Fault Diagnosis and Knowledge Extraction Using Fast Logical Analysis of Data with Multiple Rules Discovery Ability

Xiwei Bai, Jie Tan, Xuelei Wang

Fault Diagnosis and Knowledge Extraction Using Fast Logical Analysis of Data with Multiple Rules Discovery Ability

Xiwei Bai[1,2], Jie Tan[1], Xuelei Wang[1]

[1] Institute of Automation, Chinese Academy of Sciences
Beijing 100190, China
`xiwei_bai@163.com`
[2] University of Chinese Academy of Sciences
Beijing 100049, China

ABSTRACT. Data-based explanatory fault diagnosis methods are of great practical significance to modern industrial systems due to their clear elaborations of the cause and effect relationship. Based on Boolean logic, logical analysis of data (LAD) can discover discriminative if-then rules and use them to diagnose faults. However, traditional LAD algorithm has a defect of time-consuming computation and extracts only the least number of rules, which is not applicable for high-dimensional large data set and for fault that has more than one independent causes. In this paper, a novel fast LAD with multiple rules discovery ability is proposed. The fast data binarization step reduces the dimensionality of the input Boolean vector and the multiple independent rules are searched using modified mixed integer linear programming (MILP). A Case Study on Tennessee Eastman Process (TEP) reveals the superior performance of the proposed method in reducing computation time, extracting more rules and improving classification accuracy.

KEYWORDS: fault diagnosis, knowledge extraction, fast logical analysis of data, multiple rules discovery

1   Introduction

Fault detection and diagnosis (FDD) is the key technique to guarantee production safety and reduce costs, thus it has always been a research focus 1-3. Various methods with both low fault alarm rate (FAR) and high fault detection rate (FDR) were carefully designed and applied to practical industrial systems in the past few decades 4-6. FDD methods are roughly divided into three categories: model-based, expertise-based and data-based 7. Since the real industrial systems have become larger and more complex, it is almost impossible to build precise mathematical models. Meanwhile, the expert knowledge of potential faults are difficult to obtain. Therefore, the data-based FDD methods are wildly researched due to their low dependences to the structure of system and the mechanism of fault 8.

In this paper, we focus on the data-based FDD classification method. With predefined classes (include a normal class and several faulty classes), various discriminant models can be trained successfully using data-based machine learning algorithms such as support vector machine (SVM) 9, artificial neuron networks (ANN) 10, Bayesian networks 11 and fuzzy logic 12 etc. Recently, the advancement of deep learning has proven its superiority in feature extraction and classification. Researchers implemented fault diagnosis with outstanding performance using deep learning algorithms such as deep belief networks (DBN) 13, stacked autoencoder (SAE) 14, recurrent neural networks (RNN) 15 and convolutional neural networks (CNN) 16.

Although the above methods attain high classification accuracy and realize effective fault diagnosis, they all suffered from a serious problem, namely, they build complicated discriminant models and extract incomprehensible patterns, which are like black boxes to users. All connections and transformations between the input data and the final results are encapsulated. The complexity of industrial systems make it impossible for users to understand the cause and effect relationship among various features and the corresponding diagnosis. Therefore, method that derived explanatory patterns are necessary and meaningful to convince people about its effectiveness, especially for practical engineering applications, where safety and validity are the prime principles.

Due to the above reasons, logical analysis of data (LAD) was proposed to reveal the causal relationship and implement fault diagnosis 17. LAD can discover explanatory patterns based on Boolean logic and detect faults by transforming these patterns into if-then rules and developing a synthetical discriminant function. However, the traditional LAD algorithm is not perfect. On the one hand, it discovers patterns with the least degree, which equals to the number of rules accordingly. Therefore, only the most discriminative rule will be found. Owing to the fact that faults are usually caused by several independent problems, multiple independent rules should be extracted at the same time. Obviously, the traditional LAD does not own that ability. On the other hand, the data binarization step (details are illustrated in section II) of traditional LAD forces the rest step to discover patterns among a large set of high dimensional data, which is time-consuming and unnecessary. To solve these two drawbacks, a novel fast LAD with multiple rules discovery ability based on modified mixed integer linear programming (MILP) is proposed and analyzed in this paper. A case study on Tennessee Eastman Process (TEP) indicates that the proposed new LAD algorithm can discover multiple reliable discriminant rules in a comparatively shorter time and realize fault classification with high accuracy.

The rest of this paper is organized as follow: Section 2, a brief introduction of LAD. Section 3, details about the proposed fast LAD with multiple rules discovery ability. Section 4, a case study on TEP. Section 5, conclusion.

2   Logical analysis of data

Logical analysis of data (LAD) is pattern discovery and classification algorithm. It finds the optimal feature combinations to achieve binary classification based on the Boolean representation of the original data. The combination is referred as pattern, which can be transformed to if-then rules. Therefore, LAD is explanatory and can extract explicit

knowledge directly. This is of practical significance in the field of FDD because the if-then rules are easy to understood and accepted by both the administrator and the front-line worker in the industrial sites.

LAD contains three main steps: data binarization, pattern discovery and discriminant function formation 17. Data binarization transforms the original data to Boolean representations without any pre-processing or standardization operations. Pattern discovery aims to find the optimal binary variable combination that distinguish samples between two categories. Discriminant function is formed to implement binary classification.

Let S^N and S^F denote the normal and faulty data set. Through data binarization, Sample $i \in S^N$ is represented as a binary vector $a_i = \{a_{i,1}, a_{i,2}, ..., a_{i,n}\}$ with size n. We expand a_i to $a_i = \{a_{i,1}, a_{i,2}, ..., a_{i,n}, a_{i,n+1}, a_{i,n+2}, ..., a_{i,2n}\}$ with size $2n$, where $a_{i,n+j}$ is the negation of $a_{i,j}$, $j = 1, 2, ..., n$. A pattern p_l can also be represented as a binary vector $p_l^N = \{p_{l,1}^N, p_{l,2}^N, ..., p_{l,2n}^N\}$, which contains d binary variables with the value of 1, namely,

$$\sum_{k=1}^{2n} p_{l,k}^N = d \tag{1}$$

$p_{l,k}^N (k = 1, 2, ..., 2n)$ equals to 1 means $a_{i,k}$ is selected, then the real value of dimension to which $a_{i,k}$ corresponds is larger (smaller) than the cut point if k is smaller (larger) than n. Therefore, $a_{i,j}$ and its negation cannot be selected at the same time, namely,

$$p_{l,j}^N + p_{l,n+j}^N \leq 1 \tag{2}$$

To make p_l^N rejects all $i \in S^F$ and accepts as many $i \in S^N$ as possible, two constraints are introduced:

$$\sum_{k=1}^{2n} a_{i,k} p_{l,k}^N \leq d - 1, i \in S^F \tag{3}$$

$$\sum_{k=1}^{2n} a_{i,k} p_{l,k}^N + n y_i \geq d, i \in S^N \tag{4}$$

where y_i is the optimization variable that needs to be minimized. Constraints (3) forces all faulty samples not to fulfil p_l^N with at least 1 degree. Constraints (4) tries to make p_l^N cover more normal samples.

$$\min_{p_l^N, y_i, d} \sum y_i, i \in S^N$$

$$s.t. \begin{cases} \text{eq.}(1),(2),(3),(4) \\ 1 \le d \le n \\ p_{l,k}^N \in \{0,1\}, k=1,2,...,2n \\ y_i \in \{0,1\} \end{cases} \tag{5}$$

The MILP method for pattern discovery is summarized in (5).

If p_l^N cannot cover all samples, then another pattern p_{l+1}^N needs to be found by deleting all covered samples and run MILP again. All patterns form a pattern set P^N.

To obtain a pattern p_m^F that accepts $i \in S^F$ and rejects all $i \in S^N$, simply change their positions in (5). Likewise, these pattern form a pattern set P^F.

The discriminant function is formulated in the basis of both P^N and P^F. For a new extended binary vector a, the discriminant function is given as follow:

$$D(a) = \sum w_l \, \text{sgn}(p_l^N a^T - d) - \sum w_m \, \text{sgn}(p_m^F a^T - d) \tag{6}$$

where $p_l^N \in P^N$ and $p_m^F \in P^F$. $\text{sgn}(a)$ is a sign function, it equals to 1 if $a > 0$ and 0 if $a \le 0$.

3 Fast LAD with multiple rules discovery ability

Although MILP method is a useful and effective pattern discovery tool, it still has several serious defects. Let us consider the following three typical situations:

- More than one binary variables from different dimensions are needed jointly to discriminate normal and faulty data.
- Both binary variable and its negation from same dimension are needed to discriminate normal and faulty data.
- More than one binary variables from different dimensions can discriminate normal and faulty data independently.

Since MILP searches the optimal variable combination to cover as many samples as possible, the first two situations are easy to handle. However, the independent causes of fault in the third situation will never be found completely using MILP because the algorithm stops when a pattern with one critical degree is found and then focuses on the rest, unaccepted samples. The reason is obvious: this pattern can cover all samples of its kind and reject all others, thus MILP will not add more degree into it.

Normal	1	1	1	1	0	0	0	0
	1	1	1	1	0	0	0	0
Faulty	1	0	1	0	0	1	0	1
	0	0	0	0	1	1	1	1

To illustrate this problem, a simple example is shown above. Figure 1 contains two normal binary samples and two faulty samples. A pattern that contains the red or green variables can distinguish them independently. Unfortunately, MILP can only find one of them randomly and stop instantly. Therefore, we need a new MILP that can discover a pattern with all critical degrees (red and green) here, which can then be transformed into multiple independent discriminant rules.

Dealing with this situation will be very meaningful, it reveals more characteristics of the fault and reduces the FAR of the original discriminate function because more rules can determine the type of fault more accurately.

To solve this problem, a novel modified MILP is proposed in this paper to realize multiple rules discovery.

3.1 Multiple rules discovery based on modified MILP

To find identical functional degrees, we could find all potential degrees firstly, then remove those useless degrees. Mortada 19 proposed an upgraded MILP with additional item in the object function, namely, $\min \sum y_i + \delta d, i \in S^N$, where $\delta > 0$ is selected to reduce the degree of pattern, making them more explanatory. In this paper, we do the contrary. We set $\delta < 0$, then MILP will find a pattern with both high degree and high coverage of samples. The importance of finding more degrees depends on δ. Unfortunately, as long as one critical degree is found, eq. (3) will be fulfilled immediately. Degree could be increased by finding the smallest cut point of any dimension (except the dimension where the critical degree derived from) because the left part of eq. (4) will naturally raise 1, therefore, this operation will bring some useless discriminate rules in.

To fix this drawback, eq. (3) needs to be revised. Inspired by the idea that maximize the coverage of pattern in eq. (4), we propose the following new constraint to replace eq. (3):

$$\sum_{k=1}^{2n} a_{i,k} p_{l,k}^N + z_i \leq d - 1, i \in S^F \tag{7}$$

and the object function are revised as:

$$\min_{p_l^N, y_i, d, z_{i'}} \sum y_i + \delta d + \gamma \sum z_{i'}, i \in S^N, i' \in S^F \tag{8}$$

where $z_{i'}$ is optimization integer variable that increases the distinction between the faulty data and the normal pattern, $\gamma < 0$ is the weight coefficient. All $z_{i'}$ have the same range of value with $d - 1$. With this additional optimization variable, eq. (7) will not be fulfilled by one critical degree. The more critical degrees are found, the better a pattern can meet the constraint and the optimization requirement at the same time.

The hyper parameter δ and γ have opposite effects and thus should be selected carefully to reach a balance. In the ideal situation, all critical degrees with similar function are aggregated in one pattern and the useless degrees are removed. This pattern can then be transformed into multiple independent discriminant rules simply by reversing the binarization step.

It seems that the above modified MILP can achieve multiple rules discovery easily. Nevertheless, it is usually difficult to obtain an ideal result using MILP solver due to the proper value of hyper parameters are uncertain and those useless degrees are hard to eliminate. To deal with this problem, we restrict the upper limit of d and $z_{i'}$ by selecting the number of degree manually. The modified MILP is showed below:

$$\min_{p_l^N, y_i, d, z_{i'}} \sum y_i + \delta d + \gamma \sum z_{i'}, i \in S^N, i' \in S^F$$

$$s.t. \begin{cases} eq.(1),(2),(7),(4) \\ \quad 1 \le d \le ub \\ p_{l,k}^N \in \{0,1\}, k=1,2,...,2n \\ \quad y_i \in \{0,1\} \\ \quad 0 \le z_{i'} \le ub-1 \end{cases} \tag{9}$$

where ub is the upper bound. How to select an appropriate ub is discussed in the following part.

3.2   Fast LAD based on fast data binarization

The data binarization step of LAD results in large cut point set and long binary vector, making the pattern discovery step difficult and time-assuming. However, most of the cut points are useless. If those cut points are found in advance, they can be removed to reduce the computation load and realize fast data binarization.
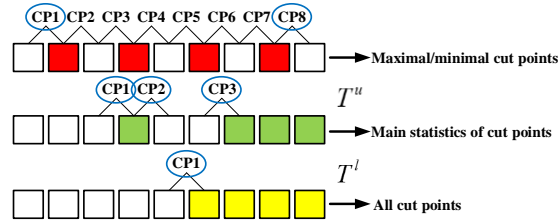

Fig. 2. The diagram of fast data binarization

The diagram of fast binarization is shown in figure 2. Assume that all collected variables follow Gaussian distribution, we estimate the mean and variance of Gaussian distribution in each dimension of the normal data. The proportion of faulty data covered between the range of positive and negative three sigma is calculated. According to the proportion, we divide the faulty data into three parts by setting an upper threshold T^u and a lower threshold T^l. Meanwhile, to make eq. (9) work, the determination of ub is necessary. This can be done according to the prior knowledge of the fault. However

in this paper, under Gaussian distribution assumption, *ub* can be determined by analysing the similarity of distribution between normal and faulty data. The fast data binarization is implemented in the basis of the following three rules:

- If the proportion is larger than T^u, this dimension is considered not discriminative, only the maximal and minimal cut points are remained.
- If the proportion is between T^l and T^u, this dimension is probably discriminative, the main statistics : mean, mode, median, maximal and minimal cut points are remained for potential usage.
- If the proportion is lower than T^l, this dimension is considered discriminative, all cut points are remained and *ub* is equal to the total number of dimension in this type.

4   A Case Study on Tennessee Eastman Process

The Tennessee Eastman Process (TEP) is a widely used chemical process simulation. TEP contains 8 components (A, B, C, D, E, F ,G and H) and 5 units (stripper, separator, reactor, compressor and condenser). Among them, A, C, D and E are the reactants; G and H are the target products; B is the non-reactive inertial component and F is the by-product. 11 manipulated variables (XMV (1-11)) and 41 measured variables (XMEAS (1-41)) are acquired from 21 designed faulty conditions (IDV (1-21)) and 1 normal condition (IDV (0)) at an interval of 3 minutes. Each condition contains 480 training sampled and 960 testing samples. In this paper, XMEAS(1-22) and XMV(1-11) are selected as input variables.

The thresholds T^l and T^u are set to 0.1 and 0.9 according to a rule of thumb. The parameters δ is set to -0.001 and γ is set to -1.

Due to the feedback mechanism of the system, the discriminative rules change as time goes by. Therefore in this paper, data in the final stable phase are used to train and test fast LAD. The training data set contains the last 100 normal data and the last 80 faulty data. The testing data set contains the last 300 normal and faulty data. Experiments were implemented on an Intel Core i7-6700 PC with 16G RAM and 3.4GHz in Matlab environment.

IDV (1, 2, 4, 7 ,17) are selected for a comparative analysis. To show the validity of modified MILP, discriminant rules discovered by both modified MILP and MILP are listed in table 1. From the table, it is obvious that the modified MILP discovered patterns with more degrees than MILP in IDV (1) and IDV (2) and thus derived more discriminant rules. To prove the correctness of the results, we analysed the data distribution of IDV (0) and IDV (1, 2, 4, 7 ,17). For IDV (1), XMEAS (1, 4, 18, and 19) and XMV (6, 9) have apparent deviations from the corresponding range of distribution of the normal data. For IDV (2), XMEAS (10, 11, 18, 19, and 22) and XMV (6, 9) are the variables that deviate from the normal state. All these deviations are captured by the modified MILP and transformed into independent discriminant rules whereas the general MILP found only one feasible rule for each fault. If another fault fulfills the rule, they might be misdiagnosed. Therefore, the modified MILP technique will reduce the misdiagnosis rate effectively. The fault classification can be implemented using any combination of these rules.

Fault No.	Modified MILP		MILP
IDV(1)	Rule group 1	XMEAS(1)>0.50 XMEAS(4)<9.07 XMEAS(18)>66.73 XMEAS(19)>246.66 XMV(3)>48.87 XMV(9)>52.23	XMV(9)>52.23
IDV(2)	Rule group 1	XMEAS(10)>0.51 XMEAS(11)>88.88 XMEAS(18)<64.45 XMEAS(19)<184.93 XMEAS(22)>78.27 XMV(6)>60.35 XMV(9)<37.00	XMV(22)>78.28
IDV(4)	Rule group 1	XMV(10)>43.01	XMV(10)>43.01
IDV(7)	Rule group 1	XMV(4)>69.08	XMV(4)>69.08
IDV(17)	Rule group 1	XMEAS(21)<94.33	XMEAS(21)<94.38 XMV(2)<54.73
	Rule group 2	XMEAS(9)<120.35	XMEAS(3)>4498.10 XMEAS(19)<235.99

Table 1. Fault discriminant rules

Fault No.	Fast LAD based on modified MILP		Traditional LAD based on MILP	
	FAR (%)	**FDR (%)**	**FAR (%)**	**FDR (%)**
IDV(1)	**0**	**94.33**	0	96
IDV(2)	**0**	**97.33**	0	99.67
IDV(4)	**0.33**	**100**	0.33	100
IDV(7)	**0**	**100**	0	100
IDV(17)	**0.33**	**98.33**	0.33	89.67

Table 2. FAR and FDR for different faults

For IDV (4) and IDV (7), two algorithms obtain identical results because these two faults have only one deviated variable, namely, XMV (10) and XMV (4). For IDV (17), MILP obtains more discriminant rules. However, only XMEAS (9,21) and XMV (10) show deviations from normal state. Among all we can learn that rule related to XMEAS (21) is the main functional rule. Other rules are not much useful. The modified MILP obtains effective rules because constraints eq. (7) increases the distinction between the faulty data and the normal pattern, therefore removes rules that are not discriminative.

The FAR and FDR of proposed fast LAD based on modified MILP and the traditional LAD based on MILP are listed in table 2. The former provides equally satisfied diagnosis results with the latter and better results for IDV (17).

The fast binarization step removes a large amount of cut points to reduce the dimensionality of the binary vector and further improves computation speed. The following comparative analysis proves its validity.
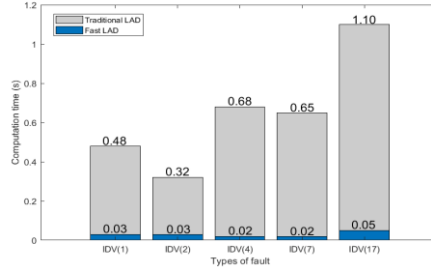
Fig. 3. The computation time comparison between traditional and fast LAD

In figure 3, apparently, the fast LAD takes much less time to obtain the final result. This experiment is run on a small data set, so the original computation time is acceptable. For a large dataset, the traditional LAD will spend very long time searching rules among massive high dimensional binary vectors while the fast LAD can remove useless dimension ahead. Therefore, it can reduce computation time effectively.

5   Conclusions

As a data-based machine learning algorithm, logical analysis of data can discover explanatory patterns and reveal the logic relations between data and results, which is very meaningful in fault diagnosis field. In this paper, we improve the traditional LAD algorithm and propose a fast LAD with multiple rules discovery ability. Based on modified MILP and fast binarization, this new algorithm can extract multiple independent fault rules within much shorter computation time. With a case study on TEP, we show the advantages of fast LAD in both fault classification and knowledge extraction.

References

1. Yin S, Ding S X, Xie X, et al. A Review on Basic Data-Driven Approaches for Industrial Process Monitoring[J]. IEEE Transactions on Industrial Electronics, 2014, 61(11):6418-6428.
2. Severson K, Chaiwatanodom P, Braatz R D. Perspectives on process monitoring of industrial systems, [J]. Annual Reviews in Control, 2016, 48(21):931-939.
3. Dai X, Gao Z. From Model, Signal to Knowledge: A Data-Driven Perspective of Fault Detection and Diagnosis[J]. IEEE Transactions on Industrial Informatics, 2013, 9(4):2226-2238.
4. Peng K, Zhang K, You B, et al. Quality-related prediction and monitoring of multi-mode processes using multiple PLS with application to an industrial hot strip mill[J]. Neurocomputing, 2015, 168(C):1094-1103.

5. Torabi A J, Meng J E, Xiang L, et al. Application of Clustering Methods for Online Tool Condition Monitoring and Fault Diagnosis in High-Speed Milling Processes[J]. IEEE Systems Journal, 2017, 10(2):721-732.
6. Zhang Y, Zhou H, Qin S J, et al. Decentralized Fault Diagnosis of Large-Scale Processes Using Multiblock Kernel Partial Least Squares[J]. IEEE Transactions on Industrial Informatics, 2010, 6(1):3-10.
7. Wen C L, Fei-Ya L V, Bao Z J, et al. A Review of Data Driven-based Incipient Fault Diagnosis[J]. Acta Automatica Sinica, 2016, 42(9):1285-1299.
8. Yin S, Li X, Gao H, et al. Data-Based Techniques Focused on Modern Industry: An Overview[J]. IEEE Transactions on Industrial Electronics, 2015, 62(1):657-667.
9. Yin Z, Hou J. Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes[J]. Neurocomputing, 2016, 174(PB):643-650.
10. Shatnawi Y, Al-Khassaweneh M. Fault Diagnosis in Internal Combustion Engines Using Extension Neural Network[J]. IEEE Transactions on Industrial Electronics, 2013, 61(3):1434-1443.
11. Cai B, Huang L, Xie M. Bayesian Networks in Fault Diagnosis[J]. IEEE Transactions on Industrial Informatics, 2017, PP(99):1-1.
12. Khan S A, Equbal M D, Islam T. A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models[J]. IEEE Transactions on Dielectrics & Electrical Insulation, 2015, 22(1):590-596.
13. Shao H, Jiang H, Zhang X, et al. Rolling bearing fault diagnosis using an optimization deep belief network[J]. Measurement Science & Technology, 2015, 26(11).
14. Sun W, Shao S, Zhao R, et al. A sparse auto-encoder-based deep neural network approach for induction motor faults classification[J]. Measurement, 2016, 89:171-178.
15. Bruin T D, Verbert K, Babuška R. Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, PP(99):1-11.
16. Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis[J]. Measurement, 2016, 93:490-502.
17. Ragab A, El-Koujok M, Poulin B, et al. Fault Diagnosis in Industrial Chemical Processes Using Interpretable Patterns Based on Logical Analysis of Data[J]. Expert Systems with Applications, 2017.
18. Ryoo H S, Jang I Y. MILP approach to pattern generation in logical analysis of data ☆[J]. Discrete Applied Mathematics, 2009, 157(4):749-761.
19. Mortada M A, Yacout S, Lakis A. Fault diagnosis in power transformers using multi-class logical analysis of data[J]. Journal of Intelligent Manufacturing, 2014, 25(6):1429-1439.