ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0035

Research of Commodity Personal Recommendation Based on User Behavior

Zijiao Tang[1], Lecai Cai[2]

[1]Sichuan University of Science& Engineering, ZiGong 643000, China

[2]Sanjiang Artificial Intelligence and Robot Research Institute, YiBin University,644000, China.

Abstract

In the traditional product recommendation system, there is a problem that the similarity of different products of a common user set cannot be distinguished. An improved method for calculating the similarity of products is proposed to solve this problem. At the same time, the accuracy and coverage index are used to evaluate the recommendation effect. The results show that the improved similarity algorithm has improved the accuracy of personalized recommendation.

Keywords

Feedback behavior, Personalized recommendation, Similarity, Evaluation indicators.

1. Introduction

In China, as people's shopping habits change, online shopping has become more and more popular. From TaoBao.com and JD.com to Pinduoduo and many emerging platforms, from keen domestic shopping to cross-border shopping, this It has brought huge business opportunities to e-commerce merchants and foreign cross-border merchants. In order to allow consumers to get more precise consumption guidelines, major e-commerce companies have introduced endless technologies to better recommend their products. At the same time, more scholars have also researched related recommendation algorithms. Li Xiaoyu [1] introduced different similarity measurement methods in the collaborative filtering recommendation algorithm and summarized the main problems and solutions faced by the algorithm. Huang Tao et al. [2] proposed that the number of common scoring items between users as similarity An important indicator for calculation. At the same time, in order to reduce the quality of recommendation caused by data sparseness, a method for measuring the similarity between users by using structural similarity in complex networks was proposed. Qiao Yu et al. [3] in order to solve the recommendation system For the cold start problem, a collaborative filtering algorithm is proposed which combines user similarity and scoring information. Aiming at the recommendation accuracy problem of sparse matrix, this paper proposes a method for calculating the similarity of products. This method calculates the difference in the proportion of common score items. The larger the proportion of common scores, the greater the similarity, otherwise the similarity small.

2. User Behavior Characteristics

In the recommendation system, user behavior characteristics describe a series of user behaviors on products. Common behaviors include purchase, collection, browsing, searching, blocking, access time, and browsing path. Behavior characteristics reflect the mapping relationship between users and products. For example, the user purchases a mobile phone and reflects the characteristics of user-product purchase. User behavior characteristics [4-5] can also be divided into explicit feedback characteristics and invisible feedback characteristics. Explicit characteristics are that users clearly

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0035

express their favorite behaviors, and implicit feedback behaviors are that users do not express their preferences explicitly Information, but indirectly feedback, and the amount of data is relatively large. For example, the user's product browsing log on the website. However, on the one hand, hidden features can objectively express the user's logic and real thoughts; on the other hand, invisible features can be mined from explicit features. User behavior characteristics are the premise of the recommendation system, which can make the recommendation system understand the user's habits in more detail, so as to find out the problems existing in product recommendation, and help product recommendation to be more personalized and intelligent.

3. Personalized Recommendation Algorithm

3.1 Algorithm Description and Workflow

The user-based recommendation algorithm [6-7] is to recommend similar products to users and emphasizes recommending new products to users. Different from this, the personalized recommendation algorithm [8] recommends to the user products with high similarity to the products he likes before, which more personalize. Description of personalized recommendation algorithm: analyze the original data set, preprocess the data according to the requirements of the recommendation algorithm; construct a user-product matrix, calculate the similarity between products [9-12], and count users who like different products at the same time Quantity; personalized recommendations for users based on product similarity and user's historical behavior. The algorithm flowchart is as follows:

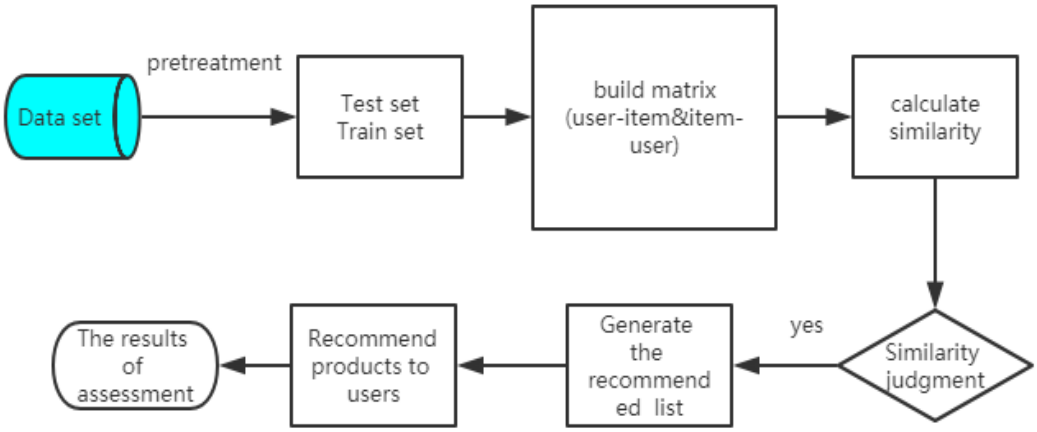

Figure 1 Recommended flowchart

3.2 Similarity Algorithm

The accuracy of product similarity calculation is very important. In actual data, the similarity is calculated by constructing a user-product matrix:

Table 1 User-commodity matrix

	User-1	User-2	…	User-k	…	…	User-n
Item-1	$V_{11}(1..\alpha)$		…	$V_{1k}(1..\alpha)$	…	…	$V_{1n}(1..\alpha)$
Item-2	$V_{21}(1..\alpha)$	$V_{22}(1..\alpha)$	…	$V_{2k}(1..\alpha)$	…	…	$V_{2n}(1..\alpha)$
…	…	…	…	…	…	…	
Item-k		$V_{32}(1..\alpha)$	…	$V_{3k}(1..\alpha)$	…	…	$V_{3n}(1..\alpha)$
…	…	…	…	…	…	…	…
Item-m	$V_{m1}(1..\alpha)$	$V_{m2}(1..\alpha)$	…	$V_{mk}(1..\alpha)$	…	…	$V_{mn}(1..\alpha)$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0035

In Table 1, the more users of any two products are evaluated, the greater the similarity; if most of the products are evaluated by the vast majority of people, this product is a popular product; if a product is popular, it There will be great similarities with other products. Therefore, the mathematical expression that defines the similarity of products is as follows:

$$sim(i, j) = \frac{N(i) \cap N(j)}{\sqrt{|N(i)||N(j)|}} \tag{1}$$

Among them, N (i) represents the set of users who have evaluated the product i, N (j) represents the set of users who have evaluated the product j, and molecule represents the set of users who have jointly evaluated the product. After the traversal calculation of the similarity of the items, the user's interest in the product is calculated. The mathematical expression is as follows:

$$Interest(u,i) = \sum_{j \in N(u) \cap S(i,num)} v_{ij} r_{uj} \tag{2}$$

Among them, N (u) represents the set of products evaluated by user u, S (i, num) represents the num recommended products similar to i by the user, and Vuj represents the user's interest in similar products j. Finally, in order to increase the variety of product recommendations, the similarity of the products is normalized. The mathematical expression is as follows:

$$normal\text{-}sim(i, j) = \frac{sim(i, j)}{\max_j v_{ij}} \tag{3}$$

3.3 Improved Similarity Algorithm

In the product-user matrix of the scoring records, the similarity between the unpopular products and the similarity of the popular products may be the same, and there may be the same similarity between the popular products and the sub-popular products. In formula (1), there is no Considering this situation, the improved algorithm differentiates the similarity of different products in a common user set by introducing a weighting factor. The weighting factor more emphasizes the proportion of products with common ratings in the two user ratings, making the similarity calculation more accurate. As shown in formula (4):

$$\partial(i, j) = \exp^{\frac{N(i) \cap N(j)}{N(i) \cup N(j)}} \tag{4}$$

4. Experimental Process and Analysis

4.1 Experimental Environment and Data

The experimental hardware platform in this article uses AMD Athlon 64 X2 processor, dual-channel 4G 1600MHZ memory, win10 operating system. The software platform uses the pycharm community edition and Python 3.7.2 as the interpreter.

In the following experimental process, the MovieLens movie rating data set published by the University of Minnesota was used. According to the specific conditions of the platform, 149,717 samples were randomly selected from the data set for experimental analysis, including 964 user ratings for 3952 movies. The moive-Ratings file is used as the experimental object. This file includes user ID, movie ID, and rating information. The data set is divided according to 87.5% of the training set and 12.5% of the test set.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0035

4.2 Experiment Analysis

the number of similar products recommended to users k in {4, 8, 12, 16, 20, 24, 28}, the actual recommended number g = 4, and adopt three indicators of accuracy, recall and coverage To evaluate the experimental results, under different K values, the comparison data of the new and old algorithms in the three indicators are shown in Table 2:

Table 2 Comparison of evaluation indicators between old and new algorithms

	Precision	Recall	coverage	N-Precision	N-Recall	N-coverage
k=12	19.44	4.5	5.84	21.91	5.3	5.39
k=16	20.52	4.88	5.66	22.22	5.17	5.86
k=20	17.9	4.32	5.54	22.05	5.19	5.26
k=24	16.98	4	4.9	24.85	5.98	5.74
k=28	22.52	5.21	5	23.15	5.56	5.58
K=32	17.75	4.14	5.55	23.92	5.57	5.52
K=64	18.06	4.18	4.42	22.99	5.36	5.49
K=128	19.75	4.51	4.93	23.77	5.4	5.7

In Table 2, the accuracy and recall ratio indexes reached maximum values at k = 28 and 24, respectively. The overall changes of these two indicators are consistent. At the same time, the accuracy and recall rate of the new algorithm have also improved significantly, indicating the addition of the new algorithm The calculation of product recommendations has played a positive role. Coverage describes the ability to recommend unpopular products. The coverage of the new algorithm gradually change as the value of K increases, and overall the ability to recommend unpopular products improves slightly.

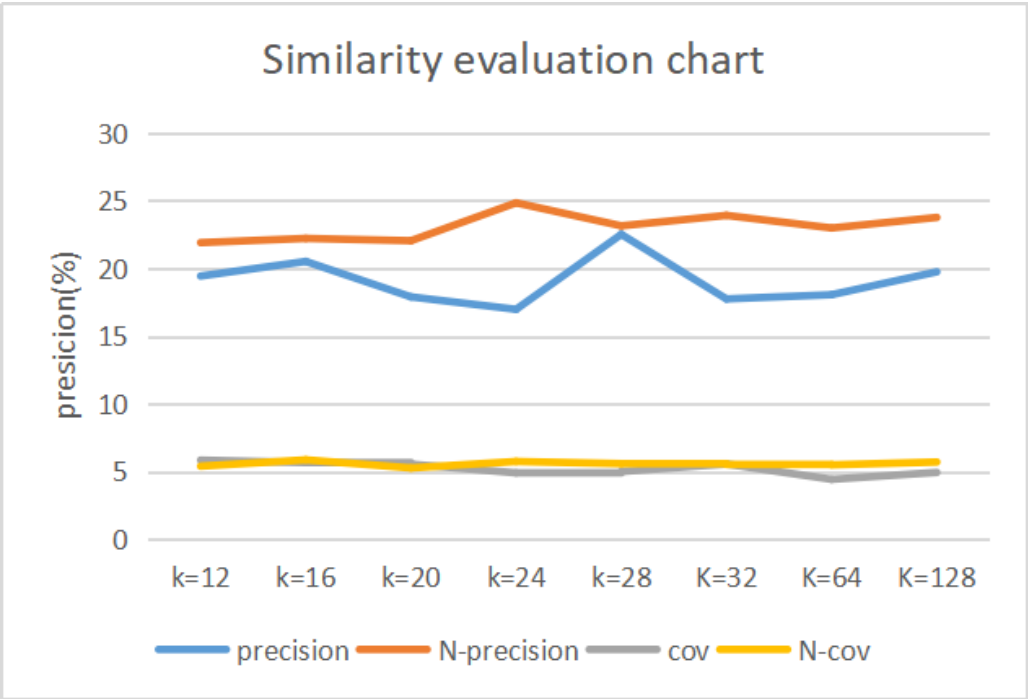

Figure 2 Similarity evaluation chart

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0035

The accuracy index describes the ratio of the number of common recommendation lists in the training set and the test set in the test set. The accuracy depends on the similarity calculation result, and the calculation of the similarity directly affects the accuracy. In order to test the effectiveness of the new algorithm for similarity calculation, set k to a value in the range [12-128], as shown in Figure 2. As a weight factor is added to the traditional algorithm, it accounts for the proportion of common score The distinction is made, so the accuracy of the new algorithm is improved, which verifies the effectiveness of the improved similarity algorithm.

5. Conclusion

Based on the problem of similarity calculation in traditional personalized recommendation system, this paper proposes an improved similarity algorithm. Through experimental comparison and analysis, it is proved that the new algorithm improves the recommendation accuracy to a certain extent, but the coverage has not improved significantly. The next phase of work will look at how to improve coverage.

References

[1] Li Xiaoyu,Research of Collaborative Filtering Recommendation[J]. Computer & Digital Engineering, 2019. 47(09): pages 2118-2122+2136.

[2] Huang Tao, Huang Ren and Zhang Kun,Improved Collaborative Filtering Recommendation Algorithm[J]. Computer Science, 2016. 43(S1): pages 400-403.

[3] Qiao Yu and Li Lingjuan,Collaborative filtering algorithm based on user similarity and rating information [J]. Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition), 2017. 37(03): pages 100-105.

[4] Wang Weiwei, Xia Xiufeng and Li Xiaoming,Personal interest degree model based on consumer behavior[J]. Computer Engineering and Applications, 2012. 48(08): pages 148-151+199.

[5] Jiang Zongli and Yu Li,Collaborative Filtering Recommendation Algorithm Based on User Features. Computer Systems & Applications[J], 2019. 28(08): pages 190-196.

[6] Guo Qiujun,Research on Recommendation Algorithm Based on User Behavior Feedback[D], 2019, Beijing Jiaotong University. page 68.

[7] Le man,Collaborative Filtering Recommendation Technology Based on User Behavior[D], 2019, Chongqing University of Posts and Telecommunications. page 76.

[8] Meng Liyun,Research on Personalized Recommendation of Dresses Based on Product Key Attributes[D], 2019, Zhejiang SCI-TECH UNIVERSITY. page 97.

[9] Li Zhongkun,An Improved Collaborative Recommendation Algorithm Research Based Optimized User Similarity[D], 2015, Hunan University. page 66.

[10] Xu Fengxiang, Implementation of a Collaborative Filtering Algorithm with Improved Similarity. Electronic Science and Technology, 2020(02): pages 1-7.

[11] Xu Xiaoyuan and Huang Li,Research on accuracy of distance metric learning algorithms. Modern Electronics Technique, 2019. 42(21): pages 150-153.

[12] Shen Jinxiang andBao Meiying,Optimal Recommendation Algorithm Based on User Clustering and Project Partition.Computer Systems & Applications, 2019. 28(06): pages 159-164.