Information Systems 77 (2018) 118–128

Contents lists available at ScienceDirect

Information Systems

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

Trustor clustering with an improved recommender system based on social relationships

Jaehoon Lee[a], Giseop Noh[b], Hayoung Oh[c], Chong-kwon Kim[a,*]

[a] Department of Computer Science and Engineering, Seoul National University, Seoul, South Korea
[b] College of Engineering, Division of Software Convergence, Cheongju University, Cheongju, South Korea
[c] Department of DASAN University College, Ajou University, Suwon, South Korea

ARTICLE INFO

Article history:
Received 6 June 2017
Revised 15 May 2018
Accepted 28 May 2018

Keywords:
Recommender system
Social network
Trust relationships
Trustor cluster

ABSTRACT

As we face a deluge of information in the modern world, the importance of recommender systems (RSs) that recommend relevant items to users has increased. The majority of existing RS schemes observe the prior ratings history of consumers to identify preferred items. However, current RSs suffer from the cold start problem, and their performance is dismal when new users or items appear. In order to address the cold start problem, a new type of solution that exploits social network features has been proposed. Many such social RSs analyze trustor–trustee relationships to discover latent social features shared between trustor and trustee. Since social relationships between trustors and trustees are directed, but not reciprocal, it is not guaranteed that a trustee has features in common with its trustors. Moreover, existing schemes are based on the assumption of independence between trustors who follow the same trustee, and therefore fail to recognize quintessential factors shared by the trustors.

We posit that trustors who follow the same trustee have features in common. Based on the assumption that trustors who endorse the same trustee share similar tastes, we propose a new latent feature called Matrix *S*, and develop two novel RS algorithms that learn these latent features. We conduct an extensive performance evaluation using large scale real-world datasets, and observe that our proposed methods are not only more accurate than existing schemes but also show potential extensibility.

1. Introduction

Following the enormous development of the Internet and a plethora of social web services, users today benefit from relatively convenient and efficient ways of accessing information. For example, the majority of consumers refer to product/service review systems such as Yelp before making purchase decisions. However, customers are provided with a massive volume of review information, and thus experience the problem of information overloading [1], meaning that information is presented at a rate that is too fast for an individual to process. To address this problem, a variety of recommender systems (RSs) have been proposed. An RS acts as a software agent that helps people to find favorable products/services by making personalized recommendations. From the customer's point of view, RSs help them to discover suitable products promptly, while from the provider's point of view, RSs contribute to increases in the profits of retailers and manufacturers by persuading the consumers to buy their goods. Many online commercial sites such as Amazon, Etsy, eBay, Netflix, Yelp, TripAdvisor, LinkedIn, and Facebook use RSs to recommend personalized items (e.g., products, movies, places, people, jobs, and friends).

Traditional RSs can be classified into three categories: content-based [2,3], collaborative filtering [4–12] and hybrid methods [13–16]. *Content-based* methods make recommendations by identifying the context of textual information or by comparing characteristics between content depending on the user's profile information. This type of scheme analyzes the profile of the target user and suggests items that may be appropriate for the user. The user's profile refers to the attributes of the user, extracted based on items previously purchased. An item profile consists of the characteristics of the item itself, such as category, genre, brand, and so forth. The primary advantage of the content-based approach is that items are recommended to a user by analyzing the characteristics of the user or items without purchasing information of other users or items. However, content-based schemes require well-established classification of item attributes. *Collaborative filtering* exploits information accumulated about user-item ratings, and tries to correctly predict future ratings based on prior rating information about similar users and items. This method is effective when there is enough

* Corresponding author.
E-mail addresses: jhlee@popeye.snu.ac.kr (J. Lee), ckim@snu.ac.kr, ckkim@popeye.snu.ac.kr (C.-k. Kim).

https://doi.org/10.1016/j.is.2018.05.008
0306-4379/

consistency among ratings for each item without outliers in a user-item rating matrix. *Hybrid* methods are synergistic blends of the above two approaches (content-based methods and collaborative filtering).

Simple RSs assume that users make independent decisions, and that these also do not affect the decisions of others. However, in reality, when customers purchase an item or use a service, they generally seek advice from their friends or from experts in that specific field. Experiments [17,18] have shown strong evidence of the influence of word of mouth, such that the actions of friends influence a user when selecting items/services. In recent years, there has been an exponential growth in online social networking services that can provide additional information about users, independent of the user-item rating information. Information can propagate at the speed of light within online social media, and verbal marketing is statistically proven to increase commercial profit.[1] Social information may improve the accuracy of recommendations, since people have a tendency to like what their friends like. Social networking service (SNS) makes people stay connected virtually, even when they cannot actually meet each other for various reasons. SNS is used in searching for information, exchanging opinions and sharing knowledge and ideas. Thus, the necessity of using social relationships to recommend relevant items has increased rapidly.

In this paper, we focus on the follower–followee and/or trustor–trustee types of social relationship. A prior work [19] analyzed follower behaviors and found a tendency to adhere to followees of the same political stance. The authors also showed that the political views of followers who follow a given followee are similar to each other. In a similar way to the follower–followee relationship, we first focus on the relationships between the trustors of each trustee in our datasets, and then propose trustor clustering-based RS schemes which adopt trustors' relationships as latent features. We assume that people who trust a given person possess similar tastes, and group trustors for each trustee into a cluster, aligning the attributes of each individual trustor such that they conform to those of the trustor cluster.

We conducted extensive experiments with four real-world datasets, comparing the performance of the proposed schemes with that of state-of-the-art methods, and observed that the proposed scheme improves the accuracy of recommendation.

The contributions of this paper are as follows:

- In addition to an individual user feature vector, we propose a trustor group feature vector that represents the common attributes of cluster members.
- We devise two RS models that utilize the features of the trustor group to enhance the prediction accuracy of recommendations.
- By conducting experiments on four real datasets (Ciao, Epinions, LastFM, and Douban), we validate the performance of our models and observe that they outperform prior methods.

The rest of this paper is organized as follows. Section 2 discusses related work. We explain the rating matrix, trust matrix, conceptual modeling of the trustor cluster and an overview of matrix factorization (MF) in Section 3. In Section 4, we formulate the proposed scheme for RS algorithms based on trustor clustering. Our experiments are presented in Section 5. We conclude in Section 6.

2. Related work

There have been tremendous research efforts towards the development of effective recommender systems in the past two decades. We discuss several important schemes that are closely related to our work.

Collaborative filtering (CF) is one of the most widely used techniques for providing personalized recommendations [20]. The preferences of each specific user are predicted based on the assumption that similar users or items have the same inclinations and features. CF methods can be further divided into memory-based and model-based approaches.

Memory-based techniques make predictions based on the rating behaviors of other similar users [21] or items [4,11]. By calculating the similarity between users or items, RSs recommend items that similar users like (or items that are similar to the items that the user likes). These methods are effective when user-item rating matrices are dense; however, since sparse matrices are more often the norm in many product review systems, this technique has limited applicability in a real-world environment. Moreover, there is the problem known as "cold start", which occurs when users or items first appear, since it is hard to extract the characteristic features of a user or item when there is no information about them.

Model-based techniques use rating data to train a model, and then adjust the trained model to find proper mathematical parameters. To mitigate the cold start problem, many low-dimensional matrix approximation approaches have been studied within CF [22–25]. In contrast to memory-based approaches, which exploit similarities between users or items, model-based approaches assume that these similarities are induced by some hidden lower-dimensional structure embedded in the data. MF techniques are widely used as popular latent feature models in which users and items are simultaneously represented as unknown feature vectors. These feature vectors are determined when an optimization problem that minimizes the difference between the inner products of two vectors (i.e. user latent feature vectors and the item latent feature vectors) and the original user-item rating is solved. Recently, social information was added to these methods as additional information [26–36]. It is possible that when recommending items, RSs can use not only user-item latent features but also the social relationship in a way that reflects the influences of users.

Ma et al. [26] have proposed a factor analysis approach based on a probabilistic graphical model. Their method, SoRec, combines the user-item rating matrix and user's trust matrix by sharing the user latent feature space. Even though SoRec outperforms traditional plain CF methods, the social relationship information is not utilized optimally because the user latent features are applied with the same weight in solving the matrix optimizations for ratings and trust relationships. RSTE [27] also proposed a probabilistic MF scheme based on the graphical model, which focuses on trust relationships and reflects the tastes of both a user and his/her trustees when predicting item ratings. Jamali and Ester [28] have also proposed another probabilistic MF method called SocialMF, which assumes that a user has multiple trustees, and blindly aligns the feature vector of the user with the average feature vector of the trustees, even though the user is not similar to the trustees.

SoReg [29] also tries to exploit trust relationships. Unlike SocialMF, it considers not only the trust relationships but also the similarities of the users. SoReg assumes that the tastes of a user are similar to the average tastes of their trustees. However, a trust relationship is not reciprocal, and the trustor and trustee may have different tastes; when a trustor has dissimilar trustees, SoReg may therefore distort the feature vector of the trustor. Furthermore, when there are no common item ratings among friends, it is not possible to compute similarity between users. Yu et al. [30] resolved the issue by employing user latent features to measure similarity. Wang et al. [31] proposed the use of weight transformation matrices to overcome the limitations of SoRec [26]. Their method generates a social relationship matrix by exploiting two types of weight transformation matrices, each of which adjusts user and

[1] http://allfacebook.com/facebook-word-of-mouth_b69800.