Neurocomputing 191 (2016) 388–397

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An exploration of improving prediction accuracy by constructing a multi-type clustering based recommendation framework

Xiao Ma, Hongwei Lu, Zaobin Gan*, Qian Zhao

School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 4 June 2015
Received in revised form
28 December 2015
Accepted 17 January 2016
Communicated by Peng Cui
Available online 18 February 2016

Keywords:
Recommender Systems
Multi-type Clustering
Social Trust Relationships
Collaborative Filtering

ABSTRACT

Existing clustering-based recommendation methods generally focus on the clustering of users, items or social trust relationships. Although demonstrated to be efficient and scalable to large-scale datasets, these methods are sensitive to the quality of clustering and still suffer from the problem of low accuracy. In order to solve this issue, in this paper, we propose a multi-type clustering based recommendation framework which systematically considers the trust-based user clustering, similarity-based user clustering and similarity-based item clustering to further improve the recommendation accuracy. A SVD (Singular Value Decomposition) signs-based community mining method is utilized to process the trust and distrust matrix in order to discover the trust-based user clusters. The PLSA (Probabilistic Latent Semantic Analysis)-based model is employed to explore the similarity-based user and item clusters. Then a clustering-based trust regularization term is proposed to incorporate the trust-based user clusters into the matrix factorization model. Comparative experiments on two real-world datasets demonstrate that our approach can better address the issues of data sparsity and cold start, and outperforms other state-of-the-art methods in terms of RMSE and MAE.

1. Introduction

Recommender Systems (RSs) as an indispensable type of information filtering technique dealing with information overload have attracted lots of attention in the past decades. Such systems help users explore their interests in various domains, including movies, music, books, and academics. Most of recommender systems are based on Collaborative Filtering (CF), which is a technique that automatically predicts the interests of an active user by collecting rating information from other similar users or items [1].

Typically, collaborative filtering based RSs can be classified into memory-based methods and model-based methods. Memory-based CF methods explore the entire user–item rating matrix to find similar neighbors (also known as nearest neighbors) for a given user or item. Model-based CF methods including matrix factorization [2], latent semantic models [3], Bayesian Inference [4] usually learn parameters of a model offline and make prediction for unrated items with the generated models. These methods do not need to explore the rating matrix and only store the model parameters.

Although traditional CF models have been successful in many areas, they all have to face several critical problems: data sparsity, scalability and cold-start. In order to solve these inherent problems, the clustering-based recommendation approaches have been proposed, which have been proved to improve the quality of recommender systems [5,6]. Different clustering strategies can be performed based on users, items or trust relationships, which results in several sub-matrices of the entire user–item rating matrix and groups the well-connected users or items into the same clusters. Then traditional collaborative filtering approaches can be applied to the sub-matrices, which alleviate the data sparsity and scalability problems to a large extent. Fig. 1 is a toy example of diverse types of clustering. There are six different users and items in this figure. In Fig. 1(a), users are clustered based on the similarity among users, thus users u_1, u_2, u_3 and u_4 are clustered into the similarity-based user cluster. In Fig. 1(b), users are clustered based on the trust relationships, thus users u_3, u_4, u_5 and u_6 who have trust connections between each other are clustered into the trust-based user cluster. Similarly, two similarity-based item clusters are obtained as shown in Fig. 1(c).

As shown in Fig. 1, existing clustering-based recommendation algorithms are mainly based on single type of clustering information, such as similarity-based user clustering [7], similarity-based item clustering [8], and trust-based user clustering [9,10]. Recently, Guo et al. [6] develop a clustering method through which users are iteratively clustered from the views of both ratings and

* Corresponding author.
E-mail addresses: cindyma@hust.edu.cn (X. Ma), luhw@hust.edu.cn (H. Lu), zgan@hust.edu.cn (Z. Gan), zqhuster@163.com (Q. Zhao).

http://dx.doi.org/10.1016/j.neucom.2016.01.040
0925-2312/

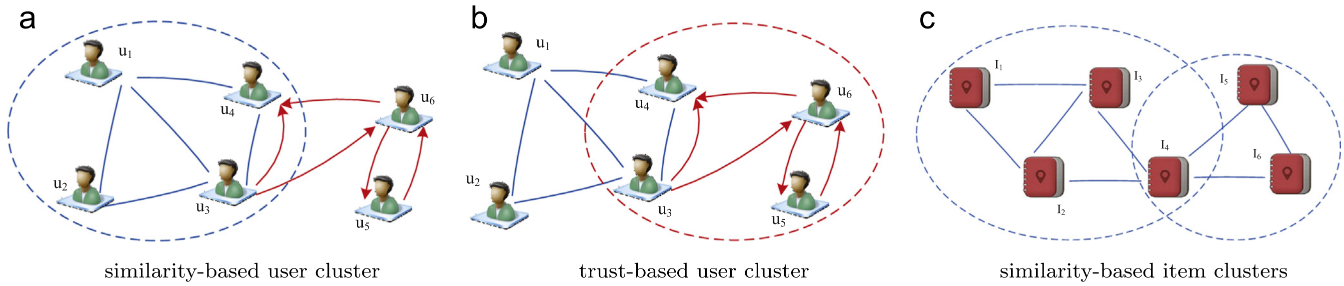

Fig. 1. A toy example of diverse types of clustering: Dotted circles denote the formed user/item clusters, undirected blue lines represent the similarity between two users, while directed red line represent the trust relationship between users. (a) User clustering based on the similarity between users; (b) user clustering based on the trust relationships between users; (c) item clustering based on the similarity between items. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

social trust relationships. There are also some works focusing on clustering users and items simultaneously (called co-clustering), which further improve the accuracy of clustering-based recommender systems [11–13]. It is known to us that recommendation performance is sensitive to the clustering results. Similar users can only be selected from the fixed size of cluster members, and in general a fewer number of similar users can be identified compared with the whole space. Therefore clustering-based methods still suffer from relatively low accuracy [6,9], which severely hinders the practical use of clustering-based methods in recommender systems.

In order to solve the issue, in this paper, we design a unified framework named as *MCRec* (*M*ulti-type *C*lustering based *Rec*ommender systems) which systematically combines multi-type clustering sources, namely, the similarity-based user clustering, the similarity-based item clustering, and the trust-based user clustering, to improve the performance of clustering-based recommender systems. The proposed MCRec can be characterized as follows.

Firstly, as co-clustering is extremely important when dealing with large, sparse data matrices [11], MCRec generates sub-matrices by clustering users and items simultaneously which are so much denser than the original user–item matrix that will greatly alleviate the data sparsity problem. Secondly, for cold start users with only a few user-generated information, traditional clustering-based methods cannot capture their personal tastes accurately. Trust clustering classifies the densely connected trust users into the same trust cluster, and provides an additional data source to enhance the clustering-based recommender systems by analyzing the preferences of their trust neighbors in the same trust community [14]. Thirdly, more information can be used to learn users' preferences by proposing a unified recommendation framework making use of multiple clustering sources. Hence intuitively, the recommendation performance will be improved, as we will demonstrate later.

Our objective is to incorporate the multi-type clustering sources into a unified recommendation framework to improve the quality of clustering-based recommender systems. Thus we will elaborate these two questions in our paper: (1) *How to combine the multi-type clustering information (as shown in Fig. 1) into a unified recommendation framework?*(2) *How can this combination further improve the accuracy of recommender systems?*

In summary, the main contributions of this paper are:

- We propose the multi-type clustering problem in the domain of clustering-based recommender systems and present a method well-suited for this problem.
- A SVD signs based community mining method is utilized to process the trust and distrust relationship matrix aiming at discovering the trust-based user clusters. And the PLSA-based model is employed to explore the similarity-based user clusters and similarity-based item clusters.
- A clustering-based trust regularization term is proposed to incorporate the trust-based user clusters into the matrix factorization model.
- Extensive of comparative experiments are conducted to explore how can this combination improve the performance of recommender systems. Real datasets based experiments demonstrate the effectiveness of our method in comparison with state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 gives an introduction of the related work on trust-based and clustering-based recommender systems. The proposed multi-type clustering framework is discussed in detail in Section 3. Section 4 includes our experimental algorithm in alleviating the data sparsity and cold start problem in rating matrix and generating more accurate recommendations. Finally, Section 5 concludes this study with future work.

2. Related work

In this section, we review the approaches to recommender systems, including trust-based recommender systems and clustering-based recommender systems.

2.1. Trust-based recommender systems

Trust-based recommender systems exploit trust information explicitly expressed by users to help model user preferences. Generally, trust-based recommender systems can be classified into memory-based methods and model-based methods. Memory-based trust-aware recommender systems use memory-based collaborative filtering methods as their basic models. They search the trust networks to obtain trust neighbors for a given user. And the correlated users for the given user can be the trust neighbors, similar neighbors, or a combination of similar neighbors and trust neighbors. For example, there are a Trust Metric module and a Similarity Metric module in the architecture of trust-aware recommender systems proposed by Massa and Avesani [15]. Therefore, the weights for identifying neighbors can be generated by trust metrics or similar metrics. They find that trust is useful in helping alleviating the issues of traditional collaborative filtering such as data sparsity and cold start. Jamali and Easter [16] propose the TrustWalker, a random walk model to combine trust-based and item-based collaborative filtering approach for recommendation, which clearly outperforms the traditional collaborative filtering approaches in terms of precision.

Model-based trust-aware recommender systems use model based collaborative filtering methods as their basic models. Matrix