Physica A 401 (2014) 15–21

Contents lists available at ScienceDirect

Physica A

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

Effect of the time window on the heat-conduction information filtering model

Qiang Guo, Wen-Jun Song, Lei Hou, Yi-Lu Zhang, Jian-Guo Liu *

Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China

HIGHLIGHTS

- The effect of the time window on the personalized recommendation algorithm is investigated.
- Only adapting recent records, the performance of the heat-conduction model could be improved greatly.
- The accuracy of small-degree users could be enhanced greatly.

ARTICLE INFO

Article history:
Received 22 June 2013
Received in revised form 22 October 2013
Available online 21 January 2014

Keywords:
Time window
Heat-conduction information filtering model
Bipartite network

ABSTRACT

Recommendation systems have been proposed to filter out the potential tastes and preferences of the normal users online, however, the physics of the time window effect on the performance is missing, which is critical for saving the memory and decreasing the computation complexity. In this paper, by gradually expanding the time window, we investigate the impact of the time window on the heat-conduction information filtering model with ten similarity measures. The experimental results on the benchmark dataset Netflix indicate that by only using approximately 11.11% recent rating records, the accuracy could be improved by an average of 33.16% and the diversity could be improved by 30.62%. In addition, the recommendation performance on the dataset MovieLens could be preserved by only considering approximately 10.91% recent records. Under the circumstance of improving the recommendation performance, our discoveries possess significant practical value by largely reducing the computational time and shortening the data storage space.

1. Introduction

The overwhelming mass information on the World Wide Web [1] makes it difficult for us to find out the valuable ones that we are really interested in Ref. [2]. In order to break through this dilemma, many recommendation algorithms, by predicting the most appealing items to the users based on their purchase history, have been proposed [3]. In such process, exploiting the context (*e.g.* location, time, gender) in which users express their preferences has been proven very valuable for improving the recommender performance [4,5]. However, the physics of the time window effect on the performance is missing.

Among the existing contextual dimensions, the time information can be considered as one of the most useful ones. In recent years, some scholars contribute themselves to integrate the time information into the recommender algorithms [6], since it plays a significant role on the improvement of the recommendation performance. Revamping two dominating collaborative filtering algorithms, Koren [7] presents a model tracking the time changing behaviors in order to distill the

* Corresponding author. Tel.: +86 15202121269.
E-mail address: liujg004@ustc.edu.cn (J.-G. Liu).

0378-4371/$ – see front matter Crown Copyright © 2014 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2014.01.012

longer-term trends from the noisy patterns. Another example of a continuous time-based model is given in Xiong et al. [8], where a Bayesian probabilistic tensor factorization (TF) model is proposed. In addition, some scholars have turned their attention to the 'less can be more' phenomenon. For instance, Zhang et al. [9] propose a hybrid method combining the time-aware and topology-aware link removal algorithms by extracting the backbone which contains the essential information for the recommender systems. Although the recommendation systems only have to deal with a small amount of data in this backbone, the relationship between the accuracy of prediction results and the time window is missing. Therefore, the physics of the time window impact on the recommendation performance is of significance for deeply understanding the online user collection behaviors and developing the effective algorithms.

In most instances, the entire available data has been applied to the temporal recommendation algorithms, due to the conventional wisdom that data size is essential for the recommender performance, in other words, the more data we have collected, the more precise the recommendation results are Ref. [10]. However we could achieve a better prediction result by only considering the most essential information such as the Top-N recommendation algorithm [11]. In addition, it makes little sense to consider the purchase behaviors happened a long time ago, because the users' collecting behavior is a dynamic process in which their preferences constantly change over time. For instance, the items collected by a certain user long time ago could not express his current preferences. Motivated by this idea that considering the information happened too long ago makes no sense, we investigate the impact of the time window on the recommender systems. We gradually expand the time window to generate a series of training sets from the benchmark datasets called Netflix and MovieLens, then we use each training set to predict the future preferences of the users proved by the testing set. The empirical results indicate that utilizing partial historical data to recommend could surprisingly promote the accuracy and diversity of the recommender systems.

The rest of the paper is organized as follows. In Section 2, we introduce the user-based heat-conduction recommendation model with ten well-known local-information-based similarity measures. In Section 3, we present our experimental work, the description of the adopted dataset, the evaluation index and the discussions of the results. In the final section, we make a conclusion and point out the possible directions for the future work.

2. Model and methods

2.1. Bipartite network and heat-conduction model

A user-object bipartite network consists of a set of user nodes denoted as $U = \{u_1, u_2, \ldots, u_n\}$, the object nodes as $O = \{o_1, o_2, \ldots, o_m\}$ and the links between these two sets, which are indicated by $E = \{e_1, e_2, \ldots, e_p\}$. The bipartite network containing n users and m objects can be represented by an adjacent matrix $A = \{a_{ij}\} \in R^{m,n}$, where $a_{ij} = 1$ if user u_j collects object o_i and $a_{ij} = 0$ otherwise.

The heat-conduction (HC) recommendation model includes two types, one is object-based model, and the other is user-based model. The object-based HC recommendation model supposes that the objects collected by the target user possess the ability to recommend highly relevant objects for him. Generally its framework is as follows: (i) construct the weighted object network denoted as the matrix W according to the known user-object relations; (ii) determine the resource vector f for each user; (iii) get the final resource distribution via

$$f^{'} = Wf; \tag{1}$$

(iv) recommend those uncollected objects with highest final resource to the target user. As to the user-based HC recommendation model [12,13], the users with similar tastes prefer to possess similar preference. Assuming that a target user u_i does not collect a certain object o_α, that is, $a_{\alpha i} = 0$. According to the user-based HC recommendation model, the probability to collect the object o_α by the user u_i is calculated by

$$v_{\alpha i} = \frac{\sum_{j=1}^{n} s_{ij} a_{\alpha j}}{\sum_{j=1}^{n} s_{ij}}, \tag{2}$$

where s_{ij} represents the similarity from the user u_j to u_i. Under this configuration, all users rated object o_α have the equal recommendation power. To the target user u_i, when the predicted scores $v_{\alpha i}$ among all the objects he/her has not collected are calculated, all these scores will be sorted in descending order, and finally those on the top will be recommended. There are numerous methods to measure the users' similarities, such as common neighbor index, Jaccard Index [14], Sørensen index [15], hybrid index [16] and so on. Although a minority of these measures are not directly proposed to the bipartite networks, they could also be introduced to measure the node similarities of the bipartite networks. Since most of previous work focus on the object-based HC recommendation model [17–20], the understanding of the similarity effect on the user-based recommendation model is lacking. In this paper, we extensively investigate ten well-known users' similarity measures based on the local information of bipartite networks. The detailed definitions would be introduced in the next section.