A Study on the Identification of Glass Artefacts Based on Multiple Logistic Regression

Yujie Han

College of Science, Minzu University of China, Beijing, China

ncss18093561226@163.com

Abstract. In this paper, traditional glass artefacts are classified into four categories, and the common factors obtained from factor analysis are used to solve a multivariate logistic regression model to avoid the 'pseudo-regression' caused by multiple covariance, and then predict the unknown types of glass artefacts. The BP neural network structure was also constructed based on the data, and the optimal solution was obtained after 1000 iterations using Bayesian regularization to minimize the sum of squares of the errors, provided that 10 neurons were used. The glass samples were identified to belong to the categories of unweathered high potassium, weathered lead barium, unweathered lead barium, unweathered lead barium, weathered lead barium, weathered high potassium, weathered high potassium and unweathered lead barium.

Keywords: Multiple logistic regression, BP neural network, glass classification.

1. Introduction

Later, with the opening of the Silk Road, cultural and economic exchanges between China and the West became more frequent, and foreign glass (such as bead-shaped glass made in West Asia and Egypt as ornaments) was introduced into China, and its production techniques were learned by local producers, resulting in the two being similar in appearance, but due to the regional characteristics of locally sourced materials, their chemical composition of the two beads differs due to the regional nature of the materials used [1-3]. Compared to ceramics, silk and tea, which were transmitted via the Silk Road and became famous overseas, the ancient glass and its production techniques are less well known, but as valuable physical evidence of the early trade between China and the West, it is still important to study its composition and production techniques [4]. Through the study of the chemical composition and decoration of ancient glass, it is possible to distinguish between indigenous and exotic glass varieties and to speculate on the time of production and their specific origin, which helps to understand ancient economic, cultural and technological developments [5]. There is now a collection of ancient glass excavated in China, which is known through examination to be classified as high potassium glass and lead-barium glass [6]. The two pieces of data that have been obtained are: information on the classification of the artefacts, the corresponding main chemical composition and their proportions [7]. To determine the type of unknown artefacts, it is first necessary to determine what types of artefacts there are, which can be divided into four categories in total: high potassium weathered glass, high potassium unweathered glass, lead-barium weathered glass, and lead-barium unweathered glass; to determine which category the unknown artefacts belong to, the chemical composition of the artefacts must be used to determine, so a multivariate (quadratic) logistic regression model can be established from the known data The probability of the unknown artefacts belonging to the type can be predicted by substituting the chemical composition of the artefacts into the regression equation, and the size of the probability can be used to determine the type to which they belong [8-9]. However, as there are so many chemical components, they may be linearly correlated with each other, and this inevitably leads to serious errors in the model, such as 'pseudo-regression' leading to incorrect model predictions, so the correlation between the chemical components also needs to be checked. To improve the sensitivity, the known data was further subdivided from four to six categories, i.e., the weathered lead and barium was further subdivided into three categories: weathered lead and barium unweathered, weathered lead and barium severely weathered. Then a BP neural network structure was built based on this data, and trained by an error back-propagation algorithm to iterate more accurate results [10].

2. The basic fundamental of classification models

In this paper, we use the converted artefact data to model and analyses the weathering probability of the glass artefacts, with the glass type, decoration and color as factors, and develop a logistic model as follows.

$$\ln P_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \tag{1}$$

Where a positive regression coefficient indicates a positive effect on the weathering of glass artefacts. If the regression coefficient is negative, the factor has a negative effect on the weathering of glass artefacts. The unknown parameters of model (1) were obtained using STATA as follows.

$$\ln P_i = -0.5102686 + 1.531911 x_1 + 0.0182532 x_2 - 0.0513332 x_3 \tag{2}$$

Where x_1 is the type of glass, x_2 is the grain of the glass, x_3 is the color and P_i indicates the probability of weathering of the glass [11].

Equation (1) shows that the probability of glass weathering is positively related to glass type and decoration, but negatively related to color, suggesting that there is indeed a relationship between glass weathering and glass type, decoration and color.

Different chemical compositions within the glass thus determining the type of glass; the high content of lead in lead-barium glass and the high content of potassium in high potassium glass can react with substances such as carbon dioxide and water in the surrounding environment, causing the glass to weather; the different ornamentation of the glass represents the shape and size of its contact area with the outside world, and the physical shape of the glass over the thousands of years it has been buried underneath determines the area and degree of chemical reaction. The color of glass depends on the chemical composition of its surface coloring agent and the acidity or alkalinity of the environment in which the glass was fired; for example, brown is generally colored with divalent manganese ions, dark blue with divalent cobalt ions, and divalent copper ions are green in an alkaline firing environment and blue in an acidic environment; depending on these different chemical compositions and firing environments, the degree of weathering can vary. The structure of BP neural network includes three layers: input layer, implicit layer and output layer. It is a multi-layer feed-forward network composed of non-linear units trained according to the error back-propagation algorithm. The principle of this model is to adjust the weights and thresholds of the network by continuous back-propagation so that the approximation to the function is the best and its error sum of squares is the smallest. The topology of the BP neural network is shown in Figure 1

In the neural network, the input layer to the hidden layer uses the sigmoid function as the activation function and the hidden layer to the output layer uses the purlin function as the activation function. A neural network model was developed between the 14 indicators and the dummy variables for the type of composite artefact and whether it was weathered. The input data to the model were: silica, sodium oxide, potassium oxide, calcium oxide, magnesium oxide, aluminum oxide, iron oxide, copper oxide, lead oxide, barium oxide, phosphorus pentoxide, strontium oxide, tin oxide and Sulphur dioxide; the output data were the dummy variables for the type of composite artefact and whether it was weathered. The neural network structure was first built according to the original data, i.e., the function relationship between the 14 indicators and the composite dummy variables was established, and the individual weight coefficients existed in each of the established nodes. The BP neural network is used to establish a non-linear functional relationship between the 14 indicators and the composite dummy variables. Considering that an important function of BP networks is the ability of non-linear mapping, this function is well suited to requirements such as function approximation, i.e., finding the relationship between two kinds of data, i.e., input vector X and target vector Y, and building a BP network to find the relationship between new X and the predicted vector, the figure shows the algorithm iterative operation in progress. the distribution of neurons in the BP neural network is shown in Figure 2. The distribution of neurons in the BP neural network is shown in Figure 2.

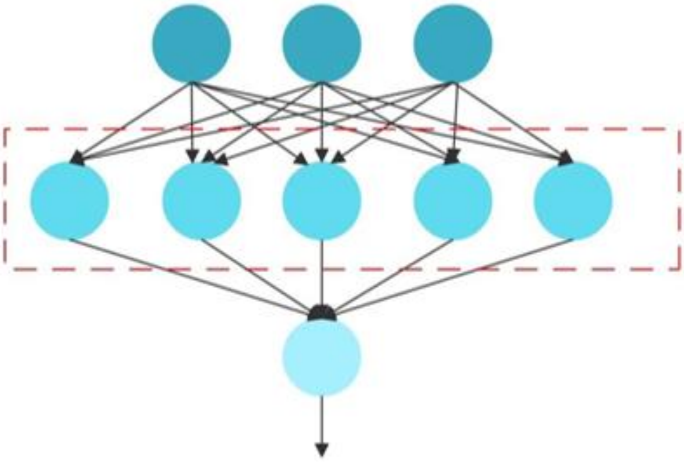

Figure 1. Topology of BP neural networks

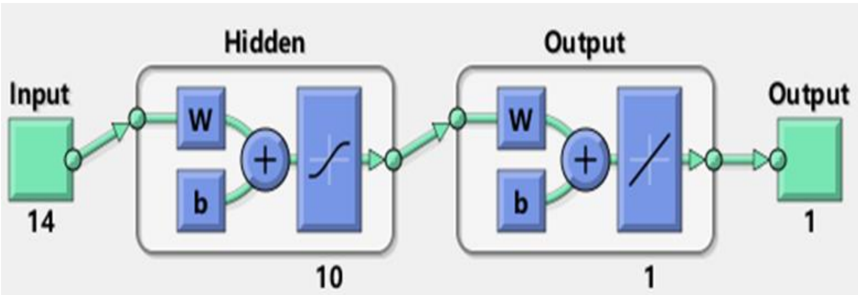

Figure 2. Schematic representation of the number of neurons in a neural network

3. Results

3.1. The establishment of simulation model

This model focuses on the classification of ancient glass objects, using cluster analysis and principal component analysis for glass objects, respectively, and classifying them to determine their main components for future glass object identification.

3.2. Analysis of experimental results

In this paper, a BP neural network is constructed using a Bayesian regularization method to obtain the optimal solution after 1000 iterations. The optimal case of the iteration values is shown in Figure 3.

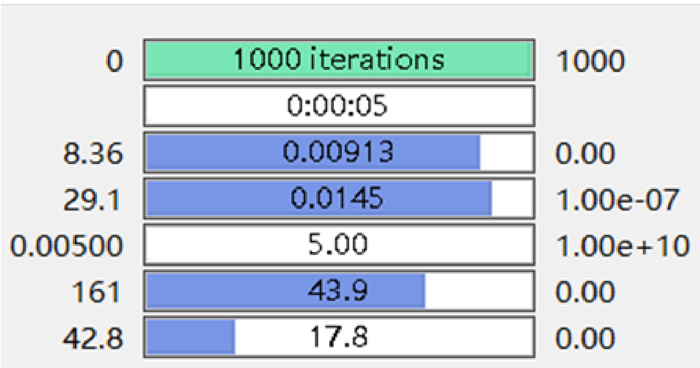

Figure 3. Schematic representation of the calculated iteration values

The number of neurons used in this calculation is 10. After 1000 training sessions (the training function is Bayesian regularized), the BP network with 10 neurons in the hidden layer approximates the function best, as it has the smallest error.

The curve fit is seen in Figure 4, the fitted curve can basically pass through the target point, indicating that the linearity of the fit is good and basically meets the fitting requirements. The fitting situation is shown in Table 1.

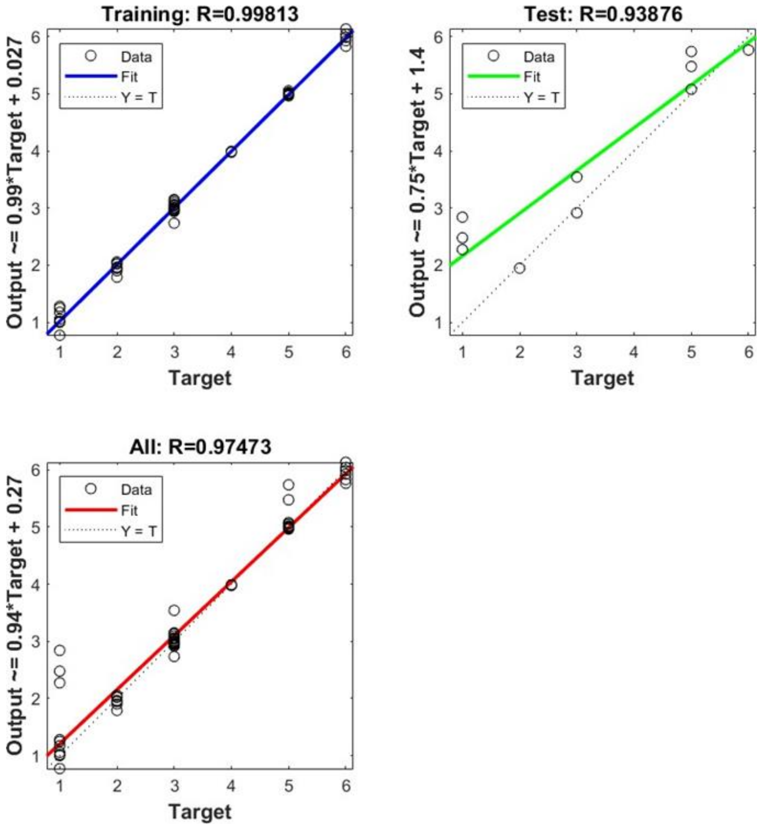

Figure 4. Regression effect of the prediction model

Table 1. Model prediction results

Heritage number	Predicted value	Prediction approximation	Types of compound variables
A1	5.3715	5	Weather-free high potash
A2	3.4208	3	Barium weathered lead
A3	-0.2511	1	Unweathered lead barium
A4	0.8884	1	Unweathered lead barium
A5	2.9295	3	Barium weathered lead
A6	5.9313	6	Weathered high potassium
A7	5.9246	6	Weathered high potassium
A8	1.1183	1	Unweathered lead barium

4. Conclusion

This study uses BP neural network for the classification of ancient glass products, and this paper can achieve effective classification for the data samples, and the logistic regression model can be used to fit the regression relationship of the model better. This paper describes the problem of classifying artefact types with a degree of accuracy and ingenuity, using a logistic model to solve the regression of discrete data with more accurate results. It provides a solid theoretical basis for the practical operation of the model when dealing with large amounts of data for prediction.

References

[1] Guo Xiaoyan. On the Historical and Cultural Factors Affecting the Development of Traditional Chinese Glass Crafts [J]. New Art, 2020, 41(11): 84-88.

[2] Chen Shuxin, Li Jingyu, Zhang Hongbin, Zhang Hui. Research on personalized multifunctional coffee table design based on BP neural network [J]. Packaging Engineering, 2022, 43(18): 247-254.

[3] Fan Y, Pei Y, Yang GD, Leng ZD, Lu WB. Peak blast vibration velocity prediction based on improved PSO-BP neural network [J]. Vibration and Shock, 2022, 41(16): 194-203+302.

[4] WANG Qing-Hua, GA Shi-Ying, HU Jian-Hua, DUANG Yuan-Hua, ZHAO Tielin. GRA-based PSO-BP neural network for oblique rolled perforated tube shape prediction [J]. Forging and pressing technology, 2022, 47(08): 88-94.

[5] Liu Yang, Cao Xueyu. Research on mechanical properties of recycled concrete prediction based on BP neural network [J]. World of Concrete, 2022, (08): 19-24.

[6] Li Canyuan, Jiao Quan, Wen Jiangli, Xun Zhicheng, Li Guangtao, Wu Di. Application of BP neural network in predicting water inflow to Miyun Reservoir [J]. Beijing Water, 2022, (04): 14-20.

[7] Ren, Xuan-Yu, Zhu, Lin-Hui. Analysis of factors influencing the willingness to use Internet healthcare based on multivariate logistic model [J]. Journal of Economic Research, 2022, (20): 46-49.

[8] Li Chunjiang,Liu Jun,Yu Ganyu. Analysis of factors influencing college students' participation in the military in higher education institutions--based on multiple logistic regression model [J]. China Student Employment, 2022, (07): 16-23.

[9] Li Haonan. Application of multivariate ordered logistic model in predicting the number of auto insurance claims [J]. Journal of the Academy of Insurance Professions, 2021, 35(06): 39-45.

[10] ZHANG Jing, WANG Wei, ZHOU Lei. Analysis of factors influencing grassland ecological protection subsidy incentive policy based on multivariate ordered logistic model--Ningxia as an example [J]. Ningxia Agriculture and Forestry Science and Technology, 2021, 62(07): 51-56.

[11] Liu Yupeng, Zhang Yanxin, Ye Zihan, Li Tong. Analysis of factors influencing dairy products consumption in Hebei Province based on multivariate logistic regression model [J]. Journal of Hebei Agricultural University (Social Science Edition), 2021, 23(03): 60-67.