Using Cluster Analysis in the Synthesis of Electrical Equipment Diagnostic Models

Ksenia Gnutova, Denis Eltyshev
Electrotechnical Department, Perm National Research Polytechnic University,
Komsomolsky Ave. 29, 614990, Perm, Russia
{lis, pab}@msa.pstu.ru

Keywords: Electrical equipment, Cluster analysis, Quality criteria, Fuzzy logic, Diagnostics.

Abstract: The article investigates the issue of improving the methods of diagnostics of electrical equipment conditions to ensure the effective assessment of equipment needs for repairs and its trouble-free, safe and economical operation. The possibility of taking advantage of different cluster analysis methods enables us to form the structure of fuzzy models of electrical equipment diagnostics. The method of synthesis of this class of models takes into account various ways to implement clustering algorithms and criteria for assessing its effectiveness. The software, which we use to study the applicability of methods for the analysis of data on temperature parameters data of transformer equipment, utilises methods such as k-means and fuzzy c-means.

1 INTRODUCTION

The reliability and quality of power supply systems of both industrial and civil use is largely determined by the ability of their constituent electrical equipment (EE) reliably perform specified functions. To solve this problem, it is necessary to organize an effective system of maintenance and repair of the EE with timely prevent potential accidents. An objective method of assessing the needs of the EE to be repaired involves the periodic (discrete) or permanent (continuous) controlling of its technical condition at the moment, at some point in the future (forecasting) and in the past (Solodyankin, 2015, Kychkin, 2016). Such control can be achieved through an integrated approach to improve the methods and technical diagnostics tools in order to ensure a safe, trouble-free and economical operation.

Construction of automated or automated information systems that will reliably diagnose the various elements of the EE and form recommendations to engineering and technical personnel, requires a specific methodology for the creation of models that can detect signs of defect states. The problem of constructing EE diagnosis models is usually quite complex, as it requires taking into account a variety of factors, including changes in the dynamics of equipment parameters and criteria for assessment of its condition, as well as the environment in which it operates (Semenov, 2004). That is why the development of effective methods of diagnosing EE in recent years increasingly applies intelligent technology such as fuzzy logic, in combination with various methods of representation of expert knowledge. To create a fuzzy model, EE diagnosis must perform the procedure and its structural parametric identification. The primary objective in this case is the right choice of parameters that affect the condition of the equipment and the formation of knowledge as well as the construction of the respective membership functions for each variable, and act as input variables of the model. To perform these operations in an automated or automatic mode, which significantly extends the capabilities of diagnostic information systems you can use EE data mining methods in particular, cluster analysis.

2 SETTING GOALS AND OBJECTIVES OF THE STUDY

The aim of this study is to investigate the possibilities of using different methods of cluster analysis to form a fuzzy diagnosis model structure - (FDM) EE.

The main tasks of the research are the following: to develop an algorithm of data clustering analysis for FDM; to perform clustering analysis of the quality assessment criteria; analysis of the influence on the quality of clustering changing the number of clusters and different ways of finding the distance between the clusters and centroids; formation recommendations.

3 ALGORITHM

Let's consider the following EE data clustering procedure in order to use the results for the synthesis of FDM (Figure 1).

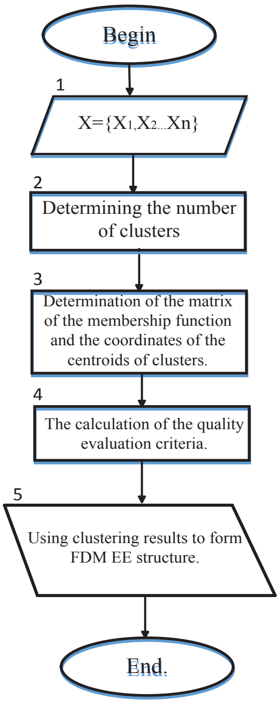


Figure 1: Clustering algorithm.

Initially, the set X input data is given, where n-number of key technical parameters of EE is controlled during its operation and affects the actual state (block 1). The next determined number of clusters is C, which is divided by the test data set. The C parameter can be set by an expert, or calculated in accordance with the established quality of the clustering criterion (block 2) (Kychkin, 2016). Using different methods of cluster analysis, we determine the matrix of membership functions and can find the cluster centroid (block 3). The last stage involves the comprehensive assessment of the cluster analysis quality (block 4).

For the comprehensive evaluation of the cluster analysis quality we will consider the following known criteria (Elizarov, 2009, Khoroshev, 2016, Eltyshev, 2016):

1. The partition coefficient:

$$PC = \frac{\sum_{i=1}^{|X|}\sum_{j=1}^{|C|} u_{ij}^2}{|X|} \tag{1}$$

where Uij is the corresponding element of the matrix accessories, X is the number of elements of the set input, C is the number of elements of the plurality of clusters. This ratio is 1 / C ≤ PC ≤ 1. The closer it is to 1, the clearer the maximum partition is. We must not forget that for a small number of clusters, the partition coefficient gives an incorrect result. To do this without changing the nature of the test, its range has been shifted so that this dependence on the number of clusters C is not associated with the beginning of a specified length, and to its end. Let's perform the experiment by taking the ratio of the partition 1/(| C |). The value range of the ratio is in the range 0 ≤ PCM ≤|C| - 1 / (| C |). The modified partition coefficient is as follows:

$$PC_M = \frac{\sum_{i=1}^{|X|}\sum_{j=1}^{|C|} u_{ij}^2}{|X|} - \frac{1}{|C|} \tag{2}$$

2. Partition entropy is as follows:

$$PE = -\frac{\sum_{i=1}^{|X|}\sum_{j=1}^{|C|} u_{ij}\ln(u_{ij})}{|X|} \tag{3}$$

where Uij is the corresponding element of the matrix accessories, the X is the number of elements of the set input, C is the number of elements of the plurality of clusters. This ratio takes the value 0≤ PE ≤ ln |C |, the best one what partition corresponds to a value close to 0. This ratio should not be used to compare solutions as well as a range of values for each clustering method will be different. Therefore, a more efficient use of the modified partition entropy is ensured (Khoroshev, 2016). The range for this criterion is not linked to the number of clusters and lies on the interval [0, 1]. The modified partition entropy is as follows:

$$PE_M = -\frac{\sum_{i=1}^{|X|}\sum_{j=1}^{|C|} u_{ij}\ln(u_{ij})}{|X|\ln|C|} = \frac{PE}{\ln|C|} \qquad (4)$$

3. The effectiveness partition is as follows:

$$PI = \sum_{j=1}^{|C|}\sum_{i=1}^{|X|} u_{ij}^2\left(d^2(c_j,\bar{x}) - d^2(x_i,c_j)\right) = \sum_{j=1}^{|C|}\sum_{i=1}^{|X|} u_{ij}^2 d^2(c_j,\bar{x}) - \sum_{j=1}^{|C|}\sum_{i=1}^{|X|} u_{ij}^2 d^2(x_i,c_j)$$
(5)

where Uij is the corresponding element of the matrix accessories, the X is the number of elements of the set input, C is the number of elements of the plurality of clusters, Cj is cluster center j, $\bar{\mathbf{x}}$ is the arithmetic mean of the input elements of the set, the set Xi is the input set, d is a distance between the elements, which can be defined in different ways (Euclidean distance, Manhattan distance, etc.) (Eltyshev, 2016).

The algorithm provides for the possibility to set the different ways of finding the distance (metric) between the clusters and their centroids when calculating the clustering options. The best known ones are the following: Euclidean distance, Manhattan distance, cosine and correlation, as well as the Hamming distance (Petrochenkov, 2015).

Automatic selection of possible metrics is in accordance with the clustering quality criterion.

4 RESEARCH RESULTS

The research of the cluster analysis algorithm (Figure 1) is carried out using the power characteristics data of a power oil-filled transformer (POT) of the average power. The object of this type is one of the defining elements of the power supply systems of any configuration, and it is important to ensure reliability of power supply to consumers, and to are a come the difficulty in determining damages and defects at an early stage of development (Solodyankin, 2015, Kychkin, 2016, Semenov, 2004). To test the algorithm, we have selected the most popular cluster analysis methods, such as fcm and k-means (Petrochenkov, 2015, Shtovba, 2007). The initial data uses real settings POT, X = {X1, X2}, where X1 is excess temperature contact of live parts, X2 is temperature difference on the surface of the tank POT and cooling system components. The initial data distribution diagram is shown in Figure 2.

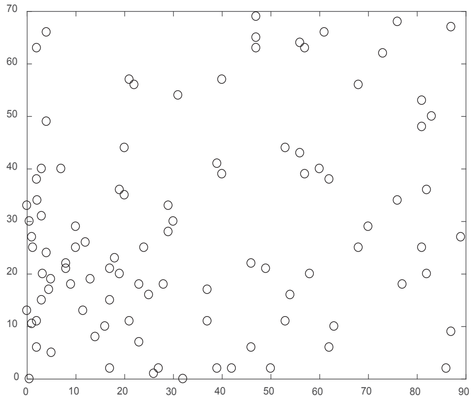
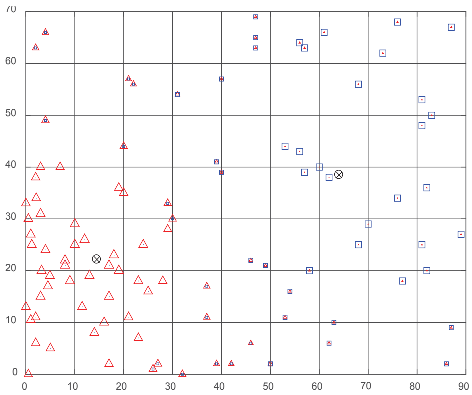


Figure 2: Distribution of raw data.

a)



b)

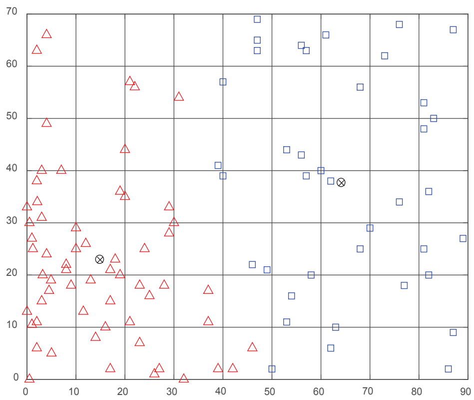
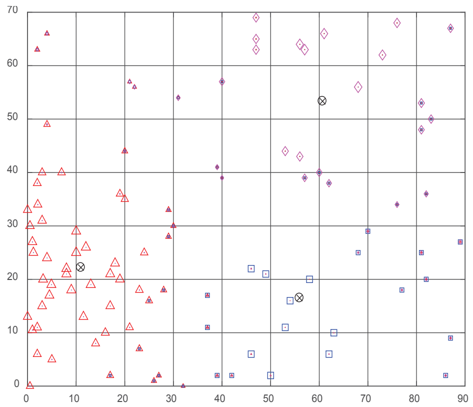


Figure 3: Results of clustering with C = 2, k-means (a) and fcm (b).

a)



b)

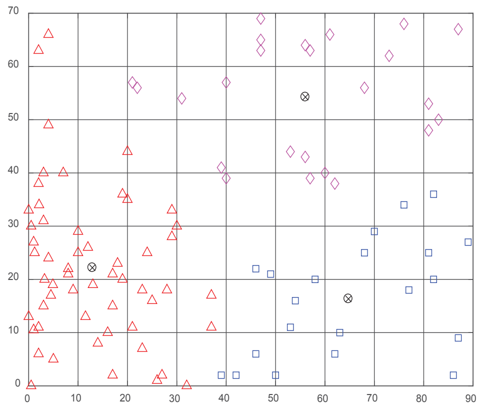
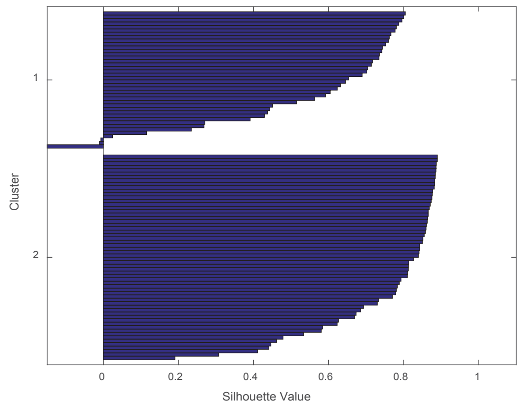


Figure 4: Results of clustering with C = 3, k-means (a) and fcm (b).

As seen in Figure 3 and Figure 4, the centroids belong to two methods when the number of clusters = 2 coincide, and when the number = 3 are different, and there are several elements that are found on the border of another cluster. The silhouette-plot (Figure 5) displays a measure of how close each point in the same cluster is to the points in the neighbouring clusters (Shtovba, 2007, Tosei Hator, 2014).

a)



b)

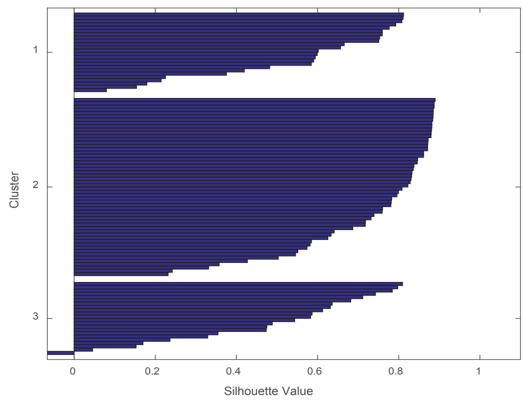


Figure 5: Silhouette-plot for C = 2 (a), C = 3 (b).

Figures 6,7 and 8 presents the results of using the quality criteria for the fcm method. We have obtained familiar indicate adequate quality evaluation data of the cluster analysis method. All the criteria are acceptable in the area. The image shows that the best decomposition occurs when the number of clusters equals to 5.

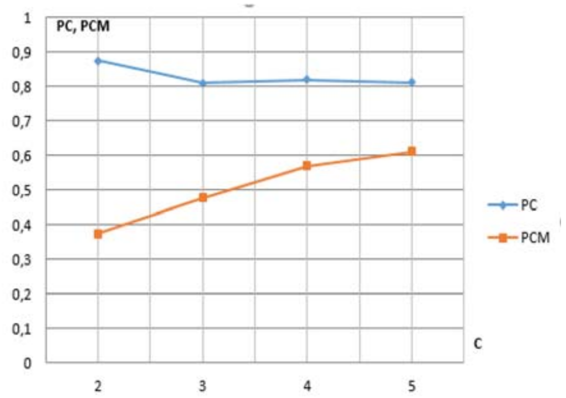


Figure 6: Graph PC, PCM = f (N).

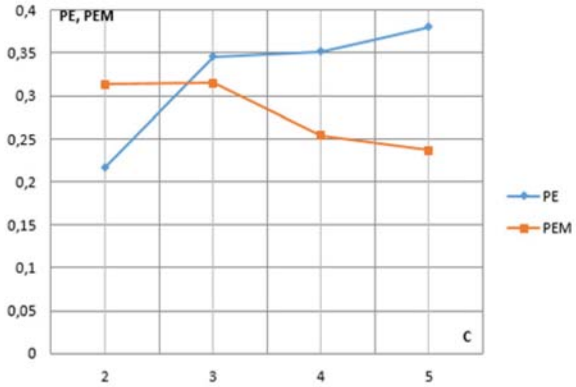


Figure 7: Graph PE, PEM = f (N).

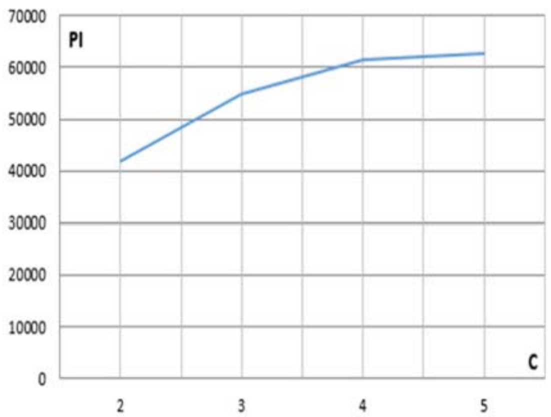


Figure 8: Graph PI = f (N).

Figures 9 and 10 presents the results of the quality criteria for the k-means method, which indicate the complexity of unambiguous assessment of the quality of the selected criteria aggregate. The belonging cluster matrix method, ranging from 0 or 1, makes it impossible to clearly and understandably assess the rate of decomposition and entropy decomposition (Petrochenkov, 2015, Shtovba, 2007). Changes in the clusters centers of coordinates affects the decomposition efficiency. The most adequate assessment method can be provided, based on the data obtained for the modified partition coefficient.

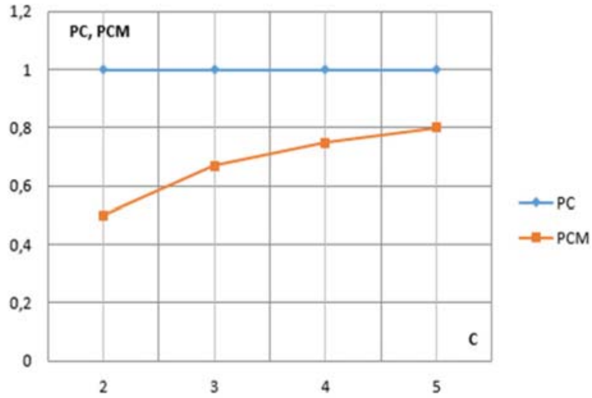


Figure 9: Graph PC, PCM = f (N).

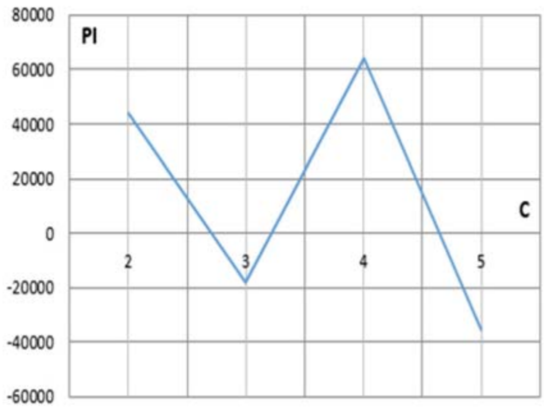


Figure 10: Graph PI = f (N).

5 CONCLUSION

The proposed article technique can be used in the construction of membership functions and rules of the knowledge base FDM. On the basis of a software implementation of clustering techniques in the analysis of known methods (k-means and fcm) made the following conclusions:

1) to determine the FDM structure use fcm method (or modifications thereof), and other methods that allow to evaluate the degree of membership of the input plurality of data items to each of the found clusters during the formation of the partition;

2) to select clustering algorithm and use the advantages of known methods of cluster analysis can

be used existing adaptive methods (Khoroshev, 2016);
3) the use of different criteria for assessing the quality of clustering does not allow an unambiguous conclusion about the optimal partition in terms of building FDM, therefore in order to find the desired number of clusters is necessary to take into account its impact on the accuracy of the classification performed using FDM.

Practical application of the methods, taking into account these factors will allow to formalize the procedure for constructing FDM and use them in automated and automatic control systems technical condition of the EE periodically, or on-line mode.

REFERENCES

Solodyankin A.A., Kolesnikov A.A., Grisha B.G., Nazarov A.A., Yanovich I.M. (2015), 'Modern methods of diagnostics and technical state assessment of the high-voltage electric power equipment', *Journal of Science and Technology*, 2, pp. 70-72.

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

Semenov V.V. (2004). *Diagnosis and monitoring of high-voltage oil-filled electrical equipment.* Ufa. Ufa State Aviation Technical University.

Elizarov S.I., Bargesyan A.A., Kupriyanov M.S., Holod I.I., Tess M.D. (2009). *Analysis of the data and processes.* 3rd edition, Saint-Petersburg: BHV-Petersburg.

Khoroshev N.I., Pogorazdov R.N. (2016), '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.

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

Petrochenkov A.B. (2015), 'Management of Effective Maintenance of the Electrotechnical Complexes of Mineral Resource Industry's Enterprises Based on Energy-information Model', *Proceedings of International Conference on Soft Computing and Measurements*. SCM 2015, pp. 122-124.

Shtovba S.D. (2007). *Design of fuzzy systems tools MATLAB* // S.D. Shtovba., Moscow: Telecom.

Tosei Hatori, Mika Sato-Ilic. (2014), 'A fuzzy clustering method using the relative structure of the belongingness of objects to clusters', *Journal of Procedia Computer Science*, 35, pp. 994 – 1002.