Influence of Fuzzy Clustering on the Accuracy of Electrical Equipment Diagnostic Models

Denis Eltyshev and Ksenia Gnutova
Electrotechnical Department, Perm National Research Polytechnic University, Komsomolsky Ave. 29, Perm, Russia
eltyshev@msa.pstu.ru, k.gnutova@gmail.com

Keywords: Diagnostic, Electrical Equipment, Fuzzy Clustering, Fuzzy logic, Membership function

Abstract: The development of electric power industry is oriented on high reliability, flexibility and efficiency of managing power grids of arbitrary configuration. For such grids an information infrastructure is required. It should consist of various software and hardware, including systems of electrical equipment monitoring and diagnostics for accumulating information about its parameters with controlling and managing its condition. As a rule, data about the electrical equipment is heterogeneous. Thus there is a necessity of certain mechanisms for data processing to provide a possibility of constructing diagnostic models in an automated mode and adapting them to power grid operating conditions. The aim of this work is to develop a mechanism for automated calculation of the electrical equipment diagnostic models parameters. It supposes using historical data analyzing algorithms that ensure high reliability of the diagnosis. Base on this the application of fuzzy clustering for constructing membership functions to assess the features of equipment condition change in fuzzy diagnostic models was considered. Different fuzzy clustering algorithms were analyzed, and a technique for processing data on the equipment operation with constructing membership functions based on fuzzy partition matrix and clusters centres was proposed. The technique allows to approximate the membership functions by typical curves and to choose the most effective variant of clustering in terms of electrical equipment diagnostic reliability. The testing of fuzzy models for assessing the condition of power transformer equipment using clustering results was performed. A high level of compliance of simulation data with the conclusions of specialized organizations performing monitoring of equipment in power supply systems for oil production facilities was obtained. The practical relevance of the results is in using the technique in the synthesis of intelligent expert-diagnostic systems for increasing the electrical equipment diagnostic reliability and reducing the duration of its unplanned downtime.

1 INTRODUCTION

The development of efficient new generation electric power systems based on the Smart Grid technology with high reliability, fault tolerance, flexibility and adaptability is associated with the introduction of modern information and telecommunication technologies [1] - [3]. One of the tasks of such systems is to provide control over the condition of electrical equipment (EE), maintain its operability and operatively manage its operation modes. Modern information and diagnostic systems for assessing the EE condition should be integrated with monitoring systems to collect and accumulate data necessary for subsequent processing and analysis. Since this possibility can not always be ensured by technical and economic factors, to obtain information about the object both stationary and mobile software and hardware can be used [1][2], [4][5]. Considering the heterogeneity and ambiguousness of the EE operating data it is advisable to use intelligent technologies for constructing diagnostic systems, based on statistical data and the experience of qualified experts [4] – [8].

2 SETTING GOALS AND OBJECTIVES OF THE STUDY

The intelligent diagnostic procedure in general can be represented as the defining of the ratio [6][7]:

$$Y\epsilon(y_1, y_2,\ldots,y_n) \rightarrow \mathbf{X}=(x_1, x_2,\ldots, x_m) \quad (1)$$

where Y is the set of classes of EE condition or types of defects in equipment elements; $\mathbf{X}$ is the vector of controlled technical parameters (diagnostic features); → is a set of rules linking the values of the parameters with the level of EE condition.

To solve the problem that is difficult to formalize, models based on fuzzy logic can be used. In this case it is necessary: to define for each variable x_i, i=1:n the number of terms for their verbal estimates (for example, the vibration level is “low”, “medium”, “high”); describe each term as a membership function (MF); construct a rule base that connect the values of the variables **X** with the class y_j, j=1:m. As a rule, these operations are carried out manually, by an expert, which does not provide the necessary flexibility of diagnostic models.

We can automate this process using the available data on the EE operation, or the data generated by the monitoring system.

In order to determine the parameters of the MF in automated mode, as well as to build the rule base, it is proposed to apply methods of data mining, in particular, fuzzy clustering [8] -[10]. This will greatly simplify the process of building expert-diagnostic systems, oriented on the assessing and management of the EE condition.

The aim of the work is to research mechanism of using fuzzy clustering algorithms for automated defining the MF in the EE diagnostic models that ensure high reliability of the diagnosis. The tasks include: developing the technique of fuzzy clustering based on the EE operating data for defining the MF parameters with approximation of MF by typical curves; estimation the influence of clustering algorithms on accuracy of EE diagnostics models.

3 DESCRIPTION OF THE RESEARCH METHODOLOGY

The procedure of the EE operating data analysis for synthesising expert-diagnostic models with automated defining of a MF can be represented as follows (Figure 1).

At the first stage (block 1) an array of initial data on the measurements of each monitored diagnostic parameter is formed.

Fuzzy clustering consists in determining the coordinates of clusters centers **C** = (c_1, c_2, ..., c_m) and the matrix **U** = [u_{ij}] showing the belonging of parameter measurements for each cluster (block 3).

Let’s consider the process of clustering using the fuzzy c-means algorithms (FCM), Gustafson-Kessel algorithms (GK) and Gath-Geva algorithms (GG) [11-13]. The difference between the methods is in the objective function and special metric, which makes it possible to allocate clusters of different shapes. For example, the FCM method minimizes the following objective function:

$$J(Z;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{m}\left\|z_k-v_i\right\|_A^2 \qquad (2)$$

where z_k is the data array, m is the exponential weight, μ_{ik} is the partition matrix, c is the number of clusters, and **A** is the diagonal matrix.

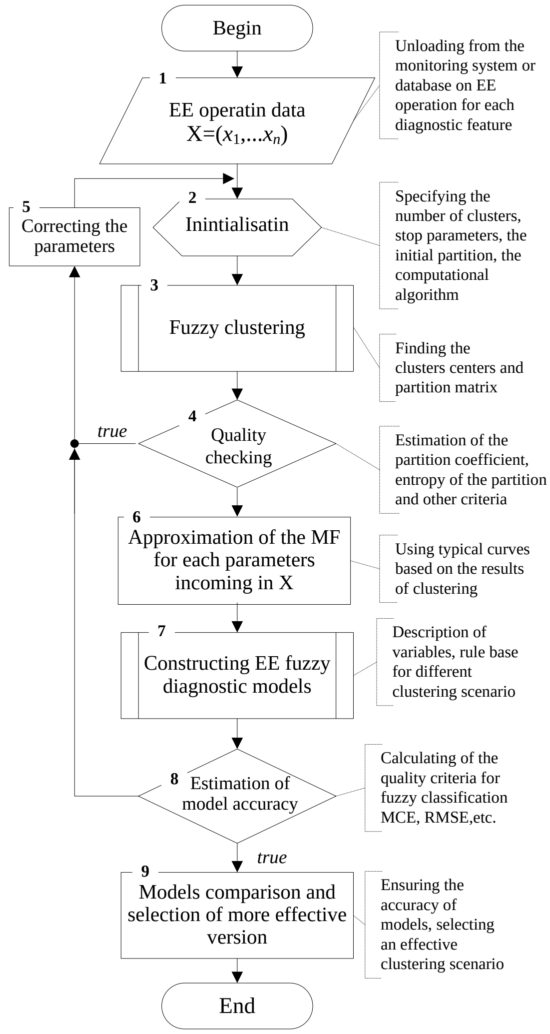


Figure 1: Block diagram of the technique for using fuzzy clustering in constructing EE diagnostics models.

The GK method uses the adaptive norm for each j-th cluster due to the individual matrix $\mathbf{A}_i$. It is a fuzzy covariance matrix of the cluster changing during the iteration process with the following objective function:

$$J(Z;U,V,\{A_i\}) = \sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{m} D_{ik\mathrm{A}_i}^{2} \tag{3}$$

where $\mathbf{A}_i$ is a fuzzy covariance matrix of the cluster.

The GG method uses the Gaussian distribution, and minimizes the sum of the squares between the data points and the prototype η_i of the being formed group.

In general the clustering algorithm can be divided into two stages: 1) initialization stage (block 2) when the membership matrix for all elements of the input set is randomly populated and the necessary parameters (number of clusters, exponential weight etc.) are selected. 2) The calculation stage (block 3), when the following steps are performed for each iteration:

1) Finding cluster centres:

$$v_i^{(l)} = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m z_k}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m},\ i = \overline{1,c} \tag{4}$$

2) Finding the covariance matrix of the cluster for the GK and GG methods:

$$F_i = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m \left(z_k - v_i^{(l)}\right)\left(z_k - v_i^{(l)}\right)^T}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m},\ i = \overline{1,c} \tag{5}$$

3) Calculation of the distances between the new cluster centres and the points of data set. For the FCM method, the Euclidean distance is used:

$$D_{ik\mathbf{A}}^2 = \left(z_k - v_i^{(l)}\right)^T \mathbf{A}\left(z_k - v_i^{(l)}\right),\ i = \overline{1,c},\ k = \overline{1,N} \tag{6}$$

For the GK method:

$$D_{ik\mathbf{A}_i}^2 = \left(z_k - v_i^{(l)}\right)^T \left[\rho_i \det(F_i)^{\frac{1}{n}} F_i^{-1}\right]\left(z_k - v_i^{(l)}\right), \tag{7}$$

For the GG method:

$$D_{ik\mathrm{A}_i}^2 = \frac{(2\pi)^{\left(\frac{n}{2}\right)}\sqrt{\det(F_i)}}{a_i}\exp\left(\frac{1}{2}\left(z_k - v_i^{(l)}\right)^T F_i^{-1}\left(z_k - v_i^{(l)}\right)\right) \tag{8}$$

$$a_i = \frac{1}{N}\sum_{k=1}^{N}\mu_{ik} \tag{9}$$

4) Recalculation of the partition matrix.
The FCM method uses the condition:
If $D_{ik\mathbf{A}} > 0$, then

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c}\left(D_{ik\mathbf{A}} / D_{jk\mathbf{A}}\right)^{\frac{2}{(m-1)}}} \tag{10}$$

otherwise

$$\mu_{ik}^{(l)} = 0, \text{ if } D_{ik\mathrm{A}} > 0 \text{ and}$$

$$\mu_{ik}^{(l)} \in [0,1],\ \sum_{i=1}^{c}\mu_{ik}^{(l)} = 1. \tag{11}$$

For GK and GG methods the partition matrix is recalculated in the same way, but $D_{ik\mathbf{A}}$ is used instead of $D_{ik\mathrm{A}_i}$.

5) Stopping the algorithm. The fuzzy clustering algorithm stops when the next condition is fulfilled:

$$\max_{k=\overline{1,M},\, i=\overline{1,c}}\left(\left|\mu_{ki} - \mu_{ki}^*\right|\right) < \varepsilon, \text{ or } \max_{i=\overline{1,c}}\left(\left|V_i - V_i^*\right|\right) < \varepsilon \tag{12}$$

To assess the effectiveness of clustering a complex analysis of the results quality is carried out (block 4). To construct the MF we use the obtained coordinates of the cluster centers and the fuzzy partition matrix (block 6). For these purposes the technique of MF approximation by typical curves (for example, Gaussian, bell-shaped, pi-like) is used.

In accordance with the results of the MF construction fuzzy models, connecting the variables **X** and *Y* (block 7), are formed, the rules base is synthesized and the accuracy of the models by the criteria of the mean classification error (MCE), root mean square error (RMSE) between model and experimental data, or complex criteria [14] is determined.

4 INVESTIGATION OF THE METHODOLOGY

Let's consider (Figure 2) a simplified example of using fuzzy clustering in the problem of assessing the thermal condition of the power oil-filled transformer equipment elements (POTE). The transformer TDNT-16000/110-U1 with a voltage of 110/35/6 kV, which is typical equipment for power supply systems of the oil-producing fields in the Perm Krai was chosen as the object of the study.

We considered a fuzzy diagnostic model with input parameters $\mathbf{X} = (x_1, x_2)$, characterizing the condition of the transformer tank, where x_1 is the excess temperature of the bolted tank bell connections; x_2 is the maximum temperature difference over the surface of the tank and the elements of the cooling system [7].

We used the results of thermal imaging control of transformers from Perm Krai oil production facilities for the period 2010-2013 as the initial data. Also we supplemented initial data by simulation modelling, taking into account the boundary values of the transformer parameters given in the technical documentation for the electrical equipment operation (sample size - 250 measurements).

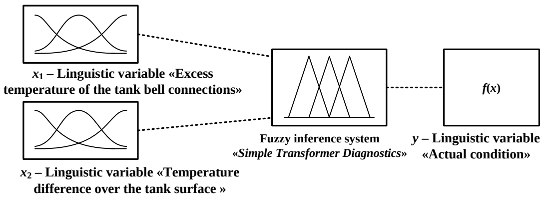


Figure 2: The structure of the fuzzy model for diagnostic POTE condition in the Fuzzy Logic Toolbox of the MATLAB.

We formed data set on the basis of monitoring protocols compiled by specialized service organizations using the thermograms of the transformer elements. In accordance with the technique of the POTE diagnostics using by the service organization and the recommendations of the normative and technical documentation we selected the classes “No defects” (1), “Developed defect” (2), “Critical defect” (3) for assessing the condition of the transformer tank *Y*. A fragment of the initial data is given in Table 1.

Table 1: Representation of the initial data for the POTE thermal diagnostic model.

№	x_1, °C	x_2, °C	Class of condition	Conclusion on the condition
1	4.9	25	1	No thermal defects
16	15	24.1	2	Developed defect (increased contact heating)
67	18	13	2	Developed defect (increased contact heating)
155	2.2	5	2	Developed defect (reduced circulation of oil in the tank)
184	32	20	3	Critical defect (contact overheating)
199	1.7	27.1	1	No thermal defects
246	7	39	3	Critical defect (increased tank heating)

We used 70% of data sample and three clusters for data processing by all fuzzy algorithms (FCM, GK, GG). The results of clustering (ordered and normalized data) are shown on Figure 3. The efficiency of clustering was estimated by partition coefficient (PC) and entropy (CE).

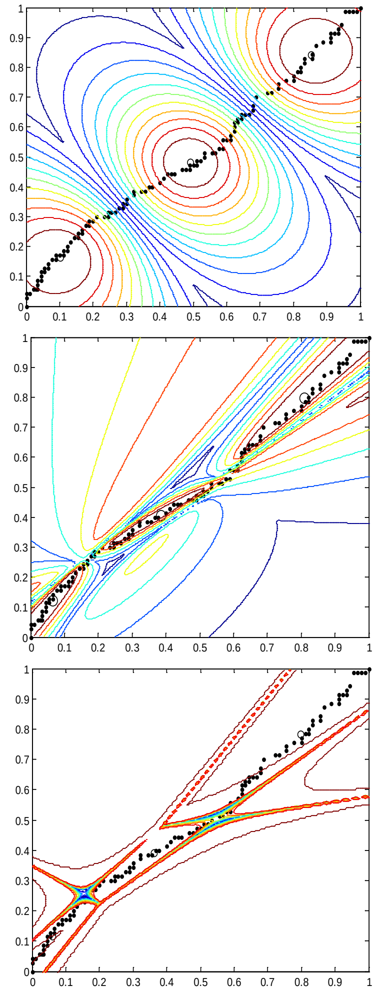


Figure 3: Fuzzy clustering of data on the POTE elements temperature, FCM (PC=0.82, CE=0.34) (a), GK (PC=0.91, CE=0.18) (b), GG (PC=0.96, CE=0.07) (c).

In order to construct fuzzy models we used the fuzzy partition matrixes for each cluster and constructed MF with three terms for assessing POTE elements temperature: “Low” (L), “Medium” (M), and “High” (H). The example of constructing the MF using fuzzy partition matrix and after approximation of the clustering results by Gaussian curves is shown in Figure 4.

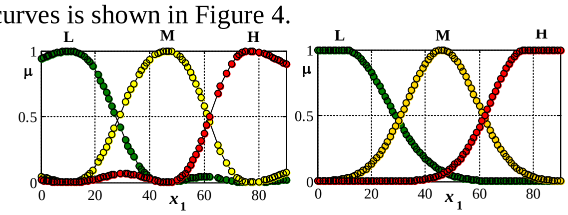


Figure 4: Constructing MF using fuzzy partition matrix (a) and approximation (b).

The rule bases of POTE fuzzy diagnostic models consist of 9th rules of following type: IF x1 = “Low” AND x2 = “Low”, then y = “No defects”.

5 RESULTS OF MODEL ACCURACY ESTIMATION

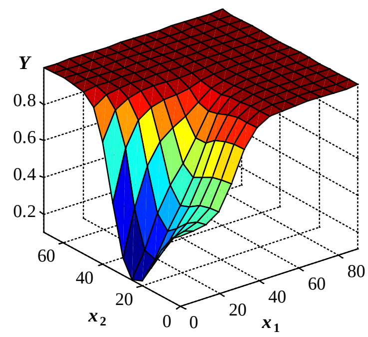
To assess the accuracy of fuzzy diagnostic models obtained by using clustering algorithms the simulation results and the conclusions of the service organizations from POTE monitoring protocols were compared.

We used 70% of data sample (190 positions) to learn fuzzy models and set the weights of the rules from the rule base by the gradient methods and complex criterion [14].

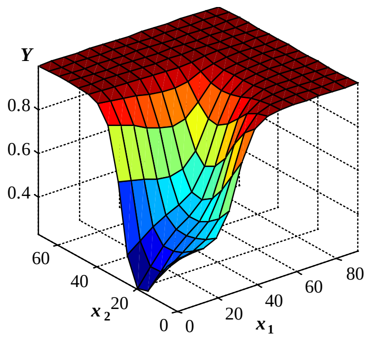
During the models adequacy verification using testing data set (30% of data sample) by χ^2 criterion we got the following results: χ^2 = 7.33 for model using FCM, χ^2 = 4.33 for model using GK, χ^2 = 6.01 for model using GG.

When a critical value of χ^2 = 34.77 the hypotheses of the models adequacy are accepted with 100% probability.

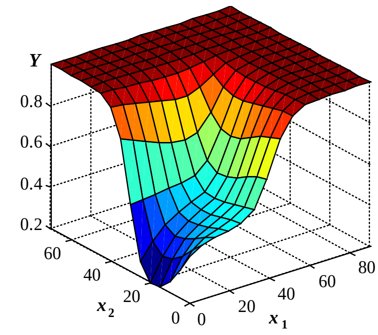
It can be seen from Figure 5 that the response surfaces of the models constructed by clustering algorithms differ little from each other.



a)



b)



c)

Figure 5: Surfaces of response of models, FCM (a), GK (b), GG (c).

To assess the effectiveness of clustering a mean fuzzy classification error (MCE) was calculated (Figure 6) on various data sets. We should note that for different cluster algorithms the number of correctly recognized classes (defects in equipment) varies from 89.2 to 92.4 %, which is a pretty good result. Thus using cluster algorithms for MF constructing provides small error of fault diagnosis of PTOE condition. For GG algorithm the error is less by 3.2% than for traditional FCM algorithm.

Reduction of errors in recognition of defects will allow in practice to reduce the equipment downtime both for the cause of the accident and according to plan. In the future, to confirm the effectiveness of the technique, it can be compared with other classification methods (for example, neural networks) in assessing the EE condition.

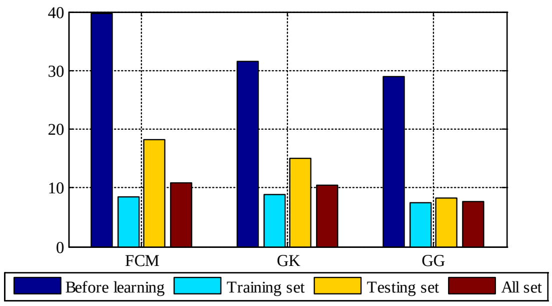


Figure 6: Results of the evaluation of the accuracy of fuzzy models of diagnostics of the POTE. The average number of misidentified MCE states, %.

It can be seen from the results that the application of clustering for the construction of the MF provides the accuracy of the diagnosis to 92.4% and more efficiently than the use of the expert method. At the same time there is an increase in the reliability of the assessment of the state by 8-12%.

6 CONCLUSIONS

Let us consider the most important results of the work:

1. The use of fuzzy clustering allows building the MF in automated mode, minimizing the expert's participation and taking into account the specifics of the electrical equipment operation contained in the available statistical data.

2. We can see deterioration or improvement in the quality of the equipment diagnostic models depending on the fuzzy clustering parameters.

3. The application of both fuzzy clustering and fuzzy modeling gives high results in the reliability of the EE condition assessment and provides adaptability of diagnostic models with possibility of the data volume increasing.

We should note that it is necessary to have an implemented technology for monitoring the EE condition and a database aggregating the results of the monitoring with processing and converting the data to a convenient form for fuzzy clustering and classification in order to provide the effectiveness of the technique. The technique is sensitive to initial data, clustering parameters, and algorithms for fuzzy models training.

We can define following directions of the further researches: formation of the rule base for diagnostic models in the automated mode; an analysis of ways to improve the accuracy of diagnostic models, including valid choice of the curve type for MF approximation and algorithms for model parameters setting; approbation of the technique on other EE diagnostics problems.

The results can be further used in the automated synthesis of expert-diagnostic models in intelligent systems for assessing the EE condition.

REFERENCES

[1] A.V. Kychkin, G.P. Mikriukov, "Applied data analysis in energy monitoring system," Journal of Problems regional energy, vol. 2(31), pp. 84-92, 2016.

[2] A.V. Kychkin, "Synthesizing A System For Remote Energy Monitoring In Manufacturing," Metallurgist, vol. 59, № 9-10, pp. 752-760, 2016.

[3] A.B. Petrochenkov, T. Frank, A.V. Romodin, A.V. Kychkin, "Hardware-in-the-loop Simulation of an Active-adaptive Power Grid," Russian Electrical Engineering, vol. 84, № 11, pp. 652-658, 2013.

[4] N.I. Khoroshev, V.P. Kazantsev, "Management Support of Electroengineering Equipment Servicing Based on the Actual Technical Condition," Automation and Remote Control, vol. 76, № 6, pp. 1058-1069, 2015.

[5] A.B. Petrochenkov, S.V. Bochkarev, A.V. Romodin, D.K. Eltyshev, "The planning operation process of electrotechnical equipment using the Markov process theory," Russian Electrical Engineering, vol. 82, № 11, pp. 592‐ 596, 2011.

[6] D.K. Eltyshev, V.V. Boyarshinova, "Intelligent Decision Support in the Electrical Equipment Diagnostics," Proceedings of the 19th International Conference on Soft Computing and Measurements, SCM 2016, pp. 157-160, 2016.

[7] D.K. Eltyshev, N.I. Khoroshev, "Diagnostics of the Power Oil-filled Transformer Equipment of Thermal Power Plants," Thermal Engineering, vol. 63, № 8, pp. 558-566, 2016.

[8] N.I. Khoroshev, R.N. Pogorazdov, "Adaptive Clustering Method in Intelligent Automated Decision Support Systems," Proceedings of the 19th International Conference on Soft Computing and Measurements, SCM 2016, pp. 296-298, 2016.

[9] K.A. Gnutova, D.K. Eltyshev, "Using Cluster Analysis in the Synthesis of Electrical Equipment Diagnostic Models," Proceedings of the 5th International Conference on Applied Innovations in IT 2017, pp. 119-124, 2017.

[10] D.K. Eltyshev, K.A. Gnutova, "Identification of the parameters of electrical equipment condition assessment models using fuzzy clustering," Proceedings of the 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017, pp. 142-144, 2017.

[11] J.C. Bezdek, R. Ehrlich., W Full, "FCM: The fuzzy c-means clustering algorithm," Computers & Geosciences, vol. 10, № 2–3, pp. 191-203, 1984,

[12] R. Babuska, P.J. van der Veen, U. Kaymak, "Improved covariance estimation for Gustafson-Kessel clustering," Proceedings of the 2002 IEEE international conference on fuzzy systems, Honolulu: Technische Universitet Eindhoven, 2002, pp. 1081-1085.

[13] I. Gath, A.B. Geva, "Unsupervised optimal fuzzy clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, vol. 11, № 7, pp.773-781.

[14] S.D. Shtovba, O.D. Pankevich, A.V. Nagorna, "Analyzing the Criteria for Fuzzy Classifier Learning," Automatic Control and Computer Sciences. 2015, vol. 49. № 3, pp. 123–132, doi:10.3103/S0146411615030098.