

Article

Feature selection for chemical process fault diagnosis by artificial immune systems☆

Liang Ming, Jinsong Zhao *

Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China



ARTICLE INFO

Article history:

Received 8 September 2017

Accepted 13 September 2017

Available online 14 October 2017

Keywords:

Artificial immune system

Genetic algorithm

Feature selection

ABSTRACT

With the Industry 4.0 era coming, modern chemical plants will be gradually transformed into smart factories, which sets higher requirements for fault detection and diagnosis (FDD) to enhance operation safety intelligence. In a typical chemical process, there are hundreds of process variables. Feature selection is a key to the efficiency and effectiveness of FDD. Even though artificial immune system has advantages in adaptation and independency on a large number of fault samples, antibody library construction used to be based on experience. It is not only time consuming, but also lack of scientific foundation in fault feature selection, which may deteriorate the FDD performance of the AIS. In this paper, a fault antibody feature selection optimization (FAFSO) algorithm is proposed based on genetic algorithm to optimize the fault antibody features and the antibody libraries' thresholds simultaneously. The performance of the proposed FAFSO algorithms is illustrated through the Tennessee Eastman benchmark problem.

© 2017 The Chemical Industry and Engineering Society of China, and Chemical Industry Press. All rights reserved.

1. Introduction

Industry 4.0, which is regarded as the fourth industrial revolution, aims at transforming today's factories into smart factories. The purpose of smart factories is to address and overcome the current challenges of shorter product lifecycles, highly customized products, enhanced work safety and stiff global competition [1]. For the chemical process industry, smart factories are expected not only to maximize the economic value of factories, but also to make the best management of processes to avoid safety incidents [2]. With the boom of science and technology, modern chemical plants are developing towards the direction of large-scale, complication and integration, which therefore increases the chance of mishaps and faults. Due to the flammable, explosive, toxic, and corrosive nature of chemical processes and the strong coupling between different parts of a production system, one partial failure may trigger the abnormality of the entire system through a chain of reactions, resulting in tremendous economic, social and/or environmental losses. Since chemical processes are generally of high risks especially when an independent protection layer fails, operation safety intelligence should be a key element of smart factories [3].

Although a number of automatic control systems, such as DCS and PLC, have been widely employed in chemical processes, abnormal situation management (ASM) still faces many challenges. Efficient ASM is

intended to detect and diagnose abnormal situations in the early period, and provide timely, reliable and automated decision supports to operators, so that they can take proper actions to prevent disturbances from turning into big incidents [3]. It is obvious that chemical process fault detection and diagnosis (FDD) plays an important role in ASM. Over the past 20 years, FDD approaches are drawing more and more attention from academia and industry, and the number of publications on FDD approaches is gradually increasing as shown in Fig. 1.

Though the research about fault diagnosis methods has a history of nearly 40 years, the current practical applications are still concentrated in mechanical faults and occasionally instrument faults. When it comes to online fault diagnosis systems for the whole chemical process, there are few related reports or available business products, and the main reasons can be concluded as follows: 1) lack of self-adaptive ability to varying process conditions because of equipment performance degradation, feed fluctuations, operating habits or something else; 2) lack of self-learning ability to faults that have not been specially modeled; and 3) dependent on a large number of fault data samples.

To overcome the disadvantages and limitations of the current fault diagnosis methods, many scholars are casting their eyes to the emerging artificial intelligence technology. The authors' research team has been studying artificial immune systems (AIS) during the past ten years, a few AIS-based strategies have been proposed [4,5].

Due to the increasing scale of modern chemical processes and wide application of advanced control systems, the number of measured variables is significantly large. For example, a fluid catalytic cracking unit (FCCU) usually contains over one thousand measured variables. If irrelevant and/or redundant variables are selected for constructing antibody

☆ Supported by the National Natural Science Foundation of China (61433001).

* Corresponding author at: Department of Chemical Engineering, Tsinghua University, Beijing 100084, China.

E-mail address: jinsongzhao@tsinghua.edu.cn (J. Zhao).

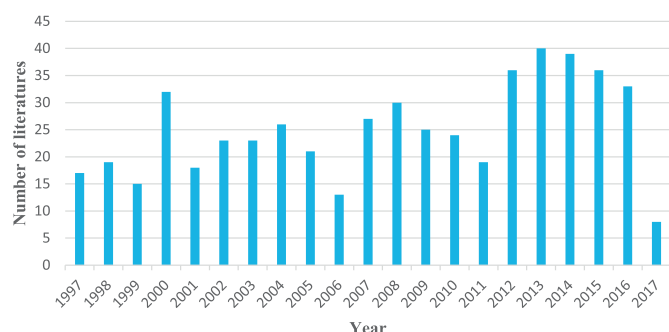


Fig. 1. Literature number on chemical process fault diagnosis from selected journals during the past 20 years.

libraries, it not only increases AIS's complexity and computing time, but also may result in the curse of dimensionality, degrading the diagnostic accuracy of AIS [6].

Feature extraction and feature selection are two main techniques to reduce dimensionality and improve FDD performance. Feature extraction techniques create a set of new features preserving relevant information by combining original features. Principal component analysis (PCA), partial least-squares regression (PLS), and linear discriminant analysis (LDA) are commonly used feature extraction techniques for process FDD. However, the new features usually lack explanation, which cannot reflect the typical variation trends of the relevant variables during a fault. Different from feature extraction, feature selection techniques retain the original meaning of the selected features. Feature selection tends to select an optimal subset of the relevant features with high separability power by removing irrelevant and redundant features. To improve the description of AIS, feature selection will be utilized to optimize AIS in this paper.

Feature selection approaches are generally categorized into 2 main types based on the evaluation methods used to test the performance of the selected feature subsets by some criteria: 1) filter and 2) wrapper [6]. Wrapper evaluates the performance using the learning algorithm during the classifier design, while filter evaluates that independently. 1) Filter has higher computational efficiency than wrapper since filter can test the performance of selected features quickly by some suitable criteria. Except for specially designed criteria, such as [7–9], which are intended to reduce the correlation or mutual information between features, many simple information statistics criteria are commonly used [10,11]. However, the optimal subset selected by filter might be very large, especially, when the feature is dependent on the classifier. Filter can remove a large number of unnecessary noisy features quickly, narrowing the search of optimal subset, therefore, it is a good feature pre-selector. 2) Wrapper uses the selected subset of features to train the classifier directly, and evaluates the performance by estimating the classification accuracy on validation data. Though wrapper is slower than filter, the subset selected is smaller, advantageous to identifying key features and simplifying FDD models. Considering the rapid development of computers and that feature selection is a preprocessing step in the offline phase, wrapper has been the hot research direction in the field of feature selection (Fig. 2). In this paper, we propose a wrapper-based approach.

If there are n features to be selected, the size of the search space is up to $O(2^n)$, which makes it burdensome to search the entire search space when n is large. For example, Tennessee Eastman (TE) process has 53 measured features, the number of possible combinations of features is 4.5×10^{15} . So search strategy is another important factor for feature selection approaches. Optimal, heuristic and randomized are three main search strategies. A number of feature selection approaches with different search strategies have been proposed, such as branch and bound [12], genetic algorithm [13], particle swarm optimization [14] and SVM-RFE [15].

Although feature selection is widely used in automated pattern recognition [16,17], its applications in multivariate statistical

process monitoring or FDD are rare. Two notable exceptions are Verron *et al.* [18] and Ghosh *et al.* [19]. AIS has advantages in adaptation and low requirement for historical samples, however, antibody library construction, including feature selection and threshold identification, used to heavily depend on experience, which is time consuming and lacks scientific foundation. In this paper, a fault antibody feature selection optimization (FAFSO) algorithm is proposed based on genetic algorithm to optimize the fault antibody features and the antibody libraries' thresholds simultaneously.

2. AIS

Human immune system has a strong ability to protect human bodies from pathogen. Innate immunity and adaptive immunity are two types of immunity [20]. AIS is developed with adaptation to the principle of adaptive immunity, bridging the gap between immunology and engineering. Clonal selection algorithm is one of the main immune algorithms. In a clonal immune algorithm, once the immune cells recognize the antigen, they are treated as parent cells and more antibodies are then generated as clones of their parents [20]. In the cloning phase, mutation is used to generate new antibodies, which can recognize the antigen triggering an immune response.

2.1. Description of antibodies and antigens

Antibodies and antigens are created by historical samples and real-time samples and the system diagnoses faults by calculating and analyzing the affinity between antibodies and antigens. To reflect the dynamic trends of process, antibodies and antigens are represented by matrices of time-sampled data [21]. Eqs. (1) and (2) show the composition of antibodies and antigens.

$$Ab_{\text{fault}}^{m \times n} = [Ab_{\text{fault}}(1); Ab_{\text{fault}}(2); \dots; Ab_{\text{fault}}(k); \dots; Ab_{\text{fault}}(m)] \quad (1)$$

$Ab_{\text{fault}}(k)$ represents the vector of selected variables at the k th sampling time after the fault encountered. N is the number of selected variables and m is the length of the antibody.

$$Ag^{l \times n} = [Ag(1); Ag(2); \dots; Ag(k); \dots; Ag(l)] \quad (2)$$

$Ag(k)$ represents the vector of selected variables at the $(l - k)$ th sampling time before now. N is the number of selected variables and l is the length of the antigen.

In this paper, the affinity is represented by distance between antibody and antigen, calculated based on Euclidean distance with sliding window.

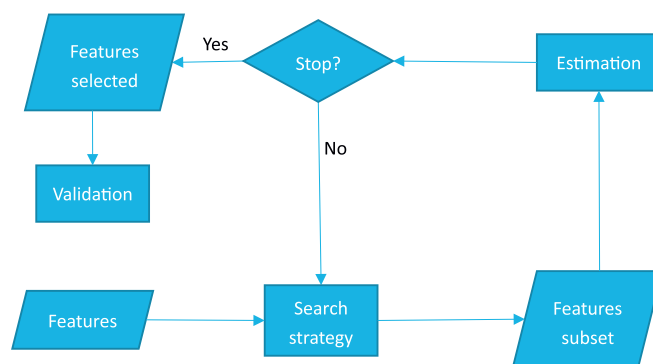


Fig. 2. Flowsheet of feature selection.

Original antibodies are generated directly from historical samples. The historical samples are first normalized to eliminate the amplitude differences between variables, shown in Eq. (3):

$$X = \frac{x - X_{\min}}{X_{\max} - X_{\min}} \quad (3)$$

where x is the original data, X_{\min} is the minimum of the historical data, X_{\max} is the maximum of the historical data and X is the normalized data.

2.2. Antibody library construction

After the original antibodies are created, mutation is employed to generate new antibodies during the clone phase.

Set two randomly selected original antibodies as A_g s and a normal historical sample as A_b , and then calculate the difference matrix between A_g and A_b respectively, which are represented as φ_1 and φ_2 . A new difference matrix φ^* can be obtained with mutation of φ_1 and φ_2 , shown in Eq. (4):

$$\varphi^* = a\varphi_1 + b(\varphi_1 - \varphi_2) \quad (4)$$

where a is a random number between 0.5 and 2, and b is a random number between -1 and 1 .

Randomly select a matrix X_n of the same length with the original antibodies from the normal historical sample, a new antibody X^* is generated based on Eq. (5).

$$X^* = X_n + \varphi^* \quad (5)$$

A threshold is necessary for each antibody library as criteria to determine whether an antigen is in the same status represented by the antibody library [4]. If an antibody library has an antibody whose difference from the antigen is smaller than the antibody library threshold, then the antigen is recognized by the antibody library. That means that the process is in the status represented by the antibody library's type.

3. FAFSO (Fault Antibody Feature Selection Optimization) Algorithm

Since feature variables comprising the antibodies and antigens and threshold are two key factors affecting the diagnostic performance of an antibody library, a fault antibody feature selection optimization (FAFSO) algorithm based on genetic algorithm (GA) is proposed to identify the best subset of variables and the best threshold that yields the best possible diagnostic performance. In the proposed optimization scheme, given a labeled training dataset and validation dataset, determine a subset of variables and threshold that make the antibody library built from the training dataset have the best diagnostic performance on the validation dataset. The training dataset consists of data of one kind of fault, whereas the validation dataset consists of data of various kinds of faults. The diagnostic performance of an antibody library is measured in terms of cumulative fault diagnosis rate and false positive rate on the validation data.

In this paper, GA is introduced to optimize the subset of variables and threshold simultaneously. Fig. 3 shows the structure of the chromosome, which is a vector with $n + 1$ elements. n is the total number of variables and the first n elements encode the variables to be selected to comprise the antibodies in an antibody library. The first n elements are represented by binary values, where 1 means the presence and 0

means the absence of a variable in the selected subset. The last element encodes the threshold of an antibody library, and is represented by real values. For example, $[1, 0, 0, 0, 1, 1, 0.2]$ indicates that the first, fifth and sixth variables are selected to comprise the antibodies and the threshold is set to 0.2.

The fitness function of GA is composed of two computing objectives: fault diagnosis rate (FDR) and false positive rate (FPR). Given a chromosome, an antibody library is built with the antibodies generated by the AIS algorithm based on the selected subset of variables and the training dataset of one specific type of fault, and the set threshold. The diagnostic performance of the antibody library is then calculated using the validation dataset based on FDR and FPR. They are defined as below, where p is the number of samples of this fault that are diagnosed as the same fault and q is the number of samples of other faults that are also diagnosed as this kind of fault. The fitness function is the sum of FDR and FPR, and the objective is to maximize the cumulative diagnostic rate.

$$\text{FDR} = \frac{p}{\text{total number of samples of this fault}} \quad (6)$$

$$\text{FPR} = 1 - \frac{q}{\text{total number of samples of other faults}} \quad (7)$$

$$\text{Fitness} = \text{FDR} + \text{FPR} \quad (8)$$

In the GA process, three genetic operators, selection, crossover and mutation are used. Proportional selection is employed to select next generation chromosome. The simulated binary crossover and polynomial mutation are adopted to generate new chromosome.

Fig. 4 shows the flowsheet of the proposed optimization scheme of an antibody library. The first generation chromosomes are generated randomly and each chromosome indicates the selected subset of variables for the antibodies and the threshold of the antibody library. A fault antibody library is then developed for each chromosome based on normalized training data, containing data of that fault type, and the algorithms described in Sections 2.1 and 2.2. Normalized validation data, which contains data of all fault types, after variable selection by each chromosome, is used as antigens to test the diagnostic performance of the developed antibody library in terms of FDR and FPR. The sum of FDR and FPR is calculated as the fitness of each chromosome. Finally, chromosomes are subjected to the genetic operation, selection, crossover and mutation, and elitism is utilized to generate the chromosomes of the next generation. The optimization process will terminate once the maximum number of generation is reached.

4. Case Study: Tennessee Eastman Process (TEP)

TEP is a simulated benchmark widely used for evaluating process control and monitoring methods [22]. Bathelt *et al.* revised the model with more measured variables and more types of faults and the revised model is available at <http://depts.washington.edu/control/LARRY/TE/download.html> [23]. In this section, the proposed AIS-GA is applied to the revised TEP, as shown in Fig. 5.

TEP consists of 53 measured variables, 22 process measurements, 19 composition measurements and 12 manipulated variables. In the experiment, all variables are sampled at every 3 min and all variables are included in analysis. Since AIS diagnoses faults on the assumption that same variables have same variation trends during same types of faults,

V_1	V_2	V_n	threshold
1	0	0	1	1	0.2

Fig. 3. Structure of chromosome.

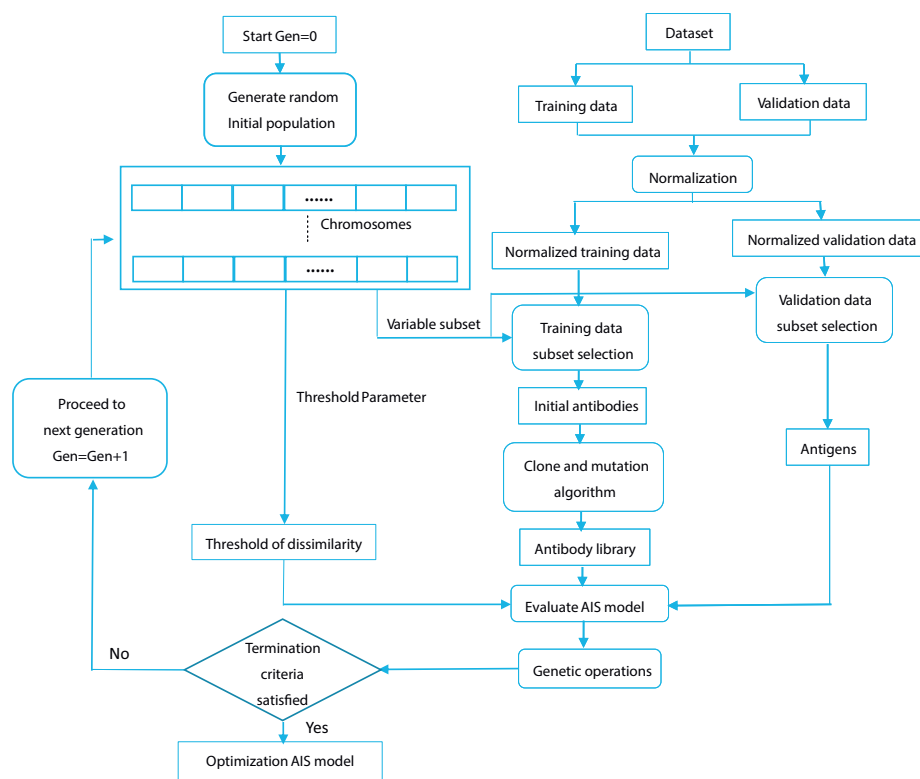


Fig. 4. Proposed GA based optimization scheme.

only faults of step type, faults 1–7, are considered in this experiment. Each type of fault is simulated for 12 times with different initial states, and the simulation time is set to 24 h. The training dataset of each fault antibody library consists of 4800 samples of that fault. The validation dataset has 480 samples of every fault. The training process tends to construct the fault antibody library based on AIS algorithm using the training dataset that has the best diagnostic performance on the validation dataset.

Once the optimal subset of variables and the threshold of every fault antibody library are determined, a modified AIS model is obtained.

During online diagnosis phase, a separate test dataset (containing 480 samples of every fault) is used to evaluate the diagnostic performance of the modified AIS model. As shown in Fig. 6, test data is first normalized and then converted into different antigens for different fault antibody libraries based on the selected subset of variables of each fault antibody library. The difference between an antigen and a fault antibody library is equal to the minimum distance between the antigen and the antibodies divided by the threshold of the fault antibody library. A test sample will be diagnosed as the fault type the minimum difference belongs to if the minimum difference is less than 1. Otherwise, it will

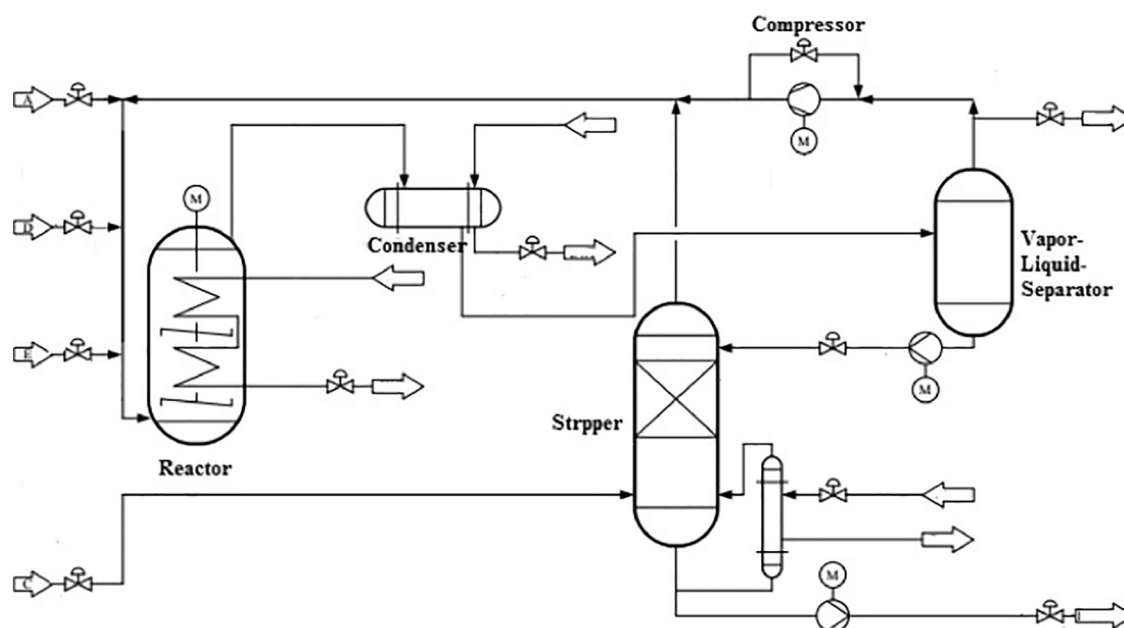


Fig. 5. P&ID of the revised process model.

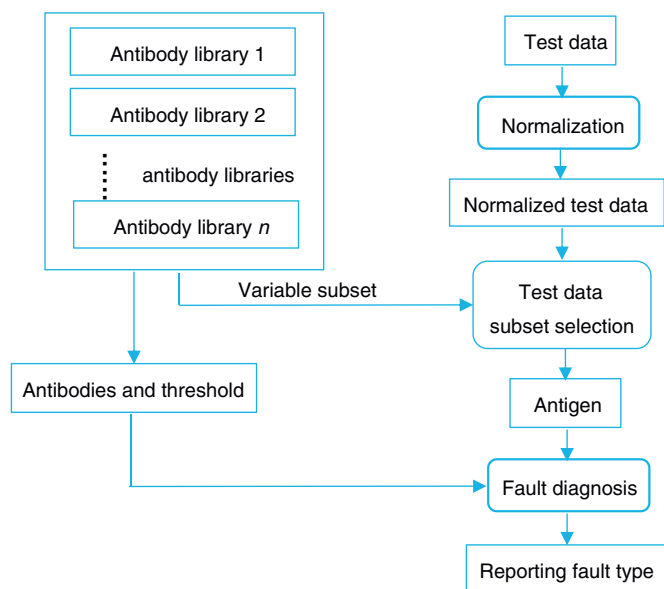


Fig. 6. Flowchart of online diagnosis process.

be diagnosed as known type. Fault diagnosis accuracy (FDA) is proposed to measure the diagnostic performance of the modified AIS model.

$$FDA = \frac{\text{number of correctly diagnosed samples}}{\text{total number of test samples}} \quad (9)$$

Table 1 lists the selected subset of variables for each antibody library and Table 2 compares the diagnostic performance of AIS and AIS-FAFSO. It reveals that the selected subsets of variables for different fault antibody libraries are different. Although the diagnostic accuracy for samples of faults 3 and 5 of AIS-FAFSO is worse than AIS, which is because the proposed scheme is meant to optimize the diagnostic performance of AIS globally, AIS-FAFSO outperforms AIS generally.

Table 1
Variable selection of different fault antibody libraries

No. of fault antibody library	No. of variables	Selected variables
Fault 1	27	XMEAS1–6, XMEAS10, XMEAS13–14, XMEAS17, XMEAS19, XMEAS21–23, XMEAS25–27, XMEAS29–30, XMEAS32, XMEAS34, XMEAS38, XMEAS41, XMEAS42, XMEAS44–45, XMEAS47, XMEAS49, XMEAS51
Fault 2	23	XMEAS2–3, XMEAS7–8, XMEAS10–11, XMEAS16, XMEAS19–20, XMEAS22–24, XMEAS26, XMEAS28, XMEAS35–37, XMEAS39, XMEAS41, XMEAS43, XMEAS45–46, XMEAS48, XMEAS50
Fault 3	23	XMEAS3, XMEAS7, XMEAS12, XMEAS14, XMEAS16, XMEAS19–21, XMEAS23, XMEAS25–27, XMEAS31, XMEAS33–34, XMEAS41, XMEAS43, XMEAS45, XMEAS47, XMEAS49, XMEAS51
Fault 4	29	XMEAS1, XMEAS4–5, XMEAS9–11, XMEAS13, XMEAS15, XMEAS17, XMEAS19–21, XMEAS23–25, XMEAS27–29, XMEAS31–33, XMEAS35–37, XMEAS39, XMEAS41, XMEAS43–45, XMEAS47–49, XMEAS51–53
Fault 5	28	XMEAS1–2, XMEAS5–6, XMEAS8, XMEAS10, XMEAS15–16, XMEAS18, XMEAS20–21, XMEAS23–24, XMEAS26–29, XMEAS31–32, XMEAS34, XMEAS36, XMEAS39–40, XMEAS42, XMEAS44, XMEAS46, XMEAS48, XMEAS50
Fault 6	31	XMEAS1, XMEAS3–7, XMEAS10, XMEAS13–16, XMEAS18, XMEAS21–27, XMEAS29, XMEAS31–34, XMEAS36–40, XMEAS42–44, XMEAS46–48, XMEAS50–52
Fault 7	26	XMEAS1–3, XMEAS5–6, XMEAS13, XMEAS15–16, XMEAS22–23, XMEAS26, XMEAS28–30, XMEAS33–34, XMEAS36–37, XMEAS39–41, XMEAS43–44, XMEAS46–47, XMEAS49–50

Table 2
Comparison of diagnostic performance of AIS and AIS-FAFSO on test data (%).

FDA	AIS	AIS-GA
Fault 1	99.78	100
Fault 2	95.87	100
Fault 3	100	93.70
Fault 4	84.16	100
Fault 5	100	90.23
Fault 6	0	100
Fault 7	100	100
Average	82.83	97.70

Table 3
Comparison of diagnostic performance of different FDD approaches (%).

FDA	a	b	c	d	e
Fault 1	100	99.87	96.37	100	100
Fault 2	99.38	97.87	97.62	99	100
Fault 3	10.25	2.37	20.62	6	93.70
Fault 4	100	100	82.75	100	100
Fault 5	34.75	99.87	96	100	90.23
Fault 6	100	99.5	100	100	100
Fault 7	100	100	100	100	100
Average	77.76	85.64	84.76	86.42	97.70

(a) Basic PCA [24]; (b) optimized variable selection based PCA [20]; (c) supervised local multilayer perceptron [25]; (d) Bayesian method; (e) AIS-FAFSO (proposed in this paper).

Table 3 shows the comparison of diagnostic performance of several FDD approaches for faults 1–7 [24–26]. Among these FDD approaches, AIS-FAFSO has a comparative diagnostic ability.

5. Conclusions

Variable selection and threshold identification of a fault antibody library in AIS used to heavily depend on experience, which is time consuming and lacks scientific foundation. In this paper, FAFSO is proposed based on genetic algorithm to optimize the subset of variables for the antibodies and the threshold of a fault antibody library automatically and simultaneously. In the experimental section, AIS-FAFSO is applied to improve the diagnostic performance of AIS globally on faults of step type in the Tennessee Eastman process. The diagnostic results of test data show that AIS-FAFSO outperforms AIS in average accuracy. However, the optimization process will be very time consuming if the training dataset and validation dataset are large, so some future work can be focused on improving computing efficiency.

References

- [1] S. Weyer, M. Schmitt, M. Ohmer, D. Gorecky, Towards Industry 4.0-standardization as the crucial challenge for highly modular, multi-vendor production systems, *IFAC-Pap. Online* 48-3 (2015) 579–584.
- [2] P.D. Christofides, J.F. Davis, N.H. El-Farra, D. Clark, K.R.D. Harris, J.N. Gips, Smart plant operations: vision, process and challenges, *AIChE J.* 53 (2007) 2734–2741.
- [3] Dai Yiyang, Hangzhou Wang, Khan Faisal, Zhao Jinsong, Abnormal situation management for smart chemical process operation, *Curr. Opin. Chem. Eng.* 14 (2016) 49–55.
- [4] Shu Yidan, Zhao Jinsong, Fault diagnosis of chemical processes using artificial immune system with vaccine transplant, *Ind. Eng. Chem. Res.* 55 (2016) 3360–3371.
- [5] Dai Yiyang, Zhao Jinsong, Fault diagnosis of batch chemical processes using a dynamic time warping (DTW)-based artificial immune system, *Ind. Eng. Chem. Res.* 50 (2011) 4534–4544.
- [6] K. Ghosh, M. Ramteke, R. Srinivasan, Optimal variable selection for effective statistical process monitoring, *Comput. Chem. Eng.* 60 (2014) 260–276.
- [7] Z.H. Sun, G. Bebis, R. Miller, Object detection using feature subset selection, *Pattern Recogn.* 37 (11) (2004) 2165–2167.
- [8] I. Inza, P. Larrañaga, R. Blanco, et al., Filter versus wrapper gene selection approaches in DNA microarray domains, *Artif. Intell. Med.* 31 (2) (2004) 91–103.
- [9] Zhou Xiaobo, Xiaodong Wang, E.R. Dougherty, Nonlinear probit gene classification using mutual information and wavelet based feature selection, *Biol. Syst.* 12 (3) (2004) 371–386.

- [10] T.S. Furey, N. Cristianini, N. Duffy, *et al.*, Support vector machine classification and validation of cancer tissue samples using microarray expression data, *Bioinformatics* 16 (10) (2000) 906–914.
- [11] I. Tabus, J. Astola, On the use of MDL principle in gene expression prediction, *EURASIP J. Appl. Signal Process.* 4 (2001) 297–303.
- [12] Liu Huiqing, Li Jinyan, L. Wong, A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns, *Genome Inform.* 13 (2002) 51–60.
- [13] Chen Xuewen, An improved branch and bound algorithm for feature selection, *Pattern Recogn. Lett.* 24 (12) (2003) 1925–1933.
- [14] L. Zhang, L.B. Jack, A.K. Nandi, Fault detection using genetic programming, *Mech. Syst. Signal Process.* 19 (2005) 271–289.
- [15] S.W. Lin, S.C. Chen, PSOLDA: a particle swarm optimization approach for enhancing classification accuracy rate of linear discrimination analysis, *Appl. Soft Comput.* 9 (2009) 1008–1015.
- [16] I. Guyon, J. Weston, S. Brnhill, *et al.*, Gene selection for cancer classification using support vector machine, *Mach. Learn.* 46 (1/2/3) (2002) 389–422.
- [17] A.K. Jain, D. Zongker, Feature selection: evaluation application, and small sample performance, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (2007) 153–158.
- [18] M.L. Raymer, W.F. Punch, E.D. Goodman, L.A. Kuhn, A.K. Jain, Dimensionality reduction using genetic algorithm, *IEEE Trans. Evol. Comput.* 4 (2) (2000) 164–171.
- [19] S. Verron, T. Tiplica, A. Kobi, Fault detection and identification with a new feature selection based on mutual information, *J. Process Control* 18 (2008) 479–490.
- [20] R.A. Goldsby, T.J. Kindt, B.A. Osborne, J. Kuby, Immunology, 5th edition W. H. Freeman and Company, San Francisco, CA, 2003.
- [21] L.N. de Castro, J.J. Timmis, Artificial immune systems: a novel paradigm to pattern recognition, in: J.M. Corchado, L. Alonso, C. Fyfe (Eds.), Artificial Neural Networks in Pattern Recognition, University of Paisley, U.K. 2002, pp. 67–84.
- [22] L.H. Chiang, E.L. Russell, R.D. Braatz, Fault Detection and Diagnosis in Industrial Systems, Springer-Verlag London Limited, Briton, 2001.
- [23] A. Bathelt, N.L. Ricker, M. Jelali, Revision of the Tennessee Eastman Process model, *IFAC (Int. Fed. Automatic Control) Pap. Online* 48 (8) (2015) 309–314.
- [24] S. Yin, S.X. Ding, A. Haghani, H.Y. Hao, P. Zhang, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process, *J. Process Control* 22 (2012) 1567–1581.
- [25] M.A.A. Rad, M.J. Yazdanpanah, Designing supervised local neural network classifiers based on EM clustering for fault diagnosis of Tennessee Eastman process, *Chemom. Intell. Lab. Syst.* 146 (2015) 149–157.
- [26] Q.C. Jiang, B. Huang, X.F. Yan, GMM and optimal principal components-based Bayesian method for multimode fault diagnosis, *Comput. Chem. Eng.* 84 (2016) 338–349.