Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 8, No. 1, 2023

Research on Glass Classification and Identification Based on Cluster Analysis and Multiple Regression Models

Jinlong Lu*, Yahan Lu

School of Information Engineering, Zhengzhou University of Industrial Technology, Henan, 451199, China
*Corresponding author: 1549955913@qq.com

Abstract: Ancient glass undergoes a large exchange of internal elements with environmental elements in burial, and its composition changes, thus affecting the correct judgement of its category. In order to classify and identify the types of ancient glass, this paper sets different types of glass as the dependent variable, and the grain, weathering or not, and their chemical composition characteristics as the independent variables, so as to form an interpretable classification law. In order to carry out the sub-classification under the same type of glass, then a chemical element content is selected and then the colour of the glass grain is observed under that content. In order to identify the type of glass, this paper builds a multiple regression model on the data of chemical composition and glass type, from which the type of glass is identified. The models established in this paper have passed the stability test.

Keywords: Glass classification; Cluster analysis; Multiple linear regression.

1. Introduction

Ancient glass is highly susceptible to weathering under the influence of the buried environment. During the weathering process, the internal elements are exchanged with the environmental elements in large quantities, which leads to changes in the proportion of its composition, thus affecting the correct judgement of its category. Therefore, it is important to investigate the influence of different chemical compositions and their contents on the weathering of glass surfaces and the degree of weathering[1-2].

2. Modelling and Analysis

2.1. Model solution

The data after pre-processing were imported into SPSSPRO for decision tree analysis to obtain the proportion of importance of the respective variables as follows Table 1:

Table 1. Proportion of importance of chemical composition

Characteristic name	Characteristic importance
Silicon Dioxide(SiO2)	0.00%
Strontium oxide(SrO)	0.00%
Tin oxide (SnO2)	0.00%
Sulphur dioxide (SO2)	0.00%
Phosphorus pentoxide (P2O5)	0.00%
Potassium oxide (K2O)	0.00%
Lead oxide(PbO)	100.00%
Barium oxide (BaO)	0.00%
Copper oxide (CuO)	0.00%
Aluminium oxide(Al2O3)	0.00%
Sodium oxide(Na2O)	0.00%
Iron oxide(Fe2O3)	0.00%
Magnesium oxide(MgO)	0.00%
Calcium Oxide(CaO)	0.00%

2.2. Analysis of results

It is known that lead oxide is much more important than the other compounds, which leads to a classification pattern for high-potassium and lead-barium glasses. The chemical composition was classified by cluster analysis, which stipulates the mean ± standard deviation as the classification interval, and divided into n categories. The following table 2 shows the subclassification rules according to the lead-barium glass compounds (due to space limitations, this paper only shows the classification rules for the Ministry of Wind compounds)[3].

Table 2. Subclassification rules for lead-barium glass compounds

	Clustering categories (mean ± standard deviation)				F	P
	classification 1(n=20)	classification 2(n=17)	classification 3(n=6)	classification 4(n=6)		
Silicon Dioxide (SiO2)	58.497±8.088	29.488±6.449	19.593±13.747	19.352±4.906	69.189	0.000***
Sodium oxide (Na2O)	1.975±2.422	0.284±0.659	0.0±0.0	0.0±0.0	4.976	0.005***
Potassium oxide (K2O)	0.161±0.126	0.226±0.408	0.185±0.303	0.053±0.131	0.586	0.627
Calcium oxide(CaO)	1.2±0.952	2.961±1.71	1.598±1.296	2.757±2.149	5.151	0.004***
Magnesium oxide (MgO)	0.739±0.573	0.713±0.703	0.0±0.0	0.788±0.638	2.665	0.059*
Sulphur dioxide (SO2)	0.183±0.818	0.0±0.0	5.92±7.49	0.0±0.0	9.188	0.000***
Note: ***, **, * represent 1 per cent, 5 per cent and 10 per cent significance levels, respectively.						

The rules for subclassification of compounds based on high-potassium glass compounds are shown in table 3 below (due to space constraints, the table shows the rules for the classification of partial wind compounds)[4].

Table 3. Subclassification rules for high-potassium glass compounds

	Clustering categories (mean ± standard deviation)			F	P
	classification 2(n=9)	classification 1(n=6)	classification 3(n=3)		
Silicon Dioxide (SiO2)	63.624±3.558	93.963±1.734	81.063±5.368	145.911	0.000***
Sodium oxide (Na2O)	0.927±1.427	0.0±0.0	0.0±0.0	1.78	0.203
Potassium oxide (K2O)	10.818±2.37	0.543±0.445	4.87±4.718	32.212	0.000***
Calcium oxide(CaO)	6.363±2.64	0.87±0.488	2.24±2.363	12.979	0.001***
Magnesium oxide(MgO)	1.133±0.672	0.197±0.306	0.917±0.809	4.48	0.030**
Aluminium oxide (Al2O3)	7.349±2.346	1.93±0.964	4.433±1.603	14.929	0.000***
Sulphur dioxide (SO2)	0.136±0.205	0.0±0.0	0.0±0.0	1.839	0.193
Note: ***, **, * represent 1 per cent, 5 per cent and 10 per cent significance levels, respectively.					

The results of the classification of the high-potassium glass are shown in Table 4 below (due to space constraints, only the first eight artefacts are shown for the clustered categories)

Table 4. Results of the classification of high-potassium glasses

type of clustering	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	Artifact number
2	69.33	0	9.99	6.32	0.87	3.93	1.74	3.87	0	1
3	87.05	0	5.19	2.01	0	4.06	0	0.78	0.25	3
2	61.71	0	12.37	5.87	1.11	5.5	2.16	5.09	1.41	3
2	65.88	0	9.67	7.12	1.56	6.44	2.06	2.18	0	4
2	61.58	0	10.95	7.35	1.77	7.5	2.62	3.27	0	5
2	67.65	0	7.37	0	1.98	11.15	2.39	2.51	0.2	6
2	59.81	0	7.68	5.41	1.73	10.05	6.04	2.18	0.35	6
1	92.63	0	0	1.07	0	1.98	0.17	3.24	0	7

The results of the lead-barium glass classification are shown in Table 5 below (due to space constraints, only the first ten artefacts are shown for the clustered categories)

Cluster analysis, also known as cluster analysis, is a multivariate statistical analysis method to classify samples or indicators according to the principle of "clustering by class", the object of which is usually a large number of samples, which can be reasonably classified according to their respective characteristics. In the cluster analysis, the classification interval of the sample is calculated based on the mean and standard deviation, because the standard deviation reflects the degree of dispersion of the sample, i.e., the larger the standard deviation, the larger the gap between the samples, which further indicates the greater sensitivity. Example: From the results of lead-barium glass, it can be seen that the standard deviation of silica is generally larger than the standard deviation of other chemical compositions, and by analogy we can derive the sensitivity of other chemical compositions.

Table 5. Results of lead-barium glass classification

type of clustering	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3	CuO	PbO	Artifact number
2	36.28	0	1.05	2.34	1.18	5.73	1.86	0.26	47.43	2
3	20.14	0	0	1.48	0	1.34	0	10.41	28.68	8
3	4.61	0	0	3.19	0	1.11	0	3.14	32.45	8
2	33.59	0	0.21	3.51	0.71	2.69	0	4.93	25.39	11
2	29.64	0	0	2.93	0.59	3.57	1.33	3.51	42.82	19
3	37.36	0	0.71	0	0	5.45	1.51	4.78	9.3	20
1	53.79	7.92	0	0.5	0.71	1.42	0	2.99	16.98	23
3	31.94	0	0	0.47	0	1.59	0	8.46	29.14	24
1	50.61	2.31	0	0.63	0	1.9	1.55	1.12	31.9	25
3	19.79	0	0	1.44	0	0.7	0	10.57	29.53	26

3. Identification Modelling and Analysis

Assume that glass type is the dependent variable m, silica, NaO, KO, CaO, MgO, Al2O3, FeO, CuO, PbO, BaO, P2O5, Strontium Oxide, SnO, Sulphur Dioxide are the independent variables $h_1, h_2, h_3 \mathrm{B}\ h_{14}$, the regression coefficient is $\alpha_1, \alpha_2, \alpha_3 \mathrm{B}\ \alpha_{14}$, and the constant term is α_0. Then the multiple linear regression equation is set:

$$m = \alpha_0 + \alpha_1 h_1 + \alpha_2 h_2 + \mathrm{B} + \alpha_{14} h_{14} \quad (1)$$

Data regression was analysed using SPSSPRO yielding regression coefficients of 0.007, 0.051, -0.03, 0.004, 0.033, 0.033, 0.016, 0.029, 0.019, 0.034, 0.015, 0.097, 0.002, 0.016.Substituting the correlation coefficients into the regression equation yields:

$$m = 0.552 + 0.007b_1 + 0.051b_2 + \mathrm{B} + 0.002b_{13} + 0.016b_{14} \quad (2)$$

Based on the results of the analysis, it can be obtained that the value of R^2 is 0.852, which shows that the regression model has a good fit. It can be used.

Figure 1 below shows the fit of this regression equation:

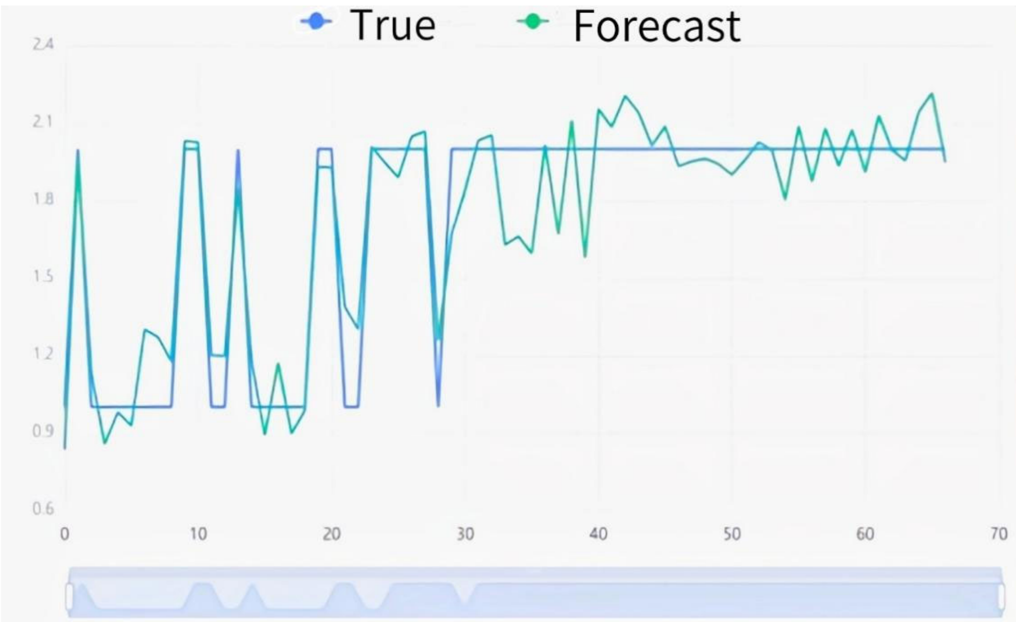

Figure 1. Plot of the fitted effect of the regression equation

Thresholds: The average of the results of the quantitative analyses of high-potassium and lead-barium is taken as the threshold value; samples below the threshold are classified as high-potassium, and samples above the threshold are classified as lead-barium. That is, when m < 1.5, the artefact is of high potassium glass type, and when m > 1.5, the artefact is of lead-barium glass.

The data were predicted and the predictions are shown in Table 6 below:

Table 6. Map of type prediction results

Artifact number	Type	Projected results
A1	High Potassium	1.393
A2	Lead Barium	1.78
A3	Lead Barium	1.883
A4	Lead Barium	1.971
A5	Lead barium	1.826
A6	High Potassium	1.169
A7	High Potassium	1.29
A8	Lead barium	1.455

Note: After testing the covariance of some samples of the model, the prediction result of the model has a slight error with the fact, and according to the analysis of the chemical composition of A8, we get that its type should be lead-barium.

Sensitivity analysis of the classification results: Since the prediction of the model has a certain error with the real value, it may be inconsistent with the reality when analysing some smaller data, and the sensitivity of the results is general.

4. Conclusions

Ancient glass is highly susceptible to weathering under the influence of the buried environment. During the weathering process, the internal elements are exchanged with the environmental elements in large quantities, which leads to changes in the proportion of its composition, thus affecting the correct judgement of its category. Therefore, it is of great research value to investigate the influence of different chemical compositions and chemical composition contents on the weathering or not of the surface of glass articles and the degree of weathering. In this regard, the cluster analysis model can be used to solve the problem. The grain, the presence or absence of weathering and their chemical compositions can be constructed as many features as possible according to the type of glass in order to form an interpretable classification law. For the purpose of subclassification, the glass types are first classified, then a chemical element content is selected then the colour of the glass grain is observed at that content, and finally the classification is made.

References

[1] DONG Han, ZOU Minghua, LI Lu et al. Classification and prediction of ancient glass compositions[J]. Journal of Xianyang Normal College,2 023, 38(04): 31-37.

[2] MA Peiying, HAN Yanlai, LI Delan et al. Analysis and classification of ancient glassware based on compositional data[J]. Mathematical Modelling and its Applications, 2023, 12(02): 63-73.

[3] T.Y. Deng,Y.H. Qiu,J.H. Chi. CLR-based compositional analysis and classification model for glass artefacts[J]. Mathematical Modelling and its Applications, 2023, 12(02): 41-51.

[4] Hsieh. Research on segmentation and classification method of ground glass lung nodules based on CT images[D]. Dalian Maritime University,2022.

[5] Shen Li-Lin. A study on the imaging characteristics of pulmonary ground glass nodules in smokers and their pathological classification prediction[D]. Wannan Medical College, 2022.