ARTICLE IN PRESS
JAL:446

Journal of Applied Logic ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

On the combination of support vector machines and segmentation algorithms for anomaly detection: A petroleum industry comparative study

Luis Martí [a,*], Nayat Sanchez-Pi [b], José Manuel Molina López [c], Ana Cristina Bicharra Garcia [a]

[a] *Institute of Computing, Universidade Federal Fluminense, Niterói (RJ), Brazil*
[b] *Institute of Mathematics and Statistics, Universidade do Estado do Rio de Janeiro, Rio de Janeiro (RJ), Brazil*
[c] *Department of Informatics, Universidad Carlos III de Madrid, Colmenarejo (Madrid), Spain*

A R T I C L E   I N F O

Article history:
Available online xxxx

Keywords:
Anomaly detection
Support vector machines
Time series segmentation
Kalman filters
Oil industry application

A B S T R A C T

Anomaly detection has to do with finding patterns in data that do not conform to an expected behavior. It has recently attracted the attention of the research community because of its real-world application. The correct detection unusual events empower the decision maker with the capacity to act on the system in order to correctly avoid, correct, or react to the situations associated with them. Petroleum industry is one of such real-world application scenarios. In particular, heavy extraction machines for pumping and generation operations like turbomachines are intensively monitored by hundreds of sensors each that send measurements with a high frequency for damage prevention. For dealing with this and with the lack of labeled data, in this paper we describe a combination of a fast and high quality segmentation algorithm with a one-class support vector machine approach for efficient anomaly detection in turbomachines. As a result we perform empirical studies comparing our approach to another using Kalman filters in a real-life application related to oil platform turbomachinery anomaly detection.

1. Introduction

The importance of anomaly detection is a consequence of the fact that anomalies in data translate to significant actionable information in a wide variety of application domains. The correct detection of such types of unusual information empowers the decision maker with the capacity to act on the system in order to correctly avoid, correct, or react to the situations associated with them.

* Corresponding author.
E-mail address: lmarti@ic.uff.br (L. Martí).

http://dx.doi.org/10.1016/j.jal.2016.11.015
1570-8683/

Please cite this article in press as: L. Martí et al., On the combination of support vector machines and segmentation algorithms for anomaly detection: A petroleum industry comparative study, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.015

Anomaly detection has extensive use in a wide variety of applications such as fraud and intrusion detection [10], fault detection in safety critical systems [16], finance [4] or industrial systems (see [35,6] for surveys on this topic).

One example in industry applications is the detection of anomalies in turbomachinery installed in offshore petroleum extraction platforms. Recent history shows us how important a correct handling of this equipment is as failures in this industry have a dramatic economical, social and environmental impact. In the particular case of industrial anomaly detection, units suffer damage due to continuous usage and the normal wear and tear. Such damages need to be detected early to prevent further escalation and losses. Data in this domain is referred to as industrial sensor data because it is recorded using different sensors and collected for analysis and it has a temporal aspect and time series analysis is also used in some works like: [15].

Due to the lack of labeled data for training/validation of models, in [20,30], we provided a solution for the detection of anomalies in turbomachinery, using a one-class SVM. This technique uses one class learning techniques for SVM [26] and learns a region that contains the training data instances (a boundary). Kernels, such as radial basis functions (RBF), are used to learn complex regions. For each test instance, the basic technique determines if the test instance falls within the learnt region. If a test instance falls within the learnt region, it is declared as normal; else it is declared as anomalous. We combined this technique with a time series segmentation to prune noisy, unreliable and inconsistent data.

In this paper we apply this approach in a real-life context oil industry-related problem and, in order to assert its validity, we compare its performance with currently used approaches, like Kalman filters and confidence intervals.

The remainder of this paper is organized in the following manner. In the next section, we discuss some theoretical foundations of the work. Subsequently, we describe our proposal in detail. After that, we present a case study for offshore oil platform turbomachinery. This case study is used to compare our approach with the application of Kalman filters. Finally in Section 5, some conclusive remarks and directions for future work are presented.

2. Foundations

The preset work addresses the problem of anomaly detection by comparing one-class SVM classifiers and Kalman filters with a novel and fast segmentation algorithm specially devised for this problem. In this section we present the theoretical pillars supporting the proposal.

2.1. Anomaly detection

Fault and damage prevention is known as the problem of finding patterns in data that do not conform to an expected behavior [8]. Unexpected patterns are often referred as anomalies, outliers or faults, depending on the application domain. In broad terms, anomalies are patterns in data that do not conform to a well-defined 'normal' behavior and, hence, can be labeled as 'anomalies' [8].

Anomaly detection techniques are also classified in three categories based on the nature of the undesired anomaly: point anomaly, contextual anomaly, collective anomaly. Point anomaly is when an individual data instance can be considered as anomalous with respect to the rest of the data. If this data instance is anomalous in a specific context then it is considered to be contextual anomaly. Furthermore, if a collection of related data instances is anomalous with respect to the rest of entire dataset, where these individual instances are not anomalies by themselves but their occurrence together is anomalous, then it is termed as collective anomaly.

Applications of anomaly detection techniques vary depending on the user, the problem domains, and even the data. Anomaly detection techniques have been proposed in literature, based on distribution, distance, density, clustering and classification.

Please cite this article in press as: L. Martí et al., On the combination of support vector machines and segmentation algorithms for anomaly detection: A petroleum industry comparative study, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.015