A Modified TSK Network and Its Application in Classification

Linh Tran Hoai

Hanoi University of Technology, 1 Dai Co Viet str.,
Hanoi, Vietnam

evnbk@hn.vnn.vn

Abstract

The paper presents a modified structure of Takaga-Sugeno-Kang (TSK) network with a fully automated building and learning algorithm. The modification has resulted in a great reduction of nonlinear parameters of the network (almost three times). The modified network can be initiated using Gustafson-Kessel clustering algorithm. After initiation all parameters are further fine-tuned by an gradient learning algorithm. With the proposed method of building and learning modified TSK network, users can easily generate automatically an effective TSK network for practical problems without needing deep knowledge of the fuzzy reasoning theory. As a numerical experiment, the solution has been tested in the problem of gas recognition as a fuzzy reasoning system with very high accuracy.
Conclusions: The modified TSK network is a new fuzzy reasoning system which is more effective than the classical one. The simpler structure leads to shorter time of parameters adaptation.
Keywords: Modified TSK Network, Artificial Nose, Fuzzy Rules Number Generation.

I. Introduction

The neuro-fuzzy reasoning systems are now applied widely in almost every domain of technical problems, such as: approximation, classification, model identification or process control [2,5]. Typically the fuzzy system contains many of the so called inference rules of reasoning, whose answers are then aggregated and defuzzified to give the final answer of the system. The classical TSK network belongs to these systems, which is characterized by the lack of defuzzification block [1,2]. The general structure of TSK network is presented on Fig. 1. The network can be characterized by triple (*N*; *M*; *K*), where *N* - the dimension of input vectors, *M* - the number of inference rules and *K* - the output signals. Each inference rule has a vector form:

$$if \quad \mathbf{x} \ \ is \ \ \mathbf{A}_i \ \ then \ \ y = f_i(\mathbf{x}) \qquad (1)$$

In the case of many rules, when a new input vector **x** is presented as the input of the network, the *i*th rule produces its answer $y = f_i(\mathbf{x})$ with a so called fire strength $w_i(\mathbf{x})$. This fire strength should be maximum 1 if the input **x** is at the center $\mathbf{A}_i$ of the rule ($w_i(\mathbf{x}) \approx 1$ when $\|\mathbf{x} - \mathbf{A}_i\| \to 0$) and should decrease when **x** tends away from $\mathbf{A}_i$ ($w_i(\mathbf{x}) \approx 0$ when $\|\mathbf{x} - \mathbf{A}_i\| \to \infty$). The answer of each rule in a TSK network has a linear form

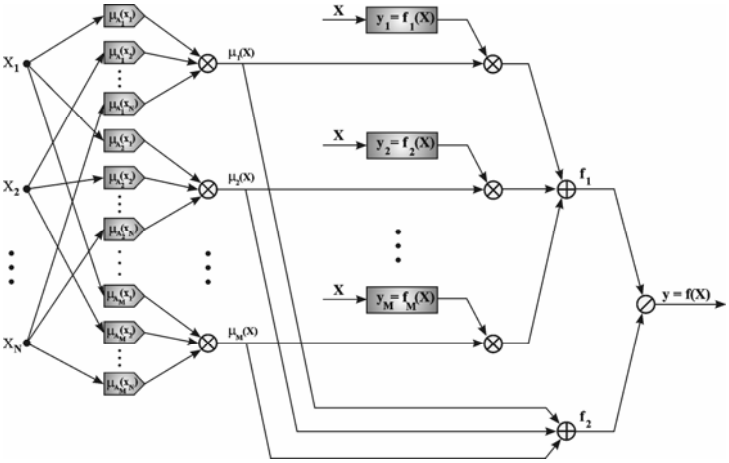

Fig. 1. *The structure of classical TSK network*

$$y_i = f_i(x) = \sum_{i=1}^{N} p_{ij} x_j + p_{i0} \tag{2}$$

The premise "*if* $\mathbf{x}$ *is* $\mathbf{A}_i$" is implemented as the fuzzifier, producing the fire strength, taken here in the generalized bell function form [2,4,6]

$$\mu_A(x) = \prod_{i=1}^{N} \mu_A(x_i) = \prod_{i=1}^{N} \frac{1}{1 + \left(\frac{x_i - c_i}{\sigma_i} \right)^{2b_i}} \tag{3}$$

where c_i, σ_i and b_i are the center, width and exponent coefficient, respectively, of the function for *i*th input variable x_i. Finally, at the existence of *M* rules, the neuro-fuzzy TSK system output signal $y(\mathbf{x})$, upon the excitation by the vector $\mathbf{x}$, is described by the equation [2]

$$y(x) = \frac{\sum_{k=1}^{M} \mu_A^{(k)}(\mathbf{x}) \cdot f_k(\mathbf{x})}{\sum_{r=1}^{M} \mu_A^{(r)}(\mathbf{x})} = \frac{\sum_{k=1}^{M} \left[\prod_{j=1}^{N} \mu_A^{(k)}(x_j) \right] \cdot \left[p_{k0} + \sum_{j=1}^{N} p_{kj} x_j \right]}{\sum_{r=1}^{M} \left[\prod_{j=1}^{N} \mu_A^{(r)}(x_j) \right]} \tag{4}$$

The adjusted parameters of the system are the nonlinear parameters (c_i, σ_i, b_i) of the fuzzifier functions and the linear parameters (weights p_{ij}) of TSK functions. The network structure corresponding to the equation (4) is referred as the neuro-fuzzy TSK network [1,2]. The difficult problem of learning TSK network is its complexity in the cases of multi-input vectors, which is the most popular ones. In traditional model with *N* inputs, 1 output and *M* fuzzy rules, the number of linear parameters equals $M \cdot (N+1)$. At the same time each fuzzifier for every input signal in every rule is characterized by 3 nonlinear coefficients c_i, σ_i and b_i, i.e. the network has $3 \cdot M \cdot N$ nonlinear parameters to be adapted in the learning process. In this paper we propose a simplified network with fewer parameters and new method of calculating the fire strength of fuzzy rules, which has similar effectiveness of operation in learning and testing modes.

II. Modified structure and learning algorithm of TSK network

In classical model, the fire strength of the *r*-th fuzzy rule is depended on the distance between the input vector and the center $\mathbf{c}^{(r)}$ of the rule, a vector of adaptive widths $\sigma^{(r)}$ and a vector of exponent coefficients $\mathbf{b}^{(r)}$ for all input signals. These coefficients are needed to scale the data in each dimension, but at the same time they also complicate the structure of network and increase the number of nonlinear parameters. To reduce the number of these parameters we have used another distance measure, which combines all the scaling in one. This measure is defined in general by [4,5,6]

$$d^2\left(\mathbf{x},\mathbf{c}^{(r)}\right)=\left(\mathbf{x}-\mathbf{c}^{(r)}\right)^T\cdot\mathbf{A}_r\cdot\left(\mathbf{x}-\mathbf{c}^{(r)}\right) \tag{5}$$

where $\mathbf{A}_\mathrm{r}$ is a positive, symetric, scaling matrix. It can be easily recognized that (5) is an extension of euclidian measure, where $\mathbf{A}_r=\mathbf{I}$. With such measure, the new fire strength of the *r*-th rule can be calculated using the relation containing only one global pair of width σ_r and exponent b_r coefficients common for all *N* variables. The scaling matrix $\mathbf{A}_\mathrm{r}$ fulfills the role of scaling for each dimension. The modified fuzzifier function is then defined by

$$\mu_A^{(r)}(\mathbf{x})=\frac{1}{1+\left(\dfrac{d\left(\mathbf{x},\mathbf{c}^{(r)}\right)}{\sigma_r}\right)^{2b_r}} \tag{6}$$

The modified model of TSK network is presented on Fig. 2. It has only $M\cdot(N+2)$ nonlinear adapted $\mathbf{c}^{(i)}$, σ_i and b_i for i=1,..., M (the matrixes $\mathbf{A}_\mathrm{i}$ can be deterministically calculated on the basis of the locations of centers). This means a significant reduction of number of nonlinear parameters (instead of $3\cdot M\cdot N$). For this new defined measure, we have proposed an algorithm of automatic building TSK network including the structure identification (setting up the number of inference rules), parameter initialization (start values for nonlinear parameters $\mathbf{c}^{(i)}$, $\mathbf{A}_\mathrm{i}$, σ_i and b_i and the linear parameters for TSK functions).

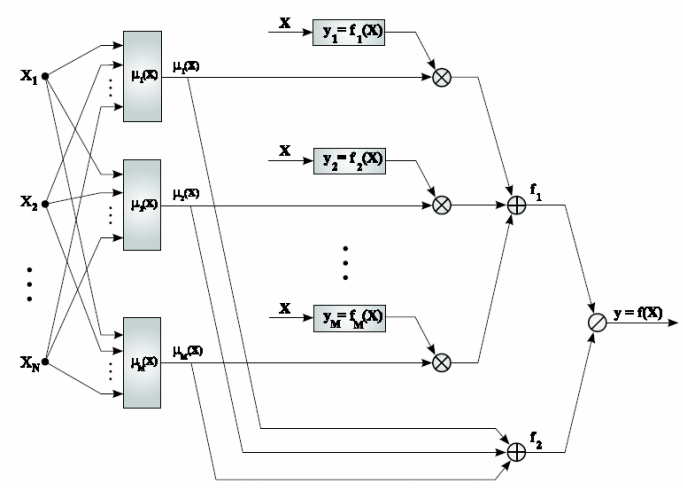

Fig. 2. *The structure of the modified TSK network*

The schema of the algorithm is presented on Fig. 3. As it can be seen, the first stage of the algorithm is the determination of the number *M* of the clusters and their locations $\mathbf{c}^{(i)}$ using fuzzy clustering algorithm. In the next stage, the nonlinear parameters σ_i, b_i and the linear ones p_{ij} are initialized. After initialization, these parameters are adapted with the hybrid algorithm [6]. The model is then validated on the basis of the error functions in learning and testing modes. If the errors

are too high (especially in testing mode), the model is rejected and the algorithm is repeated with another range of the centers number. In the whole process, the only parameter need to be set up by the user is the maximum of the range, in which the number of centers is looked up.

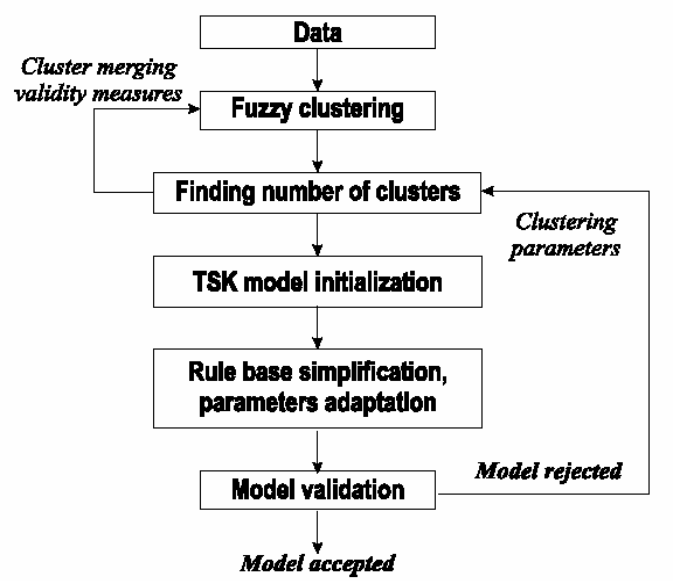

Fig. 3. *The schema of automatic building of TSK network*

III. Automatic generation of rules using fuzzy self-organization and supervised learning algorithm

The most important problem in TSK network is the determination of the number of rules that should be used in the modeling of data. More rules means better representation of data, but at the same time it will also increase the complexity of the neural network and will result in higher cost of data processing. Too complex neural network may also lead to the decrease of the generalization ability and the deterioration of quality of the network operation in the testing mode. Hence the most important problem is to reduce the number of inference rules, by eliminating the combinations corresponding to the empty data space regions. We have solved this problem here by applying fuzzy self-organization of the data. Therefore procedures for automatic calculation of the number of clusters, irrespective of dimensionality and distribution of data are required. In this paper we present the Gustafson - Kessel (GK) clustering algorithm to determine the locations of a number of centers, a quality global measure to fix the number of the centers, and the algorithm of setting up the initial values of fuzzifier functions [3,4,6].

III.1 The GK clustering algorithm

In the GK algorithm, the distance from the input vector $\mathbf{x}_j$ to the center $\mathbf{c}^{(i)}$ is defined by the equation (5). The degree, with which the vector $\mathbf{x}_j$ belongs to the center $\mathbf{c}^{(i)}$, is

$$u_{ij} = \frac{1}{1+d_{ij}^2} \tag{7}$$

The fuzzy clustering algorithms search for the partition matrix and cluster centers such that the objective function *E* is minimized [2,3,4]

$$E = \sum_{i=1}^{C}\sum_{j=1}^{p} u_{ij}^{m} d_{ij}^{2} \tag{8}$$

subject to $\sum_{i=1}^{C} u_{ij} = 1$ for j = 1, 2, ..., p and $0 \le u_{ij} \le 1$ for i = 1, 2, ..., C; j = 1, 2, ..., p. The algorithm has proved to behave well on different classification problems of data distributed within flat, elongated clusters.

III.2 Determination of the number of clusters

To control the number of clusters we will apply here the so called validity measures [3] that assess different qualities of the clusters, on the basis of which the optimal number of clusters can be determined. Four different validity measures have been applied to the solution of the problem.

- *fuzzy hypervolume measure* $V_h = \sum_{i=1}^{C} \sqrt{\det(\mathbf{F}_i)}$ where $\mathbf{F}_i$ - the cluster covariance matrix, calculated by $\mathbf{F}_i = \dfrac{\sum_{j=1}^{p} u_{ij}^{m} \left(\mathbf{x}_j - \mathbf{c}^{(i)}\right)\left(\mathbf{x}_j - \mathbf{c}^{(i)}\right)^T}{\sum_{j=1}^{p} u_{ij}^{m}}$
- *average partition density* $D_A = \dfrac{1}{C}\sum_{i=1}^{C} \dfrac{S_i}{\sqrt{\det(\mathbf{F}_i)}}$ where the parameter S_i (i = 1, 2, ..., C) is calculated only for these vectors $\mathbf{x}_k$ that lie within a hyperellipsoid, whose radii are the standard deviations of the cluster features and are defined as $S_i = \sum_k u_{ik}$ for such *k*, that $\left(\mathbf{x}_k - \mathbf{c}_i\right)^T \mathbf{F}_i^{-1} \left(\mathbf{x}_k - \mathbf{c}_i\right) < 1$.
- average within-cluster distance $D_w = \dfrac{1}{C}\sum_{i=1}^{C} \dfrac{\sum_{k=1}^{p} u_{ik}^{m} d_{ik}^{2}}{\sum_{k=1}^{p} u_{ik}^{m}}$
- *average cluster flatness* $t_A = \dfrac{1}{C}\sum_{i=1}^{C} t_i$ where $t_i = \dfrac{\lambda_{iN}}{\lambda_{i1}}$ and λ_{ij} - the *j*-th eigenvalue of the cluster covariance matrices $\mathbf{F}_i$, arranged in a descending order, i.e., $\lambda_{i1} \ge \lambda_{i2} \ge ... \ge \lambda_{iN}$.

A good partition is indicated by small values of V_h and t_A and high values of D_A and D_w. An example of values of the validity as a function of center number is shown on Fig. 4.

In our solution we aim at simultaneous satisfaction of all these four quality measures. To get one quality measure we define one *heuristic global quality factor* α in the following way

$$\alpha = a_1 V_h - a_2 D_A - a_3 D_w + a_4 t_A \tag{9}$$

where a_i (for *i = 1, 2, 3, 4*) are positive scaling coefficients. The minus signs for a_2 and a_3 mean we prefer the smaller values of the heuristic global quality factor. The local minima of this measure indicate the suboptimal numbers of cluster centers. The procedure of finding those minima relies on calculating the quality factor for different number of clusters and choosing those corresponded to the local minima.

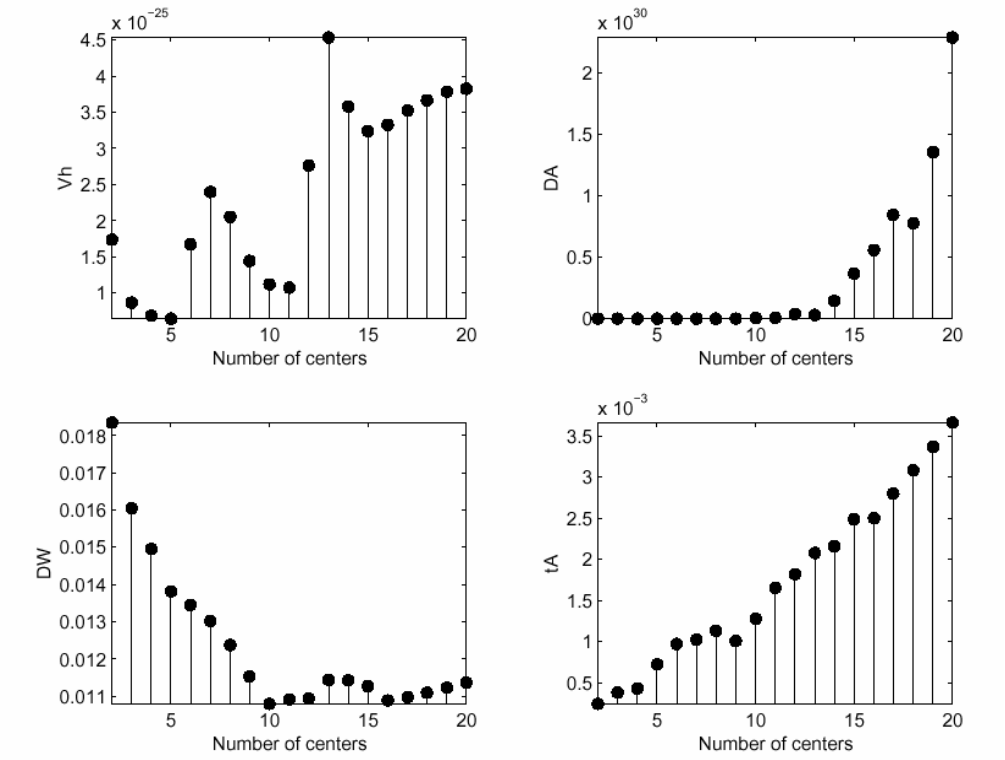

Fig. 4. *Values of 4 validity measures as a function of center number*

Figure 5 presents the values of the heuristic global quality factor α as a function of center number. The curve has some minima (located at center number equal 5, 11 or 16). These values indicate the potential center number of the input data.

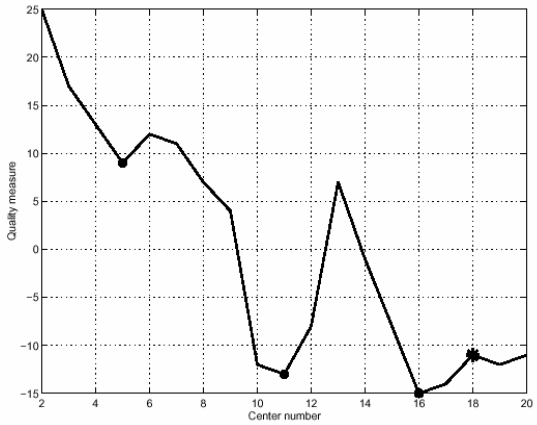

Fig. 5. *Values of final heuristic measure as a function of center number*

The clustering procedure applied for grouping the data is an important step in building inference rules for fuzzy TSK system. Each cluster center is associated with the center of the appropriate inference rule and the number of cluster indicates the number of applied rules.

III.3 Initialization the inference functions

The previous subsections have presented the algorithm of determination the number of centers and their first locations $\mathbf{c}_i$. Therefore the other two parameters of inference function, i.e. the width σ_i and the exponent b_i coefficients, are still needed to be initialized. From our observations, it has resulted

that the start value of the exponent coefficient b_i equal 1 is sufficient for most of cases. It means that the antecedent part of each rule can start with the bell form. The other situation is for the width coefficient σ_i. In making the inference rules more sensible on the input range, this coefficient is much stronger than the exponent one. A small value of σ_i makes the rule to be more 'crisp', but a bigger value makes the rule to be more active for wide range of input. Its value should depend on the data distribution and the distances between center locations. We have proposed a method to initialize the values of σ_i. The algorithm is following

- For every center $\mathbf{c}_i$, calculate the distances to all other centers using (5).
- Calculate the effective ratio R_i^{ef} as the average distance from center $\mathbf{c}_i$ to some nearest neighbors (usually up to 5).
- Set the initial value σ_i to $R_i^{ef}/2$ (this value will assign the value of membership function equal 0.2 for the data points at distance R_i^{ef}).

III.4 Parameters adaptation

After initialization, the modified TSK network needs further fine-tuning to increase its performance. This tuning can be done by a learning process, which is very typical for the neural networks. In adaptation of the parameters of the network we have applied the hybrid algorithm [2], which contains two stages: learning the linear parameters and learning the nonlinear parameters. The hybrid learning algorithm of TSK neuro-fuzzy network belongs to the supervised ones, which perform the minimization of the cost function [2]

$$E = \frac{1}{2}\sum_{l=1}^{p}\left(y(\mathbf{x}^{(l)}) - d^{(l)}\right)^2 \qquad (10)$$

In the first stage, the linear parameters are adapted using the SVD algorithm and the pseudo-inversion of a matrix [6]. This method will lead us to the global optimal values of linear parameters subject to the actual values of nonlinear parameters. The nonlinear parameters are adapted using the steepest descent method of minimization of this function. For example, the relation for adapting the center $\mathbf{c}^{(k)}=[c_{k1},\ldots, c_{kN}]$ parameter is as follows [1,2]

$$c_{kj}^{(n+1)} = c_{kj}^{(n)} - \eta_c \frac{\partial E^{(n)}}{\partial c_{kj}^{(n)}} \qquad (11)$$

where n means the consecutive iteration. Due to the small effectiveness of gradient method in comparing with SVD decomposition in learning process, each stage of learning linear parameters is associated with a number of gradient learning iterations (in our experiments the ratio was 20/1).

IV. Numerical experiments

The modified TSK network has been applied for the estimation of the concentration of the gas mixture components by applying the matrix of semiconductors sensors. The technical data of performed experiments are as follows:

- number of gases in the mixture - 4 (carbon oxide, methane, propane/butane and methanol).
- number of sensors - 5 (TGS-815, TGS-822, TGS-842 Figaro sensors and NAP-11A, NAP-11AE Nemoto sensors).
- 216 known gas mixtures for learning purposes.

- 204 gas mixtures for testing purposes.

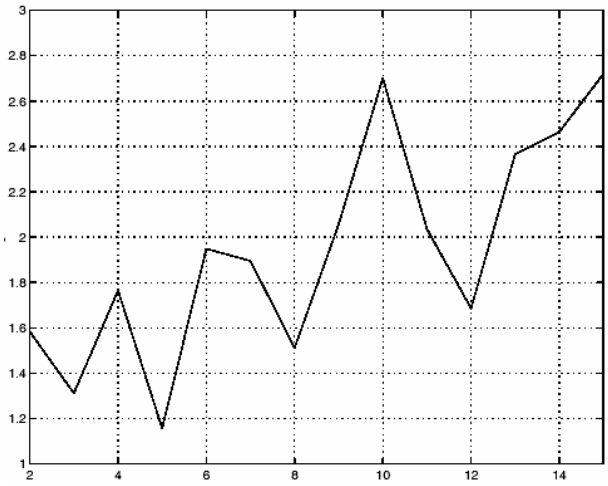

Fig. 6. *Values of final heuristic measure for the gas mixture problem*

At five sensors, the number of inputs of the network is also 5. In the learning data set, we have prepared different combination of components presence (from mixture with all 4 components to the mixture with none of those under consideration components), as well as different concentrations at each component. The GK clusterization procedure applied to the data set has allowed obtaining the suboptimal number of inference rules, which in this case was equal 3, 5, 8 and 12 as seen in the Fig. 6. In further experiments we have chosen those numbers. In Table 1 the number of misclassifications in learning and testing process at different numbers of interference rules is presented.

Table 1. *Results of learning and testing the modified TSK network with different number of rules*

No of rules	No of learning error	Learning error percentage	No of testing error	Testing error percentage
3	8	3.70%	12	5.88%
5	5	2.31%	3	1.47%
8	2	0.93%	7	3.43%
12	0	0.00%	6	2.94%

As it can be seen from this table, by increasing the number of interference rules we can decrease the learning error. However in testing phase, too many rules will lead to the "over learning" effect. From this table it is resulted that 5 is the optimal number of rules (it is also the minimum of the heuristic value α).

V. Conclusions

The paper has presented a modification of the classical TSK network and the algorithm of building an effective structure of this modified network including selecting the optimal number of cluster centers and initializing the parameters of fuzzifier functions. The modified network has been implemented and successfully tested. The results of numerical experiments of the classification of toxic gas components have been presented and discussed. The obtained accuracy of classification is very good after a fully automatic algorithm. The only parameter required in this algorithm is the maximum of the range, where the center number is searched.

References

[1] Haykin, S.: "*Neural networks, a comprehensive foundation*", Macmillan College Publishing Company, New York, 1994.

[2] Jang, J. S., Sun, C. T., Mizutani, E.: "*Neuro-fuzzy and soft computing*", Prentice Hall, New Jersey, 1997

[3] Pal, N. R., Bezdek, J. C.: "*On cluster validity for the fuzzy c-means model*", IEEE Transactions on Fuzzy Systems, vol. 3, 1995, pp. 370-379

[4] Gustafson, D., Kessel, W.: "*Fuzzy clustering with a fuzzy covariance matrix*", Proc. IEEE CDC, San Diego, pp. 761-766

[5] Osowski, S., Brudzewski, K.: "*Fuzzy self-organizing hybrid neural network for gas analysis system*", IEEE Transactions on Instrumentation and Measurement, vol. 49, 2000, pp. 424-429

[6] Linh, T. H.: "*The modification of TSK network in neuro-fuzzy systems*", XXV IC SPETO, 2002, pp. 525-528

[7] Osowski, S., Linh, T. H.: Neuro-fuzzy TSK network for gas measurements, ISTET, 2001, Linz, pp. 242 - 247.

Tran Hoai Linh was born in 1974 in Hanoi, Vietnam. He received the M.Sc. in Applied Informatics, Ph.D. and Dr. Sc. in Electrical Engineering from the Warsaw University of Technology in 1997, 2000 and 2005, respectively.

He is currently a Researcher and Lecturer in the Institute of Instrumentations and Industrial Informatics, Department of Electrical Engineering, Hanoi University of Technology. His professional research interests are artificial methods and applications in classification and estimation problems.