

# Smart Marketing Data Collection and Analysis based on Logistic Regression Algorithm

Lei Zhang

Huaibei Vocational and Technical College, Anhui Huaibei 235000, China

Zanglei22z12@gmail.com

**Abstract**—Based on the essence of logistic regression algorithm, this paper discusses the specific aspects of cluster analysis in smart marketing from five aspects: customer segmentation, experimental market selection, sampling plan design, sales area determination, and market opportunity research. According to the application, a prediction model based on the radial basis function (RBF) is established to predict the actual network data flow. Simulation experiments show that, compared with the prediction of BP neural network, the prediction method of RBF neural network based on chaos theory has fast learning speed and high prediction accuracy.

**Keywords**—Smart Marketing, Data Collection, Logistic Regression, Cloud Computing

## I. INTRODUCTION

At this stage, I believe that every enterprise needs data and all like to collect data. The digitization of information provides the most effective way to ensure that companies operate more efficiently and make accurate decisions. In a joint survey conducted by big data AI companies and RockFuel, it is found that the number of people who use big data for sales is significantly higher than that of companies that do not use data analysis for marketing, and the value generated by big data analysis is far greater than its original value. actual value. These use data to analyze consumer companies and have more accurate judgments on discovering market trends and consumer consumption trends [1-6].

The collection and use of a large number of individual consumer raw data will produce new ways to understand consumer behavior and formulate marketing strategies. E-commerce is a consumer-oriented and individualized management method. Online store marketing makes highly targeted small group marketing and even individual marketing possible. Its biggest feature is that it is consumer-oriented. Consumers will have greater freedom of choice than in the past. They can find satisfying products on a global scale according to their individual characteristics and needs, regardless of geographical restrictions. By entering the corporate website or virtual store of interest, consumers can obtain more information and make shopping more personalized. The sales strategies of online stores based on online platform sales are obviously different from those of physical stores. Although online stores sold on online platforms also rely on publicity and promotion, in the end they are all realized through clickstream. Through the empirical analysis of Taobao platform, Chen Hongjiao and others gave the relevant factors affecting sales volume, and analyzed which indicators have a significant impact on the average monthly sales of online stores [7-14].

Liu Shuo and others divided store sales into three categories: high, medium, and low, established an orderly multi-categorical response variable model, and gradually screened out five variables that have a significant impact on sales. At this stage, I believe that every enterprise needs data and all like to collect data. The digitization of information provides the most effective way to ensure that companies operate more efficiently and make accurate decisions. In a joint survey conducted by big data AI companies and RockFuel, it is found that the number of people who use big data for sales is significantly higher than that of companies that do not use data analysis for marketing, and the value generated by big data analysis is far greater than its original value. actual value. These use data to analyze consumer companies and have more accurate judgments on discovering market trends and consumer consumption trends. The collection and use of a large number of individual consumer raw data will produce new ways to understand consumer behavior and formulate marketing strategies. As the amount of data accumulates, the data becomes super powerful and becomes more and more difficult to understand. In order to make better use of these data, people have explored a major data method change, namely: changing the scientific data analysis method from fitting data to market-oriented theory, and then to using data frame theory (Firestein, 2012) [15-21].

The application of this technology and method has greatly improved people's work efficiency, and also made big data more valuable and easier to be used by humans. This is the charm of big data, allowing us to "perceive the future." Zhou Jia studied the influencing factors of consumer purchase intention in online transactions, and empirically analyzed the influencing factors of consumer purchase intention through OLS and Tobit regression models. Wei Yun studied the factors influencing the competitiveness of Yunnan agricultural products e-shops, and analyzed the factors affecting the competitiveness of Yunnan agricultural products e-shops from the macro and micro levels. Chen Weiyang analyzed the influencing factors of customer satisfaction of B2C clothing online marketing, and exploratory proposed several promotion strategies and solutions for B2C clothing online marketing. Wu Yue'e and Huang Yi also did related research on the factors affecting customer purchases under the online platform. The above research basically studies the influencing factors and indicators under the network platform, and does not make corresponding analysis on the influence and structure of the factors and indicators. This article summarizes the most important indicators that affect the purchase rate of online customers as search proportions, mall proportions, click-through rates, number of treasures, and through trains, and specifically study the actual impact and positive

correlation of these indicators on customer purchase rates. It is still the role of negative correlation [22-24].

## II. THE PROPOSED METHODOLOGY

### A. Logistic Regression Algorithm

Logistic regression is used to study the probability  $p$  of a certain social phenomenon, such as the probability of a customer buying a certain product, and what factors the probability  $p$  is related to (these independent variables can be continuous or categorical), but Logistic regression requires the dependent variable to be a binary variable, with only two values of 0 and 1. The dependent variable  $p$  ( $y = 1$ ) is the object to be studied. The independent variables that affect  $p$  are denoted as  $x_1, x_2, \dots, x_k$ , Logistic regression equation can be written as: The method of estimating Logistic regression model and estimating multiple regression model is different of.

$$K_g(x) = xK_g^A + (1-x)K_g^B - ax(1-x) \quad (1)$$

$$\frac{\partial n}{\partial t} = \frac{1}{q} \operatorname{div} \bar{J}_n + G_n - R_n \quad (2)$$

Multivariate regression uses least squares estimation to minimize the sum of squares of the difference between the true value of the explanatory variable and the predicted value. The non-linear special diagnosis of Logistic transformation makes the model adopt the iterative method of maximum likelihood estimation to find the "most likely" estimate of the coefficient. In this way, when calculating the entire model, the likelihood value is used instead of the sum of squared deviations. The graph of the common Logistic transform function is shown in Figure 1. On the one hand, logistic regression is a relatively fast and effective method. First, determine whether the test data is projected to one of the two predefined classes. Second, perform a quick training on the classification of the training set, which represents a series of features and classes. These two points fully show that logical regression is suitable for calculating the attack condition probability of individual independent attributes. In addition, whenever an attack is determined to exist, the newly identified attack features can be used to quickly re-extract Logistic Regression Classifiers (LRCs) for later attack detection. On the other hand, belief propagation first involves the calculation of conditional probabilities, in order to further obtain the confidence of the existence of an attack in the virtual environment.

Confidence propagation is a relatively complete and effective method for attack detection, which can ensure that the obtained confidence accurately reflects the probability contribution of different characteristic attributes. The purpose of data mining is to solve actual business problems. The model established in the mining modeling stage has high performance from an analytical point of view, but whether the model has practical meaning, can solve business problems and is used to guide practice. The model needs to be evaluated before release. Model evaluation should check the various steps of modeling, focusing on checking whether there are some important business issues that have not been fully considered. If the established model cannot meet the business

needs, you must return to the previous stages, reselect the data, and build the model. The evaluation criteria of the current model mainly consider the hit rate, coverage rate and promotion rate of the preference model.

### B. Smart Marketing Data Collection

In order to adapt to the diversity and variability of the environment, companies need to continue to innovate and create more value. For example, by combining the user's location information and search history, Google determines whether the advertisement displayed on the mobile phone during the user's Google search service effectively prompts the customer to visit the store. Companies can use consumer information obtained from big data to provide faster and more accurate decision-making basis for innovation. Amazon uses big data to predict when users will make a purchase and ship the products to the shipping warehouse closest to the consumer in advance. Amazon uses the value brought by big data to recreate the distribution strategy. This radical innovation has enabled the company to save costs, improve customer satisfaction, and further improve the company's business efficiency. Innovation is essential for every company. Therefore, innovative companies must design a new system platform suitable for analyzing big data to realize the value of big data, so as to implement innovative marketing activities. However, any innovative marketing activities will be imitated in a highly competitive business environment. Therefore, enterprises must accelerate the speed of transforming big data into valuable information, which is essential for companies that want to use big data to gain a competitive advantage.

$$Z(g) = \sum_{i=0}^{k-1} \sum_{j=1}^d d_{j,i} \cdot 2^{i \cdot d + j - 1} \quad (3)$$

$$a_i = z_i - (z_{i-1} \times 2^{kd} - 1) \quad (4)$$

There is a wealth of product categories in the online market, and consumers' personalized consumption concepts are also increasing. Online consumption is more convenient and flexible than physical store consumption, and it can better meet the discerning needs of customers. This change in marketing concepts can help companies gain a head start in competition. In marketing, only by reasonably categorizing the sub-markets owned by the enterprise into several large areas can we effectively formulate marketing strategies and tactics that meet the characteristics of the areas and appoint appropriate area managers. For example, a company has 30 sub-markets across the country, and each market has different index values on variables such as population size, per capita disposable income, regional retail sales, and sales of a certain product of the company. Using these variables as clustering variables, combining the subjective wishes of decision makers and the objective criteria provided by relevant statistical software, the 30 sub-markets can be clustered into several large areas. Next, you can formulate reasonable strategies and tactics for different areas, and appoint appropriate area managers. With the advent of the Internet age, a monopoly such as Ford's sale of a single black car is no longer possible to bring high profits.

Consumers in this era only need to use search engines to find personalized products with special specifications and colors that meet their requirements. For example, consumers can select a clothing style on the website of a clothing company, modify it.

### C. Smart Marketing Data Collection Based on Logistic Regression Algorithm

LRCs determine the existence of the attack by analyzing the four attribute characteristics of `in_connect`, `out_connect`, `unKnown_exec` and `port_change`. Since the guest VM that is under attack often establishes an external connection with the remote attacker, which may cause the above attribute characteristics to change, each attribute represented by the node can form a Bayesian network as shown in Figure 1. In graphical models such as Bayesian networks and Markov random fields, belief propagation is used to calculate the random distribution of target nodes.

The Bayesian network in uncertainty theory can provide a representation of the relationship between different features when determining the existence of an attack. Here, if a node  $v$  in the Bayesian network is given, the boundary probability of neighboring nodes can be used to calculate the corresponding state the lower confidence  $BEL(v)$ . In order to face the ever-stronger competitors, Ford Motor uses consumer data analysis to carry out their product changes. Ford obtained information about the owners of approximately 4 million cars through sensors and remote applications. After analyzing the data collected by the car's speech recognition system, Ford realized that noise interference to the driver accounted for the largest proportion of the driver's dislike of car speech. Ford actively rectified it, introducing automatic noise reduction technology and redefining the microphone position.

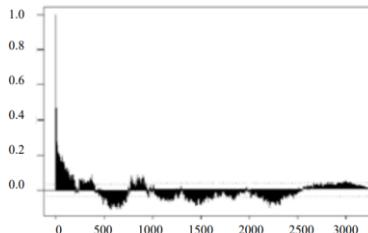


Fig.1. Smart Marketing Data Collection

This innovative move by Ford has significantly increased the proportion of users using the car's voice recognition system, and the degree of satisfaction has also been recognized by consumers. The DSA algorithm first loads all known malicious port numbers and benign port numbers from the distributed Cassandra database, and then uses these two port types to train LRCs. Obviously, the algorithm can determine the probability of malicious attacks on unknown ports before passing the unknown port information to the confidence propagation model. After repeated training, the logistic regression classifier can determine any feature attribute as "malicious attack" or "benign", and finally pass its probability to the final confidence propagation process.

### III. EXPERIMENT

The logistic regression algorithm is shown in the figure.

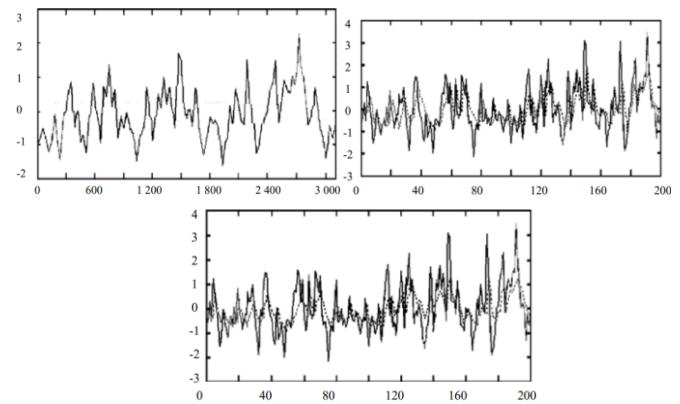


Fig.2. Logistic regression algorithm

Smart marketing data analysis is shown in the figure.

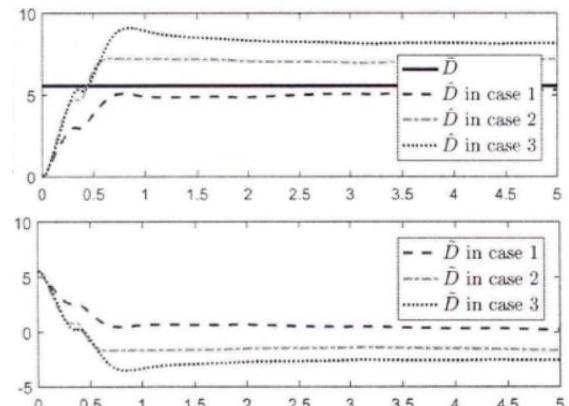


Fig. 3. Smart marketing data analysis

Smart marketing collection based on logistic regression algorithm is shown in the figure.

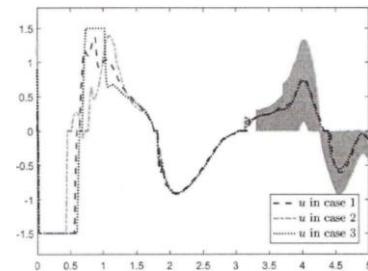


Fig. 4. Smart marketing collection

### IV. CONCLUSION

After the fragmentation of physiology knowledge, if the teaching concept is not changed, therefore, the teaching philosophy of teachers and the learning motivation of students should be effectively changed, and the teaching of physiology knowledge should be organically combined with the content of other basic subjects through collective lesson preparation, trial lectures and pre-teaching measures. Let teachers better integrate physiology knowledge into the organ system teaching model; teaching activities through PBL problem learning, group discussions, independent learning, early clinical internships, hospital volunteers and other forms of teaching activities will help students.

## REFERENCES

- [1] Song Hui. Big Data Consumer Market Analysis and Marketing Transformation[J]. Chinese Market, 2020(17):3.
- [2] Song Hui. Big data consumption market analysis and marketing transformation[J]. 2021(2020-17):199-201.
- [3] Ren Xitao, Wang Jian. Research on Enterprise Marketing Innovation Based on Big Data Collection and Analysis Technology[J]. 2021(2020-31):116-120.
- [4] Guo Zengwen. The application of electric energy collection system in the marketing analysis of power supply enterprises[J]. Digital World, 2018.
- [5] Ye Lizhou. The application and practice of user portraits based on big data analysis in the precision marketing of cigarettes [C]// Constructing a new system to stimulate new vitality-Guangxi Tobacco Society 2019 Excellent Proceedings. 2020.
- [6] Wang Ke. Research on the Marketing Strategy of M Company's Automatic Identification and Data Collection Products [D]. Nanjing University of Science and Technology, 2018.
- [7] Ren Xitao, Wang Jian. Research on Enterprise Marketing Innovation Based on Big Data Collection and Analysis Technology[J]. Chinese Market, 2020(31):5.
- [8] Bi Xiaoming. A sharing system based on agricultural big data:; CN110309478A[P]. 2019.
- [9] Tian Chaohui. Research and Analysis of Automotive Precision Marketing Based on Big Data——Taking the Construction of Audi Automotive User Profiles as an Example[J]. Journal of Fujian University of Technology, 2019, 17(4):7.
- [10] Li Xiaoyi. Research on the application status and trends of artificial intelligence in the field of marketing [J]. China Strategic Emerging Industries, 2018, 000(026):75.
- [11] Ma Yanjun. Discussion on the Marketing Strategy of Enterprises in the Big Data Era[J]. China Business Forum, 2019, 000(007): 90-91.
- [12] Hu Chengli. Discussion on the Marketing Strategy of Enterprises in the Big Data Era[J]. Information Weekly, 2019(9):1.
- [13] Zhang Yuxiang. Design and application of electrical energy information acquisition software [D]. Dalian University of Technology, 2019.
- [14] Yang Wei. Research on Value Marketing Strategy of Data Trend. Beijing University of Posts and Telecommunications, 2019.
- [15] Li Jinliang. Analysis of Caopu Company's Competition Status and Marketing Strategy Research [D]. Tianjin University, 2019.
- [16] Sun Chenguang. Analyze the application of electricity consumption information collection system in electric power marketing[J]. Encyclopedia Forum Electronic Magazine, 2018.
- [17] Xingcheng Cheng. Analysis of Marketing Opportunities and Challenges in the Background of Big Data[J]. Modern Marketing (Late Period), 2018, 000(011):56.
- [18] Qin Xu. Marketing analysis based on big data[J]. Economics, 2019, 2(4).
- [19] Li Zhen. The application of electric energy harvesting system in the marketing of power supply enterprises[J]. Science and Wealth, 2019, 000(016): 30.
- [20] Huang Lin, Wang Kaige. How do commercial banks optimize digital marketing methods in the digital age[J]. Bank of China, 2020(9): 3.
- [21] Chen Yiqing, Yang Lian'an, Gao Bowen, et al. Intelligent analysis of precision marketing of agricultural products based on big data and GIS[J]. Groundwater, 2019, v.41; No.201(06):84-86.
- [22] He Jinzhi, Liang Lisha, Hou Keyu, et al. Design and empirical research on data collection system of product attention in offline footwear stores[J]. Leather Science and Engineering, 2020, 30(1): 4.
- [23] Yang Heng. A marketing information collection device:; CN208351543U[P]. 2019.
- [24] Yu Yang. Analysis on the Marketing Strategy of Enterprises in the Big Data Era[J]. Taxation, 2020(2):1.