Financial Engineering and Risk Management (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2024.070323
ISSN 2523-2576 Vol. 7 Num. 3

Analysis of Credit Customer Delinquency Based on BP Neural Network Model

Xinyu Wu*, Jielin Shang

PLA Dalian Naval Academy, Dalian, China
**Corresponding author: 772774131@qq.com*

Keywords: BP neural network; credit; delinquency; prediction

Abstract: With the advancement of the economic market in the new century, the issues of credit evaluation and risk prediction have received increasing attention. The advent of the big data era has led to the widespread development and application of neural networks. As an excellent classification tool, artificial neural networks do not require fixed premises or assumptions about inputs and outputs before modeling. They possess self-learning and self-adaptation capabilities, exhibit strong nonlinear mapping abilities, and have fault tolerance mechanisms, making them a powerful tool for solving credit issues. This paper utilizes R software to clean, process, and analyze credit customer data provided by a German credit database. A BP neural network model is established and evaluated based on criteria such as accuracy and AUC value. The model demonstrates good fitting effects and is used to predict the corresponding data set.

1. Introduction

We can consider deep learning as the application of deep neural networks (DNN) for machine learning, which involves using neural networks with deep structures to develop powerful models . In 2005, Paulo J. Lisboa and Azzam F.G. Taktak published a systematic review in the authoritative journal Neural Networks, proposing the application of neural networks for decision support in cancer diagnosis [1]. In 2015, Fani E. Asimakopoulou and colleagues applied neural networks in electrical engineering to estimate ground resistance using an artificial neural network model [2]. In 2017, Tulasi K. Paradarami and others incorporated metadata information from business reviews into a traditional system developed using content-based collaborative filtering methods, constructing a hybrid recommendation system with neural networks [3][4]. Building on these studies, this paper utilizes credit customer data from a German credit database to establish a BP neural network model and predict delinquency.

2. Model Establishment

Regarding the structure of neural network models, determining the number of hidden layers and the number of nodes in these layers is crucial and indispensable. This determination was once a key constraint in the development of neural networks.

Ignaccolo[5] pointed out that when all nodes use sigmoid functions, a single hidden layer is

sufficient to solve any classification problem. Networks with two hidden layers can achieve any desired decision boundaries for classification. However, the final fitting effect of a network with two hidden layers is not significantly better than that of a network with only one hidden layer, especially in small networks. Moreover, in practical operations, to reduce errors, it is generally easier to increase the number of nodes in the hidden layer rather than the number of hidden layers[6]. Therefore, this paper decides to establish a model with only a single hidden layer. The empirical formula is summarized as follows:

$$n_1 = \sqrt{nm},$$

$$n_1 = \log_2^n,$$

$$n_1 = \sqrt{n+m} + a,$$

In this context, $a \in [1,10]$ is an integer, n_1, m and n are the number of neurons in the hidden layer, output layer, and input layer, respectively. Finally, we can further select the sigmoid function by comparing the actual values of the output terms.

3. Evaluation Metrics

This paper selects AUC (Area under the Curve) and KS (Kolmogorov-Smirnov) value as the evaluation metrics. First, let's introduce the confusion matrix(See Table 1):

Table 1: Confusion matrix

True Values	Predicted Values	
	0	1
0	True Negative (TN)	FalsePositive (FP)
1	False Negative (FN)	True Positive (TP)

Subsequently, precision P and recall R are defined as follows:

$$P = \frac{TP}{TP + FP},$$

$$R = \frac{TP}{TP + FN}.$$

According to the defined formulas, it is evident that the difference between precision and recall lies in their denominators[7]. When returning to the data, the distinction is that they target different populations. Precision (P) is based on the prediction results, indicating the proportion of true positive samples among those predicted as positive. Simply put, precision measures whether the identified samples are correct, while recall(R) measures whether all relevant samples are identified. These two metrics are typically inversely related: for a given model, a higher precision often corresponds to a lower recall, and vice versa. Given this situation, a harmonic mean of precision and recall, known as the F1 score, is defined as follows:

$$\frac{2}{F_1} = \frac{1}{P} + \frac{1}{R},$$

$$F_1 = \frac{2TP}{2TP + FP + FN},$$

Based on the aforementioned symbols, the definitions of TPR and FPR are as follows:

$$TPR = \frac{TP}{T} = \frac{TP}{TP + FN},$$

$$FPR = \frac{FP}{F} = \frac{FP}{FP + TN},$$

TPR (True Positive Rate): Represents the proportion of true positives in the actual positive instances (T), also known as Sensitivity.FPR (False Positive Rate): Represents the proportion of false positives in the actual negative instances (F).Clearly, the larger the TPR and the smaller the FPR, the better the model's prediction results.

A two-dimensional curve constructed using the defined FPR and TPR as the horizontal and vertical coordinates, respectively, is called the ROC (Receiver Operating Characteristic) curve. Alternatively, TNR (True Negative Rate) can also be used as the horizontal coordinate:

$$TNR = 1 - FPR,$$

TNR, also known as Specificity, represents the proportion of true negatives in the actual negative instances. On the graph(See Figure 1), each point's coordinates are (TPR, 1 - FPR). AUC (Area Under the Curve) refers to the area under the ROC curve. Given that the diagonal line corresponds to the random probability model of "coin flipping," any useful model should have its ROC curve above this diagonal. Consequently, the AUC value ranges from 0.5 to 1, with values closer to 1 indicating better model performance.

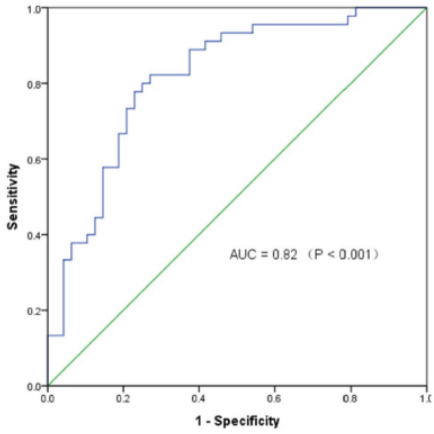

Figure 1: Sample diagram of ROC and AUC

In practice, there is no standard or unified opinion on the threshold for AUC values. It is generally judged based on industry-specific experience[8].The KS value is actually derived from the Kolmogorov-Smirnov statistic, which is based on the Kolmogorov-Smirnov test. It essentially measures the model's ability to separate positive and negative samples. The specific calculation process is as follows:Convert each sample's prediction result into a probability or a score. Rank these probabilities or scores from the smallest to the largest. Then, compute the cumulative distribution for both positive and negative samples. The KS value is the maximum absolute difference between these two cumulative distributions.The KS value ranges from 0 to 1, and generally, the larger the value, the better the model's ability to distinguish between positive and negative samples. Similar to the AUC value, there is no official standard for what constitutes a good KS value; it is typically judged by experience. Generally, a KS value above 0.41 indicates a fairly good model. However, in credit scoring and other models expecting a normal distribution, a KS value exceeding 0.9 is considered unusable.

4. Case Analysis

4.1 Data Description

The sample size for this study comprises 1000 observations from a German credit database, accessible via the following URL:

http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/german.data

Data analysis and modeling are conducted using the R programming language[9]. The dataset includes a qualitative dependent variable indicating whether credit customers defaulted on their loans. Specifically, there are 700 customers who did not default (coded as 1) and 300 customers who did default (coded as 2). Additionally, the dataset contains 20 independent variables representing various attributes of the customers. These independent variables include both qualitative and quantitative variables, with 7 being quantitative and 13 being qualitative. An overview of some of these variables is provided below(See Table 2):

Table 2: Variable description of case tow

Variable Nature	Variable Name	Variable Type	Variable Levels
dependent Variables	Credit Default Status	Qualitative	1 = No default, 2 = Default
Independent Variables: Qualitative Variables (13) Quantitative Variables (7)	Status of existing checking account	Qualitative	A11-A14
	Credit Purpose	Qualitative	A40-A410
	Job	Qualitative	A171-A174
	Foreign worker	Qualitative	A201 : Y A202 :N
	Age	Quantitative	[19,75]
	Duration in months	Quantitative	[4,72]

4.2 Model Training

In this study, considering the dataset's lack of clear delineation between training and testing sets, a 3:1 split was employed to partition the data into training and testing sets. Subsequently, 750 records were randomly selected for the training set, while the remaining 250 records were designated as the testing set, utilizing simple cross-validation for model selection. The performance of the models was evaluated using AUC (Area Under the Curve) and KS (Kolmogorov-Smirnov) statistics as metrics for assessing goodness of fit.

Regarding the determination of model structure, the input and output layers were set to have 47 and 1 neurons, respectively.A single hidden layer was selected, with all other parameters held constant as in Case Study 1, models were constructed with varying numbers of hidden layer neurons: 5, 6, 7, 8, 9, and 10. The corresponding results are presented below(See Table 3):

Table 3: Experimental results

Neuron	Count Accuracy
5	0.696
6	0.732
7	0.7
8	0.684
9	0.72
10	0.696

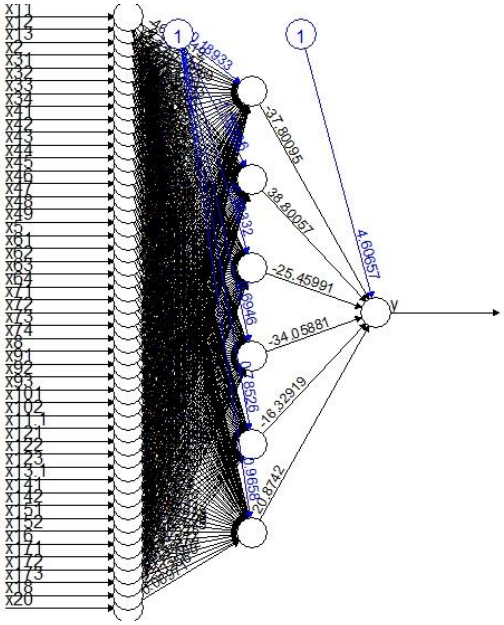

Figure 2: Model Topology Disgram

Based on the experiment, it is evident that the highest accuracy is achieved when the hidden layer neurons are set to 6 [10]. Therefore, the number of hidden layer neurons is determined to be 6. Subsequently, five-fold cross-validation is conducted, yielding an average AUC value of 0.761 and a KS value of 0.407, indicating excellent model performance. Below are the topology diagram(See Figure 2), ROC curve, AUC value, and KS value (See Figure 3)for the initial fit with 6 hidden layer neurons.

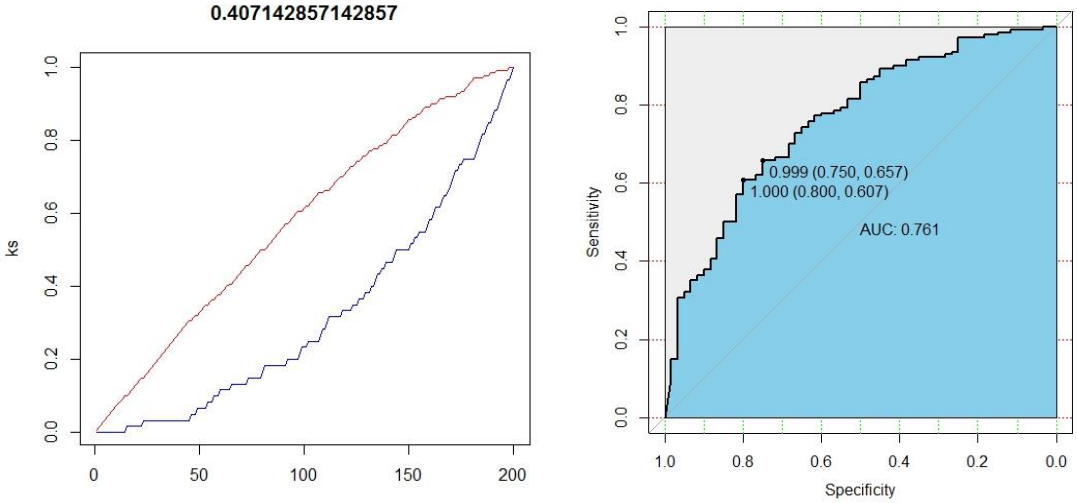

Figure 3: ROC curve and KS value

5. Conclusions

The case study presented in this paper utilizes real credit data from a company. Building upon data cleaning processes, the dataset's original training set is flexibly redistributed into new training and prediction sets at certain proportions for learning purposes. Both the AUC and KS values demonstrate the model's outstanding performance, further validating its efficacy in credit assessment.

References

[1] William Caballero,Ramo ́n Giraldo and Jorge Mateu. A universal kriging approach for spatial functional data [J]. Stochastic Environmental Research and Risk Assessment 2013 (7):1553–1563.
[2] Antonio Cuevas, Manuel Febrero and Ricardo Fraiman, On the use of the bootstrap for estimating functions with functional data [J]. Computational Statistics and Dat a Analysis 2006(51):1063–1074.
[3] Giraldo R, Delicado P, Mateu J. Continuous time-varying kriging for spatial prediction of functional data: An environmental application [J]. Journal of Agricultural, Biological, and Environmental Statistics 2010(1):66–82.
[4] Giraldo R, Delicado P and Mateu J. Ordinary kriging for function- valued spatial data. Environmental and Ecological Statistics 2011(3): 411–426.
[5] gnaccolo R, Mateu J and Giraldo R. Kriging with external drift for functional data for air quality monitoring [J]. Stochastic Environmental Re- search and Risk Assessment 2014(28): 1171–1186.
[6] Lina Schelin, Sara Sjöstedt-de Luna. Kriging Prediction Intervals Based on Semiparametric Bootstrap [J] Mathematical Geosciences 2010(42): 985–1000.
[7] Maria Franco-Villoria, Rosaria Ignaccolo. Bootstrap based uncertainty bands for prediction in functional kriging [J]. Spatial Statistics, 2017(8):130-148
[8] Cressie N. Statistics for Spatial Data [M]. Revised Edition Wiley, New York,1993.
[9] AM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. R package version 1.8.6[EB/OL]. http://CRAN.R-project.org/package=mgcv
[10] Brunsdon C, Fotheringham S, and Charlton M. "Geographically Weighted Re-gression-Modelling Spatial Non-Stationarity," [J]. The Statistician, 1998(47): 431–443.