

## Analysis of Personalized Recommendation Algorithm Based on Content

Lin Zhu<sup>1,\*</sup>, Sheng Dai<sup>1,b</sup>

<sup>1</sup>Guangdong University of Science and Technology, Department of Computer Science, Dongguan, Guangdong 523083, China.

\*hope9037@163.com, <sup>b</sup>dai1879424@163.com

### Abstract

In order to solve the information overload and improve the data utilization rate, the recommendation system came into being. This article mainly explains the content-based recommendation algorithm in the recommendation system, details its concept and specific process, analyzes the key problems and common solutions, and finally summarizes the advantages and disadvantages of the algorithm and the future development direction.

### Keywords

Recommendation algorithm, Feature extraction, Similarity.

### 1. Introduction

With the development of various electronic devices and mobile network technologies, applications such as e-commerce and information search have become an important part of our lives. According to 44 "Statistical Reports on the Development of China's Internet", as of June 2019, the number of Internet users in China reached 854 million, of which 668 million were online news users and 639 million were online shopping users. In order to meet the personalized needs of users of this size and at the same time to solve the situation of excessive product information, the recommendation system came into being and has been widely and deeply applied, which has become the focus of the industry [1]. There are three main methods used by mainstream recommendation systems: content-based recommendation algorithms, collaborative filtering recommendation algorithms, and hybrid recommendation algorithms. This article mainly analyzes and studies the content-based personalized recommendation algorithm.

### 2. The concept of content-based recommendation algorithm

Content-based recommendation is based on the user's past data records to predict the content that the user may be interested in, and then make recommendations [2]. The core idea is to analyze and refine the user's data to obtain the user's characteristic model and user's behavior representation, and then match the target object and the user, and recommend the target object with a high degree of matching. The simplified process is shown in the figure below, see Fig.1.

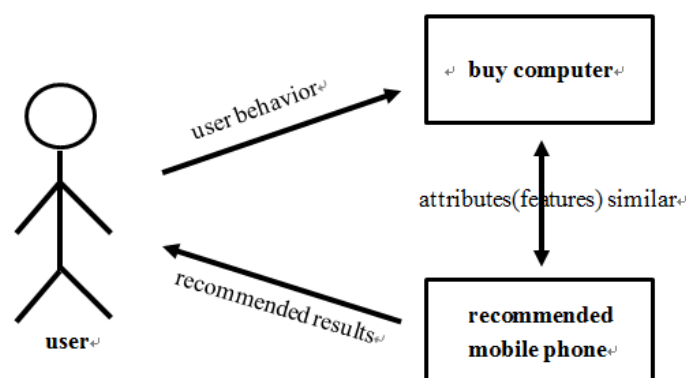


Fig.1 content recommendation diagram

The content-based recommendation algorithm can quickly recommend only by analyzing the user's personal historical data, which can greatly avoid the impact of the cold start problem. The algorithm focuses on comparing the consistency of future predictions with current behavior, and produces recommendation results that are highly similar to the present.

### 3. Process of content-based recommendation algorithm

The main process of the content-based recommendation algorithm is divided into three steps: feature extraction, constructing the characteristics of the user and the target; feature learning, mining the user's preference feature model; performing recommendations, comparing user preferences and the target to generate recommendations. The process is shown in the figure below, see Fig.2.

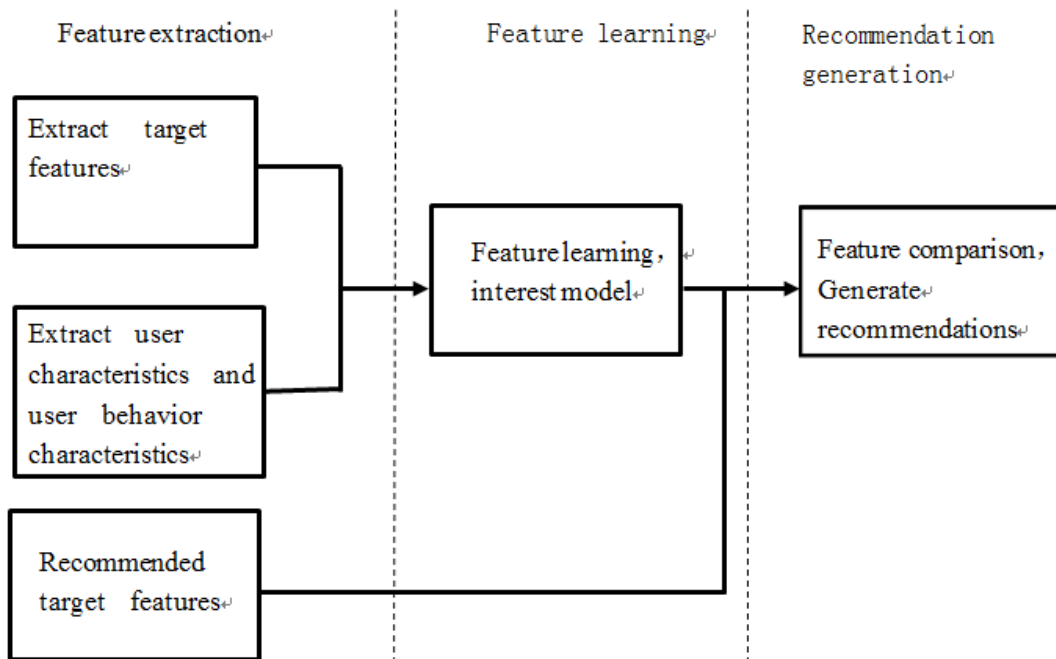


Fig.2 content recommendation process

Specific steps are as follows:

(1) Feature extraction, extracting user characteristics and user behavior characteristics, and characterizing user interest preferences at the same time, obtaining a data model of interest targets. This step plays a decisive role in the recommendation and directly affects the accuracy of the recommendation. The most critical issue in this process is how to accurately express the data and get the correct feature representation.

The types of data are complex and diverse, including numerical, nominal, and evaluation<sup>[3]</sup>. When these data are used to represent the user's preferred features, they need to be processed to produce a more consistent representation of the features and at the same time benefit mathematical calculations. Commonly used data processing methods include data specification, statistical methods (such as TF-IDF) and other methods. When representing features, vector representation is usually used to model features.

(2) Feature learning analyzes the characteristics of users and objects of interest to produce connections between users and objects of interest<sup>[4]</sup>. The key question in this step is what method to use to analyze and process the characteristic data. When analyzing data, methods such as machine learning and data mining are often used, such as decision tree algorithm and naive Bayes algorithm.

(3) In the process of recommendation generation, we compare the characteristics of interested objects and potential objects, and select objects with high similarity for recommendation. The key problem in this process is the calculation of similarity. The commonly used calculation methods include cosine angle, Euclidean distance, jacquard coefficient and Pearson correlation coefficient.

In the recommendation system, user interest migration is an important research content. The content that users are interested in will change with time. In order to get better recommendations, it is necessary to accurately represent the changes of users' interest. At present, the research of interest change is mainly based on collaborative filtering algorithm. There are few researches on the interest migration of content-based recommendation algorithm. Literature [4] proposes a recommendation algorithm based on the interest weighting of intermediate goods, which reflects the change of users' interest by increasing the time weight of the interest of the latest intermediate goods. Generally speaking, how to express the change of interest conveniently and concisely is the key and difficult point of interest migration, and is the key point of more perfect recommendation system.

#### **4. Advantages and disadvantages of content-based recommendation algorithms**

The advantages and disadvantages of content-based recommendation algorithms are obvious. The advantage is that it is hardly affected by the cold start problem, as long as the user has a small amount of data, he can make recommendations; at the same time, the content-based recommendation algorithm only analyzes the user's own data and does not involve other users, so its data processing is relatively simple. The main disadvantage is that the accuracy of the recommendation depends heavily on the correct representation of the data. It needs to be supplemented by questionnaires and statistics to obtain more comprehensive and accurate data. At the same time, content-based recommendations cannot tap the user's potential interest. Since the recommendation algorithm is based on the comparison between the future and the current similarity, the ones with high similarity will be recommended, so it is not possible to generate new ones with a large gap from the original features. recommend. This also leads to the reason why the content-based recommendation algorithm is in the auxiliary position in the recommendation system, and it is often used together with other algorithms such as collaborative filtering.

#### **5. Summary**

Although the content-based recommendation algorithm has inherent defects and highly depends on the quality of the content, it is also widely used in practical applications. Its role is undeniable. As long as there are available attributes in the data, the content recommendation algorithm will be used. By combining other algorithms, the data can be screened well, thereby improving the efficiency and accuracy of recommendations. With the development of knowledge recommendation, deep learning and other technologies, it is believed that more and more efficient recommendation algorithms will appear in the future.

#### **Acknowledgements**

University level scientific research projects: University level scientific research project of Guangdong University of science and technology in 2019(Youth Project): Research on personalized recommendation technology based on big data.

#### **References**

- [1] Wei Yu, Dehua Xu. Overview and Prospect of recommended algorithm [J]. Science and technology and innovation, 2019(04):50-52.
- [2] Han Cong Huang Fu, Zhaodi Xiao. Analysis of personalized recommendation algorithm based on user behavior data analysis [J]. Electronic design engineering, 2019, 27 (07):38-41.
- [3] Weichao Liu, You Yang, Ping Yu. Review of Content-based News recommendation system [J]. Fujian computer, 2019, 35 (09):71-74.
- [4] Zhiping Yang, Qian Xu, Ming Ma. Content recommendation algorithm based on intermediate commodity interest weighting [J]. Journal of Beihua University (NATURAL SCIENCE EDITION), 2017, 18 (06):831-834.