International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET207445

Analysis and Research on Increased Probability Matrix Factorization Techniques in Collaborative Filtering

Kongari Mounika[1], B. V. N. Krishna Suresh[2]

[1]M. Tech Scholar Department of CSE, NRI Institute of Technology Visadala, Guntur(Dt), Andhra Pradesh, India

[2]Assistant Professor Department of CSE, NRI Institute of Technology Visadala, Guntur(Dt), Andhra Pradesh, India

Article Info
Volume 7 Issue 4
Page Number : 182-187
Publication Issue :
July-August-2020

Article History
Accepted : 01 Aug 2020
Published : 05 Aug 2020

ABSTRACT

The matrix factorization algorithms such as the matrix factorization technique (MF), singular value decomposition (SVD) and the probability matrix factorization (PMF) and so on, are summarized and compared. Based on the above research work, a kind of improved probability matrix factorization algorithm called MPMF is proposed in this paper. MPMF determines the optimal value of dimension D of both the user feature vector and the item feature vector through experiments. The complexity of the algorithm scales linearly with the number of observations, which can be applied to massive data and has very good scalability. Experimental results show that MPMF can not only achieve higher recommendation accuracy, but also improve the efficiency of the algorithm in sparse and unbalanced data sets compared with other related algorithms.

Keywords : Matrix Factorization, Collaborative Filtering, Recommendation system, SVD, PMF

I. INTRODUCTION

Traditional collaborative filtering approaches can neither handle large data sets, nor solve the problem of data scarcity. In this paper, to solve the problems mentioned above, basic methods of matrix factorization are discussed, and the collaborative filtering (CF) technologies based on matrix factorization algorithms are deeply analyzed. Collaborative filtering is one of the most widely used techniques in recommendation systems, which filters some unrelated information using the degree of information association, and retains the useful parts to users. Tapestry, GroupLens, Ringo and Video Recommender are relatively early recommender systems [1].Tapestry is the first proposed recommender system based on collaborative filtering, in which the target users need to clearly list other users whose behavior are more similar to theirs. GroupLens is a kind of automated collaborative filtering recommender system based on user ratings for recommending movies and news. Recommender systems such as Ringo and Video usually recommend music and movies via e-mail. Breese [2] et al., preceded an in-depth analysis of various collaborative filtering recommendation algorithms and their improvements. Amazon.com, Jester and WWW are some other recommender systems based on collaborative filtering [3]. Currently, collaborative filtering algorithm is mainly classified into three categories: memory-based collaborative filtering, model-based collaborative filtering and hybrid recommendation. The most frequently used model in collaborative filtering recommender systems is the k-Nearest Neighbor (kNN), which includes two kinds of technologies of usersbased recommendation and items-based recommendation [4]. Examples of model-based approaches include factor model [5], Bayesian classification model [6], clustering model [7] and graph model [8]. In the model-based approaches, training datasets are firstly used to learn a predictive model offline, and then the model is applied to online systems to make recommendations. The key of the algorithm is how to learn an effective predictive model [9]. The problem to be solved in this paper is mainly about data sparsity and scalability, which is ubiquitous in traditional collaborative filtering recommendation algorithms and accordingly a modified probability matrix factorization algorithm (MPMF) is presented. Firstly, basic method of matrix factorization (MF) was discussed deeply and the collaborative filtering technologies based on MF algorithm were analyzed extensively. The MF algorithms such as singular value factorization (SVD) and the probability matrix factorization (PMF) etc., were analyzed from the aspects of their performance, and the existing problems of these algorithms were also stated. Secondly, the impact of user feature vectors dimensions and item feature vectors dimensions on the recommendation accuracy and recommendation efficiency was observed by means of experiments, as well as the consistency of the recommendation results of PMF algorithms both in the training set and data sets. Finally, the optimal parameters of MPMF algorithm were determined by experiments. Experimental results in public dataset Netflix indicated that MPMF algorithm was superior to traditional algorithms of collaborative filtering recommendation in both running efficiency and accuracy.

II. RELATED WORK

The Netflix dataset is very sparse with its sparse degree of 1%. In order to solve the problem of data sparsity and algorithmic scalability in the collaborative filtering algorithms, the researchers at home and abroad have conducted a series of studies and proposed a variety of solutions. Chen [10] et al., have conducted collaborative filtering recommendation of products that target users may be interested in combined with the revenue of recommendation results. Yang [11] et al., put forward a collaborative filtering method based on inferences. Xue [12] et al., proposed a method based on clustering in 2005, which divided users into k classes and chose all users in the nearest class from target users as neighbors and calculated the similarity between them; however the uncertainty of the number of clusters still remains a problem in this method. Tomoharu [13] et al., recommended goods that users are most likely to purchase through collaborative filtering algorithm based on maximum entropy principle. Park [14] et al., improve the recommendation accuracy of Yahoo! by combining collaborative filtering and search engine. Recently an important research trend in recommender systems is the application of both latent semantic models and matrix factorization techniques in collaborative filtering systems. What they have in common is the complementing of the scoring matrix through dimensionality reduction methods. Matrix factorization techniques can provide more accurate recommended items to users through the reduction of dimensions of the sparse matrix. At present matrix factorization methods mainly include nonnegative matrix factorization (NMF) [15], singular value decomposition (SVD) [15], probability matrix factorization (PMF) and Bayesian probability matrix factorization (BPMF). In 1999, both Lee and Seung published nonnegative matrix factorization algorithm in Nature [17], which immediately drew the attention of many researchers in collaborative filtering systems. The main idea of NMF is that high dimensional matrix can be decomposed into the product of two or more low-dimensional matrices in order to study the nature of high dimensional matrix in low-dimensional space [18]. Sarwar et al., published an article on the application of the singular value decomposition algorithm to the design of the recommender system in 2000 [19]. In SVD, the feature values of data set are denoted by singular value and are ranked according to their importance, which achieves the purpose of dimensionality reduction by means of discarding unimportant feature vectors. SVD algorithm did not get much attention in the field of recommender systems when it was proposed after a few years because of its two disadvantages- over-fitting and lack of precision. In 2006, Simon Funk proposed an improved SVD method in his Blog, which attracted the attention of the matrix factorization method throughout the academic field [20]. The matrix factorization method proposed by Simon Funk was subsequently called latent factor model (LFM) by Netflix Prize winner Koren. Matrix factorization can lead to better recommendation accuracy after many tests of KDDCup [21] and Netflix [22] competition.

III. THE DEFINITION OF FUNDAMENTAL MATRIX FACTORIZATION MODEL

The notations of this paper as summarized in Table 1.

Table 1. Mathematical Notations

Notation	Description
N	Number of users
M	Number of items
R	User-item rating matrix of N users on M items
$U \in RD \times N$	User's latent matrix of feature vectors
$V \in RD \times M$	Item's latent matrix of feature vectors
U_i	Specific user's latent feature vectors (column vector)
V_j	Item-specific latent feature vectors (column vector)
U	Current user
V	Current item
r_{ij}	Real rating by user i on item j

IV. THE IMPROVED PROBABILITY MATRIX FACTORIZATION ALGORITHM

MPMF PMF algorithm is a kind of matrix factorization technique based on probability. Matrix factorization tends to be transformed into an optimization problem with its local optimal solution by iteration. Experiment 1 (as seen in Section 6.2) showed that dimension D has a great impact on execution time. How to reduce the dimension of feature vectors while ensuring the accuracy of recommendation algorithm remains an issue of vital importance, since the running time that the user might accept could not be too long for lots of the recommendation systems. In real systems the key problem affecting the recommendation accuracy is to determine dimensions of both user feature vectors and item feature vectors, the value of which researchers usually give according to their experiences, while how to get an optimal value has been a hard problem. The solution of PMF recommendation algorithm is done by the gradient descent method, during which the rate of error declining becomes increasingly slow, as a result PMF requires more time of iterations and training. In order to solve these problems in PMF algorithms we propose an improved probability matrix factorization algorithm named MPMF. The algorithm consists of the following four steps: maximizing the log-posterior probability, solving implicit feature vectors, normalization process and limiting the dimension of feature vectors. At last we compare our method MPMF with two popular methods: PMF and traditional collaborative filtering algorithms.

V. EXPERIMENTAL ANALYSIS

5.1. Experiment Scheme In this section, we conduct four experiments to compare recommendation quality of our approach with other state-of-the-art collaborative filtering methods. The four experiments are as follows: Experiment 1. Comparison of algorithm performance: we analyzed the effect of the feature vector dimension D on execution time of PMF by changing its value, which was completed in Section 6.2. Experiment 2. We analyzed different impacts of dimension D on RMSE both in the training set and in the testing set, which was completed in Section 6.3. Experiment 3. We analyzed the impact of the feature vector dimension D on RMSE by changing its value , which was completed in Section 6.4. Experiment 4. Prediction accuracy comparison: we compared the recommendation accuracy of improved MPMF with basic Netflix algorithm, SVD algorithm and PMF algorithm, which was completed in Section 6.5. The impacts of feature vector dimension D on PMF algorithm is illustrated through experiment 1. The optimal value of the dimension D is determined through experiment 2 and 3. And the recommendation accuracy of the algorithm MPMF is tested through experiment 4.

5.2. The Impacts of Dimension D on Running Time of PMF We designed Experiment 1 to verify the impacts of dimension D of both the user feature vector and the item feature vector on execution time and the experimental results is shown in Figure 2. As can be seen from the figure, with the value of dimension D increases, the running time of PMF algorithm becomes increasingly long and the operating efficiency really declines substantially.

5.3. Analysis of Time Complexity of MPMF The time complexity of matrix factorization algorithm derives mainly from gradient descent, and the computational cost mainly comes from the objective function C and the corresponding gradient descent formula. Matrices U and V both are sparse. The time complexity of objective functions from equations (6) is VU nDDOn , in which U n and V n represent the number of non-zero elements in matrix U and V respectively. Hence, the time complexity is VU nDDOn at each iteration.

VI. CONCLUSION

In this paper, aiming at solving defects of traditional recommender systems, firstly we gave a deep discussion of basic methods of matrix factorization, and then we put forward a kind of improved probability matrix factorization algorithm called MPMF, which has been turned out to be a better way of addressing data sparseness problem with the characteristics of easier programming and lower time complexity. Finally, we verified through experiments that the prediction accuracy of algorithm MPMF is quite high with good scalability on real data sets, and analyzed the impacts of feature vector dimension on recommendation precision and efficiency. We are ready to deploy and implement an actual system and hope to introduce social relationship information among users in the following work. We are to propose a collaborative filtering method which integrates user ratings and information of user social relations in order to further enhance the recommendation accuracy. Besides, how to determine adaptively the optimal parameters still remains a job of further exploration and research.

V. REFERENCES

[1]. L. Lu, M. Medo and C. H. Yeung, “Recommender systems”, Physics Reports, vol. 519, no. 1, (2012), pp. 149.

[2]. J. S. Breese, D. Heckerman and C. Kadie “Empirical analysis of predictive algorithms for collaborative filtering”, Proceedings of the Fourteenth conference on Uncertainty in

artificial intelligence, Morgan Kaufmann Publishers Inc., (1998), pp. 43-52.

[3]. F. Ricci, L. Rokach and B. Shapira , “Introduction to recommender systems handbook”, Recommender Systems Handbook, Springer US, (2011), pp. 1-35.

[4]. M. Jahrer, A. Töscher and R. Legenstein , “Combining predictions for accurate recommender systems”, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, (2010), pp. 693-702.

[5]. R. Bell, Y. Koren and C. Volinsky, “Modeling relationships at multiple scales to improve accuracy of large recommender systems”, Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, (2007), pp. 95-104.

[6]. L. M. De Campos, J. M. Fernández-Luna and J. F. Huete, “Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks”, International journal of approximate reasoning, vol. 51, no. 7, (2010), pp. 785-799.

[7]. H. Wu, Y. J. Wang and Z. Wang, “Two-Phase collaborative filtering algorithm based on co-clustering”, Journal of Software, vol. 21, no. 5, (2010), pp. 1042-1054.

[8]. C. C. Aggarwal, J. L. Wolf and K. L. Wu, “Horting hatches an egg: A new graph-theoretic approach to collaborative filtering”, Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, (1999), pp. 201-212.

[9]. G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions”, Knowledge and Data Engineering, IEEE Transactions on, vol. 17, no. 6, (2005), pp. 734-749.

[10]. M. C. Chen, L. S. Chen and F. H. Hsu, “HPRS: A profitability based recommender system”, Industrial Engineering and Engineering Management, (2007), pp. 219-223.

[11]. J. M. Yang and K. F. Li, “An inference-based collaborative filtering approach”, Dependable, Autonomic and Secure Computing, (2007), pp. 84-94.

[12]. G. R. Xue, C. Lin and Q. Yang, “Scalable collaborative filtering using cluster-based smoothing”, Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, (2005), pp. 114-121.

[13]. T. Iwata, K. Saito and T. Yamada, “Modeling user behavior in recommender systems based on maximum entropy”, Proceedings of the 16th international conference on World Wide Web. ACM, (2007), pp. 12811282.

[14]. S. T. Park and D. M. Pennock, “Applying collaborative filtering techniques to movie search for better ranking and browsing”, Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, (2007), pp. 550-559.

[15]. Q. Gu, J. Zhou and C. H. Q. Ding, “Collaborative Filtering: Weighted Nonnegative Matrix Factorization Incorporating User and Item Graphs”, SDM, (2010), pp. 199-210.

[16]. S. Zhang, W. Wang and J. Ford, “Using singular value decomposition approximation for collaborative filtering”, E-Commerce Technology, (2005), pp. 257-264.

[17]. D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization”, Nature, vol. 401, no. 6755, (1999), pp. 788-791.

[18]. D. D. Tu, C. C. Shu and H. Y. Yu, “Using unified probabilistic matrix factorization for contextual advertisement recommendation”, Journal of Software, vol. 24, no. 3, (2013), pp. 454-464.

[19]. B. Sarwar, G. Karypis and J. Konstan, “Application of dimensionality reduction in recommender system-a case study”, MINNESOTA UNIV MINNEAPOLIS DEPT OF COMPUTER SCIENCE, (2000).
[20]. S. Funk, “Netflix update: Try this at home”, http://sifter.org/~simon/journal/20061211.html, (2006).
[21]. T. Chen, W. Zhang and Q. Lu, “SVDFeature: A Toolkit for Feature-based Collaborative Filtering”, Journal of Machine Learning Research, vol. 13, (2012), pp. 3619-3622.
[22]. R. M. Bell and Y. Koren, “Lessons from the Netflix prize challenge”, ACM SIGKDD Explorations Newsletter, vol. 9, no. 2, (2007), pp. 75-79.
[23]. Y. Koren, R. Bell and C. Volinsky, “Matrix factorization techniques for recommender systems”, Computer, vol. 42, no. 8, (2009), pp. 30-37.
[24]. A. Mnih and R. Salakhutdinov, “Probabilistic matrix factorization”, Advances in neural information processing systems, (2007), pp. 1257-1264.
[25]. Y. N. Fang, Y. F. Guo and X. T. Ding, “An improved singular value decomposition recommender algorithm based on local structures”, Journal of Electronics & Information Technology, vol. 35, no. 6, (2013), pp. 1284-1289.
[26]. N. D. Lawrence and R. Urtasun, “Non-linear matrix factorization with Gaussian processes”, Proceedings of the 26th Annual International Conference on Machine Learning, ACM, (2009), pp. 601-608.
[27]. R. Salakhutdinov and A. Mnih, “Bayesian probabilistic matrix factorization using Markov chain Monte Carlo”, Proceedings of the 25th international conference on Machine learning. ACM, (2008), pp. 880-887.

Cite this article as :

Kongari Mounika, B. V. N. Krishna Suresh, "Analysis and Research on Increased Probability Matrix Factorization Techniques in Collaborative Filtering", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 4, pp. 182-187, July-August 2020. Available at doi : https://doi.org/10.32628/IJSRSET207445
Journal URL : http://ijsrset.com/IJSRSET207445