Data Imputation Using a Trust Network for Recommendation

Won-Seok Hwang, Shaoyu Li, Sang-Wook Kim*
Dept. of Electronics and Computer Engineering
Hanyang University, Korea
hws23, chris, wook@hanyang.ac.kr

Kichun Lee
Dept. of Industrial Engineering
Hanyang University, Korea
skylee@hanyang.ac.kr

ABSTRACT

Recommendation methods suffer from the data sparsity and cold-start user problems, often resulting in low accuracy. To address these problems, we propose a novel imputation method, which effectively densifies a rating matrix by filling unevaluated ratings with probable values. In our method, we use a trust network to estimate the unevaluated ratings accurately. We conduct experiments on the Epinions dataset and demonstrate that our method helps provide better recommendation accuracy than previous methods, especially for cold-start users.

Categories and Subject Descriptors

H.2.8.d [**Information Technology and Systems**]: Database ApplicationsData Mining

Keywords

Recommendation System, Trust Network, Data Imputation, Matrix Factorization

1. INTRODUCTION

A number of recommendation methods have been proposed for predicting ratings of items to which a target user has not evaluated yet [1]. Among them, the probabilistic matrix factorization (PMF) model is known to be quite effective [6][3]. However, these methods suffer from low accuracy when most of users rated only a few items, i.e., the rating matrix is very sparse, called the *data sparsity problem* [1]. In particular, it has been an important issue to accurately predict the ratings for *cold-start users* who rated only a small number of items [1].

In order to address the data sparsity and cold-start user problems, existing *imputation methods* [4] replace unevaluated ratings (missing values) with *probable values.* However, most of those methods overlook additional information such as a social network. Also, they cannot be applied to the PMF models [6] because they are designed only for their specific methods.

In this paper, we propose a novel imputation method that (1) exploits a *trust network* as an additional information and (2) can be applied to most of existing recommendation methods including the PMF models. Our method estimates the unevaluated rating of an item for a user by aggregating the corresponding ratings given by his/her *reliable neighbors.* In order to find the reliable neighbors, the method refers to a trust network, a kind of a social network, representing trust relationships among people, because users who are connected in the trust network tend to have similar preferences [2][5]. In addition, the proposed method fills with the probable values only for the items evaluated by a sufficient number of reliable neighbors. Owing to the proposed method, we can get the *desified* rating matrix, thereby achieving higher accuracy by applying most of recommendation methods including the PMF models.

*Corresponding author

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577363.

2. PROPOSED METHOD

This paper addresses data imputation for recommendation, which densifies a *rating matrix* whose element indicates a rating of an item given by a user. The proposed method fills the unevaluated rating of an item for a user with the most probable value estimated by aggregating the ratings given by *reliable neighbors* who have similar tastes to him/her.

We find the reliable neighbors of each user by examining the *trust network.* In a trust network, a trust relationship represents a user (trustor) trusts another user (trustee). In our method, the reliable neighbors of a user represent his/her *trustees* and *trustors* in the trust network. The rationale behind considering the trustees is based on the *assumption* that a trustor would be interested in those items which his/her trustees are interested in, in short, that *a trustor and trustee would have similar interests* [2][5]. If this assumption holds, the reasoning in the reverse direction is also possible. That is, a trustee would have interests in those items which his/her trustors are interested in. This is the reason why we consider the trustors as well as the trustees as reliable neighbors.

In our method, the accuracy of the estimation is dependent on the number of reliable neighbors who provide their ratings. Therefore, the estimation of probable values by only a few neighbors could be imprecise. In order to solve this problem, we skip estimating those unevaluated ratings for the items which are rated by an *insufficient* number of reliable neighbors for user u during the imputation process, and fill with probable values only for the rest of items, called the *candidate item set* of user u, $C_\theta(u)$. To build $C_\theta(u)$, we sort the items for u based on the number of his/her neighbors who evaluated each item in descending order, and then, select only the top θ percent of items.

The proposed imputation method estimates the unevaluated rating $r'_{u,i}$ of item i for user u as follows:

$$r'_{u,i\in\{i|i\in C_\theta(u)\}} = \bar{u} + \frac{1}{k}\Sigma_{v\in\{v|r_{v,i}\neq null, v\in N(u)\}}(r_{v,i} - \bar{v}) \quad (1)$$

where $r_{v,i}$ denotes the rating of user v on item i, and $\bar{u}$ ($\bar{v}$) represents the average rating of user u (v). $N(u)$ denotes a set of reliable neighbors of user u. Also, $r_{v,i} \neq null$ means user v evaluated item i. As we can see in Equation 1, the proposed method considers not all of the users in $N(u)$ but only those users who rated item i in $N(u)$. The denominator k indicates the number of users v in the set $\{v \mid r_{v,i} \neq null, v \in N(u)\}$.

With the proposed imputation method, we can get the densified rating matrix $R = (r_{u,i})_{M\times N} + (r'_{u,i})_{M\times N}$, where M and N denote the numbers of users and items, respectively, and $r_{u,i}$ and $r'_{u,i}$ indicate the original and estimated ratings of user u on item i, respectively. This densified matrix has exactly the same form as the original rating matrix, so any recommendation methods including the PMF models [6] can use the densified matrix for improving their accuracy.

3. EVALUATION

We used a real-world dataset, Epinions [5] for our experiments. The dataset contains 49,289 users, 139,738 items, 664,824 ratings on the items, and 487,002 trust statements. Also, there are 16,910 cold-start users, who gave ratings on less than 5 items.

In the experiments, we used *root mean square error (RMSE)* as an evaluation metric. We performed 5-fold cross validation. In each fold, we used 80% of rating data as a training set and the remaining 20% as a test set.

We first examined how the prediction accuracy changes with different θ. Figure 1 shows the result obtained with PMF as a recommendation method. We observe that RMSE is significantly reduced with data imputation and is the lowest when θ is 30%.

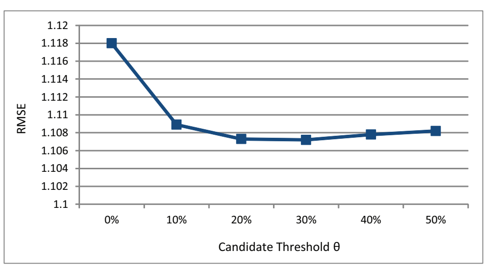

Figure 1: Accuracy with different θ.

Next, we compared the accuracy of the PMF model equipped with our imputation method with those of existing recommendation methods. In this experiment, we chose three existing recommendation methods: the user-based recommendation method [1], original PMF model [6], and SocialMF [2]. We set the parameter λT as 0.5 for SocialMF, which determines the influence of the social network in rating prediction. We performed 5-dimensional matrix factorization in the original PMF model, SocialMF, and our proposed method, and we set the parameters λU and λV as 0.01 for those methods. Besides performing experiments on all users, we also performed the same experiments for only the cold-start users.

Table 1 shows the results. We observe that the PMF model equipped with the proposed method is more accurate than others. Also, we see that the PMF model equipped with the proposed method and SocialMF outperform the original PMF model, which indicates that the use of a trust network is effective for improving recommendation accuracy. Besides, the proposed method is more accurate than SocialMF because it exploits *trustors* as well as trustees. For cold-start users, all the methods show accuracy worse than that for all users because the cold-start users provide less information. In summary, the proposed method provide higher accuracy than other recommendation methods. In addition, the differences between the proposed method and existing methods in terms of RMSE are bigger for the cold-start users compared with those for all users.

Table 1: Accuracy comparison

Method	RMSE	
	All Users	Cold-Start Users
User-based	1.185	1.350
PMF	1.118	1.210
SocialMF	1.115	1.209
Ours	**1.107**	**1.199**

4. CONCLUSIONS

We proposed a novel imputation method that is useful for most of recommendation methods. In order to estimate the unevaluated ratings accurately, the proposed method uses the trust network and fills only a part of unevaluated ratings. We performed experiments by using a real-world dataset and show that the PMF model equipped with our imputation method outperforms existing methods in recommendation for cold-start users as well as all users.

5. ACKNOWLEDGMENTS

This research was supported by the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-4009) and IT/SW Creative research program (NIPA-2013- H0503-13-1029). Also, it was supported by the Ministry of Culture, Sports, and Tourism (MCST) of Korea under Korea Copyright Commission in 2013.

6. 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 Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[2] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proc. of the 4th ACM Conf. on Recommender Systems, Recsys*, pages 135–142, 2010.

[3] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[4] H. Ma, I. King, and M. R. Lyu. Effective missing data prediction for collaborative filtering. In *Proc. of the 30th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval, SIGIR*, pages 39–46, 2007.

[5] P. Massa and P. Avesani. Trust-aware recommender systems. In *Proc. of the 2007 ACM Conf. on Recommender Systems, RecSys*, pages 17–24, 2007.

[6] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. *Advances in Neural Information Processing Systems, NIPS*, 20:1257–1264, 2008.