

NEW METHODS FOR DESIGNING AND REDUCTION OF NEURO-FUZZY SYSTEMS

Krzysztof Cpałka
IT Institute, Academy of Management, Lodz, Poland
krzysztof.cpalka@kik.pcz.pl

Abstract

In the paper, we propose novel methods for designing and reduction of neuro-fuzzy systems without the deterioration of their accuracy. The reduction and merging algorithms gradually eliminate inputs, rules, antecedents, and the number of discretization points of integrals in the center of area defuzzification method. Our algorithms have been tested using well known classification benchmark.

Key words: algorithm of best global eliminations, algorithm of best local eliminations, consecutive eliminations algorithm, consecutive mergings algorithm, interpretability, logical approach, neuro-fuzzy system

1 Introduction

In literature various neuro-fuzzy systems have been developed (see e.g. [7, 9-11, 13]). They combine the natural language description of fuzzy systems and the learning properties of neural networks. Recently several algorithms have been proposed to increase interpretability and accuracy and decrease complexity of fuzzy rule-based systems. For various methods of designing fuzzy rule-based systems the reader is referred to [1-6, 8].

In the paper we propose new methods for designing flexible neuro-fuzzy systems. The first method, algorithm of consecutive eliminations, is oriented to the most possible simplification of the system structure. The second method, the algorithm of the best local eliminations, is oriented to the most possible simplification of the system structure with simultaneous increase accuracy of the system. The third method, the algorithm of the best global eliminations, is oriented to the searching, across all the parameters of the system, of an element whose reduction is the most advantageous from the point of view of the accuracy. In subsequent stages of our algorithms we reduce number of discretization points, number of inputs, number of rules and number of ante-

cedents. Moreover, in this paper we present consecutive mergings algorithm, oriented on merging of similar input and output fuzzy sets in neuro-fuzzy system. The methods are tested on the Wisconsin breast cancer problem.

2 Description of neuro-fuzzy system

In this paper we consider multi-input, single-output neuro-fuzzy system mapping $\mathbf{X} \rightarrow \mathbf{Y}$, where $\mathbf{X} \subset \mathbf{R}^n$ and $\mathbf{Y} \subset \mathbf{R}$.

The fuzzifier performs a mapping from the observed crisp input space $\mathbf{X} \subset \mathbf{R}^n$ to the fuzzy sets defined in \mathbf{X} . The most commonly used fuzzifier is the singleton fuzzifier which maps $\bar{\mathbf{x}} = [\bar{x}_1, \dots, \bar{x}_n] \in \mathbf{X}$ into a fuzzy set $A' \subseteq \mathbf{X}$ characterized by the membership function:

$$\mu_{A'}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \bar{\mathbf{x}} \\ 0 & \text{if } \mathbf{x} \neq \bar{\mathbf{x}} \end{cases} \quad (1)$$

The fuzzy rule base consists of a collection of N fuzzy IF-THEN rules in the form:

$$R^{(k)} : \left[\begin{array}{l} \text{IF } [x_1 \text{ is } A_1^k] \{w_{1,k}^r\} \text{ AND } \dots \\ \text{AND } [x_n \text{ is } A_n^k] \{w_{n,k}^r\} \\ \text{THEN } y \text{ is } B^k \end{array} \right] \{w_k^{\text{agr}}\}, \quad (2)$$

where $\mathbf{x} = [x_1, \dots, x_n] \in \mathbf{X}$, $y \in \mathbf{Y}$, $A_1^k, A_2^k, \dots, A_n^k$ are fuzzy sets characterized by membership functions $\mu_{A_i^k}(x_i)$, B^k are fuzzy sets characterized by membership functions $\mu_{B^k}(y)$, $k = 1, \dots, N$, $w_{i,k}^r \in [0,1]$, $i = 1, \dots, n$, $k = 1, \dots, N$, are weights in antecedents of the rules, whereas $w_k^{\text{agr}} \in [0,1]$, $k = 1, \dots, N$, are weights in aggregation of the rules.

The fuzzy inference determines a mapping from the fuzzy sets in the input space \mathbf{X} to the fuzzy sets in the output space \mathbf{Y} . Each of N rules (2) determines a fuzzy set $\bar{B}^k \subset \mathbf{Y}$ given by the compositional rule of inference:

$$\bar{B}^k = A' \circ (A^k \rightarrow B^k), \quad (3)$$

where $A^k = A_1^k \times A_2^k \times \dots \times A_n^k$. Fuzzy sets \bar{B}^k , according to the formula (3), are characterized by membership functions expressed by the *sup-star* composition:

$$\mu_{\bar{B}^k}(y) = \sup_{\mathbf{x} \in \mathbf{X}} \left\{ \mu_{A^k}(\mathbf{x})^T * \mu_{A_1^k \times \dots \times A_n^k \rightarrow B^k}(\mathbf{x}, y) \right\}, \quad (4)$$

where $*$ can be any operator in the class of t-norms. It is easily seen that for a crisp input $\bar{\mathbf{x}} \in \mathbf{X}$, i.e. a singleton fuzzifier (1), formula (4) becomes:

$$\begin{aligned} \mu_{\bar{B}^k}(y) &= \mu_{A_1^k \times \dots \times A_n^k \rightarrow B^k}(\bar{\mathbf{x}}, y) \\ &= \mu_{A^k \rightarrow B^k}(\bar{\mathbf{x}}, y), \\ &= I(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(y)) \end{aligned} \quad (5)$$

where $I(\cdot)$ is an “engineering implication” (Mamdani approach) or fuzzy implication [9-11]. The aggregation operator, applied in order to obtain the fuzzy set B' based on fuzzy sets \bar{B}^k , is the t-norm or t-konorm operator, depending on the type of fuzzy implication. In this paper we use the fuzzy implication.

The defuzzifier performs a mapping from a fuzzy set B' to a crisp point \bar{y} in $\mathbf{Y} \subset \mathbf{R}$. Discrete form of the COA (centre of area) method is defined by following formula:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \mu_{B'}(\bar{y}^r)}{\sum_{r=1}^N \mu_{B'}(\bar{y}^r)}, \quad (6)$$

where \bar{y}^r denotes centres of the membership functions $\mu_{B'}(y)$, i.e. for $r = 1, \dots, N$:

$$\mu_{B'}(\bar{y}^r) = \max_{y \in \mathbf{Y}} \{ \mu_{B'}(y) \}. \quad (7)$$

For further investigations we choose neuro-fuzzy systems of a logical type with an Reichenbach S-implication used in formula (5). Moreover, we incorporate weights in antecedents of the rules $w_{i,k}^r \in [0,1]$, $i = 1, \dots, n$, $k = 1, \dots, N$, and weights in aggregation of the rules $w_k^{\text{agr}} \in [0,1]$, $k = 1, \dots, N$. A general architecture of neuro-fuzzy system of the logical type includes certainty

weights to aggregation of rules w_r^{agr} , $r = 1, \dots, N$, and to connectives of antecedents $w_{i,k}^r$, $k = 1, \dots, N$, $i = 1, \dots, n$:

$$\bar{y} = f(\bar{\mathbf{x}}) = \frac{\sum_{r=1}^N \bar{y}^r \cdot \text{agr}_r(\bar{\mathbf{x}}, \bar{y}^r)}{\sum_{r=1}^N \text{agr}_r(\bar{\mathbf{x}}, \bar{y}^r)}, \quad (8)$$

where

$$\begin{aligned} \text{agr}_r(\bar{\mathbf{x}}, \bar{y}^r) &= T_{k=1}^N \{ I_{k,r}(\bar{\mathbf{x}}, \bar{y}^r); w_r^{\text{agr}} \} \\ &= T_{k=1}^N \{ S \{ I_{k,r}(\bar{\mathbf{x}}, \bar{y}^r), 1 - w_r^{\text{agr}} \} \} \end{aligned} \quad (9)$$

and

$$I_{k,r}(\bar{\mathbf{x}}, \bar{y}^r) = S \{ N(\tau_k(\bar{\mathbf{x}})), \mu_{B^k}(\bar{y}^r) \}. \quad (10)$$

Moreover, the firing strength of rules is given by

$$\begin{aligned} \tau_k(\bar{\mathbf{x}}) &= T_{i=1}^n \{ \mu_{A_i^k}(\bar{x}_i); w_{i,k}^r \} = \\ &= T_{i=1}^n \{ S \{ \mu_{A_i^k}(\bar{x}_i), 1 - w_{i,k}^r \} \}. \end{aligned} \quad (11)$$

In this paper, starting with a description (8-11), we develop new methods for design of neuro-fuzzy systems. The methods are based on the concept of the weighted triangular norms [11].

3 Description of neuro-fuzzy system with new formula of discretization

Many neuro-fuzzy architectures developed so far in the literature are based on the formula (6) with the assumption that number of terms in a corresponding formula (6) is equal to the number of rules N . In this paper we relax that assumption and replace formula (6) and (8) by

$$\bar{y} = \frac{\sum_{r=1}^R \bar{y}_r^B \cdot \mu_{B'}(\bar{y}_r^B)}{\sum_{r=1}^R \mu_{B'}(\bar{y}_r^B)} = \frac{\sum_{r=1}^R \bar{y}_r^B \cdot \text{agr}_r(\bar{x}, \bar{y}_r^B)}{\sum_{r=1}^R \text{agr}_r(\bar{x}, \bar{y}_r^B)}, \quad (12)$$

where $R \geq 1$, and $\bar{y}_r^B, r = 1, \dots$, are discretization points of the fuzzy set B' . In this paper we do not equalize number of fuzzy rules (N) with number of discretization points (R) and generally it is possible that $R > N$ or $R < N$ or $R = N$. Thanks to that we obtain a more general formula describing neuro-fuzzy systems and we have much more opportunities, adjusting parameters N and R , to design and reduce such systems. To our best knowledge such approach has not yet been studied in the literature.

4 Algorithms of reduction and merging of neuro-fuzzy systems

Now we will develop new algorithms of reduction and merging of neuro-fuzzy system (12). The algorithms are based on analysis of weights in antecedents of the rules $w_{i,k}^f \in [0,1]$, $i = 1, \dots, n$, $k = 1, \dots, N$, and weights in aggregation of the rules $w_k^{\text{agr}} \in [0,1]$, $k = 1, \dots, N$.

4.1 Algorithm of consecutive eliminations

The algorithm of consecutive eliminations (CEA) is oriented to the most possible simplification of the system structure. Its idea is based on consecutive eliminations of contradicted, non active and unimportant elements of the system starting from discretization points, and next inputs, whole rules and, finally, antecedents of rules. If a specific reduction, e.g. reduction of a concrete input, is acceptable (accuracy of the system is not worse than before the reduction), then the reduction is accepted, otherwise it is cancelled. The flowchart of the algorithm is depicted in Figure 1.

4.2 Algorithm of the best local eliminations

The algorithm of the best local eliminations (ABLE) is oriented to the most possible simplification of the system structure with simultaneous increase accuracy of the system. Its idea is based on finding within each group of parameters (discretization points, inputs, whole rules and antecedents of rules) such element that assures the best reduction from the view point of the system accuracy. If such an element is found, the reduction is performed and

the search is repeated within the same group of parameters, otherwise another group of parameters is analyzed. This idea takes into account the fact that reduction of the next rule may have better influence on the system accuracy than reduction of the current rule. The flowchart of the algorithm is depicted in Figure 2.

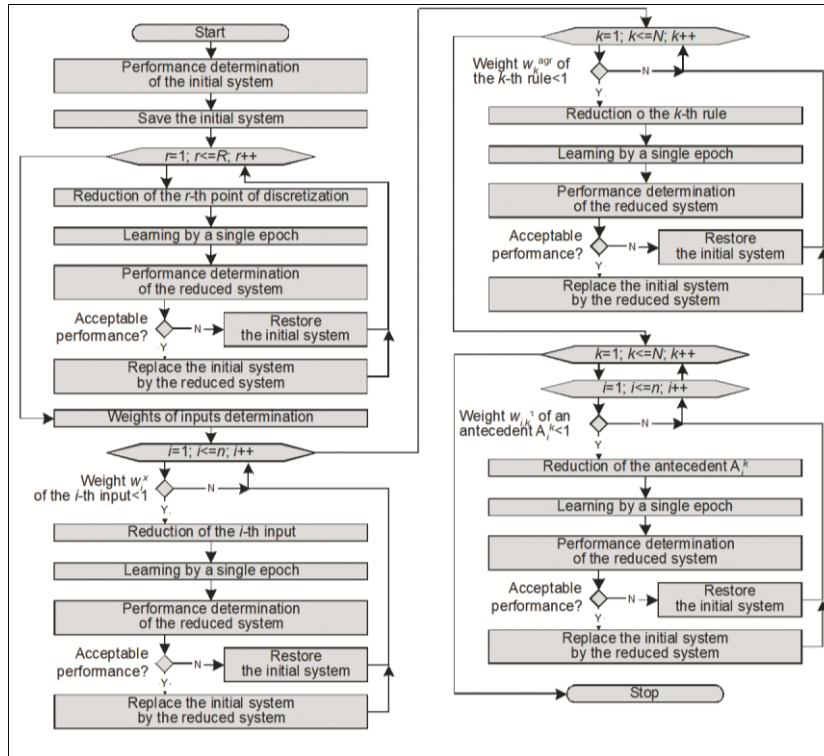


Figure 1. Algorithm of consecutive eliminations (CEA)

4.3 Algorithm of best global eliminations

The best global eliminations algorithm (ABGE), is oriented to the searching, across all the parameters of the system, of an element whose reduction is the most advantageous from the point of view of the accuracy. If there is such an element, its reduction is performed and the search is repeated, if not, then the reduction algorithm is stopped. This idea takes into account the fact that if e.g. an element of the linguistic model causes the biggest mistakes in the system's performance, then the reduction should be started from that element, temporary ignoring elements with less adverse impact on the system. The flowchart of the algorithm is depicted in Figure 3.

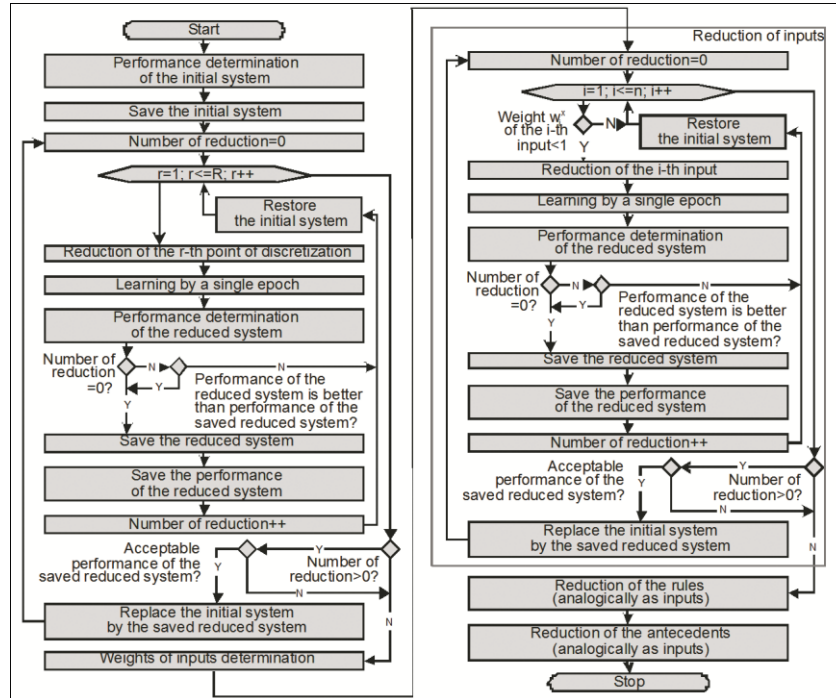


Figure 2. Algorithm of the best local eliminations (ABLE)

The flowcharts in Figures 1-3 comprises 4-parts. First, we determine performance of the initial system (before the reduction process); for example, in a case of the classification we determine a percentage of mistakes of the system. The weights $w_i^x \in [0,1]$, $i = 1, \dots, n$, are calculated using

$$w_i^x = \frac{1}{N} \sum_{k=1}^N w_{i,k}^r . \quad (13)$$

In subsequent stages we reduce number of discretization points, number of inputs, number of rules and number of antecedents.

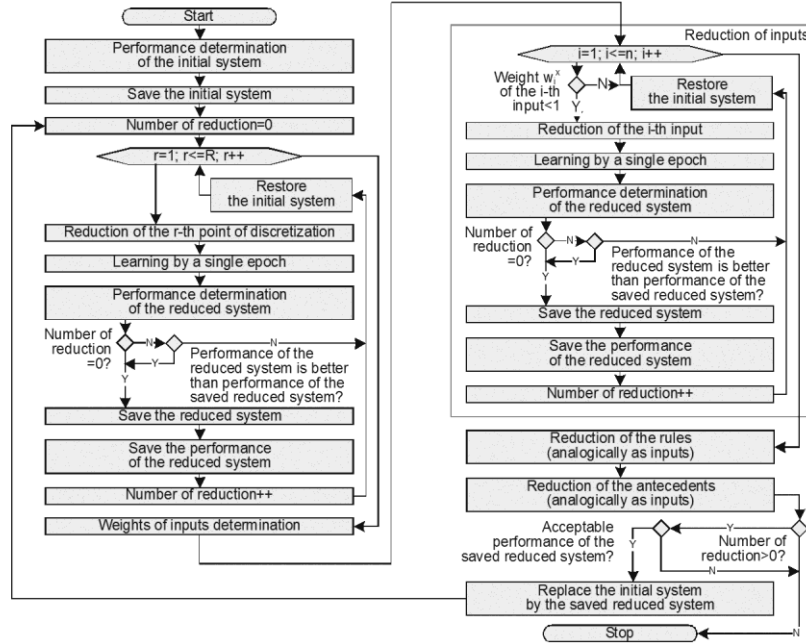


Figure 3. Algorithm of the best global eliminations (ABGE)

4.4 Consecutive mergings algorithm

The consecutive mergings algorithm (CMA) is initialized by the performance determination (number of correctly classified samples) of system (12) before merging. Next, we compare all combinations of input fuzzy sets corresponding to particular input features. The comparison is based on the similarity measure

$$similarity(A_i^{k_1}, A_i^{k_2}) = \frac{\left| \min \left\{ \begin{array}{l} Gauss(\bar{x}_i; \bar{x}_{i,k_1}^A, \sigma_{i,k_1}^A), \\ Gauss(\bar{x}_i; \bar{x}_{i,k_2}^A, \sigma_{i,k_2}^A) \end{array} \right\} \right|}{\left| \max \left\{ \begin{array}{l} Gauss(\bar{x}_i; \bar{x}_{i,k_1}^A, \sigma_{i,k_1}^A), \\ Gauss(\bar{x}_i; \bar{x}_{i,k_2}^A, \sigma_{i,k_2}^A) \end{array} \right\} \right|}, \quad (14)$$

where $|\cdot|$ denotes the cardinality of a fuzzy set, $A_i^{k_1}$ i $A_i^{k_2}$, $i=1, \dots, n$, $k_1=1, \dots, N$, $k_2=1, \dots, N$, are fuzzy sets described by Gaussian membership

functions. Their centers are located in points \bar{x}_{i,k_1}^A i \bar{x}_{i,k_2}^A , and widths are denoted by σ_{i,k_1}^A i σ_{i,k_2}^A .

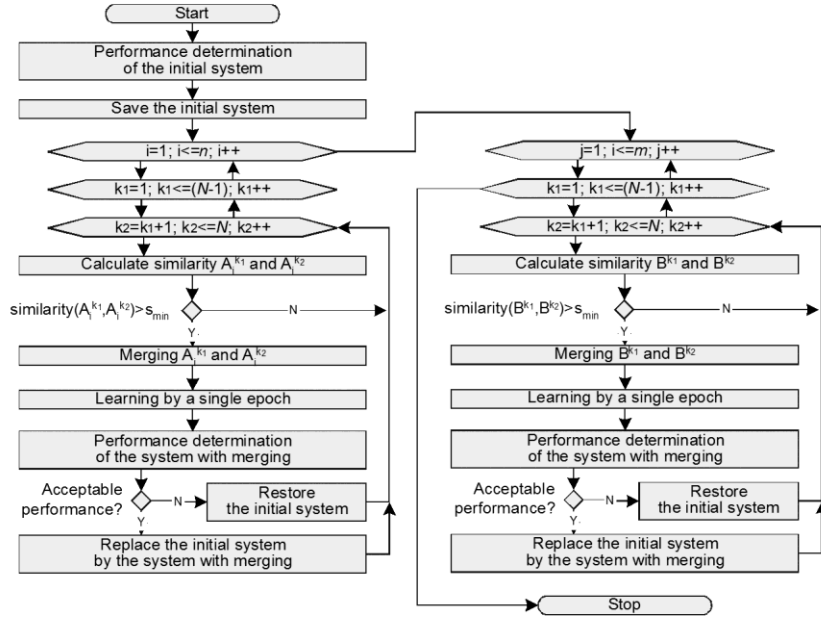


Figure 4. Consecutive mergings algorithm (CMA)

For each pair of input fuzzy sets we determine the value of similarity measure. If that value exceeds the threshold s_{\min} then the input fuzzy sets are merged. More precisely, Gaussian fuzzy sets $A_i^{k_1}$ and $A_i^{k_2}$ are replaced by fuzzy set A_i^k , which is also Gaussian with the center and width given by

$$\bar{x}_{i,k}^A = \frac{\bar{x}_{i,k_1}^A \cdot w_{i,k_1}^r + \bar{x}_{i,k_2}^A \cdot w_{i,k_2}^r}{w_{i,k_1}^r + w_{i,k_2}^r} \quad (15)$$

and

$$\sigma_{i,k}^A = \frac{\sigma_{i,k_1}^A \cdot w_{i,k_1}^r + \sigma_{i,k_2}^A \cdot w_{i,k_2}^r}{w_{i,k_1}^r + w_{i,k_2}^r}. \quad (16)$$

The importance of the antecedent A_i^k being a result of merging antecedents $A_i^{k_1}$ and $A_i^{k_2}$ is described by

$$w_{i,k}^r = \frac{w_{i,k_1}^r + w_{i,k_2}^r}{2}. \quad (17)$$

After each merging, a simplified neuro-fuzzy system is trained by a single epoch and then tested. Testing allows for evaluation of the influence of merging on accuracy of the simplified system. If merging does not worsen the accuracy then the simplified system replaces the previous one. Otherwise, merging is canceled and the initial system is restored. In a similar way we merge the output fuzzy sets B_i^k , $k=1, \dots, N$. The only difference is that the centers and widths of merged fuzzy sets are given by

$$\bar{y}_k^B = \frac{\bar{y}_{k_1}^B + \bar{y}_{k_2}^B}{2} \quad (18)$$

and

$$\sigma_k^B = \frac{\sigma_{k_1}^B + \sigma_{k_2}^B}{2}. \quad (19)$$

5 Simulation results

The neuro-fuzzy system (12) is simulated on the Wisconsin breast cancer problem [12]. In our experiments we apply system (12) with 12 different combinations of number of rules ($N=1, \dots, 4$) and number of discretization points ($R=2, \dots, 4$). We assume that:

- Learning of neuro-fuzzy system (12) is based on the cross-validation procedure 10CV and the momentum back propagation algorithm with $\alpha = 0.25$, $\mu = 0.10$, and 2000 epochs.
- Reduction of neuro-fuzzy system (12) is based on the CEA algorithm, ABLE algorithm, and ABGE algorithm.
- Merging of similar input and output fuzzy sets in neuro-fuzzy system (12) is based on the CMA algorithm with $s_{\min} = 0.5$.

The Wisconsin breast cancer data contains 699 instances (of which 16 instances have a single missing attribute) and each instance is described by nine attributes (clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland

chromatin, normal nucleoli, mitoses). We removed those 16 instances and used the remaining 683 instances. Out of 683 data samples, 444 cases represent benign breast cancer and 239 cases describe malignant breast cancer. The problem is to classify whether a new case is a benign (class 1) or malignant (class 2) type of cancer.

The experimental results for the Wisconsin breast cancer problem are depicted in Table 2 (the best improvement of the system accuracy is indicated in bold), and Figure 5. In Table 2 we show the percentage of mistakes in the learning and testing sequences before and after reduction (reduction and merging). In Figure 5.a we show degree of parameter number reduction [%], in Figure 5.b degree of learning time reduction [%], in Figure 5.c percentage of neuro-fuzzy systems having a particular input (attribute) after the reduction process, in Figure 5.d percentage of inputs (attributes) corresponding to a particular neuro-fuzzy system after the reduction process.

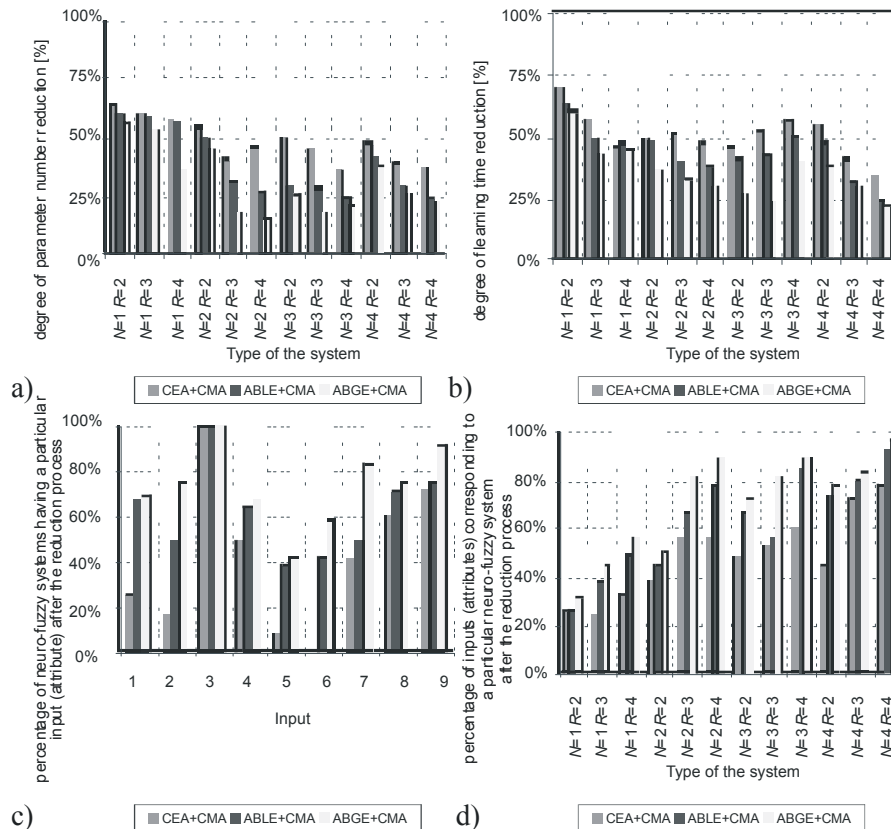


Figure 5. Experimental results for Wisconsin breast cancer problem

Analyzing Table 2 and Figure 5, we come to the following conclusions:

- Accuracy of the system after reduction (reduction and merging) is not worse as before reduction. Moreover, in many cases accuracy after the reduction was higher. The removal of system antecedents and or rules causing bad performance may have been the cause of this result (see Table 2).
- Increasing the number of discretization points resulted in a significant increase of the accuracy of system (12). A great advantage of our method is that it is possible to increase accuracy without losing interpretability (see Table 2).
- The reduction of system (12) resulted in a decrease of parameters and weights which were the subject of training; from little to as much as 60% (see Figure 5.a). Consequently, it shortened learning time from little to 70% (see Figure 5.b).
- The reduction eliminated from few to up to 70% inputs which did not influence the system accuracy (see Figure 5.d). At the same time we see that some inputs were never reduced (see Figure 5.c).
- The reduction resulted in significant improvement of the knowledge transparency in the final system.

Table 2. Accuracy [%] of the system (12) before and after reduction and merging for Wisconsin breast cancer problem

<i>N</i>	Type of the system	<i>R</i>					
		2		3		4	
1	System before reduction	2.96	3.76	2.41	2.46	2.17	2.31
	System reduced (CEA+CMA)	2.91	3.76	2.40	2.46	2.15	2.31
	System reduced (ABLE+CMA)	2.88	3.62	2.35	2.17	2.09	2.17
	System reduced (ABGE+CMA)	2.89	3.47	2.33	2.17	2.12	2.17
2	System before reduction	1.75	2.02	1.77	1.88	1.74	1.88
	System reduced (CEA+CMA)	1.74	2.02	1.77	1.88	1.72	1.88
	System reduced (ABLE+CMA)	1.67	1.73	1.70	1.73	1.69	1.73
	System reduced (ABGE+CMA)	1.72	1.74	1.72	1.73	1.66	1.73
3	System before reduction	1.64	2.17	1.56	1.88	1.33	1.73
	System reduced (CEA+CMA)	1.62	2.17	1.54	1.88	1.32	1.73
	System reduced (ABLE+CMA)	1.59	2.02	1.53	1.73	1.30	1.73
	System reduced (ABGE+CMA)	1.56	2.02	1.49	1.73	1.27	1.73

4	System before reduction	1.36	1.88	1.17	1.73	1.07	1.59
	System reduced (CEA+CMA)	1.36	1.88	1.17	1.73	1.07	1.59
	System reduced (ABLE+CMA)	1.33	1.88	1.14	1.73	1.06	1.59
	System reduced (ABGE+CMA)	1.32	1.88	1.12	1.73	1.06	1.45

6 Final remarks

In the paper we described new methods for designing and reduction of neuro-fuzzy systems. In simulations we reduced the number of discretization points, the number of inputs, the number of rules and the number of antecedents. Moreover, we automatically detected and merged similar input and output fuzzy sets. From simulations it follows that the reduction process of neuro-fuzzy structures based on adjustable weighted triangular norms do not worsen the performance of these structures. Our methods allows to decrease the number of parameters in neuro-fuzzy structures and consequently the learning time.

Acknowledgement

This work was partly supported by the Polish Ministry of Science and Higher Education (Habilitation Project 2007-2010, Polish-Singapore Research Project 2008-2011).

References:

1. Alcalá R., Alcalá-Fdez J. and Herrera F., Aug. 2007, A Proposal for the Genetic Lateral Tuning of Linguistic Fuzzy Systems and Its Interaction With Rule Selection, *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 4, pp. 616-635.
2. Alonso J.M., Cordon O., Guillaume S. and Magdalena L., 2007, Highly Interpretable Linguistic Knowledge Bases Optimization: Genetic Tuning versus Solis-Wetts. Looking for a good interpretability-accuracy trade-off, *In Proc. of the 2007 IEEE Int. Conf. on Fuzzy Systems*, pp. 1-6.
3. Botta A., Lazzarini B., Marcelloni F. and Stefanescu D., 2007, Exploiting Fuzzy Ordering Relations to Preserve Interpretability in Context Adaptation of Fuzzy Systems, *In Proc. of the 2007 IEEE Int. Conf. on Fuzzy Systems*, pp. 1-6.
4. Casillas J., Cordon O., Herrera F. and Magdalena L., 2003, *Interpretability Issues in Fuzzy Modeling*, Springer.
5. Cpalka K., 2009, A New Method for Design and Reduction of Neuro Fuzzy Classification Systems, *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 701–714.

6. Cpałka K., Rutkowski L., 2006, *A New Method for Designing and Reduction of Neuro-Fuzzy Systems*, 2006 IEEE International Conference on Fuzzy Systems, IEEE World Congress on Computational Intelligence, Vancouver, BC, Canada.
7. Czogała E., Łęski J., 2000, *Fuzzy and Neuro-Fuzzy Intelligent Systems*, Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York.
8. Pizzileo B. and Kang Li, 2007, A New Fast Algorithm for Fuzzy Rule Selection, *In Proc. of the 2007 IEEE Int. Conf. on Fuzzy Systems*, pp. 1-6.
9. Rutkowski L., 2004, *Flexible Neuro-Fuzzy Systems*. Kluwer Academic Publishers.
10. Rutkowski L. and Cpałka K., Feb 2005, Designing and learning of adjustable quasi triangular norms with applications to neuro fuzzy systems, *IEEE Trans. on Fuzzy Systems*, vol. 13, pp. 140 151.
11. Rutkowski L., Cpałka K., May 2003, Flexible neuro-fuzzy systems, *IEEE Transactions on Neural Networks*, vol. 14, pp. 554 574.
12. UCI respository of machine learning databases, Available online: <http://ftp.ics.uci.edu/pub/machine-learning-databases/>.
13. Yager R. R., Filev D. P., 1994, *Essentials of fuzzy modelling and control*. John Wiley & Sons.