Physica A 436 (2015) 658–664

Contents lists available at ScienceDirect

Physica A

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

Predicting online ratings based on the opinion spreading process

Xing-Sheng He [a], Ming-Yang Zhou [a], Zhao Zhuo [a], Zhong-Qian Fu [a,*], Jian-Guo Liu [b,*]

[a] Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230027, PR China

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

HIGHLIGHTS

- We present a rating prediction method based on the user opinion spreading process.
- The user similarity is defined by measuring the amount of opinion a user transfers to another one.
- The algorithmic accuracy are improved greatly comparing with the item average method.

ARTICLE INFO

Article history:
Received 19 January 2015
Received in revised form 6 May 2015
Available online 18 May 2015

Keywords:
Recommender system
Opinion spreading process
Bipartite network
Collaborative filtering

ABSTRACT

Predicting users' online ratings is always a challenge issue and has drawn lots of attention. In this paper, we present a rating prediction method by combining the user opinion spreading process with the collaborative filtering algorithm, where user similarity is defined by measuring the amount of opinion a user transfers to another based on the primitive user-item rating matrix. The proposed method could produce a more precise rating prediction for each unrated user-item pair. In addition, we introduce a tunable parameter λ to regulate the preferential diffusion relevant to the degree of both opinion sender and receiver. The numerical results for Movielens and Netflix data sets show that this algorithm has a better accuracy than the standard user-based collaborative filtering algorithm using Cosine and Pearson correlation without increasing computational complexity. By tuning λ, our method could further boost the prediction accuracy when using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as measurements. In the optimal cases, on Movielens and Netflix data sets, the corresponding algorithmic accuracy (MAE and RMSE) are improved 11.26% and 8.84%, 13.49% and 10.52% compared to the item average method, respectively.

1. Introduction

With the rapid development of Internet and World Wide Web, lots of sources and data are generated and available for the public [1,2]. These superabundance information bring people convenience and excitement as well as a problem which is widely known as information overload [3]. For example, people are overwhelmed by massive information when they try to purchase a product, search for an article, ask for advice. What is more, enormous new information come out every single day, which makes the problem even harder. Meanwhile, especially in the advertisement field, customers are no longer

* Corresponding authors.
E-mail addresses: zqfu@ustc.edu.cn (Z.-Q. Fu), liujg004@ustc.edu.cn (J.-G. Liu).

http://dx.doi.org/10.1016/j.physa.2015.05.066
0378-4371/

satisfied with the stereotyped, less innovative suggestions and craving for specific personalized services. As a consequence, there is an urgent need of a feasible tool to address such an issue and recommender systems have interested a lot of researchers and scholars [4]. Miscellaneous methods have been proposed to build recommender systems under different circumstances, including the collaborative filtering (CF) method [1,2,5], spectral analysis [6,7], latent factor model [8,9], hybrid algorithms [10–12], and so on.

Recently, some physical approaches have also found applications in recommender systems (e.g. heat conduction (HC) [13] and probability diffusion (PD) [14]). Due to the low computational complexity and high efficiency, these physical approaches have drawn a lot of attention. Zhou et al. studied the effect of initial configuration on network-based recommendation and found that decreasing the initial resource located on popular objects can further improve the accuracy [15]. Liu et al. introduced a modified collaborative filtering (MCF) by obtaining user–user correlations based on a diffusion process, moreover, by considering the second-order correlations, MCF outperforms standard collaborative filtering among accuracy, popularity and diversity [16]. Zhou et al. also solved the apparent diversity–accuracy dilemma of recommender systems by using a free tunable parameter combining the HC and PD processes together, which is of high accuracy and diversity [12]. Guo et al. presented an improved user-based HC model which takes users' positive and negative opinions into account. By predicting users' interested and disliked object lists independently, then filtering out the dislike objects from the interesting lists, this method gains great improvement in accuracy [17]. The aforementioned works show the superiority of these physical approaches in prediction accuracy, computational complexity and recommendation diversity. However, these physical dynamics inspired methods only give a recommendation list for a target user but are unable to give a concrete rating prediction for an unrated user-item pair.

Recommender systems usually utilize various kinds of user-item interactive information of its users for a set of items (e.g., songs, movies, cuisines, travel destinations, products, academic articles). These information can be obtained implicitly, typically by monitoring users' historical behaviors like books read, web pages visited, songs heard, products purchased. Not only behavior information, some recommender systems may use users' demographical information, such as gender, location, age, occupation as well. However, through the implicit feedback, users' interest in products cannot be obtained completely. The most fundamental and convenient input data for a recommender system is high-quality explicit feedback from users regarding their interests to products. The feedback can be measured by ratings and a higher rating usually means a more positive attitude to an item. A particular recommender system collects ratings from their customers and places the data in a matrix with one dimension representing users and the other dimension representing items of interest. Most previous methods [5,12,15–20] take a coarse-graining procedure and transform the input matrix into unary form: An object is considered to be collected/favored by a user if and only if the given rating is above a threshold. However, this method does have some drawbacks: First, given a rating interval where the maximum number denotes the most positive opinion and the minimum number indicates the most negative attitude, users' feedback ratings do not evenly distribute along the rating interval [17]: In general, users tend to give high ratings. Second, various users may have various rating criteria, some of them are inclined to give high ratings whereas others are perfectionists that they hardly give a full mark to any items they have viewed. In addition, besides the recommendation list, it is more important to generate a concrete prediction rating score for the uncollected user-item pairs. In this paper, by taking the aforementioned factors into account, we present a new user similarity measurement based on opinion spreading process and develop an improved method to enhance the prediction accuracy. In the optimal cases, comparing with the ones generated by standard CF using Pearson correlation, the prediction accuracy measured by Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) could be increased by 2.29% and 3.07% for Movielens when the neighbor size is 50, 3.17% and 3.55% for Netflix when the neighbor size is 80, respectively.

2. Method description

Suppose there are m users and p items in a recommender system, the user and item sets can be denoted as $U = \{u_1, u_2, \ldots, u_m\}$ and $O = \{o_1, o_2, \ldots, o_p\}$. A recommender system with explicit feedback can be fully depicted by a rating matrix $\mathbf{R} = \{r_{i\alpha}\} \in R^{m,p}$, where $r_{i\alpha}$ is the rating of user i on item α and $r_{i\alpha}$ can be any real number, but are often integers in the range [1, 5]. Meanwhile, we can easily derive a particular adjacency matrix $\mathbf{A} \in R^{m,p}$, where the element $a_{i\alpha} = 1$ if user i has given a rating on item α and 0 otherwise from the origin rating matrix $\mathbf{R}$. For the sake of clarity, we use Latin letters for users and Greek letters for items to distinguish different types of nodes. k_i and k_α thus denote the degree of user node i and item node α, respectively.

We first describe the user opinion spreading (UOS) method. For user i and user j who have given a rating on item α, assuming they both hold a positive or negative opinion to item α (we will discuss how to define positive and negative opinion based on ratings later), it can be concluded that the two users reach an agreement on item α, thus the two users may have similar taste and their opinions may be crucial to each other. On the other hand, if user i favored item α whereas user j despised, we can assume user i and user j have a discrepancy on item α and thus their opinions may be irrelevant to each other.

As a consequence, if two users i and j have ever rated a same item α, we assume there is an opinion spreading path from i to j through item α and vice versa. Users at the terminals of the path therefore can transmit their opinions to each other. But there is a question that how to judge a user's attitude towards a particular item according to his/her rating on that item. In many previous works [5,12,15–20], a coarse-graining method was applied to distinguish the favored/disfavored ratings