Research on composition analysis and type identification of ancient glass products based on data mining

Chang Su¹, Jingjing Wang²

¹School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai, 200433, China
²School of Mathematics, Shanghai University of Finance and Economics, Shanghai, 200433, China

Keywords: Chi-square test, decision tree, cluster analysis, BP neural network, grey correlation analysis

Abstract: The ancient glass products are easy to be weathered, which makes a large number of internal and external elements exchange, leading to the change of their composition proportion, which further affects the composition analysis and identification. This paper establishes a mathematical model based on the weathering phenomenon of ancient glass products in China and the related chemical composition data, to predict, analyze and solve the classification laws and classification results of different glasses, the statistical laws and correlation of chemical composition content, and other problems.

1. Introduction

The main raw material of glass is quartz sand (silicate). During the refining process, the chemical composition of the final refined glass is different due to the different added cosolvents. At the same time, due to the influence of the storage environment, ancient glass is very easy to be weathered, and its chemical composition proportion will also change accordingly, affecting the correct judgment of glass type[1]. Therefore, determining mine the chemical composition system of glass under the influence of the environment is an important aspect of the study of ancient Chinese glass. In the context of the big data era, the article uses data mining methods to analyze the chemical composition of ancient Chinese glass, and combines the existing achievements of other scholars at home and abroad on ancient glass research, to predict the chemical composition of ancient Chinese glass from the perspective of scientific and technological research, Complete the prediction of ancient glass categories, based on the actual data.

2. Methods

After preprocessing the data, use SPSS² to conduct the R × C contingency table chi-square test on whether the surface is weathered and glass type, whether the surface is weathered and glass decorative pattern, and whether the surface is weathered and glass color, according to the significance level presented by the test results, judge whether the glass type, texture, and color have a significant impact on the surface weathering; Excel is used to analyze the statistical laws of the processed data. For the content data of high potassium glass and lead barium glass, as well as the chemical
composition with or without weathering, the statistical laws of the chemical composition with or without weathering are obtained by comparing and analyzing the average value and median value, and are tested with standard error. At the same time, according to the test data of weathering points, the ratio of the average chemical composition content before and after weathering of the two types of glass is used to predict the corresponding chemical composition content before weathering.

Normalize the pretreated data, and then set the proportion of the training set and test set through SPSS to analyze and verify the chemical composition content of high potassium glass and lead barium glass by a decision tree, and obtain reasonable classification standards and rules. Then, the method of cluster analysis is applied to the sub-classification of different glass types, and appropriate methods are selected according to different data conditions judged by Matlab program: pedigree clustering method and K-means clustering method to obtain more scientific classification results. Then, the rationality and sensitivity of classification results are analyzed according to the "elbow method" by using the evaluation function image.

Based on the above results, analyze and identify the chemical components of the unknown glass relics, establish a BP neural network model to verify the results twice, and conduct sensitivity analysis by substituting the original data values and other methods to change the corresponding variable indicators.

Finally, for different types of glass cultural relics samples, use Matlab to carry out grey correlation analysis, take the chemical components with considerable content after normalization as the parent sequence to solve the correlation coefficient, and analyze the correlation between chemical components within the category, as well as the difference of the correlation between the same chemical components among the categories.

3. Model establishment and solution

3.1 Influence factors and chemical composition prediction of glass before and after weathering

3.1.1 R × C contingency table chi-square test

In this paper, the relationships between the surface weathering of cultural relics and their glass type, decoration, and color are analyzed by using the R×C contingency table chi-square test, where R is the number of rows and C is that of columns in the contingency table. When the significance of the chi-square value (i.e., p-value) is less than 0.05, the original hypothesis is rejected, and the two variables are considered to be significantly correlated with 95% certainty. The chi-square test analysis shows that based on the surface weathering and type, the significance p-value is 0.020, showing significance at the level, and rejecting the original hypothesis. Therefore, the glass type has a significant impact on surface weathering, while the color has little impact on the surface weathering of glass relics.
3.1.2 Statistical analysis of the chemical composition content of each type of glass with or without weathering

According to the pretreated data, the data are divided into four parts - high potassium weathering, high potassium nonweathering, lead barium weathering, and lead barin non weathering, according to the type of glass and whether there is weathering. The average value, standard error, and median corresponding to the content of various chemical components in these four parts are calculated using Excel. See Table 1 for an example: whether there is weathering statistical data for high potassium glass, and whether there is weathering data for lead barium glass.

Table 1: Statistical data of weathering of high potassium glass

|       | No weathering | Weathering |
|-------|---------------|-------------|
|       | Average value | Standard error | Median | Average value | Standard error | Median |
| SiO₂  | 67.98         | 2.53        | 65.53 | 93.96         | 0.71          | 93.51  |
| K₂O   | 9.76          | 0.86        | 9.83  | 0.54          | 0.18          | 0.67   |
| CaO   | 5.33          | 0.89        | 6.10  | 0.87          | 0.20          | 0.83   |
| MgO   | 1.08          | 0.20        | 1.17  | 0.20          | 0.13          | 0.00   |
| Al₂O₃ | 6.62          | 0.72        | 6.19  | 1.93          | 0.39          | 1.72   |
| Fe₂O₃ | 1.67          | 0.32        | 2.11  | 0.27          | 0.03          | 0.28   |
| CuO   | 2.45          | 0.48        | 2.35  | 1.30          | 0.19          | 1.55   |
| SO₂   | 0.10          | 0.05        | 0.00  | 0.28          | 0.09          | 0.28   |
| Na₂O  | 0.53          | 0.27        | 0.00  |               |               |        |
| PbO   | 0.41          | 0.17        | 0.16  |               |               |        |
| BaO   | 0.60          | 0.28        | 0.00  |               |               |        |
| P₂O₅  | 0.91          | 0.13        | 1.02  |               |               |        |
| SrO   | 0.04          | 0.01        | 0.02  |               |               |        |

According to Table 1, the comparison and analysis of high potassium weathering and high potassium nonweathering are carried out to compare the median and average values of the chemical composition of high potassium glass in non weathering and the chemical composition after weathering. The same is true for lead barium glass. Finally, the statistical rules of whether there is weathering chemical composition on the surface of high potassium glass and lead barium glass samples are as follows:

1. After weathering, the silica content of high potassium glass increased significantly, while the magnesium oxide content decreased; The content of alumina, iron oxide and copper oxide decreased; The content of sulfur dioxide increased; The contents of sodium oxide, lead oxide, barium oxide, phosphorus pentoxide, and strontium oxide decreased (to zero).

2. After weathering, the content of silica in lead barium glass increased significantly; The contents of potassium oxide, aluminum oxide, barium oxide, and sodium oxide increased; The contents of calcium oxide, iron oxide, lead oxide, phosphorus pentoxide, and strontium oxide decreased; The content of copper oxide decreased slightly; The content of magnesium oxide is stable without obvious change.

3.1.3 Prediction of chemical composition before weathering

According to the above, the proportion of average values of various chemical compositions before and after weathering of two types of glass products is used to predict the chemical composition content of weathering points before weathering. The specific prediction results are shown in Table 2.
Table 2: Forecast Data of High Potassium Glass before Weathering

| Cultural relic sampling point | SiO₂ | K₂O | CaO | MgO | Al₂O₃ | Fe₂O₃ | CuO | P₂O₅ |
|-------------------------------|------|-----|-----|-----|-------|-------|-----|------|
| 07                            | 67.02| 0.00| 6.56| 0.00| 6.79  | 1.07  | 3.12| 1.97 |
| 09                            | 68.75| 10.60| 3.80| 0.00| 4.53  | 2.01  | 2.93| 1.13 |
| 10                            | 70.01| 16.53| 1.29| 0.00| 2.78  | 1.64  | 1.59| 0.00 |
| 12                            | 68.22| 18.15| 4.41| 0.00| 5.01  | 1.83  | 3.12| 0.49 |
| 22                            | 66.82| 13.30| 10.17| 3.51| 12.01| 2.20  | 1.04| 0.68 |
| 27                            | 67.08| 0.00| 5.76| 2.96| 8.61  | 1.26  | 2.91| 1.16 |

Since the sodium oxide, lead oxide, barium oxide, strontium oxide, and sulfur dioxide at the weathering point of high potassium glass are all zero, the predicted value before weathering is also zero, which is not shown in the datasheet. Similarly, sodium oxide and strontium oxide of lead barium glass are also zero, which are not shown in the following Table 3.

Table 3: Prediction Data of Lead Barium Glass before Weathering

| Cultural relic sampling point | SiO₂ | K₂O | CaO | MgO | Al₂O₃ | Fe₂O₃ | CuO | PbO | BaO | P₂O₅ | SrO |
|-------------------------------|------|-----|-----|-----|-------|-------|-----|-----|-----|------|-----|
| 02                            | 77.35| 0.64| 1.07| 1.08| 7.66  | 1.81  | 0.19| 22.81| 0.00| 0.32 | 0.11 |
| 11                            | 71.61| 0.31| 1.61| 0.65| 3.60  | 0.00  | 3.66| 12.21| 15.04| 0.85 | 0.22 |
| 19                            | 63.19| 0.00| 1.34| 0.54| 4.78  | 1.30  | 2.61| 20.59| 5.51 | 0.80 | 0.11 |
| 34                            | 76.28| 0.36| 0.36| 0.00| 2.17  | 0.46  | 1.12| 22.38| 10.30| 0.03 | 0.13 |
| 36                            | 84.36| 0.20| 0.17| 0.00| 2.14  | 0.31  | 0.51| 20.01| 11.15| 0.01 | 0.13 |
| 38                            | 70.21| 0.00| 0.31| 0.00| 3.44  | 0.28  | 0.54| 23.71| 10.08| 0.04 | 0.24 |
| 39                            | 55.96| 0.00| 0.51| 0.00| 0.67  | 0.00  | 0.65| 29.35| 7.43 | 0.10 | 0.36 |
| 40                            | 35.63| 0.00| 0.86| 0.00| 0.60  | 0.19  | 0.00| 33.76| 6.89 | 0.16 | 0.40 |
| 41                            | 39.36| 0.64| 2.27| 2.50| 4.45  | 1.75  | 0.14| 21.22| 10.05| 0.67 | 0.28 |
| 43(1)                         | 26.46| 0.00| 2.40| 0.81| 3.01  | 0.74  | 3.97| 28.78| 7.51 | 0.00 | 0.38 |
| 43(2)                         | 46.26| 0.00| 2.93| 0.87| 4.56  | 1.36  | 1.12| 21.52| 3.36 | 1.16 | 0.28 |
| 49                            | 61.38| 0.00| 2.10| 1.35| 7.20  | 1.81  | 0.52| 16.44| 6.28 | 1.00 | 0.27 |
| 50                            | 38.33| 0.00| 1.46| 0.43| 2.50  | 0.32  | 0.84| 21.16| 14.62| 0.57 | 0.39 |
| 51(1)                         | 52.47| 0.00| 1.64| 1.09| 7.02  | 1.16  | 1.02| 19.35| 9.20 | 0.73 | 0.23 |
| 51(2)                         | 45.52| 0.00| 2.35| 1.33| 3.36  | 0.41  | 0.56| 24.69| 0.00 | 0.79 | 0.00 |
| 52                            | 54.88| 0.00| 1.04| 0.50| 1.55  | 0.22  | 0.52| 22.80| 8.89 | 0.52 | 0.26 |
| 54                            | 47.50| 0.47| 1.46| 1.17| 5.55  | 0.00  | 0.62| 26.67| 7.25 | 0.38 | 0.52 |
| 54*                           | 36.48| 0.00| 1.02| 1.02| 4.88  | 0.00  | 1.00| 28.11| 0.00 | 1.28 | 0.66 |
| 56                            | 62.15| 0.00| 0.55| 0.00| 2.47  | 0.00  | 0.59| 19.84| 15.91| 0.23 | 0.00 |
| 57                            | 54.20| 0.00| 0.60| 0.00| 2.92  | 0.00  | 0.86| 21.69| 17.81| 0.00 | 0.00 |
| 58                            | 64.79| 0.49| 1.60| 0.72| 4.71  | 0.84  | 2.32| 18.92| 7.89 | 0.81 | 0.14 |

*Indicates a severely weathered point

Figures in brackets indicate different sampling positions of the same glass sample
3.2 Glass classification

3.2.1 Establishment and solution of the decision tree model

The chemical composition content data and whether there is weathering (represented by 0-1 binary dummy variable) are imported into SPSS. In addition, the training set and test set data are divided by a ratio of 7:3, and a decision tree classification model is established. The structure of the decision tree is shown in Figure 1.

![Figure 1: Structure of glass type division decision tree](image)

It can be seen from the figure that there are 42 data in the training set, which are divided into 2 categories, 27 and 15 respectively, and the gini coefficients are all 0. In conclusion, under the condition of stable prediction of the decision tree model, the lead oxide content can be used as the basis for dividing the high potassium glass and the lead barium glass: when the lead oxide content is less than 6.965, the glass type is high potassium glass; When the lead oxide content is more than 6.965, the glass type is lead barium glass [4].

3.2.2 Cluster analysis

To eliminate the impact of different sizes of indicators of different properties in the new processing table in the appendix, the data is scaled and mapped to the range of (0,1), and the data is processed in the same way and dimensionless, to better carry out comparative evaluation and comprehensive analysis. Based on normalization, the analysis shows that the subcategory division in this chapter conforms to the characteristics of cluster analysis, that is, the process of dividing a dataset into multiple groups or clusters composed of several similar objects, which maximizes the similarity of the same group or cluster and minimizes the similarity between different clusters. Hierarchical clustering, as one of the most commonly used methods of cluster analysis, first classifies the elements to be clustered into one category, then selects the two categories with the smallest distance to merge into one, calculates the distance between the new category and other categories, and then repeatedly merges the two categories with the closest distance until all samples are merged into one category to achieve the purpose of common clustering. It can identify any shape of sample space and converge to the global optimal solution. For the high potassium glass samples in the table, the results of the pedigree clustering method performed by Matlab are shown in Figure 2 and Figure 3.

![Figure 2: Hierarchical cluster diagram of high potassium glass](image)

![Figure 3: Classification and evaluation of high potassium glass](image)
From the above figure, with the help of the "elbow rule", we can see that the degree of distortion is significantly improved when k=3. Considering that k=3 is selected as the number of clusters, we can get the sub-classification results of high potassium glass as shown in Table 4:

Table 4: Classification of High Potassium Glass Subclasses

| Classification | Cultural relic sampling point |
|----------------|-------------------------------|
| I              | 01, 03(2), 04, 05, 06, 21, 13 |
| II             | 03(1), 07, 09, 10, 12, 18, 22, 27 |
| III            | 14, 16                        |

However, the classification standard of the hierarchical clustering method for lead barium glass is not obvious. Because of the large amount of data and small differences between classes, it is considered to use other analysis methods that are faster and easier to achieve clustering, such as the K-means clustering analysis method, which takes k points in space as the center to cluster, classify the objects closest to them, and update the values of each cluster center successively through iterative method until the best clustering result is obtained.

Due to the randomness of the initial center value of the K-means model, multiple K-means models are repeatedly trained, and different k-values are selected with the help of evaluation functions until a relatively appropriate clustering category is obtained. As shown in the figure below, when k=6, the degree of distortion is greatly improved. Considering that k=6 is selected as the clustering number of lead barium glasses, the final classification results of lead barium glasses are shown in Figure 4 and Table 5:

![Figure 4: K-means cluster diagram of lead barium glass](image)

Table 5: Classification of lead barium glasses

| Classification | Cultural relic sampling point |
|----------------|-------------------------------|
| I              | 19, 39, 40, 50, 51, 52, 54, 54* |
| II             | 02, 29^, 30, 31, 41, 49, 49^ |
| III            | 25^, 28^, 32, 33, 34, 35, 36, 38, 44^, 45, 46, 53^, 55, 56, 57 |
| IV             | 11, 24, 37 |
| V              | 23^, 42^, 47 |
| VI             | 43 |

*Indicates a severely weathered point
^Indicates an unweathered point

3.2.3 Analysis of classification results

According to the reverse inference and qualitative analysis of the division results, the high potassium glass can be divided into three subcategories. It can be seen that the main division basis may be the content of silica. The essence of glass is silicate. After weathering, silicate will be dehydrated and decomposed into silica, and the composition of some metal oxides will also increase.
Therefore, the classification basis may be related to the weathering degree. These three categories are understood as "low weathering degree" (less silica content, low alkali gold oxide content) "suspected weathering" (acid oxide content decreases, alkali metal oxide content increases, but silicon dioxide content remains at a low level), and "high weathering degree" (high silicon dioxide content and high alkali metal oxide content) have obvious classification standards, and can identify the possibility of local weathering in nonweathered glass, which is consistent with the subject, and further proves the rationality and stability of the model[7].

3.3 Establishment and solution of glass type identification model

3.3.1 Judgment of glass type

According to the decision tree model in question 2, it is judged that A1~A8 are respectively high potassium glass, lead barium glass, lead barium glass, lead barium glass, lead barium glass, high potassium glass, high potassium glass, and lead barium glass.

To determine the accuracy of the decision tree model, the BP neural network method is used to verify. After the data standardization of the attachment new processing form 2, add the virtual variable type (if the glass type is "high potassium glass", the value is "0"; if the glass type is "lead barium glass", the value is "1"), build a neural network model, learn all kinds of chemical components and glass types through Matlab machine to master the rules, and judge the glass type in the attachment form 3 with the model after training. Among them, the predicted result is between -0.25 and 0.25, which approximately considers that the glass type is "high potassium glass"; Results It is considered as "lead barium glass" in the range of 0.75~1.25. The prediction result of the BP neural network model is consistent with the judgment result of the previous decision tree model, that is, the accuracy of the above judgment result of A1~A8 glass type in Appendix Table 3 is verified[3].

3.3.2 Sensitivity analysis of classification results

The sensitivity of the neural network model was tested with SPSS, and the results are shown in Table 6.

| Sample serial number | forecast result | actual result | Possibility of being lead barium glass | Possibility of high potassium glass |
|----------------------|----------------|--------------|---------------------------------------|-----------------------------------|
| A1                   | high potassium | high potassium | 0.005686334                           | 0.994313666                       |
| A2                   | lead barium    | lead barium  | 0.999999743                           | 2.57E-07                          |
| A3                   | lead barium    | lead barium  | 0.999999994                           | 5.72E-09                          |
| A4                   | lead barium    | lead barium  | 0.999991513                           | 8.49E-06                          |
| A5                   | lead barium    | lead barium  | 0.972227982                           | 0.027772018                       |
| A6                   | high potassium | high potassium | 0.001590052                          | 0.998409948                       |
| A7                   | high potassium | high potassium | 0.002769367                          | 0.997230633                       |
| A8                   | lead barium    | lead barium  | 0.999900831                           | 9.92E-05                          |

It can be seen from the above table that the probability of error in the prediction results is extremely low, and the prediction results are still stable[8].

At the same time, take the content of various chemical components as the input data, and use Matlab to calculate the sensitivity of various indicators to the prediction results. Among them, the sensitivity values of other chemical substances except for sulfur dioxide and lead oxide are extremely low, indicating that the prediction results are still stable when the content of other chemical substances changes or the data is inaccurate. Given the abnormality of the sulfur dioxide sensitivity value,
considering that the data processed by the box chart may have an impact on the calculation result of sulfur dioxide sensitivity value, the sensitivity value of sulfur dioxide to the predicted result will be reduced after machine learning again with the data not processed by the box chart.

To sum up, the prediction of the BP neural network is generally stable. The sensitivity of chemical substances other than lead oxide is low, while the sensitivity of lead oxide to the prediction results is high, that is, the verification of lead oxide content can be used as the basis for judging the glass type. Therefore, the classification results predicted by the decision tree and neural network model are reasonable and stable.

3.4 Establishment and solution of chemical composition correlation analysis model

3.4.1 Analysis of the strength of correlation

SPSS is used for grey correlation analysis. Taking high potassium glass as an example, the results are shown as shown in Table 7. The lead barium glass is treated in the same way (in which, the first row of indicators represents the parent sequence for each grey correlation analysis, the first column represents the characteristic sequence indicators, and the correlation degree of indicators is 1):

|     | SiO₂ | K₂O | CaO | Al₂O₃ | Fe₂O₃ | CuO | P₂O₅ |
|-----|------|-----|-----|-------|-------|-----|-------|
| SiO₂ | 1    | 0.587 | 0.555 | 0.639 | 0.504 | 0.602 | 0.611 |
| K₂O  | 0.608 | 1   | 0.796 | 0.752 | 0.691 | 0.66 | 0.699 |
| CaO  | 0.553 | 0.792 | 1   | 0.762 | 0.741 | 0.679 | 0.679 |
| Al₂O₃| 0.655 | 0.744 | 0.759 | 1   | 0.683 | 0.625 | 0.75  |
| Fe₂O₃| 0.537 | 0.71  | 0.763 | 0.714 | 1   | 0.714 | 0.732 |
| CuO  | 0.658 | 0.685 | 0.711 | 0.664 | 0.718 | 1   | 0.75  |
| P₂O₅ | 0.639 | 0.702 | 0.689 | 0.761 | 0.716 | 0.732 | 1     |

Due to different reference data columns, the correlation degree between the chemical components of glass cultural relics samples of the same category is slightly different. Therefore, this paper will take the average of the two correlation degree values between the chemical components of glass samples of the same category to represent the correlation degree between the chemical components of glass samples of the same category. The specific results are shown in Table 8.

|     | SiO₂ | K₂O | CaO | Al₂O₃ | Fe₂O₃ | CuO | P₂O₅ |
|-----|------|-----|-----|-------|-------|-----|-------|
| SiO₂ | 1    |     |     |       |       |     |       |
| K₂O  | 0.5975 | 1  |     |       |       |     |       |
| CaO  | 0.554 | 0.794 | 1   |       |       |     |       |
| Al₂O₃| 0.647 | 0.748 | 0.7605 | 1 |       |     |       |
| Fe₂O₃| 0.5205 | 0.7005 | 0.752 | 0.6985 | 1 |     |       |
| CuO  | 0.63 | 0.6725 | 0.695 | 0.6445 | 0.716 | 1  |       |
| P₂O₅ | 0.625 | 0.7005 | 0.684 | 0.7555 | 0.724 | 0.741 | 1     |

The magnitude of the correlation degree indicates the strength of the correlation relationship, that is, the correlation between chemical components with a high correlation degree is stronger than that between chemical components with a low correlation degree in the table.\(^{[9]}\)
3.4.3 Analysis of the difference in association relationship

Compare the correlation between similar chemical components in two types of glass samples, the result is shown in Table 9.

Table 9: Comparison of correlation between composition content of high potassium glass and lead barium glass

|                        | Lead barium glass | High potassium glass |
|------------------------|-------------------|----------------------|
| SiO$_2$/CaO            | 0.715             | 0.554                |
| SiO$_2$/Al$_2$O$_3$     | 0.797             | 0.647                |
| SiO$_2$/CuO            | 0.726             | 0.63                 |
| SiO$_2$/P$_2$O$_5$      | 0.6485            | 0.625                |
| CaO/Al$_2$O$_3$         | 0.777             | 0.7605               |
| CaO/CuO                | 0.7385            | 0.695                |
| CaO/P$_2$O$_5$          | 0.814             | 0.684                |
| Al$_2$O$_3$/CuO         | 0.7145            | 0.6445               |
| Al$_2$O$_3$/P$_2$O$_5$  | 0.6825            | 0.7555               |
| CuO/P$_2$O$_5$          | 0.7275            | 0.741                |

Through the analysis of the data in the above table, we can get the difference in chemical composition correlation between different categories:

The correlation between silica and calcium oxide, silica and alumina, silica and copper oxide, and silica and phosphorus pentoxide is smaller in high potassium glass than in lead barium glass; The correlation between calcium oxide and alumina, calcium oxide and copper oxide, calcium oxide and phosphorus pentoxide, alumina and copper oxide is smaller in high potassium glass than in lead barium glass; The correlation between alumina and phosphorus pentoxide, copper oxide and phosphorus pentoxide is greater in high potassium glass than in lead barium glass.

4. Conclusion

(1) The classification standard is obtained through the decision tree, and the prediction results of the BP neural network are used to verify each other. Moreover, the adaptive ability and fault tolerance of neural networks make the model results not only intuitive and easy to understand but also stable and reliable based on the decision tree.

(2) Two cluster analysis methods are used to classify high potassium glass and lead barium glass, making full use of the complementary advantages of the two methods, and analyzing the rationality and sensitivity of the classification results.

(3) The use of grey relational degree is a comprehensive evaluation model combining qualitative analysis and quantitative analysis, which makes the evaluation results more impressive and accurate. Only a few representative samples can have strong reliability and rationality.

References

[1] Gu Liangliang, Zhou Jing. Some problems in the study of ancient Chinese glass art [J]. Art and Design, 2021, 2(04): 112-114.
[2] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com.
[3] Mao Shisong, Wang Jinglong, Pu Xiaolong, et al. Advanced mathematical statistics [M]. Beijing: Higher Education Press, 2006.
[4] Zhou Zhihua. Machine learning [M]. Tsinghua University Press, 2016.
[5] Jain, A. K., and R. C. Dubes . "Algorithms for clustering data." Technometrics 32.2(1988): 227-229.
[6] Saroj, Kavita. Review: a study on simple k mean and modified K means clustering technique [J]. International Journal of Computer Science Engineering and Technology, 2016, 6(7): 279-281.

[7] Wang Chengyu, Tao Ying. The weathering of silicate glasses [J]. Journal of the Chinese Ceramic Society, 2003(01): 78-85.

[8] Wu Jianhui, Wabg Guoli, Yu Liqun, et al. Sensitivity analysis and application of BP neural network [C]. // The 3rd Shanxi-Hebei-Luyu Epidemiology Academic Conference and 2010 Epidemiology Zhengzhou Forum Proceedings. 2010: 160-165.

[9] Azzeh M, Neagu D, Cowling P I. Fuzzy grey relational analysis for software effort estimation. [M]. Kluwer Academic Publishers, 2010.