ARTICLE IN PRESS

Neurocomputing (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

A hybrid collaborative filtering model for social influence prediction in event-based social networks☆

Xiao Li, Xiang Cheng, Sen Su*, Shuchen Li, Jianyu Yang

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

ARTICLE INFO

Communicated by Zidong Wang

Keywords:
Event-based social network
Social influence
Matrix factorization
Neighborhood method
Collaborative filtering

ABSTRACT

Event-based social networks (EBSNs) provide convenient online platforms for users to organize, attend and share social events. Understanding users' social influences in social networks can benefit many applications, such as social recommendation and social marketing. In this paper, we focus on the problem of predicting users' social influences on upcoming events in EBSNs. We formulate this prediction problem as the estimation of unobserved entries of the constructed user-event social influence matrix, where each entry represents the influence value of a user on an event. In particular, we define a user's social influence on a given event as the proportion of the user's friends who are influenced by him/her to attend the event. To solve this problem, we present a hybrid collaborative filtering model, namely, Matrix Factorization with Event-User Neighborhood (MF-EUN) model, by incorporating both event-based and user-based neighborhood methods into matrix factorization. Due to the fact that the constructed social influence matrix is very sparse and the overlap values in the matrix are few, it is challenging to find reliable similar neighbors using the widely adopted similarity measures (e.g., Pearson correlation and Cosine similarity). To address this challenge, we propose an additional information based neighborhood discovery (AID) method by considering both event-specific and user-specific features in EBSNs. The parameters of our MF-EUN model are determined by minimizing the associated regularized squared error function through stochastic gradient descent. We conduct a comprehensive performance evaluation on real-world datasets collected from DoubanEvent. Experimental results show that our proposed hybrid collaborative filtering model is superior than several alternatives, which provides excellent performance with RMSE and MAE reaching 0.248 and 0.1266 respectively in the 90% training data of 10 000 users dataset.

1. Introduction

In the past few years, event-based social networks (EBSNs), such as Meetup[1] and DoubanEvent[2], have proliferated to be the online platforms for users to organize, attend and share social events to be held in offline physical venues [1]. EBSNs link online and offline social worlds, providing not only typical online social networking services, but also face to face offline communication by attending events. Previous studies [2–4] have shown that users could influence others to attend events in EBSNs, especially for close social ties. For instance, when an organizer publishes an event, users can express their willingness to join the event by RSVP ("yes" or "maybe")[3] and broadcast posts about their participating information to their friends (i.e., followers), who hesitate in making decisions. When a user's friends see his/her participating post, they might want to attend the event together with the user.

Social influence aims to study the behavioral change of a person because of the perceived relationships with other people in social networks. Since it has a wide range of applications [5], such as social recommendation [6] and social marketing [7], considerable works have been conducted to the influence analysis or prediction in social networks (e.g., Twitter and Facebook) [8–10]. Different from the conventional social networks, EBSNs have some unique characteristics. For example, each event in EBSNs not only contains content information but also contains organizer and location information. These

☆ The preliminary version of this paper has been accepted for publication in the proceedings of the 21st IEEE International Conference on Database Systems for Advanced Applications Workshop (DASFAA 2016 Workshop), Dallas, TX, USA, April 16–19, 2016.

* Corresponding author.

E-mail addresses: lixiao@bupt.edu.cn (X. Li), susen@bupt.edu.cn (S. Su).

[1] www.meetup.com

[2] www.douban.com

[3] The RSVP ("yes" or "maybe") indicates that a user wants to attend or is interested in an event. We assume that a user will attend the events which he/she has expressed RSVP ("yes") to.

http://dx.doi.org/10.1016/j.neucom.2016.12.024
Received 16 March 2016; Received in revised form 14 September 2016; Accepted 6 December 2016

unique characteristics might affect the social influences of users. For example, if the majority of the social friends of a user live in New York, this user might have larger social influences on events held in New York than those in San Francisco. Thus, the social influence analysis or prediction approaches used in conventional social networks might be inaccurate in EBSNs. Nevertheless, social influences of EBSN users can provide valuable insights. For an upcoming event (i.e., the event which has not been held but has been published in the EBSNs), the event organizer hopes to maximize the attendees. This goal makes him/her willing to target the influencers on this event. These influencers are able to let many friends to attend this event by sharing the event. In this case, the event organizer needs to know users' influences to their friends. Therefore, understanding users' influences is a key issue in EBSNs.

In this paper, we focus on the social influence prediction problem in EBSNs. We formulate this prediction problem as the estimation of unobserved entries of the constructed social influence matrix S, where each entry (u, e) represents the social influence of user u on event e. Notice that, we focus on predicting users' influences on upcoming events which could provide valuable information for event organizers. In particular, similar to the definition of item-level social influence in conventional social networks [11], we define user u's influence on event e as the proportion of u's friends who are influenced by u to attend event e. Different from the structure-level influence [12] and the topic-level influence [13,14], the predicted event-level influence can be used in two angles. On one hand, given an upcoming social event, we could find out the influencers to attract more friends for attending the event. On the other hand, given a user, we could recommend events for him/her to share, which can improve the interactions between the user and their friends. Matrix factorization is a straightforward approach to solve this prediction problem. By using users' observed influences, we could predict their influences on the upcoming events which have already some RSVPs ("yes" or "maybe"). However, as matrix factorization does not detect associations among the closely related items (i.e., users or events), the prediction performance of this approach might be poor. To improve the prediction accuracy, a potential approach is to integrate neighborhood methods with matrix factorization [15]. However, since the social influence matrix S is very sparse and the overlap values in S are few, it is hard to find reliable similar neighbors using the widely adopted similarity measures (e.g., Pearson correlation and Cosine similarity). Therefore, how to discover reliable event and user neighbors in S is a challenging problem.

In EBSNs, event content, event location and event organizer are the major components of an event which affect users' decisions in attending the event. Therefore, if two events are similar on these three aspects, we can consider these two events as similar events. Similarly, we consider, two users who have similar social influences on these three aspects are similar users. The rationale lies in that users' friends might have their own preferences on these three aspects, which determines whether they can be easily influenced by others. To this end, we propose an additional information based neighborhood discovery (AID) method to identify event and user neighborhood. To find the neighborhood of a targeted event e, we first capture three event-specific features (i.e., event content, event location and event organizer) and compute the similarities between e and other events on each feature. Then, we take the most similar events on each feature as a neighborhood of e. Such that, we obtain three neighborhood sets corresponding to the three features. Finally, a neighborhood aggregation strategy is proposed to derive the final neighborhood. In particular, in such strategy, we pick up the events contained in at least two neighborhood sets to make up the final neighborhood of e. Similarly, by considering three user-specific features (i.e., user influences on topics, user influences on regions and user influences on organizers), we find the neighborhood of a targeted user u in a similar way.

Based on the AID method, we present a hybrid collaborative filtering model, namely, Matrix Factorization with Event-User Neighborhood (MF-EUN) model, to predict users' social influences on upcoming events in EBSNs. The model incorporates both event-based and user-based neighborhood methods into matrix factorization and thus can take advantages of both matrix factorization and neighborhood methods. Model parameters are determined by minimizing the associated regularized squared error function through stochastic gradient descent. In summary, the major contributions of our work are listed as follows:

- We present a novel hybrid collaborative filtering model, namely, Matrix Factorization with Event-User Neighborhood (MF-EUN) model, which incorporates both event-based and user-based neighborhood methods into matrix factorization, for social influence prediction in EBSNs. To the best of our knowledge, this is the first attempt to define and solve the event-level social influence prediction problem in EBSNs.
- To find reliable similar neighbors, we propose an additional information based neighborhood discovery (AID) method by considering event-specific features (i.e., event content, event location and event organizer) and user-specific features (i.e., user influences on topics, user influences on regions and user influences on organizers) in EBSNs.
- We evaluate the performance of our prediction model on real-world datasets collected from DoubanEvent, which is the biggest event-based social network in China. Experimental results demonstrate the superiority of our MF-EUN model compared to several alternatives.

The remainder of this paper is organized as follows. We review the related work in Section 2. In Section 3, we give the definition of the social influence prediction problem, and show our model framework. In Section 4, the AID method is discussed in detail. We present our MF-EUN model for social influence prediction and show its solution in Section 5, followed by experimental evaluation in Section 6. Finally, Section 7 concludes the paper.

2. Related work

Social influence analysis: Influence is a potential factor which affects users' behaviors. Considerable works have been conducted to qualitatively validate the existence of influence [8,16]. Anagnostopoulos et al. [8] apply the statistical test (i.e., shuffle test) to identify whether influence is a source of social correlation using the time factor in a social network system. Chin et al. [2] investigate the user behaviour on attending offline events and find that social influences exist in EBSNs and users choose to attend an event partly because their friends will attend this event. Recently, some works have been proposed to quantify the social influence in different social networks [17,18,10]. Goyal et al. [18] propose both static and time-dependent model to capture influence probabilities from a log of past propagations. They consider user u influences user v on an action if they are friends and u performs this action before v. Zhang et al. [3] propose a unified metric to quantify the mutual user influence between social relation and geographical distance in location-based social networks (LBSNs). They evaluate the social influence of each user-pair in the participant set from a random walk perspective. Different from this work, we attempt to estimate each user's social influence to their friends on an event whose potential participants are unknown. Han et al. [19] propose a novel time series model for predicting the social influence of topics, which takes topical, social and geographic attributes into consideration. There are also some studies focusing on the social influence in fine-grained level. Tang et al. [14] analyze the topic-level social influence using the probabilistic model, in which they state, individuals' influences to others could vary greatly across different topics. Weng et al. [10] measure the topic-sensitive influences of users in Twitter by taking both the topical similarity between users and the