Physica A 436 (2015) 629–636

Contents lists available at ScienceDirect

Physica A

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

Ranking online quality and reputation via the user activity

Xiao-Lu Liu, Qiang Guo, Lei Hou, Can Cheng, Jian-Guo Liu *

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

HIGHLIGHTS

- We present an algorithm ranking online user reputation in terms of the user activity.
- The experimental results show that the AUC values reach 0.9065 for MovieLens.
- The results for the artificial networks show the effect of the user degree for IRUA algorithm.

ARTICLE INFO

Article history:
Received 9 February 2015
Received in revised form 26 March 2015
Available online 21 May 2015

Keywords:
Online rating system
User reputation
Object quality
Iterative algorithm

ABSTRACT

How to design an accurate algorithm for ranking the object quality and user reputation is of importance for online rating systems. In this paper we present an improved iterative algorithm for online ranking object quality and user reputation in terms of the user degree (IRUA), where the user's reputation is measured by his/her rating vector, the corresponding objects' quality vector and the user degree. The experimental results for the empirical networks show that the AUC values of the IRUA algorithm can reach 0.9065 and 0.8705 in Movielens and Netflix data sets, respectively, which is better than the results generated by the traditional iterative ranking methods. Meanwhile, the results for the synthetic networks indicate that user degree should be considered in real rating systems due to users' rating behaviors. Moreover, we find that enhancing or reducing the influences of the large-degree users could produce more accurate reputation ranking lists.

1. Introduction

Evaluating the user reputation accurately is of great significance in ranking the object quality for the online rating systems [1–5]. Nowadays, rating systems are used in many online websites, where users can evaluate objects by giving not-continuous ratings. Ranking objects by simply averaging ratings they received may be less accurate because of users' dishonesty or non-familiarity [6,7]. Therefore, building a reputation system for users is important for online social systems [8–12].

In the reputation generation algorithms, the iterative-oriented mechanisms have been widely investigated and implemented, such as the PageRank [13], HITS [14] algorithms. Laureti et al. [15] proposed an iterative refinement algorithm, namely IR algorithm, where the user's reputation is inversely proportional to the difference between his/her rating vector and the corresponding objects' calculated quality vector. The object quality and user reputation can be updated iteratively until the change of the quality between two iteration steps is lower than a certain threshold. Zhou et al. [6] developed a correlation based ranking algorithm (short for CR), where the user's reputation is determined by the Pearson correlation coefficient between his/her rating vector and the corresponding objects' calculated quality vector. Liao et al. [16] designed

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

http://dx.doi.org/10.1016/j.physa.2015.05.043

an iterative algorithm with reputation redistribution (short for IARR), presenting the reputation redistribution process to eliminate noisy information in the iterations. Meanwhile, the IARR method was modified by introducing two penalty factors (small-degree users or objects cannot have very high reputation or quality), namely IARR2 method [16], leading the objects rated by only low reputation users with less quality and the users who only rate a small number of objects with lower reputation. In empirical rating systems, users with different degrees are evolved by different mechanisms [17–19].

The traditional iterative algorithms calculate the user reputation and object quality regardless of the role of different users' activity patterns. In this paper, taking into account the user activity level, we present an improved iterative algorithm for ranking object quality and user reputation (short for IRUA), which can effectively improve the performance of the reputation measurement. The results for the empirical networks show that the AUC values of the IRUA algorithm can reach 0.9065 and 0.8705 in Movielens and Netflix data sets, respectively, which indicates that the IRUA algorithm could generate more accurate quality ranking lists than the traditional CR, IARR and IARR2 methods. Moreover, the results for the synthetic networks indirectly testify the necessity of taking into account the user activity in ranking the user reputation.

The rest of this paper is organized as follows. We present the IRUA algorithm in Section 2 and analyze the results for the empirical networks and the synthetic networks in Section 3 and Section 4, respectively. Finally, section Section 5 gives the conclusion and discussions.

2. The IRUA algorithm

The rating system can be described by a weighted bipartite network, which consists of the users denoted by U and the objects denoted by O. We denote users and objects by using the Latin and Greek letters respectively to distinguish different types of nodes. The rating given by user i to object α is denoted by $r_{i\alpha}$. The set of users rating to object α is denoted as U_α, while the set of objects rated by user i is denoted as O_i. In addition, the degree of user i and object α are denoted as k_i and k_α, respectively.

We set Q_α as the quality of object α and R_i as the reputation of user i. Initially, according to every user's degree, his/her reputation is assigned as $R_i = k_i/|O|$ (where $|O|$ is the number of objects). The quality of an object is not only determined by the weighted average rating of this object, but also relied on the maximum degree of the users who rate it, which could be expressed as

$$Q_\alpha = \max_{i\in U_\alpha} \frac{\{k_i\}}{|O|} \cdot \frac{\Sigma_{i\in U_\alpha} R_i r_{i\alpha}}{\Sigma_{i\in U_\alpha} R_i}, \tag{1}$$

where $\max_{i\in U_\alpha}\{k_i\}/|O|$ is defined as a penalty factor based on the hypothesis: If an object is rated by users who only assess a small number of objects, regarding to the high ratings, we cannot claim this object has high quality.

According to the object quality Q_α, one can calculate the Pearson correlation coefficient C_i between the user's rating vector and the corresponding object quality vector in the following way,

$$C_i = \frac{1}{k_i} \Sigma_{\alpha\in O_i} \left(\frac{r_{i\alpha} - \bar{r}_i}{\sigma_{r_i}}\right)\left(\frac{Q_\alpha - \overline{Q}_i}{\sigma_{Q_i}}\right), \tag{2}$$

where k_i is the degree of user i, σ_{r_i} and σ_{Q_i} represent the standard deviations of the rating vector of user i and the corresponding object quality vector, respectively. Meanwhile, $\bar{r}_i$ and $\overline{Q}_i$ are their mean values.

The Pearson correlation coefficient is an efficient way to quantify the similarity between two vectors. If a user has more similar ratings to the corresponding objects' calculated qualities, he/she would have a higher reputation. If the Pearson coefficient C_i is smaller than 0, the reputation R_i of user i will be assigned to 0. Therefore, C_i is limited to [0, 1]. In the IRUA algorithm, the reputation R_i of user i is correlative with the Pearson correlation coefficient as well as the user degree k_i, which determine the user reputation simultaneously,

$$R_i = \begin{cases} \dfrac{k_i^\theta}{k_{\max}^\theta} C_i & \text{if } C_i \geq 0 \\ 0 & \text{if } C_i < 0, \end{cases} \tag{3}$$

where θ is a tunable parameter. Large-degree users' reputation R_i will be enhanced when the parameter $\theta > 0$ and, on the contrary, their reputation will be reduced when $\theta < 0$.

At each time step, the object quality Q_α and user reputation R_i will be updated according Eqs. (1)–(3). The iteration will stop when the difference between the quality vectors,

$$|Q - Q'| = \frac{1}{|O|} \Sigma_{\alpha\in O} (Q_\alpha - Q'_\alpha)^2 \tag{4}$$

is lower than the threshold of $\delta = 10^{-5}$.