Analytical identification model of ancient glass based on density clustering and PSO-SVM

Aosen Sha *

School of Statistics, Dongbei University of Finance and Economics, Dalian, China

* Corresponding Author Email: s15840118633@sina.com

Abstract. In this paper, the correlation between the indicators of glass artifacts is explored in depth by establishing relevant mathematical models, giving the statistical laws of weathering for high potassium glass and lead-barium glass, and classifying a series of glass artifacts. In this paper, a support vector machine (PSO-SVM) model based on particle swarm optimization is developed. Firstly, the SVM model is proposed to achieve binary classification for the model, while the particle swarm model is established, and the hyperparameters c and g are optimized by adding adaptive particle variation operation and velocity contraction factor, and the optimal values of both are 0.248 and 2.244, respectively. and the training and validation sets are continuously changed by K-fold cross-validation, and the accuracy of the model classification is obtained as 100%. The model was then used to classify the Form 3 data, and the results were that the artifacts in groups 2, 3, 4, 5, and 8 were lead-barium glass; the artifacts in groups 1, 6, and 7 were high-potassium glass. Finally, the sensitivity analysis of the model was carried out to adjust the values of hyperparameters c and g. It was found that the model sensitivity was very low and the results were satisfactory. In addition, the spearman correlation coefficients between the individual chemical compositions of the two types of glasses were calculated, plotted thermodynamically and interpreted. Then, a gray correlation analysis model was developed to investigate the relationship between other chemical components and it by using silica as the parent sequence, and conclusions were drawn. And based on the results, the differences in the correlation between the chemical components of the two types of glasses were analyzed, and it was found that the correlation between some components in the two types of glasses and the silica content were significantly different.

Keywords: Glass; Subclass classification; PSO-SVM model; DBSCAN density clustering model spearman correlation coefficient

1. Introduction

Since ancient times, China has been using glass products to trade with Chinese and Western countries. The main chemical composition of the raw material for glass production is silica [1]. In order to improve the efficiency of the production process, it is necessary to add fluxes during refining, thus lowering the melting temperature; limestone is added as a stabilizer, and limestone is converted into calcium oxide after calcination [2]. Due to the different fluxes added, the main chemical composition of glass is different, mainly divided into lead-barium glass and high-potassium glass.

Ancient glass is prone to weathering, and the weathering process is actually a large exchange of internal elements of glass with environmental elements, resulting in a change in the proportion of the chemical composition of glass. However, for the surface unweathered glass, there are also local weathering points; for the surface weathered glass, there are also local unweathered points [3].

This study intends to address the following questions.

1) Analyze the chemical composition of glass artifacts from unknown categories; determine the categories to which they belong; and analyze the sensitivity of the classification results.

2) To analyze the correlations between the main chemical components of glass artifact samples from different categories, and to compare the differences in the correlations between the main chemical components of glass artifacts from different categories.

2. Model building and solving

2.1. Problem 1

2.1.1 Particle swarm algorithm to optimize the solution of support vector machine models

The specific steps of the particle swarm algorithm to optimize the support vector machine model are shown in Figure 1.

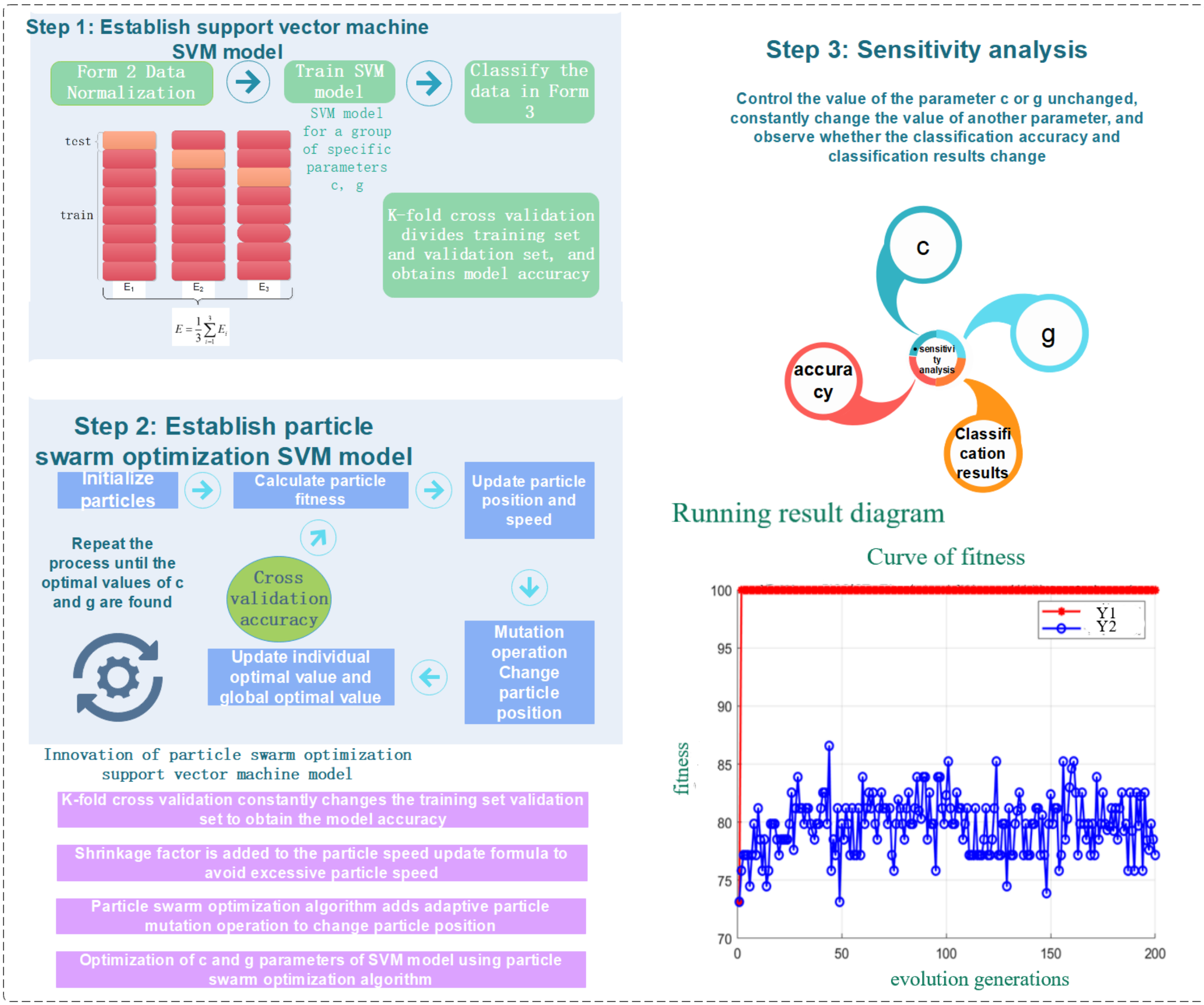

Figure 1. Particle swarm optimization support vector machine model steps

The Form 2 data is normalized and the training and validation sets are divided using tri-fold cross-validation. Set the value range of parameter c in the SVM model to [0.1,100], and the value range of parameter g to [10^{-4},10^{3}][4]. Using the particle swarm algorithm to search for the parameters c and g, this paper adds the adaptive particle mutation operation, which introduces the mutation operation into a random operator to make the position of the particle mutate. When the particle undergoes position mutation, it will change its original search direction, so that it enters other regions for searching and may find new individual optimal values and global optimal values. By iterating in this way, the mutation algorithm increases the probability of finding the global optimal solution [5].

The acceleration constants C1 and C2 are 2.05 and 2.05, respectively, and are noted as $C = c_1 + c_2 = 4.1$.

Define the velocity contraction factor as

$$\Phi = \frac{2}{|(2-C-\sqrt{C^2-4C})|} \tag{1}$$

The value used to compress the velocity to avoid the particle velocity too fast and to ensure the comprehensiveness of the particle search.

Define the i-th particle velocity update equation as

$$v_i^d = \Phi \times \left[\omega v_i^{d-1} + c_1 \cdot random(0,1) \cdot \left({P_{best}}_i^d - x_i^d\right) + c_2 \cdot random(0,1) \cdot \left({G_{best}}^d - x_i^d\right)\right] \quad (2)$$

where d is the number of current iterations, ω is the inertia weight taken as 0.9, and random(0,1) denotes the random number on the interval [0,1]. P_{kbest}, G_{kbest} are the local optimal solution and the global optimal solution of the particle in the kth generation, respectively [6].

The meaning of other parameters in the code is shown in the Table 1.

Table 1. Meaning of other parameters in the code

Parameters	Meaning	Takes values
k	Ratio of rate to position	0.6
sizepop	Maximum population size	20
Max-Gen	Maximum number of iterations	200
C_1	Acceleration constants	2.05
C_2	Acceleration constants	2.05

The optimal c value of 0.248 and the optimal g value of 2.244 were calculated using the model in the MATLAB environment, and the classification results were that artifacts in groups 2, 3, 4, 5, and 8 were lead-barium glass; artifacts in groups 1, 6, and 7 were high-potassium glass. The tri-fold cross-validation yielded 100% accuracy of the model (i.e., particle fitness), and the results were satisfactory.

Observation of the chemical composition of the artifacts, found that lead barium glass artifacts of lead oxide content are much higher than the other group, barium oxide content is mostly higher than the other group, while silica content is much lower than the other groups. High potassium glass artifacts of lead oxide and barium oxide content are 0, while containing potassium oxide and a large amount of silica. It is worth mentioning that the content of potassium oxide, lead oxide, and barium oxide in Group 1 is zero, but we can infer from the content of silica that it is more likely to belong to high potassium glass, and the reason for the zero content of potassium oxide is likely to be related to its degree of weathering. In summary, it shows that the classification results of the model are more reasonable.

2.1.2 Sensitivity Analysis

Sensitivity analysis of the classification results requires continuously changing the values of parameters c and g in the SVM model to observe the changes in the classification accuracy and classification results.

First, the size of c was controlled to be 0.248 and the size of g was continuously changed, and the resulting results are shown in Table 2.

Table 2. Sensitivity analysis of parameter g

g	Classification Accuracy	Does the classification result change
1	100%	Unchanged
4	100%	Unchanged
6	100%	Unchanged
8	98.51%	Unchanged
10	97.01%	Unchanged

Then controlling the size of g to be 2.244 and continuously changing the size of c, the result is shown in Table 3.

Table 3. Sensitivity analysis of the parameter p

c	Classification Accuracy	Does the classification result change
0.1	100%	Unchanged
1.5	100%	Unchanged
3	100%	Unchanged
5	98.51%	Unchanged
7	98.51%	Unchanged

Observing the table, it is clear that when g is constant or c is constant, changing the value of another parameter, the model results remain essentially the same with low sensitivity and satisfactory results [7].

2.2. Problem 2

In order to investigate the correlation between the chemical components of different categories of glass artifact samples, correlation analysis among the chemical components is needed. In this paper, we will use the method of gray correlation analysis, combined with the Spearman coefficient, to explain the relationship between each chemical composition.

2.2.1 Calculation of Spearman's coefficient and drawing of heat diagram

After the above analysis, it can be concluded that the distribution of each index data does not all obey normal distribution, so we choose the Spearman coefficient, which is more applicable than the Pearson coefficient, to describe the correlation between the components, and use Python to draw the heat map based on this(Figure 2) [8].

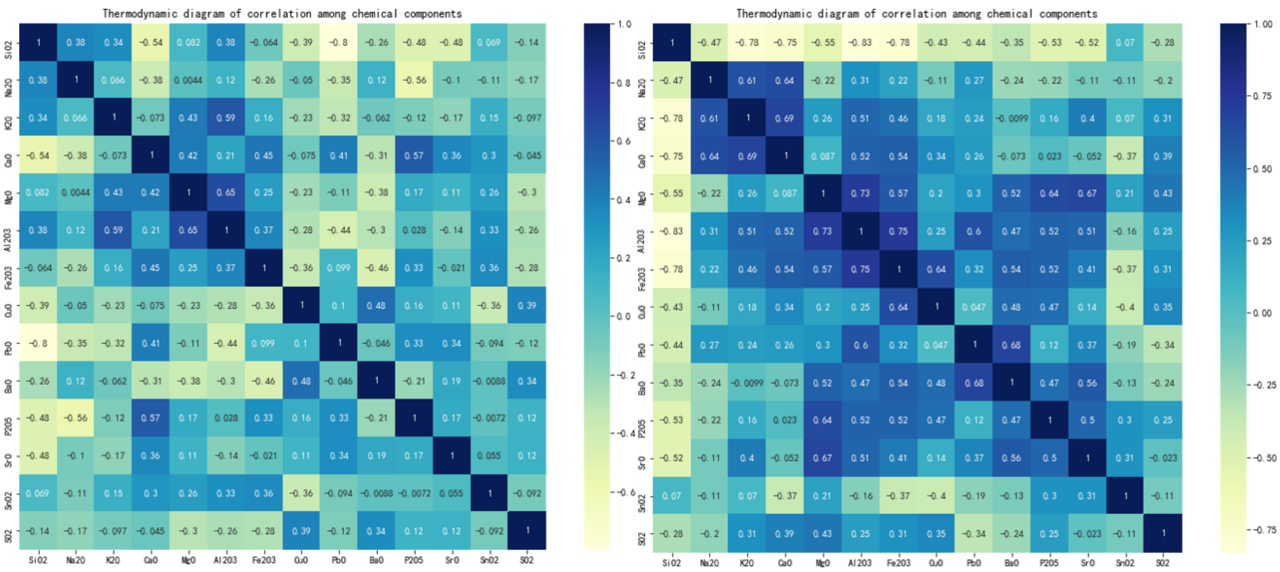

Figure 2. Spielman Heat Map

2.2.2 Analysis of the correlation between the chemical components in high potassium glass types

(1) Visual analysis of the thermogram

Observation of the thermogram shows that silica has a relationship with almost all other components, especially with potassium oxide, calcium oxide, aluminum oxide and iron oxide; in addition, the content of alumina has a strong positive correlation with iron oxide and magnesium oxide respectively.

According to the relevant literature, we can understand that many types of ores were added in the process of glass making in ancient China, and iron ore and aluminum ore are both non-renewable metal ores, which often exist at the same time; magnesium oxide can be used as flux, and aluminum and most of its compounds have high melting points, so the correlation between the two is very reasonable.

(2) Gray correlation analysis based on Spearman's coefficient

Based on the comparison of the Spearman coefficients among the chemical components above, the specificity of the silica content was found. Therefore, in this paper, the silica content was used as the parent series and the contents of other chemical components were used as sub-series for gray correlation analysis [3], and the correlation results between the contents of each component and the silica content were obtained as follows (Figure 3).

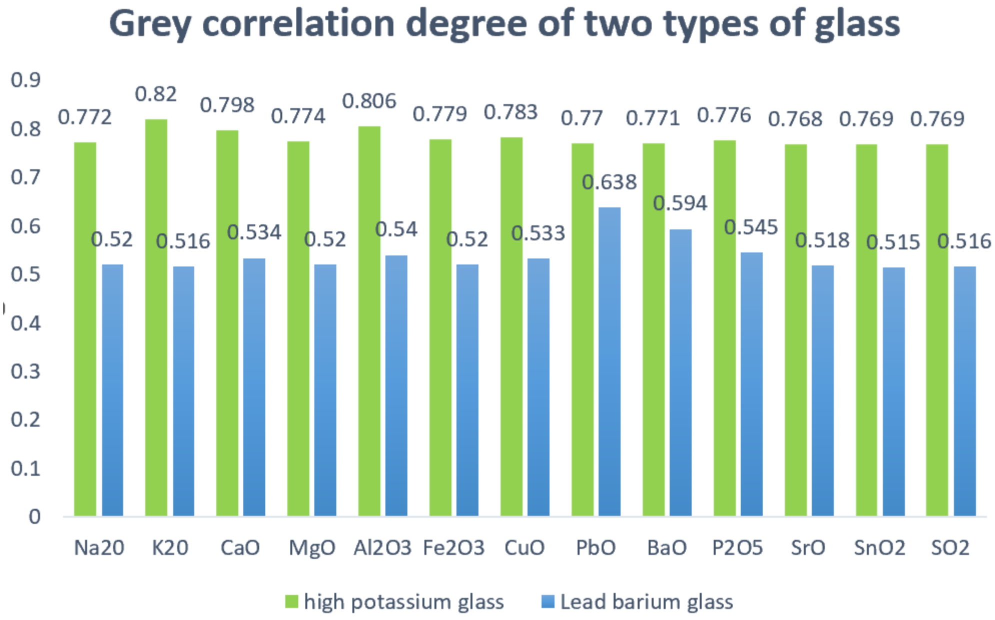

Figure 3. Two types of glass gray correlation

The results show a stronger correlation between potassium oxide, alumina, calcium oxide and silica, and this result overlaps highly with the results shown in the thermogram, indicating that the results obtained are more reasonable.

Considering that most of the above chemical components come from the main raw materials and main fluxes for the manufacture of high-potassium glass, the ratio of quartz sand (the main component is silicon) to other raw materials (limestone, aluminum ore) or fluxes (grass wood ash) may affect the performance and appearance of the final glass products under the same conditions of total raw material and product yield. As a multifunctional item, our ancient glass products are very rich in variety, style and use. Therefore, it is reasonable that such a correlation exists for these chemical components.

2.2.3 Analysis of the correlation between the chemical components in the type of lead barium glass

(1) Visual analysis of the thermogram

Observing the thermogram, it can be seen that there is a strong negative correlation between the content of silica and lead oxide, and there is a certain positive correlation between the content of magnesium oxide and alumina, and it can be presumed that the reason for this relationship is the same as that of high potassium glass [9].

(2) Gray correlation analysis based on Spearman's coefficient

Observe the above correlation results, combined with the research on the composition of lead-barium glass in question 2 of this paper, it can be seen that one of the important characteristics of lead-barium glass is the high content of lead and relatively low content of silica, so there is a very obvious negative correlation between them, the results of the heat map and the results of the gray correlation analysis both verify this conclusion.

2.2.4 Analysis of the variability of chemical composition correlations between different types of glasses

On the whole, the correlation between the chemical components of high potassium glass is slightly stronger than that of lead-barium glass.

Looking at the correlation between some chemical components of the two types of glass separately, it is easy to find that the difference between high potassium glass and lead-barium glass mainly lies in the relationship between the content of other components and the silica content (Figure 4).

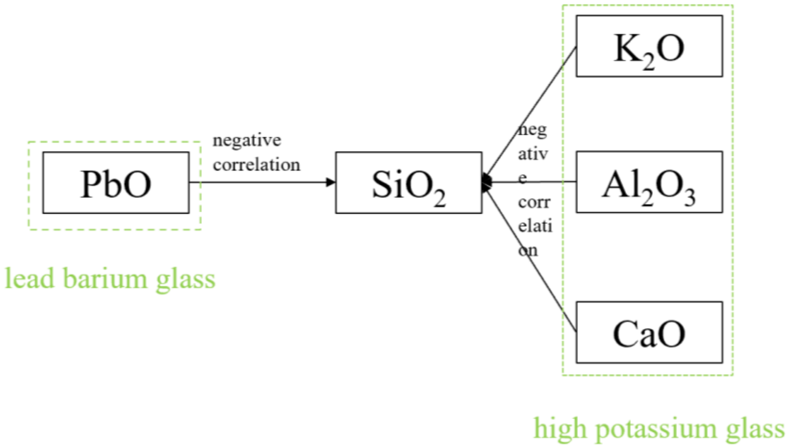

Figure 4. Variance analysis

Both lead-barium glass and high-potassium glass have a negative correlation between the content of important components and the silica content, but the types of chemical components involved in the two relationships are different, and therefore the reasons for the differences are different.

In the case of lead-barium glass, the correlation between lead oxide and silica is mainly due to the influence of the melt; in the case of high-potassium glass, the correlation between the components shown in the graph and silica is not only influenced by the melt, but also by the raw material used for manufacturing [10].

3. Conclusions

3.1. Model Advantages

1. Before model building and solving, the statistical distribution and other characteristics of the data are analyzed and checked to ensure the reasonableness of the model building.

2. The particle swarm algorithm is used to optimize the parameters of the SVM model to avoid the values of the model parameters from falling into local optimum, to speed up the convergence of the model, to improve the learning ability of the SVM model, and to enhance the accuracy of the results.

3. A gray correlation model is established, which has no excessive requirements on the number of samples and data distribution patterns, and is more applicable to the data of this problem, with less computation in the solution process.

4. The DBSCAN model is used to classify the two types of glass into subclasses, which does not need to determine the number of clusters in advance and can detect outliers in the clustering process.

5. Combined with the relevant literature to understand the importance of each indicator, the indicators are subjectively assigned using the hierarchical analysis method and objectively assigned using the entropy weight method, and the combination of the two ensures the reasonableness of the weight determination.

3.2. Model Disadvantages

1. The DBSCAN model is sensitive to changes in parameter values, and different values may have a large impact on the classification results. When the amount of data is large, the efficiency of DBSCAN model will be reduced, resulting in long convergence time.

2. In the subclass classification of the two types of glasses, the content of each chemical composition is mainly considered, and the classification basis is determined by combining with relevant literature, which is subjective.

3.3. Model Promotion

1. The clustering model established in this paper can be widely applied to the classification of other artifacts in archaeological work.

2. The clustering model established in this paper can be applied to other aspects of life. For example, when studying the development level of each region in the country, the GDP and other related indicators can be identified to cluster the regions.

References

[1] Zhang Liyan, Li Hong, Chen Shubin, Li Zhongdi, Ruan Dangzhi, Xue Tianfeng, Qian Min, Fan Sijun. Simulation method of glass composition and properties [J]. Journal of Silicate, 2022,50 (08): 2338-2350.DOI: 10.14062/j.issn.0454-5648.20220115.

[2] Ye Dingquan. High efficiency flux for making glass lithium mineral and lithium carbonate [J]. Glass fiber, 2000 (02): 42-44. DOI: 10.13354/j.cnki.cn32-1129/tq.2000.02.013.

[3] Wang Chengyu, Tao Ying. Weathering of silicate glass [J]. Journal of Silicate, 2003 (01): 78-85.

[4] Chen Haiyan,Yu Ying,Jia Yizhen,Gu Bin. Incremental learning for transductive support vector machine[J]. Pattern Recognition,2023,133.

[5] Xu Zhijun, Liu Tingting, Li Jianping, Yuan Fang. Prediction model of dynamic lateral pressure of grain storage silo wall based on support vector machine [J]. Research on Agricultural Mechanization, 2022,44 (5): 9-16.

[6] Zhang Xueqin Software defect prediction model of improved particle swarm optimization support vector machine [D]. Huzhou Normal University, 2022. DOI: 10.27946/d.cnki.ghzsf.2022.000223.

[7] Ma Chao, Lu Zhenzhou. Structural system reliability and sensitivity analysis method based on support vector machine regression [J]. Journal of Solid Mechanics, 2007 (04): 415-419. DOI: 10.19636/j.cnki.cjsm42-1250/o3.2007.04.016.

[8] Zhao Lei Signal detection algorithm based on Spearman's simple correlation coefficient in impulse noise [D]. Guangdong University of Technology, 2022. DOI: 10.27029/d.cnki.ggdgu.2022.001595.

[9] Li Qinghui, Dong Junqing, Gan Fuxi. Chemical Composition and Technological Characteristics of Early Chinese Glaze Sand and Glass Products Discussion [J]. Journal of Guangxi University for Nationalities (Natural Science Edition), 2009,15 (04): 31-41.

[10] Li Fei, Li Qinghui, Gan Fuxi, Zhang Bin, Cheng Huansheng. Proton excited X-ray fluorescence analysis of chemical composition of a batch of ancient Chinese glasses [J]. Journal of Silicate, 2005 (05): 581-586.