The Application of Student Evaluation of Online Teaching Based on Clustering Analysis Method

Lv Saidong
Institute of Security
Yunnan Normal University
Kunming,Yunnan Province,China
lsdong_km@sina.com

Xia Yaowen and Li Zhiping
Institute of Information
Yunnan Normal University
Kunming,Yunnan Province,China
xywen_km@sina.com

Abstract—**This paper analyzes the clustering analysis method in the education field in the specific application, in reading other papers at the same time, find shortcomings, the application of hierarchical clustering and fast clustering two methods mutual contrast, verify the clustering analysis method will be applied to online commentary on the rationality, and on the basis of the quantitative analysis with qualitative analysis, summed up to establish the basic principle of online evaluation.**

Keywords- Hierarchical Clustering Analysis,Fast Clustering Analysis, Online Commentary,SPSS,Evaluation Index

I. INTRODUCTION

Clustering analysis originated in taxonomy, it is study (sample or indicators) classification problem a statistical analysis method. Has been widely used in business, biology, medicine, and other fields, of course, in the education field application is very wide also.

In the Chinese know nets, wan fang database and VIP database search, cluster analysis is applied to education in the field of a total of 98 pieces of paper, but no Ph.D. Thesis, dissertation and article 83 dissertations, common. Practical application as is shown in table 1, which in the application of teaching quality evaluation is more, but after serious reading papers found that all the papers with only one clustering method, this paper studies the disadvantages, hierarchical clustering and fast clustering two methods mutual contrast verification analysis, found that using two different methods the classification results are basically consistent, tend to be verified the clustering analysis method is applied to the teaching quality evaluation of the rationality and scientific nature, and on the analysis of clustering results on the basis of summing up the establishment of the online teaching evaluation indicators of the basic principles.

TABLE I. CLUSTERING ANALYSIS IN THE FIELD OF EDUCATION OF THE SPECIFIC APPLICATION OF THE SITUATION ANALYSIS TABLE

application direction	Teaching quality evaluation	Educational administration	Score analyze	Library management	Examination paper analysis	School running conditions	Distance education	other
Paper count	33	10	10	9	4	5	11	16

II. THE ANALYSIS OF CLUSTERING ANALYSIS

As the saying goes: "birds of a feather flock together, birds of a feather flock together". Classification is the basis of the world people know. In the social, economic and natural phenomenon, there are a large number of classification problems. Clustering analysis is the study of "like attracts like" a multivariate statistical analysis method.

The basic idea of clustering analysis is based on the related degree between object classes of polymerization. In the cluster analysis before, these categories are hidden, can be divided into many categories is not know in advance. Clustering analysis principle is the same kinds of individuals have greater similarity, not the same a difference is very big.

A. Data transformation processing

In order to overcome the original data due to different units of measurement for clustering analysis the influence of the result is not reasonable. In the process of cluster analysis, the first deal with original data data transformation processing. Set the original data (1) constitute the data matrix[1]:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \qquad (1)$$

Each symbol is the following meaning:

N for the total number of samples (namely teacher total), p for the primitive variable number (namely ten evaluation norm), X_{ij} said I of the units in the first j a variable on the data value (that is, the first I a teacher in the first j indicators on the data value).

This paper briefly introduced standardized transformation method. Standardized transform the original data into standard Z score (Z score) transform method. The conversion formula is[2]:

$$X'_{ij} = \frac{X_{ij} - \overline{X}_j}{S_j} (i=1,2,\cdots,n, j=1,2,\cdots,p) \qquad (2)$$

Each symbol is the following meaning:

X'_{ij} said standardized data, $\overline{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$ said variable j the mean, standard deviation of the variable j S_j said, namely $S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2}$.

B. Choose clustering statistics

Research sample or variable density degree of quantitative index has two kinds, one kind is distance, another kind is similarity coefficient. The two kinds of index is used to reflect the sample or the variable difference size

statistics. Variable measurement scale is different, the statistics is also different[3].

1)The distance between the sample and the similarity coefficient.

- Euclidean distance: $d_{ij} = \sqrt{\sum_{k=0}^{p}(X_{ik} - X_{jk})^2}$ (3)

- Euclidian distance square: $d_{ij} = \sum_{k=1}^{p}(X_{ik} - X_{jk})^2$ (4)

2) Similarity coefficient.

- Pearson similarity coefficient:

$$d_{ij} = \frac{\sum_{k=1}^{p}(X_{ik} - \overline{X}_i)(X_{jk} - \overline{X}_j)}{\sqrt{\sum_{k=1}^{p}(X_{ik} - \overline{X}_i)^2 \sum_{k=1}^{p}(X_{jk} - \overline{X}_j)^2}} \quad (5)$$

- Angle cosine: (6)

$$d_{ij} = \cos(\theta_j) = \frac{\sum_{k=1}^{p} X_{ik} X_{jk}}{\sqrt{\sum_{k=1}^{p} X^2_{ik}} \sqrt{\sum_{k=1}^{p} X^2_{jk}}}$$

C. *Choose clustering method*

Determine the sample or variable similarity or not similarity statistics, will be to sample or variables are classified. Sample clustering and variable clustering method are many, in this paper, only this hierarchical clustering method and fast clustering method.

System clustering method is the most used a clustering method. As mentioned before, the basic point is that, first of all, each sample each as a minimum, select the distance of two kinds of merged into a new class (if the sample the relationship between the similarity coefficient, should choose similarity coefficient absolute value of the largest two merged into a new class), then calculate the new class and other classes, and the distance between the distance of two kinds of minimum to merge, so continue to, so that every time after the merger are reduce category, until all the samples are divided into a kind of so far.

Fast clustering method of clustering process first choose k observation quantity as clustering initial seeds, they are k initial clustering center; Then the amount of each observation are assigned to the k center distance minimum the class, for the first iteration form k class; Then according to the composition of each kind of observation quantity calculation of variable values, each kind of n mean in n dimensional space and form k point, this is the second iteration class center; According to this method a iterative down, until it reaches the specified iterations or suspend iterative criterion requirement, the clustering process is over[4][5].

III. PROBLEM DESCRIPTION AND DATA PREPARATION

In this paper the data is a normal university at the end of the semester students' online teaching evaluation data. The teaching evaluation indicators were ten, as shown in figure 1 shows. The students online evaluation of teaching is in the semester each student to their computer teacher the indicators of the evaluation of each index is divided into very satisfied (3 points), satisfaction (2 points), not satisfied (1 points) , original data slightly.

Figure 1. Ten teaching evaluation indicators

The figure 2, it is known that the processed data in 74 patients is effective, use this form to the data in the original data standardization, in order to eliminate the dimension is different, the V1, V2... , v10s the corresponding figure 1 ten indicators.

Figure 2. Data briefly statistical information

IV. STATISTICAL PROCESS AND RESULT ANALYSIS

A. *Hierarchical clustering statistics analysis*

In turn, click on "Analyze, Classify and Hierarchical Cluster" executive Hierarchical clustering analysis process, the clustering transform selection Z scores transformation, clustering method choice between group clustering method, clustering statistics choose Euclidean distance.

B. *Clustering process analysis*

As shown in figure 3 shows, "clustering form" the sample is given together for a class of the entire process, "order" means clustering number of steps. First of all, the third indicators and fourth indicators for a class together, and in the second clustering, and seventh indicators for a class together, at last in the fourth, the seventh and eighth clustering separately and first, five and nine indicators gathered for a major categories, other clustering process by analogy, a total of nine steps after ten indicators will gather for a categories.

	聚类表					
阶	群集组合		系数	首次出现阶群集		下一阶
	群集 1	群集 2		群集 1	群集 2	
1	3	4	7.751	0	0	2
2	3	7	12.703	1	0	4
3	5	6	13.406	0	0	6
4	1	3	17.458	0	2	7
5	9	10	18.747	0	0	8
6	5	8	21.200	3	0	7
7	1	5	24.290	4	6	8
8	1	9	25.253	7	5	9
9	1	2	73.632	8	0	0

Figure 3. Hierarchical clustering clustering process

C. *Evaluation index classification*

As shown in figure 4 shows, hierarchical clustering methods ten indexes are divided into four categories, namely, the first category: 1, 3, 4, 7, The second type: 5, 6, 8; Third class: 9, 10; The fourth class: 2;

```
C A S E      0     5    10    15    20    25
Label    Num  +---------+---------+---------+---------+---------+

V3       3    ─┐
V4       4    ─┤ ┌┐
V7       7    ──┘ │┌┐
V1       1    ────┘││
V5       5    ─────┘│
V6       6    ──────┤──────────────────────────┐
V8       8    ──────┘                           │
V9       9    ──────────┐                        │
V10      10   ──────────┘                        │
V2       2    ─────────────────────────────────────┘
```
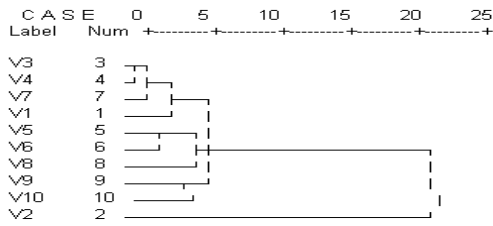
Figure 4.   Hierarchical clustering system tree

V.   FAST CLUSTERING STATISTICS ANALYSIS

A.   SPSS operation process

In turn, click on "Analyze, Classify, K - Means Cluster" executive fast clustering process, in order to and hierarchical clustering number is consistent, in the "clustering number" option which input 4, there will be ten indexes were divided into four types.

B.   Evaluation index classification

As shown in figure 5 shows,it is known that the first kind of various numerical is average, but a, nine, ten value on the high side. The second type of the second value of 5.29, so the final classification results for: the first kind: 1, 9, 10; The second type: 2; The third type of 3, 4, 6, 7, The fourth class: 5, 8;

最终聚类中心			
	聚类		
1	2	3	4
V1 1.98	2.55	2.82	2.54
V2 1.62	5.29	2.71	2.38
V3 1.74	2.45	2.75	2.42
V4 1.87	2.35	2.76	2.39
V5 1.68	2.23	2.65	2.32
V6 1.73	2.35	2.71	2.34
V7 1.79	2.48	2.75	2.42
V8 1.96	2.42	2.66	2.42
V9 1.95	2.48	2.74	2.42
V10 2.02	2.39	2.75	2.42

Figure 5.   Finally the clustering results

C.   Hierarchical clustering and fast clustering results of comparative analysis

From figure 4 and figure 5 clustering results show that the two methods of clustering results are basically the same, explain with clustering method to classify online evaluation norm has certain scientific, it also proves the accuracy of the classification results, but 1, 5, and 6 indicators appear inconsistent classification, specific analysis of the specific content of the three indexes, it is not hard to find these three indicators and other indicators have expressed content was not consistent, this just is to follow the established the basic principles of evaluation index did not appear in the classification, described the index content should not be repeated redundant phenomenon. For a class of second separate, reflected in the process of teaching students motivation evaluation, and no matter use which kinds of method, the classification evaluation index are divided into several categories according to that the index to speak clearly, and other indicators do not repeat such ambiguity. And 3, 4, 7 indicators for a class together, the three embodies the teachers' teaching ability appraisal, and the concrete observation can be found, "inspire thinking" and "guide the student to analyze and solve problem ability" is the same basic meaning, so 3 and 4 indexes appeared repeated phenomenon, in the design can also be reduced. Nine, ten for a class together, 9 item is way of teaching evaluation, and a 10 "acquire knowledge satisfaction" the index, students acquire knowledge at the same time how much by teachers' teaching and students' learning effect two aspects, satisfaction and does not suggest that teachers teach, on the contrary, satisfaction low also cannot explain teachers is not good, so that the evaluation indexes are not reasonable[6].

According to the result of cluster analysis can get teaching quality evaluation content establish basic principles: (1) the principle of independence. The indicators should be independent, the indicators should be incompatible. (2) the principle of simplicity. Each target language statement concise and comprehensive, easy to understand, should not have the double meaning, should make questionnaire simple and clear, easy to operate. (3) the consistency principle. Every specific indicators should reach the goals, the instancing the tenth indicators will not satisfy this rule, the aim is to measure teachers is good or bad, but the actual measuring result also includes students learn good and bad.

VI.   CONCLUSIONS

According to the above hierarchical clustering and fast clustering results contrast analysis, clustering analysis method that will be applied to education, especially in the field of applied to students online commentary is scientific and reasonable, and, in the data statistical analysis on the basis of summing up the establishment of teaching evaluation index on the basic principles of the more convincing, not only improve the quality of teaching, but also to promote teachers' teaching ability to further improve. Clustering analysis method has application in all fields is quite widespread, believe in education applications can also more walk more far, higher and higher.

REFERENCES

[1] Du Jiang. SPSS statistical analysis savvy [M]. People a little press. 2009.03

[2] Mao Yue. Student appraisal of teaching, the result of statistical analysis [D]. 2007.01

[3] Wang fang. Principal component analysis and factor analysis of the similarities and differences compare and application [J] statistical education. 2003 (5) : 14 - 17.

[4] Chen Xuedong. Education practice assessment of multivariate statistical solutions [J].journal of mathematical statistics and management. 2003.22 (6) : 1-4.

[5] Yang Xiaoming editor. The application of SPSS in education statistics [M]. Beijing: higher education press. 2004.278-286.

[6] Pan Xianyun. Teaching in the study of multivariate analysis trial [J]. Journal of guangzhou university (comprehensive edition). 2000.14 (3) : 86-88.