

Fuzzy Clustering Analysis Mathematical Model and Its Application in Teaching Evaluation

Shanshan Gao

Liaoning Institute of Science and Engineering, Jinzhou, 121013, China

40757633@qq.com

Keywords: fuzzy cluster analysis; mathematical model; teaching evaluation; fuzzy sets

Abstract. Fuzzy clustering analysis is based on the characteristics of objective things, degree of affinity, similarity, through the establishment of objective things fuzzy similar relationship clustering method to analysis. In this paper, establish fuzzy clustering analysis mathematical model and conduct a case study. Mathematical model according to fuzzy sets representation, data standardization, fuzzy relationship to establish process, fuzzy relation gives a variety of commonly used methods; section of case analysis first establish teaching evaluation indicator system, then follow the steps for solving mathematical model Complete eight professional teaching assessment. According to cluster analysis, extract key information from large amounts of data for policy makers, provide a strong basis for policy-makers who make the right decisions.

Introduction

Function of the cluster analysis is a set of samples in the feature space clustering between each sample according to points, between the sample point and the sample point set, and similarity measure between sample collection point, get the relationship between sample points and subsets. This approach can be quantitatively determined "affinities" between subjects, achieve the correct reasonable classification purposes. The initial classification is set out the "one or the other" thinking of view, belongs to the same thing and only belongs to the category of a class delineated, the boundaries of this classification is "clear, unequivocal, unambiguous", is part of the ordinary set theory category [1,2]. With the need for the continuous improvement of human knowledge and practical application, ordinary set theory can not completely solve the classification problems with fuzzy nature, result in a fuzzy clustering analysis. Fuzzy clustering analysis introduce the concept of fuzzy mathematics to cluster analysis, a multivariate statistical analysis methods used to study the "like attracts like", and is a branch of numerical taxonomy. This paper studies the mathematical model of fuzzy clustering analysis, and research results apply in the field of teaching assessment, servicing in order to improve the quality of teaching.

Establishment on Mathematical Model of Fuzzy Clustering Analysis

The basic idea of fuzzy clustering analysis: in general relations theory, U contain an equivalence relation R uniquely determined on a classification of U , each the equivalent of an element belong to the same class. The difference between fuzzy clustering and general clustering is that when λ drops from 1 to 0, R_λ doesn't change, thereby form a dynamic clustering. Fuzzy clustering analysis mathematical model accord to fuzzy sets representation, data standardization, process fuzzy relations are established [3,4].

(1) Fuzzy sets representation

Assumption domain X , x is an element of X . For any arbitrary $x \in X$, give mapping as follows:

$$X \rightarrow [0,1] \quad x \mapsto A(x) \in [0,1] \quad (1)$$

Set composed of the following $A = \{(x | A(x))\}$ called "ordered pair", $\forall x \in X$ is fuzzy subset X which contained by X , called fuzzy sets.

Fuzzy sets have the following methods: $A = \{(x, A(x)) | x \in X\}$; $A = \left\{ \frac{A(x)}{x} | x \in X \right\}$;

$A = \int_X \frac{A(x)}{x}$, here doesn't mean integration; when $X = \{x_1, x_2, \dots, x_n\}$ is a finite set, also be expressed as $A = \frac{A(x_1)}{x_1} + \frac{A(x_2)}{x_2} + \dots + \frac{A(x_n)}{x_n}$, "+" isn't the sum here; $A = (A(x_1), A(x_2), \dots, A(x_n))$, vector expression, $A(x) = 0$ can't be omitted.

(2) Data standardization

Data Standardization have two purposes, one eliminate various factors such as differences in the unit, magnitude, and second compress data to the range of $[0,1]$. Data standardization have many ways, this paper combine equalization and normalization method.

The average value for the j -th column data:

$$\overline{x_j} = \frac{\sum_{j=1}^n x_{ij}}{n} \quad (2)$$

The new sequence data equalization processing is:

$$\overline{x_{ij}} = x_{ij} / \overline{x_j} \quad (3)$$

The average of the results of the data are normalized, maximum value by subtracting the rows of data divided by the minimum value in a row of data minus the minimum of the row of data, the formula as follows:

$$x_{ij}^* = \frac{x_{ij} - \min(x_{i1}, x_{i2}, \dots, x_{in})}{\max(x_{i1}, x_{i2}, \dots, x_{in}) - \min(x_{i1}, x_{i2}, \dots, x_{in})} \quad (4)$$

(3) Fuzzy relationship building

Set up fuzzy relationship also called calibration, is determined by the degree of similarity between objects classified coefficient r_{ij} ($i, j = 1, 2, \dots, n$), thereby get the similarity matrix $R = (r_{ij})_{n \times n}$, is fuzzy relationship.

Assumption domain $U = \{A_1, A_2, \dots, A_n\}$, each element is a sample among them, each sample is m -dimensional vector $(x_{i1}, x_{i2}, \dots, x_{im})$, there are various methods of calculating r_{ij} . The most commonly used are the following:

The arithmetic average minimum method:

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\frac{1}{2} \sum_{k=1}^m (x_{ik} + x_{jk})} \quad (5)$$

Number of product method:

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{1}{M} \sum_{k=1}^m (x_{ik} g x_{jk}) & i \neq j \end{cases} \quad M = \max_{i \neq j} \left(\sum_{k=1}^m (x_{ik} g x_{jk}) \right) \quad (6)$$

Geometric average minimum method:

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m \sqrt{(x_{ik} + x_{jk})}} \quad (7)$$

Index similarity coefficient method:

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m \exp \left[-\frac{3}{4} g \frac{(x_{ik} - x_{jk})^2}{s_k^2} \right] \quad (8)$$

In the above formula, $s_k = \frac{1}{n} \sum_{i=1}^n (x_{ik} - \overline{x_{ik}})^2$, $\overline{x_{ik}} = \frac{1}{n} \sum_{i=1}^n x_{ik}$ $k = (1, 2, \dots, m)$

Angle Cosine method:

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} g x_{jk})}{\sqrt{\sum_{k=1}^m x_{ik}^2} g \sqrt{\sum_{k=1}^m x_{jk}^2}} \quad (9)$$

Correlation coefficient:

$$r_{ij} = \frac{\sum_{k=1}^m |x_{ik} - \bar{x}_i| |x_{jk} - \bar{x}_j|}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \quad (10)$$

In the above formula $\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}$, $\bar{x}_j = \frac{1}{m} \sum_{k=1}^m x_{jk}$.

Establishment on Indicator System of Teaching Evaluation

The primary job of teaching quality assessment is that create indicator system of evaluation, then follow the procedure for evaluating mathematical model calculating. Through field research visit, study official relation documents, reference previous research results, combine with the author's many years of management experience actual teaching, evaluation index system construction show in Fig. 1.

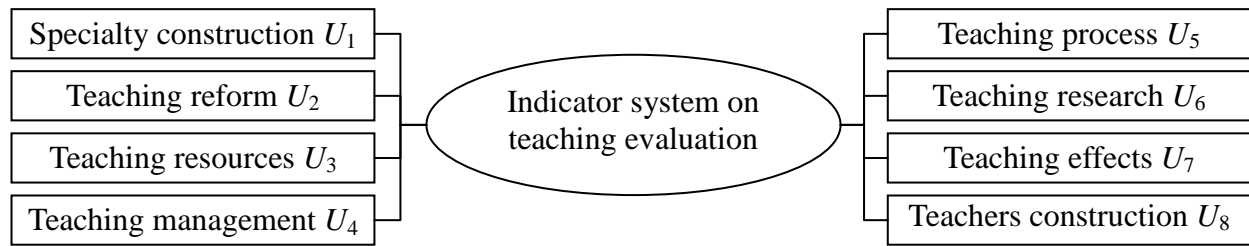


Fig. 1. Indicator system on teaching evaluation

In the index system shown in Fig. 1 includes eight indicators, in order to objectively evaluate the effectiveness, indicator system is usually secondary hierarchy, that a number of secondary indicators at each level indicators. Single indicator fuzzy evaluation contrary to secondary indicators, through the secondary indicators evaluate level indicators [5]. Consider relevant content herein are omitted secondary indicators, two indicators now level indicators should include a brief description, in order to level indicators when single indicator fuzzy evaluation. "Specialty construction" should include disciplines, training objectives, training standards, syllabus secondary indicators and so on; "Teaching reform" should include two indicators of teaching methods reform, reform of teaching methods, teaching procedure reform, teaching reform model and so on; "Teaching resources" should include two indicators of material conditions, natural conditions, social conditions and media conditions and so on; "Teaching management" should include two indicators management team teaching, teaching management system, teaching quality control, teaching management tools and the like and so on; "Teaching process" should include theoretical teaching, practical teaching, classroom rotation, second class and other two indicators and so on; "Teaching research" should include the number of high-level papers published, the number of monographs, attend conferences, two indicators excellent teaching achievement award, project approval and so on; "Teaching effects" should include the employment rate of graduates majoring on qualified rate, curriculum assessment and other two indicators of achievement and so on; "Teachers construction" should include two indicators academic degree structure, title structure, refresher training situation, visit scholars and so on [6].

Example on Fuzzy Clustering Analysis

A college of University of Economics and Management have eight undergraduate majors, they are "Management science and engineering, Business management, Enterprise management, Administrative management, Human resources management, National economic management, Technical economics, Econometrics", use fuzzy clustering analysis of these eight professional evaluation.

Step 1: Single indicator fuzzy evaluation. Using expert scoring method, statistic summary of the indicator results are shown in Table 1.

Table 1. Fuzzy evaluation results on single indicator of teaching evaluation

| Professional name | U_1 | U_2 | U_3 | U_4 | U_5 | U_6 | U_7 | U_8 |
|------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Management science and engineering | 95.31 | 82.15 | 93.35 | 86.52 | 78.36 | 90.84 | 86.84 | 88.90 |
| Business management | 76.42 | 73.84 | 59.43 | 64.34 | 48.61 | 77.52 | 65.57 | 49.46 |
| Enterprise management | 83.57 | 62.89 | 75.30 | 86.10 | 91.34 | 67.74 | 74.14 | 90.86 |
| Administrative management | 71.28 | 65.62 | 80.27 | 54.87 | 72.85 | 41.52 | 83.28 | 91.67 |
| Human resources management | 93.46 | 91.65 | 76.23 | 92.54 | 87.47 | 85.69 | 92.12 | 73.45 |
| National economic management | 67.52 | 82.08 | 45.38 | 73.83 | 91.12 | 66.46 | 38.43 | 81.37 |
| Technical economics | 85.24 | 92.14 | 73.24 | 65.37 | 72.43 | 84.55 | 68.26 | 82.04 |
| Econometrics | 76.17 | 53.36 | 64.19 | 48.64 | 67.64 | 81.31 | 72.37 | 56.33 |

Step 2: Build standardized matrix.

The new sequence data equalization processing is:

$$(\overline{x_{ij}})_{8 \times 8} = \begin{bmatrix} 1.1749 & 1.0885 & 1.3163 & 1.2096 & 1.0279 & 1.2201 & 1.1956 & 1.1582 \\ 0.9421 & 0.9784 & 0.8380 & 0.8995 & 0.6377 & 1.0412 & 0.9028 & 0.6443 \\ 1.0302 & 0.8333 & 1.0618 & 1.2037 & 1.1982 & 0.9099 & 1.0208 & 1.1837 \\ 0.8787 & 0.8695 & 1.1318 & 0.7671 & 0.9557 & 0.5577 & 1.1466 & 1.1942 \\ 1.1521 & 1.2144 & 1.0749 & 1.2937 & 1.1474 & 1.1510 & 1.2683 & 0.9569 \\ 0.8323 & 1.0876 & 0.6399 & 1.0321 & 1.1953 & 0.8927 & 0.5291 & 1.0601 \\ 1.0508 & 1.2209 & 1.0327 & 0.9139 & 0.9502 & 1.1357 & 0.9398 & 1.0688 \\ 0.9390 & 0.7070 & 0.9051 & 0.6800 & 0.8873 & 1.0921 & 0.9964 & 0.7338 \end{bmatrix}$$

The new sequence data normalization process of the latter:

$$(x_{ij}^*)_{8 \times 8} = \begin{bmatrix} 1.0000 & 0.7424 & 1.0000 & 0.8629 & 0.6962 & 1.0000 & 0.9017 & 0.9345 \\ 0.3204 & 0.5281 & 0.2929 & 0.3576 & 0.0000 & 0.7300 & 0.5055 & 0.0000 \\ 0.5776 & 0.2458 & 0.6237 & 0.8533 & 1.0000 & 0.5317 & 0.6652 & 0.9809 \\ 0.1354 & 0.3162 & 0.7273 & 0.1419 & 0.5673 & 0.0000 & 0.8354 & 1.0000 \\ 0.9335 & 0.9873 & 0.6431 & 1.0000 & 0.9095 & 0.8956 & 1.0000 & 0.5684 \\ 0.0000 & 0.7406 & 0.0000 & 0.5738 & 0.9949 & 0.5057 & 0.0000 & 0.7561 \\ 0.6377 & 1.0000 & 0.5807 & 0.3811 & 0.5575 & 0.8725 & 0.5556 & 0.7719 \\ 0.3114 & 0.0000 & 0.3921 & 0.0000 & 0.4453 & 0.8068 & 0.6322 & 0.1628 \end{bmatrix}$$

Step 3: Build fuzzy relationship. The results are as follows:

$$R = \begin{bmatrix} 1.0000 & 0.5540 & 0.8098 & 0.6736 & 0.8872 & 0.6044 & 0.8162 & 0.5563 \\ 0.5540 & 1.0000 & 0.4431 & 0.4310 & 0.5654 & 0.4413 & 0.6759 & 0.6708 \\ 0.8098 & 0.4431 & 1.0000 & 0.7303 & 0.8014 & 0.6799 & 0.7756 & 0.6017 \\ 0.6736 & 0.4310 & 0.7303 & 1.0000 & 0.6018 & 0.4884 & 0.6738 & 0.5461 \\ 0.8872 & 0.5654 & 0.8014 & 0.6018 & 1.0000 & 0.6277 & 0.8363 & 0.5678 \\ 0.6044 & 0.4413 & 0.6799 & 0.4884 & 0.6277 & 1.0000 & 0.6588 & 0.3524 \\ 0.8162 & 0.6759 & 0.7756 & 0.6738 & 0.8363 & 0.6588 & 1.0000 & 0.6596 \\ 0.5563 & 0.6708 & 0.6017 & 0.5461 & 0.5678 & 0.3524 & 0.6596 & 1.0000 \end{bmatrix}$$

Step 4: Cluster Analysis. The results are as follows:

(1) When $\lambda \geq 1$, are divided into eight categories: {Management science and engineering}, {Business management}, {Enterprise management}, {Administrative management}, {Human resources management}, {National economic management}, {Technical economics}, {Econometrics}, is put into every professional category, is the most detailed classification.

(2) When $\lambda \geq 0.8872$, are divided into seven categories: {Management science and engineering}, {Human resources management}, {Business management}, {Enterprise management}, {Administrative management}, {National economic management}, {Technical economics}, {Econometrics}.

(3) When $\lambda \geq 0.8162$, are divided into six categories: {Management science and engineering}, {Human resources management}, {Technical economics}, {Business management}, {Enterprise management}, {Administrative management}, {National economic management}, {Econometrics}.

(4) When $\lambda \geq 0.8098$, are divided into five categories: {Management science and engineering}, {Human resources management}, {Technical economics}, {Enterprise management}, {Business management}, {Administrative management}, {National economic management}, {Econometrics}.

(5) When $\lambda \geq 0.6736$, are divided into four categories: {Management science and engineering}, {Human resources management}, {Technical economics}, {Enterprise management}, {Administrative management}, {Business management}, {National economic management}, {Econometrics}.

(6) When $\lambda \geq 0.6044$, are divided into three categories: {Management science and engineering}, {Human resources management}, {Technical economics}, {Enterprise management}, {Administrative management}, {National economic management}, {Business management}, {Econometrics}.

(7) When $\lambda \geq 0.5563$, are divided into two categories: {Management science and engineering}, {Human resources management}, {Technical economics}, {Enterprise management}, {Administrative management}, {National economic management}, {Econometrics}, {Business management}.

(8) When $\lambda \geq 0.5540$, only include one category: {Management science and engineering}, {Human resources management}, {Technical economics}, {Enterprise management}, {Administrative management}, {National economic management}, {Econometrics}, {Business management}, is the most coarse classification.

Conclusion

In the fuzzy clustering analysis, first calculate the fuzzy similarity matrix, different fuzzy similar matrix will have different classification results; even with the same fuzzy similar matrix, different thresholds will produce different classification results. The Substance of Clustering problem is an optimization problem, through an iterative optimization algorithm makes the objective function system reaches a minimum, but the minimum objective function may exist a lot in the state space, only one of them is the global minimum, the other is a local minimum, iterative optimization goal is that reach global optimum rather than local optimization. In order to find an optimum solution, the algorithm often take a lot of time.

References

- [1] Miin-Shen Yang, Yi-Cheng Tian, "Bias-correction fuzzy clustering algorithms," *Information Sciences*, vol. 309, no. 10, pp. 138-162 2015.
- [2] Esmaeel Dodangeh, Yaping Shao, Maryam Daghestani, "L-Moments and fuzzy cluster analysis of dust storm frequencies in Iran," *Aeolian Research*, vol. 5, no. 8, pp. 91-99, 2012.
- [3] Road passenger Baba, "Method and application of fuzzy clustering analysis," <http://www.doc88.com/p-99094798494.html>, 2015-7-15.
- [4] A. Ramachandra Rao, V.V. Srinivas, "Regionalization of watersheds by fuzzy cluster analysis," *Journal of Hydrology*, vol. 318, no. 1, pp. 57-79, 2006.
- [5] W. X. Liu, "Construction of System Evaluation Index System and Grey Fuzzy Evaluation Model," *Computer Technology and Development*, vol. 23, no. 10, pp. 193-196, 2013.
- [6] N. H. Wang, C. C. Xv, L. Zhang, "Research of University Undergraduate Teaching Evaluation Index System," *Meitan Higher Education*, vol. 31, no. 4, pp. 63-66, 2013.