ANALYSIS OF FUZZY DATA FOR FUZZY MODELS

Witold Pedrycz
Department of Automatic Control and Computer Sci.
Silesian Technical University 44-100 Gliwice,Poland

1. Introduction

Having a look at diverse interesting results obtained for fuzzy relational equations(for recent review see e.g. [1]),we can easily notice they have been derived under assumption the given fuzzy relational equation is solvable.Unfortunatelly,it is difficult to know a priori the equation has any solution. A situation becomes much more cumbersome while studying a system of equations,

$$\begin{aligned} A_1 \circ R &= B_1 \\ A_2 \circ R &= B_2 \\ &\vdots \\ A_N \circ R &= B_N \end{aligned} \qquad (1)$$

where no simple,easy-to-use set of conditions required to make the system (1) solvable,can be expected.This is mainly due to fact of interrelationships existing in the system of equations which are difficult to detect.Even every equation has a solution separately,the overall solution may not exist at all.

This note is addressed to the problem of searching the structure of the fuzzy data appearing in the above system that enables us to study consistency of the appropriate data with respect to the form of the fuzzy relational equation,and,possibly,to eliminate some of them characterizing by the lowest values of the performance index of consistency.In this context,the use of hierarchical clustering methods will be clarified.

2. Some related results in fuzzy relational equations

First of all we summarize some useful results obtained in theory of fuzzy relational equations. As standard, a usual notation will be applied, viz.

$$A_i \circ R = B_i,\ i=1,2,\dots,N \tag{1a}$$

$$\bigvee_{k=1}^{n} [A_i(a_k) \wedge R(a_k,b_l)] = B_i(b_l) \tag{2}$$

$l=1,2,\dots,m$, $card(\mathbb{A})=n$, $card(\mathbb{B})=m$, $A_i:\mathbb{A}\to[0,1]$, $B_i:\mathbb{B}\to[0,1]$, $R:\mathbb{A}\times\mathbb{B}\to[0,1]$, $\vee=\max$, $\wedge=\min$. Denote also

$$\mathcal{R} = \{R \mid A_i \circ R = B_i\} \tag{3}$$

a set of fuzzy relations for A_i, B_i given. Now we have cf. [3]

Proposition 1

If $\mathcal{R} = \bigcap_{i=1}^{N} \mathcal{R}_i \neq \emptyset$ then a fuzzy relation

$$\hat{R} = \bigcap_{i=1}^{N} \hat{R}_i \tag{4}$$

$$\hat{R}(a_k,b_l) = \bigwedge_{i=1}^{N} \hat{R}_i(a_k,b_l) = \bigwedge_{i=1}^{N} [A_i(a_k) \to B_i(b_l)] \tag{5}$$

($a\to b = 1$, if $a\le b$; b, if $a>b$, $a,b\in[0,1]$) forms an element of $\mathcal{R}$; moreover $\hat{R} = \max \mathcal{R}$.

A form of the above relationship (4) is worthy to be noticed: the final fuzzy relation $\hat{R}$ results as intersection of the partial results $\hat{R}_i$. Denote by $\mathcal{R}_{\{i_0\}} = \bigcap_{i\neq i_0}^{N} \mathcal{R}_i \neq \emptyset$ a family of fuzzy relations which has some common elements. Let $\mathcal{R}_{\{i_0\}} \cap \mathcal{R}_{i_0} = \emptyset$. Hence in virtue of (4), inserting the pair of data (A_{i_0}, B_{i_0}) into the formula specified above we get,

$$\hat{R} = \bigcap_{i\neq i_0} (A_i \to B_i) \ \cap\ (A_{i_0} \to B_{i_0}) \tag{6}$$

and, unfortunatelly, even for $i\neq i_0$ the resulting fuzzy sets $A_i \circ \hat{R}$ may vary significantly from the fuzzy set B_i. The following propositions characterize conditions under which the fuzzy data can be treated as consistent and inconsistent, respectively.

Proposition 2.

Let $\mathcal{R}_i,\mathcal{R}_j\neq\emptyset$, $i,j=1,2,\dots,N$. If $A_i=A_j$ and $\mathrm{supp}(B_i)\cap\ \mathrm{supp}(B_j)=\emptyset$ ($\mathrm{supp}(\cdot)$ stands for a support of the relevant fuzzy set) then $\mathcal{R}_i\cap\mathcal{R}_j=$ $=\emptyset$. Moreover $A_i\circ\hat{R}=A_j\circ\hat{R}=\hat{B}$ with $\hat{B}$ being identically equal to zero, $\hat{R}$ calculated by means of (4).

Consider the second extremal case described by.

Proposition 3.

Let $\mathcal{R}_i,\mathcal{R}_j\neq\emptyset$. If $\mathrm{supp}(A_i)\cap\mathrm{supp}(A_j)=\emptyset$ then, whatever the corresponding B_i,B_j are (of course we are allowed to choose them not violating the condition specified before, viz. $\mathcal{R}_i,\mathcal{R}_j\neq\emptyset$) one has $\mathcal{R}_i\cap\mathcal{R}_j\neq$ $\neq\emptyset$.

3. Index of consistency of the fuzzy data

The above results, especially those conveyed by Propositions 2 and 3, have a direct impact on the proposed performance index. It is evident that significant inconsistencies occur if B_i and B_j differ "significantly" for the same or almost the same fuzzy sets A_i and A_j. This suggests the inconsistency existing in the pair (A_i,B_i) and (A_j,B_j) can be measured performing comparison of degrees of equality of the fuzzy sets A_i,A_j and the corresponding B_i and B_j. The higher the degree of equality of A_i and A_j, and at the same time, the lower the degree of equality of B_i and B_j, the more inconsistent the analyzed pair. And, of course, it is more difficult to solve the system of equations $A_i\circ R=B_i$, $A_j\circ R=B_j$.

Now it is expedient to introduce an index of equality of two fuzzy sets. It can be done in various ways. Here we adopt a definition that has a plausible logical interpretation, cf. [2]. The degree of equality of A_i and A_j, denoted by $[\![A_i\equiv A_j]\!]$ is given by

$$[\![A_i\equiv A_j]\!] = (A_i\subset A_j)\ \&\ (A_j\subset A_i) \tag{7}$$

where $A_i \subset A_j$ stands for the degree of containment of A_i in A_j. If "&" and "$\subset$" are expressed in terms of standard (to a certain extent) operations for fuzzy sets then (7) reads as

$$[\![A_i \equiv A_j]\!] = \{\min_{1 \le k \le n} [A_i(a_k) \rightarrow A_j(a_k)]\} \wedge \{\min_{1 \le k \le n} [A_j(a_k) \rightarrow A_i(a_k)]\} =$$

$$= \min_{1 \le k \le n} \{ \min\{[A_i(a_k) \rightarrow A_j(a_k)], [A_j(a_k) \rightarrow A_i(a_k)]\}\} \qquad (8)$$

Sometimes (7) may seem to be too restrictive (too pessimistic). On the opposite side we can consider an optimistic form of the equality index specified accordingly

$$[\![A_i \equiv A_j]\!]_{max} = \max_{1 \le k \le n} \{ \min\{[A_i(a_k) \rightarrow A_j(a_k)], [A_j(a_k) \rightarrow A_i(a_k)]\}\} \qquad (9)$$

In sequel if $\alpha_{ij} = [\![A_i \equiv A_j]\!]$ $\beta_{ij} = [\![B_i \equiv B_j]\!]$ we may expect the data set (exactly the pairs "i" and "j") is consistent. This condition may be enhanced by putting $\alpha_{ij} = [\![A_i \equiv A_j]\!]_{max}$ instead this already studied. If $\alpha_{ij} \geq \beta_{ij}$ we can hardly expect the system of equations can be solved. Summarizing two following extremal cases:

i/ $\alpha_{ij} = 1.0$, $\beta_{ij} = 0.0$ which gives the most difficult situation where the system of equations has to be solved, and

ii/ $\alpha_{ij} = 0.0$, $\beta_{ij} = 1.0$ in which the system is solvable, we introduce

$$\delta_{ij} = \alpha_{ij} \rightarrow \beta_{ij} \qquad (10)$$

It expresses a degree of feasibility the system of equations $A_i \circ R = B_i$, $A_j \circ R = B_j$ can be solved. As far the introduced degree of feasibility δ_{ij} takes into account the specified pair of the data set. To obtain a global view on the structure of the entire data set we refer to commonly used hierarchical clustering methods. It is to be underlined the objects grouped are characterized by δ_{ij} that may be viewed as the appropriate measure of similarity of the i-th and j-the j-th object.Then all δ_{ij} are arranged into a form of symmetrical matrix,

$$\underline{\delta} = \begin{bmatrix} 1 & \delta_{12} & \cdots & & \delta_{1N} \\ & 1 & \cdots & & \delta_{2N} \\ & & 1 & & \\ & & & 1 & \\ & & & \vdots & \\ & & & & 1 \end{bmatrix}$$

Starting with "N" clusters formed by individual pairs(A_i,B_i),an agglomerative procedure is used,so the clusters that are similar at the highest level are merged,and the procedure is repeated until a moment one cluster is obtained.The distance between two clusters,say $\mathfrak{X}$ and $\mathfrak{Y}$,is measured by

$$d(\mathfrak{X},\mathfrak{Y}) = \max_{\substack{(A_i,B_i)\in\mathfrak{X} \\ (A_j,B_j)\in\mathfrak{Y}}} \rho_{ij} \qquad (11)$$

with ρ_{ij} standing for a distance function $\rho_{ij}=1-\delta_{ij}$ .The hierarchy generated by the distance specified by (11) is also well known as a complete linkage method.Basing on the dendrogram derived,the most conflicting (inconsistent) elements in the data set may be selectively discarded.This means one can try to solve the reduced system of equations that is expected to have the solution,at least approximate.

References

1. A.Di Nola,W.Pedrycz,S.Sessa,W.Pei-Zhuang,Fuzzy relation equations under a class of triangular norms:a survey and new results,Stochastica,2,1984,99-145.

2. S.Gottwald,Characterization of the solvability of fuzzy equations,EIK,to appear.

3. W.Pedrycz,Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data,Fuzzy Sets and Systems,16,1985,163-175.