



Quantifying users' selection behavior in online commercial systems

Xi Wang^{a,b}, Heyang Li^{a,c}, An Zeng^{a,*}

^a School of Systems Science, Beijing Normal University, 100875 Beijing, China

^b School of Government, Beijing Normal University, 100875 Beijing, China

^c School of Statistics, Beijing Normal University, 100875 Beijing, China

HIGHLIGHTS

- We study online users' movie selection behavior.
- We find that users mostly select similar movies.
- There is still small possibility for users to select dissimilar movies.
- We propose a model to reproduce the empirical findings.

ARTICLE INFO

Article history:

Received 28 March 2018

Received in revised form 11 July 2018

Available online xxxx

Keywords:

Complex networks

Data mining

Recommender systems

ABSTRACT

In order to uncover the online user behavior patterns, this study uses massive data from online movie rental websites as an example to explore users' behavior characteristics of watching movies and puts forward a search model to fit users' viewing mode. We use complex network tools to construct and analyze the movie space. Three main conclusions are drawn. First, the average similarity between two movies a user consecutively watched is high if this user has low activity. Second, movie stickiness increases as movie popularity increases. Third, two consecutively watched movies will not be similar if these two movies are viewed at relatively long time interval. Comparing the movie space with the product space studied by Hidalgo et al. in 2007, we find that similarity is the most important factor in both networks, but jumping behaviors which do not apply to the product space exist in the movie space. Based on the above analysis, we propose a model to simulate users' behaviors of watching movies and obtain the model parameters that best fit the real data. This model reveals users' viewing mode hidden in the data. The search model may help movie websites to recommend movies for users precisely and bring commercial benefits. It is also of great significance in film promotion and development.

© 2018 Published by Elsevier B.V.

1. Introduction

With the arrival of the big data era, massive data have been accumulated in many social, biological and information systems. They make many traditional analytical methods not effective. In recent years, the rapidly developing complex network has become an important tool for empirical and modeling analysis of these big data systems [1]. Network science is pushing forward a series of issues of major theoretical and practical significance in various fields, including the underlying mechanisms to reveal the basic functions of living organisms [2], understanding human behavior characteristics and the

* Corresponding author.

E-mail address: anzeng@bnu.edu.cn (A. Zeng).

formation of social structures [3], preventing and controlling the epidemic of global infectious diseases [4], optimizing the operation of air and rail transportation systems [5], and promoting the sharing and filtering of massive information in the Internet era [6].

Among them, research of information recommendation technology and complex networks has great significance and value both in theory and application. In theory, information recommendation is an important aspect of the major scientific issues such as information mining and information filtering [7]. For application, information recommendation technology has become the core technology of most e-commerce systems and has created huge economic value [7–9]. Information recommendation usually includes three components: users, objects and recommendation methods. The recommendation method is the core and most critical part of the entire recommendation system, and determines the performance of the recommendation system. Traditional methods are mainly classified into three categories: collaborative recommendation (collaborative filtering), content-based recommendation and hybrid recommendation [10]. Due to the importance of recommendation systems, a great deal of research has been conducted on a variety of data analysis techniques such as data clustering, Bayesian networks, association rules, data classification, K-means methods, maximum entropy methods, cloud models, multi instance learning, neural networks, linear regression, etc. They are all used in the recommendation system [11]. Especially driven by contests such as Netflix [11,12] and Yahoo Music, the information recommendation algorithm is fully researched, and each recommendation algorithm is also tested accordingly. Among them, the integrated recommendation algorithm based on multiple algorithms ultimately performs best. This also shows that a single recommendation algorithm is not always a good solution to even a single score prediction problem. In addition, one challenging issue is that algorithmic performance often depends on data sets, and algorithms that perform well on one data set can be ineffective at dealing with other data sets.

In the recent ten years, complex network research has rapidly developed. Complex networks can be used to represent a large number of social, biological and communication systems [13–16], in which nodes represent individuals or organizations while edges represent the interaction between nodes. Early complex network analysis mainly focused on empirical analysis to find whether there were some common features among various networks. The most significant findings were scale-free phenomena [17] and small-world effects [18]. In addition, many feature parameters are designed to finely describe the topological characteristics of the network, such as the fractal dimension used to characterize the self-similarity of networks [19], and the betweenness centrality used to characterize the potential load in traffic network connections [20]. The assortativity coefficient is used to characterize the connection mode of the network [21,22], the modularity is used to characterize the strength of the community structure [23], and the subgraph centrality is used to characterize the local connectivity of the network [24]. In particular, ref. [25] summarizes the various measures of complex network features.

Among complex network research, people's dynamics have been widely researched, including human mobility patterns [26], structure and tie strengths in mobile communication networks [27], leaders in social networks [28] and social group evolution [29], etc. Some famous datasets such as MovieLens, Netflix, Delicious, etc. help people learn more about online users' behavior. There are many research about online users including patterns of temporal variation [30], popularity and novelty dynamics in evolving networks [31], predicting the volume of comments on online news stories [32], popularity prediction in microblogging networks [33,34] and trending prediction in online networks [35]. Although online user behavior patterns have been extensively researched, the selection behavior of users in online commercial systems has been rarely studied. For example, the characteristics of user selecting movies, music, books, etc. In this paper, data of users watching movies are used as examples. This study explores online users' behavior patterns of watching movies by using research methods in complex networks. At the same time, this study will produce practical results by providing a search model based on complex networks, and put forward suggestions on the development of the movie.

2. Data description

We used data from two online movie rental websites, cleaned and processed as MovieLens¹ dataset and Netflix² dataset. MovieLens dataset contains the information of 1682 movies watched by 943 users. Netflix dataset contains the information of 16569 movies watched by 4960 users.

Both degree and degree distribution are essential for complex network research. The degree of a movie denotes how many times this movie has been watched and can also be defined as movie popularity. The degree of a user denotes how many movies this user has watched and can also be defined as user activity. Given that the degree distribution of movies or users is not smooth enough, we draw the cumulative degree distributions of movies and users in the two datasets to show the degree distribution clearly in fewer fluctuations. As is shown in Fig. 1, the cumulative degree probability appears to grow logarithmically with the decreasing degree, which is well-fitted by an exponential distribution.

Based on the analysis of the movie degree, this article defines the concept of the similarity between movies [36]. A_i represents the set of users who have seen movie i . Similarity Φ_{ij} between movie i and movie j is defined as:

$$\Phi_{ij} = \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \quad (1)$$

¹ <https://grouplens.org/datasets/movielens/>

² <https://www.kaggle.com/netflix-inc/netflix-prize-data>

Download English Version:

<https://daneshyari.com/en/article/7374628>

Download Persian Version:

<https://daneshyari.com/article/7374628>

[Daneshyari.com](https://daneshyari.com)