ISSN 2616-5880 Vol. 4, Issue 2: 1-5, DOI: 10.25236/AJMC.2023.040201

A study of the composition of ancient glass based on a principal component analysis model

Tianyu Zhu*

College of Physics and Electromechanical Engineering, Hubei University of Education, Wuhan, China
**Corresponding author: 869354913@qq.com*

Abstract: *In the early years of the country, China used the Silk Road to trade glass, learn how to make it and make it. As a result, the glass made in China was similar in appearance to exotic glass, but differed in chemical composition. The main objective of this paper is to develop a compositional analysis and identification model for ancient glass objects, which can be used to study the degree of surface weathering of glass objects and its relationship to glass type, decoration and color, as well as the types of glass objects that belong to unknown categories. The chemical composition content of the pre-weathering artifacts was further predicted and the correlation between the chemical components was analyzed. This paper introduces a glass classification method based on principal component analysis and a Gaussian mixture model with the expectation maximization principle for the identification of glass artifacts. Then a confusion matrix was introduced to perform sensitivity analysis on the model, and the correct and recall rates were 0.911 and 0.844, respectively, with good scores.*

Keywords: *principal component analysis, gaussian mixture model, glass classification*

1. Introduction

The Silk Road was an important route for economic and cultural exchange between East and West, which strongly contributed to the economic development of China and the development process of world civilization, and was of great significance in terms of ethnic integration and artistic and cultural exchange [1-2]. Ancient glass is an important historical source for exploring the Silk Road. Ancient glass is susceptible to weathering from environmental influences, and the burial environment is usually harsh, with glass objects of different compositions subject to weathering to varying degrees and complex geographical factors, and the issue of the chemical composition and type of artefact before weathering has been the focus of several domestic research institutes. The type of glass, glass decoration, glass color and other factors have different degrees of influence on weathering, and when weathering is severe it is not even possible to identify the type of glass artefact, which seriously affects the study of glass artefacts and their historical significance, and the detection and analysis of the chemical composition of weathered glass artefacts has become an important issue to be resolved. The study of the types of cultural relics and the model of differentiated types can prevent further weathering of cultural relics, which has a significant role and important theoretical and practical significance for the excavation and conservation of cultural relics in China. The analysis of the chemical composition of glass objects excavated in Sichuan Province [2], and the resulting analysis of the chemical composition of glass samples can be roughly dated to the period to which they belong. This method, based on the analysis of the chemical composition of glass, not only provides a valuable scientific basis for the dating of ancient Chinese glass, but also points out the similarities and differences between Chinese glass and Central Asian and Western Asian glass, providing scientific data for the study of East-West exchanges in ancient times. A collection of ancient artefact data exists which needs to be analyzed for its composition and identified for its type [3-4].

2. The basic fundamental of classification models

The main purpose of principal component analysis is the desire to use a small number of variables to explain most of the variables in the information, and for all the variables originally presented, the variables with high correlation are transformed into variables that are independent of each other, and these new variables maintain the original information as much as possible in terms of reflecting the information on the subject. And these independent variables are called principal components and are used to explain the composite indicators [5].

ISSN 2616-5880 Vol. 4, Issue 2: 1-5, DOI: 10.25236/AJMC.2023.040201

PCA is a common method of multivariate statistical analysis which not only reduces the dimensionality of the matrix, but also allows the analysis of various types of samples through principal component score plots. Based on the currently available data for the available sample data, it can be found that silica is the most widely distributed, followed by lead oxide and barium oxide, which have their own characteristic areas, and the rest of the chemical elements are low, probably due to the fact that the metal is very reactive and easily loses electrons to react with oxygen, as well as disappearing in another form [6-7].

In order to classify the glass types (high potassium, lead barium) accurately, GMM was used to discriminate the glass types. Due to the large amount of information on the original chemical elemental components of the data provided by GC-MC and the small correlation between some chemical elements and the elemental characteristics of the glass, PCA was first used to extract valid information from the original data, and GMM classification was carried out before. The distribution of the probability density of the samples was estimated using GMM to obtain the probability of the samples on each category, and the one with the highest probability was usually selected as the sample category to which the sample belongs, we use the EM algorithm to estimate the GMM parameters.

In building the classification model, the 71 glass artefact samples collected were randomly divided into two datasets, with 40 in the training set (17 high potassium and 23 lead-barium) and 31 in the prediction set (18 high potassium and 13 lead-barium), each dataset contained samples of 2 different types of glass, the training set was used to build the classification model, the prediction set was not involved in modelling and was only used to validate the performance of the classification model. The sample data included the first 8 principal components of the GCMS data PCA. Due to the small number of glass samples, the randomness of one classification result is large, so the data set was randomly assigned into training and prediction sets 100 times, and the average result of 100 cycles was used as the final classification result, and the GMM score was set to 4. The average correctness of GMM for the training and prediction sets was 0.911 and 0.844, respectively, among these 100 classification results, in which the classification would the best correct result was 100%, and this result occurred 19 times with a correct rate of over 95%. The samples that were prone to classification errors were those that were not weathered, and the results indicate that the approach taken in this experiment is reliable [8].

A Gaussian mixture model (GMM) is a linear combination of multiple Gaussian functions to represent the probability density function of the data, assuming that different classes of samples consist of different Gaussian probability distributions [9-10]. Each Gaussian component density is determined by a vector of means and a covariance matrix, which is usually defined as follows:

$$p(x) = \sum_{k=1}^{K} \pi_k p(x|k) \qquad (1)$$

where: k is the number of models (i.e. the number of sample categories); π_k is the weight of the kth Gaussian, whose mean is u_k ,and the variance is σ_k σ, the maximum expectation (EM) algorithm is usually used to estimate some parameters such as variance, mean and weights in the Gaussian mixture model [11].

Each column of the confusion matrix represents the predicted category, the total of each column indicates the number of data predicted to be in that category each row represents the true attribution category of the data, the total of data in each row indicates the number of data instances in that category; the value in each column indicates the number of true data predicted to be in that category [12-13].

Accuracy rate (AC): is the most commonly used classification performance metric. It can be used to indicate the accuracy of a model, i.e., the number of correct identifications by the model/total number of samples. In general, the higher the accuracy of the model, the better the model is. The accuracy is calculated as follows.

$$AC = \frac{TP+TN}{TP+FN+FP+TN} \qquad (2)$$

Recall (RE): Also known as the check-all rate, recall is the ratio of the number of samples correctly identified by the model as positive classes to the total number of positive classes in the actual positive samples that the classifier can predict [14]. In general, the higher the RE, the more positive class samples are correctly predicted by the model and the better the model works, calculated as follows.

ISSN 2616-5880 Vol. 4, Issue 2: 1-5, DOI: 10.25236/AJMC.2023.040201

$$Re = \frac{TP}{TP + FN} \tag{3}$$

F1 Score (F1S): mathematically defined as F1 Score, is the evaluation score of the model [15]. The value of F1S is from 0 to 1, where 1 is the best and 0 is the worst.

$$F1S = \frac{2 * RE * AC}{RE + AC} \tag{4}$$

3. Results

3.1 The establishment of simulation model

This study identifies the type of artefact given in the form and gives the sensitivity of the classification model. This paper introduces a classification of glass based on principal component analysis and a Gaussian mixture model with the principle of maximization of expectations for the identification of glass artefacts. This paper addresses the glass chemical composition and glass type (type of high potassium glass and lead-barium glass), and considering that the problem is a dichotomous classification problem, the results are set to 0 and 1. Since there are 14 oxides of chemical elements studied in this paper, by preprocessing the data with principal component analysis, this paper selects a Gaussian model for classification according to the maximization of the expected improvement of principal component analysis, and divides the data into a training set and a test set Prediction was performed. The confusion matrix was introduced into the sensitivity analysis of the model. The correctness and recall of the confusion matrix will be good to evaluate the performance of the model.

3.2 Analysis of experimental results

Table 1: Graph of predicted glassware results

High Potassium Glass	Barium lead glass
A1	A2
A6	A3
A7	A4
A8	A5

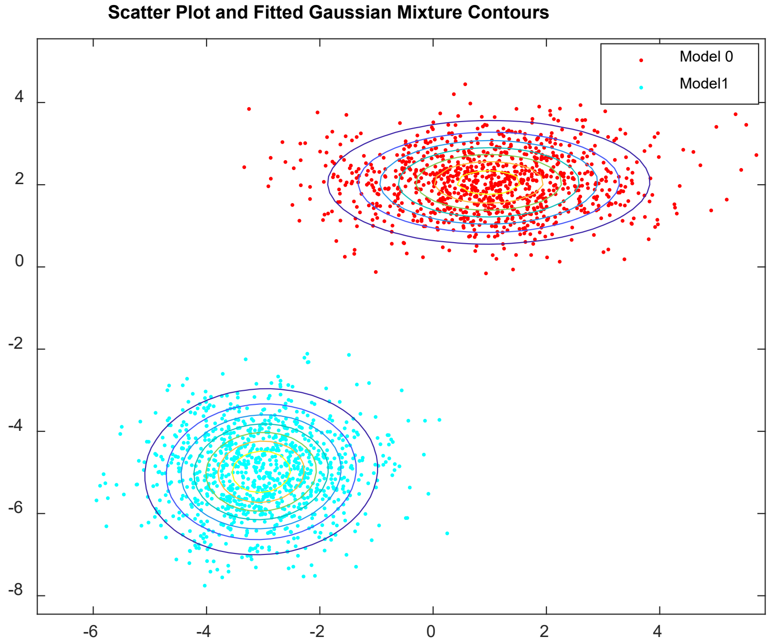

Figure 1: Gaussian mixture clustering diagram

The maximum expectation (EM) algorithm is usually used to estimate some parameters such as variance, mean, and weights in the Gaussian mixture model, and the prediction results are shown in the following table 1.

From the comparison between prediction data and actual data, the BP neural network has better prediction performance and relatively small error, which can meet the demand completely, and has fast prediction speed and convenient operation. This paper introduces the Gaussian Mixture Model (GMM), which can be seen as an optimization of the k-means model. It is both a technical tool commonly used in industry and a generative model. A Gaussian mixture model attempts to find a mixture representation of the probability distribution of a multidimensional Gaussian model, and thus fits an arbitrarily shaped data distribution. In the simplest scenarios, GMMs can be clustered in the same way as k-means. The clustering diagram for the Gaussian mixture model is shown in Figure 1. The predicted results were 65 weathered and 6 unweathered, which were substituted to obtain RE = 0.915, and the model prediction was insensitive; the model score was 0.915.

4. Conclusions

This paper describes the classification models for high potassium glass and lead-barium glass with a degree of accuracy and ingenuity, uses principal component models and Gaussian mixture models for glass distribution respectively, effectively constructs a glass classification model, and chooses to use a confusion matrix for sensitivity analysis of the model, proving that the model has good correctness and recall. In this paper, a Gaussian mixture model GMM is used for classification prediction, and PCA is used synthetically for data pre-processing to enhance the robustness of the model. This model also has implications for the planning and selection of glass craft production.

The model studied in this paper is very flexible and breaks the traditional purely discrete statistical modeling by cleverly integrating discrete and continuous into one. Given the limited resources available on the Internet, this paper combines existing statistical models, such as principal component analysis for data pre-processing, significance testing to justify the model, and a combination of Gaussian mixture models and discriminant analysis to provide a mathematical basis for archaeological classification. In this paper, the color, type, surface weathering, and ornamentation of glass are treated as variables to derive a quantitative mathematical model for the quality identification of glass. It also predicts the chemical elements before weathering with good results, which will help archaeologists to draw on the model to solve practical archaeological problems in the future. The optimal scores of the model parameters are also derived under the corresponding conditions, thus having good theoretical and practical significance, and the modeling method employed has good generalization value in practical applications and can be used as a reference to some extent.

References

[1] Wu B., Yu D. Deep convolutional neural network-based classification and detection of cell phone glass cover surface defects [J]. Software Engineering, 2021, 24(12):6-10.

[2] Zhang Jian, Hao Nannan. Gemological classification and identification of meteorite glass[J]. Shandong Chemical Industry, 2021, 50(17):160-161.

[3] Liu C. X., Deng P. F., Pan S. H., Luo W. D., Zhang S. L., Li T. M.. Classification and research status of fireproof glass [J]. Guangzhou Chemical Industry, 2021, 49(15):16-18.

[4] Yang X. Detection and classification of glass defects based on machine vision[D]. Fujian College of Engineering, 2021.

[5] Li Yingli. Research on glassware excavated in China from the 10th to the 12th century[D]. Liaoning Normal University, 2021.

[6] Ren Zhanzhan. Research on glassware of Wei, Jin, North and South Dynasties, Sui and Tang [D]. Suzhou University, 2020.

[7] Huang Jianyi. Research on the detection and classification method of cell phone panel glass defects based on digital image processing [D]. Xi'an University of Electronic Science and Technology, 2020.

[8] Li Xiangjing. Study on the correlation between the proportion of solid component volume in pulmonary ground glass shadow and the pathological classification of lung adenocarcinoma[D]. Henan University, 2020.

[9] Kong Yang. Study of glass precrystallization behavior of fluoroaluminosilicate system[D]. Qilu University of Technology, 2019.

[10] Du Wenxiu, Yang Juan, Gao Shuai, Wu Benhua, Wang Yinmao, Gao Guozhong. Introduction to the

ISSN 2616-5880 Vol. 4, Issue 2: 1-5, DOI: 10.25236/AJMC.2023.040201

characteristics and classification of fireproof glass [J]. Anhui Architecture,2015,22(04):179-181.
[11] He Yuanyuan. Research on contemporary decorative art glass design in China [D]. Wuhan University of Technology,2015.
[12] Huang Liguang, Chen Jinfang. Classification and status of electric glass melting kilns [J]. Glass, 2012, 39(12):37-39.
[13] Sun H, Liu M, Li L, et al. A new classification method of ancient Chinese ceramics based on machine learning and component analysis [J]. Ceramics International, 2019, 46(6).
[14] Daniel Ramos, Grzegorz Zadora. Information-theoretic feature selection using data obtained by scanning electron microscopy with energy dispersive X-ray spectroscopy for the classification of glass traces [J]. Analytica Chimica Acta, 2011, 705(1).
[15] Zetlein Michael B. Glass classification by elemental composition. A numerical assessment [J]. Forensic Science International, 1979, 13.