Chaos, Solitons and Fractals 103 (2017) 588–595

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Detecting outliers in complex nonlinear systems controlled by predictive control strategy

Biao Wang[a], Zhizhong Mao[a], Keke Huang[b,*]

[a] Department of Control Theory and Control Engineering, Northeastern University, Shenyang 110819, China

[b] School of Information Science and Engineering, Central South University, Changsha 410083, China

ARTICLE INFO

Article history:
Received 29 March 2017
Revised 26 June 2017
Accepted 20 July 2017

Keywords:
Outlier detection
Complex nonlinear system
Model predictive control
Ensemble learning
One-class classification

ABSTRACT

Detecting outliers in complicated nonlinear systems that are controlled by model predictive control is a significant work for engineering applications. Based on the features of data in practical systems, we propose a one-class classification ensemble method incorporating the notion of Feature Subspace with Bagging. Clustering and PCA (Principal Component Analysis) are integrated to obtain a more informative feature space, where Feature subspaces and bootstrap replications are implemented orderly to generate more accuracy and diverse base learners. A detector is constructed based on the above methodology, and a model updating strategy is also provided. By means of comparison with competitive methods, the effectiveness of the proposed detector has been verified.

1. Introduction

Nonlinearity is the most prominent feature of complex systems in multidisciplinary discipline, such as chemistry [1], metallurgy [2], aerospace [3], biology [4], physics [5] etc. Referring to the control problem of nonlinear systems, model predictive control (MPC) has been extensively researched. The basic control strategy in MPC is the selection of a set of future control moves (control horizon) and minimizing a cost function based on the desired output trajectory over a prediction horizon with chosen length [6]. This requires a reasonably accurate prediction model that captures the essential nonlinearities of the process under control, to predict multi-step ahead dynamic behavior. Typically, the prediction model is constructed once at the initial phase of a MPC scheme. Nevertheless, slow drifts in unmeasured disturbances and changes in process parameters will lead to significant mismatch in plant and model behavior as time progresses. Furthermore, a MPC scheme is usually designed under the assumption that sensors and actuators are free from faults, which is usually impossible for real-world systems, where soft faults, such as biases in sensors or actuators, are frequently encountered. Finally, sensors or actuators may also fail during operation, which results in loss of degrees of freedom for control. Such occurrences can lead to a significant degradation in the closed loop performance of the MPC scheme and may also lead to instability. We represent the structural schematic diagram of dynamic matrix control (DMC) [7] in Fig. 1, from which we could find that forecasting model plays a crucial role in DMC scheme. Then qualities of measurements of y_R, u, y_M, d, y will have great influence directly or indirectly on the final performance of DMC. As a result, we can conclude that the primary reason for these phenomena is that abnormal measurements (outliers) would participate directly or indirectly in the construction of prediction model, which leads to model-plant mismatch naturally. Therefore, timely detecting outliers and preventing them to construct the prediction model are extremely significant for MPC schemes of nonlinear systems. Following the definition of outlier by [8], we regard outliers as patterns in data that do not conform to a well-defined notion of normal behavior, which is usually represented by normal process data in the sense of practical nonlinear systems. For most the published work in the literatures, addressing outliers for nonlinear systems controlled by MPC scheme has rarely been considered, which motivates us for this study.

In this paper, we develop a novel direction that is referred to as outlier detection for nonlinear systems. According to the literature we have referred to, model predictive control was often used as an effective control strategy for those complex nonlinear systems. However, MPC has never been the only one, other methods like adaptive control has also been successfully applied in several nonlinear systems. In other words, our proposed outlier detection scheme may be appropriate for many nonlinear systems and more data-based control strategies including MPC. However, in this paper, our focus is *complex nonlinear systems controlled by MPC*. It is also worth stating that only the detection method

* Corresponding author.
E-mail address: huangkekefendou@126.com (K. Huang).

http://dx.doi.org/10.1016/j.chaos.2017.07.018
0960-0779/

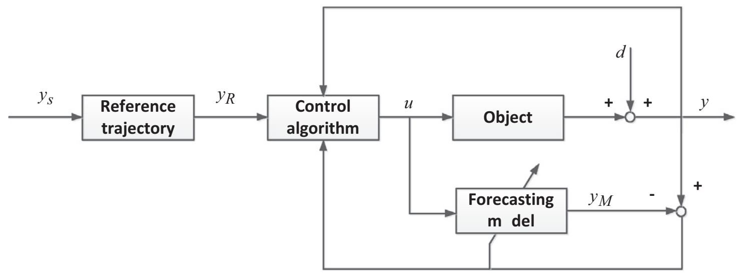

Fig. 1. Structural schematic diagram of dynamic matrix control (DMC).

is focused and the problem of control has not been researched here. Actually, the problem of outlier detection has drawn a significant amount of interests from artificial intelligence, data mining and machine learning in the past decade, reflecting in the publication of massive research papers, such as distance-based methods, density-based methods, classification-based methods, cluster-based methods [9] etc. In contrast to these traditional methods, in this paper, we investigate the characteristics of data and propose a novel and hybrid one-class classifier ensemble learning solution. Our method incorporates the notion of Bagging and Feature Subspace in order to deliver simultaneously accurate and diverse base learners. Our proposed feature selection method that integrates the thought of clustering with PCA can generate a new feature space, which is more concise and informative than the original feature space. Several feature subsets can be generated from this new feature space and more subsets can be further obtained via bootstrap sampling. With sub-models training on these subsets and an aggregation rule, unseen measurements can be grouped into two categories, namely normal samples and outliers. Finally, we propose a simple but appropriate strategy to update the detection method in order to capture the dynamics in process data.

The rest of this paper is structured as follows. The characteristic of data in practical chaotic systems is summarized in Section 2. Section 3 expands upon the proposed method. A series of experiments are carried out in Section 4. Finally, some conclusions are drawn in Section 5.

2. Features of data and method

In this section, we analyze the features of process data in practical nonlinear systems, based on which essential features of corresponding outlier detection method can be also be concluded.

(1) Data unlabeled. In most MPC schemes of nonlinear systems, data used for constructing the forecasting model are measurements of process data that are collected online. Labeling for these real-time measurements is a time-consuming task and even an impossible task.
(2) Noisy should be another prominent feature of data in practical systems. Due to the complicated operating conditions, noisy data is inevitable.
(3) High-dimensional data is increasingly pervasive in many areas with the development of advanced sensor technology. Due to the increasing scale of the nonlinear systems, requirements for control are becoming more rigorous. As a result, control performance can reach a high level only with more state variables participating into the control framework.
(4) In addition, inherent methodology of MPC also requests the detection to be implemented in real-time, which indicates that the detection phase should be fast enough.

As thus, we can conclude that a competent outlier detection method for practical nonlinear systems should: (1) be constructed with unlabeled data (unsupervised learning); (2) be robustness to

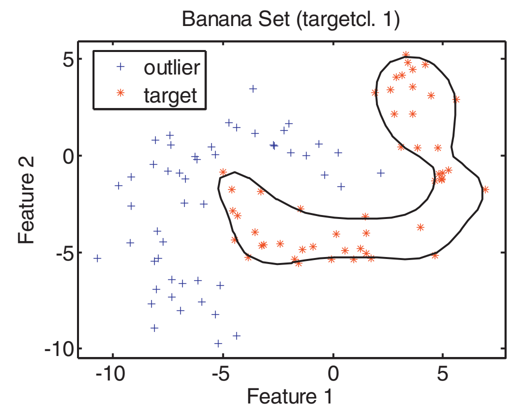

Fig. 2. An example of one-class classifier; "*" denotes target class, and "+" denotes outlier class; the solid line represents the trained boundary.

noise in training set; (3) cope with high-dimensional data; (4) be implemented in real-time; (5) be fast enough. Although the algorithms for outlier detection can also be applied for fault diagnosis and detection (FDD), the ultimate objective is totally different from that of FDD. The aim of a FDD is similar with that of process monitoring, i.e. guaranteeing the product quality, while the aim of this paper is to provide convenience for process control. Both two processes are of great importance for industrial processes [10]. Furthermore, our proposed method can also facilitate the implementation of FDD like the scheme in [11].

3. Methodology

According to the analysis in Section 2, we propose a hybrid and novel one-class classifier ensemble method, where the notion of Bagging [12] and Feature Subspace are combined to deal with the intricate data in practical nonlinear systems. The following subsections will expand upon methodology concerning our outlier detection method.

3.1. One-class classification

One-class classification (OCC) can be regarded as a special case of classification problem where the collection of counter-examples is hard even impossible, which is extremely appropriate for outlier detection under the condition of unlabeled data. In addition, the detection phase of classification-based method is simple and fast enough to be implemented in real-time. Generally, the task of OCC is to define a boundary around the target class (often normal examples), such that it accepts as much of the target objects as possible, while it minimizes the opportunity of accepting outlier objects. Fig. 2 illustrates a one-class classifier trained with a banana-shaped dataset. A number of OCC classifiers have been proposed in the literature and can be grouped into three categories.