WK - CLUSTERING ALGORITHMS

LI SANG HO

Jilin teacher's institute of craft, Changchun, China

Li YENG SUN

Jilin university, Changchun, China

This paper presents the wk-clustering algorithms. It is reputed that this will be a simpler and more convenient method for hierarchical cluster analysis and fuzzy cluster analysis.and

1. Introduction

There are many hierarchical clustering methods to use at present, but when there are many samples sizes, then there is a great amount of computain. For surmounting these difficulties, we advance the wk-clustering algorithms.

By a proximity relation are mean a reflexive, symmetrical fuzzy relation By a similarity relation are mean a reflexive, symmetrical, max-min transitive fuzzy relation.

As everyone knows, a proximity relation on finite universes can be represented as matrix R:

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ \vdots & & & \vdots \\ r_{1n} & r_{2n} & \cdots & 1 \end{bmatrix} \qquad (1.1)$$

Write $R^{*}=RUR^{2}U\ldots UR^{m}U\ldots \triangleq \begin{bmatrix} 1 & r^{*}_{12} & \cdots & r^{*}_{1n} \\ \vdots & & & \vdots \\ r^{*}_{1n} & r^{*}_{m} & \cdots & 1 \end{bmatrix}$

then, R* is the transitive closure of R.

As everyone knows proximity coefficients betwoen samples in the cluster analysis can be represented as a matrix (1.1), and the distances between samples can be represented as a matrix:

$$T = \begin{bmatrix} 0 & t_{12} & \cdots & t_{1n} \\ \vdots & & & \vdots \\ t_{1n} & t_{2n} & \cdots & 0 \end{bmatrix} \qquad (1.2)$$

2. Definition

Definition 2.1 let $\lambda_i = \bigvee_{\substack{j=1 \\ j\neq n}}^{n} (r_{ij})$, $\lambda_i^* = \bigvee_{\substack{j=1 \\ j\neq i}}^{n} (r_{ij}^*)$, and $\lambda_i = r_{ik}$. write

$B_i(\lambda_i) = \{k, i\}$. $C = \{B_i(\lambda_i)\}$.

Definition 2.2 let

$$G = \{ i \mid B_i(\lambda_i) \in C \}$$

then $H \subset G$ is called a subteam of R iff:

1) For all $h \in H \subset G$, there are $g \in H \setminus \{h\}$ and $l \in B_h$, such that $l \in B_g$

2) $\left(\bigcup_{h \in H} B_h \right) \bigcap \left(\bigcup_{\alpha \in G \setminus H} B \right) = \emptyset$.

In particular if H is set of simple point and iff 2) then H is called a subteam of R.

write $A_H = \bigcup_{h \in H} B_h$, then A_H is called an associated class of H.

Definition 2.3. let $\mathcal{A}$ denote the family of all associated classes of the subteam of R, and write $e = \mathrm{card}(\mathcal{A})$, then, we call the matrices

$$X_{H_\alpha} = [r_{ij}]$$

the associated submatrices of R, where $i, j \in A_{H_\alpha}$, $\alpha = 1, \ldots, e$

Definition 2.4 let $A_{ii} = X_{H_I}$, then A_{ij} denotes the submatrices which is made of the intersections between rows of A_{ii} and columns of A_{jj}. and d_{ij} denotes the greatest element of A_{ij}. We call the matrix $D = [d_{ij}]$ the deputy matrix of R.

Example 2.1 let

$$\begin{bmatrix} 1 & 0.9 & 0.7 & 0.1 & 0.2 & 0.4 \\ & 1 & 0.6 & 0.3 & 0.4 & 0.5 \\ & & 1 & 0.5 & 0.2 & 0.6 \\ & & & 1 & 0.8 & 0.3 \\ & & & & 1 & 0.4 \\ & & & & & 1 \end{bmatrix}$$

then $\lambda_1=0.9$, $\lambda_2=0.9$, $\lambda_3=0.7$, $\lambda_4=0.8$, $\lambda_5=0.8$, $\lambda_6=0.6$, $B_1(0.9)=\{1, 2\}$, $B_2(0.9)=\{2,1\}$, $B_3(0.7)=\{3.1\}$, $B_4(0.8)=\{4, 5\}$, $B_5(0.8)=\{5, 4\}$ $B_6(0.6)=\{6, 3\}$; $C=\{\{1, 2\},\{3, 1\},\{4, 5\},\{6, 3\}\}$, $G=\{1, 3, 4, 6\}$ $H_1=\{1, 3, 6\}$, $H_2=\{4\}$, $A_{H_1}=\{1, 2, 3, 6\}$, $A_{H_2}=\{4, 5\}$

$$A_{11} = \begin{bmatrix} 1 & 0.9 & 0.7 & 0.4 \\ & 1 & 0.6 & 0.5 \\ & & 1 & 0.6 \\ & & & 1 \end{bmatrix} \quad A_{22} = \begin{bmatrix} 1 & 0.8 \\ & 1 \end{bmatrix} \quad A_{12} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \\ 0.5 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 0.5 \\ & 1 \end{bmatrix}$$

Definition 2.5 Given $R=\{r_{ij}\}$, then we obtain that G and

$$\lambda g_1 , \ldots , \lambda g_p \quad (g_p \in G) \qquad (2.1)$$

and we call the (2.1) the fundamental elements of R. If $p<n-1$, then, from the deputy matrix of R, i.e., the matrix D, we can obtain the fundamental elements of D. And so on and so fortb, we always can seek n-p-1 elements of R, and arrange these in order from large to small, and denote as:

$$k_1, k_2, \ldots , k_{n-p-1} \qquad (2.2)$$

we call the

$$1, k_1, k_2, \ldots , k_{n-p-1} \qquad (2.3)$$

the natural permutation of the fundamental elements of R.

3. Theorem

We can obtain the follwing theorem:

Theorem 3.1 Given $R=\{r_{ij}\}_{n\times n}$, then

1) n-p=e,

2) Card (A_H)=Card(H)+1,

3) $\sum_{i=1}^{c}$ Card (A_{H_i})=n.

Theorem 3.2 The trasitive closure of fuzzy proximity matrices R makes up fundamental elements of R.

Theorem 3.3 The result of fuzzy clustering only relates to the ordering of the fudamental elements of R.

Theorem 3.4 Any levels of the hierarchical clustering correspond to a fundamental element of R, and the natural permutation without repetition of the fundamental elements of R determines a class of the hierarchical clustering under the level

Up to now, we have only discussed the condition on the (1.1), but shall can be discussing dually the condition on the (1.2). This time, if the diagonal elements "1", "large" and the mulliplying rule of the matrices, i.e., $r_{ij}^{(2)} = \sup_k \inf \{ r_{ik}, r_{kj} \}$

change dually into "0", "Small" and

$$t_{ij}^{(2)} = \inf_k \sup \{ t_{ik}, t_{kj} \}$$

then, can obtain similar results.

4. WK - Clustering algorithms

Now, we can present the wk-clustering algorithms, and its steps are as fo follows:

Step 1. Calculate the proximity coefficients or the distances between samples.

Step 2. Select the natural permutation of the fundamental elements of the proximity matrix or the distance matrix (not certainly fuzzy).

Step 3. According to the natural permutation of the fundamental elements directly obtain a cluster graph, and classify.

Example 3.1 let

$$R = \begin{pmatrix} 1 & 0.1 & 0.9 & 0.3 & 0.8 & 0.1 & 0.2 & 0.1 \\ & 1 & 0.1 & 0.3 & 0.1 & 0.1 & 0.1 & 0.1 \\ & & 1 & 0.5 & 0.6 & 0.2 & 0.2 & 0.1 \\ & & & 1 & 0.5 & 0.1 & 0.1 & 0.1 \\ & & & & 1 & 0.2 & 0.2 & 0.1 \\ & & & & & 1 & 0.7 & 0.3 \\ & & & & & & 1 & 0.4 \\ & & & & & & & 1 \end{pmatrix}$$

then $B_1(0.9)=\{1.3\}$, $B_2(0;3)=\{2.4\}$, $B_3(0.9)=\{3.1\}$ $B_4(0.5)=\{4, 5\}$

$B_5(0.8)=\{5, 1\}$ $B_6(0.7)=\{6, 7\}$ $B_7(0.7)=\{7,6\}$ $B_8(0.4)=\{8,7\}$

$C=\{\{1, 3\}, \{2, 4\}, \{4, 5\}, \{5.1\}, \{6.7\}, \{8, 7\}\}$

$G=\{1, 2, 4, 5, 6, 8\}$, $H_1=\{1, 2, 4, 5\}$, $H_2=\{6, 8\}$

$A_{H_1}=\{1, 2, 3, 4, 5\}$, $A_{H_2}=\{6, 7, 8\}$

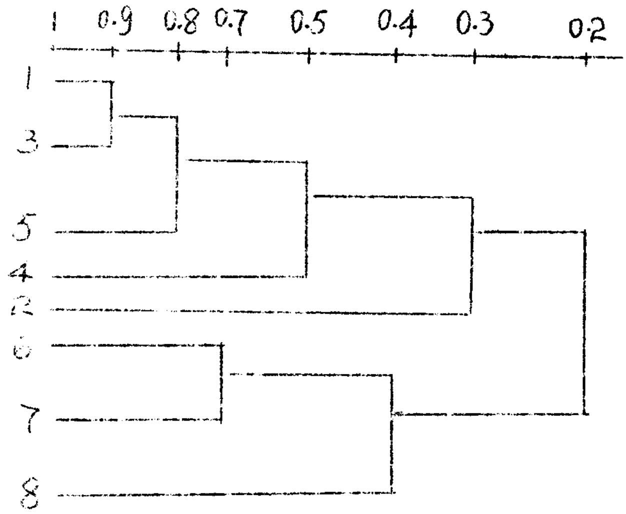
$$X_{H_1} = A_{11} = \begin{pmatrix} 1 & 0.1 & 0.9 & 0.3 & 0.8 \\ & 1 & 0.1 & 0.3 & 0.1 \\ & & 1. & 0.5 & 0.6 \\ & & & 1 & 0.5 \\ & & & & 1 \end{pmatrix} \qquad X_{H_2} = A_{22} = \begin{pmatrix} 1 & 0.7 & 0.3 \\ & 1. & 0.4 \\ & & 1 \end{pmatrix}$$

$$A_{12} = \begin{pmatrix} 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0.2 \\ & 1 \end{pmatrix}$$

The natural permutation of fundamental elements of R:

Value	0.9	0.8	0.7	0.5	0.4	0.3	0.2
You index	1	5	6	4	8	2	1
column index	3	1	7	5	7	4	7

Obtain a cluster graph are as follows:



References

1. Zadeh, L.A. Similarity relations and Fuzzy orderings, Information sciences, 3, 177-200 (1971)
2. Wang Pei-Zhuang, Introduction of fuzzy mathematics, Practice and knowledge on Mathematics, 2, 3, 1980(China)
3. Feng Kaitai and Pan Enpei, Clustering analysis, Geological publishing House, Beijing, 1982.
4. C.V.Negoita and D,A,Ralescu, Applications of fuzzy sets to systems analysis, 169-179
New York: John Wiley, 1975
5. Bezdek, J.C. Cluster Validity with fuzzy sets, J. cybern, 3, No.3 58-73, 1974
6. Li Seng Ho, A new algorithms for computing the transitive closure of fuzzy proximity relation matrices, BUSEFAL, 16, 44-53, 1983.