the same time, then increase the accuracy of recommendations.

3.1. Get the user-item rating matrix

The difference of one-dimensional recommendation and multi-dimensional recommendation is that the latter has more user rating information and can be used effectively during the recommendation process. The process can be represented as:

$$R(u,i)=(r_1,r_2,...,r_k) \tag{3}$$

Where r1-rk are the user's rating on k-dimensions of each item. The user-item multidimensional rating matrix is shown in Table 2.

Table2. User-multidimensional rating matrix

	i_1			i_2			...	i_n		
	r_1	...	r_k	r_1	...	r_k	...	r_1	...	r_k
u_1	4	...	5	3	...	5	...	5	...	5
u_2	5	...	5	4	...	5	...	3	...	4
...	...	...	...	...	...	...	...	...	...	...
u_m	4	...	5	3	...	5	...	4	...	5

3.2. Calculate the rating similarity

This paper uses multidimensional distance to measure the similarity between users. The smaller distance between two users means the higher similarity between two users. The calculation of similarity is mainly divided into the following three steps:

(1) The calculation of multi-dimensional rating distance at the same project of two users.

$$d_{rating}=\sqrt{\sum_{i=1}^{n}\left|R(u,i)-R(m,i)\right|^2} \tag{4}$$

(2) The calculation of average rating distance of two users.

$$d_{user}(u,m)=\frac{1}{|I(u,m)|}\sum d_{rating}(R(u,i),R(m,i)) \tag{5}$$

Where $I(u,m)$ the number of products is scored jointly by user u and user m, $d_{user}(u,m)$ represents the scoring distance of two users, and d_{rating} is means the scoring distance of two users on item dimension i.

(3) The calculation of user rating similarity.

$$s(u,m)=\frac{1}{1+d_{user}(u,m)} \tag{6}$$

3.3. Calculate initial multidimensional rating

The mean-based prediction algorithm proposed in this paper mainly uses the rating regularization technique to predict the user's initial multidimensional score on the items. It is reduce the impact of user preferences on the score prediction initially.

$$s=R(m,i)-\overline{R(m)} \tag{7}$$

$$z=\frac{1}{\frac{1}{N(u)}[\overline{R(m)}\cdot\sum_{m\varepsilon N(u)}sim(u,m)]} \tag{8}$$

$$R(u,i)=\overline{R(u)}+z\cdot\sum_{m\varepsilon N(u)}s\cdot sim(u,m) \tag{9}$$

Where z is a regularization factor to regularize scores of 1-5, s is to deconcentrate to solve the user's scoring habits on the prediction results, $R(m,i)$ represents the rating of user m at item i.

3.4. Learn the aggregation function

In order to reflect the user's personalization, the paper presents the user's preference by setting the weight in each dimension of the project, and uses the support vector regression machine (SVR) to realize the relationship learning between the total score and each dimension score.

This paper assumes that in the original space, the sample set S is linearly inseparable, so a non-linear mapping is used to map the data into a high-dimensional space, so that there is a very good linear regression feature in the feature space H, Linear regression, and then returned to the original space [15]. Given the training sample data sets $S=\{(x_i,y_i,i=1,2,3,...,q)\}$, (x_i,y_i) refers to the user's score pair, x_i refers to the user's multidimensional score set $(r_1,r_2,...,r_k)$, y_i refers to the user's total score in the corresponding project r_0, q represent the sum of the number of user comment items. Relational learning is achieved in the high-dimensional scoring space, according to the given score on the construction of the optimal linear function.

$$f(x)=(\omega\cdot x)+b \tag{10}$$

The import of ε insensitive loss function, slack variable ξ_i, ξ_i^* and penalty factor c. The original function problem transform into optimization problem.

$$\begin{cases}\min\limits_{\omega,\xi_i,\xi_i^*}\frac{1}{2}\|\omega\|^2+c\sum_{j=1}^{q}\left(\xi_i+\xi_i^*\right)\\ s.t.\quad y_i-f\left(x_i\right)\le\varepsilon+\xi_i\\ \qquad f\left(x_i\right)-y_i\le\varepsilon+\xi_i^*\\ \qquad \xi_i,\xi_i^*\ge 0, i=1,2,\dots,q\end{cases} \tag{11}$$

Using the Lagrange α_i, α_i^*, η_i and η_i^*, transform the quadratic programming problem into dual problem [16].

$$L\left(\omega,b,\xi_i,\xi_i^*\right)=\frac{1}{2}\|\omega\|^2+c\sum_{j=1}^{q}\left(\xi_i+\xi_i^*\right)-\sum_{i=1}^{q}\alpha_i\left(\varepsilon+\xi_i-y_i+f\left(x_i\right)\right)-\sum_{i=1}^{q}\alpha_i\left(\varepsilon+\xi_i^*+y_i-f\left(x_i\right)\right)-\sum_{i=1}^{q}\left(\eta_i\xi_i+\eta_i^*\xi_i^*\right) \tag{12}$$

Derive the partial derivatives of ω, b, ξ_i, ξ_i^*, then we can get an optimal Lagrange coefficient [17].

$$\begin{cases} \max\limits_{\omega} \left\{ -\frac{1}{2}\sum_{i,j=1}^{q}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)x_i^T x_i+\sum_{i=1}^{q}\left(\alpha_i-\alpha_i^*\right)-\varepsilon\sum_{i=1}^{q}\left(\alpha_i+\alpha_i^*\right) \right\} \\ s.t. \qquad \sum_{i=1}^{q}\left(\alpha_i-\alpha_i^*\right)=0, 0\le \alpha_i, \alpha_i^* \le c, i=1,2,\ldots,q \end{cases} \tag{13}$$

Where (13) contains the inequality constraint, so it is need to use the KTT (Karush-Kuhn-Tucker) to get α_i, α_i^* and b.

$$f(x)=\sum_{i=1}^{q}\left(\alpha_i-\alpha_i^*\right)K\left(x_i,x\right)+b \tag{14}$$

Therefore, a kernel function is introduced to get the regression decision function, $K(x_i,x)$ is kernel functions in SVR [18].

4 Experiments

In order to verify the performance of the MDAA algorithm, the experimental compares it with the traditional collaborative filtering recommendation algorithm on two data sets which are got from TripAdvisor. Dataset1 and Dataset2 are users' feedback for Paris Hotel and Beijing Hotel from 2015 to 2017. The detailed information is shown in Table3.

Table3. Datasets Statistic information

Datasets	User's numbers	Hotel's numbers	Dimensional's numbers
Dataset1	1695	917	6
Dataset2	943	1682	6

4.1 Evaluation

The accuracy of the recommendation system is the basic indicator for evaluating the recommendation algorithm. RMSE (Root Mean Squared Error) measures the accuracy of the prediction by calculating the deviation between the predicted user score and the actual user score [8].

$$RMSE=\sqrt{\frac{1}{\left|E^P\right|}\sum\nolimits_{(u,i)\in E^P}\left(r_{ui}-r_{ui}^{'}\right)^2} \tag{15}$$

Where r_{ui} represent the true rating user u on item i, $r_{ui}^{'}$ is the prediction rating of user u on item I, E^P is test data set.

4.2 Experiment design

In the experiment, the user rating value is between 0 and 5, and the score of 0 indicates that the user does not rate the hotel. This paper mainly compares the recommended results between MDAA recommendation and traditional recommendation algorithms in two aspects:

(1) Comparing the recommended results of single-dimension recommendation algorithm, multi-dimension recommendation algorithm and MDAA recommendation algorithm;

(2) Comparing the prediction effects of traditional prediction algorithm and the improved mean prediction algorithm in this paper.

4.3 Result Analysis

In the experiment, each algorithm is applied to two datasets, comparing the traditional recommendation algorithm with MDAA recommendation algorithm, traditional prediction algorithm and mean prediction algorithm.

The interpretation of character in the figures is shown in Table 4.

Table4. The Interpretation of Character

Character	Interpretation
T-Prediction	Original prediction algorithm
M-Prediction	Mean-prediction algorithm
M-Dataset	The dataset based mean-prediction algorithm
T-Dataset	The dataset based original prediction algorithm
Single-dimension	Single-dimensional recommendation
Seven_metric	Multi-dimensional recommendation based metric
SVR_	SVR kernel function

In order to better reflect the recommended results of the MDAA algorithm, this paper uses RMSE error backwash method to achieve the proposed algorithm comparison.

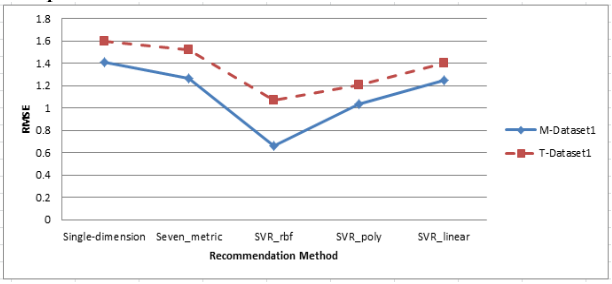

Fig.1. Comparison of T and M-Prediction in Dataset1

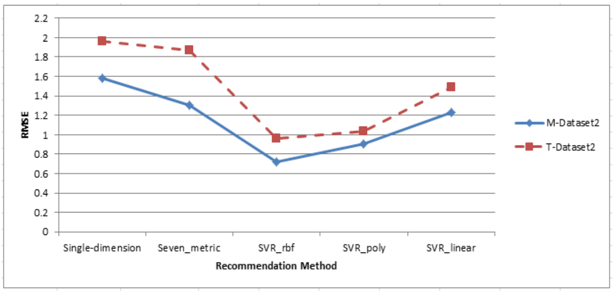

Fig.2. Comparison of T and M- Prediction in Dataset2

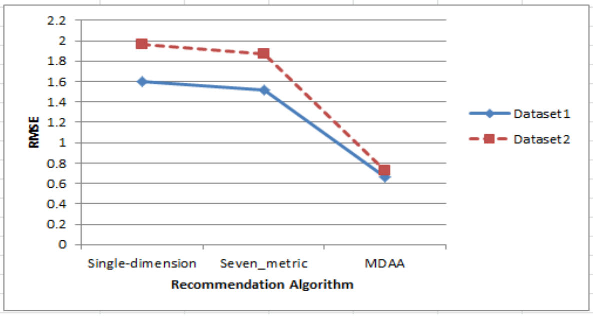

Fig.3. Comparison of single-dimension, multi-dimension and MDAA recommendation

Analysis of Figure 1, Figure 2 and Figure 3, we can get the following conclusions:

(1) In Figure 1 and Figure 2, the line charts obtained by the M-Prediction algorithm are located below the line chart obtained by the T-Prediction algorithm in both data sets, draw a conclusion, not only single dimension recommendation but also multi-dimension recommendation, the M-Prediction algorithm proposed in this paper all improve the accuracy of the prediction.

(2) In Figure 3, comparing the experiment results obtained by the three recommended algorithms, it is obvious for us to find that on the experimental data sets, the MDAA recommendation algorithm has a lower recommendation error than the original single-dimension recommendation algorithm.

5 Conclusions

In order to make up the shortcoming of the traditional collaborative filtering recommendation algorithm that ignores the user's preference. A recommendation algorithm which considers the impact of data sparsity, non-positive scoring information and user personalization on recommendation was proposed in this paper. This algorithm can improve the recommendations accurate by introduction the multidimensional user-item rating matrix, aggregation function and average forecasting technique.

The experiment result shows that compared with the traditional collaborative filtering recommendation algorithm, the MDAA algorithm can provide a good user experience by consider the user's personalization. The next improvement direction of this paper is to consider the external factors such as time and region comprehensively and incorporate the economic model [19] into the recommendation process to achieve a better recommendation.

Acknowledgments

This research is supported by the International Science and Technology Cooperation and Exchange plan Project of Shaanxi province (2016kw-047).

References

1. Wang Guoxia, Liu Heping. Survey of personalized recommendation system. *Computer Engineering and Application*, 2012, 48(7):66-76
2. Yang Bo, Zhao Pengfei. Review of the art of Recommendation Algorithms. *Journal of Shanxi University*(Nat. Sci. Ed.), 2011, 34(3):337-350
3. Xiong Xiangyun, Fu Yunchen, Liu Zhaoqing. Multi-dimensional recommendation algorithm based on bipartite network projection[J]. *Computer Applications and Software*, 2014,(08):253-256
4. Wei S Y,Ye N,Zhang S,et al . Collaborative filtering recommendation algorithm based on item clustering and global similarity. *Proceedings of the Fifth International Conference on Business Intelligence and Financial Engineering*,Lanzhou,2012:69-72
5. Xing Zhe, Liang Jingfan, Zhu Qing. Multi-dimensional Adaptive Collaborative Filtering Recommendation Algorithm. Journal of Chinese Computer Systems, 2011
6. Huang Tao, Huang Ren, Zhang Kun. Improved Collaborative Filtering Recommendation Algorithm. *Computer Science 2016*, 43(6A):400-403
7. Gediminas Adomavicius, NikosManouselis, YoungOk Kwon. Multi-criteria recommender systems. *Recommender Systems Handbook, Second Edition*, 2015:847-880
8. Zhu Yuxiao, LV Linyuan. Evaluation Metrics for Recommendation System. *Journal of University of Electronic Science and Technology of China 2012*, 41(2):163-175
9. Wenjin Wu, Wen Zhang, Ye Yang, et al. DREX: Developer Recommendation with K-Nearest-Neighbor Search and Expertise Ranking. *Proceedings of the 2011 18th Asia-Pacific Software Engineering Conference*, 2011:389-396
10. Natalya I Kovalevs. Recommended practices for improving the competitiveness of the Russian education services market under the conditions of the international educational integration. *Review of European Studies*, 2015
11. Artūras Kaklauskas , Edmundas Kazimieras Zavadskas. Advanced Methods for Computational Collective Intelligence. *Advanced Methods for Computational Collective Intelligence Studies in Computational Intelligence* , 2013
12. Polatidis, Nikolaos, Georgiadis, et al. A dynamic multi-level collaborative filtering method for improved recommendations. *Computer Standards & Interfaces*, 2017, 51:14-21
13. Gediminas Adomavicius, YoungOk Kwon. New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems*, 2007, 22(3): 48-55
14. Ricci, H. Werthner. Case-Based Querying for Travel Planning Recommendation. *Information Technology and Tourism*, 2002, 4: 215–226

15. M. Ghaedi, Mahmoud reza Rahimi, A.M. Ghaedi, et al. Application of least squares support vector regression and linear multiple regression for modeling removal of methyl orange onto tin oxide nanoparticles loaded on activated carbon and activated carbon prepared from Pistacia atlantica wood. *Journal of Colloid and Interface Science*, 2016, 461: 425-434
16. G. Santamaría-Bonfil, A. Reyes-Ballesteros, C. Gershenson. Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 2016, 85: 790-809
17. Ye Ren, Ponnuthurai Nagaratnam Suganthan, Narasimalu Srikanth. A Novel Empirical Mode Decomposition With Support Vector Regression for Wind Speed Forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(8): 1793-1798
18. Subhransu Maji, Alexander C. Berg, Jitendra Malik. Efficient Classification for Additive Kernel SVMS. *Pattern Analysis and Machine Intelligence*, 2013, 35(1): 66-77
19. Ting-Peng Liang, Xin Li, Chin-Tsung Yang, et al. What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, 2015: 236-260