Design and Application of Personalized Recommendation Algorithm Model Based on E-Commerce Platform Data

Qi Ding[a] and Zhigang Zhu[b]
City Institute, Dalian University of Technology, Dalian, Liaoning, China

Keywords: E-Commerce Platform Data, Personalized Recommendation, Collaborative Filtering Algorithm (CF), Data Model, Python.

Abstract: The algorithm models of K-means, Item-CF and User-CF based on Python environment can well realize various functions of personalized recommendation service system of e-commerce platform, and can formulate different recommendation service strategies for different user groups, which can effectively solve the problem of adaptation between e-commerce platform and users' needs and improve users' purchase efficiency and experience. Therefore, this paper takes the running data of e-commerce platform as the research object, relies on data processing class libraries such as Numpy and Pandas in Python environment, builds a personalized recommendation engine, and forms a personalized recommendation service system adapted to the call of Web Server through systematic encapsulation. Personalized recommendation service system will be between user I/O interface and e-commerce platform, and adopt MVC technology framework as the core design, and design API interfaces that can be called according to different application scenarios, so as to achieve a high degree of integration between recommendation system and e-commerce platform, meet the recommendation service strategy formulation requirements of e-commerce platform, and at the same time, it has good technical expansion performance in improving personalized recommendation.

1 INTRODUCTION

E-commerce, which combines the application advantages of computer science and technology, network information technology and remote communication technology, is a kind of commercial activity centered on commodity exchange in the virtual network environment. With all-round intelligent terminal equipment, high-speed and convenient network communication and safe and transparent mobile payment function, consumers enjoy the great convenience brought by constantly updated and abundant shopping information, but also have to suffer from information overload problems such as excessive information bombing, information expansion and information kidnapping. (Du, 2018)

On the one hand, the information overload problem of E-commerce originates from the explosive growth of E-commerce shopping platforms; on the other hand, it is caused by the increasingly perfect development of E-commerce platforms and the increasing data and information capacity. For the E-commerce platform itself, the function deployment is becoming more and more comprehensive, the whole system is becoming more and more complex, and the data interaction behind it is becoming more and more frequent. At the same time, there are more and more stores and brands carried by E-commerce platforms, and the number of commodities has reached the order of one billion. Diversified platform selection, rich search, comparison and screening functions, massive commodity information and a large amount of marketing and promotion information can help users to get the commodities they need, but it is often counterproductive, prompting users to have negative emotions such as anxiety and entanglement, which affects the formulation of consumption decisions. In addition, after more and more consumers have met their basic needs, their consumption behaviors are more mature and their consumption psychology is more stable, so that their consumption needs are

[a] https://orcid.org/0000-0001-7956-4041
[b] https://orcid.org/0000-0002-5782-4444

Ding, Q. and Zhu, Z.
Design and Application of Personalized Recommendation Algorithm Model Based on E-Commerce Platform Data.
DOI: 10.5220/0012030000003620
In *Proceedings of the 4th International Conference on Economic Management and Model Engineering (ICEMME 2022)*, pages 310-315
ISBN: 978-989-758-636-1

personalized and heterogeneous, which puts forward higher requirements for the experience of E-commerce platform, and has a great impact on the formulation of consumption decisions. How to break through the influence of E-commerce platform information overload on consumers' shopping decisions is a hot issue in the development of E-commerce to improve consumer experience and shopping efficiency. (Shang, 2021) In view of this, this paper believes that E-commerce platform can build a personalized and professional recommendation service system. With the application advantages of a large number of machine learning algorithms in data processing, it can quickly realize the design and deployment of recommendation engines in different scenarios based on user data, consumer behavior data and commodity data, support users with different roles to call different recommendation strategies through the Web server of E-commerce platform, and provide personalized information services and consumer decision support for E-commerce platform consumers.

2 OVERVIEW OF KEY TECHNOLOGIES

2.1 Data Source

The core function of E-commerce platform recommendation service system is to mine users' interests and preferences, predict users' implicit needs and explicit needs, and finally make personalized recommendations for users according to various actions triggered and executed by consumers on the platform. As a prerequisite for the function realization of the data source recommendation system, its quality and processing mode directly determine the running effect of the subsequent algorithm model. In the process of data source selection, common elements involve four parts: user information, commodity information, user consumption behavior and user scene. The examples covered are shown in Table 1.

Table 1: Common data information table of the E-commerce platform.

No.	Data elements	Example
1	Userinfo	Age, gender, education background, occupation, family composition, etc
2	Commodity information	Categories, brand, price, origin, weight, color, specification, unit, shelf life, etc
3	User consumption behavior	Browse, find, click, play, add shopping cart, collect, comment, retweet, etc
4	User's scene	Geographic location, system interface, time, specific festivals, major events, etc

The user information and commodity information come from the internal database of the E-commerce platform, while the user's consumption behavior and scene come from the user log file of the E-commerce platform. In addition, according to the characteristics of the E-commerce platform itself, the four data elements include structured data, semi-structured data and unstructured data. Semi-structured data and unstructured data can't be directly input into the recommended algorithm model for operation, so ETL (Extract-Transform-Load) tool is needed to complete data preprocessing, and after feature engineering, it finally meets the operation standard.

Generally, data preprocessing involves three stages: extraction, conversion and loading, aiming at integrating scattered, messy and inconsistent data in E-commerce platform, and transforming unstructured or semi-structured data into structured data, which will facilitate the subsequent data application. Extraction is the process of data collection, that is, the aggregation of all kinds of data required by the recommendation service system. For structured data, you can directly call or build a cross-reference map to complete the collection, while the log files need to be buried in the client interface and obtained by the log collection Web server. In the transformation stage, all kinds of data information should be cleaned, format adjusted, missing filled, deleted and repeated, and finally a data with uniform format, high structure, high data quality and good compatibility can be obtained. (Chen, 2016) In the final loading stage, the converted data can be transmitted to the recommendation system for storage and provided to the feature engineering stage of the recommendation algorithm for processing.

In order to make the recommendation results of recommendation service system more in line with users' real needs and satisfy users' personalized and differentiated consumption psychology, it is necessary to carry out feature engineering processing according to various data to form a unique user portrait model. The central idea of user portrait model is to extract users' hidden interests and preferences, and form labels to help the accuracy of personalized

recommendation services. The user portrait model will integrate four data elements: user information, commodity information, consumer behavior and user scene. The modeling results are shown in Figure 1, where U, A, I and T represent the collection of information elements, while u, a, i and t represent the specific information content.

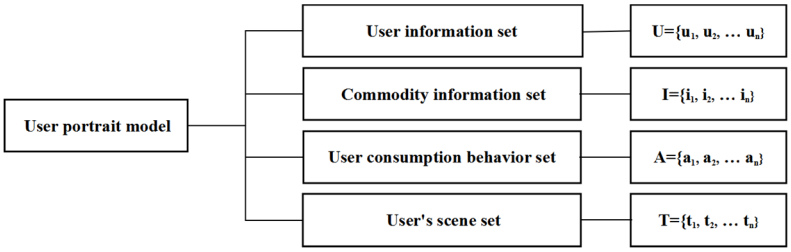

Figure 1: User portrait model structure diagram.

2.2 Recommendation Algorithm

Recommendation algorithm is the core content of recommendation service system, and it is the key to realize personalized recommendation technology. The quality of recommendation algorithm design fundamentally determines the effect of recommendation service, and also relates to the application effect of recommendation service system in E-commerce platform. In the actual application process, there are many algorithms that can meet the personalized recommendation service of E-commerce platform, and their design ideas, starting points and data analysis and processing processes are different. Common algorithm models include commodity association rule algorithm, demographic information rule algorithm and collaborative filtering recommendation algorithm.

Commodity-based association rule algorithm is a conventional recommendation technology, which can be widely applied to various E-commerce platforms. The principle is to generate association rules by analyzing the internal relations of all commodities added to the shopping cart in users' consumption behavior. When a certain commodity appears in the shopping cart of users, an associated commodity recommendation list is automatically generated, thus improving the cross-selling ability of commodities. The specific association rules are shown in Formula 1, where S represents the goods that trigger the rules, H represents the recommended goods, P represents the support, and C represents the confidence.

$$S \Rightarrow H(\text{p\%},\text{C\%}) \quad (1)$$

The algorithm based on demographic information rules is a relatively simple recommendation technology, which is suitable for new E-commerce users who have no historical consumption behavior. The user groups are classified only by the differences of user information, and then corresponding recommendation strategies are formulated according to different classifications. Most of the user information is text fields, which belong to unstructured data. When classifying information, it is necessary to use TF-IDF, Naive Bayes and other methods to complete feature extraction before realizing classification.

Collaborative filtering-based recommendation algorithm is the most widely used recommendation technology at present. The theoretical basis of collaborative filtering is that each user's interest preferences remain relatively stable within a certain time range, and each user is not independent, and can form a small group by virtue of the similarity of interest preferences. Therefore, the interest preferences of other members can be predicted by the interest preferences of one group member. (Liu, 2022) The implementation steps of recommendation algorithm based on collaborative process include data acquisition, nearest neighbor search and recommendation set generation. The user images formed after data collection and processing can represent users' interests and preferences. In the process of nearest neighbor search, two algorithms, user-based collaborative filtering (User-CF) and commodity-based collaborative filtering (Item-CF), can be constructed by using Euclidean Distance, Pearson Correlation and Cosine Similarity. After the algorithm is calculated, the similarity is sorted to form a recommendation list, and the results of TOP-N items are pushed to users.

Compared with the three recommendation algorithms, the association rule algorithm based on commodities is more convenient, but it can't adapt to the analysis and processing of massive data, and the degree of personalized recommendation is low, and the recommendation accuracy is insufficient. However, the algorithm based on demographic information rules has a rough overall processing

process, and the actual effect is not good. The integrity and authenticity of user information have a great influence on the recommendation results. Collaborative filtering recommendation algorithm has a high degree of automation and intelligence, which can adapt to large data sets and complete complex personalized recommendations, but it lacks good performance in dealing with the cold start of new users. On the whole, we can combine the rules of demographic system with collaborative filtering recommendation algorithm to achieve the purpose of complementary advantages, and finally improve the quality of personalized recommendation service.

2.3 Development Process

According to the above application requirements, we complete the configuration and deployment of the personalized recommendation service system development environment. The development content of the system is divided into two parts, one is the construction of hardware equipment and development environment, the other is the construction and training of each recommendation algorithm model in Python environment by using data processing class libraries such as Numpy and Pandas, forming a personalized recommendation system that can support Web Server calls.

First of all, in terms of hardware device selection, according to the system application requirements, the CPU selects intel core-i7-10700F @ 2.90GHz, 8 cores, 16GB memory and 1TB hard disk space. In the process of building the software environment, the operating system is Linux CentOS 4.7, and the compiling environment is Visual Studio 2019. Pyton 4.2.0 is selected as the development environment, and Anaconda integrated distribution can be selected for installation and deployment to shorten the time. In addition, the deployment and import of Numpy and Pandas need to be completed with the help of PyCharm tool, and after verification and configuration, the configuration of development environment can be completed.

Secondly, under PyCharm, according to the implementation process of personalized recommendation service, we complete data import, data preprocessing, user portrait model construction and related algorithm construction in turn. For example, in the process of constructing collaborative filtering algorithm, the design and development of User-CF and Item-CF will be completed according to user information and commodity information respectively. The key code for implementing the User-CF algorithm model is shown in Figure 2, and the similarity calculation method adopts cosine similarity evaluation.

```
import numpy as np
from math import sqrt
def yuping(lst_u,R,u_u,M,N,n):
    for i in range(M):
        if lst_u[i]!=0:
            a+=1#
            aver_1 +=lst_u[i
            aver_1=aver_1/a
    for o in range(N):
        if R[o][m]!=0:
            for i in range(M):
                if R[o][i]!=0:
                    b+=1
                    aver_2 +=R[o][i]
            aver_2=aver_2/b
            fenzi+=u_u[n][o]*(R[o][m]-aver_2)
            fenmu+=abs(pex(lst_u,R[o],M))
    return aver_1+fenzi/fenmu
```

Figure 2: Implementation code of user-based collaborative filtering algorithm (User-CF)

After the recommended algorithm model is built, a suitable training path is made in the train.py file, and a large amount of data is imported to complete the model training. After the whole system is developed, it will be packaged and published on the Web Server server of E-commerce platform. After the corresponding ports are configured, users can use the system from the client browser. Through the introduction of the above key technical theories, the construction environment, related software and implementation process of personalized recommendation algorithm are determined, and the technical feasibility of the whole project of personalized recommendation service system based on E-commerce platform is also clarified.

3 DETAILED FUNCTION IMPLEMENTATION

3.1 Data Collection Module

When the user logs in to the E-commerce platform, the personalized recommendation system will automatically start, and obtain the user information, the user's historical consumption behavior and the scene information of the user. After data preprocessing, a user portrait model is formed, and the corresponding interest preference features are input into the user feature database to be saved. Figure 3 shows the key code of missing data execution processing in the data preprocessing stage.

```
dfisnull(). any().sum()
df.drop('user_ geohash',axis=1,inplace =True)
df['date']= df['time'].str[0: 10]
df['time']= df['time'].str[11:]
df['time']= df['time'].astype(int)
df['hour']  =  pd.cut(df['time'],bins=[-1,5,10,13,18,24],labels=['in
the small hours','morning', 'noon', 'afternoon', 'evening'])
```

Figure 3: Missing key code of data processing in data preprocessing stage.

This process makes a basic distinction between new customers and regular customers. Because the new customers lack historical consumption data, the algorithm called in the follow-up personalized recommendation is different from the regular customers, further improving the three-dimensional and comprehensive personalized recommendation service.

3.2 Personalized Recommendation Module

When users realize information retrieval and browsing through E-commerce platform, personalized recommendation module will automatically call various recommendation algorithm models to complete personalized product information recommendation. The realization of this function will also be differentiated according to the roles of new customers and regular customers. Regular customers can combine User-CF and Item-CF recommendation algorithms, and improve the recommendation quality by weighting, transformation, combination, feature combination, cascading and other combinations. In addition, for new customers, personalized recommendation can be completed by using the rule algorithm based on demographic information alone, as shown in Figure 4, which is the key code for completing user clustering according to K-means algorithm in Python environment.

```
class KMeans:
    def __init__(self, k, times):
        self.k = k
        self.times = times
        for t in range(self.times):
            for index, x in enumerate(X):
                dis = np.sqrt(np.sum((x - self.cluster_centers_) ** 2, axis=1))
                self.labels_[index] = dis.argmin()
            for i in range(self.k):
                self.cluster_centers_[i] = np.mean(X[self.labels_ == i], axis=0)
    def predict(self, X):
        X = np.asarray(X)
        result = np.zeros(len(X))
        for index, x in enumerate(X):
            dis = np.sqrt(np.sum((x - self.cluster_centers_) ** 2, axis=1))
            result[index] = dis.argmin()
        return result
```

Figure 4: K-means clustering algorithm based on demographic information rules.

3.3 Personalized Interface

The personalized interface will complete the display of personalized recommendation results, and the final results will be presented in the form of product list. At the same time, this list is fed back to customers through the server, so that customers can know the personalized recommendation information in real time, improve their online purchase desire, and improve the sales performance of E-commerce platform. (Wang, 2019) After testing, the collaborative filtering algorithm and demographic information rule algorithm supported by the system can well meet the recommendation service function requirements of the general e-commerce platform. As shown in Table 2, the system shows the performance of Item-CF and User-CF algorithms in the test data set.

Table 2: The performance of Item-CF and User-CF algorithms in the test data set.

Number of neighbors (K)	Item-CF		User-CF	
	Similarity (Nsim)	Neibr ratio (Neibr)	Similarity (Nsim)	Neibr ratio (Neibr)
10	0.6241	0.3501	0.5607	1.0000
20	0.5884	0.3581	0.5274	0.9988
40	0.5657	0.3461	0.4766	0.9733

4 CONCLUSIONS

In order to realize the personalized recommendation system of E-commerce platform, the author takes the personalized recommendation algorithm model as the core, selects collaborative filtering algorithm as the main network model, and matches the association rule algorithm and demographic information rule algorithm. Under Python development environment, the author relies on Numpy and Pandas data processing class library to realize the construction and training of personalized recommendation service intelligent machine, and supports the call of E-

commerce platform Web server. The system starts automatically when users log in, predicts the products suitable for users and recommends them according to user behavior data information, promotes the effective transformation of users' implicit needs, enhances the cross-selling ability of E-commerce system, and provides higher consumption experience for online shoppers.

ACKNOWLEDGEMENTS

The Project of innovation and entrepreneurship for college students of Liaoning province (No.S202213198005).

REFERENCES

Chen Zongyan, Yan Jun. (2016) Collaborative Filtering Recommendation Algorithm Based on Sparse Data Preprocessing. Computer Technology and Development.06:59-64.

Du Huiyuan. (2018) Research on the Impact of Information Overload of E-commerce Platform on Consumers' Purchase Decision. University of International Business and Economics. 05.

Liu Hualing, Guo Yuan, et al. (2022) Research Progress of Similarity Algorithm in Collaborative Filtering. Computer Engineering and Applications.04:27-33.

Shang Mengying, Bai Zong, et al. (2021) Design and Implementation of E-commerce Personalized Recommendation System. Electronic Technology & Software Engineering.12:128-129.

Wang Min. (2019) Design and Implementation of E-commerce Personalized Recommendation System Based on Collaborative Filtering. Guangzhou University.05.