Smart Alarming Methods: an overview, highlight on statistical methods

Jean-Paul Valois[1], Christophe Blondeau[2], Simplice Dossou-Gbete[3], Laurent Bordes[4]

[1] TOTAL, F64018- Pau Cedex (France)
jean-paul.valois@total.com
[2] TOTAL, F64018- Pau Cedex,(France)
christophe.blondeau@mines-nancy.org
[3,4] Dépt. Mathématiques, Univ. Pau, BP 576, F64012 PAU Cedex (France)
simplice.dossou-gbete@univ-pau.fr
laurent.bordes@univ-pau.fr

Abstract. Methods of Smart Alarming intend to detect as soon as possible novelty or anomaly in Data Streams. A review is proposed to highlight the key points of using them. In case of univariate data, the more suitable method is not the same as for stationary variable or non-stationary variable. Multivariate data set are often dealt with using unsupervised learning based methods, either with factor analysis (mostly PCA) or clustering algorithms. Each of these methods must be applied in a specific situation: prior knowledge of possible anomalies should be needed or not, learning data set can be large sized or not, and so on. Some examples are outlined. Discussion underlines the importance of having a prior knowledge of variable behaviour, and to consider the global flow chart, including eventually a data preprocessing.

Keywords: Smart alarming, Novelty detection, Anomaly detection.

1 Introduction

Several industrial contexts produce data streams [1], [2], and practical needs can be to diagnose as soon as possible any change of the system under monitoring. A lot of smart alarming or novelty detection methods have been designed to detect and characterize the changes. The aim is to detect the novelty before it becomes obvious, and thus prevent its consequences e.g. [3], [4].
Applications have been designed to control the industrial production [5], to forecast the Stock Exchange (graphic approach), to analyze biometric images [6], and so on. Very few reviews have appeared [7], [8], [9], [10]. Methods can be split [7] into parametric [11] if a known family of distribution is assumed to model the learning data set, or non-parametric otherwise. Both cases often result in a probability distribution, the test data set (most recent values) is deemed to be a novel when it falls into low probability region or over a fixed threshold. Methods are thus ranked according to their involved algorithms. Our topic is different, we intend to highlight some practical key points that could appear using these methods or that should be first considered in order to conveniently design a smart alarming project. In this scope the paper proposes (part 2) a literature overview, then (part 3) a few (outlined here shortly) examples.
As an automatic "black box" procedure is often the final product, information concerning novelties is often graphically displayed; the methodology to do this is a noticeable point, e.g. [12], [13], which is not considered in our paper.

2 An overview of methods

The methods from the references survey have been classified into four items: stationary or non-stationary data, unsupervised multivariate learning base methods using PCA or clustering.