Research Article

Research on E-Commerce Customer Feature Extraction Question Answering System Based on Artificial Intelligence Semantic Analysis

Wenbo Niu [ORCID]

Xi'an International University, Xi'an 710077, China

Correspondence should be addressed to Wenbo Niu; 2017041026@stu.gzucm.edu.cn

In order to analyze e-commerce customer behavior and preference, a migration identification method of consumer behavior tendency is proposed. Data mining technology is adopted to mine social data in individual online we-media platforms and to mine individual personal attributes and preferences from their unconscious social language. Its methods are through the customer identification model construction related research, consumer preference identification and analysis related research, based on data mining technology of consumer preference identification and analysis, and the introduction of feature extraction method: semantic analysis. According to the data, there are 2,990 customer interest consumption forecasts, 1,836 customer social network data consumption forecasts, and 3,652 customer preference consumption forecasts. In order to screen out the main factors that have the greatest impact on consumer behavior from all kinds of consumer behavior propensity factors, the multiple step-based regression method is adopted for factor selection. Because of the large difference in the multidimensional dynamic vector, the corresponding consumer behavior tendency changes greatly, so the migration identification method of consumer behavior tendency is feasible.

1. Introduction

Current e-commerce market research contains a large amount of e-commerce planning and forecasting information. Many large B2B service providers predict the next generation of products and services by studying user trends. This will lead the buyer to the most attractive market in the near future [1]. For example, the study of shopping cart systems has a great impact on how service providers better cater to market needs, technologies, interests, and behaviors. High-quality e-commerce market research can reduce unnecessary expenses of enterprises and is a method to obtain the cognition of products and users [2]. At present, there are many aspects of e-commerce market research, including cross-selling and referral selling. At the end of the day, companies spend a lot of money to study these problems in order to compete with their competitors for customers. In the course of e-commerce practice, it is actually very easy to locate the "best customer" through some technical means. Out of a large number of potential customers, only a small number of customers eventually become the so-called "best customers." Among them, only a few potential customers become new customers. Only a few new customers are active customers. Only a small number of active customers become repeat customers and purchase again. These are called stable customers [3]. Only a few of the stable customers increase their purchases and deepen their relationship with the site, and these people become the "best customers." Today, in the era of big data, data mining technology is used to obtain online user behavior data and conduct classification and integration, at the same time, combined with user information, in this way, more specific customer images can be obtained. Such a full sample can not only objectively reflect the characteristics of their consumer behaviors, but also, to a certain extent, correlate and predict their consumer behavior tendency [4]. Based on this, this paper proposes to use the customer social behavior data on the we-media platform to build a customer identification model.

Pang et al. proposed a method for calculating semantic similarity of Chinese words based on WordNet, which has become the main basis for calculating emotional polarity strength of Chinese words [5]. Peng et al. studied the problem of feature recognition of comment objects. They proposed to use the association rule method to mine the stable and hidden feature attributes of comments and put forward the overall evaluation of goods according to the semantic tendency of sentences [6]. Sumerta put forward the conclusion that through regression analysis, clothing sales can be improved through reasonable improvement of quality, optimization of fabric selection, and improvement of logistics service [7]. As one of the most important information can reflect user intention, the analysis of consumer behavior on the website information and mining was performed, to be able to get the user's interest and purchase intentions, to build consumer interest and demand model, to build a targeted change sales strategy, to adjust service to meet customer needs, and to improve the user viscosity. However, the current e-commerce market research based on consumer behavior analysis is by no means smooth sailing. In terms of statistical analysis, human dynamics in e-commerce market has been studied from the perspective of complex systems, but there are still a lot of problems to be empirically analyzed. In the aspect of feature analysis, the current research work is mainly based on the experience of researchers or questionnaires to decide the selection of features. With the increase of the number of features, the noise also increases correspondingly, resulting in inaccurate results. For example, a large amount of e-commerce information exists in objects with different attributes, such as products and customers. Existing cluster analysis studies can only use the relationship between single objects for analysis and mining, ignoring the rich information contained in the data, resulting in different analysis results. In the current research, the migration identification method of consumer behavior tendency is proposed [8]. Data mining technology is adopted to mine social data in individual online we-media platforms and to mine individual personal attributes and preferences from their unconscious social language. In this method, relevant research methods are built through customer identification model, consumption preference identification and analysis, consumption preference identification and analysis based on data mining technology, and text feature extraction method: semantic analysis method [9]. According to the data, there are 2,990 customer interest consumption forecasts, 1,836 customer social network data consumption forecasts, and 3,652 customer preference consumption forecasts. In order to screen out the main factors that have the greatest impact on consumer behavior from all kinds of consumer behavior propensity factors, the multiple step-based regression method is adopted for factor selection. Because of the large difference in the multidimensional dynamic vector, the corresponding consumer behavior tendency changes greatly, so the migration identification method of consumer behavior tendency is feasible.

2. Methods

2.1. Related Research on Customer Identification Model Construction. Data selection is the key to build the customer identification model. Only the most objective data can build

the model with the strongest identification ability. The data sources in the current research on constructing customer identification model are divided into two categories: questionnaire data and database data. The research on obtaining data through questionnaires is a relatively traditional way and the research period is relatively short. There are many research studies on customer value identification based on the questionnaire, and the data of listed companies and a customer value identification model based on BP neural network is constructed. Then, the relationship between customer value and enterprise profit is analyzed. The questionnaire survey was used to target market customer needs to identify and build a model. In recent years, there have been an increasing number of research studies using database data, most of which use data mining technology to mine customer demand data from the database previously composed of production, operation, sales, and other links, and establish identification models to solve practical problems [10]. For example, the decision tree and logistic algorithm are used to mine the customer database of mobile Fexin business and establish the customer prediction model. By using data mining technology to mine supermarket customer database, loyalty-profit customer segmentation model was established, different types of customer groups were identified, and the changing rule of customer value was predicted. Taking the marketing data of Sanqiang group as the object, the "user portrait" database mining is used to establish precision marketing segmentation model, reconstruct consumer demand, and accurately identify and locate consumer groups. Such data improve the shortcomings of subjective questionnaire data and insufficient sample size, but such data still have limitations. There are various driving factors for consumers' consumption behaviors, and both subjective and objective factors coexist, when determining customer attributes. The consumption behavior data caused by the subjective active factors of customers are the best so as to accurately analyze customer attributes; therefore, it is not objective to deduce the customer attributes only through the established consumption behavior, and it is impossible to fully grasp the real attributes of customers. Figure 1 shows the changes in the size of online shoppers in recent years.

2.2. Research on Identification and Analysis of Consumption Preference. With the emergence of various information collection terminals, massive data generated by online and offline exchanges are stored in enterprise databases as the source data for analysis of consumer preferences [11]. Because of the huge amount of data, data mining has become the main method to identify and analyze consumer preferences based on database data extraction. From the point of view of data mining, methods are divided into two categories: supervised data mining methods and unsupervised data mining methods.

2.2.1. Supervised Data Mining Methods. Common supervised data mining methods include Logit analysis method and decision tree analysis method. The difference between
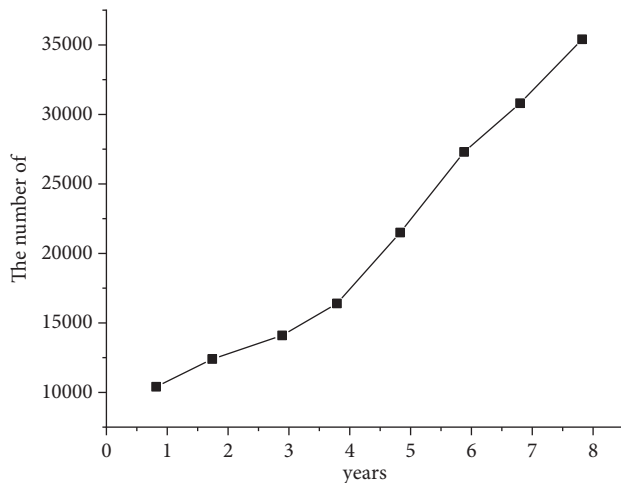
FIGURE 1: Changes of online shopper planning.

the standard Logit model and the hybrid Logit model is that the former's assumptions ignore the differences between consumers; in the latter, the coefficient random error is used to express the preference difference among individuals, the decision-making behavior of consumers is modeled, and the consumer preference is reflected by the coefficient. Decision tree analysis method has better processing effect on discrete data, a clothing style preference model was constructed based on the ID3 decision tree classification algorithm, and the customer style preference problem was transformed into a decision tree induction problem.

2.2.2. Unsupervised Data Mining Methods. Common unsupervised data mining methods include association rules and clustering. The former is the most basic analysis method, the premise of traditional association rules is that consumers' preferences are fixed and the research subject is commodities; therefore, the bidirectional association rule method with consumers as the main body is proposed. Although the method improves accuracy, the data do not take into account the data other than purchasing behavior, which can only be obtained through the intelligent e-commerce IoT of GPS, infrared sensors, etc. All the information in the three processes of prepurchase, purchase, and post-purchase is obtained; therefore, two-way association rules are more suitable for analyzing consumer preferences based on e-commerce data. Cluster analysis can best reflect the characteristics of unsupervised method in data mining, which can distinguish different consumer groups from the consumer database, and summarizes the consumption patterns or habits and preferences of each type of consumers, but there are many deficiencies, especially in the analysis of real data. It is difficult to determine the input parameter value, and the slight change of the parameter value leads to the large variation of the clustering result of consumer preference, so the clustering algorithm relying only on global parameters cannot describe the real structure of consumer preference well .

3. Results and Analysis

3.1. Identification and Analysis of Consumption Preference Based on Data Mining Technology

(1) Build a customer identification model.

The established customer preference identification model uses data mining technology to collect, sort out, and analyze the basic information and online behavior data of customers on the we-media platform so as to identify and lock customers according to consumption characteristics and other indicators. Construction of individual customer identification database: when building a customer identification model, the first step is to mine the behavior characteristic data of individual customers (original target) to form the basic label, that is, to build the individual customer identification database. Through the study and sorting out of the previous research studies of scholars, the "consumer graph" constructed in precision marketing based on the five dimensions proposed in the portrait of social media with high conversion rate is studied. Based on the "4C" theory and combined with the data features used, the framework of individual customer identification database is proposed, which mainly includes two aspects: basic customer information and online behavior data, as shown in Figure 2.

Community research is carried out on the basis of the establishment of individual customer identification database. This step makes the customer identification model more comprehensive and radiates its friends, family members, and colleagues through individual customers. Each individual customer as the center of the circle to form a community, the intersection of multiple communities will be able to make the individual more comprehensive label.

Cross-sectional research is to lock and mine social data on the we-media platform of people related to the original goal (such as individuals with frequent interactions and users with clear groups), enrich the basic label of the original target, and describe it from different latitudes and granularities. With the popularity of social software, individuals have diversified their use of social software. In order to distinguish work and life, they will apply for multiple accounts on the same social software; therefore, longitudinal research is to mine the data of different social accounts of the original target, including different accounts on the same platform and accounts on different platforms (by default, an individual has only two accounts at most on different we-media platforms). Comprehensive identification can be performed based on the login device or login IP address. The social network analysis method is introduced above, which is a quantitative method to formally describe the social network. A social
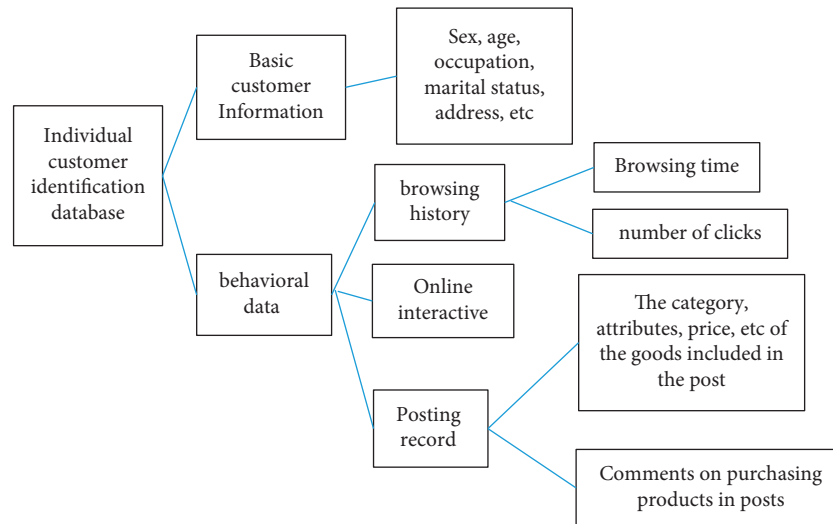
FIGURE 2: Individual customer identification database.

network consists of a collection of nodes and wires, each node representing an individual (the original goal), connections indicate relationships, which can be relatives, friends, colleagues, or even strangers with the same commodity preferences. Individuals are connected through relationships and eventually form social networks, each label sharing mutual reference, then, form each three-dimensional user. The original target is at the heart of the social network, users who are more connected and more active online than other customers, the larger and more complex the social network, the more comprehensive the customer information will be. The social network structure is shown in Figure 3.

(2) Identify and analyze preferences.

Based on the customer identification model, data related to consumption preferences are mined to identify and analyze consumption preferences. The steps and methods are as follows: first, consumer interest mining based on online social data was performed. For the discovery of customer interest, in the process of online customer participation in information dissemination, different interests and concerns will be presented at different times, which can be called dynamic interest migration. Firstly, the characteristics of feature items as customer interests in the dynamic migration process are analyzed, and the text feature extraction method is proposed. On this basis, the finite mixed probability model is used to study the identification method of customer interest and the merging method of new interest content is also used. The network customer is the main body of information transmission, and the customer with large centrality plays a vital role in information transmission. The preliminary research found that in online social network, the customer nodes with a large degree of connection can help to spread information to more customers, and the
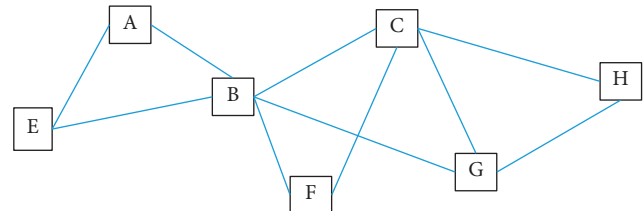

FIGURE 3: Social network structure diagram.

customer nodes with a large number of medium can help to expand the range of information transmission, customer nodes with large aggregation are helpful to local information propagation, and the positive and negative attributes of node connection can influence the choice of information propagation path. Based on these structural attributes, important customer nodes with different centrality can be selected so as to provide important customer attributes for the establishment of specific types of customer behavior models. Second, customer consumption behavior pattern mining based on online social data was performed. In online social networks, the behavior patterns of some subsets of customer nodes are similar to each other due to the similar roles, concerns and interests of certain types of customer nodes. Therefore, on the basis of large-scale customer node behavior data, based on the convergence of customer node behavior patterns and the similarity of behavior characteristics, the antinoise clustering algorithm is used to analyze the complete set of customer nodes, and several subsets of customer nodes with similar behavior patterns are obtained. Furthermore, the frequent item association rule mining algorithm is selected to analyze a subset of customer nodes, and the main behavior patterns of customers in the subset can be obtained. Thirdly, customer preference recognition and analysis based on customer consumption behavior pattern was

Table 1: Customer consumption behavior forecast table.

Prediction of the direction	Prediction of the outcome
Customer interest consumption	2990
Customer social network data	1836
Customer preference	3652

performed. Through semantic analysis, customer interest and behavior data obtained from online social media and consumer behavior tendency data obtained from other types of data sources are analyzed. The change and development process of different types of consumption behaviors was recorded in chronological order and formed a time series. Based on spatial reconstruction, the measured values were constructed into a group of multidimensional dynamic vectors [12]. The large difference in the time series means that the consumption behavior tendency contained in the original data changes greatly. This analysis is therefore able to identify shifts in consumer behavior, and due to the introduction of new types of data, such as online social networks, some changes that cannot be reflected in the consumer behavior data released based on the traditional questionnaire survey can also be identified so as to make the identification results more accurate and perfect.

(3) In order to implement the prediction of new commodity buyer behavior based on decision tree induction, this chapter uses the sales transaction data of Jingdong Mall. Taking into account the changes in the initial form of Jingdong Mall, this paper only selects the newly listed products in memory category in 2009 and 2010. In order to avoid the impact of the website's anniversary promotion on customers' purchasing behavior, the paper finally selects the newly listed products from July 2009 to February 2010. After the new product is put on the shelves, the personal page links of the top 20 customers who bought the product first will be obtained according to the purchase time of the customers. Then, the personal purchase history will be entered into the personal purchase history to obtain all the purchase history of the new product on the website. The training set was constructed according to the characteristics obtained by the above method, and the data of customer purchasing behavior was predicted, as shown in Table 1.

4. Conclusions

In this study, we used data mining technology to mine and integrate social data in customer we-media platform and built customer identification model so as to mine and identify consumption preference. In addition, this paper puts forward the research ideas and suggestions of using text feature extraction method, antinoise clustering algorithm, and semantic analysis method to categorize and summarize

consumption preferences and explore their change rules. According to the forecast results, there are 2990 customer interest consumption forecasts, 1836 customer social network data consumption forecasts, and 3652 customer preference consumption forecasts. It indicates that in order to screen the main factors that have the greatest impact on the consumer behavior from all kinds of behavioral propensity factors, multiple step-based regression method is adopted for factor selection. Because of the large difference in the multidimensional dynamic vector, the corresponding consumer behavior tendency changes greatly, so the migration identification method of consumer behavior tendency is feasible. With these research bases, we can further expand our research direction to the current popular applications of recommendation systems in the field of e-commerce market research so as to improve the accuracy of recommendation systems and provide good user experience.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] H. Mei and X. Li, "Research on e-commerce coupon user behavior prediction technology based on decision tee algorithm," *International Core Journal of Engineering*, vol. 5, no. 9, pp. 48–58, 2019.

[2] Z. Yanrong, Z. Zhijie, Y. Jing, and W. Kun, "Research on e-commerce consumer behavior prediction based on rough sets," *International Journal of U- & E-Service*, vol. 8, no. 4, pp. 69–76, 2015.

[3] C. Chen, C. Hou, J. Xiao, and X. Yuan, "Purchase behavior prediction in e-commerce with factorization machines," *IEICE - Transactions on Info and Systems*, vol. 99, no. 1, pp. 270–274, 2016.

[4] Wang, "A decision method for returns logistics based on the customer's behaviour in e-commerce," *Procedia Computer Science*, vol. 60, pp. 1506–1515, 2015.

[5] J. Pang, "Research on e-commerce applications based on neural network," *Journal of Computational and Theoretical Nanoscience*, vol. 13, no. 8, pp. 5227–5230, 2016.

[6] M. Peng, Y. Qin, C. Tang, and X. Deng, "An e-commerce customer service robot based on intention recognition model," *Journal of Electronic Commerce in Organizations*, vol. 14, no. 1, pp. 34–44, 2016.

[7] I. K. Sumerta, "Online consumer behavior on using social media on e-commerce, based on the aisas model approach. case study; bukalapak, tokopedia and blili.com," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1, pp. 234–242, 2019.

[8] R. Aringhieri, D. Duma, and V. Fragnelli, "Modeling the rational behavior of individuals on an e-commerce system," *Operations Research Perspectives*, vol. 5, pp. 22–31, 2017.

[9] L. B. Davis, S. X. Jiang, S. D. Morgan, I. A. Nuamah, and J. R. Terry, "Analysis and prediction of food donation behavior for a domestic hunger relief organization," *International Journal of Production Economics*, vol. 182, pp. 26–37, 2016.

[10] Y. Saito and E. Kusukawa, "Impact of information sharing regarding customer returns ratio on optimal sales strategy under e-commerce," *Industrial Engineering and Management Systems*, vol. 14, no. 2, pp. 111–121, 2015.

[11] N. Y. Sari and S. Hermawati, "The effect of shopping lifestyle, hedonic shopping motivation, and sales promotion on impulse buying behavior in e-commerce (case study of berrybenka consumer)," *Jurnal Ilmiah Ekonomi Bisnis*, vol. 25, no. 1, pp. 45–54, 2020.

[12] R. R. Izzati, N. S. Muntiah, and N. Hidayah, "An analysis of factor that influence the interests in behaviors of using accounting information systems based on e-commerce," *Jurnal AKSI (Akuntansi dan Sistem Informasi)*, vol. 5, no. 1, pp. 1–5, 2020.