Physica A 398 (2014) 199–209

Contents lists available at ScienceDirect

Physica A

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

Compare two community-based personalized information recommendation algorithms

Yuan Wen [a], Yun Liu [a,*], Zhen-Jiang Zhang [a], Fei Xiong [a], Wei Cao [b]

[a] *Key Laboratory of Communication & Information Systems, Beijing Municipal Commission of Education, Beijing JiaoTong University, Beijing 100044, PR China*

[b] *China Information Technology Security Evaluation Center, Beijing 100085, PR China*

HIGHLIGHTS

- Community-based method is suitable in personalized recommendation.
- We compare two different approaches of communities' formation.
- Improved similarity formula can improve the diversity in recommendation.
- Non-strictly divided communities method has greater accuracy and diversity.

ARTICLE INFO

Article history:
Received 5 September 2013
Received in revised form 6 November 2013
Available online 27 December 2013

Keywords:
Personalized recommendation
Communities' formation
Similarity model
Accuracy of recommendation
Complex network
Link prediction

ABSTRACT

In recent years, bipartite-networks-based recommendations have attracted the attention of many researchers. Many of them are committed to improving the recommendation algorithms such as network-based inference (NBI) or probability spreading (ProbS). However, usually one or two parameters are tunable in these algorithms for optimizing the recommendation results. In these situations the optimal parameters are often applicable to specific data sets. Thus we consider using a community-based personalized recommendation, which has characteristics of simple and universal applicability. In this article, we investigate the effects of two different approaches to communities' formation based on traditional similarity formula and two improved similarity formulae proposed by us. The experimental results show that the approach of non-strictly divided communities presents greater accuracy and diversity in personalized information recommendations.

1. Introduction

In recent years, with the development of the technology, the network has played a very important role in people's daily lives [1–4]. With the further development of the network, more and more online applications appear and the network becomes more and more intelligent. For example, the network can analyze the degree of association between objects, and then predict any unknown link based on the already known data [5–8]. These predictions and recommendations are becoming widely used, such as Amazon's recommendations for books [9], and the TiVo digital video system for TVs and movies [10]. When people buy things on Amazon, they usually can get recommended products that are related to the things they bought. Sometimes the recommended products are exactly what customers want [11]. In other fields, such as the medical field, medical scientists try to discover the unknown link between drugs and a large number of targets.

* Corresponding author. Tel.: +86 15120074247.
E-mail addresses: 09111026@bjtu.edu.cn (Y. Wen), yunliubjtu@126.com, liuyun@bjtu.edu.cn (Y. Liu), zhangzhenjiang@bjtu.edu.cn (Z.-J. Zhang), xiongf@bjtu.edu.cn (F. Xiong), caow@itsec.gov.cn (W. Cao).

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2013.12.037

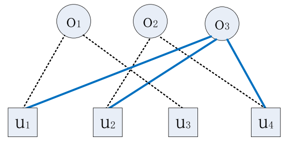

Fig. 1. Links in a user-object network.

It is often necessary to make choices to targets with proper drugs but without sufficient clinical trial data. At this point, recommendation systems play a significant role and provide link predictions for such situations [12,13].

There are already many successful recommendation methods that are designed with different characteristics [14]. One of the most widely used methods is based on the collaborative filtering technique, which is the best recommendation method in the past because of the simple mechanism and good recommendation results [15]. But recently a lot of efforts have been devoted to designing better recommendation algorithms based on bipartite graphs [16–20]. For example, Zhou et al. have introduced a bipartite-network-based recommendation called network-based inference (NBI) that successfully increases the recommendation effect [21,22]. Furthermore, some researchers use a scaling-based algorithm, which is independent of the length of recommendation list, to recommend objects on bipartite networks [23]. Some researchers use hybrid algorithms which combine the heat spreading (HeatS) and probability spreading (ProbS) algorithms to improve recommendation accuracy and diversity [24,25]. In the methods above, many researchers focus on adjusting the tunable parameters to optimize the recommendation system. However, these optimized parameters are usually suitable only for specific data sets, without extensive availability. We consider the use of community-based recommendation because it is simple and universally applicable.

In this article, we compare the effects of two approaches to communities' formation. We improve the precision evaluation formula in order to make the criteria more reasonable. Under two different approaches to communities' formation, we test and analyze the different recommendation results based on three different similarity formulae. At last the test results show that the approach of non-strictly divided communities can generate more accuracy and recommendation personalization.

2. Methods

2.1. The idea

This article is motivated by the idea that similar users often have similar interests. Furthermore, friends may like same objects. For example, there exist many reasons that two persons can become a pair of friends, but the most important thing is that they have similar interests. Similarly if two users like similar objects, they can be regarded as a pair of friends; if several users like similar objects, these persons constitute a community. In the bipartite network, a community with enough large number of users can recommend objects to its members according to most other members' collections.

Fig. 1 shows the links in a user-object network. The solid lines represent that u_1, u_2 and u_3 constitute a community. The dashed lines represent that u_1–u_3 and u_2–u_4 are two pairs of friends. Broadly speaking, a pair of friends can be seen as a community.

The main task of recommendation of any user is to predict the possible links based on other users' collections in a community. According to the similarity between users, we can decide which users can constitute a community. In this article, we improve the similarity formula and use different methods of communities' formation to test which one is the best for constituting communities.

2.2. Improving the similarity formula

(1) Traditional similarity formula (TS)

Generally, recommendation systems are designed on the basis of bipartite user-object graph $G(u, o, e)$, which contains users $U = \{u_1, u_2, \ldots, u_i, \ldots, u_m\}$, objects $O = \{o_1, o_2, \ldots, o_p, \ldots, o_n\}$, and links $e = \{e_{ip} : u_i \in u, o_p \in o\}$. A link is drawn between u_i and o_p if user i has collected object p (when the rating is no less than 3 if the scale is from 1 to 5). The similarity between two users u_i and u_j can be calculated as [26]:

$$Sim_{TS}(u_i, u_j) = \frac{|C(u_i) \cap C(u_j)|}{|C(u_i) \cup C(u_j)|}. \quad (1)$$

Here $C(u_i)$ and $C(u_j)$ indicate the numbers of collected objects by user i and user j respectively. Similarity between user i and user j strongly depends on the number of commonly collected objects by two users, but it is inversely proportional to the sum of their respective numbers of collections as Eq. (1) shows. In fact, when calculating the similarity between users, it should not have any relation with users' respective numbers of rated objects. In many real networks, how many movies a user rates only has a relation with the minimum number of rating required by rating system. For example in a data set called Movielens (described below in detail), every user is asked to rate no less than 20 movies, but which movie a user rates is a random thing. Thus we propose another similarity formula which has a relation only with commonly rated objects by users.