

A Hybrid Recommendation Algorithm Based On SVD And Weighted Slope One

Zhou Wenji*

East China Normal University
School of Information Science Technology
Shanghai, China
Wenji1393@126.com
* Corresponding Author

Yu Qingsong

East China Normal University
School of Information Science Technology
qsyu@cc.ecnu.edu.cn
Shanghai, China

Abstract—With the arrival of the era of big data, recommendation algorithm has become a hot research field in internet life and social business. However, the traditional recommendation algorithms have problems of cold start, scoring matrix data sparse, affecting the accuracy of recommendation. This paper proposes a hybrid recommendation algorithm based on the SVD (singular value decomposition) and Slope One algorithm, experiments results show that the algorithm proposed by this paper has superior performance compared with the traditional collaborative filtering recommendation algorithm.

Keywords—SVD; Slope-One; Recommendation Algorithm; Collaborative Filtering

I. INTRODUCTION

The rapid development of the Internet led people in the 21st century into the era of big data. Under this background, the internet life has produced large amounts of data. Whether it is shopping, learning, entertainment, data redundancy always makes people hard to obtain information quickly and efficiently. With the emergence of recommendation algorithm, this problem has been greatly improved^[1]. Recommendation algorithm has important application value, which is not only a challenging research topic in computer field, but also attracted many researchers from many fields, such as mathematics, physics, cognition, artificial intelligence, management, marketing etc.

Collaborative filtering recommendation algorithm is the most classic and popular one. The traditional collaborative filtering algorithm can be divided into content-based collaborative filtering and item-based collaborative filtering. The basic assumption of the collaborative filtering recommendation algorithm is that it can be recommended to the user by finding the attractive content according to the user's preference. Based on the similarity of the target users and its neighbor users, it makes prediction of target users in the recommended object preference^[2].

Although the collaborative filtering recommendation algorithm has achieved excellent effect and great success, it still exists some problems, such as cold start, scoring sparse matrix^[3]. In fact, whether the e-commerce website or the movie recommendation site, they provide millions kinds of products. Most of the users are not likely to give an evaluation of most products, only 1% or less of the goods have been evaluated. As a result, this leads to the most common problem in a recommendation system, data sparse problem. Data sparse is

also the most severe problem encountered in collaborative filtering algorithm, which will greatly affect the accuracy of the recommendation system. Cold start problem is one of the common problems encountered in practical application. According to the principle of the collaborative filtering algorithm, it finds out the similar users by analyzing the historical information of them. We can foresee that when a new user or a new item join the recommendation system, collaborative filtering algorithm can't find the user's similar users due to the new one which has not established contact to other elements.

To solve these problems, this paper puts forward a kind of hybrid recommendation mechanism based on singular value decomposition and Slope One. Experimental results show that the recommendation accuracy and quality are improved by the integration of two efficient algorithms.

II. SLOPE ONE ALGORITHM

Slope One algorithm is a prediction algorithm based on the score, it is one of the collaborative filtering algorithms. Professor Daniel Lemire put forward the algorithm in 2005. The principle idea of it comes from a simple linear model

$$f(v) = v + b \quad (1)$$

the constant b in the formula represents the average difference between the two items^[3,5].

According to the b from antecedent formula we can use it to predict the score of the new item. Suppose user U_s 's rating for item I_i and item I_j is R_{si} and R_{sj} ; user U_t 's rating for item I_i is R_{ti} , and user U_t 's rating for item I_j remains unknown, we mark it as R_{tj} .

TABLE I. THE SCORE PREDICTION

U/I	...	I_i	...	I_j	...
...
U_s	...	R_{si}	...	R_{sj}	...
...
U_t	...	R_{ti}	...	$R_{tj}=?$...
...

According to the linear regression, the traditional Slope One algorithm shows that

$$R_{ij} = (R_{sj} - R_{si}) + R_{ti} \quad (2)$$

$$dev_{j,i} = \sum_{u \in n} \frac{u_j - u_i}{n} \quad (3)$$

$$\text{Hence, } P(R_{uj}) = \frac{\sum_{k \in n} (dev_{j,i} + U_i)}{n} \quad (4)$$

where the n is the number of the items in the collection of the score, and the $dev_{j,i}$ is the average value of error from the formula(1) as b . $P(R_{uj})$ represents the prediction score^[8].

III. WEIGHTED SLOPE ONE ALGORITHM

If we only consider about the scoring error of two items, obviously it makes the prediction results inaccurate. 100 users of an item score and 1000 users for its score generated by the proportion of influence is apparently not the same. More users evaluate a project, its score will be more representative^[3,4]. Therefore, people improve the traditional Slope One algorithm by weighted.

$$P'(R_{uj}) = \frac{\sum_{i \in S(u) - \{j\}} (b_{jk} + R_{uk}) W_{j,i}}{\sum_{i \in S(u) - \{j\}} W_{j,i}} \quad (5)$$

The weighted slope one algorithm has the following advantages:

- a) *The algorithm is not complicated. It is easy to implement because of the simple process. There is no need to call too many modules, which may affect the system.*
- b) *High efficiency query response. Based on the proposed algorithm, the recommendation system runs fast and can respond to users' requests immediately.*
- c) *Reasonable accuracy. The slope one algorithm performs better than the traditional collaborative filtering algorithm.*

IV. SVD (SINGLE VALUE DECOMPOSITION)

Generally, recommendation system keeps a very sparse rating matrix. Multi-dimensional sparse matrix can reduce data sparseness degree by dimension reduction. Matrix decomposition method is a commonly used matrix dimension reduction method, and it also has the function of data noise filtering. Singular Value Decomposition(SVD) is a commonly used matrix decomposition technique which can effectively extract matrix eigenvalue, reveal the internal structure of the matrix, and also avoid excessive sparse data in a certain degree^[11].

SVD divides a $m \times n$ matrix R into 3 matrices.

$$R = U \times S \times V^T \quad (6)$$

$$X = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \quad (7)$$

R represents scoring matrix, decomposed into an orthogonal matrix U , a diagonal matrix S and an orthogonal matrix V . Intuitively, the singular values in S are "weighting factors" for U and V ^[6,7].

The value of each element in diagonal matrix S decreases, these elements show the change weight of the direction of the eigenvector corresponding to the feature vector. Here, we keep K largest elements, and produce a restructuring matrix of S_k ^[10]. Accordingly, we have restructuring matrix of U_k and V_k in the same way.

From formula(5), we have

$$R_k = U_k \times S_k \times V_k^T \quad (8)$$

Hence, we now have got a matrix R_k which is a Similarity matrix of R and its rank is k . R_k will take place of R as a new scoring matrix.

V. USER SIMILARITY

A. Pearson Correlation Coefficient

In the recommendation system, there are many ways to calculate the similarity of users. This paper takes Pearson Correlation Coefficient.

It is a linear regression model in statics, it can measure linear relationship between two vectors^[2].

$$W_{uv} = \frac{\sum_{i \in n} (R_{ui} - R_u) \cdot (R_{vi} - R_v)}{\sqrt{\sum_{i \in n} (R_{ui} - R_u)^2} \sqrt{\sum_{i \in n} (R_{vi} - R_v)^2}} \quad (9)$$

Here, n represents the user set, R_{ui} represents the score of item i from user u , R_u is the average value of them. Similarly, R_{vi} and R_v represent the corresponding value.

VI. SVD-WEIGHTED SLOPE ONE ALGORITHM

In order to solve the problem of the sparsity of the scoring matrix and improve the efficiency of the algorithm, this paper puts forward a hybrid recommendation algorithm based on SVD and slope one (SVD-Weighted Slope One algorithm). The algorithm can be described as follows:

A. *Input scoring matrix R .*

B. *Reduce the dimensionality of R by formula(5), which is mentioned as single value decomposition.*

- C. Determine the value of k , then we get three matrix as U_k, S_k, V_k^T . After calculation, we have a new similar matrix of R .
- D. The similarity between users can be calculated by improved Pearson Correlation Coefficient in formula (8).
- E. Dynamically select the number of user neighbors, and constitute a neighbor matrix.
- F. We obtain the forecast result according to the weighted Slope-One algorithm.

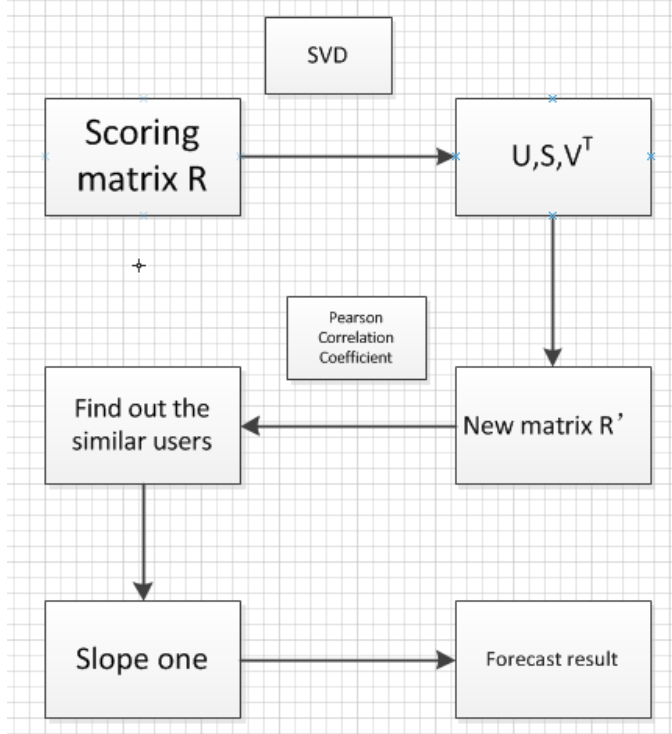


Figure 1. sketch map of the hybrid algorithm

VII. EXPERIMENT

To verify the performance of the hybrid algorithm, we implement the algorithm with python language in the environment of IDEL and MATLAB2010b.

A. Experimental data set

The experimental data is the standard data set Movielens which is collected and established by the GroupLens research group of University of Minnesota. As a movie data set for product research, it contains more than 18326 of users together with their score for almost 9261 of movies. The evaluation is divided into 1 to 5 points, the higher the score, the higher the user's evaluation of the movie. Data sets includes user information, movie information and score information.

B. evaluation criterion

1) RMSE(root mean squared error)

RMSE is the evaluation criteria for the Netflix competition. The smaller the RMSE value, the higher the accuracy of the algorithm.

$$RMSE = \sqrt{\frac{\sum_{u,i \in N} (r_{ui} - r'_{ui})^2}{n}} \quad (10)$$

n represents the number of the items in the collection of the score, r_{ui} represents users' actual score for the items, and the r'_{ui} represents users' predictive score for the items.

2) MAE(mean absolute error)

At present, among the recommended quality evaluation criteria of recommendation algorithms, the most commonly used is the Mean Absolute Error (MAE). It judges the accuracy of algorithm by calculating the deviation between the predictive interest of the user to an item and the real interest[9]. Set the actual score set for $\{q_1, q_2, \dots, q_N\}$, predictive score set for $\{p_1, p_2, \dots, p_N\}$, then the MAE can be defined as

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (11)$$

C. Experiment and result analysis

Firstly, we divided the entire data set into the training set and the test set according to the ratio of 80% and 20%. In this paper, the SVD-Weighted Slope one algorithm is compared to User-CF, Slope One algorithm and Weighted Slope One algorithm.

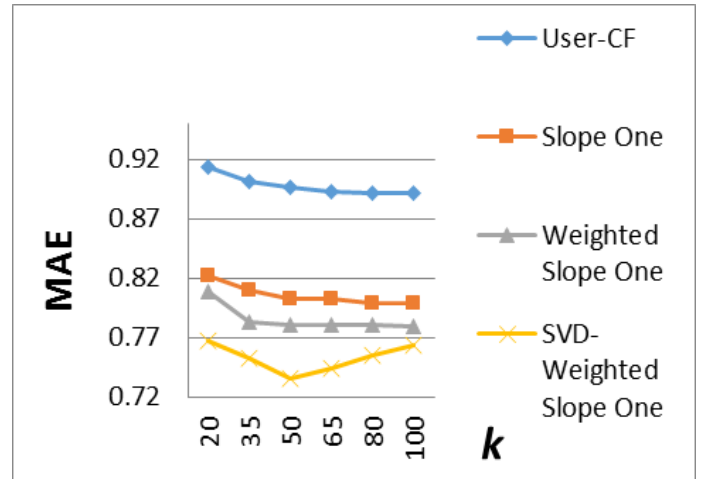


Figure 2. MAE values of 4 algorithms under different k

To investigate the effect of the training set size on the performance of the hybrid algorithm, we divided the data set into different proportions of test set and training set. From Fig.3 and Fig.5, we can see that the MAE value and RMSE value decrease rapidly with the growth of the training set proportion.

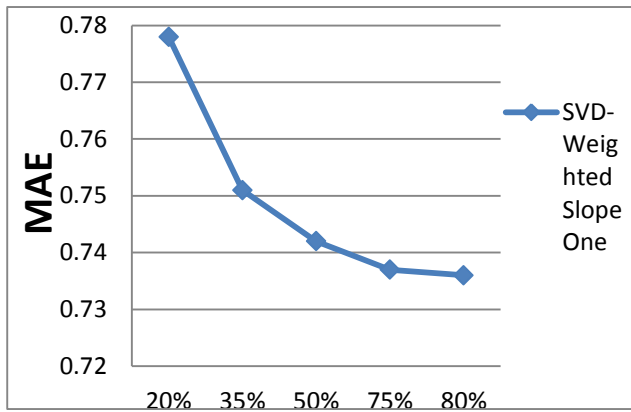


Figure 3. Influence of training set size on MAE value

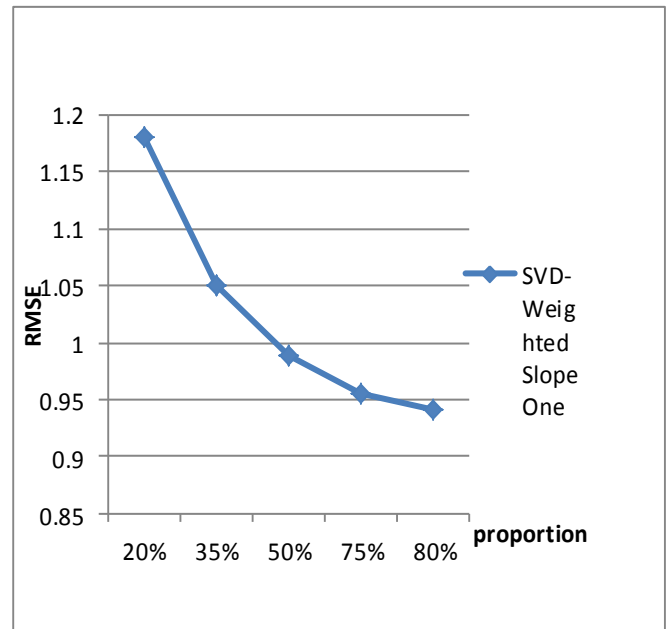


Figure 5. Influence of training set size on RMSE value

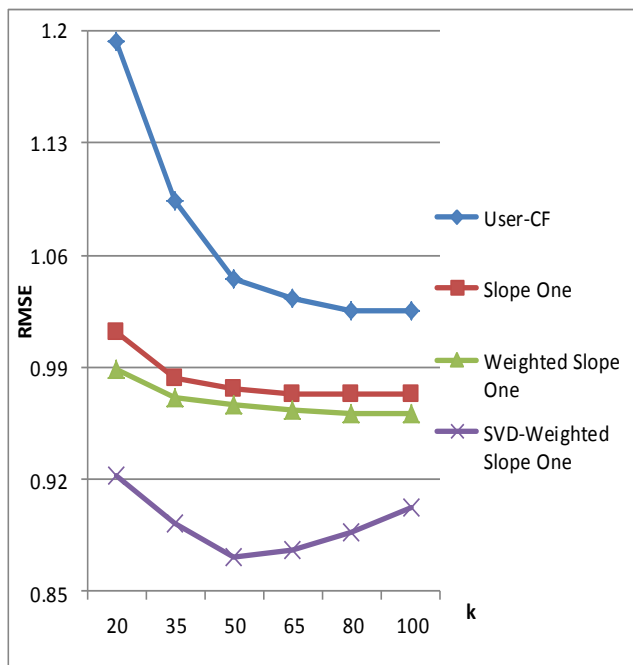


Figure 4. RMSE values of 4 algorithms under different k

From Fig.2 and Fig.4, it can be proved that the proposed algorithm in this paper is significantly improved in the performance than the traditional CF and Slope One algorithm. When the nearest neighbor number is 50, SVD-Weighted Slope One keeps the minimum MAE(0.736) and the minimum RMSE(0.872).

VIII. CONCLUSIONS

Nowadays, the development of the recommendation algorithm is becoming more and more mature. Whether it is based on the content, based on the item or the user, each direction of the algorithm research has made great progress.

This paper presents a new hybrid algorithm based on Slope One and Singular Value Decomposition algorithm aiming at the problem of sparse matrix data. For the characteristics of the two algorithms, we fuse them to produce a new one. According to the experiment analysis, the algorithm can improve the quality of recommendation.

REFERENCES

- [1] Gediminas Adomavicius; Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the State-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(06): 734-749. DOI: 10.1002/pon.1586.
- [2] Badrul Sarwar, George Karypis, Joseph Konstan et al. Item-Based Collaborative Filtering Recommendation Algorithms[C]. // 10th International World Wide Web Conference. 2001: 285-295.
- [3] Kai Yu, Xiaowei Xu, Martin Ester et al. Feature weighting and instance selection for collaborative filtering: an information-theoretic approach[J]. Knowledge and information systems, 2003, 5(2): 201-224.
- [4] MUKUND DESHPANDE, GEORGE KARYPIS. Item-Based Top-N Recommendation Algorithms[J]. ACM transactions on information systems, 2004, 22(1): 143-177.
- [5] Li J, Sun L, Wang J. A slope one collaborative filtering recommendation algorithm using uncertain neighbors optimizing[M]. // Web-Age Information Management. Springer Berlin Heidelberg, 2012: 160-166.
- [6] Zhou, Xun, He, Jing, Huang, Guangyan et al. A Personalized Recommendation Algorithm Based on Approximating the Singular Value Decomposition (ApproSVD)[C]. // The 2012 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology. vol. 2. 2012: 458-464.

- [7] Tingda Lu,Yan Wang,William Perrizo et al.Extension Study on Item-Based P-Tree Collaborative Filtering Algorithm for Netflix Prize[C].//25th International Conference on Computers and Their Applications 2010 (CATA-2010).2010:180-185.
- [8] Lemire D, Maclachlan A. Slope One Predictors for Online Rating-Based Collaborative Filtering[J]. in SIAM Data Mining (SDM05, 2007:21--23.
- [9] Yaquan Xu,Haibo Wang,Gary Kochenberger et al.A SVD/SVM MODEL FOR E-COMMERCE RECOMMENDER SYSTEMS[C].//Decision Sciences Institute Annual Meeting; 20071117-20; Phoenix,AZ(US).2007:106-111.
- [10] Xiangyu Zhao,Zhendong Niu,Wei Chen et al.Interest before liking: Two-step recommendation approaches[J].Knowledge-based systems,2013,48(Aug.):46-56.
- [11] Huseyin Polat,Wenliang Du.SVD-based collaborative filtering with privacy[C].//Proceedings of the 2005 ACM symposium on Applied computing.2005:791-795.
- [12] J. Albaladejo,C. Dominguez,L. Escolar et al.A hybrid structured methodology for developing computer-based industrial computer systems[C].//Proceedings of the 2012 international conference on frontiers in education: computer science & computer engineering. v.1.2012:321-327.
- [13] Dongsheng Li,Qin Lv,Xing Xie et al.Interest-based real-time content recommendation in online social communities[J].Knowledge-based systems,2012,28:1-12.
- [14] Qi ZHANG,Hao YANG,Yuguang WEI et al.Selection of Destination Ports of Inland-Port-Transferring RHCTS Based on Sea-Rail Combined Container Transportation[C].//Innovation and sustainability of modern railway: Third international symposium on Innovation and sustainability of modern railway (ISMIR 2012), September 20-21, 2012, Nanchang, China.2012:675-680.
- [15] Zi-Ke Zhang, Lu Yu, Kuan Fang et al.. Website-oriented recommendation based on heat spreading and tag-aware collaborative filtering[J]. Physica A: Statistical Mechanics and its Applications, 2014, 399.
- [16] Qinqiao Song, Guangtao Wang, Chao Wang. Automatic recommendation of classification algorithms based on data set characteristics[J]. Pattern Recognition, 2012, 45(7).
- [17] Ralf Krestel, Peter Fankhauser. Personalized topic-based tag recommendation[J]. Neurocomputing, 2011, 76(1).
- [18] Nooshin Hosseini. Mo1908 Predicting Colorectal Cancer Survival: A Data Mining Approach[J]. Gastroenterology, 2014, 146(5).
- [19] Dmitry Sokolov, Dimitri Plemenos, Karim Tamine. Methods and data structures for virtual world exploration[J]. The Visual Computer, 2006, 22(7).
- [20] Hao Wen, Liping Fang, Ling Guan. A hybrid approach for personalized recommendation of news on the Web[J]. Expert Systems with Applications, 2012, 39(5).
- [21] J Lian, Y Liu, Z j Zhang et al.. Personalized recommendation via an improved NBI algorithm and user influence model in a Microblog network[J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(19).