Ancient Glass Products Identification and Composition Analysis Based on Support Vector Machine

Pinyue Guo*, #, Shiyi Yao#, Yuchi Ma

Jinan University, Guangzhou, Guangdong, 511443

*Corresponding author: guopinyue@stu2020.jnu.edu.cn

#These authors contributed equally.

Abstract. Studying the chemical composition and weathering-related properties of ancient glass holds immense significance as it served as a vital trading commodity along the early Silk Road in China. The texture transformations of ancient glass are closely linked to its chemical composition and storage conditions. Throughout the weathering process, internal elements continually interact with the surrounding environment, leading to alterations in their chemical proportions. To address this, this article has developed a classification method utilizing support vector machines for identifying glass types and analyzing their compositions. This article research provides archaeologists with a convenient tool for investigating ancient glasses, significantly enhancing the identification process.

Keywords: Non-Parametric Test, Clustering Analysis, Principal Component Analysis, Support Vector Machine, Ancient Glass Identification.

1. Introduction

Ancient glass artifacts, serving as significant trade goods along the early Silk Road in China, possess substantial historical research value. During the manufacturing process, silica and cosolvents are blended and melted, subsequently cooled and solidified at high temperatures to create glass products. The addition of various cosolvents yields diverse chemical composition ratios in the glass. Over extended periods of storage, the burial environment influences the glass, resulting in a certain degree of weathering. During this phase, internal elements interact with external elements, leading to alterations in the chemical composition proportions of the glass, deviating significantly from its original chemical makeup. Therefore, the study of glass's chemical composition holds great importance in the identification of glass artifacts.

Support machine learning as one of the most significant technologies in machine learning, have been widely applied in classification, regression and several other areas. In recent years, there are many improved methods proposed. For instance, for better solving the nonparallel cases, Liu L, Li P, Chu M, et al. proposed RNPSVM [1]. In order to solve the non-linear problems and analyzing the fuzzy dataset, Huang X, et al. proposed FSVM [2]. In discriminating between the elements of two massive datasets, O.L. Mangasarian proposed SORSVM [3]. In order to lower computational times, Haiyan Chen et al. proposed ILTSVM [4]. Aim at constructing two non-parallel hyperplanes to classify data and extracting features, NPHSVM was proposed by Feixiang Sun et al. [5]. FSVM is proposed by Sain H, Kuswanto H, Purnami S, et al.to solve the fuzzy time series dataset. [6] Though support vector machine has achieved great breakthrough, it is seldom used in the ancient glass classification.

On account of the importance of ancient glass on the Silk Road and unpredictable chemical contents in ancient glass due to the weather, this article proposes that use Cluster based on support vector machine. Here, the main work of this paper is given as follows:

(1) Verified the correlation between the material characteristic and weathering degree and study the chemical composition relationships between different glass types.

(2) Analyze the statistical patterns of weathering chemical composition content on the surface of samples based on two types of glass: high potassium and lead barium and establish an appropriate

prediction model to predict the chemical composition content of weathering detection points before weathering.

(3) Provide a method that using SVM based on clustering and principal component analysis to classify the ancient glass, and make a contribution to expand the application ancient glass.

2. Correlation and Difference Analysis

To determined glasses made from different materials, this article first studies the correlation and difference between two common materials with their material characteristics and chemical composition

2.1. Correlation

2.1.1 Material Characteristic

Based on available data, the weathering rate of high potassium glass is exceptionally high, reaching 70%, whereas lead barium glass exhibits a weathering rate of approximately 30%. This indicates that high potassium glass is more susceptible to weathering compared to lead barium glass. Notably, pattern B exhibits a weathering rate of 100%, significantly surpassing the weathering rates of pattern A and pattern C, which are around 50%. The weathering rate of pattern B is notably higher compared to other pattern types. It is worth mentioning that due to the uneven distribution of glass color, the correlation between weathering and color types is not significant.

2.1.2 Chemical Composition

For the chemical component potassium oxide, the undifferentiated content of high potassium is about 5% -15%, and the weathered content of high potassium is about 0% -5%, The undifferentiated content of most lead and barium is about 0% -0.5%, while the weathered content of most lead and barium is about 0% -0.5%. The above figure shows that potassium oxide has weathered lower than undifferentiated in high potassium glass, showing significant differences; In lead barium glass, the undifferentiated and weathered content ranges are consistent, and there is no significant difference.

2.2. Difference

2.2.1 Material Characteristic

This article observed the indicators and found that there was no continuity between them. In this paper, chi square test was used to evaluate the difference of disordered categorical variable. By analyzing the degree of deviation between the actual observation values of the sample and the theoretical values, the smaller the chi square value, the smaller the degree of deviation between the two; On the contrary, the situation is the opposite. Calculate using this formula for real value:

$$X^2=\sum\frac{(f_o-f_e)^2}{f_e} \tag{1}$$

Table 1. Statistical Test Between Material Type and Weathering

		Weathering		Total	X^2	Adjusted X^2	P
		N	Y				
Material Type	HPG	12	6	18	6.880	5.452	0.009***
	LBG	12	28	40			
Total		24	34	58			

Table 1 shows that due to the significance p-value of 0.009, it is less than 0.05. Refusing the original hypothesis, there is a significant difference between surface weathering and the type of cultural relic glass.

Similarly, this article can find there is no significant difference in surface weathering and cultural relic decoration types and also no significant difference in surface weathering and cultural relic color.

2.2.2 Chemical Composition

This article uses an independent t-test to analyze the influence of various chemical components on the degree of weathering:

$$S_c{}^2 = \frac{\sum X1^2 - \frac{(\sum X1)^2}{n1} + \sum X2^2 - \frac{(\sum X2)^2}{n2}}{n1+n2-2} \quad (2)$$

$$S_{\overline{X1}-\overline{X2}}{}^2 = \sqrt{S_c{}^2\left(\frac{1}{n1}+\frac{1}{n2}\right)} \quad (3)$$

$$t = \frac{(\overline{X1}-\overline{X2})-0}{S_{\overline{X1}-\overline{X2}}{}^2} \quad (4)$$

Among them, X1 and X2 are two different sets of variables; Is the average value of the sample, and n1 and n2 are the sample sizes, respectively.

With statistical test, due to p being less than 0.05, the original hypothesis is rejected. There are significant differences between the weathering of silicon dioxide, potassium oxide, calcium oxide, magnesium oxide, aluminum oxide, and iron oxide, and their chemical composition content has a significant impact on whether high potassium glass is weathered in both material type.

From the sample alone, the weathering ratio of high potassium glass is higher than that of lead barium glass. Silicon dioxide, potassium oxide, calcium oxide, magnesium oxide, alumina, and iron oxide have a significant impact on the weathering of high potassium glass, while silicon dioxide, sodium oxide, calcium oxide, phosphorus pentoxide, alumina, and lead oxide have a significant impact on the weathering of lead barium glass.

3. Predict Model

3.1. Set up Predict model

According to statistical rules, this article selects silicon dioxide, aluminum oxide, and copper oxide for chemical composition statistics to calculate the average weathering changes in Table 2.

Table 2. chemical composition statistics to calculate the average weathering changes

	Weathing	Detection Rate	Detection Interval	Weathering	Change Rate
SiO2					
HPG	Y	100%	92.35~96.77	93.96	1.382171227
	N	100%	59.01~87.05	67.98	0.723499361
LBG	Y	100%	3.72~53.33	24.91	0.455726308
	N	100%	31.94~75.51	54.66	2.194299478
Al2O3					
HPG	Y	100%	0.81~3.5	1.93	0.291540785
	N	100%	3.05~11.15	6.62	3.430051813
LBG	Y	100%	0.45~13.65	2.97	0.665919283
	N	100%	1.42~14.34	4.46	1.501683502
CuO					
HPG	Y	100%	0.55~3.24	1.56	0.530612245
	N	100%	0.47~5.09	2.94	1.884615385
LBG	Y	100%	0.19~10.57	2.67	1.543352601
	N	100%	0.11~8.46	1.73	0.647940075

Due to the low rate of change, the difference before and after weathering is within a controllable range. This article can predict it using the following chemical formula:

*pre weathering chemical composition content=post weathering content * reverse weathering change rate.*

3.2. Predict Result

Table 3. Chemical composition content before weathering point of high potassium glass

Sample Number	Relic Number	Sample Point	SiO2	Al2O3	CuO
9	07	07	67.01774585	6.791502591	6.106153846
12	09	09	68.74690932	4.527668394	2.921153846
13	10	10	70.01303321	2.778341969	1.583076923
15	12	12	68.21875479	5.007875648	3.109615385
25	22	22	66.81516603	12.00518135	1.036538462
31	27	27	67.08286079	8.609430052	2.902307692

Table 4. Chemical composition content before weathering point of lead barium glass

Sample Number	Relic Number	Sample Point	SiO2	Al2O3	CuO
2	02	02	79.60918507	8.604646465	0.168464419
10	08	08	44.19319149	2.012255892	6.74505618
11	08	08*	10.11572059	1.666868687	2.034531835
14	11	11	73.70651947	4.03952862	3.194344569
22	19	19	65.03903653	5.361010101	2.274269663
29	26	26	43.42518667	1.051178451	6.848726592

Table 3 and Table 4 illustrate the predict result of chemical composition content before weathering point of high potassium glass and lead barium glass, separately.

4. SVM Classification Model

4.1. The basic fundamental of support vector machine

4.1.1 The principle of non-linear support vector machine

Support Vector Machine (SVM) is a type of learning machine based on statistical learning theory. The basic principle to apply SVM into classification is as follows. First, using the kernel function to map the input vectors into a higher dimensional feature space. Next, construct an optimal hyperplane which can separate the datasets into different classes.

Assume that the training data is the form as $\{x_i\}_n, i = 1,2,\ldots,n;\ x \in R^n$, with corresponding labels $y \in \{+1,-1\}$ where -1 and +1 are used to present the different two classes. This article goal is to seek an optimal binary classifier from the training datasets, which can make as small as possible mistakes in the testing datasets. And in this paper, the non-linear separable case will be introduced.[7][8]

In order to prevent overfitting, the relaxation variable $\delta \geq 0$ is introduced into represent the deviation of the sample distance from the support plane, which can be expressed as the following optimization problem:

$$\min \boldsymbol{\varphi}(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||^2 + C\ [\textstyle\sum_{i=1}^{n} \boldsymbol{\delta}_i] \tag{5}$$

$$\text{s.t.}\quad \mathbf{y_i}[\mathbf{x_i} \cdot \mathbf{w} + \mathbf{b}] - \mathbf{1} + \boldsymbol{\delta}_i \geq \mathbf{0},\ \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{n}$$

where C is the penalty coefficient, which represents the degree of punishment for misclassified samples.

In order to transform inequality constraints into equality constraints, Lagrange multiplier is used to solve the problem. By introducing multiplier $\alpha_i \geq 0$, the objective function is transformed into:

$$L(w,b,\alpha) = \frac{1}{2}||w||^2 + C[\sum_{i=1}^{n}\delta_i]\text{-}\sum_{i=1}^{n}\alpha_i\{\ y_i[x_i \cdot w + b] - 1\} \tag{6}$$

According to the derivative:

$$\frac{\partial}{\partial w}L(w,b,\alpha)=w - \sum_{i=1}^{n}\alpha_i y_i x_i = 0 \tag{7}$$

$$\frac{\partial}{\partial b}L(w,b,\alpha) = \sum_{i=1}^{n}\alpha_i y_i = 0 \tag{8}$$

Then apply the equation (7) and equation (8) into (6), this article obtains:

$$L(w,b,\alpha) = \sum_{i=1}^{n}\alpha_i\text{-}\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j(x_i \cdot x_j) \tag{9}$$

And in order to solve the problem better, this article transforms it into dual problem:

$$\max\ \ L(w,b,\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i \cdot x_j\right) \tag{10}$$

$$\text{s.t. } \sum_{i=1}^{n}\alpha_i y_i = 0\,,\ C \geq \alpha_i \geq 0, i = 1,2,\ldots,n$$

If α_i^* is the optimal solution, then $w^* = \sum_{i=1}^{n}\alpha_i^* y_i x_i$.

According to the KKT condition, when the product of Lagrange multipliers and the constraint is equal to 0, the optimal value is $\alpha_i[y_i(x_i \cdot w + b) - 1] = 0,\ i = 1,2,\ldots,n.$ For most points $\alpha_i = 0$ that are not in the hyperplane, then α_i^* which is not 0, corresponds to the support vector, and w^* is a linear combination of the support vectors.

4.1.2 The selection of Kernal function

Assume that the nonlinear mapping: $\pi: R^n \rightarrow H$maps the samples in the input space to the high-dimensional feature space H, and the training algorithm only needs to use the inner product in the feature space when constructing the optimal hyperplane, denoted by $\pi(x_i) \cdot \pi(x_j)$. According to Mercer theory [8], this article can find the kernel function $K(x_i, x_j) = \pi(x_i) \cdot \pi(x_j)$. Then the problem can be optimized as:

$$\max\quad L(w,b,\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j)) \tag{11}$$

$$\text{s.t. } \sum_{i=1}^{n}\alpha_i y_i = 0\,,\ C \geq \alpha_i \geq 0, i = 1,2,\ldots,n$$

Gaussian RBF (radical basis function) function is chosen as kernel function in this paper because of its excellent results for real-world application.[9][10]

4.2. The application of SVM in ancient glassware classification

4.2.1 The establishment of SVM classification model

The ancient glassware identification is a binary classification. This article uses the percentage of chemical composition in the ancient glassware as the input features, corresponding with K and Ba as the label.

4.2.2 The data pre-processing

Before the experiment, this article needs to pre-process the datasets. First, assume the black in the dataset indicates that this component is not detected, then the missing chemical component is filled as 0.

Then this article numbered each cultural relics sampling point sequentially and calculated the sum of each component of the detection point. Detection points 19 and 21 were found to be 79.41% and 71.89% respectively, both of which were less than 85%, so they were regarded as invalid points.

Next, in order to map the data into the standard 0-100%, this article normalized the data to facilitate the preprocessing and analysis the data.

Maximum and minimum standardization formula:

$$x' = \frac{x - xmin}{xmax - xmin} \tag{12}$$

4.3. The results

By the confusion matrix heat map of the training dataset and testing dataset, the average accuracy of the training dataset and testing dataset can be achieved 100%, which are shown in the Figure 1 and Figure 2.

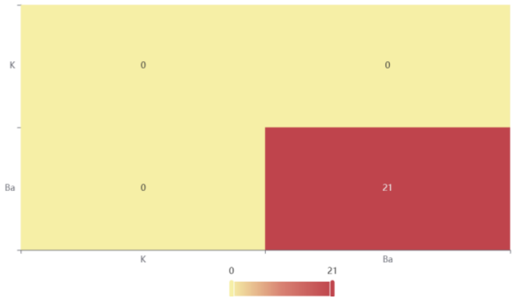

Figure 1. Training dataset

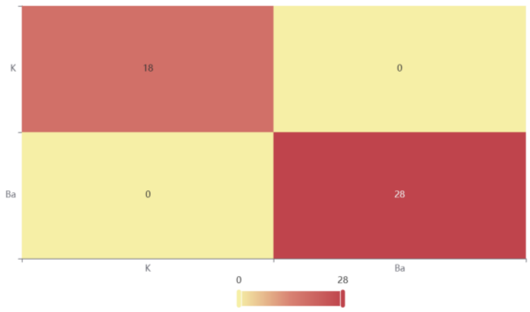

Figure 2. Testing dataset

With the high accuracy of the SVM classification model, this article uses it to predict the unknown ancient glassware type. And the predicted results as the following Table 5.

Table 5. The predicted results of unknown ancient glassware

No.	The predicted probability of K	The predicted probability of Ba	Predicted result
A1	0.868802717	0.131197283	K
A2	0.080611623	0.919388377	Ba
A3	0.002177486	0.997822514	Ba
A4	0.026496451	0.973503549	Ba
A5	0.388843448	0.611156552	Ba
A6	0.810271249	0.189728751	K
A7	0.789245265	0.210754735	K
A8	0.042621873	0.957378127	Ba

5. Conclusion

This article conducted an analysis of the statistical patterns of weathering-related chemical composition content on the surface of glass samples, specifically focusing on two types: high potassium and lead barium glass. This article goal was to establish a prediction model capable of predicting the chemical composition content of weathering detection points prior to weathering.

To further explore this topic, this article proposed a clustering and principal component analysis-based Support Vector Machine (SVM) approach for the classification of ancient glass. Initially, clustering analysis was employed to reduce data dimensionality and quantitatively analyze the chemical content. The results revealed a significant distinction between K glass and Ba glass. Subsequently, principal component analysis was utilized to obtain scores and determine the substantial contribution rate of principal components in explaining variables. Finally, an SVM model was established, exhibiting excellent performance on both the training and test datasets. The results demonstrated that SVM is highly accurate and efficient in the classification of ancient glass. Moreover, this article model exhibits high flexibility and lower computational complexity, making it suitable for low-cost and accurate identification of ancient glass products.

While SVM has proven effective in classification tasks, this article believes that integrating existing efficient methods such as minimizing the leave-one-out error bound [11] into this article model could further enhance its performance. Additionally, this article finds the extension to multiclass classification, regression, and semi-supervised learning [12] intriguing and are actively considering their implementation.

References

[1] Liu L, Li P, Chu M, et al. Robust nonparallel support vector machine with privileged information for pattern recognition[J]. International Journal of Machine Learning and Cybernetics, 2023, 14(4): 1465-1482.

[2] Huang X, Zhang S, Lin C, et al. Quantum Fuzzy Support Vector Machine for Binary Classification[J]. Computer Systems Science & Engineering, 2023, 45(3).

[3] O. L. Mangasarian and D. R. Musicant. Successive overrelaxation for support vector machines. IEEE Transactions on Neural Networks, 10:1032–1037, 1999.

[4] Chen H, Yu Y, Jia Y, et al. Incremental learning for transductive support vector machine[J]. Pattern Recognition, 2023, 133: 108982.

[5] F. Sun and X. Xie, "Deep Non-Parallel Hyperplane Support Vector Machine for Classification," in IEEE Access, vol. 11, pp. 7759-7767, 2023.

[6] Sain H, Kuswanto H, Purnami S, et al. Fuzzy support vector machine for classification of time series data: A simulation study[J]. Decision Science Letters, 2023, 12(3): 487-498.

[7] Roushangar K, Ghasempour R. Supporting vector machines[M]//Handbook of Hydroinformatics. Elsevier, 2023: 411-422.

[8] Zhang X D, Zhang X D. Support vector machines[J]. A Matrix Algebra Approach to Artificial Intelligence, 2020: 617-679.

[9] Paine A E, Elfving V E, Kyriienko O. Quantum kernel methods for solving regression problems and differential equations[J]. Physical Review A, 2023, 107(3): 032428.

[10] Ali I M S, Hariprasad D. Hyper-heuristic salp swarm optimization of multi-kernel support vector machines for big data classification[J]. International Journal of Information Technology, 2023: 1-13.

[11] Xie X, Li Y, Sun S. Deep multi-view multiclass twin support vector machines[J]. Information Fusion, 2023, 91: 80-92.

[12] Yu G, Ma J, Xie C. Hessian scatter regularized twin support vector machine for semi-supervised classification[J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105751.