Research Journal of Applied Sciences, Engineering and Technology 5(20): 4895-4899, 2013
DOI:10.19026/rjaset.5.4338
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: September 27, 2012 Accepted: December 13, 2012 Published: May 15, 2013

Research Article

Fuzzy Clustering Based 41 Universities Innovation Capability Evaluation of Jiangsu Province

[1, 3]Hou Jun, [2]Ding Lingling, [2]Li Qianmu, [1]Liu Kui and [2, 3]Qi Yong
[1]Department of Humanities and Social Sciences,
[2]Department of Computer Science, Nanjing University of Science and Technology,
Nanjing 210094, China
[3]Changshu/Lianyungang/Wuxi Research Institute, Nanjing University of Science and Technology,
Changshu 215513, China

Abstract: Innovation capability evaluation plays a vital role on promoting and ensuring the persistent and in-depth work of scientific research activities in university. Some scholars have proposed a variety of qualitative innovation capability evaluation methods, however, how to separate from subjective judgments and how to describe the vagueness of capability information are still the most important problems. This study proposed fuzzy clustering method and Based on this method 41 universities innovation capability evaluation of Jiangsu province is processed. This study provided a novel means to solving practical innovation capability evaluation problems.

Keywords: Fuzzy clustering, innovation capability, performance evaluation

INTRODUCTION

With the increasing of scientific research input, the utilization of scientific research funding and related resources draws more and more widespread attention of the society. The university innovation capability refers to making a faithful, objective and justice comprehensive evaluation, with qualitative and quantitative analysis of the university scientific research input, output, efficiency, benefits and impact in a specific period using scientific and standardized evaluation methods. At present, many scholars have proposed a variety of qualitative methods on innovation capability assessment such as AHP, PCA and Weighting Statistics, but these above methods can't completely separate from subjective judgments. With the constant enhancement of the complexity and uncertainties of the questions as well as the fuzziness of human thinking, the information of innovation capability becomes vague. To solve this problem, this study builds a fuzzy set to describe the vagueness of assessment information. As a vague concept, when people cannot identify all the factors to inspect, only the selected part of the factors to be evaluated as a representative. Traditional qualitative methods are very difficult to process these compressed data, so a fuzzy clustering method is proposed in this study. With it, assessment information can take advantage of the membership functions which built by the fuzzy sets to quantify and describe the effectiveness of how many may be regarded as a good degree of attribution. Li *et al*. (2012) study a risk assessment method on the implement of risk education. Yin and Tat (2011) study the fuzzy theory and applications. Thomas *et al*. (2010) have a research of an investigation into the factors that prevent secondary schools becoming health promoting. Qi and Li (2007) have a research of the evaluation of university independent innovation ability. Li *et al*. (2007) study the evaluation strategy on the university scientific research innovation competitiveness.

THE INNOVATION CAPABILITY EVALUATION METHOD BASED ON FUZZY CLUSTERING

As shown in, selecting the fuzzy clustering index set which affects the ability in the ability evaluation. The factor set of evaluation objects is three level indicators: $u^1 = \{u_1^1, u_2^1, u_3^1\}$ and seven two indicators $u^2 = \{u_1^2, u_2^2, \cdots, u_7^2\}$.

Cluster analysis is a multivariate analysis method in mathematic statistics; it uses mathematical method to quantitatively determine the Affinity-disaffinity relationships of the samples in order to achieve the objective divided. The fuzzy clustering based on fuzzy equivalence relations is a commonly used means of the innovation capability evaluation.

In the fuzzy clustering, an appropriate grouping must satisfy the following three fuzzy equivalence relations:

Corresponding Author: Li Qianmu, Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, China

- Reflexivity. Any object and itself is similar, the mathematical definition is $r_{ii}=1$
- Symmetry. If two objects O_i and O_j is similar, then O_i and O_j is similar too, the mathematical definition is $r_{ij} = r_{ji}$
- Transitivity. If two objects O_j and O_j is similar and O_j, O_k is also similar, then and is similar, the mathematical definition is $R \bullet R \subseteq R$

So the clustering results obtained by fuzzy clustering analysis based on fuzzy equivalence relations can be more in line with the objective laws.

The method of fuzzy clustering analysis based on fuzzy equivalence relations can be divided into the following several steps:

Step 1: Constructing a data matrix: According to the clustering object, we set n objects that to be clustered as the rows and the m property parameters of each object as the columns to constructing the data matrix X:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

And then, standardizing the matrix X and normalizing the data to the interval [0, 1] to get the matrix G:

$$G = \begin{bmatrix} g_{11} & \cdots & \cdots & g_{1m} \\ g_{21} & \cdots & \cdots & g_{2m} \\ \vdots & & & \vdots \\ g_{n1} & \cdots & \cdots & g_{nm} \end{bmatrix}$$

There are many standardized methods; the commonly used are standard deviation transform and differential transform two:

- **Standard deviation transforms:**

$$(i = 1,2,\cdots n; j = 1,2,\cdots m)$$

$$a_{ij} = (x_{ij} - \overline{X}_j) / \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{X}_j)^2}$$

The condition is. $\overline{X}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, j = 1,2,\cdots m$. After such treatment, the sample data is relatively small and it's not conducive to data analysis, so we should do the linear transformation to the data $g_{ij} = g_{ij} \times 10 + 50$.

- **Differential transforms:**

$$(j = 1,2,\cdots,m)$$

$$g_{ij} = \frac{x_{ij} - \min\limits_{1 \le i \le n}\{x_{ij}\}}{\max\limits_{1 \le i \le n}\{x_{ij}\} - \min\limits_{1 \le i \le n}\{x_{ij}\}}$$

Step 2: Computing fuzzy similarity matrix: The elements in fuzzy similarity matrix R denote the fuzzy similarities between the objects, we consider that the fuzzy similarities between x_i and x_j is $r_{ij} = R(x_i, x_j)$. The methods of computing $r_{ij} = R(x_i, x_j)$ include the scalar product method, the cosine amplitude method, the correlation coefficient method and the maximum and minimum method, Euclidean distance method and so on. In practical applications, Euclidean distance method is used of the most:

Firstly, we need to calculate $d(x_i, x_j) = \sum_{k=1}^{m} |g_{ik} - gik|$ and then $r_{ij} = 1 - c \cdot d(x_i, x_j)$, where the selection of parameter c should make it that $0 \le r_{ij} \le 1$. In this study, we set c as the reciprocal of the max value of all $\sum_{k=1}^{m} | g_{ik} - g_{jk} |$.

Step 3: Analyzing to get the fuzzy clustering: Because the word of 'fuzzy' determines the fuzzy clustering analysis must be conducted with a given fuzzy parameter α, we introduce the concept of R^{α}. The calculation method of the element in R^{α} is as follows:

$$r_{ij}^{\alpha} = \begin{cases} 1 & r_{ij} \ge \alpha \\ 0 & r_{ij} < \alpha \end{cases}, \alpha \in [0,1]$$

According to the definition of the fuzzy similarity matrix R, we can know that R satisfies the reflexivity and symmetry, but not necessarily satisfies transitivity. Therefore, R^{α} which calculated by R also meets the reflexivity and symmetry, but not necessarily satisfies transitivity.

At present there are three main algorithms: the equivalent closure method, the direct clustering method, the maximum spanning tree method, can obtain R^{α} which satisfies the transitivity.

The equivalent closure method is used to calculate the power of R, including $R^2, R^4, \cdots, R^k, R^{2k}, \ldots R^k = R^{2k}$ and R^K is the fuzzy equivalence relation of R, the next step is calculating the matrix. This method is derived from the theory of 'transitive closure and have a strict mathematical foundation. The method would be a good way when the clustering objects are not many. But with the increase of the number of objects to be clustered, due to the need to calculate the matrix of power can yet be regarded as a good method, but as to the increase of the number of object clustering, due to the need for calculation the power of matrix R and its computational complexity along with increased exponentially, so it is not a very effective algorithm.

The maximum spanning tree method which based on the matrix R's corresponding fuzzy diagram uses the

principle of graph theory to get the maximum spanning tree and select a given fuzzy parameter to get the clustering relationship. Although the method is very intuitive, it's not suitable for programming. Therefore the study uses a clustering algorithm with complexity $O(n^2)$ to track the pass sequence of the clustering objects and the algorithm is based on direct clustering.

The main steps are as follows:

- Calculate the upper matrix R^{α}, according to the fuzzy similarity matrix R. Due to the symmetry, it only needs to calculate the elements (the main diagonal elements without calculation) of the lower triangular part in R^{α} and then put the non-zero elements' subscripts in this part into the following standard two-dimensional array record [t][2] where t = n (n−1)/2
- Setting g = 1 and then store g in the array record b[n]; Searching the array record [t] [2] row by row, if an element of a certain line in the array record equals to g and the other element of the row is in the array b, then the element will be sorted in array b
- Marking g times take the elements in the array b, repeating the step 3 until there is no new element need to store into the array b
- Storing the elements of the array b into the two-dimensional array classified [n][n]

Taking a number from 1 to n and make sure that the number doesn't exist in the array $Classified\ [n][n]$, repeating steps 3 to 5 until the total number of the elements in $Classified\ [n][n]$ equals n the number of unspecified object, the elements of the array $Classified[n][n]$ is the clustering results of fuzzy clustering analysis.

FUZZY CLUSTERING BASED 41 UNIVERSITIES INNOVATION CAPABILITY EVALUATION OF JIANGSU

In this section, we get the raw data of the 41 universities in Jiangsu Province as an example. Setting the universities as the rows and the 15 indexes of every university as the columns:

Table 1: Fuzzy similarity matrix

r_{ij}	1	2	…	40	41
1	1	0.363636	…	0.15427	0.410468
2	0.363636	1	…	0.022039	0.297521
3	0.426997	0.393939	…	0.407713	0.752066
4	0.4573	0.38843	…	0.358127	0.732782
5	0.440771	0.201102	…	0.584022	0.763085
6	0.22865	0.110193	…	0.829201	0.5427
7	0.358127	0.162534	…	0.465565	0.528926
8	0.407713	0.203857	…	0.741047	0.69146
9	0.289256	0.137741	…	0.818182	0.603306
10	0.352617	0.173554	…	0.592287	0.575758
11	0.294766	0.07438	…	0.85124	0.559229
12	0.305785	0.110193	…	0.809917	0.570248
13	0.192837	0.068871	…	0.892562	0.586777
14	0.195592	0.090909	…	0.898072	0.517906
15	0.31405	0.190083	…	0.77686	0.556474
16	0.225895	0.07989	…	0.900826	0.509642
17	0.146006	0.008264	…	0.922865	0.429752
18	0.129477	0.07989	…	0.898072	0.498623
19	0.170799	0.049587	…	0.917355	0.487603
20	0.157025	0.016529	…	0.936639	0.446281
21	0.165289	0.024793	…	0.931129	0.454545
22	0.146006	0	…	0.92011	0.429752
23	0.258953	0.15427	…	0.515152	0.539945
24	0.267218	0.134986	…	0.903581	0.573003
25	0.220386	0.093664	…	0.702479	0.53168
26	0.192837	0.049587	…	0.942149	0.476584
27	0.176309	0.063361	…	0.933884	0.493113
28	0.162534	0.432507	…	0.429752	0.479339
29	0.129477	0.052342	…	0.570248	0.501377
30	0.209366	0.112948	…	0.870523	0.548209
31	0.170799	0.033058	…	0.92562	0.454545
32	0.134986	0.022039	…	0.92562	0.567493
33	0.203857	0.057851	…	0.936639	0.490358
34	0. 132231	0.052342	…	0.961433	0.545455
35	0.23416	0.107438	…	0.911846	0.550964
36	0.206612	0.126722	…	0.325069	0.545455
37	0.137741	0.162534	…	0.837466	0.567493
38	0.181818	0.038568	…	0.947658	0.473829
39	0.176309	0.030303	…	0.914601	0.46281
40	0.15427	0.022039	…	1	0.515152
41	0.410468	0.297521	…	0.515152	1

- Constructing the data matrix
- Using the standard deviation transform method to standardize data
- Using distance method to construct fuzzy similarity matrix, as shown as the Table 1:

Obviously, the fuzzy similar matrix in Table 1 meets the reflexivity and symmetry, that is $r_{ii} = 1$ and $r_{ii} = r_{ii}$; Selecting different values of α can get different groupings. The grouping results of this example due to different α are shown as follow:

If α =0.77||0.78||0.79 , the result is {2 //1 //3 //4 //5 //6 8 9 11 12 13 14 15 16 17 18 19 20 21 22 24 25 26 27 30 31 32 33 34 35 37 38 39 40 //7 //10 //23 //28 //29 //36 //41}

If the value of α is 0.77, setting the value equals 1 when r_{ij} >0.77 and the value equals 0 when r_{ij} <0.77 to construct the array R^{α} of fuzzy similar matrix R. After clustering, we get that the innovation capability in 2006 of Nanjing University of Information Engineering, Nanjing Normal University, China Pharmaceutical University, Nanjing University of Posts and Telecommunications, Nanjing Medical University, Nanjing University of Traditional Chinese Medicine, Nanjing Institute of Nanjing Forestry University, Nanjing University of Finance and Economics, Nanjing Institute of Physical Education, Nanjing Institute of the Arts, Nanjing Audit University, Nanjing Xiaozhuang University, Jiangsu Police Officer University, Jinling Institute of Technology, Jiangsu University, Jiangsu University of Science and Technology, Jiangsu Polytechnic University, Changzhou Institute of technology, Jiangsu Teachers University of Science and Technology of Suzhou, Changshu Institute of Technology, Nantong, Yancheng2006 double hit capability Institute of Technology, Yancheng Teachers University, Huaihai Institute of Technology, Xuzhou Normal University, Xuzhou Medical University, Xuzhou Engineering and Huaiyin Teachers University and other universities and universities belong to the same level:

If α =0.86||0.87, then the result is {2 //1 //3 //4 //5 //6 24 27 35 9 11 14 16 17 19 20 21 22 26 30 31 32 33 34 38 39 40 13 18 12 37 //7 //8 //10 //15 //23 //25 //28 //29 //36 //41}

Similarly, if the value of α is 0.87, we get the result that the innovation capability in 2006 of Nanjing University of Information Engineering, China Pharmaceutical University, Nanjing University of Posts and Telecommunications , Nanjing Medical University, Nanjing University of Traditional Chinese Medicine, Nanjing Engineering Institute, Nanjing University of Finance and Economics , Nanjing Institute of Physical Education, Nanjing Arts Institute, Nanjing Audit University, Nanjing Xiaozhuang University, Jiangsu Police Academy Jinling Institute of Technology, Jiangsu University, Jiangsu University of Science and Technology, Changzhou Institute of Technology, Jiangsu Teachers University of Science and Technology of Suzhou , Changshu Institute of Technology, Nantong, Yancheng Institute of Technology, Yancheng Teachers University, Huaihai Institute of Technology, Xuzhou Normal University , Xuzhou Medical hospital, Xuzhou Engineering and Huaiyin Teachers University and other universities and universities belong to the same level:

If α =0.97, then the result is {2 //1 //3 //4 //5 //6 //7 //8 //9 //10 //11 //12 //13 //14 //15 //16 //17 20 21 22 31 38 19 26 33 39 //18 //23 //24 //25 //27 //28 //29 //30 //32 //34 //35 //36 //37 //40 //41 }

If the value of α is 0.97, the innovation capability in 2006 of Nanjing Institute of Physical Education, Nanjing Xiaozhuang University, Jiangsu Police Officer University, Jinling Institute of Technology, Changshu Institute of Technology, Xuzhou Medical University, Nanjing Audit University, Changzhou Institute of Technology, Yancheng Institute of Technology and Xuzhou Engineering belong to a level .

CONCLUSION

In the specific operation, the value of α is selected by the evaluation operators according to the policy guidance and evaluation purposes, also we can analyze the universities and universities on the same level deeply depending on the different evaluation needs. The traditional decision-making theorists believe that the research or application of existing fuzzy theory in the evaluation and analysis emphasize on the value deduced and less explore that whether the fuzzy concept in the evaluators' hearts can really be represented by the fuzzy number, the derivation of it has too many stop-gap (ad hoc) and the lack of a rigorous foundation, so we cannot make sure that the results is still able to reflect the actual behavior. Therefore, the fuzzy clustering can sometimes get the unclear and unreasonable results. This study designs the application method of fuzzy theory in the innovation capability evaluation field to the readers, to provide a means for solving practical problems.

ACKNOWLEDGMENT

The research for this study has been funded by the Jiangsu High Education Scientific Research Project, the National Natural Science Foundation of China (60903027, 6127419), the Jiangsu Natural Science Foundation (BK2011370), the Aerospace Innovation Fund (CALT201102) and the Lianyungang Science & Technology Project (CG1124).

REFERENCES

Li, Q., J. Hou, T. Liu, X. He, K. Liu and Y. Qi, 2012. A risk assessment method on the implement of risk education. Energy Educ. Sci. Tech. Part B, 14: 127-139.

Li, Q.M., Y. Qi and J. Hou, 2007. An evaluation strategy on the university scientific research innovation competitiveness. China Soft Sci., 26: 152-157.

Qi, Y. and Q. Li, 2007. Research on the evaluation of university independent innovation ability. Sci. Res. Manage., 12: 3-8.

Thomas, M., D. Benton, K. Keirle and R. Pearsall, 2010. An investigation into the factors that prevent secondary schools becoming health promoting. Welsh J. Educ., 19: 64-77.

Yin, L. and N.M. Tat, 2011. Fuzzy Theory and Applications. National University of Defense Technology Press, Changsha.