Unsupervised Methods for Anomalies Detection through Intelligent Monitoring Systems

Alberto Carrascal[1], Alberto Díez[1], Ander Azpeitia[1]

[1] Fundación Fatronik-Tecnalia, Paseo Mikeletegi 7, Parque Tecnológico, 20009 Donostia, Spain

Abstract. The success of intelligent diagnosis systems normally depends on the knowledge about the failures present on monitored systems. This knowledge can be modelled in several ways, such as by means of rules or probabilistic models. These models are validated by checking the system output fit to the input in a supervised way. However, when there is no such knowledge or when it is hard to obtain a model of it, it is alternatively possible to use an unsupervised method to detect anomalies and failures. Different unsupervised methods (HCL, K-Means, SOM) have been used in present work to identify abnormal behaviours on the system being monitored. This approach has been tested into a real-world monitored system related to the railway domain, and the results show how it is possible to successfully identify new abnormal system behaviours beyond those previously modelled well-known problems.

Keywords: Unsupervised Anomaly Detection, Unsupervised Classification, Intelligent Monitoring Systems, Clustering.

1 Introduction

Because of the recent technological revolution occurred in industrial sector, it turns increasingly difficult to raise any appropriate manual maintenance process. Thus, the amount of information about the state of the system is being monitored is continuously increasing, exceeding the capacity of maintenance technicians. While the industry is undergoing a technological revolution, new reactive, proactive and predictive maintenance approaches are being developed.

The success of the majority of the monitoring and intelligent diagnosis systems relies on the use of the knowledge regarding existing domains (Knowledge Based Systems, or KBS) [1]. In this kind of domains, the main difficulty regarding failure and anomaly detection is how to make the expert knowledge explicit and how to model it. Knowledge modelling based on rules is one of the most common approaches [2]. Nevertheless, there exist domains where this approach can not be applied, due to either, non-existing previous expert knowledge or overly complex knowledge base management [3].

Supervised learning models do not successfully resolve this problem as they require previous knowledge about which should be the system output when new data

come in, and also a high external support will be needed [4]. On the contrary, unsupervised learning models classify monitored system data by means of some similarity measure without any external support. Failure detection is achieved by comparing and identifying new cases with past breakdowns, whilst anomalies are detected whenever there is no mapping with any previous case. Therefore, the problem of identifying failures and anomalies can be transformed into an unsupervised classification problem [5].

The monitoring domain presented in this work concerns the railway domain. It is an especially critical domain in which is really important to assure the safety for every journey, for both passengers and cargo; which implies that all the components embedded into the train accomplish some reliability standards. In such domain, an exhaustive control of life cycle parameters of train components has to be carried out, guaranteeing correct operation working for all of them throughout their service life-time.

2 Unsupervised Methods

Unsupervised methods have been used in many contexts and domains, involving different unsupervised learning problems. The main goal of these techniques is to perform a clustering of similar datasets, which are supposed to have the same pattern. Such pattern could be very significant in order to classify or to identify behaviours linked to the data, or in order to detect or to infer possible failures or anomalous conditions; different from supervised learning (and reinforcement learning), where the learner is only provided with unlabelled examples.

In the study presented in this paper, the performance of different techniques have been tested, to illustrate the differences between them and to analyze their behavior in a real monitoring system. It is hardly important to underline that, methods selected for this study (HCL, K-Means and SOM) are a representative subset of the unsupervised classification approaches.

HCL (Hierarchical Clustering) is an algorithm that builds clusters iteratively in a hierarchical structure. The iterative process can be either agglomerative or divisive. Normally, agglomerative strategy is more commonly used [6]. Differences between methods arise because of the different ways of defining similarity (distance) between clusters. Several agglomerative techniques such as complete linkage, single linkage, average linkage, weighted average linkage, median linkage, centroid linkage and Ward's linkage are possible.

HCL graphic results (dendograms) are complex and confusing when the amount of data is considerable. This fact, together with the fact that the complexity of this algorithm is on the order of $O(n^2)$, depending on the configuration, makes difficult to employ this technique when the amount of data is over thousands of samples. This is the main reason why in those cases it is more advisable to apply other approaches that are easier to interpret and less computationally expensive, such as K-Means and SOM [7], [8].

K-Means classification algorithm performs a partition of data space into *k* clusters. Each cluster is represented by an element, the centroid or a mean point, whose initial

value can be randomly set or estimated by applying some kind of heuristic. In an iterative process, the elements are assigned to the partition with the least distance between them and the centroid of the partition. After elements assignation into clusters, cluster centroids are recalculated with the elements that belong to its partition. The process converges to a solution with a linear complexity *O(n)*, which is not always the global optimum. The success of K-Means approach is strongly determined by the choice of the *k* value, the metric employed and the initial centroids values.

The Self-Organizing Maps (SOM) approach allows representing into a low-dimensional map a high-dimensional data set, so that the similarity between analyzed data can be easily identified [9]. SOM map is composed by neurons grouped according to a topology (hexagonal and rectangular topologies are the most common ones). Each neuron has associated a weighted vector that allows mapping entry data into each neuron on the basis of a given measure. To achieve this, the data are presented to the network iteratively, so that at each time step, the winning neuron, or the neuron that has associated the weighted vector most similar to the sample, modifies its associated vector to increase its similarity with given data. The vector associated to the winning neuron, and to the neighbouring neuron vectors according to the topology used, is modified by means of a decreasing function of the distance between nodes on the map grid. Neighbourhood functions most commonly used are Gaussian and Bubble functions.

SOM provides a non-linear, ordered, smooth classification of high-dimensional input data, preserving neighbourhood relations. This capacity for managing high dimensional data with good results and performance, makes this approach possible to be applied in many complex domains, such us engineering, bioinformatics and genetics, communications, etc. [10], [11], [12].

3 Monitored System

Monitored system based on intelligent diagnostics that has been used for present study, is related to one of the most critique last generation train component, made by CAF [13] company: self-propelled, dual voltage electric train units with a variable gauge system (ATPRD). As safety measure, ATPRD incorporate ATMS (Acceleration and Temperature Monitoring System) equipment, developed by CAF; which allows knowing temperatures and accelerations at any time inside the train motion units, called bogies. The importance of these component measurements is critical, since the failures that can occur on the trains are mainly associated to anomalous behaviours inside the bogies.

There are several sensors to monitor the acceleration and temperature of the bogies, strategically replicated and placed over the train. Every 5 minutes during a train journey, sensors acquire readings of those parameters which are forwarded by means of a GSM connection with the train. Such information is registered and stored in a database to provide the needed data input to our approach.

Sensors are distributed in 8 bogies per train, as it is showed in Figure 1. Each bogie has 32 sensors which can be divided in five groups: internal and external wheel

groups with 16 temperature sensors installed on the wheels (4 per wheel), cylindrical and conic hollow shaft groups with 8 temperature sensors installed on the hollow shafts (HS hereafter), and finally, reduction gear group with 5 temperature sensors.

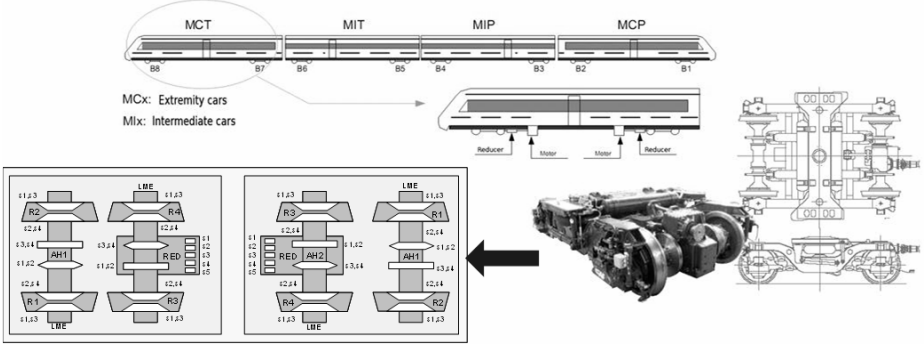

Fig. 1. Bogies and sensors distribution in the ATPRD unit.

3.1. Derived variables

In order to easily identify journey anomalies, a model representation that easily allows comparing the main characteristics of different journeys as simple data vectors was needed. To that, due to the lack of any thermodynamic model of the train, a set of quantitative variables has been defined to characterize the bogies behaviour each journey.

Derived parameters also eliminate temporal dependency of collected sensor data; so, for each journey, each derived parameter has an associated value. This representation also allows comparing journeys of different duration, since each journey is identified by a vector of values of constant dimension. Derived parameters used in this study are the following:

- Correlation of a sensor with its pair: Pair sensors are sensors physically located very close, so that the correlation between their readings is used to validate its correct operation.
- Volatility of a sensor: this variable measures the variability of a certain sensor as the absolute differences mean.
- Sensors mean group.
- Mean square of sensors group: the mean of a group of sensors squared.
- Maximum absolute value of a subgroup of sensors: this derived parameter allows identifying atypical values measured by the sensors.
- Percentiles (10, 20, 80, and 90) of a group of sensors: P_{10} and P_{20} allow identifying low sensor values, whereas P_{80} and P_{90} are intended to identify high sensor values.

Considering the number of sensors and groups the total number of derived parameters obtained is 65. This way, different bogie behaviours are characterized by 65-dimensional real vectors.

4. Results

Test data used in this study have been collected from 12 ATPRD units, monitored during eleven months (from January to October 2008), obtaining a total of 9.100 vectors that represent the behaviour of the different bogies over units monitored. Euclidean distance has been adopted as similarity measure used by every analysed unsupervised classification method. Gaussian normalization has been applied in order to minimize the impact related to the different domains, means and variances over each component vector.

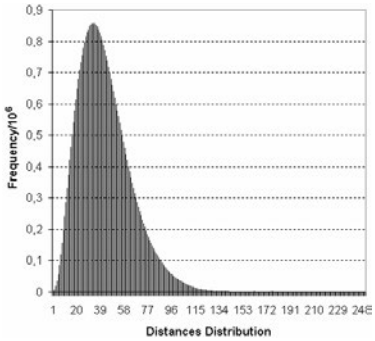

Fig. 2. Whole data distances distribution

A first similarity analysis on existing data shows a high level of regularity regarding the data patterns. This regularity is shown in figure 2 where the whole data distances distribution can be shown. There exists an expected non-stochastic behaviour in the distances distribution, with a clear deviation to little distances. This confirms that rail lines regularity (concerning journey duration, velocity profile, et cetera) is reflected by the regularity of the vectors that represent train bogies behaviours.

Classification of the different bogie behaviours dealing with the similarity of the vectors that represent them has been carried out. HCL method has been applied with an agglomerative strategy and average linkage. As shown in Figure 3, the complexity of the graphic representation of the output obtained by applying this algorithm is an important handicap when interpreting the results. Only the most clearly divergent cases can be easily isolated. A more detailed analysis of marginal cases identified by HCL technique clearly shows a small group of five anomalous bogie behaviours. These cases have been contrasted with the sensor values in those dates, concluding that they are related to anomalous behaviours on different group of sensors: Wheels, HS and Reduction gear group. Figure 4 (A) shows an example of this anomalous behaviour on Wheels group. The graphic illustrates the signal related to train speed (bottom signal), in relation with the signals related to the sensors of current group. The irregular signal (upper signal) that is uncorrelated with the other ones is obviously associated to a malfunction of the sensor that represents. Regarding these five anomalous samples, the calculated derived parameter values are notably different

from the other cases analyzed by the HCL method. As has been checked, the main causes of the sensors failures are wrong connections and water invasion.

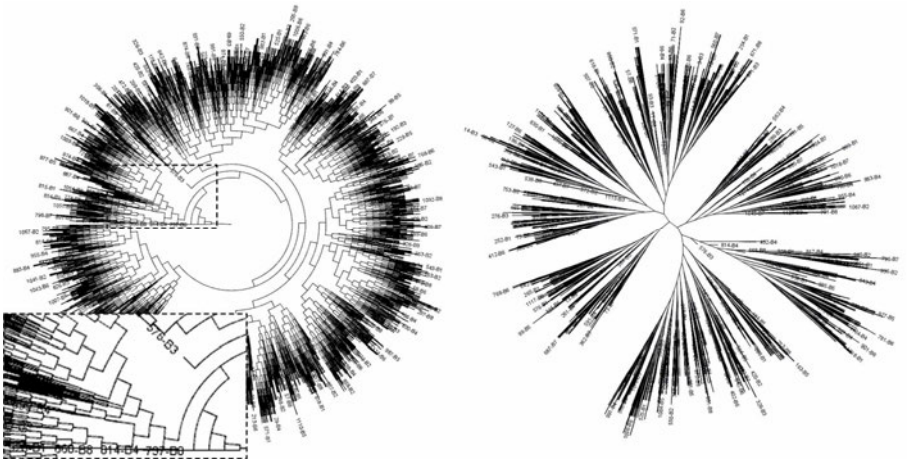

Fig. 3. HCL graphical output

In contrast to HCL approach, unsupervised classification methods, such as K-Means and SOM, allow a more intuitive interpretation of generated clusters. Both techniques require the specification of the maximum number of clusters to be considered. After some experimental tests, this maximum value was fixed to 225 (15x15-dimensional SOM map).

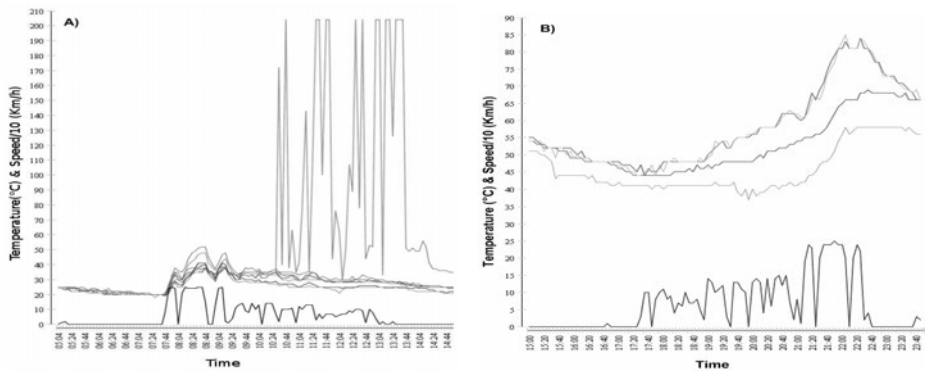

Fig. 4. (**A)** Bogie anomalous behaviour example related to wheels sensors group failure. (**B)** Example of Bogie anomalous temperature profile.

A rectangular topology, Gaussian neighbouring function and a learning rate value of 0.5 were chosen to configure the Kohonen network. In the same way, clusters medium size was determined after ten test executions of each method. This executions rate was considered as enough in order to obtain a reliable clusters size distribution. As illustrated in the figure 5, around 30% of the data is grouped into clusters with size between 100 and 500 elements. The rest of the data is distributed into lower size

clusters, with a more uniform behaviour in the case of K-Means. Small clusters (with a size between 0 and 5), are more frequent in K-Means approach, showing more sensibility to little variations on the data.

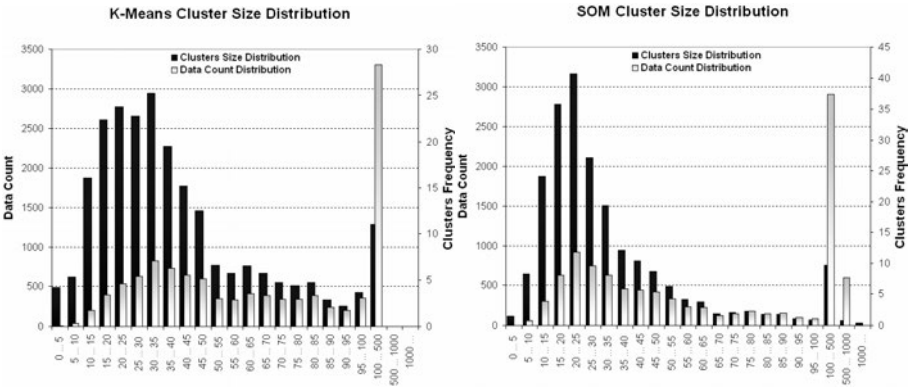

Fig. 5. K-Means and SOM clusters size distribution.

In order to identify failures on bogies behaviour, a more detailed analysis of smallest cluster was performed, realizing that in the case of SOM approach, data classified into small clusters are severely modified with each method execution. Nevertheless, K-Means results are more homogeneous. Anomalous elements identified by HCL technique are all located into lower size clusters obtained by applying K-Means. On the contrary, this clusters distribution was not achieved with SOM.

Regarding small clusters given by SOM or mainly by K-Means, the results have been really interesting, detecting other anomalous situations (see Figure 4.B). These behaviours are related to sensor failures but also to anomalous journeys caused by external conditions such as journey sections with unusual speed, exceptional climatic conditions, non-scheduled stops, etc.

5. Conclusions

In order to address the problem related to the process of anomaly identification in intelligent monitored systems, unsupervised methods are a very useful alternative when there is no previous expert knowledge about the application domain. Tests performed highlighted how anomalous situations of interest were detected by means of classic unsupervised methods: HCL, SOM and K-Means.

The railway sector analyzed in present paper shows a very regular behaviour. A significant set of cases is enough to cover all possible normal behaviours associated to the data. Anomalous behaviours that imply failures can be detected by means of unsupervised classification techniques.

In experiments performed, HCL model shows very good results regarding detection of anomalous cases. However, HCL is computationally expensive and the

interpretation of its results demands increasing effort as the size data grows. SOM model is computationally less expensive and its graphical output is more intuitive. Nevertheless, owing to the input noise tolerance associated to artificial neural networks approaches and to the topology preservation, the obtained classifications show less accuracy level than those obtained when using K-Means approach. Further, the K-Means method improves the detection of slight data variations related to bogies behaviours.

Train bogies anomaly behaviours detected in this paper have been contrasted with maintenance information. From the analysis, these anomalies have been found to match failures in sensors or train journeys strongly influenced by external conditions, which confirms the validity of our approach.

Acknowledgments. Present work has been financed by Fatronik-Tecnalia. The data used for the study has been provided by NEM Solutions and CAF.

References

1. Gonzalez, A.J., Dankell D.D.: Engineering of knowledge-based systems. Prentice Hall (1993)
2. Brachmand, R.J., Levesque, H.J.: Knowledge Representation and Reasoning, MIT Press, Cambridge, MA (2003)
3. Preece, A.D.: Validation of Knowledge-Based Systems: The State-of-the-Art in North America. J. Study of Artificial Intelligence Cognitive Science and Applied Epistomology. 11(4) (1994)
4. Alpaydin, E.: Introduction to Machine Learning. Adaptive Computation and Machine Learning, MIT Press (2004)
5. Eskin, E., Arnold, A., Prerau, M., Portnoy, L., Stolfo, S.: A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data. Applications of Data Mining in Computer Security. LNCS, vol. 6(4), Barbara, Daniel; Jajodia, Sushil (Eds.) (2002)
6. Williams, C.: A MCMC approach to hierarchical mixture modeling, Advances in Neural Information Processing Systems. Vol. 12, pp. 680--686 (2000)
7. Garrett-Mayer, E., Parmigiani, G.: Clustering and Classification Methods for Gene Expression Data Analysis. Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 70. (2004)
8. Yin L., Huang CH., Ni J.: Clustering of gene expression data: performance and similarity analysis. BMC Bioinformatics, Vol. 7(Suppl 4):S19 (2006)
9. Kohonen, T. Self-organizing maps. Springer-Verlag, Berlin (1997)
10. Carrascal, A., Couchet, J., Ferreira, E., Manrique, D.: Anomaly Detection using prior knowledge: application to TCP/IP traffic. Artificial Intelligence in Theory and Practice - IFIP International Federation for Information Processing, vol. 217, pp. 139--148 (2006)
11. Kohonen, T., Oja, E., Simula, O., Visa, A., Kangas, J.: Engineering Applications of the Self-Organizing Map. Proceedings of the IEEE. Vol. 84, Issue 10, pp. 1358- - 1384, (1996)
12. Huang, S., Ward, M. O., Rundensteiner, E. A.: Exploration of dimensionality reduction for text visualization. Technical Report TR-03-14, Worcester Polytechnic Institute, Computer Science Department (2003)
13. Construcciones y Auxiliar de Ferrocarriles, http://www.caf.net/caste/home/index.php