Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, v. 6, n. 1, 2018.

Trabalho apresentado no XXXVII CNMAC, S.J. dos Campos - SP, 2017.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Proposal of an approach to data driven in fault diagnosis using fuzzy clustering techniques

Adrián Rodríguez Ramos[1]
José M. Bernal de Lázaro[2]
Orestes Llanes-Santiago[3]
Departamento de Automática y Computación, Universidad Tecnológica de la Habana "José Antonio Echeverría", La Habana, Cuba
Antônio José da Silva Neto[4]
Instituto Politécnico da Universidade do Estado do Rio de Janeiro (IPRJ/UERJ)

Abstract. In this work an approach to design data driven based fault diagnosis systems using fuzzy clustering techniques is presented. In the proposal, the data was first preprocessed using the Noise Clustering algorithm. This permits to eliminate outliers and reduce the confusion as a first part of the classification process. Secondly, the Kernel Fuzzy C-means algorithm was used to achieve greater separability among the classes, and reduce the classification errors. The proposed approach was validated using the nonlinear continuous stirred-tank reactor benchmark problem. The results obtained indicate the feasibility of the proposal.

Keywords. Fault diagnosis, fuzzy clustering, FCM, NC, KFCM.

1 Introduction

In industries the faults in equipments can have an unfavorable impact in the availability of the systems, the environment and the safety of operators. For such reason, the faults need to be detected and isolated, being these tasks associated to the fault diagnosis systems [1]. By performing an analysis of the different techniques developed in the recent years for control and fault diagnosis tasks, it is significative the increment in the use of the fuzzy clustering methods [2, 3].

The Fuzzy C-Means (FCM) algorithm, [4], obtains very good results with noise free data but are highly sensitive to noisy data and outliers [5]. Other similar techniques such as Possibilistic C-Means (PCM) [6] and Possibilistic Fuzzy C-Means (PFCM) [7] work better in presence of noise in comparison to FCM. Noise Clustering (NC) [8], Credibility

[1]adrian.rr@electrica.cujae.edu.cu
[2]jmbernal@crea.cujae.edu.cu
[3]orestes@tesla.cujae.edu.cu
[4]ajsneto@iprj.uerj.br

Fuzzy C-Means (CFCM) [9], and Density Oriented Fuzzy C-Means (DOFCM) algorithm [10] were proposed specifically to work efficiently with noisy data.

The clustering output depends upon various parameters such as distribution of data points inside and outside the cluster, shape of the cluster and linear or non-linear separability. The effectiveness of the clustering method relies highly on the choice of the distance metric adopted. Researchers have proposed other distance measures such as, for example, Kernel based distance measure in data space and in high dimensional feature space, such that non-hyper spherical/non-linear clusters can be detected [11]. Another problem usually present in fuzzy clustering methods is that their performance depend significantly on the initialization of their parameters.

In order to overcome these problems, in this work a fault diagnosis methodology, using fuzzy clustering techniques in the preprocessing and classification processes, is proposed to fault diagnosis.

The organization of the work is as follows: in section 2 a description of the proposed classification methodology using fuzzy clustering techniques is presented. The section 3 presents the experiment design, and the case study used to validate the proposed methodology. In section 4 an analysis of the obtained results is presented. Finally, the conclusions are posed.

2 Proposed classification methodology using fuzzy clustering

The classification scheme proposed in this work is shown in Fig. 1. In the first step, a set of N observations (data points) $X = [x_1, x_2, ..., x_N]$ are classified into c+*1* classes using the NC algorithm. The first c classes represent the faults to be diagnosed, as well as the normal operation conditions of the process, and they contain the data points to be used in the next step of the classification methodology. The other remaining class contains the data points identified as outliers, and they are not used in the next step.

In the second step, the Kernel Fuzzy C-Means (KFCM) algorithm [11] is applied. This algorithm receives the set of observations classified by the NC algorithm in the c classes as a set of observations to be classified. The KFCM algorithm maps these observations into a higher dimensional space in which the classification process obtains better results. Finally, it can be implemented a final step for optimizing the parameters of the NC and KFCM algorithms. In this work, the step 3 of optimization was not applied.

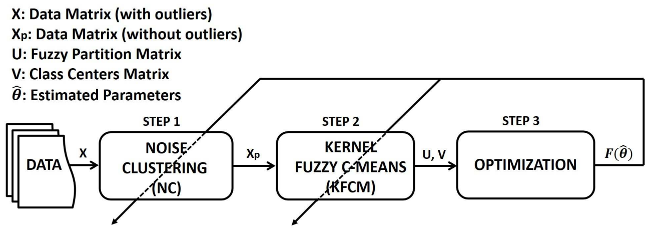

Figure 1: Classification scheme using fuzzy clustering.

3 Study case and experimental design

In order to apply the proposed methodology to fault diagnosis, the Continuous Stirred-Tank Reactor (CSTR) benchmark problem was selected [12]. The CSTR is composed of a tank, a cooling jacket, a stirrer, a pump, and instrumentation such as control valves, level, temperature and flow transmitters, and PI controllers. The mass and energy balance equations for describing the dynamics of the process can be seen in [12]. The controlled variables are the reactor temperature (T), and the tank level (h). The manipulated variables are the flow of coolant supply (Q_C) and the reactor output flow (Q). The CSTR and its feedback control system are shown in Fig. 2.

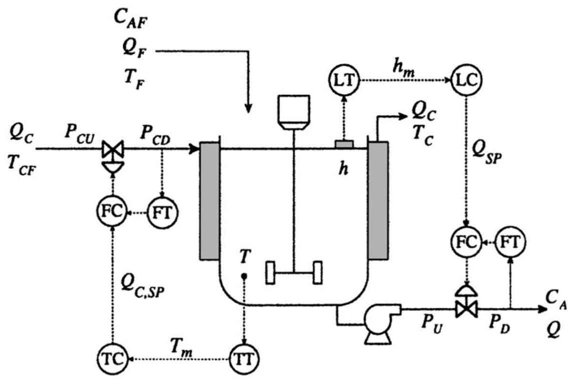

Figure 2: Continuous Stirred-Tank Reactor (CSTR) system with cascade control.

A set of measurements of the 14 process variables shown in Table 1, was stored with a sample time of 1 second. For each one of the six process states (Normal operation and the five faults shown in Table 2) were stored 1000 observations, leading to a total of 6000 observations. To this data set were added 600 new observations evenly distributed among the classes in order to represent the possible outliers for each class. Furthermore, white noise was added in the simulation to measurements and process variables in order to simulate the variability present in real world processes.

Table 1: Measured process variables.

Description	**Symbol**	**Description**	**Symbol**
Reactor Concentration	C_A	Feed concentration	C_{AF}
Reactor temperature	T	Feed temperature	T_F
Coolant temperature	T_C	Feed coolant temperature	T_{CF}
Reactor level	h	Level control signal	hC
Reactor Outlet flow	Q	Flow control signal	QC
Reactor Coolant flow	Q_C	Temperature control signal	TC
Reactor feed flow	Q_F	Coolant Flow control signal	Q_CC

The four experiments presented in Table 3 were performed. In the first and the second experiments, the step 1 (outliers determination) of the proposed diagnosis scheme was not applied. In the first experiment, the FCM algorithm was applied in the step 2 (classifica-

DOI: 10.5540/03.2018.006.01.0419

Table 2: Faults simulated in the CSTR benchmark problem

Fault	**Description**	**Nominal Value**
F6	Step change in Q_F	10 L/min
F7	Ramp change in C_{AF}	6.10^{-4}(mol/L)/min
F8	Ramp change in T_F	0.1 K/min
F9	Ramp change in T_{CF}	0.1 K/min
F11	Step change in downstream pressure P_D	5 psi

tion), and in the second experiment the KFCM algorithm was used. For the experiments 3 and 4 the NC algorithm was applied in the step 1, and the FCM and KFCM algorithms were applied in the second step, respectively. The values of the parameters used for the applied algorithms are: Number of iterations = 100, $\epsilon = 10^{-5}$, $m = 2$, $\sigma = 10$ (only used for the KFCM algorithm) and $\lambda = 0.01$ (only used for the NC algorithm), these values in the parameters are used in [10, 13]. The objective of these experiments is to study the joint effect of the data preprocessing to eliminate the outliers for the classification process and the kernel algorithm in the classification process.

Table 3: Experiments performed.

Experiment	Stage 1	Stage 2
1	-	FCM
2	-	KFCM
3	NC	FCM
4	NC	KFCM

4 Analysis of the results

4.1 Experiment 1

Table 4 shows the confusion matrix for experiment 1 where NOC: Normal Operation Condition, and the faults F6-F9 and F11 are described in Table 2. The main diagonal is associated with the number of observations successfully classified (OSC). Since the total number of observations per class is known, the accuracy (TA = OSC/Total), and the overall error (E = 1 - TA) can also be computed. The last row shows the average (AVE) of TA and E.

These results indicate the difficulty of the FCM algorithm to obtain satisfactory results in the classification in the presence of outliers. In addition, it is evident a great confusion between faults 7 and 8.

4.2 Experiment 2

As shown in Table 5, KFCM algorithm has similar difficulties to FCM in the classification process in the presence of outliers. This problem affects the correct classification of the different operating states.

Table 4: Confusion matrix for experiment 1: FCM (NOC: 1100, F6: 1100, F7: 1100, F8: 1100, F9: 1100, F11: 1100)

	NOC	F6	F7	F8	F9	F11	TA (%)	E (%)
NOC	**1001**	15	21	9	36	18	91.00	9.00
F6	5	**1009**	23	15	33	15	91.72	8.28
F7	1	24	**709**	319	31	16	64.45	35.55
F8	5	19	22	**1018**	24	12	92.54	7.46
F9	22	5	14	35	**1008**	16	91.63	8.37
F11	6	15	28	8	29	**1014**	92.18	7.82
AVE							**87.25**	**12.75**

Table 5: Confusion matrix for experiment 2: KFCM (NOC: 1100, F6: 1100, F7: 1100, F8: 1100, F9: 1100, F11: 1100)

	NOC	F6	F7	F8	F9	F11	TA (%)	E (%)
NOC	**1026**	19	14	24	5	12	93.27	6.73
F6	19	**1013**	18	5	30	15	92.09	7.91
F7	5	11	**737**	301	27	19	67.00	33.00
F8	24	19	22	**1018**	5	12	92.54	7.46
F9	22	18	10	4	**1028**	18	93.45	6.55
F11	29	15	7	28	6	**1015**	92.27	7.73
AVE							**88.44**	**11.56**

4.3 Experiment 3

Step 1
Table 6 shows that the NC algorithm classifies as outliers 595 observations (O class), which represents a 99.17% accuracy in classification. In addition, NC algorithm obtains good results in the classification of the other states although these classification results will not be used in the next step.
Step 2
After applying the algorithm FCM in the step 2 to classify the observations obtained from the step 1 the results are the same to those shown in Table 6 for the classes NOC, F6, F7, F8, F9 and F11. This is because the main difference between NC and FCM algorithms is the capacity to classify outliers. Therefore, when the data to classify are *clean* of outliers the results are similar.

4.4 Experiment 4

Step 1
The classification results observed with step 1 in this experiment are the same of as those obtained in experiment 3 (Table 6).

Table 6: Confusion matrix for experiment 3: NC (NOC: 1000, F6: 1000, F7: 1000, F8: 1000, F9: 1000, F11: 1000, O: 600)

	NOC	F6	F7	F8	F9	F11	O	TA (%)	E (%)
NOC	**1000**	0	0	0	0	0	0	100	0
F6	0	**994**	0	1	0	0	5	99.40	0.60
F7	0	0	**674**	326	0	0	0	67.40	32.60
F8	0	1	1	**998**	0	0	0	99.80	0.20
F9	0	0	0	1	**995**	4	0	99.50	0.50
F11	0	0	0	3	0	**997**	0	99.70	0.30
O	0	1	3	0	0	1	**595**	99.17	0.83
AVE								**94.99**	**5.01**

Step 2

Table 7 shows the confusion matrix. In this case better classification results are achieved compared with FCM algorithm, as a result of the better separability of the classes due to the application of the kernel function.

Table 7: Confusion matrix for experiment 4: NC-KFCM (NOC: 1000, F6: 1000, F7: 1000, F8: 1000, F9: 1000, F11: 1000)

	NOC	F6	F7	F8	F9	F11	TA (%)	E (%)
NOC	**1000**	0	0	0	0	0	100	0
F6	0	**999**	0	1	0	0	99.90	0.10
F7	0	0	**995**	5	0	0	99.50	0.50
F8	0	0	1	**999**	0	0	99.90	0.10
F9	0	0	0	1	**999**	0	99.90	0.10
F11	0	0	0	0	0	**1000**	100	0
AVE							**99.87**	**0.13**

5 Conclusions

In the present work a new classification scheme to fault diagnosis using fuzzy clustering techniques is proposed. In the proposal, the NC algorithm is used in a first step of preprocessing data to remove the outliers, and the KFCM algorithm is used in a second step of data classification to make use of the advantages introduced by the kernel function in the class separability, in order to obtain better classification results. Some experiments were performed and their results show the feasibility of the proposal. A possible third step could be used to optimize the parameters of the used algorithms.

6 Acknowledgements

The authors acknowledge the financial support provided by FAPERJ, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, CNPq, Conselho

Nacional de Desenvolvimento Científico e Tecnológico, CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, from Brazil.

References

[1] R. Isermann, Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors and fault-tolerant systems, Springer, 2011.

[2] M. Seera, C. P. Lim, C. K. Loo, H. Singh, A modified fuzzy min-max neural network for data clustering and its application to power quality monitoring, Applied Soft Computing 28 (2015) 19 – 29.

[3] A. T. Jahromi, M. J. Er, X. Li, B. S. Lim, Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis, Neurocomputing 196 (2016) 31 – 41.

[4] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.

[5] A. Gosain, S. Dahika, Performance analysis of various fuzzy clustering algorithms: A review, 7th International Conference on Communication, Computing and Virtualization 79 (2016) 100 – 111.

[6] R. Krishnapuram, J. M. Keller, A possibilistic approach to clustering, IEEE Transactions on Fuzzy Systems 1 (1993) 98 – 110.

[7] N. R. Pal, K. Pal, J. M. Keller, J. C. Bezdek, A possibilistic fuzzy c-means clustering algorithm, IEEE Transactions on Fuzzy Systems 13 (2005) 517 – 530.

[8] R. N. Dave, Characterization and detection of noise in clustering, Pattern Recognition Letters 12 (1991) 657 – 664.

[9] K. K. Chintalapudi, M. Kam, A noise resistant fuzzy c-means algorithm for clustering, In: IEEE Conference on Fuzzy Systems Proceedings 2 (1998) 1458 – 1463.

[10] P. Kaur, A density oriented fuzzy c-means clustering algorithm for recognising original cluster shapes from noisy data, Int. J. Innovative Computing and Applications 3 (2011) 77 – 87.

[11] D. Q. Zhang, S. C. Chen, A novel kernelized fuzzy c-means algorithm with application in medical image segmentation, Artificial Intelligence in Medicine 32 (2004) 37 – 50.

[12] M. Johannesmeyer, A. Singhal, D. Seborg, Pattern matching in historical data, AIChE Journal 48 (2002) 2022 – 2038.

[13] P. Kaur, A. Soni, A. Gosain, Robust kernelized approach to clustering by incorporating new distance measure, Engineering Applications of Artificial Intelligence 26 (2013) 833 – 847.

DOI: 10.5540/03.2018.006.01.0419