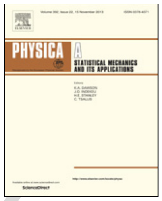
Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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



Identifying online user reputation of user–object bipartite networks

Xiao-Lu Liu [a], Jian-Guo Liu [a,b,*], Kai Yang [a], Qiang Guo [a], Jing-Ti Han [b]

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

[b] *Data Science and Cloud Service Centre, Shanghai University of Finance and Economics, Shanghai 200433, PR China*

HIGHLIGHTS

- Based on the Bayesian analysis, we present a parameter-free algorithm for ranking online user reputation.
- The user reputation is calculated based on the probability that their ratings are consistent with the main part of all user opinions.
- The AUC values of the presented algorithm could reach 0.8929 and 0.8483 for the MovieLens and Netflix data sets.
- The computation complexity of the presented algorithm is a linear function of the network size.

ARTICLE INFO

Article history:
Received 31 May 2016
Received in revised form 15 August 2016
Available online xxxx

Keywords:
Online rating system
User reputation
Object quality
Beta distribution

ABSTRACT

Identifying online user reputation based on the rating information of the user–object bipartite networks is important for understanding online user collective behaviors. Based on the Bayesian analysis, we present a parameter-free algorithm for ranking online user reputation, where the user reputation is calculated based on the probability that their ratings are consistent with the main part of all user opinions. The experimental results show that the AUC values of the presented algorithm could reach 0.8929 and 0.8483 for the MovieLens and Netflix data sets, respectively, which is better than the results generated by the CR and IARR methods. Furthermore, the experimental results for different user groups indicate that the presented algorithm outperforms the iterative ranking methods in both ranking accuracy and computation complexity. Moreover, the results for the synthetic networks show that the computation complexity of the presented algorithm is a linear function of the network size, which suggests that the presented algorithm is very effective and efficient for the large scale dynamic online systems.

1. Introduction

How to evaluate the user reputation in terms of their rating behaviors is important for the online rating systems [1–4]. Nowadays, online rating systems provide channels for users to show their preferences. However, not every user gives ratings subjectively since each user has his/her specific tastes and motivations [5–7]. Therefore, how to identify the online user reputation in terms of their ratings or selecting behaviors is important for building a reputation system [8–11].

Recently, the iterative ranking algorithms have been widely explored [12,13]. Zhou et al. [14] designed an iterative algorithm based on the correlation between the user rating and object quality vectors (short for the CR algorithm). The user reputation and object quality can be updated iteratively until the change between two iteration steps is smaller than a

* Corresponding author at: Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China.
E-mail address: liujg004@ustc.edu.cn (J.-G. Liu).

http://dx.doi.org/10.1016/j.physa.2016.10.031
0378-4371/

Please cite this article in press as: X.-L. Liu, et al., Identifying online user reputation of user–object bipartite networks, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.10.031

threshold. Liao et al. [15] developed an iterative algorithm with reputation redistribution (short for IARR), by redistributing the reputation to eliminate noisy information in the iterations. To filter out the influence of the unreliable users, Liao et al. [15] proposed the IARR2 method by introducing two penalty factors which assign smaller reputations to the users who rate small number of objects. The non-iterative online user reputation ranking algorithms are also discussed. Gao et al. [16] proposed a group-based ranking method (namely GR method) by grouping users according to their ratings. Then users' reputations could be determined by the corresponding group sizes.

In social networks analysis [17], by propagating ratings provided by multiple advisors, Teacy et al. [18] employed a probability density function to estimate the reputation of a selling agent. Zhang et al. [19] adopted the beta probability distribution to model the advisor's public reputation, which is estimated as the probability that he/she will provide fair ratings. Additionally, a rating will be regarded as the fair rating if it is consistent with the majority of the other ratings for one specific seller provided by other buyers. The expected value of the probability that a user will give fair ratings is calculated as his/her public reputation, which could be extended from the social networks to user–object bipartite networks to evaluate user reputation and object quality.

By introducing the Bayesian analysis, we present an parameter-free algorithm to rank online user reputation via the beta probability distribution, namely RBPD algorithm, where the user reputation is estimated as the probability that he/she will provide fair ratings to objects. Combining with users' personalities, the users' ratings are characterized to the positive or negative opinions. Finally, we use the expected value of the probability that the user will give fair ratings to calculate the reputation instead of the iteration process. Implementing our method for empirical networks and synthetic networks, the results show that the RBPD algorithm produces more accurate reputation ranking lists and the computation complexity is a linear function with the network size.

2. The RBPD algorithm

The rating system can be described by a weighted bipartite network [20–22], which consists of the users denoted by set U and the objects denoted by set O. The number of users, objects and ratings are denoted by $|U|$, $|O|$ and $|E|$, respectively. We use the Latin and Greek letters to represent the users and objects, respectively. The rating $r_{i\gamma}$ given by user i to object γ is the weight of the link in the bipartite network and all the ratings could be described as a rating matrix $\mathbf{A}$. The user set U_γ is defined as the users who rate to object γ, and the object set O_i is defined as the objects which are rated by user i. Moreover, the degree of user i and object γ are denoted as k_i and ρ_γ, respectively.

2.1. The online user reputation evaluation

The reputation of user i is denoted by R_i. We use the Bayesian analysis to model the user reputation. Bayesian analysis [18] adopts a binary event to measure each of users' ratings: Fair rating (denoted by 1) or unfair rating (denoted by 0). The definition of fair rating for bipartite networks could be introduced in the following way. User i provides a rating $r_{i\gamma}$ to object γ, the rating will be judged to determine whether it is consistent with the majority of the other opinions to object γ given by other users. Determining consistency with the majority of opinions can be achieved by identifying if the rating's opinion accounts for more than 50% of all opinions [19]. We define a rating $r_{i\gamma}$ as the fair rating if it is consistent with the majority of all users' opinions, otherwise as the unfair rating.

There are two kinds of opinions to the objects: Positive and negative ones. We use a coarse-graining method to distinguish them. The quantity $r'_{i\gamma}$ is defined as the extent of fanciness via the rating $r_{i\gamma}$, from which one can discover the opinion from user i to object γ. Considering the user personality that different users tend to have different rating criteria, where some users tend to give high ratings and others tend to give low ratings, we use a normalized method to transform a rating to the extent of fanciness in the following way,

$$r'_{i\gamma} = \begin{cases} 2(r_{i\gamma} - r_i^{\min})/(r_i^{\max} - r_i^{\min}) - 1 & r_i^{\max} \neq r_i^{\min} \\ 0 & r_i^{\max} = r_i^{\min}, \end{cases} \tag{1}$$

where $r_i^{\max}$ and $r_i^{\min}$ denote the maximum and minimum ratings user i gives, respectively. In this way, all the ratings given by one specific user would be transferred to $[-1, 1]$, where the maximum and minimum ratings are mapped into 1 and -1. Specifically, for the users who always give the same ratings, their ratings are normalized to 0. The normalized rating matrix is denoted by $\mathbf{A}'$, where the element is the rating's extent of fanciness. The positive and non-positive values could be interpreted as the positive and negative opinions give by users. For all the ratings, after observing whether they are fair or not, the results are denoted by matrix $\mathbf{B}$, where the element is Y or N (an fair rating is denoted by Y, others are denoted as N).

The reputation R_i of user i is defined as the probability θ_i that user i will provide fair ratings to objects, which lies in [0,1]. Because there is only partial information about users, the best way to estimate the probability θ_i is to use its expected value,

$$R_i = E(\theta_i). \tag{2}$$

The expected value $E(\theta_i)$ of the probability θ_i is up to the probability density function, where the beta probability distribution [18] is commonly used as a prior distribution for random variables that take on continuous values in the interval [0,1]. For user i, whether the ratings are fair or not can be expressed by the following vector,

$$D_i = \left[X_i(1), X_i(2), \ldots, X_i(k_i)\right], \tag{3}$$