Nonlinear Pattern Classification Using Fuzzy System and Hybrid Genetic-Imperialist Algorithm

Krystian Łapa and Krzysztof Cpałka

Abstract An approach proposed in this paper allows to select neuro-fuzzy classifiers taking into account new interpretability criteria. Those criteria are focused not only on complexity of the system, but also on semantics of the rules. The approach uses capabilities of new hybrid population algorithm which is a combination of the genetic algorithm and the imperialist competitive algorithm. This combination allows to select not only the parameters of the neuro-fuzzy system, but also the structure of it. In simulations typical issues of classification were used.

Keywords Neuro-fuzzy classifier · Population-based algorithm · Interpretability

1 Introduction

The process of creation of methods for nonlinear classification is focused mostly on reaching high accuracy. The other important goal is focused on achieve a good clarity and interpretability of classification rules, which allows to better understand considered problem. These both aims are contradictory, so the balance between accuracy and interpretability of classifier is often investigated in the literature (see e.g. [6, 7, 8, 18]).

Nonlinear classification can be based on many types of approaches. Among them, for example, a neuro-fuzzy systems (see e.g. [13, 17]) can be found. In these systems the knowledge in the form of *if...then...* rules is gathered. These rules contain linguistic variables and variables corresponding to fuzzy sets and their parameters. Methods created to increasing interpretability of neuro-fuzzy system rules take an important place in the literature. The interpretability arises not only

K. Łapa (✉) · K. Cpałka
Institute of Computational Intelligence, Częstochowa University of Technology,
Częstochowa, Poland
e-mail: krystian.lapa@iisi.pcz.pl

K. Cpałka
e-mail: krzysztof.cpalka@iisi.pcz.pl

from complexity of the system, but also from semantics of the rules (see e.g. [2, 7, 19]). In this research area it is worth to list methods focused on: (a) Definition and implementation of new criteria of interpretability of fuzzy rules (see e.g. [1, 7]). (b) Appropriate aggregation of these criteria (see e.g. [8, 18]) and using multi-objective methods (see e.g. [1, 18]). (c) Use of population-based algorithms to obtain interpretable systems (see e.g. [12]) etc.

In this paper we propose a new approach which allows to select fuzzy classifiers taking into account different interpretability criteria (including, among others, semantics). This approach is based on hybrid population-based algorithm, which is a fusion between genetic algorithm (see e.g. [17]) and imperialist competitive algorithm (ICA) (see e.g. [3]). The genetic part of the algorithm allows for automatic selection of the structure of neuro-fuzzy system, use of the imperialist algorithm allows to simultaneously select the parameters of these structures. Algorithm ICA was chosen as a part of the proposed hybrid method because: (a) it was created taking inspiration from social evolution, (b) it is a multi-population algorithm and it provides migration and competition of sub-populations in order to improve obtained solutions, (c) it is distinguished by two interesting operators: assimilation and revolution. It is worth to mention that the system presented in our previous paper [14] was used for classification process. Our approach is additionally focused on trade-off between accuracy and interpretability of the system and allows to present accuracy-interpretability dependences using estimated Pareto front (see e.g. [17]).

This paper is organized as follows: in Sect. 2 a description of proposed system and its tuning process for nonlinear classification is presented. In Sect. 3 a interpretability criteria to increase interpretability for neuro-fuzzy systems are shown. The results of simulations are presented in Sect. 4, finally the conclusions are described in Sect. 5.

2   Description of Neuro-Fuzzy System for Classification and Algorithm for Its Tuning

2.1   Description of the System

We consider multi-input, multi-output neuro-fuzzy system mapping $\mathbf{X} \to \mathbf{Y}$, where $\mathbf{X} \subset \mathbf{R}^n$ and $\mathbf{Y} \subset \mathbf{R}^m$. The flexible fuzzy rule base consists of a collection of N fuzzy if-then rules in the form:

$$R^k : \left[\left(\begin{array}{l} \text{IF} \left(\bar{x}_1 \text{ is } A_1^k \right) \Big| w_{k,1}^A \text{AND} \ldots \text{AND} \left(\bar{x}_n \text{ is } A_n^k \right) \Big| w_{k,n}^A \\ \text{THEN} \left(y_1 \text{ is } B_1^k \right) | w_{1,k}^B, \ldots, \left(y_m \text{ is } B_m^k \right) | w_{m,k}^B \end{array} \right) \Big| w_k^{\text{rule}} \right], \qquad (1)$$

where n is a number of inputs, m is a number of outputs, $\bar{\mathbf{x}} = [\bar{x}_1, \ldots, \bar{x}_n] \in \mathbf{X}$, $\mathbf{y} = [y_1, \ldots, y_m] \in \mathbf{Y}$, $A_1^k, \ldots, A_n^k$ are fuzzy sets characterized by membership functions $\mu_{A_i^k}(x_i)$, $i = 1, \ldots, n$, $k = 1, \ldots, N$, $B_1^k, \ldots, B_m^k$ are fuzzy sets characterized by membership functions $\mu_{B_j^k}(y_j)$, $j = 1, \ldots, m$, $k = 1, \ldots, N$, $w_{k,i}^A \in [0,1]$, $i = 1, \ldots, n$, $k = 1, \ldots, N$, are weights of antecedents, $w_{j,k}^B \in [0,1]$, $k = 1, \ldots, N$, $j = 1, \ldots, m$, are weights of consequences, $w_k^{\text{rule}} \in [0,1]$, $k = 1, \ldots, N$, are weights of rules. The flexibility of rule base results from using weights of the antecedences and consequences of the rules. Using of weights need a proper defined aggregation function, which definition can be found in our previous work (see [5]). In logical approach output signal $\bar{y}_j$, $j = 1, \ldots, m$, of the neuro-fuzzy system can be described by the formula:

$$
\bar{y}_j = \frac{\sum_{r=1}^{R} \bar{y}_{j,r}^{\text{def}} \cdot \overset{N}{\underset{k=1}{T^*}} \left\{ S^* \left\{ 1 - \overset{n}{\underset{i=1}{T^*}} \left\{ \mu_{A_i^k}(\bar{x}_i); w_{k,i}^A \right\}, \mu_{B_j^k}\left(\bar{y}_{j,r}^{\text{def}}\right); 1, w_{j,k}^B \right\}; w_k^{\text{rule}} \right\}}{\sum_{r=1}^{R} \overset{N}{\underset{k=1}{T^*}} \left\{ S^* \left\{ 1 - \overset{n}{\underset{i=1}{T^*}} \left\{ \mu_{A_i^k}(\bar{x}_i); w_{k,i}^A \right\}, \mu_{B_j^k}\left(\bar{y}_{j,r}^{\text{def}}\right); 1, w_{j,k}^B \right\}; w_k^{\text{rule}} \right\}}, \quad (2)
$$

where $\bar{y}_{j,r}^{\text{def}}$, $j = 1, \ldots, m$, $r = 1, \ldots, R$, are discretization points, R is a number of discretization points (points in Y, in which the fuzzy inference from the rule base (1) is discretized, resulting from, among others, use of typical for neuro-fuzzy systems defuzzification operations, which allow to determine the real value of the system output signal), $T^*\{\cdot\}$ and $S^*\{\cdot\}$ are weighted triangular norms (see e.g. [17]). In particular, t-norm with weights of arguments can be denoted as follows (see e.g. [17]):

$$
T^*\{a_1, a_2; w_1, w_2\} = T\{1 - w_1 \cdot (1 - a_1), 1 - w_2 \cdot (1 - a_2)\} \overset{\text{e.g.}}{=} (1 - w_1 \cdot (1 - a_1)) \cdot (1 - w_2 \cdot (1 - a_2)),
$$

$$(3)$$

where t-norm $T\{\cdot\}$ is a generalization of the usual two-valued logical conjunction (studied in classical logic), w_1 and $w_2 \in [0,1]$ mean weights of importance of the arguments $a_1, a_2 \in [0,1]$. T-conorm with weights of arguments can be denoted analogously:

$$
S^*\{a_1, a_2; w_1, w_2\} = S\{w_1 \cdot a_1, w_2 \cdot a_2\} \overset{\text{e.g.}}{=} 1 - (1 - w_1 \cdot a_1) \cdot (1 - w_2 \cdot a_2). \quad (4)
$$

For more details see our previous papers, e.g. [17].

2.2 Description of the Tuning Algorithm

The purpose of the algorithm described in this section is an automatic selection of the structure and parameters of the rules in form (1) (number of inputs,

antecedences, consequences, rules) and system in form (2) (discretization points). In this process interpretability criteria defined in Sect. 3 are used. Considered algorithm is a fusion between genetic algorithm (which allows to select the structure of the system) with imperialist competitive algorithm (which allows to select the parameters of the system).

Encoding of parameters and **structure.** The parameters of system (2) are encoded in the following individuals (Pittsburgh approach, in which a single individual of the population encodes the entire neuro-fuzzy system):

$$
\mathbf{X}_{ch}^{\text{par}} = \left\{
\begin{array}{c}
\bar{x}_{1,1}^{A}, \sigma_{1,1}^{A}, \ldots, \bar{x}_{n,1}^{A}, \sigma_{n,1}^{A}, \ldots \bar{x}_{1,Nmax}^{A}, \sigma_{1,Nmax}^{A}, \ldots, \bar{x}_{n,Nmax}^{A}, \sigma_{n,Nmax}^{A}, \\
\bar{y}_{1,1}^{B}, \sigma_{1,1}^{B}, \ldots, \bar{y}_{m,1}^{B}, \sigma_{m,1}^{B}, \ldots \bar{y}_{1,Nmax}^{B}, \sigma_{1,Nmax}^{B}, \ldots, \bar{y}_{m,Nmax}^{B}, \sigma_{m,Nmax}^{B}, \\
w_{1,1}^{A}, \ldots, w_{1,n}^{A}, \ldots, w_{Nmax,1}^{A}, \ldots, w_{Nmax,n}^{A}, w_{1,1}^{B}, \ldots, w_{m,1}^{B}, \ldots, w_{1,Nmax}^{B}, \ldots, w_{m,Nmax}^{B}, \\
w_{1}^{\text{rule}}, \ldots, w_{Nmax}^{\text{rule}}, \bar{y}_{1,1}^{\text{def}}, \ldots, \bar{y}_{1,Rmax}^{\text{def}}, \ldots, \bar{y}_{m,1}^{\text{def}}, \ldots, \bar{y}_{m,Rmax}^{\text{def}}
\end{array}
\right\}
$$
$$
= \left\{ X_{ch,1}^{\text{par}}, \ldots, X_{ch,L}^{\text{par}} \right\},
$$

(5)

where $L = Nmax \cdot (3 \cdot n + 3 \cdot m + 1) + Rmax \cdot m$ is the length of the parameters $\mathbf{X}_{ch}^{\text{par}}, ch = 1, \ldots, \mu$ for the parent population or $ch = 1, \ldots, \lambda$ for the temporary population, $\left\{ \bar{x}_{i,k}^{A}, \sigma_{i,k}^{A} \right\}, i = 1, \ldots, n, k = 1, \ldots, N$, are parameters of Gaussian membership functions $\mu_{A_i^k}(x_i)$ of the input fuzzy sets $A_1^k, \ldots, A_n^k$ (were used in our simulations), $\left\{ \bar{y}_{j,k}^{B}, \sigma_{j,k}^{B} \right\}, k = 1, \ldots, N, j = 1, \ldots, m$, are parameters of Gaussian membership functions $\mu_{B_j^k}(y_j)$ of the output fuzzy sets $B_1^k, \ldots, B_m^k$, $Nmax$ is the maximum number of rules, $Rmax$ is the maximum number of discretization points. The process of selecting the structure of the system is done using additional parameters $\mathbf{X}_{ch}^{\text{str}}$. Their genes take binary values and indicate which rules, antecedents, consequents, inputs, and discretization points are selected. The parameters $\mathbf{X}_{ch}^{\text{str}}$ are given by:

$$
\mathbf{X}_{ch}^{\text{str}} = \left\{
\begin{array}{c}
x_1, \ldots, x_n, A_1^1, \ldots, A_n^1, \ldots, A_1^{Nmax}, \ldots, A_n^{Nmax}, B_1^1, \ldots, B_m^1, \ldots, B_1^{Nmax}, \ldots, B_m^{Nmax}, \\
\text{rule}_1, \ldots, \text{rule}_{Nmax}, \bar{y}_{1,1}^{\text{def}}, \ldots, \bar{y}_{1,Rmax}^{\text{def}}, \ldots, \bar{y}_{m,1}^{\text{def}}, \ldots, \bar{y}_{m,Rmax}^{\text{def}}
\end{array}
\right\}
$$
$$
= \left\{ X_{ch,1}^{\text{str}}, \ldots, X_{ch,L^{\text{str}}}^{\text{str}} \right\},
$$

(6)

where $L^{\text{str}} = Nmax \cdot (n + m + 1) + n + Rmax \cdot m$ is the length of the parameters $\mathbf{X}_{ch}^{\text{str}}$. Their genes indicate which rules $(\text{rule}_k, k = 1, \ldots, Nmax)$, antecedents $(A_i^k, i = 1, \ldots, n, k = 1, \ldots, Nmax)$, consequents $(B_j^k, j = 1, \ldots, m, k = 1, \ldots, Nmax)$, inputs $(\bar{x}_i, i = 1, \ldots, n)$, and discretization points $(\bar{y}^r, r = 1, \ldots, Rmax)$ are taken to the system. We can easily notice that the number of inputs used in the system and encoded in the individual ch can be determined as follows:

$$n_{ch} = \sum_{i=1}^{n} \mathbf{X}_{ch}^{\text{str}} \{x_i\}, \tag{7}$$

where $\mathbf{X}_{ch}^{\text{str}}\{x_i\}$ means parameters of the individual $\mathbf{X}_{ch}^{\text{str}}$ associated with the input x_i (as previously mentioned, if the value of the gene is 1, the associated input is taken into account during work of the system). The number of rules (N_{ch}) used in the system and encoded in the individual *ch* may be determined analogously.

   Evolution of the parameters and structure. The idea of the proposed algorithm is shown in Fig. 1. In **Step 1** of the algorithm, an initial population (in a size of N_{pop}) is created and evaluated (each individual is called *colony*). It is worth to mention that for each *colony* both the real value parameters $\mathbf{X}_{ch}^{\text{par}}$ and the structure parameters $\mathbf{X}_{ch}^{\text{str}}$ are initialized. From initial population N best *colonies* are chosen, and on the basis of each of them *empires* (subpopulations) are created. Best *colony* in every *empire* is called *imperialist*. The remain $N_{pop} - N$ colonies are spread in a specified way among the *empires*. In **Step 2** of algorithm a assimilation and revolution process [which is responsible to tune real parameters of the system (2)] are made. These processes purpose is to move *colonies* toward *imperialist* in their *empires*. Extension of this step relies on using mutation operator from genetic algorithm, which is used to modify the structure of the system (2). The mutation operator has been designed to be proportional to the value of the evaluation function of the *colonies* (best *colony* have 0 % chances to be modified, worst *colony* have 100 % chances to be modified). In **Step 3** an evaluation of the modified *colonies* is made. If a *colony* gets a better value than *imperialist* of its *empire* then the *imperialist* is replaced by this *colony*. It is worth to mention that the fitness function defined in our paper promotes these *colonies* which are characterized, among others, by the simplest structures. In **Step 4** of the algorithm, an *empire* competition (based on the power of empires) takes place. The *empire* which win competition (*empire* selected using roulette wheel method basing on probability calculated by using power of *empires*) gets the weakest *colony* of the weakest *empire*. If *empire* lost all *colonies*, it is removed from the algorithm. In the **Step 5** a

Fig. 1 The basic idea of proposed hybrid algorithm
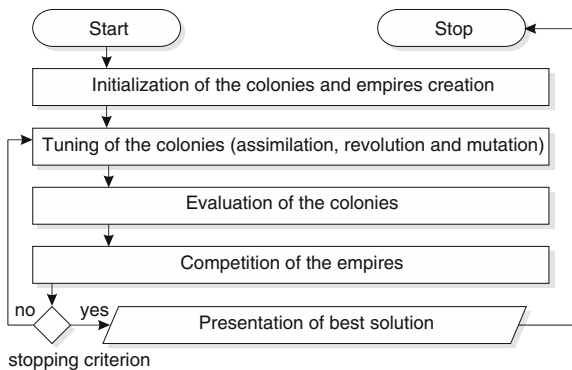
stop condition is checked (e.g. if number of iterations reaches maximum value). If stop condition is met, algorithm ends (and best *colony* of best *empire* is presented), otherwise algorithm goes back to step 2. More details about algorithms used in proposed hybrid genetic-imperialist algorithm can be found e.g. in [3, 17].

Chromosome population evaluation. Each individual $\mathbf{X}_{ch}$ of the parental and temporary populations is represented by a sequence of parameters $\left\{\mathbf{X}_{ch}^{\mathrm{par}}, \mathbf{X}_{ch}^{\mathrm{str}}\right\}$, given by formulas (5) and (6). First parameters take real values, whereas the second ones take integer values from the set $\{0, 1\}$. The system aims to minimize value of the following fitness function:

$$\mathrm{ff}(\mathbf{X}_{ch}) = T^* \big\{ \mathrm{ffaccuracy}(\mathbf{X}_{ch}), \mathrm{ffinterpretability}(\mathbf{X}_{ch}); w_{\mathrm{ffaccuracy}}, w_{\mathrm{ffinterpretability}} \big\}, \tag{8}$$

where $T^*\{\cdot\}$ is the algebraic weighted t-norm (see e.g. [17]), $w_{\mathrm{ffaccuracy}} \in (0, 1]$ is a weight of the component $\mathrm{ffaccuracy}(\mathbf{X}_{ch})$ and $w_{\mathrm{ffinterpretability}}$ is a weight of the component $\mathrm{ffinterpretability}(\mathbf{X}_{ch})$. The component $\mathrm{ffaccuracy}(\mathbf{X}_{ch})$ determines the accuracy of the system (2) (in a form of classification error). The component $\mathrm{ffinterpretability}(\mathbf{X}_{ch})$ determines complexity-based (component $\mathrm{ffint}_A(\mathbf{X}_{ch})$) and semantic-based (components $\mathrm{ffint}_B(\mathbf{X}_{ch}) - \mathrm{ffint}_E(\mathbf{X}_{ch})$) interpretability of the system (2) encoded in the tested individual:

$$\mathrm{ffinterpretability}(\mathbf{X}_{ch}) = $$
$$T^* \left\{ \begin{array}{l} \mathrm{ffint}_A(\mathbf{X}_{ch}), \mathrm{ffint}_B(\mathbf{X}_{ch}), \mathrm{ffint}_C(\mathbf{X}_{ch}), \mathrm{ffint}_D(\mathbf{X}_{ch}), \mathrm{ffint}_E(\mathbf{X}_{ch}) \\ \mathrm{ffint}_E(\mathbf{X}_{ch}), \mathrm{ffint}_F(\mathbf{X}_{ch}), \mathrm{ffint}_G(\mathbf{X}_{ch}); w_{\mathrm{ffintA}}, w_{\mathrm{ffintB}}, w_{\mathrm{ffintC}}, w_{\mathrm{ffintD}}, w_{\mathrm{ffintE}} \end{array} \right\}, \tag{9}$$

where $w_{\mathrm{ffintA}} \in (0, 1]$ denotes weight of the component $\mathrm{ffint}_A(\mathbf{X}_{ch})$, etc. The individual components of the formula (9) are defined in the next section.

3 An Interpretability Criteria for Neuro-Fuzzy System for Nonlinear Classification

In this section a new interpretability criteria for neuro-fuzzy system for nonlinear classification are described. Each criterion is a component of fitness function responsible for interpretability (9) of the system. The criteria are defined as follows:

(a) The component $\mathrm{ffint}_A(\mathbf{X}_{ch})$ determines complexity of the system (2) i.e. a number of reduced elements of the system (rules, input fuzzy sets, output fuzzy sets, inputs, and discretization points) in relation to length of the parameters $\mathbf{X}_{ch}^{\mathrm{str}}$ (it allows to increase complexity-based interpretability):

$$\text{ffint}_A(\mathbf{X}_{ch}) = \frac{\left(\begin{array}{c}\sum_{i=1}^{n} \mathbf{X}_{ch}^{\text{str}}\{x_i\} \cdot \sum_{k=1}^{Nmax} \mathbf{X}_{ch}^{\text{str}}\{\text{rule}_k\} \cdot \mathbf{X}_{ch}^{\text{str}}\{A_i^k\} \\ + \sum_{j=1}^{m} \sum_{k=1}^{Nmax} \mathbf{X}_{ch}^{\text{str}}\{\text{rule}_k\} \cdot \mathbf{X}_{ch}^{\text{str}}\{B_j^k\} + \sum_{j=1}^{m} \sum_{r=1}^{Rmax} \mathbf{X}_{ch}^{\text{str}}\{\bar{y}_{m,r}^{\text{def}}\}\end{array}\right)}{N_{ch} \cdot (n_{ch} + m) + m \cdot Rmax},$$

$$(10)$$

where $\mathbf{X}_{ch}^{\text{str}}\{x_i\}$ means a parameter of $\mathbf{X}_{ch}^{\text{str}}$ associated with the input x_i, etc.

(b) The component $\text{ffint}_B(\mathbf{X}_{ch})$ reduces overlapping of input and output fuzzy sets of the system (2) encoded in the tested individual. This criterion aims to the situation where crossover point between two nearest fuzzy sets have $\mu(x)$ value at c_{ffint} (set to 0.5) and it prevents from situations where nearest fuzzy sets overlaps each other:

$$\text{ffint}_B(\mathbf{X}_{ch}) = \frac{\sum_{i=1}^{n_{ch}} \sum_{k=1}^{\text{noifs}(i)-1} \left(2\left|c_{\text{ffint}c} - \hat{y}_{i,k}^1\right| + \hat{y}_{i,k}^2\right) + \sum_{j=1}^{m_{ch}} \sum_{k=1}^{\text{noofs}(j)-1} \left(2\left|c_{\text{ffint}c} - \hat{y}_{j,k}^1\right| + \hat{y}_{j,k}^2\right)}{2\left(\sum_{i=1}^{n_{ch}}(\text{noifs}(i) - 1) + \sum_{j=1}^{m_{ch}}(\text{noofs}(j) - 1)\right)},$$

$$(11)$$

where $\text{noifs}(i)$ stands for number of active fuzzy sets of i input, $\text{noofs}(j)$ stands for number of active fuzzy sets of j output, $\hat{y}_{i,k}^1, \hat{y}_{i,k}^2$ are $\mu_{A_i^k}(x)$ value of crossover points between two input fuzzy sets and $\hat{y}_{j,k}^1, \hat{y}_{j,k}^2$ are $\mu_{B_j^k}(x)$ value of crossover points between two output fuzzy sets. Those values can be calculated for inputs (and analogically for outputs) as:

$$\hat{y}_{i,k} = \exp\left(-0.5\left(\mathbf{x}_{ch}^{\text{supp}}\{\bar{x}_{i,k}^A\} + \mathbf{x}_{ch}^{\text{supp}}\{\bar{x}_{i,k+1}^A\}\right) / \left(\mathbf{x}_{ch}^{\text{supp}}\{\sigma_{i,k}^A\} \pm \mathbf{x}_{ch}^{\text{supp}}\{\sigma_{i,k+1}^A\}\right)^2\right),$$

$$(12)$$

where $\mathbf{X}_{ch}^{\text{supp}}$ stands for additional set of system parameters [which is build temporary on a base of $\mathbf{X}_{ch}$ from Eq. (11)] with sorted (by position of their centres) list of non-reduced fuzzy sets (for details see [5]).

(c) The component $\text{ffint}_C(\mathbf{X}_{ch})$ increases the integrity of the shape of the input and output fuzzy sets with the inputs and outputs of the system (2) encoded in the tested individual. This criterion aims to achieve fuzzy sets with similar sizes under the same inputs and outputs:

$$\text{ffint}_C(\mathbf{X}_{ch}) = \frac{1}{n_{ch} + m}\left(\begin{array}{c}\sum_{i=1}^{n} \mathbf{X}_{ch}^{\text{str}}\{x_i\} \cdot \sum_{k1=1}^{Nmax} \mathbf{X}_{ch}^{\text{str}}\{\text{rule}_{k1}\} \cdot \text{shx}_{i,k1}(\mathbf{X}_{ch}, i, k1) \\ + \sum_{j=1}^{m} \cdot \sum_{k1=1}^{Nmax} \mathbf{X}_{ch}^{\text{str}}\{\text{rule}_{k1}\} \cdot \text{shy}(\mathbf{X}_{ch}, j, k1)\end{array}\right),$$

$$(13)$$

where $\text{shx}_{i,k1}(\mathbf{X}_{ch})$ (and analogically $\text{shy}_{j,k1}(\mathbf{X}_{ch})$) is a function for calculating proportion between fuzzy sets defined as follows:

$$\text{shx}(\mathbf{X}_{ch}, i, k1) = 1 - \frac{\min\left(\mathbf{X}_{ch}^{\text{par}}\left\{\sigma_{i,k1}^A\right\}, \frac{1}{N_{ch}}\sum_{k2=1}^{N\max}\mathbf{X}_{ch}^{\text{str}}\{\text{rule}_{k2}\}\mathbf{X}_{ch}^{\text{par}}\left\{\sigma_{i,k2}^A\right\}\right)}{\max\left(\mathbf{X}_{ch}^{\text{par}}\left\{\sigma_{i,k1}^A\right\}, \frac{1}{N_{ch}}\sum_{k2=1}^{N\max}\mathbf{X}_{ch}^{\text{str}}\{\text{rule}_{k2}\}\mathbf{X}_{ch}^{\text{par}}\left\{\sigma_{i,k2}^A\right\}\right)},$$

$$(14)$$

where $\mathbf{X}_{ch}^{\text{par}}\left\{\sigma_{i,k}^A\right\}$ stands for a gene of the individual $\mathbf{X}_{ch}^{\text{par}}$ associated with the parameter $\sigma_{i,k}^A$ (the width of the Gaussian function), $\mathbf{X}_{ch}^{\text{par}}\left\{\sigma_{j,k}^B\right\}$ means parameter of the $\mathbf{X}_{ch}^{\text{par}}$ associated with the parameter $\sigma_{j,k}^B$.

(d) The component $\text{ffint}_D(\mathbf{X}_{ch})$ increases complementarity (adjusting position of the input fuzzy sets and data $\bar{x}_{z,i}$) of system (2) encoded in the tested individual:

$$\text{ffint}_D(\mathbf{X}_{ch}) = \frac{1}{Z \cdot n_{ch}}\left(\sum_{z=1}^{Z}\sum_{i=1}^{n}\mathbf{X}_{ch}^{\text{str}}\{x_i\} \cdot \max\left(1, \left|1 - \sum_{k=1}^{N\max}\mathbf{X}_{ch}^{\text{str}}\{\text{rule}_k\} \cdot \mu_{A_i^k}(\bar{x}_{z,i})\right|\right)\right).$$

$$(15)$$

(e) The component $\text{ffint}_E(\mathbf{X}_{ch})$ increases readability of the antecedents and weights (it aims to reach specified values of weights—0, 0.5 and 1) of rules of system (2) encoded in the tested individual:

$$
\begin{aligned}
\text{ffint}_E(\mathbf{X}_{ch}) = 1 - \frac{1}{2N_{ch}}\Bigg(& \frac{1}{n_{ch}}\sum_{k=1}^{N\max}\mathbf{X}_{ch}^{\text{str}}\{\text{rule}_k\}\sum_{i=1}^{n}\mathbf{X}_{ch}^{\text{str}}\{x_i\} \cdot \mu_w\left(w_{i,k}^A\right) \\
& + \sum_{k=1}^{N\max}\mathbf{X}_{ch}^{\text{str}}\{\text{rule}_k\} \cdot \mu_w\left(w_k^{\text{rule}}\right)\Bigg),
\end{aligned}
$$

$$(16)$$

where $\mu_w\left(w_{i,k}^A\right)$ is a function defining congeries around values 0, 0.5 and 1 (in simulations we assumed that $a = 0.25, b = 0.50$ and $c = 0.75$). This function is described as follows:

$$\mu_w(x) = \begin{cases} (a-x)a^{-1} & \text{for} \quad x \geq 0 \quad \text{and} \quad x \leq a \\ (x-a)(b-a)^{-1} & \text{for} \quad x \geq a \quad \text{and} \quad x \leq b \\ (c-x)(c-b)^{-1} & \text{for} \quad x \geq b \quad \text{and} \quad x \leq c \\ (x-c)(1-c)^{-1} & \text{for} \quad x \geq c \quad \text{and} \quad x \leq 1 \end{cases}. \quad (17)$$

4   Simulation Results

In our simulations we considered five typical problems from the field of non-linear classification [15]: (a) wine recognition problem, (b) glass identification problem, (c) Pima Indians diabetes problem, (d) iris classification problem, (d) Wisconsin breast cancer problem. For each problem a 10-fold cross validation was used, and the process was repeated 10 times. Moreover, for each simulation problem a seven

Table 1 Values of the weights of the components ffaccuracy($\mathbf{X}_{ch}$) and ffinterpretability($\mathbf{X}_{ch}$) [see formula (8)] for various variants considered in simulations: case I–case V

Name of the weight	Case I	Case II	Case III	Case IV	Case V	Case VI	Case VII
$w_{\text{ffaccuracy}}(\mathbf{X}_{ch})$	1.00	1.00	1.00	1.00	0.75	0.25	0.10
$w_{\text{ffinterpretability}}(\mathbf{X}_{ch})$	0.10	0.25	0.75	1.00	1.00	1.00	1.00

Table 2 The accuracy (%) of the neuro-fuzzy classifier (2) for learning phase, testing phase and average value of them both for simulation variants case I–case VII

Problem	Sequence	Case							Other authors testing results
		I	II	III	IV	V	VI	VII	
Wine recognition problem	Testing	86.00	84.12	86.24	83.29	82.06	77.65	78.63	85.00–98.61 [10, 16]
	Learning	93.22	94.21	93.95	93.61	93.00	91.70	92.39	
	Average	89.61	89.16	90.09	88.45	87.53	84.67	85.51	
Glass identification problem	Testing	69.76	69.81	68.57	68.03	61.01	48.45	46.31	49.99–74.00 [4, 10, 16]
	Learning	73.13	72.96	72.29	71.07	68.33	63.94	62.46	
	Average	71.45	71.39	70.43	69.55	64.67	56.20	54.38	
Pima Indians diabetes problem	Testing	75.39	72.32	75.26	74.52	66.28	65.37	65.42	45.90–80.00 [4, 11]
	Learning	78.46	78.39	78.02	76.37	70.50	66.60	67.51	
	Average	76.93	75.36	76.64	75.44	68.39	65.98	66.46	
Iris classification problem	Testing	92.78	92.44	92.33	92.89	92.78	85.52	86.86	81.80–97.84 [4, 9]
	Learning	97.48	97.47	97.26	97.53	97.10	97.07	96.39	
	Average	95.13	94.96	94.80	95.21	94.94	91.30	91.62	
Wisconsin breast cancer problem	Testing	96.49	96.47	96.42	96.12	95.97	95.37	95.37	90.00–97.24 [9, 16]
	Learning	97.57	97.67	97.57	97.56	97.34	96.97	96.96	
	Average	97.03	97.07	96.99	96.84	96.65	96.17	96.17	

variants of learning were applied. Each variant had different set of weights of fitness function (8)—see Table 1. Weights of remaining criteria were set as follows: $w_{\text{ffintA}} = 0.50, w_{\text{ffintB}} = 1.00, w_{\text{ffintC}} = 1.00, w_{\text{ffintD}} = 0.20, w_{\text{ffintE}} = 0.50$. The following parameters associated with ICA algorithm were set as follows: number of colonies $N_{pop} = 100$, number of empires $N = 10$, number of iterations to 1000, the revolution rate to 0.3. The mutation probability of genetic operator was set to 0.2.

The conclusions from simulations can be summarized as follows: (a) Using a low value of the weights (such as 0.2) for components of the function (9) caused a reduction in the readability of the relationship between the values of interpretability criteria and the accuracy of the system (see Fig. 3-row 4). (b) Using extreme weight cases (Case I and Case VII) often has no effect on improvement of the system (see Table 2) and it can cause deterioration of the solutions (in comparison to other cases). Solutions founded for these cases may appear under estimated Pareto front
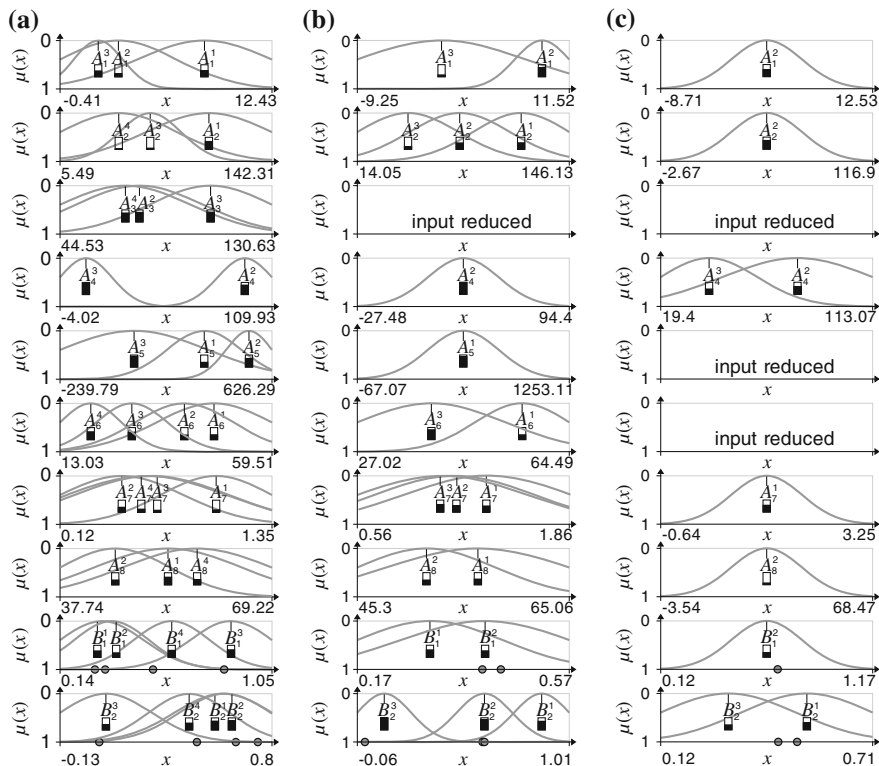
Fig. 2 Example input and output fuzzy sets of the neuro-fuzzy system (2) for the Pima Indians diabetes problem for three various settings of the function (8): **a** case II, **b** case IV, **c** case VI. The position of the discretization points was marked as *black circles*, the weights of the fuzzy sets was marked by *rectangles*. The degree of coverage of the rectangle translates to value of the weight (fully covered rectangle stands for weight 1, and non-covered rectangle stands for weight 0)

(see Fig. 3). (c) Using proposed interpretability criteria allows to achieve semantic clear rules of the system (2) (see Fig. 2). (d) Considering seven cases of weights allowed to determine the estimated Pareto fronts, which make possible to select the interpretability-accuracy trade-off (compromise) by the user (see Fig. 3). (e) Number of reduced inputs and rules depends from the simulation problem (see Fig. 3-row 6 and 7). For example for classification problem (c) system can reduce up to 3 inputs (see Fig. 2) without significantly lost in the accuracy of the system. (f) Achieved results are comparable (in a field of accuracy) with results achieved by other authors using different methods (see Table 2). It should be emphasized that the purpose of the paper was not to achieve the best possible accuracy in comparison with the accuracy obtained by other methods. The purpose of the paper was to increase the legibility of knowledge represented in the form of fuzzy rules with acceptable accuracy of the system. It seems that this objective has been achieved.
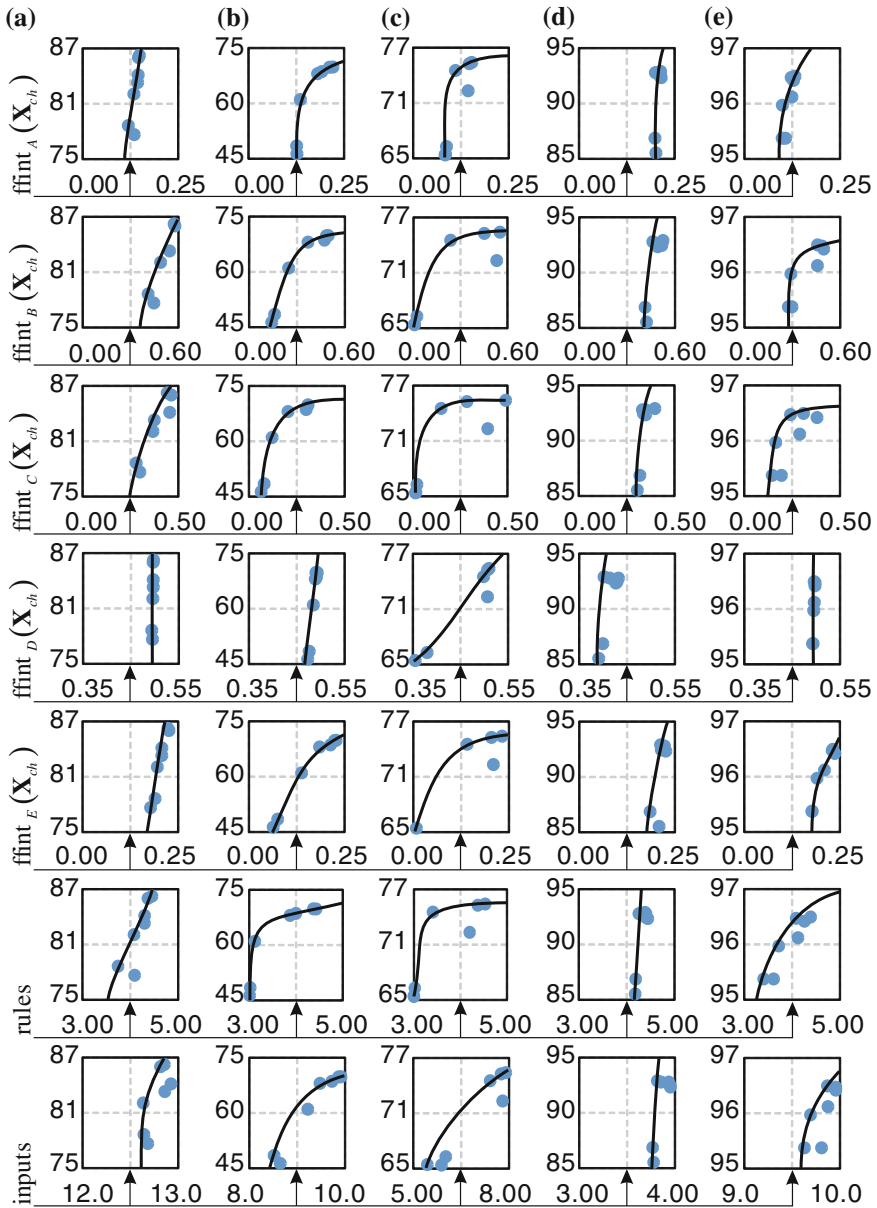
Fig. 3 Dependence between accuracy (%) of neuro-fuzzy classifier (2) (average for learning and testing phase) and values of interpretability components ffint$_A$($\mathbf{X}_{ch}$) − ffint$_G$($\mathbf{X}_{ch}$) for considered variants of the simulations case I–case VII for following simulation problems: **a** wine recognition problem, **b** glass identification problem, **c** Pima Indians diabetes problem, **d** iris classification problem, **e** Wisconsin breast cancer problem

5   Conclusions

In this paper a new approach for non-linear classification was proposed. It is based on possibilities of neuro-fuzzy system and new hybrid genetic-imperialist algorithm. The purpose of this algorithm was to select both the structure and the structure parameters of the estimated classifier with different interpretability criteria taken into consideration. Those criteria are focused not only on the complexity of the system, but also on semantic part of the system. Simulation results performed for typical problems of classification confirmed the correctness of the proposed approach.

References

1. Alonso, J.M.: Embedding HILK in a three-objective evolutionary algorithm with the aim of modeling highly interpretable fuzzy rule-based classifiers. Eur. Centre Soft Comput. 15–20 (2010)
2. Alonso, J.M., Cordon, O., Quirin, A., Magdalena, L.: Analyzing interpretability of fuzzy rule-based systems by means of fuzzy inference-grams. In: 1st World Conference on Soft Computing, pp. 181.1–181.8 (2011)
3. Atashpaz-Gargari, E., Lucas, C.: Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. IEEE Congress on Evolutionary Computation 7, pp. 4661–4666 (2007)
4. Bostanci, B., Bostanci, E.: An evaluation of classification algorithms using Mc Nemars Test. Adv. Intell. Syst. Comput. **201**, 15–26 (2013)
5. Cpałka, K., Łapa, K., Przybył, A., Zalasiński, M.: A new method for designing neuro-fuzzy systems for nonlinear modelling with interpretability aspects. Neurocomputing **135**, 203–217 (2014)
6. Fazzolari, M., Alcalá, R.: Francisco Herrera, A multi-objective evolutionary method for learning granularities based on fuzzy discretization to improve the accuracy-complexity trade-off of fuzzy rule-based classification systems: D-MOFARC algorithm. Appl. Soft Comput. **24**, 470–481 (2014)
7. Gacto, M.J., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. Inf. Sci. **181**, 4340–4360 (2011)
8. Gacto, M.J., Alcalá, R., Herrera, F.: A Multiobjective evolutionary algorithm for tuning fuzzy rule based systems with measures for preserving interpretability. In: Proceedings of the Joint International Fuzzy Systems Association World Congress and the European Society for Fuzzy Logic and Technology Conference (IFSA/EUSFLAT 2009) (2009)
9. Hossen, J., Sayeed, S., Yusof, I., Kalaiarasi, S.M.A.: A framework of modified adaptive fuzzy inference engine (MAFIE) and its application. Int. J. Comput. Inf. Syst. Ind. Manage. Appl. **5**, 662–670 (2013)
10. Jensen, R., Cornelis, C.: Fuzzy-rough nearest neighbour classification. In: Transactions on Rough Sets XIII, pp. 56–72. Springer, Berlin (2011)
11. Kalaiselvi, C., Nasira, G.M.: A novel approach for the diagnosis of diabetes and liver cancer using ANFIS and improved KNN. Res. J. Appl. Sci. Eng. Technol. **8**(2), 243–250 (2014)

12. Kumar, G., Rani, P., Devaraj, C., Victoire, D.: Hybrid ant bee algorithm for fuzzy expert system based sample classification. Comput. Biol. Bioinf. IEEE/ACM Trans. **11**(2), 347–360 (2014)
13. Łapa, K., Przybył, A., Cpałka, K.: A new approach to designing interpretable models of dynamic systems. Lect. Notes Artif. Intell. **7895**, 523–534 (2013)
14. Łapa, K., Zalasiński, M., Cpałka, K.: A new method for designing and complexity reduction of neuro-fuzzy systems for nonlinear modelling. Lect. Notes Artif. Intell. **7894**, 329–344 (2013)
15. Machine Learning Repository [Online]. Available from: https://archive.ics.uci.edu/ml/datasets.html Accessed 6 June 2015
16. Qu, Y., Shang, C., Shen, Q., Parthalain, M., Wei, W.N.: Kernel-based fuzzy-rough nearest neighbour classification. In: Fuzzy Systems (FUZZ), 2011 IEEE International Conference on, pp. 1523–1529 (2011)
17. Rutkowski L., 2008, Computational Intelligence, Springer
18. Shukla, P.K., Tripathi, S.P.: A review on the interpretability-accuracy trade-off in evolutionary multi-objective fuzzy systems (EMOFS). Information **3**, 256–277 (2012)
19. Zalasiński, M., Łapa, K., Cpałka, K.: New algorithm for evolutionary selection of the dynamic signature global features. Lect. Notes Artif. Int. **7895**, 113–121 (2013)