Precise Marketing Data Mining Method of E-Commerce Platform Based on Association Rules

Zhang, Hong-ni; Dwivedi, Ashutosh Dhar

Document Version
Accepted author manuscript

Published in:
Mobile Networks and Applications

DOI:
10.1007/s11036-021-01886-3

Publication date:
2022

License
Unspecified

Citation for published version (APA):
Zhang, H., & Dwivedi, A. D. (2022). Precise Marketing Data Mining Method of E-Commerce Platform Based on Association Rules. *Mobile Networks and Applications*, *27*(6), 2400-2408. https://doi.org/10.1007/s11036-021-01886-3

Precise Marketing Data Mining Method of E-commerce Platform Based on Association Rules

Hong-ni ZHANG[1], Ashutosh Dhar Dwivedi[2,*]

[1]Shandong Polytechnic, Jinan, 250000,China

788@sdp.edu.cn

Centre for Business Data Analytics, Department of Digitalization, Copenhagen Business School, 2000, Frederiksberg, Denmark

*add.digi@cbs.dk

Abstract. Today, marketing data management and scheduling capabilities for mass e-commerce platform are required. In this paper, association rules are introduced in the precision marketing data mining for the e-commerce platform. First, the multi-source of precision marketing data of e-commerce platform is analyzed, where the collection and attribute analysis of multi-source data is carried out by optimizing the backstage hardware of e-commerce platform. Then, the association rules are introduced by using the means of information fusion. Meantime, the distribution similarity is used to mine the precision marketing data of e-commerce platform. Finally, the mining results are verified by residual and post checking to ensure the mining effectiveness. Experimental results show that the output of mining parallel scheduling ability can be effectively achieved precise marketing data mining with high robustness.

Keywords: Association rules; E-commerce platform; Precision marketing; Data mining.

More and more enterprises through e-commerce platform to carry out transactions, settlements and other business activities, online consumption and investment activities are also increasing the number of people. People's daily activities, including basic necessities of life, booking tickets, education and other activities, can be met through the network, e-commerce and consumer integration more and more closely. By June 2014, e-commerce transactions in China had exceeded RMB5.8 trillion, with more than 670 million netizens and 310 million online shoppers. However, while hundreds of millions of consumers shop on the Internet every day, tens of thousands of businesses do business on the Internet. For each business, how to segment the market, accurately locate the products and market, seize the old customers, find new customers, and stand out in many enterprises has become a top priority for every enterprise, the problem placed in front of the home is how to find out the hidden, unknown and valuable information from a large amount of data in the database[1-3].

In response to such e-commerce issues, the literature [4] found that a growing number of Chinese consumers are turning to the online market for foreign seafood. When buying seafood online, customers are unable to actually evaluate the product, and the market page, rather than the seafood label, conveys all the information about the product. As a result, some progress has been made in analysing e-commerce data using DNA barcode technology. Document [5] Design the platform for monitoring the operation data of subsea cable tunnel, introduce the functions of video monitoring system, fiber temperature

measurement system, subsea cable stress monitoring system and tunnel comprehensive monitoring and early warning system, and point out that the centralized monitoring mode should be adopted for comprehensive monitoring to improve the application efficiency and reliability of operation data.

Data mining is a hot topic in the field of database and artificial intelligence. It has been widely used in various fields and industries, including production, marketing, customer relationship management in the business field, investment and financing evaluation in the financial field and stock trading, route planning in the logistics field and weather forecast, etc. In the field of education, the management in the field of higher education is of significance to the analysis of the employment situation of students and graduates. The raw data needed by data mining can be structured data, or semi- structured data, such as graphics, images, text, etc., or heterogeneous data in the network. Algorithms used in data mining include classification analysis, regression analysis, clustering analysis, web mining, early warning analysis and so on. These algorithms simulate people's thinking logic, such as induction and deduction, mining data from different angles to meet customer segmentation information needs, customer behavior prediction, feature discovery, risk early warning and so on.

1 E-commerce platform precision marketing mode

In the precise marketing mode, the marketing subjects shall, by collecting consumer data to conduct computer statistics, establish and store big data database, formulate marketing programs, accurately analyze the preferences of customers for consumer goods, optimize the consumption process according to customer needs, establish a good trust relationship with consumers, actively promote products, attract the attention of consumers, constantly enhance the brand awareness of consumers, and improve and develop new products. Then, according to the existing marketing conditions to meet the needs of consumers, big data shall be used to classify and analyze consumer groups. When the existing marketing methods fail to meet the needs of consumers, high- tech means shall be used in the marketing model to carry out a series of marketing activities such as online direct marketing and video marketing, and different types of marketing channels shall be opened up, such as online platform marketing and online marketing, the specific contents are as follows:

(1) Products: When producing consumer goods, enterprises pay attention to product quality and performance, and increase consumers' awareness of and trust in the products. In product marketing and publicity, enterprises are committed to promoting the accurate circulation of their products, building their own image with the content of their products, and increasing the sales performance of their products; when implementing sales activities, enterprises use big data value capability to transmit product information to consumer groups by using computer network platforms, recommend and build the image of their own brands to consumers, and formulate their own marketing plans based on the needs of consumer groups and with the characteristics of their products; and they establish consumer files mainly to record the daily activities of customers and the way to purchase products, as well as the evaluation of purchased products, which can enable enterprises to better understand consumer groups, develop products suitable for consumer groups from the perspective of consumer group needs, and promote the accurate marketing of products.

(2) Channels: in accordance with the precision marketing strategy set up by social networks, an important platform for precision marketing consumer groups to snap up products on social networks has been established to find various market channels in marketing, strengthen market flexibility, and constantly expand and develop new marketing channels. Make use of Internet platforms to create

personalized websites belonging to their own enterprises, strengthen the columns for interaction with consumers, enable enterprises to effectively communicate with consumers in a timely manner, facilitate enterprises to recommend their own products to consumers, enhance the trust between consumers and enterprises, and open the consumption mode of consumer goods; enterprises publicize their own product information on the official websites established, promote their own products, make use of Internet platforms to interact with consumer groups, timely grasp the feedback information from consumer groups, and adjust the marketing strategy in a timely manner during the marketing process, improve the product quality of enterprises in the production of products, and meet the needs of consumer groups. When promoting consumer goods on the network platform, enterprises combine online promotion with offline activities to create an excellent image of promoting their own product brands and inspire more online consumers to actively participate in the activities; at the same time, businesses use websites to disseminate information and increase product publicity efforts to strive for better and faster dissemination effects.

(3) Price: First, the marketing price shall be accurately set in the marketing of consumer goods, and the precise marketing mode shall reach the customers' hands at the fastest turnover rate, thus reducing the transit expenses and lowering the sales price; secondly, the construction of the website allows the consumers to fully understand the performance and price of the products they want to purchase when they purchase the products, and the two parties directly communicate and communicate with each other in a timely manner on the website platform so as to accurately solve the price-related problems of the products; while the marketers use the network platform to understand the consumers and give the consumers a satisfactory and reasonable price.

(4) Promotion: the enterprises have changed the traditional consumption mode on the sales platform, promoted the online promotion of consumer goods, and in the process of spreading the product information of the enterprises, created and improved the brand image of the products by means of implantable advertisement so as to let people enhance the product image of the enterprises while having leisure and entertainment; the enterprises have used the websites to spread the information, intensified the publicity of the products and intensified the broadcasting effect; the enterprises have accurately attracted the consumer groups to the social network, used the existing consumer groups to carry out secondary communication, promoted the products under the conditions of guaranteeing the quality and profits, and expanded the user groups of the products.

To sum up, enterprises in the marketing of consumer goods and network as a medium to communicate with consumer groups, understand the consumer groups, consumers need product information, in order to better promote product sales. The formulation of precise marketing strategy for consumer goods has been widely publicized, has played an active role in promoting the products, has not only saved costs for enterprises, but also improved consumers' trust in the products of enterprises, has satisfied consumers' desire for purchasing products, has stimulated consumers' desire for purchasing products, has improved the business level of enterprises in marketing and customer service, has improved the overall marketing and service capabilities of enterprises, and has truly realized the low-cost and high-efficiency marketing strategy.

2 Analysis and collection of precise marketing data attribute of e-commerce platform

2.1 Hardware optimization of e-commerce platform background

According to the above analysis results of precision marketing mode, precision marketing data of E-commerce platform is multi-source. According to the requirement of multi-source data collection, the original system hardware framework is optimized to improve the level of data collection and increase the usability, so as to ensure the realizability of the subsequent system software module development.

The original system hardware framework is shown in Figure 1.

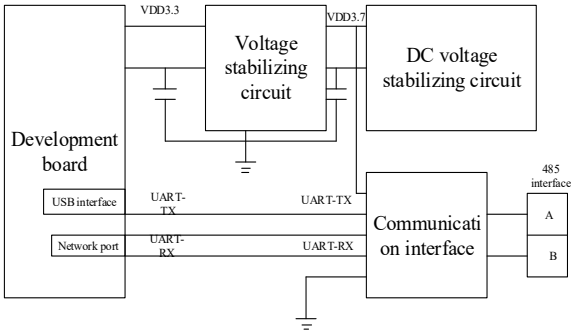

Fig 1. Original system hardware framework

Using the above results as the basis of hardware design, the selection of system hardware is completed

First of all, in this design, the main control chip of the data acquisition system is designed, and ARM chip is embedded in the original system hardware structure, so as to realize the functions of multi-source data acquisition, preprocessing and temporary storage.

ARM chip is a RISC processor that supports multiple instructions processing at the same time, in which multiple semiconductors can be set. In this design, S3C2440 chip is selected as the core processing chip: in this chip, the advanced control bus structure is set, which contains 200 pins and is sealed in FBGA mode. The core of the chip is composed of arm9tdmi processor core, instruction high-speed and data efficient cache, MMU storage management and AMBA bus. In order to ensure the stability of the system, the corresponding terminal controller is installed in the chip to support the system emergency interrupt service.

When the terminal controller is running, considering the economy of the system design, the system needs to provide two preprocessing modes to work at the same time, that is, the dual output power chip is used to provide two-way voltage, and the dual output power chip is set as tps67d300 [6]. The chip has a single power supply dual output performance, high voltage accuracy, output current below 1a, with a typical quiescent current. In addition to the above power circuit design, the system acquisition circuit is optimized. The specific optimization results are shown in Figure 2.

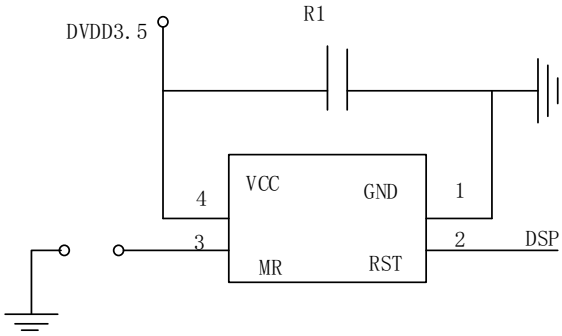

Fig 2. Acquisition circuit supporting dual preprocessing mode

In Figure 2, according to the characteristics of the selected main control chip, the acquisition of this system has two different dual preprocessing modes. By comparing the results, the button acquisition mode is selected to complete the data acquisition. When the main control chip is powered on, the system will collect the chip once. The chip is set in the form of multiple pins. A watchdog circuit and a 16 bit adder counter are installed under the chip. When the count value is the maximum, the watchdog will send out the corresponding output pulse to realize the acquisition of the main control chip. When the data acquisition is finished, all the current operations are terminated to make the main control chip enter the initial state, refresh the system operation behavior data in real time, collect all the data and transfer it to the memory.

Using the original system hardware development board, set the corresponding chip peripheral circuit, the circuit is mainly composed of power supply circuit, core chip circuit and communication circuit. The power supply circuit uses 3.7V DC power supply as the main power supply. Through the voltage stabilizing circuit, the 3.7V power supply voltage is converted into 3.5V power supply voltage and supplies power for the development board. Based on the main control chip of the system is S3C2440 chip [7], the clock circuit, acquisition source, power circuit and structure are set around the chip to make S3C2440 chip the smallest control system in the system. The corresponding data communication structure is set at the bottom of the chip, and the system driver and receiver are used as the terminal of the system data receiving and transmitting. In this system, two groups of general timing devices are set, including 10 timers. The system can be flexibly timed to ensure the stability of data acquisition time.

Next, the data storage device is optimized. The memory chip adopts small volume and large space storage medium (which can be expanded to 1TB), which almost occupies no space and reduces the waste of space volume [8].

Using IPv6 flow label, the background storage network of e-commerce platform is set as the architecture of IPv6 network application platform, which is a kind of strengthened network environment. It can improve the time delay of network infrastructure and hardware energy efficiency, which is the basis of background hardware optimization design. The environment setting of IPv6 network application platform is shown in Figure 3.

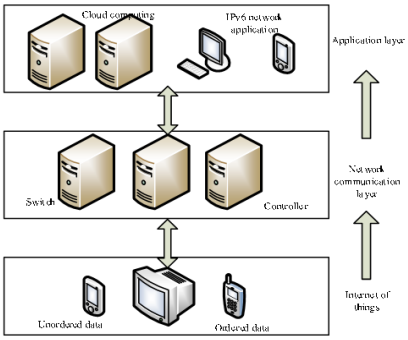

Fig 3. IPv6 network application platform

In Figure 3, using the ipv6128 bit address space, the other parts of the background of the e-commerce platform are controlled through the control circuit, and the data collected by the information collector is transmitted to the central processor of the computer server through the communication device. After data preprocessing, the data is transmitted orderly through the input unit, and the processed data is analyzed through the output sheet The output information or instruction is transmitted to the corresponding peripheral terminal through the communication device to execute the instruction task [9], so that it can meet the address requirements of the data-driven quality, so as to improve the use effect of the data-driven quality control method.

The hardware selection results are integrated into the traditional system hardware, and the hardware installation is completed according to the system hardware framework designed in this paper. The installed system hardware framework is used as the development environment of the specific mining process.

2.2 Attribute analysis of precision marketing data

In e-commerce platform, there are a lot of valuable information about customer trajectory. Generally, the text similarity between precision marketing data and customer behavior data is high, and precision marketing data features with location sequence attributes can be used to represent spatiotemporal embedded precision marketing data. Compared with the algorithm of small data, big data with more effective and simple algorithm makes people no longer expect accuracy, but hybridity. How to mine valuable relevant information from the hybridity of big data has become a problem that the e-commerce platform must think about. The time and location transfer in the corresponding customer trajectory is very important for understanding the customer's shopping behavior, considering different factors The time factor of degree, location transfer time, location sequence and other influencing factors, namely customer behavior data attribute, can analyze the precision marketing data attribute of e-commerce platform in real time. Therefore, e-commerce platform precision marketing customer attribute data not only reproduces the whole picture of consumers, reflects the needs of consumers, but also is the basis of e-commerce platform mining consumer demand and value, consumer segmentation, the implementation of precision marketing and other activities.

In this design, the customer trajectory model is established based on the generative confrontation network (GAN) to analyze the precision marketing data attributes of e-commerce platform.

As a new generation of network mode, GaN has become a new hot spot of deep learning and artificial technology intelligence, and has shown great application and development prospects in image and image processing, voice processing, information security and other fields. The network can train the model by observing the common pattern of data samples and label probability. The trained model can generate new data according to the distribution of samples for supervised learning and unsupervised learning. Details

are all in invisible deep learning, and generative model plays an important role in deep learning. This model can capture two groups of data with high correlation. It does not need to obtain the target information on the class tag, learn the relevant characteristics of the actual data, present the distribution characteristics of the sample data, generate new data similar to the training sample, and the new data contains a large number of genera Sexual information.

According to the customer attributes (as shown in Table 1) and the research results of the storage performance of the optimized master chip, in this design, the customer trajectory model is established based on the generative confrontation network (GAN) to analyze the data attributes of e-commerce platform precision marketing.

As a new generation of network mode, GaN has become a new hot spot of deep learning and artificial technology intelligence, and has shown great application and development prospects in image and image processing, voice processing, information security and other fields. The network can train the model by observing the common pattern of data samples and label probability. The trained model can generate new data according to the distribution of samples for supervised learning and unsupervised learning. Details are all in invisible deep learning, and generative model plays an important role in deep learning. This model can capture two groups of data with high correlation. It does not need to obtain the target information on the class tag, learn the relevant characteristics of the actual data, present the distribution characteristics of the sample data, generate new data similar to the training sample, and the new data contains a large number of attribute information Interest.

The construction process of customer trajectory model based on GaN is as follows:

Combining with the relevant prior knowledge, the customer stability analysis needs to take the customer's recent use characteristics and consumption characteristics as the prediction basis, and the required information mainly includes customer data table, customer derived information table, etc. [10], as shown in Figure 4:

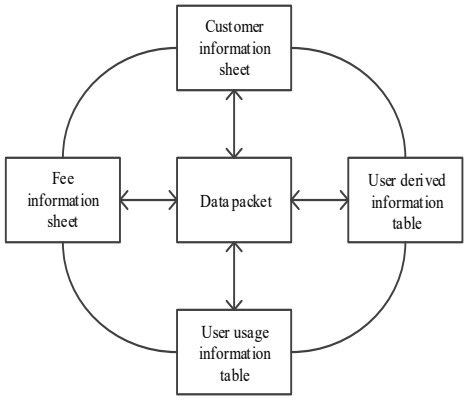

Fig 4. Information needed for stability calculation

In Fig. 4, the generated row vector $k+\varepsilon$ of the stored data packet constitutes the $(k+\varepsilon)\times k$ -order matrix $G_{(k+\varepsilon)\times k}$, which is described as the generation matrix of $k+\varepsilon$ stored data packet, namely $G_{(k+\varepsilon)\times k}=[g_1,g_2,\cdots,g_{k+\varepsilon}]^T$. By virtue of the generation matrix, the $k+\varepsilon$ kinds of storage packets can also be described as:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_{k+\varepsilon} \end{bmatrix} = G_{(k+\varepsilon)\times k} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix} \quad (1)$$

Through the linear algebra theory, it can be seen that if the generating matrix $G_{(k+\varepsilon)\times k}$ has a reversible $k \times k$ -order submatrix, that is, G columns of full rank, then the formula group in formula (1) only has a unique solution. After solving the equation group, k kinds of unknown source packets $X_1, \cdots, X_k$ can be calculated. Therefore, the probability that $\mathrm{Sin}\, k$ nodes can obtain k kinds of unknown source data packets through random $k+\varepsilon$ kinds of stored data packet operations is equivalent to the probability that $G_{(k+\varepsilon)\times k}$ column of the generation matrix of these $k+\varepsilon$ kinds of stored data packets is full rank, that is, $Rank\left(G_{(k+\varepsilon)\times k}\right) = k$.

As the data mining process of e-commerce platform precision marketing is a dynamic control process, the controller form of the hardware part is set as tight format dynamic linearization controller [11,12]. This controller is a relatively smooth nonlinear function. At very different time h, if the data control partial derivative $\partial c(\bullet) / \partial e(k+1)$, the constraint conditions of the controller can be shown as follows:

$$\left|\Delta r(k)\right| \le \alpha_1 \left|\Delta r(k+1)\right| + \beta_1, \forall k \in n \quad (2)$$

$\Delta r(k) = r(k) - r(k+1)$; $\Delta r(k+1) = r(k+1) - r(k)$.

Select SQL Server database, it is a database platform that can be used for large-scale online transaction processing, data warehouse, e-commerce and other applications. It can not only share servers in e-commerce platform, query all kinds of data, but also reduce the cost of management, production and other aspects for e-commerce platform. In order to improve the control ability of e-commerce platform precision marketing data-driven boundary, effective bounded constraints are set to ensure that the controller value remains unchanged at this time, $\left|\Delta r(k+1)\right| \ne 0$, then the customer trajectory model [13] can be obtained, and the formula is as follows.

$$\Delta r(k) = \varepsilon(k) \Delta r(k+1) \quad (3)$$

Table 1. Customer attributes of e-commerce platform precision marketing

Data type	Attribute	Source

Demographic characteristics	Gender age occupation marital status family address	Investigation and research
Behavior characteristic	Demand induction: internal cause (regional season); external cause (color, style, etc.)	Investigation and research
	Information collection: browsing time, connection source, click collection times	Opinions of research experts
	Comparative choice: product category color price quality brand	Opinions of research experts
	Purchase decision: purchase time and payment method	Expert Opinion
	Post purchase evaluation: Product Service	Investigation and research
Psychological feature	Preference: product price	Investigation and research
	Attitude: satisfaction and loyalty	Investigation and research
	Value: perceived value	Investigation and research

3 E-commerce platform precision marketing data mining method based on association rules

3.1 E-commerce platform precision marketing information fusion and distribution similarity analysis

Data mining means the decision support process of finding patterns in the collection of some facts or observation data. In this process, it is necessary to normalize the multi-source e-commerce platform precision marketing information, that is μ_{ik}^{m}, to create k-classification target function between data basins by using information fusion means, so as to cluster e-commerce platform precision marketing information:

$$J_m\left(U,V\right)=\sum_{k=1}^{n}\sum_{i=1}^{c}\mu_{ik}^{m}\left(d_{ik}\right)^2 \tag{4}$$

In the formula, $\left(d_{ik}\right)^2$ is the i-fold function of the clustering coefficient of the marketing information. The historical data m is analyzed quantitatively and optimized by K-value. The quantitative recursive feature extraction results of the precision marketing information of the e-commerce platform are obtained

$$x_n=a_0+\sum_{i=1}^{M_{AR}}a_i x_{n-i}+\sum_{j=0}^{M_{MA}}b_j\eta_{n-j} \tag{5}$$

a_0 is the precision marketing information of the original e-commerce platform, p_0 is the sampling amplitude, x_{n-i} is the scalar time series, and b_j is the oscillation decay value during scheduling.

In order to improve the precision marketing information quantitative mining ability of e-commerce platform, the association rule method is introduced. The biggest limitation of association rules is that

the items in the item set can't be used as the aftereffect, and the items in the classification identification set can't be used as the antecedent. Therefore, it's necessary to transform the antecedent and aftereffect problems of association rules into solving the K-means clustering objective function, making it a simple least square estimation problem, that is, calculating the precision marketing information map of e-commerce platform table β is the consistent scheduling value of resource convergence vector, let $\|Y - X\beta\|$ be the lowest, and let $\|\cdot\|$ be the F-norm in Euclidean norm:

$$P_{loss} = 1 - \frac{1 - p_0}{\rho} = \frac{p_0 + \rho - 1}{\rho} = \sum_{n=1}^{N} p_{K,n} \tag{6}$$

Setting p is the interference feature vector of precision marketing information of e-commerce platform [14], which transforms the evaluation of precision marketing information mining ability of e-commerce platform into the least square solution

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)} + n(t) \tag{7}$$

Where, $x(t)$ is the real part of data information compactness and $y(t)$ is the imaginary part of data information compactness.

Using the surrogate data method, amplitude randomization processing is adopted for the precision marketing information of e-commerce platform to obtain $x'(k)$. The interference functional is implemented for the mining ability evaluation experience distribution information in Category k to obtain the subclass set of category k. The precision marketing information distribution of e-commerce platform is recorded as:

$$U_{util} = \gamma \bar{X} \tag{8}$$

The principal component feature quantity of mining ability evaluation is constructed, and the similarity of mining resource distribution is calculated:

$$Sim_1(d_i, d_{1j}) = \frac{\sum_{k=1}^{M} W_{ik} \times W_{1jk}}{\sqrt{\sum_{k=1}^{M} W_{ik}^2} \bullet \sqrt{\sum_{k=1}^{M} W_{1jk}^2}} \tag{9}$$

Among them, d_i is the prior distribution feature vector of e-commerce platform precision marketing data mining ability evaluation, d_{1j} is the K-means clustering center vector of the first layer of information.

The linear correlation feature fusion algorithm is introduced to complete the clustering and merging of index parameters of mining effect evaluation [15-16], and the output e-commerce platform precision marketing data fusion analytic formula is as follows:

$$P(w|x)=P(x|w)/P(x) \tag{10}$$

If the quantitative recursive feature $(N(i) \bmod L)<m$, the probability density property of the mineable data distribution is $p(i)=\left\lfloor \frac{N(i)}{L} \right\rfloor$,

It is necessary to constrain the mining work in advance, and the content of the constraint must ensure that the mining work has a small amount of calculation and a strong effect [17-18].

Using $confidence(Q \Rightarrow W)$ to represent the probability that feature set Q contains feature set W, and $confidence(W \Rightarrow Q)$ is contrary to the above meaning, intrusion detection mining is carried out in three aspects of location relevance, direction relevance and length relevance of e-commerce platform precision marketing data.

(1) Mining location relevance

The precision marketing data mining of e-commerce platform can be obtained by calculating the centroid of transmission channel. The centroid of Q and W channels of precision marketing data set is represented by g_1 and g_2, and the distance between the two centroids is $|g_1,g_2|$. Then the mining result of location association degree is the same as that of segment cluster $(\bar{Q},\bar{W})$ of two centroids, and the location association degree is mined [19-20].

(2) Direction association mining

Direction association mining represents the angle (s_1,s_2) between the transmission direction of precision marketing data set Q and W, and its cosine value is described as follows:

$$\cos(s_1,s_2)=\frac{s_1,s_2}{|s_1||s_2|} \tag{11}$$

According to formula (17), the larger the transmission direction angle of intrusion monitoring precision marketing data set Q and W, the smaller the value of $\cos(s_1,s_2)$. If the value of (s_1,s_2) is greater than 180 degrees, the value of $\cos(s_1,s_2)$ is negative. In order to reduce the impact of location association degree mining results on direction association degree mining results, distributed precision marketing data mining method based on NoSQL uses sine value description method of $[1-\cos(s_1,s_2)]$ instead of traditional description method of $[1-\cos^2(s_1,s_2)]$ to mine direction association degree accurately.

$$sim(dist)=avg(|s_1||s_2|)[1-\cos(s_1,s_2)] \tag{12}$$

(3) Mining length relevance

Length association mining is the calculation of different thinking for location association mining, and it is also the addition budget for its mining results. Its core is the process of calculating the length of transmission channel of two data sets. Therefore, the mining results can be expressed as follows:

$$sim(length) = (\bar{Q}, \bar{W}) \frac{|Q - W|}{\max(Q - W)} \quad (13)$$

(4) Mining clustering

For the intrusion monitoring data after dimension reduction, the constraint conditions given by the combination of formula (11), (12) and (13) are used to cluster, and the final mining results are obtained. Suppose that F represents the mining frequency, then F_s is the mining clustering, that is, the final mining result：

$$F_s = \frac{F_j}{F_{\max}} \lg \frac{A}{A_s} \quad (14)$$

F_j is the probability that the location, direction and length attributes of precision marketing data appear together, $F_{\max}$ is the maximum value of the correlation degree of the above attributes, and A and A_s are the total number of samples before and after mining respectively.

After obtaining the mining data, residual test and posterior test are needed

(1) Residual test

The residual sequence $e^{(0)}(k)$, relative error sequence Δ_k and average relative error $\overline{\Delta}$ of $x^{(0)}(k)$ and $\hat{x}^{(0)}(k)$ are obtained respectively

$$e^{(0)}(k) = x^{(0)}(k) \_ \hat{x}^{(0)}(k) \quad (15)$$

$$\Delta_k = \left| \frac{e(k)}{x^{(0)}(k)} \right| \times 100\% \quad (16)$$

$$\overline{\Delta} = \frac{1}{n} \sum_{k=1}^{n} \Delta_k \quad (17)$$

(2) The posterior error of the test.

The average value $\overline{x}$ of the original data and the average value $\overline{e}$ of the residual error are obtained:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x^{(0)}(k) \quad (18)$$

$$\bar{e} = \frac{1}{n-1}\sum_{k=2}^{n} e^{(0)}(k) \tag{19}$$

The variance ${s_1}^2$ of the original data, the variance ${s_2}^2$ of the residual error, the ratio C of the mean square error and the small error probability P are obtained:

$${s_1}^2 = \frac{1}{n}\sum_{k=1}^{n} [\, x^{(0)}(k) - \bar{x}\,]^2 \tag{20}$$

$${s_2}^2 = \frac{1}{n-1}\sum_{k=2}^{n} [\, e^{(0)}(k) - \bar{e}\,]^2 \tag{21}$$

$$C = s_2 / s_1, \quad P = p\left\{\left|e^{(0)}(k) - \bar{e}\right| < 0.6745 s_1\right\} \tag{22}$$

Order $\xi_k = \left|e^{(0)}(k) - \bar{e}\right|$, $s_0 = 0.6745 s_1$, and get $P = p\{\xi_k < s_0\}$.

At this time, if the data mining meets the requirements of residual test and posterior test at the same time, it means that it has completed the e-commerce platform precision marketing data mining.

4 Experimental analysis

In order to verify the effectiveness of the proposed data mining method for precision marketing of e-commerce platform based on association rules, the simulation test is carried out in windows10 with 8g running memory, core i5-7400 CPU and python 3.6 programming code. The specific test equipment parameters are shown in Table 2.

Table 2. Parameter setting of test equipment

Parameter name	Parameter content	Parameter name	Parameter content
equipment	150mm*100mm*120mm	maximum	64TB
size	2.50kg	Capacity	90W AC transformer
Equipment	RAID0,1,5,6,10	Adapter	240V
Weight	512MB	Input voltage	BT/PT
RAID	1.5GHz	Download	Wireless network adapter
Pattern	SATA II	Network settings	Support heat dissipation

Taking KTT data set as the training sample set of marketing data mining, the data scale is 50 000 KB, the data distribution area is 200-800, and the statistical time cycle is 12 d. According to the above experimental parameter setting, the method of reference [4] is compared with that of reference [5], and

the relative error and average error are set as test indexes. The higher the value of test indexes is, the lower the accuracy of data mining results is; on the contrary, the higher the accuracy is

Table 3. Comparison of relative errors of different methods

Test type	Relative error / (%)		
	Method of this paper	Methods of literature [4]	Methods of literature [5]
A1	0.257	0.274	0.307
A2	0.158	0.185	0.214
A3	0.135	0.164	0.185
A4	0.204	0.225	0.255
A5	0.145	0.174	0.212
A6	0.164	0.191	0.221
A7	0.217	0.232	0.262
A8	0.235	0.285	0.324
A9	0.245	0.301	0.354
A10	0.107	0.131	0.167

Table 4. Average error comparison of different methods

Test type	Average error / (%)		
	Method of this paper	Methods of literature [4]	Methods of literature [5]
A1	0.267	0.295	0.317
A2	0.174	0.201	0.238
A3	0.155	0.182	0.225
A4	0.204	0.231	0.275
A5	0.157	0.197	0.241
A6	0.163	0.182	0.213
A7	0.216	0.254	0.295
A8	0.235	0.278	0.314
A9	0.252	0.285	0.332
A10	0.114	0.142	0.161

Comprehensive analysis of the experimental data in Table 3 and table 4 shows that the relative error and average error values of this method are significantly lower than those of the other two methods, which indicates that this method has high accuracy. This is because this method analyzes the multi-source of precision marketing data of e-commerce platform, and based on the optimization results of background hardware of e-commerce platform, the multi-source data is collected Set and attribute analysis, its mining process is more targeted, mining accuracy is guaranteed.

The process of data mining is a process of repeated cycles. If each step does not achieve the expected goal, it needs to go back to the previous steps, readjust and execute. Not every work of data mining needs every step listed here. For example, when there are no multiple data sources in a work, step (2) data integration can be omitted. However, in this process, the omitted data source may lose part of the data attribute metadata, because, in order to more comprehensively verify the effectiveness of the method in this paper, data mining is set up The method of literature [4] is compared with that of literature [5]. The specific experimental results are shown in Figure 5

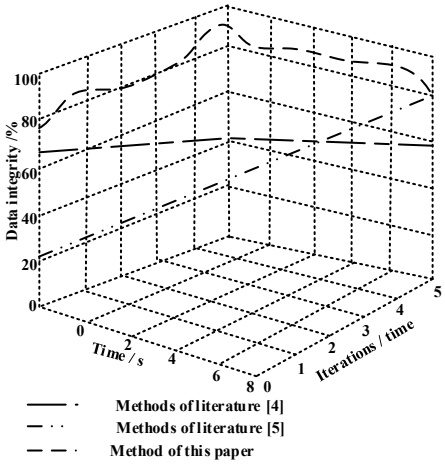

Fig 5. Comparison results of data mining integrity of different methods

Analysis of the experimental data in Figure 4 shows that the data mining integrity of different methods will be affected by the number of test samples. The data mining integrity of this method is the highest among the three methods; the data mining integrity of the method in literature [4] is the second; the data mining integrity of the method in literature [5] is the lowest. This is because this method uses information fusion means and distribution similarity analysis The analysis method optimizes the association rule algorithm and efficiently realizes the e-commerce platform precision marketing data mining.

5 Conclusion

According to the precision marketing mode of e-commerce platform, this paper analyzes the multi-source of precision marketing data of e-commerce platform. By optimizing the background hardware of e-commerce platform, multi-source data collection and attribute analysis are carried out. By using information fusion means and distribution similarity analysis, association rule method is optimized, precision marketing data mining of e-commerce platform is carried out, and residual error test and posterior error test are completed. The relative error and average error are significantly lower, and the integrity of data mining is also significantly higher, which ensures the effect of precision marketing data mining on e-commerce platforms.

Reference

1.He F, Xu Q, Wang N, Zhu C. A method to obtain the wind field characteristics of super-large aperture radio telescope site based on single-point wind tower and numerical simulation[J]. Research in Astronomy and Astrophysics, 2020, 20(12):199 (9pp).

2.Sonashree S , Manickam J . Development of l-Amino-Acid-Based Hydroxyl Functionalized Biodegradable Amphiphilic Polyesters and Their Drug Delivery Capabilities to Cancer Cells[J]. Biomacromolecules, 2020,21(1):171-187.

3.Yu, W. Discovering Frequent Movement Paths From Taxi Trajectory Data Using Spatially Embedded Networks and Association Rules[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(3):855-866.

4.Xiong X, Yuan F, Huang M, Cao M, Xiong X. Comparative Evaluation of Web Page and Label Presentation for Imported Seafood Products Sold on Chinese e-Commerce Platform and Molecular Identification Using DNA Barcoding[J]. Journal of food protection, 2020, 83(2):256-265.
5.Qin D . Data Monitoring Platform for Submarine Cable Tunnel Operation Based on Data Mining[J]. Journal of Coastal Research, 2019, 93(sp1):960.
6.Sun X , Shi Z , Zhu J . Multi-objective Design Optimization of an IPMSM for EVs Based on Fuzzy Method and Sequential Taguchi Method[J]. IEEE Transactions on Industrial Electronics, 2020, PP(99):1-1.
7.Emamian R , Ebrahimi M , Karimi-Maleh H . Electrochemical Platform Based on Synergic Effect of Fe3O4/SWCNTs and 1-ethyl-3-methyl Imidazolium Chloride as Sensor for Determination of Xanthine and Theophylline in Food Samples[J]. Journal of The Electrochemical Society, 2018, 165(14):B762-B766.
8.Kwak J , Zhang Y , Yu J . Legitimacy building and e-commerce platform development in China: The experience of Alibaba[J]. Technological forecasting and social change, 2019, 139(FEB.):115-124.
9.Wan X , Chen J . The relationship between platform choice and supplier's efficiency-evidence from China's online to offline (020)e-commerce platforms[J]. Electronic Markets, 2019, 29(2):153-166.
10.Yue Y, Li B. Effects of E-Commerce Platforms on Firm Export——Evidence from China's Industrial Enterprises[J]. China Economist, 2019, 14(05):114-127.
11.Wang O , Somogyi S , Charlebois S . Food choice in the e-commerce era: A comparison between Business-To-Consumer (B2C), Online-To-Offline (O2O) and New Retail[J]. British Food Journal, 2020, 122(4):1215-1237.
12.Zhang H , Dong J . Prediction of Repeat Customers on E-Commerce Platform Based on Blockchain[J]. Wireless Communications and Mobile Computing, 2020, 2020(12):1-15.
13.Lin Y., Li Y., Yin X., and Dou Z., “Multisensor Fault Diagnosis Modeling Based on the Evidence Theory,” IEEE Transactions on Reliability, 2018, 67(2): 513-521.
14.Liu S., Wang S., Liu X., et al. Fuzzy Detection aided Real-time and Robust Visual Tracking under Complex Environments. IEEE Transactions on Fuzzy Systems, 2021, 29(1), 90-102
15.Zhang J , Wu T , Fan Z . Research on Precision Marketing Model of Tourism Industry Based on User's Mobile Behavior Trajectory[J]. Mobile Information Systems, 2019, 2019(4):1-14.
16.S. Liu. Introduction of Key Problems in Long-Distance Learning and Training, Mobile Networks and Applications, Vol. 24, No. 1, pp. 1-4, 2019
17.Cui F , Hu H , Xie Y . An intelligent optimization method of E-commerce product marketing[J]. Neural Computing and Applications, 2021, 5(1):1-14.
18.Shazad B, Khan H, Zahoor-ur-Rehman, Farooev M, Mahmood A, Mehmood I, Rho S, Nam Y. Finding Temporal Influential Users in Social Media Using Association Rule Learning[J]. Intelligent automation and soft computing, 2020, 26(1):87-98.
19.Shuai L., Chunli G., Fadi A., et al. Reliability of response region: A novel mechanism in visual tracking by edge computing for IIoT environments, Mechanical Systems and Signal Processing, 2020, 138, 106537
20.Lin Y., Wang C., Wang J. X., et al, A Novel Dynamic Spectrum Access Framework Based on Reinforcement Learning for Cognitive Radio Sensor Networks, Sensors, 2016, 16(10): 1675.