2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Genetic Fuzzy Classifier with Fuzzy Rough Sets for Imprecise Data

Janusz T. Starczewski *Member, IEEE*, Robert K. Nowicki *Member, IEEE*, Bartosz A. Nowak
Institute of Computational Intelligence, Czestochowa University of Technology,
Al. Armii Krajowej 36, 42-200 Czestochowa, Poland
Email: {janusz.starczewski; robert.nowicki; bartosz.nowak}@iisi.pcz.pl

***Abstract*—The main problem addressed in this paper is to handle adequately imprecision of input data by means of a combination of fuzzy methods with the rough set theory. We will make use of fuzzy rough sets derived as rough approximations of fuzzy antecedent sets by non-singleton fuzzy premise sets in a fuzzy classifier. Adaptation of the parameters of this system will be done by the standard genetic algorithm.**

I. Introduction

FUZZY systems comprise the following components: fuzzifiers, bases of knowledge, inference modules, aggregation modules and defuzzifiers. The variety of combinations of methods for inference, aggregation, fuzzification and defuzzification lead to numerous practical designs of neuro-fuzzy systems; however, the fuzzifier is habitually ignored, independently whether imprecision is present in inputs or not.

In this paper, we intend to apply non-singleton fuzzification in order to enrich the structure of neuro-fuzzy systems. The application of the non-singleton fuzzification affects the reduction of sensitivity to changes of input values. Nevertheless, the main motivation to choose the non-singleton fuzzification arises from taking into consideration the imprecision of the measurements or from the use of a fuzzy system in the case of corrupted or noisy data [19], [26], [27].

Generally, fuzzification is a mapping from an input space to a fuzzy set. When we obtain, as a result of fuzzification, a fuzzy set $\mu_{A'}(x)$ defined on some set X rather than a membership grade at single point x, we mean non-singleton fuzzification. Our approach regards non-singleton fuzzification as the mapping to a generalized membership function $\mu_{A'}(x, \bar{x})$ in order to analyze the whole spectrum of fuzzified $\bar{x}$ values ahead of time of reasoning. A need of an a'priori knowledge about the imprecision of inputs is crucial to determine an adequate fuzzification of membership functions.

In [19] (see also [18]), Mouzouris and Mendel have demonstrated that it is possible to interpret the non-singleton fuzzy inference system as a singleton fuzzy system by enclosing fuzzification with antecedents from the rule base. They have restricted themselves to the so called Mamdani fuzzy logic systems in which inference relies on fuzzy conjunctions of antecedents and consequents in the rule base. In this paper, we will extend this idea in two directions. Firstly, we will generalize fuzzification of non-singleton types with the use of rough sets. Secondly, we will equip neuro-fuzzy systems with genuine implications to achieve so called logical-type neuro-fuzzy systems. We will prove that the non-singleton fuzzification in the case of particular logical-type fuzzy system architectures can be implemented using modified antecedent fuzzy sets in the classical singleton fuzzification architectures. Besides, this will allow us to notice unexpected similarities between the standard formulation of fuzzification and fuzzy rough sets.

Although, fuzzy sets [38] and rough sets [23] are semantically different techniques for representation of uncertain data, they can be used together in two known combinations. As rough fuzzy sets they have been used in controllers [4], in classification [31] and in neural architectures for classification with missing data [20], [21]. As fuzzy rough sets [8], [11], [25] they have been applied to dimensionality reduction [12], [13], [32], to neural networks [33], [34], and to neurocomputing [17].

We will choose the concept of fuzzy rough sets combining fuzzy sets with fuzzified rough partition that originates from works of Dubois and Prade [5], [6]. Formally, if Φ is a fuzzy partitioning of universe U, fuzzy sets F_i are its partitions, and A is a fuzzy subset of U, i.e., $A \subseteq U$. The fuzzy rough set is defined as a pair $(\Phi_* A, \Phi^* A)$, where the set $\Phi_* A$ is a Φ–lower approximation of the fuzzy set A, and the set $\Phi^* A$ is its Φ–upper approximation. Accordingly, the membership functions of fuzzy sets $\Phi_* A$ and $\Phi^* A$ are defined as follows:

$$\mu_{\Phi^* A}(F_i) = \sup_{x \in U} \min\left(\mu_{F_i}(x), \mu_A(x)\right), \tag{1}$$

$$\mu_{\Phi_* A}(F_i) = \inf_{x \in U} \max\left(1 - \mu_{F_i}(x), \mu_A(x)\right). \tag{2}$$

The main problem addressed in this paper is to handle adequately imprecision of input data by means of a combination of fuzzy methods with the rough set theory. We will make use of fuzzy rough sets derived as the rough approximations of fuzzy antecedent sets by non-singleton fuzzy premise sets.

Owing to this alternative, formulas (1) and (2) describe the broader and narrower fuzzy sets as a composition of a fuzzification and a fuzzy antecedent set with respect to the marginal t-norm and conorm, i.e. $\min$ and $\max$. Thus the first of these expressions is a particular case of non-singleton fuzzification applied to Mamdani fuzzy logic systems, i.e. relying on conjunction rules. Surprisingly, the second equation realizes the non-singleton fuzzification in logical-type fuzzy systems, relying on strong implications. Therefore, a fuzzy rough set can be viewed as an extension to non-singleton fuzzification. Besides, we will reveal that the same fuzzy rough sets can be used both to the conjunction-type and the logical-type fuzzy systems, which has potential in advance of the flexible compromise fuzzy systems [28]. However,

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

the development of efficient structures of new systems and neuro-fuzzy rough systems, presented in this paper, is not straightforward since upper and lower rough approximations form interval functions.

Our approach differs from the commonly known approaches to non-singleton fuzzification in matter that sets in the rule base and fuzzification sets are no longer ordinary fuzzy sets, but an interpretation of fuzzification in terms of fuzzy rough approximations lead us to an interval fuzzy inference system. The interval uncertainty emerges naturally as a result of the fuzzified rough approximation of fuzzy antecedent sets. Therefore, the fuzzy-rough fuzzification method, we will describe in this paper, is well-grounded both in the fuzzy and the rough set theory. Moreover, this method may be considered technically as a new generation method for type-2 fuzzy sets in fuzzy logic systems [1], [2], [10], [14], [35], [37].

The rest of this paper is ordered as follows: In Section II, preliminaries about fuzzy rough systems applied to classification are delivered. Several original formulae to implement rough fuzzy non-singleton fuzzification in a fuzzy logic system are provided. Section III explains details about genetic optimization of the rough fuzzy system structure. Section IV supplies application examples, and finally Section V draws conclusions.

II. Fuzzy Rough Systems

The problem of embedding non-singleton fuzzification into the antecedent part of a fuzzy logic system has been solved in [22]. In this section, fuzzy rough approach to non-singleton fuzzification in the fuzzy logic system framework will be considered.

A. Fuzzy Rough Sets Forming Non-Singleton Fuzzification

Now, we can apply the fuzzy rough approximation of an antecedent fuzzy set A^k assuming that the fuzzy partitioning Φ is determined by imprecision of input data. This imprecision may be induced by non-singleton fuzzification, such that a premise fuzzy set A' plays a role of a fuzzy partition set F_i in the definition of the fuzzy rough set given by (1) and (2).

The shape of the membership function of A' is specified by a fuzzification method. In common problems, we have an a'priori knowledge about the imprecision of input data, so thus we can assign spreads of a fuzzifying function in each dimension. Therefore, emphasizing that A' is also an explicit function of $\bar{\mathbf{x}}$, we can involve the information about fuzzification by substituting Φ–upper and Φ–lower approximations of A^k, denoted by A^{*k} and A^k_*, for the conventional compatibility between a fuzzy premise A' and a fuzzy antecedent A^k, i.e.,

$$\mu_{A^{*k}}(\bar{\mathbf{x}}) = \sup_{\mathbf{x}\in\mathbf{X}} \min\left(\mu_{A'}(\mathbf{x},\bar{\mathbf{x}}), \mu_{A^k}(\mathbf{x})\right), \tag{3}$$

$$\mu_{A^k_*}(\bar{\mathbf{x}}) = \inf_{\mathbf{x}\in\mathbf{X}} \max\left(1-\mu_{A'}(\mathbf{x},\bar{\mathbf{x}}), \mu_{A^k}(\mathbf{x})\right). \tag{4}$$

Considering non-singleton fuzzification as fuzzy partitioning, we anticipate that (3) has the same form as the standard fuzzification [22] if only the strongest t-norm is employed. The upper approximation together with the lower approximation may be regarded as an extension of traditional non-singleton fuzzification in the conjunction-type (Mamdani) fuzzy systems. Besides, it may seem surprising that if we apply de Morgan laws, the lower approximation (4) is a particular (the strongest) form of fuzzification in the logical-type fuzzy systems based on S- or QL-implications. Although the fuzzy rough lower approximations of antecedents are convergent with using non-singleton fuzzification with either strong or quantum logic implications, the reasoning process is till now based on fuzzy conjunctions.

To go into details, let us assume that $\bar{\mathbf{x}}$ is independent of $\mathbf{x}$. In order to find all the supreme minima in equation (3), we have to vary $\mu_{A'}(\mathbf{x},\bar{\mathbf{x}})$ in the whole spectrum of possible $\bar{\mathbf{x}}$ values. For each $\bar{\mathbf{x}}$, an upper fuzzy rough grade can be produced. In Figure 1a), the construction of upper fuzzy rough grades (indicated by squares) is demonstrated for three exemplary $\bar{\mathbf{x}}$, i.e., $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$ and $\bar{\mathbf{x}}_3$.

Similarly, in order to find all the lowest maxima in equation (4), we have to vary the fuzzy complement of $\mu_{A'}(\mathbf{x},\bar{\mathbf{x}})$ in the whole spectrum of possible $\bar{\mathbf{x}}$ values. The construction of lower fuzzy rough grades is presented in Figure 1b).

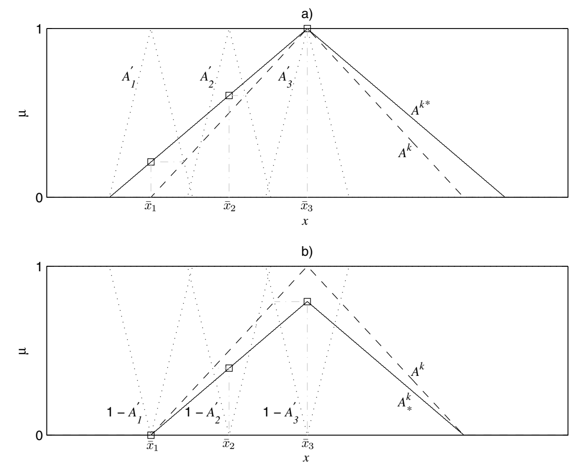

Fig. 1. Construction of (a) — upper and (b) — lower antecedent fuzzy rough sets; A^k — fuzzy antecedent membership function (dashed line), $A'_i(x) = A'(x,\bar{x}_i)$ — examples of non-singleton fuzzy premise membership functions (*dotted lines*) $i=1,2,3$, A^{k*} and A^k_* — upper and lower antecedent fuzzy rough membership function (*dashed lines*)

Figure 2 shows in a concise form examples, in which antecedent fuzzy sets A^k are transformed into pairs of upper and lower membership functions $\left\{A^k_*, A^{k*}\right\}$. Gaussian and triangular fuzzy premises A' are combined with Gaussian and triangular fuzzy antecedents A^k.

It can be immediately observed that upper approximations in fuzzy rough systems are concurrent to fuzzification in the conjunction-type fuzzy systems. Unexpectedly, lower approximations in fuzzy rough systems coincide with fuzzification in the logical-type fuzzy systems.

1) Gaussian fuzzification of Gaussian antecedents: Suppose we have two Gaussian membership functions, $\mu_{A'_i}$ and $\mu_{A_{k,i}}$, and assume an aggregating t-norm to be the algebraic product. The antecedent membership function embedding

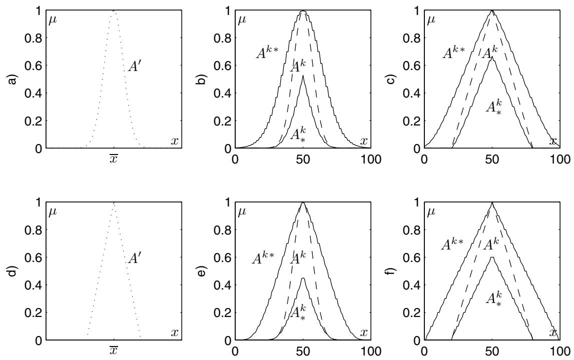

Fig. 2. Examples of antecedent fuzzy rough sets: (a) — Gaussian fuzzification, (b) — Gaussian antecedent (*dashed line*) and its upper and lower fuzzy rough sets for Gaussian fuzzification (*solid lines*), (c) — triangular antecedent (*dashed line*) and its upper and lower fuzzy rough sets for Gaussian fuzzification (*solid lines*); (d) — triangular fuzzification, (e) — Gaussian antecedent (*dashed line*) and its upper and lower fuzzy rough sets for triangular fuzzification (*solid lines*), (f) — triangular antecedent (*dashed line*) and its upper and lower fuzzy rough sets for triangular fuzzification (*solid lines*)

Gaussian fuzzification by $\mu_{A_i'}(\mathbf{x})$ can be evaluated as follows:

$$\mu_{\tilde{A}_k}(x_i') = \sup_{x_i \in X_i} \left(\mu_{A_i'}(x_i', x_i)\mu_{A_{k,i}}(x_i)\right) \tag{5}$$

$$= \sup_{x_i \in X_i} \begin{pmatrix} \exp\left(-\frac{1}{2}\left(\frac{x_i - x_i'}{\sigma_i}\right)^2\right) \\ \cdot \exp\left(-\frac{1}{2}\left(\frac{x_i - c_{k,i}}{\sigma_{k,i}}\right)^2\right) \end{pmatrix} \tag{6}$$

$$= \sup_{x_i \in X_i} \exp \begin{pmatrix} -\frac{1}{2}\left(\frac{x_i - x_i'}{\sigma_i}\right)^2 \\ -\frac{1}{2}\left(\frac{x_i - c_{k,i}}{\sigma_{k,i}}\right)^2 \end{pmatrix}, \tag{7}$$

where $\sigma_i, \sigma_{k,i}$ define spreads of fuzzy sets $A_i', A_{k,i}, c_{k,i}$ defines centre of fuzzy set $A_{k,i}$.

Using differentiation, $\mu_{A_n'}(x_n, x_n')\mu_{A_{k,n}}(x_n)$ attains its supremum at

$$x_i^* = \frac{(\sigma_i)^2 c_{k,i} + (\sigma_{k,i})^2 x_i'}{(\sigma_i)^2 + (\sigma_{k,i})^2}. \tag{8}$$

After simple algebra, we obtain the following membership function which remains Gaussian, i.e.,

$$\mu_{\tilde{A}_k}(x_i') = \exp\left(-\frac{1}{2}\left(\frac{x_i' - c_{k,i}}{\tilde{\sigma}_{k,i}}\right)^2\right), \tag{9}$$

where

$$\tilde{\sigma}_{k,i} = \sqrt{(\sigma_i)^2 + (\sigma_{k,i})^2}. \tag{10}$$

The lower membership function is approximated by two pieces of Gaussians interpolated at points $(c_{k,i} - \tilde{\sigma}_{k,i}, c_{k,i}, c_{k,i} + \tilde{\sigma}_{k,i})$, i.e.,

$$\underline{\mu}_{A_k}(x_i') = \begin{cases} \exp\left(-\frac{1}{2}\left(\frac{x_i - \mathcal{M}_{k,i}}{\tilde{\sigma}_{k,i}}\right)^2\right) & \text{if } x_i' < c_{k,i} \\ \exp\left(-\frac{1}{2}\left(\frac{x_i - \mathcal{N}_{k,i}}{\tilde{\sigma}_{k,i}}\right)^2\right) & \text{otherwise.} \end{cases} \tag{11}$$

where

$$\mathcal{M}_{k,i} = \begin{matrix} c_{k,i} \\ +\tilde{\sigma}_{k,i} \end{matrix} \sqrt{-2\ln\left(-\frac{1 - \frac{\frac{\sigma_i^2}{\sigma_{k,i}^2}}{\left(1+\frac{\sigma_i^2}{\sigma_{k,i}^2}\right)^{\left(\frac{\sigma_{k,i}}{\sigma_i}\right)^2+1}}}{}\right)} \tag{12}$$

$$\mathcal{N}_{k,n} = \begin{matrix} c_{k,i} \\ -\tilde{\sigma}_{k,i} \end{matrix} \sqrt{-2\ln\left(-\frac{1 - \frac{\frac{\sigma_i^2}{\sigma_{k,i}^2}}{\left(1+\frac{\sigma_i^2}{\sigma_{k,i}^2}\right)^{\left(\frac{\sigma_{k,i}}{\sigma_i}\right)^2+1}}}{}\right)} \tag{13}$$

2) Triangular Fuzzification of a Triangular FS: Assuming that both fuzzifying fuzzy sets and sets to be fuzzified are triangular, the possibility and necessity fuzzifications remain triangular. Starting from a symmetric premise membership (fuzzification) function $\mu_{A_i'}(x_i) = \Big/ \min\left(\frac{x_i - x_i' + \Delta_i}{\Delta_i}, \frac{x_i' - x_i + \Delta_i}{\Delta_i}\right)\Big/$ and an asymmetric antecedent membership function $\mu_{A_{k,i}}(x_i) = \Big/ \min\left(\frac{x_i - c_{k,i} + \delta_{k,i}}{\delta_{k,i}}, \frac{c_{k,i} - x_i + \gamma_{k,i}}{\gamma_{k,i}}\right)\Big/$, the possibility function embedding triangular fuzzification is triangular as well,

$$\overline{\mu}_{A_k}(x_i') = \Bigg/ \min\begin{pmatrix} \frac{x_i' - c_{k,i} + \widetilde{\delta}_{k,i}}{\widetilde{\delta}_{k,i}}, \\ \frac{c_{k,i} - x_i' + \widetilde{\gamma}_{k,i}}{\widetilde{\gamma}_{k,i}} \end{pmatrix}\Bigg/, \tag{14}$$

where

$$\widetilde{\delta}_{k,i} = \delta_{k,i} + \Delta_i, \tag{15}$$
$$\widetilde{\gamma}_{k,i} = \gamma_{k,i} + \Delta_i. \tag{16}$$

Further, we recall the triangular necessity function given by

$$\underline{\mu}_{A_k}(x_i') = \Bigg/ \min\begin{pmatrix} \frac{x_i' - c_{k,i} + \delta_{k,i}}{\widetilde{\delta}_{k,i}}, \\ \frac{m_{k,i} + \gamma_{k,i} - x_i'}{\widetilde{\gamma}_{k,i}} \end{pmatrix}\Bigg/. \tag{17}$$

This triangular function has a center $\underline{c}_{k,i}$ calculated by

$$\underline{c}_{k,i} = \frac{\Delta_i\left(\gamma_{k,i} - \delta_{k,i}\right)}{2\Delta_i + \delta_{k,i} + \gamma_{k,i}} + m_{k,i}\ . \tag{18}$$

and a height given by

$$\underline{h}_{k,i} = \frac{\gamma_{k,i} + \delta_{k,i}}{2\Delta_i + \delta_{k,i} + \gamma_{k,i}}\ . \tag{19}$$

B. Triangular Fuzzification of a Symmetric Triangular FS

If we impose symmetry on the antecedent membership function, i.e., $\mu_{A_{k,i}}(x_i) = \Big/ \min\left(\frac{x_i - m_{k,i} + \delta_{k,i}}{\delta_{k,i}}, \frac{c_{k,i} - x_i + \delta_{k,i}}{\delta_{k,i}}\right)\Big/$, the possibility function embedding also symmetric triangular fuzzification is described as follows:

$$\overline{\mu}_{A_k}(x_i') = \Bigg/ \min\begin{pmatrix} \frac{x_i' - c_{k,i} + \widetilde{\delta}_{k,i}}{\widetilde{\delta}_{k,i}}, \\ \frac{c_{k,i} - x_i' + \widetilde{\delta}_{k,i}}{\widetilde{\delta}_{k,i}} \end{pmatrix}\Bigg/, \tag{20}$$

and the necessity presents in its form

$$\underline{\mu}_{A_i}(x_i') = \Bigg/ \min\left(\begin{array}{c}\frac{x_i'-c_{k,i}+\delta_{k,i}}{\widetilde{\delta}_{k,i}},\\ \frac{c_{k,i}-x_i'+\delta_{k,i}}{\widetilde{\delta}_{k,i}}\end{array}\right)\Bigg/ \ , \tag{21}$$

where $\widetilde{\delta}_{k,i}$ is given by (15). This triangular function has a centre at $c_{k,i}$ and a corresponding height given by

$$\underline{h}_{k,i} = \frac{\delta_{k,i}}{\delta_{k,i}+\Delta_i} \ . \tag{22}$$

C. Fuzzy Rough Classification System

In the case of classification, the membership of an object x to a class ω_j (specified by the corresponding consequent) is fuzzy ($\overline{z}_j^k = \mu_{\omega_j}(x)$). Consequently, the rules can be written as [16]

$$\begin{array}{c} R^k\colon \text{IF } x_1 \text{ is } A_1^k \text{ AND } x_2 \text{ is } A_2^k \text{ AND}\ldots \\ \ldots \text{ AND } x_n \text{ is } A_n^k \text{ THEN } \mathbf{x}\in\omega_1(\overline{z}_1^k), \mathbf{x}\in\omega_2(\overline{z}_2^k),\ldots \\ \ldots, \mathbf{x}\in\omega_m(\overline{z}_m^k) \end{array}, \tag{23}$$

where observations x_i are independent variables, $k=1,\ldots,N$ is the number of N rules, and $\overline{z}_j^k$ is the membership degree to the j–th class ω_j according to rule k.

An optimization procedure to obtain the maximal and minimal centroids for a rough (or interval-valued) fuzzy set on the assumption that crisp memberships of objects to classes are given, i.e., the k-th rule consequent that object either belongs to the j-th class or not is binary, $\overline{z}_j^k \in \{0,1\}$, was given in [20], [21].

Theorem 1 (Rough membership of class, [21]): Let us consider the neuro-fuzzy classifier defined by the equation

$$z_j = \frac{\sum_{\substack{k=1 \\ k:\ \overline{z}_j^k=1}}^{N} \widetilde{A}^k}{\sum_{k=1}^{N} \widetilde{A}^k} \tag{24}$$

where $\widetilde{A}^k$ is a rough approximation of a fuzzy set A^k given by its upper and lower approximations, A_*^k and A^{k*}, respectively, and the single-rule membership of object to the j-th class is binary

$$\overline{z}_j^k = \begin{cases} 1 & \text{if } \mathbf{x}\in\omega_j \\ 0 & \text{if } \mathbf{x}\notin\omega_j \end{cases} \tag{25}$$

for all rules $k=1,\ldots,N$ and all classes $j=1,\ldots,m$. Then, the lower and upper approximations of the membership of the current object to class ω_j is given by

$$\overline{z}_{j*} = \frac{\sum\limits_{\substack{k=1 \\ k:\ \overline{z}_j^k=1}}^{N} \mu_{A_{\mathrm{L}}^k}(\mathbf{x})}{\sum\limits_{k=1}^{N} \mu_{A_{\mathrm{L}}^k}(\mathbf{x})} \tag{26}$$

and

$$\overline{z}_j^* = \frac{\sum\limits_{\substack{k=1 \\ k:\ \overline{z}_j^k=1}}^{N} \mu_{A_{\mathrm{U}}^k}(\mathbf{x})}{\sum\limits_{k=1}^{N} \mu_{A_{\mathrm{U}}^k}(\mathbf{x})} \ , \tag{27}$$

where A_{L}^k and A_{U}^k are defined as follows

$$A_{\mathrm{L}}^k = \begin{cases} A_*^k & \text{if } \overline{z}_j^k = 1 \\ A^{k*} & \text{if } \overline{z}_j^k = 0 \end{cases} \tag{28}$$

and

$$A_{\mathrm{U}}^k = \begin{cases} A^{k*} & \text{if } \overline{z}_j^k = 1 \\ A_*^k & \text{if } \overline{z}_j^k = 0 \end{cases} \ . \tag{29}$$

The most useful result concerns the defuzzification in the case of binary memberships of objects to classes. Consequently, the crucial thing is that this result does not require any arrangement of $\overline{z}_j^k$ as the Karnik-Mendel type-reduction for type-2 fuzzy sets does.

Theorem 2 ([36] based on [20]): Having rough approximations, the upper A_{j*}^k and the lower A_j^{k*}, of a binary set $\overline{z}_j^k \in \{0,1\}$ representing the single-rule class membership (25), where k is the index for rules $k=1,\ldots,N$ and j is the index for classes $j=1,\ldots,m$, the lower and upper approximations of the membership of an object to class ω_j is given by

$$\overline{z}_{j*} = \frac{\sum_{k=1}^{N} A_{j*}^k \overline{z}_j^k}{\sum_{k=1}^{N} A_{j*}^k \overline{z}_j^k + \sum_{k=1}^{N} A_j^{k*} \neg\overline{z}_j^k} \ , \tag{30}$$

$$\overline{z}_j^* = \frac{\sum_{k=1}^{N} A_j^{k*} \overline{z}_j^k}{\sum_{k=1}^{N} A_{j*}^k \neg\overline{z}_j^k + \sum_{k=1}^{N} A_j^{k*} \overline{z}_j^k} \ . \tag{31}$$

For the proof see [36].

Observe that N can be extended to N_j, and in such manner, the number of rules does not necessarily has to be the same for each class.

Therefore, the resulting neuro-fuzzy rough architecture for classification can be seen in Figure 3.

The proper interpretation of defuzzificated values is a problem of fuzzy rough classification. Let $\overline{z}_{j*}$ be a lower membership grade of an object x to class ω_j and $\overline{z}_j^*$ be its upper membership grade in the form of equations (26) and (27), respectively. In this case, we may fix two numbers (thresholds) z_{IN} and z_{OUT} such that $1 > z_{\text{IN}} \geq z_{\text{OUT}} > 0$. Consequently, a crisp decision can be made in the following way

$$\begin{cases} x\in\omega_j & \text{if } \overline{z}_{j*} \geq z_{\text{IN}} \text{ and } \overline{z}_j^* > z_{\text{IN}} \\ x\notin\omega_j & \text{if } \overline{z}_{j*} < z_{\text{OUT}} \text{ and } \overline{z}_j^* \leq z_{\text{OUT}} \\ \text{Perhaps } x\in\omega_j & \text{if } z_{\text{IN}} > \overline{z}_{j*} \geq_{\text{OUT}} \text{ and } \overline{z}_j^* > z_{\text{IN}} \\ \text{Perhaps } x\notin\omega_j & \text{if } \overline{z}_{j*} < z_{\text{OUT}} \text{ and } z_{\text{OUT}} < \overline{z}_j^* \leq z_{\text{IN}} \\ \text{undefined} & \text{otherwise.} \end{cases} \tag{32}$$

If we assume for convenience that $z_{\text{IN}} = z_{\text{OUT}} = \frac{1}{2}$, formula (32) takes the following form

$$\begin{cases} x\in\omega_j & \text{if } \overline{z}_{j*} > \frac{1}{2} \text{ and } \overline{z}_j^* > \frac{1}{2} \\ x\notin\omega_j & \text{if } \overline{z}_{j*} < \frac{1}{2} \text{ and } \overline{z}_j^* < \frac{1}{2} \\ \text{undefined} & \text{otherwise.} \end{cases} \tag{33}$$

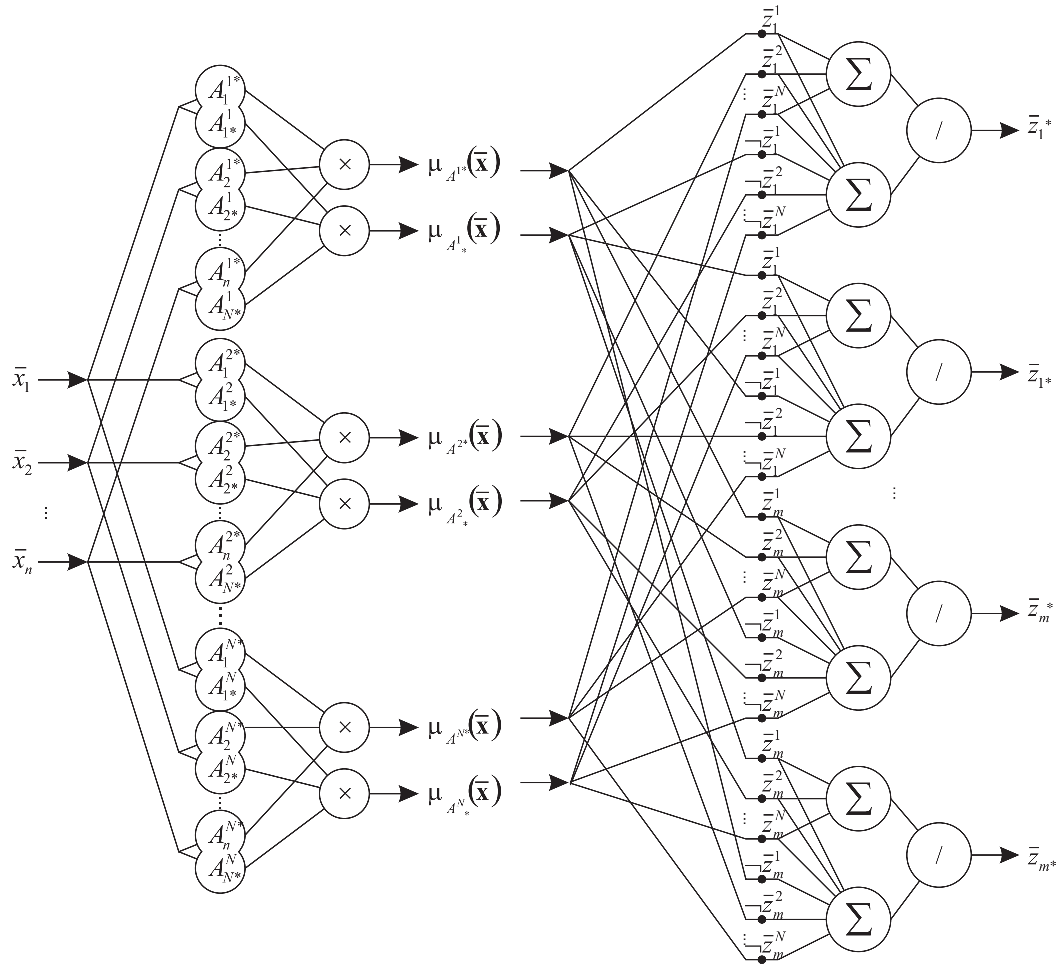

Fig. 3. Fuzzy rough system for classification

III. Genetic Optimization

We have solved the problem of structure optimization by using the standard genetic algorithm with binary coding. The other methods of optimization as evolutionary learning [7], particle swarm optimization [24] have been also considered, and they will be likely applied in the future work. The algorithm has been employed to find values of the antecedent set parameters, i.e. centres $c_{k,i}$ and spreads $\sigma_{k,i}$. The architecture of the classifier, the number of rules, and the scalar consequents (0 or 1) have been fixed a'priori. The length of chromosomes has been steady and depended on the number of rules and input features. Each parameter of the system has been coded in 6 bits. The only original element of the algorithm has been the specific initialization. We have used the PDS-FCM [9] clustering algorithm to place in the first population a single super-genotype. The rest of genotypes have been random. This perhaps surprising strategy has yielded the most accurate results. As expected, the final winner genotype has been the descendant of the super-genotype from the first generation. However, it has been also the result of the nondeterministic tournament selection together with the gentle mutation $(0.2 - 1.0\%)$. Recombination has been performed by two crossing points. In this way, the genetic algorithm with a small size of population (10 genotypes) has allowed to tune the solution obtained from the PDS-FCM algorithm.

IV. Simulations

In all experiments, we have used 3 Rules per each class. Individuals have been coded as numeric values using 6 genes by default and 16 for optdigits and pendigits databases. Populations have consisted of 10 individuals.

We have performed 50 iterations of clustering and genetic algorithm. The mutation level has been equal to 1% for glass, ionosphere, page-blocks, Wisconsin databases and 0.2% for the rest of databases. Tests have been elaborated using 10-folds cross validation. The elaborated fuzzy rough system has been compared against the standard non-singleton fuzzy logic system.

The first training set, Dermatology, contains 34 attributes, 33 of which are linear valued and one of them is nominal. Patients were first evaluated clinically with 12 features. Afterwards, skin samples were taken for the evaluation of 22 histopathological features. The values of the histopathological features are determined by an analysis of the samples under a microscope.

The Glass benchmark is concerned with determining whether the glass was a type of "float" glass or not.

The Ionosphere radar data was collected by a system in Goose Bay. This system consists of a phased array of 16 high-frequency antennas. The targets were free electrons in the ionosphere. Received signals were processed using an autocorrelation function whose arguments are the time of a pulse and the pulse number. There were 17 pulse numbers for the Goose Bay system. Instances in this database are described by 2 attributes per pulse number, corresponding to the complex values returned by the function resulting from the complex electromagnetic signal.

Iris flower is a standard task for classification and pattern recognition studies. 150 instances are described by four features (x_1 — sepal length in cm, x_2 — sepal width in cm, x_3 — petal length in cm, x_4 — petal width in cm). All the flowers are classified into three species: setosa, versicolor, and virginica. Each species is uniformly represented by 50 flowers.

Optical Recognition of Handwritten Digits Data Set (Optdigits) were collected from a total of 43 people. All input attributes are integers in the range 0...16. The last attribute is the class code 0...9

In the Page-blocks classification, 5473 examples came from 54 distinct documents. Each observation concerned one block. All attributes are numeric. Data are in a format readable by C4.5.

The Parkinsons dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals (”name” column). The main aim of the data is to discriminate healthy people from those with PD, according to ”status” column which is set to 0 for healthy and 1 for PD.

The Pen-Based Recognition of Handwritten Digits Data Set (Pendigits) was collected as 250 samples from 44 writers. All input attributes are integers in the range 0...100. The last attribute is the class code 0...9.

The Pima Indians Diabetes data (PID) contain two classes: healthy (500 patterns) and diabetes cases (268 patterns). All patients were females at least 21 years old, of Prima Indian heritage. Each pattern is characterized by 8 attributes such as number of times being pregnant or oral glucose concentration etc.

In the Vowel dataset, nine male speakers uttered two Japanese vowels /ae/ successively. A single utterance by a speaker forms a time series whose length is in the range 7-29 and each point of a time series is of 12 features (12 coefficients).

The Wisconsin Breast Cancer dataset groups medical cases into two classes: Benign — consisted of 458 patterns, and Malignant — 241 patterns. Each pattern is characterized by 10 numerical attributes. The original dataset contains 16 instances with missing values which have been removed in order to make the fuzzification of antecedents able.

The proposed method has been evaluated using two measures. The first one describes how often an sample has been misclassified into each class.

$$\varphi = 1 - \frac{1}{m \cdot M} \sum_{s=1}^{M} \sum_{j=1}^{m} \begin{cases} 1 & \text{if} \begin{pmatrix} \bar{z}^*_{j,s} > 0.5 \\ \wedge \bar{z}_{j,s*} > 0.5 \\ \wedge x_s \in \omega_j \end{pmatrix}, \\ 1 & \text{if} \begin{pmatrix} \bar{z}^*_{j,s} < 0.5 \\ \wedge \bar{z}_{j,s*} < 0.5 \\ \wedge x_s \notin \omega_j \end{pmatrix}, \\ 0 & \text{else,} \end{cases} \tag{34}$$

where m is the number of classes, M is the number of samples, x_s is s-th test sample, ω_j is j-th class, $\bar{z}^*_{j,s}$ and $\bar{z}_{j,s*}$ are respectively upper or lower approximations of membership of s-th test sample to ω_j.

The second criterion for evaluation is a mean of samples that has been neither not classified nor classified to a particular class,

$$\psi = \frac{1}{m \cdot M} \sum_{s=1}^{M} \sum_{j=1}^{m} \begin{cases} 1 & \text{if } \bar{z}^*_{j,s} \leq 0.5 \wedge \bar{z}_{j,s*} \geq 0.5, \\ 0 & \text{else.} \end{cases} \tag{35}$$

The proposed solution has been evaluated with various spreads for input values

$$\sigma_i = \sigma^{coef} * (x_{i,max} - x_{i,min}) * (2\log(2))^{-0.5}, \tag{36}$$

where σ^{coef} is constant coefficient, $x_{i,max}, x_{i,min}$ are respectively maximal and minimal value of attribute x_i. It is worth to mention that $e^{-0.5\left(\frac{1}{(2\log(2))^{-0.5}}\right)^2} = 0.5$.

The simulation has been performed assuming three various levels of input imprecision. It has been reflected in the spread value. The results are showed in Table I. All the results are presented both for the non-singleton neuro-fuzzy classifiers (NF) and the rough neuro-fuzzy classifiers (RNF). In the case of NF systems, the values express the number of misclassified samples (in %) from testing sequences. Thus, the rest of them has been classified correctly. The RNF classifiers have given not two but three or even five possible answers (see eq. (33) and (32)). In the test we have used three-valued version of systems and in the table we have placed the number of misclassified samples and number of samples with the classifier answer marked as ”undefined” (see eq. (33)). For that group samples, accuracy of classification has been deemed as ”uncertain”. The rest of them has been classified correctly. When we focus on misclassified samples, we can observe that in most cases the number of incorrectly classified samples is lower for the fuzzy rough system than for the standard non-singleton fuzzy logic system. Thanks to the third output value, which is available in the RNF system, the classifier can stem the unequivocal answer, in other words, it can answer ”I don't know”. Obviously, it happens at the expense of both incorrect and correct answers.

V. Conclusions

Non-singleton fuzzification serves to handle the imprecision of input's measurements or noisy input data. From the point of view presented in this paper, it is a source of uncertainty for antecedent membership functions in fuzzy rough systems.

TABLE I. CLASSIFICATION ACCURACY

data base	network	measure	σ^{coef} 0.00%	1.25%	3.75%
dermatology	NF	φ	0.033	0.005	0.004
	RNF	φ	0.033	0.004	0.002
		ψ	0.000	0.031	0.106
glass (2 classes)	NF	φ	0.121	0.116	0.135
	RNF	φ	0.121	0.088	0.107
		ψ	0.000	0.014	0.033
ionosphere	NF	φ	0.157	0.137	0.128
	RNF	φ	0.157	0.151	0.066
		ψ	0.000	0.042	0.119
iris	NF	φ	0.282	0.018	0.009
	RNF	φ	0.282	0.000	0.000
		ψ	0.000	0.044	0.102
optdigits	NF	φ	0.014	0.013	0.012
	RNF	φ	0.014	0.007	0.007
		ψ	0.000	0.121	0.224
page-blocks	NF	φ	0.041	0.034	0.030
	RNF	φ	0.041	0.040	0.040
		ψ	0.000	0.000	0.001
parkinsons	NF	φ	0.217	0.113	0.087
	RNF	φ	0.217	0.041	0.021
		ψ	0.000	0.293	0.308
pendigits	NF	φ	0.023	0.023	0.020
	RNF	φ	0.023	0.017	0.008
		ψ	0.000	0.013	0.041
PID	NF	φ	0.316	0.320	0.261
	RNF	φ	0.316	0.305	0.225
		ψ	0.000	0.014	0.111
vowel	NF	φ	0.091	0.091	0.060
	RNF	φ	0.091	0.091	0.076
		ψ	0.000	0.000	0.048
wisconsin	NF	φ	0.043	0.036	0.031
	RNF	φ	0.043	0.030	0.016
		ψ	0.000	0.014	0.042

We have considered a neural realization of the fuzzy rough systems dedicated to classification with the original defuzzification method. Additionally, we have demonstrated that instead of fuzzifying premises, we can exclusively perform required changes in antecedent membership functions.

The simulated classification examples have shown that fuzzy rough systems can even produce no false classification, performing only certain or possible assignments. It seems promising that fuzzy rough systems give uncertain answers rather than wrong answers. Without difficulty, not classified cases can be redirected to a new more particular investigation.

In the real systems, especially such as medical diagnosis, no-classification is more dependable than wrong strict classifications. Our future goal is to optimize the percentage of correct classifications providing that incorrect classification rate is always equal zero. We have also high expectations for applying the rough neuro-fuzzy classifiers in ensemble [15] as well as for employing some form of flexibility in the rough neuro-fuzzy systems [3], [29], [30].

ACKNOWLEDGMENT

The project was funded by the National Science Centre under decision number DEC-2012/05/B/ST6/03620.

REFERENCES

[1] O. Castillo and P. Melin. *Type-2 fuzzy logic: theory and applications*, volume 223 of *Studies in Fuzzines and Soft Computing*. Springer, 2008.

[2] B.-I. Choi and F. C.-H. Rhee. Interval type-2 fuzzy membership function generation methods for pattern recognition. *Information Sciences*, 179:2102–2122, June 2009.

[3] K. Cpałka and L. Rutkowski. Flexible takagi-sugeno fuzzy systems. In *Proceedings of IEEE International Joint Conference on Neural Networks, 2005. IJCNN'05.*, volume 3, pages 1764–1769. IEEE, 2005.

[4] E. Czogala and H. Roderer. On the control of allpass components using conventional, fuzzy and rough fuzzy controllers. In *Proceedings of IEEE International Conference on Fuzzy Systems, International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium*, volume 3, pages 1405–1412, 1995.

[5] D. Dubois and H. Prade. Rough fuzzy sets and fuzzy rough sets. *Internat. J. General Systems*, 17(2-3):191–209, 1990.

[6] D. Dubois and H. Prade. Putting rough sets and fuzzy sets together. In R. Słowiński, editor, *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, pages 203–232. Kluwer, Dordrecht, 1992.

[7] M. Gabryel and L. Rutkowski. Evolutionary learning of mamdani-type neuro-fuzzy system with evolutionary strategies. In L. Rutkowski and et al., editors, *Artificial Intelligence and Soft Computing — ICAISC 2010, Part I*, volume 4029 of *Lecture Notes in Artificial Intelligence*, pages 354–359. Springer-Verlag, Berlin Heidelberg, 2006.

[8] S. Greco, M. Inuiguchi, and R. Słowiński. Fuzzy rough sets and multiple-premise gradual decision rules. *International Journal of Approximate Reasoning*, 41(2):179–211, 2006.

[9] R. J. Hathaway and J. C. Bezdek. Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics*, 31:735–744, 2001.

[10] C. Hwang and F. C.-H. Rhee. Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means. *IEEE Transactions on Fuzzy Systems*, 15:107–120, 2007.

[11] M. Inuiguchi and T. Tanino. New fuzzy rough sets based on certainty qualification. In S. K. Pal, L. Polkowski, and A. Skowron, editors, *Rough-Neural Computing: Techniques for Computing with Words*, pages 277–296. Springer-Verlag, Berlin, Heidelberg, New York, 2004.

[12] R. Jensen and Q. Shen. Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. *IEEE Transactions on Knowledge and Data Engineering*, 16:1457–1471, 2004.

[13] R. Jensen and Q. Shen. Fuzzy-rough sets assisted attribute selection. *IEEE Trans. Fuzzy Syst.*, 15(1):73–89, 2007.

[14] N. N. Karnik and J. M. Mendel. Applications of type-2 fuzzy logic systems to forecasting of time series. *Information Sciences*, 120:98–111, 1999.

[15] M. Korytkowski, L. Rutkowski, and R. Scherer. From ensemble of fuzzy classifiers to single fuzzy rule base classifier. In Leszek Rutkowski, Ryszard Tadeusiewicz, LotfiA. Zadeh, and JacekM. Zurada, editors, *Artificial Intelligence and Soft Computing ICAISC 2008*, volume 5097 of *Lecture Notes in Computer Science*, pages 265–272. Springer Berlin Heidelberg, 2008.

[16] L. I. Kuncheva. *Fuzzy Classifier Design. Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000.

[17] P. Lingras. Fuzzy-rough and rough-fuzzy serial combinations in neurocomputing. *Neurocomputing*, 36(1-4):29–44, 2001.

[18] J. M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall PTR, Upper Saddle River, NJ, 2001.

[19] G. C. Mouzouris and J. M. Mendel. Nonsingleton fuzzy logic systems: theory and application. *IEEE Transactions on Fuzzy Systems*, 5(1):56–71, 1997.

[20] R. Nowicki. On combining neuro-fuzzy architectures with the rough set theory to solve classification problems with incomplete data. *IEEE Trans. Knowl. Data Eng.*, 20(9):1239–1253, 2008.

[21] R. Nowicki. Rough-neuro-fuzzy structures for classification with missing data. *IEEE Trans. Syst., Man, Cybern. B*, 39, 2009.

[22] R. Nowicki and J. T. Starczewski. On non-singleton fuzzification with dcog defuzzification. In L. Rutkowski and et al., editors, *Artificial Intelligence and Soft Computing — ICAISC 2010, Part I*, volume 6113 of *Lecture Notes in Artificial Intelligence*, pages 168–174. Springer-Verlag, Berlin Heidelberg, 2010.

[23] Z. Pawlak. Rough sets. *International Journal of Computer and Information Science*, 11:341–356, 1982.

[24] P. S. Prampero and R. Attux. Magnetic particle swarm optimization. *Journal of Artificial Intelligence and Soft Computing Research*, 2(1):59–72, 2012.

[25] A. M. Radzikowska and E. E. Kerre. A comparative study of fuzzy rough sets. *Fuzzy sets and systems*, 126:137–155, 2002.

[26] D. Rutkowska, R. Nowicki, and L. Rutkowski. Singleton and non-singleton fuzzy systems with nonparametric defuzzification. In *Computational Intelligence and Application*, pages 292–301. Springer-Verlag, 1999.

[27] D. Rutkowska, L. Rutkowski, and R. Nowicki. On processing of noisy data by fuzzy inference neural networks. In *Proceedings of the IASTED International Conference, Signal and Image Processing*, pages 314–318, Nassau, Bahamas, October 1999.

[28] L. Rutkowski. *Flexible neuro-fuzzy systems: structures, learning and performance evaluation*. Kluwer Academic Publishers, 2004.

[29] L. Rutkowski and K. Cpałka. Compromise approach to neuro-fuzzy systems. In Peter Sincak, Jan Vascak, Vladimir Kvasnicka, and Jiri Pospichal, editors, *Intelligent Technologies - Theory and Applications*, volume 76, pages 85–90. IOS Press, 2002.

[30] L. Rutkowski and K. Cpałka. Neuro-fuzzy systems derived from quasi-triangular norms. In *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on*, volume 2, pages 1031–1036 vol.2, July 2004.

[31] M. Sarkar. Rough fuzzy functions in classification. *Fuzzy Sets and Systems*, 132(2):353–369, 2002.

[32] M. Sarkar and B. Yegnanarayana. Rough fuzzy set theoretic approach to evaluate the importance of input features in classification. In *Proc.International Conference on Neural Networks, ICNN-97*, pages 1590–1595, Texas, USA, June 1997. June 9-12.

[33] M. Sarkar and B. Yegnanarayana. Application of fuzzy-rough sets in modular neural networks. In *Proc. IEEE World Congress on Computational Intelligence*, pages 741–746, Alaska, USA, May 1998. may 4-9.

[34] M. Sarkar and B. Yegnanarayana. Fuzzy-rough neural networks for vowel classification. In *Proc. IEEE International Conference on Systems, Man and Cybernetics*, pages 11–14, California, USA, October 1998.

[35] J. T. Starczewski. Efficient triangular type-2 fuzzy logic systems. *International Journal of Approximate Reasoning*, 50:799–811, 2009.

[36] J. T. Starczewski. Generalized uncertain fuzzy logic systems. In *Advanced Concepts in Fuzzy Logic and Systems with Membership Uncertainty*, volume 284 of *Studies in Fuzziness and Soft Computing*, pages 137–179. Springer Berlin Heidelberg, 2013.

[37] Ö. Uncu and I. B. Türkşen. Discrete interval type 2 fuzzy system models using uncertainty in learning parameters. *IEEE Transactions on Fuzzy Systems*, 15(1):90–106, 2007.

[38] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.