PHYSA: 17553 Model 3Gsc pp. 1–9 (col. fig: NIL)

ARTICLE IN PRESS

Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

A vertex similarity index for better personalized recommendation

Ling-Jiao Chen [a,b], Zi-Ke Zhang [c], Jin-Hu Liu [a,b], Jian Gao [a,b,*], Tao Zhou [a,b]

[a] CompleX Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

[b] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

[c] Institute of Information Economy, Hangzhou Normal University, Hangzhou 310036, People's Republic of China

HIGHLIGHTS

- A vertex similarity index CosRA is proposed, which combines both advantages of cosine index and resource-allocation index.
- Results on real rating data suggest the overall better performance of CosRA-based recommendation method.
- Further experiments show that CosRA index is parameter-free with a significant advantage in real applications.

ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form 10 August 2016
Available online xxxx

Keywords:
Vertex similarity
Recommender systems
Personalized recommendations
Information filtering

ABSTRACT

Recommender systems benefit us in tackling the problem of information overload by predicting our potential choices among diverse niche objects. So far, a variety of personalized recommendation algorithms have been proposed and most of them are based on similarities, such as collaborative filtering and mass diffusion. Here, we propose a novel vertex similarity index named CosRA, which combines advantages of both the cosine index and the resource-allocation (RA) index. By applying the CosRA index to real recommender systems including MovieLens, Netflix and RYM, we show that the CosRA-based method has better performance in accuracy, diversity and novelty than some benchmark methods. Moreover, the CosRA index is free of parameters, which is a significant advantage in real applications. Further experiments show that the introduction of two turnable parameters cannot remarkably improve the overall performance of the CosRA index.

1. Introduction

The development of the Internet and e-commerce makes our lives more convenient as billions of products are available online [1]. Meanwhile, the problem of information overload plagues us everyday as it is much harder to dig out relevant objects than ever [2]. Thus far, personalized recommendation was thought to be the most promising way to efficiently solve the problem of information overload [3,4]. Personalized recommendation benefits both buyers and sellers, and it is now playing an increasing role in our online social lives. Many online platforms (Amazon, eBay, AdaptiveInfo, Taobao, etc.) have introduced personalized recommendation systems [5], which predict users' potential choices by analyzing historical behaviors of users, attributes of objects, and so on [6]. For example, Amazon.com recommends books by analyzing users'

* Corresponding author at: CompleX Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China.
E-mail address: gaojian08@hotmail.com (J. Gao).

http://dx.doi.org/10.1016/j.physa.2016.09.057
0378-4371/

Please cite this article in press as: L.-J. Chen, et al., A vertex similarity index for better personalized recommendation, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.09.057

purchase records [7], and AdaptiveInfo.com recommends news by using users' reading histories [8]. In recent years, personalized recommendation has found wide applications [9] in recommending movies [10,11], videos [12], research articles [13], driving routes [14], locations [15,16] and so on.

So far, a variety of personalized recommendation algorithms have been proposed [17–21], among which user-based (UCF) and item-based collaborative filtering (ICF) are the most representative ones [22]. UCF and ICF are respectively based on the weighted combination of similar users' opinions and the similarity between items [23]. Recently, many diffusion-based algorithms are proposed by introducing some physical dynamics into the recommender systems, such as mass diffusion (MD) [24] and heat conduction (HC) [25]. The simplest version of MD can be considered as a two-step resource-allocation process in bipartite networks [26]. Later, Zhou et al. [27] and Jia et al. [28] proposed two algorithms by giving new strategies in the initial resource distribution, Zhou et al. [29] proposed a hybrid method that combines both MD and HC, Lü et al. [30] proposed a preferential diffusion method by considering node weights in redistributing resources, and Liu et al. [31] proposed a weighted heat conduction algorithm by considering edge weighting. Reviews of previous literatures can be found in Refs. [17,18].

Essentially, the aforementioned collaborative filtering and diffusion-based methods are based on similarities [32,33]. In collaborative filtering, the most commonly used index is cosine similarity [34–36]. However, it strongly tends to recommend popular objects, resulting in accurate yet less-diverse recommendations [37]. In diffusion-based methods, the diffusion is indeed a resource-allocation process, and the node similarity is characterized by the resource-allocation (RA) index [38,39]. The RA index gives high priority to assign resources to large-degree nodes, which leads to high accuracy but low diversity of MD [40]. In fact, the cosine index and RA index are complementary to each other, and thus to combine the two can possibly improve the overall performance. How to design a suitable similarity index for better recommendation is still an open issue and such index can be applied in characterizing many network structures and functions [41,42].

In this paper, we propose a vertex similarity index, named CosRA, for better personalized recommendation. Based on the CosRA index which combines advantages of both the cosine index and the RA index, we further propose a personalized recommendation algorithm. Extensive experiments on four real data sets suggest that the CosRA-based method performs better in accuracy, diversity and novelty than some benchmark methods. Moreover, we provide some insights on the mechanism of the CosRA index and extend it to a more general form by introducing two turnable parameters. Interestingly, results suggest that the original CosRA index is almost optimal, and its effectiveness cannot be remarkably improved by adjusting the parameters. Such feature is significant since a parameter-free index is more applicable than a parameter-dependent index. Our work sheds lights on the importance of a suitable vertex similarity index in enhancing the overall performance of personalized recommendation.

2. Vertex similarity index

A recommender system can be naturally described by a user–object bipartite network $G(U, O, E)$, where $U = \{u_1, u_2, \ldots, u_m\}$, $O = \{o_1, o_2, \ldots, o_n\}$ and $E = \{e_1, e_2, \ldots, e_z\}$ are sets of users, objects and links, respectively. To distinguish object-related and user-related indices, we respectively use Greek and Latin letters for them. Meanwhile, the bipartite network $G(U, O, E)$ can be naturally represented by an adjacency matrix A, whose element $a_{i\alpha} = 1$ if there is a link connecting node U_i and node O_α, *i.e.*, user i has collected object α, otherwise $a_{i\alpha} = 0$. The main purpose of recommendation algorithms is to provide a target user with a ranking list of his uncollected objects. For user i, the recommendation list with length L is denoted as o_i^L. That is to say, o_i^L is a set of L objects with the highest recommendation scores for user i.

First, we introduce two widely used similarity indices in recommendation algorithms, namely, the cosine index and the RA index. Taking two objects α and β as an example, the cosine index between them is defined as

$$S_{\alpha\beta}^{Cos} = \frac{1}{\sqrt{k_\alpha k_\beta}} \sum_{i=1}^{m} a_{i\alpha} a_{i\beta}, \tag{1}$$

where k_α and k_β are the degrees of objects α and β, respectively. In fact, the cosine index measures the similarity between two objects' rating vectors of an inner product space. Meanwhile, the resource-allocation process is equivalent to the one-step random walk in the user–object bipartite networks starting from the common neighbors [39]. Specifically, the RA index between two objects α and β is defined as

$$S_{\alpha\beta}^{RA} = \sum_{i=1}^{m} \frac{a_{i\alpha} a_{i\beta}}{k_i}, \tag{2}$$

where k_i is the degree of user i. Indeed, the RA index is the entry of the transformation matrix in the simplest version of the MD process [26], which is a variant on an earlier version of the probabilistic spreading algorithm [29].

Then, we introduce the proposed CosRA similarity index. On the one side, both of the degrees of the two objects should be considered, and the effect of popular objects should be restricted in calculating the similarity. On the other side, the effect of small-degree users should be enhanced to decrease the advantage of large-degree nodes in the network. Based on these