Contents lists available at ScienceDirect

Physica A

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

Evolution properties of online user preference diversity

Qiang Guo [a], Lei Ji [a], Jian-Guo Liu [a,b,c,*], Jingti 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
[c] Department of Physics, Fribourg University, CH-1700 Fribourg, Switzerland

HIGHLIGHTS

- The evolution pattern of the online user preference is investigated.
- The user rating preference would become diverse and then get centralized finally.
- The correlation between the user ratings and the object qualities keeps increasing.

ARTICLE INFO

Article history:
Received 15 May 2016
Received in revised form 25 September 2016
Available online xxxx

Keywords:
Online user collective behaviors
User preference
Evolution properties
Markov model

ABSTRACT

Detecting the evolution properties of online user preference diversity is of significance for deeply understanding online collective behaviors. In this paper, we empirically explore the evolution patterns of online user rating preference, where the preference diversity is measured by the variation coefficient of the user rating sequence. The statistical results for four real systems show that, for movies and reviews, the user rating preference would become diverse and then get centralized finally. By introducing the empirical variation coefficient, we present a *Markov* model, which could regenerate the evolution properties of two online systems regarding to the stable variation coefficients. In addition, we investigate the evolution of the correlation between the user ratings and the object qualities, and find that the correlation would keep increasing as the user degree increases. This work could be helpful for understanding the anchoring bias and memory effects of the online user collective behaviors.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Understanding the effects of the online collective behavior patterns is important for the computational social science [1], decision-making processes [2–4], and the marketing strategy [5–8]. Recently, the exploration of the online user preferences has become a promising issue for predicting the success of culture products and presenting precise marketing strategy [9–12]. Han et al. [13] investigated the adaptive interests of online user preferences, which is helpful for solving the diversity–accuracy dilemma [14]. Zhang et al. [15] proposed an evolutionary hypergraph model to reconstruct the online user-tag-resource network, showing the user preferences to save resources with tags they are interested in. An evolving network model then was introduced to reproduce the user selection patterns on tags, music and movies [16]. Thanks to the development of web 2.0, online users could not only select, browse, but also review and share what they like. Their preferences and tastes are reflected by their collective rating or selecting behaviors [17,18]. Many efforts have been devoted

* 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.2016.10.050
0378-4371/© 2016 Elsevier B.V. All rights reserved.

ARTICLE IN PRESS

Table 1
The basic statistical properties of the data sets. N and M is the number of users and objects respectively. The number of links E represents all ratings delivered by users during the time span, denoted by the number of days from the start date to the end date.

Data sets	N	M	E	Start date	End date	#day
MovieLens	71 567	10 681	10 000 054	01-09-1995	01-05-2009	5110
Netflix	478 190	17 770	17 197 883	01-01-2003	12-31-2004	731
Amazon	99 621	645 055	2 005 409	05-31-1996	09-15-2005	3394
Epinions	120 492	755 760	13 668 320	01-10-2001	05-30-2002	505

to detect the online user preference patterns by analyzing these user generated contents [12,19–21]. Rybski et al. [22,23] analyzed the communication activity and the correlation between the long-term correlation and inter-event clustering. Yang et al. found that there is an anchoring bias effect during the user rating process [24]. Hou et al. [25] investigated the memory effect of the online user rating series [26,27] and found that the *Markov* model could regenerate online user collective behaviors, which indicates that a user's next action depends only on his/her current behaviors [28–31].

The above works mainly focused on the correlation between the next action and the current one regardless to the user lifespan. How do the online user preferences change throughout the user life span? Whether the Markov model could be commonly used to generate the online behaviors? Inspired by these questions, we investigate the evolution properties of online user preference diversity, where the preference diversity is measured by the variation coefficient. Firstly, we rescale each user life span to an standard time interval (see Methods), which has been used for the domain analysis of various online settings [32–35]. Comparing with the null model, the empirical results indicate that, for movies and reviews, the diversity of user preference increases initially and then decreases to a small value. More importantly, we find that the *Markov* process could only describe some empirical user rating behaviors when the variation coefficients are stable. Furthermore, we investigate the correlation between user rating value and the object quality. The statistical results suggest that, generally speaking, the correlation between the rating value and the object quality decreases.

2. Data description

Four different data sets are introduced in this paper, differing in the subject matter and the span of the time window, as shown in Table 1. These data sets are widely applied to modeling the patterns of online user preferences [24–26]. Each record in these data sets reads the user–object pair, followed by the rating the user gave to the object as well as the corresponding timestamp. The *MovieLens*[1] and *Netflix*[2] data sets are from websites which accommodate movies of various types for users. Generally, the users would not only watch movies but also give their ratings. The *Epinions* data set [36] compiles millions of user ratings on expert product reviews (e.g., reviews on electronics, cars and books). The *Amazon*[3] data set retains huge volume of user rating records on books. In the *MovieLens* data set, the user ratings are float number from 0.5 to 5, which is differentiated into 10 levels with 0.5 score as the difference of adjacent levels. In the other data sets, ratings are integers from 1 to 5. In this paper, we only take into account the users who rated at least 100 objects with active time span no less than 40 days. Meanwhile, we assume the object quality can be denoted by the ratings average over all of the users.

3. Methods

In the above data sets, users can give explicit ratings to the object, expressing to what extent they like or dislike the object. The higher a rating is, the higher the user evaluates the object. Processing the user rating sequence is reasonably regarded as the direct and effective way to reflect the properties of online user preferences. For instance, consecutively delivering higher or lower ratings has been used to identify the anchoring bias and memory effect of user online rating and selecting behaviors [24,25]. In this paper, the evolution properties of online user preference diversity could be measured by the variation coefficient V, a widely used distribution-based measure in the field of complex systems [37], defined as the ratio of the standard deviation σ to the mean μ. Specifically, we normalize each user's time span into the same time interval [0, 1], where 0 and 1 represent the starting and ending time of the online user activities. For a user, say u, who delivered m_u ratings during his/her time span, the variation coefficient V could be calculated in the following way. Firstly, the normalized time span is divided into T intervals with the same length, which could be denoted as $[0, \frac{1}{T})$, $[\frac{1}{T}, \frac{2}{T})$, $\ldots$, $[\frac{T-2}{T}, \frac{T-1}{T})$, $[\frac{T-1}{T}, 1]$, respectively. All the $m_u(t)$ ratings before time t consist of the sequence $RS(t) = \{r_1, r_2, \ldots, r_{m_u(t)}\}$, where r_i is the ith rating given by user u. Therefore, for each time t, the coefficient of variation $V_u(t)$ for user u can be expressed as,

$$V_u(t) = \frac{\sigma_u(t)}{\mu_u(t)}, \tag{1}$$

[1] http://www.grouplens.org.
[2] http://www.netflix.com.
[3] http://www.amazon.com.

Please cite this article in press as: Q. Guo, et al., Evolution properties of online user preference diversity, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.10.050

Download English Version:

https://daneshyari.com/en/article/5103531

Download Persian Version:

https://daneshyari.com/article/5103531

Daneshyari.com