Research on glass classification and identification based on support vector machine

Jianan Zhang[1, *, #], Jie Ji[2, #], Yubing Lu[2, #]

[1]College of Science, Minzu University of China, Beijing, China, 100081

[2]School of Economics, Minzu University of China, Beijing, China, 100081

*Corresponding author: zhangjianan202012@163.com

#These authors contributed equally.

Abstract. Ancient glass is very susceptible to the environment of buried environment and weathering, in the soil its internal elements and the environment a large number of exchange, so that the proportion of glass components change, in order to identify the type of glass artifacts, we established a glass classification and identification model based on support vector machine (SVM). Firstly, 67 valid sample points of known categories were classified and sorted, and the data were integrated in the order of lead-barium glass, high-potassium glass, and unknown cultural relics. Use matlab to bring the sorted data into the SVM classification model to obtain the type of unknown cultural relics. Finally, the sensitivity analysis of classification was carried out by reducing the number of known types of cultural relics and adjusting the calcium oxide (CaO) content of lead-barium glass and high-potassium glass. The results showed that the glass artifacts numbered A6 and A7 were classified as high potassium weathering. A1 is high potassium unweathered; A2 and A5 are lead-barium weathering; A3, A4 and A8 are lead barium unweathered. Finally, sensitivity analysis shows that the overall stability of the model is good.

Keywords: Support vector machine, sensitivity analysis, glass identification.

1. Introduction

Glass was first introduced to China through the Silk Road[1-2], and it is also an important "witness" of China's early foreign trade and trade[3]. The main raw material of glass is quartz sand, and the main chemical component is silica (SiO2)[4]. China's ancient glass absorbs its technology and takes local materials, using lead ore as a flux, and the fired lead barium glass is usually considered to be China's own invented glass variety[5-7]. However, because the weathering of glass is inevitable, this is also an urgent problem to be solved in the glass industry. This paper studies the analysis and identification of the chemical composition of ancient glass[8], and adopts a series of analysis methods to explore the relationship between the types and chemical components of glass, which is of great significance to archaeological research[9-10].

2. Establishment and solution of glass identification model

2.1. Model building

Support Vector Machine (SVM) classification models have been widely used in the field of pattern recognition. The main idea is to find a hyperplane that allows it to correctly separate as many types of data points as possible, while keeping the two types of data points furthest from the classification surface, as shown in the figure 1.

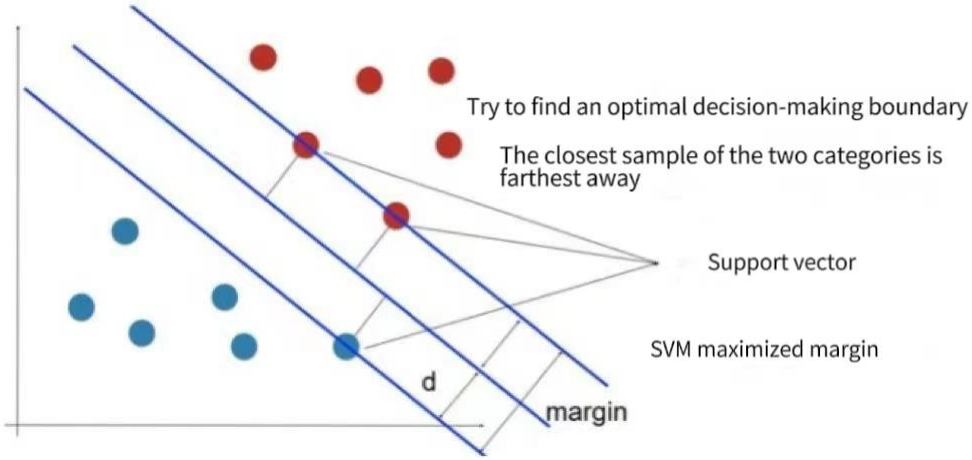

Figure 1. Schematic diagram of SVM classification model

According to the given training set T = {[a1, y1],[a2, y2], ⋯ ,[al, yl]} ∈ (Ω × y) l, where ai ∈ Ω = R n , Ω is called the input space, and each point in the input space ai consists of n attribute features, yi ∈ y = {−1,1}, i = 1, ... , l. Find a real-valued function g(x) on Rn so that the problem of inferring the y value corresponding to any pattern x is a classification problem using the classification function f(x) = sign(g(x)).

Considering the training set T, if ∃ω ∈ R n,b ∈ R and positive numbers are ε such that all AIs that make yi = 1 have (ω · ai) + b ≥ ε (where (ω · ai) represents the inner product of vectors ω and ai), and all AIs that make yi = −1 have (ω · ai) + b ≤ −ε, then the training set T linearity is said to be separable, and the corresponding classification problem is linearly separable. Note that the two types of sample sets are M+ = {ai |yi = 1,[ai,yi] ∈ T},M− = {ai |yi = −1,[ai,yi] ∈ T}. Define the conv of M+ as

$$\text{Conv}(M^{+})=\left\{a=\sum_{j=1}^{N+}\lambda_j a_j \middle| \sum_{j=1}^{N+}\lambda_j=1,\lambda_j\geq 0, j=1,\cdots,\text{N+};\text{aj}\in \text{M}+\right\} \quad (1)$$

The conv of M− (M−) is

$$\text{Conv}(M^{+})=\left\{a=\sum_{j=1}^{N-}\lambda_j a_j \middle| \sum_{j=1}^{N-}\lambda_j=1,\lambda_j\geq 0, j=1,\cdots,\text{N}-;\text{aj}\in \text{M}-\right\} \quad (2)$$

where N+ represents the number of sample points in the +1 type sample set M+, N− represents the number of sample points in the −1 type sample set M−, and the sufficient condition for the linear divisibility of the training set T is that the convex hulls of the two types of sample sets M+ and M− of T are separated. The convex hulls of the positive and negative point sets are located on both sides of the hyperplane (ω · x) + b = 0, so the two convex hulls are separated.

The hyperplane in space Rn can be written as (ω · x) + b = 0, with the parameter (ω, b) multiplied by any one.

The nonzero constant gives the same hyperplane, defined as satisfying the condition:

$$\begin{cases} y_i((\omega\cdot \text{ai})+\text{b})\geq 0, \text{i}=1,\cdots,\text{l} \\ \min\limits_{\text{i}=1,\cdots,\text{l}} |(\omega\cdot \text{ai})+\text{b}=1. \end{cases} \quad (3)$$

The hyperplane of is the canonical hyperplane of the training set T. When the training set T is linearly separable, there is a unique gauge hyperplane.

$$\begin{cases} (\omega\cdot \text{ai})+\text{b}\geq 1, \text{i}, \text{yi}=1 \\ (\omega\cdot \text{ai})+\text{b}\leq -1, \text{yi}=-1. \end{cases} \quad (4)$$

For sample points of class yi = 1, its interval from the gauge hyperplane is

$$\min_{yi=1} \frac{|(\omega \cdot ai) + b|}{||\omega||} = \frac{1}{||\omega||}. \tag{5}$$

For sample points of class yi = −1, the interval from the canonical hyperplane is

$$\min_{yi=-1} \frac{|(\omega \cdot ai) + b|}{||\omega||} = \frac{1}{||\omega||}. \tag{6}$$

Then the interval between ordinary support vectors is $\frac{2}{||\omega||}$. Optimal hyperplane means maximization$\frac{2}{||\omega||}$., As shown in the figure 2, $(\omega \cdot x) + b = \pm 1$ This is called a classification boundary, Therefore, the problem of finding the optimal hyperplane can be transformed into the following quadratic programming problem.

$$\min \frac{1}{2}||\omega||^2, s.t.\ yi((\omega \cdot ai) + b) \geq 1, i = 1, \cdots, l. \tag{7}$$

The problem is characterized by the objective function $\frac{1}{2}||\omega||$ 2, And the constraints are all linear selection α∗ a positive component αj ∗, and calculated from this

$$b* = \sum_{i=1}^{l} yi\alpha i * (ai \cdot aj), \tag{8}$$

Then construct the categorical hyperplane (ω∗· x) +b∗ = 0, and from this the decision function is obtained.

$$g(x) = \sum_{i=1}^{l} yi\alpha i * (ai \cdot x) + b * \tag{9}$$

Get a classification function.

$$f(x) = \text{sign}(\sum_{i=1}^{l} yi\alpha i * (ai \cdot x) + b *) \tag{10}$$

This classifies unknown samples. When the two classes of samples of the training set T are linearly separable, all sample points are distributed outside the classification boundary except for the ordinary support vector distributed over the two classification boundaries (ω · x) + b = ±1. The hyperplane constructed at this time is a hard-spaced hyperplane. When the two classes of samples of the training set T are approximately linearly separable, it is permissible that there are constraints that do not meet them:

$$yi((\omega \cdot ai) + b) \geq 1 \tag{11}$$

When the two classes of samples of the training set T are linearly separable, all sample points are distributed outside the classification boundary except for the ordinary support vector distributed over the two classification boundaries (ω · x) + b = ±1. The hyperplane constructed at this time is a hard-spaced hyperplane. When the two classes of samples of the training set T are approximately linearly separable, it is permissible that there are constraints that do not meet them:

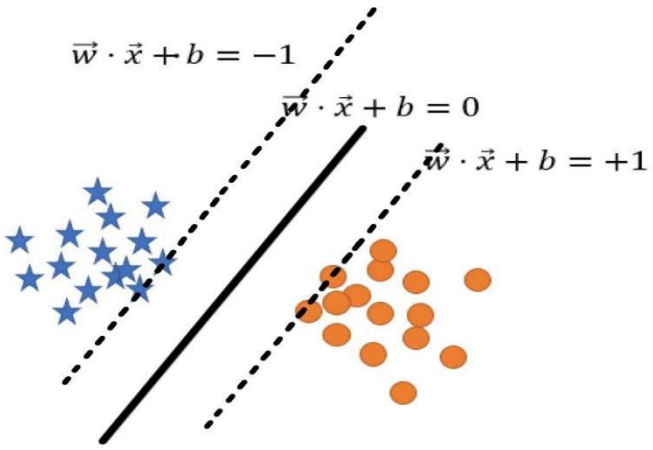

Figure 2. Schematic diagram of a soft-spaced hyperplane

The method of softening is by introducing relaxation variables.

$$\xi_i \geq 0, i = 1, \cdots, l, \tag{12}$$

to get the "softening" constraint.

$$yi((\omega \cdot ai) + b) \geq 1 - \xi_i \, i = 1, \cdots, l, \tag{13}$$

When ξi is sufficiently large, the sample point always satisfies the above constraints, but it is also necessary to try to avoid ξi taking too large a value, for which it is punished in the objective function to obtain the following quadratic programming problem.

$$\min \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{l} \xi i, \tag{14}$$

$$\text{s.t.} \begin{cases} y_i\big((\omega \cdot \text{ai}) + \text{b}\big) \geq 1 - \xi i \\ \xi i \geq 0, i = 1, \cdots, l. \end{cases} \tag{15}$$

where C > 0 is a penalty parameter.Its Lagrange function is as follows.

$$\frac{1}{2}||\omega||^2 + C\sum_{i=1}^{l} \xi i - \sum_{i=1}^{l} ai(y_i\big((\omega \cdot \text{ai}) + \text{b}\big) - 1 + \xi i) - \sum_{i=1}^{l} \gamma i \xi i \tag{16}$$

where γi ≥ 0 and ξi ≥ 0. The duality of the original problem is as follows.

$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{l} \sum_{i=1}^{l} \text{yiyj}\alpha i \alpha j(\text{ai} \cdot \text{aj}) + \sum_{i=1}^{l} \alpha i \tag{17}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{l} \xi i \, y_i ai = 0 \\ 0 \leq \alpha i \leq \text{C}, i = 1, \cdots, l. \end{cases} \tag{18}$$

Solve the above optimization problem to get the optimal solution $\alpha* = [\alpha 1 *, \cdots \alpha l*]^T$, calculation.

$$\omega^* = \sum_{i=1}^{l} \text{yi}\alpha i * \tag{19}$$

A positive component of the $\alpha * 0 < \alpha j* < C$ is selected and calculated from this.

$$\text{b} * = \sum_{i=1}^{l} \text{yi}\alpha i * (\text{ai} \cdot \text{aj}) \tag{20}$$

And from this, the classification function is obtained.

$$\text{f(x)} = \text{sign}(\sum_{i=1}^{l} \text{yi}\alpha i * (\text{ai} \cdot \text{x}) + b *) \tag{21}$$

Thus, the unknown sample is classified, and it can be seen that when C = ∞, it is equivalent to a linearly separable situation.

2.2. Solving of the model

Using the data of known samples, lead barium glass and high potassium glass are screened out and combined with the data that has not yet been classified into a new data table, and the data is imported into the SVM algorithm using matlab software to achieve the final classification, with i = 1, ···, 30 to represent 75 samples, and the jth index of the ith sample point is AIJ.yi = 1 indicates lead barium glass, and yi = −1 indicates high potassium glass. The calculation results are as follows:

Mean vector for 67 sample points.

$$\boldsymbol{\mu}=[\boldsymbol{\mu}1,\ \cdots\ ,\boldsymbol{\mu}8]=[50.9947,0.7097,2.2005,2.6730,0.7379,4.3959,0.9421,1.7871\ ,23.2788,6.6855,2.5494,0.2358,0.0789,0.3844] \quad (22)$$

Standard deviation vector of 67 sample points

$$\boldsymbol{\sigma}=[\boldsymbol{\sigma}1,\ \cdots\ ,\boldsymbol{\sigma}8]=[23.9273,1.4157,4.2371,2.4340,0.6601,3.1351,1.1963,2.0751,20.2396,7.6613,3.4111,0.2573,0.3397,2.0345] \quad (23)$$

All sample point data were normalized using the following formula: aij = aij−μj σj , i = 1, ··· , 30, j = 1, ··· , 8. Correspondingly, xj = xj−μj sj , (j = 1,2, ··· , 8) is a standardized indicator variable. Note x = [x 1, ···, x 8] T. The row vector of the 27 classified sample points after normalization is bi = [ai1, ··· , ai8] (i = 1, ··· , 27). The support vector machine model of the linear kernel function is used to classify and find the support vector is $\boldsymbol{b}2$, $\boldsymbol{b}4$, $\boldsymbol{b}7$, $\boldsymbol{b}8$, $\boldsymbol{b}12$, $\boldsymbol{b}17$, $\boldsymbol{b}18$, $\boldsymbol{b}19$, $\boldsymbol{b}32$, $\boldsymbol{b}52$, $\boldsymbol{b}54$, $\boldsymbol{b}56$, $\boldsymbol{b}57$, $\boldsymbol{b}59$, $\boldsymbol{b}60$, $\boldsymbol{b}62$, $\boldsymbol{b}64$, $\boldsymbol{b}65$, $\boldsymbol{b}66$

The linear classification function is $c(\tilde{x}) = \sigma i\ \beta i\ K(\boldsymbol{b}i\ ,\ \tilde{x}) + b = K(\boldsymbol{b}2\ ,\ \tilde{x}) + K(\boldsymbol{b}4\ ,\ \tilde{x}) + K(\boldsymbol{b}7\ ,\ \tilde{x}) + K(\boldsymbol{b}8\ ,\ \tilde{x}) + K(\boldsymbol{b}12\ ,\ \tilde{x}) + 0.0394K(\boldsymbol{b}17,\ \tilde{x}) + 0.5419\ K(\boldsymbol{b}18\ ,\ \tilde{x}) + 0.1247K(\boldsymbol{b}19\ ,\ \tilde{x}) + 0.1780K(\boldsymbol{b}32\ ,\ \tilde{x}) + 0.3149K(\boldsymbol{b}52\ ,\ \tilde{x}) + 0.2942\ K(\boldsymbol{b}54\ ,\ \tilde{x}) + K(\boldsymbol{b}56\ ,\ \tilde{x}) + 0.0321K(\boldsymbol{b}57\ ,\ \tilde{x}) + K(\boldsymbol{b}59\ ,\ \tilde{x}) + 0.4277K(\boldsymbol{b}60\ ,\ \tilde{x}) + 0.7972K(\boldsymbol{b}i62,\ \tilde{x}) + 0.0179K(\boldsymbol{b}64,\ \tilde{x}) + K(\boldsymbol{b}65\ ,\ \tilde{x}) + K(\boldsymbol{b}66\ ,\ \tilde{x}) + 1.6015$

The test was performed using known sample points, and the error was 0.075, which passed the test.The final classification results are shown in the following table 1.

Table 1. Classification result table

Artifact number	Category	The category label to which it belongs	Surface weathering
A1	High potassium glass	-1	No weathering
A2	Lead barium glass	1	weathering
A3	Lead barium glass	1	No weathering
A4	Lead barium glass	1	No weathering
A5	Lead barium glass	1	weathering
A6	High potassium glass	-1	weathering
A7	High potassium glass	-1	weathering
A8	Lead barium glass	1	No weathering

2.3. Sensitivity analysis

In the analysis of the chemical composition of unknown glass cultural relics, we use the attached data for SVM classification, the following study in the case of keeping other parameters unchanged, swapping the calcium oxide (CaO) content of lead barium glass and high alumina glass, observing the difference between the results obtained and the results before the exchange, the cultural relics with different classification results are marked in red (the same below), and the classification results are shown in the following table 2:

Table 2. Influence of chemical composition on classification results

Artifact number	Category	The category label to which it belongs	Surface weathering
A1	High potassium glass	-1	No weathering
A2	Lead barium glass	1	weathering
A3	Lead barium glass	1	No weathering
A4	Lead barium glass	1	No weathering
A5	Lead barium glass	1	weathering
A6	Lead barium glass	1	weathering
A7	High potassium glass	-1	weathering
A8	High potassium glass	-1	No weathering

The results showed that two cultural relics were classified differently from those before the chemical content was not exchanged, which showed that the SVM classification model was sensitive to chemical composition content.

In the following study, while keeping other parameters unchanged and randomly reducing 10 lead-barium glass cultural relics and 5 high-potassium glass cultural relics, the differences between the results obtained and the classification results before deletion are shown in Table 3 below:

Table 3. Effect of sample size on classification results

Artifact number	Category	The category label to which it belongs	Surface weathering
A1	High potassium glass	-1	No weathering
A2	Lead barium glass	1	weathering
A3	Lead barium glass	1	No weathering
A4	Lead barium glass	1	No weathering
A5	Lead barium glass	1	weathering
A6	High potassium glass	-1	weathering
A7	High potassium glass	-1	weathering
A8	Lead barium glass	1	No weathering

It is known that in the case of a large number of cultural relics, the SVM classification model is stable for the number of cultural relics. This is because the two glass components are classified in a fixed way, and when the data is universal, the results obtained by using the SVM classification model are basically error-free, which further verifies the rationality and reliability of the model.

3. Conclusions

In this paper, 67 valid sample points of known categories are classified and sorted, and the data are integrated in the order of lead-barium glass, high-potassium glass, and unknown cultural relics. Then, MATLAB is used to bring the collated data into the support vector machine (SVM) classification model, and the type of unknown artifact is calculated. Finally, the sensitivity analysis

of classification was carried out by reducing the number of known types of cultural relics and adjusting the calcium oxide (CaO) content of lead-barium glass and high-potassium glass.

References

[1] Ji Luoyuan, PJ Cherian.Peacock blue-glazed pottery specimens excavated from Kerala, India[J].Bulletin of The Palace Museum,2022(06):55-67+148.DOI:10.16319/j.cnki.0452-7402.2022.06.004.

[2] Do Khai Ly,Su Miao, Yang Yuanyuan. Communication and Inclusion: The Influence of Chinese Culture on Decorative Arts along the Silk Road[J]. Journal of Donghua University(Social Sciences),2022,22(02):37-46.DOI:10.19883/j.1009-9034.2021.0319.

[3] Pan Ling, Tan Wenyu. Western cultural factors in the Xianbei relics of Hulunbuir: A discussion on the "Grassland Silk Road" during the Han Dynasty[J]. Archaeology,2022(05):110-120+2.)

[4] Cui Ning, Wang Gong. Sino-foreign exchanges on the grassland Silk Road in Tongliao region of the Liao Dynasty[J].Journal of Inner Mongolia University for Nationalities(Social Sciences Edition),2022,48(01):23-28.DOI:10.14045/j.cnki.nmsx.2022.01.001.

[5] Duck-shaped glass note Witness of the Steppe Silk Road[J].Communist Party Member,2022(02):2.)

[6] Shang Yue. A grassland silk road connecting Eastern and Western cultures[N]. Liaoning Daily,2022-01-05(012). DOI:10.28534/n.cnki.nlnrb.2022.000037.

[7] Jiang Bo. Ports, shipwrecks and trade goods: archaeological discovery and research of the Maritime Silk Road[J].Studies in the History of Maritime Communication,2021(04):8-22.DOI:10.16674/j.cnki.cn35-1066/u.2021.04.002.

[8] Huang Qiaohao. From the local to the exotic, feel the "Ryu" brilliance of sea silk glassware[J].Collection. Auction, 2021(06): 56-61.

[9] An Jiayao. Silk Road and Glassware[J].Cultural Relics Tiandi,2021(12):87-92.)

[10] Silk Road Treasures:Exhibition of Cultural Relics from the Collection of Ikuo Hirayama Silk Road Art Museum[J].Collection,2020(08):173.)