Physica A 421 (2015) 44–53

Contents lists available at ScienceDirect

Physica A

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

Information filtering via balanced diffusion on bipartite networks

Da-Cheng Nie [a], Ya-Hui An [a], Qiang Dong [a,b,*], Yan Fu [a,b], Tao Zhou [a,b]

[a] Web Sciences Center, School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 611731, PR China

[b] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, PR China

HIGHLIGHTS

- The optimal hybrid algorithm of MD and HC processes is investigated.
- BD algorithm gives recommendations with superior accuracy and diversity.
- BD algorithm recommends more unpopular objects to users.

ARTICLE INFO

Article history:
Received 5 May 2014
Received in revised form 18 September 2014
Available online 12 November 2014

Keywords:
Recommendation algorithm
Mass diffusion
Heat conduction
Bipartite network

ABSTRACT

The recent decade has witnessed the increasing popularity of recommender systems, which help users acquire relevant commodities and services from overwhelming resources on Internet. Some simple physical diffusion processes have been used to design effective recommendation algorithms for user–object bipartite networks, such as mass diffusion (MD) and heat conduction (HC) algorithms, which have different advantages respectively on accuracy and diversity. In this paper, we explore how to combine MD and HC processes to get better recommendation performance and propose a new algorithm mimicking the hybrid of MD and HC processes, named balanced diffusion (BD) algorithm. Numerical experiments on three benchmark data sets, *MovieLens*, *Netflix* and *RateYourMusic*, show that BD algorithm outperforms three typical diffusion-like algorithms on the three important metrics, accuracy, diversity and novelty. Specifically, it not only provides accurate recommendation results, but also yields higher diversity and novelty in recommendations by accurately recommending unpopular objects.

1. Introduction

Web 2.0 and its applications have achieved significant developments in the past few years, which bring us more convenience as well as overwhelm us with the information ocean on Internet [1]. This is the so-called *Information Overload* problem [2]. Nowadays, online shopping becomes more and more popular in our daily life. For instance, there are millions of books (e-books) on Amazon.com, and the turnover of Taobao.com exceeded 35 billion RMB (China's currency, about 6 billion US dollars) on the shopping festival day of Nov 11, 2013 [3]. In this case, we find that it is very difficult to choose the relevant ones from countless candidates on these e-commerce websites, and thus an automatic way that can help us to

* Corresponding author at: Web Sciences Center, School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 611731, PR China.
E-mail address: dongq@uestc.edu.cn (Q. Dong).

http://dx.doi.org/10.1016/j.physa.2014.11.018
0378-4371/

make right decision under the information overload circumstance becomes a significant issue in both academic and industrial communities.

The emergence of search engines partially alleviates this dilemma; a user inputs the keywords and then search engines return the results accordingly. However, search engines always return the same results to different users if they key in the same words. When users resort to search engines, they have already known what they want and can easily find keywords to describe it. But in the most occasions, users do not know what they really want, or it is hard to find appropriate words to describe it. Therefore, recommender systems have been designed to solve this problem. We can see that in recent years, recommender systems have greatly promoted the development of E-business, and vice versa [4].

Collaborative filtering (CF) [5–12] is the most frequently used technology in recommender systems, which makes use of the object collecting history of users to predict the potential objects of interest to the target user, including user-based CF [10] and object-based CF [8,12]. However, the original CF methods give recommendation by computing the similarity between user preferences, which will make the recommendation results more and more similar among users [13,14]. What is more, CF algorithms cannot deal with the cold start problem [15], i.e., when a new user or object is added to the system, it is difficult to obtain recommendations or to be recommended. Therefore, the content-based [16] methods have been proposed to solve this problem, which generate recommendation results by computing the similarity between user profiles, but user profiles are usually difficult to acquire due to the constraint of information retrieval techniques. Generally speaking, CF methods and content-based methods will generate similar recommendation results with poor diversity and novelty. For the recent developments of recommender systems, the readers are referred to the comprehensive review by Lü et al. [17].

To improve the diversity and novelty of recommendation results, many other personalized recommendation algorithms have been proposed, including trust-aware methods [18,19], social-impact methods [20–22] and tag-aware methods [23]. Recently, based on two physical processes, the mass diffusion (MD) algorithm and heat conduction, many effective recommendation algorithms have been designed on user–object bipartite networks, including the MD algorithm and HC algorithm [14,24,25]. The MD algorithm is essentially a resource redistribution process between objects via neighboring users [24,26], which achieves high accuracy but low diversity. Zhang et al. [27] introduce a voting system in the diffusion process to get better recommendation results. The HC algorithm is like a heat conduction process from objects to neighboring users and back to objects again, which has high diversity but low accuracy. Ideally, a good recommendation algorithm should exhibit both of high accuracy and high diversity.

In Ref. [14], Zhou et al. proposed an algorithm to nonlinearly combine the MD and HC processes (HHP for short), which solves the apparent diversity–accuracy dilemma of recommender systems. Liu et al. [28] proposed a biased heat conduction (BHC for short) algorithm, which simultaneously enhances the accuracy and diversity by decreasing temperatures of small-degree objects in the heat conduction process. Lü et al. [29] proposed a preferential diffusion algorithm, taking into account the heterogeneity of users' degrees. All of the above-mentioned algorithms derived from MD and/or HC processes demonstrate good accuracy and diversity. However, the strategy of combining MD and HC processes into one recommender system to get the optimal accuracy and diversity remains to be an open problem.

In this paper, we explore how to combine MD and HC processes to get better recommendation performance and propose a new algorithm mimicking the hybrid of MD and HC processes, named balanced diffusion (BD) algorithm. Numerical experiments on three benchmark data sets, *MovieLens*, *Netflix* and *RateYourMusic*, show that the BD algorithm outperforms three typical diffusion-like algorithms on the three important metrics, accuracy, diversity and novelty. Specifically, it not only provides accurate recommendation results, but also yields higher diversity and novelty in recommendations by accurately recommending unpopular objects.

2. Methods

A recommender system can be represented by a bipartite network $G(U, O, E)$, where $U = \{u_1, u_2, \ldots, u_m\}$, $O = \{o_1, o_2, \ldots, o_n\}$, and $E = \{e_1, e_2, \ldots, e_q\}$ represent the m users, n objects, and q links between the m users and n objects, respectively. The system could be fully described by an adjacency matrix $A = \{a_{l\alpha}\}_{m,n}$, where $a_{l\alpha} = 1$ if there exists a link $e_{l\alpha}$ between user u_l and object o_α and $a_{l\alpha} = 0$ otherwise.

We assume that a user collects an object because he/she likes it, then the essential task of a recommender system becomes to generate a ranking list of the target user's uncollected objects. All the recommendation algorithms inspired by diffusion-like process work by initially assigning all the objects a certain amount of resources, denoted by the vector $\mathbf{f}$ (where f_α is the resource of object o_α), and then reallocating these resources via the transformation $\mathbf{f}' = W\mathbf{f}$, where W is called the resource transfer matrix.

The original recommendation algorithm mimicking the mass diffusion process is called the MD algorithm, also referred to as Network-Based Inference (NBI) [24] and ProbS [14]. For a target user u_l, the initial resource vector $\mathbf{f}$ on the objects is defined as $f_\alpha = a_{l\alpha}$, where $a_{l\alpha} = 1$ if user u_l has collected object o_α, otherwise $a_{l\alpha} = 0$. The element $w_{\alpha\beta}$ of the transfer matrix W is written as

$$w_{\alpha\beta} = \frac{1}{k_{o_\beta}} \sum_{l=1}^{m} \frac{a_{l\alpha} a_{l\beta}}{k_{u_l}}, \tag{1}$$

where $k_{o_\beta} = \sum_{i=1}^{m} a_{i\beta}$ and $k_{u_l} = \sum_{r=1}^{n} a_{lr}$ denote the degrees of object o_β and user u_l, respectively.