A supervised cooperative clustering scheme for diagnosing process faults in an industrial plant

Mohammad Anvaripour, Sima Soltanpour, Roozbeh Razavi-Far, Mehrdad Saif, Q.M. Jonathan Wu
Department of Electrical and Computer Engineering
University of Windsor
Ontario, Canada
Email: {anvaripm, soltanps, roozbeh, msaif, jwu}@uwindsor.ca

***Abstract*—This paper presents a novel supervised clustering technique including different clustering algorithms which cooperate together to span the decision space in a supervised manner. It uses a variety of clustering methods for an efficient partitioning. An evolutionary algorithm is used to tune the key parameters of the cooperative scheme which minimizes an error-based objective function on the training dataset. The proposed supervised scheme is developed for diagnosing faults in the Tennessee Eastman process, which is a standard benchmark for fault detection and diagnosis. Experimental results show that the proposed technique can efficiently diagnose the process faults.**

***Keywords*—cooperative clustering, genetic algorithm, classification, fault diagnosis**

I. Introduction

The increasing demands for the safety and the reliability of the systems result in increasing interests in fault detection and diagnosis systems. In the last few decades, numerous approaches have been successfully applied for fault detection and diagnosis in industries [1]. Data driven techniques have been extensively used for this purpose due to their accuracy and adaptability [2].

Contemporary data-driven diagnostic systems are intelligent classifiers that work in a supervised manner by using the label data from both normal and faulty classes to increase the sensitivity to the specific type of the fault [3][4]. The goal is then to label any unseen pattern in test dataset by means of the obtained knowledge from mapping the labeled data to their class or decision space during the training session [5].

Among these techniques, clustering algorithms including k-means [6], fuzzy c-means (FCM) [7] and Gaussian mixture model (GMM) [8] have been used for fault detection task. Although, these techniques are capable of identification of anomalies and detecting faults, they are not designed nor evaluated to diagnose the type or location of the faults, i.e., they cannot classify or isolate different classes of faults due to their unsupervised nature. These unsupervised clustering techniques detect the faults by thresholding the input observations representing any deviation from the normal behavior [3].

Thanks to evolutionary algorithms, clustering techniques have been adopted for multi-class classification and fault diagnosis in a supervised manner [9-11]. An evolutionary algorithm is used to optimize Mahalanobis metrics of FCM clustering algorithms, constructing a FCM classifier with a priori knowledge about clusters [10]. These techniques are extended in the form of bagged ensemble of FCM classifiers [11], [12], but only resampling techniques are used to create the diversity among the individual FCM models. To improve the performance, these techniques can only relay on increasing the diversity by resampling the training data, and thus, their performance cannot be improved where the individual FCM classifiers are doomed to misclassify the points around the complex decision boundaries. This is due to the nature of the partitive FCM algorithm.

There exist other clustering algorithms based on the density [13] and the connectivity [14] of the patterns. Moreover, combination of several clustering methods is one of the ongoing research directions. Cluster ensemble techniques combine different clustering methods to improve the quality and robustness of the results. These techniques are based on two steps, clusters generation and consensus function [15]. To improve the quality of clusters, cluster ensemble techniques resort to an aggregation or consensus function. However, they are computationally expensive due to creating and merging multiple partitions. On the other hand, cooperative clustering (CC) techniques have higher quality performance and lower computational complexity [16].

In this work, a supervised cooperative clustering scheme has been developed which makes use of various clustering algorithms for the sake of increasing the variety. This scheme contains three major modules: (a) individual unsupervised clustering algorithms; (b) cooperation module; (c) genetic algorithm optimization module. The general scheme of the supervised cooperative clustering is illustrated in Figure 1. The individual clustering techniques cooperate together for partitioning the data while their parameters are adjusted by minimizing an error-based objective function during the training with labeled data using genetic algorithm. Thus, the proposed novel scheme can be used as a powerful supervised algorithm for multiclass classification tasks, which is applied for diagnosing process faults in the Tennessee Eastman Process (TEP).

The rest of the paper is organized as follow. The problem statement and proposed approach is presented in the section 2. Tennessee Eastman Process (TEP) is presented in section 3. Experimental results and conclusion are provided in sections 4 and 5, respectively.

978-1-5090-0623-6/16/$31.00

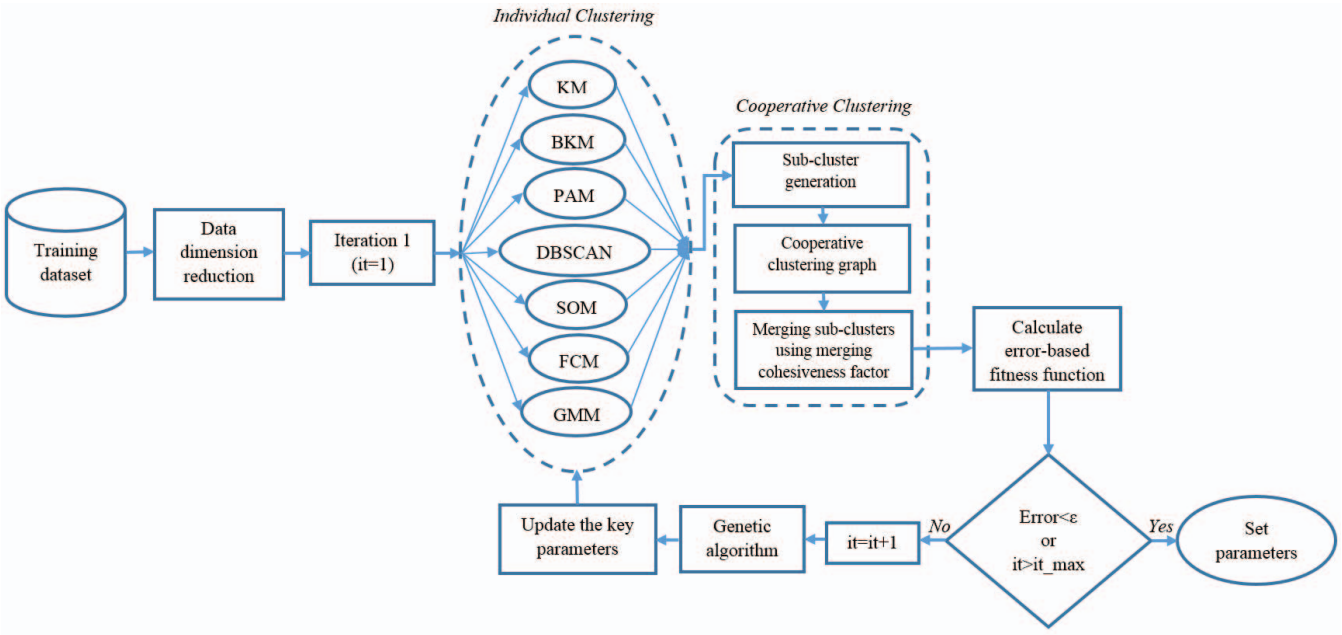

Fig. 1. General diagram of the proposed supervised cooperative clustering approach.

II. Problem Statement and The Proposed Approach

Thanks to an evolutionary algorithm, here, a novel cooperative clustering technique has been developed for fault diagnosis in a supervised manner. This scheme is based on various clustering algorithms which cooperate together in partitioning, aiming to classify the pattern of normal and faulty classes by resorting to genetic algorithm which tries to minimize an error-based objective function on the training data and find the best values for the key parameters. The selected individual clustering techniques and the cooperative clustering scheme are firstly presented as they are the prerequisites for the precise explanation of the proposed supervised cooperative clustering scheme.

A. Individual Clustering algorithms

In this paper different types of clustering methods is re-introduced and applied cooperatively to form better clusters. Methods have been chosen such a way to cover different types of clustering algorithms to gain their advantages over the others. Clustering techniques as an unsupervised approach aims to cluster data in different partitions. These partitions contain with meaningful groups of objects that share common characteristics. Different clustering methods are used including partitive and iterative based methods such as KM [17], BKM [18], PAM [19] and FCM [20], density based method such as DBSCAN [13], distribution based method such as GMM [21], and competitive learning such as SOM [22].

1) K-means: In this algorithm initial cluster's centroids are selected to the dataset and all the objects are assigned to these centroids by calculating the distance between objects to the centroids. The centers of the clusters are found and will be updated by iterative K-means until minimizing the squared error as well as sum of distances of the points objective function to the centers.

2) Bisecting K-means: In this method the whole dataset is considered as one cluster and then partitioned in two parts by K-means algorithm. The process is continued to achieve the desired number of clusters.

3) Partitioning around medoids (PAM): In this approach a representative object are selected in each cluster at each iteration. To find medoids, some objects are selected to minimize an objective function which is the distances of all the objects in a cluster to the cluster medoid. It is a significance approach in noisy and small datasets but it is not useful for the large dataset.

4) Fuzzy C Means (FCM): FCM which is indeed a fuzzy version of K-means assigns each point to each cluster by means of a degree of membership. In this algorithm every point has a degree of membership to all clusters, instead of being completely members of just one cluster.

5) DBSCAN: DBSCAN is an effective clustering method to discover clusters in the dataset using local density of points [13]. DBSCAN contains two adjustable parameters, Minpts and Eps which Minpts are the number of points in a circle with radius of Eps. The underlying idea is to have at least one point near to the point belongs to a cluster with the distance less than Eps. In DBSCAN clustering, we have two kinds of points, border and core points. Border points belong to the neighborhood within Eps radius of a core point and a core point contains minimum number of points MinPts within its neighborhood with Eps radius.

6) Gaussian Mixture Model (GMM): In GMM multivariate normal density components are combined that each point belongs to a cluster according to a probability of belonging

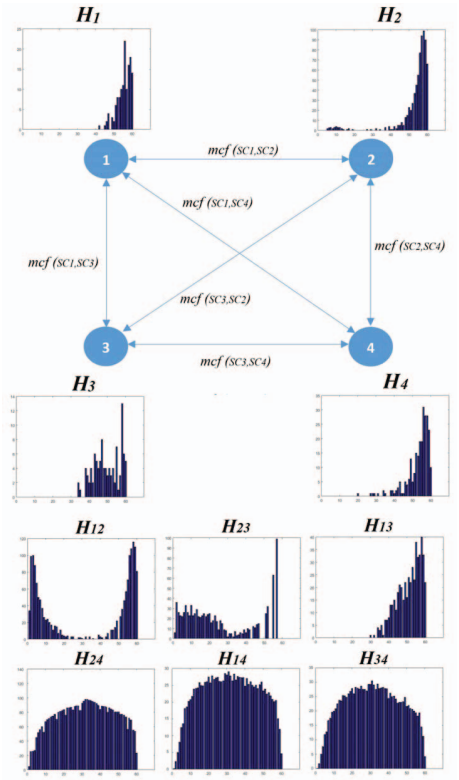

Fig. 2. Cooperative Clustering graph in CC method. The two sub-clusters with higher mcf are merged. H_i indicates the similarity histogram of sub-cluster i and H_{ij} indicates the similarity histogram between two sub-clusters i and j.

to each cluster. By selecting the component to maximize the posterior probability, clustering is performed. GMM has been used in fault clustering tasks [8].

7) Self Organizing Map (SOM) : In this method input space is projected on prototypes of a low-dimensional regular grid to be effectively utilized to visualize and explore properties of the data [22]. Where the number of units are large, similar units are grouped together.

B. Cooperative clustering (CC)

Cooperative clustering (CC) uses different clustering algorithms cooperatively to achieve the high quality clusters. Thus, input to the cooperative clustering is the data to be partitioned and M different clustering algorithms which they have to cluster the data into the c clusters. The dataset is defined by X of size $n \times k$ matrix $X = \{x_i\}$, $i = 1, ..., n$ where n is equal to the number of patterns and each of them has k attributes or features. Similarity (or distance) Matrix (SM), represents pair wise cosine similarity between patterns. Since SM is a symmetric matrix, we only consider $n \times (n-1)/2$ elements. In addition, each member of SM has value between -1 and 1.

Membership, $mem(x)$, of each point $x \in X$ in the dataset can be calculated by the weights multiply to the assigned label of each individual clustering algorithm as follows:

$$mem(x)|_{c_1,c_2,...,c_M} = w_1 mem(x)|_{c_1} \times ... \times w_M mem(x)|_{c_M} \quad (1)$$

where $mem(x)|_{c_i}$ is the class label of pattern x using i-th clustering method c_i, $i = 1, ..., M$, where M is the number of clustering algorithms and w_i, where $i = 1, ..., M$, is the weight of each clustering algorithm in voting a label, which can be obtained during the training. Given two different data patterns x_1 and x_2, these two patterns can only belong to the same cluster or sub-cluster if $mem(x_1) = mem(x_2)$, see equation (1). Since each clustering method may assign different label to one pattern or object and have different results, the number of sub-clusters is more than or equal to the data clusters.

To represent each sub-cluster, similarity histogram as a cosine statistical representation, is used as in [16]. Constant similarity value intervals are used to determine the number of bins (N), which is a preset parameter. Since cosine similarity is between -1 and 1, size of each bin (b) in the histogram is equal to $2/N$. For every bin (b), histogram contains number of similarities that fall into the same interval. Cooperative contingency is considered as a graph where sub-clusters (SCs) and their relationships are its nodes and edges, respectively. The coherency of merging is applied to merge two sub-clusters. Fig. 2 is an example of the cooperative contingency graph with four SCs. Histograms are updated after merging as follows:

$$H_{pq}(b) = H_p(b) + H_q(b) + |Sim(x_1, x_2)|$$
$$\forall x_1 \in SC_p, x_2 \in SC_q, b = 0, ..., N-1 \quad (2)$$

where $(b - (N/2)) \times bs < Sim(x_1, x_2) \leq (b - (N/2)) \times bs + bs$. In the above equation,$bs$ is the bin size, $Sim(x_1, x_2)$ is the similarity of two patterns x_1 and x_2, H_{pq} is the new cluster histogram obtained by merging two sub-clusters p and q, $H_p(b)$ is the number of similarities for the bin number equal to b^{th} bin from similarity histogram H_p. Number of additional similarities from merging is defined by $|Sim(x_1, x_2)|$. Merging cohesiveness factor of two sub-clusters (SCs), $(mcf(SC_p, SC_q))$, is equal to the weighted number of the bin similarities $(H_{pq}(b))$ considering a threshold δ over the total count of similarities $(n_{Sim}(SC_p, SC_q))$ which can be reformulated as follows:

$$mcf(SC_p, SC_q) = \frac{\Sigma_{b=b\_th}^{N-1}((b \times bs) - 1 + (bs/2)) \times H_{pq}(b)}{n_{Sim}(SC_p, SC_q)} \quad (3)$$

where b_th is the bin corresponding to the similarity threshold δ.

At each iteration two SC with the maximum mcf are merged to form a new SC and similarity histograms of the new SCs are calculated and cooperative contingency graph is formed. This process is continued until the desired number of clusters are extracted.

Two SCs are similar to each other, if their mcf is high enough and also their similarity histogram has more bins close to 1 on the right hand side of the similarity histogram. As

Fig. 2 illustrates, H_{13} has more bins in the right side of the histogram and highest bins close to 1. It shows the higher similarities between these two SCs, 1 and 3, and thus they can be merged together to make a new sub-cluster.

C. Supervised Cooperative Clustering (SCC)

As mentioned above, clustering is an unsupervised learning task where the information about class labels are not available a priori. To diagnose faults in industrial plants, a supervised CC, (SCC), is developed. In this scheme some parameters of the clustering techniques and their weights are optimized in a supervised fashion by means of genetic algorithm [23].

GA makes use of the 'stochastic uniform selection' mechanism to select a preset number of individuals from the population. It also uses the 'scattered crossover' operator to create offspring for the next generation, and the 'uniform selection' operator for reinsertion of offspring in the population.

SCC makes use of different types of clustering algorithms cooperatively, which need to be optimized. There is a list of parameters: K-means: c, w; BKM: c, w; PAM: c, w; FCM: c, m, w; DBSCAN: $Eps, Minpts, w$; GMM: G, w; SOM: D, w; CC: b_{th}, where c is the number of initial clusters that each individual clustering method has to cluster the data to the number of c clusters, w is the weight that is multiplied to the labels, Eps and $Minpts$ are parameters of the DBSCAN, G is the number of Gaussian components, D is the number of layer of the SOM, and b_{th} is the bin corresponding to the similarity threshold in the CC method (See equation 3). To perform SCC, these parameters are updated in a supervised manner to obtain final clusters as close as possible to the a priori known classes based on the available labeled data. The supervision is then the process of minimizing the error function $E(\Gamma, \Gamma^*)$ using genetic algorithm between the a priori known class label Γ and the obtained class label Γ^* as follows:

$$E(\Gamma, \Gamma^*) = \frac{\text{Count of misclassified patterns}}{\text{Total number of patterns}} \qquad (4)$$

The overall training step of the proposed supervised clustering is as follows:

1. Initialize first iteration to 1 ($it = 1$) and randomly initialize all other parameters.

2. At each step of the genetic algorithm, M different clustering methods partition n training data into c clusters. By applying partitioning results of M clustering methods to cooperative clustering with the preset parameters vector, data is clustered into the desired number of classes at the iteration t $\Gamma^*(t) = (\Gamma_1^*(t), ..., \Gamma_c^*(t))$.

3. Compute the error function using equation (4).

4. If E is less than the predefined threshold ε, or the number of iterations meet the maximum value, stop; otherwise, increment it by 1 and update parameters vector $\{c^t, m^t, Eps^t, Minpts^t, G^t, D^t, b\_th^t, w_1^t, w_2^t, \ldots, w_M^t\}$ to $\{c^{t+}, m^{t+}, Eps^{t+}, Minpts^{t+}, G^{t+}, D^{t+}, b\_th^{t+}, w_1^{t+}, \ldots, w_M^{t+}\}$.

5. Return to step 2. Upon convergence the classifier is constructed with best fit of the model parameters.

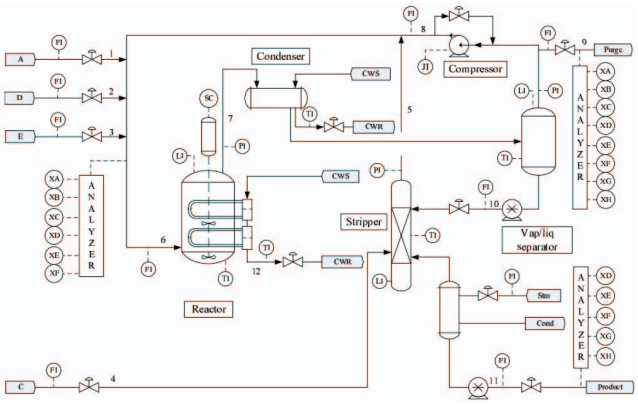

Fig. 3. Tennessee Eastman Process Benchmark [25].

Once a new test pattern becomes available its label can be predicted by means of the SCC model whose parameters are adjusted during the training phase.

III. Tennessee Eastman Process (TEP)

TEP [24] is a model of a realistic industrial process. The model is widely studied as a challenging benchmark for control and monitoring. There are five major units, i.e., reactor, condenser, separator, compressor and stripper in the model for the whole process. The process contains eight components including inert and byproduct: A, B, C, D, E, F, G, and H, in which G and H are two desired products. A, B, C, D, E, and F are gaseous reactants which are fed to the reactor to produce liquid products G and H. At the same time, the two additional byproduct reactions occur too. The process has 41 measurements, i.e., XMEAS(1-41), and 12 manipulated variables, i.e., XMV(1-12). The measurements or XMEAS include 22 continuous process measurements and 19 sampled process measurements. There are 20 different faults in the model that they have been designed based on the details in [25] by XMEAS and XMV. The data set used in this paper are given in [25] and is popular for process monitoring and fault diagnosis in researches. The data set includes 22 training sets and 22 testing sets of faults and normal process. One training and one testing set was obtained under normal operational condition and the other 21 sets were collected under 21 different faulty conditions for 48 operation hours and 960 samples are obtained for each test data. Each fault data is 960×52 dimension. For each of 21 faulty testing sets, the fault was introduced at 8th operation hour where it is 161th sample. Fig. 3 illustrates schematic diagram of the TEP simulation. Fig. 4 shows the TEP data of the two classes of normal and fault 2 the first two dimension of the data (XMEAS 1 and XMEAS 2).

IV. Experimental Results

In this section, the proposed Supervised Cooperative Clustering (SCC) approach is evaluated for fault detection and diagnosis by Tennessee Eastman Process (TEP) challenging dataset which is described in section 3. To form the SCC,

TABLE I
FAULT DETECTION RATE USING SPARSE FEATURES (SR) AND PRINCIPAL COMPONENTS (PCA).

Fault	Fault Type	SR	PCA
1	A/C Feed Ratio, Stream 4	99.95%	98.95%
2	B Composition, Stream 4	99.6 %	99.1 %
3	D Feed Temperature Step Stream 2	96.3%	96.3%
4	Reactor Cooling Step Water Inlet Temperature	99.37%	98.37%
5	Condenser Cooling Step Water Inlet Temperature	99.75%	99.35%
6	A Feed Loss Step Stream 1	90%	89.2%
7	C Header Pressure Stream 4	99.97%	99.96%
8	A, B, C Feed Composition Random Variation Stream 4	88%	83%
9	D Feed Temperature Random Variation Stream 2	90%	91%
10	C Feed Temperature Random Variation Stream 4	99.75%	99.12%
11	Reactor Cooling	98.5	98.6%
12	Condenser Cooling	95%	95.8%
13	Reaction Kinetics Slow Drift	98.2%	98.8%
14	Reactor Cooling Sticking Water Valve	99.4%	93.4%
15	Condenser Cooling Sticking Water Valve	98%	96.9%
16	Unknown	99.5%	99.5%
17	Unknown	95%	95%
18	Unknown	98%	97%
19	Unknown	99.5%	98.5%
20	Unknown	95%	94%
	Average	**96.94**%	**96.09**%

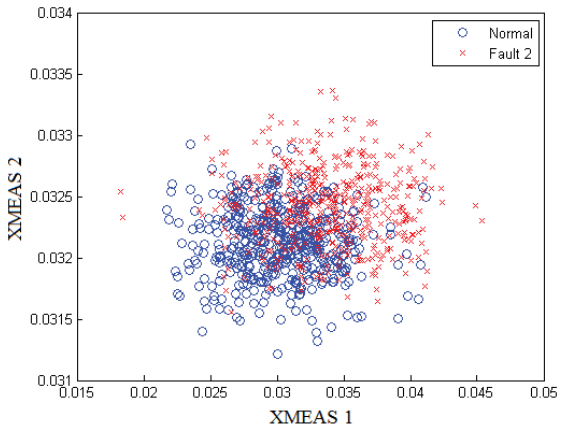

Fig. 4. TEP data of the two classes of normal and fault 2 in two dimension of the XMEAS 1 and XMEAS 2.

some clustering methods such as K-means, BKM, PAM, FCM, DBSCAN, GMM and SOM are used. Principle component analysis (PCA) [26] and Sparse Features (SR) [27] have been applied to reduce the dimension and extract significant features for the use of SCC.

A. Fault Detection

1) Clustering and fault detection using PCA: PCA is utilized for projecting the high-dimensional feature into a more meaningful and lower dimensional space. The TEP data is in 52 dimensions. To have a better performance the dimension has been reduced to 7 using PCA. The resulting data to be applied for clustering has 960 $\times$ 7 dimensions where the first 480 rows are normal and the remaining samples are faulty. Fault detection has been conducted for all the faults. Table 1 reports the detection accuracy by PCA dimension reduction technique. Fig. 5 illustrates the decision boundaries of the two classes of normal and fault 2 in two dimensions of the first two components of the PCA. Fig. 5 shows that SCC has more accurate decision boundaries than the other individual clustering methods.

2) Clustering and fault detection using Sparse Features: Sparse features (SR) [27] is an effective approach in which a linear combination of basis vectors approximates a sample and reduce the dimension. Redundant and non-orthogonal basis vector can be learned using SR. One of main advantages of this technique is its robustness to redundancy and noise. For a given basis matrix (a) sparse coefficients of samples are firstly extracted and then used for feature selection. Following expression represents formulation of the SR model [27]:

$$(a|B,s,k) = s_1b_1 + s_2b_2 + ... + s_kb_k + \varepsilon = Bs + \varepsilon \quad (5)$$

where dimension of a, B, s and ε is $d \times 1, d \times k, k \times 1$, and $d \times 1$ respectively. k is sparse features dimension. $B = [b_1, b_2, ..., b_k]$ and b_i are dictionary and dictionary atom respectively, also s is a sparse coefficient vector, and ε is an error term. In the above equation, B, s, and k are parameters of the model. The error term is defined using a Gaussian distribution with zero mean and isotropic covariance as $\varepsilon \sim N(0, \Sigma)$ and $\Sigma = \omega I$ where ω is equal to positive scalar and I is identity matrix. Also $b_i \sim N(0, Z)$, $Z = I$ with zero mean and identity covariance I. By considering Laplacian prior with zero mean and isotropic covariance for the coefficient vector and the Gaussian distributed likelihood, the posterior is stated as follows:

$$p(s|B,a,\Sigma,\Gamma) = p(a|B,s,\Sigma,\Gamma)p(s|B,\Sigma,\Gamma)/p(a) \quad (6)$$
$$\propto p(a|B,s,\Sigma)p(s|\Gamma)$$

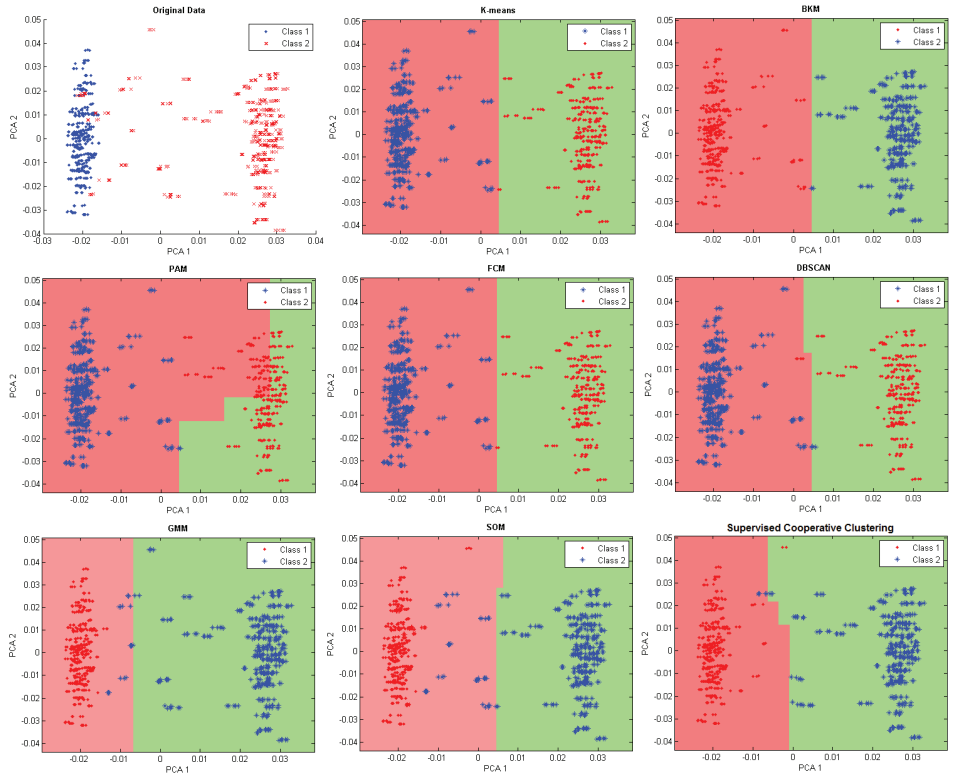

Fig. 5. Decision boundaries over the first two principle components.

$$p(s|\Gamma) = L(0,\Gamma) = e^{-\|s\|_1/\gamma}/(2\gamma)^k \tag{7}$$

$$p(a|B,s,\Sigma) = N(Bs,\Sigma) = e^{-\|a-Bs\|_2^2/2\omega}/(2\pi)^{d/2}\omega^{d/2} \tag{8}$$

By substitution equations (7) and (8) in equation (6) we have

$$L(s) = logp(a|B,s)+logp(s) = -\|a-Bs\|_2^2/2\omega-\|s\|_1/\gamma+v \tag{9}$$

where v is a constant value. We extract sparse coding model from data by minimizing expression 6 which is equivalent to maximizing the posteriori.

$$argminf(s) = 0.5\|a-Bs\|_2^2 + \lambda\|s\|_1 \tag{10}$$

where $\lambda = \omega/\gamma$ controls the trade off between sparsity and error. The dimension of the data is reduced to 7 in this paper using SR.

Fig. 6 illustrates the classification results using SR features. These plot shows a better separation of the data of different classes on the first two sparse features which enhances the classification accuracy. Table 1 reports fault detection rate using Sparse features.

As it can be seen from the Table 1 there exist some differences between the PCA and SR features results. Although the process of SR is taking more times than PCA, it has better performance comparing to PCA.

B. Fault Classification

For the classification tasks, the two types of dimension reduction methods have been applied. Three scenarios are

TABLE II
CLASSIFICATION PERFORMANCE OF THE FIRST SCENARIO.

	SR	PCA
Fault 1	92.8	91.5%
Fault 7	92	91%
Fault 10	100%	98%

TABLE III
CLASSIFICATION PERFORMANCE OF THE SECOND SCENARIO.

	SR	PCA
Normal	100%	100%
Fault 1	92%	91.8%
Fault 7	94%	94.3%
Fault 10	100%	100%

TABLE IV
CLASSIFICATION PERFORMANCE OF THE THIRD SCENARIO.

	SR	PCA
Normal	100%	100%
Fault 1	92%	91.8%
Fault 2	96.5%	95.3%
Fault 7	94%	94%
Fault 10	98%	97%

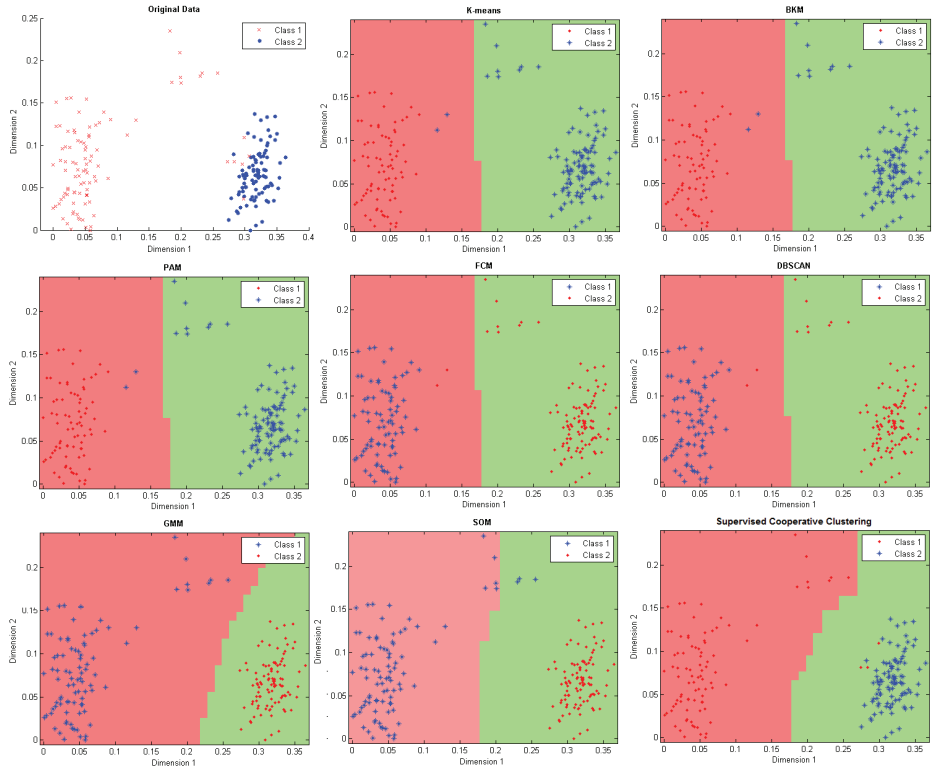

Fig. 6. Decision boundaries over the first two sparse features.

implemented for fault classification and diagnosis. At the first scenario, three kinds of faults, 1, 7, and 10 have been selected for classification. Training data includes the pattern of these three faults in the following form:

$$\begin{cases} X_{train} = \left(X_{train1}, X_{train7}, X_{train10}\right) \\ X_{test} = \left(X_{test1}, X_{test7}, X_{test10}\right) \end{cases} \tag{11}$$

where X_{train} denotes the training patterns and is a matrix of size 1440×52. Test data, X_{test}, includes the test patterns and is a matrix of size 2400×52. Table 2 shows the fault classification and diagnosis results of these faults that are A/C feed ratio, C header pressure loss, and D feed temperatures that all of them are coming from the stream 4. As the Table 2 reports, classification using SR features outperform PCA components.

At the second step, normal class has been added to these three fault classes. In this experiment, 4 classes of normal, fault 1, fault 7, and fault 10 are classified. The training matrix would be 1920×52 and test matrix would be of size 3200×52.

$$\begin{cases} X_{train} = \left(X_{Normal}, X_{Train1}, X_{Train7}, X_{Train10}\right) \\ X_{test} = \left(X_{Normal}, X_{test1}, X_{test7}, X_{test10}\right) \end{cases} \tag{12}$$

Table 3 illustrates the reasonable classification performance of the proposed approach for four classes in the presence of the normal data. Similar to the previous scenario, classification using SR features has better performance.

At the third scenario, fault 2 has been added to the previous four classes. In this experiment, 5 classes of normal and faults are classified. The training matrix would be 2400×52 and test matrix would be of size 4000×52.

$$\begin{cases} X_{train} = \left(X_{Normal}, X_{Train1}, X_{Train2}, X_{Train7}, X_{Train10}\right) \\ X_{test} = \left(X_{Normal}, X_{test1}, X_{test2}, X_{test7}, X_{test10}\right) \end{cases} \tag{13}$$

Table 4 illustrates the acceptable classification performance of the proposed approach for four classes in the presence of the normal data. Similar to the previous scenario classification using SR has better performance.

All the above experiments indicate that the proposed approach has a good classification performance in fault diagnosis and is reliable to be implemented in industrial setting for controlling and monitoring of the process, even in the presence of the multiple faults.

C. Performance of the base classifiers during the training

Fig. 7 illustrates the weights of the base classifiers, w, during 30 generation of the genetic algorithm for classification of the normal, fault 1, fault 7, and fault 10. The weights of each classification algorithm have been updated iteratively in each generation. It shows that the weight of the GMM is increasing during the training process. Fig. 7 shows that base classifiers take random initial weights, but the weights to GMM dominates other weights at last step of the classification process during second scenario.

Best and mean penalty values of the genetic algorithm for the 30 generations during the training, for classification of the

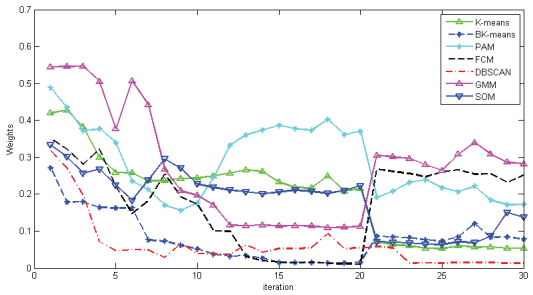

Fig. 7. Changes of the weights assigned to individual clustering methods during the optimization and training of the normal and faults 1, 7, and 10.

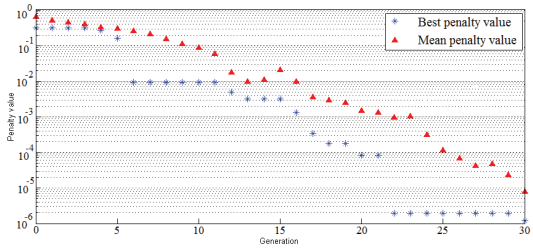

Fig. 8. Best and mean penalty value of the genetic algorithm during 30 generations of the training for classification of the normal and faults 1, 7, and 10.

normal and faults 1, 7, and 10 are presented in Fig. 8. The best fit is decreasing along the 30 generations of the SCC and the average penalty value decreases to match with the best fit, which shows that genetic algorithm properly optimizes the parameters during the training.

V. Conclusion

In this paper a method for fault detection and diagnosis has been proposed by means of clustering techniques which cooperate together and optimize in a supervised fashion. This method achieves high classification performance. Different kinds of unsupervised clustering methods have been used in the proposed cooperative scheme to generate more reliable decision boundaries. Genetic algorithm has been applied to optimize the parameters of the cooperative clustering including bin corresponding to the threshold, the number of the initial clusters, and the key parameters of individual clustering methods such as parameters in FCM, DBSCAN, GMM, and SOM. The advantage of the proposed scheme is to have multiple clustering algorithms which act cooperatively and also generate multiple decesion boundaries which can reduce the total error. Sparse features and PCA have been utilized for dimension reduction. The proposed scheme has been examined by Tennessee Eastman Process. The attained results show the good performance of the proposed approach in fault detection and diagnosis including multiple faults. Further research direction would be to study the proposed approach on different datasets and to improve the optimization scheme.

References

[1] X. Wang, U. Kruger, G. Irwin, G. McCullough, and N. McDowell, "Nonlinear PCA with the local approach for diesel engine fault detection and diagnosis", IEEE Trans. Control Syst. Technol., vol. 16, no. 1, pp. 122-129, Jan. 2008.
[2] G. Li, S. J. Qin, and D. Zhou,"A New Method of Dynamic Latent-Variable Modeling for Process Monitoring", IEEE Trans. Industrial Electronics., vol. 61, no. 11, pp. 6438-6445, 2014.
[3] M. Grbovic, W. Li, N. A. Subrahmanya, A. K. Usadi, and S. Vucetic, "Cold Start Approach for Data-Driven Fault Detection", IEEE Trans. on Industrial Informatics, vol. 9, no. 4, pp. 2264-2273, 2013.
[4] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines", IEEE Trans Neur. Net., vol. 13(2), pp. 415-425, 2002.
[5] A. K. Jain, R. P. W. Duin and J. Mao, "Statistical Pattern Recognition: A Review", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 4-37, Jan. 2000.
[6] Q. P. He, J. Wang, and S. J. Qin, "A new fault diagnosis method using fault directions in Fisher discriminant analysis", AIChE, vol. 51, no. 2, pp. 555-571, 2005.
[7] P. Teppola, and P. Minkkinen, "Possibilistic and fuzzy c-means clustering for process monitoring in an activated sludge waste-water treatment plant", Journal of Chemometrics, vol. 13, no. 3-4, pp. 445-459, 1999.
[8] I. Kiss, P. Haller, and A. Beres, "Denial of Service attack detection in case of Tennessee Eastman challenge process", Procedia Technology, vol. 19, pp. 835-841, 2015.
[9] C. Yin, X. Zhao and S. Mu, "A fast multiclass classification algorithm based on cooperative clustering", Neural Processing Letters, vol. 38, no. 3, pp. 389-402, 2013.
[10] E. Zio, and P. Baraldi, "Identification of nuclear transents via optimized fuzzy clustering", Ann. Nucl. Energy, vol. 32(10), pp. 1068-1080, 2005.
[11] P. Baraldi, R. Razavi, and E. Zio, "Bagged ensemble of fuzzy c-means classifiers for nuclear transient identification". Ann. Nucl. Energy, vol. 38, no. 5, pp. 1161-1171, 2010.
[12] P. Baraldi, and R. Razavi-Far, and E. Zio, A method for estimating the confidence in the identification of nuclear transients by a bagged ensemble of FCM classifiers, 7th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT), vol. 1, pp. 283-293, 2010.
[13] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", 2nd Int. Conf. Knowl. Discov. & Data Min., vol. 96(34), pp. 226-231, 1996.
[14] T. Zhang, "Data clustering for very large datasets plus applications", PhD dissertation, Dept. Computer Science, University of Wisconsin, Madison, 1997.
[15] A. Strehl, and J. Ghosh, "Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions", Journal of Machine Learning Research, vol. 3, pp. 583-617, 2003.
[16] R. Kashef and M. S. Kamel, "Cooperative clustering", Pattern Recognition, vol. 43, no. 6, pp. 2315-2329, 2010.
[17] J. A. Hartigan, and M. A. Wong, "A k-means clustering algorithm", Applied Statistics, vol. 28, no. 1, pp. 100-108, 1979.
[18] S. M. Savaresi, and D. Boley, "On the performance of bisecting K-means and PDDP", in: Proceedings of the First SIAM International Conference on Data Mining, pp. 114, 2001.
[19] L. Kaufmann, and P. Rousseeuw, "Finding Groups in Data", Wiley, New York, 1990.
[20] J. C. Bezdek, R. Ehrlich and W. Full, "The fuzzy C-means clustering algorithm", Computers and Geosciences, vol. 10(2), pp. 191-203, 1984.
[21] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction", Proceeding at International Conference on Pattern Recognition (ICPR), pp. 28-31, 2004.
[22] J. Vesanto, E. Alhoniemi, "Clustering of the Self-Organizing Map", IEEE Transaction On Neural Networks, vol. 11, no. 3, pp. 586-600, 2000.
[23] M. Melanie, "An Introduction to Genetic Algorithms", Massachusetts: MIT Press, 1999.
[24] J. J. Downs, and E. F. Vogel, "A plant wide industrial process control problem", Computers and Chemical eng., vol. 17(3), pp. 245-255, 1993.
[25] L. H. Chiang, E. L. Russel, and R. D. Braatz, "Fault detection and diagnosis in industrial systems", New York, Springer, 2003.
[26] I. Jolliffe, "Principal Component Analysis", 2nd Edition, Springer, 2002.
[27] Y. Li, and A. Ngom, "Supervised dictionary learning via non-negative matrix factorization for classification", IEEE conference on Machine Learning and Applications (ICMLA), pp. 439-443, 2012.