Retraction

Retraction: Construction of E-commerce Personalized Information Recommendation System in the Era of Big Data
(J. Phys.: Conf. Ser. 2074 012085)

Published 9 September 2022

This article has been retracted by IOP Publishing following an allegation that raises concerns this article may have been created, manipulated, and/or sold by a commercial entity. In addition, IOP Publishing has seen no evidence that reliable peer review was conducted on this article, despite the clear standards expected of and communicated to conference organisers.

The authors of the article have been given opportunity to present evidence that they were the original and genuine creators of the work, however at the time of publication of this notice, IOP Publishing has not received any response. IOP Publishing has analysed the article and agrees there are enough indicators to cause serious doubts over the legitimacy of the work and agree this article should be retracted. The authors are encouraged to contact IOP Publishing Limited if they have any comments on this retraction.

Retraction published: 9 September 2022
Construction of E-commerce Personalized Information Recommendation System in the Era of Big Data

Yaosheng Wang1,*

1Nanchang Vocational University, Jiangxi 330500

*Corresponding author e-mail: 38943770@ncvn.edu.cn

Abstract. With the continuous expansion of the scale of e-commerce, personalized recommendation technology has been widely used. However, the traditional recommendation system has been unable to meet the current needs of data processing, and good big data processing ability has become the basic requirement of the new personalized recommendation system. In addition, traditional recommendation systems are often limited to tangible goods recommendation, and pay less attention to e-commerce logistics service recommendation. In this paper, through the in-depth study of information personalized recommendation service in e-commerce environment, combined with the application background of big data: Taking the user dissimilarity matrix as the recommendation model, we propose IU usercf and UDB slope one recommendation algorithm. The two algorithms based on incremental update recommendation model have good scalability, can effectively deal with big data, and have high prediction accuracy. The proposed algorithm is applied to the actual system, taking e-commerce logistics service as the recommendation object and IU-usercf as the recommendation algorithm, the personalized recommendation system for e-commerce logistics service is constructed. The e-commerce logistics service recommendation system explores the application practice of recommendation algorithm under big data, and enriches the application scenarios of personalized recommendation technology.

Keywords: Personalized Recommendation, BigData, E-commerce, Incremental Update

1. Introduction

The era of big data has come. In fact, it has become an urgent need in the real world to seek effective big data processing technologies, methods and means. In the field of e-commerce, with the upsurge of online shopping, the number of e-commerce users has risen sharply, and the goods and services provided are also various. The e-commerce system will produce a large amount of information data every moment, and its data scale has shown the characteristics of big data for a long time. Taking user behavior information as an example, user behavior information is an important basic data for personalized recommendation system modeling, which belongs to the category of data collected by e-commerce system [1].

The rapid development of e-commerce has also brought the prosperity of e-commerce logistics.
The market demand for logistics services is increasing, and the third party logistics industry has been unprecedentedly developed. At present, there are tens of thousands of logistics enterprises in the domestic market, and each enterprise has its own characteristics in logistics services. In terms of providing high-quality services, national or even global large-scale logistics enterprises are not necessarily better than some regional small and medium-sized logistics enterprises. However, due to the problem of "information overload", users are submerged in the mass of information, it is difficult to find, and they are not willing to spend more time to find more suitable logistics services. Therefore, the vast majority of users are forced to choose some large logistics enterprises in the choice of logistics services, and give up the opportunity to obtain better services. At the same time, with the increasingly intensive service network of various logistics enterprises, the line combination is more and more, and the price, speed, security and other factors are often different. Coupled with the large number of user groups, the massive logistics service information and evaluation in e-commerce system are changing all the time. Each order of users also needs related logistics services, and the logistics information has presented big data features. Personalized recommendation system can collect, sort, classify and model the user's history, behavior and other related information resources according to the user's potential needs and interests, and then recommend the goods or services to the user. However, with the unprecedented development of e-commerce scale, the amount of data to be processed by personalized recommendation system has also shown the characteristics of big data [2].

To sum up, this paper first introduces the relevant theories and technologies of personalized recommendation and big data; proposes an incremental update recommendation model; proposes two recommendation algorithms based on the incremental update model; and finally constructs a personalized recommendation system for e-commerce logistics service.

2. Related Concepts

2.1. Personalized Recommendation System
Since the birth of recommender system, many scholars and institutions have carried out extensive and in-depth research on recommender system, and have defined recommender system from different perspectives. But if only from the perspective of e-commerce, it can be defined as: recommender system is an application software system that can recommend goods for e-commerce website. Based on the user's historical information or current behavior, using statistical analysis, machine learning and other means, it can select the goods or collection of goods that the user may be interested in from a large number of goods and recommend them to the user, so as to improve the quality of products Website visitors become buyers, improve cross selling ability and increase user stickiness [3].

2.2. Definition of Big Data
The concept of big data itself is abstract, and different research institutions have different understanding of it [4]. So far, the definition of big data has not been unified. In short, big data refers to the data with extremely large scale. As Wikipedia defines big data, big data refers to a data set that takes more time to acquire, manage and process data than can be tolerated by using common software tools. However, from the perspective of quantity, it is obviously too one-sided to distinguish it from the previous concepts of "massive data" and "super large scale data". Therefore, mm puts forward the famous big data 3V model, and thinks that big data has three characteristics: volume, diversity and velocity [5].

2.3. Incremental Update Recommendation Model
(1) Only update the places that need to be updated;
(2) Compared with the full update, the update is faster and the processing capacity is greatly reduced;

In order to adopt incremental update, it is necessary to set rules and specify the update strategy before updating [6].

Big data is a kind of data acquisition, management and processing using conventional tools or
software, which consumes more time than can be tolerated.

Generally, the big data processing technology will rely on distributed computing platforms such as Hadoop, but for the algorithm model itself, the incremental updating improvement can improve the efficiency of the algorithm itself in processing big data. The distributed computing platform adopts the idea of "divide and rule" for big data processing, which divides big data into pieces, and then carries out distributed parallel processing [7]. Incremental update, in simple terms, is a small amount of multi-batch processing, such as using the form of tag to mark the changed part of big data, and then extract one or more or even all of the tag data for processing, and continuously traverse to complete the processing of the whole big data. Incremental update and distributed computing platform complement each other in big data processing, which can greatly improve the processing efficiency [8].

At present, there is little research on incremental updating of recommendation model in the field of personalized recommendation, but incremental updating does play a very important role in big data processing and has a broad application prospect. The traditional collaborative filtering algorithm mainly uses the similarity measurement method for modeling, so this chapter first introduces the traditional similarity measurement method, and finds out the problem of model updating. Then, the incremental update recommendation model is proposed, and on this basis, it is further combined with Hadoop cloud computing platform to more perfectly realize the efficient processing of big data from the inside out of the algorithm model [9].

The basis of collaborative filtering algorithm recommendation is similarity matrix model. The similarity matrix is calculated by similarity measurement method based on user rating matrix.

Cosine similarity:

$$sim(i, j) = \cos(i, j) = \frac{i \cdot j}{\|i\| \cdot \|j\|}$$ \hspace{1cm} (1)

Modified cosine similarity:

$$sim(i, j) = \frac{\sum_{C \in L_k} (R_{i,c} - \overline{R}_i)(R_{j,c} - \overline{R}_j)}{\sqrt{\sum_{C \in L_k} (R_{i,c} - \overline{R}_i)^2 \cdot \sum_{C \in L_k} (R_{j,c} - \overline{R}_j)^2}}$$ \hspace{1cm} (2)

Pearson correlation coefficient:

$$sim(i, j) = \frac{\sum_{C \in L_k} (R_{i,c} - \overline{R}_i)(R_{j,c} - \overline{R}_j)}{\sqrt{\sum_{C \in L_k} (R_{i,c} - \overline{R}_i)^2 \cdot \sum_{C \in L_k} (R_{j,c} - \overline{R}_j)^2}}$$ \hspace{1cm} (3)

2.4. Incremental Update Recommendation Model

Item dissimilarity: the item dissimilarity of items I and j is calculated as follows:

$$dev_{i,j} = \sum_{x \in S_{j,x,i}} \frac{u_{j,x} - u_{i,x}}{\text{card}(S_{j,x,i})}$$ \hspace{1cm} (4)

User dissimilarity: user dissimilarity of user a and user B is calculated as follows:

$$dev_{a,b} = \sum_{x \in S_{a,x,b}} \frac{|u_{a,x} - u_{b,x}|}{\text{card}(S_{a,x,b})}$$ \hspace{1cm} (5)

After the dissimilarity between two users is calculated, the dissimilarity matrix model is obtained.
This model is an incremental update recommendation model, which is also the model to be built in this chapter [10]. If a new score is added, the score vector changes. At this time, the system marks the changed score vector, and the algorithm program calls the marked vector, then calculates and locally updates the matrix D. The specific updating strategies are as follows:

$$dev_{b,a} = dev_{b,a} + |b_i - a_i|$$

(6)

3. U-usercf and UDB Slope one Algorithms

3.1. Iu-usercf Algorithm

The prediction steps of iu-usercf algorithm are as follows: select the nearest neighbor by setting a threshold, for example, select the user whose dissimilarity is less than 1 and co-occurrence times is more than 5 as the nearest neighbor:

$$(dev_{uu} < 1 \land cout_{uu} > 5) \Rightarrow u' \in NN_u$$

(7)

The next step is to generate corresponding recommendations for target users according to the score of neighbor users. The calculation method is as follows:

$$P_{u,j} = \frac{\sum_{u \in NN_u} count_{uu} \times r_{ui}}{\sum_{u \in NN_u} count_{uu}}$$

(8)

3.2. Slope One Principle and Implementation

(1) Basic algorithm

Suppose that user y evaluates item F and does not evaluate item J. The Curse indicates that user y scores item F. now considering all items evaluated by user y, the following prediction formula can be obtained:

$$P(y) = \frac{\sum_{i \in R_j} (dev_{j,i} + y_i)}{card(R_j)}$$

(9)

(2) Weighted algorithm

Considering this situation, if 100 users over rate item 1 and item 2 at the same time, and another 1000 users over rate item 3 and item 2 at the same time, it is obvious that the item dissimilarity calculated based on the latter has higher reliability. Therefore, the weighted algorithm takes the number of users who have evaluated items I and j at the same time as the weight. At the same time, the more users who have evaluated items I and j, the more reliable the dissimilarity degree of items I and j is. A score prediction based on this dissimilarity degree is also more reliable. Therefore, a higher weight is given. The formula is defined as follows:

$$P^w(y)_j = \frac{\sum_{i \in R_j} (dev_{j,i} + y_i)C_{i,j}}{\sum_{i \in R_j} C_{i,j}}$$

(10)

4. Experimental Results and Analysis

4.1. Experimental Results of IU-usercf and UDB Slope One Algorithm
Table 1. The influence of co occurrence times on the prediction results of iu-usercf

| MAE | COUNT | Dataset_1 | Dataset_2 | Dataset_3 | Dataset_4 | Dataset_5 | Average |
|-----|-------|-----------|-----------|-----------|-----------|-----------|---------|
| DEV<=0.5 | >=0 | 0.81844777 | 0.8323203 | 0.8233749 | 0.8378361 | 0.8301466 |         |
|      | >=5  | 0.8009542 | 0.8098048 | 0.79624516 | 0.8072254 | 0.8112707 | 0.8051001 |
|      | >=10 | 0.7587955 | 0.7730422 | 0.75954413 | 0.7688363 | 0.77084947 | 0.76621354 |
|      | >=15 | 0.74596864 | 0.7342955 | 0.7198973 | 0.7351986 | 0.740198 | 0.7351116 |
|      | >=20 | 0.72505516 | 0.71084374 | 0.7053737 | 0.7092792 | 0.7246949 | 0.7150493 |
|      | >=25 | 0.7006608 | 0.68108946 | 0.6847732 | 0.7086302 | 0.7158696 | 0.6981446 |
|      | >=30 | 0.6916839 | 0.67614317 | 0.66839576 | 0.6740827 | 0.6890039 | 0.6798619 |

Table 2. The influence of user dissimilarity on iu.usercf prediction results

| MAE | COUNT | Dataset_1 | Dataset_2 | Dataset_3 | Dataset_4 | Dataset_5 | Average |
|-----|-------|-----------|-----------|-----------|-----------|-----------|---------|
| DEV<=0.5 | <=0.3 | 0.54270154 | 0.42857143 | 0.47622162 | 0.4810306 | 0.44295302 | 0.47429562 |
|      | <=0.5 | 0.72505516 | 0.71084374 | 0.7053737 | 0.7092792 | 0.7246949 | 0.7150493 |
|      | <=0.7 | 0.7313588 | 0.7300477 | 0.7269871 | 0.7211506 | 0.7379606 | 0.72950095 |
|      | <=0.9 | 0.74092805 | 0.7390904 | 0.7317639 | 0.7351338 | 0.74417704 | 0.73821867 |
|      | <=1.1 | 0.7628976 | 0.759573 | 0.7549254 | 0.7541421 | 0.76492697 | 0.75929296 |
|      | <=1.3 | 0.78249216 | 0.78242606 | 0.7769431 | 0.77437675 | 0.78507 | 0.78026164 |
|      | <=1.5 | 0.79609597 | 0.7987922 | 0.79355 | 0.78852975 | 0.8010643 | 0.7956065 |
As shown in Figure 1, under the condition of constant dev, with the increase of count, the MAE value becomes smaller and smaller, and the prediction becomes more and more accurate. This is because, in the case of a certain degree of user dissimilarity, with the increase of co-occurrence times, the neighbors found are more and more similar to the target users, but the number of neighbors will also be less and less. The more similar the neighbors are, the more accurate the prediction is. However, due to the small number of very similar neighbors, the number of items that can be recommended for target users is also limited. Therefore, in practice, we should choose a moderate co-occurrence frequency according to the characteristics of the data set, rather than the larger the better.

4.2. Algorithm Comparison

It can be seen from Figure 2 that the MAE score of UDB-Slope One algorithm is the lowest no matter on the dense data set movielens or the sparse data set book crossing, that is, the prediction accuracy is the highest. It can be seen that when using UDB slope one algorithm for prediction and recommendation, the user dissimilarity matrix model can provide high-quality neighbor set for the
algorithm from two aspects of user dissimilarity and co-occurrence times. Therefore, UDB slopeone algorithm can significantly improve the quality of recommendation system.

4.3. Construction of Personalized Recommendation System for E-commerce Logistics Service

Function module design:

(1) User behavior record module

The module is used to manage user behavior information, mainly the evaluation information of logistics service. It provides functions of adding, deleting, modifying, checking and marking information. The management information is mainly used to build and update the recommendation model.

(2) Model building module

For the initial input logistics service score data, the user dissimilarity matrix calculation formula is used to build the recommendation model. The module only needs to run in the initial stage of the system, and can be suspended after the model is built, and only the model update module is used.

(3) Model update module

For the newly generated score data, the logistics service score vector containing the new score data is called in the form of marking, and the incremental update formula is used to update the user dissimilarity matrix model locally.

(4) Recommendation algorithm module

The recommendation algorithm module calculates and forecasts the user's score of specific logistics services based on the user dissimilarity matrix according to the relevant information contained in the current user's order behavior, and recommends the logistics services that meet the requirements of commodity type and route and have high prediction score to the user.

System architecture design:

The system adopts B/S architecture, users access the server through browser. The front-end interface includes user interface and administrator interface. Among them, the front-end user interface is mainly integrated into the shopping interface of e-commerce system to realize the logistics service recommendation function for users. The administrator interface is mainly used by the administrator to maintain the related information of the recommendation system. The server is set up on the cloud platform, using the power of cluster for high-speed computing and storage, dealing with massive concurrent operations.

5. Conclusion

With the advent of the era of big data, large volume, fast, multi type of big data brings new challenges to personalized recommendation system. The traditional personalized recommendation technology does not consider the difficulty of big data processing, and even delays the response time of the whole e-commerce system due to the high occupation of computer resources, which seriously affects the user experience. This paper introduces the development and research status of personalized recommendation algorithm. The two algorithms based on incremental update recommendation model can not only deal with big data effectively, but also have high prediction accuracy. In order to solve the problem of less application of recommendation system in logistics service recommendation, in the fifth chapter, taking e-commerce logistics service as the recommendation object, a personalized recommendation system for e-commerce logistics service is constructed.

References

[1] Xu W, Sun J, Ma J, et al. A personalized information recommendation system for R&D project opportunity finding in big data contexts[J]. Journal of Network and Computer Applications, 2016, 59(JAN.):362-369.

[2] Tang M, Wu Z. Research on the mechanisms of big data on consumer behavior using the models of C2C e-commerce and countermeasures[J]. African journal of business management, 2015, 9(1):18-34.
[3] Ying, ZENG. On the Analysis of Library Information Ethics and the Standard Construction in the Era of Big Data[J]. Studies in Literature and Language, 2015, 11(4),1-3.

[4] Chen, Guobin. Recommendation Method of Educational Resources Under the Big Data Environment[J]. Journal of Computational and Theoretical Nanoscience, vol. 13, issue 4, pp. 2582-2587, 2016, 13(4):2582-2587.

[5] Yera R, Alzahrani A A, Martinez L. A food recommender system considering nutritional information and user preferences[J]. IEEE Access, 2019, PP(99):1-1.

[6] Jo S H, Lee J M, Park J, et al. The impact of renin-angiotensin-aldosterone system blockade on contrast-induced nephropathy: a meta-analysis of 12 studies with 4,493 patients.[J]. Cardiology, 2015, 130(1):4-14.

[7] Ravassa S, Arantxa González, Antoni Bayés-Genís, et al. Myocardial interstitial fibrosis in the era of precision medicine. Biomarker-based phenotyping for a personalized treatment[J]. Revista Espanola de Cardiologia, 2019, 73(3),13-16.

[8] Kim J S. Subway Congestion Prediction and Recommendation System using Big Data Analysis[J]. Journal of Digital Convergence, 2016, 14(11):289-295.

[9] Gao J, Yue X G, Hao L, et al. Optimization Analysis and Implementation of Online Wisdom Teaching Mode in Cloud Classroom Based on Data Mining and Processing[J]. International Journal of Emerging Technologies in Learning (iJET), 2021, 16(1):205-206.

[10] Borodinov A A. Development and research of algorithms for determining user preferred public transport stops in a geographic information system based on machine learning methods[J]. Computer Optics, 2020, 44(4):646-652.