Comparison of Different Cluttering Validity Methods in the Evaluation of Results for Finding Electrical Fault Location in Industrial MV Network Using Fuzzy Clustering Technique

Muhammad M.A.S. Mahmoud and Emin Husinli

January 10, 2020

Comparison of Different Cluttering Validity Methods in the Evaluation of Results for Electrical Fault Location in Industrial MV Network Using Fuzzy Clustering Technique

Muhammad M.A.S. Mahmoud - SM IEEE
Process Automation Engineering Department
Baku Highr Oil School
Baku, Azerbijan
muhammad.salih@socar.az

Amin Husinly
Process Automation Engineering Department
Baku Highr Oil School
Baku, Azerbijan
mmanar3@yahoo.com

Comprehensive evaluation is carried out to compare three different Cluttering Validity methods; Partition Coefficient, Partition Entropy and Proportional Exponent in the Evaluation of the results for Finding Electrical faults in industrial MV network using fuzzy clustering technique. Different data normalization methods and different range of alfa cut values for defuzzification are considered in the comparison process. The result shows that using Partition Entropy with Maximum Matrix normalization and 50% α cut gives the best effort saving of 95,15 in finding the fault.

Keywords— network distribution, fuzzy clustering, cluster validity, finding fault

I. INTRODUCTION

Finding electrical fault location in in oil & gas MV Network is necessary and main issue for continuous power delivery specially in old system, harsh environment and remote where the grid experience repeated faults. Any delay in finding the faults and repair it affects considerably reduction oil productivity. Therefore, the technique to find the faults needs to be efficient, fast and accurate as much as possible. Many years, researchers have done lot of efforts based on intelligent techniques in order to create effective solution. Fuzzy Clustering, Artificial Neural network, Expert Genetic Algorithm and System are example for intelligent programming techniques [1]. PC based techniques for fault finding are become very important for the fast results. Artificial Neural Networks (ANN) and Fuzzy Logic (FL) methods have recently gained popularity and proved successful in many practical problems [2] [3] [4].

In [5], the paper studied an existing 13.8 kilovolt distribution network which, serves an oil production field spread over an area of approximately 60 kilometers square, in order to locate any fault that may occur anywhere in the network using fuzzy c-mean classification techniques [6]. Two different methods for normalizing data and selecting the optimum number of clusters in order to classify data is introduced. Functional Coefficient Index was used to validate the clustering process. Results and conclusions are given to show the feasibility for the suggested fault location method.

In this paper we will extend the research of [5] to study the results in case Partition Coefficient, Partition Entropy and Proportional Exponent are used to validate the clustering and hence to find the faults [7][8][9].

In section II of this paper, description for the electrical network under discussion is provided to illustrate the complexity of the network. Section III discusses the data collection to construct the faults-feature matrix and the methods of normalizing these data to be suitable for clustering and validation process. Then, in section IV, indices of Partition Coefficient, Partition Entropy and Proportional Exponent is discussed. Section V discusses the fault location algorithm and procedures. The result of applying the above mentioned three clustering validity indices in the process of finding the fault in the network is also illustrated in this section. Summary and conclusion of the results is give in section VI.

II. ELECTRICAL NETWORK DISCRIPTION

An Oil Company possessed two production areas; Area1 and Area2 . Each area provided with power distribution system. Area1 power generation plant with 130 MVA capacity supplies power for Area 1 and send the needed power to Area 2 too. Also, Area 2 has local generation of 3.16 MVA capacity. Area 1 and Area 2 are electrically interconnected to transmit around 40 MVA from Area 1 to Area 2. At Area 2 (Fig. 1), the field oil-well loads are distributed among three wooden overhead transmission lines (OHTL). At Area 2 substation Smart relays are available to record any disturbance in the Area 2 network. The three OHTL are connected in mish configuration that add more completion to find the fault.

III. DATA COLLECTION AND FAULT FEATURE MATRIX

Load-Flow study is conducted to identify the pre-faultand post-fault active power and reactive power and hence the loss in respective power at each feeder for every short-circuit case.

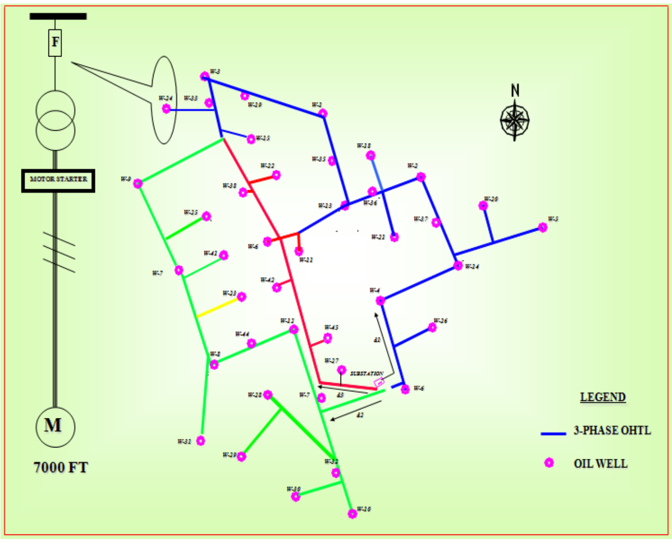

Fig. 1: Area 2 distribution network

TABLE 1: SUMMARY OF THE PARAMETERS THAT ARE SELECTED TO BUILD THE FEATURE MATRIX

Feeder 1	**Feeder 2**	**Feeder 3**
Set of nods fed from Feeder 1	Set of nods fed from Feeder 2	Set of nods fed from Feeder 3
Circuit breaker 1 status	Circuit breaker 2 status	Circuit breaker 3 status
Feeder 1 Short circuit Current red from substation	Feeder 2 Short circuit Current red from substation	Feeder 2 Short circuit Current red from substation
Phase Angel A1	Phase Angel A2	Phase Angel A3
Phase Angel B1	Phase Angel B2	Phase Angel B3
Phase Angel C1	Phase Angel C2	Phase Angel C3
Power dip in Feeder 1	Power dip in Feeder 2	Power dip in Feeder 3
VAR dip in Feeder 1	VAR dip in Feeder 2	VAR dip in Feeder 3

Short-circuit study is also carried out to identify the phase short circuit current and angle for each short circuit case and the expected circuit breaker trip status. The results of these two studies are used to construct the fault-feature-matrix for 144 nodes describing the faults for the network. Table 1 describes the parameter that have been selected to construct the fault feature matrix.

Because in the fault-feature-matrix there are wide range of values, normalization is necessary to make further calculation much easier during clustering and validation process. Two normalization methods are considered; Column-maximum and absolute matrix- maximum [5].

In normalization based on column, each value of matrix column is divided by the maximum values of the respective columns. So, it makes each value of data between 0 and 1. The second method for this is matrix normalization. However, in absolute matrix- maximum normalization, all matrix is divided by absolute maximum of whole matrix.

In the next section, data for thirteen (13) faults are selected to for testing the method of fault location. Other 131 data are used to create clusters to and to build the fault location algorithm.

Using technician experience in operating this network, it is possible to preliminary cluster this data. Based on Power dip and circuit breaker trip. Accordingly, initially data matrix can be classified into six different group.

a) 1st feeder power dip and circuit breaker trip

b) 1st feeder power dip and circuit breaker doesn’t trip

c) 2nd feeder power dip and circuit breaker trip

d) 2nd feeder power dip and circuit breaker doesn’t trip

e) 3rd feeder power dip and circuit breaker trip

f) 3rd feeder power dip and circuit breaker doesn’t trip

This technique gives us advance to improve c-means clustering performance. Nearest node will be searched in one group, not whole matrix.

In this paper we will utilize the advantage of operator experience with the use of absolute matrix-maximum normalization to develop the fault location algorithm as will be illustrated in Section V.

IV. FUZZY C-MEANS CLUSTERING ALGORITHM AND CLUSTER CALIDITY

The fuzzy c-means clustering algorithm given in [6] is used to carry out the data clustering.

The quality of a clustering is indicated by how closely the data points are associated to the cluster centers and it is the membership functions, which measure the level of association or classification. If the value of one of the membership is significantly larger than the others for a particular data point, then that point is identified as being a part of the subset of the data represented by the corresponding cluster center. But, each data point has c memberships; so, it is desirable to summarize the information contained in the memberships by a single number, which indicates how well the data point is classified by the clustering.

In [7], three cluster validity technique are proposed. First one is called *Partition Coefficient.*

$$\mathrm{PC} = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^2 \quad (1)$$

The second method is called *Partition (Classification) Entropy*

$$\mathrm{PE} = -\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij} * \log\mu_{ij} \quad (2)$$

There is also another cluster validity index which is called *Proportion Exponent.*

$$\mathrm{Pex} = \frac{1}{\mathrm{N}}\max(\mu_{\mathrm{ij}}) \quad (3)$$

For best clustering, the error given in (4) must be minimized

$$\mathrm{Error} = 1 - |(\mathrm{index})| \quad (4)$$

At the maximum value of partition coefficient, proportional exponent and minimum value of the partition entropy, efficient number of c-clusters is achieved. Closer this index to one, data is clustered more efficiently.

V. FAULT LOCATION ALGORITHM AND PROCEDURE

In this section, fuzzy c-means clustering technique is applied based on maximum matrix normalization. Through these steps result of classification and fault finding analyzed as following:

a) First absolute maximum matrix is found. Then, all the matrix is divided by this number. Accordingly all data are normalized between 0 and 1.

b) Based on the preliminary knowledge of electrical network, the possible fault location are found to be as shown in the foloowing table 2:

TABLE 2: PRELIMINARY CLUSTERING

Cases Description	Possible location
1st feeder power dip and circuit breaker trip	12
1st feeder power dip and circuit breaker doesn’t trip	6
2nd feeder power dip and circuit breaker trip	14
2nd feeder power dip and circuit breaker doesn’t trip	6
3rd feeder power dip and circuit breaker trip	9
3rd feeder power dip and circuit breaker doesn’t trip	7

c) All data is clustered based on FCM technique. Cluster validity technique is applied in order to determine the most efficient number of cluster centroid. In this case Partition Coefficient is implemented.

d) Euclidian distance is calculated between test data and the full data in the selected group is checked, based on each corresponding partition, cluster centroid for each data point is determined.

e) α -cut defuzzification is selected in a range in which 100% sucessful trials (5) is achieved wherethe nearest node to the fault is located in selected cluster.

$$\text{Succesfull Trials}(\%) = \frac{\text{Number of succesfull trial}}{\text{Number of testing case}(13)} * 100\% \quad (5)$$

Effort saving for each case is calculated based on the formulas (6) and (7):

$$\text{Effort Savings} = \left(1 - \frac{\text{Number of possible location}}{\text{all nodes}(144)}\right) * 100\% \quad (6)$$

$$\text{Average effort savings}\% = \frac{\text{Effort Savings}*100}{\text{Number of testing case}(13)}\% \quad (7)$$

It is worth to highlight here that α -cut virtual zone around the centroid as shown in Fig. 2, and accept all nodes inside the zone as a possible fault location. If test case is not inside of this zone, α -cut shall is to be increased in 0.1 increment, and so on until all test cases are found within α -cut circle. A few times code has been run to optimum the best choice for α – cut.

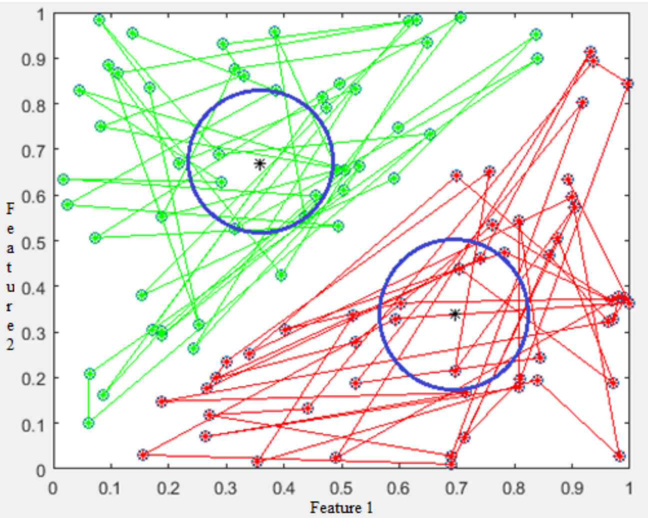

Fig. 2: α -cut presentation

Based on the above Number of possible fault locations have been achieved based on the three different validity indices methods and these techniques divided two group on their own: column normalization and matrix normalization. Different results and effort saving are calculated and compared. In addition, effect of the α -cut are examined t00oo.

Matlab program is written to implement the above six steps and the results are analyzed and summarized as follows tables:

TABLE 3: EFFORT SAVED USING MATRIX MAXIMUM NORMALIZATION (PARTITION COEFFICIENT)

Test cases	Number of Possible locations	Optimum number of clusters	Fault located in the cluster?	Saving Effort
1	9	11	YES	94%
2	11	11	YES	92%
3	4	5	YES	97%
4	13	12	YES	91%
5	8	6	YES	94%
6	1	5	YES	99%
7	12	11	YES	92%
8	14	12	YES	90%
9	3	5	YES	98%
10	12	13	YES	92%
11	10	12	YES	93%
12	3	6	YES	98%
13	8	6	YES	94%
Percentage of Successful trails		**100%**		
Average effort savings		**94.15384615%**		
α -cut coefficient		**0.6**		

TABLE 4: EFFORT SAVED IN COLUMN MAXIMUM NORMALIZATION (PARTITION ENTROPY -1)

Test cases	Number of Possible locations	Fault located in the cluster?	Effort Savings
1	27	YES	81%
2	27	YES	81%
3	11	YES	92%
4	33	YES	77%
5	9	YES	94%
6	10	YES	93%
7	6	YES	96%
8	3	YES	98%
9	11	YES	92%
10	33	YES	77%
11	33	YES	77%
12	9	YES	94%
13	6	YES	96%
Percentage of Successful trails		**100%**	
Average effort savings		**88.30769231%**	
α- cut coefficient		**0.4**	

TABLE 5: EFFORT SAVED IN COLUMN MAXIMUM NORMALIZATION (PARTITION ENTROPY-2)

Test cases	Number of Possible locations	Is nearest node located in the cluster?	Effort Savings
1	27	YES	81%
2	24	YES	83%
3	10	YES	93%
4	33	YES	77%
5	9	YES	94%
6	9	YES	94%
7	6	YES	96%
8	3	YES	98%
9	10	YES	93%
10	33	YES	77%
11	33	YES	77%
12	9	YES	94%
13	6	YES	96%
Percentage of Successful trails		**100%**	
Average effort savings		**88.69230769%**	
α cut coefficient		**0.3**	

TABLE 6: EFFORT SAVED IN MATRIX MAXIMUM NORMALIZATION (PARTITION ENTROPY-1)

Test cases	Number of Possible locations	Optimum number of clusters	Fault located in the cluster?	Saving Effort
1	6	9	YES	96%
2	3	9	YES	98%
3	3	6	YES	98%
4	**9**	11	YES	94%
5	7	7	YES	95%
6	1	6	YES	99%
7	10	9	YES	93%
8	14	11	YES	90%
9	3	6	YES	98%
10	11	11	YES	92%
11	11	11	YES	92%
12	3	7	YES	98%
13	8	8	YES	94%
Percentage of Successful trails		**100%**		
Average effort savings		**95.15384615%**		
α- cut coefficient		**0.5**		

TABLE 7: EFFORT SAVED IN MATRIX MAXIMUM NORMALIZATION (PARTITION ENTROPY-2)

Test cases	Number of Possible locations	Optimum number of clusters	Fault located in the cluster?	Saving Effort
1	8	9	YES	94
2	11	9	YES	92
3	4	6	YES	97
4	14	11	YES	90
5	8	7	YES	94
6	1	6	YES	99
7	11	9	YES	92
8	14	11	YES	90
9	3	6	YES	98
10	12	11	YES	92
11	10	11	YES	93
12	3	8	YES	98
13	8	7	YES	94
Percentage of Successful trails		**100%**		
Average effort savings		**94.0769231%**		
α -cut coefficient		**0.6**		

TABLE 8: EFFORT SAVED IN MATRIX MAXIMUM NNORMALIZATION (PARTITION ENTROPY-1)

Test cases	Number of Possible locations	Fault located in the cluster?	Effort Savings
1	27	YES	81%
2	27	YES	81%
3	11	YES	92%
4	33	YES	77%
5	9	YES	94%
6	9	YES	94%
7	6	YES	96%
8	3	YES	98%
9	10	YES	93%
10	33	YES	77%
11	33	YES	77%
12	9	YES	94%
13	6	YES	96%
Percentage of Successful trails		**100%**	
Average effort savings		**88.46153846%**	
α -cut coefficient		**0.4**	

TABLE 9: EFFORT SAVED IN COLUMN MAXIMUM NORMALIZATION (PROPORTIONAL EXPONENT-2)

Ttest cases	Number of Possible locations	Fault located in the cluster?	Effort Savings
1	27	YES	81%
2	22	YES	85%
3	9	YES	94%
4	28	YES	81%
5	9	YES	94%
6	9	YES	94%
7	6	YES	96%
8	3	YES	98%
9	10	YES	93%
10	32	YES	78%
11	20	YES	86%
12	9	YES	94%
13	6	YES	96%
Percentage of Successful trails		**100%**	
Average effort savings		**90%**	
α -cut coefficient		**0.3**	

TABLE 10: EFFORT SAVED IN MATRIX MAXIMUM NORMALIZATION (PROPORTIONAL EXPONENT)

Test cases	Number of Possible locations	Optimum number of clusters	Nearest node exist in fault locations	Saving Effort
1	5	10	YES	97%
2	11	10	YES	92%
3	4	5	YES	97%
4	13	11	YES	91%
5	8	6	YES	94%
6	1	4	YES	99%
7	11	10	YES	92%
8	14	11	YES	90%
9	3	5	YES	98%
10	12	11	YES	92%
11	10	11	YES	93%
12	3	6	YES	98%
13	8	6	YES	94%
Percentage of Successful trails		**100%**		
Average effort savings		**94.38461538%**		
α -cut coefficient		**0.6%**		

The results of above tables are summaries in Table 11, which shows that the best result obtained from the case of "Partial Entropy" with matrix maximum normalization and 50% alfa-cut

TABL 11: SUMMARY OF THE RESULTS

Index	Normalization	α -cut coefficient	Average effort saving
Partition Coefficient [5]	Column	60%	75%
Partition Coefficient [5]	Matrix	100%	87%
Partition Coefficient	Matrix	60%	94.15%
Partition Entropy	Column	40%	88.31%
Partition Entropy	Column	30%	88.7%
Partition Entropy	Matrix	60%	94.07%
Partition Entropy	**Matrix**	**50%**	**95.15%**
Proportional Exponent	Column	40%	88.46%
Proportional Exponent	Column	30%	90%
Proportional Exponent	Matrix	60%	94.38%

VI. Result analysis and conclusion

The main purpose of this paper is to compare the results of the method that was discussed in [5] using only Partition Coefficient cluster validation with another cluster validation indices in order to find efficient way to detect the fault location in oil field area. In this paper we compared 3 different cluster validity indices; Partition Coefficient, Partition Entropy and Proportional Exponent with the result obtained from [5]. In line with [5], two different normalization method are implemented. Operator experience are used to improve the clustering process. α -cut defuzzification technique is used and the value is selected in a range in which 100% successful trials is achieved. The results for all cases are coppered including the result obtained in [5]. It shows that that best result is obtained by the new procedure if "Partial Entropy" with matrix maximum normalization and 50% alfa-cut are used.

Acknowledgment

I would like to present this paper to my daughter Noor, my sons Ahmad and Kareem and Baku Higher Oil school, my second family.

References

[1] Muhammad M.A.S. Mahmoud, Zafar Qurbanov, "Review of fuzzy and ANN fault location methods for distribution power system in oil and gas sectors", International Federation of Automatic Control (IFAC 21018) , ELSEVER

[2] Muhammad M.A.S. Mahmoud, "Detection of singl line-to-ground faults through Impedence in Mesh distripution Network", Modern Electrical Power System MEPS15-IEEE, Wroclw, July 2015.

[3] Meshal A Al-Shaher, Manar M. sabry, Ahmad s saleh, Fault location in multi-ring distribution network using artificial neural network. Electrical Power System Research.vol. 64 CRL, pp 47-53, 2006.

[4] Muhammad M.A.S. Mahmoud, "New area in fuzzy applications" Fuzzy controls recent advanced in theoory and application (Book) Chaper 17, ISBN 978-953-51- 0759-0, PP 385 – 440. INTECH, 2012.

[5] Muhammad M.A.S. Mahmoud, "3-Phase fault finding in oil field MV distribution network using fuzzy clustering technique " , Journal of Energy and Power Engineering vol.7, pp. 155-161 2013.

[6] L.X. Wang, A Course in Fuzzy Systems and Control, Prentice-Hall International, Inc., USA, 1997

[7] M.P. Windham, Cluster validity for the fuzzy c-means clustering algorithm, IEEE Trans. Pattern Anal Machine Intell. PAMI-4 (4) (1982) 357-363.

[8] J.C. Bezdek, S.K. Pal, Fuzzy Models For Pattern Recognition, The Institution of Electrical and Electronics Engineering, Inc., USA, 1992.

[9] A New Cluster Validity Index for Fuzzy Clustering Sreeram Joopudi, Suraj S. Rathi, S. Narasimhana and Raghunathan Rengaswamy