Physica A 436 (2015) 236–245

Contents lists available at ScienceDirect

Physica A

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

Improve the algorithmic performance of collaborative filtering by using the interevent time distribution of human behaviors

Chun-Xiao Jia [a], Run-Ran Liu [a,b,*]

[a] Institute of Information Economy, Hangzhou Normal University, Hangzhou, 310036, People's Republic of China

[b] Postdoctoral workstation of alibaba group, Hangzhou, 311121, People's Republic of China

HIGHLIGHTS

- The effects of interevent time of human behaviors on CF algorithm are investigated.
- The newly proposed time-related CF algorithm outperforms the standard CF algorithm.
- This work validates the findings of interest-driven model of human dynamics.

ARTICLE INFO

Article history:
Received 9 January 2015
Received in revised form 7 March 2015
Available online 12 May 2015

Keywords:
Collaborative filtering
Interevent time distribution
Bipartite network

ABSTRACT

Recently, many scaling laws of interevent time distribution of human behaviors are observed and some quantitative understanding of human behaviors are also provided by researchers. In this paper, we propose a modified collaborative filtering algorithm by making use the scaling law of human behaviors for information filtering. Extensive experimental analyses demonstrate that the accuracies on *MovieLens*and *Last.fm* datasets could be improved greatly, compared with the standard collaborative filtering. Surprisingly, further statistical analyses suggest that the present algorithm could simultaneously improve the novelty and diversity of recommendations. This work provides a creditable way for highly efficient information filtering.

1. Introduction

Recently, with the development of Internet and world wide web [1], especially for the mobile Internet, we are confronted with the problem of information overload [2]. Therefore, more fast and convenient information services are put forward by users online. In order to break through this dilemma, various recommender algorithms [3–11], which attempt to predict users interests and potential shopping tendencies, have been proposed. Recommender algorithms have been widely found applications in many websites, e-commerce websites recommend produces for the buyers to improve their volume, social websites recommend new friends to improve the site stickiness, news websites recommend news for their users to serve users better in the future. In comparison to the traditional tools, such as search engines (which require precise keywords to describe what he/she needs) and category navigation (where the contents are classified by topics), recommender systems

* Corresponding author at: Institute of Information Economy, Hangzhou Normal University, Hangzhou, 310036, People's Republic of China.
E-mail address: runranliu@163.com (R.-R. Liu).

http://dx.doi.org/10.1016/j.physa.2015.05.060
0378-4371/

provide us with a different way to filter information and return personalized results to different users by analyzing user historical activities, which probably can recommend some surprising items beyond the limits of user's knowledge. Although the recommender systems have been found in many fields, the current generation of recommendation systems still requires further improvements to make recommendation methods more effective.

The current recommender algorithms include collaborative filtering (CF) [12–14], content-based methods [15,16], spectral analysis [17,18], principal component analysis [19], the widest applied of which is CF method. The CF method has two categories in general, one is user-based method (UCF) [12], and the other one is called object-based method (OCF) [20], which are based on the similarities between users and those between objects respectively. The UCF method recommends the target user the objects collected by the users sharing similar tastes, and the OCF method recommends those objects similar to the ones that the target user preferred in the past. Recently, researchers make use of some physical dynamics on user–object bipartite networks, such as random walks and heat conduction [4–6,21–23], to measure object similarity, which are found to be highly effective in generating recommender algorithms. Inspired by these methods, subsequent research also find that recommender algorithms can be also improved by adjusting network structures, such as extracting information backbone [24], or adding ground users [25]. Recently, time context information has found applications at some stage of the prediction process for recommendations [26] and link predictions [27], and they can provide differentiated recommendations depending on the target recommendation time. However, the interevent time of user behaviors are not fully valued in the evaluation of object–object similarities or user–user similarities. Taking movie recommendation as an example, if two users have watched a movie at the same time, they may have similar tastes. However if they have watched a movie with a long time interval, it sounds far-fetched to say they have similar tastes. In a particular time period, users may have a relatively stable interest and they may collect similar objects driven by this interest. In a word, the time interval of user behaviors may play an important role in the evaluation of user–user or object–object similarities. Very recently, many scaling laws of interevent time distribution of human behaviors are observed [28,29]. The subsequent research also reveals that there exists a strongly positive correlation between user's activity and the total number of user's actions [30], and a significantly negative correlation between the user's activity and the width of the interevent time distribution [31,30]. At the same time, some theoretical models have been proposed to understand the corresponding phenomena, including task-driven [28] and interest-driven models [32]. To some extent, these findings uncover the collective online habits and behaviors of users, which must play important roles on the prediction of user's online behaviors. In this paper, we have proposed a modified CF method by making use of the time interval of human behaviors for information filtering.

2. Method

A recommender system could be described by a bipartite network [33] in which there are two kinds of nodes: m objects and n users. All users' historical records are represented by the edges connecting users and objects: if an object o_i is collected by a user u_j, there is an edge between o_i and u_j, and the corresponding element a_{ij} in the adjacent matrix A is set as 1, otherwise it is 0. The main task of a recommender system is to generate a ranking list of the target user's uncollected objects based on the observed information and to recommend the top-ranked objects to the target user. For the UCF method, the similarities between any pair of users are calculated according to the intersection of their collection lists. Analogously, for the OCF method, the similarities between any pair of objects are calculated according to the intersection of their users. There are at least three ways previously proposed to measure similarity, such as Sorensen's index of similarity [34], cosine similarity [35,36], and Jaccard index [37], the most widely used of which is cosine similarity. For instance, the similarity between users u_i and u_j can be represented as the following equation by the cosine similarity representation:

$$s_{ij} = \frac{\sum_{l=1}^{m} a_{li}a_{lj}}{\sqrt{k_i(u)k_j(u)}}, \tag{1}$$

where $k_i(u)$ and $k_j(u)$ represent the degrees of users u_i and u_j respectively. In UCF, the predicted score v_{ij} (to what extent u_j likes o_i), is given as:

$$v_{ij} = \frac{\sum_{l=1, l\neq i}^{n} s_{jl}a_{il}}{\sum_{l=1, l\neq i}^{n} s_{jl}}. \tag{2}$$

For OCF method, the similarity between two objects o_i and o_j can be represented as

$$s_{ij} = \frac{\sum_{l=1}^{n} a_{il}a_{jl}}{\sqrt{k_i(o)k_j(o)}}, \tag{3}$$