Financial Engineering and Risk Management (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2023.060904
ISSN 2523-2576 Vol. 6 Num. 9

E-commerce Recommendation Algorithm Based on Big Data Analysis and Genetic Fuzzy Clustering

Jiwang You[1,a,*]

[1]*Nanjing University Business School, Nanjing, Jiangsu, China*
[a]*yjw2010work@126.com*
[*]*Corresponding author*

Keywords: Genetic Algorithm, Fuzzy Clustering, Recommendation Algorithm, EC

Abstract: With the continued expansion of the EC scale, personalized recommendation technology is widely used. However, traditional referral systems cannot meet current data processing needs, and the presence of highly powerful big data analytics capabilities is a fundamental prerequisite for new personalized referral systems. The paper focuses primarily on EC recommended algorithm research based on big data analysis and fuzzy clustering gene analysis. Based on the literature data, understand the basic theoretical issues related to EC boosting calculations and analyze the methods of genetic fuzzy group analysis. The EC promotion algorithm is designed and the designed algorithm is tested. In conclusion, the algorithm given in this work has a low MAE like the other two algorithms, so its configuration quality is high.

1. Introduction

In the field of EC, due to the explosive spread of electronic goods, the number of EC users has also grown rapidly, and the products and services they have brought are also diverse [1-2]. Therefore, huge amounts of information and data are generated in the EC system, and the scale of the quantity has also been huge [3-4]. Taking user behavior information as an example, user behavior information is the key basic element of personalized modeling recommendation system, and it also belongs to the main data type collected by EC system [5-6]. According to incomplete statistics from user behavior data analysis companies that focus on EC business, users generally have to browse five web pages or 36 web pages before finally choosing a product, social media or search engine, and interaction with the engine quite a lot of time. If all the data collected by the system are integrated, it means that there is about one terabyte of active data every day. Then for larger EC websites, the volume of user behavior data generated every day can be imagined [7-8].

Regarding the research and development of EC promotion algorithms, some researchers have proposed that recommendation algorithms are the core technology of the entire promotion system. The suggestion system can discover potential usage interests and tendencies through user feedback behavior. In order to better understand behavior and user preferences, there are currently many recommended algorithms used to model user behavior and preferences, and explicit feedback behavior usually requires some external data to improve the modeling, such as user social relationships, tag information etc [9]. However, the current research results are still relatively lack

of how to organize these data, and have produced more accurate suggestions. In implicit feedback activities, user behaviors and preferences are usually time-sensitive. Considering time series preferences and overall user preferences, it is a difficult task to study the internal connection of various time series user preferences. Therefore, it is necessary to choose different recommendation mechanisms for various user feedback behaviors. Modeling the situation is critical [10]. Some researchers also provide a logical StyMarkov ensemble model, which can be constructed as a Markov chain with singers, and learn to use the vector representation of songs to predict the singers. Like the sequence suggestion model, the overall suggestion model does not examine the behavior of the user's sequence, but only makes recommendations for the user's entire music purchase history. The collaborative filtering algorithm can be used to generate push results. The time push mode can capture the time deviation, but needs to consider the time-sensitive characteristics preferred by the user, while the overall push mode can record the user's overall settings, but ignores the time setting. If you want to better solve the problem, you can integrate the two [11]. In summary, there has been quite a lot of research on EC promotion algorithms, but further research is needed in data processing.

This paper studies the genetic fuzzy clustering EC recommendation algorithm based on big data analysis, and then analyzes the related overview of EC recommendation and genetic fuzzy clustering algorithm on the basis of literature data, and designs the algorithm and analyzes the design the algorithm performs detection and draws relevant conclusions through the detection results.

2. Research on EC Recommendation Algorithm

2.1. Overview of EC Recommendations

(1) Main features of EC

1) User interests change quickly

The convenience and flexibility of e-commerce allows users to access anytime, anywhere, regardless of time and place, ensuring significant interest [12]. For example, users can browse and find other people's products on their mobile devices, compare prices when shopping at malls, and make quick changes to show their direct interest.

2) User behavior data contains information from multiple sources

When users trade on e-commerce platforms, the behavioral data generated includes heterogeneous information from multiple sources (browsing, adding shopping carts, favorites, purchases, etc.). This can reflect the user's geographic location and current scene information, time information, and recent user behavior, which has a greater impact on the expected outcome of the recommendations.

(2) The basic requirements for establishing EC recommendation

1) User behavior can be modeled by making full use of user behavior data. In the mobile EC environment, a large amount of behavioral data hides unique behavioral patterns and user interest patterns. Therefore, personalized recommendation algorithms should be able to make full use of user behavior data, user behavior models, and user tracking capabilities. It accurately measures the user's interest in a particular product and provides them with a list of high-quality suggestions.

2) It can meet the current interests of users in real time, and meet the requirements of mobile EC settings in real time. In a mobile environment, the needs and preferences of users are often sudden, and the time and place scenes will also change. Existing table decomposition algorithms, similarity tables based on scores, and interest models based on text mining technology mainly reflect the long-term preferences of users, and cannot respond to users' current interests in real time.

3) It can integrate the multi-source information contained in the user behavior data, and use the

location information to find the nearest neighbor set, and alleviate the cold start problem of new users. The most important feature of mobile EC is location relevance. Accurate location helps to accurately represent the model that users are interested in, and it can also help understand the preferences of user groups on the site. When new users enter the recommendation system, even if they have no action records or historical data, they can obtain their location information and make suggestions based on their scene and their surroundings.

2.2. Genetic Fuzzy Clustering Algorithm

(1) The idea of K-means grouping algorithm

K-means algorithm is an unsupervised learning method, an indirect classification method based on the number of samples and similarity measures. The calculation takes k as a parameter, divides n objects into k complexes, and the similarity rate is calculated according to the average value of the objects in the cluster (also regarded as the center of mass of the cluster). The calculation first selects k random objects, each object represents the center of the cluster, and distributes each remaining object to the closest cluster, depending on the distance between the object and the center of each cluster, and then estimates the new center value of each group, repeating the above process until the reference function converges. The k-means algorithm is a standard point-to-point correction algorithm and an iterative dynamic grouping algorithm. The main point is that the sum of squares of errors is a reference function. Class-centered point-by-point change: If a pixel sample belongs to a specific class group according to a specific principle, the mean value of the class group should be recalculated, and the new mean value should be used as the collection center to change the data image of each next group. Class center batch: After all pixel samples have been sorted according to the class center of a specific group, the average value of each category is calculated as the set center for the next sorting.

(2) Fusion of genetics and clustering algorithm

In the process of data mining, the quality of grouping is very important to the entire knowledge discovery process. In addition to accuracy, grouping requires high performance and large amounts of data. Therefore, a good grouping algorithm must satisfy prior knowledge independence, as few parameters as possible, accuracy, speed and scalability. Traditional grouping algorithms based on grouping criteria are essentially local search algorithms. They use an iterative hill climbing method to find the best value. The k-means algorithm is a classification method in the classification of grouping methods. The division method first creates an initial division, the number of divisions to be constructed, k, and then there is an iterative relocation technique that improves the division by moving objects between parts. There are three problems with cluster mining. One is that it takes a long time to process a large amount of data, the other is that it is easy to fall into the minimum value, and the third is that it is sensitive to the center of the original cluster.

In view of the above problems, the study found that genetic algorithm is a method to find the best solution by simulating the natural evolution process. Its main feature is that it can effectively use implicit parallelism and global information, and the results can reflect the large area of the search area, which is convenient real-time processing, excellent robustness. Therefore, combining these features with the grouping algorithm can solve the deficiencies of the grouping algorithm.

3. EC Recommendation Algorithm based on Big Data Analysis and Genetic Fuzzy Clustering

3.1. Data Composition

The data that can be used for personalized EC mainly includes website click traffic data and mobile device data. In daily life, through specific registration, the personal data of EC users can be

combined with user identity information to form a complete set of personalized EC recommendation data. Record the number of online clicks and time of each user for effective marketing and promotion. Service providers can use this data to carefully analyze user access patterns and provide more targeted services.

Personalized user behavior data mainly includes social networks, linked sites, webpage dwell time, search keywords, mobile click applications, LBS-based user behavior data, etc (see Figure 1).

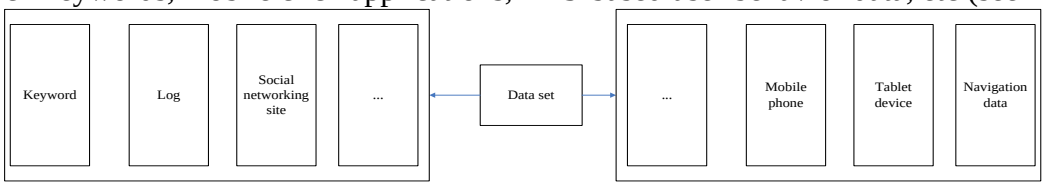

Figure 1: User behavior data based on LBS

3.2. Data Processing

(1) Quantification of interactive behavior

This article divides interactive behavior data into explicit data and implicit data. The data displayed is the evaluation data after the user purchases the product. Silent data includes information such as user purchases, browsing, favorites, favorites, and searches. The interactive part designed in this paper quantifies these hidden data into scoring information in a weighted manner, and the magnitude of user preferences represented by the quantified interactive behavior.

The design of the interactive behavior preprocessing part of this article regards the user's highest comprehensive score as the user's preference for the product (overall score). For example, after a user browses product A, collects product A, browses other products, and finally returns to the market to purchase product A, the product's evaluation is positive. Currently, four types of interactions are being created. Login, bookmark, buy, like browse pages and search product details.

(2) Unstructured data processing

Product information modeling is mainly the processing of unstructured information. Researchers in the field of information retrieval have done a lot of research on how to deal with unstructured information. In the process of implementing the recommended algorithm in this article, unstructured data processing uses a vector space model commonly used in the field of information retrieval.

3.3. Genetic Fuzzy Clustering Algorithm Process

(1) Chromosome interpretation and population initialization

Automatically determining the number of clusters is the first issue to be considered when designing a clustering algorithm. This section uses the chromosome length of the genetic algorithm to describe the number of clusters, and encodes the center of the cluster as a real number. Suppose X(t) is the population of generation t, and the code name X{(t) of the i-th atom can be expressed as $X_1(t) = x_{11}, x_{12}, \ldots, x_{1k_1}$. Among them, K represents the number of cluster centers on the chromosome. Obviously, the length of the chromosome is L = K * d.

(2) Fitness function

Evaluate the advantages and disadvantages of the grouping scheme. In other words, selecting the objective function of the grouping is the key to realize the optimization algorithm grouping. This document uses the DB-Index standard as the grouping objective function.

$$DB = \frac{1}{k}\sum_{i=1}^{K} R_{i,qt} \tag{1}$$

In the formula, K represents the number of clusters.

(3) Crossover operator

To prevent the element at the center of the cluster from being split to create meaningful new atoms, the following intersection operator is used. Define two intersecting parent atoms as sum $X_1(t)=x_{11},x_{12},...,x_{1k_1}$ and $X_2(t)=x_{21},x_{22},...,x_{2k_1}$, and randomly select integers $[1,k_1-1]$ and $[1,k_2-1]$ from k_1 and k_2, respectively, as the crossing positions of both parents. When two parents mate to obtain offspring of chromosomes $X_1'(t)=x_{11},x_{12},...,x_{1k_1}$ and $X_2'(t)=x_{21},x_{22},...,x_{2k_1}$, the code length is $L_1'=(k_1+K_2-k_2)d$ and $L_2'=(k_2+K_1-k_1)d$, respectively.

The crossover strategy introduced above not only dynamically adjusts the number of subgroups contained in an individual in the group, but also regroups the center of the group. This greatly increases the ability of the algorithm to obtain true clustering results.

(4) Mutation operator

Sudden change is the key to the algorithm generated from local optimization. The following mutation strategies are used to mutate to produce better individuals:

Variant $d(c_i,c_j)=\min\{d(c_p,c_q)\big|p,q=1,2,..K,p\neq q\}\wedge d(c_i,c_j)\leq\sigma,\sigma\geq 0$ considers the discretion between cluster centers. If the distance between the two centers of the cluster c on the chromosome satisfies, then the two centers of the cluster are the merger of any population. The merged new cluster center is:

$$C_{new}=0.5\times(c_i+c_j) \tag{2}$$

Among them, K is the number of chromosome clusters. Obviously the new chromosome length becomes L=(K-1)d.

4. Algorithm Test

The test data uses the user rating registration information submitted by MovieIens on the Internet. The test data set selected in the site database includes: 1000 users evaluated about 1200 recorded articles from 1700 projects. The movie user rating system includes five evaluation levels {1,2,3,4,5}, 1 is the lowest evaluation, and 5 is the highest evaluation. The entire data set is subdivided into a training set and a test set at a ratio of 70% to 30%.

Table 1: Algorithm test results

	Traditional algorithm	User-item based collaborative filtering algorithm	The algorithm of this article
5	0.730	0.76	0.944
10	0.720	0.74	0.942
15	0.710	0.735	0.938
20	0.708	0.732	0.936
25	0.705	0.730	0.934
30	0.710	0.732	0.930
35	0.711	0.733	0.928

In order to test the effectiveness of the algorithm given in the article, the algorithm in the article is compared with the collaborative filtering algorithm and the traditional recommendation algorithm. The number of neighbors gradually increased from 5 to 35 at five intervals. The test results are shown in Table 1.

It can be seen from Figure 2 that as the number of adjacent users increases, the MAE of each algorithm has a downward trend, but the MAE of the algorithm proposed in this paper is lower than the other two algorithms, resulting in higher configuration quality. Therefore, the collaborative

filtering algorithm based on genetic fuzzy clustering can provide users with more effective personalized suggestions and enhance the personalized service functions of EC websites.

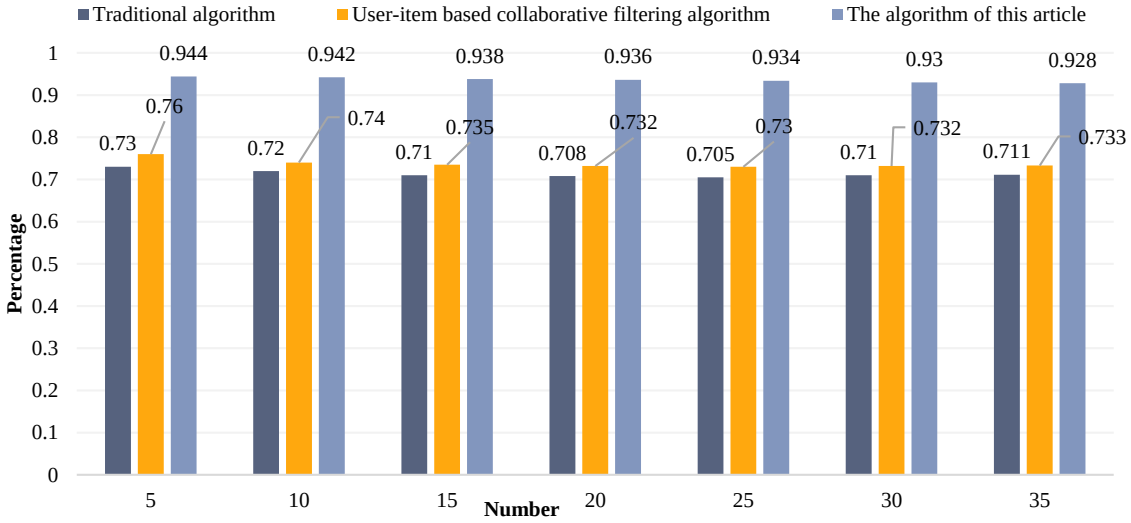

Figure 2: Algorithm test results

5. Conclusions

In this paper, the EC recommendation algorithm based on genetic fuzzy clustering is researched. After understanding the relevant theories, the EC recommendation algorithm based on genetic fuzzy clustering algorithm is designed, and the designed algorithm is verified by experiments. The experimental results are it is concluded that the algorithm proposed in this paper can better realize various personalized recommendations.

References

[1] Lu Q, Guo F. A novel e-commerce customer continuous purchase recommendation model research based on colony clustering. International Journal of Wireless & Mobile Computing, 2016, 11(4):309-317.

[2] Hu Q Y, Zhao Z L, Wang C D, et al. An Item Orientated Recommendation Algorithm from the Multi-view Perspective. Neurocomputing, 2017, 269(dec. 20):261-272.

[3] Liu X. An improved clustering-based collaborative filtering recommendation algorithm. Cluster Computing, 2017, 20(2):1281-1288.

[4] Zheng G, Yu H, Xu W. Collaborative Filtering Recommendation Algorithm with Item Label Features. International Core Journal of Engineering, 2020, 6(1):160-170.

[5] Cui L, Huang W, Qiao Y, et al. A novel context-aware recommendation algorithm with two-level SVD in social networks. Future Generation Computer Systems, 2017, 86(SEP.):1459-1470.

[6] Feng W, Zhu Q, Zhuang J, et al. An expert recommendation algorithm based on Pearson correlation coefficient and FP-growth. Cluster Computing, 2019, 22(3):1-12.

[7] Zhu H, Tian F, Wu K, et al. A multi-constraint learning path recommendation algorithm based on knowledge map. Knowledge-Based Systems, 2018, 143(MAR.1):102-114.

[8] Yang F, Wang H, Fu J. Improvement of recommendation algorithm based on Collaborative Deep Learning and its Parallelization on Spark. Journal of Parallel and Distributed Computing, 2021, 148(2):58-68.

[9] Zhou X, Su M, Feng G, et al. Intelligent Tourism Recommendation Algorithm based on Text Mining and MP Nerve Cell Model of Multivariate Transportation Modes. IEEE Access, 2020, PP (99):1-1.

[10] Fang X, Wang J, Sheng D, et al. Recommendation algorithm combining ratings and comments. AEJ - Alexandria Engineering Journal, 2021, 60(6):5009-5018.

[11] Akter S, Wamba S F. Big data analytics in E-commerce: a systematic review and agenda for future research. Electronic Markets, 2016, 26(2):173-194.

[12] Fan Y, Ju J, Xiao M. Reputation premium and reputation management: Evidence from the largest e-commerce platform in China. International Journal of Industrial Organization, 2016, 46(May):63-76