Application of BP Neural Network Algorithm in Classification of Red Wine Grades

Sheng Gao, Jie Fang and Ning Miao

Department of Information Science and Technology, Pearl River College of Tianjin University of Finance and Economics, Faculty of Data Engineering, Xiangrui street, Jingjin new town, Baodi District, Tianjin, China
Email: gslypswu@163.com

Abstract. The rapid development of society promotes the improvement of people's living standard. Red wine has entered thousands of households, at the same time, high-quality red wine is more and more loved by people. As an important standard to distinguish the quality of red wine, the grade of red wine plays an important reference role in the evaluation of red wine. Therefore, accurate and efficient classification of red wine grade is particularly important. In this study, the BP neural network model is established, and the MIV algorithm is introduced to screen the chemical properties of red wine, further optimizing the BP neural network. The optimized BP neural network is applied to the classification of red wine grades to complete the efficient classification of red wine grades. The experimental results show that the correct rate of red wine classification can be effectively improved by using the red wine data screened through MIV algorithm. This meets the requirements of accuracy in classification of red wine grades.

1. Introduction
With the improvement of life quality, people pay more and more attention to obtain high quality red wine. As an important standard to distinguish the quality of red wine, red wine grade is of great significance in people's appreciation of red wine. Therefore, it is very important to study the accurate and efficient classification of red wine grade. In recent years, foreign researches on classification of red wine mainly focus on refining the composition of red wine, and classifying the grade of red wine by statistical analysis method. Feher Ioana et al. [1] took the content and region of chemical elements in red wine as the classification attribute of red wine grade, and classified the red wine grade by Fuzzy Principal Component Analysis and Linear discriminant analysis. Milovanovic Miodrag et al. [2] took the carboxylic acid level in red wine as the main attribute of red wine classification, and classified the red wine grades by statistical methods: principal component analysis and self-organization mapping. Most of the chemical components in red wine are the main basis attributes of red wine classification, and the accuracy of classification results is high. However, this method needs to extract the chemical components in red wine, and the experimental data is not easy to obtain. Nattane Luiza Costa et al. [3] applied the method of support vector machine in data mining to classification of red wine grades. At the same time, it was concluded that the key factors affecting red wine grades were anthocyanin content and red wine color. The method of data mining is used in classification of red wine grade, the data is easy to get and the classification accuracy is high. In China, the classification of red wine mainly focuses on the research of classification methods. Shao Guoqiang et al [4] applied Bayes method in data mining to classification of red wine grades with an accuracy of 86.98%. Liu Pan [5] improved Bayesian method and applied it to classification of red wine grades, with an accuracy of...
89.802%. The defect of Bayesian method is that it needs to input prior knowledge. In order to avoid this defect, Bi Yanliang et al. [6] improved the genetic algorithm in data mining to optimize the neural network structure and apply it to red wine classification, with an accuracy of 90.35%. The deficiency of this study is that in order to reduce the complexity of the neural network, it is necessary to reduce the dimension of red wine data, break the integrity of attributes in the source data, and affect the preparation rate of red wine classification. Based on the above experience, this study uses MIV value to screen red wine attributes, to retain the attributes in the source data as much as possible, to ensure the integrity of the data, and to improve the neural network to improve the accuracy of red wine classification.

2. Optimized BP Neural Network Model

2.1 Establishing BP Neural Network Model

Neural network is to simulate the structure and operation principle of biological nervous system. The adaptive units are combined into interconnected networks. BP neural network is a multi-layer feedforward neural network with no loop or loop inside the network. its structure includes input layer (layer 0), hidden layer (layer 1) and output layer (layer 2). BP neural network operation process includes a large quantity stage: the first stage is the forward propagation of the signal, which transmits the signal from the input layer to the output layer; the second stage is the back propagation of the error, which adjusts the weight and bias of the signal from the hidden layer to the input layer. BP nerve the network topology is shown in figure 1.

![BP neural network topology](image)

**Figure 1.** Topological structure of BP neural network.

BP first phase of the neural network is implemented as follows:

I. The input layer receives external information and quantifies it

\[ V = (V_1, V_2, \ldots, V_n), V_j = (W_{j1}^1, W_{j2}^1, \ldots, W_{jn}^1), \]

\[ W_{jn} \] denotes weight value of the \( n \) attribute of the \( j \) variable representing the discrepancy.

II. Enter the \( W_{jn} \) into the next layer, get the linear results, and activate the neurons in the lower layer, the activation formula is following.

\[ a_j^1 = f(z_j^1) \]

\[ z_j^1 = \sum_k W_{jk}^1 a_k^{1-1} + b_j^1 \]
\( a_j^L \) represents the output of the j neuron in layer L, \( W_{jk}^L \) denotes the signal weight of the k neuron in layer (L-1) points to the first neuron in layer L. \( \delta_j^L \) represents the deviation of the j neuron in layer L, and \( f \) represents the activation function. In order to speed up the operation of the BP neural network while avoiding the non-convergence of the network, this study selected ReLU as the activation function and \( \lambda \) as the gradient descent parameter, such as formula (4).

\[
 f(x) = \begin{cases} 
 x & x > 0 \\
 \lambda x & x \leq 0
\end{cases}
\]  

(4)

III. Repeat step (2) until each layer network completes the output while outputting the layer output cost function values to assess the loss or risk of BP neural networks. Cost function of m sample data such as formula (5).

\[
 L(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_0(x^i) - y^i)^2
\]  

(5)

\[
 h_0(x^i) = \theta_0 + \theta_1 x
\]  

(6)

Where \( \theta_0 \) and \( \theta_1 \) are the parameters of m sample data to synthesize straight lines \( h_0(x^i) \) represent the actual output of the i sample data through the BP neural network, and the expected output of the i data. and the smaller the cost function value, the more successful the BP neural network is.

To optimize the structure of BP neural network and introduce the idea of back propagation, by repeatedly adjusting the weight \( W \) and bias values \( b \) in each layer of BP neural network, the cost function value of output layer output is minimized. BP specific steps of neural network back propagation are as follows:

I. Calculation of neuron error in input layer.

\[
 \delta_j^L = \frac{\partial L(\theta_0, \theta_1)}{\partial z_j^L} \sigma \left( z_j^L \right)
\]  

(7)

\( \delta_j^L \) represents the error of the j neuron in layer L, \( z_j^L \) the linear output of the first neuron in layer L.

II. Calculate the error of each neuron in the hidden layer.

\[
 \delta_j^L = \sum_k W_{kj}^{L+1} \delta_k^{L+1} \sigma \left( z_j^L \right)
\]  

(8)

III. Calculate the change rate of each neuron weight and bias in each layer network.

\[
 \frac{\partial L(\theta_0, \theta_1)}{\partial b_j^L} = \delta_j^L
\]

\[
 \frac{\partial L(\theta_0, \theta_1)}{\partial W_{kj}^L} = a_{k}^{L-1} \delta_j^L
\]  

(9)

IV. Reverse update weights and bias based on gradient descent principle.

\[
 b^L \leftarrow b^L - \alpha \frac{\partial L(\theta_0, \theta_1)}{\partial b^L}
\]

\[
 W^L \leftarrow W^L - \alpha \frac{\partial L(\theta_0, \theta_1)}{\partial W^L}
\]  

(10)

\( \alpha \) is the learning step in gradient descent.

This study after repeated back propagation of the BP neural network, the BP neural network is finally adjusted to the optimal.

2.2 MIV Algorithm is Introduced to Train BP Neural Network

MIV algorithm is the best index to evaluate variable correlation in neural networks. The features of variables in input BP neural networks are screened by MIV algorithm to improve the correctness of BP neural network classification. Specific steps are as follows:

I. A BP neural network trained using raw data sets.
II. By increasing the number of features of an independent variable in the original data set by 10%, the simulation results $R_1$ formed.

III. By reducing the number of features of this independent variable by 10%, the simulation results $R_2$ formed.

IV. The average value of the difference between the $R_1$ and the $R_2$ is obtained, and the MIV value of the independent variable is obtained.

V. Repeat II, III, IV step to calculate the MIV values of all variables in the original dataset.

VI. Comparing the MIV absolute values of these variables, we construct the characteristic table of variables that affect the correct rate of network output.

VII. According to the variable feature order, the variable features with great influence on the BP neural network are selected.

3. Classification of Red Wine Grades Based on Optimized BP Neural Network Model

Based on the optimized BP neural network model, red wine classification can be divided into four steps: the first step is data preparation and preprocessing; the second step is to establish BP neural network model and train optimization; the third step is to use MIV algorithm to screen the characteristics of red wine data; the fourth step is to import the screened red wine data into the neural network model to achieve red wine classification.

3.1 Data Preparation and Preprocessing

The red wine data set is obtained by downloading the UCI database. It records the chemical composition analysis of different quality grades of red wine. There are 1599 samples in the data. Each sample contains 11 characteristic components. According to the chemical composition, they are fixed acidity, volatile acidity, citric acid, residual sugar, chloride, free sulfur dioxide, total sulfur dioxide, density, acidity sulfate, alcohol. Red wine is divided into six grades of 3 to 8. For the convenience of statistical classification, the two grades of 3 and 4 are divided into the first category, the two grades of 5 and 6 are divided into the second category, and the two grades of 7 and 8 are divided into the third category. The first 1199 samples of red wine were used as training data, and the remaining 400 samples were used as test data. In view of the above two problems, the quality and integrity of the original data can be improved through data noise reduction processing, and the original data feature dimension can be normalized through data normalization processing. The data normalization formula selected in this study is as follows:

$$\frac{x - \text{min}}{\text{min} - \text{max}} = \frac{x - \text{min}}{(\text{max} - x) + (x - \text{min})}$$

(11)

Where min is the minimum value of data feature and Max is the maximum value of data feature.

![Figure 2. Part of red wine data after pretreatment](image_url)
3.2 Establishing BP Neural Network Model and Training Optimization
According to the content of Chapter 2 of this paper, BP neural network is established. The parameters of the network are optimized by back propagation, and the training algorithm is selected at the same time. The training algorithm is to help train the neural network model. There are three alternative algorithms. The first is the Levenberg–Marquardt algorithm, which can provide numerical solutions to minimize the number nonlinearity. This algorithm can solve the problem that the inverse matrix does not exist or the initial value is too far from the local minimum by modifying the parameters. The advantages of this algorithm are short training time and weak generalization ability. The second is Bayesian regularization, which is to find a better estimation method to reduce the occurrence of over fitting. The advantage is high accuracy, the disadvantage is long training time. The third is the quantitative conjugate gradient method, which overcomes the disadvantage of slow convergence and has the advantages of small storage and high stability. Combined with the data characteristics, the first training algorithm is selected to train the neural network.

3.3 Screening Data Characteristics of Red Wine
According to the MIV algorithm introduced in 2.2 of this paper, we can filter the red wine data. By calculating the MIV value of each independent variable, it is convenient to compare the size and screen out the variables with great influence.

![Figure 3. MIV value](image)

As can be seen from the above figure, the first variable ranks in the first three, and the tenth and eleventh variables are fixed acidity, sulfate and alcohol content. The input of these three characteristic attribute variables has a greater impact on the output of the results, and the impact of the rest variables tends to be negative or small enough to be ignored. Therefore, three variables are selected from 11 variables to simulate the neural network.

3.4 Classification of Red Wine Grades
The training data are trained by BP neural network, and then the test data are predicted and classified. In this process, the best classification results are obtained by continuously adjusting the number of hidden neurons, the maximum number of training times, training accuracy, learning rate and other training parameters of BP neural network. In this paper, because the training data is divided into two groups: deleted variables and not deleted variables, the training parameters of these two groups are different after adjustment, and it is more convenient to compare the results after classification.

4. Analysis of Experimental Results

4.1 Comparison of BP Neural Network Training Results
The training diagram of BP neural network is as follows:
Figure 4. BP neural network training diagram

The training data without deleting variables and the test data without deleting variables are trained by BP neural network. On the left is the training diagram of BP neural network without deleting variables, on the right is the training diagram of BP neural network with deleting variables. It can be seen from the above figure that the training data without deleting variables are trained 18 times in BP neural network, 21 times in BP neural network from the 12th training to the best mean square deviation of 0.12711, and 0.12112 from the 15th training to the best mean square deviation of 0.12112. Because the closer the mean square deviation is to 0, the better the fitting is, so the fitting result after deleting variables is slightly better One o'clock. Moreover, from the curve fitting in the graph, we can see that the trend of data fitting after deleting variables is more stable and accurate.

4.2 Comparison of Accuracy of Classification Results

The collected training data are divided into two groups: deleted variables and retained original variables. They are trained by BP neural network, and then the trained neural network is used to predict and classify the test data. The classification results are shown in the following figure:

Table 1. Classification accuracy of optimized BP neural network.

| Data related operations          | Number of training data | Number of test data | Number of data properties | Classification accuracy |
|---------------------------------|-------------------------|---------------------|---------------------------|-------------------------|
| Before using MIV algorithm to filter | 1199                    | 400                 | 11                        | 88.75%                  |
| After using MIV algorithm to filter | 1199                    | 400                 | 4                         | 90.50%                  |

It can be seen from the above figure that the accuracy of prediction and classification for the data with deleted variables is 88.75%, and the accuracy of prediction and classification for the data with reserved variables is 90.50%. The classification results are almost the same, indicating that the fixed acidity, sulfate and alcohol in red wine have the greatest impact on the quality level.

In the classification method, the innovation is realized, and the correlation between variables and MIV are combined to improve the screening efficiency. MIV value is widely used in the process of independent variable selection in other fields, which has a good reference value. As a popular intelligent algorithm, neural network has been widely used for its superior processing ability to nonlinear system, but there are also some defects. In this paper, based on BP neural network, the MIV value of independent variables is calculated, and the variables that have little influence on the results are deleted, then it is applied to red wine quality classification. The experimental results show that the classification effect is almost the same. However, the classification accuracy still needs to be improved. In the future research, the network weight threshold and other parameters will be adjusted appropriately to improve the classification accuracy. At the same time, the theory and quality
intelligent detection which can be applied in other fields have certain practical significance for some enterprises.

5. References

[1] Feher Ioana, Magdas Dana Alina, Dehelean Adriana, Sarbu Costel. Characterization and classification of wines according to geographical origin, vintage and specific variety based on elemental content: a new chemometric approach. [J]. Journal of food science and technology, 2019, 56(12).

[2] Milovanovic Miodrag, Žeravík Jiří, Oboříl Michal, Pelcová Marta, Lacina Karel, Cakar Uros, Petrovic Aleksandar, Glatz Zdeněk, Skládal Petr. A novel method for classification of wine based on organic acids. [J]. Food chemistry, 2019, 284.

[3] Nattane Luíza Costa, Laura Andrea García Llobodanin, Inar Alves Castro, Rommel Barbosa. The use of data mining to classify Carménère and Merlot wines from Chile. [J]. Expert Systems, 2019, 36(2).

[4] Shao Guoqiang, Jie Longmei, Jiang Qingfeng. Research on the application of Bayes method in red wine classification [J]. Intelligent computer and application, 2016, 6 (03): 29-31.

[5] Liu Pan. Classification of red wine quality grades based on RBF and naive Bayes [J]. Electronic technology and software engineering, 2019 (04): 144-145.

[6] Bi Yanliang, Ning Qian, Lei Yinjie, Wang Wei. Optimization of BP neural network based on improved genetic algorithm and its application to classification of red wine quality grades [J]. Computer measurement and control, 2016, 24 (01): 226-228.