Academic Journal of Materials & Chemistry
ISSN 2616-5880 Vol. 4, Issue 6: 19-24, DOI: 10.25236/AJMC.2023.040604

Identification of Ancient Glass Based on Multiple Rinear Regression

Cui Wang

School of Management, Shandong University of Technology, Zibo, Shandong, 255000, China

Abstract: *The weathering of ancient glass under the influence of burial environment will result in the change of the composition ratio of glass. In order to identify the types of ancient glass products, this paper first establishes four multiple regression equations through a batch of existing weathered glass sample detection data, and obtains the chemical composition content rule of ancient glass surface weathering. Then, the difference between the absolute value of the predicted value and the actual value is the minimum as the basis for the identification of the unknown category of glass relics. Finally, the perceptron model optimized by particle swarm optimization is used as the auxiliary evidence to compare the results of the two, analyze the different categories, and finally get a more accurate answer.*

Keywords: *Multiple linear regression, Perceptron model, Classification of ancient glass*

1. Introduction

Weathering is one of the most common phenomena of ancient glass and one of the most important factors affecting quantitative analysis. Since ancient glass has been buried underground for a long time, different degrees of corrosion and weathering will occur due to different burial environment and composition system, and the composition of the unweathered part on the surface and inside is very different [1]. Bethany Matthews [2] et al. studied the characterization of ancient glass alteration by various environments, taking Broborg in Hilford, Sweden as an example. Feng Bailing [3] studied the development and construction of a special database suitable for glass research. Yin Yulong [4] studied the composition analysis and identification of ancient glass products by means of optimal scaling method, systematic clustering and Fisher multiple classification criteria. Lu Jiajia [5] classifies ancient glass through the method of base integration feature selection and random forest. However, there are few studies on the identification of unknown classes of ancient glass. For the identification of unknown ancient glass types, this paper first constructed multiple linear regression equation based on sample data to describe the law of weathering chemical composition content on the surface of glass samples, and then constructed a multiple linear regression model to determine the actual situation of unknown data with the goal of minimizing the difference between the actual value and the expected value. However, considering that there may be errors in the identification of a single model, in order to ensure the scientificity and rationality of the identification, by learning the original data and analyzing the category of unknown data set, the results of the two are compared. If the results are consistent, the output is direct; if the results are inconsistent, the different places are analyzed and identified in detail, and the third way is verified again according to the particularity of the components.

2. Establishment of glass identification types

2.1 Statistical law analysis based on multiple linear regression

There is a batch of chemical composition data of glass relics of unknown category. In order to identify their types, this paper firstly divides all samples into four categories according to weathering or not and glass type: Lead barium weathering, lead barium weathering without weathering, high potassium weathering, high potassium weathering without weathering, four multivariate linear regression equations were established to analyze the statistical rule of weathering chemical composition on the surface of cultural relics samples.

In these composition data, all samples contain silica (SiO_2), so silica (SiO_2) is more representative. In this paper, silica was selected as independent variable and 13 other chemical components were selected as dependent variable for multiple linear regression.

ISSN 2616-5880 Vol. 4, Issue 6: 19-24, DOI: 10.25236/AJMC.2023.040604

2.1.1 Establishment of multiple linear regression model

Taking high potassium weathering as an example, the upper limit of *i* is 6. When it is the other three types, only the upper limit of *i* is changed to: (high potassium no weathering 9) (lead barium no weathering 22) (lead barium no weathering 11).

The observed values $y, x_1, x_2, \cdots x_{13}$ are respectively $b_i, a_{i1}, \cdots a_{i13}, i = 1, 2, \cdots, 6,$ and

$$X = \begin{bmatrix} 1 & a_{11} & a_{12} & \cdots & a_{1,13} \\ 1 & a_{21} & a_{22} & \cdots & a_{2,13} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{61} & a_{62} & \cdots & a_{6,13} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0.59 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_6 \end{bmatrix} = \begin{bmatrix} 92.63 \\ 95.02 \\ \\ 92.72 \end{bmatrix} \tag{1}$$

The estimated value of $c_0, c_1, c_2, \cdots c_{13}$ obtained by the least square method, that is, the estimated value $\hat{c}_j$ should be selected so that when $c_j = \hat{c}_j$, *j*=0,1,2…,13, the error sum of squares reaches the minimum.

$$Q = \sum_{i=1}^{6} \varepsilon_i^2 = \sum_{i=1}^{6} \left(b_i - \hat{b}_i\right)^2 = \sum_{i=1}^{6} (b_i - c_0 - c_1 a_{i1} - c_2 a_{i2} - \cdots - c_{13} a_{i13})^2 \tag{2}$$

To this end,

$$\frac{\partial Q}{\partial c_j} = 0, j = 0, 1, 2 \cdots, 13 \tag{3}$$

The normal equations are obtained, and the estimated value of $c_0, c_1, c_2, \cdots c_{13}$ solving the normal equations is obtained.

$$\left[\hat{c}_0, \ \hat{c}_1, \ \hat{c}_2, \ \cdots \ \hat{c}_{13}\right] = (X^T X)^{-1} X^T Y \tag{4}$$

For the multivariate linear model with dependent variable silica (SiO_2), for the data with constant values in 13 chemical components, the multivariate linear model will delete these variables in the analysis.

2.1.2 Stepwise regression analysis

When establishing the regression model, not every factor has a great influence on *y*. Therefore, this paper selects factors by stepwise regression method (elimination of superior selection method). The specific process is as follows:

Step1: A multiple linear regression equation with silica (SiO_2) as the dependent variable was established.

Step2: Test the significance of regression coefficient, and take *t* value corresponding to the maximum probability value P_{max};

Step3: Determine whether $P_{max} \leq 0.05$, if it meets the requirement, go to Step5, if not, go to Step4;

Step4: H_0 is acceptable, that is, the linear relationship between this index and the dependent variable is not significant. Remove the index and return to Step1.

Step5: Then H_0 can be rejected, and the linear relationship between all indexes and dependent variables is significant. Output the equation and end.

Through the above calculation, four multiple linear regression equations can be obtained, and the corresponding multiple linear equations of two weathering states of the same kind of glass can be used to represent the statistical law of whether the cultural relic samples have weathered chemical components, which can intuitively show the difference before and after.

2.1.3 The solution of statistical law based on multiple linear regression

SPSS26.0 software was used to solve the problem, and the multiple linear regression equations corresponding to the four cases were obtained.

ISSN 2616-5880 Vol. 4, Issue 6: 19-24, DOI: 10.25236/AJMC.2023.040604

As for lead barium weathering, the analysis shows that there are no indicators removed because all are constant values, that is, multiple linear regression is still carried out on 13 chemical components.

In order to show that there is a close relationship between silica (SiO_2) and 13 other chemical components in lead-barium weathering, this paper uses SPSS26.0 to test the significance of the obtained multiple regression model:

(1) Decision coefficient

Table 1: Model summary

Model	R	R squared	R squared after adjustment	Errors in standard estimates
1	1.000[a]	.999	.997	.92733

Table 1 shows that the complex correlation coefficient R=1.000 and the multiple decision coefficient R^2=0.999. The adjusted R^2=0.997, the larger the adjusted R^2 value, the better the model fitting effect, 0.997 is close to 1, the fitting effect is very good.

(2) Significance test of regression equation

Table 2: Variance analysis table

Model		Squares	Degrees of freedom	Mean square	F	Significance
1	Regression	7129.400	13	548.415	637.730	.000[b]
	Residual error	6.880	8	.860		
	Total	7136.280	21			

As can be seen from Table 2, F=637.730，F0.05（13，22）＜F0.05（10，22）=2.297＜637.730.The regression equation is significant, that is, there is an obvious functional relationship between the independent variable and the dependent variable.

(3) Significance test of regression coefficient

Table 3: Regression coefficient table

Model		**Unstandardized coefficient**		**Standardized coefficient**	**t**	**Significance**
		B	Standard error	Beta		
1	constant	101.551	2.323		43.715	.000
	Na_2O	-1.266	.198	-.145	-6.390	.000
	K_2O	2.554	1.584	.034	1.612	.146
	CaO	-1.696	.517	-.119	-3.279	.011
	MgO	-1.302	.568	-.051	-2.293	.051
	Al_2O_3	-.950	.168	-.186	-5.642	.000
	Fe_2O_3	-1.843	.524	-.063	-3.519	.008
	CuO	-.689	.348	-.048	-1.978	.083
	PbO	-1.045	.042	-.832	-25.106	.000
	BaO	-1.336	.098	-.603	-13.676	.000
	P_2O_5	-1.213	.214	-.206	5.667	.000
	SrO	4.412	1.587	.052	2.779	.024
	SnO_2	-4.636	6.222	-.012	-.745	.478
	SO_2	-.438	.159	-.108	-2.764	.025

Table 3 shows the T-test result of whether the partial regression coefficient of the model for 13 variables is 0.t values were 43.715, -6.390, 1.612, -3.279, -2.293, -5.642, -3.519, -1.978, -25.106, -13.676, -5.667, 2.779 and -0.745, respectively, and the probability p value was less than the significance level 0.05.However, potassium oxide (K_2O) was 0.146, tin oxide (SnO_2) 0.478, magnesium oxide (MgO) 0.051, and copper oxide (CuO) 0.083.At this point, it is considered that the linear correlation between components less than 0.05 and the dependent variable is significant, while potassium oxide (K_2O), tin oxide (SnO_2), magnesium oxide (MgO) and copper oxide (CuO) are not significant. Therefore, the revised multiple regression equation is as follows:

$$y_1 = 101.551 - 1.266x_1 - 1.696x_3 - 0.95x_5 - 1.843x_6 - 1.045x_8 - 1.336x_9 - 1.213x_{10} + 4.412x_{11} - 0.438x_{13} \quad (5)$$

The regression equation obtained by this study shows that the coefficients of x_{11},x_6,x_3 and x_1 are larger, among which the coefficient of P_2O_5 is the largest, and it is a positive correlation coefficient, that is, the more P_2O_5 content, the more SiO_2content, P_2O_5 content has the greatest influence on the content of SiO_2.x_6,x_3,x_1,x_9, which are relatively large, are negatively correlated with SiO_2, that is, the more there

ISSN 2616-5880 Vol. 4, Issue 6: 19-24, DOI: 10.25236/AJMC.2023.040604

are, the less SiO_2 there is.

In the same way, the multiple linear regression equation of lead barium without weathering, high potassium weathering and high potassium without weathering is calculated.

Regression equation of lead barium without weathering:

$$y_2 = 68.001 - 2.101x_1 + 15.5x_2 + 13.390x_3 + 0.19x_4 + 0.216x_5 - 2.623x_6 - 1.062x_9 - 2.409x_{10} - 15.937x_{11} - 1.233x_{13} \quad (6)$$

This equation is the regression equation of lead barium without weathering. By studying its parameters, it is found that x_2, x_3, x_4 and x_5 are all positively correlated, among which x_2, has the highest coefficient, reaching 15.5, indicating that x_2 and sodium oxide have the greatest positive influence on it, and other elements are negatively correlated with SiO_2, in which x_{11} is P_2O_5, that is, the more content, the less SiO_2.

Regression equation of high potassium weathering:

$$y_3 = 102.966 - 0.223x_3 - 7.247x_4 - 13.443x_6 - 3.160x_7 + 3.975x_{10} \quad (7)$$

This regression function expresses the regression equation of high potassium weathering. As for the effect of SiO_2 content, only five elements have an effect on it. x_6 (alumina) has the largest correlation coefficient, which is negative correlation, that is, the content is inversely proportional to the content of SiO_2 and has a great influence.

Regression equation of high potassium without weathering:

$$y_4 = 62.69 - 0.849x_1 + 0.551x_2 + 2.798x_4 - 6.449x_6 - 2.503x_8 + 8.199x_{10} - 55.066x_{11} + 0.837x_{13} \quad (8)$$

For the regression equation of high potassium weathering free, x_1, x_6 and x_8 are negatively correlated with SiO_2, while the rest are positively correlated. The coefficient of x_6 is the largest, that is, the content of Al_2O_3, which has a greater influence on SiO_2.The coefficient of x_{11} is the largest, that is, the content of P_2O_5 has the greatest influence on the content of SiO_2.

2.2 Identification of glass classes

Each data set to be measured is substituted into the four multiple linear regression equations obtained, namely, the surface weathering is substituted into the two equations of lead barium weathering and high potassium weathering. The difference between the predicted value and the actual value of silica is calculated, and the category with a small difference is selected as the type of glass cultural relics, so as to determine the type of glass. The calculation process is as follows:

$$Y_1 = \overrightarrow{C_1} \cdot \overrightarrow{X_1}^T + 101.551 \quad (9)$$

$$Y_3 = \overrightarrow{C_3} \cdot \overrightarrow{X_3}^T + 102.996 \quad (10)$$

$$S_k = arg\min\{|E_{3k} - Y_{1k}|, |E_{3k} - Y_{3k}|\} \quad (11)$$

$$S_k = \begin{cases} 1 \quad S_k = |E_{3k} - Y_{1k}| \\ 0 \quad S_k = |E_{3k} - Y_{3k}| \end{cases} \quad (12)$$

Y_1 represents the silica content of lead-barium weathered glass, Y_3 represents the silica content of high-potassium weathered glass, $\overrightarrow{C_1}$ and $\overrightarrow{C_3}$ represent the coefficients of each component in the multiple linear regression models of lead-barium weathered glass and high-potassium weathered glass. $\overrightarrow{X_1}^T$ and $\overrightarrow{X_3}^T$ represent the transposition of vectors represented by the chemical components of lead barium weathered glass and high potassium weathered glass.

S_k refers to the type situation of the weathered relics of unknown category, E_{3k} refers to the actual silica content of the surface weathered samples, Y_{1k} refers to the predicted silica content of lead-barium weathered glass, and Y_{3k} refers to the predicted silica content of high-potassium weathered glass.

Similarly, the unweathered surface samples are taken into the two equations of unweathered lead barium and unweathered high potassium, and the difference between the predicted value and the actual value of silica is calculated. The category with a smaller difference is selected as the type of glass cultural relics, so as to determine the type of glass.

$$Y_2 = \overrightarrow{C_2} \cdot \overrightarrow{X_2}^T + 68.001 \tag{13}$$

$$Y_4 = \overrightarrow{C_4} \cdot \overrightarrow{X_4}^T + 62.69 \tag{14}$$

$$Y_4 = \overrightarrow{C_4} \cdot \overrightarrow{X_4}^T + 62.69 \tag{15}$$

$$T_k = \begin{cases} 1 & T_k = |E_{4k} - Y_{1k}| \\ 0 & T_k = |E_{4k} - Y_{2k}| \end{cases} \tag{16}$$

Y_2 represents the silica content of lead-barium non-weathered glass; Y_4 represents the silica content of high potassium non-weathered glass;$\overrightarrow{C_2}$ and $\overrightarrow{C_4}$ represent the coefficients of each component in the multiple linear regression model of lead-barium and high-potassium weathering free glass;,$\overrightarrow{X_2}^T$ and $\overrightarrow{X_4}^T$ represent the transpose of vectors represented by the chemical components of lead barium and high potassium glass without weathering.

T_k refers to the type situation of unweathered relics of unknown class, E_k refers to the true silica content of unweathered samples on the surface, Y_{2k} refers to the predicted silica content of lead barium unweathered glass, and Y_{4k} refers to the predicted silica content of high potassium unweathered glass.

3. Results

3.1 Identification results of glass types

The predicted silica value of the theory was calculated through excel, and the data is shown in Table 4:

Table 4: Predicted value of SiO_2

Surface weathering	High potassium	Lead barium	Surface weathering	High potassium	Lead barium
Weathering free	61.49	142.03	Weathering	157.99	33.24
Weathering free	-85.55	147.9	Weathering	72.54	68.03
Weathering free	17.56	70.11	Weathering	93.04	97.08
Weathering free	6.44	60.65	Weathering	96.31	93.39

The smaller difference between the two can be used to determine whether weathering or not, and the specific results are shown in Table 5:

Table 5: Multiple linear regression classification case

Cultural relics	A1	A2	A3	A4	A5	A6	A7	A8
type	High potassium	Lead barium	Lead barium	High potassium	Lead barium	High potassium	Lead barium	Lead barium

3.2 Analysis of results

In order to discuss the accuracy of the multiple linear regression model, the strategy learned by the perceptron improved based on particle swarm optimization will be used in this paper to test the unknown category of sample data as a test set.

The results predicted by the improved perceptron model based on particle swarm optimization are shown in Table 6:

ISSN 2616-5880 Vol. 4, Issue 6: 19-24, DOI: 10.25236/AJMC.2023.040604

Table 6: Perceptron classification situation

Cultural relics	A1	A2	A3	A4	A5	A6	A7	A8
type	High potassium	Lead barium	Lead barium	Lead barium	Lead barium	High potassium	High potassium	Lead barium

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.

By comparing the two results, it can be found that there is a difference between A4 and A7.

However, by observing the silica content, it can be found that the silica content of A4 is close to the average silica content of lead barium glass rather than the average of high potassium glass. Similarly, the silica content of A7 is closer to the average for high potassium glass than for lead barium glass. To sum up, the final results are shown in Table 7:

Table 7: Final classification situation

Cultural relics	A1	A2	A3	A4	A5	A6	A7	A8
type	High potassium	Lead barium	Lead barium	Lead barium	Lead barium	High potassium	High potassium	Lead barium

4. Conclusions

Based on the chemical composition data of known glass cultural relics, this paper obtains the rule of chemical composition content of whether the surface of glass cultural relics is weathered or not, and then analyzes the chemical composition of unknown glass cultural relics and identifies their types. Finally, by constructing multiple linear regression model and supporting vector machine improved by particle swarm optimization algorithm, more accurate type identification results are obtained. When studying the rule of chemical composition content on the surface of glass relics weathering or not, the most representative index is the independent variable, and the other indexes are the dependent variable to establish a multiple linear or nonlinear regression model, which is of popularization significance.

References

[1] Liu Song, Li Qinghui, Gan Fuxi. Influence of Surface Factors on Portable X-ray Fluorescence Analysis of ancient glass samples [J]. Spectroscopy and spectral analysis, 2011, 31(07): 1954-1959.

[2] Bethany Matthews, Bruce Arey, Carolyn Pearce, Albert Kruger. Characterization of Glass Alterations in Ancient Glass from Various Environments from Broborg, a Vitrified Swedish Hillfort [J]. Microscopy and Microanalysis, 2020, 26(S2).

[3] Feng Bailing. The Construction and application of the database of Ancient Glass Beads unearthed in China [D]. Northwest University, 2021. DOI: 10.27405/d.cnki.gxbdu. 2021.001676.

[4] Yin Yulong. Composition Analysis of ancient glass Products by Association Prediction [J]. Contemporary Chemical Industry Research, 2023, No. 126(01): 122-126.

[5] Lu Jiajia. Classification model of ancient glass based on integrated feature selection and random forest [J]. Journal of Ceramics, 2023, 51(04): 1060-1065. DOI: 10.14062/j.issn. 0454-5648.20220790.