Brazilian Journal of Operations & Production Management
Volume 10, Number 2, pp. 79-90.

Fuzzy rule-based classifier for Fault Prediction in a Thermoelectric Unit

Otacílio José Pereira
PhD student in Industrial Engineering at Federal University of Bahia), Master in Computer Science and graduated in Computer Engineering. In 2010 he became a research in UFBA/Brazilian Oil project on Pattern Recognition and Artificial Intelligence applied to Industrial Plants; Lecturer in computer science and engineering at University of Salvador (UNIFACS). Experience as software engineer and database designer worked in projects for mining companies and eucalyptus pulp industry.
otaciliojpereira@gmail.com

Cristiano Hora de Oliveira Fontes
Universidade Federal da Bahia, Escola Politécnica, Rua Aristides Nóvis, 02, Federação, Salvador, BA, CEP 40210-630, Brasil – 55 71 32839505
cfontes@ufba.br

Carlos Arthur M. Teixeira Cavalcante
Universidade Federal da Bahia, Escola Politécnica, Rua Aristides Nóvis, 02, Federação, Salvador, BA, CEP 40210-630, Brasil – 55 71 32839505
arthurtc@ufba.br

Sérgio Torres Sá Barretto
Petróleo Brasileiro S.A. (PETROBRAS), TIC-BA/SITI, Av. ACM, 1113, EDIBA, Itaigara, Salvador, BA, CEP 41856-90, Brasil
sa.barretto@petrobras.com.br

Luciana de Almeida Pacheco
Petróleo Brasileiro S.A. (PETROBRAS), PTEC-BA/CGEDA, Av. ACM, 1113, EDIBA, Itaigara, Salvador, BA, CEP 41856-90, Brasil
luciana.pacheco@gmail.com

Abstract

Pattern recognition from data is a potential alternative for the extraction of knowledge about processes and it may be useful for predicting failures, control and support decision making, among others. The knowledge extracted can be used to implement models based on Artificial Intelligence such as Fuzzy Inference Systems (FIS). Tools from Information Technology (IT) and automation techniques can also be used in data-based approaches to enable the storage and handling of large amounts of historical process data. This paper presents the implementation of a fuzzy inference system for fault prediction in a gas turbine of a thermoelectric unit. The first step comprised the pattern recognition through the clustering of multivariate time series obtained from the Plant Information Management System (PIMS). The second step comprised the development of a FIS using a data-based approach to define the membership functions and rules. The results showed the potential of the fuzzy model to predict the probability of failure during the start of the turbine this presenting a feasible alternative to support decision-making at operational level.

Keywords: fuzzy inference systems; pattern recognition; multivariate time series; fault prediction.

Received: August, 2013 | Accepted: November, 2013.

1 Introduction

Pattern recognition from data is a potential alternative for the extraction of knowledge or the "personality" of a process and it may be useful for predicting failures, control and support decision making, among others. One way to achieve this is to use models based on Artificial Intelligence, such as Fuzzy Inference Systems (FIS). A FIS aims to emulate human reasoning by employing logical sentences provided by an expert or by data analysis (CINTRA, 2008).

Advances in 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 Databases (KDD) and Data Mining (DM) that use data to obtain useful knowledge, adding value to businesses in strategic sectors such as ecommerce, medicine and the economy (FAYYAD, 1996; PIATETSKY, 2007). In industrial processes, improvements in instrumentation and automation technology has meant greater availability of data in production plants. While these technologies for data acquisition have been consolidated, the analysis of this information with the appropriate knowledge generation is still an active field of research. This scenario encourages the use of data-based approaches for the development of systems based on knowledge extracted from data of process variables.

There are many works on pattern recognition in univariate time series (LIAO, 2005) and (KEOG, 2002) as well as the standard approaches (LIAO, 2005) and the feasibility to deal with this kind of problem using point-prototype clustering models (BEZDEK, 2005). However, pattern recognition in multivariate time series represents a more complex problem (non-point prototyping problem) with intrinsic features (SINGHAL, 2005). This kind of problem cannot be solved using directly 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 similarity metrics among objects.

Some works cope with clustering and pattern recognition using uni and multivariate time series together with applications. Liao (LIAO, 2005), Keog and Kasetty (KEOG, 2002) and Fu (FU, 2011) present good reviews involving clustering, classification, segmentation and pattern recognition on univariate and multivariate time series (MTS) with many features of extraction approaches such as Principal Component Analysis (PCA), correlation analysis, wavelets and others. 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. Singhal and Seborg (SINGHAL, 2002) and Yang and Shahabi (YANG, 2004) highlight the use of PCA as a similarity metric in pattern recognition evolving MTS and they present case studies associated to nonlinear dynamic systems (batch fermentation in nonisothermal continuous stirred tank reactor). Vlachos (VLACHOS, 2002) presents an application of clustering MTS using the longest common subsequence as a similarity metric to discover multidimensional trajectories. D'Urso and Mahalaj (DURSO, 2009) use autocorrelation functions. Vlachos (2003) and D'Urso and Mahalaj (DURSO, 2012) use wavelets for the extraction of features. A review of recent works such as D'Urso and Mahalaj (DURSO, 2012) suggests that clustering, pattern recognition and feature extraction associated to multivariate time series have also been studied recently and its application in industrial process (fault detection, control tuning, optimal control, among others) may pose challenges but also knowledge discovery.

Additionally, once the step of recognition patterns and groups has been carried out, the design of classifiers using FIS (fuzzy inference systems) implies in the development of structure and specific approaches to the problem of classification of multivariate time series. Some challenges associated to the design of Fuzzy rule-based classifiers include the generation of membership functions, rule extraction, the structure and kind of fuzzy inference system and the specific approaches for multivariate time series.

The classifier design is a final step in pattern recognition. Generally speaking, a classifier is any function capable of providing the degrees of membership of an object to all clusters. In other words, it is a model that classifies an object according to the classes (clusters) recognized (BEZDEK, 2005). Delgado et al. (1997) present different approaches to use clustering results obtained in point prototype or non-point prototype partitions to extract pieces of a fuzzy rule-based classifier (rules, membership functions). The approaches can be characterized by two strategies, namely the integrating and non-integrated. The former copes with the integrated input space and multivariable membership functions must be estimated based on the clustering results. The second considers the input space separately through the projections of the fuzzy clusters in each input domain. Andrade et al. (2011) present a fuzzy rule-based classifier for fault detection and diagnosis in an alternative gas compressor which is part of a gas processing unit. The authors developed a Type-2 fuzzy inference system to classify univariate time series according to each pattern recognized.

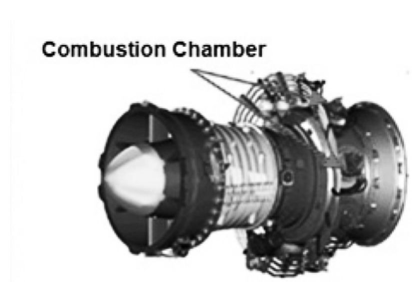
This paper presents the development and implementation of a fuzzy inference system for fault prediction in a gas turbine of commercial scale. This turbine is the main part of a thermoelectric unit belonging to an industrial park belonging to the Brazilian Oil Company. The FIS is based on a preliminary step that comprised the pattern recognition from data considering the cases with and without failure during the starting of the turbine. The patterns were recognized through clustering of Multivariate Time Series obtained from the Plant Information Management System (PIMS) (CARVALHO, 2003). The following sections present the case study and the whole method including the modeling of the FIS.

2 Case study

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

The TPP-RA has three gas turbines (GT) Rolls Royce RB211-G62 DF (ROLLS-ROYCE, 2010), each coupled to a 27 MW electric gener-

(a)



(b)

Figure 1: (a) UTE-RA (SÁ BARRETO, 2009) and (b) Gas turbine RB211-G62 DF (ROLLS-ROYCE, 2010)

ator and the plant has a total generating capacity of 137 MW together with the production of 260.3 t/h of steam. Trips may occur in the turbines (SARAVANAMUTTOO, 1996) and can be caused by factors such as surge, vibration and temperature dispersion. This work focused on the prediction of failures during the starting of the turbine caused by temperature dispersion, based on patterns previously recognized. This trip occurs when the temperature of one of the 17 temperature sensors, distributed radially around the combustion chamber, reaches a temperature of ± 150°C different from the average of the others (ROLLS-ROYCE, 2010).

3 Classifier design method

The modeling of the fuzzy system is part of a broader method composed of two phases (Figure 2). The first phase comprised the acquisition of knowledge, represented in the form of operation patterns, from time series obtained from the Plant Information Management System (PIMS) of the thermoelectric unit.

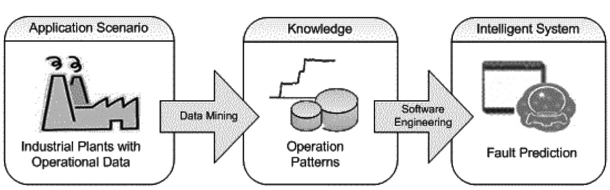


Figure 2: Two phases of the method

3.1 Pattern Recognition in Multivariate Time Series for Fault Prediction

The first phase of the method is illustrated in Figure 3. The main process variables associated with the turbine operation comprised the flow of natural gas, the inlet temperature of the natural gas and the temperature of the exhaust gases. The pattern recognition comprised the following steps:

1. Generation of samples. Occurrences of normal starts and starts with trip caused by temperature dispersion were chronologically identified through operation reports during the period 2008 to 2010. Two groups were established, one group with 18 starts with trip and another group with 57 occurrences of normal starts. This step was supported by the linear scan algorithm (VLACHOS, 2005) which enabled the automatic capture of samples of start events of the turbine directly in the database.

2. Analysis of similarity within and among groups. This step involved the quantification of the level of homogeneity in each group generated (starts with and without trip) and the distinction between them. The calculation of similarity between multivariate time series (WANG, 2007; SINGHAL, 2005; YANG, 2004) was used instead of the common approaches applied in univariate cases (LIAO, 2005; VLACHOS, 2005) in which the Euclidean distance, for example, can be used directly. In this paper the SPCA (Similarity Factor Principal Component Analysis) (YANG, 2004) was used as a similarity metric which provides a dimensionless index determined from the angles between the principal components of each object.

3. Clustering and pattern recognition. Inspired in the Fuzzy C-Means (LIAO, 2005; GAN, 2007; HOPPNER, 1999), an algorithm adapted for the clustering of multivariate time series and also based on optimization was developed. This algorithm created different groups of normal starting of the turbine, the degree of membership of each object to each of these groups and the centers or patterns of each group.

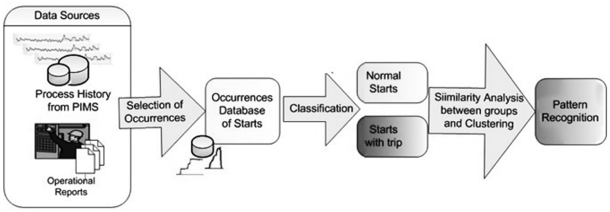


Figure 3: First phase of the method

Two patterns associated to normal starts were recognized (N1 and N2) and one pattern was recognized for the starts with trip (T1). The similarities, quantified by SPCA indicator, enable a polygon (in this case a triangle) to be established that identifies the applicability of each pattern in the fuzzy system configuration. The similarity between N1 and T1 has the value of 0.987, while between T1 and N2 the similarity has the value of 0.956 and between N2 and N1, 0.904 (Figure 4). Therefore, a startup of the turbine close to pattern N2 should be considered safe operation with a lower probability of failure while a startup close to N1 and T1 (and hence distant N2) may be considered unsafe operation, more susceptible to failure. Pattern N2 was adopted as a reference pattern for predicting faults.

The method applied in this work was capable of recognizing unsafe pattern associated to the startup of the turbine. The recognition of a safe pattern (desirable way to start the turbine) represents a potential result and improves the knowledge about this piece of equipment considering that the phenomena related to the trip caused by temperature dispersion are really poorly understood both by operational area as well as by the equipment manufacturer. This highlights one of the contributions of this work and its potential industrial application.

3.2 Fuzzy Classifier Implementation for Fault Prediction

The rationale used to predict failures is described in Figure 5. The key concept is the comparison (similarity or distance and approximation) of a sample in real time operation with the pattern identified as reference (safe pattern N2). If the start, in real time, is close to the safe pattern (N2) and approaching this, the probability of failure will therefore be considered low. Otherwise, it is assumed that there is a higher probability of failure or at least of the turbine not starting safely.

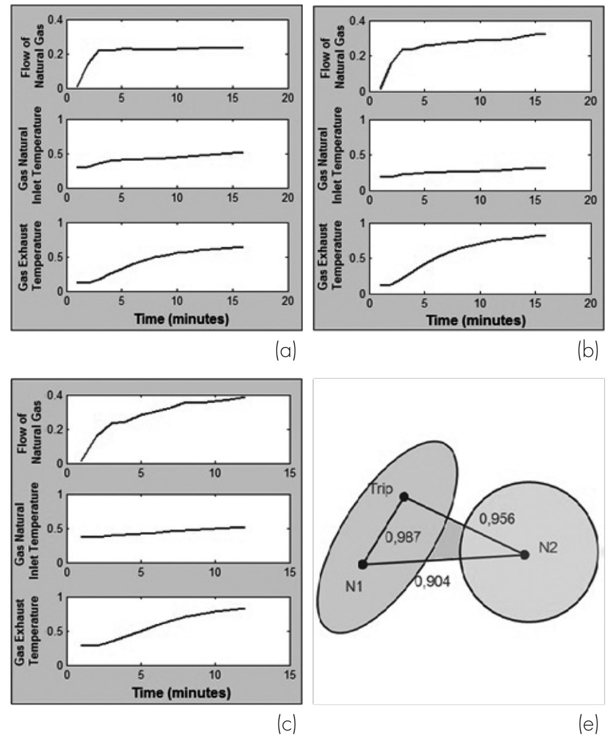


Figure 4: (a) and (b) patterns of normal starts (N1 and N2); (c) pattern of trip (T1) and (d) Triangle of distances between patterns

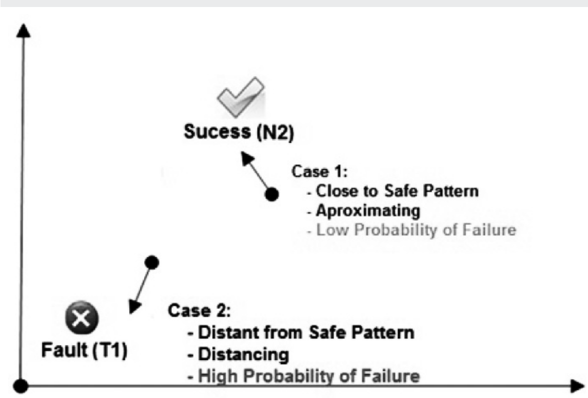


Figure 5: The fuzzy inference system

The architecture of the FIS is presented in Figure 6. The two antecedents of the fuzzy system comprise the distance (SPCA) between the real time "window" and the reference pattern and the closeness between them. These two inputs are the antecedents of the fuzzy rule-based classifier and the consequent comprises the probability of failure of the gas turbine at a given moment in time.

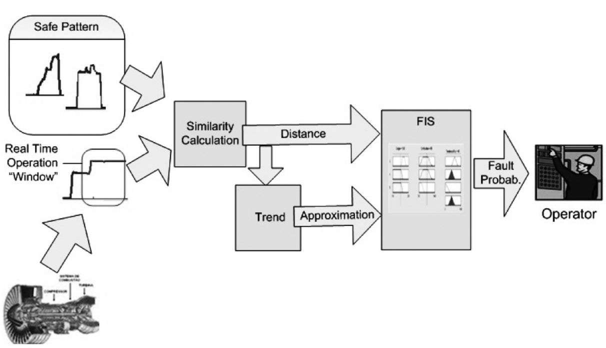


Figure 6: Fuzzy System Architecture

One of the normal start patterns (N2) was chosen as reference pattern because it was more distinct than the other patterns and represents a suitable pattern for safe start.

The antecedents of the FIS were configured by a clustering process commonly used in data-based approach for fuzzy systems. Based on the reference pattern (N2), samples of starting the turbine were collected and the distance and closeness speed between each object and the reference pattern were obtained. Given the dynamic nature of this analysis, the distances and speeds for each object at every minute considering a total window of 16 minutes were obtained. A subset of these variables was selected as the basis to define the fuzzy sets of the antecedents. The fuzzy sets and respective membership functions were obtained using the Fuzzy C-Means clustering method. This analysis led to four clusters (four fuzzy sets) associated to the distance ("Very distant", "Distant", "Close" and "Very close") and three clusters associated to the closeness ("Distancing", "Stopped" and "Approximating"). Figure 7 shows the membership functions associated to the fuzzy sets related to the linguistic variable distance.

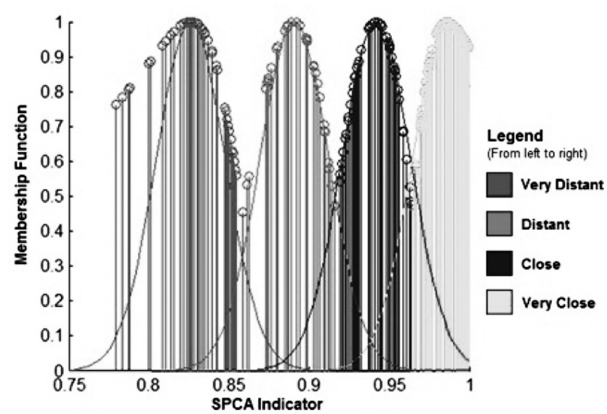


Figure 7: Membership functions – linguistic variable distance

The rules of the FIS were obtained from an analysis of the objects in a crossing perspective involving the antecedents. In order to define the consequent (probability of failure) in each rule, Figure 8 presents an analysis of quadrants obtained from each combination among the fuzzy sets of antecedents. In this particular case the distances and approach speed are related to the objects in a "window" of 8 minutes. The analysis of the rules involves the observation of each quadrant that is just the intersection of these variables. For example, the quadrant marked in Figure 8 represents objects whose the distances are "Very Close" and the approach speed are "Stopped", which is the case in which the object is very close and stationary with respect to the pattern of safe operation (N2). Furthermore, this quadrant presents much more objects associated to the cluster N2 and suggests the rule: "If object is **Very close** and **Stopped** then the probability of failure is **Low**".

The results for the failure probability were established empirically considering three triangular fuzzy sets (low, medium and high) around 20%, 50% and 80%, respectively.

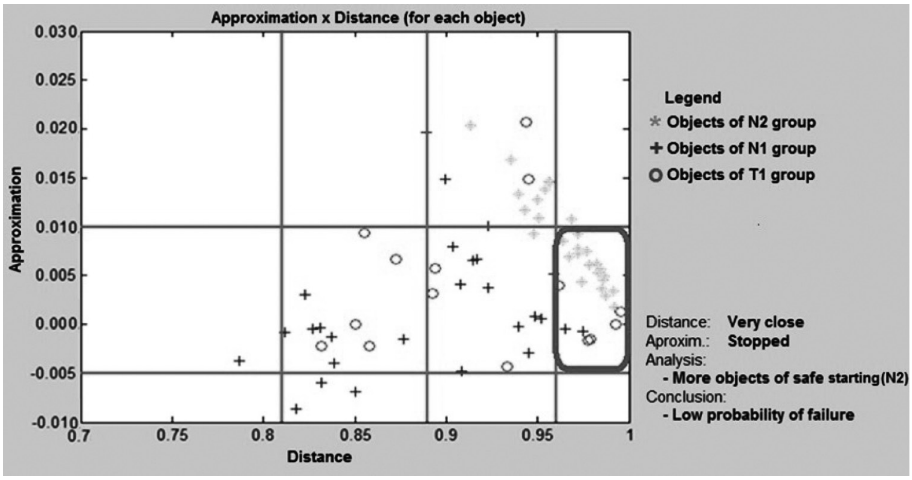


Figure 8: Quadrant analysis for the rule definition

3.3 Clustering of Multivariate Time Series – modified FCM

Fuzzy C-Means (FCM) is a well-known method belonging to the C-Means families of batch clustering models (Bezdek et al., 2005, Bensaid et al., 1996), suitable for clustering objects represented by time series (Liao, 2005). The FCM method is made up of the following optimization problem:

$$\text{Min } J_m(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m \|x_k - v_i\|^2 \tag{1}$$

Subject to $\sum_{i=1}^{c} u_{ik} = 1 \quad \forall k$

Where c represents the number of clusters and n is the number of objects. μ_{ik} $\left(u_{ik} \in [0,1],\ k = 1,\dots,n \quad i = 1,\dots,c\right)$ is the degree of membership of the k-th object (x_k) to the i-th cluster and v_i is the prototype (or center) of the i-th cluster. m ($m \geq 1$) is the degree of fuzzification.

The optimization problem presented by equation (1) has an analytical solution considering point-prototype clustering in which the Euclidian Distance can be used as a metric of similarity among objects (Liao, 2005). In this case, general expressions for v_i and μ_{ik} are obtained applying first order necessary conditions to the optimization problem for both probabilistic and possibilistic approaches. These general expressions and the general algorithm used for the FCM method cannot be applied to the clustering of multivariate time series (non point-prototype problem). In this work a modified version of FCM is developed. This approach comprises the numerical resolution of the optimization problem (1) using the SPCA index to measure the distance (norm) between each object and the prototype (center of the clusters). At each iteration, the degrees of membership $\left(u_{ik} \in [0,1],\ k = 1,\dots,n \quad i = 1,\dots,c\right)$ are updated using one of the equations derived from the application of first order necessary conditions:

$$u_{ik} = \left(\frac{1}{\|x_k - v_i\|^2}\right)^{\frac{1}{m-1}} \Big/ \sum_{j=1}^{c}\left(\frac{1}{\|x_k - v_j\|^2}\right)^{\frac{1}{m-1}} \tag{2}$$

The SPCA index is based on the principal components for each series that comprise the object. In this work, the first k principal components whose variances represent 95% of the total variance are chosen. Then, the algorithm calculates the similarity between these k first principal components and the definition of the index is presented as follows (SINGHAL, 2005) :

$$S_{PCA}(A,B) = trace(LM^T ML^T) = \sum_{i=1}^{k} \sum_{j=1}^{k} \cos^2 \theta_{ij} \tag{3}$$

A and B are original matrices that represents the multivariate time series, L and M are the matrices that contain the first k principal components of A and B. θ_{ij} is the angle between the i-th principal component of A and the j-th principal component of B. The value of the SPCA index is within the range [0,1]. The closer the unit, the greater the similarity between the objects.

4 Results

The data base used to define the rule-base included 75 objects (28, 29 and 18 objects belonging to clusters N2, N1 and T1, respectively). The 12 rules of the FIS are shown in Table 1.

Table 1: Fuzzy Inference Rules derived from data analysis

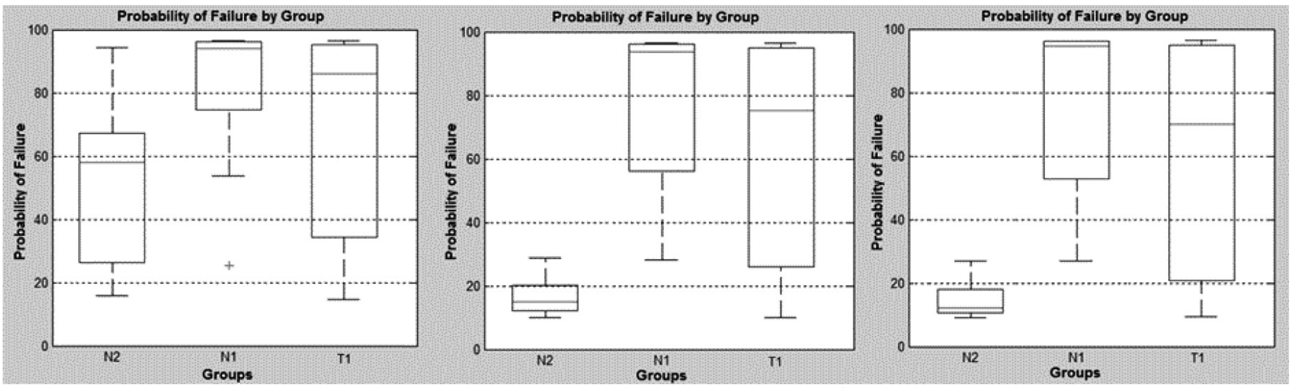
Distance	Approximation	Fault Probability
Very Distant	Distancing	High
Very Distant	Stopped	High
Very Distant	Approximating	High
Distant	Distancing	Medium
Distant	Stopped	High
Distant	Approximating	High
Close	Distancing	Low
Close	Stopped	Medium
Close	Approximating	High
Very Close	Distancing	Low
Very Close	Stopped	Medium
Very Close	Approximating	Low

Two samples (training and test) were collected in order to evaluate the performance of the fuzzy classifier. The training and test data comprised objects (time series) from 2008 to 2010 and only 2011, respectively. The former was used to adjust some parameters of the fuzzy model (membership functions).

Figure 9 presents three box-plots with the distribution of the failure probability of objects belonging to each cluster using the training data considering the instants 5, 8 and 11 minutes. The results suggest that the time period of 5 minutes is not enough or suitable to identify or predict failure for the objects in their groups. On the other hand, for the windows of 8 and 11 minutes a clear distinction between the safe starts (N2) and the other two (N1 and T1) can be seen. Figure 10 presents the box-plots using the test data and confirms the results shown in Figure 9, especially regarding the behavior associated to pattern N2.

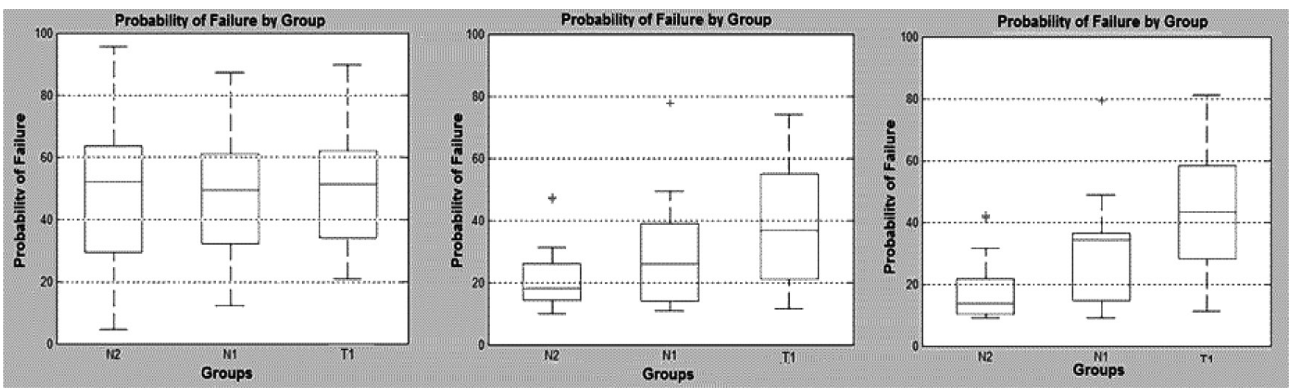
Figure 11 presents the dynamic behavior of the failure probability (training data) considering three objects in each cluster. Despite the sharp rise in the first 5 minutes, the failure probability in the objects belonging to cluster N2 decreases to low values (around 20%) indicating fewer failures due to the trip caused by temperature dispersion. For the objects belonging to cluster T1 the probability remains at higher levels (above 60%). The same behavior can also be seen in Figure 12 (test data) showing the generalization capability of the fuzzy classifier.

The results, based on the real data from the gas turbine, show the efficient and ability of the fuzzy rule-based classifier to predict trips caused by temperature dispersion. Considering the intrinsic difficulty and lack of knowledge about the phenomenon that causes the temperature dispersion in the combustion chamber, the proposed fuzzy model is a potential tool to support decision-making at operational level. Furthermore, the safe pattern recognized can be useful for optimal control purposes providing the desired trajectory of the process variables (flow of natural gas, inlet temperature of the natural gas and temperature of the exhaust gases) to be tracked during the startup of the turbine.



(a) (b) (c)

Figure 9: Box-plot analysis of similarity of the objects in their groups in (a) 5 minutes, (b) 8 minutes and (c) 11 minutes (training data)



(a) (b) (c)

Figure 10: Box-plot analysis of similarity of the objects in their groups in (a) 5 minutes, (b) 8 minutes and (c) 11 minutes (test data)

5 Conclusions

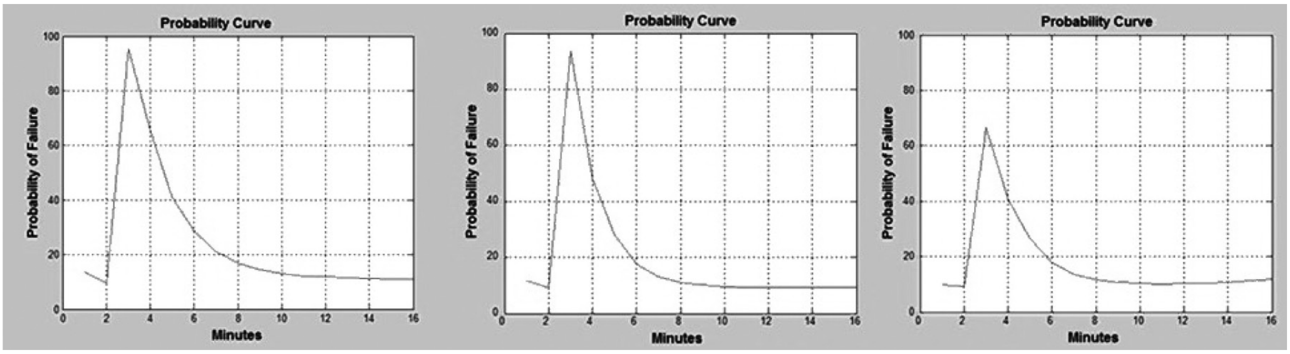
This paper presents the development and implementation of a fuzzy inference system for the dynamic prediction of failure in a commercial scale gas turbine, based on operating patterns recognized directly from data from the Plant Information Management System. An extensive procedure is applied in order to model the classifier (rule-base) considering as antecedents the distance and the approximation of the starting curve from a previously recognized safe pattern.

The results obtained using two sample data sets (training and testing) show the potential of the fuzzy classifier in the prediction of failure in the turbine as well as the use of this tool in real time operation. The fuzzy rule-based classifier also provides a useful and efficient way to support decision making at management level enabling operational maneuvers to avoid stopping the operation of the power station as a whole. One of these could be to alter the turbine in operation.

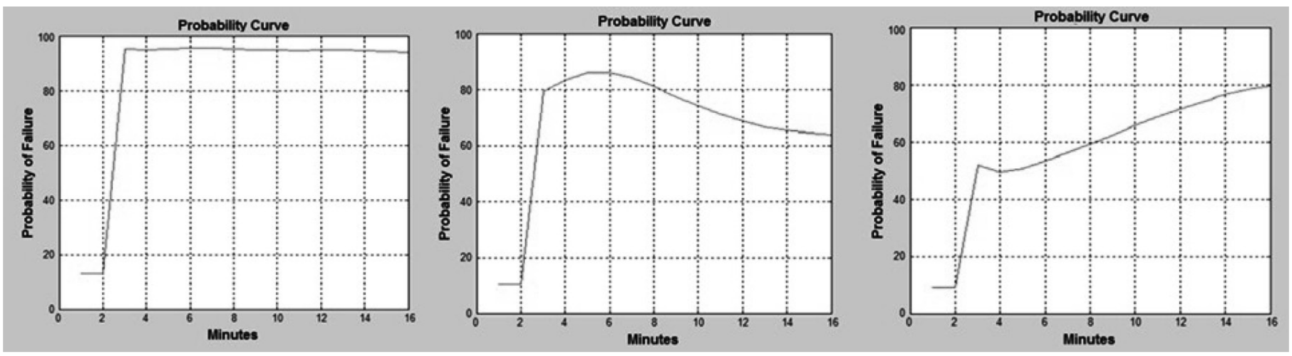
References

ANDRADE, V. E. FONTES, C. H. EMBIRUCU, M. (2011) An Interval Type-2 Fuzzy Logic Approach for Instrument Fault Detection and Diagnosis. Proceedings of the World Congress on Engineering (WCE 2011), Vol. 2.

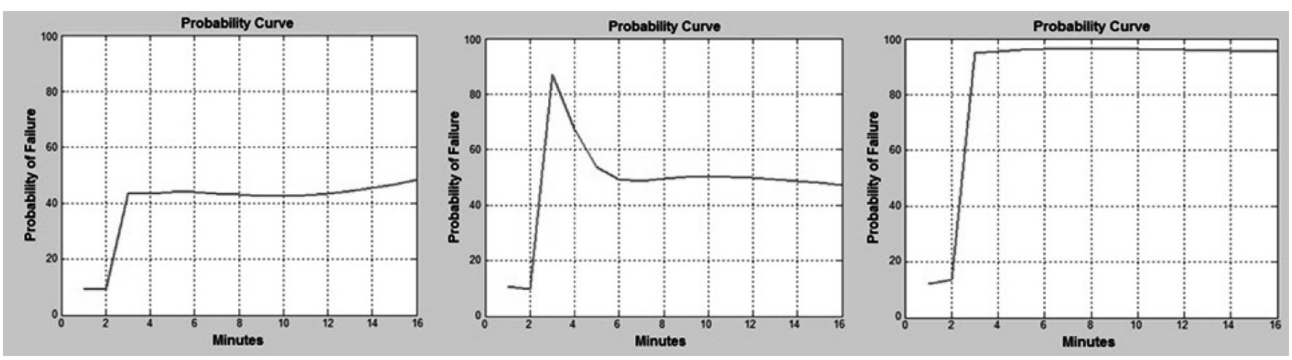
BEZDEK, J. C. (2005) Fuzzy Models and Algorithms for pattern recognition and image processing, Springer.



Samples of objects in group N2



Samples of objects in group T1



Samples of objects in group N1

Figure 11: Probability curves (training data)

CARVALHO, F. B., TORRES, B. S., FONSECA, M. O., SEIXAS FILHO, C. (2003) Sistemas PIMS – Conceituação, Usos e Benefícios. VII Seminário de Automação de Processos da Associação Brasileira de Metalurgia e Materiais – ABM, Santos/SP.

CINTRA, M. E. CAMARGO, H. A. MONARD, M. C. (2008). A study on techniques for the automatic generation of membership functions for pattern recognition. In Congresso da Academia Trinacional de Ciências (C3N) (Vol. 1, pp. 1-10).

DELGADO, M. GOMEZ-SKARMETA, A. F. MARTIN, F. (1997) A Fuzzy Clustering-Based Rapid Prototyping for Fuzzy rule-Based Modeling. IEEE Transactions on Fuzzy Systems (Vol. 5, no. 2, pp. 223-233)

D'URSO, P. MAHARAJ, E. A. (2009) Autocorrelation-based fuzzy clustering of time series, Fuzzy Sets and Systems, 160, pp. 3565–3589

D'URSO, P. MAHARAJ, E. A. (2012) Wavelets-based clustering of multivariate time series Fuzzy Sets Syst, 193,33 –61.

FAYYAD, U. M. PIATETSKY, G. SMYTH, P. UTHURUSAMY, R. (1996) Advances in Knowledge Discovery and Data Mining. California: American Association for Artificial Intelligence e MIT Press.

FU, T. (2011) A review on time series data mining. Engineering Applications of Artificial Intelligence (2011) 164-161.

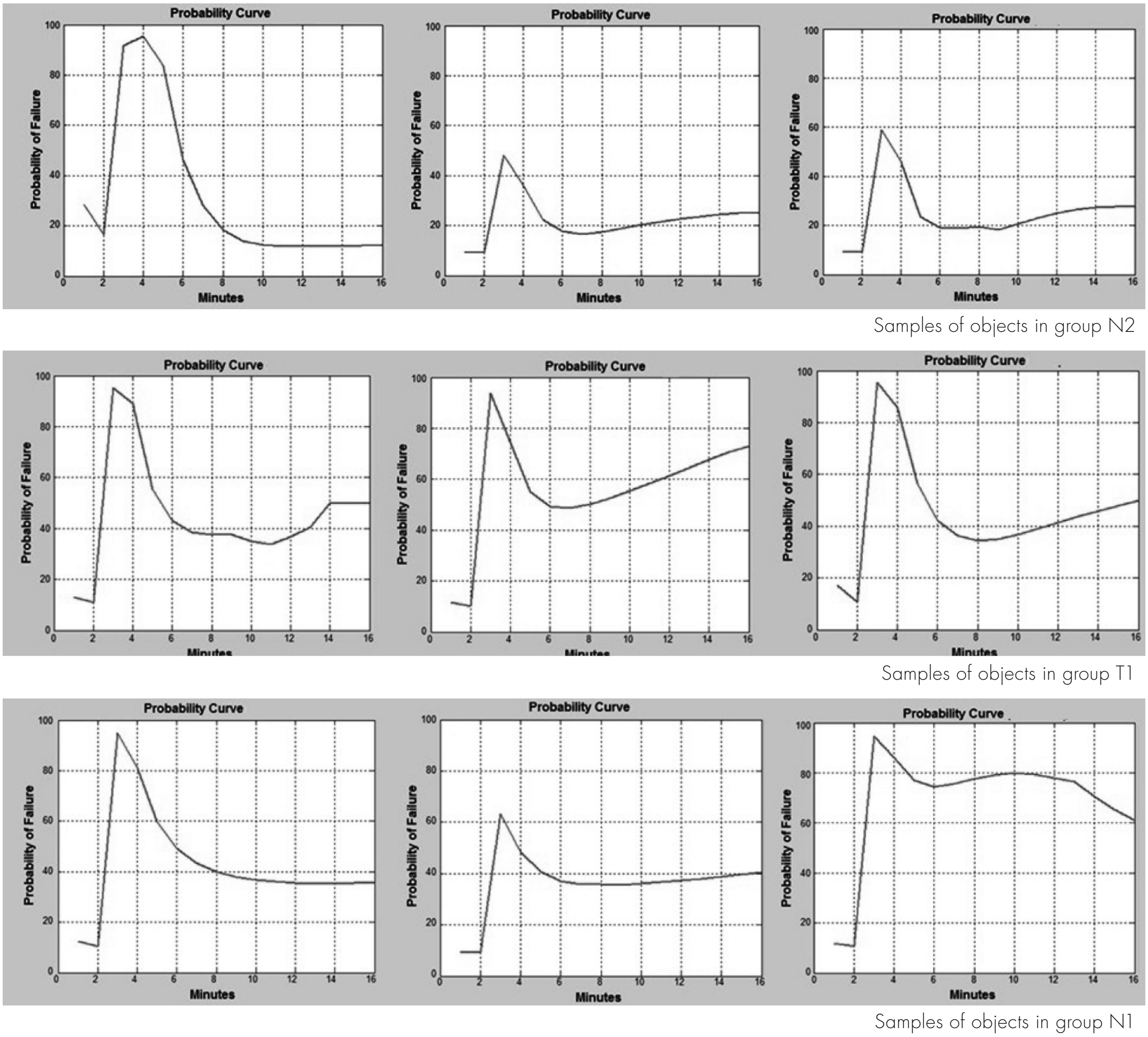


Figure 12: Probability curves (test data)

GAN, G. MA, C. WU, J. (2007) Data clustering : theory, algorithms, and applications. ASA-SIAM series on statistics and applied probability.

HOPPNER, F. KLAWONN, R. KRUSE, R. RUNKER, T. (1999) Fuzzy Cluster Analysis. Wiley, Chichester.

KEOG, E. J. KASETTY, S. (2002) On the need for time series data mining benchmarks: a survey and empirical demonstration, in: 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26.

LIAO, T. W. CELMINS, A. K. ROBERT, J. HAMMELL, I. (2001). A fuzzy c-means variant for the generation of fuzzy term sets. Fuzzy Sets and Systems, 135(2):241–257.

LIAO, T. W. (2005) Clustering of time series data - a survey, Science Direct. Pattern Recognition 38, p. 1857 – 1874.

MANLEY-COOKE, P. RAZAZ, M. (2007). An efficient approach for reduction of membership functions and rules in fuzzy systems. Fuzzy Systems Conference - FUZZ-IEEE,1:1609–1613.

PIATETSKY, G. (2007). Data mining and knowledge discovery 1996 to 2005: overcoming the hype and moving from “university” to “business” and “analystics”, Data Mining Knowledge Discovery, 15, 99-105.

ROLLS-ROYCE. (2010) Material de Treinamento do Conjunto RB 211-G62 DF.

SÁ BARRETO, S. T. (2009) Desenvolvimento de Metodologia para Atualização em Tempo Real de Modelos Matemáticos de Processos Decisórios. 2009. Dissertação (Mestrado em Mecatrônica) – PPGM, Universidade Federal da Bahia, Salvador.

SARAVANAMUTTOO, H. I. H.; ROGERS, G. F. C.; COHEN, H. (1996) Gas Turbine Theory. 5.ed. Dorchester: Prentice Hall.

SINGHAL, A. SEBORG, D.E. (2002) Pattern Matching in Multivariate Time Series Databases Using a Moving-Window Approach, Industrial and Engineering Chemistry Research, 41, 3822-3838.

SINGHAL, A. SEBORG, D.E. (2005) Clustering Multivariate Time-series Data, Journal of Chemometrics, 19, 427-438.

VLACHOS, M. KOLLIOS, G. GUNOPOLOS, G. (2002) Discovering Similar Multidimensional Trajectories, in Proc. Of the 8th ICDE, San Jose, CA, 2002, pp. 673–684.

VLACHOS, M. LIN, J. KEOGH, E. GUNOPOLOS, D. (2003) A Wavelet-Based Anytime Algorithm for K-Means Clustering of Time Series. In Proc. Workshop on Clustering High Dimensionality Data and Its Applications, pp.23-30.

VLACHOS, M. (2005) A practical Time-Series Tutorial with MATLAB. European Conference on Machine Learning, 2005.

WANG, X. WIRTH, A. WANG, L. (2007) Structure-based statistical features and multivariate time series clustering, Seventh IEEE Conference on Data Mining, 351-360.

YANG, K. SHAHABI, C. (2004) A PCA-based Similarity Measure for Multivariate Time Series. Second ACM International Workshop on Multimedia Databases, ACM-MMDB 2004, Washington, DC, USA, November 13, 2004.