The Principal Component Analysis and Cluster Analysis of Trace Elements in Gentian — Source link

Xin Zhao, Mingwei Xu, Guoqing Sun, Yang Jiao ...+3 more authors

Institutions: Huizhou University, College of Information Technology, City College of New York

The Principal Component Analysis and Cluster Analysis of Trace Elements in Gentian

Xin Zhao, Mingwei Xu, Guoqing Sun, Yang Jiao, Haijiao Yu, Quanming Li, Guogang Zhao

HAL Id: hal-02179966

https://hal.inria.fr/hal-02179966

Submitted on 11 Jul 2019

The Principal Component Analysis and Cluster Analysis of Trace Elements in Gentian

Xin Zhao[1], Mingwei Xu[2], Guoqing Sun[2], Yang Jiao[3], Haijiao Yu[4], Quanming Li[4], Guogang Zhao[4(✉)]

[1]School of Mathematics and Big Data, Huizhou University, Huizhou, China, 516007 China;
zhaoxin@hzu.edu.cn

[2] College of Information Technology, Jilin Agricultural University, Changchun, 130118 China;
364685279@qq.com, 397820020@qq.com

[3]Sixth Middle School in Changchun, Changchun, 130000,China;
61516131@qq.com

[4]The City College of JLJU, Changchun, 130000,China;

haihai951kl@163.com, keeper007@126.com, zhaogg2010@foxmail.com

Abstract. In this paper, using principal component analysis and cluster analysis method, combining with the statistics software MATLAB. The Qingyuan rough gentian, Zuojia rough gentian, Zuojia leaf gentiana system clustering analysis of experimental data. Main ingredients swertia glycosides, Swertia bitter glycosides , gentian bitter glycosides and oleanolic acid and ursolic acid, etc. etc. Cluster analysis results showed that the three principal component contribution rate of the Qingyuan rough gentian is 89.24% , the three principal component contribution rate of the Zuojia rough gentian is 89.85% , the three principal component contribution rate of the Zuojia leaf gentian is 93.56%. By cluster analysis to Qingyuan rough gentian, Zuojia rough gentian, Zuojia leaf gentiana aristata 3 groups of data were divided into 6, 5, 4 classes, and determine the appearance characteristics of the high quality gentian, provide the basis for breeding to select high quality gentian.

Keywords: Gentian · Principal component analysis · System clustering analysis

1   Introduction

Gentian for commonly used Chinese traditional medicine, has clear heat dry wet and the efficacy of purging liver fire. Gentian including gentian bitter glycosides and swertia glycosides, Swertia bitter glycosides , and oleanolic acid and ursolic acid, etc. Commonly used in damp heat jaundice, red eyes, mouth, bitter, pain such as disease treatment. Of gentiana bitter glycosides and swertia glycosides, r swertia bitter glycosides, oleanolic acid and ursolic acid content of five elements, principal component analysis and cluster analysis in order to select high-quality gentian provided the scientific basis and theoretical basis for breeding.

2   Experiments and Methods

2.1   Experimental Samples

Data sources for the analysis in Jilin Agricultural University College of Chinese medicinal materials of gentian principal component extraction of experimental data, data 200 experimental data, through data cleaning, after screening the final selection of 90 data, and divided into three groups.

2.2   Experimental Methods

Principal component analysis is the use of data dimension reduction to exclude chemical information overlaps and correlation of multivariate linear statistical techniques, principal component analysis can be used to study the samples the dependencies between different data, looking for a comprehensive factor, for the comprehensive evaluation of the sample. This paper uses the MATLAB software to carry

on the principal component analysis. First of all, the article to gentian bitter glycosides and swertia glycosides, swertia bitter glycosides, and oleanolic acid and ursolic acid,content of five elements principal component analysis. According to the various components of gentian, choice x1: swertia glycosides;X2: swertia bitter glycosides; X3: gentian bitter glycosides;X4: ursolic acid ; X5: oleanolic acid content for further five indexes such as principal component analysis, principal component ranking and contribution rate are obtained.

Clustering analysis is a method of mathematical statistics, suitable for sample attribution is not clear situation. Based on many properties of the sample, according to the sample of the nature of closeness degree for automatic classification of multivariate statistical analysis methods. First define the distance between the sample in each type of sample, will be the nearest two kind of merger, to calculate the distance between new class and other classes, and according to the minimum distance classification, Repeat this process, at every time reducing 1 class , until all become a kind of sample. The 200 item experimental gentian data from college of Chinese Medicine Materials of jilin agricultural university, going through data cleaning, screening, finally chooses 83 item data, and divided into three groups.

3   Results and Discussion

3.1   The result of principal component analysis

By gentian element characteristic value and the variance contribution rate of data table, we can get three principal component contribution rate. Three principal component contribution rate is respectively: Qingyuan rough gentian three principal component contribution rate of 89.24%, Zuojia rough gentian three principal component contribution rate of 89.85%, Zuojia leaf gentiana before gentiana aristata three principal component contribution rate 93.56%. So on the premise of less loss of information can be described with three principal components of trace element content in gentiana all of the data, it satisfies the basic demands of principal component analysis. According to the score ranking, No. 30 samples of qingyuan rough gentian, has the characteristics of green stem leaves and purple flowers is of high quality. No.26 samples of Zuojia rough gentian, has the characteristics of purple stem leaves and blue flowers is of high quality. No. 22 samples of Zuojia leaf gentian, has the characteristics of green stem leaves and blue flowers is of high quality.

Table 1. The contribution of Qingyuan rough gentian

eigenvalue	feature vectors	contribution
2.1494	(0.3407,0.2956,-0.0322,0.6109,0.6499)	0.4299
1.3556	(0.1462,0.6471,0.7049,-0.2080,-1405)	0.2711
0.9569	(-0.8132,-0.1215,0.4173,0.3428,0.1799)	0.1914
0.4811	(-0.4270,0.6921,-0.5723,-0.0165,-0.1038)	0.0962
0.0570	(0.1378,0.0104,0.0202,0.6824,-0.7175)	0.0114

Table 2. The ranking of the principal component of Qingyuan rough gentian

Serial number	score	rank	Serial number	score	rank	Serial number	score	rank
1	-0.1060	27	11	0.0002	13	21	-0.0282	19
2	-0.0365	20	12	-0.0260	18	22	0.2199	2
3	0.0692	6	13	-0.0686	23	23	0.0684	7
4	-0.0189	16	14	-0.0230	17	24	0.0365	10
5	-0.0853	26	15	-0.0623	22	25	0.0742	5

2

6	0.0447	9	16	0.1298	3	26	-0.0047	14
7	-0.1201	29	17	-0.0782	24	27	0.0946	4
8	-0.1863	30	18	-0.0531	21	28	0.0058	12
9	-0.1115	28	19	-0.0142	15	29	0.0675	8
10	-0.0815	25	20	0.0292	11	30	0.2644	1

Table 3. The contribution of Zuojia rough gentian

eigenvalue	feature vectors	contribution
2.3406	(-0.4797, -0.4685, -0.4564, 0.4848 0.3272)	0.4681
1.5342	(0.1129, 0.4455, 0.3852, 0.4658, 0.6507)	0.3068
0.6192	(0.8198, -0.0804, -0.5333, 0.1490, 0.1218)	0.1238
0.3096	(-0.2831, 0.7557, -0.5801, -0.0227, -0.1086)	0.0619
0.1964	(-0.0708, -0.0672, -0.1492, -0.7248, 0.6655)	0.0393

Table 4. The ranking of the principal component of Zuojia rough gentian

Serial number	score	rank	Serial number	score	rank	Serial number	score	rank
1	-0.1294	20	11	0.2448	5	21	-0.3305	29
2	-0.1423	22	12	-0.1189	19	22	-0.1336	21
3	-0.3906	30	13	0.1824	8	23	0.0810	11
4	-0.2006	25	14	0.2355	7	24	0.0209	15
5	0.1372	10	15	0.0508	13	25	-0.2495	26
6	0.3090	3	16	-0.2509	27	26	0.3406	1
7	0.0276	14	17	0.1688	9	27	0.0204	16
8	-0.2790	28	18	-0.1697	23	28	0.0553	12
9	0.2402	6	19	0.3153	2	29	-0.0346	17
10	-0.0812	18	20	0.2645	4	30	-0.1835	24

Table 5. The contribution of Zuojia leaf gentian

| eigenvalue	feature vectors	contribution
2.6302	(-0.0628, -0.5306, -0.3829, 0.5362, 0.5295)	0.5260
1.0917	(-0.5542, 0.2787,0.6283,0.3264, 0.3374)	0.2183
0.9565	(0.8263, 0.2190, 0.3368, 0.2661, 0.2915)	0.1913
0.2172	(0.0775, -0.7698, 0.5868, -0.1877, -0.1479)	0.0434
0.1044	(0.0130, -0.0131, 0.0293, 0.7071, -0.7063)	0.0209

Table 6. The ranking of the principal component of Zuojia leaf gentian

Serial number	score	rank	Serial number	score	rank	Serial number	score	rank
1	0.0573	10	11	-0.0745	14	21	-0.0713	12
2	0.1012	7	12	-0.1426	18	22	0.4794	1
3	-0.2020	20	13	-0.1723	19	23	0.2077	2
4	-0.0742	13	14	-0.1084	17			

|					
5	-0.2568	23	15	0.2054	3
6	-0.0853	15	16	0.0907	8
7	0.0497	11	17	-0.0876	16
8	-0.2194	21	18	0.1914	4
9	-0.2502	22	19	0.1691	5
10	0.1299	6	20	0.0628	9

3.2 Clustering analysis

According to 83 item data, through the different sections of the clustering analysis of gentiana different shape characteristics, find out the similar sex between them (i.e., the degree of close relatives). Using matlab to gentiana standardzing trace elements in the original data. According to the clustering analysis of the processes associated with select observation object clustering, using the Euclidean distance is analyzed, using Average linkage connections between every two sample, drawing in sequence. The corresponding results are shown in figure.
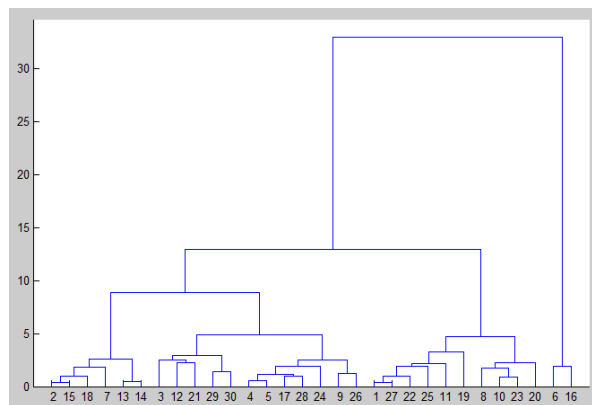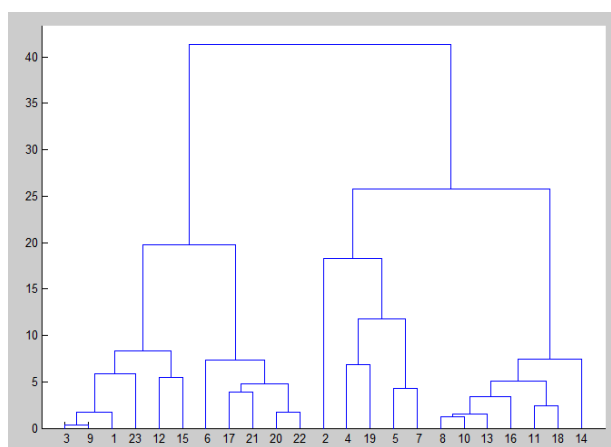

(a)　Clustering analysis results of Qingyuan rough gentian

Cluster analysis results showed that the 30 item data of Qingyuan rough gentian can be divided into six categories. (1,27,22,25,11,19) belong to the same class, (2,15,18,7,13,14)belong to the same class, (3.12,21,29,30) belong to the same class, (4,5,17,28,24,9,26) belong to the same class,(6,16) belong to the same class,(8,10,23,20) belong to the same class, (2, 15, 13, 14, 4, 5, 1, 27) have main component elements content is low.
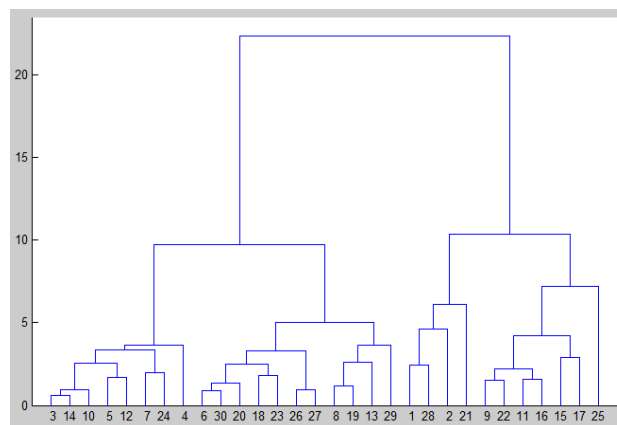
(b)　Clustering analysis results of Zuojia rough gentian

4

Cluster analysis results showed that the 30 item data of Qingyuan rough gentian can be divided into five categories. (1,28,2,21) belong to the same class, (3,14,10,5,12,7,24)belong to the same class, (6,30,20,18,23,26,27)belong to the same class,(8,19,13,29)belong to the same class, (9,22,11,16,15,17,25) belong to the same class, (3,14) have main component elements content is low.

(c) Clustering analysis results of Zuojia leaf gentian

Cluster analysis results showed that the 23 item data of Qingyuan rough gentian can be divided into four categories. (1,3,9,23,12,15) belong to the same class, (2,4,19,5,7)belong to the same class, (6,17,21,20,22) belong to the same class, (8,10,13,16,11,18,14) belong to the same class, (3,9) have main component elements content is low.

5   Conclusions

To sum up, both by principal component analysis and cluster analysis to find out containing trace elements of high quality gentian, and through its shape characteristics for identification. In order to further choose to breeding new varieties of gentiana provide guidance. By principal component analysis and cluster analysis methods to analyze gentian trace elements, can provide abundant information for traditional Chinese medicine formula and theoretical guidance.

6   Acknowledgement

References

1.   Zhou libing, Jiang ziqin, Wu qixun. Qinghai region white spines of trace elements in the leaves and principal component analysis and cluster analysis. Anhui agricultural science, i . 38(13), 6649-6650, 6652(2010)
2.   Zhao rui zhi, Grad leung, Qiu xiaohui. In Gentian medicinal materials the extraction process of gentiana bitter glycosides . China Pharmacy, 12(2005)
3.   Zhou changhui, Gao yanzi, Qi zhengxing, Wu qixun, Chen ying. The tea trace elements of principal component analysis and cluster analysis. Journal of nanyang institute of technology, 4(2), 104-106(2012)
4.   Wang aiping, Zhu luyi. Different regions of sichuan dangshen clustering analysis of trace elements[J]. Trace elements in guangdong science, 16(6), 57-61(2009)
5.   Wang huiqin, Xie mingyong, Yang miaofeng. The safflower trace elements in different regions factor analysis and cluster analysis[J]. Journal of xiamen university, 45(1), 72-75(2006)
6.   Zhang yifan. The trace elements in traditional Chinese medicine radix scutellariae principal component analysis and cluster analysis. The guide newspaper of traditional Chinese medicine, 10(2012)

7. Cao hongcui, Wu qixun. The trace elements in Chinese traditional medicine licorice principal component analysis and systematic cluster analysis. the Proceedings of gansu union university, 3(2010)