

Retraction

Retracted: Construction of User Behavior Web Data Mining Model for Internet Marketing

Wireless Communications and Mobile Computing

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. Wei, "Construction of User Behavior Web Data Mining Model for Internet Marketing," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2813718, 10 pages, 2022.

Research Article

Construction of User Behavior Web Data Mining Model for Internet Marketing

Ming Wei 

Guanghua School of Management, Beijing 100000, China

Correspondence should be addressed to Ming Wei; 2101220932@stu.pku.edu.cn

Received 28 July 2022; Revised 24 August 2022; Accepted 8 September 2022; Published 22 September 2022

Academic Editor: Jun Ye

Copyright © 2022 Ming Wei. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of the Internet and the rapid change of the era of big data, data mining has played a great role in exploring and discovering potential valuable information from the vast sea of data and has become one of the most popular research and practice fields. At the same time, the emerging marketing model of network marketing is sweeping the market, and network marketing is increasingly replacing the traditional marketing model with its unique advantages, such as wide coverage, fast dissemination, and being more flexible and targeted. Preserving, researching, and analyzing the behavior of network marketing users are the difficulties of current network marketing. It is the focus of this paper to explore how to use Web data mining technology to model the user behavior of Internet marketing. For the above problems, this paper used the weight calculation algorithm of users' daily behavior feature items and the sequential pattern discovery algorithm based on multiple factor constraints to scientifically construct the online marketing user behavior model. The experimental results have shown that the sequential pattern discovery algorithm based on multiple factor constraints could save 30%-40% of the running cost when building a user behavior model. At the same time, the user behavior model constructed by this algorithm has improved the simulation and prediction accuracy of user behavior by 37%, which showed that the use of Web data mining to build network marketing user behavior model could provide an objective basis for the strategic development of network marketing.

1. Introduction

In recent years, the data in the web has been increasing at a rate of 1 million web pages per day. The widespread popularity of computer networks has made the Internet develop into a huge distributed information space with hidden potential value knowledge and immeasurable data. The reason is that people's lives are inseparable from the Internet. Various shopping software, entertainment software, take-away software, and travel software are flooded in people's daily lives, and everyone's clothing, food, housing, and transportation have left traces on the Internet. Today, with the rapid development of the Internet, the mass production and collection of various forms of information have caused an explosion of information. For ordinary network users, how to quickly and accurately obtain valuable network information, find out what is useful to them, and find out more valuable knowledge has become an important topic

currently faced. On the other hand, as the Internet is becoming more and more popular today, because of its low cost, convenience and speed, and not being limited by time and space, e-commerce has developed extremely rapidly since its inception and has become popular all over the world.

At the same time, the advantages of network marketing, a marketing model that combines the Internet with modern digital electronic methods, are gradually emerging. Based on the above discussion, it is necessary to build a user behavior model for network marketing with the help of web data mining today when network marketing is popular. It obtains useful information from a large amount of data, finds out customer purchasing preferences, and provides personalized services to increase customer retention time. It taps potential users to expand market share, which is of great strategic significance to help e-commerce take off.

This paper uses Web data mining technology to study the construction of online marketing user behavior model.

It uses the weight calculation algorithm of users' daily behavior feature items and the sequential pattern discovery algorithm based on multiple factor constraints, which is aimed at providing a more accurate and rich database for user behavior research in the face of Internet marketing, so as to better improve the construction of behavior models. Constructing a scientific and accurate user behavior model can better promote the implementation of network marketing strategy and the selection of marketing methods and objects and then have a more profound impact on the development of e-commerce. In this paper, the experiment and result analysis of sequential pattern discovery algorithm based on multiple factor constraints show that the user behavior model constructed by this algorithm can save about 40% of the running cost. The innovation of this paper are as follows: (1) a sequential pattern discovery algorithm constrained by multiple factors is used for the construction of the user behavior model, which makes the research of this model more scientific and accurate. (2) Based on Web data mining technology and the construction of user behavior model, it provides more scientific algorithm support for network marketing and even e-commerce.

2. Literature Review

On the booming and ubiquitous Internet, a large amount of online user behavior data is generated every day, so the analysis of network user behavior has developed rapidly from scratch. Sarker et al. proved that user behavior has certain regularity through a large number of experiments, but did not find a suitable method to mine the features [1]. After analyzing various algorithms and software, Qian et al. found that the behavior of Internet users has its unique regular characteristics, and this behavioral characteristics play an important role in predicting the future behavior of most Internet users [2]. Many scholars in China and other countries have done a lot of work on how to improve the efficiency of data mining. However, most of the current methods have certain defects, such as high algorithm complexity and being easy to fall into the local optimal solution or large amount of calculation. Shin believed that the behavior of network users is a process with a certain subjective initiative. Due to the concealment of network users themselves, it is difficult for network users to obtain their real identities [3]. Singh et al. proposed a new PPSA scheme to protect users' sensitive data from being acquired or leaked by external analysts or aggregation service providers [4]. Esdar et al. found in their research that user behavior has a significant guiding effect on e-commerce websites, which can help e-commerce websites improve web page interaction design and enhance user experience [5]. Although the current academic analysis of user behavior is varied, most of the existing analysis only emphasizes strong privacy protection. User behavior analysis for online marketing only protects its privacy is very limited, which is difficult to meet the need of behavior analysis.

With the rapid development of science and technology and the continuous improvement of people's living standards, network terms such as the Internet and web have also

begun to appear in thousands of households. In order to have a deeper understanding of consumers' purchasing preferences and consumption habits in the Internet age, web-based data mining is proposed. Gang was the first to explain web data mining. Web data mining is to start from the large-scale raw data stored in the web background information database, mining the information that is helpful to marketers and can be used directly, and finally display it to marketers in a concise and intuitive way [6]. Agirre believed that data mining also known as database knowledge discovery is mainly composed of seven parts: data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge representation [7]. The main purpose of Web data mining is to convert these data into useful information to help companies make better marketing strategies. Most of the current research focused on text mining, but there was no report on image data mining. When Lixia et al. researched algorithms in data mining, they found that the decision tree algorithm is currently the most widely used and adaptable data mining algorithm. The decision tree algorithm belongs to the inductive reasoning algorithm, which can use the approximate discrete value function to process the noisy data and then analyze to obtain the expression [8]. On the other hand, Kazanidis et al. pointed out that there are many methods of data mining, including classification, regression analysis, cluster analysis, feature analysis, and web data mining. Different methods are applied in different occasions, and they have different characteristics [9]. Jeffery et al. suggested predicting user access requests. Based on the user's previous access situation, the content of the page that the user will visit next was dynamically generated [10]. Due to its low accuracy, traditional data mining requires users to make multiple operation attempts. However, the task of Web data mining would rely on the self-operation and operation of the Internet, which can improve the intelligence of interactive learning and coordination. Therefore, its addition not only improved the efficiency of data mining but also saved valuable time for users.

3. Construction Method of User Behavior Web Data Mining Model for Internet Marketing

3.1. Internet Marketing User Behavior

3.1.1. e-commerce. e-commerce is a new economic and commercial activity in the information environment, and it is also the main carrier and platform for the development of the information economy. Since its birth in 1997, e-commerce has gone through more than 20 years of exploration and development. China's Internet construction and e-commerce development have played a pivotal role in the world [11]. Since 2013, China's e-commerce transaction volume has been rising steadily, from 1 billion in 2013 to 7 billion in 2020 with a rapid development, as shown in Figure 1.

As shown in Figure 2, the average annual growth rate of China's e-commerce transaction volume was also generally on the rise, reaching its peak in 2018 and up to 57.8%. The e-commerce market has broad development space and huge

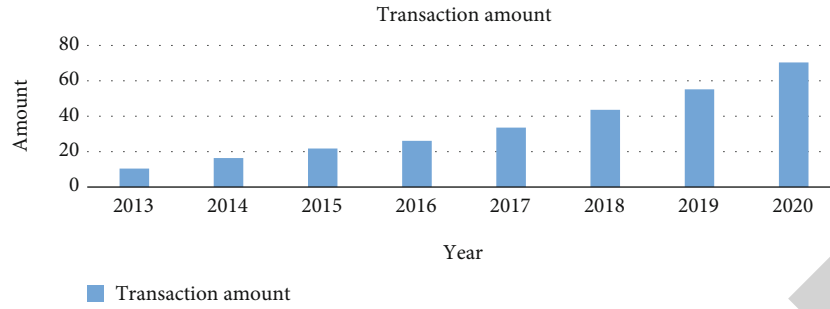


FIGURE 1: China's e-commerce transaction scale from 2013 to 2020 (trillion yuan).

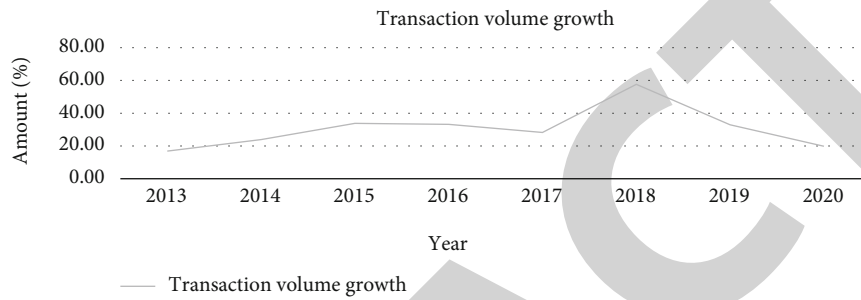


FIGURE 2: Growth rate of e-commerce transactions in China from 2013 to 2020.

TABLE 1: Scale and growth rate of e-commerce employment in China from 2015 to 2017.

Year	Employment scale (10,000 people)	Year-on-year growth rate
2015	2698	16.54%
2016	3255.74	21.78%
2017	3790.43	15.74%

potential, and is becoming a new engine for the rapid and sustainable development of China's national economy.

At present, e-commerce has been shining brightly in economic life, and it is playing an increasingly important role in manufacturing, commerce and trade and financial industry. e-commerce urges people to make rational use of "fragmented time," cultivate people's new consumption habits, and promote the rapid growth of consumer demand. For example, thanks to the development of e-commerce, people can easily realize the freedom of shopping by simply moving their fingers on their mobile phones without having to squeeze out a large chunk of time for shopping. By the end of 2016, the number of online shopping users in China reached 467 million, and the transaction size of the online retail market was 5,155.6 billion yuan, a year-on-year increase of 26.2%. The development of e-commerce has created a large number of jobs, as shown in Table 1. By the end of 2016, China's e-commerce services directly or indirectly created 37.6043 million jobs, and e-commerce has also become the preferred field for mass entrepreneurship and innovation.

3.1.2. *Internet Marketing.* The development of e-commerce gave birth to network marketing. This is due to the develop-

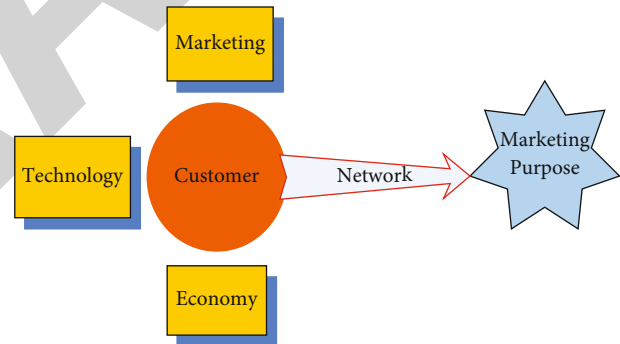


FIGURE 3: Basic structure of network marketing.

ment of e-commerce, which enables people to shop easily. The traditional marketing combined with traditional commerce has lost its rooted soil, and new marketing models will naturally appear to replace it. Network marketing was born in the 1990s. It is a new type of marketing method based on the Internet and effectively using the network medium to spread a large amount of data information [12]. It can effectively use the social media and digital information that has been popularized at present, so that marketing goals can be achieved faster and better. Network marketing is the result of three combined forces of marketing, technology, and economy. At present, the most common understanding of network marketing is a series of activities that realize the purpose of enterprise marketing by focusing on the customer and using the network as the medium. Its specific principle is that enterprises carry out network marketing around customers and use the network as a medium to spread marketing to promote the realization of network marketing purposes, as shown in Figure 3.

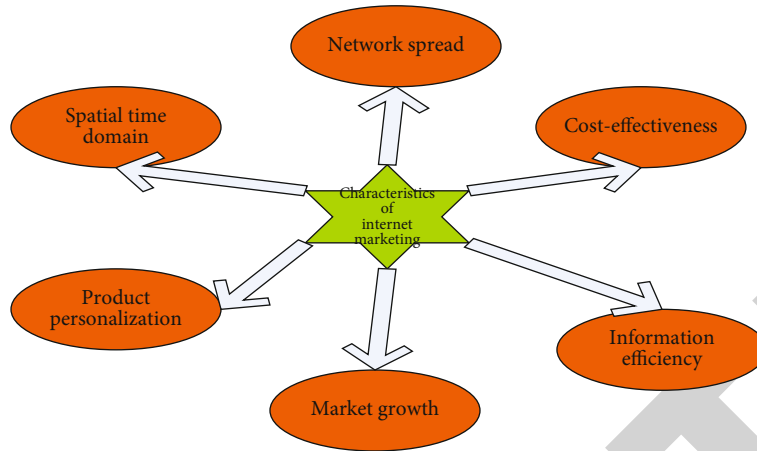


FIGURE 4: Characteristics of Internet marketing.

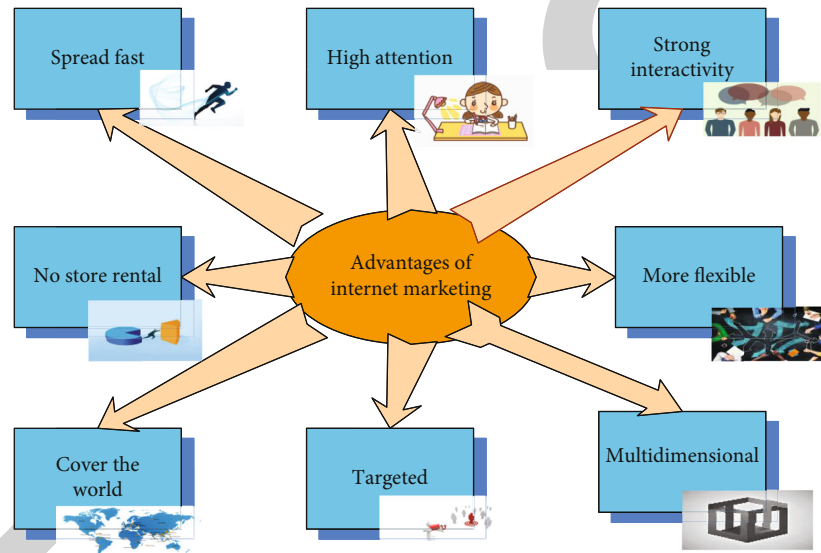


FIGURE 5: Advantages of Internet marketing.

It should be pointed out that “Internet marketing” does not equal “e-commerce.” Internet marketing is a means to achieve the purpose of e-commerce, and e-commerce is the ultimate purpose of Internet marketing. Network marketing occupies an important position in the whole process of e-commerce, and e-commerce is an advanced stage pursued by the development of network marketing [13]. Compared with the traditional marketing model, network marketing has the following characteristics, as shown in Figure 4.

Network marketing does not necessarily mean that the purpose of e-commerce can be achieved. However, in order to achieve e-commerce, Internet marketing must be a prerequisite. For the development of enterprises, network marketing will play an irreplaceable and important role in traditional marketing. In the long run, network marketing has the following eight advantages, as shown in Figure 5.

3.1.3. Types of Internet Marketing Users. For network marketing, the most important foundation is the user [14–16].

User behavior is the various behaviors that users do during the process of entering the e-commerce platform until exiting, usually including browsing clicks, decision-making, and evaluation. Capturing user behavior and adjusting the network marketing strategy is the key to the success of network marketing. As a very important part of network marketing, customer analysis is based on customer classification to carry out marketing activities, which is more targeted and has better marketing effects.

Through research and analysis, this paper divided the types of users of network marketing into three categories.

- (1) For homeless users, the number of types of homeless users and all the values are relatively low. From this trend, it can be seen that homeless users are not very interested in online marketing activities. They are only in a wait-and-see state, so the marketing focus of this user is to stimulate the user’s interest in shopping

TABLE 2: Differences between different user behaviors.

	Nomadic user	Follow other users	Active user
Clicks	0 ~ 20	21~200	201~500
Browsing dwell time(s)	≤20	22~50	51~60
The number of purchases	≤5	6~100	≥200

- (2) Follow other users. As the name implies, it is to pay attention to the behavior of other users. This type of user is generally less proactive, and most of them pay more attention to the consumption of others. Therefore, in the face of such users, the focus of online marketing is to expand the marketing scale for the consumption points that they are concerned about to consolidate such users
- (3) Active users are more active on the e-commerce platform, which is commonly known as the “hand-picking party.” They are keen on consumption and cannot resist the means of online marketing. In the face of this type of users, new online marketing methods should be constantly created to attract their attention, while consolidating and developing new active users

Different types of users have different user behaviors, and the types of users can be analyzed from the user’s click behavior, browsing time, purchase records, and other behaviors. For example, active users are keen to see now and buy now, so their traffic data is high. However, on the other hand, because they skip the shopping cart step, they have relatively low data on adding to the cart, as shown in Table 2.

Users’ browsing behaviors can be classified into four categories, such as click, add to cart, favorite, and purchase [17]. In fact, by calculating the number of times in each period and making a line graph, the user’s browsing behavior can be further observed, as shown in Figure 6. Among them, Figure 6(a) shows that the number of times the user clicks on the item changes with time. Figure 6(b) shows the change in the number of users adding to the shopping cart over time.

According to the above chart, Table 3 gives the clustering results divided into 8 segments, which comes from the ordered clustering analysis of the Q language. Q language is a special language for cluster analysis, which is characterized by the ability to analyze time-sorted data and more intuitively show the hidden laws behind the data.

3.2. Construction of User Behavior Feature Model Based on Internet Marketing

3.2.1. Extraction of User’s Daily Behavior Feature Items. In order to better describe the behavioral characteristics of users, the various behaviors of users can be regarded as a

set. Here, the user behaviors are defined as $\text{actitem} = \{\text{act}_{id}, \text{sem}\}$, act_{id} is the ID of the user behavior to which they belong, and sem is the behavior represented by the behavior feature item.

In view of the above settings, this paper obtains the formula for calculating the behavior degree as follows:

$$\text{ActItem}_{\text{act}_i} = \left(\frac{\sum_{\forall \text{act} \in \text{act}_i} c_{\text{act}}}{\sum_{\forall \text{act}} c_{\text{act}}} + \frac{\sum_{\forall \text{act} \in \text{act}_i} d_{\text{act}}}{\sum_{\forall \text{act}} d_{\text{act}}} \right). \quad (1)$$

Among them, $\sum_{\forall \text{act} \in \text{act}_i} c_{\text{act}}$ represents the number of times that the behavior occurs in the face of online marketing, $\sum_{\forall \text{act}} c_{\text{act}}$ represents the number of times the user’s behavior occurs in the entire complete trajectory, $\sum_{\forall \text{act} \in \text{act}_i} d_{\text{act}}$ represents the duration of the current behavior, and $\sum_{\forall \text{act}} d_{\text{act}}$ represents the total time of the user’s behavior in the entire complete trajectory.

The weight of a feature item is calculated according to the entire calculation formula. If it is greater than a certain threshold, it means that this behavior is more important to the user. If it is greater than this threshold, it indicates that this behavior of customers is relatively frequent, and it is an important data source for researching customer behavior and building customer behavior models. Thus, it is included in the user’s behavior set, and its corresponding behavior feature items are collected into the behavior feature item set.

3.2.2. Calculation of the Weight of the User’s Daily Behavior Feature Item. Through the above research, the behavior feature items of the user’s daily behaviors can be obtained. However, in the process of modeling, the extracted behavior feature items need to be calculated with a certain weight. The weight value indicates the influence of the feature item on the user trajectory data model. There are many ways to calculate the degree and weight:

- (1) Word frequency: the word frequency is defined as the comparison value between the number of documents D containing the feature item and the number of all documents, which can be expressed as the following formula:

$$DF = \frac{D}{D_{\text{total}}}. \quad (2)$$

- (2) Mutual information (MI): mutual information is used to measure the amount of information that an event contributes to other events. In the selection process of feature items, mutual information can be used to measure the degree of correlation between keyword k and category L_i , as shown in

$$MI(K, D_i) = \frac{p((K, D_i))}{P(T) * P(D_i)}. \quad (3)$$

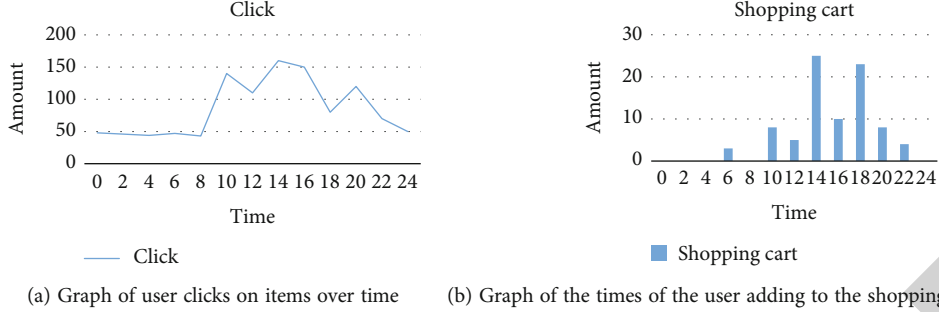


FIGURE 6: Line chart of the number of times users clicked items and added to carts.

TABLE 3: Time series clustering results of user clicks.

	Q1 = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
	Q2 = {11, 12, 13, 14, 15, 16, 17, 18}
	Q3 = {19, 20, 21, 22, 23}
	Q4 = {24, 25, 26, 27, 28, 29, 30}
8	Q5 = {31, 32, 33, 34, 35}
	Q6 = {36, 37, 38}
	Q7 = {39, 40, 41, 42, 43, 44, 45, 46}
	Q8 = {47, 48}

Among them, K represents the keyword, D represents the document D containing the feature item, and P represents the contribution information of an event.

- (3) TF-IDF method: TF-IDF is a weighting technique often used in user information retrieval and data mining to evaluate the importance of a word to a certain document set or a certain corpus. Through TF-IDF, the user's daily behavior and activity types can be counted, as well as the representativeness and discrimination of each word to the user. Its calculation formula is as follows:

$$tf_{i,j} = \frac{m_{i,j}}{\sum_k m_{k,j}}. \quad (4)$$

Reverse document frequency measures the general importance of words, which is of great significance in the field of natural language processing, especially for language models and text classification. The IBF of a specific word can be calculated by dividing the total number of documents by the number of documents containing the word and then taking the logarithm of the quotient. The calculation formula is as follows:

$$ibf_i = \log \frac{|B|}{|\{j : s_i \in b_j\}|}. \quad (5)$$

- (4) Analytic hierarchy process: it solves problems by decomposing, by dividing a large problem into dif-

ferent levels. The highest level is the target level, the lowest level is the specific solution, and the rest are various indicators that affect decision-making [18]. The type of model that this paper needs to build is a hierarchical vector model. In the initial stage of model creation, AHP is used. Through AHP, the model is divided into three layers. The first layer is users, the second layer is behavior, and the third layer is feature items.

The main flow chart of AHP is shown in Figure 7.

When calculating the importance of a word, it simply counts the number of occurrences of the word. However, in actual situations, it is necessary to consider not only the number of occurrences of the user's behavioral feature items but also the weight of each user's behavioral feature items. So $m_{i,j}$ in Formula (4) is calculated as

$$m_{i,j} = \sum_{k=1}^n \text{item}_{k,i} * w_{k,i}. \quad (6)$$

Then, the same can be obtained, and Formula (4) can be changed to

$$tf_{i,j} = \frac{\sum_q \text{item}_q * w_q}{\sum_k \sum_q \text{item}_{k,q} * w_{k,q}}. \quad (7)$$

$tf_{i,j}$ represents the importance of user j 's behavioral feature i , and the expression for reverse behavioral frequency ibf_i becomes

$$ibf_i = \log \frac{|B|}{1 + \sum_j |\{j : t_i \in b_j\}| * w}. \quad (8)$$

Based on the above operations, the weight calculation formula of the behavior feature item of user j can be finally obtained as follows:

$$w_{i,j} = tf_{i,j} * ibf_i = \frac{\sum_q \text{item}_q * w_q}{\sum_k \sum_q \text{item}_{k,q} * w_{k,q}} * \log \frac{|B|}{1 + \sum_j |\{j : t_i \in b_j\}| * w}. \quad (9)$$

3.2.3. Representation of User Trajectory Data Model. The user model reflects the characteristics of the user's daily behavior to a certain extent and also reflects the user's point

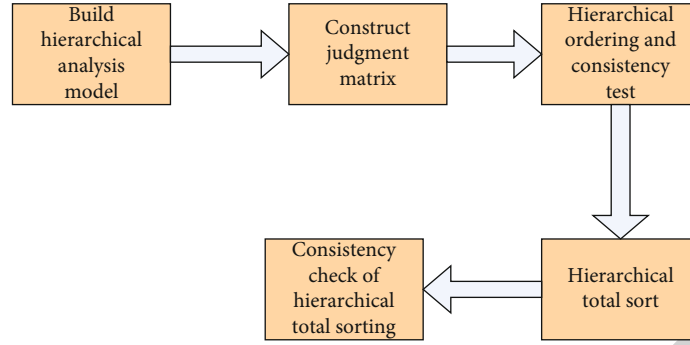


FIGURE 7: Main flow chart of AHP.

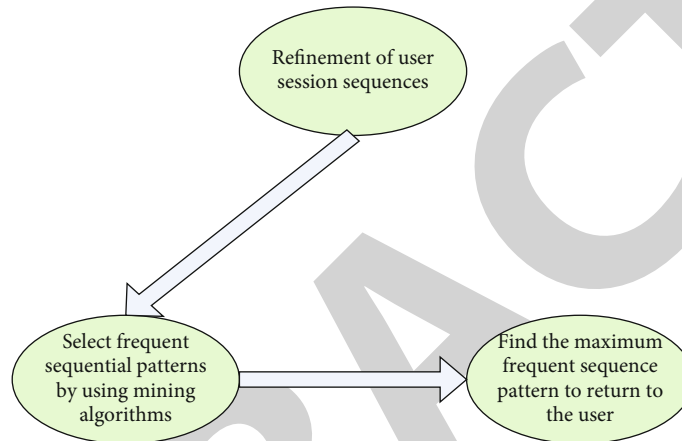


FIGURE 8: Sequence pattern mining process diagram.

of interest. According to the feature extraction algorithm mentioned above, an improved hierarchical vector model is used to describe the user's daily behavior and activity characteristics [19].

If the user has U with m characteristic behaviors and n different main behaviors, then, the user's trajectory data model can be expressed by the following formula:

$$\text{Model}_u = \{(S_1, W_1, M_1), (S_1, W_1, M_1), (S_n, W_n, M_n)\}. \quad (10)$$

3.3. Sequential Pattern Discovery Algorithm Based on Multiple Factor Constraints

3.3.1. Sequential Pattern Mining Algorithm. The so-called sequence refers to the combination of elements arranged in a certain order. In everyday life, there are examples of this everywhere, such as the steps of the user's account opening procedures in the bank, the sequence of the consumer's consumption behavior in the store, and the pages that the user visits the website and browses in turn.

The so-called sequential pattern mining is the process of finding all frequent sequences with support greater than or equal to the threshold for a data set and a performance index, that is, the minimum support threshold [20]. The sequence pattern mining process is shown in Figure 8.

3.3.2. Inadequacies of Traditional Apriori-Like Algorithms. As a classic algorithm for sequential pattern mining, the Apriori-like algorithm is one of the most commonly used algorithms in log mining and occupies an important position in the research and application process. However, due to some limitations of its own, the actual effect of this algorithm is not very good, which is mainly reflected in two aspects including the efficiency of the algorithm and the effectiveness of the mining results [21].

First, the space-time overhead of this algorithm is very large, and the algorithm efficiency is not high. The basic idea and implementation steps of the Apriori-like algorithm for sequential pattern discovery are based on the traditional Apriori algorithm. In practical applications, the scale of this database is often very large, which brings huge time overhead. This kind of algorithm is an iterative method, which will repeatedly execute the above process, so the huge space-time overhead often becomes a heavy burden for mining work and affects the efficiency of mining.

Secondly, in terms of the validity of the mining results, although the coverage of frequent sequence pattern mining is relatively high, it cannot distinguish the sequence patterns with high meaningful value from the useless patterns very well. It cannot really distinguish the difference between sequences, cannot really reflect the user's interest, and cannot filter the results carefully and effectively.

Facing the problem of low validity of mining results, an intuitive solution is to introduce other evaluation mechanisms. Based on this idea, this paper introduced a new constraint factor to improve the original method.

3.3.3. Algorithm Improvement Based on Multiple Factor Constraints

(1) *Page Interest Factor*. Page interest is a measure, which is a measure of the user's interest in a page. For the measurement of page interest, there are generally two methods of composition, one is based on the number of user visits, and the other is based on the time of user visits [22].

First of all, this paper defines the user's page interest in page r_i as B_i . If the total number of user visits to this page is n and the total number of visits to all internal pages is N , there are

$$B_i = \frac{n}{N}. \quad (11)$$

Assuming that for any page $r_i \in R$ and any user $U_j \in U$ who has visited the page, if the total time is t_i^j when U_j stays on r_i , the number of visits is m_i^j , and the average visit time of U_j on the page is $\text{ave}_j(r_i) = t/m_i^j$, then there are

$$\text{ave}_j(r_i) = \frac{t_i^j}{m_i^j}. \quad (12)$$

For any page $r_i \in R$, the number of users who have visited the page is M_i , and the average access time of page r_i is $\text{AVE}(r_i)$; then,

$$\text{AVE}(r_i) = \frac{\sum_{j=1}^{M_i} \text{ave}_j(r_i)}{M_i}. \quad (13)$$

Next, the page interest degree factor is defined which is used in this article: for any $r_i \in R$, its page interest degree is recorded as A_i ; then there are

$$A_i = \frac{\text{AVE}(r_i)}{\max_AVE(R)}. \quad (14)$$

In actual use, it would be found that the page interest factor calculated by different pages may vary greatly. Therefore, the optimization formula in this paper is

$$A_i = 0.3, \frac{\text{AVE}(r_i)}{\max_AVE(R)} \leq 0.3. \quad (15)$$

When the page interest degree is between 0.3 and 1, there is the following formula:

$$A_i = \frac{\text{AVE}(r_i)}{\max_AVE(R)}, 0.3 \leq \frac{\text{AVE}(r_i)}{\max_AVE(R)} \leq 1. \quad (16)$$

After introducing the page interest factor, another important factor is considered that affects the user's visit behavior, that is, page importance.

TABLE 4: Data table of sequential patterns based on multiple factors.

Page sequence	Weighted support	Page sequence	Weighted support
A	75	EG	16
B	150	AG	0
C	56	BG	0
G	27	CA	0
AB	6.75	CEG	15.4
BE	0	BCE	17.1
CG	0	BCEG	6.5

TABLE 5: Sequence pattern data table based on traditional Apriori-like algorithm.

Page sequence	Weighted support	Page sequence	Weighted support
A	50	ACE	10
CE	50	BCEG	20
AB	30	ABCE	10
BEG	20	ABCEG	10

(2) *Page Importance Factor*. The importance of a page also affects the degree of user interest in the page to a large extent. There are many ways to measure the importance of a page. This article adopts a method to measure the importance of a page, which is the content link ratio. This paper defines that for any page $r_i \in R$, its content link ratio CLR_i is a ratio. If the number of links contained on page r_i is L_i and the amount of information contained on the page is G_i , then there are

$$\text{CLR}_i = \frac{G_i}{L_i}. \quad (17)$$

In order to facilitate the calculation, CLR_i need to be mapped to the (0, 1] interval. For any page $r_i \in R$, its page importance is recorded as V_i ; then there are

$$V_i = \frac{\text{CLR}_i}{\max_CLR_R}. \quad (18)$$

With the user interest factor and page importance factor, the improved Apriori-like algorithm is be further improved, so that the frequent sequence patterns mined by it can better reflect the user's browsing interest.

For the algorithm proposed above, the core criterion for frequent sequence determination is the support degree. Next, a new support degree calculation method proposed for the improvement of the above algorithm, as shown in the formula:

$$W_Support(q) = \text{Support}(q) \frac{\sum_{i=1}^m A_i V_i}{m}. \quad (19)$$

Among them, the weighting factor for revising the traditional support degree is the arithmetic mean value of the

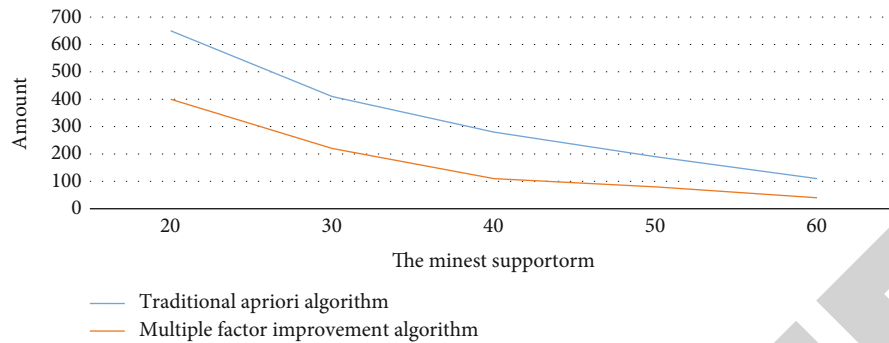


FIGURE 9: Comparison of experimental results between traditional Apriori-like algorithm and multifactor sequential pattern discovery algorithm.

products of all page interest degree factors and page importance factors in the sequence.

4. Experiment of User Behavior Modeling Based on Web Data Mining

To facilitate observation of the results, two experiments are employed to test the improved frequent sequence mining algorithm. One is a simulation experiment using fictitious data. The data comes from a simulation experiment collected by a simulated network. Compared with the real data, the amount of data is less and the calculation effect is more intuitive in order to illustrate the process of algorithm execution, and the other is an experiment using real data to show the effect of the algorithm.

4.1. Simulation Experiment of Fictitious Data. For the first experiment, assuming that each record in Table 3 occurs 10 times in the database. The total number of database transactions is 100, and the minimum support threshold is given as 10. The details are shown in Table 4.

After processing by this algorithm, 5 frequent sequences were got, the maximum frequent sequences are $\langle B, C, E \rangle$ and $\langle C, E, G \rangle$. In contrast, if the traditional Apriori-like algorithm is used to mine and analyze the same data, the execution data table is shown in Table 5. It can be seen that the improved algorithm has a very obvious filtering effect on frequent sequence patterns. It can be seen from the table that the multifactor-based sequence pattern discovery algorithm filters the experimental data from $(A, 75)$ to $(BCEG, 6.5)$. Obviously, the screening effect of this method is very good.

4.2. Real Data Experiment. The experimental data used in this paper comes from the session identification and path supplement of user behavior web data for network marketing. For the information related to the weighting factor, it is obtained through additional program statistics and saved in the form of auxiliary table files for the algorithm to directly call used to avoid excessive overhead during algorithm execution.

Using the traditional Apriori-like algorithm as a comparison, 20, 30, 40, 50, and 60 are chosen as the minimum support thresholds. After the mining work of the two

methods, the following results can be obtained, as shown in Figure 9.

The first thing to point out here is that the number in the graph refers to all the frequent sequences mined, not the maximum frequent sequence. It can be seen from the above results that with the increase of the minimum support, fewer and fewer sequences are screened. However, it is obvious that the sequential pattern discovery algorithm based on multiple factor constraints has better screening effect than the traditional Apriori-like algorithm from the beginning. In fact, many sequences like $\langle A, B \rangle$ and $\langle A, C \rangle$ in Experiment 1 were filtered out very early, and the newly added weight factor has played a good filter effect on low-value sequences.

4.3. Experimental Results. The experimental results have shown that the sequential pattern discovery algorithm based on multiple factor constraints can greatly save the model runtime cost when building a user behavior model, which is roughly 30%-40%. At the same time, the user behavior model constructed by the algorithm has a 37% increase in the accuracy of user behavior simulation and prediction. Of course, there is still no good evaluation method and standard for how to accurately measure the value of a frequent sequence pattern, which does not mean that the fewer the number of sequences mined, the better the results. This should be evaluated and revised according to the actual situation. For example, the number of sequences obtained in the study of the purchase demand of Internet marketing users is too large, but the reason for too many is that the sample of such demand data sources is complex. Too many of them are objective situations, so at this time, it is not possible to over-screen the experimental data to avoid destroying the authenticity and objectivity of the data. However, it is certain that the improved algorithm proposed in this paper can indeed play a good role in correcting the mining process, filter out a large number of useless patterns, and reduce the redundancy of mining results to a certain extent.

5. Conclusions

Web data mining is a young technology and a promising discipline. With the continuous rapid development of the Internet, various network-based services and e-commerce

activities are increasingly prosperous, and the demand for analysis of user behavior patterns would increase. As an important category of web mining, web data mining starts from recording the data that users face in the online marketing process, and directly analyzes and mines user behavior patterns, so it has a wide range of application prospects. This paper referred to the sequential pattern discovery algorithm based on multiple factor constraints in the construction of the network marketing user behavior model, which is of great significance for changing the previous network marketing. Through the introduction of such technologies as web data mining, the research on the construction of user behavior model of network marketing can be more complete and intelligent, so as to realize the update and iteration of network marketing in the era of big data.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Conflicts of Interest

The authors declare that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

References

- [1] I. H. Sarker, A. Colman, M. A. Kabir, and J. Han, "Individualized time-series segmentation for mining mobile phone user behavior," *The Computer Journal*, vol. 61, no. 3, pp. 349–368, 2018.
- [2] J. Qian, F. Qiu, W. Fan, R. Na, and S. Tang, "Privacy-preserving selective aggregation of online user behavior data," *IEEE Transactions on Computers*, vol. 66, no. 2, pp. 326–338, 2017.
- [3] D.-H. Shin and F. Biocca, "Explicating user behavior toward multi-screen adoption and diffusion," *Internet Research*, vol. 27, no. 2, pp. 338–361, 2017.
- [4] K. Singh, P. Singh, and K. Kumar, "User behavior analytics-based classification of application layer HTTP-GET flood attacks," *Journal of Network and Computer Applications*, vol. 112, pp. 97–114, 2018.
- [5] M. Esdar, U. Huebner, J. D. Liebe, and J. Huesers, "Understanding latent structures of clinical information logistics: a bottom-up approach for model building and validating the workflow composite score," *International Journal of Medical Informatics*, vol. 97, pp. 210–220, 2017.
- [6] L. Gang, "Toward a progress indicator for machine learning model building and data mining algorithm execution: a position paper," *Acm Sigkdd Explorations Newsletter*, vol. 19, no. 2, pp. 13–24, 2017.
- [7] J. Agirre, "Strategies for carbohydrate model building, refinement and validation," *Acta Crystallographica*, vol. 73, no. 2, pp. 171–186, 2017.
- [8] L. Yang, C. Shi, L. Li, and Y. Li, "High-throughput model-building and screening of zeolitic imidazolate frameworks for CO₂ capture from flue gas," *Chinese Chemical Letters*, vol. 31, no. 1, pp. 227–230, 2020.
- [9] I. Kazanidis, S. Valsamidis, E. Gounopoulos, and S. Kontogiannis, "Proposed S-Algo+data mining algorithm for web platforms course content and usage evaluation," *Soft Computing*, vol. 24, no. 19, pp. 14861–14883, 2020.
- [10] M. Jeffery, L. Egli, A. Gieraltowski, J. Lambert, and R. Sharma, "Air France Internet marketing: optimizing Google, Yahoo!, MSN, and Kayak Sponsored Search," *Kellogg School of Management Cases*, vol. 1, no. 1, pp. 1–18, 2017.
- [11] A. Bjelica, S. Aleksi, S. Golorbin-Kon, D. Sazdani, L. Torovi, and J. Cveji, "Internet marketing of cardioprotective dietary supplements," *The Journal of Alternative and Complementary Medicine*, vol. 26, no. 3, pp. 204–211, 2020.
- [12] F. Lu, M. Hu, W. Zhou, and K. Y. Wang, "Based on "Computer Internet +" rural homestay tourism product marketing," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2809541, 7 pages, 2022.
- [13] Z. Xu, M. M. Kamruzzaman, and J. Shi, "Method of generating face image based on text description of generating adversarial network," *Journal of Electronic Imaging*, vol. 31, no. 5, article 051411, 2022.
- [14] F. U. Muram, H. Tran, and U. Zdun, "Systematic review of software behavioral model consistency checking," *ACM Computing Surveys*, vol. 50, no. 2, pp. 17.1–17.39, 2018.
- [15] A. Freij, K. Walid, and M. Mustafa, "Deep learning model for digital sales increasing and forecasting: towards smart E-commerce," *Journal of Cybersecurity and Information Management*, vol. 8, no. 1, pp. 26–34, 2021.
- [16] Z. Bian, "Evaluation of WeMedia Ads marketing based on the demand of Internet economy," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 8077092, 11 pages, 2021.
- [17] J. Xiao, W. Wang, and S.-B. Tsai, "Coupling of agricultural product marketing and agricultural economic development based on big data analysis and "Internet+"," *Mobile Information Systems*, vol. 2021, Article ID 3702064, 10 pages, 2021.
- [18] M. Song, T. Niu, L. Wang, and X. Wang, "The validity of marketing strategy of tourist attractions based on experiential marketing," *Security and Communication Networks*, vol. 2022, Article ID 1825171, 10 pages, 2022.
- [19] Y. Zhu and C. Wang, "Study on virtual experience marketing model based on augmented reality: museum marketing (example)," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2485460, 21 pages, 2022.
- [20] S. Fountoukidou, J. Ham, U. Matzat, and C. Midden, "Effects of an artificial agent as a behavioral model on motivational and learning outcomes," *Computers in Human Behavior*, vol. 97, pp. 84–93, 2019.
- [21] M. Rahmani-Andebili, H. Shen, and M. Fotuhi-Firuzabad, "Planning and operation of parking lots considering system, traffic, and drivers behavioral model," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 9, pp. 1879–1892, 2019.
- [22] X. Ma and S. Lv, "Financial credit risk prediction in internet finance driven by machine learning," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8359–8367, 2019.