

An Improved Collaborative Filtering Recommendation Algorithm

Li-Yong WAN^{1,a*}, Lei XIA^{2,b}

¹ School of Electrical and Information Engineering, Nanchang Institute of science & Technology, Nanchang, 330108, China

² Information technology and training center, Nanchang Institute of science & Technology, Nanchang, 330108, China;

^awanliyong@163.com, ^bcsuzju@163.com

Corresponding Author: Li-Yong WAN

Keywords: Collaborative Filtering, Recommendation, CLIQUE, Recommended Efficiency.

Abstract. With the inflationary development of data sizes, the traditional recommendation system the recommendation efficiency of the recommendation system encounters a great challenge. This paper proposes an improved collaborative filtering algorithm. The algorithm firstly implements the clustering operation data, takes average rating and evaluated times of item as dimension of two dimensional grids, and then implements clustering based on item with CLIQUE grid clustering algorithm in accordance with item similarity. The corresponding experiment shows the method can significantly improve the recommendation efficiency of the recommendation system.

Introduction

With the development of the corresponding technology, many e-commerce websites and technology companies have focused on all kinds of the recommendation systems based on the smart algorithms [1-3]. Some researches of these algorithms are supported by many developer and researchers [4]. With the increasingly drastic market competition, almost all of e-commerce websites have improved own competition by using some different smart algorithms, so as to meet the needs of users' requirement. The collaborative filtering recommendation algorithm is one of the most successful recommendation algorithms. The basic idea of the collaborative recommendation algorithm implements recommendation for users by selecting the item of the high rating in accordance with predictive rating of rating of the nearest neighbors [5]. With the development of internet and electrical information technology, amount of commodity and database information will become more and more increasing. The traditional collaborative filtering recommendation finishes recommendation by searching the nearest neighbors of objective item in the whole database, which makes the recommendation efficiency of the recommendation system have been a great challenge.

The Collaborative Recommendation Algorithm Based on Item

The traditional collaborative filtering recommendation finishes recommendation by searching the nearest neighbor of objective item in the whole database, which implements recommendation for users by selecting the item of the high rating in accordance with predictive rating of rating of the nearest neighbor. This algorithm needs three steps to finish final recommendation. Step 1: getting the rating data table based on users' item, and producing the corresponding user-items rating matrix; Step 2: Computing the similarity of objective items and common items in accordance with the similarity algorithm, so as to get the neighbors sets of the objective item; Step 3: Computing the predictive rating of objective items by rating of neighbors' items, and implementing the final recommendation by selecting the former high items.

Computing the Similarity of User-item Matrix

The nearest items sets of objective items can be generated by similarity computing in accordance with the user-items matrix. According to above the data matrix of user rating, the items neighbor sets $T = \{t_1, t_2, \dots, t_n\}$ of sequential order of objective items can be generated by similarity computing formula. The sets T is the neighbors' items set of objective items of sets T, where $sim(i, j)$ represents the similarity between item i and objective item j , and the $sim(i, j)$ threshold is (0, 1). The existing methods of computing the similarity between items and items include Pearson correlation similarity, cosine correlation and modified cosine correlation, etc. The formula of the Pearson correlation similarity is shown in formula (1).

$$sim(i, j) = \frac{\sum_{n \in U} (R_{u,i} - \bar{R}_i)(R_{n,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}} \quad (1)$$

Where U represents user sets, $R_{u,i}$ represents that a user u rates the rating of items i . \bar{R}_i represents the average rating that all uses rate items i ; \bar{R}_j represents the average rating that all uses rate items j .

Producing the Recommendation

After getting the nearest neighbor, the rating of non-rating items will be computed. Computing the predictive rating that user u rates non-rating items t and getting the final recommendation. The computing of rating is shown in formula (2).

$$P_{u,t} = \bar{R}_t + \frac{\sum_{n \in T_i} sim(t, n) * (R_{u,n} - \bar{R}_n)}{\sum_{n \in T_i} (|sim(t, n)|)} \quad (2)$$

Where the $sim(t, n)$ represents similarity between the objectives items t and the nearest neighbor n . The items t and items n are items of items neighbors' sets T_i . The $R_{u,n}$ represents the rating that user u rates item n . The \bar{R}_t represents the average rating that user rates item t , and \bar{R}_n represents the average rating that user rates item n . According to above the computing method, the prediction that user u rates non-rating item t can be generated. Finally, selecting the several higher items of all the ratings, and submitting the recommendation results to user.

The Collaborative Filtering Recommendation Algorithm Based on Item Clustering

In recent years, there are some collaborative filtering recommendation algorithms by using clustering optimization. These algorithms are basically based on data-driven. While the CLIQUE clustering is based on space-driven, which has advantage of high-speed, the processing time is independent of data objects, which only depends on of quantitative space. It is crucial to mass data processing. This paper implements clustering of items by using grids clustering algorithm [6]. This method can generate different cluster of items, which makes the objective item quickly find out items neighbors sets by cluster of CLIQUE. The CLIQUE algorithm will generate some discrete point in accordance with setting different density threshold when working data, and divide them into discrete point cluster. There are some items that are not in cluster. For them, so long as similarity of every cluster center has acquired, and have divided into similar cluster of cluster center, the final cluster can be generated. The improved collaborative filtering recommendation algorithm is based on CLIQUE cluster algorithm, which takes full advantage of high efficiency of CLIQUE cluster algorithm.

The CLIQUE Grid Clustering

According to above the method, implementing the cluster of CLIQUE grid. Making items of the similarity fell into the corresponding cluster as far possible as.

Algorithm 1. The cluster algorithm based on items.

Input: The configuration information of the grid cluster and database *Ratings* of user rating.

Output: Dividing into s cluster.

Step 1. Querying the numbers of rating and average rating of every films from database *Ratings* of user rating, and getting the sets $M = \{m_1, m_2, \dots, m_n\}$, where $m_i = \{movieID_i, count_i, aveRating\}$

($i=1,2,3$). $movieID_i$ represents ID of the film, $count_i$ represents the number of comments, $avgRating_i$ represents the average rating of the film.

Step 2. Embedding M table into grids unit of taking the $count_i$ and $avgRating_i$ attribute. Calculating grids information first, and then scanning the grids. If the density is greater than threshold, creating the new cluster, and recursively absorb its neighbors. If the density of neighbor is greater than threshold, and its neighbor will be absorbed into the cluster, until the end of the cluster. And continue to scan the unchecked grids, repeat the above steps until all of the grids are checked. After finishing cluster, the $s+1$ clusters will be produced.

Step 3. Computing the center of the other n clusters except for $class_{s+1} = 0$, and getting table $C = \{c_1, c_2, \dots, c_s\}$, $c_i = \{movieID_i, count_i, aveRating, class\}$ ($i=1, 2, 3$). Computing the discrete items m_i (m_i represents item of $class_{s+1} = 0$) and the similarity $sim\{m_i, c_i\}$ of cluster center c_i , getting the $sim\{m_i, c_i\} = \max\{sim\{m_i, c_1\}, sim\{m_i, c_2\}, \dots, sim\{m_i, c_s\}\}$, dividing the discrete items into the most similar class. The n clusters will be produced.

The average rating of every item is allocated to the same cluster and similar items can be by the CLIQUE grid cluster algorithm, the generated cluster is expressed in sets $C = \{c_1, c_2, \dots, c_s\}$. Computing the similar neighbor of objective item by the generated cluster, and taking the objective item as the query spacing of cluster, so as to query the similar neighbor of objective item in cluster. Though the items cluster needs some time, the processing can run off-line. In addition, it has zero effect on the requirement of real-time. Before recommendation, the items have been divided into different cluster before the recommendation, the objective space of querying become very small in collaborative filtering recommendation.

Collaborative Filtering Recommendation

After getting the cluster by above the method, the final recommendation can be generated by collaborative filtering recommendation algorithm.

Algorithm 2. The nearest neighbor query.

Input: Object item t , the numbers k of the nearest neighbors, database Ratings of user rating, item cluster C ;

Output: The nearest neighbors of objective items.

Step 1. Inputting object item t , query item cluster sets $C = \{c_1, c_2, \dots, c_s\}$, getting the cluster of the objective item.

Step 2. Computing the similarity $sim\{t, c_i\}$ between cluster c_i and item t , taking the top k items of the nearest similarity as the nearest neighbor.

Getting the nearest neighbors of the objective items by the proposed method, and producing the recommendation by collaborative filtering recommendation method. Suppose $N = \{n_1, n_2, \dots, n_k\}$ was used to present the nearest neighbor sets of objective item t . And the predictive rating $p_{u,t}$ that user u rates item t by rating that user u rates the items of the nearest neighbor set N , this computing methods is shown in formula (3).

$$P_{u,t} = \bar{R}_t + \frac{\sum_{n \in T_i} sim(t, n) * (R_{n,n} - R_n)}{\sum_{n \in T_i} (|sim(t, n)|)} \quad (3)$$

The Experimental Result and Analysis

The Experimental Environment

The configurable CPUs: Core i3-2350M and 4G RAM; Operating system: Windows 8.1 professional edition; Databse: Oracle 10g. The experimental program is built with java language. The experimental dataset comes from movieLens. The dataset includes 100,000 rating records that 943 users rate 1,682 films.

The Comparison of Experimental Data

In the experiment, we compare the improved collaborative filtering recommendation of the paper with traditional collaborative filtering recommendation. Let density threshold of grids cluster to be 20, and density ratio threshold to be 0.6. And getting 3 clusters and 1 discrete point sets. The finally, three clusters can be acquired through dividing the similarity between discrete point and cluster center into the corresponding clusters. The neighbors range is from 10 to 50. The response time of recommendation of two algorithms is shown in figure 1. The comparison of the precision is shown in figure 2.

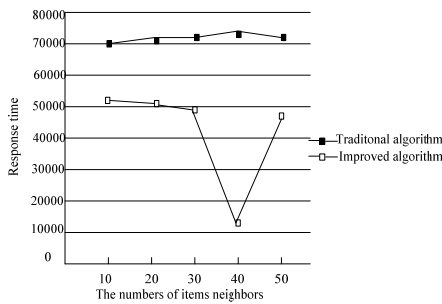


Fig.1. Comparison of the response time

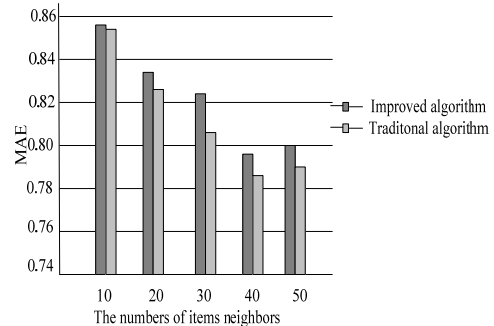


Fig.2. Comparison of the precision

Summary

With the rapid development of network technology, the magnitudes of users and items data grow rapidly, and the recommendation efficiency of the recommendation system encounters a great challenge. In practice application, with the expanding development of data sizes and we must take it into account about the problem of data processing efficiency. This paper the method of grid clustering based on spacing-driven, which divides embedding space into each individual unit of inputting object distributions. The experiments result shows that the improved algorithm improves the recommendation efficiency of the collaborative filtering algorithm to the extent.

Acknowledgement

This research was financially supported by the financial aid of the National Science-technology Support Plan Projects, Grant no. 2014BAD10B00, Science & Technology of Nanchang Institute of science & Technology, under Grant no. GJKJ-14-86.

References

- [1] A. L.DENG, Y.Y. ZHU, B.L. SHI. The Collaborative Filtering Algorithm Based on Prediction of Item Rating, *Journal of Software*, 14(2003) 1621-1625.
- [2] G.H. HE, S.G.WEI, W.T. WANG. An Improved the Cluster Neighbors Collaborative Filtering Recommendation Algorithm, *Computer Science*, 11(2004) 147-149.
- [3] Q.WANG, J.B.WANG. An Improved the Collaborative Filtering Recommendation Algorithm, *Computer Science*, 6 (2010) 226-228.
- [4] J.B. SCHAFER, J. KONSTAN and J. RIEDL. Recommender Systems in E-commerce. In: *ACM Conference on Electronic Commerce*, New York, 1999, pp. 158-166.
- [5] D.W.PENG, B.HU. A Collaborative Filtering Recommendation Algorithm Based on Users Feature and Time, *Wuhan Technology University Journal*, 2(2009) 24-28.
- [6] H. JIAWEI, K. MICHELINE. *Data Mining: Concepts and Techniques*. MORGAN KAUFMANN,2000[EB/OL].<http://www.cs.sfu.edu.ca/~han/>.