CONDITIONAL FUZZY CLUSTERING

FU GUOYAO
Nanjing Gas Turbine Research Institute
Nanjing 210037, China

Abstract: In this paper we shall present a new topic in the domain of the fuzzy clustering—conditional fuzzy clustering, and introduce an algorithm for a simple problem of the conditional fuzzy clustering.

Keywords: Clustering; fuzzy equivalent matrix; fuzzy similarity matrix.

In the fuzzy clustering problems we have presented thus far, the samples are independent each other and are on equality at the time of classifing. But in the practical activities for scientific research and producing there are also such situations that some of the samples are subjected to some restrictions, namely the clustering is under some conditions. For example, a clustering result of $(n-i)$ samples has been obtained and the fuzzy equivalent matrix B_{n-i} of order $(n-i)$ is their relational representation, the clustering result B_{n-i} is very well, it is so satisfied that we intend never to change it , or practically it is unable to change. Now the production has developed, the samples develop $n-i$ into n, we want to make fuzzy clustering for these n samples with B_{n-i} unchanged. That is a problem of conditional fuzzy clustering. Obviously to study conditional fuzzy clustering possesses theoretical and practical significance.

The fuzzy clustering method mentioned in this paper is the clustering based on fuzzy equivalent relations. The clustering is made by the transitive closure B of the fuzzy similarity matrix A, whose element is the correlation coefficient of the samples. Because the process computing the transitive closure consists of a series of unidentical variants, so there may be a big difference between B and A, hence the clustering result obtained by B perhaps does not conform to the reality represented by A. Therefore E. Ruspini, Zhu Jianying and J. Watada presented the idea of Optimal Fuzzy Equivalent Matrix (OEM), i. e. a fuzzy equivalent matrix with the smallest distance from A. In this paper Frobenius norm $\|A\|^2 = \sum_{i,j}^{n} a_{ij}^2$, $A=(a_{ij})_{n\times n}$ is accepted for distance. According to the idea of OEM the fuzzy clustering is to find a fuzzy equivalent matrix $B=(b_{ij})$ from $A=(a_{ij})$, so that

$$\begin{cases} \text{minimize } f=\sum_{i,j=1}^{n}(b_{ij}-a_{ij})^2 & (1) \\ \text{s. t.} \quad b_{ij}\geqslant \min(b_{ik},b_{jk})=b_{ik}\wedge b_{jk} \quad (i<j,i,j,k=1,2,\cdots,n) & (2) \end{cases}$$

In the subsequent text we will introduce an algorithm for the simplest situation $i=1$ of the foregoing conditional fuzzy clustering problem.

Suppose $N=\{1,2,\cdots,n\}$, S_n denotes the set of all $n\times n$ fuzzy similarity matrices, E_n the set of all $n\times n$ fuzzy equivalent matrices, $B_{n-1}\in E_{n-1}$ is a satisfied clustering relations of $n-1$ samples, the $n-$th sample is a increased sample, a_{1n}, $a_{2n},\cdots,a_{n-1,n}$ are correlation coefficients of the new sample with the original $n-1$ samples. Consequently, the similarity coefficients matrix of these n samples is

$$A=\begin{bmatrix} & & & a_{1n} \\ & \mathcal{B}_{n-1} & & a_{2n} \\ & & & \vdots \\ a_{1n} & a_{2n} & \cdots\ a_{n-1n} & 1 \end{bmatrix}$$

Hence the problem is to find a fuzzy equivalent matrix B

$$B=\begin{bmatrix} & & b_{1n} \\ & \mathcal{B}_{n-1} & b_{2n} \\ & & \vdots \\ b_{1n} & b_{2n}\ \cdots & 1 \end{bmatrix}$$

so that $\| B-A \| =\min\limits_{B'\in\Omega} \| B'-A \|$,

here $\Omega=\{B\,|\,B\in E_n,\text{and its left upper submatrix of order } (n-1) \text{ is } B_{n-1}\}$

Proposition 1. If $A=(a_{ij})\in E_n$, for $k\in N$ and $x\in[0,1]$ change a_{si}, the elements of the $s-$th row of A, and a_{is}, the elements of the $s-$th column of A into

$$a'_{si}=a'_{is}=a_{ki}\wedge x,\ 1\leqslant i\leqslant n,\ i\neq s$$
$$a'_{ss}=1$$

but keep the rest $a'_{ij}=a_{ij}$, then $A'=(a'_{ij})\in E_n$.

Proof. See[1].

Proposition 2. If $A=(a_{ij})\in E_n$, then for any column of A, $a_s=(a_{1s},a_{2s},\cdots,a_{ns})^T$ there must exist $x\in[0,1]$ and $k\neq s$ $(k\in N)$, such that

$$a_{is}=a_{ik}\wedge x,\ 1\leqslant i\leqslant n,\ i\neq s.$$

Proof. See[1].

Proposition 3. The function $g(x)=\sum_{i=1}^{n}(c_i \wedge x-d_i)^2$ has the minimal value g_0 on $[0,1]$, here the constants c_i, $d_i \in [0,1]$.
Proof. Let $0 \leqslant c_1 \leqslant c_2 \leqslant \cdots \leqslant c_n \leqslant 1$, calculate the minimal value g_k of $g(x)$, for $x \in [0,c_1],[c_i, c_{i+1}](i=1,2,\cdots,n-1)$ and $[c_n,1]$ respectively, then choose $g_0=\min_k g_k$.

According to the above propositions obtain an algorithm for computing B as follows:
※ For $x \in [0,1]$ compute the minimum value m_k, and the optimal point x_k of the function $g(x)=\sum_{i=1}^{n-1}(b_{ik} \wedge x-a_{in})^2, k=1,2,\cdots,n-1$.
※ Comparing m_k, obtain $m_s=\min_k m_k$ and x_s, define

$$b_{in}=b_{ni}=b_{is} \wedge x_s \qquad (1 \leqslant i \leqslant n-1)$$
$$b_{nn}=1$$

then

$$B=\begin{pmatrix} & & & b_{1n}\\ & \mathcal{B}_{n-1} & & b_{2n}\\ & & & \vdots \\ b_{1n} & b_{2n} & \cdots & 1\end{pmatrix}$$

is the result.

References
[1] Guoyao Fu, An algorithm for computing the transitive closure of a fuzzy similarity matrix, Fuzzy Sets and Systems 51(3) (1992), 189—194.
[2] J. Watada, H. Janaka and K. Asai, A heuristic method of hierarchical clustering for fuzzy intransitive relations. In: (R. R. Yager eds.) Fuzzy Sets and Possibility Theory, Pergamon Press, New York (1982) 148—166.
[3] Jianying Zhu, Some notable key problems in the applications of fuzzy clustering method, Fuzzy Systems and Mathematics 1(1987), 104—111.