Journal of Process Control 31 (2015) 45–54

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Decision fusion systems for fault detection and identification in industrial processes

Fuyuan Zhang[a], Zhiqiang Ge[a,b,*]

[a] State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, PR China
[b] Key Laboratory of Advanced Control and Optimization for Chemical Processes, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 3 August 2014
Received in revised form 6 March 2015
Accepted 9 April 2015
Available online 25 May 2015

Keywords:
Fault detection and identification
Decision fusion system
Data-driven model
Diversity
Dempster–Shafer evidence theory

ABSTRACT

Numerous fault detection and identification methods have been developed in recent years, whereas, each method works under its own assumption, which means a method works well in one condition may not provide a satisfactory performance in another condition. In this paper, we intend to design a fusion system by combining results of various methods. To increase the diversity among different methods, the resampling strategy is introduced as a data preprocessing step. A total of six conventionally used methods are selected for building the fusion system in this paper. Decisions generated from different models are combined together through the Dempster-Shafer evidence theory. Furthermore, to improve the computational efficiency and reliability of the fusion system, a new diversity measurement index named correlation coefficient is defined for model pruning in the fusion system. Fault detection and identification performances of the decision fusion system are evaluated through the Tennessee Eastman process.

1. Introduction

It is well known that not only the proper monitoring of the industrial process is significant and practical, but also the fast and precise identification of faults is essential for reducing the number of off-products and improving the productivity of the process. Thus, searching for the method which is effective and well-suited for monitoring is becoming more and more important. In chemical process industries, particularly, fault detection and identification is a hot research spot in the past years.

Generally, process monitoring methods can be divided into three categories [1–5]: model-based methods, knowledge-based methods, and data-based methods. Due to advantages of having few requirements of the process model and the associated expert knowledge, the data-based method has recently become the most popular one for process monitoring. Among all data-based process monitoring methods, typically used ones include principal component analysis (PCA), independent component analysis (ICA), partial least squares (PLS), artificial neural networks (ANN), etc. Although satisfactory results have been obtained in many industrial processes by using those mature methods, the equipment used in the industrial plants become more and more complicated and multifunctional, and the state of process is usually the combination of many operating conditions, which may cause performance deteriorations of those methods. In fact, it is obvious that sometimes the choice of one method under a single assumption will not achieve good results that we expect, as is shown in the work of Venkatasubramanian [4], as a result of the mismatching between the real process and the model assumption.

Therefore, it is a question here, is there a perfect method that can deal with any complex condition in a process? The answer is absolutely no. According to the No Free Lunch theorem [6], there is no algorithm which is universally superior to others, that is to say, we are not able to design a strategy that can adapt to a variety of situations, e.g. non-Gaussian data distributions, nonlinear relationships among process variables, frequent changes of operating conditions, etc.

In order to address this problem, some researchers put forward to the idea of ensemble systems [7,8]. The main purpose is to combine sorts of methods which have completely different emphases on modeling the data when dealing with the same problem through some efficient fusion algorithms. One key factor of the ensemble system is the characteristic of diversity, which means

* Corresponding author at: State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, PR China. Tel.: +86 571 87951442.
E-mail address: gezhiqiang@zju.edu.cn (Z. Ge).

http://dx.doi.org/10.1016/j.jprocont.2015.04.004
0959-1524/

each single model needs to express different views of the system, thus has different errors so that the total error can be reduced after the ensemble process. Although there is no strict definition and explicit measurement of the diversity, it has been illustrated that the more the diversity is, the better the fusion results could be [9,10]. For example, Polikar [11] has experimentally proved that the ensemble of multiple classifiers performed better than a single one where the diversity is quite significant. The other key factor of the ensemble system is about the decision making or combination for various models. In general, there are two categories: utility-based methods and evidence-based methods. A representative of the former is the voting-based method [12–14], and the latter includes Bayesian method [15], Dempster–Shafer (D–S) method [16], decision templates [17], Borda count [18], etc.

Compared to other decision making approaches, the D–S framework provides a more flexible mathematical tool for dealing with imperfect information, as well as a more simple computing procedure and concise expression of final decision. What is more, the D–S method has no limitation of the data distribution, which can bring lots of conveniences during data preprocessing. Due to those advantages, the D–S based method has been widely used for decision making in the past years [15,19–22], and has also been proved to be an appropriate approach for improving the performance of an ensemble model that deals with unreliable information [20].

In this paper, the Dempster–Shafer evidence theory is employed for the development of decision fusion systems for fault detection and identification. In order to enhance the diversity performance of the fusion system, a resampling strategy is introduced as a data preprocessing procedure, in addition to using different types of data models. Furthermore, through defining a new correlation measurement index, those classifiers which have similar characteristics are pruned from the fusion system. As a result, both of the computational efficiency and the classification reliability can be improved. Here, the fusion system which incorporates all classifiers is called as ALL fusion system, and the one with pruning strategy is represented as SELECTIVE fusion system.

The rest of the paper is organized as follows. Section 2 provides a review of preliminary knowledge about the Dempster–Shafer evidence theory. Due to the length of this paper, we have ignored detailed preliminary knowledge about selected unsupervised and supervised modeling methods, since one can easily find them in many published books and papers. Section 3 describes a complete framework of ALL and SELECTIVE fusion systems, with the definition of a new index to measure the correlations among different methods. Online fault detection and identification results are illustrated based on the proposed framework by using the Tennessee Eastman (TE) process in Section 4. Finally, conclusions are made.

2. Dempster–Shafer evidence theory

The evidence theory is initially proposed by Dempster [23] concerning lower and upper probability distribution, and Shafer [16] proved the ability of the belief functions to model uncertain knowledge. Then, the complete Dempster–Shafer theory was formulated, which enables us to combine evidences from different sources and arrives at a degree of belief which has been widely used in the field of information fusion. In this section, some basic concepts and combination rules of the Dempster–Shafer theory are introduced, one can refer to Shafer [16], Smets and Kennes [24], or Yager [25] for more detailed instructions on this subject.

2.1. Basic definitions

Definition 1. Let Ω be a finite non-empty set of N mutually exhaustive and exclusive hypotheses about some fault class domain. Then, Let us denote 2^{Ω}, the power set of Ω, composed with all the proposition of F in Ω.

$$\Omega = \{F_1, F_2, \ldots, F_N\} \tag{1}$$

$$2^{\Omega} = \{\emptyset, \{F_1\}, \{F_2\}, \ldots, \{F_N\}, \{F_1 \cup F_2\}, \{F_1 \cup F_3\}, \ldots, \Omega\}. \tag{2}$$

Definition 2. Basic probability assignment (BPA), also called the mass function or basic belief assignment, is a function mapping from 2^{Ω} to [0,1] which assigns a belief value to each element of power set. It satisfies the following two properties:

$$\begin{aligned} & m : 2^{\Omega} \to [0, 1] \\ & m(\emptyset) = 0 \\ & \sum_{A \subseteq \Omega} m(A) = 1 \end{aligned} \tag{3}$$

where $\emptyset$ is an empty set and it is called normalized BPA with $m(\emptyset)=0$, otherwise, each subset A when $m(A) > 0$, is called the focal element of m.

Definition 3. The belief function is defined as $bel : 2^{\Omega} \to [0, 1]$

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{4}$$

Definition 4. The plausible function is defined as $pl : 2^{\Omega} \to [0, 1]$

$$Pl(A) = 1 - Bel(\bar{A}) = \sum_{A \cap B \neq \emptyset} m(B) \tag{5}$$

where $\bar{A}$ is the negation of a hypothesis A.

Definition 5. $[Bel(A), Pl(A)]$ is the confidence interval which describes the uncertainty about A. If the difference between Bel and Pl increases, then the information available used for fusion will decrease. Therefore, the difference provides a measurement of uncertainty about the level of evidence.

2.2. Rule of combination

When multiple independent sources of evidence are available, such as m_1 and m_2, the combined evidence can be obtained by Dempster's rule as follows:

$$m(\emptyset) = 0, m_{1.2}(A) = m_1(A) \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C) \tag{6}$$

where $K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$, it represents the BPA when the result of the combination is an empty set, and is often interpreted as a measurement of conflict between the two pieces of evidence, which satisfies $K \neq 1$. Obviously, the larger K is, the more conflict the evidences are, and the less information is available.

Obviously, the Dempster's rule can be easily extended to more than two hypothesis, as shown in Eq. (7), i.e., by combining the BPAs of first two classifiers (m_1 and m_2) using Eq. (6) to obtain the combined BPA (m_1, m_2) and then combine the result ($m_{1.2}$) with the BPA of the third classifier (m_3) and so forth until the Tth classifier.

$$\begin{aligned} m_{1,2,\ldots,T} &= m_1 \oplus m_2 \oplus \ldots \oplus m_T = (((m_1 \oplus m_2) \oplus m_3) \oplus \ldots \oplus m_T) \\ &= ((m_{1,2} \oplus m_3) \oplus \ldots \oplus m_T) \ldots \end{aligned} \tag{7}$$

In recent years, Dempster–Shafer based fusion has been widely used in various fields, such as pattern recognition, process fault diagnosis, geographic information systems, medical diagnosis. For example, Parikh et al. [26,27] used the Dempster–Shafer evidence theory to combine the outputs of multiple primary classifiers to improve overall classification performance. The effectiveness of this approach was demonstrated for detecting failure in a diesel