Physica A 502 (2018) 288–294

Contents lists available at ScienceDirect

Physica A

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

Effect of the initial configuration for user–object reputation systems

Ying-Ying Wu [a], Qiang Guo [a], Jian-Guo Liu [b,*], Yi-Cheng Zhang [c]

[a] Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China
[b] Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, PR China
[c] Department of Physics, Fribourg University, CH-1700 Fribourg, Switzerland

HIGHLIGHTS

- We investigate the effects of the initial configuration on identifying online user reputation for the user–object bipartite networks.
- When the parameter q equals to 0.8 and 0.9, the accuracy value AUC would increase about 4.5% and 3.5% for the Netflix data set.
- Online users' reputations will increase as they rate more and more objects.

ARTICLE INFO

Article history:
Received 24 August 2017
Received in revised form 21 November 2017

Keywords:
Reputation
Initial configuration
User–object reputation systems

ABSTRACT

Identifying the user reputation accurately is significant for the online social systems. For different fair rating parameter q, by changing the parameter values α and β of the beta probability distribution (RBPD) for ranking online user reputation, we investigate the effect of the initial configuration of the RBPD method for the online user ranking performance. Experimental results for the Netflix and MovieLens data sets show that when the parameter q equals to 0.8 and 0.9, the accuracy value *AUC* would increase about 4.5% and 3.5% for the Netflix data set, while the *AUC* value increases about 1.5% for the MovieLens data set when the parameter q is 0.9. Furthermore, we investigate the evolution characteristics of the *AUC* value for different α and β, and find that as the rating records increase, the *AUC* value increases about 0.2 and 0.16 for the Netflix and MovieLens data sets, indicating that online users' reputations will increase as they rate more and more objects.

1. Introduction

Reputation systems based on the Bayesian framework have been widely used to identify user reputation for user networks, which provides a flexible framework for integrating reputation services into e-commerce applications [1,2], where the user reputation is denoted by the beta distribution [3–5]. The basic idea based on the beta distribution is to define the user reputation as the expectation value of the beta probability density function(PDF) with $\alpha = 1$ and $\beta = 1$, which applies not only to user–user networks but also to user–object bipartite networks [6,7].

Mui et al. [8] proposed a probabilistic mechanism for inference among trust, reputation, and level of reciprocity, where the user reputation was defined as a quantity embedded in the social network for evaluating the agent and encounter historical behaviors. Jøsang et al. [9] used the beta PDF to combine feedback and derive reputation ratings based on the probability expectation value. To filter out unfair ratings, Whitby et al. [10] proposed a statistical filtering technique based on the beta

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

https://doi.org/10.1016/j.physa.2018.02.147

distribution. In the recent years, the reputation systems based on Bayesian framework for user–object bipartite networks have been developed. Based on the beta distribution, Liu et al. [11] presented a parameter-free algorithm for ranking online user reputation via the beta probability distribution (RBPD), where the user reputation is calculated based on the expectation value of the probability that the user will give fair ratings. It could be not noticed that the initial fair and unfair rating distributions are set as uniform ones, denoted by the parameters $\alpha = 1$ and $\beta = 1$.

There were some Bayesian algorithms considering other forms of α and β. Jøsang et al. [12] transformed continuous ratings into discrete ratings by using the fuzzy set membership function where the parameters α and β are determined by the function and the ratings. Then Jøsang et al. [13] introduced the non-informative prior weight as a parameter to calculate user reputation scores. Although these models changed the initial configuration in the way of changing the parameters α and β, the effects of the initial configuration on reputation systems have been not discussed. Zhou et al. [14] pointed out that the configuration of initial resource distribution affects the accuracy of recommendation algorithms, even under the simplest case with binary resource. Inspired by this idea, we investigate the effects of the initial configuration on the reputation algorithm for user–object networks.

In this paper, we investigate the effect of the initial configuration for Bayesian reputation systems. The user reputation is based on the beta PDF with different parameters α and β. Firstly, we calculate user reputation scores and object quality values by the RBPD algorithm in user–object bipartite networks and take the ratio of all the opinions denoted by a fair rating parameter q into consideration at the same time. Empirical results show that comparing with the uniform initial configuration, the *AUC* value changes with the increase of the parameters α and β and the larger the parameter q is, the larger the *AUC* value is. More significantly, when the parameter q is 0.8 and 0.9, comparing with the case of $\alpha = 1$ and $\beta = 1$, the *AUC* value would increase about 4.5% and 3.5% for the Netflix data set, while the *AUC* value would increase about 1.5% for the MovieLens dat set when the parameter q equals to 0.9, meaning that the effects of the initial configuration for different data sets are different. Furthermore, we investigate the evolution characteristics of the *AUC* value for different α and β. The integrate timestamp is divided into 10 time intervals of the same length. We can find that the *AUC* value gradually increases as the rating records increase, indicating that users' rating accuracy will be enhanced as the time they stay in the system becomes longer.

2. Methods

The rating system represented by a bipartite network consists of the user set U and the object set O. We use Latin and Greek letters to distinguish users and objects, respectively. Consequently, $r_{i\gamma}$ denotes the rating given by user i to object γ and R_i denotes user i's reputation. U_γ denotes the set of users who rated a given object γ, while O_i denotes the set of objects rated by user i, and k_γ and k_i denotes the degree of object γ and user i, respectively.

2.1. Bayesian reputation systems

Bayesian reputation systems [15] take binary ratings as input: Fair rating or unfair rating. When a rating $r_{i\gamma}$ is consistent with the majority of opinions among users set U_γ, it is considered as a fair rating for bipartite networks [16], otherwise as an unfair rating. We use a fair rating parameter q to denote the ratio of all the opinions where $q \geq 0.5$. We adjust the ratio of all the opinions to find the changes of the algorithm. For example, when the parameter q is 0.5, a rating need to account for more than half of all opinions and then it will be considered as a fair rating. As the parameter q increases, the difficulty that a rating is considered as a fair rating increases. For the reputation systems based on the beta PDF, the posteriori reputation score is calculated by combining the priori reputation score with the new rating [1]. User reputation scores can be represented in the form of the probability expectation value of the beta PDF, where the parameters α and β represent the amount of fair and unfair ratings respectively.

The PDF of the beta distribution is a power function of the probability variable θ and its reflection $(1-\theta)$ as follows with the gamma function Γ:

$$Beta(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \tag{1}$$

where $0 \leq \theta \leq 1$ and the parameters $\alpha > 0$, $\beta > 0$. The expectation value of the beta distribution is denoted by $E(\theta) = \alpha/(\alpha+\beta)$. Generally, the priori distribution is defined as the uniform beta PDF with $\alpha = 1$ and $\beta = 1$. After considering user i's rating records that there are s fair ratings and f unfair ratings, respectively, the posterior distribution is the beta PDF with $\alpha = s+1$ and $\beta = f+1$. However, the uniform initial configuration may affect the accuracy of the algorithm [14]. Therefore, we assume the priori distribution of the beta PDF is not uniform and user i's reputation is denoted by the probability expectation value of the beta distribution,

$$R_i = \frac{\alpha+s}{\alpha+\beta+s+f}, \tag{2}$$

where the parameters α and β denote the amount of fair and unfair ratings given by user i when there is no rating records in the initial stage of the system.