Expert Systems with Applications 42 (2015) 3653–3661

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Extended topology based recommendation system for unidirectional social networks

Melike Yigit [a], Bilal E. Bilgin [a], Adem Karahoca [b,*]

[a] Computer Engineering Department, Bahcesehir University, Istanbul, Turkey
[b] Software Engineering Department, Bahcesehir University, Istanbul, Turkey

ARTICLE INFO

Article history:
Available online 31 December 2014

Keywords:
Social networks
Recommendation systems
Friends of Friends
Extended topology-based recommendation

ABSTRACT

The power and importance of social networks increases day by day and many social networks such as "Facebook, Twitter, Weibo and others" have more than millions of users who communicate with each other. This opportunity is triggering researchers to do studies on the social network area and supports them to do improvements for recommendation systems (RS). In this study, we propose an extension to the topology based and Friends of Friends (FoF) recommendation systems by taking into account the user actions. The proposed approach (PA) firstly classifying the data has been set into four classes and secondly an equation was computed by using the relationship of users. Our model utilizes not only the relationship of the users but also many actions and many mentions of the users to generate the recommendation to users. We evaluate the performance over precision-recall graphs and receiver operating characteristic (ROC) curves. PA extended topology based and FoF algorithm results compared with the other alternative RSs. The benchmarking study show that recommendations of extending topology based RS performs better than the extended FoF and other well-known algorithms such as graph-based and Conceptual Fuzzy Set based algorithms.

1. Introduction

1.1. Background and motivations

A social network is a structure that connects individuals via bidirectional or unidirectional links. The purpose of social networks is analyzing the relations between social entities to understand the influence of entities on each other and to identify properties of these entities. Therefore, social networks receive many contributions from many fields such as statistics, graph theory and machine learning. Graph based approach is commonly used by especially social networks to connect people via links and to construct relational structures. Therefore, understanding relations among users is an important issue to obtain meaningful information from the social structure (Tiwari, 2013). In this respect, many studies have been done to analyze social networks by considering different topics such as online social networks, social graphs representation and online marketing (Nettleton, 2013). In this respect, first analyses of social networks were started by Georg Simmel (Wasserman & Faust, 1994). He focused on relationships between social entities and defined the structure of triads (McPherson, Smith-Lovin, & Cook, 2001). Then, studies have become more specific to analyze individual relations by combining social networks with graph theory and mathematical formulations (Boyd & Ellison, 2007; Liu & Aberer, 2013; Tavakolifard & Almeroth, 2012; Wasserman & Faust, 1994). In this way, online social networks become more popular by using graph theory to connect individuals by implementing social network services, an application to define social relations between people (Nettleton, 2013).

Nowadays advances in social networks, which enable personal communication with friends, family members, classmates, and other people, have sharply increased. Recently, the continuous increase in the number of users in social networks motivates researchers to carry out the research in this field. There are lots of social networking web pages, including Facebook, Twitter, Orkut, LinkedIn or MySpace and also many companies have social network website page to make advertisements about products. Companies and people creates their own pages called as profile page by signing up and then start adding people, joining groups. Social networks have many features to increase marketing and revenues of companies by finding people who have similar interests

* Corresponding author.
E-mail addresses: melike.yigit@stu.bahcesehir.edu.tr, melike.yigit87@gmail.com (M. Yigit), bilalerman.bilgin@bahcesehir.edu.tr, bilalerman.bilgin@hotmail.com (B.E. Bilgin), adem.karahoca@eng.bahcesehir.edu.tr, akarahoca@gmail.com (A. Karahoca).

http://dx.doi.org/10.1016/j.eswa.2014.12.043
0957-4174/

with companies' services and products. According to users' interests, new products and services are offered to customers via social marketing.

Many social networks are based on bidirectional relations. However, Twitter is both bidirectional and unidirectional that means a person who follows other person does not need to be followed by those users. In Twitter, people may share posts with other users which are called retweet and posts cannot exceed 140 characters (Armentano, Godoy, & Amandi, 2011). Relations between these people and huge amount of personal information lead to increase in demand for every recommendation system, a system which makes a prediction according to users' interest rate for items or for people. Therefore, recommender systems are included by information filtering systems (Ricci, Rokach, & Shapira, 2011) to extract information of users and to obtain meaningful information by mining millions of data. Recommender systems and social networks intersect with each other with the development of online social network applications such as Facebook, and Twitter (Nettleton, 2013). First social network application, SixDegrees.com was launched in 1997 and therefore, first intersection began between recommender systems and social networks around 1997 (Boyd & Ellison, 2007). Millions of people join such kinds of applications and share their personal information, valuable for marketing and for recommender systems. Within this context, studies were initiated to make recommendations to users.

Most of the studies about the recommendation field have some disadvantages. The first disadvantage of them is that these studies are based on the relationship information. However, in fact, there is much other information related to the user actions including mentions and actions that provide further information about the user. If this information is used while making the recommendation, users with similar interests can connect with each other. The second disadvantage of the existing recommendation studies is computing the similarity values by only using the local information of users and as a result users are recommended to each other if they do not share any friends. Therefore, two users cannot be recommended to each other to become friends when only the local information is used (Meo, Nocera, Terracina, & Ursino, 2011).

1.2. Contributions of this study

This study covers an implementation of a recommender system for Twitter and compares with existing recommender systems. Proposed approach (PA) is a formula for extending topology and FOF based recommendation systems. The PA is different from other RS approaches including Armentano et al. (2011) and Liu & Aberer, 2013. The PA considers the user actions that has not considered by other studies while making the recommendations yet. Using the user actions is important to solve the drawbacks of the recommendation systems and this is also shown from our performance analysis that our extending topology based approach achieves higher precision than the existing topology based approaches. The approach named SoCo as proposed by Liu and Aberer (2013) is also completely different from our proposed method. Because we do not use the contextual information like SoCo and we classify the social network information by using our classifier method, although the SoCo combines the social network information and contextual information by using different methods such as random decision trees algorithm, and person correlation.

In this study, we propose an extending topology based approach to recommend users by using the information related to users and by applying the classifier method. The main purposes of PA are handling the disadvantages of the existing studies. Following steps show the details of the presented approach: (1) classify the user relationship information into four classes to be suggested users that are similar (the neighborhood) to the target user. Then compute a final recommendation class including user ids to recommend the target user. (2) Calculate the ratio of followers to followees of the user to recommend the target user. (3) Compute the ratios of number of mentions and the number of actions to solve the disadvantages of existing studies that are commonly used local information of the users to make the recommendation. (4) Conduct experiments on two real datasets including the Weibo and Arizona State University (ASU)'s Twitter datasets to demonstrate performance of the extending topology based recommendation algorithm. (5) Implement Friends of Friends (FoF) with PA that is the simplest recommendation approach to compare the recommendation performance of proposed approach with FoF. By performance study with two different datasets, show the effectiveness of the proposed approach.

The remainder of this paper is organized as follows: Section 2 reviews the related works about the social network recommendation systems. The following section describes the material and methods that is detailed for making the recommendations. Section 4 presents the experiments and comparison results of PA. In Section 5, the conclusion is summarized the results and offer future research works in the recommendation systems.

2. Related works

With the progression of the social networks, a variety of recommendation studies have been proposed in the literature. This section reviews improvements in the recommendation algorithms chronologically which largely rely on user interests and the friendship relations between users.

Sakaguchi, Akaho, Takagi, and Shintani (2010) proposed a system for Twitter recommendations by using Conceptual Fuzzy Sets (CFS) that enables to explain meaning of a word with word sets. They offer recommendation according to users' interests that are obtained from CFS word vector. Users can connect with the persons who have similar interests or can find new interests with CFS base Twitter recommendation system by using word vector. The importance of this study is that recommendation has been done according to user following and interests information. The advantage of this study takes care of users' interests to make recommendations to users. However, it is not more efficient than other traditional methods because of word vector size and it must be tested by more testers to obtain more accurate results (Sakaguchi et al., 2010).

Another study on graph based Twitter recommendation system has been done by Silva, Tsang, Cavalcanti, and Tsang (2010). A new clustering method through Genetic Algorithm (GA) was proposed in this study. Oro-Aro social network data set was used instead of Twitter data set for recommendation system. Graphs were designed to show users and their connections. Users were represented by nodes and their connections were indicated with links. Proposed solution consists of two steps ordering and filtering. In filtering stage, nodes with higher possibility to recommend were extracted and thus, the number of processed nodes decreased. As for ordering stage, it has some features to put more interesting nodes in the top of the recommendation list. Even though all these steps are useful for recommendation, this system depends on users and therefore, GA, an adaptive algorithm was used for this study. According to experimental results, when Silva et al. (2010) used GA instead of FOF, acceptance of the recommendations increased from 72.22% to 77.69%. These results show the importance of the efficiency of using GA for recommender systems. However, experiments were done for Oro-Aro network that has small scale network. Therefore, this study must be performed in bigger networks to test its efficiency in dense network.