Acta Technica *62* (2017), No. 4A, 623–628

Research on collaborative filtering algorithm based on nonnegative matrix decomposition model

Tianguo Tang[1]

Abstract. Collaborative filtering algorithm can deal with large amounts of data efficiently, but the accuracy of the algorithm is still insufficient. In order to optimize the performance of collaborative filtering algorithm, based on the existing single element matrix decomposition algorithm and the graph correction matrix decomposition algorithm,a nonnegative matrix decomposition collaborative filtering algorithm based on single element strategy and graph correction is proposed using non negative matrix decomposition model. It can be proved by experiment that this algorithm can combine the advantages of single element and graph correction algorithm, and can also make up for the deficiency of single element and graph correcting algorithm. It has the most obvious effect in collaborative filtering recommendation system.

Key words. Nonnegative matrix, single element, graph correction, collaborative filtering algorithm.

1. Introduction

Collaborative filtering algorithm(Mao, Xiong, Jiao, Feng, & Yeung, 2017) is divided into collaborative filtering algorithm with memory function and collaborative filtering algorithm based on model building A collaborative filtering recommendation algorithm based on matrix decompositionis the most widely used algorithm in collaborative filtering algorithm. It uses special decomposition techniques to reduce the dimensionality of feature vectors and reduce the average error, thus improving the accuracy of data filtering and high scalability.

In the late twentieth Century, Goldberg(B. Yang, Lei, Liu, & Li, 2017) invented the collaborative filtering algorithm; In the early twenty-first Century, American researchers published a representative Movielens filtering recommendation system. Nadaq Corporation in the United States has commercialized collaborative filtering recommendation algorithms for the first time (Lim, Gray, Xie, & Poleksic, 2016)and

[1]Workshop 1 - Nanchong vocational and technical college, Nanchong 637000, China; e-mail: tangtianguo999@163.com

has been marketed as a product for users. Today, the most well-known product applications of collaborative filtering algorithm are Baidu, Google and other search engines, Taobao, Jingdong and other e-commerce, QQ, WeChat and other social software (Cheng, Yin, Dong, Dong, & Zhang, 2016). However, after many years' research results, the collaborative filtering recommendation algorithm still has many problems, such as sparse differences, slow cold boot and lack of graph model, which need further analysis and study.

Therefore, this paper studies the basic contents of the collaborative filtering algorithm based on matrix decomposition. Non negative matrix decomposition model is adopted, and nonnegative matrix decomposition collaborative filtering algorithm based on single element strategy and graph correction is proposed, so as to improve the accuracy of algorithm of data selection, and ensure higher application value.

2. Nonnegative matrix decomposition collaborative filtering algorithm based on single element strategy and graph correction

Considering the time complexity, data sparseness, the dimension and scale of the graph model in the algorithm, we designed the objective function of nonnegative matrix decomposition model based on single element and correction graph:

Among them, the formula (1)adds the nonnegative correction to the user characteristic matrix, and μ is the correction parameter of the characteristic matrix of the user, and λ is the correction parameter of the characteristic matrix in the data. In practical application, it is necessary to reasonably adjust the parameter values of μ and λ, and ensure that the algorithm is in the best convergent state.

The PQ matrix of the algorithm is calculated with partial derivatives:

$$\begin{cases} \frac{\partial \mathrm{L}}{\partial \mathrm{P}} = -\left(\mathrm{R}-\mathrm{PQ}\right)\mathrm{Q}^{\mathrm{T}} + \mu \mathrm{P} \\ \frac{\partial \mathrm{L}}{\partial \mathrm{Q}} = -\mathrm{P}^{\mathrm{T}}\left(\mathrm{R}-\mathrm{PQ}\right) + \lambda \mathrm{L}_{\mathrm{V}}\mathrm{Q} \end{cases} \quad 2$$

We obtain the equality conditions of$\mathrm{p}_{\mathrm{u,k}}$ and$\mathrm{q}_{\mathrm{k,i}}$:

$$\begin{cases} \left(-\left(\mathrm{R}-\mathrm{PQ}\right)\mathrm{Q}^{\mathrm{T}} + \mu \mathrm{P}\right)_{\mathrm{u,k}} \mathrm{P}_{\mathrm{u,k}} = 0 \\ \left(-\mathrm{P}^{\mathrm{T}}\left(\mathrm{R}-\mathrm{PQ}\right) + \lambda \mathrm{L}_{\mathrm{V}}\mathrm{Q}\right)_{\mathrm{k,i}} \mathrm{Q}_{\mathrm{k,i}} = 0 \end{cases} \quad 3$$

Break L into $\mathrm{L}_{\mathrm{V}}=\mathrm{L}_{\mathrm{V}}^{+}-\mathrm{L}_{\mathrm{V}}^{-}$,$\mathrm{M}_{\mathrm{ij}}^{+}=\frac{|\mathrm{M}_{\mathrm{ij}}|+\mathrm{M}_{\mathrm{ij}}}{2}$,$\mathrm{M}_{\mathrm{ij}}^{-}=\frac{|\mathrm{M}_{\mathrm{ij}}|-\mathrm{M}_{\mathrm{ij}}}{2}$. Then, the nonnegative matrix formula (4) based on the data graph product is:

$$\begin{cases} \mathrm{P} \leftarrow \mathrm{P}\frac{\mathrm{RQ}^{\mathrm{T}}}{\mu \mathrm{P}+(\mathrm{PQ})\mathrm{Q}^{\mathrm{T}}} \\ \mathrm{Q} \leftarrow \mathrm{Q}\frac{\mathrm{P}^{\mathrm{T}}\mathrm{R}+\lambda \mathrm{QL}_{\mathrm{V}}^{-}}{\lambda \mathrm{QL}_{\mathrm{V}}^{+}+\mathrm{P}^{\mathrm{T}}(\mathrm{PQ})} \end{cases} \quad 4$$

Define the$|\mathrm{I}_{\mathrm{u}}|and|\mathrm{U}_{\mathrm{i}}|$ respectively is the total number of data user u evaluated andthe total number of data i user accepted. The formula (3) is expanded by using a single element method:

$$\begin{cases} \mathrm{p}_{\mathrm{u,k}} \leftarrow \frac{\int_{\mathrm{i}\in \mathrm{I}_{\mathrm{u}}} \mathrm{q}_{\mathrm{k,i}}\mathrm{r}_{\mathrm{u,j}}}{\int_{\mathrm{i}\in \mathrm{I}_{\mathrm{u}}}\left(\mathrm{q}_{\mathrm{k,i}}\int_{\mathrm{k}=1}^{\mathrm{f}}\mathrm{p}_{\mathrm{u,k}}\mathrm{q}_{\mathrm{k,j}}\right)+\mu_{\mathrm{p}}^{\mathrm{s}}|\mathrm{I}_{\mathrm{u}}|\int_{\mathrm{i}\in \mathrm{I}_{\mathrm{u}}}\mathrm{p}_{\mathrm{u,k}}} \\ \mathrm{q}_{\mathrm{k,i}} \leftarrow \frac{\int_{\mathrm{i}\in \mathrm{I}_{\mathrm{u}}}\mathrm{r}_{\mathrm{u,i}}\mathrm{p}_{\mathrm{u,k}}+\lambda\int_{\mathrm{i}\in \mathrm{I}_{\mathrm{u}}}\int_{\mathrm{c}\in \mathrm{I}_{\mathrm{u}}}\mathrm{q}_{\mathrm{f,c}}\left(\mathrm{L}_{\mathrm{V}}^{-}\right)_{\mathrm{c,i}}}{\int_{\mathrm{u}\in \mathrm{U}_{\mathrm{i}}}\left(\mathrm{p}_{\mathrm{u,k}}\int_{\mathrm{k}=1}^{\mathrm{f}}\mathrm{p}_{\mathrm{u,k}}\mathrm{q}_{\mathrm{k,j}}\right)+\lambda\int_{\mathrm{i}\in \mathrm{I}_{\mathrm{u}}}\int_{\mathrm{c}\in \mathrm{I}_{\mathrm{u}}}\mathrm{q}_{\mathrm{f,c}}\left(\mathrm{L}_{\mathrm{V}}^{+}\right)_{\mathrm{c,i}}} \end{cases} \quad 5$$

So far, the implementation of the nonnegative matrix decomposition (RTGNMF) algorithm based on single element and graph correction is shown as follows:

RTGNMF algorithm

InputThe evaluation matrix R $\in R^{|U|\times|I|}$matrix dimension $f$parameter μ and λmatrix $L_U$and $L_V$initializedP $\in R^{|U|\times|f|}$Q $\in R^{|f|\times|I|}$non negative
OutputEstimated evaluation value*RMSENMAE*
Set iterative control variable st=0, set *training i*
wihlenot*converge*&& t<*training i*
Set U_Nu=0U_De=0I_Nu=0I_De=0
for Evaluation value of user u on data i$r_{u,j}$*in known U-I M*
$r'_{u,i} = \int_{k=1}^{f} p_{u,k}q_{k,i}$
fordimension $k\in$ (1,f)
$U\_Nu_{u,k} = U\_Nu_{u,k} + q_{k,i}r_{u,i}$ $U\_De_{u,k} = U\_De_{u,k} + r'_{k,i}q_{u,i}$
$I\_Nu_{k,i} = I\_Nu_{k,i} + p_{k,i}r_{u,i}I\_De_{k,i} = I\_De_{k,i} + p_{k,i}r'_{u,i}$
end for
end for
fordimension $k\in$ (1,f)
foruser $m\in$ (1,U_i)
$p^{+}_{u,k} = \int_{m\in U_I} (L^{+}_{U})_{u,m}p_{m,f}p^{-}_{u,k} = \int_{m\in U_I} (L^{-}_{U})_{u,m}p_{m,f}$
foruser $u\in$ (1,U_i)
$U\_Nu_{u,k} = U\_Nu_{u,k} + \lambda p^{-}_{u,k}U\_De_{u,k} = U\_De_{u,k} + \lambda p^{+}_{u,k}$
$U\_Nu_{u,k} = U\_Nu_{u,k} + \mu^{s}_{p}|I_u|p_{u,k}p_{u,k} = p_{u,k}\frac{U\_Nu_{u,k}}{U\_De_{u,k}}$
end for
end for
end for
fordimension $k\in$ (1,f)
for data $c\in$ (1,I_u)
$q^{+}_{k,i} = \int_{c\in I_u} q_{f,c}(L^{+}_{V})_{c,i}q^{-}_{k,i} = \int_{c\in I_u} q_{f,c}(L^{-}_{V})_{c,i}$
fordata $i\in$ (1,I_u)
$I\_Nu_{k,i} = I\_Nu_{k,i} + \lambda q^{-}_{k,i}I\_De_{k,i} = I\_De_{k,i} + \lambda q^{+}_{k,i}$
$I\_Nu_{k,i} = I\_Nu_{k,i} + \mu^{s}_{q}|U_i|q_{k,i}q_{k,i} = q_{k,i}\frac{I\_Nu_{k,i}}{I\_De_{k,i}}$
end for
end for
end for
t=t+1;
end while

3. Experimental results analysis

3.1. Experimental environment and experimental data

In this paper, the 5 layer cross validation method, which is widely used in collaborative filtering recommendation algorithm, is used to verify the RTGNMF algorithm. The hardware environment of the experiment is the notebook computer, the operating system is win10 with physical memory 4G, physical storage 1TB, clocked 2.3GHz, and the software environment are matlab software.

The main experimental data of this paper is:

(1)MovieLens: it is a collection of data used by the Grouplens team, this article needs to use about 100K of data in MovieLens,the data describes 200000 evaluations of 1680 television programs by 954 users, each user evaluates at least 40 TV programs, at the same time, the data set also includes the user's age, gender, occupation, hobbies and other features.

(2)M1-latest-small??it collects data from 2000 to 2016, including evaluation of the 8500 TV programs of 700 users, because the amount of data collection in the user is less than the number of television programs, therefore, the data set is usually applied to the RTGNMF algorithm analysis under user graphs.

(3)Filmtrust:Filmtrust data set was created in 2011, and it covers all kinds of evaluation information, as well as the degree of trust between users and users.

3.2. Parameter modification and experimental analysis

Parameter modification is the key step of RTGNMF algorithm, and the RMSE parameters (Lim, Poleksic, et, al., 2016) of the algorithm will be convergent when the parameters of the RTGNMF algorithm are reasonably modified.

Check Tikhonov to single non negative matrix decomposition with MovieLens data set, the RMSE value decreased. As shown in Figure 1, when the values of μ and λ are 0.07, the convergence of the RMSE curve is poor, the fluctuation of the RMSE curve is too frequent, and the minimum RMSE value is 0.98. After the parameters of μ and λ are modified, the convergence and descent of the RMSE curve are changed. For example, when μ and λ values are all 0.02, the minimum RMSE value is 0.93.

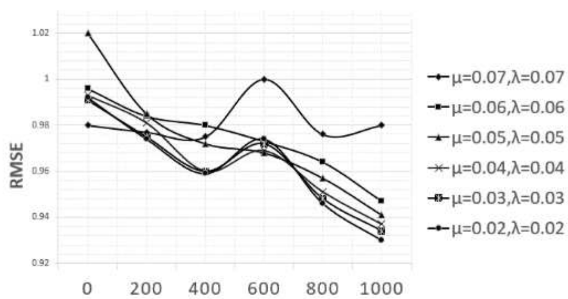

Fig. 1. RMSE curve under Data1 data set

Similarly, Draw the RMSE curve diagram under M1-latest-small and Filmtrust data set, summarizes the parameter modification effect on RTGNMF algorithm, determinereference valuesof μ and λ when RTGNMF algorithm is the best conditions.

Table 1 RTGNMF algorithm's final parameter values

Dataset	**RTGNMF algorithm**	
	μ	λ
MovieLens	0.03	0.02
M1-latest-small	0.08	0.08
Filmtrust	0.01	0.03

As shown in Figure 2, the NMF, SNMF, and RTGNMF algorithms are the RMSE values under the MovieLens, M1-latest-small, and Filmtrust data sets.

As can be seen in Figure 2, because the RTGNMF algorithm introduces the idea of graph rectification, overcome the data sparseness of single element algorithm, and play its own characteristics graph structure, its RMSE value is lower than the other two algorithms, and the recommended accuracy was superior to that of the NMF and SNMF algorithm.

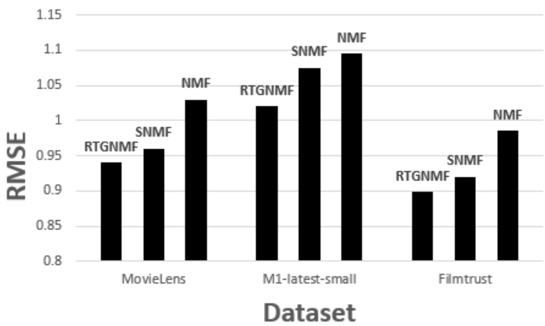

Fig. 2. RMSE values for different algorithms for the three data sets

4. Conclusion

In this paper, we study the collaborative filtering algorithm of nonnegative matrix decomposition model, and design a nonnegative matrix decomposition collaborative filtering algorithm based on single element and graph correction. This paper proposes a method based on nonnegative matrix decomposition algorithm and single element map correction, three sets of experimental data are used in the MATLAB experimental environment to analyze the influence of RTGNMF and RMSE parameters on the trend of the curve, to determine the parameter values of μ and λ when RTGNMF algorithm is in the best condition, and found that RTGNMF algorithm in precision performance index is better than that of other algorithms.

References

[1] Cheng. W,Yin. G, Dong. Y: *Collaborative Filtering Recommendation on Users' Interest Sequences.* PLoS One *11* (2016), No. 5.

[2] Lim. H, Gray. P, Xie. L: *Improved genome-scale multi-target virtual screening via a novel collaborative filtering approach to cold-start problem.* Sci Rep(2016).

[3] Lim. H, Poleksic. A, Yao. Y: *Large-Scale Off-Target Identification Using Fast and Accurate Dual Regularized One-Class Collaborative Filtering and Its Application to Drug Repurposing.* PLoS Comput Biol *12* (2016), No. 10.

[4] Mao. S, Xiong. L,Jiao. L,Feng. T: *A Novel Riemannian Metric Based on Riemannian Structure and Scaling Information for Fixed Low-Rank Matrix Completion.* IEEE Trans Cybern*47* (2012), No. 5, 1299–1213.

[5] Yang. B, Lei. Y, Liu. J: *Social Collaborative Filtering by Trust.* Applied Mathematical IEEE Trans Cybern *39* (2017), No. 8 1633–1647.

[6] B.Ẇu, H. Shen: *Analyzing and Predicting News Popularity on Twitter.* International Journal of Information Management Corresponding *36* (2015) 702–711.

[7] B. Wu,H. Shen: *Mining Connected Global and Local Dense SubGraphs for BigData.* International Journal of Modern Physics C*27* (2016).

[8] Wu. Bo, Kang. Chen,Haiying. Shen: *SPread: Exploiting fractal social community For efficient multi-coPy routing in VDTNs.*International Conference on. IEEE(2017).

Received November 16, 2016