Composition Analysis of Glass Artefacts Based on Spearman's Coefficient and Fuzzy Correlation

Chenglong Yang

Department of Leicester International Institute, Dalian University of Technology, 124000, Panjin, Liaoning Province, China

1504267885@mail.dlut.edu.cn

Abstract. The identification and analysis of the composition of glass artefacts can support the related fields in exploring the history of glass and its processes, as well as the development and progress of glass composition. This paper identifies and analyses the composition of glass artefacts by collecting relevant data and information on glass artefacts through the Spearman correlation coefficient method and fuzzy correlation. In terms of data pre-processing, data with component ratios that did not add up to between 85% and 105% were removed, and then the degree of weathering on the surface of the glass artefacts was initially determined to be closely related to the glass type by applying a chi-square test, and then the highest correlation between the weathered surface of the glass and the glass type was determined by excluding possible interference caused by ornamentation and colour through partial correlation analysis. Fuzzy correlation analysis was then applied to remove abnormal data and calculate the mean values of each chemical composition for different types of glass at different degrees of weathering to discover statistical patterns. Finally, the mean values calculated by the fuzzy correlation analysis were summed using the difference to obtain the predicted results. Six chemical components were also screened out by the Spearman correlation coefficient method, and then the difference quotients of the same components were calculated by averaging these six components for different types of glass respectively. Finally, it was found that in addition to the three elements of potassium, lead and barium, calcium and strontium also have a greater influence on glass classification. The sample data were then clustered by the hierarchical clustering method based on the correlation coefficient method of calculating distances, and the optimal number of clusters was found by calculating the clustering contour coefficients. Then, the data of the clustered subclasses were analysed and subclasses such as soda-lime glass, sulphide glass and aluminosilicate glass were classified. Finally, the clustering clarity values were defined using the entropy weighting method to measure the clustering results, and the Held mean values at different p-values were adjusted for sensitivity analysis of the clustering clarity values.

Keywords: Spearman's Coefficient, Fuzzy Correlation, Component Analysis.

1. Introduction

The main raw material for glass is quartz sand with silica as the main component. Due to the high melting point of pure quartz sand, the glass making process requires fluxes and stabilisers [1]. Of these, the flux largely influences the chemical composition of the glass and also determines the type of glass [2]. An accurate quantitative analysis of the chemical composition is therefore an important prerequisite for determining the type of ancient glass. However, after being subjected to long periods of external weathering, the proportions of the composition of ancient glass can change significantly.

The variables of glass type, decoration and colour can be individually selected and their correlation with the surface weathering of the artefacts can be analysed one by one, and the interaction between the variables can be excluded. The correlation between the surface weathering of the glass artefacts and the glass type, decoration and colour can thus be analysed, and the correlation between the chemical composition content of the artefacts and the surface weathering can be analysed on the basis of the glass type, using the data from the weathering points to predict the proportion of the chemical composition of the corresponding samples before weathering.

The differences in the content distribution of each chemical component between the two glass types are investigated. The differences in the content distribution of each chemical component within

the same glass type are then analysed to establish a relevant mathematical model for the classification of the two glass types into subclasses, and finally the model parameters are adjusted according to the results of the classification using appropriate algorithms to complete the analysis of the rationality and sensitivity of the model. The final analysis of the composition of the glass artefacts is achieved.

2. Materials and methods

2.1. Spearman's correlation coefficient method

First, we measured the strength of the correlation between each chemical component and the type of glass, and identified the main chemical components that determined the classification pattern of the glass by machine learning the feature screening correlation coefficient method [3]. The Pearson correlation coefficient method is not applicable to this feature screening as the proportions of chemical components other than silica do not obey a normal distribution. In this paper, the Spearman correlation coefficient method was used to quantify the analysis and find the main chemical components [4].

A sample with a sample size of n and calculate the average of the descending positions of the two sets of original data to obtain the rank data x_i, y_i.

The newly obtained rank difference between the two sets of data is

$$d_i = x_i - y_i \tag{1}$$

Spearson's (Spearson) correlation coefficient is

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \tag{2}$$

2.2. Fuzzy correlation analysis method

The network The fuzzy correlation analysis was based on the chi-square test and the biased correlation analysis, and the invalid data from samples taken from Artefact 15 and Artefact 17 were excluded from the biased correlation analysis based on the requirement of a cumulative sum of proportions between 85% and 105%, while the chi-square test and biased correlation analysis were integrated and the integrated tables were divided into two parts according to the glass type of the artefact (high potassium and lead barium) [5].

Biased correlations between each chemical component and weathering were analysed in two separate parts of the tables using spss software, and a group of chemical components with a significance less than 0.05 and a larger absolute value of correlation was selected from each part of the tables based on the results of the analysis. The correlations and significance between the two groups of chemical constituents and each group of chemical constituents and weathering were obtained [6].

3. Model construction and testing

3.1. Spearman's correlation coefficient method

The article uses matlab to calculate and plot the distribution pattern of the correlation coefficients between each chemical composition and glass type, where the columns 1 to 14 correspond to the 14 chemical compositions in the title annex, as shown in Figure 1.

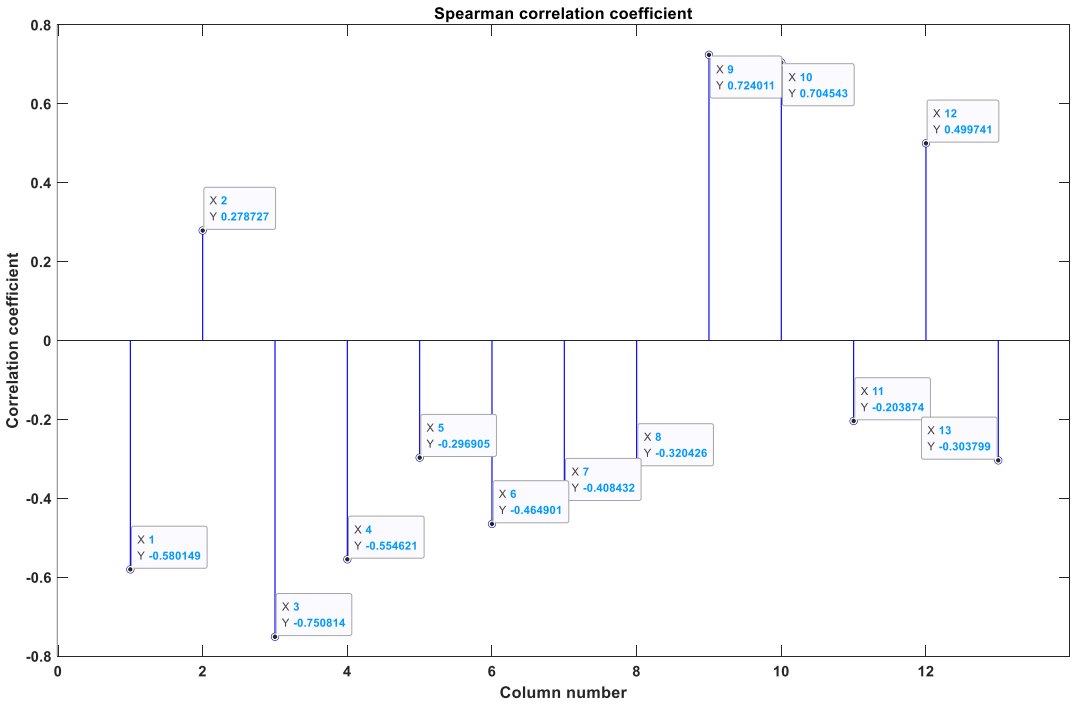


Figure 1. Distribution of correlation coefficients for each chemical composition and glass type

The six components with the highest absolute correlation coefficients were selected: silicon dioxide (SiO2), potassium oxide (K2O), calcium oxide (CaO), lead oxide (PbO), barium oxide (BaO) and strontium oxide (SrO). The arithmetic mean of each of these chemical compositions in the two types of glass products was found at m_1 m_2 (equivalent to the Held mean at the factor $p=1$).

The difference between the means of

$$D = m_1 - m_2 \tag{3}$$

Difference of means ratio.

$$R = \frac{m_1 - m_2}{m_1 + m_2} = \frac{D}{m_1 + m_2} \tag{4}$$

The range of values for the difference ratio of the means is R [-1, 1]. R The closer to 1, the more significant the proportion of the chemical in high potassium relative to the proportion in lead-barium; R the closer to -1, the more significant the proportion in lead-barium relative to the proportion in high potassium; the closer to 0, the less significant the difference in the proportion between the two types of glass R.

3.2. Fuzzy correlation model construction and testing

3.2.1 Cardinality test

Cross-tabulations of glass type, decoration and colour with surface weathering were plotted separately by SPSS and a chi-square test was done on the data from each table, while to determine whether the three variables mentioned above were correlated, we also performed a chi-square test on them two by two. The significance data obtained for each set of tests were as shown in Table 1.

Table 1. Significance data between variables

	Weathering	Glass type	Ornamentation	Colour
Glass type	0.020		0.001	0.002
Ornamentation	0.056	0.001		0.001
Colour	0.507	0.002	0.001	

3.2.2 Biased correlation analysis

In order to exclude the interference caused by the association between variables and to obtain highly accurate conclusions, biased correlation analysis was introduced in this study.

As colour, glass type, decoration and weathering are all definite class variables, they need to be converted to numerical values in advance and the rules for assignment are shown in table (no.) and table (no.) as shown in Table 2 and Table3.

Table 2. Assignment rules for glass type, decoration and weathering

Numerical values	Glass type	Ornamentation	Weathering
1	High Potassium	A	No weathering
2	Lead Barium	B	Weathering
3		C	

Table 3. Rules for the assignment of colours

Colour	Dark green	Green	Light green	Blue and green	Pale blue	Deep Blue	Purple	Black
Numerical values	1	2	3	4	5	6	7	8

3.2.3 Fuzzy correlation analysis

Using the two sets of chemical compositions derived from the chi-square test and partial correlation analysis, the data were pre-processed separately for the two parts of the table to identify outliers, one set of which is shown in the figure below as shown in Figure 2 and Figure3.

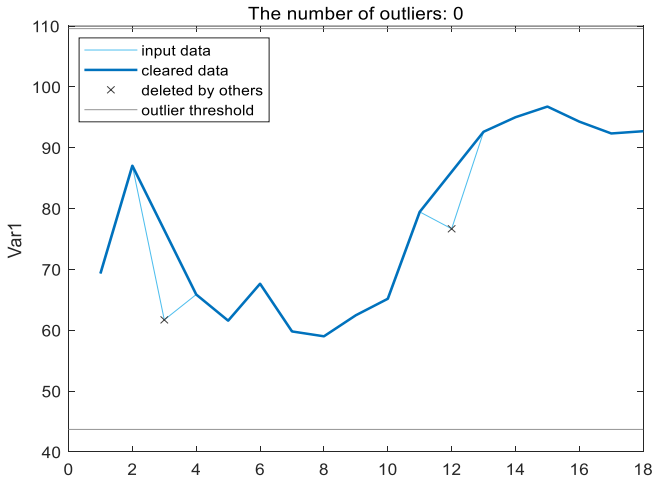


Figure 2. Data preproces results High potassium

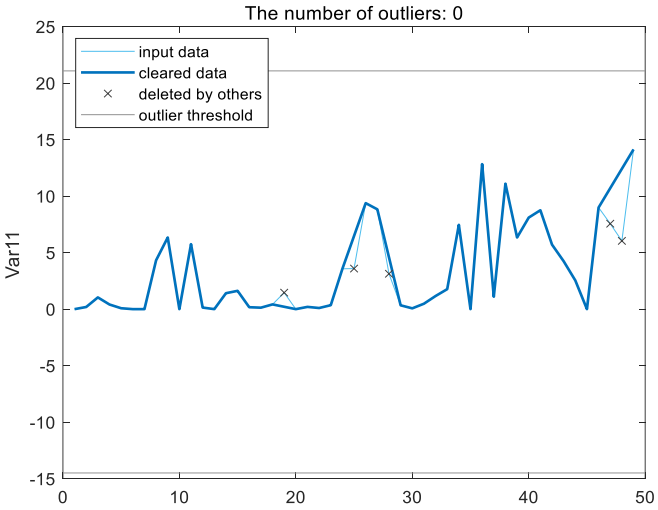


Figure 3. Data preproces results lead barium

Based on the pre-processing results, we identified three sets of outliers: data from the sample taken from Object 03 Site 2, data from the sample taken from Object 21, and data from the sample taken from the severely weathered site of Object 54.

4. Results and analysis

4.1. Analysis of Spearman's method model results

The mean values of the six components of silica, potassium oxide, calcium oxide, lead oxide, barium oxide and strontium oxide in glass products screened by using the Spearman's method model analysis and the difference between the mean values and the difference ratio are shown in Table 4.

Table 4. Means, differences and difference ratios of six components in two types glass product

Type	Silicon dioxide (SiO2)	Potassium oxide (K2O)	Calcium oxide (CaO)	Lead oxide (PbO)	Barium oxide (BaO)	Strontium oxide (SrO)
High potassium mean	69.06	10.16	5.45	0.26	0.24	0.04
Average lead and barium values	55.30	0.22	1.42	22.52	9.23	0.28
Difference D	13.76	9.94	4.04	-22.27	-8.99	-0.24
Difference ratio R	0.11	0.96	0.59	-0.98	-0.95	-0.74

As can be seen from the table: the R-value for silica is close to 0, indicating that the significance of its proportional difference between the two types of glass is low. Therefore the proportion of silica composition can only be used as a secondary factor to analyse the classification pattern of the two types of glass.

The R-value of potassium oxide is very close to 1, while the R-values of lead oxide and barium oxide are very close to -1, indicating that the variability in the proportion of these three chemical components between the two types of glass is very significant. Therefore, by determining the proportion of potassium oxide, lead oxide and barium oxide in the composition of the glass, the main factors of the first class of the glass classification law can be obtained.

The absolute values of the R-values of calcium oxide and strontium oxide, although not very close to 1, also show a multiplicative relationship in terms of the percentage content of the components in both glass classes. It can be used as a second class of major factors after potassium oxide, lead oxide and barium oxide to analyse the classification pattern of the two glass types.

The proportions of potassium oxide, calcium oxide and silica components are higher in high potassium glass compared to lead-barium glass; the proportions of lead oxide, barium oxide and strontium oxide components are higher in lead-barium glass compared to high potassium glass. Priority is given to the analysis of the classification pattern of the two glasses by the composition ratios of potassium oxide, lead oxide and barium oxide, followed by the analysis of calcium oxide and strontium oxide. The other chemical components, however, are not included as major factors in the analysis of classification patterns due to their low relevance relative to the six chemical components screened, and can be analysed as secondary factors.

4.2. Analysis of fuzzy correlation model results

By selecting After excluding outliers, a more precise biased correlation analysis of the two parts of the table was again carried out using spss, which allowed the positive, negative and magnitude of the correlation between each chemical component and weathering in the two types of artefact samples to be obtained.

The sample of artefacts with a high potassium glass type was also further divided into two groups, taking into account the degree of weathering.

a0: Glass type high potassium + surface unweathered

a1: Glass type high potassium + surface weathering

The sample of artefacts with glass type lead barium was further divided into three groups.

b0: Glass type lead barium + surface unweathered
b1: Glass type lead barium + surface weathering
b2: Glass type lead barium + heavily weathered surface

The arithmetic mean of the proportions of each chemical component was obtained separately in each group and some of the data obtained are shown in the table 5.

Table 5. Arithmetic mean of the proportions of some chemical components in each group

Glass type	Weathering	Mean value of SiO2 ratio (%)	Mean value of Na2O ratio (%)	Mean K2O ratio (%)
High Potassium	No weathering	67.742	0.834	9.960
	Weathering	93.963	0.000	0.543
Lead Barium	No weathering	54.660	1.683	0.219
	Weathering	27.056	0.244	0.133
	Severe weathering	4.165	0.000	0.200

For weathered samples, the difference between the mean value of the proportions of each chemical component in the group in which the sample is located and the mean value of the proportions of each chemical component in the unweathered group of the same glass type is first calculated, and then the proportions of each chemical component in this sample are subtracted from the corresponding difference to obtain the new proportions, and finally all the new proportions are normalised to obtain the predicted results when the sample is unweathered. The specific prediction results are shown in Table 6.

Table 6. Table of predicted results when unweathered

HerNum	HerSites	SiO2	Na2O	K2O	CaO	MgO	Al2O3	Fe2O3
07	07	0.67651	0.0085	0.0959	0.0564	0.00882	0.0695	0.0180
09	09	0.70035	0.0084	0.1018	0.0518	0.00881	0.0627	0.0195
10	10	0.71788	0.0084	0.1051	0.0476	0.00881	0.0575	0.0189
12	12	0.69434	0.0085	0.1063	0.0529	0.00883	0.0643	0.0192
22	22	0.67160	0.0084	0.1031	0.0622	0.01529	0.0847	0.0198
27	27	0.68362	0.0085	0.0968	0.0556	0.01445	0.0756	0.0185
02	02	0.61792	0.0139	0.0109	0.0085	0.01097	0.0685	0.0187
08	08	0.47131	0.0142	0.0008	0.0002	0.00000	0.0266	0.0007
11	11	0.63254	0.0148	0.0030	0.0212	0.00686	0.0418	0.0007
19	19	0.57180	0.0143	0.0008	0.0147	0.00543	0.0492	0.0140
26	26	0.46566	0.0141	0.0008	0.0000	0.00000	0.0202	0.0007
34	34	0.61558	0.0139	0.0032	0.0000	0.00000	0.0289	0.0053
36	36	0.64674	0.0352	0.0021	0.0000	0.00000	0.0284	0.0038
38	38	0.58186	0.0270	0.0008	0.0000	0.00000	0.0377	0.0035
39	39	0.52257	0.0139	0.0008	0.0000	0.00000	0.0180	0.0007
40	40	0.43149	0.0140	0.0008	0.0040	0.00000	0.0175	0.0025
41	41	0.48153	0.0150	0.0054	0.0366	0.02806	0.0490	0.0195
43	43 Part 1	0.40117	0.0144	0.0008	0.0379	0.00846	0.0361	0.0083
43	43 Part 2	0.50299	0.0146	0.0008	0.0504	0.00922	0.0486	0.0149
48	48	0.73706	0.0203	0.0036	0.0124	0.01360	0.1366	0.0100
49	49	0.58173	0.0148	0.0008	0.0322	0.01469	0.0695	0.0290
50	50	0.49806	0.0157	0.0009	0.0189	0.00463	0.0352	0.0044
51	51 Part 1	0.54023	0.0148	0.0008	0.0219	0.01184	0.0683	0.0131
51	51 Part 2	0.51709	0.0151	0.0009	0.0388	0.01483	0.0408	0.0052
52	52	0.55845	0.0278	0.0008	0.0085	0.00528	0.0263	0.0032
54	54	0.49379	0.0142	0.0040	0.0171	0.01221	0.0545	0.0007
56	56	0.59639	0.0151	0.0009	0.0000	0.00000	0.0337	0.0007
57	57	0.54199	0.0147	0.0008	0.0000	0.00000	0.0361	0.0007
58	58	0.57929	0.0143	0.0042	0.0203	0.00743	0.0487	0.0093
08	08 sevWeaSpots	0.56674	0.0173	0.0001	0.0145	0.00659	0.0454	0.0075
26	26 hea Wea Spots	0.54995	0.0170	0.0042	0.0124	0.00650	0.0455	0.0074

5. Conclusions and Discussion

5.1. Advantages of the model

A correlation analysis model combining chi-square and partial correlation analysis was developed to effectively reduce the interaction between variables while avoiding the high sensitivity caused by subjective weighting of variables.

The Spearman correlation coefficient method based on feature screening is used to screen compound variables that are more associated with glass categories. On the one hand, dimensionality reduction of the data is achieved; on the other hand, irrelevant variables are eliminated to improve the accuracy of subclassification.

5.2. Model shortcomings and improvements

The prediction model based on fuzzy correlation analysis is too homogeneous and the sample to be tested needs to be improved, which can be improved with the combined prediction of the combined model to increase the accuracy.

References

[1] Fang Xiangzhong. Cardinality distribution and cardinality test [J]. China Statistics, 2022, No. 05: 31-33 pp.

[2] Wang Xiaoyan. An introduction to rank correlation coefficient and Spearman's rank correlation coefficient [J]. Journal of Guangdong Light Industry Vocational Technology College, 2006, 04: 30-31

[3] Sun Yibing. The advantages and disadvantages of BP networks and their improvement [J]. Science and Technology Innovation Herald, 2009, 24: 24 pp.

[4] Liu J. PSO particle swarm algorithm in the generalization capability of neural networks [J]. Computer Engineering and Applications, 2009, Vol. 29: 38-40 + p.71

[5] Si Shoukui, Sun Zhaoliang. Mathematical modeling algorithms and applications, Beijing: National Defense Industry Press, 384-385 pp. 2015.

[6] Jiang Qiyuan, Xie Jinsing, Ye Jun. Mathematical Modeling (4th ed.). Beijing: Higher Education Press, 2011.