2017 2nd International Conference on Advances in Management Engineering and Information Technology (AMEIT 2017)
ISBN: 978-1-60595-457-8

A Classification Approach of Neural Networks for Credit Card Default Detection

Bu-yun ZHANG, Shi-wei LI and Chuan-tao YIN*

37th Xueyuan Road, Beihang University, Haidian District, Beijing, China

*Corresponding author

Keywords: Data classification, Deep learning, Neural network, Finance.

Abstract. By using a neural network system, which is more complex and sophisticated than a simple linear regression model, the classification simulation shall have a better performance. The data was extracted from UCI machine learning Lab which represents Taiwan credit card defaults in 2005 and their previous payment histories. This article mainly tried to determine the factors that strongly predict the future default probability with neural network comparing with linear model shows the advantage of deep learning in financial area.

Introduction

With the progress of computer technology, data mining has become an important branch of data analysis research. Among them, as the basic operation of data mining, data classification is one of the key technologies in the field of data mining. Data classification is a technology that uses the characteristics of data structure to construct a classifier, in order to classify unknown categories of data. The process of constructing classifier generally divided into two steps: training and testing. The BP neural network is one of the most basic and widely applied tool in deep learning models.

Using BP neural network, this paper sets up an analysis of the credit card default model, through the payments, credit before the default, and defaulters' personal situations, trying to build a deep learning model to predict the individual default probability. By comparing with usual linear regression model, this article shows the advantage of neural networks in financial data treatment and an improved accuracy.

Research Design

Linear Regression Model

Linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables.

$$y_i = \beta_0 \bullet 1 + \beta_1 \bullet x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i = x_i^T \bullet \beta + \varepsilon_i \quad (1)$$

The OLS method minimizes the sum of squared residuals, and leads to a closed-form expression for the estimated value of the unknown parameter β:

$$\hat{\beta} = (X^T X)^{-1} X^T y = (\sum x_i x_i^T)^{-1} (\sum x_i y_i) \quad (2)$$

Support Vector Machine

SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

We want to find the "maximum-margin hyperplane" that divides the group of points, which is defined so that the distance between the hyperplane and the nearest point x_i from either group is maximized.

Neural Network model

Artificial neural networks (ANNs) or connectionist systems are a computational model used in computer science and other research disciplines, which is based on a large collection of simple neural units (artificial neurons), loosely analogous to the observed behavior of a biological brain's axons [4][5]. Each neural unit is connected with many others, and links can enhance or inhibit the activation state of adjoining neural units. Neural networks typically consist of multiple layers or a cube design, and the signal path traverses from the first (input), to the last (output) layer of neural units.

Algorithms

Chi-square Test

It is widely used to verify if there really exists relationship between the two or more factors, which reflects on the result of the y_i when x_i changing. Here for each factor, we use the chi-squared test to verify if the factor really gives an influence on the future anticipation of the default.

Table 1. 2-Type Classification Chi-square Test.

	Future +	Future -	Sum
Factor +	A	B	A+B
Factor -	C	D	C+D
Sum	A+C	B+D	N

Hypotheses:
Ho: Factor has no effect on future default
H1: Factor has effect on future default
Degree of confidence α= 0.05

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \tag{3}$$

Here with the four-example test it becomes

$$\chi^2 = \frac{(a - \frac{a+b}{N}\bullet(a+c))^2}{\frac{a+b}{N}\bullet(a+c)} + \frac{(c - \frac{c+d}{N}\bullet(a+c))^2}{\frac{c+d}{N}\bullet(a+c)} + \frac{(b - \frac{a+b}{N}\bullet(b+d))^2}{\frac{a+b}{N}\bullet(b+d)} + \frac{(d - \frac{c+d}{N}\bullet(b+d))^2}{\frac{c+d}{N}\bullet(b+d)} \tag{4}$$

Stochastic gradient descent

Stochastic gradient descent, known as incremental gradient descent, is a stochastic approximation of the gradient descent optimization method for minimizing an objective function that is written as a sum of differentiable functions.

The estimation is expressed as below where θ_j refer to coefficients of each factor. So the model is as $h(\theta) = \sum_{j=0}^{n} \theta_j x_j$ where the lost function is therefore defined as

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - h_\theta(x_i))^2 \tag{5}$$

For each test in the function, the lost is:

$$lost(\theta, (x^i, y^i)) = \frac{1}{2}(y^i - h_\theta(x^i))^2 \tag{6}$$

and we update each time for a sample used to test the coefficients of the factors :

$$\theta_j^{'} = \theta_j + (y^i - h_\theta(x^i))x_j^i \tag{7}$$

Activation Function

The Activate function that we use here is the Sigmoid function:

$$S(x) = \frac{1}{1+e^{-x}} \tag{8}$$

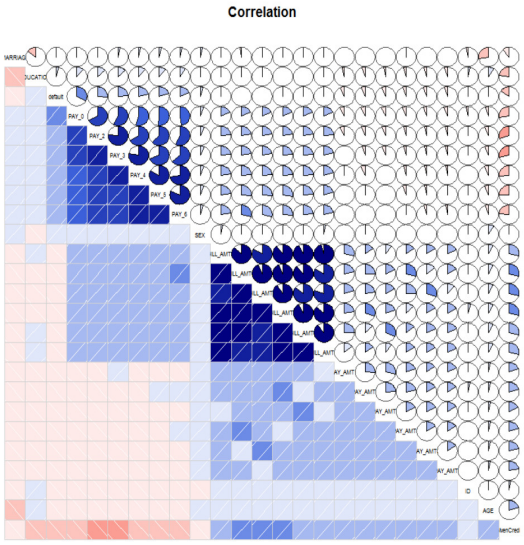

Figure 1. Linear Regression Analysis.

```
Call:
lm(formula = default ~ ., data = lineardata)

Residuals:
    Min      1Q  Median      3Q     Max
-1.19943 -0.24536 -0.18891  0.03479  1.22646

Coefficients:
             Estimate   Std.Error  t value  Pr(>|t|)
(Intercept)  2.616e-01  3.798e-03  68.875   < 2e-16 ***
GivenCredit -1.304e-07  1.965e-08  -6.633   3.34e-11 ***
PAY_0        9.332e-02  2.770e-03  33.693   < 2e-16 ***
PAY_2        1.434e-02  3.325e-03  4.313    1.62e-05 ***
PAY_3        1.099e-02  3.589e-03  3.062     0.00220 **
PAY_4        2.432e-03  3.984e-03  0.610    0.54160
PAY_5        4.420e-03  4.311e-03  1.025    0.30525
PAY_6       -7.770e-04  3.495e-03 -0.222    0.82409
PAY_AMT1    -7.176e-07  1.490e-07 -4.816    1.47e-06 ***
PAY_AMT2    -2.042e-07  1.063e-07 -1.922    0.05458 .
PAY_AMT3    -1.807e-07  1.394e-07 -1.296    0.19490
PAY_AMT4    -3.682e-07  1.536e-07 -2.397    0.01654 *
PAY_AMT5    -4.533e-07  1.548e-07 -2.928    0.00342 **
PAY_AMT6    -3.088e-07  1.333e-07 -2.317    0.02048 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3904 on 29986 degrees of freedom
Multiple R-squared:  0.1158,    Adjusted R-squared:  0.1154
F-statistic: 302.2 on 13 and 29986 DF,  p-value: < 2.2e-16
```

Figure 2. Correlation Diagram.

The activation function of a node defines the output of that node given an input or set of inputs, it is the nonlinear activation function that allows such networks to compute nontrivial problems using only a small number of nodes.

Simulation

Data selection

Here we select the card credit default data from UCI machine learning laboratory with 22 factors as right:

These factors include age, education, marriage, and financial account characteristics such as credit points.

PAY refers to account duly payment in recent months, BILL_AMT refers to amount of bills received recent months, and PAY_AMT refers to payment amount recently.

Linear Regression

After running the linear regression test, the result is shown above:

We can see by standard error that earlier account payments situation mainly decides the future default probability. If the precedent stat is already in debt, and the client is more probably to fail again on the duly payment. The residual standard error is 0.39, which is 40% of the difference between default status, and with multiple R-squared coefficient 0.1158, only 12% of the default is explained by the factors used.

Chi-square Test

By applying the Chi-square test, we found that though with a correlation coefficient less than 0.1, the four factors bellow still have importance influence on the default result, which means they should be counted as factors in further neural network research.

Table 2. Chi-square Test result.

Sex	Education	Marriage	Age
47.905	156.7958	30.4456	78.8806

Support Vector Machine Simulation

Computing the (soft-margin) SVM classifier amounts to minimizing an expression of the form below as

$$[\frac{1}{n}\sum_{i=1}^{n}\max(0,1-y_i(w\bullet x_i+b))]+\lambda\|w\|^2 \tag{9}$$

Minimizing this form can be rewritten as a constrained optimization problem with a differentiable objective function in the following way.

For each $i\in\{1,2,...,n\}$, we introduce a variable $\varsigma_i=\max(0,1-y_i(wx_i+b))$. Thus the problem of optimization becomes:

$$\min(\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\varsigma_i+\lambda\|w\|^2)$$
$$y_i(w\bullet x_i+b)\geq 1-\varsigma_i, and\varsigma_i>0 \tag{10}$$

Neural Network Simulation

Before the sample data is used, we firstly apply the preprocessing with quantitative method, including missing value cleaning and data standardizing. Our experiment mainly focuses on the difference on the performance of neural network and SVM, and the regression model, while using cross validation method to eliminate the statistical error.

Table 3. Prediction Accuracy Comparison.

Classifier	Linear Regression		SVM	Neural Network		
Type	alpha=0.2 knew	alpha=0.2 unknown	Average	Best	Worst	Average
A	**51.432%**	**16.456%**	**52.583%**	**54.371%**	**50.008%**	**52.448%**
B	**48.568%**	**83.544%**	**47.417%**	**45.629%**	**49.992%**	**47.552%**
C	**86.359%**	**98.254%**	**85.904%**	**87.218%**	**86.493%**	**87.199%**
D	**13.641%**	**1.746%**	**14.096%**	**12.782%**	**13.507%**	**12.801%**
Total Accuracy	**78.633%**	**80.160%**	**78.530%**	**79.967%**	**78.793%**	**79.487%**

A: Default predicted; B: Default unpredicted; C: Duly payment predicted; D: Duly payment wrong classified as default; alpha = approximate default rate.

Conclusion

The deep learning algorithm of the neural network is discussed in this paper on the financial data of simple applications, using nonlinear model to better optimize the existing data, comparing with the traditional regression model. It shows that neural networks have more powerful processing ability, that is suitable for large, complex financial data. Although the actual results of optimization are quite limited, there is no doubt that the use of deep learning can give a better explanation for the financial data in the data distribution. The future research on financial data can be applied onto more complex neural networks to obtain better results than usual BP networks.

References

[1] Cui L Q, Liu W J, Bao M Y. Research of data classification based on neural networks[J]. J of Liaoning Technical University, 2004, 23(4):507-509.

[2] Feng J, Starzyk J, Qiu W H. A classification of approach of neural networks based on entropy for financial data. A of Control and Decision, 2012, 221-225.

[3] Yu, F, Application in financial data prediction of neural networks based on Monte-Carlo-Adaptation rule

[4] Xu, Z G, Zhang S Y. Application of Wavelet Neural Network in high frequency financial data. Mathematics in practice and theory, 2007, volume 37.

[5] Nan X G, Meng W D. Financial alert model based on BP neural networks. Modern management science, 2007.

[6] I-Cheng The, Che-hui Lien, The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications 36 (2009) 2473-2480.