Engineering Applications of Artificial Intelligence 49 (2016) 10–18

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Brief paper

Pattern recognition in multivariate time series – A case study applied to fault detection in a gas turbine

Cristiano Hora Fontes*, Otacílio Pereira

Graduate Program in Industrial Engineering, Polytechnic School, Federal University of Bahia, Rua Aristides Novis, 2, Federação, 40.110-630, Salvador, Bahia, Brazil

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form
12 November 2015
Accepted 18 November 2015
Available online 10 December 2015

Keywords:
Data mining
Multivariate time series
Clustering
Fault detection
Gas turbines

ABSTRACT

Advances in information technology, together with the evolution of systems in control, automation and instrumentation have enabled the recovery, storage and manipulation of a large amount of data from industrial plants. This development has motivated the advancement of research in fault detection, especially based on process history data. Although a large amount of work has been conducted in recent years on the diagnostics of gas turbines, few of them present the use of clustering approaches applied to multivariate time series, adopting PCA similarity factor (SPCA) in order to detect and/or prevent failures. This paper presents a comprehensive method for pattern recognition associated to fault prediction in gas turbines using time series mining techniques. Algorithms comprising appropriate similarity metrics, subsequence matching and fuzzy clustering were applied on data extracted from a Plant Information Management System (PIMS) represented by multivariate time series. A real case study comprising the fault detection in a gas turbine was investigated. The results suggest the existence of a safe way to start the turbine that can be useful to support the development of a dynamic system for monitoring and predicting the probability of failure and for decision-making at operational level.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Advances in Information Technology (IT) have substantially improved the storage and handling of large amounts of data. It has also led to the development of methods using Knowledge Discovery in Data bases (KDD) and Data Mining (DM) that use data to obtain useful knowledge, adding value to businesses in strategic sectors such as e-commerce, medicine and the economy (Fazlollahtabar et al., 2012; Aktep and Ersoz, 2012; Xu and Liang, 2013; Bisso and Samanez, 2014). In industrial processes, technologies for data acquisition have been consolidated. This scenario encourages the use of data mining approaches for the development of systems based on the knowledge extracted from data of process variables (Strachan et al., 2007). Time series is an important class of temporal data objects and one of the mining tasks associated to this kind of data comprises pattern discovery and clustering (Fu, 2011; Izakian et al., 2015).

There are some works about pattern recognition in univariate time series (Liao, 2005; Keogh and Kasetty, 2002; Aghabozorgi et al., 2015) using standard approaches (Liao, 2005; Trebuňa and Halčinová, 2013). This kind of clustering problem can be solved using point-prototype clustering models (Bezdek et al., 2005). However, pattern recognition in multivariate time series (MTS) represents a more complex problem (non-point prototyping problem) with intrinsic features such as metrics of similarity, cluster validity and domain knowledge (Singhal and Seborg, 2005; Yang and Shahabi, 2004). This kind of problem cannot be solved directly using classical point-prototype clustering models such as Fuzzy C-Means. Furthermore, additional challenges must be considered such as the extraction of features from each data/object (set of time series) and the similarity metrics among objects.

Some works deal with clustering and pattern recognition using uni and multivariate time series together with applications. Liao (2005), Keogh and Kasetty (2002) and Fu (2011) present reviews involving clustering, classification, segmentation and pattern recognition on univariate and MTS with feature extraction approaches based on correlation analysis, wavelets and others. Other works highlight the use of PCA-based similarity metrics (SPCA) (and modified versions) in pattern recognition evolving MTS and present case studies associated to nonlinear dynamic systems (Singhal and Seborg, 2006; Dobos and Abonyi, 2012; Deng and Tian, 2013; Zhang et al., 2011). Considering the diversity of data types and their classifications, cluster analysis and recognition of prototypes (patterns) applied to the time series present some challenges and have a broad spectrum of applications (or possible applications), especially in industrial processes. A recent work (Izakian et al., 2015) proposed the use of a Fuzzy clustering

* Corresponding author.
E-mail addresses: cfontes@ufba.br (C.H. Fontes), otaciliojpereira@gmail.br (O. Pereira).

http://dx.doi.org/10.1016/j.engappai.2015.11.005
0952-1976/© 2015 Elsevier Ltd. All rights reserved.

approach (Fuzzy C-Means and Fuzzy C-Medoids) combined with a Dynamic Time Warping (DTW) technique to compare MTS (Bankó and Abonyi, 2012; Petitjean et al, 2011). Coppi et al. (2010) presented a Fuzzy approach for classifying and clustering multivariate spatial time series, considering the dissimilarity between the objects to be clustered and the features of the space of time trajectories together in the same optimization problem. Unlike the classical concept of the multivariate time series (D'urso and Maharaj, 2012), this approach is based on the concept of a spatial-time data array, where the dissimilarity is measured using the Euclidian distance considering the positions of multivariate trajectories (different variables) at each time instant. D'urso (2004) presented a fuzzy approach for clustering multivariate time series, i.e., cross sectional and longitudinal approaches, focusing on the latter, which is related to classifying objects (MTS) based on their time trajectories. The author proposed different formulations for the optimization problem of fuzzy clustering (longitudinal approach), following a concept of weighted Euclidian distance between each object and the centroid of the cluster (pattern). The weighted Euclidian is defined as the weighted sum of the Euclidian distances between each time series and the centroid at each time instant. The classifier design is another problem associated with pattern recognition using time series and some works present alternative model structures for this application (Angelov et al., 2007; Pereira et al., 2014, Simon and Hungerbuhler, 2010).

Some works have been done in recent decades into reliability and operational availability in industrial plants such as chemical and power generation plants, developing and applying methods for the Fault Detection and Diagnosis (FDD) (Nybø, 2010, Di Maio et al., 2013, Haomin et al., 2014, Baraldi et al., 2015, Mayadevi et al. 2014, Lázaro et al., 2015). Dash and Venkatasubramanian (2000) and Venkatasubramanian et al. (2003) classify FDD methods into model-based and historical data-based. Yousefi et al. (2011) present a model to detect and isolate faults in an industrial gas turbine based on the residual analysis of process variables. They adopt a model-based approach using Auto Regressive with eXogeneous Input (ARX) and Artificial Neural Networks (ANN) structures. A large amount of work has been conducted in recent years on the diagnostics of gas turbines (Dai et al., 2008, Yongbao et al., 2009, Volponi et al., 2003, Simon, 2008, Verma et al., 2006, Verma and Ganguli, 2005, Zedda and Singh, 2002, Ganguli, 2012, Zhai et al., 2007, Bertini et al., 2009), but few works present the use of clustering approaches applied to time series to detect and/or prevent failures in the process or equipment. A recent work (Wang et al., 2015) presented a fusion approach based on fuzzy clustering (Fuzzy C-Means algorithm) and support vector machine. In this case, the problem comprises significant differences between the measured outputs of 12 thermocouple temperature sensors (simultaneously), and the FCM algorithm was used to obtain the initial clustering results considering a sample with 490 averages of temperatures measured in each sensor. In the specific case of failure analysis in gas turbines, model-based approaches are applied considering empirical models identified through historical data (Rasaienia et al., 2013; Gupta et al., 2008). Even in the empirical models, the clear understanding of the phenomenon, interactions between variables, nonlinear and multivariable nature poses a challenging problem for the model-based approach. In this case, the historical based-approach represents a potential alternative to recognize behaviors or patterns of failures in equipment or production systems. One of the most commonly used techniques applied in this case comprises the Principal Component Analysis, a well-known dimensionality reduction tool capable of identifying either variables responsible for the fault and/or variables most affected by the fault. Some works present the use of the PCA in FDD (Zhang et al., 2012; Bassily et al., 2009; Li and Wen, 2014; Deng and Tian, 2013). Some works are based on the use of specific techniques associated to the recognition of failure patterns such as discrete wavelet transforms, ANN and fuzzy clustering (Lemma and Hashim, 2012; Alaei et al. 2013). Alaei et al. (2013) propose a new on-line fuzzy clustering-based algorithm developed using a mixture of adaptive Principal Component Analysis and Weighted Fuzzy C-Means (WFCM) method and the results are shown through the data base extracted from the benchmark Tennesse Eastman (TE) process (Chen et al. 2014; Yin et al., 2012).

This paper presents a comprehensive historical data-based method for the acquisition of knowledge represented by patterns of operation from industrial plant data. This method, including an extension of the Fuzzy C-Means algorithm adjusted to non-point prototyping problem, is tested and validated based on a real case study that comprised the recognition of starting patterns in a gas turbine of commercial scale for fault detection purposes. This piece of equipment is the main section of a thermoelectric unit belonging to the industrial park of the Brazilian Oil Company (Brazil). The following section presents the industrial unit and the gas turbine. Sections 3 and 4 present and discuss the method and results.

2. The case study

The industrial unit analyzed in this work is a Thermoelectric Power Plant (TPP) (Fig. 1a) which belongs to the Brazilian Oil Company. It comprises a cogeneration unit that operates in a combined cycle producing steam and electricity using natural gas as the raw material (fuel).

The TPP has three gas turbines (GT), Rolls Royce RB211-G62 DF (Rolls-Royce, 2010), each one coupled to an electric generator of 27 MW, in conjunction with other equipment to drive a total generation of 137 MW of electricity and produce 260.3 t/h steam.

The basic operation of the gas turbine comprises the injection of fresh atmospheric air through a compressor and energy is added by spraying fuel gas into combustion chambers where it

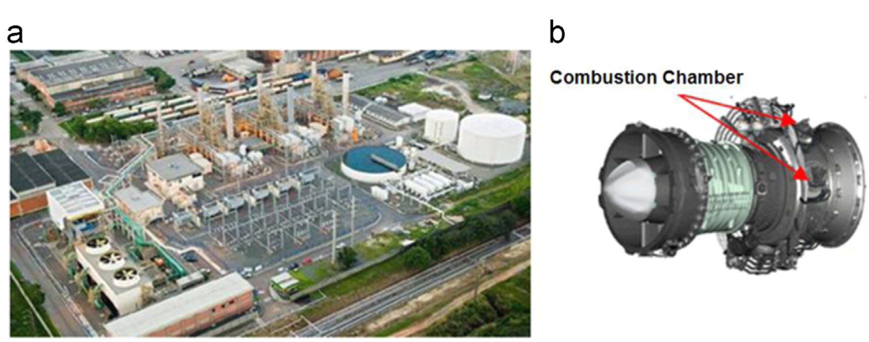

Fig. 1. (a) TPP (Pereira et al., 2014) and (b) gas turbine RB211-G62 DF (Rolls-Royce, 2010).

Download English Version:

https://daneshyari.com/en/article/380241

Download Persian Version:

https://daneshyari.com/article/380241

Daneshyari.com