International Journal of Advanced Scientific Technologies ,Engineering and Management Sciences (IJASTEMS-ISSN: 2454-356X) Volume.3,Special Issue.1,March.2017

Enhancing Collaborative Filtering Performance UsingTrustSVD

Z.Dilip Kumar, PG Scholor, SVCET, Chittoor.

Abstract-- Recommender system provide users with personalized suggestions for product or services which is a composition of software and machine learning techniques. Collaborative filtering is one of the most popular technique to implement a recommender system. Several approaches have been introduced to Collaborative filtering, yet CF suffers from two well-known issues: data sparsity and cold start, which degrades the recommendation performance. An analysis of social trust data on real-world data sets suggests that not only the explicit but also the implicit influence of both ratings and trust should be taken into consideration in a recommendation model. Trust SVD builds on top of state-of-the-art recommendation algorithm, SVD++ by further incorporating both the explicit and implicit influence of trusted and trusting users on the prediction of items for an active user. The TrustSVD, a trust-based matrix factorization technique for recommendations encompasses of an empirical trust analysis and observe that trust and ratings complement to each other, and incorporating both influence of rating and trust information and finally demonstrates that TrustSVD achieves better accuracy than other counterparts of recommendation techniques.

Index Terms—Recommender System, Collaborative Filtering, Matrix Factorization, Social Trust, Implicit influenc

1. INTRODUCTION

Recommender systems have been widely used to provide users with high-quality personalized recommendations from a large volume of choices. Robust and accurate recommendations are important in e-commerce operations and in marketing. Collaborative Filtering (CF) is one of the most popular techniques to implement a recommender system [1]. The idea of CF is that users with similar preferences in the past are likely to favour the same items (e.g., movies, music, books, etc.) in the future. CF has also been applied to tasks besides item recommendations, in domains such as image processing [2] and bioinformatics [3]. However, CF suffers from two well-known issues: data sparsity and cold start [4]. The former issue refers to the fact that users usually rate only a small portion of items, while the latter indicates that new users only give a few ratings (a.k.a. cold-start users). Both issues severely degrade the efficiency of a recommender system in modelling user preferences and thus the accuracy of predicting a user's rating for an unknown item.

To help resolve these issues, many researchers [5], [6], [7], [8], [9] attempt to incorporate social trust information into their recommendation models, given that model-based CF approaches outperform memory-based approaches [10]. These approaches further regularize the user-specific feature vectors by the phenomenon that friend often influence each other in recommending items. However, even the best performance reported by the latest work [9] can be inferior to that of other state-of-the-art models which are merely based on user-item ratings. One possible explanation is that these trust-based models focus too much on the utility of user trust but ignore the influence the item ratings themselves. To investigate this phenomenon, we conduct an empirical trust analysis based on four real-world data sets (Film Trust, Epinions, Flixster and Ciao).

1.2 DATA SETS

The four data sets used in our analysis and also our later experiments are: Epinions, FilmTrust, Flixster and Ciao. These four data sets are among the few publicly available data sets that contain both item ratings and social relationships specified by active users. They are used widely in the evaluation of previous trust-aware recommender systems. In particular, the items in Epinions and Ciao are of great variety, such as electronics, sports, computers, etc., while the items in FilmTrust and Flixster are movies only. The ratings in Epinions and Ciao are integers from 1 to 5, while those in the other data sets are real values, i.e., [0.5, 4.0] for FilmTrust, [0.5, 5.0] for Flixster, both with step 0.5. Users in these data sets can share their item ratings with each other and pro-actively connect with users of similar

taste, whereby a social network can be constructed. Statistics of the data sets are illustrated in Table 1.
By definition, the social relationships in Epinions and Ciao are trust relationships whereas those in Flixster and FilmTrust are trust-alike relationships. To explain, users in Epinions and Ciao specify others as trustworthy usually based on the evaluation of quality of others' ratings and textual reviews. Flixster adopts the concept of friendship per se where user relations are symmetric and related with movies only. Although FilmTrust adopts the concept of trust (with original values from 1 to 10), the publicly available data set contains only binary values. Such degrading may cause much noise and thus we classify the relationships as trust alike rather than trust.

TABLE 1
Statistics of the Four Data Sets

Feature	Epinions	FilmTrust	Flixster	Ciao
# users	40,163	1,508	53,213	7,375
# items	139,738	2,071	18,197	99,746
# ratings	664,824	35,497	409,803	280,391
density	0.051%	1.14%	0.04%	0.03%
# trusters	33,960	609	47,029	6,792
# trustees	49,288	732	47,029	7,297
# trusts	487,183	1,853	655,054	111,781
density	0.029%	0.42%	0.03%	0.23%

1.3 OBSERVATIONS

1. Trust information is very sparse, yet is complementary to rating information.
2. A user's ratings have a weakly positive correlation with the average of her out-going social neighbours under the concept of trust-alike relationships, and a strongly positive correlation under the concept of trust relationships.
3. A user's ratings have a weakly positive correlation with the average of her in-coming social neighbours under the concept of trust-alike relationships, and a strongly positive correlation under the concept of trust relationships.

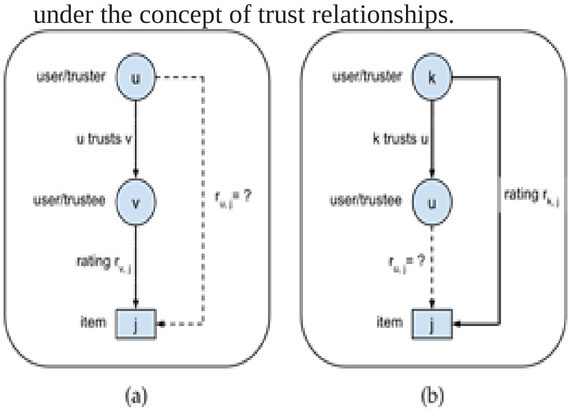

Fig. 1. The influence of (a) trustees v and (b) trusters k on the rating prediction for the active user u and target item j.

II.TRUSTSVD: A TRUST-BASED RECOMMENDATION MODEL

2.1 DEFINING PROBLEM

In social rating networks, a user can label (add) other users as trusted friend and thus form a social network. Trust is not symmetric; for example, users u_1 trusts u_3 but u_3 does not specify user u_1 as trustworthy. Besides, users can rate a set of items using a number of rating values, e.g., integers from 1 to 5. These items could be products, movies, music, etc. of interest. The recommendation problem in this work is to predict the rating that a user will give to an unknown item, for example, the value that user u_3 will give to item i_3, based on both a user–item rating matrix and a user–user trust matrix. Other well-recognized recommendation problems include for example top-N item recommendation.

Suppose that a recommender system includes m users and n items. Let $R=[r_{u,i}]_{m*n}$denote the user–item rating matrix, where each entry $r_{u,i}$represents the rating given by user u on item i. For clarity, we preserve symbols u,vfor users, and i,jfor items. Since a user only rated a small portion of items, the rating matrix R is only partially observed and oftentimes very sparse. Let $I_u= \{i|r_{u,i} \neq 0\}$denote the set of items rated by user u. Let $p_u$and q_i be a d-dimensional latent feature vector of user u and item i, respectively. The essence of matrix factorization is to find two low-rank matrices: user-feature matrix $P \in R^{d*m}$ and item-feature matrix $Q \in R^{d*n}$ that can adequately recover the rating matrix R, i.e., $R = P^TQ$, where P^T is the transpose of matrix P. The underlying assumption is that both users and items can be characterized by a small number of features. Hence, the rating on item j for user u can be predicted by the inner product of user-specific vector $p_u$and item-specific vector q_j, i.e., $r^{\wedge}_{u,j} = q_j^TP_u$. In this regard, the main task of recommendations is to predict the rating $r^{\wedge}_{u,j}$as close as possible to the ground truth $r_{u,j}$. Formally, we can learn the user- and item-feature matrices by minimizing the following loss (objective) function:

$$L_r = 1/2\Sigma_u\Sigma_{j\in Iu} (q^T{}_jp_u - r_{u,j})^2 + \lambda/2(\Sigma_u ||P_u||^2{}_F + \Sigma_j ||q_j||^2{}_F)$$

Where ||, ||$_F$ denotes the Frobenius norm, and λ is a parameter to control model complexity and to avoid over-fitting.

1.4 The TrustSVD Model

In line with the three observations of the previous section, our TrustSVD model is built on top of a state-of-the-art model known as SVD++ proposed by koren [11]. The rationale behind SVD++ is to take into consideration user/item biases and the influence of rated items other than user/item-specific vectors on rating prediction. Formally, the rating for user u on item j is predicted by:

$$\hat{r}_{u,j} = b_u + b_j + \mu + q^T_j (p_u + |I_u|^{-1/2}\Sigma_{i\in} I_u y_i)$$

where b_u, b_j represent the rating bias of user u and item j, respectively; μ is the global average rating; and $y_i$denotes the implicit influence of items rated by user u in the past on the ratings of unknown items in the future. Thus, user u's feature vector can be also represented by the set of items she rated, and finally modelled as $(p_u + | I_u|^{-1/2}\Sigma_{i\in Iu}y_i)$ rather than simply as p_u. Koren [11] has shown that integrating implicit influence of ratings can well improve predictive accuracy. We have already stressed the importance of trust influence for better recommendations, and its potential to be generalized to trust-alike relationships. Hence, we can enhance the trust-unaware SVD++ model by incorporating both the explicit and implicit influence of trust, described as follows.

Implicit influence of trusted users:

Fig 1.a shows that the trusted users of an active user have an effect on rating prediction for a certain item. We take into account this effect by modelling user preference in the same manner as rated items, given by:

$$\hat{r}_{u,j} = b_{u,j} + p_j^T(p_u + | I_u |^{-1/2}\Sigma_{i\in Iu}y_i +| T^+_u |^{-1/2}\Sigma_{v\in T^+_u}w_v),$$

Where $b_{u,j} = b_u + b_j + \mu$ hereafter represents bias terms, $w_v$is the use-specific latent feature vector of users (trustees) trusted by user u, and thus $q^T_jw_v$ can be explained by the trusted users, i.e., the influence of trustees on the rating prediction. In other words, the inner product $q^T_jw_v$ indicates how trusted users influence user u's rating on item j. An intuitive understanding has been illustrated in Fig. 1a. Similar to ratings, a user's feature vector can be interpreted by the set of users whom she trusts, i.e., $| T^+_u |^{-1/2}\Sigma_{v\in T^+_u}w_v$. Therefore, a user u is further modelled by $(p_u + | I_u |^{-1/2}\Sigma_{i\in Iu}y_i + | T^+_u |^{-1/2}\Sigma_{v\in T^+_u}w_v)$ in the social rating networks, considering the influence of both rated items, and trusted users.

Implicit influence of trusting users.

Fib. 1b shows that the trusting users of an active user can also influence the rating prediction for a certain item. In fact, *observation* 3 has indicated that such influence may be comparable with that of trusted users. Similarly, the effect can be considered by modelling user preference, given by:

$$\hat{r}_{u,j} = b_{u,j} + p_j^T(p_u + | I_u |^{-1/2}\Sigma_{i\in Iu}y_i + | T^-_u |^{-1/2}\Sigma_{k\in T^-_u}p_k),$$

Where T^-_u is the set of users who trust user u, i.e., the set of her trusters. Thus, $q^T_jp_k$ can be explained by the ratings predicted by the trusting users, i.e., the influence of trusters on the predictions. Similarly, the inner product $q^T_jp_k$ indicates how trusting users k influence user u's rating on item j. An intuitive understanding has been illustrated in Fig. 1b. Similar to ratings, a user's feature vector can be interpreted by the set of users whom trust her, i.e., $| T^-_u |^{-1/2}\Sigma_{k\in T^-_u} p_k$. Therefore, a user u is further modelled by $(p_u + | I_u |^{-1/2}\Sigma_{i\in Iu}y_i + | T^-_u |^{-1/2}\Sigma_{k\in T^-_u}p_k)$ in the social rating networks, considering the influence of both rated items and trusting users.

Combinational implicit trust influence. The implicit influence of trust neighbours on rating prediction therefor consists of two parts: the influence of both trustees and trusters. To consider both cases, we proposing following approaches.

(1)Linear combination: A natural and straight forward way is to linearly combine the two kinds of implicit trust influence, given by:

$$\hat{r}_{u,j} = b_{u,j} + q^T_j (p_u+ | I_u |^{-1/2}\Sigma_{i\in Iu}y_i + \alpha|T^+_u|^{-1/2}\Sigma_{v\in T^+_u}w_v + (1-\alpha)|T^-_u|^{-1/2}\Sigma_{k\in T^-_u}p_k).$$

Where α ∈ [0, 1] controls the importance of influence of trustees in rating prediction. Specifically, α = 0 means that we only consider the influence of trusting users; α = 1 indicates that only the influence of trusting users are considered; and α ∈ (0, 1) mixes the two kinds of trust influence together. In the case of undirected social relationships (e.g., friendship in Flixster), T^+_u will be equivalent with T^-_u, and thus the linear combination ensures that our model can be applied to both trust and trust-alike relationships.

(2)All as trusting users: In a trust relationship, a user u can be represented either by p_u as trustor or by w_u as trustee. An alternative way is to model the influence of user u's trust

neighbours, including both trusted and trusting users, in the manner of trusting users such that we can yield the following function:

$$\hat{r}_{u,j}= b_{u,j} + p_j^T(p_u + | I_u |^{-1/2}\Sigma_{i\in I_u} y_i + | T_u |^{-1/2}\Sigma_{k\in T_u} p_k)$$

where $T_u = T^+_u \cup T^-_u$ denotes the set of user u's trust neighbours. The underlying assumption is not to distinguish the roles of trust neighbours, but to treat them uniformly in terms of implicit trust influence.

(3)All as trusted users: With the same assumption, we may model the influence of all trust neighbours in the manner of trusted users. That is, we predict the user's possible rating on a target item by:

$$\hat{r}_{u,j} = b_{u,j} + q^T_j (p_u + |I_u|^{-1/2}\Sigma_{i\in I_u} y_i + |T_u|^{-1/2}\Sigma_{v\in T_u} w_v),$$

Where T_u is the set of user u's trust neighbours. Both the trusting equations will also influence the decomposition of trust relationships. However , since user-feature matrix P plays a key role in bridging both rating and trust information, the rating prediction by equation(2) may lead to better performance than that by equation(3).

With the consideration of implicit trust influence, the objective function to minimize is then given as follows:

$$L = 1/2\Sigma_u\Sigma_{j\in I_u} (\hat{r}_{u,j} - r_{u,j})^2 + \lambda/2 (\Sigma_u b^2_u + \Sigma_j b_j^2 + \Sigma_u||p_u||^2_F + \Sigma_j||q_j||^2_F + \Sigma_i||y_i||^2_F + \Sigma_v||w_v||^2_F) .$$

Where $\hat{r}$u,j is the prediction computed by (1). To reduce the model complexity, we use the same regularization parameter λ for all the variables. Finer control and tuning can be achieved by assigning separate regularization parameter variables, though it may result in greater complexity in model learning, and in comparison with other matrix factorization models.

CONCLUSION:

The proposed concept of trust-based matrix factorization model incorporates both rating and trust information. TrustSVD, takes into account both the explicit and implicit influence of ratings and of trust information when predicting ratings of unknown items. Both the trust influence of trustees and trusters of active users are involved in this model. Comprehensive experimental results shows that this approach TrustSVD outperforms both trust and ratings-based methods in predictive accuracy across different testing views and across users with different trust degrees. I conclude that this approach can better alleviate the data sparsity and cold start problems of recommender systems.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Trans. Know. Data Eng., vol. 17, no. 6, pp. 734–749, Jun. 2005.

[2] Y. Huang, X. Chen, J. Zhang, D. Zeng, D. Zhang, and X. Ding, "Single-trial ERPs denoising via collaborative filtering on ERPs images," Neurocomputing, vol. 149, pp. 914–923, 2015.

[3] X. Luo, Z. Ming, Z. You, S. Li, Y. Xia, and H. Leung, "Improving network topology-based protein interactome mapping via collaborative filtering," Knowl.-Based Syst., vol. 90, pp. 23–32, 2015.

[4] G. Guo, J. Zhang, and D. Thalmann, "A simple but effective method to incorporate trusted neighbors in recommender systems," in Proc. 20th Int. Conf. User Model., Adaptation Personalization, 2012, pp. 114–125.

[5] H. Ma, H. Yang, M. Lyu, and I. King, "SoRec: Social recommendation using probabilistic matrix factorization," in Proc. 31st Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2008, pp. 931–940.

[6] H. Ma, D. Zhou, C. Liu, M. Lyu, and I. King, "Recommender systems with social regularization," in Proc. 4th ACM Int. Conf. Web Search Data Mining, 2011, pp. 287–296.

[7] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in Proc. 4th ACM Conf. Recommender Syst., 2010, pp. 135–142.

[8] B. Yang, Y. Lei, D. Liu, and J. Liu, "Social collaborative filtering by trust," in Proc. 23rd Int. Joint Conf. Artif. Intell, 2013, pp. 2747–2753.

[9] H. Fang, Y. Bao, and J. Zhang, "Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation," in Proc. 28th AAAI Conf. Artif. Intell., 2014, pp. 30–36.

[10] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30– 37, Aug. 2009.

[11] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in Proc. 14th ACM SIGKDD Int. Conf. Know. Discovery Data Mining, 2008, pp. 426–434.