Advances in Intelligent Systems Research, volume 147
International Conference on Network, Communication, Computer Engineering (NCCE 2018)

A Review of Industrial Fault Diagnosis Based on Data-driven Methods

Wenbo Liu[1, a)], Zhiping Peng[2, b)], Delong Cui[2, c)]

[1] *Faulty of Computer Science, Guangdong University of Technology, China*
[2]*Department of Computer and Electronic Information, Guangdong University of Petrochemical Technology, China*

[a)] 402299032@qq.com
[b)] 417863130@qq.com
[c)] delongcui@163.com

Abstract. The Modern industrial devices are becoming more and more complicated and the data generated by industrial processes are increasing. It has become increasingly difficult to establish accurate mathematical models for fault diagnosis. In view of this situation, data-driven fault diagnosis methods have attracted attention and rapid development. This article reviews various methods based on data-driven fault diagnosis for industrial processes, and finally makes a forecast.

Key words: Data-driven; Fault diagnosis; Industrial process.

INTRODUCTION

The ethylene industry is an important base for the petrochemical industry. Ethylene production is an important indicator of a country's chemical industry level. Due to the dual pursuit of ethylene production and quality, modern ethylene production equipment is also moving towards large-scale and complex development. It is difficult for this large-scale and complicated industrial process to establish accurate mathematical models based on physical and chemical mechanisms to monitor and diagnose faults. With the development and maturity of machine learning and data mining technology, fault diagnosis and analysis technology based on historical process data has opened up new ways for fault diagnosis of large-scale complex industrial processes. The data-driven fault diagnosis method does not need to consider the exact mathematical model, but only needs to train the model to complete the fault analysis and diagnosis from a large amount of historical or online data generated during the operation of the system, and this method has gradually become a hot topic in the field of fault diagnosis. This article selects a typical chemical process ethylene cracker to study the application object to illustrate the application of data-driven fault diagnosis methods in complex industrial processes.

DATA-DRIVEN FAULT DIAGNOSIS METHODS

Machine learning methods are generally divided into supervised learning, non-supervised learning, and semi-supervised learning. Data-driven fault detection and diagnosis methods can also be divided into supervised learning-based fault diagnosis, unsupervised learning-based fault diagnosis, and the fault diagnosis of semi-supervised learning is shown in Figure 1. Below we will explain each fault diagnosis method in the figure.

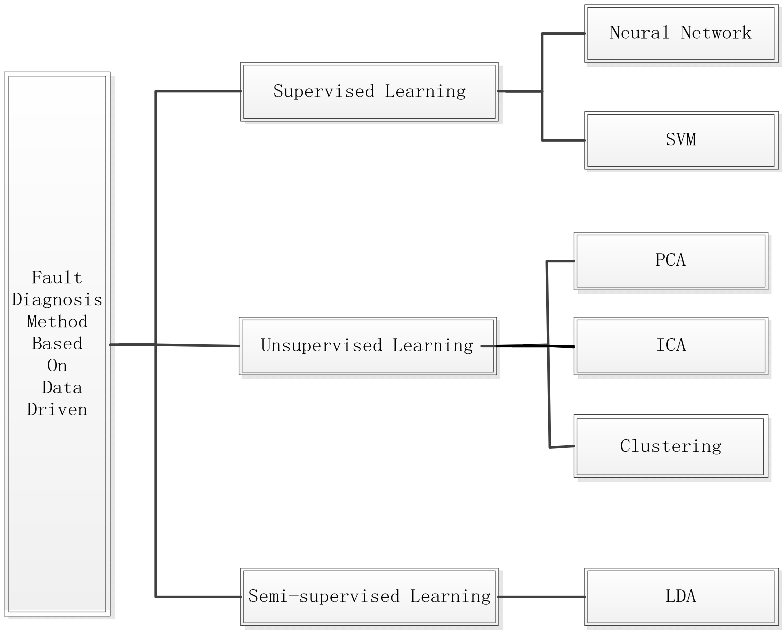

FIGURE 1. Classification chart of data-driven fault diagnosis method

Fault Diagnosis Method Based on Supervised Learning

Supervised learning methods in the field of machine learning utilize marker sample data for training system model classifiers. Neural networks and support vector machines are two important classes of supervised learning classifiers. From a certain perspective, fault diagnosis can be constructed as a pattern recognition problem whose goal is to be able to classify sample data well based on certain rules. Based on this, we will explain the fault diagnosis method based on neural network and the fault diagnosis method based on support vector machine in this part.

Fault Diagnosis Method Based on Neural Network

It is very suitable for application to the fault diagnosis system [1] because the neural network has the functions of processing complex multi-mode and associating, conjecture and memory. The fault diagnosis method based on neural network is based on training data to establish the mapping of fault recognition and classification. Then, the trained network is used for the new observed data, and the abnormal situation is judged by [2]. The learning process of neural network is to adjust the connection weight and the threshold of each functional neuron according to training data. In other words, the knowledge that [3] learns is contained in the connection weights and thresholds. So far, one of the most successful neural network learning algorithms is the well-known error inverse propagation algorithm (BP) [4]. There have been quite a lot of research on the problem of fault detection and diagnosis by using BP neural network algorithm.

In the field of chemical engineering, literature [5-9] describes the effectiveness of neural networks in the problem of fault diagnosis in detail. In document [10], a real-time online recursive learning algorithm based on functional connectivity neural network and improved BP network is presented. A soft fault diagnosis system for cracking furnace in large ethylene plant is developed based on a hybrid system consisting of neural network and expert system. The system can not only diagnose the failure mode or category of the cracking furnace, but also diagnose the "grade" of the fault at the same time.

Fault diagnosis method based on SVM

Neural network usually needs a large number of samples to train the model. Support vector machine (SVM) [11] overcomes the situation when the data sample is not enough, and it is a classifier suitable for small sample. In the case of fault diagnosis, document [12] gives the steps to establish the SVM fault diagnosis model.

The fault diagnosis method based on SVM has been applied to many aspects of the production of ethylene industry. Document [13] applies support vector machine classifier based on immune culture algorithm to fault diagnosis of industrial ethylene cracking furnace. Through field data test, classification accuracy reaches a high level, showing good application prospects. According to the problem of parameter selection in least squares support vector machine, a clonal selection algorithm is introduced in document [14]. A least squares support vector machine based on improved clonal selection algorithm is proposed. It is applied to the learning and prediction of the cracking depth of ethylene cracking furnace. The simulation shows that the method has the advantages of high precision of fast training and high precision.

Fault Diagnosis Method Based on Supervised Learning

In unsupervised learning, the marking information of training samples is unknown. The goal is to reveal the intrinsic properties and rules of data through learning from unlabeled training samples. Fault diagnosis method of unsupervised learning basic is also a continuation of the field of machine learning in unsupervised learning based on the first application of principal component analysis, independent component analysis method for feature extraction or dimensionality reduction of industrial process data, and then the analysis and diagnosis of the fault data using the method of multivariate statistics. In this part, we will expound the methods of principal component analysis (PCA), independent principal component analysis (ICA) and clustering in this kind of diagnostic method.

Fault Diagnosis Based on PCA

In theory, PCA technology is a matrix form of the high dimensional training data of the system and then a series of matrix operations. The PCA based fault diagnosis generally first passes the matrix operation to find the principal component model of the original training data sample. After finding the principal component model of data samples, we will then determine the statistics of the total data. The common statistics include SPE statistics [15] and Hotelling T2 statistics [16]. Finally, the control limit [17], which determines the operation of the process by the normal data, is to judge whether the data is normal or not by the control limit so as to complete the fault diagnosis.

Literature [18] uses TE process platform to generate simulation data and uses MATLAB software to establish fault detection and diagnosis model. The experiment shows that the fault diagnosis method based on PCA can react to the abnormal change of the process, and it can find out the cause and the details of the fault correctly. But for PCA, it does not be applied to linear and non-separable data. Literature [19] proposed the kernel PCA method (KPCA), which can be applied to data fault diagnosis.

Fault Diagnosis Based on ICA

Because today's industrial process is developing towards large-scale and complex development, there are a large number of measurement variables in the process, but these variables are not independent of each other, but are driven by only a few independent variables. Independent component analysis (ICA), [20], is to deal with this situation, its purpose is to decompose the observed data linearly, and decompose it into statistical independent components. The fault diagnosis based on ICA is to monitor and analyze the decomposed independent components in order to realize the purpose of fault diagnosis.

A new statistical process control method based on ICA is proposed in document [21]. This method is superior to traditional PCA method in multivariable statistical process control. In document [22], a nonlinear process control method based on nuclear independent component analysis (KICA) is proposed. This method is applied to the fault diagnosis in the TE process, which can effectively capture the nonlinear relationship in the process variables and show the superior fault detection ability. The [23] first analyzes some shortcomings of the original ICA method, and then proposes a ICA method based on particle swarm optimization algorithm (PSO-ICA), the fault diagnosis method is applied to the TE process, showing that PSO-ICA method can effectively capture the independent component of process variables.

Fault Diagnosis Based on Clustering

From Clustering in machine learning attempts to divide the castrated edition of data into several disjoint subsets, and each subset becomes a cluster. Through such a division, each cluster may have some potential concepts. In fault diagnosis based on clustering, clustering algorithm is usually used to cluster the normal data and fault data, and then each cluster is analyzed to achieve the purpose of fault diagnosis.

Document [24] uses interval two type fuzzy C-means clustering method to simulate ethylene cracking furnace process data, and it verifies that this method is effective. The [25] for C clustering in small samples and categories of traditional cross space sample error problem, the paper studies a dynamic kernel clustering algorithm, and proposed using this method to identify the production of ethylene cracking furnace optimal operation mode, the feasibility of dynamic kernel clustering method in table Ming on the actual industrial applications. In document [26], a fuzzy C means clustering algorithm based on particle swarm optimization is proposed to overcome the outburst of clustering results in fuzzy C mean clustering algorithm. The algorithm is applied to motor fault diagnosis. Experiments show that the algorithm makes up for the defects of fuzzy C-means algorithm, improves the efficiency and accuracy of clustering, and improves the result of fault diagnosis.

Fault Diagnosis Method Based on Semi-Supervised Learning

In real industrial production process, a lot of unlabeled sample data can be collected easily and obtaining labelled sample data requires manpower and material resources. Semi supervised learning is to allow the learners to improve their learning performance without relying on external interaction and using unlabeled samples automatically. Therefore, the fault diagnosis based on semi supervised learning has gradually become a hot spot in the field of fault diagnosis.

The [27] presents a fault diagnosis method based on online semi supervised learning, and experimental verification based on the failure mode of diesel engine under 8 different conditions were studied, the result showed the markers of fault samples is small, the addition of unlabeled fault samples can cause failure classification accuracy of 5%~8%. The [28] proposed a semi supervised SVM fault detection method based on the two stages of learning, the experimental method with support vector machine (SVM), fuzzy support vector machine (FSVM) were compared, the results show that the method in different number identification sample set under the condition that the fault detection accuracy is improved greatly compared with the other algorithm.

CONCLUSION

As the industrial process is becoming more and more complex, it is becoming more and more difficult or impossible to establish a precise mathematical model for fault diagnosis. In this paper, a data driven fault diagnosis method is reviewed. Data driven fault diagnosis only needs a lot of process data, and the knowledge discovery is carried out by machine learning and data mining. With the rapid growth of human data hitherto unknown, I believe that in the future to data driven fault diagnosis method will gradually become the mainstream trend, also urgently need to develop more of the fault diagnosis method based on data driven to deal with every kind of problem, the combination of complementary methods I believe will be a direction of development of a new fault diagnosis method the.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Fund (61772145,61672174, pdjha0334), Guangdong province science and technology projects (2015B020233019). Zhiping Peng is corresponding author.

REFERENCES

1. WH GUI, XY LIU, “Fault Diagnosis Technologies Based on Artificial Intelligence for Complex Process”, Basic Automation, 2002,9(4):1-6.
2. H Li, DY XIAO, “Survey on data driven fault diagnosis methods”, Control and Decision, 2011,26(1):1-9.
3. Pitts E, “A Logical Calculus of the Ideas Imminent in Nervous Activity”, 1943.
4. Buscema M, “Back propagation neural networks”, Substance Use & Misuse, 1998,33(2):233-270.

5. Hoskins JC, Kaliyur KM, Himmelblau DM, “Fault diagnosis in complex chemical plants using artificial neural networks”, Aiche Journal, 1991,37(1):137-141.
6. Ungar LH, Powell BA, Kamens SN, “Adaptive networks for fault diagnosis and process control”, Computers & Chemical Engineering, 1990,14(4–5):561-572.
7. Venkat V, Chan K, “A neural network methodology for process fault diagnosis”, Aiche Journal, 1989,35(12):1993-2002.
8. Venkatasubramanian V. “Inexact Reasoning in Expert Systems: A Stochastic Parallel Network Approach”, Artificial Intelligence Applications, Caia 1985, the Engineering of Knowledge-Based Systems, Proceedings of the Second Conference, Miami Beach, Florida, Usa, December, 1985.
9. Watanabe K, “Incipient Fault Diagnosis of Chemical Process via Artificial Neural Networks”, Aiche Journal, 1989,35(11):1803-1812.
10. F QIAN, JH YU, “Neural network method for fault diagnosis of industrial cracking furnace”, Global Conference on intelligent control and automation, 2000.
11. Vapnik VN, Vapnik VN, “An overview of statistical learning theory”, IEEE Transactions on Neural Networks. 1999,10(5):988-999.
12. HZ WANG, XS ZHANG, JS YU, “Fault diagnosis based on SVM”, Journal of East China University of Science and Technology,2004,30(2):179-182.
13. C SONG, HY HUANG, “Immune cultural algorithm and its applications in fault diagnosis of ethylene cracking furnace”, CIESC Journal, 2012,63(12):3996-4002.
14. Y XU, GH ZHANG, F IQAN, “Least Squares Support Vector Machines Based on Improved Clonal Selection Algorithm”, Journal of East China University of Science and Technology, 2008,34(5):729-733.
15. Jackson JE, Mudholkar GS, “Control procedures for residuals associated with principal component analysis”, Technometrics. 1979,21(3):341-349.
16. Jackson JE. “Quality Control Methods for Several Related Variables”, Technometrics. 1959,1(4):359-77.
17. ZG ZHAO, F LIU, “Algorithm for non-parametric confidence bounds based on kernel principal component analysis”, Journal of systems engineering, 2008,23(1):96-100.
18. Q ZHU, HB SHI, “Fault detection and diagnosis based on PCA for multivariable statistics”, Control engineering of China, 2006(s1).
19. Schölkopf B, Smola A, “Nonlinear component analysis as a kernel eigenvalue problem”, 2006.
20. ZQ YANG, Y LI, DW HU, “Independent component analysis a survey”, Acta Automatica Sinica 2002,28(5):762-772.
21. Kano M, “Monitoring independent components for fault detection”, Aiche Journal, 2003,49(4):969–976.
22. Lee JM, Qin SJ, “Fault detection of non-linear processes using Kernel independent component analysis”, Jorgchem. 2007,05(:4):526–536.
23. Y ZHANNG, Z YANG, “Fault detection of non-Gaussian processes based on modified independent component analysis”, Chemical Engineering Science. 2010,65(16):4630-4639.
24. Y GAO, ZL WANG, “Application of Type-2 Fuzzy Theory in Ethylene Cracking Furnace Process Monitoring”, Journal of East China University of Science and Technology, 2014,40(3):302-308.
25. ZQ GENG, QX ZHU, “Optimal pattern recognition in ethylene production via dynamic kernel clustering algorithm”, Control and instruments in chemical industry, 2005,32(2):5-8.
26. F LI, SX XIA, Q NIU, “Fault Diagnosis Method Based on Improved Particle Swarm Optimization Clustering”, Microelectronics & Computer, 2010,27(8):82-85.
27. G YIN, YT ZHANG, “Research on fault diagnosis method based on online semi supervised learning”, Journal of vibration engineering, 2012,25(6):637-642.
28. XM TAO, PD CAO, “Fault detection based on two-stage learning and semi-supervised SVM”, Journal of Vibration and Shock, 2012,31(23):39-43.