Physica A 508 (2018) 1–9

Contents lists available at ScienceDirect

Physica A

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

Symmetrical information filtering via punishing superfluous diffusion

Xuzhen Zhu [a], Hui Tian [a], Tianqiao Zhang [a,b,*]

[a] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, PR China

[b] School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, 450046, PR China

HIGHLIGHTS

- Mass diffusion based information filtering models intrinsically suffer superfluous diffusion, transferring redundant preferences to the object.
- Besides, previous mass diffusion based researches still preserve the asymmetrical diffusion mode from the collected object to the uncollected objects, leading to biased similarity estimation.
- We assume that the superfluous diffusion should be symmetrically punished and propose a symmetrical punishment model on superfluous diffusion.
- Experiments on Movielens and Netflix show more accurate, diverse, and personalized recommendation performances compared with mainstream models.

ARTICLE INFO

Article history:
Received 24 December 2017
Received in revised form 8 April 2018

Keywords:
Information filtering
Personalized recommendation
Symmetrical punishment
Superfluous diffusion

ABSTRACT

For niche recommendation, information filtering has attracted much attention from various fileds. Especially, mass diffusion based models behave prominently. Nevertheless, these models intrinsically suffer superfluous diffusion, transferring redundant preferences to the object and damaging the accuracy, diversity, and personalization of recommendation. Besides, we discover that the symmetrical diffusion can effectively improve recommendation performances. Thus, we assume that the superfluous diffusion should be symmetrically punished. Hence, we propose a symmetrical punishment model on superfluous diffusion for accurate information recommendation. Extensive experiments on two data sets *Netflix* and *Movielens* show that our proposed model outperforms mainstream indices remarkably in accuracy, diversity and personalization.

1. Introduction

The development of Internet and smart mobile devices makes our lives more convenient. Through various kinds of e-commerce systems, people are gradually accustomed to online reading news, watching movies, shopping, making friends, and so on. Meanwhile, they are also exposed to overloaded information provided by these online systems. Facing the ocean of information, people cannot effectively retrieve their interesting objects. For that, models of information filtering emerges to break this dilemma. Users are actively recommended with niche objects according to their preferences captured from online systems, such as Amazon [1], Twitter [2], AdaptiveInfo [3], TiVo [4].

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

https://doi.org/10.1016/j.physa.2018.05.065
0378-4371/

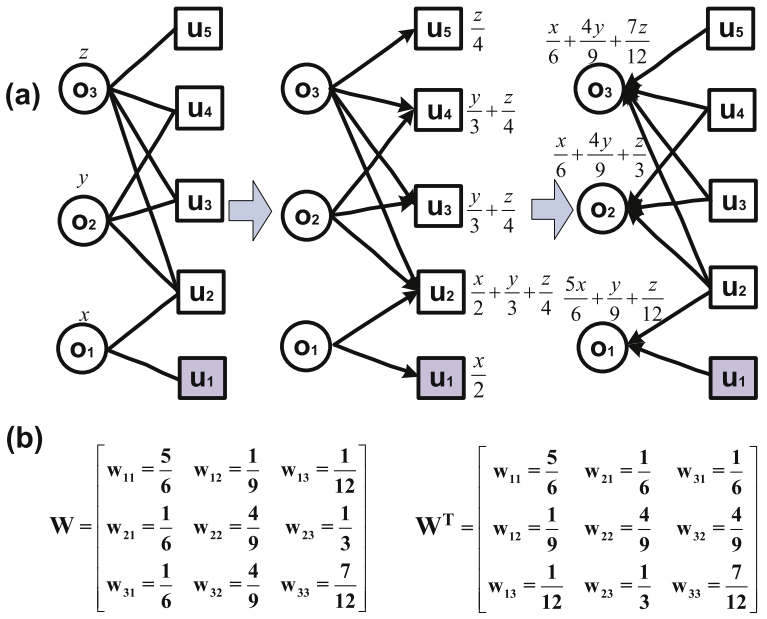

Fig. 1. Illustration of mass diffusion in NBI model. (a) The schema of mass diffusion based on BN. (b) Diffusion proportion matrices W and inverse matrix W^T obtained from NBI. The element w_{ij} means the diffusion proportion weight from the collected object o_j to the uncollected object o_i.

Drawn by abundant interests in economy and society, the investigations on models of information recommendation have been successfully focused by experts in various fields, ranging from physics to computer science [5–7]. Especially, based on bipartite network (BN), some physicists conceive information filtering models via heat conduction [8] and mass diffusion [9]. The heat conduction model has been further extended by considering biased heat conduction [10], weighted heat conduction [11], effect of time window on heat conduction [12] and congestion in recommender systems [13]. On the other hand, mass diffusion based information filtering obtains more attentions in terms of initial configuration [14], multi-channel diffusion [15], eliminating redundant correlations [16], users' tastes [17], high-order correlations [18], preferential diffusion [19], clustering coefficient [20], information core [21], code-start items [22], and consistent recommendation [23–25]. Further, for simultaneously modifying the diversity of mass diffusion model, hybrid model with heat conduction has been apparently proposed [26]. According to the hybrid idea, many subsequential models are further developed by exploring ground user [27], similarity-preferential diffusion [28] and balanced hybrid recommendation [29]. Moreover, stability of recommendation algorithms based on mass diffusion and heat conduction is also investigated [30].

Herein, we focus on the foundational mass diffusion model (i.e., network based inference, NBI) [9]. It estimates the asymmetrical similarity concerning mass diffusion proportion from the collected to the uncollected object and implies that the more proportion, the more similar they are. However, owing to sparsity and heterogeneity in BN, the asymmetrical nature of mass diffusion biases similarity estimation and largely damages the accuracy, diversity and personalization of recommendation. Besides, because of the redundant similar preferences of users to popular objects, excessive selections of the same object induce superfluous mass diffusion, further deteriorating the recommendation accuracy. Realizing symmetrical similarity estimation with synchronous depression of superfluous diffusion can bring in the further promotion of recommendation performances. In this paper, we believe that the symmetrical punishment on superfluous diffusion is appropriate for accurate similarity estimation, which corrects the disadvantage of asymmetrical diffusion and meanwhile punishes the adverse superfluous diffusion. Thus, we construct a symmetrical punishment model on superfluous diffusion (shortly called SPM) to pursue better recommendation performances.

2. Symmetrical punishment model on superfluous diffusion

In online e-commerce systems, each user has collected some objects according to real requests. In an information filtering system, the selection relationship can be described in a BN. Let user set $U = \{u_1, u_2, \ldots, u_m\}$, object set $O = \{o_1, o_2, \ldots, o_n\}$, we obtain an $n \times m$ adjacent matrix of BN $G(O, U) = \{a_{ij}\}$ with $a_{ij} = 1$ if user u_i has collected object o_j and $a_{ij} = 0$ otherwise.

Based on BN, Zhou et al. have proposed a NBI model that estimates the objects' similarity via mass diffusion [9]. As shown in Fig. 1, the mass resources proportionally diffuse from collected objects to the users at first, and then proceed to propagate to the uncollected objects. For a certain user, the elements of source vector $\overrightarrow{f} = [x, y, z]$ are assigned with 0/1 according to BN, denoting the collected or uncollected relationship between the user and objects. Generally, after the entire diffusion