*missing data, incomplete information*
*fuzzy rules, fuzzy operators, classifier*

Sylwia POŚPIECH-KURKOWSKA*

# INCOMPETE DATA IN FUZZY INFERENCE SYSTEM

The paper describes a method for dealing the incomplete data in a Mamdani fuzzy system and the influence of inference interpretation on an efficiency the fuzzy system operating on incomplete data. Two representations of missing information are discussed theoretically and then presented on an example of iris data. Various interpretations of fuzzy rules and various types of membership function are examined in order to find a solution of fuzzy system that is more robust to missing data.

## 1. INTRODUCTION

In real world we often encounter with input data vectors where some features are missing. Reasons can be various. In medicine some examinations can be risky for the patient, or too expensive, or a patient can deliberately withhold some information in questionnaire. Such incompleteness of input information is a problem a human being usually deals with very efficiently. On the contrary a fuzzy inference system where the knowledge is represented by rules (and the majority of the automatic reasoning methods), it is unable to produce a result, if the input data does not match a predefined vector of attributes. Shall we elaborate a special set of rules for every possible case of missing data? It would be impractical, and in systems with many inputs even unachievable for exploding complexity. Rather we should make a fuzzy system with a given rule base able to process incomplete information. Another question is how to choose among many interpretations of fuzzy inference process in order to design a system that is the most robust to partially missing data.

The problem of dealing with missing information has already drawn the attention of researchers. Common probabilistic solution is to replace the missing value with its estimate obtained from the conditional probability distribution given the known features. In a fuzzy setting Berthold and Huber [1] describe a method for training a classifier with incomplete data. In normal operation of the classifier a missing value on input is assigned a degree of membership equal to one. Gabrys [4] presents a generalized fuzzy min-max neural network that can deal with incomplete information both during the design and the normal operation. The missing values are represented as real valued intervals spanning the whole range of possible values. The result of the classification in

* Zakład Elektroniki Biomedycznej, Inst. Elektroniki, Politechnika Śląska ul. Akademicka 16, 44-101 Gliwice,
e-mail: sylwia.pospiech@polsl.pl

this situation is rather a reduced number of alternative classifications than one class. Description of a diagnosis support system used in various fields of medicine mentions that a user is allowed to specify some elements as ‘unknown’ [2]. In the system for assessment of skeletal development of children [5] the ability to deal with incomplete data is achieved by dividing the classifier into subclassifiers that can be activated independently.

## 2. PROCESSING OF INCOMPLETE DATA IN THEORY

### 2.1. FUZZY INFERENCE SYSTEM

Knowledge in a fuzzy inference system is represented by rules. Each rule in a Mamdani system with a multiple input and a single output has a form:

$$x_1 \text{ is } A_1 \wedge x_2 \text{ is } A_2 \wedge \ldots \wedge x_k \text{ is } A_k \Rightarrow y \text{ is } B \tag{1}$$

Where

$x_1, x_2, \ldots, x_k$ – values of input variables

$A_1, A_2, \ldots, A_k$ – fuzzy sets for input variables

$y$ - value of output variable

$B$ - fuzzy sets of output variable

Fuzzy set A is described by its membership function $\mu_A$ with values in the range [0, 1]. Statement $x_k$ *is* $A_k$ is interpreted as supremum of intersection of the fuzzy sets $x_k$ *is* $A_k$. Connective $\wedge$ (and) can be interpreted by minimum operator or by product operator. Value obtained from evaluation of the premise is called firing level of the rule. Implication $\Rightarrow$ is interpreted by minimum operator or by product operator. If there are several rules, then the inference result is obtained by aggregation of outputs of all rules. Aggregation is usually performed by maximum operator or probabilistic or (probor). Whenever a crisp output is necessary, a defuzzification is applied.

### 2.2 REPRESENTATION OF MISSING DATA

Consider a fuzzy inference system with $n$ input variables. Assume that a datum for $k$-th input is missing. How to execute the inference process in this case? Note that the only problem that needs a special solution is evaluation of premises where $k$-th variable is missing. Other effects for the inference process are caused by a way a modified firing level of the rule projects on the conclusion and the aggregation of rules.

‘Skip’ approach: if the $k$-th input is missing, then simply skip the statement $x_k$ *is* $A_k$ in the premise of the rule. Firing level of the rule in this case is calculated as:

$$\bigwedge_{\substack{1 \le i \le n \\ i \ne k}} \sup(x_i \cap A_i) \tag{2}$$

and as from definition of the membership function:

$$0 \le \sup(x_i \cap A_i) \le 1 \tag{3}$$

so from (2) and (3) it can be concluded:

$$\underset{\substack{1\le i\le n\\ i\ne k}}{\wedge} \sup(x_i \cap A_i) \ \ge \ \underset{1\le i\le n}{\wedge} \sup(x_i \cap A_i) \tag{4}$$

That means that the firing level of the rule with unknown $k$-th input is always greater or equal to the firing level of the rule with all known inputs. Statement (4) is true independent of the connective $\wedge$ interpretation, however with product as $\wedge$, firing of the rule is more likely to be higher for missing data then for complete. Note that "skip" representation is equivalent to assuming a degree of 1 for the statement $x_k$ *is* $A_k$, which is similar to approach proposed in [1].

'Null' approach: if the $k$-th input is missing, then substitute it by a 'null' fuzzy set representing our unknowledge. The null fuzzy set for the $k$-th input variable has a membership function equal to constant $c$ = 0.5 over the entire domain of the $k$-th variable. In this case the premise is evaluated:

$$\underset{1\le i\le n}{\wedge} \sup(x_i \cap A_i) = \underset{\substack{1\le i\le n\\ i\ne k}}{\wedge} \sup(x_i \cap A_i) \ \wedge \ \sup(x_k \cap A_k) \tag{5}$$

And as from definition of the membership function

$$0 \le \sup(x_i \cap A_i) \le c \tag{6}$$

If connective $\wedge$ is interpreted by minimum, then we get:

$$\min\left( \min_{\substack{1\le i\le n\\ i\ne k}} \left(\sup(x_i \cap A_i)\right), c \right) \le c \tag{7}$$

That means the rule with missing data is not fired at higher level then for complete data and the firing level is cut off at $c$ = 0.5.

If connective $\wedge$ is interpreted by product, then we get:

$$prod\left( \underset{\substack{1\le i\le n\\ i\ne k}}{prod} \left(\sup(x_i \cap A_i)\right), c \right) = c \cdot \underset{\substack{1\le i\le n\\ i\ne k}}{prod} \left(\sup(x_i \cap A_i)\right) \tag{7}$$

That means the firing level of the rule with missing data is reduced by a factor of $c$=0.5 in comparison with the rule for complete data.

## 3. DEALING WITH INCOMPLETE DATA ON EXAMPLE OF IRIS DATA

In this section we demonstrate how the proposed methods of managing missing data work and how the interpretation of inference influences the efficiency of the fuzzy classifier. Iris plant data are used for tests. This well known data set [2] published by R.A. Fisher in 1936 is often applied in tests of classifiers. The set counts 150 cases, for each case four features (sepal length, sepal width, petal length, petal width in cm) and a proper classification into one class (Iris versicolor, iris setosa, iris virginica).

### 3.1. APPLIED FUZZY INFERENCE SYSTEMS

Tests have been conducted on fuzzy systems with various types of membership functions and different interpretation of the inference process. However the basic structure of all systems is the same. Domain of each input variable is covered by three fuzzy sets representing values typical for every class. Their membership functions are generated automatically on the base of the data. Output space is divided into three fuzzy sets corresponding to classes. The same type of the membership function is used as for inputs. Parameters of the membership functions are shown in the tab.1.

| Symbol | Type of mf | Variable | Class 1 | Class 2 | Class 3 |
|---|---|---|---|---|---|
| Gaus | Gaussian | $x_1$ | 0.352 5.006 | 0.516 5.936 | 0.636 6.588 |
| | | $x_2$ | 0.381 3.418 | 0.314 2.770 | 0.322 2.974 |
| | | $x_3$ | 0.174 1.464 | 0.470 4.260 | 0.552 5.552 |
| | | $x_4$ | 0.107 0.244 | 0.198 1.326 | 0.275 2.026 |
| | | $y$ | 0.500 1.000 | 0.500 2.000 | 0.350 3.000 |
| Tri1 | Triangular | $x_1$ | 3.947 5.006 6.197 | 4.382 5.936 7.532 | 4.056 6.588 8.556 |
| | | $x_2$ | 1.741 3.418 4.891 | 1.615 2.770 3.715 | 1.813 2.974 4.213 |
| | | $x_3$ | 0.768 1.464 2.118 | 2.370 4.260 5.520 | 3.974 5.552 7.574 |
| | | $x_4$ | 0.028 0.244 0.778 | 0.837 1.326 2.037 | 1.087 2.026 2.737 |
| | | $y$ | 0.000 1.000 2.000 | 1.000 2.000 3.000 | 2.100 3.000 3.900 |
| Tri2 | Triangular | $x_1$ | 4.065 5.006 6.065 | 4.555 5.936 7.355 | 4.337 6.588 8.337 |
| | | $x_2$ | 1.927 3.418 4.727 | 1.743 2.770 3.610 | 1.942 2.974 4.075 |
| | | $x_3$ | 0.845 1.464 2.045 | 2.58 4.26 5.38 | 4.149 5.552 7.349 |
| | | $x_4$ | 0.052 0.244 0.71866 | 0.891 1.326 1.958 | 1.191 2.026 2.658 |
| | | $y$ | 0.000 1.000 2.000 | 1.000 2.000 3.000 | 2.100 3.000 3.900 |
| Trap | Trapezoidal | $x_1$ | 4.133 4.800 5.200 6.000 | 4.667 5.600 6.300 7.233 | 4.467 6.200 6.900 8.233 |
| | | $x_2$ | 2.033 3.100 3.700 4.633 | 1.833 2.500 3.000 3.533 | 2.000 2.800 3.200 4.000 |
| | | $x_3$ | 0.867 1.400 1.600 2.000 | 2.667 4.000 4.600 5.267 | 4.300 5.100 5.900 7.233 |
| | | $x_4$ | 0.067 0.200 0.300 0.700 | 0.933 1.200 1.500 1.900 | 1.267 1.800 2.300 2.567 |
| | | $y$ | 0.000 1.000 1.000 2.000 | 1.000 2.000 2.000 3.000 | 2.200 3.000 3.000 3.800 |

Tab.1 Membership functions (mf) and their parameters.

A rule base is the same in every system and consists of the following three rules:

$$
\begin{aligned}
&R1: \ x_1 \text{ is } A_{1,1} \wedge x_2 \text{ is } A_{2,1} \wedge x_3 \text{ is } A_{3,1} \wedge x_4 \text{ is } A_{4,1} \Rightarrow y \text{ is } B_1 \\
&R2: \ x_1 \text{ is } A_{1,2} \wedge x_2 \text{ is } A_{2,2} \wedge x_3 \text{ is } A_{3,2} \wedge x_4 \text{ is } A_{4,2} \Rightarrow y \text{ is } B_2 \\
&R3: \ x_1 \text{ is } A_{1,3} \wedge x_2 \text{ is } A_{2,3} \wedge x_3 \text{ is } A_{3,3} \wedge x_4 \text{ is } A_{4,3} \Rightarrow y \text{ is } B_3
\end{aligned} \tag{8}
$$

The first index at $A$ describes the input number, the second describes the class, and the index at $B$ describes the class. Inference process is interpreted by minimum, maximum, product and probabilistic or operators as described in p.2.1. Two most popular methods for defuzzification are used: center of area (COA) and mean of maxima (MOM).

### 3.2 EXAMPLE FOR A SINGLE CASE

This section demonstrates in detail how missing data influences the functioning of a fuzzy system for a single case ($x_1$= 6.2, $x_2$= 2.9, $x_3$=4.3, $x_4$=1.3). Table 2 shows the evaluation of each statement $x_i$ *is* $A_{i,k}$, and the firing for each rule for complete data and missing value for $x_3$. Rule $R_1$ is not fired, rule $R_2$ is fired at lower level for null representation, at equal or slightly higher level for skip representation, rule $R_3$ is fired at higher level for both representations. The change in firing levels is more substantial for *prod* than for *min*. Fig.1 shows the output fuzzy sets for rules, the degree of membership for $B_2$ (proper class) is smaller and the degree of membership for $B_3$ higher for both representations of missing value, that means the fuzziness of the classification increases in comparison to the result for complete data, but the defuzzified values still point the proper class

| | Complete data | | | Missing $x_3$ – skip representation | | | Missing $x_3$ – null representation | | |
|---|---|---|---|---|---|---|---|---|---|
| | $x_i$ *is* $A_{i,k}$ | Firing of $R_k$ | | $x_i$ *is* $A_{i,k}$ | Firing of $R_k$ | | $x_i$ *is* $A_{i,k}$ | Firing of $R_k$ | |
| | | Min | Prod | | Min | Prod | | Min | Prod |
| 1 | 0 0.69 0 0 | 0 | 0 | 0 0.69 - 0 | 0 | 0 | 0 0.69 0.5 0 | 0 | 0 |
| 2 | 0.84 0.86 0.97 0.95 | 0.84 | 0.66 | 0.84 0.86 - 0.95 | 0.84 | 0.68 | 0.84 0.86 0.5 0.95 | 0.50 | 0.34 |
| 3 | 0.85 0.94 0.21 0.23 | 0.21 | 0.04 | 0.85 0.94 - 0.23 | 0.23 | 0.18 | 0.85 0.94 0.5 0.23 | 0.23 | 0.09 |

Tab.2 The evaluation of the fuzzy system *tri1* with the complete data ($x_1$=6.20, $x_2$=2.90, $x_3$=4.30, $x_4$=1.30, proper class=2) and missing value for variable $x_3$. Two interpretations of *and* are used *minimum* or *product* operator.

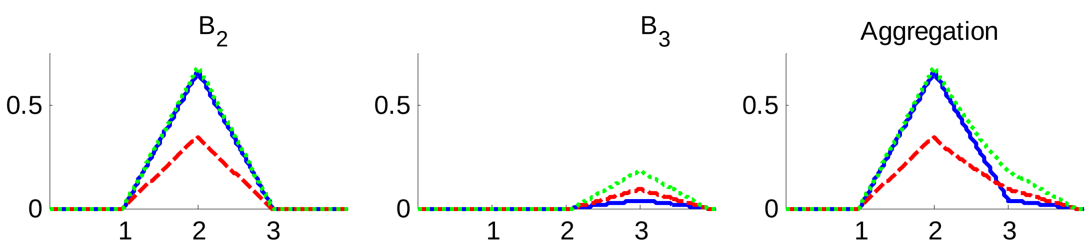


Fig.1 Outputs of rules R2, R3 and aggregation for complete data (rule R1 is not fired, see tab.2) - solid line, for missing $x_3$: skip representation - dotted line, null representation – dashed line (classifier *tri1*, implication *prod*). After MOM defuzzification 2 for all, COA: complete data 2.05, null 2.19, skip 2.18.

### 3.3 RESULTS

Fuzzy systems based on product/probabilistic or operators showed to be more sensitive to missing data then systems based on minimum/maximum operators (tab.3). However the deterioration of the classifier's performance depends strongly on which variable is missing. Features $x_3$ and $x_4$ are much more crucial for classification then features $x_1$ and $x_2$. Influence of membership function type is not considerable. The difference in efficiency of the classifiers with null and skip representation is not significant. This effect can be caused by the fact each input variable is present in premises of all rules that is a change in one rule conclusion is compensated by similar changes in the output of the other rules. Results for MOM defuzzification are similar as for COA.

| | $\wedge, \Rightarrow$ min, *aggregation*: max | | | | | | | | $\wedge, \Rightarrow$ prod, *aggregation*: Probor | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Skip | | | | Null | | | | Skip | | | | Null | | | |
| | #1 | #2 | #3 | #4 | #1 | #2 | #3 | #4 | #1 | #2 | #3 | #4 | #1 | #2 | #3 | #4 |
| Gaus | 0 | 0 | 1 | 8 | 0 | 0 | 2 | 7 | 0 | 1 | 3 | 11 | 0 | 1 | 3 | 11 |
| Tri1 | 0 | 0 | 3 | 6 | 0 | 0 | 1 | 6 | -1 | -1 | 5 | 6 | -1 | -1 | 5 | 7 |
| Tri2 | 0 | 0 | 3 | 4 | 0 | 0 | 3 | 6 | -3 | -1 | 5 | 6 | -2 | -1 | 5 | 6 |
| Trap | 0 | 0 | 1 | 6 | -1 | -1 | 1 | 6 | -1 | 0 | 4 | 7 | -1 | 0 | 4 | 7 |

Tab.3. The increase in number of wrong classifications respectively to the same fuzzy system with complete input data. A comparison of *null* and *skip* representations, *min/max* vs. *prod/probor* interpretation of inference.

## 4. CONCLUSIONS

A method for dealing with missing input values has been presented that enables a classical Mamdani fuzzy system to process incomplete data. No significant difference in overall classifier performance between the two proposed representations of missing data was observed in tests for iris plant data. The inference interpreted by minimum/maximum operators has turned out to be more robust with respect to incomplete information than the inference based on product/probabilistic sum. Future investigations should include tests for more varied rule bases and non relational interpretations of implication.

### BIBLIOGRAPHY

[1] Berthold M.R., Huber K.-P., Missing values and learning of fuzzy rules, International Journal of Uncertainty, Fuzziness and Knowledge-based Systems vol.6 (2) (1998) 171-178

[2] Boegl K., Adlassing K.-P., Hayashi Y. Rothenfluh T. Leitich H., Knowledge acquisition in the fuzzy knowledge representation framework of a medical consultation system, Artificial Intelligence in Medicine 30 (2004) 1-26

[3] Fisher R.A., Iris dataset, http://www.ics.uci.edu/~mlearn/MLRepository.html

[4] Gabrys B., Neuro-fuzzy approach to processing inputs with missing values in pattern recognition problems, International Journal of Approximate Reasoning 30 (2002) 149-179

[5] Pośpiech-Kurkowska S., Gertych A. Piętka E., Rozmyty system do szacowania rozwoju kostnego na podstawie rentgenogramów. Proc. BIB Conf., Biocybernetyka i Inżynieria Biomedyczna, 2005,