



A Novel Zero-Shot Fault Identification Based on Transfer Learning

Yuyang Gui, Minhan Yi, Hongpeng Yin^(✉), Peng Zhang, Dandan Zhao,
and Li Cai

Chongqing University, Chongqing 404100, China
yinhongpeng@gmail.com

Abstract. In industrial fault identification tasks, for the problems of inconvenient fault data collecting and lack of target fault samples, unlike many existing zero-shot fault identification methods that the manual fault attribute description requires in-depth knowledge on fault data, this paper proposes a novel zero-shot fault identification method based on transfer learning. We introduce PCA algorithm to extract potential fine-grained feature representation by constructing a discriminant attribute extractor for source and target domain, as a bridge for knowledge transfer. With source domain fault data, the shared fine-grained groups of all the fault categories can be obtained, and transferred as knowledge to the fault representation of target domain. Both the discriminant features can be formed, from the discriminant matrices in source and target domains learnt with the shared fine-grained group. Finally, zero-shot fault identification with discriminant attributes is achieved. Based on 3W datasets, the proposed method is compared with other zero-shot fault identification methods, and the experiment results show the effectiveness of the proposed method.

Keywords: Transfer learning · Zero shot · Fault identification · Fine-grained knowledge

1 Introduction

Fault identification is a technology to evaluate system device running state by device data already detected, thus it is vital to ensure the safe, steady, reliable work of industrial devices. Along with the digital and information time, data-driven technology is widely used in the processes of monitoring and diagnosing in industry [1]. In particular, fault identification based on data-driven has become the mainstream method.

Fault identification methods under data-driven directly learn the fault features from historical fault data during the system running time, thus can diagnose without exact system mathematical models. But most large-scaled industrial devices are not permitted to run until fault happens, and devices change from normal to breakdown gradually, the actual target fault samples are few.

Thus it is costing to collect samples, suppressing the applications of the traditional data-driven methods in fault diagnosis greatly [2–4].

Recently, a large quantities of fault identification methods based on transfer learning are proposed to solve the problem of fault sample lack [5, 6]. The main idea of methods based on transfer learning is, the relative and labeled data from one or more domains can be used as auxiliary knowledge to improve the model performance in target domain, and to complete the target tasks in target domain. Thus the transfer learning methods have a strong ability of representation learning and end-to-end training. Lu et al. [7] proposed a deep neural network model with domain adaptive for fault diagnosis. Yang et al. [8] proposed a polynomial kernel induced distance measurement method for deep transferring fault diagnosis. These deep transferring fault diagnosis models essentially deal with domain transition between source and target domain. Also, researchers dedicate to solving the problems of unknown fault diagnosis [9] and unlabeled fault diagnosis [10]. Gupta et al. [9] proposed an early classification method of multiple time Series (MTS) based on semantic information. Attribute learning model is used to obtain the semantic information of known fault categories, and then the semantic information of visible classes is used to predict unknown fault. Although the above method is similar to zero-shot, it is not zero-shot fault diagnosis problem in nature. For zero-shot fault diagnosis, Feng et al. [11] proposed an attribute transfer method based on fault description. By transferring the artificially defined fault descriptions to target domain, the source domain training model can be used for fault diagnosis in target domain. However, this method requires manual setting of fault attribute description and in-depth understanding of fault data.

To deal with above problems, this paper proposes a zero-shot fault identification method based on transfer learning. The proposed method is to obtain the potential fine-grained feature representation of source domain data samples, as a knowledge transferring bridge, and to construct a discriminant attribute extractor for both source and target domains. First, the proposed method utilizes PCA algorithm to extract the main fault attribute [12]. Then the shared fine-grained groups of all known fault categories are obtained according to fault date in source domain, and transferred to the representation in target domain. Both the discriminant features in source and target domain are formed by discriminant matrices learnt from the shared fine-grained groups. Finally, we can use these discriminant features to zero-shot fault diagnosing.

2 Methodology

In this chapter, the framework of the proposed MK-CSE is introduced. When new data comes, the framework embeds it into the regenerative kernel Hilbert space to obtain the optimal mapping kernel, and MK-MMD between data blocks is maximized to ensure the diversity of retained knowledge, which has better robustness and accuracy than single kernel measurement.

2.1 Zero-Shot Fault Identification Definition

Assuming that there is k classes of fault samples, then the corresponding formula is $X \in R^{a \times b}$, and the labels can be described as $Y \in R^{a \times 1}$. Where a and b are the number and dimensions of samples. Zero-shot fault identification, means that the fault data to be diagnosed has no available sample to train models (namely as target domain), while other relative and available fault data (namely source domain) is required as the auxiliary knowledge to train models instead. Thus, data from source domain is defined as $D_O = \{X_O \in R^{a_o \times b}, Y_O \in R^{a_o \times 1}\}$, while x_{oi} and y_{oi} stand for the i th sample and label from source domain, p_o is the number of fault categories in $y_{oi} = 1, 2, \dots, p_o$.

Similarly, data from target domain is defined as $D_T = \{X_T \in R^{a_t \times b}, Y_T \in R^{a_t \times 1}\}$, while x_{ti} and y_{ti} stand for the j th sample and label from target domain, p_t is the number of fault categories in $y_{ti} = 1, 2, \dots, p_t$. Then the following formula describes the zero-shot fault identification model:

$$\hat{Y}_T = func(X_T, D_O(X_O, Y_O)), s.t. \hat{Y}_T = Y_T \tag{1}$$

wherein, $func()$ is the mapping function from source domain to target domain.

2.2 Fault Fine-Grained Representation and Discriminant Feature Extraction

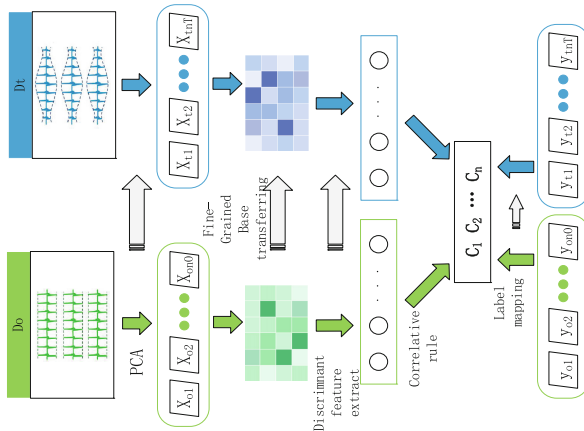


Fig. 1. Diagram of zero sample fault identification method based on Transfer Learning

The proposed zero-shot fault identification method based on transfer learning is shown as Fig. 1, divided into training stage and test stage. The blue stands for the training stage, which is aimed to learn the shared fine-grained representation of fault by using source domain data, and transfer it to the target domain, thus

the discriminant feature used to classify can be learnt. The red is the test stage, which is mainly to extract the discriminant feature according to the fine-grained representation in training stage, to achieve fault classifying. Because the labels in source domain and target domain are different, labels in source domain need to be projected into the corresponding labels under the fine-grained representation.

For the given large-scale fault data in source domain X_O , its main attribute extracted is $P(X_O) = X_O^P = \{f_{o1}^p, f_{o1}^p, \dots, f_{o1}^p\}$ after PCA analysis. Referring to the idea of Fig. 1, we use X_O^P to learn the shared fine-grained attribute and discriminant features first. Then X_O^P can be divided into a shared fine-grained base and a discriminant matrix as shown:

$$\min_{G, C_O^G} \|X_O^P - GC_O^G\|_F^2 \quad (2)$$

wherein, $G \in R^{a_o \times n}$ stands for the shared fine-grained base matrix, n is the atom number of the fine-grained base, thus we can describe X_O^P with more fault categories. And $C_O^G \in R^{n \times b}$ intends to be the discriminant matrix under the fine-grained base matrix G . We also use $tr((C_O^G)^T Y^O C_O^G)$ as restraint to strengthen the discriminant ability of the discriminant matrix. But the labels in source and target domains differ a lot after mapping, the restraint is changed into $tr((C_O^G)^T C_O^G)$. Thus Formula (2) can be rewritten as:

$$\min_{G, C_O^G} \|X_O^P - GC_O^G\|_F^2 + \|C_O^G\|_2 \quad (3)$$

Through Formula (4), we obtain the discriminant features of faults X_O^G :

$$X_O^G = X_O^P C_O^G \quad (4)$$

The procedures above demonstrate how to obtain the shared fine-grained base matrix and discriminant features from source domain. In a similar way, the matrix is transferred to target domain to get the discriminant matrix of target domain, as shown:

$$\min_{C_T^G} \|X_T^P - GC_T^G\|_F^2 + \|C_T^G\|_2 \quad (5)$$

Then the discriminant features are $X_T^G = X_T^P C_T^G$.

Optimization: For the similarity of our objective function and dictionary learning objective function, we can use the same solving processes as dictionary learning. For more precise shared fine-grained base matrix, the proposed method use the same optimization process as KSVD [13] to solve models. Fix one optimization variable in C_O^G and G , and use it to optimize the other, as a loop.

First, we initialize G , and use orthogonal matching tracking algorithm to obtain C_O^G . Then we upgrade the shared fine-grained base matrix G and the discriminant matrix C_O^G of each row in the way of row by row. g_n is the n th vector of the shared fine-grained base matrix, and c_n^T is the n th vector of the discriminant mapping matrix. Formula (2) can be rewritten as:

$$\begin{aligned} \|X_O^P - GC_O^G\|_F^2 &= \left\| X_O^P - \sum_{i=1}^N g_i c_i^T \right\|_F^2 \\ &= \|E_n - g_n c_n^T\|_F^2 \end{aligned} \quad (6)$$

wherein, $E_n = X_O^P - \sum_{i \neq n} g_i c_i^T$. The optimization of the shared fine-grained base matrix G is described as:

$$\min_{g_n c_n^T} \|E_n - g_n c_n^T\|_F^2 \quad (7)$$

In Formula (7), the optimal g_n and c_n^T obtained by SVD, can be the approximate terms of 1-rank matrix E_n . But c_n^T just by putting E_n into SVD solution is not sparse, thus we extract the positions that its corresponding c_n^T not equal to 0, and restructure a new E'_n . Then decompose it with SVD, as $E'_n = L \sum R^T$. Take the first column vector of the left singular matrix L as g_n , and the product of the first column vector and the first singular value of the right singular matrix as c_n^T . After obtaining the shared fine-grained base matrix G based on source domain data, C_T^G can be obtained by utilizing orthogonal matching tracking algorithm. The full optimization solution process is shown in Algorithm 1.

Algorithm 1: Model optimization solution

Step 1 initialize shared fine-grained base group G

Step 2 use orthogonal matching tracking algorithm on Formula(2) to obtain discriminant matrix C_O^G in source domain

Step 3 update g_n & c_n in a loop

Step 4 obtain X_O^G based on Formula(4)

Step 5 use orthogonal matching tracking algorithm on Formula(5) to obtain discriminant matrix C_T^G in target domain

Step 6 obtain discriminant features X_T^G based on $X_T^G = X_T^P C_T^G$

3 Experiments

For further evaluate the proposed method in this paper, we go on experiments with TE (Tennessee-Eastrman) process dataset. The detailed experimental settings are shown as below.

Experiment Dataset: TE process dataset is simulated based on a real industrial process. It consists of five major subsystems, including a reactor, a condenser, a vapor-liquid separator, a recycle compressor, and a product stripper. It contains 21 types of faults, each of which is composed of 41 measurement control variables and 11 process control variables. TE data is divided into training set and test set. Each type of fault in the training set contains 480 samples, with a total of 21×480 training samples. Each type of fault in the test set contains 960 samples, a total of 21×960 test samples.

Experiment Settings: First, the proposed method extracts the relative features of the faults by PCA, and obtains the main relative features of the source-domain faults X_O^P and the target-domain faults X_T^P . For each attribute of faults, we extract 20 main relative features from 52 original variables. Second, the fine grit is set to 35, both the shared fine-grained base and the source domain discriminant matrix are learnt from the source domain data. Third, use the shared fine-grained base to obtain the target domain discriminant matrix to finally obtain the discriminant features of the source and target domain. For the problem of different labels in source and target domain, the proposed method utilizes linear mapping function to change the source-domain label y_O and target-domain label y_T to labels $z_{O_k}, z_{T_k} = \{l_1, l_2, \dots, l_k\}$ which have the same information. The linear mapping function is followed:

$$y_O = wz_O \quad (8)$$

Then the mapping representation of label in target domain is $y_T = wz_T$. For zero-shot fault identification defines that there is no available training sample for the fault category to be diagnosed, the experimental Settings in this paper are the same as those in [11], and only the training set part is used. As the last 6 faults of the 21 fault classes of TE data are rarely described in the data set, 80% fault classes of the first 15 fault classes of all fault classes are used for training and 20% fault classes are used for testing. The experiment was divided into 5 groups, each group had 12 fault classes for training and 3 fault classes for testing. See Table 1 for specific categories.

In addition, this paper adopts four algorithms for fault classification, namely deep belief network (DBN) [14], support vector machine (SVM) [15], Random forest (RF) [16] and Naive Bayes (NB) [17]. The DBN is formed by two layers of 100 dimensions, the number of epochs is 100 and the size of batch is 120, the learning rate is set to 1. The LSVM parameter is set to ('-s 0 -t 0 -c 1'), and the number of decision trees for RF is set to 50. We tested the accuracy of five experiments. The calculation formula of fault classification accuracy is as follows:

$$Acc = \frac{\hat{W}_T}{W_T} \quad (9)$$

where, \hat{W}_T is the number of samples which are correctly classified, and W_T is the total number of samples of target faults in the test stage.

Results and Analysis: The diagnostic results of five groups of experiments under four algorithms are shown in Table 2. DBN classifier reaches highest at 72.64% on group B, LSVM classifier reaches highest at 84.31%, 56.11%, 44.03%, on group A, group C and group E, RF classifier reaches highest at 63.68% on group D. Since the significance characteristics of each fault are different, the fault identify accuracy varies from 44.03% to 88.47%. Although the highest fault identification accuracy of the four classifiers varies from 44.03% to 88.47%, there is still some space to improve.

Table 1. Experiment groups

Group	Train classes	Test classes
A	2 3 4 5 7 8 9 10 11 12 13 15	1 6 14
B	1 2 3 5 6 8 9 11 12 13 14 15	4 7 10
C	1 2 3 4 5 6 7 9 10 13 14 15	8 11 12
D	1 4 6 7 8 9 10 11 12 13 14 15	2 3 5
E	1 2 3 4 5 6 7 8 10 11 12 14	9 13 15

Table 2. Experiment results of 4 algorithms on 5 groups

Group	DBN	LSVM	RF	NB	Highest
A	84.31	88.47	82.92	85.83	88.47
B	72.64	69.38	66.11	69.17	72.64
C	52.29	56.11	54.58	49.79	56.11
D	56.25	55.63	63.68	61.04	63.68
E	42.29	44.03	29.65	28.61	44.03

Table 3. Comparison between our method and other SOTA methods

Group	DAE [18]	IAP [19]	SJE [20]	ESZSL [21]	Feng [11]	Ours
A	54.16	55.48	74.58	57.22	75.78	88.47
B	62.63	60.69	33.12	33.33	55.54	72.64
C	40.13	45.00	33.95	39.51	53.46	56.11
D	55.48	36.25	63.88	39.65	62.50	63.68
E	36.94	48.88	33.81	33.33	51.03	44.03
Mean	49.86	49.26	47.86	40.60	59.66	64.99

The paper also compared with DAP [18], IAP [19], SJE [20], ESZSL [21] and Feng et al. [11]. The results are shown in Table 3. The proposed method achieved the best identification results in group *A*, *B*, *C*, *D*, and the average accuracy of the five groups of experiments reached highest at 64.99%. Table 3 shows that the shared fine-grained base transferring proposed in this paper can effectively solve the problem of zero-shot fault identification.

The proposed method aims at transferring source domain to share fine-grained base matrices. Therefore, the identification results of representation matrix migration at different granularity are discussed. In this paper, comparison tests under five different fine grits are set as $k = [20, 25, 30, 35, 40]$, and the results are shown in Fig. 2. It can be seen that the overall trend is that the larger the fine grit is, the more accurate the identification result will be, which is also more consistent with the practical theoretical interpretation, that is, within a

certain range, using more dimensions to describe the fault features can make the classification effect better.

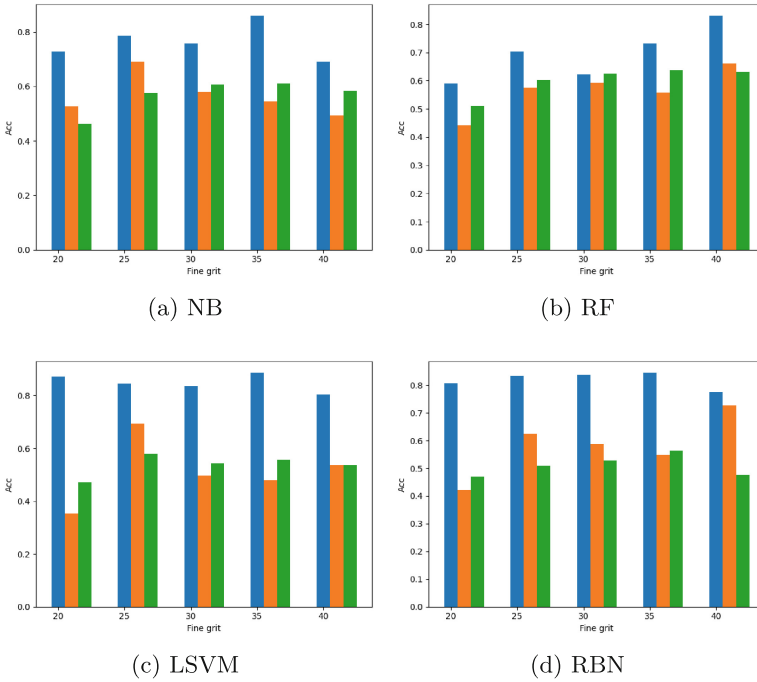


Fig. 2. Experiment results under different sizes of fine grit

4 Conclusion

In actual applications of large-scaled complex devices, for the problems that fault samples are usually few and hard to collect, and take a lot of human and material resources, we propose a zero-shot fault identification method based on transfer learning. Under the situation of no available sample, the proposed method transfers the fine-grained representation of fault data in other relative domains to target domain, thus to diagnose the faults in target domain. Compared with existing zero-shot identification methods, our method utilizes the discriminant feature extraction based on shared fine-grained base representation. The fine-grained base are flexible in dimensions, and the fine-grained base description matrices can be obtained automatically according to algorithms, without setting fine-grained description manually. On TE process dataset, the average identify accuracy is 64.99%, which gains the highest accuracy compared with other methods, and analysis results on different experiments also demonstrate the effectiveness of the proposed method.

Acknowledgment. This work is jointly supported by Graduate research and innovation foundation of Chongqing (Grant No. CYS21064), Chongqing Talents: Exceptional Young Talents Project (cstc2021ycjh-bgzxm0028), China Central Universities foundation (2019CDYGGZD001), and Scientific Reserve Talent Programs of Chongqing University (cqu2018CDHB1B04).

References

1. Liu, R., Yang, B., Enrico, Z., Chen, X.: Artificial intelligence for fault of rotating machinery: a review. *Mech. Syst. Sig. Process.* **108**, 33–47 (2018). <https://doi.org/10.1016/j.ymssp.2018.02.016>
2. Wang, F., Tan, S., Yang, Y., Shi, H.: Hidden Markov model-based fault detection approach for a multimode process. *Industr. Eng. Chem. Res.* **55**(16), 4613–4621 (2016). <https://doi.org/10.1021/acs.iecr.5b04777>
3. Zhang, C., Wang, Z., Wen, C., Liu, G., Yu, W.: Sample space based on multi-level high dimensional feature representation micro-fault diagnosis. *Acta Electron. Sin.* **48**(08), 1647–1654 (2020). <https://doi.org/10.3969/j.issn.0372-2112.2020.08.026>
4. Bao, Z., Wen, C., Ma, X.: Data preprocessing and PCA fault diagnosis method based on rate of change transformation. *Acta Electron. Sin.* **49**(11), 2234–2240 (2021). <https://doi.org/10.12263/DZXB.20201225>
5. Yu, C., Tian, R., Tan, L., Tu, X., et al.: Integrated transfer learning algorithmic for unbalanced samples classification. *Acta Electron. Sin.* **40**(7), 1358–1363 (2012). <https://doi.org/10.3969/j.issn.0372-2112.2012.07.012>
6. Ji, D., Jiang, Y., Wang, S.: Multi-source transfer learning method by balancing both the domains and instances. *Acta Electron. Sin.* **47**(03), 692–699 (2019). <https://doi.org/10.3969/j.issn.0372-2112.2019.03.025>
7. Lu, W., Liang, B., Yu, C., Meng, D., Tao, Z.: Deep model based domain adaptation for fault diagnosis. *IEEE Trans. Industr. Electron.* **64**(3), 2296–2305 (2016). <https://doi.org/10.1109/TIE.2016.2627020>
8. Yang, B., Lei, Y., Jia, F., Du, Z.: A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines. *IEEE Trans. Industr. Electron.* **67**(11), 9747–9757 (2020). <https://doi.org/10.1109/TIE.2019.2953010>
9. Gupta, A., Gupta, H.P., Biswas, B., Dutta, T.: An unseen fault classification approach for smart appliances using ongoing multivariate time series. *IEEE Trans. Industr. Inf.* **17**, 3731–3738 (2021). <https://doi.org/10.1109/TII.2020.3016590>
10. Guo, L., Lei, Y., Xing, S., Yan, T., Li, N.: Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Trans. Industr. Electron.* **66**, 7316–7325 (2019). <https://doi.org/10.1109/TIE.2018.2877090>
11. Feng, L., Zhao, C.: Fault description based attribute transfer for zero-sample industrial fault diagnosis. *IEEE Trans. Industr. Inf.* **17**, 1852–1862 (2020). <https://doi.org/10.1109/TII.2020.2988208>
12. Barshan, E., Ghodsi, A., Azimifar, Z., Jahromi, M.Z.: Supervised principal component analysis: visualization, classification and regression on subspaces and submanifolds. *Pattern Recogn.* **44**(7), 1357–1371 (2011). <https://doi.org/10.1016/j.patcog.2010.12.015>
13. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing over-complete dictionaries for sparse representation. *IEEE Trans. Sig. Process.* **54**(11), 4311–4322 (2006). <https://doi.org/10.1109/TSP.2006.881199>

14. Network, A., Related, S., Rank, T., Network, A., Related, S., Rank, T.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2006)
15. Chang, C., Lin, C.: LibSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2**(3), 1–27 (2011). <https://doi.org/10.1145/1961189.1961199>
16. Cutler, A.D., Richard, C., John, R.S.: Random forests. In: Zhang, C., Ma, Y. (eds.) *Ensemble Machine Learning*, pp. 157–175. Springer, Boston (2012). https://doi.org/10.1007/978-1-4419-9326-7_5
17. Murphy, K.P.: Naive Bayes classifiers. *Univ. Br. Columbia* **18**(60), 1–8 (2006)
18. Lampert, C.H., Nickisch H., Harmeling S., Learning to detect unseen object classes by between-class attribute transfer. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 951–958 (2009)
19. Lampert, C.H., Nickisch, H., Harmeling, S.: Attribute-based classification for zero-shot visual object categorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(3), 453–465 (2014). <https://doi.org/10.1109/TPAMI.2013.140>
20. Akata, Z., Reed, S., Walter, D., et al. Evaluation of output embeddings for fine-grained image classification. In: *IEEE Computer Vision and Pattern Recognition*, pp. 2927–2936 (2015)
21. Romera, B., Torr, P.H.: An embarrassingly simple approach to zero-shot learning. In: *Proceedings of the 32nd international conference on Machine learning*, pp. 152–2161 (2015)