International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 3 ǁ 2022 ǁ PP. 29-31

User Based Collaborative Filtering and Content-based Production Recommendation System

Jin Yang, Yizhou Han, Xinyu Xie, Tingting Liu
School of Management, Tianjin University of Technology, China

Abstract
The purpose of this paper is to design a product recommendation system for Tianjin University of Technology. The product recommendation system mainly uses the USERCF and CB recommendation algorithm and linear weighting, using reptile technology from the e-commerce platform's commodity assessment, using the USERCF computation correlation, using the CB recommended algorithm to calculate the product cosine Degree, the associated relationship and product cosine is similarly stored in the NOSQL HBase; user data and commodity feature data are stored in MySQL; and finally the TOP-N commodity bar is recommended.

Keywords: *User Based Collaborative Filtering, Content-based Recommendations, Mixing recommendations*

Date of Submission: 07-03-2022 Date of acceptance: 22-03-2022

I. INTRODUCTION

Recommended system by exploring the behavior of users, finds the user's personalized demand, thereby accurate long tail items to users who need it, helping users find them interested but difficult to find. In a online personalized recommendation system, when a user A requires personalized recommendation, you can find other users with him, and then I like those users, and the user A does not have heard the item recommended to A. This method is called a user-based synergistic filtering algorithm. The principle of content-based recommendation algorithm is based on user interested item A, using the content characteristics of users and item itself, such as user geographic location, gender, age, movie goods director, actor, release time, etc., find and A content good information B [1].

The mainstream recommended algorithms at home and abroad is mainly: content, synergy, collaboration, and mixing recommendations [2].

Based on the user's collaborative filtering algorithm is mainly divided into building a user-item score matrix, looking for a recent neighbor, generating three processes of recommendations [3][4][5].

The CB recommended algorithm only needs an important feature, ie the label [6].

Weighted mixing: The recommended results of each algorithm are combined according to a certain weight value, and finally the results of the four combinations of results are set to the user [7].

1.1.1 Initial data description

The initial data obtained includes user data and product feature data. User data includes user school numbers, user identity, and user's college, user purchasing records. Commodity feature data includes product numbers, product name, commodity characteristics. These data will store with MySQL [8].

1.1.2 Data processing

Direct use of the trade name is unable to do a recommended feature, you need to build a feature vector. Write multithreaded network crawl Get relevant data from the network. Multi-threaded reptiles are an efficient reptile, which can play a lot of potential. Specific implementation to the algorithm as follows:

i. Get the product name / bar code to the China Commodity Information Service Platform (https://www.gds.org.cn/#/home/index) Find the product details page.
ii. Use a crawler to find all the classification of the item in the product details page.
iii. Statistical related keywords.

1.1.3 Product feature vector

After getting a commodity classification label by crawler, put all keywords together to build a product keyword. For example: "Canned Coke" is classified in the product details page, no alcoholic beverage, canned, etc. From 01 vector, the vector of this product may be {drink type drink: 1, no alcoholic drink: 1, canned: 1, bottle: 0, alcoholic drink: 0 ...}. Therefore, the relationship formula of the two feature vectors X, Y is:

$$f(a,b) = \sqrt{(a-b)^2 + (a+b)^2}$$

$$d(X,Y) = \sum_{i=1}^{n} f(X[i], Y[i])$$

1.1.4 Cosine similarity

Commodities are broken down into a series of labels sufficient to explain the characteristics of the product, and establish a relationship between the user in the system (purchase, collection, view) to the goods and users. [9] Examples: User A likes canned cola, canned cola features the characteristics of drinking drinks, no alcoholic beverages and cans. According to the user's purchase of the item and the characteristics of the product, the user's preferences are calculated, and then compared with other items, the most characteristic intersection is the most intersive of the product 2 and the product 2 is recommended to the user.

The cosine similarity calculation formula of the CB recommended algorithm is as follows:

$$sim(A,B) = \frac{\sum_{i=1}^{n} A_i * B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} * \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Among them, Ai means that the user's preference for the type of goods, Bi represents the type of each item, which is non-zero 1.

1.1.5 Data storage

Since the column of the feature vector corresponds to a keyword, it is a variable amount, so the column variable data cannot be processed well for the conventional relationship database. Here is HBase, which is a non-relational database based on column-free storage, which can well meet the variable column BigTable requirements.[10]

II. RESULT AND DISCUSSION

1.2.1 Cold recommendation

Cold recommendation is recommended for historical data users. For university stores, through historical data analysis, users who do not have historical data can be labeled with appropriate artificial rules, specific:

i. Give new products with less than a small item, such as simultaneous recommendation of potato chips and beverages.
ii. According to the school's physical information, recommend different categories of goods. For example, a classmate to obesity recommends he may like high-calorie foods such as potato chips.
iii. Recommend several items in a week, such as the experimental report paper.

1.2.2 Feedback adjustment

For items recommended to the user, if the user clicks, then on the user's characteristic vector, the keyword of the item is high, which reflects the user's preference, which gives the user next time to make a recommended system It will be biased towards the item of this class. [11] At the same time, the feedback weight of the item is up to 1 (the base is 0), which when you are sorted by the product, if the relationship is the same, the feedback weight is raised in front, which reflects a product in a class of goods in a class of products. The higher the goods, the more the goods will be recommended.

1.2.3 Error recommended

Error recommendation does not have an error when it is recommended to recommend it, but deliberately recommend some products with this user without any relationship. The reason is: If a user uses a recommended system with the above-mentioned function, one time has a problem is that the user's circle will get smaller and smaller, that is, the user will be recommended to be more familiar or like. [12] And he just wants this thing. Over time, it is recommended to give the user to the user. There should be some randomness as a good recommendation system. From time to time, I will push the user to something that he has never touched. Maybe just find a new point of interest, sometimes this points of interest may not know. This is equal to expanding the recommended ability to expand the system, because the user's own point of interest is not recommended, the user may go to click, and find the user's own point of interest to enhance the system's recommendation ability, and can be a certain extent Increase sales. Of course, the error recommendation will sacrifice certain system accuracy, but the long-term look is good.

1.2.4 System Optimization

Every time the calculation is two two calculations, it takes time, so it is not possible to make an online recommendation. Now the practice is to do offline calculations, because the sales volume of the store is not particularly large, and the store is closed at 9:30, Offline recommendation calculations can be performed after the store is closed. Save to the corresponding database. [13] Even so, each recommended collection is still very big, and HBase is a relatively slowed database. There is a large delay every time you read from the database. Maintaining a Hash structure in the store, this structure stores a person ranked by about 70% of the number of goods purchases, and puts all of these people to the Hash structure in memory. Every time I recommend it to find it from memory. The cache is not hys again. It is greater than the memory of about 4G to reach the response time of around 16ms.

III. CONCLUSION

1.3.1 Recommended process

Summarize the above discussion, the specific recommendation process is:

i. Check if it is a new user, it is a small item and popular goods in accordance with the rules.
ii. Find the user's purchase record (if any), the other items that are 10% before the sales volume of history (sorting principles in feedback adjustment) are placed in the recommended collection.
iii. A person who has found a top 10% of the user-relevance, and the goods have been purchased and the goods have not been purchased to join the collection.
iv. Put the goods from the collection of goods to be recommended to put it in the recommended queue.
v. From the commodity library, find some items with this user without any relationship (2% of the total recommended amount).
vi. If the corresponding user data is read directly from the memory, the cache does not hit from the hard disk. Display the recommended queue to the user.
vii. Record the user's click to start feedback adjustment.

REFERENCES

[1]. Liang Xiang (2012). “Recommended system practice” Pp44-161.
[2]. LanXiang Liu (2021).”Getting Started with Python: Jieba Library”
[3]. YunLu Xia (2018). “Design and Implementation of Book Recommendation System”, Mongolian university, Pp.1-61.
[4]. QiLing Yan (2018). “Research on Personalization Recommendation Book”, Southwest University of Science and Technology, Pp1-61.
[5]. FangShu Shao (2018). “Personalized book recommendation based on collaborative filtering and related rules”, Zhejiang University of Commerce and Industry, Pp1-69.
[6]. ShuaiHua Nie(2015). “Design & Implementation of Intelligent Dating Website Based on Content Recommended / Collaborative Filter Recommendation Algorithm”, Huazhong Normal University, Pp1-67.
[7]. YonCai Tao, JunYan Li and Lei Shi (2016). “Geographical location-based personalized news mix recommendation research”, Small microcomputer system, Pp943-947.
[8]. C. Zhou, N. Ma, X. Cui, Z. Liu. The impact of online referral on brand market strategies with consumer search and spillover effect. Soft Computing, 2020, 24(4): 2551-2565.
[9]. C. Zhou, G. Xu, Z. Liu. Incentive contract design for internet referral services: Cost per click vs cost per sale. Kybernetes, 2020, 49(2): 601-626.
[10]. C. Li, M. Chu, C. Zhou, W. Xie. Is it always advantageous to add-on item recommendation service with a contingent free shipping policy in platform retailing? Electronic Commerce Research and Applications, 2019, 37: 1-11.
[11]. Z. Liu, C. Zhou, H. Chen, R. Zhao. Impact of cost uncertainty on supply chain competition under different confidence levels. International Transactions in Operational Research, 2021, 28(3): 1465-1504.
[12]. Z. Liu, R. Gao, C. Zhou, N. Ma. Two-period pricing and strategy choice for a supply chain with dual uncertain information under different profit risk levels. Computers & Industrial Engineering, 2019, 136: 173-186.
[13]. C. Zhou, W. Tang, R. Zhao. Optimal consumption with reference-dependent preferences in on-the-job search and savings. Journal of Industrial and Management Optimization, 2017, 13(1): 503-527.