Control Engineering Practice 80 (2018) 146–156

Contents lists available at ScienceDirect

Control Engineering Practice

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

Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation

Daniel Jung [a,*], Kok Yew Ng [b,c], Erik Frisk [a], Mattias Krysander [a]

[a] Vehicular Systems, Linköping University, Linköping, Sweden
[b] School of Engineering, Ulster University, Newtownabbey, BT37 0QB UK
[c] Electrical and Computer Systems Engineering, School of Engineering, Monash University Malaysia, Malaysia

ARTICLE INFO

Keywords:
Fault diagnosis
Fault isolation
Machine learning
Artificial intelligence
Classification

ABSTRACT

Machine learning can be used to automatically process sensor data and create data-driven models for prediction and classification. However, in applications such as fault diagnosis, faults are rare events and learning models for fault classification is complicated because of lack of relevant training data. This paper proposes a hybrid diagnosis system design which combines model-based residuals with incremental anomaly classifiers. The proposed method is able to identify unknown faults and also classify multiple-faults using only single-fault training data. The proposed method is verified using a physical model and data collected from an internal combustion engine.

1. Introduction

Fault detection and isolation are important tasks in fault diagnosis systems to identify the root cause when faults occur in the system. This is complicated by the fact that there are often many possible diagnosis candidates (fault hypotheses) that can explain the system state. In a workshop, this can result in a mechanic having to troubleshoot several components in a system before identifying the true fault, which is both costly and time-consuming (Pernestål, Nyberg, & Warnquist, 2012).

Two common approaches in fault diagnosis are model-based (Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003b) and data-driven (Yin, Ding, Xie, & Luo, 2014). Data-driven diagnosis in general classifies faults by using classifiers learned from training data using nominal data and data from different faults (Theissler, 2017). However, in many industrial applications, faults are rare events and available training data from faulty conditions is usually limited (Dong, Shulin, & Zhang, 2017; Sankavaram, Kodali, Pattipati, & Singh, 2015). Collecting a sufficient amount of data from relevant fault scenarios is a time-consuming and expensive process. Also, if there are faults that do not occur before several years of system operation time, they might not be considered during system development. Therefore, it is desirable that a diagnosis system is not only able to identify and localize known faults as they occur, but it should also be able to identify new types of faults and to improve fault classification performance over time as new data are collected.

One solution to limited training data from different fault scenarios is the use of physical models. In model-based diagnosis, fault isolation is mainly performed by matching a set of triggered residual generators with different fault signatures to compute diagnosis candidates (Cordier et al., 2004). An advantage of model-based methods, with respect to data-driven methods, is that fault isolation performance can be achieved without training data from different faults. Even though the fault has not been observed before, it is possible to point out likely fault locations based on residual information and model analysis (Pucel, Mayer, & Stumptner, 2009). However, there are often many diagnosis candidates that can explain the triggered residuals, meaning that it can still be difficult to identify the actual fault.

1.1. Problem motivation

A combined diagnosis system design has the potential of both model-based and data-driven diagnosis methodologies (Tidriri, Chatti, Verron, & Tiplica, 2016). The objective of such a hybrid diagnosis system design is to improve fault classification performance by using both physical models and data collected from previous fault occurrences. Another advantage is that performance can improve over time by incrementally retrain the data-driven classifiers as new data are collected. The idea is to first compute diagnosis candidates (fault hypotheses) that can explain the set of triggering residuals by using a fault isolation algorithm. A test quantity is evaluated to determine if a residual has triggered, i.e., has deviated from its nominal behavior, or not. The next step is to rank the different candidates, determining which candidate is the most likely, using a set of data-driven classifiers where each classifier

* Corresponding author.
E-mail addresses: daniel.jung@liu.se (D. Jung), mark.ng@ulster.ac.uk (K.Y. Ng), erik.frisk@liu.se (E. Frisk), mattias.krysander@liu.se (M. Krysander).

https://doi.org/10.1016/j.conengprac.2018.08.013
Received 18 April 2018; Received in revised form 4 July 2018; Accepted 21 August 2018
0967-0661/

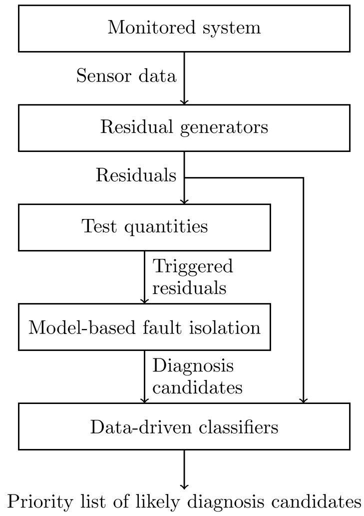

Fig. 1. A schematic of the diagnosis system design. The data-driven fault isolation is used to rank diagnosis candidates computed by the consistency-based fault isolation.

models a different fault hypothesis. The proposed diagnosis system structure is illustrated in Fig. 1. Fault isolation here refers to the problem of rejecting inconsistent diagnosis candidates while fault classification ranks how likely each of the candidates are. The purpose of the data-driven classifiers is not to reject any of the diagnosis candidates but to evaluate which of the computed candidates that are more likely by comparing residual data to previous observations of the different faults.

This paper extends on the analysis and results of the proposed hybrid diagnosis system design presented in Jung, Ng, Frisk, and Krysander (2016). A framework is formulated for combining model-based fault isolation and data-driven fault classification. Also, with respect to the previous work, the performance and robustness of the proposed hybrid diagnosis system design are evaluated using a model and collected data of an internal combustion engine.

1.2. Related research

Discussions regarding model-based and data-driven fault diagnosis methods can be found in, for example, Ding et al. (2011), Venkatasubramanian, Rengaswamy, Kavuri, and Yin (2003a), and Venkatasubramanian et al. (2003b). A survey of previous works combining model-based and data-driven fault diagnosis techniques is presented in Tidriri et al. (2016), which also points out that there are potential advantages of applying a framework to integrate the model-based and data-driven methodologies. A hybrid framework is proposed in Tidriri, Tiplica, Chatti, and Verron (2018) where different sets of residuals designed using bond graphs and sensor data are combined using a Bayesian Network (BN). The BN is used to classify the most likely fault even though there are inconsistencies between the outputs of the different residual sets and sensor data, for example if they compute different fault hypotheses. With respect to previous work, this paper proposes a hybrid fault classification strategy which computes diagnosis candidates and the likelihood of each candidate, including the likelihood of unknown faults, without increasing the risk of rejecting the true diagnosis.

In Loboda and Yepifanov (2010) and Luo, Namburu, Pattipati, Qiao, and Chigusa (2010), different sets of test quantities are designed using model-based and data-driven methods. In Shashoa, Kvaščev, Marjanović, and Djurović (2013), a model is estimated using data from a thermal power plant and a data-driven classifier is then used for fault classification. Model-based residual selection is combined with training data in Jung and Sundström (2017) to automatically identify important residuals and design test quantities, and in Cheng, Wang, and Xu (2016), residual detection performance is improved using machine learning to compensate for model uncertainties. In Jung, Khorasgani, Frisk, Krysander, and Biswas (2015), a brief comparison is made between different hybrid approaches to monitor a wind turbine. Combined methods have also been proposed for prognostics and condition-based maintenance (Chen & Pecht, 2012; Sankavaram et al., 2009). With respect to these previous works, the main focus in this paper is fault isolation and not residual design.

2. Fault isolation and model-based diagnosis

The first part of the diagnosis system in Fig. 1 follows a general model-based architecture where residuals are used to detect inconsistencies between model predictions and sensor data. In this section, it is summarized how the diagnosis candidates are computed as a set of fault hypotheses that can explain the set of triggered residuals.

In industrial systems, there are usually many potential faults that can occur that will have varying impact on the system and its performance. Let $\mathcal{F} = \{f_1, f_2, \dots, f_{n_f}\}$ denote a set of n_f *known* types of faults to be monitored by a diagnosis system. However, the set $\mathcal{F} \subseteq \mathcal{F}^*$ only represents the known subset of all possible faults $\mathcal{F}^*$ that can occur in the system. Thus, the set $\mathcal{F}$ can increase over time as new types of faults are identified.

In many cases, it is possible that multiple faults can be present in the system at the same time. Therefore, to describe the system state the term *fault mode* is used which is defined as follows.

Definition 1 (*Fault Mode*). A fault mode $F \subseteq \mathcal{F}$ is a set of faults that is present in the system.

As an example, $F = \{f_1, f_2\}$ represents the case where both f_1 and f_2 are present in the system. The nominal system state $F = \emptyset$, i.e., when the system is fault-free, is denoted the No Fault (NF) case.

2.1. Fault detection

In order to detect if a fault is present in the system, a set of residual generators $\mathcal{R} = \{r_1, r_2, \dots, r_{n_r}\}$ is computed. A residual generator is a function of sensor and actuator data which ideally is zero in the fault-free case (Svärd, Nyberg, & Frisk, 2013). A residual generator is said to be *sensitive* to a fault f_i if that fault implies that the residual is non-zero, ideally. If a residual generator is not sensitive to fault f_i, it is also said that the fault is *decoupled* from that residual generator.

Note that the definitions of residual generators and fault sensitivity describe the ideal case. However, fault detection performance is complicated by model uncertainties and measurement noise. Therefore, a change in the residual output is usually determined by evaluating a test quantity, for example statistical post-processing (Basseville, Nikiforov, et al., 1993) and thresholding of the residual.

The different residual generators are designed to monitor different parts of the system, i.e. to be sensitive to different subset of faults. The following definition of fault detectability for a given set of residual generators $\mathcal{R}$ is used (Jung & Sundström, 2017).

Definition 2 (*Fault Mode Detectability*). A fault mode $F_i \subseteq \mathcal{F}$ is structurally detectable if there exists a residual generator $r_k \in \mathcal{R}$ that is sensitive to at least one fault $f \in F_i$.

The relation between which residual generators are sensitive to which faults can be summarized in a Fault Signature Matrix (FSM). An example is shown in Table 1 where a mark at location (k, l) in the FSM indicates that residual r_k is sensitive to fault f_l. As an example, residual r_1 is sensitive to the faults f_{Waf} and f_{pim}, but not to f_{pic} and f_{Tic}.

2.2. Fault isolation

After a fault has been detected, i.e., when one or more residuals have triggered, the next step is to perform fault isolation. Fault isolation