Knowledge-Based Systems 106 (2016) 125–134

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Enhancing collaborative recommendation performance by combining user preference and trust-distrust propagation in social networks

Wei-Po Lee*, Chuan-Yuan Ma

Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 21 October 2015
Revised 17 May 2016
Accepted 18 May 2016
Available online 19 May 2016

Keywords:
Social network
Collaborative filtering
Social trust
User preference
Distrust propagation
Recommendation

ABSTRACT

Collaborative filtering (CF) is one of the most popular recommendation methods, and the co-rating-based similarity measurement is widely used in CF for predicting ratings of unfamiliar items. In addition to rating information, social trust has now been considered useful in collaborative recommendations. In this work, we present a hybrid approach that combines user ratings and social trust for making better recommendations. In contrast to other trust-aware recommendation works, our approach exploits distrust links and investigates their propagation effects. In addition, our approach combines the *k*-nearest neighbors and the matrix factorization methods to maximize the advantages of both rating and trust information. Several series of experiments are conducted, in which different types of social trust are incrementally included to evaluate the presented approach. The results show that distrust information is beneficial in ratings prediction, and the developed hybrid approach can effectively enhance the recommendation performance.

1. Introduction

Recommender systems have been promoted in different service domains for years [1,4,33]. Traditionally, recommender systems address two entities for application services: the users and the items. Such systems collect the historic ratings specified by users and then predict the rating of unfamiliar items and recommend items with highest-ratings to users. Various methods have been developed to find suitable solutions, among which the collaborative filtering (CF) recommendations attempt to automate word-of-mouth process that people receive recommendations for other specially selected people. It is generally recognized that the CF methods are more efficient and effective than content-based methods [4,23].

Algorithms for collaborative recommendations can be grouped into two classes: memory-based and model-based methods. Memory-based (also called neighborhood or *k*-nearest neighbors) methods are heuristics that make rating predictions based on the entire collection of items previously rated by the users. Model-based methods provide an alternative by transforming both users and items to the same latent factor space. This method is based on the assumption that there should be certain latent features that determine how a user rates an item. If these latent features can be discovered, often by a stochastic algorithm, then the methods are able to predict a rating regarding a certain user and a certain item. With this technique, the null entries in the original rating matrix can be filled. Both classes of methods have their advantages, depending on the specific application service required.

Standard CF methods consider only user preference (ratings) for unknown item predictions. Though rating information has proven itself useful, there are several issues yet to be addressed. For example, in the cases that the interactions among the users are not considered and cases that certain users rated a very small number of items, the rating information is not sufficient for making helpful recommendation. In addition, in real-world applications, most items are not widely rated by users; thus, there is no relevant and useful information provided. These problems may reduce the recommendation performance [2,16]. To overcome these problems, researchers have suggested additional information for developing more accurate recommender systems, for example [6,8,24,28]. Among others, the social context information that is collectable from online communities is useful and important for recognizing users' situations that can influence their decisions [34,42]. In addition, the social network theory suggests that the positions of users in a web of relationships influence their access to resources, friends, and information. Influences of friends, families and colleagues can attribute to one's intension to consume a product; therefore, information regarding social networking is also a key factor for predicting potential customers. Researchers have indicated that the relationships among acquaintances within a social network are crucial when referencing trustworthy and reliable information [14,21,32].

* Corresponding author.
E-mail address: wplee@mail.nsysu.edu.tw (W.-P. Lee).

http://dx.doi.org/10.1016/j.knosys.2016.05.037
0950-7051/

Trust information can be explicitly collected from users or implicitly inferred from users' behaviors. It is believed that, by utilizing a trust network (a social network augmented with trust ratings) in collaborative filtering methods, better results can be obtained [9,35]. With a trust-enhanced recommendation, the system can ask the users to rate other users. Thus, a user can express his level of trust in another user he has interacted with. The system can then aggregate all the trust statements in a single trust network. Here, users are represented as nodes, and trusted neighbors are connected with each other by trust links. Because the trust relationship is not symmetric, the resulting network is thus directed. The above trust relationship can be further quantified to a numerical value representing the importance of each link (i.e., arc). That is, the strength of the link indicates the trustworthiness (i.e., degree of trust) between two users. In addition to the trust information, the other type of information collectable from a trust-aware network is the distrust relationship between users. Similar to the trust information, distrust links can be explicitly specified by the users or implicit inferred by the system. However, in contrast to the trust information that has been used in many studies, there are very few works that explore the explicit use of distrust information in recommendations. Indeed, it is a challenge to model a distrust propagation in a manner both logically consistent and psychologically plausible. Thus, it is worthwhile to investigate how distrust information could be effectively utilized to improve the accuracy of recommender systems [11].

In this work, we present a hybrid approach that combines user preference and social trust for making better collaborative recommendations. Here, the user preference means the co-rating-based similarity measurement between users, and the social trust means the direct (explicitly specified) and indirect (implicitly inferred) trust relationships between users. Most importantly, in contrast to other trust-aware recommendation studies, our approach exploits distrust links and investigates their propagation effects. To validate the proposed approach, a set of analyses is first performed to show that the user preference and the social trust between users could be uncorrelated and used in a complementary manner. Then, several series of experiments, in which different types of social trust are incrementally included, are conducted to evaluate the proposed approach. The results show that distrust information is beneficial in rating prediction, and the presented hybrid approach is considerably significant. The approach can obtain suitable results in several objective experimental conditions to enhance recommendation performance.

2. Background and related works

In the procedure of developing collaborative recommender systems, the most important factors are the user information and the computational method. Often, the user information includes ratings and social trust [36,41], and the widely adopted methods can be grouped into two categories: memory-based and model-based [23,33]. This section describes briefly how the two types of information (i.e., rating and trust) are measured and how they are operated in the two categories of computational methods.

The first category (memory-based methods) performs rating predictions based on the entire collection of items previously rated by the users. That is, the value of the unknown rating $r_{u,i}$ for user u and item i is usually computed as an aggregation of the ratings of the top-k most similar users for the same item i. There are many methods to calculate the similarity among users (such as Cosine similarity and Euclidean distance), and one often used method is the Pearson correlation coefficient. The similarity between the two users, u and v, is defined as below:

$$Sim(u,v) = \frac{\sum_{i \in Ic}(r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in Ic}(r_{u,i} - \bar{r}_u)^2}\sqrt{\sum_{i \in Ic}(r_{v,i} - \bar{r}_v)^2}} \tag{1}$$

In Eq. (1), $r_{u,i}$ is the rating of user u on item i, and I_c is the set of items that users u and v have already rated (i.e., the co-rated items). $\overline{r_u}$ (or $\overline{r_v}$) is the average rating of user u (or user v) regarding all items he has rated; this coefficient is between 1 (when the preferences of both users are the same) and −1 (when their preferences are opposite). For user u, other users with the most similar preferences are chosen as a set of neighbors $Neig(u)$, and their collective opinions on a certain item m are used to predict whether u will like the item. The rating of the preference of a specific item m is defined as

$$p_{u,m} = \bar{r}_u + \frac{\sum_{v \in Neig(u)} Sim(u,v) \cdot (r_{v,m} - \bar{r}_v)}{\sum_{v \in Neig(u)} Sim(u,v)} \tag{2}$$

In this equation, $p_{u,m}$ represents the predictive rating of user u on item m, and $r_{v,m}$ is the rating of user v who is a neighbor of user u.

The neighborhood methods are popular because they are intuitive and relatively simple to implement. In addition, these methods offer useful and important properties: an explicit explanation of the recommendations and an easy inclusion of new ratings. However, standard neighborhood-based methods raise several concerns (e.g., cold-start and sparsity problems) that could affect the precision of the recommendation results. Social trust has been proposed to overcome the difficulties of the similarity-based CF method [3,25,28,30]. It is now clear that establishing a trust network among users can be helpful to the success of recommender systems. Many studies have been conducted by eliciting trust values into collaborative recommender systems. These works have improved the accuracy of predictions, offered more robustness against profile attacks and addressed the problems of sparsity and cold start [36,41].

The most popular trust measurement in a trust network is the Jaccard coefficient or its variants [7,29,34]. Without losing generality, here we use a modified normalized Jaccard coefficient as an example to explain the trust measurement. In a trust network, the weight of a link between two user nodes (i.e., the explicit trust value) is defined as

$$T_{u,v} = \frac{1}{\max T_{u,k}} \times \frac{max(|O(u) \cap O(v)|, 1)}{|O(u) \cup O(v)|}, k \in O(u) \tag{3}$$

In the above equation, $T_{u,v}$ represents the degree of trust of user u to user v, and this value is between 0 and 1 inclusively. $maxT_{u,k}$ is the maximal degree of trust among user u to all his connected neighbors, and $O(u)$ and $O(v)$ are the out-neighbors of nodes u and v, respectively. The conjunction used for the two connected nodes u and v intends to consider their common out-neighbors. To retain the explicit trust relationship specified by the user, a lowest default value of 1 is used in this equation.

Certain studies have also indicated that a user is much more likely to believe a trusted user's rather than a stranger's statements. Because a trusted user will also trust his friend's opinions, in a recursive manner, trust may propagate through the relationship network [15,18,29,45]. Therefore, in addition to the explicit trust relationships specified by the users, algorithms have been developed to search for trustable users by exploiting the trust propagation over the trust network. For example, Golberk *et al.* presented an algorithm (called TidalTrust) that performs a modified breadth-first search in the trust network to find indirect neighbors and aggregate their trust values [13]. In the work by Massa and Avesani [29], the authors adopted a depth-first search in order to search indirect neighbors. The authors used a pre-defined depth