

Recommendation Algorithm Optimization Based on Matrix Factorization

Liu Zhenzhen¹, Xu Dongping¹

1. Department of Computer Science and Technology, Wuhan University of Technology, Hubei
Wuhan, 430063 China

Email: zhenzhliu@sina.com

Keywords: Collaborative Filtering; Optimization; Matrix Factorization; Trust Relationship; Extension

Abstract. In the paper, the influence of user trust relationship on recommendation is mainly analyzed, user rating and trust relationship are also clearly defined. Meanwhile, the traditional SVD matrix factorization algorithm is extended to introduce the explicit influence (user trust value) and the implicit influence (friends trusted by users) of the user trust into the matrix factorization recommendation model. Subsequently, TFMF (Trust-fused Matrix Factorization) recommendation model is proposed in the paper to infer model training & learning in detail. The user trust relationship is not considered in implicit feedback SVD++, in other words, the item ratings marked by the trusted users can also influence the recommendation prediction. Therefore, the matrix factorization recommendation algorithm based on implicit trust relationship fusion is designed in the paper. Compared with original matrix factorization algorithm, the proposed algorithm not only has better prediction accuracy and flexibility, but also has better performance in the sparse data set due to the introduction of the trust relationship implicit feedback mechanism.

Introduction

The interpretability of the recommendation result is not only a practical application problem, but also an academic problem. Some scholars believed in early period that the interpretability of the recommendation result was crucial to user experience, because users expected to know relevant recommendation basis. Actually, the collaborative filtering has obvious advantage in this aspect. In the example for the item based collaborative filtering, when sending recommendation mails to the users, 360buy would tell the users about the recommendation reason that the users have purchased some commodities before. However, it is very difficult to explain the origin of the recommendation result to the users only through the matrix factorization, and most users more tend to accept the recommendation from the friend circle or the nearby friends rather than the recommendation from websites.

Problem Definition

The recommendation problem is defined in the paper as follows: user-item rating matrix and user trust matrix are analyzed to predict the ratings marked by the users for the unknown commodities.

A recommendation system is supposed to have m users and n commodities, and $R = [r_{u,i}]_{m \times n}$ is adopted to represent user-item rating matrix, where the rating marked by user u for commodity i can be expressed as $r_{u,i}$. For clear description, u and v are respectively defined as users, i and j are respectively defined as commodities in the paper. f -dimension implicit factor eigenvectors of user u and commodity i are respectively expressed as p_u and q_i . The core content of the matrix factorization can be described as follows: set $P^T Q \approx R$ for user eigenmatrix $P \in R^{f \times m}$ and commodity eigenmatrix $Q \in R^{f \times n}$ learned thereby, namely: make the product of the two eigenmatrixes infinitely approximate to the real rating matrix R . Therefore, the evaluation rating marked by user u for commodity i can be expressed as the inner product ($\hat{r}_{u,i} = q_i^T p_u$) of user eigenmatrix p_u and commodity eigenmatrix q_i .

According to the above description, the social network model of the users in the recommendation system supposed in the paper can be expressed as figure $\Psi = (v, \epsilon)$, where v represents m user

nodes and ε represents the direct trust relationship among the users. For ε mentioned above, matrix $T = [t_{u,v}]_{m \times m}$ can be adopted to describe the user trust relationship; the trust degree of user u to user v can be described as $t_{u,v}$; p_u and w_v are respectively defined as f -dimension implicit factor eigenvectors of users u and v . In order to associate the trust model and the rating model, the same eigenvector space is adopted for the user trust matrix and the corresponding rating matrix in the paper. Therefore, user trust matrix $P^{f \times m}$ and trust matrix $W^{f \times m}$ of the trusted user are provided in order to predict the user trust relationship through the inner product $\hat{t}_{u,v} = w_v^T p_u$ of the above two trust eigenvectors.

Trust matrix T is factorized as follows:

$$T \approx P^T W \quad (1)$$

After factor matrixes P and W of the user and the trusted user are obtained, the following formula is adopted to predict the trust value of user u to user v :

$$\hat{t}_{uv} = w_v^T p_u = \sum_{f=1}^F p_{uf} \cdot w_{vf} \quad (2)$$

Matrixes W and P can be learned through the loss function of the matrix factorization, as shown below:

$$Los(p, w) = \frac{1}{2} \sum_{(u,v) \in T_u} \left\{ (t_{uv} - w_v^T p_u)^2 + \lambda \left(\|p_u\|_F^2 + \|w_v\|_F^2 \right) \right\} \quad (3)$$

Where T_u represents the trusted user set of user u , λ is the preset penalty parameter for over-fitting prevention, and $\|\cdot\|$ is Frobenius norm.

Trust-fused Matrix Factorization

Presentation basis

SVD++ model in literature[3] was proposed and improved by Yehuda Koren of Yahoo Inc, wherein the core principle is as follows: not only the global user mean and the user-item rating deviation, but also the implicit influence of the item ratings marked by the users are considered on the basis of the traditional SVD algorithm, and relevant thought can be formally described as follows:

$$\hat{r}_{uj} = mean + b_u + b_j + q_j^T (p_u + |R(u)|^{-\frac{1}{2}} \sum_{i \in R(u)} y_i) \quad (4)$$

Similarly to relevant description in SVD, b_u and b_j respectively represent user-item rating deviation, mean represents global mean and R_u represents the set of the item ratings marked by user u . Differently, y_i is introduced to represent the implicit influence of the item rating marked by user u in the past on the commodity to be predicted. Therefore, the eigenvector of user u can be described as follows: $p_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} y_i$.

Actually, such information as user evaluation for the commodities is not well used in the traditional SVD model. For example, regardless of the ratings, a movie is always attractive to the users before they see this movie. Although SVD++ model becomes more complex after the introduction of the parameter ---- users' movie seeing history dimension, yet the model has better effect. Relevant experiment shows: compared with SVD and Asymmetric-SVD, SVD++ has lower RMSE value.

TFMF model

Implicit feedback SVD++ thought can be extended. If the data regarding the commodity checked by the user to be predicted exist in the data set, then such data can be also adopted for prediction. In consideration of the above data, user trust relationship, especially the influence of the implicit relationship, is fused into SVD++ model described in literature[3] in order to propose TFMF (Trust-fused Matrix Factorization) recommendation algorithm for the fusion of the trust dimension and the rating dimension. As a result, the explicit influence (user trust value) and the implicit influence (friends trusted by users) of the user trust are introduced into the matrix factorization recommendation model. Specifically, relevant thought is formally described as follows:

$$\hat{r}_{uj} = \mu + b_u + b_j + q_j^T (p_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} y_i + |T_u|^{-\frac{1}{2}} \sum_{v \in T_u} w_v) \quad (5)$$

In the above formula, w_v represents the implicit trust eigenvector of the user trusted by user u , T_u

represents the set of the users trusted by user u , and the quantity of users in the set is recorded as $|T_u|$ for representing the user trust scale and this parameter can be used to control the trust scale. In this way, $q_j^T w_v$ can be regarded as the influence of the rating marked by the user trusted by user u for commodity j on \hat{r}_{uj} , or the degree of the influence of the trusted user on the rating marked by user u for commodity j . Correspondingly, the user model of user u can be expressed as $p_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} y_i + |T_u|^{-\frac{1}{2}} \sum_{v \in T_u} w_v$.

As mentioned above, in order to associate the above two measurement modes, the user eigenvector in the paper is learned and factorized through user rating matrix R and user trust matrix T , wherein user eigenvectors in the user rating matrix and the trust matrix share the dimension space, namely: $R \approx P^T Q$ and $T \approx P^T W$ have the same P . In this way, user rating factor and user trust factor are integrated into a unified framework through matrix sharing. The minimum objective function can be formally described as follows:

$$Los = \frac{1}{2} \sum_{(u,j) \in I_u} (r_{uj} - \hat{r}_{uj})^2 + \frac{1}{2} \sum_{(u,v) \in T_u} (t_{uv} - \hat{t}_{uv})^2 \quad (6)$$

After the penalty factor is added to Formula (6), the following formula is adopted:

$$Los = \frac{1}{2} \sum_{(u,j) \in I_u} (r_{uj} - \hat{r}_{uj})^2 + \frac{1}{2} \sum_{(u,v) \in T_u} (t_{uv} - \hat{t}_{uv})^2 + \frac{\lambda}{2} (\sum_u b_u^2 + \sum_j b_j^2 + \sum_u |p_u|_F^2 + \sum_j |q_j|_F^2 + \sum_v |w_v|_F^2) \quad (7)$$

Experiment

Experiment method and technology are very important for data mining research and can directly determine the success or the failure of the research. The experiment method includes the following contents: how to divide training set and testing set, how to avoid accidental results and how to simulate different data sparseness.

“Cross-validation” is adopted in the paper for the experiment. According to the cross-validation principle, the data set is firstly divided into 10 subsets with the same size (each user evaluation is divided into ten equal parts). In each experiment, N ($0 < N < 10$) subsets are randomly selected from the above subsets to form the training set, and the rest subsets are taken as the testing set. The experiment is repeatedly carried out for ten times and the mean value of the ten experiment results is regarded as the final result.

In order to control the test variables, the data set is randomly divided into 10 parts in the experiment implemented for the paper, 2 parts are randomly selected as the testing set and the rest 8 parts are taken as the training set to establish the user interest model so as to predict user behaviors, wherein the corresponding evaluation indexes are statistically collected from the testing set. In order to ensure that the evaluation indexes are not over-fitting results, each experiment shall be repeated for five times in different testing sets, and the mean value of the evaluation indexes obtained from the five experiments shall be regarded as the final evaluation index.

The experiment aims at evaluating the proposed algorithm from the following aspects:

TFMF algorithm evaluation: TFMF recommendation algorithm proposed above is evaluated in Epinions data set. RMSE values and MAE values of TFMF (Trust-fused Matrix Factorization) algorithm, UserAvg (User Rating Average) algorithm, ItemAvg (Item Rating Average) algorithm, PMF (Probability Matrix Factorization) algorithm[2] and SVD++[3] algorithm are compared in the paper in order to evaluate the prediction performances of such matrix factorization recommendation algorithms as TFMF algorithm under different user trust scales.

Table 1 shows the comparison of RMSE values and MAE values of TFMF algorithm and UserAvg algorithm, ItemAvg algorithm, PMF algorithm, Basic MF algorithm, Biased MF algorithm, MF++ algorithm, etc., wherein f represents the number of the implicit factors in various MF algorithms, UserAvg and ItemAvg algorithms exclude this parameter, and the performances of various algorithms under other parameters are not particularly compared in the paper. Specifically, the algorithm parameters are set as follows: the iteration times of PMF, SVD++ and TFMF algorithms are uniformly set as $T = 20$, the learning rate of SVD++ algorithm is set as $\lambda = 0.1$,

the convergence coefficient is set as $\alpha = 0.05$, and the learning rate of TFMF algorithm is set as $\lambda = 0.8$.

Table1 The Performances of TFMF Algorithms

Number of Implicit Factors	Evaluation Index	UserAvg	ItemAvg	PMF	SVD++	TFMF
$f = 5$	RMSE	1.203	1.094	1.291	1.057	1.044
	MAE	0.930	0.928	0.980	0.818	0.805
$f = 10$	RMSE	1.203	1.094	1.198	1.058	1.045
	MAE	0.930	0.928	0.910	0.817	0.806

Conclusion

In the trust-fused matrix factorization model proposed in the paper, the explicit influence of the user trust is fused in the matrix factorization to effectively solve the cold start problem, and the implicit influence of the user trust is also fused therein to obtain higher recommendation accuracy. Meanwhile, relevant data set, evaluation indexes, experiment scheme, experiment results and analysis are also described in the paper. Specifically, the experiment is firstly carried out for TFMF recommendation algorithm to evaluate the rating prediction effect of TFMF algorithm in Epinions data set, thus to verify the recommendation accuracy of RFMF algorithm through the comparison with several common rating-based recommendation algorithms.

Reference

- [1] Wang K, Zhou X, Li T, et al. Optimizing load balancing and data-locality with data-aware scheduling[C]. Big Data (Big Data), 2014 IEEE International Conference on. IEEE, 2014: 119-128.
- [2] Zhang L, He B, Sun J, et al. Double Image Multi-Encryption Algorithm Based on Fractional Chaotic Time Series[J]. Journal of Computational and Theoretical Nanoscience, 2015, 12: 1-7.
- [3] Su T, Lv Z, Gao S, et al. 3d seabed: 3d modeling and visualization platform for the seabed[C]. Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on. IEEE, 2014: 1-6.
- [4] YishuangGeng, Jin Chen, Ruijun Fu, GuanqunBao, KavehPahlavan, Enlighten wearable physiological monitoring systems: On-body rf characteristics based human motion classification using a support vector machine, IEEE transactions on mobile computing, 1(1), 1-15, Apr. 2015