2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI) | 978-1-6654-7997-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/IWECAI55315.2022.00093

Bank Customer Churn Prediction Analysis Based on Improved WOA-SVM

Hanyue He
School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China
1424484255@qq.com

***Abstract*—At present, the business services provided by major banks to their customers are basically similar, with no distinctive features. With the development of society and the increasingly fierce competition in the industry, customers' demand for personalized services is getting higher and higher, resulting in frequent customer churn. Banks urgently need to adjust their business strategies and develop programs to meet customers' needs for personalized services so as not to lose customers and retain economic benefits. Customer churn prediction analysis is an important part of customer relationship management. The paper uses the improved whale optimization algorithm to improve the support vector machine to get the improved WOA-SVM model. It is also compared with GA-SVM, SVM, multilayer perceptron, and logistic regression methods with an example of VIP customer churn prediction in a domestic commercial bank, and found that the model obtains the best correct rate, hit rate, coverage rate, and boost factor, and is an effective model for predicting customer churn in banks.**

Keywords-component; customer churn; whale optimization algorithm; support vector machine; prediction

I. Introduction

Customer retention has a surprising impact on a bank's profitability, far outweighing the impact of size, market share, unit costs and many other factors commonly considered to be related to competitive advantage.[1] Not only does customer churn create opportunity costs due to reduced sales, but it also leads to a relative reduction in new customers attracted. A small improvement in customer retention can lead to a sizable improvement in profits,[2] so effectively identifying potential future churn allows customers to be classified so that marketing efforts can be tailored to the characteristics of different customer segments.

Data mining techniques can make marketing more accurate and rapid by creating models that predict customer behavior and discover important information hidden behind large amounts of data. To predict potential lost customers more effectively, scholars have proposed the following two main types of methods: The first type of methods is traditional classification methods such as decision trees, logistic regression, Bayesian classifiers, and cluster analysis.[3] However, these methods are not effective in dealing with large-scale, high-dimensional customer data with non-linear relationships, non-normal distribution and temporal order, and do not guarantee the generalization ability of the built models; the second type of methods is artificial intelligence classification methods, such as artificial neural networks, self-organizing mapping and evolutionary learning algorithms. This type of methods can overcome the difficulties faced by the first type of methods to a certain extent,[4] not only has the ability of nonlinear mapping and generalization, but also has better robustness and prediction accuracy. However, such methods mainly rely on the empirical risk minimization principle, which can easily lead to the degradation of generalization ability and difficulty in determining the model structure. These shortcomings greatly limit the application of these methods in practice.[5-7]

Therefore, it is still valuable to explore new customer churn prediction methods for research. Aiming at the characteristics of bank customer churn data, this paper draws on Li Lin's improved WOA-SVM model based on the structural risk minimization criterion with the overall accuracy of the model as the evaluation criterion to make predictions, and further uses the VIP customer data of a domestic commercial bank to predict its possible future churn and conducts a comparison with GA-SVM,[8] SVM, multilayer perceptron, and logistic regression methods. Comparison was made, and it was found that the improved WOA-SVM method could obtain better accuracy.

II. Overview of BLWOA-SVM model

A. Support Vector Machine (SVM)

Support Vector Machine (SVM) and its improvement algorithm is a typical machine learning algorithm for classification model.SVM aims at minimizing structural risk and has the advantages of generalization ability and strong global optimization seeking in solving nonlinear, small sample and high-dimensional pattern recognition, and has high accuracy and computational efficiency in classification evaluation. Usually, multiclassification evaluation is a typical nonlinear classification problem.SVM is implemented by mapping function $\varphi(x)$, kernel function $k(x_i,x_j)=\varphi(x_i)\cdot\varphi(x_j)$.The input samples are mapped to a high-dimensional feature space, andaccording to the Lagrangian duality, the nonlinear support vector machine is converted to solve the following convex quadratic programming problem:

$$min_\alpha \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\cdot\alpha_j\cdot y_i\cdot y_j\cdot k(x_i,x_j)-\sum_{i=1}^{M}\alpha_i \tag{1}$$

$$s.t.\sum_{i=1}^{M}\alpha_i\cdot y_i=0 \tag{2}$$

$$0\le\alpha_i\le C, i=1,2,\cdots,M \tag{3}$$

Where: α_i is the Lagrange multiplier. Using the quadratic programming method and the KKT condition, the special solution α^* of the Lagrange multiplier is obtained, and the classification decision function can be obtained as:

$$f(x) = sign\left(\sum_{i=1}^{M} \alpha_i^* y_i k(x, x_i) + b^*\right) \quad (4)$$

If a Gaussian radial basis kernel function $k(x, x_i) = \exp(-\|x - x_i\|^2 / 2g^2)$ is used, the final decision classification function is:

$$f(x) = sign\left(\sum_{i=1}^{M} \alpha_i^* \cdot y_i \cdot \exp\left(-\frac{-\|x - x_i\|^2}{2g^2}\right) + b^*\right) \quad (5)$$

where:g>0 is the bandwidth of the Gaussian kernel, i.e., the kernel function parameter.

Obviously, the merit of the multiclassification evaluation SVM is mainly determined by the values of the parameter penalty factor C and the width g of the Gaussian kernel function. It is standard practice to use metaheuristic algorithms for the selection of SVM parameters, such as particle swarm algorithms and genetic algorithms, etc. WOA has fewer adjustment parameters and convergence accuracy and speed are significantly better than genetic algorithms and particle swarm algorithms. However, it also has the disadvantages that exist in all of the swarm intelligence optimization algorithms, so it still needs to be improved. In this paper, the improved WOA is used to optimize the SVM parameters and construct the improved WOA-SVM model.

B. *Whale Optimization Algorithm (WOA)*

The whale optimization algorithm is an intelligent optimization algorithm proposed in recent years. The optimal solution is obtained based on a search that simulates the behavior of humpback whales catching prey in nature. It mainly includes three kinds of operators. After realizing local search through shrinkage envelope mechanism and spiral update mechanism, it adopts random search strategy to realize global search of the algorithm, which shows excellent performance and wide application prospect in solving optimization problems due to the characteristics of high accuracy and fast convergence. Humpback whales use two ways to capture prey, firstly surround the prey and then select the bubble net to attack, and each attack is selected in a different way to update in the optimal position with random whale individuals, and converge to the updated position.[9]

1) Surrounding prey: When a humpback whale surrounds its prey, it will choose its most suitable positioning and surround direction, and the calculation formula is as in equation (6) and equation (7).

$$\vec{D} = \left|\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)\right| \quad (6)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (7)$$

where t denotes the number of iterations, A and C are coefficient vectors, X is the position vector of the prey, and X is the position vector of the remaining whales. In the calculation process, the values of vectors A and C are adjusted to find the position of X around the optimal solution. A and C are calculated as in equations (8) and (9). Where a decreases linearly from 2 to 0; the r vector is a random vector in [0,1].

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r} - \vec{a} \quad (8)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (9)$$

2) Bubble net attack method: simulate humpback whale's bubble net behavior by two ways.

Shrinkage bracketing mechanism: This can be achieved by reducing the value of a in equation (9). When a decreases linearly from 2 to 0, the vector A is a random vector in [-a,a], i.e., the next position of the search agent is the value between the initial position and the current best position.[10] As a decreases, the range of variation of vector A also decreases, and the formula for a is given in equation (10).

$$\vec{a} = 2 * \frac{t_{max} - t}{t_{max}} \quad (10)$$

Spiral update position: firstly, the distance between individual whale position and prey position is calculated, and the spiral equation as in eq.(11) and eq.(12) is established to simulate the spiral hunting mode of humpback whales. Where b is used to define the logarithmic spiral shape and l is a random number [0,1].

$$\vec{X}(t+1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (11)$$

$$\vec{D}' = \left|\vec{X}^*(t) - \vec{X}(t)\right| \quad (12)$$

3) Searching for prey: Humpback whales have a unique way of searching for prey,which can be done randomly based on the whales' positions in relation to each other rather than the prey position, mathematically expressed as in eqs.(13) and (14), where the position of the randomly selected individual whale is denoted by Xrand. This will be updated in the search phase based on the most recently selected search agent rather than the best search agent. $A > 1$ forcing the search agent to stay away from the reference whale and allowing the WOA algorithm to perform a global search.

$$\vec{D} = \left|\vec{C} \cdot \vec{X}_{rand}(t) - \vec{X}(t)\right| \quad (13)$$

$$\vec{X}(t+1) = \vec{X}_{rand}(t) - \vec{A} \cdot \vec{D} \quad (14)$$

First,by generating a random number p of [0,1] to determine whether the algorithm enters the bubble attack mechanism, the $p > 0.5$ when the bubble attack starts, otherwise, judging whether the system enters the prey search or prey encirclement phase based on the absolute value of the coefficient vector $A < 1$ when it enters the prey encirclement phase.[11] $A \geq 1$ when it enters the prey search phase. With the increasing number of iterations, a gradually decreases from 2 to 0, and the algorithm transitions from the search prey phase to the surround prey phase.

C. *Whale Optimization Algorithm Improvement*

The standard WOA uses a random method for the initial population, which relies too much on randomness and thus limits the convergence speed, and later when the number of iterations reaches a certain number, the WOA tends to fall into local optimum and low convergence accuracy.[12] For this reason, this paper improves the convergence speed, global optimization search, and convergence accuracy of the standard WOA.

1) Rapidity improvement by incorporating Lévy flight strategy

Lévy flight was proposed by French mathematician Paul Lévy, whose probability distribution of flight steps satisfies the heavy-tailed distribution.The step size of Lévy flight is random, which is a kind of search method between long-term small-step search and short-term large-step jump change.[13] In this paper, we retain the standard WOA bracketing strategy and random search strategy, and adopt the Lévy flight strategy instead of the spiral trajectory strategy of WOA, which makes the update positions more diverse, can quickly jump out of the local optimum and improve the overall population search efficiency, and replace the WOA equation (11) with equation (15).

$$x_i^{t+1} = x_i^t + L\acute{e}vy(n) \oplus \alpha \tag{15}$$

Where. $\alpha = x_i^t - x_b$. $\oplus$ is the dot product; x_i^t is the *i*-th solution in the *t*-th generation; x_b is the current optimal solution; Lévy(s) conforms to the Lévy distribution and satisfies Lévy(s)~$|s|^{-1-\beta}$, $0 < \beta \leq 2$. The mathematical model of the Lévy distribution is given by,

$$L(s,\gamma,u) = \begin{cases} \sqrt{\frac{\gamma}{2\pi}} \exp\left[-\frac{\gamma}{2(s-u)}\right] \frac{1}{(s-u)^{\frac{3}{2}}}, 0 < u < s < \infty \\ 0 \qquad\qquad\qquad\qquad\qquad , otherwise \end{cases} \tag{16}$$

where: γ and u are greater than 0; γ is the scale parameter; u is the shift parameter; and the step size of the Lévy flight can be expressed as,

$$L\acute{e}vy(x) = 0.01 \times \frac{u \cdot \sigma}{|u|^{\frac{1}{\beta}}} \tag{17}$$

where: u and v are random numbers obeying standard normal distribution of [0,1]; β=1.5; σ is calculated as

$$\sigma = \left\{ \frac{\Gamma(1+\beta) \cdot \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \cdot \beta \cdot 2^{\frac{\beta-1}{2}}} \right\}^{\frac{1}{\beta}} \tag{18}$$

where: $\Gamma(x)$ is the gamma function, i.e. $\Gamma(x)=(xx-1)!$.

2) Global Optimization Seeking Improvement with the Introduction of Elite Reverse Learning

In order to increase the diversity of the WOA population, improve the convergence speed of the algorithm, and prevent the algorithm from entering premature maturity, this paper introduces an elite backward learning mechanism in both the initial population and each population iteration. The backward learning (BL) method adopts a two-way evaluation principle, in which not only the candidate solution of this iteration is evaluated in each iteration, but also the solution in the opposite direction of this candidate solution is evaluated, and then the solution with the best evaluation between the current solution and the backward solution is selected as the next generation candidate solution.[14] It is shown that the reverse solution has 50% more probability of being close to the global optimal solution than the current solution.[15]

The use of backward learning can greatly improve the probability of the algorithm searching for the global optimal solution, which is defined as the point where the current population a candidate solution is assumed to be an N-dimensional search space.

$$Q_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,N}) \tag{19}$$

Where: $x_{i,N} \in \mathbb{R}$; $x_{i,j} \in [x_{i,j\_min}, x_{i,j\_max}]$; $i = 1,2,\cdots,M; j = 1,2,\cdots,N$; i is a population individual, then the reverse point of Q is:

$$P_i = (y_{i,1}, y_{i,2}, \cdots, y_{i,N}) \tag{20}$$

Where: $y_{i,j} = x_{i,j\_min} + x_{i,j\_max} - x_{i,j}$; x_{i,j_min}, x_{i,j_max} are the upper and lower bounds of the jth dimension of individual i, respectively. If *f(·)* is the evaluation function of the candidate solution fitness value, for the minimization problem, if *f(P) < f(Q)*, then P is chosen as the best candidate solution for the current population, otherwise Q is chosen.

The elite inverse solution is defined as: the extreme point of the current population of ordinary individuals $Q_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,N})$ is the elite individual $Q_{i\_best} = (x'_{i,1}, x'_{i,2}, \cdots, x'_{i,N})$, Then the corresponding inverse solution is $P_{i\_best} = (y'_{i,1}, y'_{i,2}, \cdots, y'_{i,N})$,where:

$$y'_{i,j} = k \cdot (x'_{i,j\_min} + x'_{i,j\_max}) - x'_{i,j} \qquad j = 1,2,\cdots,N \tag{21}$$

where: $k \in [0,1]$ of random uniformly distributed random numbers.

$$x'_{i,j\_min} = \min(x_{i,j}), x'_{i,j\_max} = \max(x_{i,j}) \tag{22}$$

The use of dynamic boundaries facilitates the preservation of the search experience and allows the inverse solution to lie within a gradually shrinking search space. Since the possibility of jumping out of the boundary exists for the reverse solution, the reset of equation (23) is used.

$$y'_{i,j} = k' \cdot (x_{i,j\_min} + x_{i,j\_max}) + x_{i,j\_min} \tag{23}$$

where: $k' \in [0,1]$ of uniformly distributed random numbers.

D. Improved WOA-SVM model

In this paper, the whale optimization algorithm incorporating Lévy flight and elite backward learning is applied to the parameter search of SVM, and the penalty factor C and the kernel function parameter g are combined in an optimization search, so as to reduce the randomness of parameter selection to obtain the SVM classification model with optimal accuracy. , the specific steps are as follows.

a) Randomly initialize the population with individual location information as (C,g), population size H, maximum iteration t_{max}, and also divide the bank customer data into test and training sets after normalizing them to [0,1];

b) To calculate the fitness, the support vector machine learns with the population individual position information C and g. The classification error rate of the training set is the fitness of the individual, and the individual position information C* and g* with the smallest current fitness is recorded;

c) Updating population individual positions C and g. using a whale optimization algorithm incorporating Lévy flight and elite inverse learning

d) If t_{max} is not reached, then return to step b), otherwise end the iteration and return the optimal individual position information C* and g*;

e) The BLWOA-SVM bank customer churn prediction analysis model was developed using the optimal parameters C* and g*.

III. Empirical Studies

A. Experimental Preparation

The core data involved in this study are the data of VIP customers of a domestic commercial bank, including Glory Card, King Card, Star Card and Diamond Card customers. Based on the specific quantitative criteria of VIP customers at the head office level in the Management Measures of VIP Customer Service of a domestic commercial bank, combined with the attributes and characteristics of VIP customers, the criteria for customer churn is finally determined as follows: when a customer's average daily deposit balance is less than RMB 20,000 for three consecutive months after a certain observation point, the customer is defined as a churned customer.

The data in this paper comes from a domestic branch of a commercial bank in China, collecting data from January to December 2018 where the data window is taken from January to August, with a delay of September data window during which the data values of regular account balance, current account balance, treasury account balance, the number of new accounts opened in the last May, and the number of accounts cancelled in the last May are used as inputs to the model, and the customer churn status from October to December is used as output of the model. The twofold and threefold standard deviation tests were selected for abnormal data rejection. By sampling, 2965 training customers (1937 non-churning customers and 1028 churning customers) and 1273 test customers (729 non-churning customers and 544 churning customers) were obtained at the same time as the training customers.

Based on the above data, the sample set (x,y) is constructed where x is the input data and y is the category attribute of the sample for "churned customers": y=1 and for "non-churned customers": y=-1. According to the experimental analysis of Zhang Wen, the optimal kernel function can be obtained For the exponential radial basis kernel function, a radial basis kernel function with parameters u=1 and C=2 is used to construct the prediction model with good results.

B. Analysis of Experimental Results

GA-SVM, SVM, logistic regression and multilayer perceptron models are selected for comparison tests. The discriminant surface of logistic regression is $\frac{1}{1+\exp(-\beta_0-\sum_{i=1}^{d}\beta_i x_i)} = 0$, where β_0 is the constant term, d is the dimension, and β_i is the regression coefficient, which indicates the contribution of x_i to the discriminant surface; the multilayer perceptron model is a nonlinear model with an implicit layer, and the stopping condition is that the error falls into local minimal.

The model evaluation criteria refer to Lin Li, as shown in Table 1: model accuracy = (A+D)/(A+B+C+D); hit ratio = A/(A+C); coverage ratio = A/(A+B); boost factor = hit ratio / customer churn ratio in test data.

TABLE I. Description of Data

Customer Status	Predicted churn	Predicted non-churn
Actual Churn	A	B
Actual non-churn	C	D

The accuracy, hit ratio, coverage rate and boost factor of the model using BLWOA-SVM on the commercial bank VIP user dataset are 0.6239, 0.7349, 0.5361 and 1.7357, respectively. except for the slightly lower hit ratio than the multi-layer perceptron prediction model, the other indicators are significantly higher than the prediction results obtained by other methods. The coverage ratio of the multilayer perceptron prediction model is only 0.0201. It is known that the method shows overfitting phenomenon. The higher model accuracy indicates that the BLWOA-SVM model has a strong comprehensive prediction capability for the whole data set; the higher boosting factor, hit ratio and coverage ratio indicate that the use of the model can retain more potential lost customers at a smaller cost under different customer churn banking market environments. Therefore, the BLWOA-SVM model can achieve better results in customer churn in the banking industry with appropriate kernel functions and parameters selected.

TABLE II Comparison of Prediction Results of Different Prediction Models

Model	Accuracy	Hit Ratio	Coverage Ratio	Boost Factor
BLWOA-SVM	0.6239	0.7349	0.5361	1.7357
GA-SVM	0.5998	0.7196	0.2005	1.7106
SVM	0.5869	0.7103	0.1697	1.6984
Multi-layer perceptron	0.5576	0.7468	0.0201	1.6513
Logistic regression	0.5817	0.7261	0.1608	1.5534

IV. Concluding Remarks

SVM is a general learning algorithm based on statistical learning theory has a rigorous theoretical foundation, and can better solve the practical problems of nonlinearity, high and local minima that cannot be solved by traditional prediction methods in customer churn prediction in banking industry. In this paper, the improved whale optimization algorithm is used to accelerate the parameter search optimization of support vector machines. The BLWOA-SVM model is applied to bank VIP customer churn prediction. A comparative study with GA-SVM, SVM, multilayer perceptron, and logistic regression methods reveals that BLWOA-SVM has the features of simple classification surface, strong generalization ability, and high fitting accuracy from the method point of view; from the data conditions and structure, under the sample data conditions with more samples (in which support vectors are richer), richer attribute indicators, larger probability of customer churn in the sample, and fewer missing records Under the sample data conditions, BLWOA-SVM can achieve high prediction accuracy and has good application value.

References

[1] Li Xiangqing,Xia Guoen,Zhang Xianquan,Tang Qi,Ye Shuai. Customer churn prediction based on deep neural network weight integration[J]. Computer Technology and Development,2021,31(10):18-23.

[2] Shi Danlei,Du Baojun. BP neural network-based bank customer churn prediction[J]. Science and Technology Innovation,2021(27):104-106.

[3] Zhou Wanting,Zhao Zhijie,Liu Yang,Wang Jiaying,Han Xiaowei. Research on DBN prediction model for e-commerce customer churn [J/OL]. Computer Engineering and Applications:1-10[2021-12-23]. https://kns-cnki-net.webvpn.jxust.edu.cn/kcms/detail/11.2127.TP.20210919.1106.009.html.

[4] Yu S. D.,Huang X. Application of CART algorithm in airline customer churn [J]. Light Industry Science and Technology,2021,37(09):76-79.

[5] Zhang Yun,Song Shuang. Broadband customer churn prediction with non-complete data[J]. China New Communications,2021,23(14):125-126.

[6] Li G.C.,Jiang Y.L.,Ma W.B.,Xia G.E.. Customer churn prediction based on spectral regression feature dimensionality reduction[J]. Computer System Applications,2021,30(09):62-68.DOI:10.15888/j.cnki.csa.007981.

[7] Li Lin,Li Man. SVM-based analysis of bank customer churn prediction--a commercial bank VIP customer churn as an example[J]. Productivity Research,2010(09):58-59.DOI:10.19374/j.cnki.14-1145/f.2010.09.024.

[8] Zhang Steady,Zhang Lili. Analysis of bank customer churn prediction based on GA-SVM[J]. Computer and Digital Engineering,2010,38(04):55-58.

[9] Li A.D.,Liu S. Hybrid strategy to improve whale optimization algorithm [J/OL]. Computer Application Research:1-8[2021-12-23]. https://doi.org/10.19734/j.issn.1001-3695.2021.09.0449.

[10]]Zhao ZJ, Zhou XQ, Deng F, Xu C. Design of electric energy consumption monitoring system based on whale algorithm in industrial park[J]. Electronic Design Engineering,2021,29(24):102-106.DOI:10.14022/j.issn1674-6236.2021.24.021.

[11] Liu LB, Yang S, Wang ZJ, He XX, Zhao WL, Liu Shoujun, Du WG, Mi J. Coke quality prediction based on improved WOA-LSTM [J/OL]. Journal of Chemical Engineering:1-12[2021-12-23]. https://kns-cnki-net.webvpn.jxust.edu.cn/kcms/detail/11.1946.TQ.20211216.1452.005.html.

[12] Guo QC, Du XY, Zhang YY, Zhou Y. UAV 3D path planning based on improved whale algorithm[J]. Computer Science,2021,48(12):304-311.

[13] Meng Xianmeng,Cai Cui Cui. Whale optimization algorithm based on elite backward learning and Lévy flight[J/OL]. Electronic Measurement Technology:1-6 [2021-12-23]. https://kns-cnki-net.webvpn.jxust.edu.cn/kcms/detail/11.2175.TN.20211125.0843.004.html.

[14] He S, Zhang G, Chen YH, Yang SH. A WOA-SVM multi-classification algorithm incorporating Lévy flight and elite backward learning[J]. Computer Application Research,2021,38(12):3640-3645.DOI:10.19734/j.issn.1001-3695.2021.04.0164.

[15] Zhao Jing,Zhu Xijing,Meng Xiaoling,Wu Xiao. Application of improved whale optimization algorithm for time-optimal trajectory planning of robotic arm [J/OL]. Mechanical Science and Technology:1-10[2021-12-23]. https://doi.org/10.13433/j.cnki.1003-8728.20200596.