International Journal of Emerging Engineering Research and Technology
Volume 5, Issue 5, 2017, PP 30-39
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

A Hierarchical Multi-Label Classification Approach for Fault Diagnosis in Industrial Units

Mehdi Namdari*, Hooshang Jazayeri-Rad[2]

[2] *Departments of Instrumentation and Automation, Petroleum University of Technology, Ahwaz, Iran*

***Corresponding Author:** Departments of Instrumentation and Automation, Petroleum University of Technology, Ahwaz, Iran. Email: nmdri.mehdi@gmail.com*

Received Date: 29-08-2017 ***Accepted Date: 11-09-2017*** ***Published Date: 06-10-2017***

ABSTRACT

A new approach for enhancing the effectiveness of fault diagnosis algorithms based on classification tools has been suggested. The main idea in the anticipated arrangement relies on the utilization of a hierarchical categorization structure which decomposes a large-scale categorization problem into several sub-categorization tasks. Also, the proposed diagnosis system is founded on the multi-label classification which provides more than one fault predictions at a given time and hence is capable of diagnosing multiple simultaneous faults.

Keywords: *Fault diagnosis; Hierarchical classification; Multi-label classification; Tennessee–Eastman process; Multiple simultaneous faults.*

INTRODUCTION

Current industrial processes are turning out to be more and more complex. These processes are controlled in closed loops, include enormous number of constituents and utilize a large amount of variables. Fault diagnosis plays a key role to attain a high level of safety, quality and reliability in the control systems for these plants. Process fault diagnosis being the goal of the present work is linked to the process monitoring as a more comprehensive task. The aim of process monitoring is to safeguard the realization of the scheduled procedures by identifying abnormal process behaviors.

Owing to the wide ranges of the process fault diagnosis problems and the complications in their operational clarifications, numerous methods extending from analytical techniques to artificial intelligence and statistical methods have been established over the years [1]. However, complications encountered in building complete real-time models for real industrial plants which include many components, pushed the first attempts of the model-based fault diagnosis methods towards those that are commonly data-driven [2]. In these methods, without Necessitating the construction of any form of model for the process, generally the data acquired by the measurement tools are employed.

For data-based fault diagnosis the most dominant technique, which will also be employed in this work, is the supervised classification. In the supervised categorization, a set of examples with related labels or class types exists against the unsupervised classification where the raw data contain no labels. Different classification methods are also examined in this field including: the K-nearest neighbour [3], Fisher discriminant analysis [4], Bayesian networks [5], artificial neural networks [6-8] and Support Vector Machines (SVMs) in more current works [9-12].

The fault diagnosis based on classification methods have been effectively applied in various types of dynamical systems; however, there are still a number of practical limits in designing such systems for the large-scale complex chemical plants. One important concern is the consistent and practical management of multiple simultaneous faults. The conventional multi-class formulation often considered in the literature is based on the mono-label classification where an input data is restricted to be allocated to only one class. Mono-label classification resolves the simultaneous fault diagnosis problem by allowing for different combinations of single faults as new autonomous classes. Under this tactic the basic classifier should be trained for single fault classes and also for the newly introduced synthetic classes resulting from different groupings of all single fault classes [13]. However, it can clearly be concluded that this approach would be entirely prohibitive for large processes not only because of computational cost but also for the requirement of huge training information for different single and multiple faults. The multi-label classification considered in this work can be used as a substitute for the conventional mono-label formulation. This approach permits allocating each incoming data pattern to more than a single class [14].

Another challenge of the classification-based fault diagnosis procedures is the occurrence of significant number of faults. It is not unexpected that in a real complex chemical plant, due to the enormous number of installed components, hundreds or thousands types of faults with different known and unknown causes are possible. It should be noted that the general fault diagnosis problem is typically viewed as a flat classification task in the literature, where classification decisions are made within all of the existing class labels in one step. However, another procedure in the fault diagnosis problem could be the hierarchical classification where the classification task is accomplished in several steps through some levels of hierarchy. In this way, in each layer of hierarchy, some combinations of single faults can form new classes called meta-faults, and the classification process will be accomplished in subsequent levels between meta-faults until one or more single faults are predicted in the final levels. Through the hierarchical classification, the initial complicated large-scale categorization problem will be transformed into several regular problems in different levels of hierarchy. Each of these sub-problems can be resolved much more efficiently and more perfectly as well. Hence, the initial complexity will be reduced and the overall classification performance may improve significantly.

This work handles the hierarchical classification and also the multi-label fault diagnosis in a chemical plant. The main impact of this work is the proposition of an original systematic procedure to design a hierarchical structure of classifiers using process data measurements to accomplish fault diagnosis in the chemical plant. The foremost aim of the proposed methodology is to enhance the overall classification accomplishment of the learning system when compared to those approaches which are based on the flat classification configurations. This objective may be considered in cases where the number of the fault classes is large. Moreover, a more reliable and robust structure of classifiers may be attainable.

SVM Classifier

The SVM classification technique is a relatively new class of machine learning algorithm which is based on the structural risk minimization principle from the statistical learning theory [15]. SVM in its simple form is a binary classifier designed to separate a set of positive examples from the negative examples employing a maximum margin hyper-plane. The margin of a classifier can be quantified as the width up to which the boundary can be stretched on both sides before it hits one of the data samples. The samples onto which this margin hits are known as the support vectors. The SVM classifiers, due to the some of their outstanding characteristics such as simple design and implementation, evident generalization, capability in capturing nonlinearity and better dealing with low sample high input feature problems have become an increasingly prevalent method for machine learning activities including classification, regression and outlier detection in numerous applications. Detailed explanation of SVM is presented elsewhere [15].

Multi-Label Classification

Fault diagnosis issue is seldom addressed as a multi-label classification task and the traditional mono-label approaches are extensively employed in the literature instead. Nevertheless, the classification of the simultaneous faults requires simultaneous class assignment which naturally makes the fault diagnosis problem a multi-label classification issue. Following this parallelism, different developed techniques in the machine learning domain for solving the multi-label classification problems could also be applied to fault diagnosis tasks as well. The first attempt to this approach is represented in [16] where a problem transformation method using SVM as the learning algorithm is proposed. This approach transforms the original k-fault category problem into the k binary mono-label classification task through a proposed binarization technique based on the One Versus All (OVA) decomposition technique. This method is evaluated on the TEP benchmark to diagnose single and multiple fault cases and successful results have been obtained when diagnosing up to four simultaneous faults as detailed in [16]. This Multi-Label (ML) strategy, denoted to as the ML-OVA method, seems can effectively handle the multiple fault diagnosis issue in chemical processes is also adopted in this work.

HIERARCHICAL MULTI - LABEL CLASSIFICATION

Considering the ML-OVA method presented in the previous section, one natural consequence due to the employed decomposition strategy would be the generation of the imbalanced training set. This approach considers the flat classification method in which binary classifiers must separate the training samples of each single class (positive label) from all other remaining classes (negative label) in one step. Hence, the negative samples greatly outnumber the positive ones. The imbalanced data issue may significantly reduce the performance of any typical learning algorithms. As the number of classes increases, the ML-OVA method would create more imbalanced training distributions, consequently desirable results may not be obtained when directly employing this technique for large class classification problems. In order to deal whit this issue, a hierarchical version of the ML-OVA method called the HML-OVA technique is anticipated in this work. In the HML-OVA technique the original large class problem will be transformed into several small ones. As a result, a more balanced example distribution than the flat classification procedure of the ML-OVA method (where all the classes are arranged in the same level) is achieved.

Assume a typical fault diagnosis problem with five categories of faults specifically: F1, F2, F3, F4 and F5. Figure 1 illustrates the flat and hierarchical classification structures for the classifiers. The hierarchical structure in Figure 1(b) may be obtained based on a proposed procedure using the confusion matrix which will be described later. As seen in Figure 1, in the flat classification, one classifier is trained to distinguish all the considered faults in one single step. This multi-label classifier may be constructed based on the ML-OVA method and in this case it is basically constructed using a group of binary SVM classifiers (indicated as B-SVM in Figure 1). In the hierarchical classification however, more than one multi-label classifier is employed in different parts of the taxonomy. The classifiers are basically arranged in the first and second layers (denoted as Lf and Ls, respectively). Each multi-label categorizer operates as a flat categorizer in its corresponding level and may be constructed based on the ML-OVA method. The classifiers in the First Level of Hierarchy (FLH) categorize between some single faults and also some combinations of single faults called meta-faults. The Second Level of Hierarchy (SLH) also breaks each meta-fault to its comprising single faults. Generally, meta-faults can contain overlapped regions in which they have common single faults (such as fault 3 in the meta-faults 1 and 2). The training procedure for the hierarchical scheme is accomplished in this way. The FLH classifiers will use all sets of the training data and will be constructed based on the ML-OVA method i.e. corresponding to each meta-fault (single fault); one binary classifier is trained to separate the training data of that meta-fault (single fault) from all other faults. It is worth to remember that in the ML-OVA method for the normal state (named as 0 in Figure 1) a separate binary classifier is not trained. However, normal data will be used as negative label samples for training other binary classifiers. For example, in FLH as shown in Figure 1, for the meta-fault 2 binary classifier, the positive label set contains the F3, F4 and F5 samples and the negative label set contains F1, F2 and the normal samples. The same procedure is employed in the training of the SLH classifiers corresponding to each meta-fault, except that for each meta-fault the training data exclusively belonging to its sub-faults are used. For example, for the L s,1-classifier (s stands for the second level and 1 stands for the multi-fault 1) the training set is comprised of faults 2, 3 and normal samples. Two binary classifiers are then needed to be trained corresponding to faults 2 and 3 based on the ML-OVA classification system.

In the HML-OVA method, the testing procedure for a new example is performed in a top-down manner. First, the FLH determines the classification path by assigning the input sample to some of the meta-faults or single faults. Also, it is possible that none on the comprising binary SVM classifiers in the FLH predicts a positive label. In this case the example will be classified as normal state and the prediction procedure is ended. Otherwise, the SLH classifiers corresponding to the positive label meta-faults will be activated to classify the example to one or more single faults or the normal condition. The union of the all predictions will determine the set of final labels for the example. Also, note that if in a case some predictions are faulty and some other predictions are normal, the normal predictions are discarded. For example, consider that the dashed lines in Figure 1(b) be the classification path for an input example. As indicated, all the binary SVM classifiers in the FLH predicted positive label for the example, hence the sample is assigned to fault 1 and also the L s,1- and L s,2-classifiers are activated for further predictions. Next, although no binary SVM in the L s,1-classifier predicted positive label for the example (consequently the sample will be classified to the normal class by the L s,1-classifier), the L s,2-classifier assigned the example to faults 3 and 5. Hence, the set of predicted labels will be {1, 3 & 5}. Figure 1(a), shows the corresponding prediction procedure in the flat classification system.

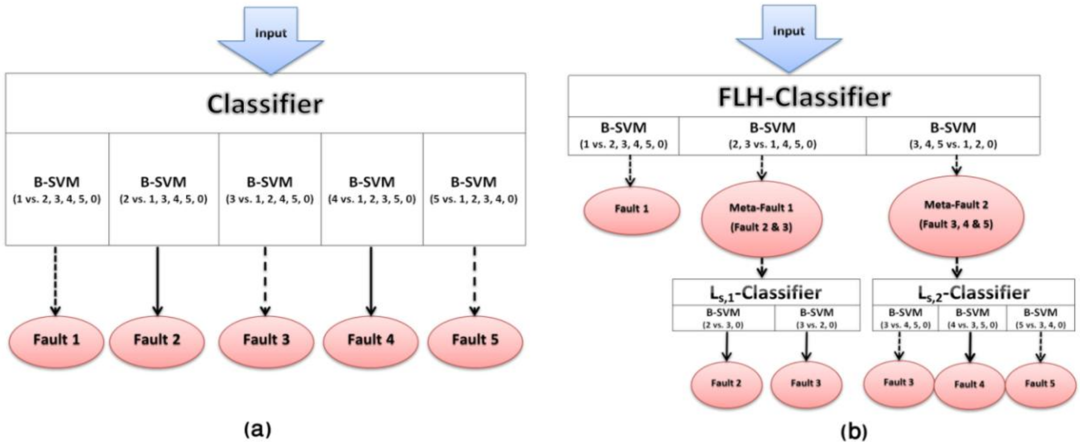


Figure1. *Classification procedures. (a) Flat structure. (b) Hierarchical structure*

In comparison to the flat classification strategy, for the proposed hierarchical classification type, more than just one multi-label classifier is employed in a way that each one deals with a smaller set of labels and a more balanced training distribution. Due to this fact, enhanced classification performance is expected. Moreover, low level classifiers are concentrated on the subsets of classes containing a reduced amount of training examples. So, local parameter tuning and feature selection for these classifiers will usually improve their accuracy and consequently these enhance the general performance of the whole learning system. Also, based on the above proposed process for the hierarchical classification technique, one may argue that the computational cost in the training step will be high. This would be true because (i) the hierarchical structure must be initially designed and (ii) more binary classifiers than the flat structure are needed to be trained (8 binary classifiers for the example hierarchical structure in Figure 1 compared to 5 for the flat construction). However, it is worth mentioning that the procedure of hierarchical mining and trainings of the binary classifiers can be executed offline. If we can determine a reasonable taxonomy and can thus improve the classification efficiency, the additional offline data preparations are defiantly worthy of being performed.

However, there is still one important open issue about how to design the hierarchical structure as the basic framework of the learning system. The design of the hierarchy generally refers to selecting the faults which should be combined to form meta-faults in the first level of the taxonomy. We propose a similarity-based approach in this work, in which the faults that have similar classification properties are combined together to form the first level meta-faults. For this purpose, the confusion matrix is employed to recognize similar faults. The confusion matrix is a standard output representation of classification problems which gives for each of the learned category, the true fault of the instances against the predicted faults. Classification performance indices such as recall and precision may be then calculated using the entry values of the matrix [14]. In this application, the confusion matrix is primarily employed to provide information about the nature of the classification problem and the relations among the faults. These invaluable indicators are then used to design the fault diagnosis system. To determine the confusion matrix, a multi-class/mono-label classifier must be trained using parts of the training data. The training information regarding to all of the fault categories needs be considered in this part. Information provided by the confusion matrix gives us an insight into the similarities of the faults. Because in setting up the classification problem similar faults will be confused amongst each other. For example, if faults A and B are two similar faults then the classifier may misclassify fault A as fault B and vise-versa. Faults A and B are thus said to be confusing faults. High misclassification values in the confusion matrix indicate similar/confusion faults which are then combined together to form the meta-faults. This procedure will be elaborated step by step in the next section using an application to the TEP dataset.

APPLICATION: TENNESSEE–EASTMAN PROCESS

The TE process benchmark is employed in this work to evaluate the proposed fault diagnosis method because of its complexity and acceptance. The TE process is the simulation of a chemical plant created by the Eastman Chemical Company to offer a representative industrial process for assessing process control and fault diagnosis methods [17]. The TEP simulator can generate one normal state, 20 types of isolated faults, as well as any multiple combinations of these faults.

Initially, only single fault experiments are carried out. Therefore, 21 simulation runs corresponding to 20 single faults and one normal state are performed. A sampling time for the recording of the observed data must be considered. This should be selected according to the characteristics of the chemical process in order to obtain a good dynamic performance for its variables. According to the discussions presented in [18], a sampling time of 3 minutes is selected here to take into account the time constants of TEP and also the instrumentation limitations for recording sampled values. Each simulation run, which took 26 hours, initiated with no fault then faults were introduced one hour after the simulation started. The total number of observations recorded for each run was 520. However, only 500 observations were collected after the introduction of each fault.

The obtained dataset are then randomly partitioned into two subsets: the training and validation sets with 60 and 40 percent of overall samples, respectively. The training set will come in useful for building each binary classifier model when applying the learning and classification algorithm. The validation set will be used for the parameter tuning purposes and also for the calculation of the confusion matrix used for the taxonomy design.

In order to generate the test dataset for the final evaluation purposes of the trained learning systems, 20 different simulation runs corresponding to 20 fault states are employed. In the test data generation procedure, the TEP simulator is not run under normal state. This is due to the implicit assumption made in this work that the fault detection task only deals with detecting the process abnormal conditions. While, in the fault diagnosis the types of faults are identified. Hence, it has been presumed that a reliable and robust fault detection model, which can isolate the normal data from abnormal ones with high accuracy such as the one-class SVM-based method proposed in [19], is previously operational in the plant. Therefore, the developed fault diagnosis systems in this work will not be examined to determine the normal state of the plant. The fault diagnosis and fault detection models may be utilized in a parallel manner so that one input example is submitted to both models. The abnormal prediction by the fault diagnosis model is discarded if the example is classified as normal by the fault detection model.

HIERARCHICAL STRUCTURE BUILDING

The first step in the implementation of the hierarchical classification system is the design of a hierarchical structure of classifiers. For this task, we provide a data-based mechanism in which no process knowledge about fault characteristics is required. The technique relies on the acquired measured data from the process. The procedure starts with learning a multi-class classifier using the training dataset of the whole fault categories. The trained multi-class classifier is then used to calculate a confusion matrix fault by fault based on the prediction on the validation dataset. The classifier learning carried out in this step is merely used for the hierarchical structure design purpose and the learned classifier would not ultimately be employed in final fault diagnosis learning system. Hence, a mediocre/quick classification technique can be applied for this purpose. We used a linear SVM classifier based on the One Versus One (OVO) multi-class strategy. In this work for all SVM learning tasks, the LIBSVM software developed by [20] is used. LIBSVM is a library for the support vector classification, support vector regression and one-class SVM. In the OVO technique, for each fault-pair a binary classifier must be trained to separate the training data of one fault from another. Since there are 20 fault classes in our problem, 20*19/2=190 independent binary SVM classifiers for each fault-pair are trained using the training data belonging to the fault-pair. To extract the confusion matrix, the validation dataset is applied to the trained binary classifiers and a majority voting strategy is then employed to combine the binary classifier outputs to obtain the classification decision for each input sample.

The obtained confusion matrix is shown if Figure 2(a). The matrix dimensions are 20 by 20. One column in this matrix is allocated for each fault to present the prediction results. The values in each column of the matrix represent the number of samples belonging to a fault classified to the same or misclassified to different categories by the learning algorithm. For example, the entry in column 5 and row 3 represents that 12 samples of the fault 5 dataset are misclassified as fault 3. Not that the sum of the elements in any column is 200 which is the sample size considered for each fault in the validation data set. Ideally, if the accuracy of classification for each sample is equal to 100%, the matrix will only contain diagonal entries corresponding to all input samples being correctly predicted to their matching true classes. In reality, the confusion matrix is often smudged with small values being distributed all over the matrix as occurred in our problem. These values however give valuable information about the nature of the classification problem in terms of class similarity.

After normalizing each column to one and ignoring the diagonal entries, the obtained values for each column can be considered as a degree of confusion between the fault corresponding to that column and the other faults. Now, a fault corresponding to a column can be combined with those faults which are predicted with high degrees of confusion to form groups of meta-faults. Here, a threshold needs to be considered so that only faults with higher confusion degrees than the threshold level are selected. This threshold is required to avoid building needless meta-faults. Therefore, without necessitating extra computational cost only those similar faults which depend on the nature of the diagnosis problem will be grouped together. We used a heuristic value of 0.025 as the confusion degree threshold in our problem. The elements of the confusion matrix which exceed this threshold are shown in bold face in Figure 2(b). For a column corresponding to a fault, the high degree confusion faults are combined to form 20 groups of faults shown in Figure 2(c). Finally, to alleviate computational complexity and avoid repetitious learning operations, those groups which their constituent faults are completely repeated in other groups are removed. For example, the fault-3 group as seen in Figure 2(c) comprises of faults 3, 5 and 9. All of these faults are simultaneously repeated in other groups such as the fault-9 group. Hence, the fault-3 group is removed. This procedure is performed to eliminate other repetitious groups. The final obtained fault taxonomy is represented in Figure 2(d).

As seen in Figure 2(d), many single faults were not contributing in the formation of meta-faults. This is due to the fact that these faults are accurately classified and are not participating as confusing faults for other faults. Consequently, these well-behaved faults are solitary placed in the first level of hierarchy and their corresponding prediction process is performed in one step. Also, as shown in Figure 2(d), some faults such as 5 and 3 are repeated in several meta-faults. This is due to the fact that these faults are repeatedly recognized as the confusion faults by other faults. The additional learning operations for these faults however may reduce the probability of the classification system to be confused in discriminating between different faults and subsequently the classification performance will be improved.

Training of the Diagnostic Scheme

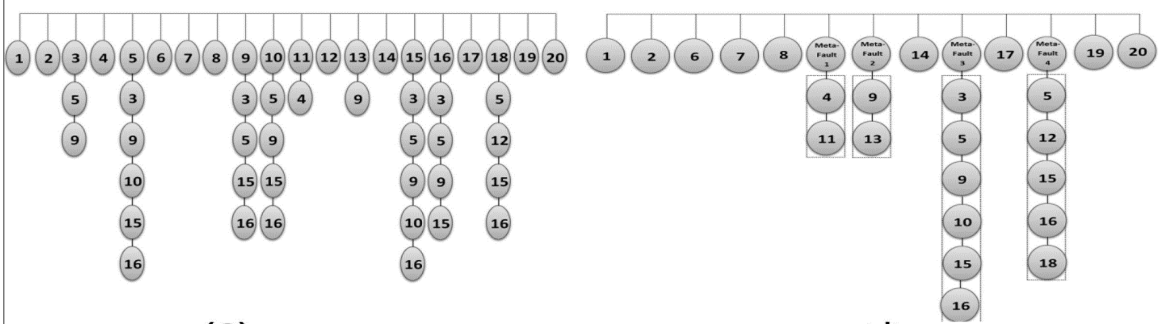
Once the taxonomy is organized the structure is trained using the HML-OVA method described in section 4. The classifiers are arranged in two layers of hierarchy. For the first level, as seen in Figure 2(d), there are altogether 13 single faults and meta-fault classifiers. Hence, according to the HML-OVA training method, 13 binary SVM classifiers must be trained for the FLH using the whole sets of the training data. For the FLH classifiers soft margin linear SVM are used as the basic binary classifier because no significant improvement was observed using non-linear kernels. This is not surprising however, because some experiments in other applications have also indicated that for large datasets and high dimensional spaces an SVM classifier generates identical testing results with or without a nonlinear mapping being employed [19]. Moreover, the value of the kernel function must be calculated for each sample in the non-linear SVM. Therefore, for large data sets the training time will become very prohibitive in comparison to the linear SVM. As a result, no mapping operation of the training samples to high dimensional spaces is performed.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	198	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	197	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
3	0	0	153	0	12	0	0	2	43	14	0	1	1	1	16	21	1	3	2	2
4	0	0	0	199	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
5	0	0	20	0	111	0	0	3	36	23	0	0	4	0	41	56	2	14	0	0
6	0	0	0	0	0	200	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	200	1	0	0	0	0	0	0	0	0	0	0	2	0
8	0	0	0	0	0	0	0	188	0	3	1	0	0	0	0	0	0	1	0	0
9	0	1	7	0	9	0	0	0	42	25	0	4	8	1	31	37	0	4	1	0
10	0	0	1	0	7	0	0	3	4	106	0	1	1	0	6	5	1	4	2	0
11	0	0	0	1	0	0	0	0	0	0	183	0	0	0	0	0	2	0	0	0
12	0	0	0	0	1	0	0	0	2	1	2	184	0	0	0	0	0	6	2	0
13	1	1	1	0	2	0	0	0	3	0	0	1	175	1	2	0	1	0	0	1
14	0	0	0	0	4	0	0	0	1	0	5	1	0	196	0	0	0	0	0	1
15	0	0	5	0	12	0	0	0	23	8	0	3	3	0	45	49	0	7	0	0
16	0	0	10	0	40	0	0	1	41	9	1	0	3	0	57	29	0	12	0	1
17	0	0	3	0	1	0	0	0	0	3	1	0	2	1	0	1	191	1	0	2
18	0	0	0	0	1	0	0	0	5	4	0	5	0	0	2	1	2	148	0	0
19	1	0	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	0	191	0
20	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	193

(a)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		0.005	0	0	0	0	0	0.005	0	0	0	0	0	0	0	0	0	0	0	0
2	0		0	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0
3	0	0		0	**0.06**	0	0	0.01	**0.215**	0.07	0	0.005	0.005	0.005	**0.08**	**0.105**	0.005	0.015	0.01	0.01
4	0	0	0		0	0	0	0	0	0	**0.035**	0	0	0	0	0	0	0	0	0
5	0	0	**0.1**	0		0	0	0.015	**0.18**	**0.115**	0	0	0.02	0	**0.205**	**0.28**	0.01	**0.07**	0	0
6	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0		0.005	0	0	0	0	0	0	0	0	0	0	0.01	0
8	0	0	0	0	0	0	0		0	0.015	0.005	0	0	0	0	0	0	0.005	0	0
9	0	0.005	**0.035**	0	**0.045**	0	0	0		**0.125**	0	0.02	**0.04**	0.005	**0.155**	**0.185**	0	0.02	0.005	0
10	0	0	0.005	0	**0.035**	0	0	0.015	0.02		0	0.005	0.005	0	**0.03**	0.025	0.005	0.02	0.01	0
11	0	0	0	0.005	0	0	0	0	0	0		0	0	0	0	0	0.01	0	0	0
12	0	0	0	0	0.005	0	0	0	0.01	0.005	0.01		0	0	0	0	0	**0.03**	0.01	0
13	0.005	0.005	0.005	0	0.01	0	0	0	0.015	0	0	0.005		0.005	0.01	0	0.005	0	0	0.005
14	0	0	0	0	0.02	0	0	0	0.005	0	0.025	0.005	0		0	0	0	0	0	0.005
15	0	0	0.025	0	**0.06**	0	0	0	**0.115**	**0.04**	0	0.015	0.015	0		**0.245**	0	**0.035**	0	0
16	0	0	0.05	0	**0.2**	0	0	0.005	**0.205**	**0.045**	0.005	0	0.015	0	**0.285**		0	**0.06**	0	0.005
17	0	0	0.015	0	0.005	0	0	0	0	0.015	0.005	0	0.01	0.005	0	0.005		0.005	0	0.01
18	0	0	0	0	0.005	0	0	0	0.025	0.02	0	0.025	0	0	0.01	0.005	0.01		0	0
19	0.005	0	0	0	0	0	0	0.005	0	0.015	0	0	0	0	0	0	0	0		0
20	0	0	0	0	0	0	0	0	0	0.005	0	0	0.005	0	0	0.005	0	0	0	

(b)



(c) (d)

Figure2. *Taxonomy extraction procedure. (a) Confusion matrix for the validation data set. (b) Degrees of confusion between each fault and others (represented in each column). (c) Groups of each fault and its confusion faults. (d) Final taxonomy.*

In the SLH, corresponding to each meta-fault one multi-label classifier must be trained to distinguish the single faults within the meta-fault. Also, as discussed before, in the training of each SLH meta-fault classifier only the training data associated to the comprising single faults of the meta-fault are utilized. Hence, in the training of an SLH classifier a lower number of labels of faults and a lower size of the training data than for an FLH classifier are required. For that reason, in the training of the SLH classifiers, it has been observed that utilization of nonlinear kernels results in a significant performance boosting. Therefore, for the binary SVM classifiers in the SLH, non-linear kernels are used to obtain a better performance for the whole learning system. For this purpose, we have chosen the RBF kernel function for all the binary SVM classifiers. This kernel function contains only one parameter, γ, to be selected and can approximate most types of kernel functions including the lineal kernel if the SVM parameters C and γ are properly chosen.

Here, one important issue arises about the appropriate parameter selections in setting up the SVM models in the hierarchical system. Proper parameter selection has a great influence on the overall classification performance. In the proposed hierarchical scheme, several multi-label classifiers operate in different classification space and hence different training parameters must be selected for each of them. This requirement however makes the training parameter selection a challenging task due to the plurality of the parameters. For example, in our problem, 4 different γ and 5 different C parameters must be selected. Here, we propose a procedure in the parameter selection phase which may lead to optimal performance of the classification system. The procedure starts from the parameter selections for the SLH classifiers. Considering the planed hierarchical structure, the performance of a classifier in the SLH can be measured independent from the other SLH classifiers and without getting involved into the training of the FLH classifiers. The parameters of an SLH meta-fault classifier are chosen in a way to obtain the best performance indices for its constituent single faults. We used a grid search on the parameters C and γ where each parameter grows exponentially (for example, C=2^(6), 2^(7), …, 2^(12) and γ=2^(-8), 2^(-7), …, 2^(0)). Various pairs of C and γ are chosen, the multi-label classifier is then trained using the selected pair of parameters for one of the SLH classifiers. Prediction on validation dataset is employed to calculate the F1 index fault by fault. See [14] for detailed information regarding to the performance evaluation indices including F1 index. Finally, the pair of parameters which results in the best value of F1 index is selected. This procedure is then repeated for each of the other SLH classifiers to find their corresponding optimal pairs of parameters.

For the FLH classifiers, as mentioned before, a linear kernel SVM classifier is considered and so the penalty parameter C is the only parameter which must be tuned. To tune the parameter C, all the SLH classifiers are primarily trained using the best parameters found before. Then, the FLH classifiers are trained using consecutive exponentially (powers of 2) increasing values of C. For each trained classifier, the validation dataset are applied to the whole classification system which includes the FLH and SLH classifiers to determine the F1 measures for all of the fault states. The same training parameter is utilized for all of the FLH classifies. In our work, the value of C which leads to the best macro-average F1 measure is equal to 2^5.

DIAGNOSIS RESULTS

The trained structure in the previous section can now be evaluated using the test data. The prediction procedure for one new test sample is detailed in section 4. Based on that procedure, the whole test data set is fed to the system and the F1 measures are computed fault by fault. A multi-label classifier based on a flat configuration is then trained. As pointed out before, in the flat classification strategy, there are no super-classes or sub-classes and all the constructed classifiers are organized in one level. Therefore, using the ML-OVA method, 20 binary SVM classifiers corresponding to 20 faults are needed to be trained.

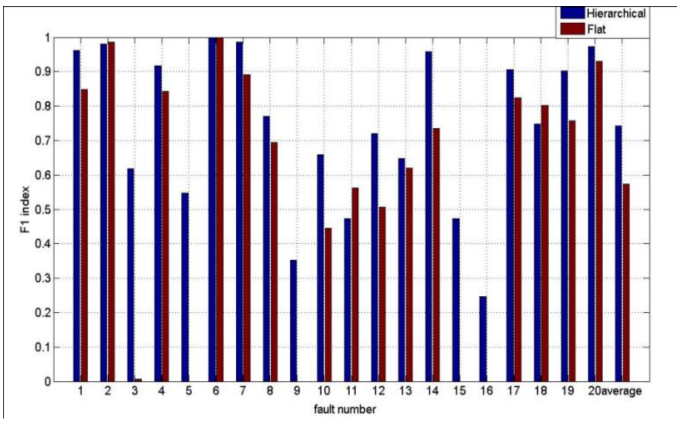


Figure3. *F1 indices for the hierarchical and flat classification systems.*

The final F1 indices obtained by the hierarchical and flat configurations are depicted in Figure 3.

As seen in this figure, in most cases, better diagnosis results are obtained by the hierarchical classification configuration than the flat one. However, via the hierarchical classification system, the main improvements are achieved for those faults which can hardly be diagnosed by the flat classification type. These faults include faults 3, 9, 15 the diagnosis of which had hardly been addressed in previous works [19]. The flat classifier was unable to diagnose faults 3, 9 and 15. However, the hierarchical system offers the F1 indices of 0.618, 0.353 and 0.472, respectively for these faults. Also, the average level of F1-measure improved to 0.742 in comparison to 0.572 obtained by the flat structure.

Yet, the diagnosis results of the hierarchical method for some faults are not satisfying enough. For example, fault 16 was not diagnosed properly. The low efficacy of the diagnosis performance for this fault and some other ones can be originated from either the FLH or the SLH classifiers. Since the main focus of this work was the demonstration of practicability of the proposed hierarchical classification method, in the classification phase of the FLH classifiers some simplifications are considered. For example, a linear kernel with a single parameter C is utilized for all the binary SVM classifies. Furthermore, to train the classifiers raw data were employed without applying any pre-filtering using the available statistical techniques. These assumptions however need to be reconsidered in a real situation in order to design a more elaborate classifier and obtain better diagnosis results.

In addition, previous studies indicated that a major improvement of the diagnosis performance can be achieved through the selection of the key variables for the classification [4]. This issue is not considered in this paper. Key variables form a subset of the whole input variables of a fault diagnosis problem. If these variables are properly selected, they improve the classification performance. Only a few variables are significant in discrimination of different faults. Hence removing other non-informative variables will have a positive effect on the sensitivity and accuracy of the classifier. This idea seems to be more practicable for the SLH classifiers which are concentrated on the subsets consisting of identical faults. These investigations however, were not within the objectives of this work and may be considered for future studies.

Conclusions

A systematic method of designing fault diagnosis systems based on the classification tools are considered in this paper. This method comprises data processing, algorithm learning and performance assessment. During this procedure, a hierarchical classification-based methodology is proposed as a consistent learning algorithm to provide improvements in the fault diagnosis performance of the large-scale industrial plants. Based on the proposed approach, rather than learning one complex global classification algorithm for discrimination between the entire original faults classes, a set of classifiers organized in a hierarchical structure are employed in a manner that each one handles a lower number of class labels. We designed a data-based procedure to construct the hierarchical structure leading to two levels of classification tasks. In the first level, the classification problem handles discrimination between some single faults and some meta-faults. These meta-faults are basically composed of combinations of some single faults which have identical classification features and are usually being confused by conventional classification methods. In the second level, the classification task deals with the decomposing of each meta-fault into its constituent single faults. The hierarchical learning task is scheduled in a manner that the trained classification system is more focused on those faults which may not be easily diagnosed.

The entire set of the TEP faults are employed in this work in order to assess the proposed procedure. For most of the fault cases, a greater diagnosis performance is realized in comparison with a case where a single layer multi-label classifier is used. However, the advantage of the procedure is more prominent for the challenging faults which cannot be appropriately diagnosed. Besides, important information about faults relations in classification perspective is offered. For example challenging fault cases are identified and separated from others. This enables this approach to use accurate separate classifiers for these cases in the second level of hierarchical classification system without concurrently handling the diagnosis of other faults using the same classifier. However, one main disadvantage of the proposed method may possibly be its computational load in learning stages which handles a hierarchical structure design, extra learning tasks and classifiers parameters adjustments. Though, it should be stated that these procedures are related to the off-line tasks and once the hierarchical classification system is trained, the prediction computation for a new test data will generally be executed in two steps of the hierarchy. Hence the testing time for the hierarchical system is similar to one single flat classifier. Considering the advantages of the proposed algorithm, it can efficiently be employed in the on-line fault diagnosis applications.

REFERENCES

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods", Computers & chemical engineering., vol. 27, (2003), pp. 293-311.

[2] M. Askarian, R. Benítez, M. Graells and R. Zarghami, "Data-based fault detection in chemical processes: Managing records with operator intervention and uncertain labels", Expert Systems With Applications., vol. 63, (2016), pp. 35-48.

[3] A. Andre, E. Beltrame and J. Wainer, "A Combination of support vector machine and k-nearest neighbors for machine fault detection", Applied Artificial Intelligence., vol. 27, (2013), pp. 36-49.

[4] L. H. Chiang, M. E. Kotanchek and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines", Computers & chemical engineering., vol. 28, (2004), pp. 1389-1401.

[5] M. Mansour, M. Wahab and W. Soliman, "Bayesian networks for fault Diagnosis of a large power station and its transmission lines", Electric Power Components and Systems., vol. 40, no. 8, (2012), pp. 845-863.

[6] M. Namdari, H. Jazayeri-Rad and N. Nabhani, "Comparing the Performance of two Neural Network Methods in a Process Fault Diagnosis System", Journal of Basic and Applied Scientific Research., vol. 3 (2013), pp. 942-947.

[7] D. Lu and J. Liu, "Application of Improved Neural Network in the Automotive Engine Fault Diagnosis", International Journal of Control and Automation., vol. 7, no. 4, (2014), pp. 355-362.

[8] C. Wu, T. Chen, R. Jiang, L. Ning and Z. Jiang, "ANN Based Multi-classification Using Various Signal Processing Techniques for Bearing Fault Diagnosis", International Journal of Control and Automation., vol. 8, no. 7, (2015), pp. 113-124.

[9] B. A. Kumar, G. Saranya, R. Selvakumar, R. S. Shree, M. Saranya and E. P. Sumesh, "Fault detection in Induction motor using WPT and Multiple SVM", International Journal of Control and Automation., vol. 3, no. 2, (2010), pp. 9-20.

[10] M. Namdari, H. Jazayeri-Rad and S. J. Hashemi, "Process fault diagnosis using support vector machines with a genetic algorithm based parameter tuning", Journal of Automation and Control., vol. 2, no. 1, (2013), pp. 1-7.

[11] X. Ye, "Fault Prediction of Electronic Equipment Based on Combination Prediction Model", International Journal of Control and Automation., vol. 8, no. 5, (2015), pp. 189-198.

[12] J. Zheng, H. Pan and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines", Mechanical Systems and Signal Processing., vol. 85, (2017), pp. 746-759.

[13] R. Eslamloueyan, "Designing a hierarchical neural network based on fuzzy clustering for fault diagnosis of the Tennessee–Eastman process", Applied Soft Computing., vol. 11, (2011), pp. 1407-1415.

[14] I. Yélamos, G. Escudero, M. Graells and L. Puigjaner, "Performance assessment of a novel fault diagnosis system based on support vector machines", Computers & Chemical Engineering., vol. 33, (2009), pp. 244-255.

[15] V. Vapnik, "The nature of statistical learning theory", Springer, New York, (2000).

[16] I. Yélamos, M. Graells, L. Puigjaner and G. Escudero, "Simultaneous fault diagnosis in chemical plants using a multilabel approach", AIChE Journal., vol. 53, (2007), pp. 2871-2884.

[17] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem", Computers & Chemical Engineering., vol. 17, (1993), pp. 245-255.

[18] L. Chiang, R. Braatz and E.L. Russell, "Fault detection and diagnosis in industrial systems", New York: Springer, (2001).

[19] S. Mahadevan and S. L. Shah, "Fault detection and diagnosis in process data using one-class support vector machines", Journal of Process Control., vol. 19, (2009), pp. 1627-1639.

[20] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines", ACM Transactions on Intelligent Systems and Technology (TIST)., vol. 2, (2011), pp. 1-27.

Citation: *N. Mehdi and J. Hooshang, "A Hierarchical Multi-Label Classification Approach for Fault Diagnosis in Industrial Units", International Journal of Emerging Engineering Research and Technology, vol. 5, no. 5, pp. 30-39, 2017.*