Physica A 417 (2015) 261–266

Contents lists available at ScienceDirect

Physica A

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

Memory effect of the online rating for movies

Yi-Lu Zhang, Qiang Guo, Jing Ni, Jian-Guo Liu *

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

HIGHLIGHTS

- There is a linear correlation between the users' rating behaviors and the real-time updated average ratings of objects.
- Users rate higher scores if the displayed average ratings are lower than 2.0, and give lower ratings if the average ratings are higher than 4.5.
- Small-degree users would rate higher than the real-time displayed average ratings, while large-degree users usually give lower ratings.
- The distributions of the users' rating bias of the large-degree users are small, yet those of the small-degree users are relatively large.

ARTICLE INFO

Article history:
Received 13 May 2014
Received in revised form 26 July 2014
Available online 20 September 2014

Keywords:
Online social networks
User's collective behaviors
Online rating
Recommendation

ABSTRACT

Online rating can directly reflect users' collective behavioral patterns which is of great concern in online social systems. In this paper, we investigate the correlations between the users' rating behaviors and the real-time updated average ratings of objects given from other users' previous ratings. We average all the ratings rated after the real-time displayed average ratings at a given interval after dividing the data into five groups according to the user degrees. By analyzing two real systems, the results show that in general there is a linear correlation with slope one between them when the displayed average ratings are between 2.0 and 4.5, but users rate higher scores if the displayed average ratings are lower than 2.0, and give lower ratings if the average ratings are higher than 4.5. Besides, small-degree users would rate higher than the real-time displayed average ratings, while large-degree users are stricter with their ratings than the others so that they usually give lower ratings whatever the movies are. Furthermore, the distributions of the users' rating bias in all the five groups show that the rating biases of the large-degree users are small, yet those of the small-degree users are relatively large. Our findings could be helpful to analyze online users' collective behaviors as well as abnormal behaviors in the networks.

1. Introduction

In the past decade, tremendous activities have been devoted to the understanding of the online users' behavioral patterns due to the rapid development of online social networks [1–6]. Particularly in the online e-commerce systems, it is significant to be better acquainted with users' purchasing behaviors and provide high-quality services. Previous works have found that people's behavior is far different from random and obeys certain predictable rules [7,8]. On the network structures, some typical statistical patterns of the online users' behaviors will appear [9–12]. For instance, Onnela et al. [9] analyzed the Facebook applications to study the role of social influence on the patterns of user behaviors and found an on–off pattern of

* Corresponding author. Tel.: +86 15202121269.
E-mail address: liujg004@ustc.edu.cn (J.-G. Liu).

http://dx.doi.org/10.1016/j.physa.2014.09.012
0378-4371/

Table 1
Basic statistics of the data sets tested. N, M, and V denote the number of users, objects and ratings respectively, and $\langle r\rangle$ is the average rating over all systems.

Data set	N	M	V	$\langle r\rangle$
MovieLens	5547	5850	698,054	3.48
Netflix	8609	5081	419,247	3.42

user selections, which is also named common and specific interests [13,14]. Bianconi et al. [10] investigated the effects of the selection mechanisms of users on the network evolution.

By further analyzing online ratings, we could design highly efficient recommender systems which are able to find out suitable products for each user as he likes [15,16]. In some systems, people's options are just confined to likes or dislikes, while in other systems, they can vote with explicit ratings usually from one to five stars. Yang et al. [17] found a so-called anchoring bias embedded in the individual decision-making processes, namely people strongly tended to give a low rating again after voting on a low-quality object, as well as to give a high rating again after voting on a high-quality object. In other words, the results indicated that a user's prior vote on some object would affect his current vote because he might take that prior vote as an anchor. Actually people's votes could be affected by prior votes [18] as well as social influences [19]. However, the user's rating is also affected by the average rating given to the target object.

The paper is to analyze the correlations between the users' rating behaviors and the real-time updated average ratings of objects displayed to them. What we value is how the users usually rate when they catch sight of the average rating of an object which originates from other users' previous ratings. We make empirical analyses based on two real data sets about movies, MovieLens and Netflix. In such networks, as a user clicks a movie, the real-time updated average rating of the movie is provided for him, and then he casts his rating. First, the data is classified into five groups according to the user degrees. In each group, The ratings given by users belonging to that group are ordered by time, and then we sum up all the ratings that are rated after the real-time displayed average ratings at a given interval and then average them to find out the relationships between users' ratings and the real-time updated average object ratings. The empirical results show that in general there is a linear correlation with slope one between users' ratings and the real-time updated average ratings when the average ratings are [2.0, 4.5], but users tend to give a higher rating when the displayed average rating is lower than 2.0, and they would like to give a lower rating if the average rating is higher than 4.5. In addition, we also find that small-degree users would rate higher than the real-time displayed average ratings, while large-degree users are stricter with their ratings than the others so that they usually give lower ratings no matter what the movies are. Furthermore, we analyze the distributions of the rating bias based on users to better understand how each user rates in the five groups. It turns out that the rating biases of the large-degree users are small, while those of the small-degree users are relatively large. And the other groups of users usually rate close to the average object ratings displayed to them.

2. Data analyses

In this paper, we consider two data sets, named MovieLens and Netflix (see Table 1 for basis statistics). The MovieLens is an online movie recommendation Web site, who invites users to rate movies and, in return, make personalized recommendations. While the Netflix web site also have DVD rental service. The MovieLens data set we choose consists of 698,054 records of 5550 movies rated by 5547 users. The Netflix data set [20] is a random sample of the whole records of user activities on Netflix.com, which contains 5081 movies, 8609 users, and 419,247 ratings. Both of the two data sets are movie rating systems with five stars, whose ratings are from one (i.e., worst) to five (i.e., best). If the user selects a movie and rates on it, a link between the user and the movie would be established.

All the ratings in each data set are ordered by the rating time as $r_1, r_2, \ldots, r_V$, where r_1 is the oldest rating, and r_V is the most recent rating. Since we want to further analyze how people would rate after they catch sight of the average rating which originates from other users' previous ratings, we first need to understand the basic rating patterns based on the MovieLens and the Netflix data sets. We define the rating bias of every rating record from user i onto object α as

$$\Delta^{i\alpha} = r_s^{i\alpha} - \langle r_{C_s}\rangle^{\alpha}, \tag{1}$$

where $\langle r_{C_s}\rangle^{\alpha}$ is the real-time updated average rating of object α that is given from other users previous ratings, $r_s^{i\alpha}$ is the rating user i rates on object α, $\Delta^{i\alpha}$ denotes the rating bias of user i on object α, $s \in [1, V]$, among which V still indicates the number of ratings.

Fig. 1 shows the distributions of the rating biases, from which we can find that the trend of distributions of the rating biases on the MovieLens and the Netflix data sets are quite similar. The majority of the rating bias is between $[-1, 1]$, and the number of ratings that have no bias is the largest in both two systems. It indicates that most people would like to give a rating close to the average rating displayed to them. However, there still exist some exceptional situations. For instance, the probability of the rating biases larger than 3 is also relatively large, which means that there is a part of users who would rate much higher than the average ratings. Moreover, some users like to give extremely low ratings even though the real-time displayed average ratings are high. Such phenomena imply that there might be some users who rate irrationally and