Physica A 469 (2017) 397–404

Contents lists available at ScienceDirect

Physica A

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

Personalized recommendation based on preferential bidirectional mass diffusion

Guilin Chen [a], Tianrun Gao [b], Xuzhen Zhu [a,*], Hui Tian [a], Zhao Yang [a]

[a] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China
[b] Beijing Information Science and Technology University, Beijing, 100192, China

HIGHLIGHTS

- Considering bidirectional mass diffusion in recommendation algorithm.
- Penalizing objects' popularity in bidirectional mass diffusion.
- An unbalanced punishment on popularity in bidirectional diffusion is shown.
- The diffusion is unsymmetric with the one originating from prior selections stronger.

ARTICLE INFO

Article history:
Received 20 April 2016
Received in revised form 7 July 2016
Available online 22 November 2016

Keywords:
Personalized recommendation
Preferential bidirectional diffusion
Bipartite network

ABSTRACT

Recommendation system provides a promising way to alleviate the dilemma of information overload. In physical dynamics, mass diffusion has been used to design effective recommendation algorithms on bipartite network. However, most of the previous studies focus overwhelmingly on unidirectional mass diffusion from collected objects to uncollected objects, while overlooking the opposite direction, leading to the risk of similarity estimation deviation and performance degradation. In addition, they are biased towards recommending popular objects which will not necessarily promote the accuracy but make the recommendation lack diversity and novelty that indeed contribute to the vitality of the system. To overcome the aforementioned disadvantages, we propose a preferential bidirectional mass diffusion (PBMD) algorithm by penalizing the weight of popular objects in bidirectional diffusion. Experiments are evaluated on three benchmark datasets (Movielens, Netflix and Amazon) by 10-fold cross validation, and results indicate that PBMD remarkably outperforms the mainstream methods in accuracy, diversity and novelty.

1. Introduction

With the rapid development of Internet technology [1], the explosive expansion of information has led to information overload [2] preventing us from efficiently retrieving information we are interested in from massive data, which has become a serious obstacle to the development of Internet technology. Facing such a situation, personalized recommendation technology [3] draws significant attention as they could alleviate the dilemma by helping users to obtain the target information quickly. So far, recommendation system has been used in many fields [4–6], but it is still difficult to meet

* Corresponding author.
E-mail address: zhuxuzhen@bupt.edu.cn (X. Zhu).

http://dx.doi.org/10.1016/j.physa.2016.11.091
0378-4371/

the increasing needs of commodity information service. Investigation shows its reasons are various, and one of the most prominent is recommendation algorithm's adaptability, including accuracy, diversity and novelty.

Fruitful personalized recommendation algorithms have been proposed according to different application scenarios and conditions [7,8]. However each method has its own weakness: Global ranking method (GRM) [8,9] will degrade user experience as it provides the same recommendation list to all users without considering users' different preferences; Collaborative filtering (CF) [10–14] is based on similarity and consequently suffers from a popularity bias problem; high computational complexity results in spectral analysis's [15] inadequacy to deal with huge-size systems; and so on. Some scholars study recommendation system based on complex networks, where recommendation system is described as bipartite network. Algorithms based on bipartite network can effectively improve the adaptability and the recommendation performance of the algorithm [16]. Facing recommendation system's long-standing challenge, theoretical physics provides us with some powerful tools such as mass diffusion [17–25] and heat conduction [26–28], which has been used to design effective recommendation algorithms on bipartite network.

So far, most of mass-diffusion-based recommendation algorithms have focused on unidirectional diffusion from collected objects to uncollected objects to distinguish the similarity between two objects [16,20], which is actually causal and biased. In physical dynamics, the essence of mass diffusion should lie in consistently powerful bidirectional diffusion abilities, so it will be unreasonable to make recommendations disregarding the opposite direction mass diffusion from uncollected objects to collected objects, which will inevitably lead to overestimated or underestimated similarity. Following this intrinsic diffusion rule, [29,30] take into account the mass diffusion from uncollected objects to collected objects as well.

Additionally, traditional mass-diffusion-based recommendation algorithms are biased towards the popular objects making the popular objects tend to be recommended frequently [16,18,20,22], which will not necessarily promote the accuracy but stealthily hurt recommendation diversity and novelty. While the recommendation performance is improved in [29,30], they still suffer from this problem. Much more practically, to recommend unpopular objects is more meaningful than to recommend popular objects. Therefore, the weight of large degree objects should be suppressed in the diffusion processes to tackle this problem, so that, diverse recommendation is obtained.

As mentioned above, based on the perspective of physical dynamics and motivated by enhancing the algorithm's ability to find unpopular and niche objects, we propose a preferential bidirectional mass diffusion (PBMD) algorithm on bipartite network by penalizing the weight of popular objects to suppress the bidirectional diffusion of popular objects. Two parameters are used to determine the appropriate extent of punishment for popular objects in bidirectional diffusion and make the algorithm flexible and adaptive. Through extensive experiments on three benchmark datasets (Movielens, Netflix and Amazon), results indicate the effectiveness of PBMD in comparison with mainstream methods.

2. Proposed method

Suppose that a recommendation system contains m users and n objects, and each user has collected some objects. Let $U = \{u_1, u_2, \ldots, u_m\}$ and $O = \{o_1, o_2, \ldots, o_n\}$ represent the users and objects respectively. The recommendation system can be described as an unweighted undirected user–object bipartite network (BN) with $m + n$ nodes. If o_j is collected by u_i, there is a link between u_i and o_j, and the corresponding element a_{ij} in the adjacent matrix A is set as 1, otherwise 0. Mathematically speaking, the essential task of a recommendation system is to generate a ranking list from the target user's uncollected objects according to the collected ones. The top L objects are recommended to this user, with L denoting the length of the recommendation list.

Through mimicking the mass diffusion resource-allocation process, referred to as Network-Based Inference (NBI) [17], where each object distributes its initial resource (the initial resource quantity is set as 1) equally to all the users who have collected it, and then each user reallocates what he/she has received to all the objects he/she has collected (also equally), the transfer weight w_{ij} (the proportion of initial resource o_j eventually distributes to o_i) can be defined as

$$w_{ij} = \frac{1}{k(o_j)} \sum_{l=1}^{m} \frac{a_{li}a_{lj}}{k(u_l)} \tag{1}$$

where $k(u_l)$ and $k(o_j)$ are the degrees of u_l and o_j, respectively representing the number of objects collected by u_l and the number of users who have collected o_j. A simple example about how to calculate transfer matrix is shown in Fig. 1.

For a target user u_j, let $R^0 = (r^0_{ij})_{n\times m}$ denote the initial resource matrix, thus r^0_{ij} can be written as $r^0_{ij} = a_{ji}$. That is to say, if o_i is collected by u_j then it has one unit resource, otherwise 0. Then the final resource distribution for each user can be represented by a matrix as below

$$R = WR^0 \tag{2}$$

where the jth column of matrix $R = (r_{ij})_{n\times m}$ represents the probability vector of u_j to collect arbitrarily object and the top-L values of uncollected objects will be recommended to u_j.

Many studies have shown that only accounting for the unidirectional mass diffusion ability from collected objects to uncollected objects, Eq. (1) will lead to an asymmetrical similarity estimation. In physical dynamics, mass diffusion intrinsically should have equally powerful bidirectional diffusion abilities. To handle the drawback of the unidirectional