Customer Classification Method of Logistics Enterprises Based on BP-AdaBoost

Xianming Shi, Guangning Li, Kang Li*, Junyan Liu and Rongcai Wang
Shijiazhuang Campus, Army Engineering University, Shijiazhuang, China

*Corresponding author email: likangsiz@126.com

Abstract: Effective customer classification is a key issue for logistics enterprises to carry out customer relationship management, which has been concerned and studied by relevant institutions. Based on BP and AdaBoost algorithm, a new customer classification method for logistics enterprises is proposed. First of all, the logistics customer value evaluation index system is designed, and the different customer characteristics are analysed. Secondly, in order to solve the problem of low accuracy of the traditional BP network, the AdaBoost algorithm is used to linearly combine the traditional BP network to produce a strong classifier. The simulation results show that the application of the new customer value classification method based on BP-AdaBoost in customer classification improves the classification accuracy.

Keywords: Customer classification, customer value, customer relationship management, AdaBoost algorithm, BP network

1. Introduction
With the rapid development of economy, especially the wide application of e-commerce in the business model, enterprise competition has gradually shifted from focusing on products and services to focusing on customers[1]. Logistics enterprises at home and abroad have gradually realized the importance of customer relationship management (CRM: Customer Relationship Management)[2]. CRM system will bring huge benefits in the overall customer service experience[3]. The good products and high-quality services that logistics customers need put forward higher requirements for logistics enterprises. In this form, logistics enterprises to maintain a good customer relationship, to provide high-quality customer service is the key for logistics enterprises to achieve competitive advantage[4]. The problem solved by CRM is to classify customers and select which customers belong to long-term customers, which belong to short-term customers, which belong to developable customers, and which belong to abandoned customers. In the case of limited enterprise resources, treat different customers in an appropriate way, so as to maximize the customer value and further improve the competitive advantage of the enterprise. Therefore, effective and reasonable customer classification can enable logistics enterprises to formulate corresponding strategies, rational allocation of enterprise resources, and maximize the interests of enterprises. The research of customer classification has become a more and more concerned topic, and it is one of the urgent problems to be solved by logistics enterprises[5].

At present, there has been some development in the research of customer classification, such as neural network method, decision tree classification method, fuzzy clustering method and so on. Ching-HsueCheng[6] proposed customer value segmentation based on RFM model and RS theory, Zhang Bin[7] proposed customer segmentation method based on KFAV, Li Yuan[8] proposed CRM system research based on decision tree classification algorithm, He Peng[9] proposed customer value classification method based on system clustering, Chen Qianshu[10] used Analytic hierarchy process to
study customer value based on RFM model. It can be seen from these studies that the customer classification method based on neural network is widely used and is more suitable for learning, classification and prediction, but the accuracy of customer classification using neural network method alone is not high and the speed is slow. For this reason, this paper proposes a customer value classification method of logistics enterprises based on BP-AdaBoost to improve the classification accuracy.

2. Design of Customer Value Evaluation Model for Logistics Enterprises

2.1 Definition and Composition of Customer Value in Logistics Enterprises

Customer value theory can be defined from the perspective of customers and enterprises. From the perspective of customers, customer value is defined as the value that customers perceive the products or services provided by enterprises. From the perspective of enterprises, customer value is defined as the value that customers contribute to the development of enterprises\cite{11,12}. The purpose of this paper is to study how to classify different customer values for the development of logistics enterprises, therefore, define customer value from the perspective of enterprises.

Different customers have different values to logistics enterprises. We can classify the customer value of logistics enterprises from a series of factors, such as customer transaction information, individual consumption behavior characteristics, customer loyalty and so on\cite{13}. Customer value research has been developed to a certain extent. Zhang Lanxia\cite{14} and others classify customer value from both quantitative and qualitative perspectives, including intrinsic value, external value, strategic value customers and so on. SGu\cite{4} and others study the customer evaluation of logistics enterprise CRM from the point of view of customer lifetime value ((CLV)). Zhang Hong\cite{15} and others think that it is necessary to predict the potential value of customers and then innovate products to improve enterprise profits. Based on the above research and combined with the characteristics of logistics enterprises, this paper designs the customer value evaluation index system of logistics enterprises according to the current value and potential value of customers.

2.2 Design of Customer Value Evaluation System for Logistics Enterprises

2.2.1 Determination of Current Value Index of Logistics Customers. The current value of logistics customers mainly includes the contribution of customers purchasing enterprise services to enterprises, which involves a series of logistics services such as storage, transportation, distribution and so on. Therefore, warehouse index, transportation index and customer credit rating index in the process of distribution are selected as the main indicators of the current value of logistics customers. These indicators can be subdivided as follows.

(1) Warehouse indicators
The warehouse has the function of storing and keeping goods, which can reflect the value of customers from the warehouse area occupied by goods, the total amount of goods in and out of the warehouse, and the number of times in and out of the warehouse.

(2) Transport indicators
Transportation is an essential link in the circulation of goods from manufacturers to customers, and the customer value is mainly reflected in the customer's average freight volume, transportation time and transportation frequency.

(3) Customer's credit rating index
From the credit rating of customers, we can see whether customers have a good reputation in the process of commercial transactions, and it is also related to the future development of customers. Customer credit rating is mainly reflected from the level of customer order performance, the level of cost settlement and the serious attitude of customers to the evaluation of logistics enterprises.

2.2.2 Determination of Potential Value Index of Logistics Customers. The potential value of the customer refers to the value that the customer may bring to the enterprise in the future cooperation with the enterprise. This paper intends to select customer loyalty, customer satisfaction and customer
growth as the main indicators of customer potential value. From the perspective of customer life cycle value theory, these three indicators can comprehensively reflect the length of customer remaining life cycle and measure the potential profit space of customers. The specific indicators are explained below.

(1) Customer loyalty indicators
Customer loyalty means that due to the influence of quality, price, service and many other factors, customers have feelings for the products or services of an enterprise, and form a preference and repeat purchase of the products or services of the enterprise for a long time. For logistics enterprises, it is to recognize the services brought by logistics enterprises and then produce sustainable logistics consumption behavior. It mainly includes logistics service price tolerance, repeat purchase level and customer service evaluation of logistics enterprises.

(2) Customer satisfaction index
Customer satisfaction refers to the index obtained by comparing the perceived effect of a product with its expected value, which reflects the intuitive feeling of the customer as the main body of the enterprise service, and has a significant impact on the interests of logistics enterprises. It is mainly evaluated from the logistics limitation level, logistics personnel service level and logistics after-sales service level.

(3) Customer growth indicators
Customer growth refers to the future development ability of customers, and the potential value of logistics customers increases with the growth of customers. It is mainly evaluated from three aspects: the possibility of customer incremental purchase, the possibility of cross-purchase and the possibility of strategic alliance.

2.2.3 Construction of Customer Value Evaluation Index system for Logistics Enterprises. Through the determination of the current customer value and potential value index of logistics enterprises, the customer value evaluation index system of logistics enterprises is constructed as shown in figure 1[12].

![Figure 1. Customer value evaluation index system of logistics enterprises.](image-url)
3. Classification Algorithm

3.1 Principle of AdaBoost Algorithm

Given a set \( \{ (x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_N, y_N) \} \), where \( x_i \) represents the feature vector of a thing, \( y_i \) represents the corresponding classification, and \( x_i = \begin{cases} -1 & \text{if } y_i \in \{-1, 1\} \end{cases} \). After \( M \) iterations of AdaBoost algorithm, \( M \) weak classifiers are obtained, that is, set \( \{ k_1, k_2, \ldots, k_M \} \), for each data item \( x_i \), all have corresponding weak classification results, that is \( k_m(x_i) \in \{-1, 1\} \). The strong classifier \( C_m \) is obtained by linearly combining the classification results of \( M \) weak classifiers[16].

After \( m-1 \) iterations, the strong classifier is:

\[
C_{m-1}(x_i) = a_1k_1(x_i) + \cdots + a_{m-1}k_{m-1}(x_i)
= \sum_{n=1}^{m-1} a_n k_n(x_i)
\]

(1)

Where \( a_n \) is the weight of the weak classifier, \( m > 1 \); After \( m \) iterations, the strong classifier is:

\[
C_m(x_i) = a_1k_1(x_i) + \cdots + a_{m-1}k_{m-1}(x_i) + a_Mk_M(x_i)
= C_{m-1}(x_i) + a_Mk_M(x_i)
= \sum_{n=1}^{M} a_n k_n(x_i)
\]

(2)

In order to judge whether the strong classifier is better than the weak classifier, it needs to be determined by the weight \( a_M \) of the \( M \) weak classifier. Here, the exponential loss sum of all data items is taken to define the error \( E \) of \( C_m \), that is:

\[
E = \sum_{i=1}^{N} \exp[-y_i C_m(x_i)]
= \sum_{i=1}^{N} \exp[-y_i[C_{m-1}(x_i) + a_Mk_M(x_i)]]
\]

(3)

The weight assigned to the data item \( x_i \) of the first iteration is 1, represented by \( w_i^{(1)} = 1 \), the weight after \( m-1 \) iterations is \( w_i^{(m)} = \exp[-y_i C_{m-1}(x_i)] \), the above formula can be simplified as follows:

\[
E = \sum_{i=1}^{N} \exp[-y_i[C_{m-1}(x_i) + a_Mk_M(x_i)]]
= \sum_{i=1}^{N} \exp[-y_i C_{m-1}(x_i)] \exp[-y_i a_Mk_M(x_i)]
= \sum_{i=1}^{N} w_i^{(m)} \exp[-y_i a_Mk_M(x_i)]
\]

(4)

The above style can be divided into:
\[ E = \sum_{y_i=k_m(x_i)} w_i^{(m)} \exp(-a_m) + \sum_{y_i \neq k_m(x_i)} w_i^{(m)} \exp(a_m) \] (5)

Since the result \( k_m \) of the weak classifier and the actual value \( y \) are either equal to 1 or equal to -1, the first item in the formula indicates that the weak classifier is classified correctly. That is to say, the result of the \( m \) iteration is the same as the actual classification \( y \), then there is \( y_i k_m(x_i) = 1 \), the second term indicates a classification error, that is, if the result of the \( m \) iteration is different from the actual classification \( y \), there is \( y_i k_m(x_i) = -1 \). Then the first item is the error summation of all correctly classified data items \( x_i \), and the second item is the error summation of all misclassified data items \( x_i \), and then the following transformations are made:

\[ E = \sum_{i=1}^{N} w_i^{(m)} \exp(-a_m) + \sum_{y_i \neq k_m(x_i)} w_i^{(m)}[\exp(a_m) - \exp(-a_m)] \] (6)

It can be seen from the above formula that the value of \( E \) depends on the size of \( \sum_{y_i \neq k_m(x_i)} w_i^{(m)} \), in order to find the minimum error, we derive the formula (5):

\[ \frac{dE}{da_m} = \sum_{y_i \neq k_m(x_i)} w_i^{(m)} \exp(a_m) - \sum_{y_i \neq k_m(x_i)} w_i^{(m)} \exp(-a_m) \] (7)

If the above formula is equal to 0, the weight is as follows:

\[ a_m = 1/2 \left\{ \ln\left( \sum_{y_i \neq k_m(x_i)} w_i^{(m)} \right) - \ln\left( \sum_{y_i = k_m(x_i)} w_i^{(m)} \right) \right\} \] (8)

Let \( \varepsilon \) denote the error rate, and its value is:

\[ \varepsilon = \sum_{y_i \neq k_m(x_i)} w_i^{(m)} / \sum_{i=1}^{N} w_i^{(m)} \] (9)

The formula (8) can be simplified as follows:

\[ a_m = 1/2 \ln[(1-\varepsilon)/\varepsilon] \] (10)

Through the above analysis, after each iteration, the weight of each training sample data is as follows:

\[ w_i^{(m+1)} = w_i^{(m)} \exp[-y_i a_m k_m(x_i)] \]

\[ = w_i^{(m)} \times \begin{cases} 1 & \text{correct classification} \\ \exp(a_m) & \text{error classification} \end{cases} \] (11)

3.2 Neural Network Logistics Customer Classification Model Based on AdaBoost Algorithm

The training sample of logistics customer value is given and normalized, the weight of each data item is set to \( 1/n \) during initialization, the BP neural network is selected as the weak classifier to train the training sample for the first time, then the corresponding customer value data is tested, and the corresponding error and the weight of the data item are calculated, and the product result of the updated weight value and the logistics customer data is used as the input of the second BP neural network. With this cycle, the classification is finally completed. The structure of the neural network logistics customer classification model based on AdaBoost algorithm is shown in figure 2.
Figure 2. Neural network logistics customer classification model based on AdaBoost algorithm.

The classification steps of neural network logistics customers based on AdaBoost algorithm are as follows:

1) Given the training sample set \( \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_N, y_N)\} \) of logistics customer value, it is used as the input of weak classifier after normalization.

2) Initialize the weight of the neural network weak classifier so that \( a_m = 1/N \), \( a_m \) represents the weight of the samples in the \( m \)-th iteration, and \( N \) is the number of samples.

3) The weak classifier \( k_m \) of neural network is trained, and the data sample of logistics customer value is used as the input of neural network.

4) Calculate the error \( \varepsilon_i \) and the average error \( \bar{\varepsilon} \) of the weak classifier of the neural network.

5) Update the sample weight and the weight of the neural network weak classifier, in which the sample data weight \( w_i^{(m+1)} = w_i^{(m)} \exp[-y_i a_M k_M(x_i)] \) and the neural network weak classifier weight \( a_m = 1/2 \ln[(1-\varepsilon)/\varepsilon] \).

6) To judge the error, and if the condition is met, the training will end, otherwise, the third step will be carried out until the error standard is reached.

7) The strong classifier \( C_m(x_i) = \sum_{n=1}^M a_n k_n(x_i) \) is obtained through the linear combination of several neural network weak classifiers. The flow of neural network logistics customer classification algorithm based on AdaBoost algorithm is shown in figure 3.
3.3 Discussion on the Convergence of the Model
The training error bound of the final classifier of AdaBoost algorithm is:

\[
\frac{1}{n} \sum_{i=1}^{n} I(x_i \neq y_i) \leq \frac{1}{n} \sum_{i=1}^{n} \exp[-y_i f(x_i)] = \prod_{m} Z_m
\]  

(12)

Where \( G(x_i) \) is the final classifier, which is the linear combination of \( f \) multiple weak classifiers, and \( Z_m \) is the linear combination of the weights of each sample point in the \( m \)-round training, and there is \( Z_m \leq 1 \), this theorem ensures the convergence of the algorithm in the iterative process.

4. Examples of Customer Classification in Logistics Enterprises
In order to verify the effectiveness of this algorithm, the above customer classification process is realized by using python programming language environment.

4.1 Customer Data Samples of Logistics Enterprises
According to the customer value of the above analysis, customer classification can be divided into two categories: customer current value and customer potential value. 100 feature-related customer data are selected from a company as training sample sets, some of the normalized data are shown in Table 1, 20 customer data are used as test samples, and normalized data are shown in Table 2. In this paper, according to the relevant information, customers are divided into four types: diamond customers, gold customers, ordinary customers and potential customers.
4.2 BP-AdaBoost Algorithm Analysis Model

Each user selects 18 characteristic factors as the input of the BP network, selects the sigmoid function as the classification function of the network, and designs four vectors to represent four kinds of customers as the output, as shown in Table 3. The number of weak classifiers is 3, and the allowable error is 0.2, in which the training times of weak classifiers is 100000 and the learning rate is 0.02. After the completion of the training, 20 data were used as test samples.
Table 3. Four types of customer representation vectors.

| Customer category     | Representation vector |
|-----------------------|-----------------------|
| Diamond customer      | (1,1)                 |
| Gold customer         | (1,0)                 |
| Ordinary customer     | (0,1)                 |
| Potential customer    | (0,0)                 |

4.3 Achievement Analysis

The network is trained with 100 customer data. After 100000 iterations, all the models converge and reach the allowable error, and the network training is over. The fitting comparison between the three weak classifiers and the generated strong classifiers is shown in figure 4. It can be seen from figure 4 that the error of the strong classifier is lower and the effect is better.

The classification results of 20 test samples are shown in Table 4, and the classification accuracy of strong classifier and weak classifier is compared as shown in Table 5. It can be seen from the table that the classification effect of strong classifier is better.

![Figure 4. Fitting distribution of weak classifier and strong classifier](image)

Table 4. Customer classification results of 20 logistics enterprises.

| Customer category     | Data number           |
|-----------------------|-----------------------|
| Diamond customer      | 3, 4, 6, 7, 8, 14, 16, 17 |
| Gold customer         | 9, 10, 13, 19         |
| Ordinary customer     | 1, 2, 5, 12, 15, 18, 20 |
| Potential customer    | 11                    |

Table 5. Classification accuracy of strong classifiers and weak classifiers

| Classifier          | Weak classifier 1 | Weak classifier 2 | Weak classifier 3 | Strong classifier |
|---------------------|-------------------|-------------------|-------------------|-------------------|
| Correct rate        | 65%               | 70%               | 80%               | 95%               |

5. Conclusion

The correct classification of customers affects the customer relationship management of logistics enterprises, and then affects the huge interests of enterprises. In this paper, the neural network weak
classifier is cascaded by AdaBoost algorithm with high precision, and a neural network logistics customer classification model based on AdaBoost algorithm is established. After training, 20 customers are classified. The analysis of the example shows that the method has a good classification effect and can be applied to the customer classification of logistics enterprises. Because of the high accuracy of AdaBoost algorithm and easy selection of cascaded weak classifiers, the classification effect is better and the scope of application is wider, but the number of iterations is not easy to set and needs further improvement.

References
[1] Fan Zhiping, Wang Jianyu, Chen Yuan. Review and Prospect of customer relationship Management (CRM). Systems Engineering 2002; 20.
[2] Li JJ, Shu H. Application research of improved neural network in customer value classification for logistics enterprises. J Converg Inf Technol 2012; 7: 399–404. https://doi.org/10.4156/jcit.vol7.issue20.47.
[3] Li Y, Huang J, Song T. Examining business value of customer relationship management systems: IT usage and two-stage model perspectives. Inf Manag 2019;56:392–402. https://doi.org/10.1016/j.im.2018.07.012.
[4] Gui S, Feng W, Zhang Z. Evaluate the performance of CRM in logistics enterprises based on AnyLogic simulation. Proc - Int Conf Manag Serv Sci MASS 2009 2009:6–9. https://doi.org/10.1109/ICMSS.2009.5303529.
[5] Zhang L, Zhang L, Chen S, Cai L, Yu Y, Hao S. Research on customer classification based on fuzzy clustering. J Comput Inf Syst 2007;3:1971–6.
[6] Cheng CH, Chen YS. Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst Appl 2009;36:4176–84. https://doi.org/10.1016/j.eswa.2008.04.003.
[7] Zhang B, Peng QY. Railway Freight Customer Segmentation Based on KFAV Model. Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal Transp Syst Eng Inf Technol 2017;17:235–42. https://doi.org/10.16097/j.cnki.1009-6744.2017.03.034.
[8] Li Yuan. Study Of CRM System Based On Decision Tree Classification Algorithms 2005.
[9] He Peng. Research On Customer Classification Of The 3rd-party Logistics. Wuhan University of Technology 2008.
[10] CHEN Qianshu, FANG Xiaoping. Research on Logistics Customer Value Based on RFM Model. 2019; 7: 19–22. https://doi.org/10.13714/j.cnki.1002-3100.2019.07.005.
[11] Wang Liwei. Research on Enterprise Marketing Strategy based on customer value. Modern Marketing (Business Edition) 2019; 2: 96.
[12] Wu Kang. Research on Customer Classification Management of ZH Logistics Company Based on Integrated Value Evaluation. Dalian University of Technology, 2018.
[13] Liu Qiuping. Research on Fuzzy Comprehensive Evaluation of Distribution customer value based on Cloud Model 2019; 28:148–9.
[14] Zhang Lanxia, Wu Guohua. Customer Classification Management in CRM. Journal of Northeastern University (Social Science) J Northeast Univ Soc Sci 2003; 5:44–9. https://doi.org/10.15936/j.cnki.1008-3758.2003.02.010.
[15] Zhang H, Liang X, Wang S. Customer value anticipation, product innovativeness, and customer lifetime value: The moderating role of advertising strategy. J Bus Res 2016;69:3725–30. https://doi.org/10.1016/j.jbusres.2015.09.018.
[16] Santos SGT de C, de Barros RSM. Online AdaBoost-based methods for multiclass problems. Artif Intell Rev 2020;53:1293–322. https://doi.org/10.1007/s10462-019-09696-6.