Using Precedent Information in Fuzzy Expert Systems[1]

D. P. Vetrov and D. A. Kropotov

Computer Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, GSP-1, 119991 Russia
e-mail: vetrovd@yandex.ru; dkropotov@yandex.ru

Abstract—Many pattern recognition tasks are connected with expert information, which can be expressed in terms of linguistic rules. A theory of fuzzy sets presents one of the most successful ways of using these rules. However, in this case, two main problems appear that cannot be completely solved by experts in some problem domains: fuzzy set formation and fuzzy rule generation. A possible solution based on the use of precedent information is proposed in this paper.

INTRODUCTION

In many pattern recognition tasks, in addition to a learning sample, there is some *a priori* information about the process under study. This knowledge (expert information) can be expressed as qualitative relations between object features and a predictable value. A theory of fuzzy sets [1] provides a convenient means for representing and using this information. Most difficult in this approach is the determination of membership functions and the formation of a system of rules of inference. Often, these issues are related to the expert's competence. Unfortunately, there are some problem domains where experts are not able to solve the problem in full measure. These bottlenecks of the theory of fuzzy sets hamper the employment of fuzzy expert systems. The solution can be found in using precedent information. In this paper, one possible way of using it during construction of a fuzzy expert system is presented.

DESCRIPTION OF EXPERT INFORMATION

Suppose we have a learning sample as initial information. We use the concept of a linguistic variable for representing expert information [1]. The general form of the expert statement takes on the following form:

$$\text{IF } X_{i_1} \in A_{k_1}^{i_1}, \dots, X_{i_l} \in A_{k_l}^{i_l}, \text{THEN } Y \in B_k. \quad (1)$$

Here, X_{i_p} is the value of the i_p feature, $A_{k_p}^{i_p}$ is a fuzzy set with the meaning of a linguistic variable value, Y is a predictable variable, and l is the order of the sentence. Thus, an expert needs only define qualitative relations, such as "IF the average temperature is high and precipitation is normal, THEN the crop will be large." This form of a sentence is close to natural language and allows experts to formulate their thoughts more freely.

Usually, for constructing such systems, an expert must determine the form and the size of fuzzy sets $A_{k_p}^{i_p}$. Note that, in most cases, this task is very complicated. In practice, however, an expert can indicate the approximate boundaries between the fuzzy sets of a feature. We use an isosceles triangle and trapezium as the forms of fuzzy sets (see Fig. 1).

Let $\mu_A(x): X \longrightarrow [0,1]$ be a membership function of a fuzzy set A.

Definition 1. The *coating* of interval $[a, b] \subset R$ is a collection of fuzzy sets $\{A_i\}_{i=1}^n$ such that $\forall x \in [a, b]$ $\exists i \in \{1, \dots, n\}$: $\mu_{A_i}(x) > 0$. If, in addition, $\mu_{A_i}(x) > \alpha$, then the coating is α-*significant*.

Let $\{[a_i, a_{i+1}]\}_{i=1}^n$ be a partition of interval $[a, b]$.

Definition 2. *Regular coating* of interval $[a, b]$ for partition $\{[a_i, a_{i+1}]\}_{i=1}^n$ is a coating $\{A_i\}_{i=1}^n$ such that

(i) $\mu_{A_i}(a_{i+1}) = \mu_{A_{i+1}}(a_{i+1})$, $\forall i = \overline{1, n-1}$,

(ii) $\forall i \in \{1, \dots, n\}$ $\exists x_* \in [a_i, a_{i+1}]$: $\mu_{A_i}(x_*) = \max_x \mu_{A_i}(x)$.

Obviously, the regular coating is α-significant for $\alpha = \min(\min_i(\mu_{A_i}(a_i)), \mu_{A_n}(a_{i+1}))$.

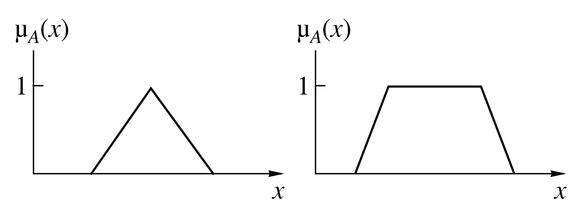

Fig. 1. Forms of fuzzy sets.

[1] This work was supported by the Russian Foundation for Basic Research, project nos. 02-07-90134, 02-07-90137, and 02-01-08007.

Received October 29, 2002

Original Text Copyright © 2003 by Pattern Recognition and Image Analysis. English Translation Copyright © 2003 by MAIK "Nauka/Interperiodica" (Russia).

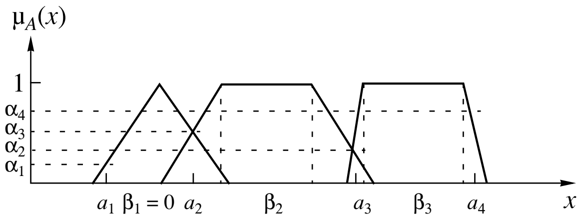

Fig. 2. Parameterization of regular coating.

Definition 3. A fuzzy set A_i of regular coating $\{A_i\}_{i=1}^n$ is β-*full*, if $\frac{\text{mes}\{x \in R | \mu_{A_i}(x) = 1\}}{a_{i+1} - a_i} = \beta$, where mes(.) is the Lebesgue measure on a real line.

It is clear that significance and fullness of regular intervals are two independent notions. Significance determines the degree of superimposing of fuzzy sets and fullness, a degree of approximation of a fuzzy set to a crisp one. Let $\alpha = \mu_{A_i}(a_i)$ and β_i be fullness of A_i. Then, a set of parameters $\{\alpha_1, \ldots, \alpha_{n+1}, \beta_1, \ldots, \beta_n\}$ uniquely determines a regular coating (Fig. 2).

Now, an expert only needs to give a partition of a feature space. A regular coating for this partition will serve as required fuzzy sets.

DESCRIPTION OF THE EXPERT SYSTEM

Suppose that we have n features and m expert rules of type (1). Let us denote the ith rule by R_i and the fuzzy set for the predictable value corresponding to this rule by resR_i. As a fuzzy implication membership function we take the function introduced by Mamdani (see [2]):

$$\mu_{A \to B}(x, y) = \min(\mu_A(x), \mu_B(y)). \quad (2)$$

Then, for the rule of type (1), an appropriate membership function will be the following:

$$\mu_{R_i}(x, y) = \min\left(\mu_{A_{k_1}^{i_1}}(x_{i_1}), \ldots, \mu_{A_{k_l}^{i_l}}(x_{i_l}), \mu_{B_k}(y)\right). \quad (3)$$

In the case where the fuzzy set for the predictable value is the same for the group of rules, a membership function is calculated according to the following formula:

$$\mu_R(x, y) = \frac{\sum_{i:\, \text{res}R_i = B} \mu_{R_i(x,y)}}{\sum_{i:\, \text{res}R_i = B} 1}. \quad (4)$$

This formula slightly differs from the classical determination of a maximum value (see [2]). Nevertheless, we consider that Eq. (4) is more flexible and can take into account the individual character of every rule of inference.

Suppose also that not only the rules of inference are known but information about their weights is likewise known. These weights can be specified by an expert or obtained from the learning sample. Let us denote the weight of the ith rule by w_i. Then, Eq. (4) is transformed to

$$\mu_R(x, y) = \frac{\sum_{i:\, \text{res}R_i = B} w_i \mu_{R_i}(x, y)}{\sum_{i:\, \text{res}R_i = B} w_i}. \quad (5)$$

Let q fuzzy sets be determined in the space of the predictable value. We denote the median of the kth fuzzy set by medB_k; i.e.,

$$\text{med}B_k = \frac{a_k + a_{k+1}}{2}.$$

Then, the point forecast can be found as

$$z = \frac{\sum_{k=1}^{q} \text{med}B_k \mu_{R_k}(x, \text{med}B_k)}{\sum_{k=1}^{q} \mu_{R_k}(x, \text{med}B_k)}. \quad (6)$$

Thus, we have the following model of forecast calculation:

$$M(\{a_k^i\}_{i=0,k=1}^{n,k_i+1}, \{\alpha_k^i\}_{i=0,k=1}^{n,k_i+1}, \{\beta_k^i\}_{i=0,k=1}^{n,k_i}, \{R_i\}_{i=1}^m, \{w_i\}_{i=1}^m). \quad (7)$$

Let us denote precedent information by $\{\hat{x}_k, \hat{y}_k\}_{k=1}^p$. Then, we can find parameters $\{\alpha, \beta\}$ of the model M by optimization on precedents with the following quality functional:

$$J = w\sum_{k=1}^{p} (\hat{y}_k - \tilde{y}_k)^2 + (1-w)\sup_k |\hat{y}_k - \tilde{y}_k|, \quad (8)$$

where $\tilde{y}_k = M(\ldots)(x_k)$ and $w \in [0, 1]$ is the parameter of optimization corresponding to the requirements on the expert system.

To reduce the amount of parameters for optimization, we assume that α and β coefficients are related to features rather than to separate sets. Then, it is possible to take $\alpha_1^i = \ldots = \alpha_{k_i+1}^i = \alpha^i$, $\beta_1^i = \ldots = \beta_{k_i}^i = \beta^i$. In addition, formula (6) gives us the opportunity of making no optimization in the space of answers. Therefore, finally, the model of expert system (7) transforms to

$$M(\{a_k^i\}_{i=0,k=1}^{n,k_i+1}, \{\alpha^i\}_{i=1}^n, \{\beta^i\}_{i=1}^n, \{R_i\}_{i=1}^m, \{w_i\}_{i=1}^m, w). \quad (9)$$

GENERATION OF THE RULES OF INFERENCE

One of the main problems of fuzzy expert systems is to obtain rules from experts. Therefore, it is necessary to be able to extract these rules from input data automatically.

Definition 4. The *representativeness* of a rule is the following value:

$$\mathrm{rep}(R) = \frac{1}{p}\sum_{j=1}^{p}\min\left(\mu_{A_{k_1}^{i_1}}(\hat{x}_{ji_1}), \ldots, \mu_{A_{k_l}^{i_l}}(\hat{x}_{ji_l})\right),$$

where $\{\hat{x}_k, \hat{y}_k\}_{k=1}^{p}$ is precedent information and the rule of inference takes on the form of (1).

Definition 5. The *effectiveness* of the rule of inference is the following value:

$$\mathrm{eff}(R) = \frac{\sum_{j=1}^{p}\min\left(\mu_{A_{k_1}^{i_1}}(\hat{x}_{ji_1}), \ldots, \mu_{A_{k_l}^{i_l}}(\hat{x}_{ji_l}), \mu_{B_k}(\hat{y}_j)\right)}{p\,\mathrm{rep}R}.$$

Then, the common scheme of rule generation will be the following:

Step 1. Construct all possible rules of the first order from available fuzzy sets, i.e., rules of the type "IF $X_{i_1} \in A_{k_1}^{i_1}$, THEN $Y \in B_k$." Calculate the representativeness of each rule and reject the rules with representativeness lower than some specified threshold C_1. For the rest of the rules, calculate effectiveness. The rules with effectiveness higher than the threshold C_2 are the desired rules of inference.

Step 2. Use the rules with representativeness higher than C_1 and effectiveness lower than C_2 to construct rules of the second order for increasing effectiveness. Add the obtained rules with representativeness and effectiveness higher than the corresponding thresholds to the set of inference rules.

Step i. Combine the rules of the $(i-1)$th order with $\mathrm{rep}R > C_1$ and $\mathrm{eff}R < C_2$ into the rules of the ith order.

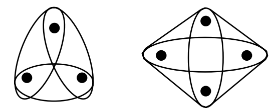

Fig. 3. Construction of rules of third and fourth orders.

The premises of each combined rule lies in the union of the premises of any two combined rules (see Fig. 3).

Add the rules with the representativeness $\mathrm{rep}R > C_1$ and effectiveness $\mathrm{eff}R > C_2$ to the set of inference rules. Reject all rules with representativeness lower than the specified thresholds. If the set of remaining rules of the ith order is not empty, go to the next step, otherwise end.

The effectiveness values of the obtained rules can be used as their weights.

CONCLUSIONS

In this paper, we present an approach to the description of expert information based on the concepts of fuzzy logic. One of the main advantages of this approach is the following: an expert no longer needs to give full information about the form and size of fuzzy sets used in the system, it is only required to make a partition of the feature space. Also, we describe here a way of generating rules from input data without experts. This method substantially simplifies the use of expert systems and broadens the area of their application.

REFERENCES

1. Zadeh, L.A., *The Concept of a Linguistic Variable and Its Application to Approximate Reasoning*, New York, 1965.
2. Perfil'eva, I.G., Applications of a Fuzzy Set Theory, in *Itogi Nauki Tekh., Ser.: Teor. Ver., Mat. Stat., Teor. Kibern.*, VINITI, 1990, issue 29, pp. 83–151.