The Role of Data Mining Technology in Advertising Marketing

Xinyi Wang*
Nanchang Institute of Science and Technology, Jiangxi, China

*Corresponding author e-mail: wangxinyi@ncpu.edu.cn

Abstract. With the rapid development of computer and other information technology and data storage technology, more and more data are stored. Data mining technology can process the data in advertising and marketing in a timely and effective manner to obtain potential and useful knowledge and information. As my country's market economy system is becoming more and more complete, and science and technology continue to advance, TV advertising has become a major communication tool for economic activities, and the TV advertising industry is booming in our country. Based on the premise of advertising marketing and data mining theory and technology, this article applies data mining technology to all aspects of the advertising marketing process. This article is based on the initial data segmentation of advertising and marketing. First, it analyzes the shortcomings of the commonly used k-means algorithm and fuzzy average algorithm in the clustering effect, and then proposes a fuzzy average algorithm based on the modified membership function and analyzes its effectiveness. And on the basis of this algorithm, analyze its role in advertising and marketing. This thesis takes the basis of advertising marketing and the technology of data mining theory as the premise, and applies data mining technology to all aspects of the advertising marketing process, especially using data mining technology to analyze and study the program ratings and characteristics. The experimental research results show that through the above analysis of data mining technology and data mining advertising and marketing, we clearly realize that advertising and marketing after people's location services are more than just finding the right people at the right time to push the most correct advertising and marketing services. The role of advertising and marketing based on Web data mining technology strengthens the interaction between users and enterprises.

Keywords: Data Mining Technology, Mean Value Algorithm, Web Data Mining Technology, Advertising Marketing Strategy Introduction

1. Algorithm Establishment
Data mining is a modern new type of business information processing technology [1]. Its main function is to extract, transform, analyze and model a large amount of business data in the business database, and extract key data to assist business decision-making [2]. In developed countries and
market economy regions, many companies begin to process business information through data mining on the basis of original information systems in order to establish their own competitive advantages and increase turnover [3]. Advertising marketing based on data mining can usually send marketing materials related to their previous consumption behavior to consumers. There are many methods of data mining, such as classification, association analysis, time series patterns, clustering, deviation analysis and forecasting. They can be applied to different fields and stages of customer-centric enterprise decision analysis and management [4-5].

Through the application of data mining technology and modern information technology, advertising and marketing method owners can provide specific information according to their own business philosophy and business needs, and can display them to customers through websites and some obvious things without causing customer disgust. It also stimulated their curiosity and gave them a detailed understanding [6]. Secondly, companies are required to continuously increase data analysis capabilities on the basis of existing data mining to promote their own rapid development, better adapt to advertisers’ marketing strategies, and achieve win-win goals [7]. Finally, in order to truly realize the effectiveness of data mining technology, it is necessary to strengthen the in-depth understanding of enterprises, enterprises and users, and summarize the data analysis experience in practice. In an enterprise, we should have data mining experts who can be responsible for data analysis and data mining [8].

In the personalized advertising and marketing method based on data mining technology, the main content of personalized marketing includes the first style, which is based on the needs of users to formulate the content they need, acts on the form of web pages they are easiest to receive, and determines information based on the needs of all users Acceptance method and time etc [9]. The existence of personalized service has brought very satisfying results for companies to close the distance with users, cultivate users, increase sales, and retain i users. On the premise that user information is protected, data mining technology will also be used to better serve user attribute data such as users' social data, consumption data, and travel for personalized marketing and improve advertising and marketing efficiency. Data mining can also be used to deliver similar users to the crowd and accurately target the target users of the product. Secondly, the participation and pertinence of online advertising is stronger, especially now that it enters the era of mobile networks, with the popularization of mobile terminals, online advertising has become the mainstream of the times [10].

2. Algorithm Establishment and Decision Tree Algorithm

2.1 Euclidean Distance Analysis Algorithm

Euclidean distance is used in many algorithms as a measure of the distance between two variables. For the relevance of the various decision-making methods of advertising and marketing, the personality of the consumer can be considered as a coordinate point. If the relevance is greater, the more they are Close; on the contrary, the smaller the two correlations, the farther they are.

The correlation formula algorithm based on Euclidean distance is as follows:

\[ \gamma(C_i, C_j) = \frac{1}{1 + \gamma(C_i, C_j)} = \frac{1}{1 + \sqrt{(x_i^1 - x_i^j)^2 + (x_i^2 - x_i^j)^2 + (x_i^k - x_i^j)^2}} \]  

(1)

Through the analysis of related related data, the related data algorithm can be obtained as follows:

\[ \rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \]

(2)

Using the correlation coefficient to express the degree of relevance of the curriculum, the formula can be rewritten as follows:

\[ \gamma(C_i, C_j) = \frac{cov(C_i, C_j)}{\sigma_{C_i} \sigma_{C_j}} = \frac{E[(C_i - \mu_{C_i})(C_j - \mu_{C_j})]}{\sigma_{C_i} \sigma_{C_j}} \]

(3)
2.2 Data Mining Algorithm

Big data objects have the complexity of the data space distribution state, such as the size of the data space distribution, inconsistent distribution, etc. The distribution pattern of data objects of the same shape and different densities, by calculating the distribution density of data objects in the data space, determine the density attraction The point (extreme point) and the density of the data object attract the density of the point, so as to realize the effective aggregation of clusters of different sizes, shapes and densities, so as to realize the effective mining and analysis of large amounts of data.

2.2.1 Dynamic Neighborhood Radius. The reachable distance of dynamic neighborhood radius adaptive density is defined as:

\[ R_A = R \frac{A_i}{A_{i+1}} \] (4)

2.2.2 Data Point Density. The formula can be expressed as:

\[ \text{density} \left( x_i \right) = \sum_{j=1}^{n} \frac{d(x_j, x_i)^2}{2\sigma^2} \] (5)

2.2.3 Density Reachable Distance. It refers to any certain data object \( x \) in the data cluster space, and the distance between data \( R \), a circular area with certain data as the center and the data distance as the radius, corresponding to the reachable density distance field of the data object:

\[ R = \text{coeff} \times \text{mean} \left( D \right) \] (6)

2.3 Decision Tree Algorithm

Decision tree is composed of directed edges and nodes, and it uses tree structure to make decision analysis. Each node type of a decision tree usually includes a root node, a number of intermediate nodes, and a number of leaf nodes. In the classification problem, the decision tree represents the process of classifying instances based on attributes, including the learning process and the classification process, that is, first use the training sample set, create a decision tree model based on the principle of minimizing the loss function, and summarize the classification rules, and then use the decision tree model to classify the new data set.

3. Model Establishment

3.1 Parameter Estimation of Logistic Regression Model

Logistic regression model is a nonlinear regression model, and the maximum similarity method can be used to estimate the parameters of the nonlinear model. Since the response variable \( Y \) obeys the binomial distribution:

\[ P(Y_i|X_i) = \pi(X_i)^y[1 - \pi(X_i)]^{1-x} \] (7)

Here \( X_i=[x_1, x_2,\ldots,x_n], i=1,2,n \), then the maximum likelihood function of the observed values of \( n \) samples is:

\[ L(\beta|X,Y) = \prod_{i=1}^{n} p(Y_i|X_i; \beta) \]

\[ = \prod_{i=1}^{n} [\pi(X_i)]^y[1 - \pi(X_i)]^{1-y} \] (8)

Log likelihood function:

\[ L(\beta) = \ln \left( L(\beta|X,Y) \right) \]

\[ = \sum_{i=1}^{n} \{y_i \ln[\pi(X_i)] + (1 - y_i) \ln[1 - \pi(X_i)]\} \] (9)

When using the Logistic regression algorithm to build a customer churn prediction model, you first need to start preprocessing the original data set, divide the data set into a training set and a test set.
according to an appropriate ratio, and determine the independent variables used for modeling (i.e. the customer's Basic information attributes, consumption information attributes, etc.) and dependent variables (that is, the customer's churn status, churn is recorded as 0, non-churn is recorded as 1). Second, use the training set to get the specific structure and regression coefficients of the regression model.

For each class J, recalculating the centroid:

$$U_j = \frac{\sum_{i=1}^{m} 1\{C^{(i)} = j\} X^{(i)}}{\sum_{i=1}^{m} 1\{C^{(i)} = j\}}$$  \hspace{1cm} (10)

Euclidean distance is the most commonly used calculation method for vector distance in mathematics. The mathematical formula of Euclidean distance in two-dimensional space is:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$  \hspace{1cm} (11)

Among them (x1, y1) and (x2, y2) are two points in space.

Manhattan distance is also called city block distance. In a two-dimensional space, the Manhattan distance formula between two points is:

$$d = |x_1 - x_2| + |y_1 - y_2|$$  \hspace{1cm} (12)

3.2 Research Object Selection

The data source used in the experiment of this thesis is the final results of the main subjects of the eight semesters of the university of Finance and Economics, a total of 150 students in two majors of 2018 and 2020. The major is: Advertising Application

Before this data mining experiment, the use of data sets for testing algorithms is a very important content, and different algorithms also need different data sets to provide support for them, so that better classification accuracy can be obtained. In order to obtain more effective data, the downloaded data must be preprocessed first. Appropriate data preprocessing methods can greatly improve the accuracy of data mining. The data object of this experimental research is the score data of each course of the 2018 majors of Advertising Applied Studies in the University of Science and Technology. The course scores are stored in the form of EXCEL tables.

4. Conclusion

4.1 Visual Application Analysis of Data Mining Algorithms

As shown in Figure 1, according to the analysis of the visualization application parameter curve based on the optimized data mining algorithm, through the application of data mining technology in advertising marketing, the interactive nature of the online advertising in the website can become stronger, more vivid and rich Creativity can also effectively save and control costs, and can accurately count the actual effects of advertising. This can not only guarantee the interests of advertising owners, but also make the website and the owners form a better cooperative relationship. However, there are
also some problems in the process of its operation, which cause certain problems and obstacles to the healthy development of advertising. First of all, in the application of data mining technology, the relevant technical staff’s understanding of data mining is not comprehensive enough. Some technical personnel believe that data mining has no practical significance for the development of online advertising, and its calculation results are often insufficient. Accurate and reliable. There are also some technical personnel who believe that by applying data mining technology for advertising and marketing, all information and knowledge can be quickly obtained. These two kinds of understandings are relatively extreme, and data technology mining requires a lot of time to operate and implement, and will incur a lot of costs. If you can't consider the actual needs, blindly using data mining will also cause a certain economic burden on advertisers.

Figure 2. Inflection point analysis of advertising marketing strategy based on Euclid algorithm

According to the data in the inflection point diagram of the Euclidean algorithm to predict the advertising marketing strategy according to Figure 2, in the early stage of the current algorithm construction, the algorithm is not stable in the prediction of the marketing quota of online learner advertising, but according to the above seven In the second experiment, we respectively compared the prediction results of each experiment with the actual results. The purpose is to improve the accuracy of the prediction model for the current high prediction data and make it more accurate.

4.2 Application Analysis of Data Mining Technology in Advertising Marketing

| Table 1. Three algorithms predict the classification results of the learners of 18 advertising marketing class 1 |
|---|---|---|
| 18 Advertising and Marketing Class 1 | Euclid | Dynamic neighborhood radius | Decision tree combination |
| The accuracy of the training data set | 71.37% | 86.79% | 93.19% |
| Calibration accuracy | 81% | 96.87% | 93.27% |
| Standard error | 0.1521 | 0.0008 | 0.0011 |

| Table 2. Three algorithms predict the results of class 2 learners in 20 advertising marketing |
|---|---|---|
| 20 Advertising and Marketing Class 2 | Euclid | Dynamic neighborhood radius | Decision tree combination |
| The accuracy of the training data set | 72.17% | 82.39% | 92.17% |
| Calibration accuracy | 86.23% | 92.31% | 96.29% |
| Standard error | 0.1321 | 0.1915 | 0.3723 |

According to the data described in Table 1, Table 2, the results of predictive analysis of the two classes of 2018 and 2020 learners are that according to the data of all the above students, after
different data preprocessing operations, Euclid’s The algorithm has the lowest accuracy rate. When applying LIBSVM, the data is processed by SMOTE and the optimal parameters are calculated at the same time. The results are more satisfactory but not stable, and the linear regression combination is better in predicting the final academic performance of online learners. Forecast accuracy rate and more stable.

The current relatively novel classification algorithm is the most commonly used method in data mining, which is similar to the discrimination mentioned in the classic diversification statistics. In classification problems, the dependent variable is generally a categorical variable. If the classification problem to be solved does not meet this condition, the continuous variable needs to be discretized into a categorical variable. Under normal circumstances, discriminant analysis can directly solve common classification problems, but if the independent variables contain more categorical variables, then discriminant analysis is no longer applicable. We can try some methods in data mining to solve this classification problem.

5. Conclusion
In summary, if you want to obtain better online advertising effects only through data mining technology, it is still impossible to achieve at the current technical level. It is also necessary for websites and enterprises to continuously strengthen the development and development of data mining technology. According to the needs of enterprises and users, it should be designed reasonably. At the same time, the cooperation between various departments of the enterprise should be strengthened to further improve the quality of online advertising and marketing. Data mining technology, as a deeper analysis and processing of data on the basis of database technology, can reveal undiscovered potential relationships hidden in a large amount of data and information useful to decision makers, so as to provide users with good decision-making assistance and support. From the emergence of data mining technology to the present, more and more companies and enterprises have established their own data warehouses, and obtained many supportable companies through the analysis and mining of a large amount of previous historical information. This is so far for advertising planning in the context of the Internet. The relevant research and analysis of the innovative development strategy of the company. From the content described in the article, it can be found that there are actually more factors affecting the innovative development of the advertising planning work network. Therefore, in the actual work process, relevant Advertising planners need to pay attention to the influence of different influencing factors on their innovation and development, and carry out continuous optimization and perfection work, so as to achieve the improvement of the quality of advertising planning work.

References
[1] Li T, Long L. Imaging Examination and Quantitative Detection and Analysis of Gastrointestinal Diseases Based on Data Mining Technology. Journal of Medical Systems, 2020, 44(1):1-15.
[2] Zhao-Yi W A N G , Zheng-De H U A N G , Ping Y A N G , et al. Regularity of Wind-dispelling Medication Prescribed by LI Dong-Yuan: A Data Mining Technology-based Study. Digital Chinese Medicine, 2020, 3(1):20-33.
[3] Li T, Long L. Imaging Examination and Quantitative Detection and Analysis of Gastrointestinal Diseases Based on Data Mining Technology. Journal of Medical Systems, 2020, 44(1):1-15.
[4] Sobh R, Singh N, Chun W, et al. Localizing to Arabic consumers: Insights from print advertising. Journal of marketing communications, 2018, 24(2):190-211.
[5] Ewins N. Variations in American marketing practices of early 19th-century ceramic importers and dealers, reflecting culture and identity. Journal of Historical Research in Marketing, 2020, 12(3):345-375.
[6] Sung B, Phau I . When pride meets envy: Is social superiority portrayal in luxury advertising perceived as prestige or arrogance?. Psychology & Marketing, 2019, 36(2):113-119.
[7] Wies S, Hoffmann A O I , Aspara J , et al. Can Advertising Investments Counter the Negative Impact of Shareholder Complaints on Firm Value?. Journal of Marketing, 2019, 83(4):58-80.
[8] Lin H Y, Yang S Y. A Smart Cloud-Based Energy Data Mining Agent Using Big Data Analysis Technology. Smart ence, 2019, 7(JUN.):1-9.

[9] Trappey A J C, Trappey C V, Fan C Y, et al. Consumer driven product technology function deployment using social media and patent mining. Advanced Engineering Informatics, 2018, 36(APR.):120-129.

[10] Jin H, Miao Y, Jung J R, et al. Construction of information search behavior based on data mining. Personal and Ubiquitous Computing, 2019(4):1-13.