Expert Systems with Applications 38 (2011) 7003–7015

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Anomaly detection in monitoring sensor data for preventive maintenance

Julien Rabatel [a,b,*], Sandra Bringay [a,c], Pascal Poncelet [a]

[a] LIRMM, Université Montpellier 2, CNRS, 161 rue Ada, 34392 Montpellier Cedex 5, France
[b] Fatronik France Tecnalia Cap Omega, Rond-point Benjamin Franklin - CS 39521, 34960 Montpellier, France
[c] Dpt MIAp, Université Montpellier 3, Route de Mende, 34199 Montpellier Cedex 5, France

ARTICLE INFO

Keywords:
Anomaly detection
Behavior characterization
Sequential patterns
Preventive maintenance

ABSTRACT

Today, many industrial companies must face problems raised by maintenance. In particular, the anomaly detection problem is probably one of the most challenging. In this paper we focus on the railway maintenance task and propose to automatically detect anomalies in order to predict in advance potential failures. We first address the problem of characterizing normal behavior. In order to extract interesting patterns, we have developed a method to take into account the contextual criteria associated to railway data (itinerary, weather conditions, etc.). We then measure the compliance of new data, according to extracted knowledge, and provide information about the seriousness and the exact localization of a detected anomaly.

1. Introduction

Today, many industrial companies must face problems raised by maintenance. The most common solution is called curative maintenance, i.e., equipment is replaced or repaired after the appearance of obvious failures, once the damage is occurred. This solution poses many problems. Curative maintenance is too belated and is particularly costly on several aspects. On the financial side first, for many companies, a few hours of downtime can result in millions of dollars in losses. It is generally much less expensive to make predictive maintenance to prevent a serious breakdown. In addition, the corrective maintenance is also a problem for security aspects. In many sensitive areas, equipment failures can cause death. For example, it is estimated that approximately 5% of motor vehicle accidents are caused by equipment malfunction or a lack of maintenance.[1] Another aspect is related to the environment and energy saving. Indeed, equipment that is worn or subject to malfunctions often consumes more energy than equipment that operates optimally. In addition, a systematic maintenance planning is not a satisfactory solution as too expensive compared to real needs. To reduce the problems of equipment maintenance, to propose ways to make maintenance both faster and more effective by anticipating serious breakdowns represents a particularly critical issue.

Such a preventive maintenance consists in detecting anomalous behavior in order to prevent further damages and avoid more costly maintenance operations. To this end, it is necessary to monitor the working equipment. Usually, monitoring data is available through embedded sensors and provides us with important information such as temperatures, humidity rates, etc. Nevertheless, data collected by sensors are difficult to exploit for several reasons. First, a very large amount of data usually available at a rapid rate must be managed to provide a relevant description of the observed behaviors. Furthermore, they contain many errors: sensor data are very noisy and sensors themselves can become defective. Finally, when considering data transmission, very often lots of information are missing.

In this paper, we focus on the field of train maintenance. Trains monitoring is also ensured by sensors positioned on the main components (wheels, motors, etc.) to provide much information (temperature, acceleration, velocity). In this context, we are subject to the difficulties we have described above: voluminous and noisy data, information transmission problems, etc. Moreover, it is important to take into account the different types of data available. We therefore wish to propose a method to exploit this information in order to assist the development of an effective predictive maintenance.

The needs in the context of train maintenance are twofold. First, it is important to provide a better understanding of monitored systems. Indeed, as they are often complex and contain many components, the experts have little knowledge about their actual behavior. This lack of knowledge makes the problem of maintenance very difficult. From another point of view it could be interesting to get an overview of the normal behavior (e.g., in case of monitoring) and then it is necessary to propose a way for characterizing such normal behaviors from a huge amount of historical data. Another challenging point that we must consider is that

* Corresponding author at: LIRMM, Université Montpellier 2, CNRS, 161 rue Ada, 34392 Montpellier Cedex 5, France.

E-mail addresses: rabatel@lirmm.fr (J. Rabatel), bringay@lirmm.fr (S. Bringay), poncelet@lirmm.fr (P. Poncelet).

[1] http://www.smartmotorist.com.

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.12.014

normal behavior strongly depends on the context. For example, a very low ambient temperature will probably affect a train behavior. Similarly, each itinerary with its own characteristics (slopes, turns, etc.) influences a journey. Consequently it is essential, in order to efficiently characterize the behavior of trains as well as to detect anomalies, to consider the surrounding context. In this paper, we will show how these elements can be directly addressed by data mining techniques and how they can be used to design a system for detecting anomalies in train behavior and help experts so that a detected problem can be dealt as promptly as possible, and the right decisions can be made.

Our approach follows the framework presented in Fig. 1. We address the two issues mentioned above: (i) the knowledge discovery process about normal train behavior and (ii) the anomaly detection in new data.

The characterization of normal behavior is divided into three steps. First, we consider the data describing the normal behavior (i.e., containing no anomalies) that were recorded in the past. These so-called historical data are segmented into different classes, which are defined by the context in which the data were recorded. This organization brings together all trips that were conducted under similar conditions. The criteria used to create the classes are, for example, the outside temperature, the itinerary, etc. Then, we extract the most frequent behaviors to characterize each of these classes. We thus obtain knowledge classes that describe very precisely normal train behavior and also provide us with essential information about the impact of contextual criteria. For example, we can answer questions such as *"What are the behaviors that are specific to a high outside temperature?"*.

After having characterized the train behavior in the first step of our framework, we wish to detect anomalies in newly recorded data. To this end, we use the previously obtained knowledge. We have developed a method to compare new monitoring data with a class of knowledge. Thus, we can detect critical events and notify experts that a maintenance operation may be necessary.

This paper is organized as follows. Section 2 describes the data representation in the context of train maintenance. Section 3 shows the characterization of normal behaviors by discovering sequential patterns. Then we present the anomaly detection for predictive maintenance approach in Section 4. Experiments conducted with real and simulated data are described in Section 5 and related work is presented in section 6. Finally, we conclude in Section 7.

2. Data preprocessing

In this section, we address the problem of preprocessing railway data. From raw data collected by sensors, we design a suitable representation for data mining tasks.

The train monitoring system exploited in this study is such that a large set of sensors is distributed on the main components of each monitored train. The key element of this system is the bogie, because failures and anomalous behaviors are most of the time associated with it. Each of the 8 bogies of a train has 32 sensors collecting information such as temperatures, accelerations and velocity. Every five minutes during a journey, all sensors collect a value stored in a central database. A complete description of these data is available in Carrascal, Díez, and Azpeitia (2009).

2.1. Sensor data for train maintenance

The data resulting from sensors for train maintenance is complex for the two following reasons: (i) very often errors and noisy values pervade the experimental data; (ii) multi-source information must be handled at the same time. For instance, in train maintenance following data must be considered.

2.1.1. Sensors

Each sensor describes one property of the global behavior of a train which can correspond to different information (e.g., temperature, velocity, acceleration).

2.1.2. Measurements

They stand for numerical values recorded by the sensors and could be very noisy for different reasons such as failures, data transfer, etc. Note that the numerical values collected by the sensors are then discretized to obtain a set of data more suited to the step of data mining described in Section 3.

2.1.3. Readings

They are defined as the set of values measured by all the sensors at a given date. The information carried out by a reading could be considered as the state of the global behavior observed at the given moment. Due to the data transfer, some errors may occur and then readings can become incomplete or even missing.

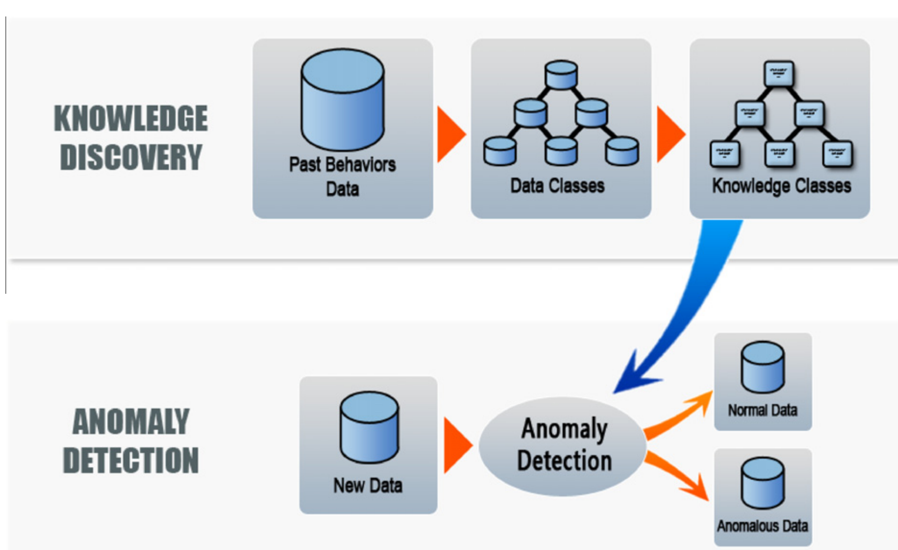

Fig. 1. General framework.