2021-1189 IJOI
https://www.ijoi-online.org/

A PERFORMANCE CLASSIFICATION OF UNIVERSITY TEACHERS ON RESEARCH, TEACHING AND SERVICES

Hsiang-Ching Lai*, I-Shou University, Taiwan (ROC)
*Corresponding Author: xiagshein@nuk.edu.tw

Chiao-Pin Bao, I-Shou University, Taiwan (ROC), cpbaao@cloud.isu.edu.tw

Chun-I Chen, I-Shou University, Taiwan (ROC), EddyChen@isu.edu.tw

Abstract

The classification of university teachers' performance is related to the rights and interests of each teacher's future development. Generally, universities allow teachers to perform self-evaluation. However, self-evaluation may show higher tolerance, lower variability, and more errors compared with the evaluation conducted by their supervisors, peers or others. Therefore, the classification of teachers' performance cannot really show the ranking of teachers' performance. The main purpose of this study is to design an effective classification method to assess university teachers' performance. In this study, 101 teachers' performance scores with respect to research, teaching and service in the university were simulated. The scores of each teacher were subtracted from the highest score of each group. The following two steps were performed: (1) grouping was implemented based on the characteristic that the weight of CCR model variable of Data Envelopment Analysis was "0"; (2) 0-1 programming assignment method was used to classify the performance in terms of research, teaching and services for teachers who failed to be grouped in the above steps. Based on the characteristic of CCR mode weight of "0" in the Data Envelopment Analysis, 95 teachers were grouped according to the CCR model after three times of grouping, and the remaining 6 teachers were grouped based on the 0-1 programming assignment method. The contribution of this study is that it replaced the z-score method with the assignment method, which features that data do not need to conform to the normal distribution and provides more functions (such as multiple-principle programming) than the z-score distribution.

Keywords: classification, Data Envelopment Analysis (DEA), assignment method

Introduction

Arreola (2004) suggested that for a diversified teacher evaluation system, it was necessary to emphasize the technology of good evaluation and the establishment of common value standards. According to Barr & Jones (1958), the prediction and measurement of teachers' performance was a process to properly assess the differences between the target characteristics of the subjects. The evaluation of teacher performance is based on teacher performance standards, which is a process of systematically collecting data on teacher performance and making further judgments and decisions (Braskamp & Ory, 1994; Iwanicki, 1990; Sun, 2007). Cashin (1996) indicated that the teacher performance evaluation system must have the following characteristics: (1) the need to define the evaluation system; (2) the interpretation of relevant plans and job responsibilities; (3) the selection of professional performance indicators; (4) the establishment of professional performance standards; and (5) the collection of relevant professional performance documents.

Since 1970, a lot of literature has studied teacher evaluation systems based on diverse principles and the evaluation criteria for university teachers at least involve the aspects of research, teaching, services and even counseling, because nowadays university teachers assume these types of roles and tasks (Johnson et al., 1998; Cui, 2018).

Arreola (2004) established a teach-research-service ternary model for university teachers, studied the parameter values of the three models, and further defined the concept and connotation of teaching, research and services. Therefore, evaluation standards and indicators at home and abroad for the performance of university teachers are based on these three dimensions.

The evaluation of university teachers needs a theory-based and systematic model. Different evaluation standards have different dimensions and measurement methods which were used to calculate different degrees of adaptability and complexity. Most of the literature on the effectiveness and correctness of teacher evaluation focuses on teacher classification (grade) and performance ranking, or emphasizes the weighted research of teaching, research and service. Chen (2005) took teacher promotion as an example to calculate and rank teachers' total performance in research, teaching and services based on multiple criteria with different units of calculation, grey related data preprocessing and VIKOR method, all of which focused on the calculation and ranking of teacher's total performance. Peng (2006) believed that research, teaching and service should be allocated with the equal weight for teacher evaluation.

Sammis et al. (2006) formulated a performance matrix model based on weighted items such as faculty classification, work content, and the time spent on teaching, research, and services. However, the viewpoint proposed by Salomon et al. (2016) that "faculty evaluation must recognize the broad definitions of teaching, scholarship, and service require more flexible and holistic evaluation approaches" has become an emerging consensus and been widely accepted for faculties' performance evaluation. Nonetheless, the academia has not reached a conclusion as how to set out proper scores of the weighted items. The disadvantages of the above methods are illustrated as follows: (1)

only focus on the total score and ranking of teacher performance; and (2) emphasize the proportion of each performance item without scientific basis. Based on Hill et al. (2013), teacher evaluation systems should focus on improving teachers' professional expertise and providing learning opportunities that can foster their competence. Liu et al. (2019) highlighted that faculty performance evaluation should focus on helping them to improve and providing feedback on their professional development.

Previous studies have paid little attention to whether teachers are properly classified into different professional categories. To provide teachers with constructive, personalized and insightful feedback on their professional development, this study proposed an effective classification method for performance evaluation.

The following sections in this paper are organized as follows: Section 2 is the literature review; Section 3 presents the research method; Section 4 discusses the case analysis; and Section 5 illustrates the conclusions.

Literature Review

Classification is used for various purposes in different areas of our daily lives. For example, business, credit forecasting, educational administration, medical diagnosis, ranking and management of various professional operations all involves classification, which provides the basis for researchers or practitioners to conduct the subsequent evaluation, ranking, and decision- making. At present, the classification research focuses on artificial intelligence, machine learning, data mining and other fields. Several typical classification methods are as follows:

As defined by Lance & Williams (1967) and Jain & Dubes (1988), the classification problem is exclusive and non-exclusive, intrinsic and extrinsic, hierarchical and non-hierarchical. Bana e Costa (1992) divided the classification problems into simple nominal classification and sequential classification. Although a lot of literature put forward different classification methods, data mining is the main classification technology:

Bayesian Classifier

Bayesian Classifier is a statistical probability classification method, which predicts the key eigenvalues. Based on the conditional probability in the categories (Duda, 1973), it mainly predicts the probability that an object belongs to a certain category and can construct the classification within a relatively short time (Friedman et al., 1997; Han & Kamber, 2006). Such techniques are significantly limited by the independence among all the sample attributes. Due to the correlation between samples, the accuracy of the classification model will decrease.

Logistic regression

The relationship between the classification dependent variable and one or more independent variables is usually measured with a 0-1 model, based on one or more predictive variables (such as features) to estimate the probability. Features can be classified or continuous (Hilbe, 2009; Han & Moraga, 1995; Khairunnahar et al., 2019), but there exist limits. Khairunnahar et al. (2019) normalized the problems existing in the classification algorithm of logistic re-

gression, but the exact relationship between weight factors and data set size still cannot be effectively obtained in their studies.

Linear Regression

Linear regression has the shortcoming of over-fitting (Bartlett et al., 2020), which means, it is too accurately matched with a specific data set to predict future observation results (Hawkins et al., 2003). In addition, it needs to deal with outliers and cannot predict more complex problems.

Support Vector Machine

It is a binary classification (Hastie, 2009), which has also been employed in the field of machine learning and graphic classification (Osuna et al., 1997). Its main disadvantage is that it is sensitive to missing data values (outliers, missing values), and requires a large amount of repeated calculation (Vapnik et al., 1995, 1997).

Decision tree

It is difficult to control the size of the tree during the decision tree construction. Many researchers have tried to improve it by using various pruning methods to avoid over-fitting, which leads to the construction of the decision tree to be divided into two steps, namely modeling and pruning. It takes much time to build a concise decision tree step by step (Bramer, 2002).

Artificial Neural Network and Genetic Algorithm

Such models address schema classification. One of the disadvantages of such an intelligent model is that the results are often complicated. Both ANN and GA have their own limitations: firstly, their effectiveness depends on the quality and accuracy of the input data sets; secondly, it requires a large amount of calculation work to minimize the over-fitting problem (Singh et al., 2009). One of the major drawbacks of the genetic algorithm is its high computational cost (Dutta, 2020).

Data Envelopment Analysis

Charnes et al. proposed the CCR performance evaluation model i n 1978 and 1979, which emphasized the "hypothesis on constant returns to scale", namely, every increase in investment leads to an increase in output.

The linear programming of the CCR model for input-oriented DEA is as follows (Charnes et al., 1978):

$$\text{Max } h_k = \sum_{r=1}^{s} u_{rk} y_{rk}$$

$$\text{s.t.} \sum_{i=1}^{m} v_i \; x_{ik} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, j = 1,2 \ldots, n$$

$$u_r, v_i \geq \varepsilon > 0, r = 1,2, \ldots, s, i = 1,2 \ldots, m$$

…… (1)

Assignment classification

The basic mathematical model of "0-1 integer linear programming" (Geoffrion & Marsten, 1972) is as follows:

$$\max\ z = \sum_{I=1}^{m}\sum_{j=1}^{n} c_{ij}\, x_{ij}$$

s. t.

$$\sum_{j=1}^{n} C_{ij}\, x_{ij} \le b_i \quad i=1,\ldots\ldots,m$$

$$\sum_{n=1}^{s} C_{ij}\, x_{ij} \le a_j \quad j=1,\ldots\ldots,n$$

$c_{ij} \in \{0,1\}$ *and* $c_j \ge 0$ for all j = 1, … , m … (2)

x_{ij} is the energy of Class *j* assigned to the *i*-th element; *m* is the number of groups classified; *n* is the element of all classifications.

Tools or algorithms have been proposed in the above literature. In fact, each algorithm has its own advantages in certain properties, but they also have limitations, and even have different problems when using other types of data. According to the research content, the following research methods are proposed in this study.

Research Method

Universities do not specify how to implement performance scores for university teachers. Usually, performance scores and the classification methods are decided by the teachers of each university based on their particular specialties. However, in terms of the overall development of the university, it is not desirable to create excessive bias in the grouping of "research", "teaching" and "service". To this end, this study proposes the following implementation process.

1. The R&D Department of the university collects the performance of teachers.
2. The CCR-DEA model (e.g. Eq. 1) is adopted to implement teacher performance classification.

(1) The scores of each teacher are subtracted from the highest scores of each group, and the three scores of each teacher are calculated as follows:
$y_{ir}=x^r - x_{ir}, i = 1, \ldots.. n,$ denotes *n* teachers; $r = A, B, C$ A, B, C are the classification of three kinds of performance; x^r is the highest score in Class *r*; x_{ir} is the *i*-th teacher's score of Class *r*.

(2) According to the three scores of each teacher $y_{ir}, r = A, B, C$ are output variables, and input variables are set as "1" to evaluate CCR efficiency.

(3) In the CCR model analysis, if the weight u_{ir} of one item is "0", the teacher will be assigned to Group *r*, where $r = A, B, C$.

(4) If none of the variable weights are "0", or there are more than two variable weights are "0", the assignment is based on the method for problem assignment.

3. If there are *S* teachers who cannot be grouped, they are classified as *S*×3 matrix, and then 0-1 assignment problem classification is implemented.

Case Analysis

Regarding research, teaching, and services, teachers in colleges and uni-

versities have different grading methods and different weights. The analysis process of the study is as follows:

This study focuses on establishing an analytical model to simulate the scores of "research", "teaching" and "service" of 101 university teachers. The teachers are graded using the CCR-DEA model, and the teacher performance classification is implemented as follows:

The scores of each teacher are subtracted from the highest scores of each group. The results of repeated grouping are shown in Tables 1-3.

Table 1. The first grouping results of the CCR model

No.	Group	No.	Group	No.	Group	No.	Group	No.	Group
1	A	21		41	B	61	B	81	C
2		22		42	C	62		82	C
3		23	B	43		63	C	83	
4		24	B	44	C	64	B	84	
5	C	25	B	45		65	A	85	A
6		26	A	46	B	66		86	C
7	B	27		47		67		87	B
8		28		48	B	68	B	88	B
9	B	29		49	B	69		89	B
10	C	30	B	50	B	70	B	90	
11		31		51		71	B	91	
12		32	C	52		72	B	92	C
13	C	33	B	53	B	73	A	93	
14		34	B	54		74		94	
15		35		55	B	75	C	95	
16	B	36		56		76	B	96	B
17	B	37	C	57		77		97	B
18	B	38		58	B	78		98	
19	B	39		59	B	79	C	99	
20		40	C	60	B	80		100	
								101	B

Note. A, B, and C represent the three categories of research, teaching and services for university teachers.

A total of 101 teachers were grouped by the repeated CCR model with the variance weight of 0, 95 teachers were assigned, and 6 teachers were not assigned. Simply put, if none of the variable weights are "0", or there are more than two variable weights are "0", the assignment is carried out according to the assignment problem". The classification scores of research, teaching and services of the 6 teachers who have not completed grouping are listed as an 6×3 matrix:

The classification results after 0-1 programming assignment method are shown in Table 4.

After simulating the scores of 101 teachers in various specialties, the

Table 2. The second grouping results of the CCR model

No.	Group	No.	Group	No.	Group	No.	Group	No.	Group
1		21		41		61		81	
2		22		42		62	A	82	
3	A	23		43	B	63		83	B
4	B	24		44		64		84	A
5		25		45		65		85	
6	B	26		46		66	A	86	
7		27		47	A	67		87	
8	A	28	B	48		68		88	
9		29	A	49		69	C	89	
10		30		50		70		90	B
11	A	31		51		71		91	B
12		32		52		72		92	
13		33		53		73		93	B
14		34		54		74	B	94	B
15		35		55		75		95	
16		36		56		76		96	
17		37		57	B	77		97	
18		38	B	58		78	B	98	B
19		39		59		79		99	B
20		40		60		80		100	B
								101	

Note. A, B, and C represent the three categories of research, teaching and services for university teachers.

weight of the variables in the repeated CCR model was "0". After grouping in research, teaching and services, 6 teachers failed to be grouped and 13 variables were obtained. They were grouped by 0-1 programming assignment method.

Conclusions

In this study, CCR model and 0-1 programming assignment method were used as the best performance classification tool for university teachers. The main advantage of the CCR model is that it can be used as a basis for the classification of performance according to the merits of each teacher. However, the CCR model often results in the phenomenon that the weight variable is "0", which is a problem for CCR researchers. In this study, the disadvantage of the variable weight being "0" turned into a useful classification method. If the 0-1 assignment method is applied completely to this study, the number of people in each category will have to be assigned by the researcher, making this classification not objective. In addition, if this study only adopted 0-1 programming assignment method for all the classification, 101 variables and 407 constraints will be generated. The huge data will cause troubles in operation. In order not to produce too many variables and constraints, this study first adopted the repeated CCR model, and classified most of the teachers according to the CCR model with the

Table 3. The third grouping results of the CCR model

No.	Group	No.	Group	No.	Group	No.	Group	No.	Group
1		21	B	41		61		81	
2	A	22		42		62		82	
3		23		43		63		83	
4		24		44		64		84	
5		25		45	B	65		85	
6		26		46		66		86	
7		27	B	47		67	C	87	
8		28		48		68		88	
9		29		49		69		89	
10		30		50		70		90	
11		31	A	51	B	71		91	
12		32		52	B	72		92	
13		33		53		73		93	
14	B	34		54		74		94	
15	B	35		55		75		95	B
16		36	B	56		76		96	
17		37		57		77	B	97	
18		38		58		78		98	
19		39	B	59		79		99	
20	B	40		60		80		100	
								101	

Note. A, B, and C represent the three categories of research, teaching and services for university teachers.

Table 4. Six teachers assigned by "0-1 programming"

No.	A	B	C	Group
12	535	195	5	A
22	301	19	99	A
35	15	160	189	C
54	150	56	176	C
56	600	67	96	A
80	56	79	153	C

variable weight of "0". The remaining teachers who could not get the weight of "0" were classified by 0-1 programming assignment method. The reason why this study employed 0-1 programming for problem assignment is that 0-1 programming method provides special integer programming for decision variables (could be captured as 0 or 1) and its calculation process is a special case of linear programming.

The grouping results of case analysis are mainly about teaching, which is in line with the fact that most teachers devote themselves to teaching, and their efforts and outstanding performance are also reflected in teaching.

Based on our previous study on the application of the CCR model and z-score to classification method, it is discovered that z-score will be confined to the normal distribution of data in terms of classification. To correspond to this disadvantage, this study proposed the classification conditions under the CCR model and assignment method without the need to satisfy the normal distribution of data. Moreover, the assignment method can be used for multiple-principle programming function, which enjoys broader application compared with z-score.

References

Arreola, R. A. (2004). Developing a comprehensive faculty evaluation system. Magna Publications.

Bana e Costa, C. A. (1992). Structuration, construction et exploitation d'un modele multicritere d'aide a la decision", PhD Thesis, Universidade Tecnica de Lisboa.

Barr, A. S., & Jones, R. E. (1958). The measurement and prediction of teacher efficiency. *Review of Educational Research, 28*(3), 256-264.

Bartlett, P. L., Long, P. M., Lugosi, G., & Tsigler, A. (2020). Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences, 117*(48), 30063-30070.

Bramer, M. (2002). Using J-pruning to reduce overfitting in classification trees. In: Bramer M., Coenen F., Preece A. (eds). Research and Development in Intelligent Systems XVIII (pp. 25-38). Springer, London.

Braskamp, L. A., & Ory, J. C. (1994). Assessing faculty work: Enhancing individual and institutional performance. San Francisco: Jossey-Bass.

Cashin, W. E. (1996). Developing an effective faculty evaluation system. IDEA Paper No. 33.

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444.

Charnes, A., Cooper, W.W., & Rhodes, E. (1979). Short communication: Measuring the efficiency of decision making units. *European Journal of Operational Research, 3*:339.

Chen, F.T. (2005). A teacher evaluation system for university faculties: A case study of a National University in Taiwan. *Journal of Technology Management, 10*(3), 121-152.

Cui, P., Schellings, G., & Beijaard, D. (2018).Teacher educator' professional learning: A literature review. *Teaching and Teacher Education, 75*(2018), 93-104.

Duda, R.O., & Hart, P.E. (1973). Pattern classification and scene analysis, Wiley New York.

Dutta, D., Sil, J., & Dutta, P. (2020). A bi-phased multi-objective genetic algorithm based classifier. *Expert Systems with Applications, 146*, 113-163.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine learning, 29*(2), 131-163.

Geoffrion, A. M., & Marsten, R. E. (1972). Integer programming algorithms: A framework and state-of-the-art survey. *Management Science, 18*(9), 465-491.

Gillman, R.A., Hensel, N., & Salomon, D.A. (2016). A new paradigm for faculty work and evaluation. Envisioning the Faculty for the Twenty-first Century. NJ: Rutgers University Press, 143-156.

Han, J., & Kamber, M. (2006). Data mining: concepts and techniques, 2nd Ed. San Francisco: Morgan Kaufmann

Han, J., & Moraga, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. In International Workshop on Artificial

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Support vector machines and flexible discriminants. In the elements of statistical learning, 417-458. Springer, New York, NY.

Hawkins, D.M., Basak, S.C., & Mills, D. (2003). Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences, 43*(2), 579-586.

Hilbe, J. M. (2009). Logistic regression models. CRC Press.

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371-384.

Iwanicki, E. F. (1990). Teacher evaluation for school improvement. In, J. Millman & L. Darling–Hammond (Eds.), The new handbook of teacher evaluation, 158-171. Newbury Park, CA: sTage.

Jain, A.K., Dubes, R.C.(1988). Algorithms for clustering data. Prentice Hall, NJ: Englewood Cliffs.

Johnson, S.T., Thompson, S., Wallace, M.B., Hughes, G., & Butty, J.M. (1998). How Teachers and University Faculty Perceive the Need for and Importance of Professional Development in Performance-based Assessment. *Journal of Negro Education, 67*, 197-210.

Khairunnahar, L., Hasib, M. A., Rezanur, R. H. B., Islam, M. R., & Hosain, M. K. (2019). Classification of malignant and benign tissue with logistic regression. *Informatics in Medicine Unlocked, 16*, 100189.

Lance, G.N., & Williams, W.T. (1967). A general theory of classificatory strategies, *The Computer Journal, 9*(4), 373-380.

Liu, Y., Visone, J., Mongillo, M., & Lisi, P. (2019). What matters to teachers if evaluation is meant to help them improve?. *Studies in Educational Evaluation, 61*, 41-54.

Marzano, R. J., & Toth, M. D. (2013). Teacher evaluation that makes a difference: A new model for teacher growth and student achievement. ASCD.

Osuna, E., Freund, R., & Girosit, F. (1997). Training support vector machines: an application to face detection. In Proceedings of IEEE computer society conference on computer vision and pattern recognition, 130-136.

Sammis, T., Mexal, J., & Fisher, J. (2006). Description of a university faculty evaluation system. *NACTA Journal, 50*(1), 44-51.

Singh, V., Khan, M., Khan, S., & Tripathi, C. K. M. (2009). Optimization of actinomycin V production by Streptomyces triostinicus using artificial neural network and genetic algorithm. *Applied Microbiology and Biotechnology, 82*(2), 379-385.

Sun, C. L. (2007). Accountability Control or Professional Development? A Dilemma of Faculty Evaluation in University, *Journal of Educational Practice and Research, 20*(2), 95-128.

Vapnik, V. N. (1995). The nature of statistical learning theory. NY: Springer Verlag

Vapnik, V., Golowich, S.E., & Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. Proceedings of the 9th International Conference on Neural Information Processing Systems, 281-287.