

An Improved Singular Value Decomposition Recommender Algorithm Based on User Trust Relationship

Junchi Pan^{1, a*}, Xingming Zhang^{1, b} and Xiaofeng Qi^{1, c}

¹National Digital Switching System Engineering & Technological R&D Center, Zhengzhou, China

^ajunchipan@163.com, ^bndsczxm@sina.com, ^ctxqixf@163.com

Keywords: recommender algorithm; collaborative filtering; trust; singular value decomposition

Abstract. Collaborative filtering is one of the most widely used recommender algorithms, whereas it is suffering the issues of data sparsity. Recommender algorithms based on trust perform better in alleviating data sparsity. However, there remain shortages in the process of mining trust relation in specific algorithms, which limit the improvement of prediction accuracy. To address this problem, the paper proposes an improved singular value decomposition algorithm, trying to integrate truster-specific and trustee-specific information and the implicit feedback of each when generating predictions. Experiments on the Epinions dataset show that the proposed algorithm performs better than state-of-the-art recommender algorithms in prediction accuracy.

Introduction

Recommender system has proved to be an effective way to solve the problem of information overload. Specially, collaborative filtering (CF) is a widely used recommendation technology. It generates recommendation for the target user by collecting the preference information of similar users and do not require rich descriptions about users and items. Generally, two types of CF techniques have been investigated: Memory-based and Model-based algorithms. Although CF techniques have been proven effective for recommendations, they still have some limitations on prediction accuracy, including data sparsity and cold-start problems [1].

With the advent of online social network, the data shared among people has provided massive information to describe the real world. Exploiting social networks information in recommender systems has become the new approach of overcoming problems in traditional CF technique [2]. Specially, researches on CF have introduced the trust relationship into recommender systems. The trust relationship is one of the most important types of social information which is based on the idea that people are often influenced by the opinion of the ones they trust in [3]. Incorporating trust into recommender system has proved to be an effective method to improve recommendation performance and to help mitigate the problem of data sparsity and cold-start [4].

However, most of the latest works on trust-based CF algorithms [5] have not taken more detailed trust information, i.e. the implicit influence of trust, into consideration. Consequently, there remains space for the improvement of prediction accuracy.

Aiming at the problems mentioned above, the paper proposes a novel trust-based recommendation model by considering truster-specific and trustee-specific information and the implicit feedback of them when generating predictions. Experimental results on the data set from Epinions.com demonstrate that our approach achieves better accuracy than state-of-the-art CF algorithms.

The rest of this paper is organized as follows: Some related works are discussed in the next section. Then we introduce the improved SVD++ algorithm. Evaluations are discussed in the next section. Finally, we give the conclusion and future work.

Related Work

Basically, two questions needs to be solved when studying trust-based CF algorithm. One is how to obtain trust data and how to evaluate the trust degree. Some researchers estimate the trust relationships between the users from their historical ratings [6,7], which essentially can be regarded as

similarity-based approach [8]. Some other researcher estimate trust from explicit trust statement [5,9,10].

Another question is how to incorporate trust into a specific algorithm. Methods can be classified into two categories. The first category incorporates trust with Memory-based algorithms. Some researchers replace similarity with trust [11]. Guo [12] calculates user similarity based on trust degree and using similarity to generated predictions. Li [6] combines similarity and trust into a unified algorithm using a weighting coefficient. However, simply combination of trust and similarity would not bring significant improvements in prediction accuracy because they have strong correlation. The second category incorporates trust with Memory-based algorithms. Ma [5] proposes the SocialMF matrix factorization algorithm, which involves the influence of trust by incorporating trust into user-specific vector. Yang [13] models truster feature and trustee feature respectively and proposes the TrustMF algorithm. Guo [4] incorporates the explicit and implicit influence of trusted users when generating predictions. However, these works do not mining the influence of trust relationship more deeply, which make it difficult to meet the needs of prediction accuracy in social network recommendation.

The Improved SVD Recommender Algorithm

Problem Definition. The paper focuses on the problem of predicting the rating of a user on an unrated item. The input data includes a user-item rating matrix and a user-user trust matrix. Suppose there are m users in the recommender system. That means there exists a trust network with m nodes. Each node represents a user and the directed edges between nodes represent the trust relationship. Let $T = [t_{u,v}]_{m \times m}$ denote the trust network, where each entry $t_{u,v}$ indicates the extent to which user u trusts user v .

Definition 1. truster-specific feature vector tr_u , describes the rating feature of user u and how user u trust other users in the system.

Definition 2. trustee-specific feature vector te_v , describes the rating feature of user v and how user v is trusted by other users in the system.

Hence, the truster feature matrix is represented by $Tr^{d \times m}$, and the trustee feature matrix is represented by $Te^{d \times m}$. The inner product of a truster-specific vector and trustee-specific vector $\overline{t_{u,v}} \approx tr_u^T te_v$ can be used to predict the trust relationship between u and v . The trust matrix can be recovered by $T \approx Tr^T Te$.

The TSVD++ Model. In this section, we propose an improved singular value decomposition model TSVD++ based on both truster-specific feature and trustee-specific feature. The TSVD++ model shown as Fig. 1 includes two sub-models: 1) TRSVD++ model that is build based on truster-specific feature; and 2) TESVD++ model that is build based on trustee-specific feature. We perform stochastic gradient descent [14] to learn the parameters in the two sub-models. Then the final parameters are obtained by liner fusion technique.

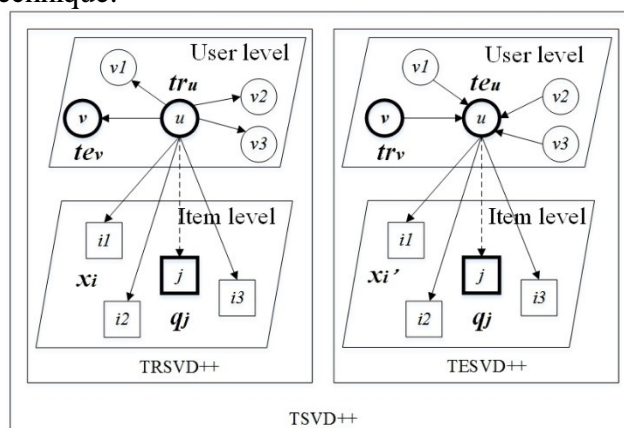


Fig. 1 TSVD++ model

1) The TRSVD++ Model. A state-of-the-art model known as SVD++ [15] takes user/item biases and the influence of user's rating behavior. Formally, the user feature is predicted by:

$$\mathbf{p}_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i. \quad (1)$$

Where \mathbf{p}_u represents how user u rates items, $|I_u|$ represents the number of items rated by u , $|I_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i$ denotes the implicit feedback of u 's rating behavior on the rating in the future.

Considering the fact that rating on some items and stating trust in other users are both kinds of opinion given by a user, we can incorporate the implicit feedback of user's trust behavior on item rating in the same way as SVD++ do, given by [4]:

$$\mathbf{p}_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i + |T_u|^{-\frac{1}{2}} \sum_{v \in T_u} \mathbf{y}_v. \quad (2)$$

where \mathbf{p}_u represents how user u rates items and how u trusts other people in the social network, $|T_u|$ denotes the number of users u trusted, $|T_u|^{-\frac{1}{2}} \sum_{v \in T_u} \mathbf{y}_v$ denotes the implicit feedback of u 's trust behavior on item rating.

For clarity, we rewrite symbol \mathbf{p}_u as \mathbf{tr}_u , and \mathbf{y}_v as \mathbf{te}_v . Thus, user u 's rating on item j is predicted as (3):

$$\overline{r_{u,j}} = \mathbf{b}_u + \mathbf{b}_j + \mathbf{m} + \mathbf{q}_j^T (\mathbf{tr}_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i + |T_u|^{-\frac{1}{2}} \sum_{v \in T_u} \mathbf{te}_v). \quad (3)$$

2) The TESVD++ Model. In social network, trust statements on other people reflects part of user's profile. Actually, the way a user be trusted by other people also reflects part of his profile. Detailedly, we take a simple trust network shown in Fig. 2 as an example. As described in the TRSVD++ model, u_1 trust u_a and u_b , then we can draw the conclusion that u_a and u_b has the ability to recommender items for u_1 . In fact, consider the user set who trust u_a and u_b respectively: $T_{u_a}' = \{u_1\}$ and $T_{u_b}' = \{u_1, u_2, u_3\}$, suppose u_1, u_2 and u_3 has the same credibility, then it can be inferred that u_b has more ability to recommender items because u_b is trusted by more people.

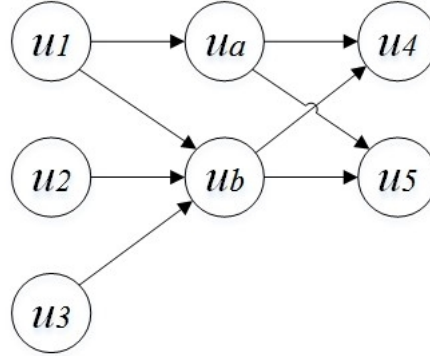


Fig. 2 A simple trust network

Taking the trustee-specific information and the implicit feedback of it into consideration, a novel user feature is predicted by:

$$\mathbf{te}_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i' + |T_u'|^{-\frac{1}{2}} \sum_{v \in T_u'} \mathbf{tr}_v. \quad (4)$$

where \mathbf{te}_u denotes how user u is trusted by others and how he rates items, $|I_u|^{-\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i'$ denotes the implicit feedback of u 's rating behavior on the rating in the future, $|T_u'|$ denotes the number of users who trust u , $|T_u'|^{-\frac{1}{2}} \sum_{v \in T_u'} \mathbf{tr}_v$ denotes the implicit feedback of how u is trusted by others. The prediction formula is given by:

$$\overline{r_{u,j}} = \mathbf{b}_u + \mathbf{b}_j + \mathbf{m} + \mathbf{q}_j^T (\mathbf{t}_u + |I_u|^{\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i' + |T_u'|^{\frac{1}{2}} \sum_{v \in T_u'} \mathbf{tr}_v). \quad (5)$$

3) Description of Algorithm. The steps of TSVD++ algorithm is described below:

Step1: Using the user-user trust matrix \mathbf{T} and the user-item rating matrix \mathbf{R} as the input. Learning the parameters in TRSVD++ model. The objective function is given as (6):

$$L = \frac{1}{2} \sum_u \sum_{j \in I_u} (\overline{r_{u,j}} - r_{u,j})^2 + \frac{I_t}{2} \sum_u \sum_{v \in T_u} (\overline{t_{u,v}} - t_{u,v})^2 + \frac{I}{2} (\sum_u \mathbf{b}_u^2 + \sum_j \mathbf{b}_j^2 + \sum_u \|\mathbf{tr}_u\|_F^2 + \sum_v \|\mathbf{te}_v\|_F^2 + \sum_j \|\mathbf{q}_j\|_F^2 + \sum_i \|\mathbf{x}_i\|_F^2) + \frac{I_t}{2} \sum_u \|\mathbf{tr}_u\|_F^2. \quad (6)$$

where $\overline{t_{u,v}} = \mathbf{te}_v^T \mathbf{tr}_u$, I and I_t are regularization parameters to avoid over-fitting. Stochastic gradient descent is applied to learn parameters \mathbf{b}_u , \mathbf{b}_j , \mathbf{q}_j , \mathbf{x}_i , \mathbf{tr}_u and \mathbf{te}_v .

Step2: Using the user-user trust matrix \mathbf{T} and the user-item rating matrix \mathbf{R} as the input. Learning the parameters in TESVD++ model in the same way as in Step1. The objective function is given as (7):

$$L = \frac{1}{2} \sum_u \sum_{j \in I_u} (\overline{r_{u,j}} - r_{u,j})^2 + \frac{I_t}{2} \sum_u \sum_{v \in T_u} (\overline{t_{u,v}} - t_{u,v})^2 + \frac{I}{2} (\sum_u \mathbf{b}_u'^2 + \sum_j \mathbf{b}_j'^2 + \sum_u \|\mathbf{te}_u\|_F^2 + \sum_v \|\mathbf{tr}_v\|_F^2 + \sum_j \|\mathbf{q}_j'\|_F^2 + \sum_i \|\mathbf{x}_i'\|_F^2) + \frac{I_t}{2} \sum_u \|\mathbf{te}_u\|_F^2. \quad (7)$$

where $\overline{t_{v,u}} = \mathbf{tr}_v^T \mathbf{te}_u$, and the parameters are \mathbf{b}_u' , \mathbf{b}_j' , \mathbf{q}_j' , \mathbf{x}_i' , \mathbf{te}_u and \mathbf{tr}_v .

Step3: Liner fusion of the parameters learned from TRSVD++ and TESVD++.

$$\mathbf{b}_u'' = \frac{\mathbf{b}_u + \mathbf{b}_u'}{2}. \quad (8)$$

$$\mathbf{b}_j'' = \frac{\mathbf{b}_j + \mathbf{b}_j'}{2}. \quad (9)$$

$$\mathbf{p}_u = \frac{\mathbf{tr}_u + \mathbf{te}_u}{2}. \quad (10)$$

$$\mathbf{q}_j'' = \frac{\mathbf{q}_j + \mathbf{q}_j'}{2}. \quad (11)$$

$$\mathbf{x}_i'' = \frac{\mathbf{x}_i + \mathbf{x}_i'}{2}. \quad (12)$$

Finally, the rating formula of TSVD++ algorithm is given as(13):

$$\overline{r_{u,j}} = \mathbf{b}_u'' + \mathbf{b}_j'' + \mathbf{m} + \mathbf{q}_j''^T (\mathbf{p}_u + |I_u|^{\frac{1}{2}} \sum_{i \in I_u} \mathbf{x}_i'' + |T_u|^{\frac{1}{2}} \sum_{v \in T_u} \mathbf{te}_v + |T_u'|^{\frac{1}{2}} \sum_{v \in T_u'} \mathbf{tr}_v). \quad (13)$$

Experiments and Results

Dataset. We use the dataset from Epinions.com [16] in the experiments. The data set contains 40163 users and 139738 items. The number of ratings is 664824. The number of trusters and trustees are 33960 and 49288, and the number of trust relationship is 487183. The ratings range from 1 to 5. The trust statements are binary, including trust (1) and distrust (-1).

Evaluation Metrics. We use Mean Absolute Error (MAE) and root mean square error (RMSE) to evaluate the performance of different recommendation algorithms. In general, lower MAE and RMSE mean better predictive accuracy.

$$MAE = \frac{1}{|S_{test}|} \sum_{(u,j) \in S_{test}} |r_{u,j} - \overline{r_{u,j}}|. \quad (14)$$

$$RMSE = \sqrt{\frac{1}{|S_{test}|} \sum_{(u,j) \in S_{test}} |r_{u,j} - \overline{r_{u,j}}|^2}. \quad (15)$$

Where S_{test} is the testing set and $|S_{test}|$ is the number of user-item pairs.

Comparison Methods. To evaluate the performance of TSVD++ method, 3 categories of comparison CF methods are selected:

Memory-based baselines: UserMean and ItemMean, mean rating of users and items respectively.

Model-based CF: SVD++ [15], PMF [17].

Trust-based CF: TrustSVD [4], SocialMF [5], TrustMF [13].

The optimal experimental settings for comparison methods are suggested by previous work, as shown in Table 1.

Table 1. Parameter Setting for Comparison Methods

Methods	Optical Parameters
PMF	$I_u=I_v=0.001$
SVD++	$I=0.35$
SocialMF	$I=0.001, I_r=1$
TrustMF	$I=0.001, I_r=1$
TrustSVD	$I=0.9, I_r=0.5$

In the evaluations, 5-folder crossing validation is used. In each time, we select 80% of data as training set and the rest of 20% for test. The results are the arithmetical average of the 5 tests.

Validation on Prediction Accuracy. In the experiments, we set the number of latent features $d=5$ and $d=10$ [4, 10, 13], learning rate=0.001 [4]. To prevent infinite iteration, we set the experimental maximum number of iterations to be 100. The optimal parameter settings for TSVD++ are $I=0.08$, $I_r=0.6$.

The experimental results are shown in Table 2. TrustSVD performs the best for all comparison method. These results are consistent with Guo’s conclusion [4]. On the contrary, our approach TSVD++ outperforms the comparison methods. Specifically, 1.10% improvements in MAE when $d=5$, and 0.82% when $d=10$. 1.01% improvements in RMSE when $d=5$, and 0.69% when $d=10$. Some explanations can be made: In the comparison methods, the baseline and model based algorithms do not make use of the trust information. SocialMF exploits the explicit trust statements but fails to consider the implicit feedback of them. TrustMF incorporates both truster-specific and trustee-specific information but without their implicit feedback. TrustSVD exploits truster-specific information and their implicit feedback but fails to consider trustee-specific feature. However, our approach integrates truster-specific and trustee-specific information and the implicit feedback of each when generating predictions. In a word, our approach takes advantage of the mentioned comparison methods, and performs better in predictive accuracy.

Table 2. Performance comparison in Epinions dataset

	Metrics	UserMean	ItemMean	PMF	SVD++	SocialMF	TrustMF	TrustSVD	TSVD++	Improvement
d=5	MAE	0.936	0.933	0.986	0.824	0.832	0.823	0.810	0.802	1.10%
	RMSE	1.300	1.099	1.297	1.063	1.076	1.076	1.050	1.041	0.82%
d=10	MAE	0.936	0.933	1.015	0.824	0.833	0.826	0.811	0.803	1.01%
	RMSE	1.300	1.099	1.205	1.063	1.090	1.102	1.051	1.044	0.69%

Conclusion and Future Work

This paper proposed an improved SVD-based recommendation algorithm TSVD++ that incorporates truster-specific and trustee-specific information and the implicit feedback of them in the process of rating prediction. The experimental results in real world dataset prove that TSVD++ outperforms conventional CF methods in predictive accuracy. In the future, we plan to improve the model by adapting the weighted regularization technique. In the meantime, we will also look into better method in parameter fusion to reduce the complexity of the algorithm.

Acknowledgment

The work is funded by The National High Technology Research and Development Program of China (Grant No. 2014AA01A704) and The Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 61521003)

References

- [1] Adomavicius, Gediminas, and Alexander Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no.6, pp. 734–749, 2005.
- [2] Bao, Hui Fang Yang, and Jie Zhang, "Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation," *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)[C]*. Quebec, pp. 30–36, 2014.
- [3] Sinha, Rashmi R., and Kirsten Swearingen, "Comparing recommendations made by online systems and friends," In *Proceedings of the Joint DELOS-NSF Workshop on Personalization and Recommender systems in Digital Libraries*, Dublin, Ireland, pp. 64–69, 2001.
- [4] Guo, Guibing, Jie Zhang, and Neil Yorke-Smith, "TrustSVD: collaborative filtering with both the explicit and implicit influence of user trust and of item ratings," *Proceedings of the 29th AAAI conference on artificial intelligence(AAAI)[C]*. Austin, 2015.
- [5] Ma, H., Zhou, D., Liu, C., Lyu, M. R., and King, I, "Recommender systems with social regularization," *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM)[C]*. Hong Kong, pp. 287–296, 2011.
- [6] Li, Jiyun, Caiqi Sun, and Juntao Lv, "TCMF: trust-based context-aware matrix factorization for collaborative filtering," *IEEE 26th International Conference on Tools with Artificial Intelligence(ICTAI)[C]*. Limassol, pp. 815-821, 2014.
- [7] Jia, Dongyan, Fuzhi Zhang, and Sai Liu, "A robust collaborative filtering recommendation algorithm based on multidimensional trust model," *Journal of Software*, vol.8, no.1, pp. 11-18, 2013.
- [8] ZHANG G F, "Research on the key issue of trust-based personalized recommender systems in E-commerce," Nanchang: Jiangxi university of finance & economics, 2009.
- [9] Golbeck, Jennifer Ann, "Computing and applying trust in web-based social networks," Ph.D. dissertation, University of Maryland, CollegePark, 2005.
- [10] Jamali, Mohsen, and Martin Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," *Proceedings of the 4th ACM Conference on Recommender Systems (RecSys)[C]*. Barcelona, pp. 135–142, 2010.
- [11] Chen, S., Luo, T., Liu, W., and Xu, Y, "Collaborative filtering with fine-grained trust metric," *Proceedings of CIDM'09*, Nashville, TN, USA, Mar. 2009.
- [12] Guo, Guibing, Jie Zhang, and Daniel Thalmann, "Merging trust in collaborative filtering to alleviate data sparsity and cold start," *Knowledge-Based Systems*, pp. 57-68, 2014.
- [13] YANG B, Lei Y, LIU D. "Social collaborative filtering by trust," *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)[C]*. BeiJing, pp. 2747–2753, 2013.
- [14] Yang, B., Lei, Y., Liu, D., and Liu, J, "An improved singular value decomposition recommender algorithm based on local structures," *Journal of Electronics & Information Technology*, vol.35, no.6, pp. 1284-1289, 2013.
- [15] Koren, Yehuda, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 426–434, 2008.
- [16] P.Massa, "The extended epinions data set,"[Online]. Available: <http://www.trustlet.org/datasets/>
- [17] Mnih, Andriy, and Ruslan Salakhutdinov, "Probabilistic matrix factorization," *Advances in neural information processing systems*, pp. 1257-1264, 2007.