

Component Analysis and Identification of Ancient Glass Products Based on Correlation Analysis

Shirui Li *

College of physics, Sichuan University, Chengdu, China

* Corresponding author: shirui.li@foxmail.com

Abstract. The Silk Road was an important channel for cultural exchanges between ancient China and the West, and glass was a precious material evidence of early trade exchanges. In order to explore the changes of relevant properties and chemical compositions of these glass relics before and after weathering, it is necessary to calculate their correlation degree and predict the changes of compositions. In this paper, the correlation and correlation between the components are calculated by establishing the grey correlation analysis method, independent sample T test method, random forest algorithm, K-means algorithm and other models, so as to analyze and predict the correlation properties and chemical components of glass relics before and after weathering. First, preprocess the data, remove the invalid data, and take the glass type, decoration and color as the comparison sequence. Whether the cultural relic surface is differentiated into a reference sequence, use the gray correlation analysis method to calculate that the gray correlation between the glass type and whether it is differentiated is the largest, which is 0.7939, indicating that the glass type is most closely related to whether it is weathered. According to the type of glass, the average, median and standard deviation are calculated, and the chemical composition with strong correlation with weathering is determined by independent sample t-test. The prediction results are obtained by predicting the principal components of weathering points.

Keywords: Glass composition analysis, Correlation analysis, Clustering algorithm.

1. Introduction

The Silk Road was an important channel for cultural exchanges between ancient China and the West, and glass was a precious material evidence of early trade exchanges. Early glass was often made into ornaments and introduced into China. After absorbing its technology, ancient Chinese glass was made locally, similar to its appearance but different in chemical composition. Due to different fluxes added during refining, the glass can be divided into lead barium glass and potassium glass.

Ancient glass is easily weathered by burial environment. In the weathering process, due to a large amount of exchange between internal elements and external elements, the proportion of their components changes, which affects the correct judgment of their categories. The research on the relevant properties of ancient cultural relics is a major topic in the field of archaeology in China, which is of great significance. There are a number of relevant data on ancient glass products in China. Archaeologists have divided them into two types: high potassium glass and lead barium glass.

2. Establishment and solution of model

2.1. The Relationship between the Surface Weathering of Glass Relics and Their Glass Types, Patterns and Colors

The accumulated glass composition ratio and the data between 85% and 105% are regarded as valid data. First, sum up the proportion of the corresponding main components, and the data that the cumulative proportion of the components is between 85% and 105% are regarded as invalid data. Through calculation, it is found that the cumulative sum of the component proportions of cultural relics 15 and 17 is outside the range of valid data, so they are omitted. Table 1 shows the cumulative sum of the proportion of some cultural relics.

Table 1. Cumulative Sum of Components of Some Cultural Relics

Cultural relic number	Cumulative sum of component proportions
01	97.61
02	99.89
03	100
04	96.06
05	96.51
06	98.92
07	99.7
08	99.82
...	...
15	79.49
16	98.41
17	71.89

Conditional quantification: quantifies text information to make it easier to process data information. For example, the high potassium of glass type is 1, and the lead barium is 2. The patterns A, B and C are 1, 2 and 3 respectively. Different colors are 1-7 in turn.

2.2. Establishment of model

When exploring the relationship between the surface weathering of glass relics and their glass types, patterns and colors, the gray correlation analysis method was used. Grey correlation analysis is a multi-factor statistical analysis method. That is, in a gray system, you can understand the relative strength of a project affected by other factors. The evaluation and analysis are carried out according to the grey correlation degree between factors. If the change trend of factors over time is similar, the correlation degree between the two is high; On the contrary, the correlation is low. Through the analysis of the correlation between the comparison sequence and the reference sequence, the system is comprehensively evaluated. [1] In order to evaluate different influencing factors, it is necessary to determine the reference sequence and comparison sequence, and solve their correlation coefficient and correlation degree. For one reference sequence and several comparison sequences, the correlation coefficient solution model [2] is:

$$\xi_{oi}(k) = \frac{\min_i \min_k \Delta_{0i} + \eta \max_i \max_k \Delta_{0i}}{\Delta_{0i} + \eta \max_i \max_k \Delta_{0i}} \tag{1}$$

Where, η Is the resolution coefficient, which is 0-1 and 0.5 in most cases, the correlation degree is given after the correlation coefficient is weighted and averaged, as shown in the formula:

$$R_{oi}(k) = \frac{1}{n} \sum_{k=1}^n \psi_k \xi_{oi}(k) \tag{2}$$

Where, ψ_k is the weight, and the value range is $0 \leq \psi_k \leq 1$.

Determine the reference sequence and comparison sequence. Set whether the surfaces of these glass relics are differentiated into reference sequences X_0 : $X_0=(X_0^{(1)}, X_0^{(2)})$, glass type as comparison sequence X_1 , ornamentation as comparison sequence X_2 , and color as comparison sequence X_3 : $X_0=(X_0^{(1)}, X_0^{(2)})$, $X_1=(X_1^{(1)}, X_1^{(2)}, X_1^{(3)})$, $X_2=(X_2^{(1)}, X_2^{(2)}, X_2^{(3)}, X_2^{(4)}, X_2^{(5)}, X_2^{(6)}, X_2^{(7)})$.

2.3. Solution of model

According to the correlation coefficient and correlation degree solution formula of equation (1) and equation (2), take glass type, decoration and color as the comparison sequence, obtain the correlation degree with weathering, and analyze the relationship between them.

The correlation coefficient diagram is drawn by spss software as follows in Fig 1.

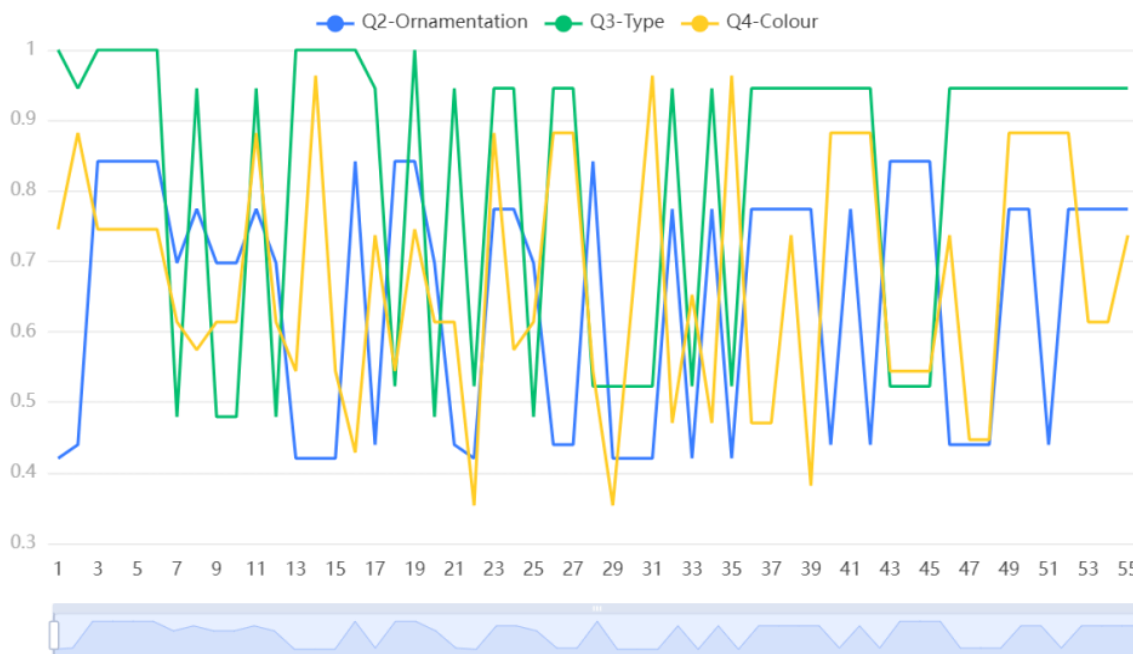


Figure 1. Correlation coefficient diagram

Use matlab to solve the correlation coefficient and correlation degree equation, import the quantized data Form 1 into the comparison sequence and reference sequence, and calculate the correlation degree between glass type, decoration and color and weathering. The correlation degree table 2 is obtained. Among them, the gray correlation degree of glass type is the highest, 0.7939, indicating that glass type has the greatest influence on whether the glass is weathered.

Table 2. Grey correlation degree between properties and weathering

Comparison sequence	Grey correlation degree	ranking
Glass type	0.7939	1
Ornamentation	0.6909	2
colour	0.5378	3

The magnitude of each correlation degree represents the close relationship between it and weathering.

2.4. Statistical rules for analyzing the content of weathering chemical components

2.4.1. Data related calculation and visual processing

In mathematics, we often analyze data by obtaining important statistics such as average, median and variance of a group of data. Firstly, each statistic of the data set is obtained and the correlation between the content of each chemical component and whether it is differentiated is analyzed separately by using the independent sample T test method, so as to show the difference of the content of the components before and after the surface weathering of the cultural relics.

According to the type of glass, the average, median and standard deviation of chemical composition content before and after weathering on the surface of cultural relics were calculated. The following table 3 and 4 is obtained:

Table 3. Summary of Average Values of High Potassium Components

Surface weathering	(SiO ₂)	(Na ₂ O)	(K ₂ O)	(CaO)	(MgO)	(Al ₂ O ₃)	(Fe ₂ O ₃)	(CuO)	(PbO)	(BaO)	(P ₂ O ₅)	(SrO)	(SnO ₂)	(SO ₂)
No weathering	67.984	0.695	9.331	5.333	1.079	6.620	1.932	2.453	0.412	0.598	1.402	0.042	0.197	0.102
weathering	93.963	0.000	0.543	0.870	0.197	1.930	0.265	1.562	0.000	0.000	0.280	0.000	0.000	0.000

Table 4. Summary of Average Number of Lead and Barium Components

Surface weathering	(SiO ₂)	(Na ₂ O)	(K ₂ O)	(CaO)	(MgO)	(Al ₂ O ₃)	(Fe ₂ O ₃)	(CuO)	(PbO)	(BaO)	(P ₂ O ₅)	(SrO)	(SnO ₂)	(SO ₂)
No weathering	53.444	0.772	0.258	1.232	0.492	3.195	0.933	1.557	23.594	10.499	0.904	0.297	0.065	0.282
weathering	33.615	0.953	0.143	2.346	0.701	3.838	0.556	1.996	36.872	10.487	4.155	0.366	0.056	0.987

It can be seen from the above table that the average content of silicon dioxide in both high potassium glass and lead barium glass changes significantly before and after weathering. Visualize it and get the following statistical as shown in Fig 2 and Fig 3

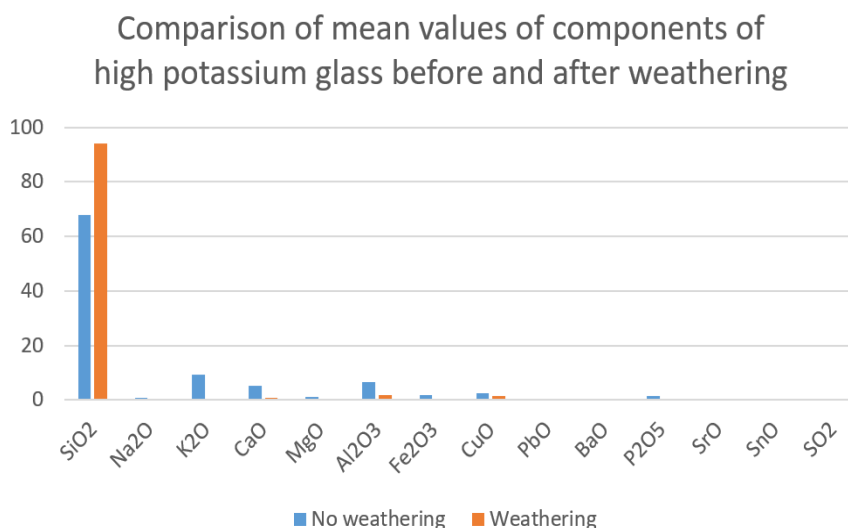


Figure 2. Comparison of average composition content of high potassium glass before and after weathering

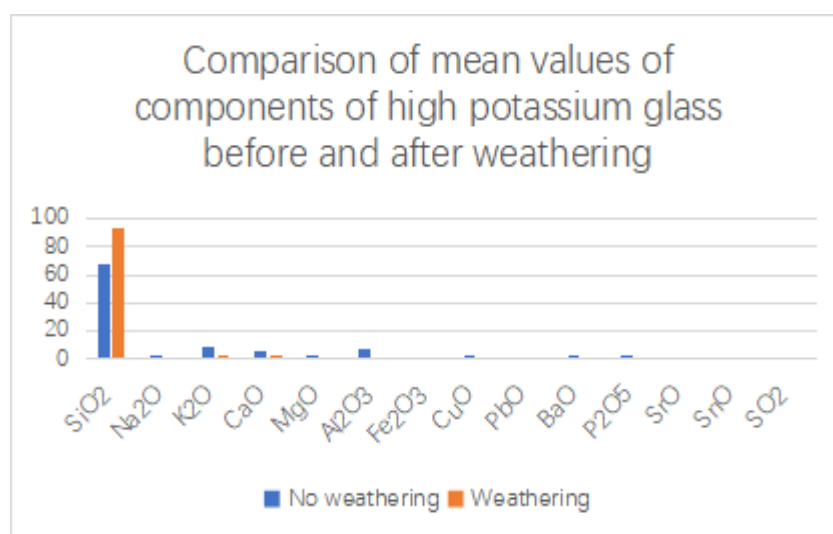


Figure 3. Comparison of average lead barium glass content

3. Correlation analysis by independent sample T test

The original hypothesis H_0 is the same as the single sample t-test. The values of the two-sample t-test also have three different situations: significantly not equal to, significantly greater than, and significantly less than [3]. Determine the test statistics, whether the high potassium and lead barium are weathered, and the content of each component generally conforms to the bell-shaped normal distribution curve, variance σ_1 and σ_2 . After determining the value, estimate the sampling distribution of the content of weathering components in high potassium and lead barium, namely σ It can be expressed as:

$$t = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3}$$

In the formula, x and y are the mean values of the contents of weathered and non-weathered components of high potassium/lead barium respectively, s_1 and s_2 are the standard deviations of the contents of weathered and non-weathered components of high potassium/lead barium respectively, and n_1 and n_2 represent the quantities in the data of weathered and non-weathered samples of high potassium/lead barium respectively.

The spss software has confirmed that the samples conform to the normal distribution, and the t-test can be used to obtain the component statistics and make the difference judgment [4]. Observe the F value of statistics and the P value of significance. Calculate the corresponding probability P value according to the distribution to which t statistics obey. The analysis results of independent sample t-test will give two t-statistics for selection, and it is necessary to select appropriate statistics according to the corresponding test results. Set the significance level α Compare with the probability P value and make statistical decisions. The analysis results of independent sample T section are shown in Table 5:

Table 5. T Test Results of Content of Some Components in High Potassium Type

Variable name	Variable value	sample size	average value	standard deviation	t	P	Mean value difference	Cohen's d value
SiO ₂	No weathering	12	67.984	8.755	7.095	0.000***	25.979	3.547
	weathering	6	93.963	1.734				
	total	18	76.644	14.467				
Na ₂ O	No weathering	12	0.695	1.287	1.303	0.211	0.695	0.651
	weathering	6	0	0.000				
	total	18	0.463	1.089				
K ₂ O	No weathering	12	9.331	3.920	5.391	0.000***	8.788	2.696
	weathering	6	0.543	0.445				
	total	18	6.402	5.308				
CaO	No weathering	12	5.333	3.092	3.461	0.003***	4.463	1.731
	weathering	6	0.87	0.488				
	total	18	3.845	3.308				
MgO	No weathering	12	1.079	0.676	3.011	0.008***	0.882	1.505
	weathering	6	0.197	0.306				
	total	18	0.785	0.712				
	weathering	6	1.93	0.964				

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively

Observe the mean value of different chemical components before and after weathering, P value of F test results and Cohen's d value of their difference range in the above table, and judge whether the correlation P value between each chemical component and weathering is significant ($P < 0.05$) and the difference range. The p value of weathering is small in the test of silicon dioxide, potassium oxide, calcium oxide, magnesium oxide, aluminum oxide and iron oxide, so the statistical results are significant, indicating that these chemical components have significant differences in weathering, while other chemical components have low correlation with weathering [5], as shown in Fig 4.

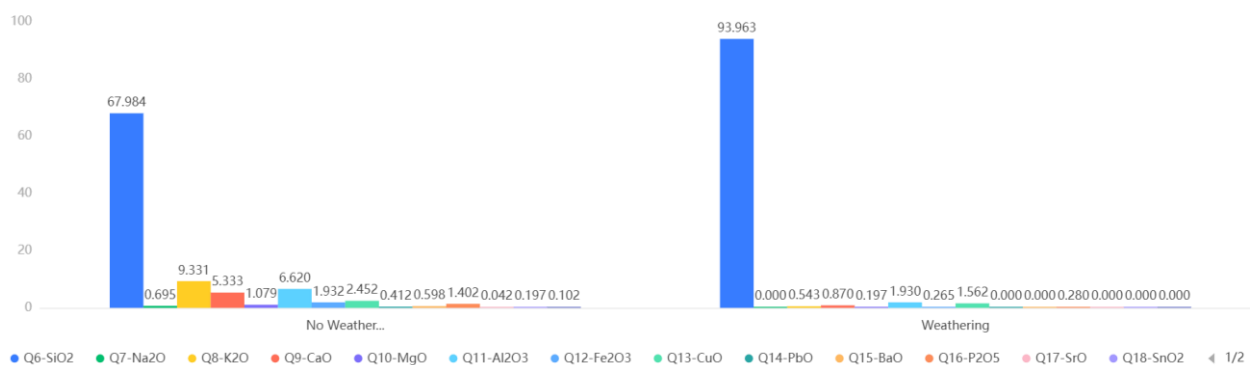


Figure 4. Histogram of T test results

Similarly, according to the above method, the independent T test is conducted for each chemical component of lead barium type and whether it is differentiated, and the table is obtained. It can be seen that the content of silicon dioxide, sodium oxide, calcium oxide, lead oxide, phosphorus pentoxide and strontium oxide in the lead barium type has a strong correlation with weathering, while other chemical components have a low correlation with weathering.

3.1. Prediction of chemical composition content of weathering point before weathering

3.1.1. theoretical analysis

As for the independent T-test results of the chemical composition content, it can be seen that whether the high potassium type and lead barium type are weathered has a high correlation with some chemical components, but a low correlation with some components. That is, the content of some chemical components changes significantly before and after weathering, so we can ignore these chemical components with very low correlation when predicting the data before weathering at weathering points. Therefore, when predicting the content of chemical components before weathering, we mainly predict the components that have significant correlation with differentiation. For high potassium type, the content of silicon dioxide, potassium oxide, calcium oxide, magnesium oxide, aluminum oxide and iron oxide before weathering is mainly predicted. For the type of lead and barium [6], the content of silicon dioxide, sodium oxide, calcium oxide, lead oxide, phosphorus pentoxide and strontium oxide are mainly predicted.

According to the degree of differentiation, it can be divided into four grades: no weathering, weathered glass with no weathering point, weathered glass with no weathering point, weathered glass with severe weathering point, and calculated the average value according to different glass types [7]. According to the proportional relationship between the content of each chemical component at different degrees of differentiation and the average value of the content of chemical components without weathering point, the content of each chemical component at each weathering point before weathering is determined.

3.1.2. Mean value analysis and component prediction

The average content of silicon dioxide, potassium oxide [8], calcium oxide, magnesium oxide, aluminum oxide and iron oxide before and after weathering in high potassium type is analyzed, and the following table 6 is obtained:

Table 6. Comparison of main components of high potassium glass before and after weathering

Surface weathering	SiO ₂	K ₂ O	CaO	MgO	Fe ₂ O ₃	Al ₂ O ₃
No weathering	67.770	9.567	5.735	1.055	1.789	6.406
weathering	93.963	0.543	0.870	0.197	0.265	1.930
ratio	0.7212	17.6188	6.5920	5.3553	6.7509	3.3192
Difference	-26.193	9.024	4.865	0.858	1.524	4.476

According to the ratio relationship of chemical components before and after weathering in the table, the content of chemical components before weathering of high potassium glass weathering point is predicted, and the final results are shown in Table 7.

Table 7. Partial prediction results of data before weathering of high potassium glass

Cultural relic number	type	Surface weathering	SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	Al ₂ O ₃	Fe ₂ O ₃
7	1	2	66.800	0.000	9.024	7.053	0.858	6.572	1.148
9	1	2	68.530	0.000	10.395	4.087	0.858	4.381	2.160

Note: 2 marked in the column of surface weathering in the above figure is the weathered point, and the bold data is the prediction data

The content of silicon dioxide, sodium oxide, calcium oxide, lead oxide, phosphorus pentoxide and strontium oxide before and after weathering in the lead barium type is analyzed [9], and the following table 8 is obtained:

Table 8. Comparison of average chemical composition of lead and barium

Surface weathering	SiO ₂	Na ₂ O	CaO	Al ₂ O ₃	P ₂ O ₅	SrO
No weathering	54.928	0.836	0.970	3.119	0.920	0.287
weathering	27.726	0.268	2.557	3.075	4.507	0.374
Severe weathering	8.480	0.000	2.067	1.980	9.243	0.757
Unweathered point	56.789	2.551	1.552	6.263	1.371	0.237

According to the ratio relationship of chemical components before and after weathering in the table, the content of chemical components before and after weathering of lead barium glass weathering point is predicted [10], and the final result is as shown in Table 9.

Table 9. Prediction Results of Lead and Barium

type	Surface weathering	Cultural relics sampling point	SiO ₂	Na ₂ O	CaO	Al ₂ O ₃	P ₂ O ₅	SrO
Lead barium	weathering	02	71.874	0.568	0.888	5.812	0.729	0.146
Lead barium	weathering	08	39.899	0.568	0.561	1.359	0.733	0.284
Lead barium	weathering	11	66.545	0.568	1.332	2.728	1.915	0.284
Lead barium	weathering	19	58.720	0.568	1.111	3.621	1.802	0.146

4. Conclusion

In this paper, the relationship between the surface weathering of glass relics and their glass types, patterns and colors is analyzed, that is, the relationship models between various variables and surface weathering are established respectively. Grey relational analysis is a multi-factor statistical analysis method. The idea of grey correlation analysis can be used to determine whether the surface of cultural relics is divided into reference sequence by taking glass type, decoration and color as comparison sequence. To explore the relationship between the surface differentiation of glass relics and their glass types, patterns and colors.

According to the type of glass, the statistical rule of whether there are weathering chemical components on the surface of cultural relics is analyzed. That is, the chemical composition content of high potassium glass and lead barium before and after weathering are statistically analyzed. Analyze the average, mean, standard deviation and other important indicators of the component content before and after weathering to judge the content change before and after weathering. At the same time, use

the independent sample T test method to separately analyze the correlation between the content of each chemical component and whether it is weathered, so as to show the difference of the component content before and after weathering on the surface of the cultural relics sample.

Finally, according to the data after weathering, the chemical composition content before weathering is predicted. According to the different weathering degree, analyze the relationship between the chemical composition content and predict the content of each weathering point before weathering.

References

- [1] Zhang Kexue, Yang Haijiang, He Manchao, et al Comprehensive evaluation of intelligent degree of working face based on grey correlation analysis [J] *Modern Tunnel Technology*, 2022, 59 (1): 69 - 79. DOI: 10.13807/j.cnki.mtt.2022.01.007.
- [2] Fu Cheng, Guo Wei, Guo Tianyue, Yang Hang, Li Qingyuan. A method for identifying dominant channels based on grey correlation TOPSIS [J]. *Natural Gas and Oil*, 2022, 40 (03): 68 - 74.
- [3] Dong Xiuyue. Study on the optimal method of paired t-test and grouped t-test [J]. *Journal of Mathematical Medicine and Pharmacy*, 2010, 23 (01): 11 - 14.
- [4] Zhang Kexue, Yang Haijiang, He Manchao, et al Comprehensive evaluation of intelligent degree of working face based on grey correlation analysis [J] *Modern Tunnel Technology*, 2022,59 (1): 69-79. DOI: 10.13807/j.cnki.mtt.2022.01.007.
- [5] Fu Cheng, Guo Wei, Guo Tianyue, Yang Hang, Li Qingyuan. A method for identifying dominant channels based on grey correlation TOPSIS [J]. *Natural Gas and Oil*, 2022, 40 (03): 68 - 74.
- [6] Dong Xiuyue. Study on the optimal method of paired t-test and grouped t-test [J]. *Journal of Mathematical Medicine and Pharmacy*, 2010, 23 (01): 11 - 14.
- [7] Shen Qinpeng, Zhang Xia, Zhang Tao, et al Establishment of aroma classification model of flue-cured tobacco based on chemical components of tobacco [J] *Hubei Agricultural Science*, 2015, 54 (5): 1220-1226. DOI: 10.14088/j.cnki.issn0439 - 8114.2015.05.049.
- [8] Guo Li Construction and research of precision learning model based on multi-mode sample t-test [J] *Journal of Shaoguan University*, 2020,41 (6): 13-17. DOI: 10.3969/j.issn.1007-5348.2020.06.004.
- [9] Yang Song, Cui Chun, Chu Wenjuan, et al A prediction model for the release of conventional components of medium cigarette smoke based on material parameters [J] *Tobacco Science and Technology*, 2022, 55 (7): 47 - 55.
- [10] Li Hao, Ling Xianwu, Zhao Jucheng, et al Analysis of asphalt pavement interlayer temperature based on grey correlation degree [J] *Municipal Technology*, 2022, 40 (4): 55-61. DOI: 10.19922/j.1009-7767.2022.04.055.