Commun Nonlinear Sci Numer Simulat 51 (2017) 115–123

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

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

Research paper

Information mining in weighted complex networks with nonlinear rating projection

Hao Liao[a], An Zeng[b], Mingyang Zhou[a,*], Rui Mao[a], Bing-Hong Wang[c]

[a] Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, PR China

[b] School of Systems Science, Beijing Normal University, Beijing, 100875, PR China

[c] Department of Modern Physics, University of Science and Technology of China, Hefei 230027, PR China

ARTICLE INFO

Article history:
Received 8 September 2016
Revised 19 March 2017
Accepted 24 March 2017
Available online 7 April 2017

Keywords:
E-commerce system
Complex networks
Social interaction
Rating projection

ABSTRACT

Weighted rating networks are commonly used by e-commerce providers nowadays. In order to generate an objective ranking of online items' quality according to users' ratings, many sophisticated algorithms have been proposed in the complex networks domain. In this paper, instead of proposing new algorithms we focus on a more fundamental problem: the nonlinear rating projection. The basic idea is that even though the rating values given by users are linearly separated, the real preference of users to items between the different given values is nonlinear. We thus design an approach to project the original ratings of users to more representative values. This approach can be regarded as a data pretreatment method. Simulation in both artificial and real networks shows that the performance of the ranking algorithms can be improved when the projected ratings are used.

1. Introduction

The coming big data era brings us a critical problem in the information explosion age: how to extract the valuable information from the big data at hand. This problem is especially crucial in online systems where the available data are overwhelmingly abundant due to the rapid expansion of the Internet [1–4]. To filter out irrelevant online items (e.g. books, movies or others) for users [7], the recommender system, such as the collaborative filtering methods are widely applied [5,6]. Besides the relevance, the quality of items is also of great importance to online users. Therefore, many online websites, such as Amazon.com and Netflix.com build the online reputation system [8–11] in which users can give their opinions to an item by assigning certain rating value to it. The purpose of the reputation system is to help users uncover the true quality of items. After obtaining the rating data, some algorithms are needed to generate the ranking of items. The most straightforward way is to simply use the arithmetic average of ratings to rank items' quality. However, since this method has low ranking accuracy and is sensitive to spamming behavior, many other ranking algorithms have been proposed recently [12,13].

Other types of ranking algorithms compute users' reputation and items' quality self-consistently. More specifically, these algorithms usually update users' reputation in an iterative way and aggregate the ratings based on the reputation of users [14]. A representative one of these algorithms is called iterative refinement (IR) [16]. In IR, a user's reputation is inversely

* Corresponding author.
E-mail address: zhoumy2010@gmail.com (M. Zhou).

http://dx.doi.org/10.1016/j.cnsns.2017.03.018
1007-5704/

proportional to the mean difference between his rating vector and objects' estimated quality vector (i.e., weighted average rating based on user reputation). The estimated quality of objects and reputation of users are iteratively updated until the values reach a stationary point. This method is further modified by assigning trust to each individual rating [15,17,18]. Recently, Zhou et al. [19] takes the robustness of the algorithm into account and propose to calculate a user's reputation by the Pearson correlation [20] between his ratings and objects' estimated quality. This method is usually referred to as the Correlation-based Ranking (CR) method and it can be resistant to the malicious spamming behaviors of some users.

However, a fundament problem in the reputation system has been neglected for a long time. For most online reputation systems, the rating values are discrete and linearly separated. For example, some well-known websites such as Amazon.com and Netflix.com use the 5-star rating system: users are allowed to rate items with integers from 1 (worst) to 5 (best) [21,22]. However, the real preference of users to items between different rating values can be actually nonlinear. For instance, the difference between ratings 4 and 3 might not be equal to that between 5 and 4. Based on this idea, we design a rating projection method which allows us to project the original rating values to more representative ones. We consider both artificial and real networks. The projected ratings are then used as input to several ranking algorithms and significant improvement in the ranking accuracy is observed.

2. The nonlinear rating projection method

The reputation systems can be normally described by weighted bipartite networks consisting of online users and items. If a user rated an item, there is a link between them, and the link weight is the rating value that the user gives to the item. Here, we consider a common case where users rate items by using the integer scale from 1 (worst) to 5 (best). In this 5-star rating system, 3 means neutral. However, when a user rates item α with 4 and item β with 5, it doesn't mean that the user like β two times more than α. This problem also exists when one compares rating 1 and 2. Based on this idea, we design a rating projection method. Since rating 1, 3 and 5 respectively stand for the worst, neutral and best, we preserve these three ratings. The method transfers the rating values 2 and 4 to new values R_2 and R_4 via

$$R_2 = P_1 * 1 + (1 - P_1) * 3 \tag{1}$$

$$R_4 = P_2 * 3 + (1 - P_2) * 5 \tag{2}$$

where P_1 has a meaning of the ratio between rating 1 and 3. P_2 has similar meaning of the ratio between rating 3 and rating 5. To simplify the rating projection, we find a simple way to represent above projection method by using p_1 and p_2 as tunable parameters as follows

$$R_2 = 1 + p_1 * 2 \tag{3}$$

$$R_4 = 3 + p_2 * 2, \tag{4}$$

where p_1 and p_2 are tunable parameters. Clearly, when $p_1 = 0.5$ and $p_2 = 0.5$ are assigned, it gives the original rating values. However, the rating values becomes nonlinear when $p_1 \neq 0.5$ and $p_2 \neq 0.5$. In the following, we will investigate the performance of different ranking algorithms when the projected ratings are used.

3. Ranking algorithms

In this work, we mainly consider four ranking algorithms in our experiments. We first introduce some notations for these ranking algorithms. The users are denoted by set U and items are denoted by set O. To better distinguish different types of nodes in the bipartite network, we use Latin letters for users and Greek letters for items. The rating given by a user i to an item α is denoted by $r_{i\alpha}$. Moreover, we define the set of items selected by user i as O_i and the set of users selecting item α as U_α, and the degree of users and objects are respectively k_i and k_α. We also denote the quality of item α and the reputation of user i as Q_α and R_i, respectively.

(i) The first algorithm is the so-called *mean* method. In this method, the quality of an item is simply the mean ratings the item received. As this method is easy to calculate, it is widely used by many websites. However, as this method does not consider user's reputation, it is sensitive to malicious manipulations.

(ii) The *iterative refinement* (IR) calculates user reputation and item quality in a self-consistent way. It considers a user's reputation as inversely proportional to the mean squared error between his/her rating vector and the corresponding objects' weighted average rating vector [16]. The estimated object quality values Q_α is defined as

$$Q_\alpha = \frac{\sum_{i \in U_\alpha} R_i r_{i\alpha}}{\sum_{i \in U_\alpha} R_i}, \tag{5}$$

and the estimated reputation of user R_i is computed as

$$R_i = \left(\frac{1}{|O_i|} \sum_{\alpha \in O_i} (r_{i\alpha} - Q_\alpha)^2 + \varepsilon \right)^{-\beta}, \tag{6}$$