Automation and Machine Learning (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/autml.2023.040309
ISSN 2516-5003 Vol. 4 Num. 3

# *Composition Analysis and Identification of Ancient Glassware Based on Multimodal Analysis*

**Jianguo Deng[1, *], Jiahang Liang[1], Ruiqing Yang[1], Yuan Wang[1], Zhenxing Zhang[2]**

[1]*College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin, 300222, China*

[2]*College of Sciences, Tianjin University of Science and Technology, Tianjin, 300222, China*

[*]*Corresponding author: dengjianguo426@163.com*

***Keywords:*** Glass, Logistics regression model, Chi-square test, Stepwise regression model, Severity

***Abstract:*** Based on the data of cultural relics, this paper constructs the models of Logistics regression and stepwise regression, and discusses the relationship between the changes of chemical composition before and after weathering and the factors of glass relics. First of all, the relationship between weathering and glass type, pattern and color are analyzed by Logistics regression model and chi-square test. The results show that the relationship between weathering and glass type is strong, the relationship between weathering and grain decoration is weak, and the relationship between weathering and color is weak. Then, this study established a stepwise regression model and cross-verified to reveal the statistical law of weathering chemical composition. Six independent variables of weathering, lead oxide, sodium oxide, potassium oxide, barium oxide and alumina were retained by stepwise regression analysis. Finally, the component content prediction model before weathering is established by using the mean method to predict the corresponding compound content of the weathering point before weathering under the same glass type. In addition, with the help of the multiple regression equation, the cultural relic data are substituted into the regression equation to predict the classification results, and the sensitivity is analyzed by variance test. A judgment index, bad degree, is defined to judge the sensitivity. The prediction results show that 6 of the 8 cultural relics are less bad, the sensitivity of the prediction results is weak, and the classification results are better.

## 1. Introduction

Glass was introduced into China from West Asia and Egypt through early trade, and ancient glass with similar appearance but different chemical composition was made from local materials through the study of its technology. The main raw material of glass is quartz sand ($SiO_2$)[1-3]. In refining, fluxes such as plant ash and natural bubble alkali should be added, and limestone should be added as stabilizer. When different fluxes are added, the chemical composition of the glass is different, such as using plant ash as flux to form potassium glass[4]. Ancient glass will be weathered by the influence of burial environment, and the composition proportion of glass will change in the process of weathering, so it will affect the type of judgment[5].

However, the degree of weathering on the surface of cultural relics will be different, with obvious weathering areas, general weathering areas and individual cases without weathering. At present, a batch of cultural relics have been divided into two types: high potassium glass and lead-barium glass.

This paper intends to solve the following problems:

(1) Analyze the relationship between the weathering of the surface of cultural relics and the type, pattern and color of glass, and the statistical law of the content of weathering chemical components on the surface of cultural relics, and predict the content of chemical components before weathering according to the data of weathering points.

(2) Analyze the components of the unknown categories of cultural relics, identify their types, and analyze the sensitivity of the results.

## 2. Research methods.

In order to analyze the relationship between the surface weathering of cultural relics and glass type, pattern and color, there are multiple dependent variables and independent variables, Logistics regression model and chi-square test are established according to the relevant data, and the SPSS software is used to solve the problem. Secondly, the statistical law of weathering chemical composition content should be found, and in order to establish the optimal multiple linear regression equation to explain the relevant statistical law, a stepwise regression model is established.

Finally, the content of chemical components before weathering is predicted according to the detection data of weathering point.

According to some unknown types of cultural relics information, it is necessary to identify the glass type by analyzing the chemical composition, establish a multiple regression equation model, use MATLAB software to replace the data to solve the results, and analyze the sensitivity according to the classification results.

## 3. Establishment and solution of the model

### 3.1 Establishment and solution of the model of chemical composition change before and after weathering

#### 3.1.1 Analysis of the relationship between surface weathering and glass type, pattern and color

##### 3.1.1.1 Data preprocessing

It is considered that the cumulative proportion of each chemical composition and the data between 85% and 105% are valid data, so it is necessary to initially eliminate the data outside this range, and the data of No. 15 and No. 17 cultural relics can be removed according to the proportion sum.

##### 3.1.1.2 Establishment and solution of Logistics regression Model

(1) Establishment of Logistics model.

Logistics regression belongs to nonlinear regression. When the dependent variable is classified variable and the relationship between independent variable and dependent variable is not linear, Logistics regression analysis can be used.

It is a multiple regression analysis method to study the relationship between binomial classification or multinomial classification and some influencing factors.

(2) The solution of Logistics model.

First of all, the dependent variable is set as the weathering of the surface of cultural relics, that is, weathering and non-weathering; the independent variables that affect the weathering are glass type, color and decoration respectively.

The Logistics model is solved by SPSS software, and the results are as follows (Table 1):

Table 1: Solution of Logistics model

| | | Score | Degree of freedom | Significance |
|---|---|---|---|---|
| Variable | Ornamentation | 5.720 | 2 | 0.057 |
| | Type | 3.861 | 1 | 0.049 |
| | Color | 7.011 | 7 | 0.428 |
| Overall statistics | | 27.682 | 10 | 0.002 |

According to the solution results, the significance of glass type is $P < 0.05$, and the significance of glass decoration and color is $p > 0.05$.

It's not significant.

That is to say, it is found that the relationship between the surface weathering of cultural relics and the type of glass is strong, the relationship with grain decoration is slightly weakened, and it has almost nothing to do with the color of glass.

The results of the relationship between the weathering of cultural relics and the type, color and decoration of glass have been obtained above, and then the correctness of the Logistics model is verified by SPSS software. The test results are as follows:

Table 2: Check the correctness of the results

| Actual measurement | | Forecast | | |
|---|---|---|---|---|
| | | Surface weathering | | Correct percentage |
| | | Weathering | unweathering | |
| Surface weathering | Weathering | 27 | 3 | 90.0 |
| | unweathering | 6 | 16 | 72.7 |
| Overall percentage | | | | 82.7 |

As can be seen from the above table 2, the correct percentage of the results obtained by using the Logistics model is 82.7%, and the correctness of the results is high.

In order to further illustrate the accuracy of the results and the feasibility of the model, the single factor chi-square test was carried out on the pattern, color and type of glass respectively.

#### 3.1.1.3 Chi-square test.

Chi-square test is a hypothesis test method specially used to solve the statistical analysis of counting data, which belongs to the category of non-parametric test.

The main purpose of this paper is to compare the correlation analysis of two or more sample rates and two classification variables.

Chi-square test is divided into fit test and independence test [2]. Because this question requires the relationship between weathering and glass type, color and decoration, it belongs to the question of whether there is a correlation between two or more factors with multiple classifications. So it's an independence test.

The chi-square test is carried out on the surface weathering and glass decoration, type and color by using SPSS software. This topic is the correlation degree between the two classification variables, and it is not an ordered classification variable, so the cross table in SPSS software can be used.

The test results of surface weathering and ornamentation are as follows (Table 3):

Table 3: Chi-square test of ornamentation

| Chi-square test | | | |
|---|---|---|---|
| | Value | Degree of freedom | Progressive significance (bilateral) |
| Pearson chi-square | 5.720[a] | 2 | 0.057 |
| Likelihood ratio (L) | 7.900 | 2 | 0.019 |
| Number of valid cases | 52 | | |

Note: the expected count of 2 cells (33.3%) is less than 5.
The minimum expected count is 2.54.

The chi-square test results of surface weathering and color are as follows (Table 4):

Table 4: Chi-square test of colors

| Chi-square test | | | |
|---|---|---|---|
| | Value | Degree of freedom | Progressive significance (bilateral) |
| Pearson chi-square | 7.011[a] | 7 | 0.428 |
| Likelihood ratio (L) | 8.822 | 7 | 0.266 |
| Number of valid cases | 52 | | |

The expected count of 12 cells (75.0%) is less than 5.

The minimum expected count is .42.

The chi-square test results of surface weathering and types are as follows (Table 5):

Table 5: Chi-square test of types

| Chi-square test | | | | | |
|---|---|---|---|---|---|
| | Value | Degree of freedom | Progressive significance (bilateral) | Precise significance (bilateral) | Precise significance (unilateral) |
| Pearson chi-square | 3.861[a] | 1 | 0.049 | | |
| Continuity correction b | 2.758 | 1 | 0.097 | | |
| Likelihood ratio (L) | 3.853 | 1 | 0.050 | | |
| Fisher accurate test | | | | 0.070 | 0.049 |
| Number of valid cases | 52 | | | | |

Note: the expected count of 0 cells is less than 5.
The minimum expected count is 6.77.

Because the chi-square test requires that the minimum expected frequency of each cell is greater than 1, the chi-square test of surface weathering and types is valid.

The calculation method of Pearson's chi-square is:

$$\sum \frac{(\text{Theoretical value} - \text{expected value})^2}{\text{Expected frequency}} \tag{1}$$

According to the above test results, it is known that the surface weathering is significantly related to the type of Pearson's square ($p < 0.05$), so there is a strong correlation between the weathering and the type of glass.

### 3.1.2 Statistical law of weathering chemical composition.

#### 3.1.2.1 Establishment of stepwise regression model.

There are two kinds of stepwise regression analysis: forward stepwise regression and backward stepwise regression. This topic adopts backward stepwise regression, which is contrary to forward stepwise regression. It is necessary to eliminate the useless data after introducing variables. The specific analysis steps are shown in figure 1 below:

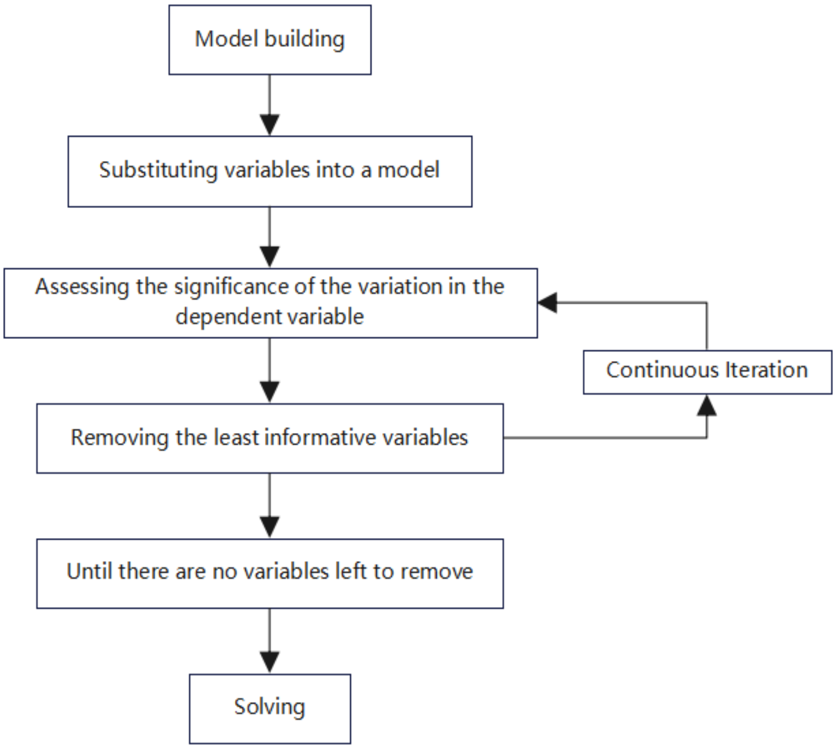


Figure 1: Analysis steps of stepwise regression model.

Establish all chemical constituents A1, a2...

Ai regression equation for dependent variable glass type b, and then F test for these independent variables: take the minimum value

$$F_{z1}=\min(F_1,F_2,\ldots F_i) \tag{2}$$

If Fz1 > Fm is obtained, it shows that there are no independent variables to be eliminated, and the optimal regression equation can be obtained.

Otherwise, the az1 is removed.

And let az1=ar, continue to build A1, a2...

The regression equation of ar and dependent variable b, continue to do F test, take the minimum value:

$$F_{z2'}=\min(F_{1'},F_{2'}\ldots F_{r-1'}) \tag{3}$$

If there is Fz2 > Fm, it shows that the equation is optimal without eliminating variables, on the contrary, it is necessary to eliminate az2 and make az2=ar-1, that is, iterate continuously according to the previous method, until there are no variables in the equation that need to be eliminated, and the final optimal regression equation is obtained.

### 3.1.2.2 Solution of stepwise regression model.

It is necessary to analyze the statistical law of weathering chemical composition content on the surface of cultural relics according to the types of cultural relics. first of all, 5 groups of data were randomly selected from 68 groups of complete data as the test set (here, 16, 43, 24, 11, 62 rows of data were randomly selected by using random number seeds), and the other 63 were used as training sets.

Using Stata software, because the types of dependent variables and independent variables with or without weathering belong to qualitative variables, it is necessary to construct virtual variables (L1, L2, F1, f2 with or without weathering), and then establish a stepwise regression model to solve them. Because there are four combinations of virtual variables, four multiple regression equations will be obtained.

$$y = \beta_1 + \beta_2 f_1 + \beta_3 x_1 + \beta_4 x_2 + \beta_5 x_3 + \beta_6 x_4 + \beta_7 x_5 \tag{4}$$

$$\beta = [\beta_1 \ \ \beta_2 \ \ \beta_3 \ldots \beta_7] \tag{5}$$

$$\beta = \begin{matrix} 0.0836 & 0.2468 & 0.0153 & 0.0266 & -0.0474 & 0.0141 & 0.0257 \\ 0.3304 & -0.2468 & 0.0153 & 0.0266 & -0.0474 & 0.0141 & 0.0257 \\ 0.9164 & -0.2468 & -0.0153 & -0.0266 & 0.0474 & -0.0141 & -0.0257 \\ 0.6696 & 0.2468 & -0.0153 & -0.0266 & 0.0474 & -0.0141 & -0.0257 \end{matrix} \tag{6}$$

Where:

The value of high potassium in L1 is 0, the value of lead and barium is 1, the value of high potassium is 1, and the value of lead and barium is 0.

The weathering value of F1 is 0, the unweathered value is 1, the unweathered value is 1, and the unweathered value is 0.

After stepwise regression, the independent variables with no significant correlation with dependent variables were eliminated, and six independent variables with or without weathering, lead oxide, sodium oxide, potassium oxide, barium oxide and alumina were retained.

### 3.1.2.3 Test of stepwise regression model.

First, the joint significance test of the regression equation obtained by stepwise regression was carried out, in which the regression coefficients of the original hypothesis Haugh 0 were all 0.

Table 6: Significance test of regression equation

| | (f1,l1) | (f1,l1) | (f1,l1) | (f1,l1) |
|---|---|---|---|---|
| Prob > F | 0 | 0 | 0 | 0 |
| F statistics | 64.41 | 64.41 | 64.41 | 64.41 |

From the results of Table 6, it can be seen that the P value of the joint significance test is 0, less than 0.01, so the original hypothesis Hauppo can be rejected with a confidence interval of more than 99%.

Therefore, the regression equation obtained by stepwise regression analysis meets the requirements.

The advantages and disadvantages of the four multiple regression equations are compared. By making the four regression equations predict the virtual variable values of the corresponding categories of the five groups of test sets, the residuals and the sum of squares of the five groups of fitting values and actual values are obtained respectively. The results are as follows Table 7.

By comparison, it is found that the sum of residual squares of the multiple regression equations corresponding to the stepwise regression in the four cases are exactly the same, indicating that the stepwise regression in these four cases is the same, so the type is taken as the dependent variable. The multiple regression equation is obtained by stepwise regression with or without weathering, lead oxide, sodium oxide, potassium oxide, barium oxide and alumina as independent variables.

Table 7: 5 residual and sum of squares of fitted values and actual values

| | **(f1,l1)** | **(f1,l1)** | **(f1,l1)** | **(f1,l1)** |
|---|---|---|---|---|
| The square of residual error obtained by testing 11 rows of data | 0.0037 | 0.0037 | 0.0037 | 0.0037 |
| The square of residual error obtained by testing 16 rows of data | 0.0014 | 0.0014 | 0.0014 | 0.0014 |
| The square of residual error obtained by testing 24 rows of data | 0.0088 | 0.0088 | 0.0088 | 0.0088 |
| The square of residual error obtained by testing 43 rows of data | 0.0174 | 0.0174 | 0.0174 | 0.0174 |
| The square of residual error obtained by testing 62 rows of data | 0.0105 | 0.0105 | 0.0105 | 0.0105 |
| The sum of squares of residuals obtained from the test of prediction set data | 0.0419 | 0.0419 | 0.0419 | 0.0419 |

### 3.1.3 Prediction of component content before weathering

#### 3.1.3.1 Prediction model of component content before weathering

According to the fact that the weathering of the surface of cultural relics is related to the chemical composition and the type of glass, the glass is divided into two groups according to the type, and the weathering of the surface of cultural relics is related to the content of its chemical composition. Therefore, the content of chemical composition can be counted to judge the surface weathering of cultural relics, and the content of the same type of compounds before and after weathering can be compared and predicted by the mean method.

That is to say, the average value of the content of the same type of compound before and after weathering is calculated, that is, the content of the same type of compound before and after weathering.

$$\overline{\omega}_i = \frac{1}{2}(\omega_{i,\ 1} + \omega_{i,\ 0}) \tag{7}$$

Indicates the content of type I compounds before and after weathering, while 1 and 0 represent no weathering and weathering.

According to the calculated mean value, the content of each compound without weathering can be predicted by substituting it into the cultural relics to be predicted.

#### 3.1.3.2 Calculate the content of pre-weathering components from the model

According to the formula, the values of high potassium and lead-barium ω _ I are obtained. The results are as follows (Table 8):

Table 8: The value of high potassium, lead and barium $\overline{\omega}_i$

| High potassium | Lead and barium |
|---|---|
| 74.6575 | 42.14978 |
| 0.52125 | 0.904565 |
| 7.155625 | 0.193478 |
| 4.163125 | 2.085 |
| 0.809375 | 0.671957 |
| 5.313125 | 3.762391 |
| 1.51375 | 0.691957 |
| 2.294375 | 1.889348 |
| 0.30875 | 30.805 |
| 0.44875 | 10.12326 |
| 1.12125 | 2.954348 |
| 0.03125 | 0.341087 |
| 0.1475 | 0.061957 |
| 0.07625 | 0.851739 |

Because:

$$\omega_{i,\ 1} = 2\overline{\omega}_i - \omega_{i,\ 0} \tag{8}$$

Therefore, according to the formula, the content of each compound without weathering can be predicted.

### 3.2 Establishment and solution of classification model.

#### 3.2.1 Establishment and solution of Multivariate regression equation Model.

(1) the establishment of multiple regression equation model.

The type of glass should be judged according to the content of chemical composition, and the surface weathering is known, so the multiple regression model is adopted.

(2) the solution of multiple regression equation model.

The prediction results of multiple regression equation model are obtained by using MATLAB substitution data. The prediction results are as follows (Table 9):

Table 9: Results of multivariate regression model

| Cultural relic number | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| Prediction result | Lead and barium | Lead and barium | Lead and barium | Lead and barium | Lead and barium | High potassium | High potassium | Lead and barium |

#### 3.2.2 Sensitivity analysis of classification results.

The multiple regression model is used to predict the classification results, and the variance test can be used to analyze the sensitivity of the results.

Since the classification result is defined as 0 and 1, the variance SSE is defined as the square of the difference between the predicted value ε and the standard value ε 0 between 0 and 1, that is:

$$SSE = (\varepsilon - \varepsilon_0)^2 \tag{9}$$

Therefore, when the prediction result is between 0 and 1, that is, 0.5, the variance is the largest, which is defined as the extremely bad variance $SSE_0$.

$$SSE_0 = 0.5^2 = 0.25 \tag{10}$$

The bad degree δ is defined as the quotient of variance and extreme bad variance, and it is considered that when $\delta < 50\%$, the bad degree is less, the sensitivity is weak, and the classification result is better.

$$\delta = \frac{SSE}{SSE_0} \tag{11}$$

The bad degree of the result of cultural relic classification is calculated.

As can be seen from Table 10, 6 of the 8 cultural relics are less than 50% bad, that is, 75% of the predicted classification results are less bad, less sensitive, and the classification results are better(Table 10).

Table 10: The severity of the classification results

| Cultural relic number | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| Severity (%) | 93.6207 | 44.0177 | 0.067 | 0.5385 | 50.9134 | 1.377 | 11.181 | 7.9696 |

## 4. Conclusion

This paper concludes that a comprehensive analysis of the relationship between the chemical composition of glass relics and their weathering factors can be achieved through the construction of Logistics regression and stepwise regression models. The study reveals that the degree of weathering is significantly associated with the type of glass, while the relationships with grain decoration and color are weaker. Moreover, the stepwise regression model demonstrates that six independent variables, including lead oxide, sodium oxide, potassium oxide, barium oxide, and alumina, have a considerable impact on the weathering chemical composition. The paper also presents a component content prediction model for the pre-weathering state, which can assist in estimating the compound content before weathering under the same glass type. The sensitivity of the prediction results is analyzed using a bad degree judgment index, and the variance test shows that the classification results are relatively accurate. However, the sensitivity of the prediction results is found to be weak, with only 6 out of 8 relics being classified as less bad. Future research could focus on improving the sensitivity of the prediction model to better preserve and protect cultural relics. Overall, the proposed models provide a valuable tool for understanding the impact of weathering on glass relics and contribute to the preservation of these valuable cultural heritage objects.

## References


*[1] Katja S,Racem R B,Taras G, et al. Household dishwashing detergents efficiently inactivate HSV-1, but not the non-enveloped viruses HAV and MNV, while all viruses were removed from glass by manual scrubbing[J]. Food Microbiology, 2024, 118.*

*[2] Mingxu F,Xiaoqin Z,Jinzhou S. Experimental investigation on cutting force and machining parameters optimization in in-situ laser-assisted machining of glass–ceramic[J]. Optics and Laser Technology,2024,169.*

*[3] Yongliu Z,Yuan L,Hui S, et al. Improving the adhesion of diamond-like carbon films on glass substrate by chemical etching and oxygen plasma treatment[J]. Vacuum,2023,218.*

*[4] Sibel O,Merve O Z,M. A D. Modulation of optical oxygen sensitivity of H2(TPP) by using Eu3+, Gd3+ and Yb3+ ions doped silicate-based bioactive glass powders immersed in simulated body fluid[J]. Optik,2023,294.*

*[5] Flaviano M,Dries S. Exact solution to the fully connected XY model with Gaussian random fields by the replica method[J]. Physica A: Statistical Mechanics and its Applications,2023,629.*