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.

2.1 Stationary data

In the simplest cases, the learning data set can be characterized by a constant (e.g. manufacturer requirement…) so the used threshold appears as an objective and acceptable limit. This widely used method is fully deterministic and does not involve any statistical concept. It is typically an univariate method ; multivariate cases can also be considered, variables are then taken separately, in this case a complex mixing of logical constraints must been designed in order to refer all the possible faults. Domains of using are e.g. oil industry [14], [15], [4] or safety nuclear plants [5]. These methods are easy to design in the univariate case. They are well suited to cases where the signal/noise ratio is higher, otherwise false-alarms or no-detection could occur.
A second widely used family of methods takes into account the probability of finding the latest value(s). An anomaly is deemed if either the test values fall into a low density region of the considered distribution, or they lie out of the 2 or 3 σ interval. The learning data set should not include any anomaly. Several works reviewed in [7] have been devoted to the situation where it is difficult to specify the probability function.
A non-parametric approach takes into account the quantiles, displaying these as box-plots graphs [16]. The "median method" [17], [18], [19] used in medicine involves an other non-parametric way to define the probability of occurrence of further anomalies; so well specified and described anomalies are need in the learning data set.
A special case takes into account the probability for an equipment (engine…) to fall out of order [20], this does not strictly refer to the smart alarming method because the aim is to renew the equipment before it is out of order.
Threshold method is in fact a frequent concept in other more advanced methods (especially those of maintenance policies).

2.2 Non-stationary data

Trend methods aim to account for the dynamic of the system in comparing the latest set of values with the learning data set. For instance, a regression is performed. Statistical tests can consider latest values, comparing them with the limits of the regression confidence interval, either or monitoring along time the regression parameter themselves.
An interesting algorithm of this family is the 'change-point' detection method [21], used in oil industry [23], [3] or in finance [22]. This method takes into account a large period before the latest point and splits the curve into homogeneous several time sections. The last anomalous event, if any, is therefore detected after a lag time.

2.3 Unsupervised learning based methods

2.3.1 Factor Analysis and PCA

PCA has been noticed for a long time as able to detect multivariate outliers [24]. Further research [25] first runs PCA using a without fault data set. Then the latest data are added. One way is to take them as a passive set (supplementary points) in a new PCA run; the author supposes that the "secondary" factors should be the most modified. The scores are compared with reconstructed data, an anomaly is deemed in case of residuals differing from a white noise. Besides, this method besides determines which variable is responsible for the deviation.
A second way is to include the latest data set in the active data set. The previous PCA modeling (correlations variables – factors) is more or less modified; differences are tested and monitored along time. A domain of using is the survey of atmospheric pollution.
In both ways stationary variables are needed and the learning data set is supposed not to have any outlier.

2.3.2 Clustering methods

A large variety of clustering algorithms have been used (K-means, hierarchical, Self-Organizing Maps,…), and sometimes improved [9]. Clustering is first performed using the learning data set; it should be convenient that this includes all the expected anomalies. A second run includes the test data set. Imputation of latest points can result into two different situations. They could be included in the previous

classes (and in this case could either deemed to standard situation or to an anomaly); otherwise the actual situation is a new one and one cannot conclude whether last values should be deemed as a new normal situation or as an unknown anomaly. Besides the user could be faced with a more or less intense rearrangement of the previous classes; analysis of the consensus can make the interpretation less easy.
The distance of the test sample from a class mean – e.g. using a nearest neighbor algorithm – can be considered [26], [6], [27], [28].
Clustering methods have been used in oil industry [29]. To better use these methods, all the possible anomalies should be identified. Nevertheless it is difficult to diagnose an *incertae sedis* situation. These methods do not describe the anomaly (which variable changed, and amplitude of this deviation).

2.4 Advanced methods

A lot of more advanced methods have been experienced in the scope of Smart Alarming, and have been reviewed in [7], [8], [9], like fuzzy logic, neural networks, support vector machine, wavelets...Discussion about outlines of practice has been here limited to the mostly used or standard methods.

3 Outlines of examples

A few methods have been selected in each of the main ways of novelty detection which have been highlighted in section 2. Each way has been exemplified with either synthetic or real data sets. 'Trade off" validation method [18], [30] has been performed in convenient cases.
Although this set of examples is limited, it draws attention to some discussion points.
- The behavior of the variable(s) - stationary or not, signal/noise ratio of expected anomalies, and so on - must be considered first, the most suitable method can be simple or more advanced.
- A custom-made method should sometimes be designed according to the physical meaning of data, their data type, and of course the practical needs.
- The flow chart should eventually include a preprocessing step, including for instance extraction of some data features or filtering: smoothing (like LOESS [31], KALMAN [32], or MEWMA [33] filters) can offer an alternative when the learning data set cannot be afforded without anomalies.

4 Conclusion

Some smart alarming methods have been reviewed and ranked according to statistical and practical considerations. Outlines of practice have been deduced considering most frequently used or standard methods, main points could probably be appropriate using more advanced algorithms. Anyway the global flow chart must be considered, eventually including preprocessing steps (transformation of the variable to better detect the anomalies). Smart alarming needs to be carefully designed, to avoid false alarms, and to not miss real ones. In all cases, a general point seems to have a thorough knowledge of the variable's behavior, and of its physical meaning; in some cases this can include a precise diagnostic and description of all expected anomalies. The dimension of the learning data set should be considered also, this can be of low dimension (univariate as an extreme case) or of high dimension: some methods cannot be designed with a smaller training data set. The user should decide if a long term drifting of the process must be handled as anomaly or not, in the last case continuous upgrading of the model must be excluded. Besides, the suitable method should not be the same whether the expected anomalies are previously well known or if they have not been described, or if a quite new normal situation could occur.

References

1 Garofalakis, M., Gehrke, J., Rastogi, R.: Querying and mining data streams: you only get one look. In Proceedings of the 2002 ACP SIGMOD International Conference on Management of Data, 635-635, ACM Press (2002)
2 Aguilar-Ruiz, J.S.: Recent advances in data stream mining, 38èmes Journées de Statistiques, SFdS, Paris, 29 may-2 june (2006)

3 McCann, D., White, D., Rodt, G.: Computerized flow monitors detect small kicks. - Oil and Gas Journal, 90, 8, Feb 24, 62-65 (1992)
4 Weishaupt, M.A., Omsberg, N.P., Jardine, S.I., Patterson, P.A.: Rig Computer system improves Safety for deep HP/HT wells by kick detection and well monitoring, Society of Petroleum Engineers, n°23053, presented at the Offshore Europe Conference, Aberdeen, Scotland, Sept. 3-6 (1991)
5 AMDEC, Procédures pour l'Analyse des Modes de Défaillance, de leurs Effets leurs Criticités, instruction armée américaine MIL-P-1629, Nov-9 (1949), http://chaqual.free.fr/outils/amdec/histoireamdec.html (2000)
6 Tarassenko, L., Nairac, A., Townsend, N., Cowley, P., Novelty detection in jet engines, IEE Colloquium on Condition Monitoring Imagery, External Structures and Health, 41-45 (1999).
7 Markou M., Singh S.: Novelty detection, a review - part 1: statistical approaches, Signal Processing, 83(12): 2481-2497 (2003)
8 Markou M., Singh S.: Novelty detection, a review - part 2: neural network based approaches, Signal Processing, 83(12):2499-2521 (2003)
9 Hodge, V.J., Austin, J.: A survey of outlier detection methodology, Kluwer Academic Pub. Netherlands Artificial Intelligence Review, 22, 2, 85-126 (2004)
10 Odin, T., Addison, D.: Novelty detection using neural network technology, Proc. COMADEN conf. (2000)
11 Desforges, M.J., Jacob, P.J., Cooper, J.E.: Applications of probability density estimation to the detection of abnormal conditions in engineering, Proc. Institute of Mechanical engineers, vol. 212, 687-703 (1988)
12 Gillick, B., Gaber, M.M., Krishnasswamy, Sh., Zaslavsky, A.: Visualisation of Cluster Dynamics and Change Detection in Ubiquitous Data Stream Mining. Faculty of Information Technology, Monash University, ICML 2006, 3rd International Workshop on Knowledge Discovery from Data Streams, Pittsburgh, June (2006), http://www.cs.cmu.edu/~jroure/iwkdds/schedule.html
13 White, D., Lowe, C.: Advances in Well Control Training and Practice, 3rd annual IACD et al Europe Well Contr. Conf. (Noordwijkerhout, Neth, 6-2 April 92) Proc Pap 9, 12 pp (1992)
14 Jardine S.I., McCann, D.P., Barber, S.S.: An advanced system for the early detection of sticking pipe, IADC/SPE Drilling Conference, New-Orleans, Feb 18-21 (1992)
15 Leder , P.C., McCann, D.P., Hatch, A.J.: New Real Time Anti-collision Alarm Improves Drilling Safety, Society of Petroleum Engineers, n°36016 (1996), pub. Journal of Petroleum Technology, 241-242 (1996)
16 Laurikkala, J., Juhola, M., Kentala, E.: Informal identification of outliers in medical data, Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2000), Berlin, August (2000)
17 Goldman, G., Waterson, C.K, Lucas, W.: Smart anesthesia monitoring system: Rule-based intubation detection, IEEE/Engineering in Medicine and Biology Society Annual Conference Part 3 (1988)
18 Krol, M., Reich, D.L.: The Algorithm for detecting Critical conditions during anaesthesia, IEEE Computer Society, 12th IEEE Symposium on Computer-Based Medical Systems, CBMS'99, 208-213 (1999)
19 Reich, D.L. , Osinski, T.K., Bodian, C., Krol, M., Saric, R. K., Roth R.: An algorithm for assessing intraoperative mean arterial pressure lability - Anaesthesiology , 87 , 1 , 156-161 (1997)
20 Lannoy, A., Procaccia, H.: Evaluation et maîtrise du vieillissement industriel, Ed. Tec&Doc, Lavoisier, coll. EDF R&D, 361pp (2005)
21 Davison, A.C., Hinkley, V.D.: Bootstrap methods and their application, Technometrics, 42, 2, 216-17 (2000)
22 Taylor, W.: Change-Point Analyzer 2.0 shareware program, Taylor Enterprises, Libertyville, Illinois. Web: http://www.variation.com/cpa (2000)
23 Budleson, B.: Early detection of drill string washouts reduce fishing jobs, World oil, oct. (1990)
24 Gnadadesikan, R.: Methods for statistical data analysis of multivariate observations, Wiley, New-York (1977)
25 Harkat, F.: Détection et localisation de défauts par analyse en composantes principales, thèse de doctorat, Institut National Polytechnique de Lorraine, France, 30 june (2003).
26 Schölkopf, B., Williamson, R.C., Smola, A.J., Shawe-Taylor, J., Platt, J.: Support vector method for novelty detection. In Neural Information Processing Systems (2000).
27 Manson, G., Pierce, G., Worden, K.: On the long-term stability of normal condition for damage detection in a composite panel, Proc. 4th Intern. Conf. on Damage Assessment of Structures, Cardiff, UK, June (2001)
28 Tax, D.M.J., Duin, R.P.W., Outlier detection using classifier instability, in Advances in Pattern Recognition, the Joint IAPR International Workshops, 593-601, (1998)
29 Zangl, G., Hannerer, J.: Data Mining, Applications in the Petroleum Industry. Round Oak Publishing (2003)
30 Seagull, F., Sanderson, P.: Anaesthesia alarms in surgical context -Proceedings of the Human Factors and Ergonomics Society, v 2, 1048-1052 (1998)
31 Cleveland, W.S.: Visualizing data, 360pp., Hobart Press, New Jersey (1993)
32 Kalman, R.E.: A New Approach to Linear filtering and prediction Problems, Transactions of the AMSE, Journal of Basic Engineering, 35-65, March (1960)
33 Cheng, M.L., Away, Y, Hasan, M.K.: The algorithm and design for real-time Hotelling's T^2 and MEWMA control chart in MSPC (2002), interstat.statjournals.net/YEAR/2004/articles/0407001.pdf (2004)