Physica A 390 (2011) 2414–2420

Contents lists available at ScienceDirect

Physica A

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

Information filtering via weighted heat conduction algorithm

Jian-Guo Liu [a,b,*], Qiang Guo [a], Yi-Cheng Zhang [a,c]

[a] Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China
[b] CABDyN Complexity Centre, Säid Business School, University of Oxford, Park End Street, Oxford OX1 1HP, United Kingdom
[c] Department of Physics, University of Fribourg, Chemin du Musée 3, CH-1700 Fribourg, Switzerland

ARTICLE INFO

Article history:
Received 26 September 2010
Received in revised form 10 February 2011
Available online 2 March 2011

Keywords:
Recommender systems
Bipartite networks
Heat conduction

ABSTRACT

In this paper, by taking into account effects of the user and object correlations on a heat conduction (HC) algorithm, a weighted heat conduction (WHC) algorithm is presented. We argue that the edge weight of the user–object bipartite network should be embedded into the HC algorithm to measure the object similarity. The numerical results indicate that both the accuracy and diversity could be improved greatly compared with the standard HC algorithm and the optimal values reached simultaneously. On the Movielens and Netflix datasets, the algorithmic accuracy, measured by the average ranking score, can be improved by 39.7% and 56.1% in the optimal case, respectively, and the diversity could reach 0.9587 and 0.9317 when the recommendation list equals to 5. Further statistical analysis indicates that, in the optimal case, the distributions of the edge weight are changed to the Poisson form, which may be the reason why HC algorithm performance could be improved. This work highlights the effect of edge weight on a personalized recommendation study, which maybe an important factor affecting personalized recommendation performance.

1. Introduction

The amount of information in the world is increasing far more quickly than our ability to process it. All of us have known the feeling of being overwhelmed by the number of new books, journal articles, and conference proceedings coming out each year [1–5]. Technology has dramatically reduced the barriers to publishing and distributing information. Being an effective tool to address this problem, the recommender system has caught an increasing amount of attention from researchers to engineers, and it has become an essential issue in Internet applications such as e-commerce systems and digital library systems [6]. Recommender systems are one of the fastest growing segments of the Internet economy today. They help reduce information overload and provide customized information access for targeted domains. It is being widely used in many application settings to suggest products, services, and information items to potential consumers. A personalized recommender system includes three parts: data collection, model analysis and the recommendation algorithm, among which the algorithm is the core part. Motivated by the significance in economy and society, various kinds of algorithms have been proposed, including collaborative filtering (CF) approaches [7–15], content-based analyses [16,17], tag-based algorithms [18–20], link prediction approach [21], hybrid algorithms [22–24], and so on. For a review of current progress, see Refs. [4,5] and the references therein.

Very recently some physical dynamics, such as the heat conduction (HC) process [25] and probability diffusion method [26] have been successfully applied in personalized recommendation. In the standard HC algorithm, the objects one target

* Corresponding address: Business School, University of Shanghai for Science and Technology, No. 516 Jungong Road, Yangpu District, 200093 Shanghai, PR China.
E-mail address: liujg004@ustc.edu.cn (J.-G. Liu).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2011.02.023

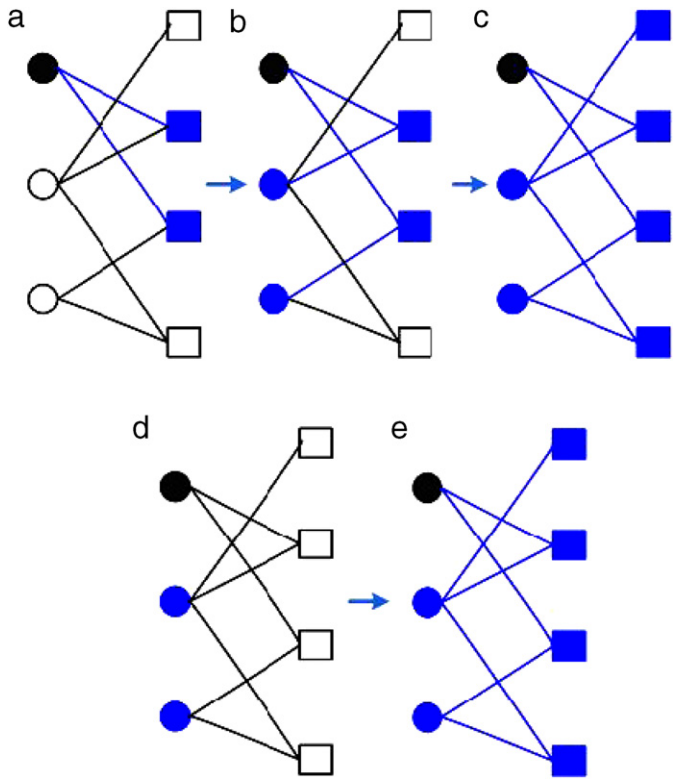

Fig. 1. (Color online) Illustration of network-based recommendation (a–c) and collaborative filtering (d, e) algorithms on the user–object bipartite network. Users and objects are shown as circles and squares, with the target user indicated by the black circle. The network-based algorithm supposes the objects one user has collected are able to recommend new objects (a–c), while the collaborative filtering algorithm is implemented based on the neighbor users' opinions (d, e).

user has selected were set as the heat resource with constant temperature 1, all other unselected object temperatures are set as 0. The heat is diffused from the object set to user set, then it flows back again. To one heat resource with temperature 1, the unselected objects' final temperature is defined as the direct similarity to the resource node. Since the HC algorithm is implemented based on user–object bipartite networks, it's also called a *network-based algorithm*. Fig. 1 demonstrated the main idea of the network-based (a–c) and CF (d, e) algorithms. Combining the similar neighbors' opinions, CF algorithms recommend new objects to the target user. Unlike the standard CF algorithm, the network-based algorithms suppose that the objects one user has collected/selected have the power to recommend new objects. The numerical results have indicated that these physical approaches have both high efficiency and low computational complexity [9–12,25–29]. However, all of the above algorithms are implemented in the topological user–object bipartite networks, in which the edge between a user and an object indicates that the user is fond of this object. In other words, all of the objects and users with far different degrees have been treated equally. By introducing the collaborative similarity, Shang et al. [30] found the correlation between users' degree and taste diversity, which indicates that the edge contains unequal information in a bipartite network. Liu et al. [14] investigated the edge weight effect on CF algorithm and found that, in the optimal case, the edge weight obeys the power-law distribution. Inspired by the above works, we argue that the edge weight should be taken into account to analyze the effects on HC algorithm performance. Since each edge contains the target user's special interest or habit information, the edge diversity could approximately reflect the user's tastes.

In this paper, we suppose that the preference level endowed in each edge relies on the degrees of the two connected nodes, which should be taken into account to improve HC algorithm performance. Implementing the HC algorithm on the weighted user–object bipartite network, a modified algorithm named the weighted HC algorithm (WHC) is present. The numerical results indicate, when the edge weight obeys the Poisson distribution, both the accuracy and diversity could be improved greatly and reach the optimal values approximately simultaneously.

2. Heat conduction recommendation algorithm

A recommender system consists of users and objects, and each user has collected some objects. Denote the object-set by $O = \{o_1, o_2, \ldots, o_m\}$ and user-set by $U = \{u_1, u_2, \ldots, u_n\}$, the recommender system can be fully described by a bipartite network with $m + n$ nodes, where an object is connected with a user if and only if the object has been collected by this user. Connections between any two users or two objects are not allowed. Based on the bipartite user–object network,