Research on Knowledge Fusion of E-commerce User Data

Meng Cong

Jinan Vocational College, Jinan, Shandong, China

Keywords: E-commerce; User data; Knowledge fusion

Abstract: Traditional data processing thinking mode and technology can not meet the needs. Big data technology is entering our life with an irresistible trend. In this paper, the basic principles of the O2 O e-commerce model are elaborated, and the user data characteristics, data mining and analysis of the E-commerce model under the big data environment are analyzed, as well as the application of the big data in the E-commerce model. Then, the platform data is used to analyze the relevance of the data and recommend the products. Finally, the development opportunities of the business model in the big data environment are proposed.

1. Introduction

The rapid development of computers and the popularity of the Internet provide residents with the premise of online shopping. The rapid development of mobile network, rapid logistics and transportation, and the convenience of electronic payment further promote the development of online shopping. The improvement of relevant laws, regulations and e-commerce system makes more residents trust online shopping. In China's e-commerce market, C2C market is developing steadily, B2B market is still dominant. B2C market benefits from the development of mobile e-commerce, its growth is rapid, and the market of O2 O is expanding constantly. Residents'shopping habits have changed from having to go to physical stores to online shopping more convenient and affordable, and the concept of online shopping has changed from curiosity about online shopping and pursuing lower prices to the requirements of commodity quality, shopkeeper service quality, shopping user experience and the convenience and speed of logistics. Thus, consumers'demand for online shopping is increasing. How to gain a place in the current fierce competition of e-commerce, win more consumers, maintain the continuous growth of turnover and so on are a series of problems faced by all e-commerce enterprises.

On the one hand, a large number of transactions have brought huge profits to commercial enterprises, on the other hand, a large amount of data has also brought them some problems. At present, the e-commerce industry agrees that e-commerce has entered the era of big data, but facing the huge amount of data, the efficiency of data capture and utilization of e-commerce enterprises is very low. The commercial value of data is far from being developed. In the era of data explosion, how to make use of data and bring business value to enterprises, and how to change traditional business thinking will be the difficult problems that every e-commerce enterprise will face.

2. Data Mining Procedure and Method Analysis of Oxygen Dioxide Business Model in Large Data Environment

2.1 Data Mining Framework Analysis

Owing to the characteristics of the user data of O2 O, e-commerce enterprises can not use traditional technology to efficiently analyze their information to obtain useful business value. Nowadays, how to derive the effective value from the huge and complex user information has become one of the important means of competition among the O2 O e-commerce enterprises. Big data mining is one of the tools. It is a technology that can get some rules and conclusions by analyzing massive data.

The framework of user mining for O2 O e-commerce includes three aspects: data source, data preprocessing and data analysis. Data preprocessing includes data organization, data collection and

  DOI: 10.25236/aisct.2019.038

data storage. Data application adopts the object-oriented approach of O2 O, including platform-oriented, user-oriented and business-oriented applications.

2.2 Analysis of Data Mining Procedure for Oxygen Business Model

In the whole process of data mining, the focus is to solve such a problem, that is, to analyze the characteristics of various groups of users in the O2O e-commerce, so as to analyze the characteristics of individuals, analyze certain laws, and then use them to transform into useful business value.

2.3 User Data Collection

User data sources of O2O mainly include user data of various O2O platforms, social networks and mobile devices. Data is generated in the form of “streams” in the process of users using the O2O platform and other social platforms. Because these three user data sources are overlapping and similar, some of the data content presented will be overlapping. Therefore, when collecting data, they are usually classified into three categories: transaction, interaction and observation data.

(1) User data preprocessing

Because the original data can't be used directly, there are some problems such as noise, redundancy and missing, so data should be processed to supplement some missing values in data preparation, so as to improve the value of data utilization. The next step is the data transformation process. Structured data is better processed. It filters structured data, extracts valid data and eliminates invalid data. In data, unstructured and semi-structured data can not be analyzed by traditional data processing tools. Therefore, in the process of data transformation, these semi-structured and unstructured data will be transformed into machine language or index, and then into weighted logic or fuzzy logic, mapping to standard values through different keywords.

Then there is the data extraction process, which is to detect the correlation and correlation between data. Most users'characteristics can be represented by the relativity of data, which can be used for user personalized recommendation. Use relevant data to gain effective value and bring more business benefits to enterprises.

(2) User Data Mining and Its Application in O2O E-commerce

Businessmen usually use data mining to predict future trends. Different data mining models are selected according to different needs of businesses. In data mining, there are three main types of models: association rule analysis, clustering analysis and classification analysis. At present, user model is also a common model, which classifies people's age, gender, interest, and race. In the user data mining of O2O e-commerce, the commonly used methods are association rule analysis, change and deviation analysis, classification and clustering analysis, etc. The results of data mining will be interpreted and applied [2].

General data mining applications include personalized recommendation, Web mining and search, anomaly detection, large data visualization analysis and so on.

3. Analysis of Data Mining in the Model of Oxy Electronic Commerce

Deep data mining and data analysis can benefit all participants in the O2 O e-commerce model. For O2 O users, more accurate data analysis can provide them with more affordable, timely and personalized services. For O2 O merchants, the analysis of data mining can enable them to grasp market dynamics more timely and effectively and respond quickly to various changes. For the O2 O platform, effective data mining can enable them to develop more accurate marketing strategies to obtain more business information. Using a large amount of data of user's consumption behavior, accurately locate the user's consumption habits, interests and so on, thus predicting the user's next consumption behavior, and using this platform will make specific content marketing for users. This kind of precise marketing is different from traditional marketing. It can save more cost and get better marketing effect. It can win more consumers and more business resources. In addition, the platform can optimize the platform website with the help of user data mining. For example, by mining the characteristics of platform visitors, we can modify the structure and appearance of the

website; by bundling commodities with a certain degree of support and trust, we can promote sales; the platform can also add links between closely related pages, rationally arrange the caching strategy of server pages, and improve user browsing satisfaction. In addition, through data mining, the O2 O platform discovers and retains customers to stabilize customer relationships. Data mining technology can also be used to analyze fraudulent businesses to prevent network fraud.

4. Application of Data Mining in the Model of Oxy Electronic Commerce

4.1 Data Acquisition and Preprocessing

The data of this experiment is generated through the research data. A total of 88 163 purchase data of users are collected, and the corresponding items have been corresponding to the number. Hadoop model and Apriori algorithm are used to analyze data association. Because of the clutter, repeatability and incompleteness of the data, the data will be pre-processed before processing the data.

32 39 41 48 174 175 176 177 178
32 38 39 47 48 179 180 181 182 183
39 184 185 186
36 38 41 48 140 187 188
39 48 186 189 190 191 192 193 194 195 196 197 198 199 200
39 201 202 203 204 205 206 207 208 209
39 65 193 210 211 212 213 214 215
179 216 217 218 219 220 221 222 223 224
225 226 227
39 41 48 228 229 230 231
36 38 39 232 233 234 235 236 237 238 239 240 241 242
39 243 244 245
39 41 48 246 247 248 249 250
39 48 65 251 252 253
48 230 254

Figure 1 Raw data

4.2 Hadoop model and related introduction of Apriori algorithm

Hadoop is an open source framework developed by the Apache Foundation and is a well-known cloud computing model in data mining. Hadoop is typically used to process large amounts of data and can also be used to write and run large distributed data processing. Apriori algorithm is the most influential algorithm in the field of association rule mining. Many current algorithms, such as Apriori Tid, DHP, Apriori Hybrid, etc. are all improvements or extensions based on Apriori algorithm. The Apriori algorithm utilizes an iterative method of layer-by-layer search to find itemsets that frequently appear in a transaction database.

Apriori algorithm description

Step 1: Generate a set L1 of frequent 1-item sets.

By scanning the transaction database D, the support frequency of each 1-item set is calculated, and the set L1 of the frequent 1-item set is obtained.

Part 2: Connecting Steps

In order to generate a set LK of frequent K-item sets, a set CK of the candidate K-item sets is required. CK is a set of items that may become a frequent K-item set, which is obtained by JOIN operation from two items in the set of frequent (K-1)-item sets L_{K-1}, ie $L_{K-1} \infty L_{K-1}$. If p, $q \in L_{K-1}$, $p=\{p_1, p_2, ..., p_{K-2}, p_{K-1}\}$, $q=\{q_1, q_2, ..., q_{k-2}, q_{k-1}\}$, and when $1 \leq K-1$, pi = qi, when i = k-1, $p_{k-1} \neq q_{k-1}$, then $p \cup q = \{p_1, p_2, ..., p_{k-1}, q_{k-1}\}$ is candidate k - Elements in the set Ck of the item set.

The third step: pruning step

Since Ck is a superset of Lk, elements of Ck that are not frequent itemsets need to be eliminated.

If a set X is a frequent item set, then it is a subset of all frequent item sets. So, if a candidate's K-set (k-1) is not a member of LK-1, then the K-item set is uncommon and can be removed from Ck. After pruning, the database is scanned, the frequency of the item set in Ck is calculated, the non-conforming frequency item set is eliminated, and finally the set Lk of frequent K-item sets is generated. By iterating through the loop, repeating the steps to connect the pruning step until the new frequent itemsets (non-empty) can no longer be generated, they get the frequent itemsets that satisfy the minimum support [2].

4.3 Data Processing

Experimental platform: Windows+jdk+inteliij idea

4.4 data processing results

In the case where the degree of support is 5%, the candidate set of conditions is satisfied.

```
[32 38 39] => [32, 38, 39, 48] [X]-1840 => [X,Y]-1236 CONFIDENCE-RATE = 67.17391
[32 38 48] => [32, 38, 39, 48] [X]-1646 => [X,Y]-1236 CONFIDENCE-RATE = 75.09113
[32 38] => [32, 38, 39, 48] [X]-2833 => [X,Y]-1236 CONFIDENCE-RATE =  43.628662
[32 38] => [32, 38, 39] [X]-2833 => [X,Y]-1840 CONFIDENCE-RATE =  64.948814
[32 38] => [32, 38, 48] [X]-2833 => [X,Y]-1646 CONFIDENCE-RATE =  58.100952
[32 39 41] => [32, 39, 41, 48] [X]-2359 => [X,Y]-1646 CONFIDENCE-RATE = 69.77533
[32 39 48] => [32, 39, 41, 48] [X]-5402 => [X,Y]-1646 CONFIDENCE-RATE = 30.470196
[32 39] => [32, 38, 39, 48] [X]-8455 => [X,Y]-1236 CONFIDENCE-RATE =  14.618568
[32 39] => [32, 38, 39] [X]-8455 => [X,Y]-1840 CONFIDENCE-RATE =  21.76227
[32 39] => [32, 39, 41, 48] [X]-8455 => [X,Y]-1646 CONFIDENCE-RATE =  19.46777
[32 39] => [32, 39, 41] [X]-8455 => [X,Y]-2359 CONFIDENCE-RATE =  27.90065
[32 39] => [32, 39, 48] [X]-8455 => [X,Y]-5402 CONFIDENCE-RATE =  63.89119
[32 41] => [32, 39, 41, 48] [X]-3196 => [X,Y]-1646 CONFIDENCE-RATE =  51.501877
[32 41] => [32, 39, 41] [X]-3196 => [X,Y]-2359 CONFIDENCE-RATE =  73.81101
[32 41] => [32, 41, 48] [X]-3196 => [X,Y]-2063 CONFIDENCE-RATE =  64.54944
```

Figure 2 data processing results

4.5 Data Analysis

(1) From the results, it can be concluded that under the assumption of 60% confidence, when users purchase 32 and 38 products, the probability of users buying 1236 products while buying 32 and 38 products is higher than that of 1840 and 1646 products because the confidence of 1236 products is higher than that of 1840 and 1646 products.

(2) From the results, it can be concluded that under the assumption of 60% confidence, when users purchase 32 and 39 products, the probability of users buying 5402 products is higher than that of 1236, 1840, 1646 and 2359 products, because the confidence of users buying 5402 products is higher than that of 123, 1840, 1646 and 2359 products.

(3) From the results, it can be concluded that, under the assumption of 60% confidence, when users purchase products No. 32 and No. 41, the probability of users purchasing goods No. 2359 while purchasing goods No. 32 and No. 41 is higher than that of goods No. 1646 and No. 2063.

References

[1] Yuan Yuan, Zhao Ruihua. Analysis of the application value of big data to O2O e-commerce. [J]. Communication World, 2016, 2.

[2] Zhao Wei. Analysis and Prospect of O2O Mode in E-commerce Era [D]. Shanghai: University of International Business and Economics, 2015.

[3] Gao Haijian. Research on e-commerce industry based on big data perspective [D]. Beijing: Capital University of Economics and Business, 2015

[4] Qin Yanhua. Discussion on the Development of E-commerce under the Big Data Era[J]. Information and Computer, 2016 (2): 19-20

[5] Wang Huimin. Value Creation and Model Innovation of E-commerce under the Background of Big Data[J].Commercial Economic Research, 2015 (7):76-77

[6] Mo Yuqing. The significance of the “Internet +” in the government work report of the two sessions [N]. Computer and Network, 2015.

[7] Liu Xiaohong. Research on O2O Operation Mode of China's E-commerce--Taking H Company as an Example [D]. Jilin: Jilin University, 2015.

[8] Wang Jiajia. Big Data Strategy for Physical Retailing in New Markets [N]. China Business News, 2013, 01 (A03) .

[9] Zilong Fang, Pengju Li, The Mechanism of “Big Data” Impact on Consumer Behavior [J]. American Journal of Industrial and Business Management, 2014, 4:45-50.