Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (19): 4010-4013, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.4010.4013

Decision Making of College Based on the Fuzzy Clustering and Inclusion Degree

Yu Peng
Library, Chongqing University of Science and Technology, Chongqing, 401331, China

Abstract: About the decision making of college, we adopt fuzzy clustering analysis to construct a model in this study. According to what they like, major, rejoin, marks and so on, they choose a number of colleges. Using model and conditions, we make cluster analysis. For the universities and professionals that we haven't chosen, we calculate inclusion degree, so as to determine which category they belong to.

Key words: Fuzzy clustering analysis, inclusion degree, data mining, university choice

INTRODUCTION

With the improvement of the education system, the number of the university entrance exam also is increasing. From the national statistical data, we can find a interesting phenomenon that "famous universities do not have enough to eat, the local colleges and universities do not have enough to eat". On the other hand, there is a large gap between east universities and west universities. The reason of appearing this kind of circumstance is that students cannot fully understand the university and choose university only with the score as reference (Zhang, 2007). Therefore for university rankings analysis with the light of important significance (Liu, 2004).

Clustering analysis is one of the multivariate analysis and it is an important branch in the unsupervised models. According to some standards, it divides no classes marked samples into several subset. It makes similar samples to a class as much as possible and vice versa. As a kind of unsupervised classification method, clustering analysis has been widely used in pattern recognition, data mining, computer vision, fuzzy logic control and so on. In the most of the real problems, the classifications often have ambiguity and there is no clear boundary between them. In the cluster analysis, it is more reasonable to introduce fuzzy technology (Mitra *et al*., 2002), so that is fuzzy clustering analysis method.

In this study, we use the fuzzy clustering analysis method on the choice of college clustering analysis. (Chen *et al*., 1996). According to the comprehensive evaluation indexes, we divide some of the universities into three categories: Which can be chosen and given up. By calculating inclusion degree, we can further define the students how to choose other universities (Gao, 2004).

SELECTION MODEL ESTABLISHED

The advantage of this selection model is that the users themselves select oneself relative aspects from numerous evaluation standards and generate evaluation criteria. The selection model starts to cluster analysis as select standards of the user and recommend to the user of the results of cluster analysis. The whole process is shown as shown in Fig. 2.

The key of this selection model lies in the cluster analysis. Clustering analysis mass data generally need four steps, as Fig. 1 shows.

Step 1: Data standardization:

- **Establish data matrix:** A domain is the object of classification for $U = \{x_1, x_2,..., x_n\}$ and each object has M index to express its Properties, namely: $x_i = \{x_{i1}, x_{i2},..., x_{im}\}$ $(i = 1, 2,..., n)$, So, we get the original data matrix:

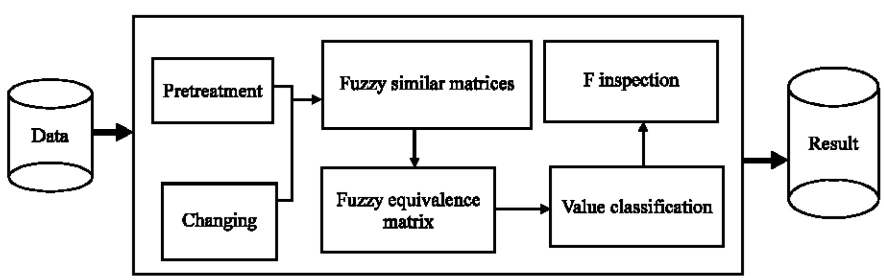

Fig. 1: Fuzzy cluster analysis

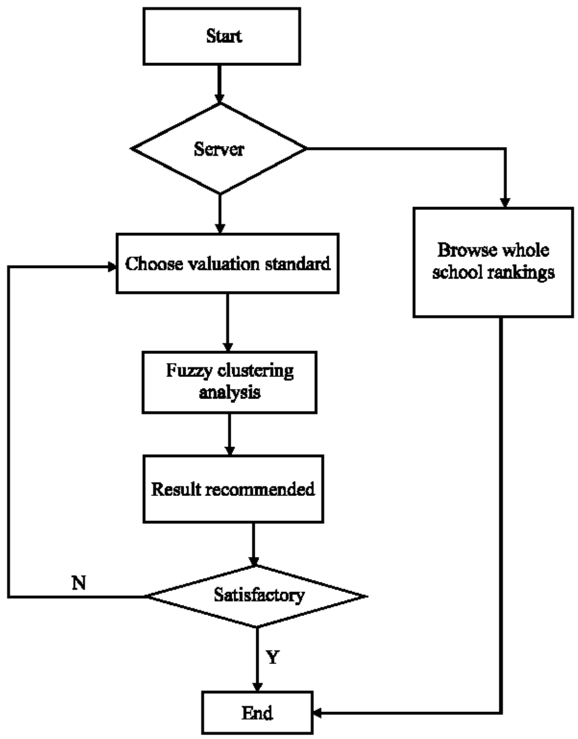

Fig. 2: Select model process

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

- **Data standardization:** In the actual problem, different data generally have a different dimension. In order to compare different dimension, we usually need to make the data transform appropriately. Even so, the data we get may not be in [0, 1]. The methods of data standardization usually have several ways, such as, standard deviation transform, translation transform and logarithmic transformation. In this study, we adopt logarithmic transformation

Step 2: **Establish fuzzy similar matrices:** Establishing similar matrix method, mainly concentrated in the similarity coefficient method, distance method and subjective comments method. This study adopts Maximum and minimum method, get fuzzy similar matrices

Step 3: Adopt the clustering method based on boolean matrix method. The specific procedure is as follows: first of all, Calculate fuzzy similar matrix of λ- cutting matrix and We write it for R_λ. The second step, If R_λ is the equivalent, we can get U in the λ- level of classification. Otherwise, in a certain arrangement, it has the form of special son matrix. At this time, we only need to Put 0 rewrite of 1 at the son of special matrix until it cannot produce the above form of child matrix. And we get R^*_λ for equivalent matrix. So according to R we can get x level classification

CALCULATE INCLUSION DEGREE

We mark college choice system for B getting by fuzzy cluster analysis. So for a given new university x_i, we can get its equivalent class:

$$[x_i]_B = \{x\epsilon U|f_a(x_i), a\epsilon A\}$$

When the objects given are plentiful, $[x_i]_B \neq 0$, $\forall D_j(i\leq j\leq k)$, we can calculate a Inclusion Degree. According to the maximum subjection principle, $D = \max\{d|d = D(D_j/[x_i]_B)\}$, When D reaches maximum, xi belong to D_j class (Fisher, 1987).

In the college choice system, according to the awareness of the university, professional, students' interest and so on We will divide them into three categories: D_1 is category that we can choose, D_2 is category that we may consider and D_3 is category that we should give up. So, to a new object x, $D(D_i/[x]_B)$, (i = 1, 2, 3), x belongs to a classification as follows:

$$x \in \begin{cases} D_1 & D = D(D_1/[x]_n) \\ D_2 & D = D(D_2/[x]_n) \\ D_3 & D = D(D_3/[x]_n) \end{cases}$$

If $D = D(D_1/[x]_B)$, then x belongs to D_1 that is, namely you don't consider x; If $D = D(D_2/[x]_B)$, then x belongs to D_2, that is, you may consider x; If $D = D(D_3/[x]_B)$, then x belongs to D_3 that is, you can choose x.

REASON FOR CLASSIFICATION

Theorem: If $\lambda_1 \geq \lambda_2$, then $|U/R_{\lambda_1}|\geq|U/R_{\lambda_2}|$ among them U is a limited domain and R is a similar relation of any two classification in U.

From the theorem can be seen: U/Rλ is increased along with λ increases. That is if λ is bigger, the object categories are more. When λ = 0, all the objects are only one category and when λ = 1, each object is one actegory. $\lambda\epsilon(\lambda_1, \lambda_h]$, all the objects can be divided into k categories, among them:

$$\lambda_1 = \max\{\lambda\epsilon[0, 1]|g(\lambda) = k-1\}\ (1 \leq k\leq n)$$

$$\lambda_h = \max\{\lambda \in [0, 1] | g(\lambda) = k\}$$

When k = n and k = 1, all the objects can be divided into n categories and one category; if λ_1 and λ_h are not sure, the objects cannot be divided into k categories.

CASE STUDY

Problem: Gang Li, domestic poverty, but also wants to choose to their future development better city, so he selected from all the choice inside evaluation items W (teaching evaluation, school employment rate, the urban development, the admission status). According to the user's chosen standard, extraction of all schools list of data related information, preprocessing. In order to explain, we choose 8 selected item is presented. Formulas of 1, get fuzzy similar matrices as the following matrix shown:

$$R = \begin{bmatrix} 1 & 0.18 & 0.29 & 0.35 & 0.58 & 0.56 & 0.8 & 0.45 \\ 0.18 & 1 & 0.39 & 0.54 & 0.68 & 0.64 & 0.69 & 0.25 \\ 0.29 & 0.39 & 1 & 0.91 & 0.75 & 0.85 & 0.5 & 0.39 \\ 0.35 & 0.54 & 0.91 & 1 & 0.8 & 0.64 & 0.52 & 0.71 \\ 0.58 & 0.68 & 0.75 & 0.8 & 1 & 0.95 & 0.24 & 0.19 \\ 0.56 & 0.64 & 0.85 & 0.64 & 0.95 & 1 & 0.49 & 0.58 \\ 0.8 & 0.69 & 0.5 & 0.52 & 0.24 & 0.49 & 1 & 0.69 \\ 0.45 & 0.25 & 0.39 & 0.71 & 0.19 & 0.58 & 0.69 & 1 \end{bmatrix}$$

Using least-square method for R closure as equations T(R) shown:

$$T(R) = \begin{bmatrix} 1 & 0.5 & 0.48 & 0.61 & 0.51 & 0.7 & 0.48 & 0.25 \\ 0.5 & 1 & 0.48 & 0.5 & 0.54 & 0.65 & 0.56 & 0.45 \\ 0.48 & 0.48 & 1 & 0.7 & 0.6 & 0.5 & 0.54 & 0.62 \\ 0.61 & 0.5 & 0.7 & 1 & 0.85 & 0.61 & 0.56 & 0.51 \\ 0.51 & 0.9 & 0.6 & 0.85 & 1 & 0.7 & 0.62 & 0.5 \\ 0.7 & 0.65 & 0.5 & 0.61 & 0.7 & 1 & 0.5 & 0.62 \\ 0.48 & 0.56 & 0.54 & 0.56 & 0.62 & 0.5 & 1 & 0.7 \\ 0.25 & 0.45 & 0.62 & 0.51 & 0.5 & 0.62 & 0.7 & 1 \end{bmatrix}$$

When $\lambda >= 0.9$, Divided into eight groups: {A}, In this study, we make decision analysis by the fuzzy clustering and inclusion degree. And cite choice example to prove our idea. Through the calculation of inclusion degree, we determine the classification of new object. Through this method, the fuzzy clustering and Inclusion Degree can be widely used in many fields (Hu and Feng, 2009):

- When $\lambda >= 0.9$, Divided into 8 groups: {A}, {B}, {C}, {D}, {E}, {F}, {G}, {H}
- When $0.9 > \lambda >= 0.7$, Divided into 6 groups: {A, F}, {C, D, E}, {B}, {F}, {G}, {H}

Table 1: Data statistics

College	Teaching evaluation (0.3)	Employment (0.3)	City dev (0.2)	Admit (0.2)
A	0.2	0.3	0.1	0.1
B	0.3	0.3	0.1	0.2
C	0.3	0.2	0.1	0.2
D	0.1	0.2	0.2	0.1
E	0.2	0.1	0.2	0.1
F	0.3	0.3	0.1	0.2
-	-	-	-	-
G	0.2	0.1	0.1	0.1

- When $0.7 > \lambda >= 0.5$, Divided into two groups: {A, B, D}, {C, E, D, F, H}
- When $0.5 > \lambda > 0.25$, Divided into one groups: {A, B, C, D, E, F, G, H}

If the evaluation of λ is in (0.5, 0.7]. We put it into two categories: one category is D_1 = {A, B, D} which we can't consider. The other category is D_2 = {A, B, D} which we should choose.

The attribute value is divided into the following several equivalent interval (Table 1):

$$a1: [15, 18) \rightarrow 1, [19, 22) \rightarrow 2, [23, 37) \rightarrow 3$$

$$a2: [1, 1.5) \rightarrow 1, [1.5, 2) \rightarrow 2$$

$$a3: [0.0, 0.5) \rightarrow 1, [0.5, 1) \rightarrow 2$$

Given the new object x7 = (27, 0.6, 0.8), the discrete variable x7 = (3, 2, 2). [x7] A = {x3, x4}.

We get:

$$D(D_1/[x_7]_A) = \frac{2}{2} = 1$$

So, this is the first kind of circumstance: x_7 belong to D_1 that is, university x_7 can be chosen. Then we can get:

$$D(D_2/[x_7]_A) = \frac{0}{2} = 0$$

So, this is the second kind of circumstance: x_7 belong to D_1. This result is consistent with argument. And It is also consistent with the reality (Yue and Yan *et al*., 2010).

CONCLUSION

In this study, we make decision analysis by the fuzzy clustering and Inclusion Degree. And cite choice example to prove our idea. Through the calculation of inclusion degree, we determine the classification of new object. Through this method, the fuzzy clustering and Inclusion Degree can be widely used in many fields.

REFERENCES

Chen, M.S., J. Han and P.S. Yu, 1996. Data mining: An overview from a database perspective. IEEE Trans. Knowledge Data Eng., 8: 866-883.

Fisher, D., 1987. Improving inference through conceptual clustering. Proceedings of the 6th National Conference on Artificial Intelligence, July 13-17, 1987, Seattle, Washington, DC., pp: 461-465.

Gao, X.B., 2004. The fuzzy clustering analysis and application. Xian xian: University of electronic science and technology press.

Hu, D. and K. Feng, 2009. One method of new attr ibute reduction based on discretization o f cont inuous attr ibutes. Appl. Res. Comput., 1: 64-65.

Liu, L., 2004. University rankings for the value analysis of university development. Sci. Technol. Manage., 13: 26-56.

Mitra, S., S.K. Pal and P. Mitra, 2002. Data mining in soft computing framework: A survey. IEEE. Trans. Neural Networks, 13: 3-14.

Yue, H.L. and D.Q. Yan, 2010. New algorithm for discretization based on information entropy. Comput. Sci., 4: 231-233.

Zhang, R., 2007. The discussed of value and problems of Our university rankings. Higher Eng. Edu. Res., 3: 41-42.