

Fault Description Based Attribute Transfer for Zero-Sample Industrial Fault Diagnosis

Liangjun Feng  and Chunhui Zhao , Senior Member, IEEE

Abstract—In this article, a challenging fault diagnosis task is studied, in which no samples of the target faults are available for the model training. This scenario has hardly been studied in industrial research. But it is a common problem that massive fault samples are not available for the target faults, which limits the successes of conventional data-driven approaches in practical application. Here, we introduce the idea of zero-shot learning into the industry field, and tackle the zero-sample fault diagnosis task by proposing the fault description based attribute transfer method. Specifically, the method learns to determine the fault categories using the human-defined fault descriptions instead of the collected fault samples. The defined description consists of arbitrary attributes of the faults, including the fault positions, the consequences of the fault, and even the cause of the fault, etc. For the attribute knowledge of target faults, they can be prelearned and transferred from some readily available faults occurred in the same process. Afterwards, the target faults can be diagnosed based on the defined fault descriptions without the need for any additional data based training. Besides, the supervised principle component analysis is adopted in our method to extract the attribute related features to offer an effective attribute learning. We analyze and interpret the feasibility of the fault description based method theoretically. Also, the zero-sample fault diagnosis experiments are designed and conducted on the benchmark Tennessee–Eastman process and the real thermal power plant process to validate the effectiveness. The results show that it is indeed possible to diagnose target faults without their samples.

Index Terms—Fault description, fault diagnosis, industry process, machine learning, transfer learning, zero-shot learning.

I. INTRODUCTION

FAULT diagnosis and classification are essential tasks for large-scale industrial maintenance [1], [2], and massive data-driven approaches have been proposed to improve the diagnosis ability for various faults [3], [4]. Traditionally,

classification-based fault diagnosis consists of three stages: data acquisition, feature extraction, and fault classification [5]. For the first data acquisition step, numerous sensors are installed to collect the fault signals during the daily operation. The feature extraction is usually done by matrix decomposition methods [6], time-frequency analysis [7], and the recent convolutional neural networks [8], [9] to realize the dimensionality reduction and denoising. In the last classification step, the extracted features are used as the input to get machine learning models [10] for fault diagnosis and classification. However, the three-step process is more like the rule in laboratory rather than the expectation in practice, since it is a common industrial scenario where no or few data are available for the target faults which appear during the daily process operation. Considering many faults can be destructive and cause huge losses, few factories would be allowed to run to failure and collect samples to train a fault diagnosis system [11]. Machine usually experiences a gradual degradation process from health to failure, which also reveals that it is time-consuming and expensive to obtain sufficient fault samples for the data-driven methods [12].

To overcome the trouble in collecting samples for certain faults, an acceptable way is to apply the knowledge learned from some easily obtained or historical faults (training faults) to those faults which are difficult or expensive to collect (target faults) [13], [14]. Hence, Long *et al.* [15] integrated the deep neural network with the maximum mean discrepancy item to present a deep transfer learning for fault diagnosis, in which none of samples of the target faults were necessary for highly accurate performance. The sparse autoencoder was also utilized by Lu *et al.* [16] for the deep domain adaptation of training and target faults. However, for deep transfer learning [17], the same kinds of faults have already been prepared well in the training domain. Although there is no demand for the fault samples in target domain, the deep transfer learning actually tackles the domain shift (between the training and target domain) problem instead of the concerned zero-sample problem. Besides, the fault tree analysis [18], [19] was also a classic method for industry scenario, which diagnosed the system failure based on some basic faults. It was successfully applied to understand the causes of system failure, and found the best way to reduce the risk, or confirmed the occurrence rate of a fault [20], [21]. However, when the fault tree analysis was applied to the quantitative analysis, the frequencies of basic faults were usually required by the expert system. It means that the records (samples) for some target faults had been prepared, which did not meet the requirement of zero-sample setting.

Manuscript received January 30, 2020; accepted April 12, 2020. Date of publication April 20, 2020; date of current version November 20, 2020. This work is supported in part by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization under Grant U1709211 and in part by Zhejiang Key Research and Development Project under Grant 2019C03100 and Grant 2019C01048. Paper no. TII-20-0509. (Corresponding author: Chunhui Zhao.)

The authors are with the State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: liangjunfeng@zju.edu.cn; chhzhao@zju.edu.cn).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TII.2020.2988208

Recently, the zero-shot learning proposed by Lampert *et al.* [22] presents promising performance for the zero-sample animal recognition task, in which no training images are available for the animals appeared at test stage. Instead of directly training the classification algorithms, Lampert utilized the nonlinear support vector machine to learn the attributes of animals, i.e., the “colors” and “shapes,” and exploited the well-trained machines to predict the attributes of test images of other animals. Based on the transfer of learned attributes, the new animals could be detected. The method was termed as direct attribute prediction (DAP). Lampert *et al.* [23] also learned the animal attributes from the labels to present the indirect attribute prediction (IAP) method. Besides, the compatibility functions and square loss were utilized in zero-shot learning to learn the mapping between the images and attributes. Akata *et al.* [24] conducted the zero-shot classification of an image proceeds by finding the label yielding the highest joint compatibility score. Romera-Paredes and Torr [25] proposed an embarrassingly simple approach for zero-shot learning, which modeled the relationships between features, attributes, and classes as a two linear layers network based on the square loss. The idea of zero-shot learning that learns an object from its attributes [26] provides a possible solution to the zero-sample fault diagnosis problem. However, when it turns to the fault diagnosis task, there are no images available to obtain the “colors” or “shapes” attributes for various faults. The visual attributes are inapplicable for the industrial sensor signals, and more effective auxiliary information is required for the zero-sample fault diagnosis task.

Intuitively, when we learn a new kind of fault in practice, the characteristics and semantic description of the fault will be noticed first rather than the samples. For examples, from the description “a device that is connected with a tube, a coupling and a valve to convey gas, liquid, or fluid with solid particles,” we can learn the object “pipe” without seeing it at all. Also, humans know there is a “pipe blockage” fault, when they are told that “the fluid flow in pipe is stopped or abnormally slow.” Hence, it is feasible to diagnose and recognize various faults based on the well-concluded descriptions instead of the samples. The human-defined fault descriptions are convincing and intuitive evidence for the identification of fault categories. Each description of industrial faults usually consists of several attributes, including the effects of the fault, the position of the fault, and the cause of the fault, etc. Obviously, there are numerous possible attributes, and the fact that the human-defined attributes which transcend class boundaries can be exploited for an efficient cross-class learning [22], [23]. For examples, both of the fault “abnormal vibration of pulverizer bearing” and the fault “temperature rise of pulverizer coil” occur at the “pulverizer.” To learn and recognize the “pulverizer” attribute, we can therefore make use of existing training data by merging samples of several other fault categories which also occur at the “pulverizer,” and apply the knowledge to the target faults. The fact that the human-defined attributes can be shared by various faults is this possibility to obtain attribute knowledge from the descriptions of different fault categories.

Besides, the high-dimension and low-density information is another essential characteristic of the industrial sensor data [27], [28]. Statistical feature extraction algorithms [29], [30]

are usually required for an effective data mining and property revealing. The most popular and classic method is the principle component analysis [31], [32], which conducts the feature extraction in an unsupervised manner. Numerous case studies and former works have validated its effectiveness. However, for the fault diagnosis task [33], which is generally defined in the supervised learning paradigm, the supervised principle component analysis [34] can be more helpful for the extraction of target related features and presenting an effective diagnosis.

In this article, the fault diagnosis scenario in which no training samples are provided for the target fault categories is studied by the fault description based attribute transfer method. Instead of the conventional diagnosis paradigm, the fault description which consists of arbitrary attributes is provided for each fault as the auxiliary information in our method. The fault description layer is embedded between the fault sample layer and the fault category layer. Based on the fine-grained and class-shared attributes in the fault description layer, a cascade diagnosis system can be built to transfer the attribute knowledge from training faults to target faults to conduct the zero-sample fault diagnosis. We also adopt the supervised principle component analysis as the feature extractor in the fault description based method to provide the attribute related features for a more efficient learning process.

The main contributions of this article are summarized as follows.

- 1) It is the first time that the zero-sample fault diagnosis task is concluded and solved, which attempts to diagnose target faults without their samples.
- 2) The fault description based method which sets the fault descriptions as the auxiliary knowledge source is proposed to achieve the attribute transfer from training faults to target faults for the zero-sample diagnosis.

The remainder of this article is organized as follows. In Section II, we formalize the concerned zero-sample diagnosis problem. In Section III, the fault description based attribute transfer method and feasibility analysis are presented. In Section IV, the zero-sample fault diagnosis task is conducted on the benchmark Tennessee–Eastman process and the real thermal power plant (TPP) process. Finally, Section V concludes this article.

II. PROBLEM FORMULATION

A. Configuration of Descriptions in Vector Space

As mentioned above, we conclude the fault description for each kind of fault to offer the fine-grained class-level information. The description consists of arbitrary attributes, including the effects of the fault, the specific fault position, and the cause of the fault, etc. Each attribute is a dimension in vector space, and the description of a fault is denoted as $\mathbf{a}' \in \mathbb{R}^{C'}$, where C' is the number of attributes. For L kinds of faults, the description matrix can be denoted as $\mathbf{A}' \in \mathbb{R}^{L \times C'}$. The one-hot coding technique is adopted here to make \mathbf{A}' a sparse matrix $\mathbf{A} \in \mathbb{R}^{L \times C} = \text{one-hot}(\mathbf{A}')$, where C is the dimension of one-hot coding. All the elements in \mathbf{A} are either 1 or 0, which denotes the attribute exists or not in the description of a certain fault category.

B. Formulation of Zero-Sample Fault Diagnosis

We aim to diagnose and recognize p kinds of faults. The set of target faults is denoted as $T = \{t_1, \dots, t_p\}$. However, there are no training samples available for T . Some accessible fault categories may be utilized here, which are denoted as $S = \{s_1, \dots, s_q\}$, where q is the number of accessible faults. But T and S are disjoint from each other, i.e., $T \cap S = \emptyset$. The samples of S are denoted as $\mathfrak{S} = \{\mathbf{X}_S \in \mathbb{R}^{N_S \times D}, \mathbf{Y}_S \in \mathbb{R}^{N_S}\}$, where N_S is the number of samples, and D is the feature dimension. The zero-sample fault diagnosis learns a mapping f from S to T , which can be formulated as

$$\text{Min } C\text{Loss}(\mathbf{Y}_T, \hat{\mathbf{Y}}_T), \text{ and } \hat{\mathbf{Y}}_T = f(\mathbf{X}_S, \mathbf{Y}_S | \mathbf{X}_T) \quad (1)$$

where \mathbf{X}_T and \mathbf{Y}_T denotes the samples and labels of the target faults at test stage, respectively, and $C\text{Loss}$ denotes arbitrary classification loss.

After introducing the fault descriptions for both of T and S , the objective function in (1) can be rewritten as

$$\text{Min } C\text{Loss}(\mathbf{Y}_T, \hat{\mathbf{Y}}_T), \text{ and } \hat{\mathbf{Y}}_T = f(\mathbf{X}_S, \mathbf{Y}_S, \mathbf{A} | \mathbf{X}_T) \quad (2)$$

where $\mathbf{A} = [\mathbf{A}_S, \mathbf{A}_T] \in \mathbb{R}^{L \times C}$ is the attribute description matrix, and $L = p + q$. It is worth mentioning that both of the attribute description matrix \mathbf{A}_S of S and \mathbf{A}_T of T are available for the model training, since the attribute descriptions are class-level rather than sample-level and common knowledge rather than specialized expert knowledge which can readily be obtained. Compared with the conventional data-driven fault diagnosis problem, the proposed task is a more meaningful and challenging task, which aims to overcome the trouble in collecting numerous samples for the target fault categories.

III. METHODOLOGY

As defined in Section II, our interest is to diagnose and classify the faults which have no samples during the training stage. Here, we begin with the proposed fault description based method, and then provide the feasibility analysis.

A. Fault Description Based Attribute Transfer

For the ordinary machine learning models, such as support vector machine or decision tree, they usually learn one parameter vector (or other representation) α_i for each kind of fault category based on the fault samples or extracted features. Clearly, it is infeasible for them to directly realize the learning from S to T for the zero-sample fault diagnosis task. Because the target classes $T = \{t_1, \dots, t_p\}$ were not presented during the training stage, and there are no parameter vectors obtained for T . This dilemma is described in Fig. 1.

In other words, to identify the disjoint target faults, the learning system needs to extract target faults related information from training faults. Since no training data are available for the target faults, this coupling cannot be extracted from samples directly. Hence, the fault descriptions \mathbf{A} are utilized to offer additional auxiliary information and achieve the attribute transfer from training faults to target faults.

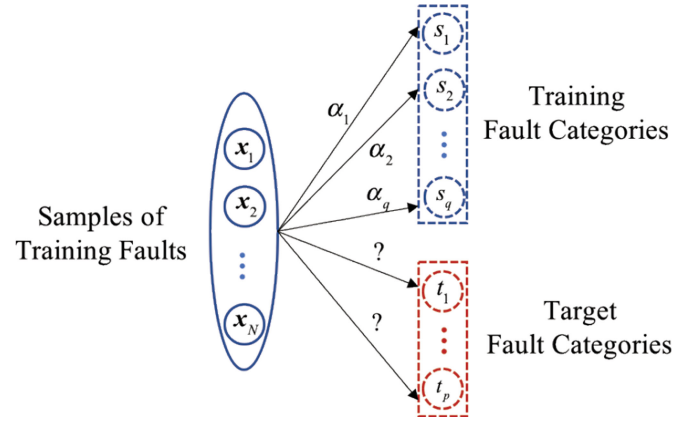


Fig. 1. Dilemma of the ordinary machine learning models for the zero-sample fault diagnosis task, in which none of the parameter vectors α can be obtained for target faults during the training stage.

The basic idea of the fault description based method for zero-sample fault diagnosis task is presented in Fig. 2. The first step of the cascade model is feature extraction. We merge the training label \mathbf{Y}_S with the training fault descriptions \mathbf{A}_S to obtain the training attribute label $\mathbf{Z}_S = [\mathbf{z}_1^S, \dots, \mathbf{z}_C^S]^T \in \mathbb{R}^{N \times C}$, which denotes each fault sample is described by a C -dimension attribute vector. The supervised principle component analysis is applied on the pair $\{\mathbf{X}, \mathbf{z}_i^S\} (i = 1, \dots, C)$ to extract the attribute related features. The features of training sample \mathbf{x} is denoted as \mathbf{b} , and the transform is denoted as φ . The second step is the attribute learning and transfer stage based on the extracted features. Instead of directly learning a mapping between features and labels, it trains an attribute learner $\alpha_i (i = 1, \dots, C)$ for each kind of attribute a_i under the supervised manner during the training stage. **At test time, these attribute learners allow the predictions of attribute values (either 1 or 0) for each test sample of target fault.** Note that as long as the attributes are described in the same dimensions for both of the training and target faults, none of additional training is needed for more attribute learners for target faults. Based on the predicted attribute vectors of test samples, the third step can be conducted, in which the final categories of the test samples are inferred. Because the fault descriptions of target faults $\mathbf{A}_T \in \mathbb{R}^{p \times C}$ are known, the inference rules β from fault descriptions to fault categories are ready for test samples, such as the nearest neighbor searching, etc.

Here, the cascade model in Fig. 2 is also described in the probabilistic manner for an accurate presentation. The feature \mathbf{b} is extracted from the sample \mathbf{x} by the transform mapping φ . The second step which makes inference from features to fault descriptions can be presented as $p(\mathbf{a} | \mathbf{b}) = \prod_{i=1}^C p(a_i | \mathbf{b})$, where $\mathbf{a} \in \mathbb{R}^C$ is the description of a fault in \mathbf{A} . For the third step, it obtains the final category t from \mathbf{a} . Here, the attribute vector \mathbf{a} for the t th class is denoted as $\mathbf{a}^t = [a_1^t, \dots, a_C^t]$. Based on the Bayes' rule, we denote the inference as

$$p(t | \mathbf{a}) = \sum_{t=1}^p \frac{p(\mathbf{a}^t | t) p(t)}{p(\mathbf{a}^t)} [\mathbf{a} = \mathbf{a}^t] \quad (3)$$

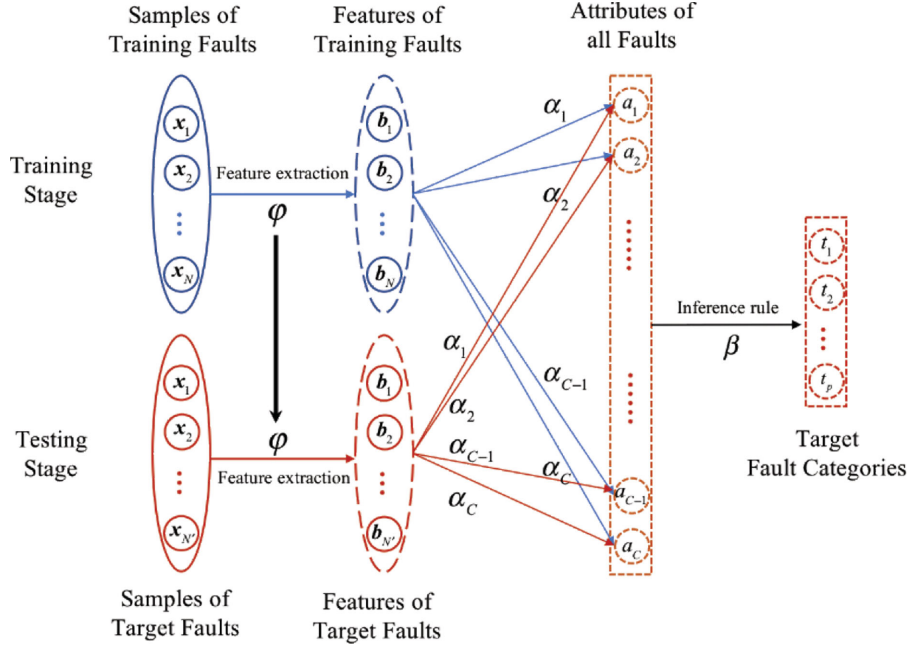


Fig. 2. Basic idea of the fault description based method. N is the number of samples of training faults, N' is the number of samples of target faults at test stage, φ is the transform mapping of supervised principle component analysis, \mathbf{b} is the feature of \mathbf{x} , $\alpha_i (i = 1, \dots, C)$ is the attribute learners (regressors or classifiers) for each attribute, $\mathbf{a}_i (i = 1, \dots, C)$ is the attributes in $\mathbf{A} \in \mathbb{R}^{L \times C}$, and β is the inference rule from the attribute descriptions to target fault categories.

where $[a = a^t] = 1$ if $\mathbf{a} = \mathbf{a}^t$ and otherwise it is 0. Since the attribute matrix of target faults \mathbf{A}_T is known, we have $p(a^t|t) = 1$. In the absence of more specific knowledge, the class prior $p(t)$ is assumed identical. For the factor $p(a^t) = \prod_{i=1}^C p(a_i^t)$, the empirical means $p(a_i^t) = \frac{1}{q} \sum_{j=1}^q a_i^{s_j}$ over the training fault categories can be assumed as attribute priors. Combining the three steps, the posterior of a test fault category can be obtained from a test sample as

$$p(t|\mathbf{x}) = p(t|\mathbf{a})p(\mathbf{a}|\mathbf{b}) = \frac{p(t)}{p(\mathbf{a}^t)} \prod_{i=1}^C p(a_i^t|\mathbf{b}) \quad (4)$$

where $\mathbf{b} = \varphi(\mathbf{x})$. To assign the best output class from all target fault categories t_1, \dots, t_p to a test sample \mathbf{x} , the maximum posterior estimation can be used

$$f(\mathbf{x}) = \arg \max_{j=1, \dots, p} \prod_{i=1}^C \frac{p(a_i^{t_j}|\varphi(\mathbf{x}))}{p(a_i^{t_j})} = \arg \max_{j=1, \dots, p} \frac{p(\mathbf{a}^{t_j}|\mathbf{b})}{p(\mathbf{a}^{t_j})} \quad (5)$$

where \mathbf{a}^{t_j} denotes the fault description vector of the j th target fault category and $a_i^{t_j}$ is the i th element of \mathbf{a}^{t_j} . Equation (5) reveals that our description based method actually conducts the fault diagnosis between the features \mathbf{b} and the fault description \mathbf{a} . And hence, the fault features and descriptions are essential for the zero-sample task.

B. Feasibility Analysis

Let $\mathbf{X}_S = [\mathbf{x}_1^S, \dots, \mathbf{x}_N^S]$ and $\mathbf{X}_T = [\mathbf{x}_1^T, \dots, \mathbf{x}_{N'}^T]$ denote the samples of training faults and target faults. $\mathbf{B}_S = [\mathbf{b}_1^S, \dots, \mathbf{b}_N^S]$ and $\mathbf{B}_T = [\mathbf{b}_1^T, \dots, \mathbf{b}_{N'}^T]$ are the features, $\mathbf{Y}_S \in \mathbb{R}^N$ and $\mathbf{Y}_T \in \mathbb{R}^{N'}$ are their labels. $\mathbf{A} = [\mathbf{A}_S, \mathbf{A}_T] \in \mathbb{R}^{L \times C}$ denotes

that each fault category is described by a C -dimension attribute vector, and hence the attribute label $\mathbf{Z}_S = [\mathbf{z}_1^S, \dots, \mathbf{z}_N^S] \in \mathbb{R}^{N \times C}$ and $\mathbf{Z}_T = [\mathbf{z}_1^T, \dots, \mathbf{z}_{N'}^T] \in \mathbb{R}^{N' \times C}$ can be obtained by merging of \mathbf{A} with \mathbf{Y}_S and \mathbf{Y}_T . Without loss of generality, the attribute learners $g_S = \{\alpha_1^S, \dots, \alpha_C^S\}$ and $g_T = \{\alpha_1^T, \dots, \alpha_C^T\}$ for training and target faults are assumed as the linear mapping from the features \mathbf{B}_S and \mathbf{B}_T to the attribute labels \mathbf{Z}_S and \mathbf{Z}_T .

The transfer technique of the proposed fault description based method is the second step shown in Fig. 2, in which the attribute learners of training fault categories g_S are exploited as those of the target fault categories g_T . Here, we analyze the feasibility of the sharing of attribute learners, i.e., $g = g_S = g_T$.

Definition: For $\forall \mathbf{b}_i^T$, assume $\exists \mathbf{u}_i \in \mathbb{R}^N$, $\mathbf{b}_i^T = \mathbf{u}_i \mathbf{B}_S$. Similarly, for $\forall \mathbf{z}_i^T$, $\exists \mathbf{v}_i \in \mathbb{R}^N$, $\mathbf{z}_i^T = \mathbf{v}_i \mathbf{Z}_S$. The set $U = \{\mathbf{u}_i\}$ and $V = \{\mathbf{v}_i\}$ are the *relational knowledge* of \mathbf{B}_T and \mathbf{Z}_T on \mathbf{B}_S and \mathbf{Z}_S , respectively.

Both of the features \mathbf{B} and the attribute knowledge \mathbf{Z} are the descriptions of faults. The relational knowledge set U actually encodes a kind of dependence of \mathbf{B}_T on \mathbf{B}_S in feature space, as well as V encodes that in the attribute label space. Based on the definition, we have the following lemma for the sharing of attribute learners.

Lemma: If $\forall j, \mathbf{u}_j = \mathbf{v}_j$, the linear map $g_S : \mathbf{b}^S \rightarrow \mathbf{z}^S$ learned from the samples of training fault categories $\{(\mathbf{b}_i^S, \mathbf{z}_i^S)\}$ can be used as $g_T : \mathbf{b}^T \rightarrow \mathbf{z}^T$ to make predictions for the test samples of target fault categories $\{(\mathbf{b}_i^T, \mathbf{z}_i^T)\}$.

Proof: Let $\mathbf{b}_j^T = \sum_{i=1}^N \mathbf{u}_j \mathbf{b}_i^S$, and $g_S(\mathbf{b}_j^T) = g_S(\sum_{i=1}^N \mathbf{u}_j \mathbf{b}_i^S) = \mathbf{u}_j \sum_{i=1}^N g_S(\mathbf{b}_i^S) = \mathbf{v}_j \sum_{i=1}^N \mathbf{z}_i^S = \mathbf{z}_j^T = g_T(\mathbf{b}_j^T)$.

Hence, according to the lemma, when the relational knowledge U and V are equal to each other, the sharing of attribute learners is totally feasible under the linear case. Intuitively,

TABLE I
FAULTS OF TEP USED IN THE ZERO-SAMPLE FAULT DIAGNOSIS

No.	Fault state	Disturb
1	A/C feed ratio, B composition constant	Step change
2	B composition, A/C ratio constant	Step change
3	D feed temperature	Step change
4	Reactor cooling water inlet temperature	Step change
5	Condenser cooling water temperature	Step change
6	A feed loss	Step change
7	C header pressure loss	Step change
8	A, B, C feed composition	Random variants
9	D feed temperature	Random variants
10	C feed temperature	Random variants
11	Reactor cooling water inlet temperature	Random variants
12	Condenser cooling water valve	Random variants
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking

the relational knowledge U in the feature space is potentially determined by the samples of training and target fault categories, which cannot be changed. However, the relational knowledge V in the attribute label space is decided by the defined \mathcal{A} , which is assigned to be learned from the fault descriptions and improved by trial and error.

IV. EXPERIMENTS

In this section, two case studies are provided. One is designed using the benchmark Tennessee–Eastman process. We detail the experiment settings and model implementations on the fully studied data set for the better understanding of the proposed zero-sample fault diagnosis task and method. The other is conducted on the real thermal power plant process, in which we show the application and the model-reality match of the fault-description based attribute transfer method.

A. Tennessee–Eastman Process

1) Introduction for Data set: Tennessee–Eastman process (TEP) contributed by Downs and Vogel [35], [36] is a fully studied fault diagnosis data set in industry, and hence it is helpful to fairly present the newly proposed zero-sample diagnosis task and the effectiveness of the fault description based attribute transfer method. Also, we present the fine-grained attribute descriptions of TEP used in experiments to show the auxiliary information for the zero-sample fault diagnosis task.

TEP consists of five major subsystems, including a reactor, a condenser, a vapor-liquid separator, a recycle compressor, and a product stripper. The data set offers 21 kinds of faults, and each kind of fault is described by 41 measured variables and 11 manipulated variables. 480 samples are collected for the training of each fault. Since the last 6 kinds of faults have little description in the data set, the first 15 kinds of faults are utilized here to conduct the zero-sample fault diagnosis. The 15 fault categories are introduced in Table I. The studied 15 faults of TEP are different from each other. When some of them have zero sample for the model training, it is hard for the conventional methods to conduct the fault diagnosis. Hence, the proposed zero-sample fault diagnosis is actually meaningful and practical.

TABLE II
SPECIFIC INFORMATION OF ATTRIBUTES FOR TEP

No.	Attributes
Att#1	Input A is changed
Att#2	Input C is changed
Att#3	A/C ratio is changed
Att#4	Input B is changed
Att#5	Related with pipe4
Att#6	Temperature of input D is changed
Att#7	Related with pipe2
Att#8	Disturbance is step changing
Att#9	Input is changed
Att#10	Temperature of input is changed
Att#11	Occurred at reactor
Att#12	Temperature of cooling water is changed
Att#13	Occurred at condenser
Att#14	Related with pipe 1
Att#15	Disturbance is random varying
Att#16	Model parameters are changed
Att#17	Disturbance is slow drift
Att#18	Related with cooling water
Att#19	Related with valve
Att#20	Disturbance is sticking

2) Details of Model Implementation: The fault descriptions are the basis of the proposed solution for the zero-sample fault diagnosis. The configuration of the attribute descriptions in vector space for TEP, i.e., the attribute matrix \mathcal{A} , are shown in Fig. 3, and the specific attribute names are shown in Table II. Each kind of fault is described by 20 fine-grained attributes. Comparing Table II with Table I, it should be noticed that the attribute descriptions can readily be concluded from the statements in Table I. Based on the fault descriptions, we diagnose target faults without their samples for model training.

The first step of the description-based method in Fig. 2 is the feature extraction. The supervised principle component analysis is adopted to extract the attribute related features. Specifically, each pair $\{\mathbf{X}_S, \mathbf{z}_i^S\}$ ($i = 1, \dots, C$) is fitted by the supervised principle component analysis [34]. And then the extractor is applied to the test data \mathbf{X}_T . For each attribute, 20 features are extracted from the original 52 variables. Since there are 20 different attributes, a data set with 400 features is prepared for the following steps after the feature extraction. The second step is the training of attribute learners. Three different machine learning algorithms are adopted here, including the linear support vector machine (LSVM), the nonlinear random forest (RF), and the probabilistic naïve Bayes (NB). We use the implementation of scikit-learn [37] for the three models for a fair comparison. The parameter of slack term of LSVM is set as 1, and the number of decision trees of RF is set as 50. Last, the nearest neighbor searching is used as the inference rule to decide the final fault categories, and the ordinary Euclidean distance is used.

There are 15 kinds of faults utilized in TEP, and 80%–20% split is adopted for the train/test division, which means 12 kinds of faults are set as the training faults and the 3 other faults are set as the target faults. To test the performance on the whole data set, the TEP data set is divided into 5 groups, and each group has 3 faults for test and 12 faults for training. The train/test division is shown in Table III. The number of training samples is 5760 (12*480), and the number of the test samples is 1440

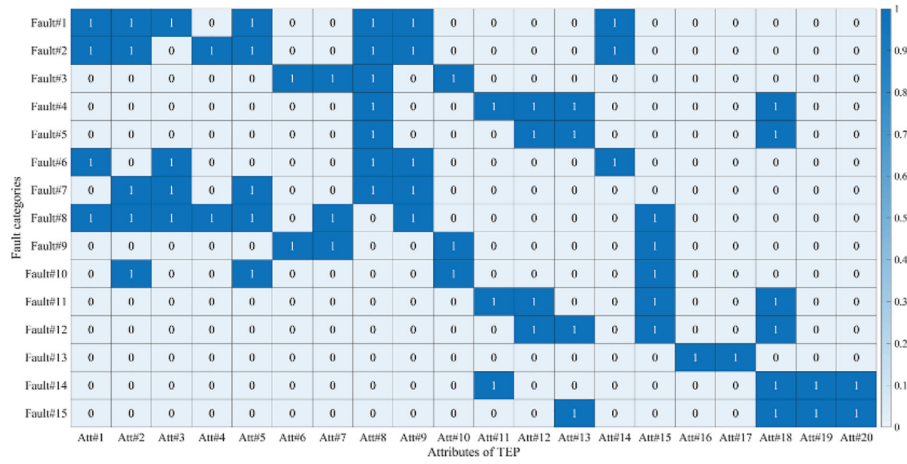


Fig. 3. Fault description matrix A for TEP data set. Each kind of fault is described by 20 fine-grained attributes. The “1” in figure denotes the fault has this attribute, and the “0” denotes not. The specific names of attributes are shown in Table II.

TABLE III
FIVE GROUPS OF TRAIN/TEST SPLIT FOR TEP

No.	Training Faults	Target Faults
A	2, 3, 4, 7-13, 15	1, 6, 14
B	1, 2, 3, 5, 6, 8, 9, 11-15	4, 7, 10
C	1-7, 9, 10, 13, 14, 15	8, 11, 12
D	1, 4, 6-15	2, 3, 5
E	1-8, 10, 11, 12, 14	9, 13, 15

TABLE IV
RESULTS OF ZERO-SAMPLE FAULT DIAGNOSIS FOR TEP(%)

Group.	LSVM	RF	NB	Heighest	Mean
A	58.68	88.40	80.27	88.40	75.78
B	58.12	51.87	62.63	62.63	57.54
C	44.51	48.33	67.56	67.56	53.46
D	53.34	61.73	72.43	72.43	62.50
E	42.56	43.12	67.43	67.43	51.03

(3*480). We report the accuracy of the five groups experiments. The accuracy is given by

$$\text{Acc} = \frac{N_C}{N} \quad (6)$$

where N_C is the number of correctly classified samples, and N is the number of samples of target faults at test stage.

3) *Results of Zero-Sample Fault Diagnosis*: The results of zero-sample fault diagnosis are shown in Table IV. The highest accuracies varying from 62.63% to 88.40% with the change of the train/test split. The confusion matrices of the best results in the five groups experiments are shown in Fig. 4. Clearly, the performance of the proposed fault description based attribute transfer method is significantly higher than the chance level of 33.33%, which proves the initial aims of the fault description based method: it is possible to diagnose and classify different kinds of faults without their training samples based on the human-defined fault descriptions. As for the accuracies of some specific fault categories, such as 26% in Group D and 21% in Group E in Fig. 4, they are caused by the difficulty of the zero-sample task, and the few-sample experiments in next section will

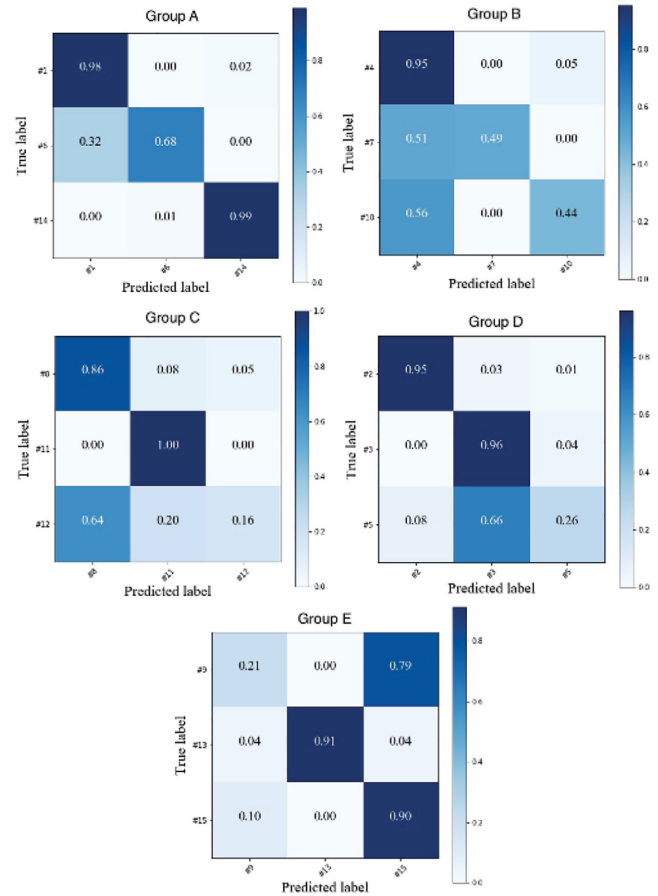


Fig. 4. Confusion matrices of the best results of the five groups for TEP.

show that our results are actually quite competitive. Besides, it is worth mentioning that although the proof for the feasibility of sharing of attribute learners are limited to the linear case, the higher performance of nonlinear classifier and probabilistic classifiers, i.e., RF and NB, shows that the method is generally applicable.

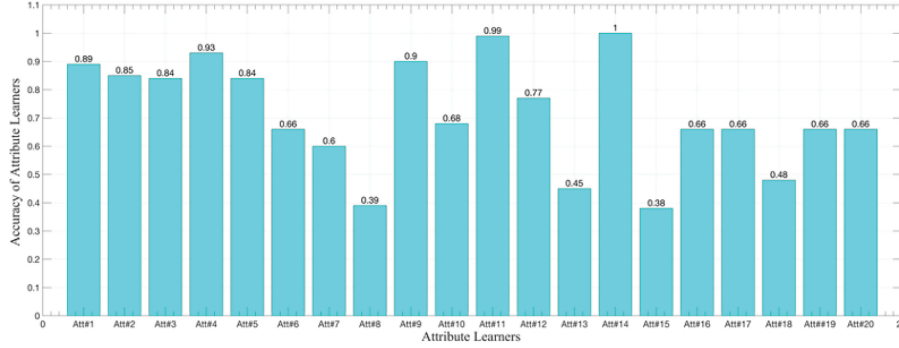


Fig. 5. Accuracies of 20 attribute learners for the group A with RF as attribute learners.

TABLE V
COMPARISON WITH ZERO-SHOT LEARNING ON TEP(%)

Group.	A	B	C	D	E	Mean
DAP	54.16	62.63	40.13	55.48	36.94	49.86
IAP	55.48	60.69	45.00	36.25	48.88	49.26
SJE	74.58	33.12	33.95	63.88	33.81	47.86
ESZSL	57.22	33.33	39.51	39.65	33.33	40.60
Ours(LSVM)	58.68	58.12	44.51	53.34	42.56	51.44
Ours(RF)	88.40	51.87	48.33	61.73	43.12	58.69
Ours(NB)	80.27	62.63	67.56	72.43	67.43	69.86

Note: Both of the deep features of convolutional networks and the image-based attributes required by SJE and ESZSL are unavailable under the fault diagnosis scenario. Their unsatisfactory performance on groups B, C, and E reveals their poor generalization ability for the zero-sample fault diagnosis task.

To the best knowledge of the authors, no previous data-driven-based methods achieve the fault diagnosis and classification without their samples for model training. As discussed in Section I, typical techniques including the deep transfer learning, and fault tree analysis, are not applicable for the zero-sample fault diagnosis setting. Hence, to provide the comparison under the same setting, the zero-shot learning methods, i.e., DAP and IAP proposed by Lampert *et al.* [22], [23], are utilized. Note that DAP and IAP are designed for the image classification task, and the “shapes” or “colors” attributes designed by Lampert cannot be utilized here, but the fault descriptions concluded in this article are used. The image features, i.e., HoG and SIFT, are also inapplicable here. We make DAP and IAP learn from the original data. The slack term of the nonlinear support vector machine (Gaussian kernel) of DAP and IAP is set as 1. Hence, the comparison is actually based on our contribution, i.e., the fault descriptions. And although both of the ideas of DAP and ours are the attribute transfer, different attribute learners and features are utilized by our method. The results are shown in Table V. When the linear LSVM is set as the attribute learners, our method presents competitive accuracies with DAP and IAP through the five groups. And when the nonlinear attribute learners, i.e., RF, and the probabilistic attribute learner, i.e., NB, are exploited, our method obtains significantly higher accuracies. Also, we attempt to compare our methods with SJE by Akata *et al.* [24] and ESZSL by Romera-Paredes and Torr [25]. However, since both of the deep features of convolutional networks and the image-based attributes required by SJE and

ESZSL are unavailable under the fault diagnosis scenario, their performance suffers from the serious degradation on groups B, C, and E, which reveals their weak generalization ability under the industry scenario and the difficulty of the zero-sample fault diagnosis task.

Another performance which should be assessed is the accuracies of the attribute learners, which is the basis of the diagnosis of zero-sample faults. In our method, the model learns the fault attributes first, and then the model knows the faults. The accuracies of 20 attribute learners for the group A with RF as attribute learners are shown in Fig. 5. Most of the attribute learners perform much better than the random guess (50% accuracy), and some of them even present above 90% accuracies for the group A, which directly illustrates the effectiveness of the proposed fault description based method.

4) *Comparison With Few-Sample Learning:* For comparison, we also conduct the few-sample learning experiments based on the group A and group C, in which 1, 10, 50, 200, 500 samples are used for the model training. The training samples are randomly selected from the other group data of TEP, and the test data are the same as those in the zero-sample fault diagnosis. The compared algorithms include LSVM, RF, NB, XGBoost (XGB) [38], AdaBoost (ADA), K-neighbors (KNN), gradient boosting machine (GBDT) [39], and light gradient boosting machine (LGBM) [40]. For all the compared models, the implementations by their authors or scikit-learn package [37] are utilized for a fair comparison. The LSVM model of scikit-learn package used in experiments adopts the “1-v-1” strategy by default for the multiclass classification problem. As for the parameters, we uniformly adopt the default settings by their authors without any adjustment, which can usually present good performance. The results of the few-sample learning experiments are shown in Tables VI and VII.

As shown in Tables VI and VII, the conventional machine learning algorithms perform definitely poor when only a few fault samples are provided, and some of them perform even worse than the random choice (33.33%) when only 1 sample is used. Although TEP is a fully studied and classic fault diagnosis data set, the compared methods still need at least 200 samples to present the competitive results as ours, which shows the difficulty of the task of zero-sample fault diagnosis. Actually, it is unfair for our methods to compare with the few-sample learning,

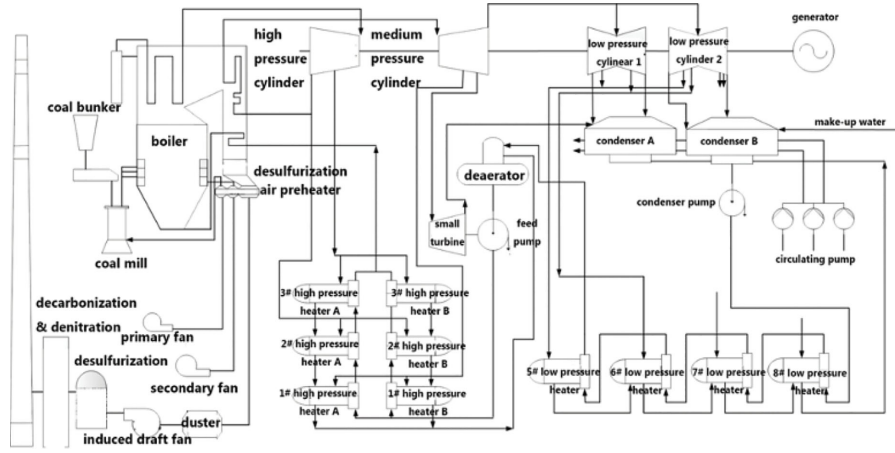


Fig. 6. Systematic configuration for 1000 MW ultrasupercritical thermal power unit.

TABLE VI
COMPARISON WITH FEW-SAMPLE LEARNING ON GROUP A

Acc(%)	Number of training samples for each fault				
	1	10	50	200	500
LSVM	23.54	17.98	39.23	49.79	87.15
RF	25.00	55.55	50.90	57.01	99.37
NB	23.54	25.41	34.72	84.86	96.80
XGB	33.33	49.16	53.33	69.86	99.65
ADA	25.90	53.68	38.81	74.58	89.58
KNN	23.54	26.94	52.50	48.95	86.11
GBDT	25.48	55.20	27.15	47.63	99.72
LGBM	33.33	33.33	48.95	61.59	100.0
Ours	88.40				

TABLE VII
COMPARISON WITH FEW-SAMPLE LEARNING ON GROUP C

Acc(%)	Number of training samples for each fault				
	1	10	50	200	500
LSVM	32.15	23.05	27.43	49.51	62.01
RF	27.84	27.56	35.55	60.34	79.44
NB	32.15	29.44	41.31	77.29	79.37
XGB	33.33	25.00	29.37	56.25	82.15
ADA	25.83	26.59	28.75	58.05	70.90
KNN	32.15	31.66	31.80	53.12	67.70
GBDT	32.50	29.79	31.80	57.84	82.29
LGBM	33.33	33.33	30.20	59.79	81.66
Ours	67.56				

since the zero-sample setting and one-sample setting are totally different and more difficult. The concluded fault descriptions indeed offer additional fault information for the diagnosis.

5) *Robustness to the Noises*: The noises are typical and widespread disturbance in the industrial process. Hence, the noise experiment is conducted to show the robustness of the proposed method, in which the training and test faults are added with noises. The added noises are assigned to the Gaussian distributions with zero mean and varying variance. NB is adopted as the attribute learners. The results are shown in Table VIII.

Generally, the proposed method performs well when the variance of noise is less than 0.5. When the variance of noise is larger than 0.5, the performance is degraded. And sometimes,

TABLE VIII
RESULTS OF THE NOISE EXPERIMENTS ON GROUPS A AND C

Group	A		C	
Noise	Acc(%)	Att(%)	Acc(%)	Att(%)
—	80.27	70.51	67.56	60.69
$G(0.1)$	80.06	70.51	64.93	58.59
$G(0.3)$	82.56	70.50	65.00	58.55
$G(0.5)$	83.68	70.52	66.04	58.42
$G(0.7)$	73.33	69.72	63.26	57.93
$G(1.0)$	73.26	69.99	60.09	57.76
$G(5.0)$	65.55	67.21	57.63	56.83

$G(v)$ denotes the Gaussian noise, where v is the variance. And the mean of the noise is zero. “Acc” denotes the accuracy of zero-sample fault diagnosis, and “Att” is the corresponding average accuracy of 20 attribute learners. NB is adopted as the attribute learners.

the system performs even better with a little noise, such as the performance on group A when the variance of noise is set as 0.3. The proposed method diagnoses faults based on the recognition of different attributes, which is a basic binary classification task. As shown in Table VIII, both the average accuracies of 20 attribute learners for group A and group C are stable under the noisy settings, which help to obtain the robust results for the zero-sample task.

B. Thermal Power Plant Process

1) *Introduction for Data set*: The thermal power plant (TPP) fault data set is acquired from a real industrial process of a 1000 MW ultrasupercritical thermal power unit which is shown in Fig. 6. The thermal process consists of two main subsystems, i.e., the boiler system and the steam turbine system. During the generating process, the boiler system first heats the water to the steam with high pressure and temperature. And then the steam is transported to the turbine system to drive the electrical generator. The whole power unit achieves the transformation from chemical energy to electrical energy. Since the thermal power generation is a large-scale process, collecting fault samples by unit shutdown could cause huge losses. Hence, it is of great significance to diagnose certain fault categories under the zero-sample setting.

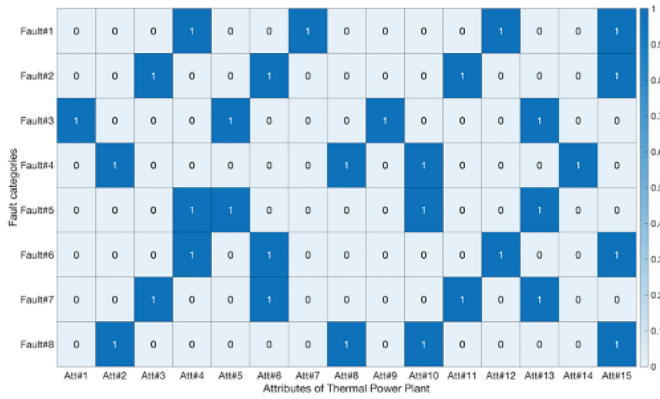


Fig. 7. Fault description matrix A for TPP. Each kind of fault is described by 15 fine-grained attributes. The “1” in figure denotes the fault has this attribute, and the “0” denotes not. The specific names of attributes are shown in Table IX.

TABLE IX
INFORMATION OF ATTRIBUTES FOR THE TPP

No.	Attributes
Att#1	Occurred at the 5# low pressure heater
Att#2	Occurred at the coal mill
Att#3	Occurred at the small turbine
Att#4	Occurred at the condenser pump
Att#5	Related with valve
Att#6	Related with bearing
Att#7	Related with coil
Att#8	Related with grand
Att#9	The shift problem
Att#10	The pressure problem
Att#11	The vibration problem
Att#12	The temperature problem
Att#13	Abnormal index value
Att#14	Low index value
Att#15	High index value

TABLE X
FOUR GROUPS OF TRAIN/TEST SPLIT FOR TPP

No.	Training Faults	Target Faults
A	4, 5, 6, 7, 8	1, 2, 3
B	1, 2, 3, 7, 8	4, 5, 6
C	2, 3, 4, 5, 6	7, 8, 1
D	1, 3, 4, 6, 7	2, 5, 8

The data set offers eight kinds of faults which occur over the whole unit. Each kind of fault is illustrated by 68 measured signals and 4320 samples. The faults are described with 15 kinds of attributes. The fault description matrix A for the thermal power plant data set is shown in Fig. 7 and the specific attribute information is shown in Table IX. To test the robustness of the proposed fault description based attribute transfer method, four different groups of train/test split are adopted and shown in Table X. Each group contains five kinds of faults for model training and three kinds of faults for test. The model implementation is the same as that for TEP.

2) *Results of Zero-Sample Fault Diagnosis:* The results of zero-sample fault diagnosis for the TPP process are shown in Table XI. The confusion matrices for the best results of the

TABLE XI
RESULTS OF ZERO-SAMPLE FAULT DIAGNOSIS FOR TPP(%)

Group.	LSVM	RF	NB	Heighest	Mean
A	92.53	99.48	90.82	99.48	94.28
B	99.90	83.31	89.69	99.90	90.97
C	83.91	80.44	90.87	90.87	85.07
D	92.21	99.96	92.63	99.96	94.93

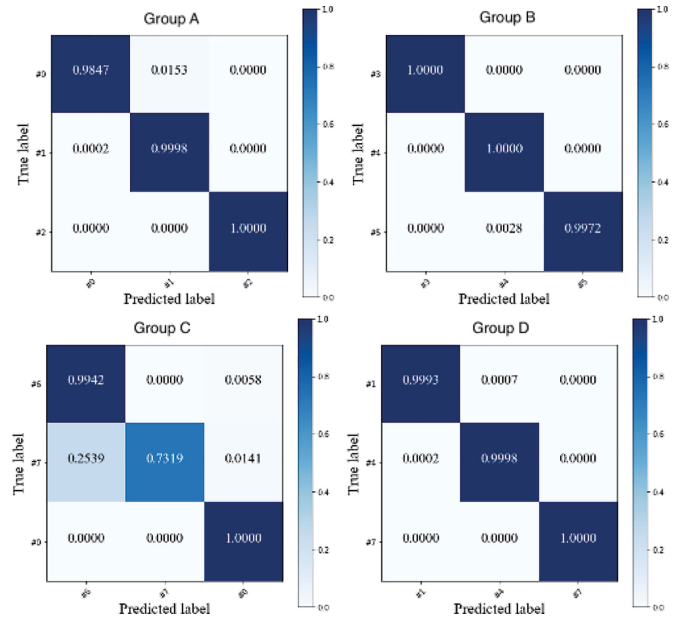


Fig. 8. Confusion matrices of the best results of the four groups for the TPP process.

four groups experiments are shown in Fig. 8. Generally, the results of zero-sample fault diagnosis for TPP are much higher than those for TEP. For the four different groups, the average accuracies vary from 85.07 to 94.93%, whereas those for TEP vary from 51.03 to 75.78%. This may be explained by the fault description matrix A . The definition of A for the TPP process reveals more detailed information in comparison with that of TEP. For example, the specific process variables, i.e., the “temperature,” the “pressure,” and the “vibration,” and the specific properties of process variables, i.e., the “low” and the “high,” are provided in Table IX. The detailed information can be easily learned from the sensor signals and applied to recognize attributes accurately. And hence, all the three machine learning algorithms, i.e., LSVM, RF, and NB, can obtain good performance on the target faults. The accuracies of the 15 attribute learners for group A with RF as attribute learners are shown in Fig. 9. The average accuracy of 20 attribute learners for the group A of TEP with RF as attribute learners is 71.45%, while that of TPP is 75.06%. This validates the effectiveness of the proposed fault description based attribute transfer method, in which the fault descriptions and attributes are well-learned first, and then the faults can be well-diagnosed. Also, since many machines in the TPP process are rotary machines, such as the turbine, the generator, and the coil mill, the process generally works under the noisy environment. Hence, the high

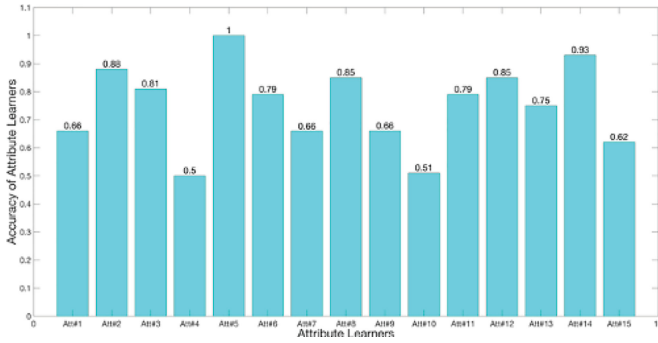


Fig. 9. Accuracies of 15 attribute learners for the group A with RF as attribute learners for the TPP process.

TABLE XII
COMPARISON WITH ZERO-SHOT LEARNING ON TPP(%)

Group.	A	B	C	D	Mean
DAP	96.20	81.32	89.21	93.67	90.10
IAP	92.28	96.33	81.33	90.09	90.01
SJE	66.78	99.33	68.20	99.33	83.41
ESZSL	62.78	96.28	72.37	94.96	81.59
Ours(RF)	99.48	83.31	80.44	99.96	90.79
Ours(NB)	90.82	89.69	90.87	92.63	91.00
Ours(LSVM)	92.53	99.90	83.91	92.21	92.13

Note: Both of the deep features of convolutional networks and the image-based attributes required by SJE and ESZSL are unavailable under the fault diagnosis scenario. Their unsatisfactory performance on groups A and C reveals their poor generalization ability for the zero-sample fault diagnosis task.

performance shown in Table XI and Fig. 9 also validates the robustness and model-reality match of the proposed method.

We also compare our method with other zero-shot learning methods. The experiment setting is the same as that of TEP. The results are shown in Table XII. For the TPP process, the linear LSVM presents higher average accuracies through four different groups, which reveals the linear characteristic of the process. The SJE and ESZSL show unstable performance on group A and group B, since the designed deep features are inapplicable here. Also, DAP and IAP present lower accuracies in comparison with ours. DAP directly learn from the fault samples in experiment, while the designed model is assigned to learn from the attribute related features by the supervised principle component analysis. Since both of DAP and ours are based on the transfer of attributes, the results may reveal the importance of the features for the zero-sample fault diagnosis task, which can be indicated by (5).

V. CONCLUSIONS

Considering the difficulty of collecting fault samples, we formulate the zero-sample fault diagnosis task in this article. And the fault description based attribute transfer is designed as the first attempt to diagnose target faults without their samples. In comparison with the few-sample learning on the benchmark data set, the advantage that zero sample of target faults is required can be observed. The high accuracies on the real TPP process and the feasibility analysis also show that with the well-designed attribute descriptions for faults, it is indeed possible to achieve

the zero-sample diagnosis. And there are many works that can be developed and improved upon in future study. For examples, first, the solution of zero-sample diagnosis that generating samples for the target faults based on their fault descriptions using some popular generative models, i.e., the generative adversarial networks, can be considered; second, in this article, the supervised principle component analysis was utilized as the basic method to extract the attribute related features for the zero-sample diagnosis, and more meaningful and interpretable methods can be developed to conduct the task better.

REFERENCES

- [1] Y. Ma, S. Bing, H. Shi, and Y. Yang, "Fault detection via local and nonlocal embedding," *Chem. Eng. Res. Des.*, vol. 94, pp. 538–548, 2015.
- [2] Q. Liu, S. J. Qin, and T. Chai, "Decentralized fault diagnosis of continuous annealing processes based on multilevel PCA," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 687–698, Jul. 2013.
- [3] W. Fan *et al.*, "Hidden Markov model-based fault detection approach for multimode process," *Ind. Eng. Chem. Res.*, vol. 55, no. 16, pp. 4613–4621, 2016.
- [4] S. Zhao, B. Huang, and L. Fei, "Fault detection and diagnosis of multiple-model systems with mis modeled transition probabilities," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5063–5071, Aug. 2015.
- [5] Q. Liu, Q. Zhu, S. J. Qin, and T. Chai, "Dynamic concurrent kernel CCA for strip-thickness relevant fault diagnosis of continuous annealing processes," *J. Process Control*, vol. 67, pp. 12–22, 2017.
- [6] G. Xin and H. Jian, "An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process," *Neurocomputing*, vol. 32, pp. 1023–1034, 2015.
- [7] X. Q. Deng, X. M. Tian, and X. Y. Hu, "Nonlinear process fault diagnosis based on slow feature analysis," in *Proc. 10th World Congr. Intell. Control Autom.*, 2012, pp. 3152–3156.
- [8] L. Cui *et al.*, "Sensor-based vibration signal feature extraction using an improved composite dictionary matching pursuit algorithm," *Sensors*, vol. 14, no. 9, pp. 16715–16739, 2014.
- [9] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, 2019.
- [10] Y. L. Murphey, M. A. Masrur, Z. H. Chen, and B. Zhang, "Model-based fault diagnosis in electric drives using machine learning," *IEEE/ASME Trans. Mechatronics*, vol. 11, no. 3, pp. 290–303, Jun. 2006.
- [11] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse auto-encoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2416–2425, Apr. 2019.
- [12] J. Zhao *et al.*, "The modeling procedures for model-based diagnosis of slowly changing fault in hybrid system," *Adv. Mater. Res.*, vol. 186, pp. 403–407, 2011.
- [13] Y. Pan *et al.*, "An approach for HVCB mechanical fault diagnosis based on a deep belief network and a transfer learning strategy," *J. Elect. Eng. Technol.*, vol. 14, no. 1, pp. 407–419, 2019.
- [14] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly-accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.
- [15] W. Long, G. Liang, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2019.
- [16] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017.
- [17] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [18] W. S. Lee, D. L. Grosh, F. A. Tillman, and C. H. Lie, "Fault tree analysis, methods, and application: A review," *IEEE Trans. Rel.*, vol. RE-34, no. 3, pp. 194–203, Aug. 1985.
- [19] A. K. Reay and J. D. Andrews, "A fault tree analysis strategy using binary decision diagrams," *Rel. Eng. Syst. Saf.*, vol. 78, no. 1, pp. 45–56, 2002.
- [20] R. M. Sinnamon and J. D. Andrews, "Improved efficiency in qualitative fault tree analysis," *Qual. Rel. Eng. Int.*, vol. 13, no. 5, pp. 293–298, 1997.
- [21] C. Samir, *Fault Tree Analysis*. Hoboken, NJ, USA: Wiley 2006.

- [22] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 951–958.
- [23] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 453–465, Mar. 2014.
- [24] Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele, "Evaluation of output embeddings for fine-grained image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2927–2936.
- [25] B. Romera-Paredes and P. H. Torr, "An embarrassingly simple approach to zero-shot learning," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 2152–2161.
- [26] S. Kang, D. Lee, and C. D. Yoo, "Face attribute classification using attribute-aware correlation map and gated convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 4922–4926.
- [27] C. H. Zhao and Y. X. Sun, "Comprehensive subspace decomposition and isolation of principal reconstruction directions for online fault diagnosis," *J. Northwest A F Univ.*, vol. 23, no. 10, pp. 1515–1527, 2013.
- [28] C. H. Zhao and F. R. Gao, "Fault subspace selection approach combined with analysis of relative changes for reconstruction modeling and multifault diagnosis," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 3, pp. 928–939, May 2016.
- [29] J. C. Wang *et al.*, "Dimension reduction method of independent component analysis for process monitoring based on minimum mean square error," *J. Process Control*, vol. 22, no. 2, pp. 477–487, 2012.
- [30] C. H. Zhao, W. Q. Li, and Y. X. Sun, "A sub-principal component of fault detection modeling method and its application to online fault diagnosis," in *Proc. 9th Asian Control Conf.*, 2013, pp. 1–6.
- [31] J. H. Chen *et al.*, "Principle component analysis based control charts with memory effect for process monitoring," *Ind. Eng. Chem. Res.*, vol. 40, no. 6, pp. 1516–1527, 2001.
- [32] M. Misra *et al.*, "Multivariate process monitoring and fault diagnosis by multi-scale PCA," *Comput. Chem. Eng.*, vol. 26, no. 9, pp. 1281–1293, 2002.
- [33] M. Li and X. Wu, "Fault diagnosis for fans of coal based on CBR hybrid threshold method," in *Proc. 7th Int. Conf. Fuzzy Syst. Knowl. Discovery*, 2010, pp. 2008–2011.
- [34] E. Barshan *et al.*, "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds," *Pattern Recognit.*, vol. 44, no. 7, pp. 1357–1371, 2011.
- [35] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, 1993.
- [36] W. Yu and C. Zhao, "Online fault diagnosis in industrial process using multi-model exponential discriminant analysis algorithm," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 1317–1325, May 2019.
- [37] S. Ashish and J. Ritesh, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, 2012.
- [38] T. Q. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD*, 2016, pp. 785–794.
- [39] H. J. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [40] G. L. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," *Adv. Neural Inf. Process. Syst.*, 2017, pp. 3149–3157.



Liangjun Feng received the B.Eng. degree in automation from North China Electric Power University, Beijing, China, in 2017 and is currently working toward the Ph.D. degree in control science and engineering with the College of Control Science and Engineering, Zhejiang University, Hang-Zhou, China.

During his B.Eng. degree, he researched with the Laboratory of Guotian Yang and studied as an exchange student with the University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA. His current research interests include machine learning, artificial intelligence, and pattern recognition.

Mr. Feng was the recipient of the Beijing outstanding graduate honor.



Chunhui Zhao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in control science and engineering from Northeastern University, Shenyang, China, in 2003, 2006, and 2009, respectively.

Since January 2012, she has been a Professor with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. She has authored or coauthored more than 120 papers in peer-reviewed international journals. From January 2009 to January 2012,

she was a Postdoctoral Fellow with the Hong Kong University of Science and Technology and the University of California, Santa Barbara, Los Angeles, CA, USA. Her current research interests include statistical machine learning and data mining for industrial application.

Prof. Zhao was the recipient of the National Top 100 Excellent Doctor Thesis Nomination Award, New Century Excellent Talents in University, China, and National Science Fund for Excellent Young Scholars, respectively.