



Received: 07 October 2015
Accepted: 17 May 2016
Published: 22 June 2016

Corresponding author: Sidong Xian,
School of Science, Chongqing University
of Posts and Telecommunications,
Chongqing 400065, P.R. China
E-mail: sidx@163.com

Reviewing editor:
John Lee, Hong Kong Institute of
Education, Hong Kong

Additional information is available at
the end of the article

EDUCATIONAL ASSESSMENT & EVALUATION | RESEARCH ARTICLE

Principal component clustering approach to teaching quality discriminant analysis

Sidong Xian^{1*}, Haibo Xia^{2,3}, Yubo Yin¹, Zhansheng Zhai¹ and Yan Shang²

Abstract: Teaching quality is the lifeline of the higher education. Many universities have made some effective achievement about evaluating the teaching quality. In this paper, we establish the Students' evaluation of teaching (SET) discriminant analysis model and algorithm based on principal component clustering analysis. Additionally, we classify the SET by clustering the result of extracting the indexes through the principal component analysis (PCA), then we also test the rationality of the rating using Fisher's discriminant function. Finally, the model and algorithm are proved to be effective and objective according to the empirical analysis.

Subjects: Education; Information Science; Mathematics & Statistics

Keywords: principal component analysis; clustering analysis; discriminant analysis; students' evaluation of teaching; index system

1. Introduction

The teaching quality is the general trend of higher education nowadays. Correspondingly, how to improve the teaching quality in colleges or universities, is being regarded as an innovation of the instruction system. On the one hand, the college which try to structure the reasonable culture knowledge and skill system is helpful for cultivating the undergraduate with the ability of innovation; on the other hand, the students studying the relevant course could learn much knowledge to satisfy the requirements of personal development. So, the course teaching which is the core link in the higher education teaching can directly affect the accomplishment of the personnel training and teaching goal.



Sidong Xian

ABOUT THE AUTHORS

Sidong Xian is currently a professor with the school of Science, Chongqing University of Posts and Telecommunications, Chongqing, China. He is presiding over the Graduate Teaching Reform Research Program of Chongqing Municipal Education Commission (No. YJG143010) and the Teaching Reform Research Program of Chongqing University of Posts and Telecommunications (No. XJG201328). And he directed his students Haibo Xia, Yubo Yin, Zhansheng Zhai and Yan Shang to study the problem of teaching evaluation in the project. Yubo Yin and Zhansheng Zhai have participated in data collection and processing, Haibo Xia and Yan Shang have analyzed and modeled the data.

PUBLIC INTEREST STATEMENT

The quality of teaching, which affected honor of major colleges and universities in all over the world, was a significant ingredient in personnel training term. This perspective article describes the way of evaluating the teaching quality, based on data gathered via analyzing index system related with the teaching quality, which was used to illustrate impacts of the teaching. It was found that the way is not only the primary part of the teaching quality evaluation, but also a suitable solution to evaluate. Clearly, teaching and studying are the key factor to decide the teaching quality, and the course is the important link of teaching and studying. Evaluating these effects can improve education in colleges. Exploration of the quality of teaching issues can also help universities to enhance their competitiveness and the teaching level.

The course teaching quality evaluation is the vane of the teaching quality evaluation in college. Many colleges generally regard the teaching evaluation as a way to improve the teaching quality. The teaching quality evaluation, first coming into being in Harvard University in 1920s and blooming in the later period in the popularization of the higher education (Li, 2012), has its main principal: the school serve the students and the goal of it is to realize the personal expectation. Importantly, evaluating the teaching must mainly consider whether the teaching can satisfy the students' demand, so the students are suitable to make evaluation. From 1960s, the teaching evaluation is in a wide use in college, middle school, and primary school in many developed countries (Chen, 2001; Fernandez & Mateo, 1992). It has become a unique feature in education evaluation and improved the teaching quality effectively (Fernandez & Mateo, 1992; Luo, 2013; Zhou, Zhang, & Chen, 2013).

In late 1980s, SET began to be introduced in China and applied in some colleges. With the start of the quality project of the higher education, the domestic colleges pay more attention to the monitoring system of teaching quality evaluation by using the SET. And there are five kinds of SET, which are as follows. The first one is the students' teaching evaluation group system represented by Nanjing University; the second is the student messenger system represented by Wuhan University. In addition, the student questionnaire partly checked system represented by Tsinghua University and the all students evaluating the teaching system represented by North China Electric Power University are also part of SET; the last is the student teaching messenger and total students directly evaluating mixed system represented by Chongqing University of Posts and Telecommunications. Additionally, many scholars have studied the SET both qualitatively and quantitatively. For example, Chen and Zhao (2004) propose to study the SET from four aspects relevant to teaching, namely teaching method, content of courses, teaching idea, and teaching style. Chen and Zhou (2008) referenced the existent SET system and using the fuzzy hierarchy analysis to compound the total hierarchy weight, get the weight of each hierarchy of general objective, and put forward the SET model. Because of the SET information's limited, incomplete, and characterized by random uncertainty, Xu (2006) using the grey system theory (Deng, 1982, 1989) and defining the whitenization weight function of the index, gets the SET grey clustering model and rates different courses.

Nowadays, SET is concerned with two questions: one is whether the system of SET is measurable or scientific; the other is that the index system often has the multicollinearity which few models and algorithm can overcome. Si (2006) use the PCA to transfer the multi-index problem to fewer indexes problem. Though it can overcome the multicollinearity of the indexes to some extent, new indexes using PCA directly for evaluating it will cause some situation inconsistent with the real condition. In order to solve this problem, in this paper, PCA and clustering algorithm are employed to establish the SET model based on principal component and clustering, then used Fisher's discriminate function to test the reliability of the evaluation. Consequently, given/based on the scientific criterion for college instruction, the SET system may better display its function in instruction management and quality evaluation.

2. Measurement to the SET and its index system

A scientific SET index system should investigate the teaching quality from multiple perspectives. In order to evaluate the teaching quality accurately and effectively, both of the evaluation system and method are importance. SET is affected by many factors. It not only involves teaching attitude, teaching content, teaching method and the teachers' personal qualities, but also including the studying attitude of the students, school spirit, teaching conditions, social morality, and some relevant policy. It will also be related with the variety of many other factors. In order to fit the evaluation demand in different periods for people, it is very important to set the evaluation index scientifically. After having referenced many SET index systems, according to the principal of scientific, orientation, and operability, we establish the evaluation index system (see Table 1) based on the data of course standards, teachers' performance, and students' experience. We also give some explanation for them as follows. The course standards are instructional documents of teaching in defining the course quality, course goal, content goal, and implementing recommendations of a specific course. In these parts of the basic conception, course goal and implementing recommendations, it

Table 1. The index system of SET

Latent variable	Measurable (variable) index
Course standards	selection of teaching materials of this course (x_1)
	The study materials provided of this course (x_2)
	The arrangement of teaching content of this course (x_3)
	The arrangement of assessment of this course (x_4)
Teachers performance	The extent of the teacher's help in studying this course (x_5)
	The teachers class performance (x_6)
	The teaching method the teacher uses (x_7)
	The teachers help after class (x_8)
Students experience	The class teaching of this course (x_9)
	The difficulty of this class (x_{10})
	The extent of learning in this course (x_{11})

elaborates in detail, especially making some requirements about the basic study for all students. At the same time, defining the course standards and grasping the teaching rules are helpful for improving the teaching quality (Yiannis & Sotierios, 2014). Teaching is a science, also an art.

Using various kinds of teaching methods flexibly, finishing the teaching task students' and getting good teaching results are the basic skills of a teacher. The teachers' teaching performance not only affects the teaching quality, but also has an impact on the students' learning interests deeply. Such good classroom performance is necessary for a good teacher. Student is the object of the teaching work service, and the students' experience during study is the fundamental basis in evaluating a course. So keeping the students' interest can make the evaluation results better, and it can also improve the teaching quality.

3. SET discriminate model based on PCA and clustering

Composing the PCA correlation matrix through the teaching quality index variables, we calculate the score of each PC according to the linear combination of the principal component decided by the PCA. Then we cluster the PC score of each course and establish the discriminate function based on principal component score (Julian & Mauro, 2014). Using the discriminate function classifies the course and verify the results (Ye & Wei, 2009; Zhang, 2007; Chen & Hou, 2007; Ding & He, 2004).

3.1. Architecture of PCA

Principal components analysis(PCA)(Yeung & Ruzzo, 2001) is a well-known technique for extracting the main tendencies of data on reduced dimension space, which is projected space from original high-dimensional data space.

Suppose we are given a set $X = \{x_1, x_2, \dots, x_p\}$ of n samples and its p index variables. We can transfer the original variables to the factor variables which are the linear combination of the original variables. That is to say, x_1, x_2, \dots, x_p are synthesized to $m(0 < m < p)$ variables f_1, f_2, \dots, f_m , and these m uncorrelated variables are called the principal component 1, 2, ..., m , respectively. The expression is as follow:

$$\begin{cases} f_1 = a_{11}x_1 + a_{21}x_2 + \dots + a_{p1}x_p, \\ f_2 = a_{12}x_1 + a_{22}x_2 + \dots + a_{p2}x_p, \\ \dots, \\ f_m = a_{1m}x_1 + a_{2m}x_2 + \dots + a_{pm}x_p \end{cases} \quad (1)$$

In the analysis of the practical problems, we often choose the biggest principal components to decrease the number of the variables, grasp the main contradiction and simplify the relationships between variables. Choosing the principal component is to confirm the load $a_{il}(i = 1, 2, \dots, p, l = 1, 2, \dots, m)$ of the original variable $x_j(j = 1, 2, \dots, p)$ in each principal component.

According to the theory of principal components analysis, the steps of principal components analysis are follow:

Firstly, to calculate the index mean value and standard deviation of each sample, then calculate the correlation coefficient matrix R .

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

where $r_{ij} = \frac{\text{cov}(x_i^*, x_j^*)}{\sqrt{\text{Var}(x_i^*)\text{Var}(x_j^*)}} = E(x_i^* x_j^*)$, $x_{ki}^* = \frac{x_{ki} - \bar{x}_i}{s_i}$, $\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki}$, $s_i = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)^2}$, $(i, j = 1, 2, \dots, p, k = 1, 2, \dots, n)$.

Secondly, extract the principal component of the R . Namely calculate eigenvalue $\lambda_i(i = 1, 2, \dots, p)$ and eigenvector $e_i(i = 1, 2, \dots, p)$ of the correlation coefficient matrix R . And rank the eigenvalue in order. Then we calculate the contribution rate and the cumulative contribution rate as follow.

$$\alpha_i = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}, \quad L(m) = \frac{\sum_{i=1}^m \lambda_i}{\sum_{j=1}^p \lambda_j}. \tag{2}$$

We choose the number of principal component when $L(m) \geq 0.90$.

At last, calculate the load of each variable in principal component and get the expression as (1).

$$l_{ij} = p(f_i, x_j) = \sqrt{\lambda_i} e_{ij} (i, j = 1, 2, \dots, p). \tag{3}$$

3.2. Clustering analysis

Clustering is the method to study the mutual relationship between data logically and physically, and its analysis results can not only reveal the intrinsic relations and differences between data, but also provide the important foundation for the further data analysis and knowledge discovery. The essence of clustering analysis is to classify the data into several clusters according to the distance until the difference between the data in the same cluster is as tiny as possible and the difference between the data in the different clusters is as big as possible. The key of clustering is to cluster the similar things into the same cluster. According to the Euclidean distance of two sets during clustering, the modeling is as follows.

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}, \quad (i, j = 1, 2, \dots, p). \tag{4}$$

Where x_{ik} is the index i in sample k , n is the number of sample. d_{ij} is the distance between sample i and sample j . We combine the clusters of the minimum distance until the number of cluster can meet the requirement.

According to the theory of cluster analysis, the steps of hierarchical clustering algorithm are as follows:

Step 1. Initially classify. Let $k = 0$, and each mode is a cluster, namely $G_i^{(0)} = \bar{x}_i, i = 1, 2, \dots, m$.

Step 2. Calculate the distance between each cluster, and get a symmetrical distance matrix $D^{(k)} = (d_{ij})_{m \times n}$, where m is the number of cluster.

Step 3. Find the minimum elements in Step 2, and let it be the distance between $G_i^{(k)}$ and $G_j^{(k)}$. Then we combine the cluster $G_i^{(k)}$ and $G_j^{(k)}$ to one cluster, and we can get the new cluster $G_1^{(k+1)}, G_2^{(k+1)}, \dots, G_{m-1}^{(k+1)}$.

Step 4. Check the result of cluster whether satisfy the objection, if not, then return to Step 2.

3.3. Fisher's discriminate analysis

Discriminate analysis (Sun & Li, 2014) can judge which cluster the new sample belongs to when studying the classification in accordance. Its basic principal is to determine the undetermined coefficient of the discriminate function through analyzing the mass data of the objects and calculating the recognition indexes. When the mean vector of population is high in multicollinearity, the Fisher's discriminate method is simple, and we can judge through a few discriminate functions. It also doesn't have any special requests about the distribution of the population, so it has a wide use.

Suppose there are k populations G_1, G_2, \dots, G_k , and we extract n_1, n_2, \dots, n_k samples. Let $n = n_1 + n_2 + \dots + n_k$, and $x_\alpha^{(i)} = (x_{\alpha 1}^{(i)}, x_{\alpha 2}^{(i)}, \dots, x_{\alpha p}^{(i)})$ is the observation vector of sample α in population i . We suppose the discriminate function is $y(x) = c_1 x_1 + c_2 x_2 + \dots + c_p x_p = C^T X$, where $\bar{y}^{(i)} = C^T \bar{X}^{(i)}$ is the mean value of sample in G_i and sample variance is $\sigma_i^2 = C^T S^{(i)} C$, where $\bar{X}^{(i)}, S^{(i)}$ is the sample mean vector and sample covariance matrix, respectively. In order to make coefficient vector C to be the biggest, we need to let $\lambda = \frac{\sum_{i=1}^k n_i (\bar{y}^{(i)} - \bar{y})^2}{\sum_{i=1}^k q_i \sigma_i^2}$ be the biggest, where q_i is the positive weighting coefficient set by ourselves, and it is usually chosen as the prior probability. Let $q_i = n_i - 1$, and substitute $\bar{y}^{(i)} = C^T \bar{X}^{(i)}, \sigma_i^2 = C^T S^{(i)} C, \bar{y} = C^T \bar{X}$ to get: $F = \frac{C^T A C}{C^T E C}$, where $E = \sum_{i=1}^k q_i S^{(i)}$ is the dispersion matrix of group, and A is the sample covariance matrix between populations. After these calculating, we can get the model: $\max F = \frac{C^T A C}{C^T E C}$. Let $\frac{\partial F}{\partial C} = 0 \implies AC = \lambda EC$, we can get λ and C is the generalized eigenvalue and the corresponding eigenvector of A, E . Let $\lambda_1, \lambda_2, \dots, \lambda_m (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0)$ to express all the non zero eigenvalue, and C_1, C_2, \dots, C_m is the corresponding eigenvector. Then we can structure m discrimination functions:

$$y_l(x) = C^{(l)} x, \quad l = 1, 2, \dots, m. \tag{5}$$

and we can get y_k belong to the cluster i , if $|y_k - \bar{y}_i| < |y_k - y_q| (k, q = 1, 2, \dots, m, q \neq 1)$.

According to the theory of discriminate analysis, the calculating steps of discriminate analysis are as follow:

Step 1. Use the selected training sample to calculate the cluster mean value $\bar{X}_p^{(k)}$ and the mean difference d_p . Calculate the coefficient matrix by $s_{pj} = \sum_{i=1}^{n_1} (x_{ip}^{(1)} - \bar{X}_p^{(1)})(x_{ij}^{(1)} - \bar{X}_j^{(1)}) + \sum_{i=1}^{n_2} (x_{ip}^{(2)} - \bar{X}_p^{(2)})(x_{ij}^{(2)} - \bar{X}_j^{(2)}) + \sum_{i=1}^{n_3} (x_{ip}^{(3)} - \bar{X}_p^{(3)})(x_{ij}^{(3)} - \bar{X}_j^{(3)})$. Then we can get the discrimination function through:

$$\begin{cases} s_{11}C_1 + s_{12}C_2 + \dots + s_{1p}C_p = d_1, \\ s_{21}C_1 + s_{22}C_2 + \dots + s_{2p}C_p = d_2, \\ \dots, \\ s_{p1}C_1 + s_{p2}C_2 + \dots + s_{pp}C_p = d_p \end{cases} \tag{6}$$

Step 2. Use the data of the training sample to substitute into the discrimination function, calculate the center of gravity of each cluster. According to $y_0 = \frac{n_1 \bar{y}^{(1)} + n_2 \bar{y}^{(2)}}{n_1 + n_2}$, we can calculate the critical point of discriminate y_0 , and then we can judge through the discriminate criterion.

Step 3. Check the result of discrimination whether satisfy the objection, then calculate the misjudgment rate and discriminant accuracy; if not, then return to Step 1.

3.4. A new discriminate model based on principal component clustering and its algorithm

3.4.1. Discriminate model based on principal component and clustering for teaching quality grading

According to the SET index system combined with PCA model, clustering model and discriminate analysis model, we can establish the discriminate model based on principal component clustering (DPCC) for teaching quality grading shown as follow.

$$\begin{cases} d_{ij} = \sqrt{\sum_{i=1}^p (h_{ik} - h_{jk})^2} & (a), \\ d_j = \sum_l^k s_{jl} f_l & (b) \end{cases} \quad (7)$$

where $h = \sum_{i=1}^k \alpha_i f_i$ (k is the number of the principal component), $f_i = \sum_{l=1}^p a_{li} x_l$ ($i = 1, 2, \dots, m$).

3.4.2. An algorithm of the DPCC

Based on the above analysis and model (7), an algorithm of the New Discriminate Model based on Principal Component Clustering (DPCC) was developed as follow.

Procedure DPCC:

Input: A: data of the students' evaluation of teaching(SET).

Output: D: an update classification model and result

1. Initialization phase.

2. While not end of the SET do

2.1. Once a new labeled data SET in, store its elements and class information, increase the size of the class by one which the new data belongs to.

2.2. According to formula 2, we can get the cumulative contribution rate, and we choose the principal component by the rate over 0.90. We can get the score of principal component factor through the load of each index in the principal component, and the principal component expression as formula 1.

2.3. According to the score of principal component factor of each course solved by formula (1), we can cluster them by agglomerative hierarchy clustering method through formula (7a). According to the mean principal component score in each cluster, we can rate each cluster by the size of the score, from high to low. The rating result is also the same.

2.4. According to the results by formula (1) and formula (7a), we can use formula (7b) to get the discriminate function of each cluster, and substitute the data of training sample back into discriminate functions and calculate the center of gravity of each cluster. And we can judge them by discriminate criterion.

3. Endwhile.

4. Return D.

Table 2. Eigenvalue of the correlation matrix

PC	Eigenvalue	Variance contribution	CVCR
f_1	9.979	0.90715	0.90715
f_2	0.522	0.04741	0.95456
f_3	0.198	0.01804	0.97260

Notes: PC = principal component; CVCR = cumulative variance contribution rate.

Table 3. Factor score matrix

Indexes	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
f_1	0.940	0.966	0.990	0.967	0.987	0.965	0.984	0.992	0.982	0.742	0.936
f_2	-0.032	-0.158	0.069	0.019	-0.096	-0.069	-0.006	-0.008	-0.148	0.669	-0.084
f_3	0.089	-0.167	-0.046	0.100	-0.028	-0.193	-0.103	0.034	0.021	-0.013	0.316

3.5. Case study

The authors have recently observed the SET of the Chongqing University of Post and Telecommunications (CQUPT), which has 13 major departments, 47 undergraduate programs, 38 postgraduate programs, and 2 PhD programs. CQUPT, founded in 1950, is one of the few universities that has been blue-printed and given priority to develop by the State. It has also grown into a prominent university and is acknowledged by its excellence in the field of information industry both within and outside China. The university has built many well-established relationships with over 50 universities from other countries such as USA, Canada, UK, Germany, and Korea, etc.

Firstly, we set the questionnaire and obtain the data from all students in fall semester in 2012. Then, we normalize the data every variable, and get the correlation matrix, choose the principal component by formula 2. At last, we compute the value of evaluation of every course use the DPCC by the algorithm of the DPCC, and analyze the synthetical score situation with SPSS and get the eigenvalue of the correlation matrix shown as Table 2 and factor score matrix shown as Table 3.

According to Table 3, we can get

$$\begin{cases} f_1 = 0.940x_1 + 0.966x_2 + 0.967x_3 + \dots + 0.742x_{10} + 0.936x_{11} \\ f_2 = -0.032x_1 - 0.158x_2 + 0.069x_3 + \dots + 0.669x_{10} - 0.084x_{11} \\ f_3 = 0.089x_1 - 0.167x_2 - 0.046x_3 + \dots - 0.013x_{10} + 0.316x_{11} \end{cases} \quad (8)$$

The score of each course is shown in Table 4 solved by the corresponding PC.

Table 4 is the score of PC solved by each PC factor matrix. Now, we can cluster the score of PC of each course, and the clustering dendrograms are shown as Figures 1 and 2.

Table 4. Score of the course

Course	A	B	C	D	E	F	G	H	I	J	K	L	M
Score of one PC	-0.079	-2.8672	5.9698	6.8237	-11.5069	10.2784	12.7567	2.8717	-8.7014	-4.6539	15.1127	-4.9058	-21.0988
Score of three PCs	0.4986	-2.6249	6.2879	6.2039	-10.6665	10.5260	13.2355	2.6509	-9.8082	-5.4677	15.0192	-5.1602	-20.6946

Figure 1. The clustering dendrogram based on 1 PC.

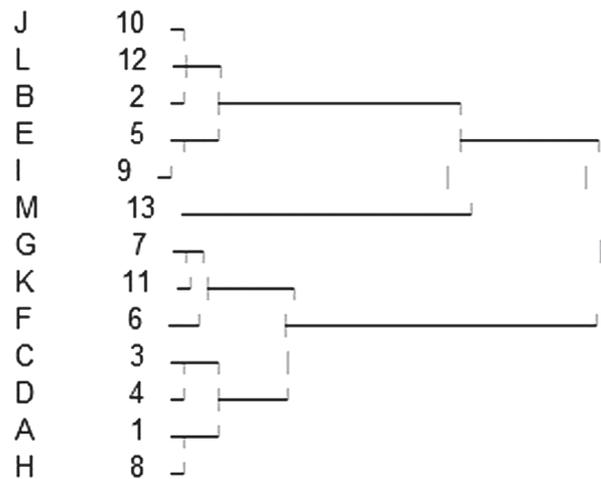
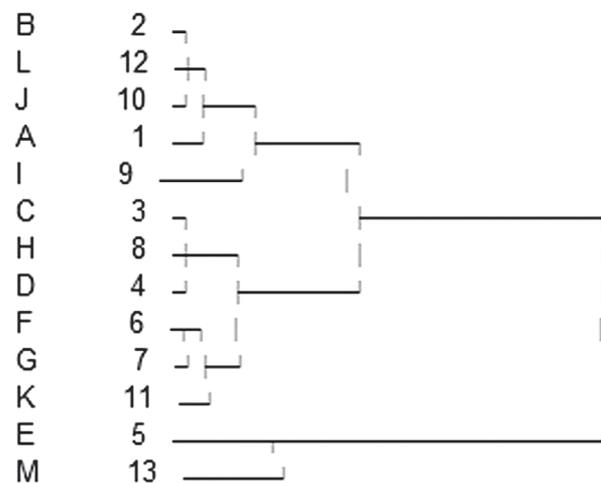


Figure 2. The clustering dendrogram based on original dat.



According to the 2 dendrograms, we can find choosing 1PC or 3 PCs, the results have no difference when we classify these courses into 3 clusters. So, only one PC has reflected the information of the original variables basically. Then we can classify the courses according to the clustering results. On the basis of the dendrogram, we can classify A, C, D, F, G, H, K into a cluster, B, E, I, J, L into a cluster, and M is a cluster.

We can calculate the mean value of each samples score in each cluster, and rank the 3 clusters. According to the score high to low, we rate them “good,” “medium,” and “bad,” expressing by “1, 2, 3,” respectively. The clustering result is shown as Table 5, and we can see there are six in first cluster, six in the second cluster, and one in the third cluster.

According to the score of PC factors of each course and the clustering results in Table 5, we can discriminate the 13 courses and determine the reliability of each clustering analysis result. The coefficient of the Fisher’s discriminate function is shown in Table 6.

So, the Fisher’s discriminate function of the first cluster is $d_1 = 0.343f_1 - 2.415$, the second cluster is $d_2 = -0.292f_1 - 2.050$ and the third cluster is $d_3 = -0.943f_1 - 11.044$. According to the model established before, we can substitute the PC score of each course into the three Fisher’s discriminate functions. we judge the cluster of the sample by the discriminate criterion. If $|y_k - \bar{y}_i| < |y_k - y_q| (k, q = 1, 2, \dots, m, q \neq 1)$, we can get y_k belong to the cluster i . The discriminate results can be seen in Table 7, and there are six in the first cluster, six in the second cluster, and one

Table 5. The classification results after the clustering analysis

Course	RC
A	1
B	2
C	1
D	1
E	2
F	1
G	1
H	1
I	2
J	2
K	1
L	2
M	3

Note: RC=Result of cluster.

Table 6. The coefficient of the Fisher’s discriminate function

Principal component	Cluster		
	1	2	3
f_1	0.343	-0.292	-0.943
Constant	-2.415	-2.050	-11.044

Table 7. The classification result after the Fisher’s discriminate analysis

Cluster		DCR			Total
		1	2	3	
Result of cluster	1	6	1	0	7
	2	0	5	0	5
	3	0	0	1	1
Percentage	1	85.7	14.3	0	100
	2	0	100	0	100
	3	0	0	100	100

Note: DCR = Discriminate classification results.

in the third cluster. In Table 7, the data in the diagonal is the number of the course classified correctly and there are 12 courses are classified correctly. Compared with the clustering results, we can find 0.923 of the classification and rating in clustering analysis result is correct, and, that is to say, the clustering analysis and the discriminate function analysis have comparatively good agreement.

4. Conclusions

Being the primary part of the teaching quality evaluation, SET is a complex problem including multilayer and multi-index. In this paper, according to the new SET index system, we establish discriminate model based on the PCA and clustering. Firstly, PCA is used to simplify the complex data of the instruction quality and reduce the quantities of synthetic indexes to substitute the more primitive variables. Then, we rate the courses rationally through clustering analysis. Lastly, we structure the Fisher’s discriminate

function evaluation model. The model has not only overcome the multicollinearity phenomena in the index system, but also tested the result by comparing the clustering analysis with discriminate analysis. It renders the SET model more objective and pave the way for students teaching quality evaluation.

In the future, we are willing to continue studying on the extension of the case. Such as the more possibility of robust can be obtained with using the different method. We can also apply the DPCC to other universities.

Acknowledgements

The authors express their gratitude to the editor and the anonymous reviewers for their valuable and constructive comments. YJG143010XJG201328CYS15165

Funding

This work was supported by the Graduate Teaching Reform Research Program of Chongqing Municipal Education Commission [grant number YJG143010], the Chongqing research and innovation project of graduate students [grant number CYS15165] and the Teaching Reform Research Program of Chongqing University of Posts and Telecommunications [grant number XJG201328].

Author details

Sidong Xian¹
E-mail: sidx@163.com
ORCID ID: <http://orcid.org/0000-0002-9871-4444>
Haibo Xia^{2,3}
E-mail: 512843667@qq.com
Yubo Yin¹
E-mail: 763675540@qq.com
Zhansheng Zhai¹
E-mail: 961522195@qq.com
Yan Shang²
E-mail: shang.yan@foxmail.com

1 School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, P.R. China.
2 School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, P.R. China.
3 School of Electrical Engineering, Chongqing University, Chongqing 400044, P.R. China.

Citation information

Cite this article as: Principal component clustering approach to teaching quality discriminant analysis, Sidong Xian, Haibo Xia, Yubo Yin, Zhansheng Zhai & Yan Shang, *Cogent Education* (2016), 3: 1194553.

References

- Chen, Q. Y. (2001). Evaluation of classroom teaching of information technology in primary and middle schools. *China Educational Technology*, 3, 28–30.
- Chen, Q. B., & Hou, X. L. (2007). Clustering and principal component analysis of germplasm resources of onion. *Jiangsu Journal of Agricultural Sciences*, 23, 376–378.
- Chen, F. H., & Zhao, X. Y. (2004). The teaching evaluation in college internal. *Modern Information*, 2, 12–15.
- Chen, Y. H., & Zhou, S. M. (2008). The application of fuzzy analytical hierarchy process in the college teaching evaluation. *Pioneering with Science & Technology Monthly*, 8, 115–116.
- Deng, J. L. (1982). Control problems of grey system. *Systems & Control Letters*, 1, 288–294.
- Deng, J. L. (1989). Introduction to grey system theory. *The Journal of Grey System*, 1, 1–24.
- Ding, C., & He, X. (2004). K-means clustering via principal component analysis. In *Proceedings of the 21st International Conference on Machine Learning* (pp. 225–232). Banff.
- Fernandez, J., & Mateo, M. A. (1992). Student evaluation of university teaching quality: Analysis of a questionnaire for a sample of university students in Spain. *Educational & Psychological Measurement*, 52, 675–686.
- Julian, E. W., & Mauro, Z. (2014). Main difference between volatiles of sparkling and base wines accessed through comprehensive two dimensional gas chromatography with time-of-flight mass spectrometric detection and chemometric tools. *Food Chemistry*, 164, 427–437.
- Luo, M. (2013). Application of AHP-DEA-FCE model in college English teaching quality evaluation. *International Journal of Applied Mathematics & Statistics*, 51, 101–108.
- Li, X. F. (2012). *The ideal and reality of students' evaluation: The study based on the fourth generation evaluation theory* [D]. Beijing: Beijing Normal University.
- Si, F. J. (2006). The application of principal component analysis in teaching evaluation. *Intelligence*, 26, 78–79.
- Sun, Z., & Li, J. (2014). Kernel inverse fisher discriminant analysis for face recognition. *Neurocomputing*, 134, 46–52.
- Xu, Q. (2006). The grey cluster model of classroom teaching evaluation. *Journal of Jilin Normal University*, 5, 51–53.
- Ye, X. F., & Wei, Y. W. (2009). Model of the quality evaluation of flue-cured tobacco based on principal component analysis and cluster analysis. *System Science and Comprehensive Studied in Agriculture*, 25, 268–270.
- Yeung, K. Y., & Ruzzo, W.L., (2001). Principal component analysis for clustering gene expression data. *Bioinformatics*, 17, 763–774.
- Yiannis, N., & Sotirios, G. D. (2014). On the student evaluation of university course and faculty member's teaching performance. *European Journal of Operational Research*, 238, 199–207.
- Zhang, H. (2007). The case teaching method of the principal component and clustering analysis. *Statistics and Decision*, 20, 163–164.
- Zhou, L., Zhang, N., & Chen, Q. (2013). Principal component model for comparative evaluation of e-Learning system: An empirical investigation. *International Journal of Applied Mathematics & Statistics*, 42, 330–336.



© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



Cogent Education (ISSN: 2331-186X) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

