Information Technology for Extracting the Accurate, Compact and Interpretable Mamdani-type Rule Base

Serhiy Shtovba [1] [0000-0003-1302-4899], Viktor Mazurenko [2] [0000-0002-7589-8797], Mykola Petrychko [2] [0000-0001-6836-7843]

[1] Vasyl' Stus Donetsk National University, 600-richchia str., 21, Vinnytsia, 21021, Ukraine
s.shtovba@donnu.edu.ua

[2] Vinnytsia National Technical University, Khmelnytske Shose, 95, Vinnytsia, 21021, Ukraine
viktor.mazurenko@gmail.com
petrychko.myckola@gmail.com

Abstract. The paper dedicates to development of an information technology for the design of Mamdani-type fuzzy systems derived from experimental datasets. The proposed information technology provides synthesis of accurate, compact and interpretable fuzzy rule bases. A distinctive feature of the information technology is the following four-stage operation scheme: 1) generating a list of adequate candidate-rules; 2) selection of the rules; 3) reduction of the antecedents of the selected rules; 4) parametric tuning of a fuzzy rule base. The criteria of accuracy and interpretability are involved in the first and the fourth stages, and the criteria of accuracy and compactness are used in the second and the third stages. New models for interpretability protection of fuzzy rule bases during the tuning are also proposed. Examples of the application of this information technology for solving 7 identification problems with real experimental data are given.

Keywords: fuzzy identification, rule selection, fuzzy rule base, accuracy, interpretability, compactness, fuzzy inference, genetic algorithm, Pareto front.

1 Introduction

Fuzzy rule-based systems are a popular tool for modeling complex dependencies in engineering, agronomy, economics, ecology, medicine, biology, politics, sports and in other fields [1, 2]. The semantic part of a fuzzy system is a rule base, i.e. a set of if–then rules that describes the relationship between inputs and outputs with usage of terms such as *Low*, *Average*, *High* etc. Those linguistic terms are formalized by fuzzy sets [3]. There are several types of fuzzy rule-based systems: Mamdani, Sugeno, Tsukamoto etc. Among them, the most popular and human-friendly is Mamdani-type system.

The paper dedicates to the design of Mamdani-type fuzzy systems based on utilizing experimental dataset. Such design usually is performed in two stages: fuzzy structural identification and fuzzy parametrical identification [4]. The structural identification consists of determining the input and output variables of models, the formation of term-sets for linguistic variables and describing the dependence by fuzzy production rules. The first two procedures are common to any identification method. The last two procedures are specific for fuzzy identification. As a result of fuzzy structural identification, we obtain a rough model, which outlines the target dependence. At the stage of fuzzy parametric identification, the model parameters are tuned, generally, the membership functions and weights of the rules are changed.

The quality of the fuzzy rule bases is evaluated according to the criteria of accuracy, compactness, and interpretability [4–9]. There are a lot of research relating to extraction fuzzy rule base from datasets [4, 7, 9–14], but a question is still open how to guarantee obtaining an accurate, compact and reasonable (transparent) fuzzy rule base in expert-free regime.

The aim of the paper is to develop an information technology that provides synthesis from experimental data of accurate, compact and interpretable fuzzy rule bases. A distinctive feature of the technology is the following four-stage operation scheme: 1) generating a list of adequate candidate-rules; 2) selection of rules; 3) reduction of the antecedents of the selected rules; 4) parametric tuning of a fuzzy rule base. The first 3 stages correspond to structural identification, and the fourth stage is parametric one. The criteria of accuracy and interpretability are involved in the first and the fourth stages, and the criteria of accuracy and compactness are used in the second and the third stages. New models for interpretability protection of fuzzy rule bases during the tuning are also proposed. The proposed information technology is tested on typical tasks from UCI Machine Learning Repository.

2 Mamdani-type fuzzy rule base

Antecedent and consequent in a Mamdani-type rule are represented by fuzzy sets. Each Mamdani-type rule can be interpreted as a zone with fuzzy boundaries in some factor space. For each fuzzy zone, the output variable takes a fuzzy constant value.

Let us write the Mamdani-type fuzzy rule base as follows [10]:

$$if\ \left(x_1 = \tilde{a}_{i1}\ \ and\ \ x_2 = \tilde{a}_{i2}\ \ and\ ...\ and\ \ x_n = \tilde{a}_{in}\right),\ then\ \ y = \tilde{d}_i,\ with\ weight\ w_i \tag{1}$$

where $X = \left(x_1, x_2, \ldots, x_n\right)$ and y denotes the input variables and the output variable;

$\tilde{a}_{ij}$ denotes the fuzzy term, for example, *Low*, *Average*, *High*, that assess a linguistic value of x_j in the *i*-th rule, $i = \overline{1, N}$, $j = \overline{1, n}$;

N denotes a number of the rules;

$\tilde{d}_i$ denotes a consequent of the *i*-th rule in the form of a fuzzy term;

$w_i \in [0; 1]$ denotes the weight of the *i*-th rule.

Each fuzzy term $\tilde{a}_{ij}$ belongs to term-set $\left\{l_{j1}, l_{j2}, ..., l_{jk_j}\right\}$, $j = \overline{1, n}, k_j \le N$.

Some rules in the base (1) may be short. For the short rule, arbitrary linguistic value of some variable does not influence on the consequent. In this case, the corresponding variable is excluded from the rule antecedent or the term *Don't care* is assigned to it [15].

Let us denote the membership function of terms $\tilde{a}_{ij}$ as $\mu_{ij}(x_j)$, $x_j \in \left[\underline{x}_j; \overline{x}_j\right]$, and of terms $\tilde{d}_i$ as $\mu_{d_i}(y)$, $y \in \left[\underline{y}; \overline{y}\right]$. We use the Gaussian membership function:

$$\mu(x) = \exp\left(-\frac{(x-b)^2}{2c^2}\right), \tag{2}$$

where b denotes the core of the fuzzy set, and $c > 0$ acts as a concentration factor.

The logical inference for the input vector $X^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$ is carried out according to Fig. 1. Firstly, membership degrees $\mu_{ij}(x_j^*)$ of the input values x_j^* to fuzzy terms $\tilde{a}_{ij}$ from rule base (1) are calculated by the formula (2). As a result, we obtain the input vector in the form of a bi-fuzzy set $\tilde{\tilde{X}}^* = \left(\frac{\mu_{i1}(x_1^*)}{\tilde{a}_{i1}}, \frac{\mu_{i2}(x_2^*)}{\tilde{a}_{i2}}, \ldots, \frac{\mu_{in}(x_n^*)}{\tilde{a}_{in}}\right)$. The support of a bi-fuzzy set is fuzzy [16]. In our case it equals to term-set $\left\{l_{j1}, l_{j2}, \ldots, l_{jk_j}\right\}$, $j = \overline{1, n}$.

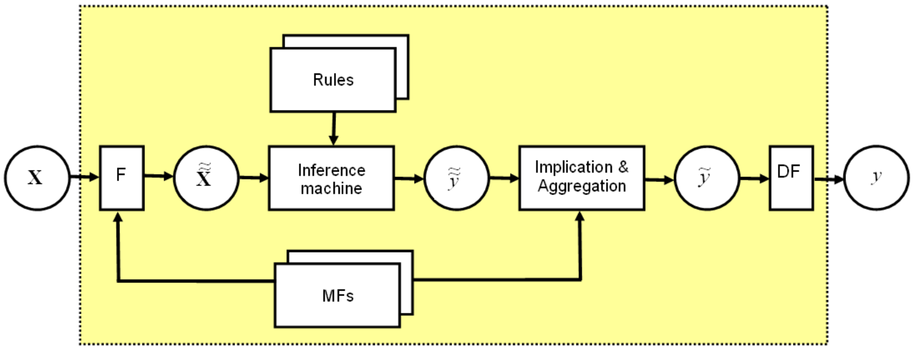

Fig. 1. The logical inference with Mamdani-type fuzzy rule base
(F – fuzzification, MFs – membership functions, DF – defuzzification)

The inference machine outputs the result in form of the following bi-fuzzy set

$$\tilde{y}^* = \left(\frac{\mu_{d_1}(X^*)}{\tilde{d}_1}, \frac{\mu_{d_2}(X^*)}{\tilde{d}_2}, \ldots, \frac{\mu_{d_N}(X^*)}{\tilde{d}_N}\right) \tag{3}$$

where $\mu_{d_i}(X^*) = w_i \cdot \min\left(\mu_{i1}(x_1^*), \mu_{i2}(x_2^*), \ldots, \mu_{in}(x_n^*)\right)$, $i = \overline{1, N}$.

Mapping the output bi-fuzzy set (3) to fuzzy set with support $\left[\underline{y}; \overline{y}\right]$ is carried out in the following way:

$$\tilde{d}_i^* = \mathrm{imp}\left(\tilde{d}_i,\ \mu_{d_i}(X^*)\right), i = \overline{1, N}, \tag{4}$$

where imp denotes the implication, which is realized by the operation of the minimum. The geometric interpretation of this implication is a cut of the graph of the membership function $\mu_{d_i}(y)$ at the level $\mu_{d_i}(X^*)$, $i = \overline{1, N}$ (Fig. 2).

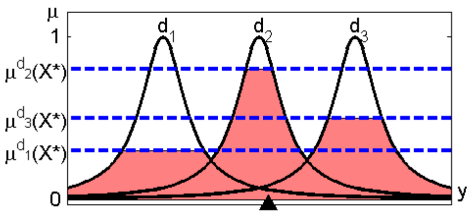

Fig. 2. Implication, aggregation and defuzzification in Mamdani inference

The resulting fuzzy set is obtained by aggregation of fuzzy sets (4):

$$\tilde{y}* = agg\left(\tilde{d}_1^*, \tilde{d}_2^*, ..., \tilde{d}_N^*\right), \tag{5}$$

which is implemented by the operation of maximum over membership functions.

A crisp output value $y*$ corresponding to the input vector X^* is found by defuzzifying the fuzzy set (5) using centroid method:

$$y^* = \frac{\int_{\underline{y}}^{\bar{y}} y \cdot \mu_y(y)\, dy}{\int_{\underline{y}}^{\bar{y}} \mu_y(y)\, dy}.$$

To prevent the effect of narrowing the output range through centroid defuzzification, we expand the support of fuzzy terms $\tilde{d}_i$, $i = \overline{1, N}$ as in [10].

3 Fuzzy identification quality criteria

It is assumed that the dataset reflecting relation between features $X = (x_1, x_2, ..., x_n)$ and output y is presented as follows:

$$(X_r, y_r),\ r = \overline{1, M}, \tag{6}$$

where X_r denotes the input vector in the r-th row of the dataset;

y_r denotes the corresponding output value.

The task of fuzzy identification is to extract the rule base (1) from dataset (6) with the best quality. We will evaluate the fuzzy model quality according to the criteria of accuracy, compactness, and interpretability. The accuracy and compactness of the model are traditional criteria for assessing the quality of identification. For fuzzy models, interpretability, i.e. ability to explain to customers in natural language how the model functions, is also important. The possibility of a meaningful interpretation is an important advantage of fuzzy models, allowing them to compete with other technologies for identifying complex dependencies.

The accuracy of fuzzy model $F(X)$ is assessed by the root mean squared error $RMSE = \sqrt{\frac{1}{M}\sum_{r=\overline{1,M}}\left(y_r - F\left(X_r\right)\right)^2}$ or normalized error $NRMSE = \frac{RMSE}{\overline{y} - \underline{y}}$.

To assess the compactness of the model, we use the following indicators: N – the total number of rules and A – the total length of all antecedents, i.e., the number of terms in all the antecedents.

Interpretability refers to the possibility of a meaningful explanation of the structure and parameters of the model. A fuzzy model is interpretable if the following conditions are satisfied [10]:

• the rule base is not contradictory or redundant, that is, it does not contain rules with the same antecedents;

• the rule base is consistent with the number of terms, that is, each term appears in at least one fuzzy rule;

• an arbitrary input vector produces a non-empty output fuzzy set;

• each term is represented by a normal and convex fuzzy set;

• each term-set is meaningfully interpreted.

A term-set is interpretable if the placement of fuzzy sets on the support is reasonable. For example, the term *Average* is between the terms *Low* and *High*. At the same time, the height of *Low-High* intersection is lower than height of *Low-Average* intersection and also lower than height of *Average-High* intersection. As an example, Fig. 3 shows an uninterpreted term-set with the following problems:

A) strong similarity of membership functions of neighboring fuzzy sets *Low* and *Below Average*, which may contradict the rule base;

B) a loss of linear ordering of the term-set through different curve of membership functions, for example, on the interval (65; 82) fuzzy set *Average* is larger than fuzzy set *Above Average*, and on interval (0; 3) fuzzy set *Below average* is larger than fuzzy set *Low*, although it should be the other way around;

C) a bias of the extreme term core – a decrease in the value of the variable x from 8 to 0 reduces the membership grade to the term *Low*, although it should be the other way around;

D) a cover spot – value $x \in (84;88)$ does not belong to any fuzzy set, hence, for any rule base, the result of inference for $x \in (84;88)$ will be an empty fuzzy set.

Let us define a fuzzy term-set interpretability in the following linguistic form: *Low lower Average, Average lower High, High lower Very High* etc. In a more formalized form, we tie the term-set interpretability with the following three conditions.

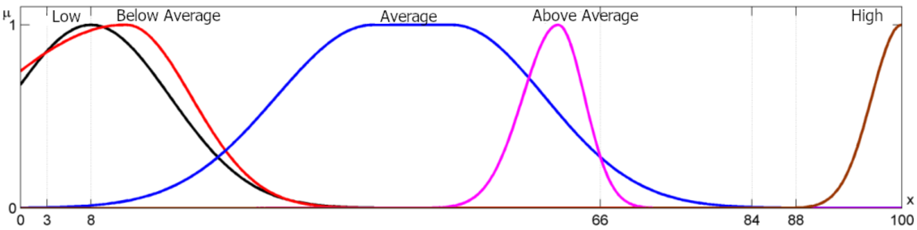

Fig. 3. Typical interpretability problems with a term-set

1. The number of terms should not be too large so that the expert can match each fuzzy set with a reasonable linguistic assessment. In [4], the term-set cardinality is bounded above by a "magic" number 7 ± 2 [17]. Based on our experience in the design of fuzzy systems, it is more expedient to use term-sets with cardinality up to 7.

2. Fuzzy sets of different terms should not be equivalent or almost equivalent. Hence, the graphs of the membership functions of neighboring terms, for example, *Low* and *Average* should definitely differ by an eye.

3. The-term set must be linearly ordered. Let us denote by l_i a term with number *i*, $i = \overline{1, K}$. This term is used as a linguistic assessment of variable x over an interval $\left[\underline{x}; \overline{x}\right]$. Let term l_i corresponds to fuzzy set $\tilde{l}_i$ with membership function $\mu_i(x)$. Linear ordering condition for term-set $\{l_1, l_2, ..., l_K\}$ is written as follows:

$$\begin{cases} \forall x \in \left[\underline{x}, \sup\left(\text{core}(\tilde{l}_i)\right)\right] & : \ \mu_i(x) \geq \mu_{i+1}(x) \\ \exists x \in \left[\underline{x}, \sup\left(\text{core}(\tilde{l}_i)\right)\right] & : \ \mu_i(x) > \mu_{i+1}(x) \\ \forall x \in \left[\inf\left(\text{core}(\tilde{l}_{i+1})\right), \overline{x}\right] & : \ \mu_i(x) \leq \mu_{i+1}(x) \\ \exists x \in \left[\inf\left(\text{core}(\tilde{l}_{i+1})\right), \overline{x}\right] & : \ \mu_i(x) < \mu_{i+1}(x) \end{cases}, \ i = \overline{1, K-1} \tag{7}$$

To preserve the linear ordering of term-set $\{l_1, l_2, ..., l_K\}$ of the variable x over the interval $\left[\underline{x}; \overline{x}\right]$, let us introduce the following constrains on the parameters of membership functions (2):

$$\begin{cases} b_1 = \underline{x} \\ b_K = \overline{x} \\ b_i < b_{i+1}, \ i = \overline{1, K-1} \\ \forall s_i \in (\underline{x}; \overline{x}) \rightarrow s_i \in (b_i; b_{i+1}), \ i = \overline{1, K-1} \end{cases}, \tag{8}$$

where (b_i, c_i) denote the parameters of the Gaussian membership function (2) of the fuzzy set $\tilde{l}_i$, $i = \overline{1, K-1}$;

s_i denotes an abscissa of crossing point of the graphs of membership functions of neighboring fuzzy sets $\tilde{l}_i$ and $\tilde{l}_{i+1}$, $i=\overline{1, K-1}$.

The first two lines in (8) determine that the cores of extreme terms are fixed at the variable boundaries. These conditions protect against C-type problem with interpretability. The third line provides the linear ordering of the cores of fuzzy sets $\left\{\tilde{l}_1, \tilde{l}_2, ..., \tilde{l}_K\right\}$. The first three lines (8) together hold the core of any fuzzy set into range $[\underline{x}; \overline{x}]$. The fourth line (8) demands that any inside-interval crossing point of the neighboring membership functions locates between the cores of these fuzzy sets. The fourth line protects against B-type problem with interpretability.

The crossing points of neighboring fuzzy sets are calculated by the following simple formulas:

$$s_i = \frac{b_i c_{i+1} + b_{i+1} c_i}{c_{i+1} + c_i} \tag{9}$$

$$s_i = \frac{b_i c_{i+1} - b_{i+1} c_i}{c_{i+1} - c_i} \tag{10}$$

One point of pair (9)–(10) locates into the range $(b_i; b_{i+1})$. Let us denote it as q_i, and another point of the pair as v_i. It allows to simplify line 4 in (8) into the following two lines:

$b_1 < q_1 < b_2 < q_2 < ... < q_{k-1} < b_K$;

$v_i \notin (\underline{x}; \overline{x})$.

To protect against D-type problem with interpretability, we fix the minimal level of the height h_i of the intersection of neighboring fuzzy sets $\tilde{l}_i$ and $\tilde{l}_{i+1}$, $i=\overline{1, K-1}$. This height is calculated by formula (2) in case $x=q_i$ as follows:

$$h_i = \max\left(\exp\left(-\frac{1}{2}\left(\frac{b_{i+1}-b_i}{c_i \pm c_{i+1}}\right)^2\right)\right). \tag{11}$$

As a result, system (8) is transformed to the following form:

$$\begin{cases} b_1 = \underline{x} \\ b_K = \overline{x} \\ b_1 < q_1 < b_2 < q_2 < ... < q_{k-1} < b_K \\ v_i \notin (\underline{x}; \overline{x}), \;\; i=\overline{1, K-1} \\ h_i \ge h^*, \;\; i=\overline{1, K-1}, \end{cases} \tag{12}$$

where h^* denotes the minimum permissible height of neighboring fuzzy sets intersection.

4 Structural identification

One of the most important tasks of structural identification is the selection of fuzzy rules from some pre-formed set of candidates. Candidate rules can be formed by an expert or obtained by processing relevant experimental data. Ideally, a fuzzy rule base should be interpretable, compact, and adequate. It is impossible to achieve this in real problems, because in practice it is rational to choose an interpretable rule base with the right balance between compactness and accuracy. A necessary condition for such balance is getting the interpretable rule base from the Pareto front in the coordinates "model compactness – model accuracy".

Let us introduce the following notations: R is a candidate-list of rules and $y = F(R', X)$ is a fuzzy model that ties inputs X with the output y with usage of $R' \subseteq R$ fuzzy rules ($R' \subseteq R$).

A typical approach to the rules selection [15, 18] is a choice of $R' \subseteq R$ that provides:

$$\begin{cases} RMSE(R') \to \min \\ N(R') \le N^* \end{cases} \tag{13}$$

or

$$\begin{cases} N(R') \to \min \\ RMSE(R') \le RMSE^* \end{cases}, \tag{14}$$

where N^* and $RMSE^*$ denotes the permissible levels of compactness and accuracy.

The typical approach forms a large region of feasible solutions, a significant part of which is located far from the Pareto front (Fig. 4a and 4b). This slows down the search for optimal solution that locates at the Pareto front. To reduce the searching space, we proposed a method [19] for choosing the rule base in the vicinity of the Pareto front. We mark this vicinity by the following linear constraint that describes the compactness-accuracy trade-off:

$$RMSE(R') \le k_0 + k_1 \cdot N(R'), \tag{15}$$

where $k_0 > 0$ and $k_1 < 0$ denotes the parameters, choosing which one can form the feasible region in the vicinity of the Pareto front.

Taking into account (15), problems (13)–(14) are transformed to the following form:

$$\begin{cases} RMSE(R') \to \min \\ RMSE(R') \le k_0 + k_1 \cdot N(R') \end{cases}; \tag{16}$$

$$\begin{cases} N(R') \to \min \\ RMSE(R') \le k_0 + k_1 \cdot N(R') \end{cases}. \tag{17}$$

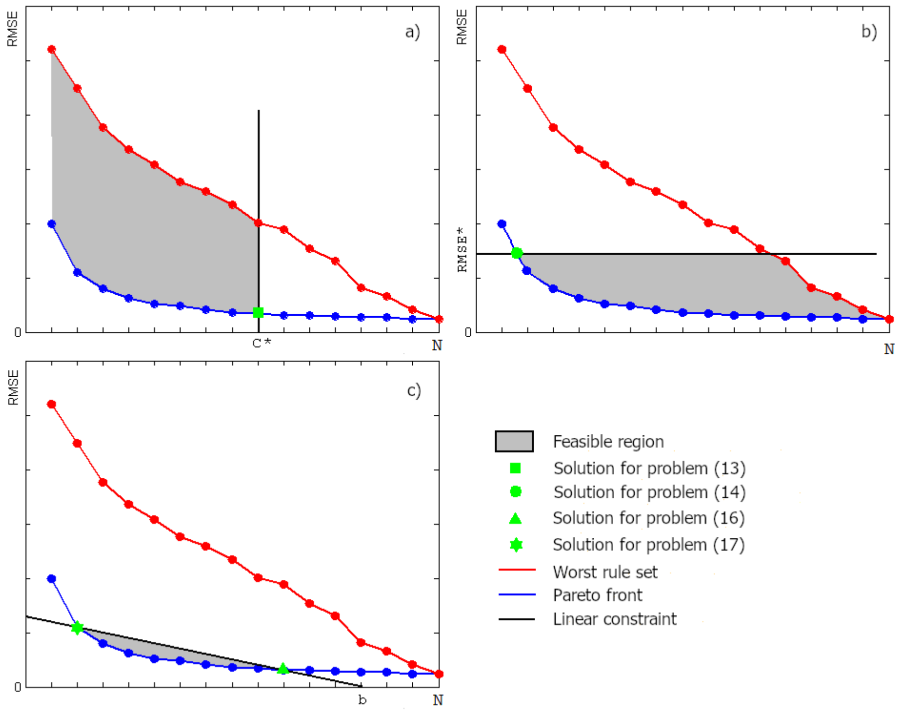

Fig. 4. Feasible region: a) for problem (13); b) for problem (14); c) for problems (16)–(17)

The reduction of the search space for problems (16)–(17) is illustrated on Fig. 4c.

To match the coefficients in the linear constrains in (16)–(17) it is possible to use the endpoints of the Pareto front. They correspond to almost empty and to almost full rule bases. Also, it is easy to find out the upper estimation of the Pareto front (Fig. 5) using a greedy algorithm based on the ideas of the approximate Sahni method for knapsack problem [20]. The computational complexity of this greedy algorithm is quadratic.

The search for optimal solutions is performed by a genetic algorithm with chromosome encoding according to the Pittsburgh scheme [21]. Each chromosome represents a fuzzy base with its own set of rules R'. Each gene of this chromosome corresponds to one candidate rule. A gene has the value 1, if the relative rule is picked up. If the rule is not selected, then the gene has the value 0. For example, the chromosome

$(1, 0, 1, 0, 0, 1, 0)$ encodes a fuzzy knowledge base with three rules numbered 1, 3, and 6. The initial population is randomly generated, but with the inclusion of suboptimal solutions found by the greedy algorithm.

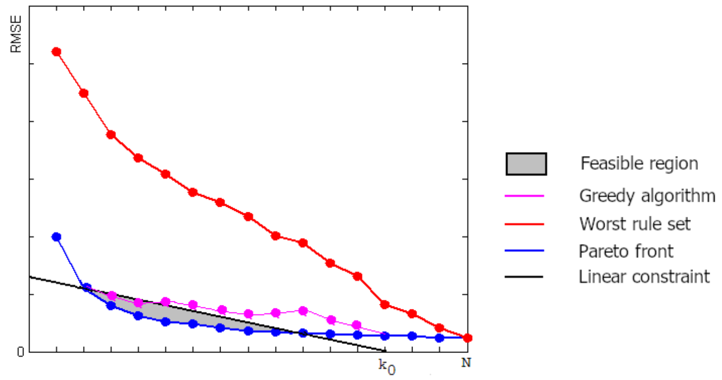

Fig. 5. Upper estimation of Pareto front by a greedy algorithm

After obtaining the optimal set of rules, we carry out the reduction of antecedents. The goal of this procedure is to replace one or more terms in the antecedent with the term *Don't care*. Such a replacement is equivalent to deleting the corresponding terms from the rule antecedents, i.e. transition from long rules to short rules. For example, the transition from rule

$$If\ \left(x_1 = Low\ \ and\ \ x_2 = Average\ and\ \ x_3 = High\right),\ then\ y = Low$$

to rule

$$If\ \ x_2 = Average,\ then\ \ y = Low\,.$$

During this procedure the compactness of the rule base is improving by the criterion A – the total length of all the antecedents. Sometimes this leads to a reduction of the cardinality of term-sets, as well as to the merging the several rules into one.

Antecedents' reduction is also carried by the genetic algorithm. The gene takes the value 1, if the rule uses the term from the best fuzzy knowledge base, and the value 0 if *Don't care* term is used. The larger the number of *Don't care* terms in a rule, the shorter and more compact the fuzzy rule base. Optimization is carried out according to a criterion $RMSE$ with a ban on conflicting rules. Conflicting rules mean rules with the same antecedents. After optimization, a special procedure cleans the rule base – it deletes the rules with only *Don't care* terms and compresses the term-sets with inactive terms.

5 Parametric identification

A parametric identification is tuning the weights of the rules and parameters of membership functions that provide the minimum $RMSE$ on the test sample. From mathematical point of view, the parametric identification is a kind of continuous optimization task. Parametric identification does not change the structure of the model, therefore, the criteria of compactness is not used during optimization. However, unconstrained optimization can lead to an uninterpretable rule base.

We form the vector of controlled variables $P=\left(W, P_1, P_2, ..., P_n, P_y\right)$, where W denotes weights of the rules, $\left(P_1, P_2, ..., P_n\right)$ denotes the parameters of membership functions of terms for the input variables $\left(x_1, x_2, ..., x_n\right)$ and P_y denotes the parameters of membership functions of terms for the output variable y.

For a term-set t $\{l_1, l_2, ..., l_K\}$ of a variable x on an interval $\left[\underline{x}; \overline{x}\right]$, the following parameters are configurable:

$\left(b_2, b_3, ..., b_{K-1}\right)$ – the cores of non-extreme terms $l_2, l_3, ..., l_{K-1}$;

$\left(q_1, q_2, ..., q_{K-1}\right)$ – the crossing points of adjacent fuzzy sets on the interval $\left[\underline{x}; \overline{x}\right]$;

c_1 – the concentration coefficient of the membership function for term l_1.

Thus, for a variable x, the vector of tuning parameters is set up as follows: $P_x=\left(q_1, b_2, q_2, b_3, ..., b_{K-1}, q_{K-1}, c_1\right)$. This allows to use the constraint for interpretability protection in a convenient form (12), in contrast to more cumbersome and subjective expression in [5, 10].

Based on P_x, the concentration coefficients for the membership functions of $\tilde{l}_i$ are calculated as follows: $c_i=\pm c_{i-1}\dfrac{q_{i-1}-b_i}{q_{i-1}-b_{i-1}}$, $i=\overline{2, K}$.

6 Information Technology

The proposed models and methods are implemented in the form of fuzzy identification information technology. Fuzzy identification is carried out in 4 stages, according to the concept of Generation – Selection – Reduction – Tuning.

At the first stage *Fuzzy Rules Generation*, fuzzy rules are generated from experimental data by the direct pass method. This method is based on the ideas of Wang–Mendel method [22], so we only distinguish that the term chosen is not with the maximum grade for any one row of the data sample, but the term with the maximum average membership for all data from the corresponding zones of factor space. If the analogue of the Wang–Mendel method is the fuzzy classification algorithm with a single winner rule, then the analogue of the propose method is the fuzzy classification algorithm with voting rules [23]. Usually fuzzy classifiers with voting rule scheme provide better accuracy, therefore, a similar scheme is chosen to rule generation.

The second stage is *Fuzzy Rules Selection*. It is implemented using a binary genetic algorithm with conditions (16)–(17). Optionally, constraint parameters can be estimated from the learning curves that produced by the greedy algorithm.

The third stage is *Antecedents Reduction*. It is implemented by the same binary genetic algorithm using alternative coding of each rule with *Don't care* terms.

The last stage is *Tuning*. It changes rule weights and membership function using gradient and quasi-Newton optimization methods. For interpretability protection the optimization is carried out with constrains (12).

7 Experiments

Fuzzy identification experiments were carried out for 7 real tasks (Table 1). Each dataset was divided into training and test samples. The training set includes odd lines of source data, and the test set includes even lines.

Table 1. Used data sets from UCI Machine Learning Repository

№	Tasks	Samples	Inputs
1	Auto-MPG	398	9
2	Boston Housing	506	13
3	Combined Cycle Power Plant	9568	4
4	Condition Based Maintenance of Naval Propulsion Plants	11934	16
5	Airfoil Self-Noise	1503	6
6	SkillCraft1 Master Table	3395	20
7	Physicochemical Properties of Protein Tertiary Structure	45730	9

For rules generation, a linguistic partition was used with 3 terms per input and with 5 terms per output. The fuzzy rule generation based on clustering by FCM was chosen as a concurrent. FCM synthesizes a compact but weak interpreted rule base.

Genetic selection of rules was carried out according to (16) with the following coefficients: $k_0 = 0.3$ and $k_1 = -0.0075$. These values produce a fuzzy rule base with less than 40 rules and with *NRMSE* below 0.3. The reduction of antecedents was carried out according to (13) with the following constraint on the total length of the antecedents – $\frac{A}{A_{\max}} \le 0.9$. Quasi-Newton method is used as optimization routine for tuning the rule base. The optimization lasted 150 iterations, and two cases are inspected – with and without interpretability protection.

At first, large rule bases were generated; the number of rules for some tasks exceeded several hundred (Fig. 6). After selection, the number of rules was significantly reduced to a level comparable to FCM bases.

During antecedents reduction the number of rules didn't change, but some rules became short.

To show the effect of antecedents reduction the red bars on Fig. 6 present the following scaled number of rules $Rules' = \frac{A \cdot N}{A_{\max}}$. In some cases it was possible to halve the length of antecedents. For SkillCraft task, FCM was not used due to missing some data in the data set.

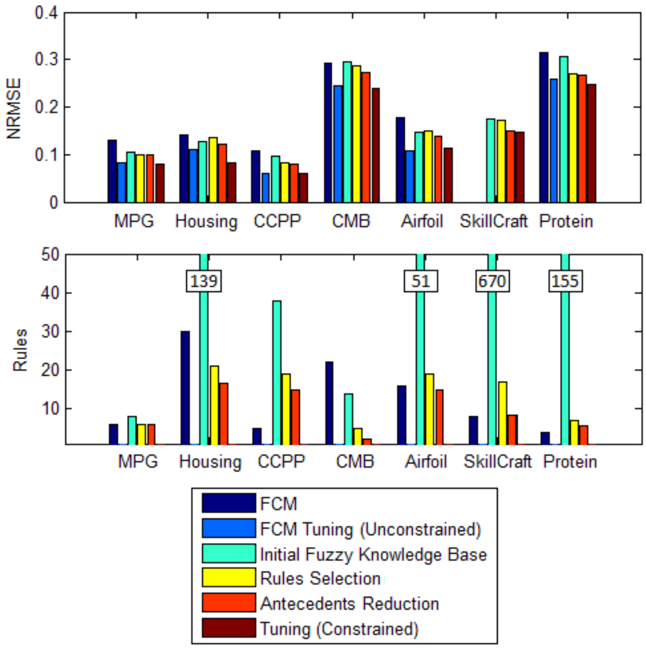

Fig. 6. Results of fuzzy identification using the FCM method and the proposed method

In general, after tuning, the fuzzy rule bases that synthesized by proposed method are a little more accurate than in the case of the FCM method (Fig. 6). Moreover, the proposed method produces not only accurate rule bases, but also interpretable. In addition, the rule bases that were synthesized by the proposed method, have a lot of short rules. For example, for CMB task, the length of the antecedents is reduced by 4 times. Selection of rules and reduction of antecedents often reduce the total number of terms also (Fig. 7). The less terms are in the rule base, the simpler the task of membership functions tuning. A significant reduction in the number of terms took place for tasks with a large number of inputs: SkillCraft, Housing and CMB.

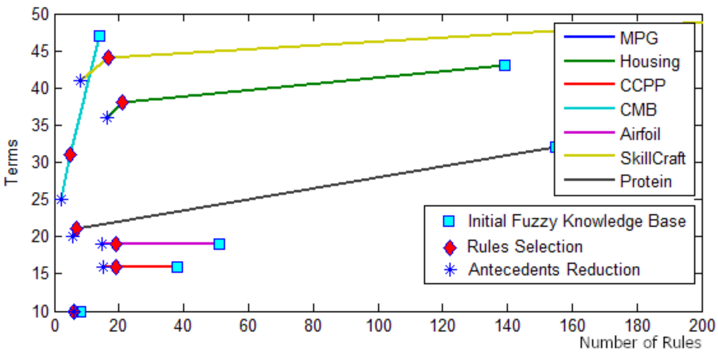

Fig. 7. The traces of compactness of fuzzy rule bases

The experiments showed that the proposed constraints for interpretability protection do not reduce the accuracy (Fig. 8). The difference between accuracy in the case of tuning with the constraints and without them is only a few hundredths for six tasks. This confirms that for a correctly synthesized fuzzy rule base, interpretability does not harm accuracy.

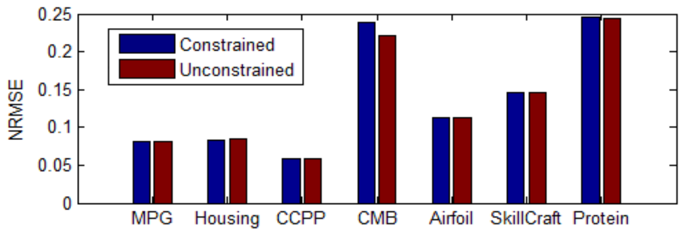

Fig. 8. Accuracy of tuned fuzzy rule bases

8 Conclusion

The proposed information technology of fuzzy identification provides the synthesis of accurate, compact and interpretable rule bases. Information technology includes 4 stages: the generation of reliable fuzzy candidate rules, selection of rules, reduction of antecedents and tuning the membership functions and rule weights. Computer experiments on 7 tasks from the UCI Machine Learning Repository showed, that the proposed information technology synthesizes fuzzy rule bases with accuracy and compactness at the level of competitive technologies. At the same time, unlike competitors, the proposed technology provides the interpretability of fuzzy rule bases.

References

1. Precup R. A.: Survey on industrial applications of fuzzy control. Computers in Industry. **62**, 213-226 (2011). doi:10116/j.compind.2010.10.001.
2. Wong, Bo K., Vincent S. Lai.: A survey of the application of fuzzy set theory in production and operations management: 1998–2009. International Journal of Production Economics. **129**, 157-168 (2011). doi:10.1016/j.ijpe.2010.09.013.
3. Zadeh L.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Syst. Man Cybernet. **3**, 28-44 (1973). doi:10.1109/TSMC.1973.5408575.
4. Rotshtein, A., Rakytyanska, H.: Fuzzy evidence in identification, forecasting and diagnosis. Springer, Heidelberg (2012). doi:10.1007/978-3-642-25786-5.
5. Rotshtein, A. P., Shtovba, S. D.: Influence of defuzzification methods on the rate of tuning a fuzzy model. Cybernetics and Systems Analysis. **38**, 783-789 (2002). doi: 10.1023/A:1021851228684 .
6. Ishibuchi, H., Nakashima T., Murata, T.: Three-objective genetics-based machine learning for linguistic rule extraction. Information Science. **136**, 109–133. (2001). doi: 10.1016/S0020-0255(01)00144-X

7. Cordon O.: A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems. International Journal of Approximate Reasoning. **52**, 894–913 (2011). doi: 10.1016/j.ijar.2011.03.004.
8. Gacto, M. J., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences. **181**, 4340–4360 (2011). doi: 10.1016/j.ins.2011.02.021.
9. Guillaume, S., Charnomordic, B.: Learning interpretable fuzzy inference systems with FisPro. Information Sciences. **181**, 4409–4427 (2011). doi:10.1016/j.ins.2011.03.025.
10. Shtovba S.: Ensuring Accuracy and transparency of Mamdani fuzzy model in learning by experimental data. Journal of Automation and Information Sciences. **39**, 39-52 (2007). doi: 10.1615/JAutomatInfScien.v39.i8.50.
11. Jin, Y., Cao, W., Wu, M., & Yuan, Y.: Accurate fuzzy predictive models through complexity reduction based on decision of needed fuzzy rules. Neurocomputing, **323**, 344-351 (2019). doi: 10.1016/j.neucom.2018.10.010.
12. Tsakiridis, N. L., Theocharis, J.B., Zalidis, G.C.: DECO3RUM: A differential evolution learning approach for generating compact Mamdani fuzzy rule-based models. Expert Systems with Applications. **83**, 257-272 (2017). doi: 10.1016/j.eswa.2017.04.026.
13. Ojha, V., Abraham, A., Snášel, V.: Heuristic design of fuzzy inference systems: A review of three decades of research. Engineering Applications of Artificial Intelligence. **85**, 845-864. (2019). doi: 10.1016/j.engappai.2019.08.010.
14. Mamaghani, A.S., Pedrycz, W. Structural optimization of fuzzy rule-based models: Towards efficient complexity management. Expert Systems with Applications. **152**, 1-12 (2020). doi: 10.1016/j.eswa.2020.113362.
15. Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Transactions on Fuzzy Systems. **3**, 260–270 (1995). doi: 10.1109/91.413232.
16. Liu, B.: Theory and practice of uncertain programming. Studies in Fuzziness and Soft Computing, Vol. 103, Physica-Verlag, (2002). doi: 10.1007/978-3-7908-1781-2.
17. Miller G.A.: The magic number seven plus or minus two: some limits on our capacity for processing information. Psychological Review. **63**, 81–97 (1956). doi: 10.1037/h0043158.
18. Ishibuchi, H., Murata, T., Turksen, I. B.: Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets and Systems. **89**, 135–150 (1997). doi: 10.1016/S0165-0114(96)00098-X.
19. Shtovba, S., Mazurenko, V., Savchuk, D.: Genetic algorithm for selecting rules of the fuzzy knowledge base balanced according to the accuracy and compactness criteria. Scientific Works of Vinnytsia National Technical University. **3** (2012).
20. Martello, S., Toth, P.: Knapsack problems: algorithms and computer implementations. John Wiley & Sons, New York, (1990).
21. Cordon, O., Gomide, E., Herrera, E., Homannc, E., Magdalena, L.: Ten years of genetic fuzzy systems: current framework and new trends. Fuzzy Sets and Systems. **141**, 5–31 (2004). doi: 10.1016/S0165-0114(03)00111-8.
22. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man and Cybernetics. **22**, 1414-1427 (1992). doi: 10.1109/21.199466.
23. Ishibuchi, H., Nakashima, T., Morisawa, T.: Voting in fuzzy rule-based systems for pattern classification problems. Fuzzy Sets and Systems. **103**, 223–238 (1999). doi: 10.1016/S0165-0114(98)00223-1.