

Comparative study on chemical composition of ancient glass based on machine learning and deep learning

Dan Bai *

Criminal Justice College of Ecupl, East China University of Political Science and Law, Shanghai, China, 201620

* Corresponding Author Email: 2031365470@qq.com

Abstract. In this paper, a BP neural network was established based on the deep learning framework to determine the category of ancient glass under the influence of weathering. Compared with random forest and support vector machine, the accuracy reached 100%, in which random forest and BP neural network predicted the same results. The convolutional neural network model was proposed to solve the classification problem, and the applicability of the convolutional algorithm in one-dimensional data was verified. Then, through the visualization of the chemical composition correlation of different categories of glass cultural relics, it is concluded that high-potassium glass has a strong correlation with silica, and lead-barium glass has a strong correlation with lead oxide. For the difference of chemical composition correlation between different types of glass cultural relics, the conclusion that silica in high-potassium glass is negatively correlated with other components was obtained by constructing knowledge map analysis.

Keywords: weathering degree of glass, Spearman correlation coefficient, random forest, knowledge map

1. Introduction

China's glass has been making culture for thousands of years, according to archaeological data show that the earliest appeared in the late Spring and Autumn period. After absorbing Western technology, Chinese ancient glass is made from local materials. The main raw material is quartz sand. Because of its high melting point, flux of different chemical composition will be added in smelting to reduce the melting temperature. The processed glass is highly susceptible to moisture and air in the storage and transportation process, which leads to the weathering of the glass surface. Weathering through a series of physical and chemical reactions leads to changes in the proportion of its chemical composition. However, the weathering of glass and the change of the proportion of chemical composition have a great influence on the judgment of its type. With the development of glass industry, the judgment of its type has attracted more and more attention. The research on the weathering of glass and the content of chemical composition has become a development trend [1].

In this paper, we first analyze the chemical composition of glass relics of unknown categories and identify their types. In this paper, we believe that the existing small sample data can be used to establish classification model fitting through machine learning. Since there are many features involved in the classification of glass artifacts and the correlation of some chemical components is low, it is necessary to extract the features. In the experimental stage, this paper will compare the training and testing results of random forest, support vector machine, BP neural network and the classification results of unknown categories of glass artifacts, and evaluate their sensitivity combined with statistical rules. Next, for different categories of glass cultural relics, the correlation relationship between their chemical components is analyzed, and the differences between the correlation relationships are compared [2]. In the analysis stage, this paper believes that the correlation between individual chemical components can be analyzed first, and then the correlation relationship of the same chemical components can be summarized.

2. Training and testing of the model

The data were trained and tested using BP neural network (BP-MLP), random Forest (RF) and support vector machine (SVC) classification models. The fitting degree of each model is shown in figure 1[3].

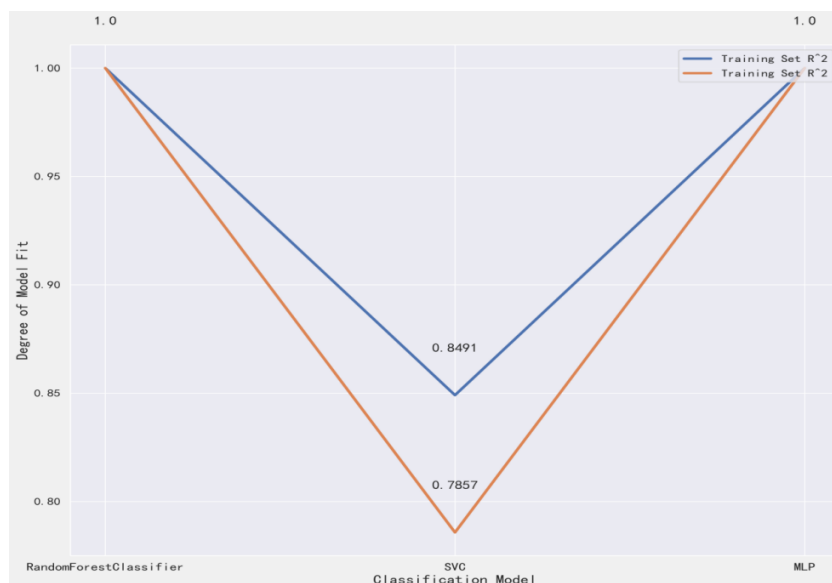


Figure 1. Comparison of fit degree of three classification models

As can be seen from Figure 1, in small sample data, the accuracy of random forest and BP neural network reaches 100% in both training and testing sets, while the accuracy of SVM training and testing sets is 84.91% and 78.57%, respectively. In solving the classification problem of this question, the classification effect of SVM is significantly lower than that of other two models. Therefore, random forest and BP neural network with 100% accuracy are selected to predict the unclassified glass relics data in this paper[4].

The random forest model is a classification model trained and predicted by multiple decision tree models. The decision rule of any decision tree is shown in figure 2[5].

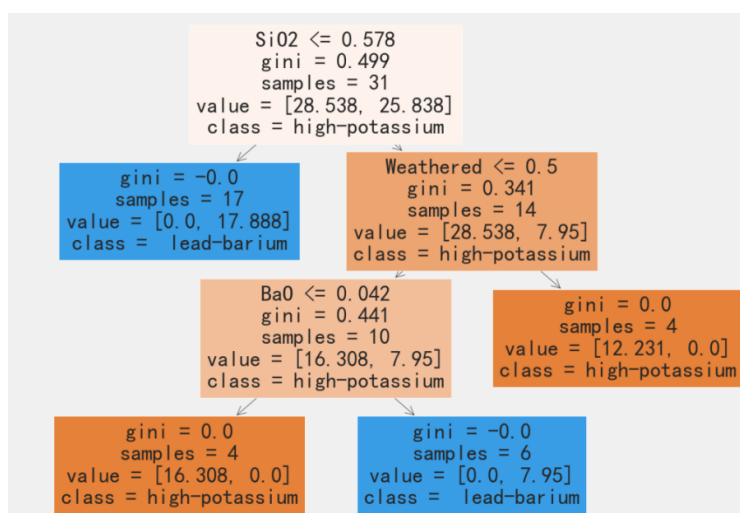


Figure 2. Decision rules for selecting any decision tree

In this paper, the weight difference between the features of the small sample dataset is obvious, so the decision tree can effectively classify under the dichotomous classification task. However, the neural network model, due to the few features of the dataset and the small number of samples, does not show great advantages. However, in terms of model optimization, the BP neural network with high accuracy is still optimized in this paper, and the gradient is trimmed through the model design

of the deep learning custom layer to avoid gradient explosion. However, it is a pity that the optimization effect is difficult to reflect on limited data sets. In reality, the neural network model is better than the rule-based decision tree model and random forest model for more categories of glass products and more data sets[6].

3. Prediction results of the model

In terms of prediction, the prediction results of the three models are shown in Table 1, 1 and 0 represent high-potassium glass, 1 represents lead-barium glass; Are the prediction results of 8 unclassified glass relics from A1 to A8 in turn [7].

Table 1. Prediction results of each model

Model prediction	Results
RandomForest (RandomForestClassifier)	[0 1 1 1 1 0 0 1]
Support Vector Machine (SVC)	[1 1 1 1 1 1 1 1]
BP Neural Network (BP-MLP)	[0 1 1 1 1 0 0 1]

It can be seen from Table 1 that the random forest model and BP neural network model with 100% accuracy on the test set and training set both predict the same result.

The robustness of BP Neural network and random forest has been compared above, but Convolutional Neural Networks (CNN) can also be used for accelerated learning of chemical components with certain correlation. The detailed model information is shown in Table 2[8].

Table 2. Model details

Layer (type)	Output Shape	Param #
Conv1d-1	[5, 16, 14]	48
ReLU-2	[5, 16, 14]	0
MaxPool1d-3	[5, 16, 7]	0
Conv1d-4	[5, 32, 6]	1056
ReLU-5	[5, 32, 6]	0
MaxPool1d-6	[5, 32, 1]	0
Flatten-7	[5,32]	0
Linear-8	[5, 2]	66

Two convolutional layers and two pooling layers are used for calculation, and the accuracy is also percentage, as shown in Figure 3[9].

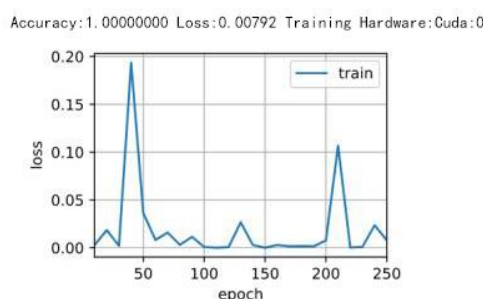


Figure 3. Prediction results using two convolution layers and two pooling layers

4. Data analysis

4.1. Data correlation analysis

After the above comparative experiments, it can be obtained that the neural network still maintains a high accuracy after removing lead oxide (PbO) and silicon dioxide (SiO₂), which have a high

contribution of principal components. Although the stability of the prediction results is decreased, it can still show the connection between the other features. In the comparison experiment of classification model, principal component analysis has been carried out on each element, so in the solution of this chapter, this paper is more inclined to show the relationship between chemical elements more clearly. Due to the small number of features in the dataset, the correlation between the observations of Spearman's coefficient can be calculated pairwise for each chemical component. In addition, due to the different changes in the content of each chemical component of glass relics in different categories, the correlation and positive and negative are also different. In the data analysis, this paper will divide the glass into high potassium type and lead and barium type according to the type of glass [10].

Firstly, this paper draws the correlation thermal map between the chemical components of high-potassium glass and lead-barium glass for visual analysis, and the results are shown in Figure 4.



Figure 4. Correlation thermograms between chemical components of high-potassium glass (left) and lead-barium glass (right)

Secondly, in order to be more intuitive to show the changes of the correlation coefficient of chemical composition characteristics, respectively, the high potassium type and lead barium glass in the chemical composition and the rest of the 13 chemical composition the spearman correlation coefficient between the plot of for, for a total of 28 pieces of the combined image correlation coefficient curves of the chemical composition, as shown in figure 5, because the space is limited, Due to limited space, only two curve images are shown here.

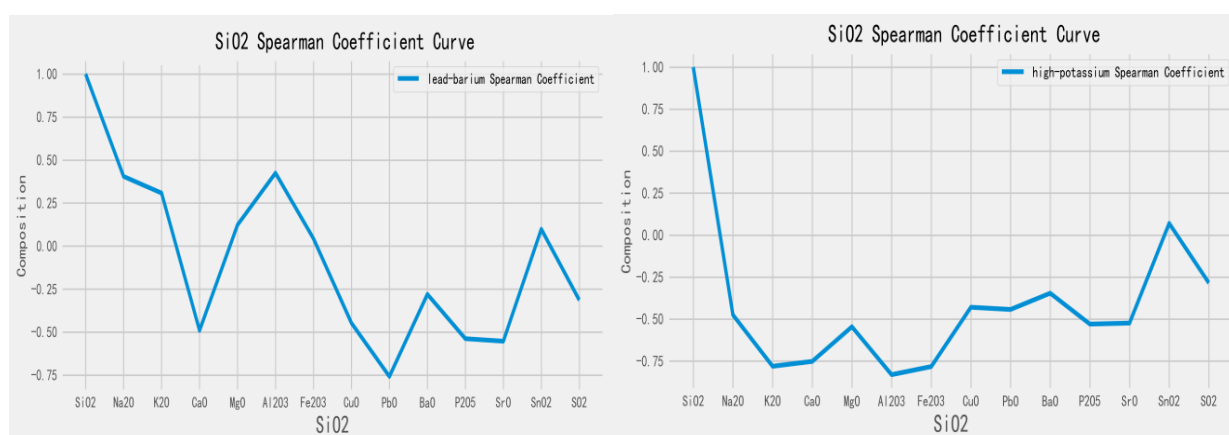


Figure 5. Spearman coefficient curves of silica (SiO₂) and 13 chemical compositions for high-potassium type glass (left) and lead-barium type glass (right)

In order to analyze the correlation relationship, this paper combined Spearman correlation coefficient thermogram and curve to analyze the chemical composition of different types of glass, and obtained the following results:

High-potassium type glass:

(1) As silica (SiO_2) with the largest proportion before and after weathering, it can be directly observed by visual analysis that its correlation with potassium oxide, aluminum oxide, calcium oxide three chemical components is extremely high and negative correlation, and the Spearman correlation coefficient is above 80%;

(2) The correlation between copper oxide and the other 13 chemical components was low;

(3) All the other chemical components mentioned above were correlated with the other chemical components to varying degrees. For example, iron trioxide and phosphorus pentoxide, potassium oxide and calcium oxide, selenium oxide and phosphorus pentoxide have relatively higher correlation, and the Spearman coefficient between them is between 70%-80%.

Lead-barium type glass:

(1) Relatively speaking, the correlation between copper oxide and barium oxide is the highest, the correlation coefficient is 0.72; The correlation between barium oxide and silica was second, the correlation coefficient was 0.63.

(2) From the analysis of thermal map, it can be seen intuitively that the correlation between the chemical components of this type of glass is generally not high except for the chemical components presented above.

4.2. Data difference analysis

Knowledge map is to use the visualization of map image showing the data of the relationship between the semantic web, glass cultural relics for the purpose of this article research data, can be divided according to the correlation, will be higher than the positive correlation coefficient of average and below average as a strong correlation between the negative correlation coefficient, by way of knowledge map display, as shown in figure 6.

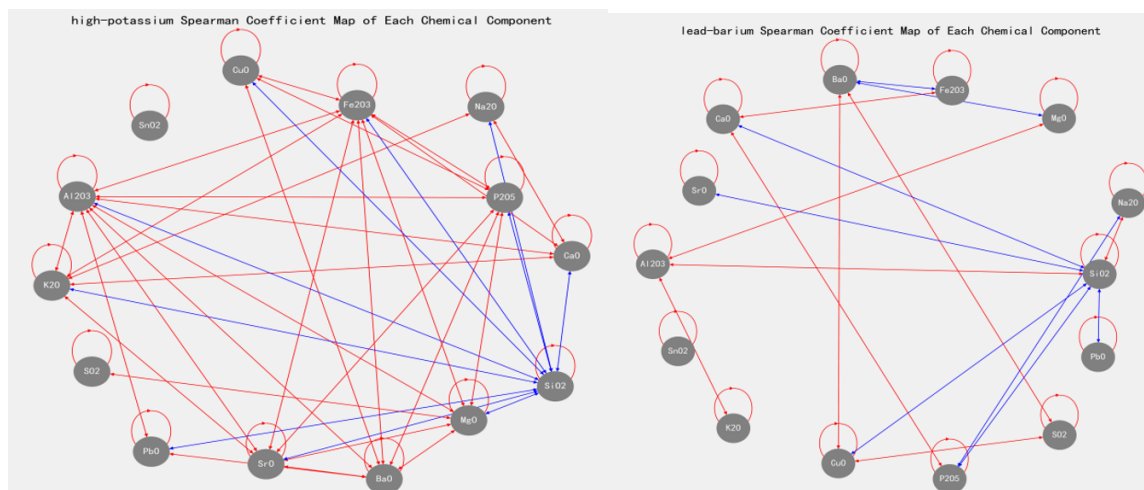


Figure 6. Spearman correlation coefficient knowledge map of each chemical composition of high-potassium type glass(left) and lead-barium type glass(right)

Based on the advantages and applicability of knowledge graph, this paper constructs the knowledge graph of different types of glass chemical elements to conduct the difference analysis of problem 4, as shown in Figure 6. The blue line represents the negative correlation of nodes at both ends and is less than the mean value of the negative correlation coefficient, while the red line represents the positive correlation of nodes at both ends and is greater than the mean value of the positive correlation coefficient.

The difference of the relationship between the chemical components of the two types of glass can be reflected by the number and type of connections between the chemical elements at each node of

the map. Through the analysis of the knowledge maps of the two types of glass, it can be found that the chemical components negatively correlated with SiO₂ can be obviously mined from the map of high potassium, and there is no chemical component positively correlated with SiO₂.

The advantage of the knowledge map is that it can show the corresponding relationship based on the demand. Therefore, the complexity of the graph can be adjusted by limiting the threshold of correlation coefficient. As shown in Figure 7, by setting the threshold values of positive and negative correlation coefficients to 0.6 and -0.6, chemical composition combinations with strong correlations can be quickly found. However, the chemical composition with no edges connected to it indicates that its changes are relatively independent.

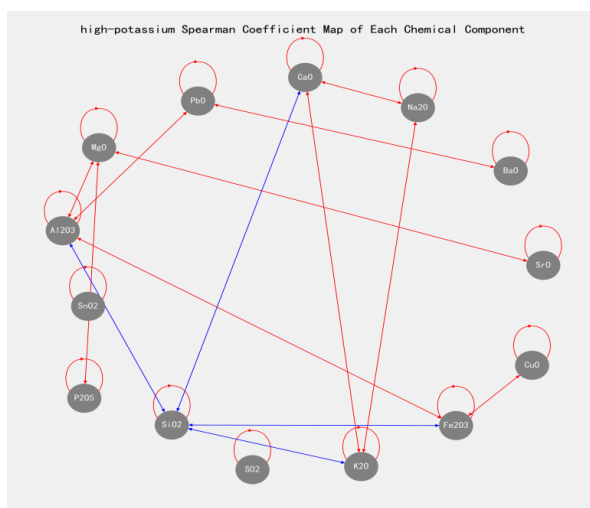


Figure 7. Spearman coefficient knowledge map of chemical composition of high-potassium type glass after threshold change

5. Conclusion

In order to determine the category of ancient glass under the influence of weathering, this paper establishes BP neural network based on deep learning framework. Compared with random forest and support vector machine, the accuracy reached 100%, and both random forest and BP neural network predicted the same results. The convolutional neural network model was proposed to solve the classification problem and verify the applicability of the convolutional algorithm in one-dimensional data. Then, through the visualization of the chemical composition correlation of different categories of glass cultural relics, it is concluded that high-potassium glass has a strong correlation with silica, and lead-barium glass has a strong correlation with lead oxide. For the difference of chemical composition correlation between different types of glass cultural relics, the conclusion that silica in high-potassium glass is negatively correlated with other components was obtained by constructing knowledge map analysis.

References

- [1] Frabasil Joaquin, Rivas Dana Velasquez, Bertot Gustavo M., Gualco Luciana, Ricardo Carabajal, Carozza Patricia, Gaggioli Daniela, Colmenares Gabriela, Sokn Silvia. Correlation analysis between lysosomal enzyme activities and the different types of leukocytes in dried blood spots[J]. Molecular Genetics and Metabolism, 2022, 135(2).
- [2] Lu Yier, Ye Chenyang, Yuan Ying. Phenotypic characteristics and T cell receptor properties in melanoma: deciphering the correlation at single-cell resolution[J]. Signal Transduction and Targeted Therapy, 2022, 7(1).

- [3] G.E. X, Zhou A, Chen J. POSB261 Cancer Patients' Willingness to Pay per Quality-Adjusted Life Year in ASIA-Pacific: A Systematic Review and Correlation Analysis with Population Metrics[J]. Value in Health, 2022, 25(1S).
- [4] Walker Trent D, Kaffenberger Benjamin H. The Adaptation of Commonly Available Academic Statistical Software to Illustrate Skin Disease Through a Heat Map. [J]. Journal of the American Academy of Dermatology, 2021, 7(1).
- [5] Yokoyama Akihito, Okazaki Hiroshi, Makita Naoyuki, Fukui Ayako, Piao Yi, Arita Yoshifumi, Itoh Yohji, Tashiro Naoki. Regional differences in the incidence of asthma exacerbations in Japan: A heat map analysis of healthcare insurance claims data. [J]. Allergology international: official journal of the Japanese Society of Allergology, 2021, 71(1).
- [6] Sahani Mrutyunjaya, Rout Susanta Kumar, Dash Pradipta Kishore. FPGA implementation of epileptic seizure detection using semisupervised reduced deep convolutional neural network [J]. Applied Soft Computing Journal, 2021, 11, 10.
- [7] Tan Jiannan, Oyekan John. Attention Augmented Convolutional Neural Network for acoustics based machine state estimation[J]. Applied Soft Computing Journal, 2021, 1, 10.
- [8] Xiao Yong-Liang, Li Sikun, Situ Guohai, You Zhisheng. Unitary learning for diffractive deep neural network[J]. Optics and Lasers in Engineering, 2021, 13, 9.
- [9] Kamran Iqra, Naz Saeeda, Razzak Imran, Imran Muhammad. Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease[J]. Future Generation Computer Systems, 2021, 117, 11.
- [10] Lee C.K.M., Ng Kam K.H., Chen Chun-Hsien, Lau H.C.W., Chung S.Y., Tsoi Tiffany. American sign language recognition and training method with recurrent neural network [J]. Expert Systems with Applications, 2021, 167, 12.