Customer Value Classification Model Based on Improved TOPSIS Method and BP-Adaboost Algorithm

Huajian Liu [1, #, *], Hongrui Zhu [2, #], Boyi Yang [1, #]

[1] School of Statistics and Data Science, Jiangxi University of Finance and Economics, Nanchang, China, 330013

[2] School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China, 430073

* Corresponding Author Email: 13389656280@163.com

[#]These authors contributed equally.

Abstract. Identifying customer groups with different consumption values can help enterprises understand and analyze customers' purchasing behaviors and preferences, so as to effectively improve their operating profits. In this paper, the customer value evaluation index system is constructed on the basis of existing research on customer value classification using entropy weight method and improved TOPSIS method, and different grades are classified according to the comprehensive score of customers. Subsequently, the customer comprehensive score is used as the output layer, and the BP-Adaboost algorithm is used to construct the customer value classification model. The empirical results show that the improved TOPSIS method is more advantageous in terms of error control and classification accuracy, and the BP-Adaboost algorithm outperforms the traditional BP neural network in the customer value classification model, which improves the classification precision and accuracy. This study provides a powerful tool for dealing with the customer value classification problem, which helps enterprises make more accurate decisions in customer relationship management and resource optimization allocation.

Keywords: Customer value classification, Improved TOPSIS, BP-Adaboost algorithm, BP neural network.

1. Introduction

With the development of market economy and information technology, organizations face the challenge of managing many customers [1]. Customers are usually not loyal to one brand, so companies need to adopt differentiation strategies to manage different customer groups [2]. Customer segmentation helps firms to gain a deeper understanding of their customers' buying behaviors and preferences by dividing the market into small groups with specific attributes through social, behavioral, and consumer characteristics [3]. Reviewing the existing research on customer value segmentation, customer value segmentation focuses on exploring customer-oriented segmentation methods [4], which include four main types: demographic segmentation, lifestyle segmentation, behavioral segmentation, and benefit segmentation.

Customer behavioral value classification focuses on predicting the future behavioral patterns of customers by analyzing their historical behavior [5]. This classification method assumes that there is a correlation between the customer's historical behavior and future behavior, so that possible future customer behavior can be predicted by examining the customer's past behavior [6]. The RFM model was first proposed by Hughes in 1994 to describe and differentiate customers [5]. RFM analysis is a customer segmentation tool that scores each customer by evaluating three key metrics: recent purchasing behavior, frequency of purchase, and purchase amount. These scores are then multiplied to form a composite score that is used to rank all customers [7]. RFM is now widely used for customer value categorization and has been expanded by different scholars based on their research. Yu Yue and Li Leiming, on the basis of the traditional RFM model, used the average single consumption amount to replace M in the original model, and added the new variable T to determine the customer's consumption trend and the degree of activity, and ultimately realized the accurate identification of the consumption characteristics of gas station customers and supporting price strategy through the hierarchical analysis method (AHP) [8]; Zhuo Ling and Sun Xin proposed the RVS model, with the number of times the customer initiates a call and represents V, and the average call market of customers represents S, and combined with the K-Means clustering method to realize the accurate classification of digital cluster users [9]; Chen Danhong et al. proposed the RFMC model, adding the function of crowdsourcing, and through the AHP analysis method, the accurate identification of gas station customers' consumption characteristics and the supporting pricing strategy were finally realized [10]. The RFMC model was proposed, crowdsourcing customer credit C was added to the model, and the superiority of traditional RFM model and RFMC model in crowdsourcing customer classification was compared [10].

Customer benefit-value categorization is a market identification method first proposed by Haley in 1963. This approach is based on factors that have a causal relationship rather than merely descriptive characteristics. The advantage of interest segmentation is that it goes beyond the surface behaviors, attitudes, and motivations of customers and digs deeper into the real interests behind the customers, thus identifying the market more accurately [11]. BP neural network is widely used in the field of artificial intelligence because of its powerful nonlinear approximation ability. It is trained by back-propagation algorithm to optimize the network weights. Adaboost algorithm improves the classification accuracy by combining multiple weak classifiers to form a strong classifier. The BP-Adaboost algorithm combines the advantages of both, enhances the generalization ability, improves the training accuracy, and solves the problem of decreasing classification speed [12]. At present, many scholars have applied the BP-Adaboost algorithm to customer value classification: Yan Chun and Zhang Xinyu combined the improved K-Means algorithm and BP-Adaboost to effectively solve the problem of life insurance customer classification and churn prediction [13]; Gu Yulei et al. aimed at the shortcomings of BP-Adaboost algorithm that could not accurately calculate the comprehensive score of the supplier to design the supplier evaluation method based on the TOPSIS method of supplier evaluation model [14]; Zhu Jiran et al. proposed a low-voltage power user value evaluation method based on AHP and BP-Adaboost algorithm, and utilized two parameters of the coefficient of determination and accuracy for verification [15].

In summary, although the existing literature has carried out useful explorations of customer value classification in multiple dimensions, the following problems still exist: First, although the RFM model for customer behavioral value classification has innovations in quantitative dimensions, such as RFMC, the weight calculation of quantitative indicators still adopts the hierarchical analysis method or the subjective weight method for calculating the weights; second, the BP-Adaboost algorithm in customer benefit value classification cannot accurately calculate the comprehensive score, and the traditional TOPSIS method is still defective. Third, there are few literatures comparing the advantages and disadvantages of BP and BP-Adaboost in the classification of customer value.

Based on this, this paper first proposes a customer value index system based on entropy weight method and improved TOPSIS method and divides the customer value of different star ratings according to the results of the comprehensive score. Then, combining the nonlinear fitting ability of BP neural network and the advantage of Adaboost algorithm in improving the generalization ability of the model, it is proposed to use the algorithm of fusion BP-AdaBoost to construct a strong predictor of customer value, and comprehensively compare the effect of before and after the improvement of the TOPSIS method, as well as the effect of a single predictor (BP) and the fusion predictor (BP-Adaboost). Based on the predicted comprehensive score, the customer value class is divided, which guides the enterprise to carry out product targeting promotion and service risk control; it also helps the enterprise to identify the characteristics of different classes of customers and formulate differentiated service strategies. The research framework diagram of customer value comprehensive evaluation is shown in Figure 1.

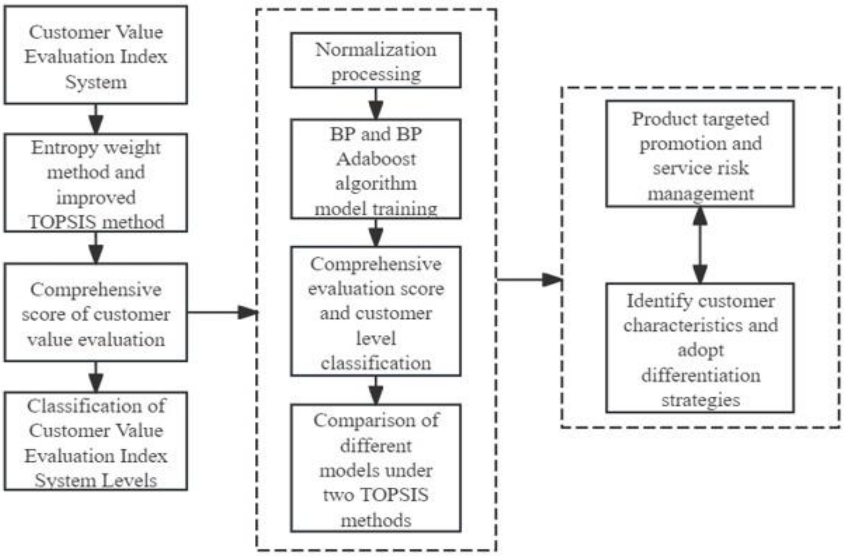

Figure 1. Research framework for comprehensive customer value assessment.

2. Customer value evaluation index system based on entropy weight method and improved TOPSIS method

2.1. Introduction to the measurement methodology

Entropy value method is an objective assignment method based on data discretization, the more information, the higher the weight. TOPSIS method ranks the advantages and disadvantages by comparing the distance between the observed object and the ideal solution. This paper combines the entropy value method and the improved TOPSIS method to avoid subjective weight judgment and improve the objectivity and reliability of decision-making results.

2.2. Customer value evaluation index system

This study is based on the traditional RFM model of purchase frequency (Frequency) and purchase amount (Monetary) to expand, and based on the research of Xuxuefeng et al [16] from the four dimensions of consumption ability, visit frequency, marketing acceptance, consumer loyalty to study the customer value identification, the specific index system is constructed as shown in Table 1 below.

Table 1. Client value evaluation index system.

Tier 1 indicators	Tier 2 indicators	Tier 3 indicators
Customer value evaluation index system	Spending power	Annual income
		Spending on consumer activities
		Click-to-content ratio
	Visiting Frequency	Click-to-purchase conversion rate
		Number of website visits
		Average number of pages visited per session
		Average time spent on website per visit (minutes)
	Marketing Acceptance	Number of times content is shared on social media
		Number of times an email is opened
		Number of times links in emails were clicked
	Consumption Loyalty	Number of times a customer has made a previous purchase
		Accumulation of loyalty points by customers

2.3. Improved TOPSIS method for synthesizing evaluations

The traditional TOPSIS method is a common method of comprehensive evaluation within a group, which can effectively obtain information from the original material, and is able to evaluate the relative advantages and disadvantages among the available objects based on the method of ranking a finite number of evaluation objects according to their proximity to the idealized goal, if the evaluation object is the closest to the positive ideal solution (the optimal solution) and the farthest from the negative ideal solution (the worst solution), then this evaluation object is Optimal.

Although the traditional TOPSIS method has been widely used in several fields, we find that it suffers from certain inconsistencies in differentiation when ranking customer values. Specifically, when using the Euclidean distance as the evaluation criterion, there may be situations where the distance of an individual from a positive or negative ideal solution increases or decreases at the same time, which makes it difficult to differentiate the merits of different individuals [17]. In order to solve this problem, this paper proposes an improved method, i.e., the Manhattan distance is used to replace the traditional Euclidean distance, and the specific steps are described:

Let the positive ideal solution be $J^+ = (a_1^+, a_2^+, \dots, a_n^+)$, the negative ideal solution is $J^- = (a_1^-, a_2^-, \dots, a_{n)}^-$, the attribute value of target A_i is$(a_{i1}a_{i2}\dots,a_{in})$. The distances of the objective A_i to the positive and negative ideal solutions are respectively:

$$\mathrm{D}_i^+ = \sum_{j=1}^{n} \left|a_j^+ - a_{ij}\right| \tag{1}$$

$$\mathrm{D}_i^- = \sum_{j=1}^{n} \left|a_{ij} - a_j^-\right| \tag{2}$$

In this method, the sum of the distances between each goal and the positive and negative ideal solutions remains constant. This means that if the distance of a target from a positive ideal solution is shortened, then its distance from a negative ideal solution is increased accordingly. In this way, we have solved the problem of inconsistent differentiation that we previously encountered when using Euclidean distances.

2.4. Classification of customer value evaluation index system hierarchy

Through the combined calculation of the entropy weight method and the improved TOPSIS method mentioned above, we get the comprehensive score of customer value. Next, according to the isometric grouping criteria in statistics we categorize the customer value class into five-star customers, four-star customers, three-star customers, two-star customers, and one-star customers. Five class divisions are defined as shown in Table 2 below.

Table 2. Definition of classifications.

Grade	Definition of grading
Five-star customer	Composite score of 85 or above
Four-star customer	Composite score between 70 and 85
Three-star customer	Composite score between 55 and 70
Two-star customer	Combined score between 40 and 55
One-Star Customer	A composite score of less than 40

3. Customer value classification model based on BP-Adaboost

3.1. Principles of the underlying algorithms

A BP neural network is a supervised predictive model that learns by matching a series of input data with the corresponding outputs. This model exhibits strong fitting ability in dealing with nonlinear problems and can theoretically accommodate any nonlinear function [13]. However, BP neural networks may encounter the problem of local optimization during training, which can reduce the accuracy of prediction [15]. Meanwhile, the AdaBoost algorithm continuously reduces the prediction error through an iterative process, which enhances the generalization performance of the model. Therefore, this paper proposes an algorithm that combines BP neural networks and AdaBoost by training multiple predictors and adjusting the weights of the samples with lower prediction accuracy while decreasing the weights of the samples with higher prediction accuracy in order to improve the overall prediction accuracy.

3.2. Customer Value Classification Modeling Process

In this paper, BP-Adaboost algorithm is used to evaluate the customer value score, the specific steps are shown in Figure 2 below:

First, a training set is composed of samples from the customer value evaluation index system $X = \{x_1, x_2, \dots, x_m\}$, according to the entropy weight method and the improved TOPSIS method comprehensive evaluation of the customer value comprehensive score, so as to classify the customer value class. Each classification sample in the training set X is given a separate weight, and the initial weights are equal.

$$D_1 = \{d_{11}, d_{12}, \dots, d_{1m}\} \tag{3}$$

Next, the Nth BP neural network weak predictor I_N is trained with samples to make predictions and based on its results $G_N = \{y_1', y_2', \dots y_m'\}$ with expectation Y to derive the classification error samples, and calculate the classification error rate of the Nth classifier from the classification error sample weights.

$$e_N = \sum D_N(l), l = 1,2, \dots, m\big(G_N(l) \neq Y(l)\big) \tag{4}$$

Next, the weight coefficient W_N of this classifier is calculated based on the classification error of I_N, and the training set sample weight D_{N+1} of the next classifier I_{N+1} is adjusted based on the classification error rate e_N. And B_N is the sample weight normalization factor.

$$W_N = \frac{1}{2} ln\left(\frac{1-e_N}{e_N}\right) \tag{5}$$

$$\mathrm{D_{N+1}(i+1)} = \frac{D_N(i)}{B_N} e^{[-W_N Y(i) G_N(i)]} \tag{6}$$

$$B_N = \sum_{i=1}^{m} D_N(i) e^{[-W_N Y(i) G_N(i)]} \tag{7}$$

Finally, steps two and three are continued until the training of the N weak classifiers is completed. Use the trained and completed N weak classifiers to identify the samples of the customer value evaluation index system, combine the results of the N classifiers for back-normalization, and derive the comprehensive customer value score f. Classify the customer value class according to the score f.

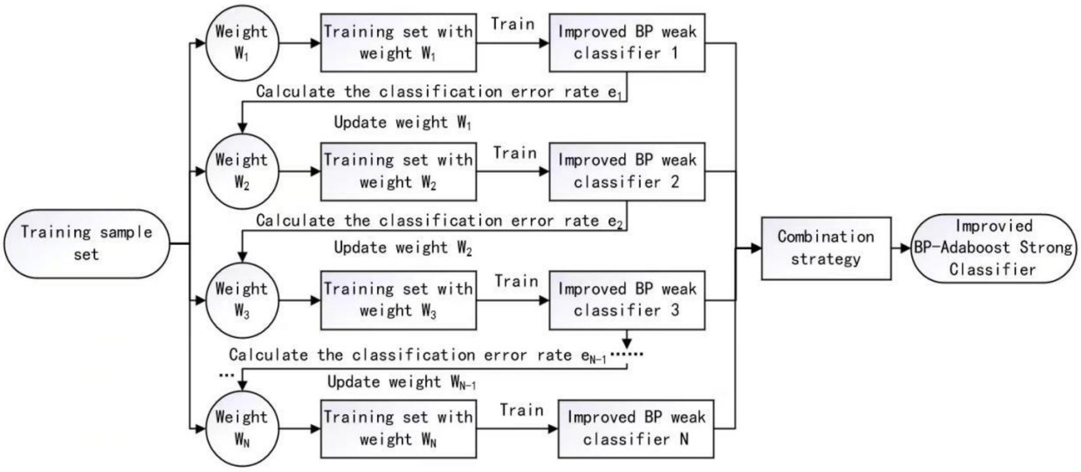

Figure 2. Model diagram of BP-Adaboost algorithm.

4. Empirical studies

4.1. Data sources and processing

The experimental dataset for this study is derived from the AliCloud Tianchi database and the Kaggle database contains 8,000 exhaustive records, each detailing the customer's annual revenue, campaign spend, marketing content click-through rate, and a host of other key features. In order to ensure that features of each numerical type are fairly compared in the analysis process, independent of differences in their original scales, a normalization measure was used to standardize and unify the data by normalizing all relevant numerical features into the interval 0 to 1.

4.2. Comparison of experimental results of predictors BP and BP-Adaboost

In the experiments of this paper, we classify customer value by constructing and training BP neural network and BP-Adaboost strong predictor. In order to comprehensively evaluate the overall performance of the model, we use three parameters, mean absolute error MAE, mean square error MSE and accuracy rate, as the indexes for evaluating the model.

The BP-Adaboost algorithm integrates 10 BP neural network weak classifiers with annual income, spending on consumer activities, clicked content ratio, purchase conversion rate, number of visits to the website, average number of visits to the session page, length of visit, number of social media shares, email open rate and number of clicks on the link as input layers, and evaluates the customer value as the output layer through the entropy weighting-improved TOPSIS method. The sample weights are dynamically adjusted according to the prediction results and finally fused into a strong predictor. In the experiment of 8000 samples, 7000 are used for training and 1000 are used for testing, as can be seen in Fig. 3, the prediction error of BP-Adaboost strong predictor on the test data is generally lower than that of the weak predictor except for a very small number of them, and the absolute value of the error is even lower, and its MAE is only 0.0383, which is much lower than that of a single BP neural network, 0.1602, which is a strong proof that the superiority of BP-Adaboost in customer value classification.

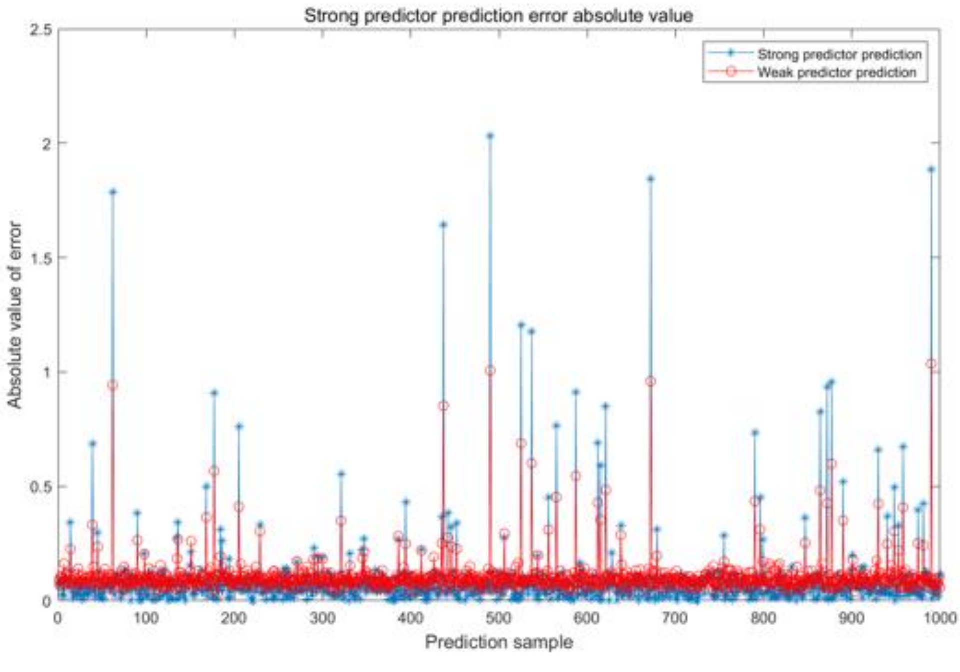

Figure 3. Comparison of experimental results for predictors BP and BP-Adaboost.

4.3. Comparison of all experimental results and conclusions

By comparing the two TOPSIS methods with BP neural network and BP-Adaboost algorithm, we get the results shown in Figure 4 below.

By comparing the experimental results, the improved TOPSIS method shows better classification performance than the traditional TOPSIS method when combined with both BP neural network and BP-Adaboost algorithm. Specifically, the mean square error (MSE) of the improved TOPSIS method under the BP-Adaboost algorithm is 0.0085, which is much lower than that of the traditional TOPSIS method of 0.30688, indicating its advantage in error control. In terms of classification accuracy and mean absolute error (MAE), the improved TOPSIS method combined with the BP-Adaboost algorithm reached 0.998 and 0.0383, while the traditional TOPSIS method combined with the BP-Adaboost algorithm was 0.969 and 0.3072, respectively, which further confirmed the superiority of the improved method.

Meanwhile, the BP-Adaboost algorithm outperforms the BP neural network in terms of classification effect regardless of which TOPSIS method is used. The MSE of the improved TOPSIS method combined with the BP-Adaboost algorithm is 0.0085, which is lower than 0.01400 when combined with the BP neural network, showing the superiority of BP-Adaboost in error control. The BP-Adaboost algorithm is also significantly better than the BP neural network in terms of classification accuracy and MAE, and the classification accuracy of the improved TOPSIS method combined with the BP-Adaboost algorithm is 0.998, which is much higher than that of the BP neural network at 0.986; the MAE is 0.0383, which is much lower than that of the BP neural network at 0.1602, which proves the superiority of the BP- Adaboost algorithm's superiority in customer value classification task.

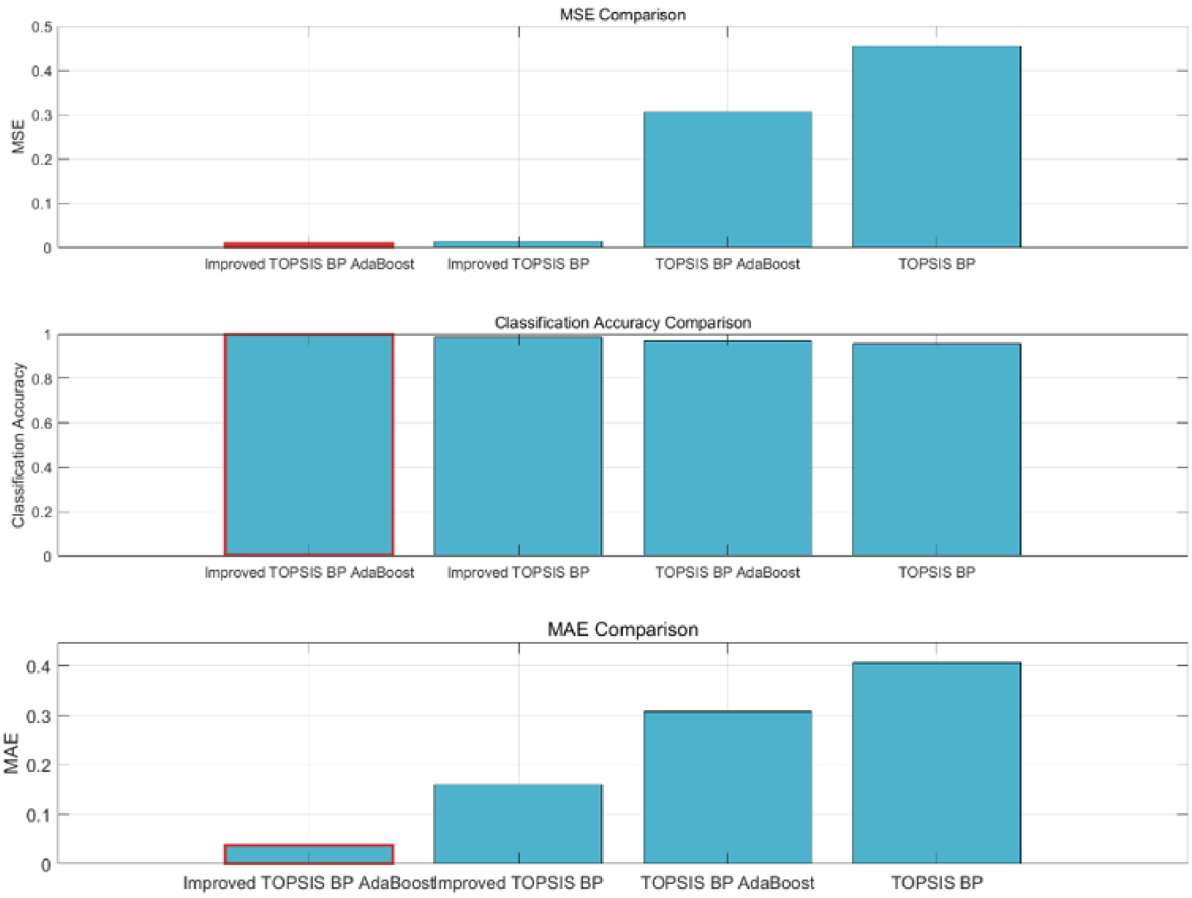

Figure 4. Comparison chart of experimental results of customer value classification.

5. Conclusions

In summary, the combination of the improved TOPSIS method and the BP-Adaboost algorithm shows the most excellent performance in customer value classification, which is manifested by the lowest MSE, the highest classification accuracy and the smallest MAE. This combination not only improves the classification accuracy, but also effectively reduces the error, which has a very high practical value and potential for promotion. The innovation of the improved TOPSIS method in weight assignment and comprehensive score calculation complements the powerful generalization ability of the BP-Adaboost algorithm, and together they provide a powerful tool for dealing with the customer value classification problem, which helps enterprises make more accurate decisions in customer relationship management and resource optimization allocation.

Customer value classification and customer relationship management aim to identify and cultivate high-value customers while attracting potential customers, and accurate customer value classification is the key to realize efficient marketing. The customer value classification model constructed based on the improved TOPSIS method and BP-Adaboost algorithm in this paper has a higher error control rate and prediction accuracy than the traditional TOPSIS and BP neural network in the empirical study, and with the experimental results of this paper as a reference, we put forward the following suggestions for customer value classification:

Optimize customer classification strategy, identify different value customers through refined classification, and dynamically adjust the criteria to maintain accuracy and timeliness. Implement differentiated marketing strategies: provide high-value customers with personalized experiences and exclusive services to enhance brand loyalty; customize cost-effective products and services for medium-value customers to meet the demand for quality and value; and for low-value customers, enhance brand recognition and market share through precision marketing.

Enhance customer service quality, improve service processes and establish a multi-channel service system. Utilize customer relationship management information system and big data analysis to deeply explore customer value, and monitor and adjust marketing strategy and service mode in real time.

Regularly evaluate and optimize customer management and marketing processes to improve operational efficiency. Introduce advanced technologies to improve the accuracy of customer value classification and prediction. Identify and personalize the marketing of potential high-value customers by analyzing behavioral data to improve conversion rates and customer value.

This study constructs a customer value evaluation index system from customer behavior and interest value classification, and selects four dimensions of consumption ability, visit frequency, marketing acceptance, and consumption loyalty for customer value identification, which can be regarded as a partial study of customer value. In the future, other dimensions of customer value classification can be considered comprehensively and the customer value evaluation index system can be established from more aspects.

References

[1] Zong Y, Xing H. Customer stratification theory and value evaluation—analysis based on improved RFM model [J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(3): 4155-4167.

[2] Li Y, Chu X, Tian D, et al. Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm [J]. Applied Soft Computing, 2021, 113: 107924.

[3] Rahim M A, Mushafiq M, Khan S, et al. RFM-based repurchase behavior for customer classification and segmentation[J]. Journal of Retailing and Consumer Services, 2021, 61: 102566.

[4] Anitha P, Patil M M. RFM model for customer purchase behavior using K-Means algorithm [J]. Journal of King Saud University - Computer and Information Sciences, 2022, 34(5): 1785-1792.

[5] Liu Yingzi, Wu Hao A review of research on customer segmentation methods [J] Journal of Management Engineering, 2006 (1): 53-57.

[6] Smaili M Y, Hachimi H. New RFM-D classification model for improving customer analysis and response prediction [J]. Ain Shams Engineering Journal, 2023, 14(12): 102254.

[7] Yu Hui, Liao Xiaohong Overview of Customer Segmentation Methods [J] Management and Technology of Small and Medium sized Enterprises (Next Issue), 2014 (11): 17-18.

[8] Yu Yue, Li Leiming Customer Value Classification and Precision Marketing Strategies for Gas Stations: Pricing Model Analysis Based on Improved RFAT Model [J] Price Theory and Practice, 2018 (11): 158-161.

[9] Zhuo Ling, Sun Xin A digital cluster user classification method based on an improved RFM model [J] Computer Application Research, 2020, 37 (9): 2822-2826.

[10] Chen Danhong, Peng Zhanglin, Wan Dequan, etc User value identification and segmentation of crowdsourcing platforms: based on an improved RFM model [J] Computer Science, 2022, 49 (4): 37-42

[11] Xu Didi Literature Review on Customer Segmentation Theory: Segmentation Dimensions and Their Applications in the Banking Industry [J] Economist, 2015 (9): 176-179.

[12] Zhu Chenfei, Huang Shuhua, Wang Huaicong, etc An Improved BP AdaBoost Algorithm and Its Application Research [J] Modern Electronic Technology, 2019, 42 (19): 64-67+72.

[13] Yan Chun, Zhang Xinyu Research on Life Insurance Customer Loss Prediction Algorithm Based on Improved K-means and BP Adaboost [J] Journal of Shandong University of Science and Technology (Natural Science Edition), 2022, 41 (1): 54-65.

[14] Gu Yulei, Ma Hui, Wang Yuqin, etc Evaluation of Aviation Equipment Suppliers Based on BP Adaboost and TOPSIS [J] Journal of Shandong University (Engineering Edition), 2024, 54 (1): 63-73.

[15] Zhu Jiran, Zhang Di, Zhang Zhidan, etc Low voltage power user value evaluation method based on AHP and BP Adaboost [J] Journal of Electric Power Science and Technology, 2022, 37 (5): 155-163.

[16] Xu Xuefeng, Xing De'en, Zong Xuanjun Research on Customer Value Segmentation Model Based on PCA and SOM Neural Network Algorithms [J] Electrical Automation, 2017, 39 (3): 49-52+56.

[17] Hao Qikai, Xun Panpan, E Chaoran, etc Air combat threat assessment based on coefficient of variation method and improved TOPSIS method [J] Journal of Artillery Launch and Control, 2024: 1-10.