日本オペレーションズ・リサーチ学会
2004年春季研究発表会

Fuzzy Metric Clustering and Dynamic Programming

01003460 城西大学 岩村覚三 IWAMURA Kakuzo
東京電機大学 堀口正之 HORIGUCHI Masayuki
01101550 千葉大学 * 蔵野正美 KURANO Masami

Given a set N, of n entries, and the metric d_{ij} of each $i, j \in N$, we wish to partition this set into a number of subsets(called clusters) such that it minimized the maximum of the within-group distance([1]). We often encounter the case where each metric d_{ij} is not known precisely and perfectly. In order to deal with this vagueness, we formulate the fuzzy metric clustering problem referring to [3] and consider the applicability of Dynamic Programming Algorithm to our problem.

Definition.

Let R_+ denote the set of all non-negative real numbers. A monotone fuzzy number on R_+ is a fuzzy set $\tilde{a} : R_+ \to [0,1]$ satisfying the following properties:

(i) $\tilde{a}(t_1) \leqq \tilde{a}(t_2)$ for $t_1, t_2 \in R_+$ with $t_1 \leqq t_2$,

(ii) $\tilde{a}(t)$ is right-continuous and has at most finite discontinuous points,

(iii) there exists $t \in R_+$ with $\tilde{a}(t) = 1$.

Let $\mathcal{F}_+$ denote the set of all monotone fuzzy numbers on R_+. A partial order, called fuzzy max order, on $\mathcal{F}_+$ is defined as $\tilde{a} \preccurlyeq \tilde{b}$ $(\tilde{a}, \tilde{b} \in \mathcal{F}_+)$ if $\tilde{a}(t) \geqq \tilde{b}(t)$ for all $t \in R_+$. Then, it clearly holds that for $\tilde{a}, \tilde{b} \in \mathcal{F}_+, \tilde{a} \preccurlyeq \tilde{b}$ is equivalent to $\tilde{a}^{-1}(\alpha) \leqq \tilde{b}^{-1}(\alpha)$ for all $\alpha \in [0,1]$.

Here, we define the fuzzy maximum of $\{\tilde{a}_i : i = 1, 2, \ldots, k\} \subset \mathcal{F}_+$ by the following:

$$\widetilde{\max}\{\tilde{a}_i : 1 \leqq i \leqq k\}(y) = \sup_{\substack{y=\max\{x_1, x_2, \ldots, x_k\} \\ x_i \in R_+ (1 \leqq i \leqq k)}} \min_{1 \leqq i \leqq k} \tilde{a}_i(x_i) \quad (y \in R_+). \tag{1}$$

It is well-known that $a_i \preccurlyeq a_1$ $(2 \leqq i \leqq k)$ if and only if $a_1 = \widetilde{\max}\{\tilde{a}_i : 1 \leqq i \leqq k\}$.

Let $N = \{1, 2, \ldots, n\}$. The map $\tilde{d} : N \times N \to \mathcal{F}_+$ is called a fuzzy metric function if the following conditions (i)–(iii) are fulfilled:

(i) $\tilde{d}(i,j) = 0$ *iff* $i = j$,

(ii) $\tilde{d}(i,j) = \tilde{d}(j,i)$ *for* $i, j \in N$,

(iii) $\tilde{d}(i,j) \preccurlyeq \tilde{d}(i,k) + \tilde{d}(k,j)$ *for* $i, j, k \in N$.

Usually, for each element $i \in N$, there is a multi-dimensional fuzzy data, from which the fuzzy metric $\tilde{d}(i,j)$ $(i, j \in N)$ will be constructed.

Given a subset $A \subset N$, the fuzzy maximum of the fuzzy metric $\tilde{d}(i,j)$ $(i, j \in A)$ is

$$\tilde{d}(A) = \widetilde{\max}\left\{\tilde{d}(i,j) : i, j \in A\right\}.$$

For any fixed m with $1 < m < n$, $J = (J_1, J_2, \ldots, J_m)$ is called m-partition of N if each J_i is non-empty cluster of N, $J_i \cap J_j = \emptyset$ $(i \neq j)$ and $\cup_{i=1}^m J_i = N$. Let $\mathcal{J}_m$ be the set of all m-partition of N.

For each $J = (J_1, J_2, \ldots, J_m) \in \mathcal{J}_m$, the fuzzy maximum of the within-group fuzzy maximum $\tilde{d}(J_k)$ is given by

$$\tilde{D}(J) = \widetilde{\max}\left\{\tilde{d}(J_k) : 1 \leqq k \leqq m\right\}.$$

Note $\tilde{D}(J) \in \mathcal{F}_+$ for $J \in \mathcal{J}_m$. Then, given a fuzzy metric function $\tilde{d}$ and the number of clusters m, the problem is described as follows:

Problem A.

$$\mathop{Optimize}_{J \in \mathcal{J}_M} \tilde{D}(J).$$

We say that J^* is Pareto-optimal if $\tilde{D}(J) \not\succcurlyeq \tilde{D}(J^*)$ means $\tilde{D}(J) = \tilde{D}(J^*)$.

Analysis. (Omitted.)

D.P. Algorithm for finding Pareto-optimal solutions.

We say that a vector $P = (n_1, n_2, \ldots, n_m)$ is a distribution form (cf. [1,2])if $\sum_{i=1}^{m} n_i = n$ and n_i is a positive integer with $n_1 \geqq n_2 \geqq \cdots \geqq n_m$.

For each distribution form, the D.P. algorithm is applied, which is described in the following numerical example.

A Numerical Example. Consider the problem of clustering the elements of $N = \{1, 2, 3, 4\}$ into two clusters, that is, $n = 5$ and $m = 2$. Table 1 gives the fuzzy metric $\tilde{d}(i, j)$, where for any $a, b \in R_+$ with $a < b$, let

$$[a, b](t) = \begin{cases} 0, & 0 \leqq t < a \\ (t-a)/(b-a), & a \leqq t \leqq b \\ 1, & b < t. \end{cases}$$

i \ j	1	2	3	4
1		[5, 10]	[4, 9]	[6, 8]
2			[5, 8]	[7, 10]
3				[7, 8]
4				

Table 1: Fuzzy metric $\tilde{d}(i, j)$ $\left(\tilde{d}(i, j) = \tilde{d}(j, i), \tilde{d}(i, i) = \tilde{0}\right)$

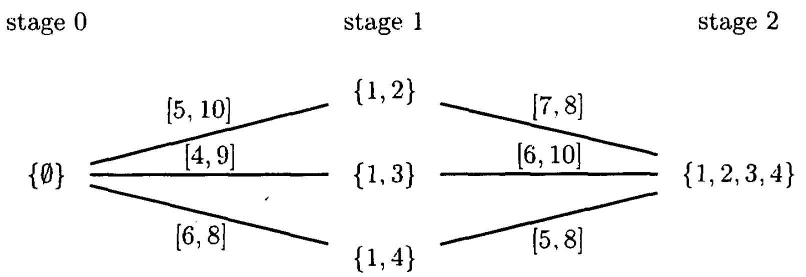
We have the following distribution forms:

$$P_1 = (3, 1), P_2 = (2, 2).$$

The possible clusters with the distribution form P_i $(i = 1, 2)$ are as follows:

$$P_1 : \{1, 2, 3\}\{4\}, \{1, 2, 4\}\{3\}, \{1, 3, 4\}\{2\}, \{2, 3, 4\}\{1\}.$$
$$P_2 : \{1, 2\}\{3, 4\}, \{1, 3\}\{2, 4\}, \{1, 4\}\{2, 3\}.$$

For a given distribution form P, the stages for D.P. algorithm correspond to the formulation of each cluster and the set of states in each stage is the set of subsets clustered so far. We need a stage 0, in which the empty cluster $\emptyset$ is the only state. For the distribution form $P_2 = (2, 2)$, we have the following diagram of a D.P. network and recursive formula.



$\tilde{f}_2^{P_2}(\{1,2,3,4\}) = \tilde{0}, \tilde{f}_1^{P_2}(\{1,2\}) = [7,8], \tilde{f}_1^{P_2}(\{1,3\}) = [6,10], \tilde{f}_1^{P_2}(\{1,4\}) = [5,8], \tilde{f}_0^{P_2}(\{\emptyset\}) = \mathrm{Ext}\{[5,10] \vee \tilde{f}_1^{P_2}(\{1,2\}), [4,9] \vee \tilde{f}_1^{P_2}(\{1,3\}), [6,8] \vee \tilde{f}_1^{P_2}(\{1,4\})\} = \mathrm{Ext}\{[5,10] \vee [7,8], [4,9] \vee [6,10], [6,8] \vee [5,8]\} = [6,8] \vee [5,8] = [6,8]$, where $\vee$ in the above means $\widetilde{\max}$ and $\mathrm{Ext}(\tilde{D}) = \{\tilde{x} \in \tilde{D} | \tilde{x} \preccurlyeq \tilde{y}, \tilde{y} \in \tilde{D}$ means $\tilde{x} = \tilde{y}\}$ and $\tilde{D} \subset \mathcal{F}_+$. Thus, the Pareto-partition for P_2 is $\{1,4\}, \{2,3\}$ with the fuzzy value $[6,8]$. Similarly, we can find that the Pareto-partition for P_1 is $\{1,2,3\}\{4\}$ with the fuzzy value $[5,10]$. Therefore, the Pareto values for Problem A is $\mathrm{Ext}\{\tilde{f}_0^{P_1}(\{\emptyset\}), \tilde{f}_0^{P_2}(\{\emptyset\})\} = \{[5,10], [6,8]\}$ with Pareto-partitions $\{1,4\}\{2,3\}$ and $\{1,2,3\}\{4\}$.

References.

[1] T. S. Arthanari and Yadolah Dodge. *Mathematical programming in statistics.* John Wiley & Sons Inc., New York, 1981. With a foreword by C. R. Rao, Wiley Series in Probability and Mathematical Statistics.

[2] R. E. Jensen. A dynamic programming algorithm for cluster analysis. *Oper. Res.*, 17:1034–1057, 1969.

[3] Hideki Kaminura and Masami Kurano. Clustering by a fuzzy metric. *Fuzzy Sets and Systems*, 120(2):249–254, 2001.