ISSN 2616-5880 Vol. 5, Issue 3: 104-109, DOI: 10.25236/AJMC.2024.050316

Research on classification of glass types of ancient cultural relics based on support vector machine model

Yifan Fu[1,*,#], Hang Guo[2,#], Liuhui Hong[2], An Gao[1], Wenxuan Zhou[2]

[1]Saxo Fintech Business School, University of Sanya, Sanya, 572022, China
[2]School of New Energy and Intelligent Networked Automobile (School of Science and Technology), University of Sanya, Sanya, 572022, China
[*]Corresponding author: 15501938775@163.com
[#]These authors contributed equally.

Abstract: *With the deepening of archaeological research, as a precious historical relic, the weathering of glass cultural relics has attracted more and more attention. In view of the effect of weathering on the chemical composition of glass cultural relics, two key questions are proposed in this paper: first, the chemical characteristics of the weathered glass before weathering are revealed by predicting the chemical composition of the weathered glass of lead-barium and high potassium type, that is, "composition reduction"; Secondly, a robust glass subclass classification model is constructed. The proposed model can not only accurately classify glass cultural relics when weathering causes the chemical composition to shift, but also ensure that the predicted deviation of the original composition of the weathered cultural relics does not significantly affect the classification results. By using the analytic hierarchy process (AHP) to select the key chemical components that can distinguish the glass sub-categories, the cultural relics are divided into four sub-categories: lead-barium - high sodium, lead-barium - low sodium, high potassium - low calcium and high magnesium, high potassium - high calcium and low magnesium. The corresponding classification model is established by support vector machine (SVM). The experimental results show that AHP model shows excellent classification accuracy and robustness for both unweathered data and weathered relics, which verifies its important application value in archaeological work.*

Keywords: *Glass cultural relics, chemical composition analysis, classification model*

1. Introduction

With the deepening of archaeological research, as a precious historical relic, the weathering of glass cultural relics has attracted more and more attention[1]. In view of the effect of weathering on the chemical composition of glass cultural relics, two key questions are proposed in this paper: first, the chemical characteristics of the weathered glass before weathering are revealed by predicting the chemical composition of the weathered glass of lead-barium and high potassium type[2], that is, "composition reduction"; Secondly, a robust glass subclass classification model is constructed. The proposed model can not only accurately classify glass cultural relics when weathering causes the chemical composition to shift, but also ensure that the predicted deviation of the original composition of the weathered cultural relics does not significantly affect the classification results[3].

Because the composition of glass is complex and easy to be affected by the environment, resulting in glass weathering and a large number of internal elements and external elements exchange, greatly affecting the archeologists' judgment of its type, therefore, the composition analysis, identification and subclassification of ancient glass have become a huge obstacle for archaeologists to study[4]. Based on principal component analysis, the classification rules of high potassium and lead barium glasses were studied, and the main chemical components were selected by hierarchical clustering method. Based on the data analysis of the chemical composition of ancient glass products, the classification method of glass relics was explored [5-6].

ISSN 2616-5880 Vol. 5, Issue 3: 104-109, DOI: 10.25236/AJMC.2024.050316

2. Define the data scope

Before category division, we first need to define the scope of analysis data[7], the same rule applies to the subsequent sub-category division; 1. Use only data from unweathered points; 2. Included variables: sodium oxide (Na_2O), calcium oxide (CaO), magnesium oxide (MgO), aluminum oxide (Al_2O_3), iron oxide (Fe_2O_3)² [8].

Table 1: Composition prediction of the weathered part of lead-barium glass before weathering

Heritage sampling site	Sodium oxide	oxidation	Potassium oxide	Calcium oxide	Magnesium oxide	Aluminum oxide
2	69.02	0.13	1.34	0.69	1.11	7.33
8	42.12	0.14	0.06	0.48	0.04	1.
08Severe weathering point	13.49	0.19	.	1.44	0.06	2.18
11	65.25	0.13	0.27	.95	0.68	3.51
19	63.88	0.14	0.06	0.9	0.63	5.1
26	41.53	0.14	0.06	0.47	0.04	0.99
26Severe weathering point	10.35	0.18	0.7	.29	0.06	2.21
34	67,61	0.12	12	21	0.04	2.06
36	66.7	.18	0.16	0.	0.03	1.81
8	62.12	4.3	0.05	2	0.04	3.26
39	59.89	0.15	0.06	0.39	0.05	0.7
40	46.59	0.18	0.07	0.81	0.06	084
41	47.65		0.76	1.98	3.5	78
43 part1	37.24	0,2	0.08	.4	1.8	.54
43part2	56.95	0.17	0.07	2.59	1.24	6
48	73,08	1.81	0.29	.6	1.05	12.58
49	62.5	0.14	0.06	1.54	1.58	7.85
50	47,64	10	0.07	1.31	0.62	3.8
51part1	56.21	0.15	0.06	1.26	1.35	8.06
51part2	59.34	0.18	0.07	2.2	2	4.60
52	58.1	4.55	0.06	0.79	0.62	1.76
54	52.57	0.16	0.51	1.16	1.5	6.58
54Severe weathering point	51.03	.2	0.08	0.02	1.64	732
56	60,74	0.14	0.06	0.39	0.04	2.59
57	54.92	0.14	0.06	0.44	0.04	3.17
58	63,55	0.14	0.48	1.13	0.82	4.
Iron oxide	Copper oxide	Lead oxide	Barium oxide	Phosphorus pentoxide	Strontium oxide	Sulfur dioxide
1.23	0.12	18.58	0.04	0.31	0.09	0.02
0.03	5.12	12.35	35.79	0.34	0.18	1.26
.04	7	19.55	49.1	1	0.37	10.25
0.03	2.34	10.16	15.55	0.83	0.17	0.02
1	8	19	6.32	0.86	0.1	0.02
0.03	547	12.76	37.08	0j	0,23	0.96
0.04	2.45	17.14	54.03	0.76	0.41	10.34
0.31	0.7	18.11	10.35	0.03	0.1	0.02
0.19	2	14.44	10	0.01	0.09	0.02
0.19	.4	19.15	10.1	0.04	0.18	0.02
0.03	0.49	28.67	9.03	0.12	0.33	0.02
0.18	0.03	40.31	10.22	0.22	0.45	0.03
1.51	0.12	23.45	13.8	0.87	0.29	0.02
0.79	2.93	36.98	11.99	0.01	0.46	0.03
1.27	0.97	24.18	4.69	1.53	0.29	0.02
0.49	0.01	4.43	5.49	0.07	0.08	0.01
2.07		15.28	7.26	1.09	0.24	0.02
0.3	0.73	24.01	20.62	0.76	0.42	0.02
05	0	18.93	11.19	0.84	0.21	0.02
0.41	0.51	29.38	0.06	1.1	0.03	0.03
0.18	0.39	22.04	10.68	0.58	0.24	0.02
0.03	0.48	26.94	91	0.45	0,49	0,.02
0.04	0.98	35.9	0.07	1.91	0.8	0.03
0.03	0.4	17.7	17.64	0.24	0.02	0.02
0.03	0.61	20.06	20.48	0	0.02	0.02
0.63	1.6	16.94	0.78	0.85	0.12	0.02

ISSN 2616-5880 Vol. 5, Issue 3: 104-109, DOI: 10.25236/AJMC.2024.050316

Table 2: Composition prediction of the weathered parts of high potassium glass before weathering

Heritage sampling site	Sodium oxide	oxidation	Potassium oxide	Calcium oxide	Magnesium oxide	Aluminum oxide
7	77,59	0.13	0.98	4,41	0.32	8.24
9	73,6	0.12	13,41	2.36	0.3	5.08
10	72.82	0.12	20.31	0.78	0.29	3.03
12	66.96	0.11	21.04	2.52	0.27	5.15
22	61.41	0.11	14.44	5.44	4.1	11.56
27	75,76	0.13	0.96	3.78	4.25	10.19
Iron oxide	Copper oxide	Copper oxide	Barium oxide	Phosphorus pentoxide	Strontium oxide	Strontium oxide
0.75	4.35	0.19	0.16	2.72	0.07	0.08
1.3	1.92	0.18	0.15	1.44	0.06	0.08
1.03	1.01	0.17	0.14	0.16	0.06	0.07
1.08	1.88	0.16	0.14	0.57	0.06	0.07
1.22	0.59	0.15	0.13	0.74	0.05	0.07
0.86	2.02	0.19	0.16	1.57	0.06	0.08

Combined with the data in Table 1, considering the limited amount of data available and the poor performance of machine learning based on decision trees, we first tried Logistic regression[9]. The model of the Logistic regression presents a significant overfitting, which implies that the sample is linearly separable. After observing the scatter diagram of the distribution of each component, barium oxide (BaO) and lead oxide (PbO) are selected as the main basis for the classification of high potassium glass and lead-barium glass. In order to give a more clear classification basis, we select support vector machine (SVM) suitable for small samples and high dimensional pattern recognition. Combined with the data in Table 2, support vector machine (SVM) is based on the principle of minimizing structural risk, and takes interval maximization as the learning strategy. Its algorithm is essentially an optimization algorithm for solving convex quadratic programming. After hierarchical clustering, there are 37 unweathering sampling point data of ancient glass samples, and the data dimension reaches 14 dimensions (14 chemical components)[10]. The current classification goal is to divide high-potassium glass and lead-barium glass into two categories, and further distinguish sub-categories of high-potassium glass and lead-barium glass. Multi-class or multi-binary support vector machines have a strong application value in this case, as shown in Figure 1.

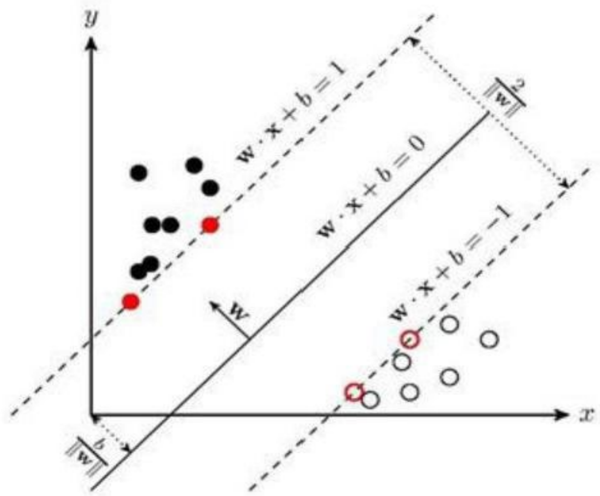

Figure 1: Schematic diagram of SVM

We assume that the dataset $(X_1,Y_1),(X_2,Y_2),..,(Xn,Y_7)$, where:

$$y(x) = w^2\psi(x) + b, \psi(x) \tag{1}$$

From this, we can deduce:

$$yi \cdot y(x;\ ) > 0 \tag{2}$$

In this case, the distance between each point and the dividing line can be expressed as:

ISSN 2616-5880 Vol. 5, Issue 3: 104-109, DOI: 10.25236/AJMC.2024.050316

$$\text{dist}=\frac{y_i(w^T\cdot\psi(x_i)+b)}{|w|} \tag{3}$$

Based on the structural risk minimization principle and interval maximization strategy, the model optimizes the function. Thereafter, the SVM introduces the Lagrange multiplier method to solve the objective function and obtain the optimal decision boundary.

The following kernel functions exist in support vector machines.

Linear kernel function:

$$K(x_i, x_j) = x_i^T x_j \tag{4}$$

Since the scatter plot composed of barium oxide (BaO) and lead oxide (PbO) contents in high-potassium glass and lead-barium glass is linearly separable, we directly choose the linear function as the kernel function of support vector machine, and the following results are obtained by analysis.

$$\xi Pbo = -3.51\xi BaO + 5.97 \tag{5}$$

Therefore, the classification law of high potassium glass and lead-barium glass can be summarized as follows:

$$\begin{gathered}\xi PsO + 3.51\xi BaO - 5.97 > 0,\ \text{The sample belongs to lead-barium glass}\\ \xi p6o + 3.51\xi BaO - 5.97 < 0,\ \text{The sample belongs to high potassium glass}\end{gathered} \tag{6}$$

After defining the two categories of high potassium glass and lead barium glass, we need to explore the internal sub-category classification standard of the two. Considering the large number of characteristic dimensions of data, the small number of samples and the sparse distribution, this paper first uses the hierarchical clustering method for exploratory analysis. Hierarchical clustering is a kind of clustering algorithm, which creates a hierarchical nested clustering tree by calculating the similarity between different categories of data points. In the clustering tree, the original data points of different categories are the lowest level of the tree, and the top level of the tree is the root node of a cluster. Euclidean distance:$d(x, y) = \sqrt{2} = {}_1(x;\ -y)^2$calculates the relative distance between variables. Ward's minimum variance method: Minimize the entire within-group variance. At each step, a pair of clusters with the smallest inter-cluster distance is merged. Cohesive clustering: Also known as AGNES (Cohesive nesting). Work in a bottom-up manner. Each object is initially considered as a single-element cluster (leaf), and at each step of the algorithm, the two most similar clusters are combined into new larger clusters (nodes). Iterate this process until all points are just members of a single large cluster (root). The results can be plotted as a dendrogram. Suppose n samples have been classified, and the number of classifications is $k, G_1, G_2, \ldots, Gk$ is denoted by k classes,G; The number of samples in the class is denoted by n_1, G; The center of gravity of the class is denoted by X, G; The ith sample of the class is denoted by Xj $(t = 1,2, \ldots, k)$, then G; The sum of squared deviations of samples in the class is expressed as:

$$W_t = \sum_{i=1}^{n_t}\left(X_i^t - \overline{X}^t\right)^T\left(X_i^t - \overline{X}^t\right) \tag{7}$$

Where XYDX is an m-dimensional vector, and the sum of squared total deviations of class k is:

$$W = \sum_{t=1}^{k} W_t = \sum_{t=1}^{k}\sum_{i=1}^{n_t}\left(X_i^t - \overline{X}^t\right)^T\left(X_i^t - \overline{X}^t\right) \tag{8}$$

Divide the n samples into one category, and W=0; When merging some two classes, the increase of W should be minimized; Repeat until all the samples are grouped together. The increased $\triangle W$ after the merger of two types is taken as the square distance between them, that is, $Dm^2 = W, -(W, +W)$ represents the square distance between Gp and G_2, then:

ISSN 2616-5880 Vol. 5, Issue 3: 104-109, DOI: 10.25236/AJMC.2024.050316

$$W_r = \sum_{i=1}^{n_r}\left(X_i^r - \overline{X}^r\right)^T\left(X_i^r - \overline{X}^r\right)$$
$$= \sum_{i=1}^{n_p}\left(X_i^p - \overline{X}^p\right)^T\left(X_i^p - \overline{X}^p\right) + \sum_{i=1}^{n_q}\left(X_i^q - \overline{X}^q\right)^T\left(X_i^q - \overline{X}^q\right) \tag{9}$$

Among them:

$$\overline{X}^r = \frac{1}{n_r}\left(n_p\overline{X}^p + n_q\overline{X}^q\right)$$
$$D_{pq} = W_r - \left(W_p + W_q\right) = \frac{n_p n_q}{n_r}\left(\overline{X}^p - \overline{X}^q\right)^T\left(\overline{X}^p - \overline{X}^q\right) = \frac{n_p n_q}{n_r} d_{pq}^2 \tag{10}$$

When G_2 and Gq are merged into G, the distance between G and other types of Gg is:

$$D_{rk}^2 = \frac{n_p n_q}{n_r}\left(X_r - \overline{X}^k\right)^T\left(X_r - \overline{X}^k\right) = \frac{n_p + n_k}{n_r + n_k}D_{pk}^2 + \frac{n_q + n_k}{n_r + n_k}D_{qk}^2 - \frac{n_k}{n_r + n_k}D_{pq}^2 \tag{11}$$

High potassium glass hierarchical clustering: observe the clustering hierarchical relationship, set the threshold (height=7) at the vertical line of the maximum distance, and obtain three clusters: cluster 1:3(position 1), 13, 14, 16; Cluster 2:1, 3(position 2), 4, 5, 6(position 2), 21; Cluster three: 6(position 1), 18. It can be found that the different clusters have a good discrimination degree in the three components of sodium oxide (Na_2O), calcium oxide (CaO) and magnesium oxide (MgO). Cluster 1: high sodium and high calcium; Cluster 2: low sodium and high calcium; Cluster 3: Low calcium and high magnesium. By observing the clustering hierarchy, the threshold value (height=12) was set at the vertical line of the maximum distance, and two clusters were obtained: cluster 1:42(position 2), 33, 23, 29, 32, 47, 30, 25, 44, 45, 28, 53, 35, 31; Cluster 2:49, 55, 20, 37, 42(position 1), 24, 30. It can be found that different clusters have a good discrimination degree in one component of sodium oxide (Na_2O) : cluster 1: high sodium; Cluster two: low sodium.

As the method of exploratory analysis, the hierarchical clustering method provides a reference for us to select the subcategories and corresponding indicators. In the scatter plot, it can be found that some components also have good discrimination, showing polar distribution (such as calcium oxide in lead-barium glass), but Na oxide has a greater impact on the distance calculation in the model, so the discrimination of this component is not a good reference for the sub-class discrimination of the model. At the same time, it is necessary to combine categorical variables such as color, surface weathering and chemical practices when distinguishing the subclasses, which are not considered in this model. In summary, combined with the results of the hierarchical clustering method, based on the simple and efficient goal, we set the following sub-categories: high potassium glass: high Na/low Na, high Ca/low Mg/low Mg/high Ca; Lead-barium glass: high sodium/low sodium;

We show that the classifier not only has a good classification effect on the unweaved data of the training model, but also has a very good classification effect on the unweaved data "restored" by the weaved data. And in the subsequent sensitivity analysis, it can be seen that the classifier can still achieve effective classification even if certain disturbance is applied to the intercept of the classification line.

Therefore, the existing sub-category classifiers of cultural relics have good robustness, which is most critically reflected in the fact that even though there will be some deviation in the "component reduction" of weathered cultural relics, this deviation will not significantly affect the classification results. Therefore, the classification model can effectively classify weathered glass relics, as shown in Table 3.

Table 3: Caption

The subclasses	Include a sample	Key Features
Gaona	23,25,36,38,42,44,45,47,48,52,53,54,55	Most of them are blue
Low sodium tolerance	2,8,11,19,20,24,26,28,29,30,31,32,33,34,35,37,39,40,41,43,46,49,50,51,54,56,57,58	
Low calcium and high magnesium	6,18	Most are blue and not easily weathered
High calcium and low magnesium	1,3,4,5,7,9,10,12,13,14,16,21,22,27	

ISSN 2616-5880 Vol. 5, Issue 3: 104-109, DOI: 10.25236/AJMC.2024.050316

3. Conclusion

Through the chemical composition analysis and modeling research of glass cultural relics, this study has achieved the following key results and conclusions: Firstly, the prediction of the chemical composition of glass cultural relics before weathering has been successfully realized, that is, through the analysis of the weathered glass, the chemical composition before weathering can be more accurately predicted. This provides conservation workers with a non-destructive predictive means to better understand and protect the original characteristics of glass artifacts. Secondly, a robust glass subclass classification model is developed, which can effectively distinguish different types of glass artifacts. This model still maintains high classification accuracy under the influence of weathering, which proves its reliability and practicality in practical application. The establishment of this model provides a scientific methodological support for the classification and protection of cultural relics, especially in the face of complex environment and long-term weathering. Finally, the results of this study are not only theoretical breakthroughs, but also have important practical applications. It provides new technical means and theoretical support for cultural relic protection workers, helps to better manage and protect the precious glass cultural relic heritage, and promotes the long-term maintenance and inheritance of cultural heritage.

References

[1] Gao, G. & Wang, Q. Based on support vector machine (SVM) classification of ancient glass products [J]. Journal of jiangxi normal university, 2024, (3): 19-22.
[2] Han Li, WANG Junyin. Composition analysis and type identification of ancient glass relics [J]. Journal of Yuxi Normal University,2024, 40(03):78-86.
[3] Shi Wei. Research on the composition of ancient Glass Products based on statistical analysis method [J]. Shandong business vocational college journal, 2024, 24 (02): 110-114.
[4] Wang B G, Jiang J J. Identification of ancient glass species based on composition data analysis and fuzzy pattern recognition [J]. Science & Technology Innovation and Application, 2024,14(07): 41-46.
[5] Qian H, Yue S, Xu L, et al. Ancient glass system discrimination method based on multivariate linear regression [J]. Journal of gansu science and technology, 2024, 40 (02): 56-61.
[6] Shao Guangming, Xia Xianqi, Yin Hejie. Based on CART and cluster analysis classified prediction model of ancient glass research [J]. Journal of tonghua normal university, 2024, (02) : 31-35
[7] Xie, H., Zhang, J., Guo, L. et al. Research on Mathematical Modeling of Properties and Chemical Composition of ancient Glass Products [J]. Journal of Dezhou University, 2023, 39(06):6-14.
[8] Chen, X., Ye, et al. Classification and identification of ancient glass cultural relics based on K-Means clustering and SVM algorithm [J]. Journal of Natural Science of Harbin Normal University, 2023, 39(04):70-79.
[9] Jiang, S., Chu, Z. l., Li, J. et al. Ancient relics glass chemical element correlation analysis [J]. Journal of chemical engineering and equipment, 2023, (7): 23-25.
[10] Shi, B., Zhao, X., Wang, H. et al. Based on multilayer perceptron network of ancient glass products category prediction [J]. Journal of lanzhou liberal arts college journal (natural science edition), 2023, 5 (3): 57-62.