



# Application of Principal Component Cluster Analysis in the Quality of Cordyceps Sinensis

Xin Zhao, Hongjian Yang, Yuqi Sheng, Yang Jiao, Haijiao Yu, Yao Qin,  
Guogang Zhao

## ► To cite this version:

Xin Zhao, Hongjian Yang, Yuqi Sheng, Yang Jiao, Haijiao Yu, et al.. Application of Principal Component Cluster Analysis in the Quality of Cordyceps Sinensis. 10th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Oct 2016, Dongying, China. pp.279-283, 10.1007/978-3-030-06155-5\_27 . hal-02179989

**HAL Id: hal-02179989**

**<https://inria.hal.science/hal-02179989>**

Submitted on 12 Jul 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Application of Principal Component Cluster Analysis in the Quality of Cordyceps Sinensis

Xin Zhao<sup>1</sup>, Hongjian Yang<sup>2</sup>, Yuqi Sheng<sup>2</sup>, Yang Jiao<sup>3</sup>, Haijiao Yu<sup>4</sup>, Yao Qin<sup>4</sup>, Guogang Zhao<sup>4</sup>(✉)

<sup>1</sup>School of Mathematics and Big Data, Huizhou University, Huizhou, China, 516007,China;  
zhaoxin@hzu.edu.cn

<sup>2</sup> College of Information Technology, Jilin Agricultural University, Changchun, 130118, China;  
1612192868@qq.com, 53029098@qq.com

<sup>3</sup>Sixth Middle School in Changchun, Changchun, 130000, China;  
61516131@qq.com

<sup>4</sup>The City College of JLU, Changchun, 130000, China;  
haihai951kl@163.com, 470695861@qq.com, zhaogg2010@foxmail.com;

**Abstract.** In this paper, the kinds and contents of amino acids in Cordyceps sinensis from different habitats of Tibet components using principal component analysis and cluster analysis, the principal component cluster analysis to evaluate the nutritional value of different localities, and provide scientific basis for the further research and development and utilization of Cordyceps resources in Tibet area. The principal component analysis method is a method of using the central idea of dimensionality reduction, the analysis method of multi indicators into a multivariate data several comprehensive index of a few statistics. The principal component analysis method can guarantee the minimizing loss of original data information, with less comprehensive variables instead of multiple variables of the original. Cluster analysis can be used to classify samples of multiple variables by using comprehensive information, the classification results are intuitive, clustering dendrogram clearly show the results of numerical classification, clustering analysis results than the traditional classification method is more detailed, comprehensive, reasonable.

**Keywords:** Cordyceps sinensis · amino acid · principal component analysis · cluster analysis

## 1 Introduction

Cordyceps sinensis is Cordyceps sinensis Sacc parasitic on the drying complex of stroma of the parasitic bat moth larvae on the insects and larval body<sup>[1]</sup>, it is a traditional rare tonic medicine in China, the earliest found in the "Materia Medica from the new", it is the nature of "gamping", have the effect of "protect lung and tonifying kidney, eliminating phlegm, relieving cough". Modern pharmacological research show that the Cordyceps sinensis has to adjust the body's immune<sup>[2]</sup>, fall hematic fat, fall blood sugar, anti-aging, anti-tumor and many kinds of pharmacological activities<sup>[3-4]</sup>, the amino acid is material base of Cordyceps sinensis to strong effect and enhance immune function<sup>[5]</sup>, and it is closely related to the function of the auxiliary treatment of nervous system<sup>[6]</sup>, digestive system disease and antibacteria. Zhang Shishan<sup>[7]</sup> research showed that some amino acids in Cordyceps sinensis has sedative effects. The amino acid of Cordyceps sinensis not only has a

wide range of pharmacological effects and important nutrients, so the amino acid research has positive significance. China's Tibet the main producing areas of Cordyceps sinensis, at present, for Tibet within the jurisdiction of different habitats component analysis of amino acid in Cordyceps sinensis has not been reported. Therefore, this paper intends to analysis and evaluation of nutritional value of amino acids in Cordyceps sinensis from different regions of Tibet, in order for further research and development by China's Tibet Cordyceps sinensis resources to provide a scientific basis.

## 2 Experiments and Methods

### 2.1 Experimental Samples

The origin of Tibet Cordyceps sinensis contains 17 kinds of amino acids, 7 kinds of essential amino acids, 4 kinds of delicious amino acids, 9 medicinal amino acids. The range of total amino acids between the 23.4%~24.52%, essential amino acids accounted for total amino acids between the 26.47%~27.52%, fresh flavor amino acids accounted for total amino acids between the 35.12%~37.44% and drug effective amino acids accounted for total amino acids in the range between the 55.03%~57.54%. In this paper, we use the data of Yan Dong and Yang Xinyu<sup>[13]</sup>, The method is improved by using Principal component cluster analysis.

### 2.2 Principal component cluster analysis

In the principal component analysis method, when the variance contribution of the first principal component was higher (above), we can think the principal component can reflect the information with the original variables provide information almost, this is the time to sort and evaluation according to the scores of the first principal component.

(1) Erroneous zone of comprehensive score of principal component.

In many cases, the variance contribution rate of the first principal component is not high enough, need to consider a number of main components  $F_1, F_2, \dots, F_r$ , in order to sort the sample, commonly used method is based on the variance of each principal component  $F_i$  contribution rate  $a_i$  for the weight, the comprehensive score of the principal component is:  $H = a_1 F_1 + a_2 F_2 + a_3 F_3 + \dots + a_r F_r$ , and according to the formula to calculate the main component of the sample and sort them.

The comprehensive score calculation looks very reasonable and it seems to increase the information content, that is, to increase the variance contribution rate, in fact, it is just an illusion, through to calculate the variance of  $H$  we can found that it is possible

to completely opposite conclusions.

$$\text{Var}(H) = \sum_{i=1}^r a_i^2 \text{Var}(F_i) = \sum_{i=1}^r a_i^2 \lambda_i \leq \lambda_1 \quad (1)$$

From this process we find that the variance of the composite score is smaller than the variance of the first principal component, which indicates that the information content of the comprehensive principal component score is less than that of the first principal component.

## (2) principal component clustering

The scheduling problem in multi index system evaluation, If the variance contribution rate of the first principal component  $F_1$  is not large enough, that is, the first principal component expression of the original data is not large enough, only by the first principal component score on the evaluation of the sample will be one-sided, At this time, the principal component analysis and cluster analysis can be used to combine, and use the Principal component cluster analysis.

Clustering analysis is a method of samples or variables are classified according to the intimacy of their nature, although the cluster analysis can be a good classification of multiple indicators of the data, but it can't get the evaluation result of each good or bad degree.

The idea of principal component clustering is take the principal component analysis firstly, then several principal components of the sample clustering analysis, combined with the first principal component score ranking to classify the samples. A new comprehensive evaluation method is obtained, specific practices are as follows:

$$F_k = a_{1k}X_1 + a_{2k}X_2 + a_{pk}X_p, k = 1, 2, \dots, r. \quad (2)$$

Carry on the system clustering on the selected new datamatrix  $(F_1, F_2, \dots, F_r)$ . Then calculate the mean score of all kinds of samples to determine the sort of class; Finally, according to the first principal component score of the sample, the order of all kinds of samples was determined, and the comprehensive evaluation was obtained.

## 3 Results and Discussion

### 3.1 The data is processed by the method of cluster analysis

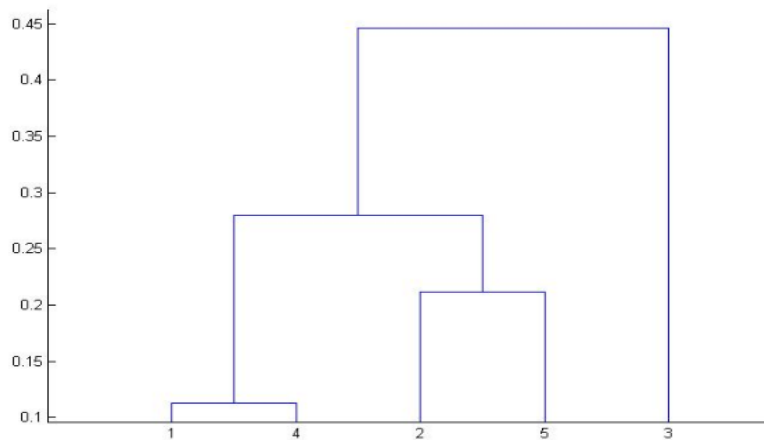


Fig. 1. Data processing results

Through MATLAB to carry on the cluster analysis for 5 kinds of Cordyceps sinensis,

Figure 1 Results that 1, 2, 4, 5 are clustered into a large class, 3 together for a class, The quality difference of amino acid of 1,2,4,5 was small, and the difference between the 3 and 1,2,4,5 amino acid was very large.

### 3.2 Carry on Principal Component Clustering:

Carry on principal component analysis about, We work out a linear combination of each of the principal components represented by each variable.

$$Z_1 = -0.193X_1 - 0.352X_2 - 0.315X_3 + 0.0961X_4 - 0.3485X_5 + \dots - 0.3438X_{16} - 0.1667X_{17} \quad (3)$$

$$Z_1 = 0.0092X_1 + 0.0575X_2 + 0.0974X_3 + 0.2436X_4 - 0.0305X_5 + \dots + 0.0705X_{16} - 0.3337X_{17} \quad (4)$$

$$Z_1 = -0.4808X_1 - 0.0853X_2 - 0.2361X_3 - 0.4658X_4 + 0.0543X_5 + \dots - 0.0562X_{16} - 0.0051X_{17} \quad (5)$$

**Table 1.**The score of amino acid content of Cordyceps sinensis

	Principal score one	component score two	Principal score three
Naqu Cordyceps	-4.592403298	2.582416524	-1.953775208
changdu Cordyceps	-4.326408906	2.528693027	-1.797035154
Linzhi Cordyceps	-4.855238038	2.373217091	-1.584953726
Shannan Cordyceps	-4.584720453	2.587483542	-2.066390962
Shigatse Cordyceps	-4.470809315	2.374504833	-1.810989538

## 4 Conclusions

By the clustering analysis method of 5 kinds of Cordyceps amino acid quality clustering results and found Linzhi Cordyceps contained amino acid quality with respect to the Nagqu Cordyceps, Changdu Cordyceps, Shannan Cordyceps, Shigatse Cordyceps higher. This shows that amino acids quality of Cordyceps sinensis in Linzhi have the large difference with Naqu Cordyceps, changdu Cordyceps, Shannan Cordyceps, Shigatse Cordyceps. Among the 17 kinds of amino acids, seven of them are essential amino acids, and ten of them are nonessential amino acids, from linear combination by principal component analysis we can draw five kinds of amino acid content of Cordyceps sinensis accounted for the proportion of all, essential amino acids of changdu Cordyceps is highest in five of the origin of Cordyceps sinensis, in umami amino acids, five of the origin of Cordyceps sinensis umami amino acid content was higher, Shannan Cordyceps umami amino acid content highest, which determines the delicious taste of Cordyceps sinensis, in drug effective amino acid, Cordyceps medicinal amino acids is very rich, containing 9 kinds of drug effective amino acids, in five of the origin of Cordyceps medicinal amino acids accounted for total 55.03%~57.54%, Shannan Cordyceps medicinal amino acids content of the highest, it is aweto nourishing and strengthening the immune by providing the material basis.

## References

1. Chinese Pharmacopoeia Commission. Pharmacopoeia of Peoples Republic of China [M]. Beijing: Chinese Medical Science and Technology Press, 106(2010)
2. Hu Min, Pi Huimin, Zheng Yuanmei. The Chemical Constituents And Pharmacological Effects of Cordyceps Sinensis[J]. Li Shizhen Med Mater Med Res, 19(11), 2804-2806(2008)
3. Zhang Bohua, Chen Jing. Research progress of Pharmacological Actions of Cordyceps Sinensis [J]. Information of Traditional Chinese Medicine, 29 (1), 134-135(2012)
4. Lei Wansheng, Xie Lianbin. Chen peace. The Research Situation of Cordyceps Sinensis [J]. The Journal of Naval Medicine, 27 (3), 262-269(2006)
5. Chen Yanyan, Qiu Ximin, Huang Siyu. Determination of Amino Acid Content in Different Parts of Cordyceps Sinensis [J]. Food Science, 32 (8), 259-261(2011)
6. Ji Sha, Shi Xiaobing, Yi Jun. The General Situation of The Study on the Chemical Constituents of Cordyceps Sinensis [J]. Journal of Fujian University of Traditional Chinese Medicine, 9 (2), 46-47(1999)
7. Zhang Shishan, Zhang Dansen, Zhu Tongjun. Analysis of Amino Acid in Cordyceps Sinensis Pharmacology [J]. Acta Pharmaceutica Sinica, 26 (5), 326-330(1991)
8. Ministry of Health, PRC. GB/T5009. 124——2003, Determination of Amino Acids in Foods [S], Beijing: China Standard Press(2004)
9. Zhu Shengtao, Wu Kun. Evaluation of Protein Nutritional Value——Amino Acid Ratio Coefficient Method [J]. Journal of Nutrition, (10), 187-190(1988)
10. Pelit PL, Yang VR. Nutritional Evaluation of Protein Food [M]. Beijing: People's Medical Publishing House, 55-58(1984)
11. Wang Li, Song Zhifeng, Huang Huang. Content Determination of Amino Acids in Cordyceps Sinensis From Different Habitats by HPLC [J]. Chinese Medicine Drug, 32 (6), 984-987(2010)
12. Sophie, Su Jun, Jiang Yan Cheng. A Comparative Study on the Content of Xinjiang Cordyceps and Mannitol in Cordyceps Sinensis, Polysaccharide and Amino Acids [J]. Xinjiang Agricultural Sciences, 45 (1),36-45, (3), 517-521(2008)