Annual Reviews in Control 38 (2014) 32–44

Contents lists available at ScienceDirect

Annual Reviews in Control

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

Anomaly detection/detectability for a linear model with a bounded nuisance parameter

Fouzi Harrou [a,*], Lionel Fillatre [b], Igor Nikiforov [c]

[a] Chemical Engineering Program, Texas A&M University at Qatar, Doha, Qatar
[b] Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France
[c] ICD, LM2S, University of Technology of Troyes, UMR 6281, CNRS, 12, rue Marie Curie CS 42060, 10004 Troyes Cedex, France

ARTICLE INFO

Article history:
Received 21 April 2013
Accepted 1 March 2014
Available online 13 April 2014

ABSTRACT

Anomaly detection is addressed within a statistical framework. Often the statistical model is composed of two types of parameters: the informative parameters and the nuisance ones. The nuisance parameters are of no interest for detection but they are necessary to complete the model. In the case of unknown, non-random and non-bounded nuisance parameters, their elimination is unavoidable. Some approaches based on the assumption that the nuisance parameters belonging to a subspace interfere with the informative ones in a linear manner, use the theory of invariance to reject the nuisance. Unfortunately, this can lead to a serious degradation of the detector capacity because some anomalies are masked by nuisance parameters. Nevertheless, in many cases the physical nature of nuisance parameters is (partially) known, and this a priori knowledge permits to define lower and upper bounds for the nuisance parameters. The goal of this paper is to study the statistical performances of the constrained generalized likelihood ratio test used to detect an additive anomaly in the case of bounded nuisance parameters. An example of the integrity monitoring of GNSS train positioning illustrates the relevance of the proposed method.

1. Introduction

1.1. The state of the art

Monitoring systems is becoming increasingly important to maintain reliability and safe system operation. It can be defined as the set of actions carried out to detect, isolate faulty measurement sources and then remove them before they affect the system performance (Isermann, 2006). The role of detection is to identify any anomaly event indicating a distance from the system behavior compared to its nominal behavior. Furthermore, anomaly isolation determines the location of the detected anomaly. In this paper, however, the focus will be on anomaly detection only. The most important issue in the monitoring safety-critical systems is to detect anomalies that infect the monitored system (Basseville & Nikiforov, 2002; Fillatre & Nikiforov, 2007; Fouladirad & Nikiforov, 2005; Lacresse, Grall, & Nikiforov, 2005). Fault detection is essential for proper and safe system operation. The need for monitoring techniques that can accurately and quickly detect abnormal situations and anomalies has greatly attracted the attention of researchers and engineers. Over the past few decades, several monitoring techniques have been developed (Chaitanya, 2011; Qingsong, 2004; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003a, Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003c). Anomaly detection methods can be classified into three main categories: data-based or model-free methods, model-based or analytical methods, and knowledge-based methods (Venkatasubramanian et al., 2003a, 2003c). Anomaly detection using knowledge based methods is usually a heuristic process. The approaches in this category are mostly based on causal analysis, expert systems (Kim et al., 2005), possible cause and effect graphs (PCEG) (Wilcox & Himmelblau, 1994), failure modes and effects analysis (FMEA) (Wirth, Berthold, Krämer, & Peter, 1996), Hazop-digraph (HDG) (Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003b), or Bayesian networks (Sylvain & Abdessamad, 2008). The main limitation of these methods is that they are more suitable to small-scale systems with a small number of variables, and thus may not be appropriate to monitor complex systems. Furthermore, data-based anomaly detection methods rely on the availability of historical data of the inspected system under normal operation mode (Venkatasubramanian et al., 2003c). These data are first used to build an empirical model of the system, which is then used to detect anomalies in future data. Data-based monitoring methods include the latent variable methods, e.g., partial least square

* Corresponding author. Tel.: +974 66741841; fax: +974 44230065.
E-mail addresses: fouzi.harrou@tamu.edu (F. Harrou), lionel.fillatre@i3s.unice.fr (L. Fillatre), nikiforov@utt.fr (I. Nikiforov).

http://dx.doi.org/10.1016/j.arcontrol.2014.03.003
1367-5788/

(PLS) regression, principal component analysis (PCA), canonical variate analysis (CVA), independent component analysis (ICA) (Chaing, Russel, & Braatz, 2001; Venkatasubramanian et al., 2003c), neural networks (Subbaraj & Kannapiran, 2010), Fuzzy systems (Dexter & Benouarets, 1996), and pattern recognition methods (Mohammadi & Asgary, 2005). However, the applicability of data-based methods mainly depends on the availability of quantity and quality of the data. On the other hand, model-based or analytical anomaly detection methods rely on comparing the system measured variables with the information obtained from a mathematical model of the monitored system, which is usually developed based on some fundamental understanding of the system under anomaly-free conditions. These methods received more attention by engineers and practitioners than the data-based methods because of their mathematical and systematic characteristics. The residuals, which are the difference between the measurements and the model predictions, is an anomaly indicator about the monitored system. Using the information carried by the residuals, anomalies can be detected and isolated (Kinnaert, 2003; Nyberg & Nyberg, 1999). When the monitored system is under normal operating conditions (no anomaly), the residual is zero or close zero due to modeling uncertainties and measurement noise. However, when an anomaly occurs the residuals deviate significantly from zero indicating the presence of a new condition that is significantly distinguishable from the normal faultless working mode (Kinnaert, 2003; Nyberg & Nyberg, 1999). The model-based monitoring approaches include the observer-based approaches (Clark, Fosth, & Walton, 1975; Patton, Clark, & Frank, 1989; Xu, 2002), parity space approaches (Chow & Willsky, 1984; Frank, 1990; Patton & Chen, 1991; Ragot, Maquin, & Kratz, 1993; Staroswiecki, 2001), and interval approaches (Adrot, 2000; Adrot, Ploix, & Ragot, 2002; Benothman, Maquin, Ragot, & Benrejeb, 2007). Of course, the effectiveness of these model-based monitoring methods depends on the accuracy of the models used.

In the framework of parametric model-based anomaly detection approaches (Venkatasubramanian et al., 2003a), the anomalies detection is based on the model which describes the monitored system. The model is usually developed based on some fundamental understanding of physics of the process, under fault-free conditions. Often, this model is composed of two types of parameters: the informative parameters and the nuisance ones. Typically, the informative parameters including the anomalies, often called the parameters of interest, define the statistical hypotheses, and the nuisance parameters are of no interest for detection but they are necessary to complete the model (Basseville & Nikiforov, 2002). It is supposed that the nuisance parameters are unknown but non-random. In particular, this means that the nuisance can be intentionally chosen by adversary (or attacker) to maximize the negative impact of anomaly on the monitored system (by masking the anomaly, for example). The nuisance parameter has no desirable impact on the performance indexes. When designing a test for deciding between two hypotheses in the presence of a nuisance parameter, the goal is to achieve performance indexes independent from the actual value of the nuisance parameters. Therefore, the only solution is to reduce or eliminate the impact of the nuisance parameters on the decision function. Ideally, we would like to design a decision function which depends only on the parameter of interest.

How to deal with nuisance parameters is an important problem in the framework of statistical anomaly detection. Several elimination methods are available to handle these parameters. It is impossible to describe all of them in this paper. We overview some key ideas to reduce or eliminate the effect of nuisance parameters. In the likelihood approach, the problem is to find a likelihood function for the parameter of interest only (Basseville & Nikiforov, 2005). Another approach consists of constructing a pseudo-likelihood function for the parameter of interest (Basu, 1977; Liseo, 1999; Pawitan, 2001; Reid & Fraser, 2003; Severini, 2000). A pseudo-likelihood function depends only on the data and the interest parameter. Conditional and marginal likelihood functions are examples of what are called pseudo-likelihood functions. These approaches use the conditioning or marginalization to eliminate the nuisance parameters from the likelihood function and they are only available when the model has a particular structure. Furthermore, even when a marginal or conditional likelihood functions exists, calculation of the likelihood function is often difficult (Pawitan, 2001; Severini, 2000). An alternative approach is to use the modified profile likelihood function which is available for general models (Pawitan, 2001; Severini, 1998, 2000). The modified profile likelihood may be derived as an approximation to either a marginal or conditional likelihood when either of those likelihoods exists. Furthermore, the calculation of the modified profile likelihood function does not require the existence of a marginal or conditional likelihood and, hence, it has been adopted for general use. Also there is a Bayesian approach which is based on the knowledge of a priori law on nuisance parameters (Berger, Liseo, & Wolpert, 1999; Basu, 1977; Liseo, 1999). It is necessary to fix an a priori law of nuisance parameter, compute the posterior, integrate out the nuisance parameter from the posterior to arrive at the posterior distribution (integrated likelihood function) which depend only on the parameter of interest. The form of the posterior distribution depends on the a priori law used. In literature (Basu, 1977; Liseo, 1993, 1999; Robert, 2006) various a priori laws on the nuisance parameters were presented, among the priori laws more often used are the uniform law, the Haar measure, the reference prior and the law of Jeffrey (Eaton & Sudderth, 2010). The Bayesian approach is especially efficient if a reliable a priori information is available on the nuisance parameters. The invariant approach (Borovkov, 1998; Ferguson, 1967; Lehmann, 1996) is based on the nuisance rejection and, therefore, does not use any a priori information on the distribution of nuisance parameter.

1.2. Motivation

The elimination of nuisance parameters is universally recognized as a major problem of statistics (Basu, 1977; King, 1996). In the case of non-Bayesian approach, the elimination of the unknown but non-random nuisance parameters is unavoidable. This elimination presents several difficulties. If no a priori information is available on the nuisance, it is assumed that the nuisance parameters belong to a certain subspace of the observation space. Often the theory of invariant tests is used to reject the nuisance parameters in such a case, especially if the nuisance parameters interfere with the informative ones in a linear manner. The rejection of the nuisances by using the theory of invariance lead to a serious degradation of the detector capacity and the problem of anomaly detectability appears, i.e. some anomalies can be masked by the nuisance parameters due to the procedure of their rejection (Fillatre & Nikiforov, 2007). Nevertheless, in many cases the physical nature of nuisance parameters is (partially) known, and this knowledge may allow us to define lower and upper bounds to limit the variations of these parameters (the power of an engine is limited, the altitude of an aircraft is always positive...) (Lacresse & Grall, 2001). This information have to be integrated in the statistical decision rule to improve the power function of the detector and to reduce the subset of undetectable anomalies. It was shown that taking into account the lower and upper bounds of the nuisance parameters the anomaly detector performs better than in the case of un bounded nuisance parameters (Harrou, Fillatre, & Nikiforov, 2008).