Exploiting ratings and social relationships effectively for social recommendation

Samrah Bablani[1], Irfan Ali Bhacho[1], Nouman Qadeer Soomro[2]

[1] Department of Computer Systems Engineering, MUET, Jamshoro, Pakistan.
[2] Department of Software Engineering, MUET, SZAB Campus, Khairpur Mir's, Pakistan.
sam.sb023@gmail.com, irfan.ali@faculty.muet.edu.pk, nouman.soomro@muetkhp.edu.pk

Abstract. Recommender systems are relatively new tools which are becoming choice of users to help them find relevant information. The idea behind recommendation engine is to predict what people might like and to discover relationships among them. However, recommender encounters problem in terms of data sparsity/cold-start users. In this paper we propose an enhanced approach for social networks which is based on matrix factorization model. It exploits ratings of social relationships only when cold-start users have stated less or no ratings else consider the original ratings of a user unlike existing approaches which exploit the social relationships, nevertheless. It is done by minimizing the Objective function. And hence produce effective recommendations. Different experiments performed on real-World Flixster dataset prove that our proposed approach exceeds other methods in recommendation.

Keywords: Recommender system · Matrix factorization-based model · Cold-start · Social network · Lower RMSE.

1 Introduction

E-Commerce is the place with overwhelming number of choices. Hence filtering and prioritizing need to be done to efficiently deliver the related information. Recommender systems mange the issue of information excess by filtering necessary and relevant information out of largely generated information on the internet as per user's preference, fondness or observed behavior about the item. Recommender systems are basically information cleaning systems with the ability to predict the preference of a user of an item based on his user's profile. In recent years, recommender systems have become insanely common and popular. They are utilized in nearly all e-commerce and social websites which include areas like books, movies, research articles, news, queries, restaurants, financial services, life insurance, online dating, social tags and other general genres. Every user u gives ratings of some number to set of items. The recommender partakes the task to predict the rating of item *i* which is not evaluated for a user. Or to suggest some products for the specified user *u* based on the already existing scores or ratings [3,5].

Two main algorithms work with recommender system. *Content-based filtering*. This algorithm compares between content of items and user profile.

Content of the item refers to assigning attributes to items so that content of each item is known. This is how Content-based filtering recommends items. e.g. Sci-fi adventures with strong female leads or Quirky rom-coms. On the other hand, user profile consists of objects that user has interacted with, watched, read, listened to or browsed past. In *Collaborative filtering*, recommendation is made when bunch of users whose assessment profiles are utmost alike to that of user u, rate item i. This algorithm is most effective when enough ratings are expressed by users to have shared ratings with other users. e.g. a site may recommend customer to purchase book if he has already purchased other books from there. This is done by comparing the shared ratings and historical preferences of users [1,3,4,5]. However, there are issues with collaborative filtering. First being of *Data sparsity/ Cold-Start user Problem* in which when user has stated no or limited number of ratings, it becomes difficult to recommend un-rated items. And other of ***Scalability*.** Fast computation and time is required to calculate recommendations where there are lots of choices [2,3,5].

Possible Solutions for above problems could be that most of the existing approaches exploit social networks to solve cold-start user problem. Existing approaches assume that a user and his/her social pears have similar preferences [2,3,5]. In *Online Social Network*, Users explicitly establish relationships with each other. e.g. Facebook, YouTube, Amazon etc. *Social Recommendations* are Algorithmic suggestions made by Social network for people to follow, read, like etc [7,10]. Existing recommendation approaches validate that exploiting social network significantly solves the cold start user problem.

But still there are *Issues with existing Recommendation approaches* [1,3,4,5,7,8,9] that (1) existing approaches exploit all the social relationships, if they have similar preferences. However, it is not always the case since despite being friends, in real world/ online social networks can have dissimilar preferences. (2) Moreover, if a user has rated large number of items then exploiting his/her all social relationships may increase computational complexity**.**

We propose simple yet effective matrix factorization model-based approach. Firstly, we compute similarity score between each user and their social peers so that only similar social peers for each user can be exploited. Then we introduce parameter "α" (alpha) that controls the use of user's ratings and social peers' information such that it exploits ratings of the user if he has given large number of ratings otherwise ratings of his/her social peers.

The rest of the paper is set as follows. In section2, we review related work regarding Recommender system. In section 3, we give detailed proposed approach. In section 4, we present experimental results and discussion. In section 5, with conclusion of paper, some future directions of work will be provided.

2 Related Work

In this section, we review existing social recommendation approaches and state their problems. Also, we will specifically talk about matrix factorization method since our approach is based on it.

The objective of the *Matrix factorization* is to get two low-dimensional matrices U and V such that rating matrix $\hat{R}$=UV. U and V are obtained by minimizing or maximizing objective function:

$$min_{U,V} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(r_{ij} - U_i^T . V_j\right)^2 + \lambda_U \|U_i\|^2 + \lambda_V \left\|V_j\right\|^2 \quad (1)$$

Such that each entry r_{ij} shows rating of each user i on item j and $\|.\|^2$

represents Frobenius norm.
Eq. 1 is divided into parts. The first part is referred to as factorization part which minimizes the difference between user's ratings and their respective predicted ratings. The regularization part contains λ_U and λ_V called regularization parameters used to avoid over-fitting [2,7]. If rating matrix R has dimension $d_1 * d_2$, then U will have dimension D*d_1 and V D*d_2. Here D if not chosen right leads to the problem of overfitting [3,5,11].
In recent years, many approaches have been proposed and most of them [1,2,4,7] address the common issue of data sparsity and cold-start user problem. Different approaches formulate and combine different techniques at their best to offer solution of the above issues and hence provide high quality recommendations.
Some algorithms use matrix factorization-based methods [1,4] in a way that combine users' sparse rating data and sparse social trust network in terms of their trust relationship for better provision of users' preferential patterns and valued recommendations [1]. While other approaches use separate algorithm like SVD integrated with matrix factorization technique to predict the unknown values or ratings for item by considering the explicit and implicit ratings and trust information. Explicit and implicit are two data gathering ways. For explicit, user provides data on some concrete rating scale (rating of book from 1 to 5 stars). For implicit data collection, system logs the behavior and actions of the user on the site and convert into preference. Combining the two can give optimized result [3,4,12]. In some cases, there is set good trade-off between trust-based and item-based approaches for recommendation i.e. not only ratings of target item but that of similar items are also taken into account [10]. Another similar approach which fuses users' taste and their trusted friends' favor together to negate the current recommender systems' assumption that all the users are independent and identically distributed. Hence this approach models system with more accuracy and reality to make it consistent with real world recommendation by introducing probabilistic factor analysis. It can be applied on large datasets [9].

3 Methodology

Describe Our methodology follows four main objectives. (1) Propose scheme that exploits social relationships optimally. (2) Experiments on real-World dataset. (3) Implementation of existing approaches. (4) Comparative analysis of proposed and existing approaches.

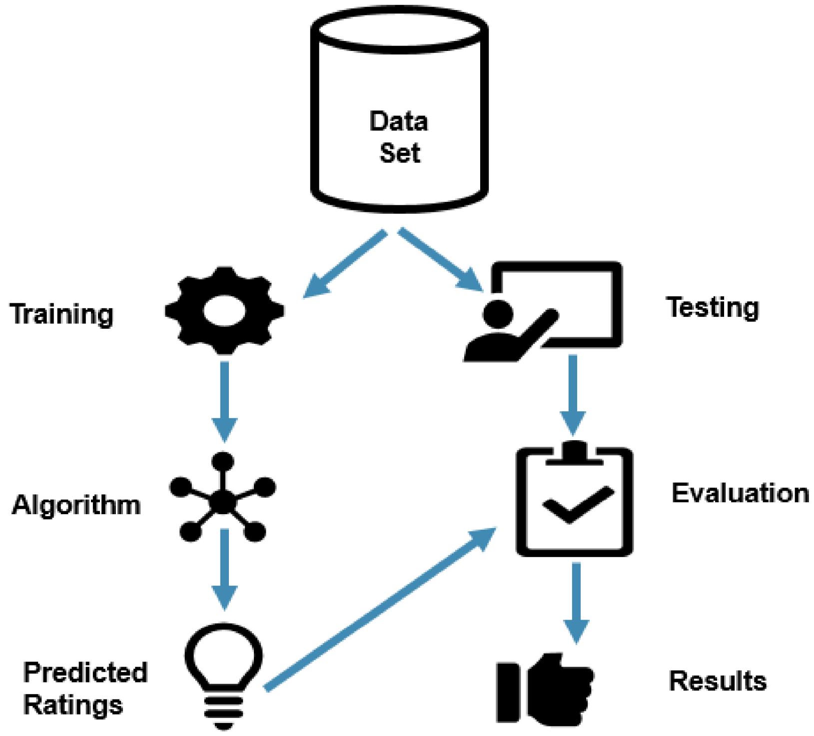

Figure 1: Methodology

This section presents details of our proposed approach. To address issue (1), we compute similarity score between each user and their social peers using Pearson correlation coefficient (PCC). If user i and u have rated two sets of items I(i) and I(u) and I= I(i)∩I(u) then

$$\text{Sim}(i,u)=\frac{\sum_{j\in I}(r_{ij}-R'_{i}).(r_{uj}-R'_{u})}{\sqrt{\sum_{j\in I}((r_{ij}-R'_{i})^{2}}.\sqrt{\sum_{j\in I}((r_{uj}-R'_{u})^{2}}} \qquad (2)$$

where $R'_u$is the average of user u's all the ratings on items [2]. After similarity score is calculated, only top-K most similar social peers are exploited. To address issue (2), we introduce parameter "α" (alpha) in an Objective function that controls the use of user's ratings and social peers' information such that it exploits the ratings of user only if he has given large number of ratings. If not, then ratings of his social peers.

$$min_{U,V}\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha\left(r_{ij}-U_i^T.V_j\right)^2+(1-\alpha)\|U_u-U_u^T\|^2+\lambda_U\|U_i\|^2+\lambda_V\left\|V_j\right\|^2 \quad (3)$$

The latent vectors can be found by taking partial derivative of Objective function w.r.t U and V called *Stochastic Gradient Descent*.

$\mathbf{U} \rightarrow \eta \frac{\partial F}{\partial U}$

$\mathbf{V} \rightarrow \eta \frac{\partial F}{\partial V}$

Where η is the learning rate. We find other parameters in the proposed approach as follows: K- assumed values: 5,10,15,20. α- we compute for each user. α=Number of ratings of user/ Highest number of ratings by a user i.e. from 0 to 1. λ_U and λ_V= 0.01 as assigned by most existing approaches.

4 Results & Discussion

We conducted multiple experiments to validate the success and usefulness of our approach over existing approaches. We give experimental settings like dataset characteristics, evaluation matrix, competitor approaches in this section for ground support before finally presenting our experimental results.

a. Dataset. We have used Flixster dataset for our experiments. It is publicly available. The statistics for dataset are given below in table 1. We have used 70% of the ratings as training set and 30% of the ratings as a test set.

Dataset	Users	Items	Ratings	Scale	Social Peers	Social Network type	Items
Flixster	147,612	48,794	8,196,077	[0.5,5.0]	787,213	Friendship	Movies

Table 1

b. Evaluation Metrics. We applied most recognized metrics to evaluate the predictive ratings, namely Root Mean Square Error (RMSE). The smaller the RMSE value, the more specific and accurate a recommendation.

$$\text{RMSE}=\sqrt{\frac{\sum_{u,j}(r'_{u,j}-r_{u,j})^2}{N}} \qquad (4)$$

where N is the number of test ratings.

c. Competitor Approaches. We compare our proposed approach with four existing approaches: *MF* [1],[3],[5],[7], *TrustSVD* [4], *SocialMF* [8], *TrustMF* [9]. We consider Matrix Factorization (MF) as our base for recommendation. It only utilizes rating information. TrustSVD is a trust-based Matrix factorization technique. It takes influence of both trust and rating into recommendation model. In SocialMF, to inject social influence in model, user dependent features get connected to direct neighbors and hence trust and social networks get propagated. TrustMF is a model-based method to map users into low-dimensional latent feature spaces in terms of their trust relationship.

d. **Experimental Results.** Figure 2 shows RMSE of proposed approach for various values of K and α. Figure 3 shows comparison of proposed approach and existing ones.

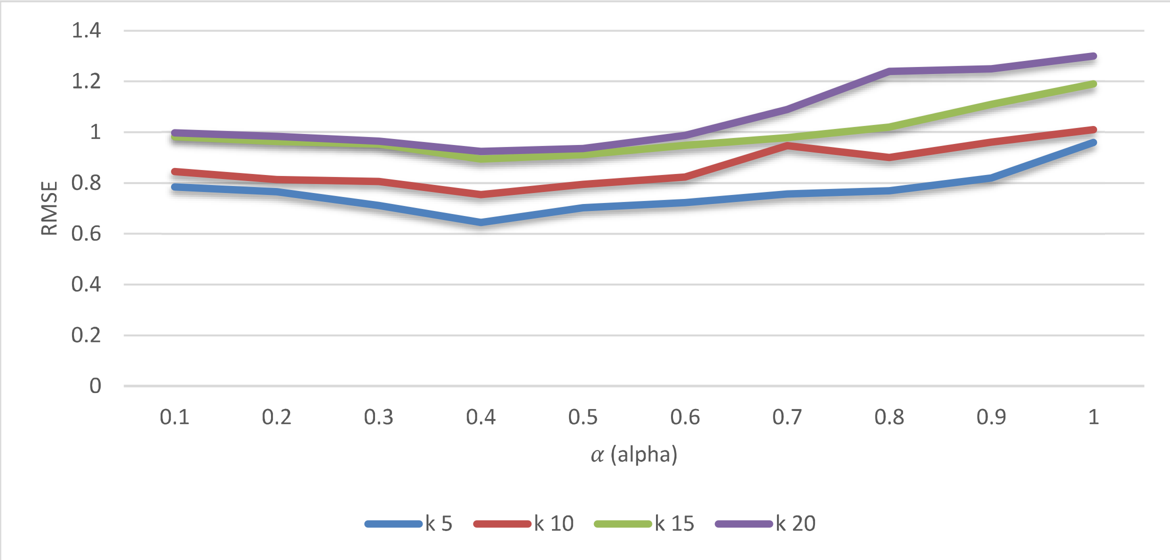

Figure 2: RMSE of Proposed approach for various values of K and α.

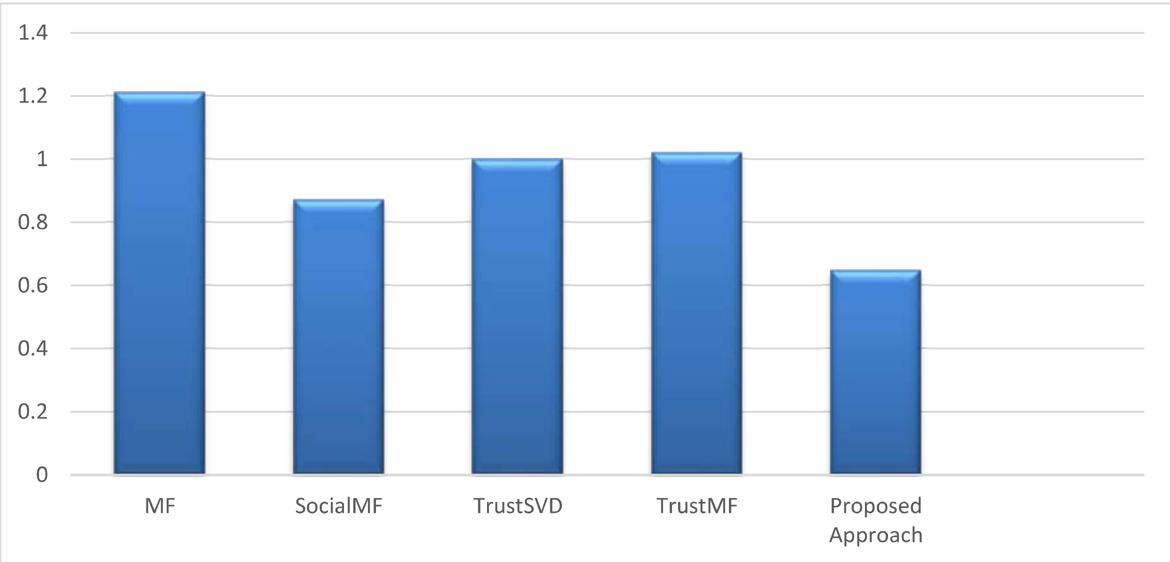

Figure 3: Comparative analysis of Proposed approach and existing ones.

It is visible in figure 2 that the value of RMSE is smaller for lower value of K. The range of the alpha is from 0 to 1. Alpha controls the use of user's ratings and the ratings of her social peers. If a user has provided very small number of ratings, then according to equation (3), alpha's value becomes 0 and only the ratings of a user's social peers will be exploited. If a user has provided a large number of ratings then alpha's value becomes near to 1, as a result, only users' ratings will be exploited. The RMSE of proposed approach does not directly depend on alpha. We have just performed experiments to find the optimal value of alpha. The RMSE decreases till the value of alpha is near to 0.5, then it starts increasing because a larger value (i.e., between 0.5 and 1) indicates that MF model will learn the preferences of the target user from the ratings of her social peers. Since a user and her social peers may have dissimilar preferences so that's why a user's true preference may not be learned by exploiting ratings of her social peers. As a result, the RMSE of the proposed approach decreases as the value of Alpha increases. K represents the number of social peers being exploited for a user. If, we exploit large number of social peers of each user then clearly the computational time of the proposed approach will also increase. In order to find the optimal value of K, we performed experiments for the various values of K. From initial experimental results, we found out that as the value of K increases then RMSE also increases and proposed approach performed best (lowest RMSE) when K=5. Therefore, in the proposed approach, we have set the value of K to 5. In figure 3, it can be clearly seen that RMSE value for our proposed approach is lowest than all other competitor existing approaches. In this paper, we have only focused on improving the accuracy of recommendation approach that exploits the information i.e. ratings of social peers. RMSE is the most widely used metric for evaluating the accuracy of recommendation approaches. Therefore, we have only computed RMSE value of proposed approach and the existing approaches. Moreover, lower value of RMSE indicates that recommendation approach performs better for making accurate recommendations to users. Most of the existing approaches exploit ratings of all the social peers of a user. Since a user and her social peers may have dissimilar preferences so that is the reason why the existing approaches achieve higher RMSE value. This evidently implies that our proposed approach outperforms existing methods for optimal recommendation.

5 Conclusion

In this paper we proposed an improved model-based approach for recommendation in Social networks. Our approach is a matrix factorization-based method. It deals better with Cold-start user problem than existing approaches such that RMSE of our approach is at a least than other competitor approaches by dint of our proposed approach exploit user’s ratings or social relationships on demand. All in all, our approach recommends better than others in terms of effectiveness.

References

1. Yang, Bo, et al. "Social collaborative filtering by trust." *IEEE transactions on pattern analysis and machine intelligence* 39.8 (2016): 1633-1647.
2. Ali, Irfan, Jiwon Hong, and Sang-Wook Kim. "Exploiting implicit and explicit signed trust relationships for effective recommendations." Proceedings of the Symposium on Applied Computing. ACM, 2017.
3. Aggarwal, Charu C. Recommender systems. Cham: Springer International Publishing, 2016.
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." Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.
5. Ricci, Francesco, Lior Rokach, and Bracha Shapira. "Recommender systems: introduction and challenges." Recommender systems handbook. Springer, Boston, MA, 2015. 1-34.
6. Yang, Xiwang, Harald Steck, and Yong Liu. "Circle-based recommendation in online social networks." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.
7. Jamali, Mohsen, and Martin Ester. "A matrix factorization technique with trust propagation for recommendation in social networks." Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010.
8. Jamali, Mohsen, and Martin Ester. "Trustwalker: a random walk model for combining trust-based and item-based recommendation." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.

ACM, 2009.

9. Ma, Hao, Irwin King, and Michael R. Lyu. "Learning to recommend with social trust ensemble." Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009.
10. Ma, Hao, et al. "Sorec: social recommendation using probabilistic matrix factorization." Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008.
11. Cambridge Spark. “Tutorial: Practical Introduction to Recommender Systems.” Medium, Cambridge Spark, 12 Oct. 2018, blog.cambridgespark.com/tutorial-practical-introduction-to-recommender-systems-dbe22848392b.
12. “Overview.” Recommendation Systems :: General Collaborative Filtering Algorithm Ideas,www.cs.carleton.edu/cs_comps/0607/recommend/recommender/collaborativefiltering.html.