Advances in Safety, Reliability and Risk Management – Bérenguer, Grall & Guedes Soares (eds)

SPAMUF: A behaviour-based maintenance prediction system

Pedro Bastos
Instituto Politécnico de Bragança, Bragança, Portugal

Isabel Lopes
Universidade do Minho, Escola de Engenharia, Guimarães, Portugal

Luís Pires
Instituto Politécnico de Bragança, Bragança, Portugal

ABSTRACT: In the last years we have assisted to several and deep changes in industrial manufacturing. Many industrial processes are now automated in order to ensure the quality of production and to minimize costs. Manufacturing enterprises have been collecting and storing more and more current, detailed and accurate production relevant data. The data stores offer enormous potential as source of new knowledge, but the huge amount of data and its complexity far exceeds the ability to reduce and analyze data without the use of automated analysis techniques.

The paper addresses an organizational architecture that integrates data gathered in factories on their activities of reactive, predictive and preventive maintenance. The research is intended to develop a decentralized predictive maintenance system (SPAMUF—Prediction System Failures for Industrial Units Globally Dispersed) based on data mining concepts. Predicting failures more accurately will enable taking appropriate measures to increase reliability.

Keywords: Management, maintenance, data mining, knowledge discovery.

1 INTRODUCTION

The industrial production has suffered considerable changes, becoming more complex, contributing to this a need for increased efficiency, greater flexibility, product quality and lower costs (Bansal, *et al.* 2004). As markets become ever more dynamic, grows the need to introduce concepts of flexibility and agility, enabling companies to deliver customized products reacting promptly to fluctuating demands.

In order to launch to the market diversified products, with a competitive price and high quality, companies have incorporated some techniques and postures that allowed them to converge to the type of production that the market wants. The companies incorporation of the Just in Time (JIT) philosophies and agile production allows them to provide diversified products, and if needed, in large quantities.

Maintenance process is usually performed by integration of maintenance and process engineering functions at the phase of selection and application of machines and equipment; and also through pro-active actions on those machines and equipments that will necessarily passes by preventive and predictive maintenance (Palmer 1999). Sometimes changes at the project are also considered. In literature it is possible to find three generic types of maintenance (Chu, *et al.* 1998; Pinjala, *et al.* 2006):

- Corrective maintenance, consisting in repair actions when some equipment or machine fails. The equipment is in action until the moment that it fails. At that moment it will be repaired or replaced. The main disadvantages of this approach included fluctuant and unpredictable production, high levels of non-conforming products and scraps as well as high levels of maintenance interventions motivated by catastrophic failures (Swanson 2001);
- Preventive maintenance, characterized for periodic maintenance operations in order to avoid equipment failures or machinery breakdowns, through optimal preventive maintenance scheduling using a wide range of models describing the degrading process of equipment, cost structure, and admissible maintenance actions (Yao, *et al.* 2004);
- Predictive maintenance uses some indicators to “feel” when some breakdown is close to happening. This type of maintenance intends to make interventions on machinery before harmful events may occur.

The need to satisfy companies requirements leads to high pressure in the factories maintenance systems. The maintenance function, considered non-value-adding, it is ever more asked to contribute for the costs reducing, while keeping the machines in excellent working condition (Bansal, *et al.* 2006).

Present industrial maintenance systems are becoming obsolete, not enabling a more capable response to increasing demands of the production system. It is urgent to define maintenance systems capable of shaping the requirements of production environments in order to maximize their ability to respond to failure.

Nowadays, the amount of data generated and stored during industrial activities exceeds the capacity to analyze them without the use of automated analysis techniques. As a consequence of that increase of information, the data processing using traditional methods has become more difficult and complex (Goebel & Gruenwald 1999). The conventional tools of data analysis have limited capacity to detect patterns and discover the existing knowledge in data, because they only use statistical methods (Michalsky, *et al.* 1998).

In the course of time, there was a need for the existence of a new generation of computational tools and techniques in order to assist humans in extracting useful information from data, in other words, knowledge. Thus, in the late 80's emerged the area of Knowledge Discovery in Databases (KDD), using models and data mining techniques for extract useful knowledge, patterns and tendencies previously unknown, in a autonomous and semi-automatic way (Apte, *et al.* 2002).

Data mining is applied in several areas of industrial engineering, such as the area of design engineering, production systems, systems and decision support, fault detection and quality improvement, and customer relationships management. In turn, their application in the field of industrial maintenance began in the 90's.

Harding, *et al.* (2006) present a prototype system called EXPERT-MM (Batanov, *et al.* 1993), that consists in databases containing failure machines events and the behavior of relevant equipments at failure time, used in the design of the maintenance management systems. This prototype works historical failure data and provides suggestions for an appropriate preventive maintenance schedule. A data-based design of optimal maintenance methods has also been proposed by Hsu & Kuo (1995). They suggested that 100% inspection should start after the manufacture of a certain number (n) of parts and when the percentage of bad parts reaches a certain threshold value. Preventive maintenance should then start to bring the process under control again. When the process has been controlled and an additional n parts have been manufactured, the procedure can be repeated. Another work by Sylvain, *et al.* (1999) used different data mining techniques, including decision trees, rough sets, regression, and neural networks to predict component failures based on data collected from the sensors of an aircraft. Their results also led to the design of preventive maintenance policies before the failure of any component. Romanowski & Nagi (2001) applied data mining in a maintenance domain to identify which subsystems were responsible for low equipment availability. They recommended a preventive schedule and found that sensors and response frequency provide the most information about faults. They used a decision tree to model the data.

The purpose of this paper is to plan and apply a decentralized predictive maintenance system based in the application of different data mining techniques over maintenance data, generated by different machines in the same or in different production lines of industrial units globally dispersed. The goal is to foresee a failure and generate a set of notifications based in preventive maintenance actions. Predicting the possibility of breakdowns with bigger accuracy will increase systems reliability.

2 SYSTEM FUNCIONALITY

The aim of the system to be developed is to integrate the occurrences of faults in similar machines from different factories, creating a system (SPAMUF) of distributed databases which allows, using data mining, the prediction of failures in a way to perform timely interventions in equipment and consequent increase of availability and productivity (Fig. 1).

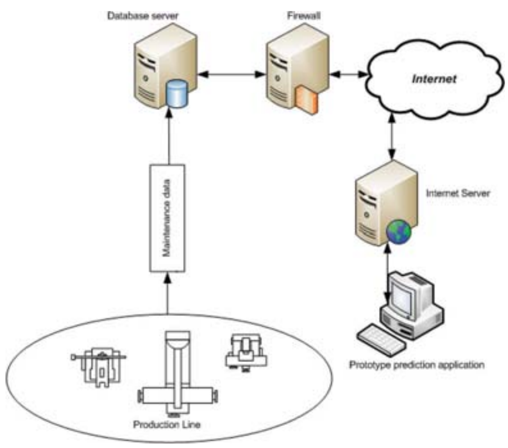

Figure 1. Overview of the project.

Currently, we have a project partner, which is an important international organization in the industry of electronic components for the car industry. The organization stores historical data from three years of corrective and preventive maintenance held on a certain number of machines in some production lines. However, the existing databases are inconsistent and without any type of normalization, because most of data collected from different maintenance actions are registered manually in the existing informatic system by different maintenance teams. This will be solved by defining a database that will serve as the basis for applying the prediction prototype, where the data will be collected through the implementation of automated agents. In order to build this new database, some requirements have been identified:

- define the most suitable subset of attributes:
  - maintenance teams;
  - type of maintenance performed.
- normalize the existing attributes:
  - equipment identification;
  - type of equipment;
  - production line and area identification.
- create new attributes:
  - fixed lists to designate the type of failure;
  - date/hour of failure;
  - date/hour of the beginning of the intervention;
  - date/hour of the end of the intervention.

However, the existing database was used to start applying certain data mining techniques in order to verify the existence of any important relationship and set all the requirements to implement in the new database.

The data collecting process has to be performed by agents, which will be responsible for adapting and transforming the information. Even when data on the factory floor is collected through maintenance operators (using a formalized internal registry, for example), this information has to be interpreted by software agents, to adapt and transform the data structures into a semantically viable knowledge base.

Data mining techniques will focus on data to discover implicit and hidden knowledge in order to generate predict patterns of behavior and events. The possibility of events occurrence will be provided through a prediction system for each plant. Data relating to the intervention process, the used material, the consequences of non-intervention and scenario generation will be provided over the form of decision support system.

Usually, enterprises do not share data produced from their maintenance interventions. This research intends to create an organizational architecture that makes the integration of data produced in

Figure 2. Website interface (http://www.esa.ipb.pt/~bastos/spamuf/).

factories on their maintenance activities. A decentralized predictive maintenance system is being developed based on data mining concepts.

All the results obtained by the system will be available on a web platform (Fig. 2) which identifies the users ensuring a secure environment.

All the members that cooperate on the system must have mutual trust and a blind confidence on the infrastructure. The resultant information is critical. Therefore, it should be ensure that the information which flows from the agents to the database is a secure and trustful one. It is also important that the database server only accepts information that comes from authorized agents. This validation is very important for stopping a malicious agent/user from flooding the database with erroneous information that will make the prediction model untruthful. A SSL certificates for authentication where used, these are digitally signed documents which bind the public key to the identity of the private key owner. By this way each, project partner will have access to privileged and confidential information. Digital certificates mainly serve two purposes:

- To establish the owner's identity;
- To make the owner's public key available.

A digital certificate is issued by a trusted authority—a certificate authority (CA)—and it is issued only for a limited time. When its expiration date passes, the digital certificate must be replaced. SSL uses digital certificates for key exchange, server authentication, and optionally, client authentication (Oppliger, *et al.* 2008).

All the results will be returned through the data mining tool used in the prototype prediction application directly to the Web platform.

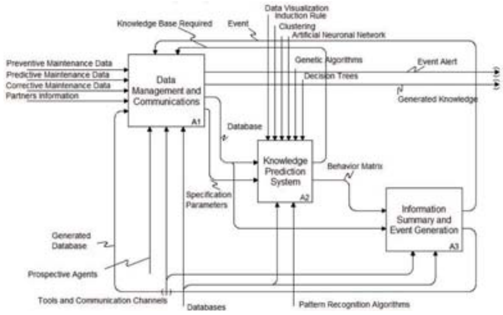

Figure 3. Main system in IDEF0 format.

IDEF0 method is used to specify model functions, representing high-level activities of a process and its decomposition in lower hierarchical sub activities or processes. IDEF0 models portrait a view of the process in terms of Inputs, Controls over the process, Outputs, and Mechanisms acting on the process, named ICOM's. IDEF0 Function Modeling is designed to model decisions, actions, and activities of an organization or system. IDEF0 model notation uses functions and activities abstracted from temporal sequence. The diagrams in this notation show activation (of functions, processes), not flow. More information can be found at Feldmann & Tieso (1998).

The global system will be based on three main processes, as shown in the Figure 3: data management and communications (A1), knowledge prediction system (A2) and information summary and event generation (A3).

The A1 activity will be responsible for data collecting, which will take place from a local perspective to a higher layer. As output, this activity will produce a database and a set of specification parameters previously used in the data preparation.

The A2 activity is the main module of knowledge production and inference of behavior patterns related to each equipment of a factory unit. This activity will generate as an output a behavior matrix that will be the input of the synthesis of information module and event generation (A3), which in turn use the resources of the A1 activity to generate events that consists in proactive failures notifications. This output function aims to advise the person responsible for maintenance in order to act over the equipment before malfunctioning.

The A1 activity is composed by four sub-activities (Fig. 4).

The Data Selection and Data Preparation activity (A11) is responsible for selecting and analyzing data from different types of maintenance as well as information partners in order to generate a set of constraints that control the phase of Definition of the Knowledge Base (A12).

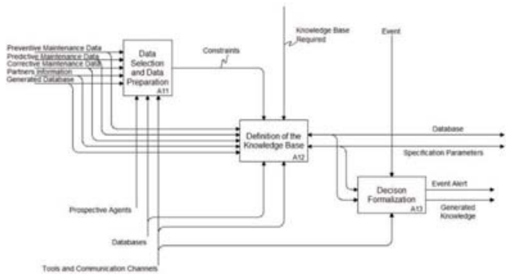

Figure 4. A1 sub-activities.

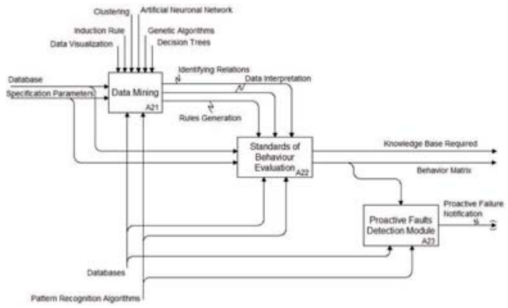

Figure 5. A2 sub-activities.

The process of data gathering (A11) will be performed automatically by prospective agents using other mechanisms such databases and tools and communication channels. Data collected on the production line through the agents will be adapted and transformed to data structures in order to create a knowledge base semantically viable.

The definition of the knowledge base sub-activity (A12) will generate the database used in the main system and its specification parameters, which in turn serve as input for the decision formalization sub-activity (A13), generating the outputs of the system.

The knowledge prediction system module (A2) (Fig. 5), consists in three activities, the first one is data mining activity (A21), that uses the generated database and the specification parameters that will affect the whole process and the application of different artificial intelligence methods.

In this sub-activity, depending on the type of data to be analyzed will be applied techniques such as:

- Decision trees, structures in the shape of a tree representing sets of decisions. These decisions create rules for classification of data sets. The specific methods of the decision trees include Classification Trees and Regression (CART Classification and Regression Trees) and Chi Square

Interactive and Automatic Detection (CHAID Automatic Interaction Detection Chi Square);

- Artificial neuronal network, one of the most known and used techniques in Data Mining (Berson, *et al.* 2000). It consists in a set of simple processing elements (nodes), with a large number of interconnections. The whole structure is based on recreation of human brain, that is to say, the ability to learn and self correct itself;
- The induction rule that consist in the ability to detect trends and patterns in data groups, and is one of the Data Mining techniques better known. The main objective of this method is to obtain dependencies between values or attributes. Generally the results are presented in the form of rules such as $X \rightarrow Y$ (Toloo, *et al.* 2009);
- Genetic algorithms which is a technique widely used in optimization problems and are based on principles of natural laws of evolution proposed by Charles Darwin in 1859 (Beasley, *et al.* 1993);
- Clustering that is a technique whose purpose is to detect the existence of different groups within a data set, and in case of existence, determine or characterize them;
- Visualization techniques will be used in order to represent the data mining results through graphs or diagrams in order to provide a better representation of complex information.

The output of this activity consists in identifying the relationships, data interpretation and rules generation.

In order to see an example of the project outputs, we take some fields (Breakdown number, Machine, Production line, Date, Type of breakdown and Maintenance group) of the database of our pattern and after some pre-processing data, we could apply neuronal networks on different fields in order to obtain possible and relevant relations between them and visualize it and consequentially perform some interpretation (Fig. 6).

All this information will control module for the standards of behavior evaluation (A22), which serve as the basis for the generation of a behavior matrix.

This matrix thus controls the proactive detection module failures (A23) to generate proactive notification of faults to be made available to partners.

The A3 module (Fig. 7), information summary and event generation, will receive as input the output of the A1 activity, database, and the A2 activity output, behavior matrix.

This module consists in three activities. Activity A31, information reception module, is responsible for generating a synthesis of information of all the information processed and will serve as input to A32 activity, event generation, responsible for standardizing all the information that will be used in activity A33, event dissemination.

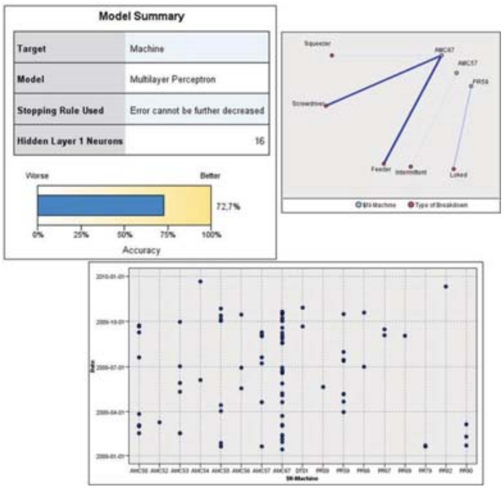

Figure 6. Example of artificial neuronal network application.

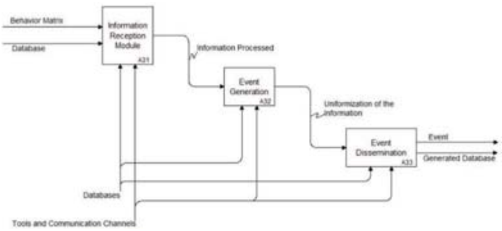

Figure 7. A3 sub-activities.

3 CONCLUSIONS

A globalized world is a place with several challenges and also fulfilled by vast opportunities. In this paper it was presented an architecture that explores the benefits of this globalization and uses new technological means to incorporate more accuracy in a prediction system used by industrial maintenance teams. Organization managers and maintenance leaders are more concerned with highlighting areas of existing or potential maintenance problems in order to be able to improve performance and minimize the operational cost of maintenance. Applying data mining techniques on the available industrial maintenance data may help to discover useful rules that allow locating some critical issues that will have substantial impact on improving all the maintenance processes. Before using rules to change operations, it is important to examine the rules. For this a domain expert is required. Unexpected rules that do not make sense may also signal others, more nefarious. The proposed system will help enterprises to collect, extract and create knowledge in a way that enterprises will predict with more accuracy the moment to realize maintenance actions and thus improve the productivity of manufacturing process. The innovative point of this system is the capability of collecting and treats data dispersed in different facilities that result from maintenance interventions in different environments. The existing huge amount of data from maintenance actions are not fully used to increase the efficiency of maintenance prediction. Data mining seems to be the step forward that will change the actual state.

Currently we are in the stage of database development and implementation of automatic agents to collect data in a way to generate the database that supports the system.

As future work, as soon a consistent database will be available, different data mining techniques will be tested in a way to verify which presents more accuracy and better results, so we can perform the validation of the results and incorporate them in the web platform created.

REFERENCES

Apte, C.; Liu, B.; Pednault, E.P.D.; Smyth, P. (2002). Business Applications of Data Mining. *ACM* 45(8): 49–53.

Bansal, D.; Evans, D.J.; Jones, B. (2004). A real-time predictive maintenance system to a production machine systems. *International Journal of Machine Tools and Manufacture* 44: 759–766.

Bansal, D.; Evans, D.J.; Jones, B. (2006). BJEST: A reverse algorithm for the real time predictive maintenance system. *International Journal of Machine Tools and Manufacture* 46: 1068–1078.

Batanov, D.; Nagarur, N.; Nitikhumkasem, P. (1993). Expert-MM: A Knowledge Based System for Maintenance Management. *Artif. Intell. Eng.* 8: 283–291.

Beasley, D.; Bull, D.R.; Martin, R.R. (1993). An Overview of Genetic Algorithms, Part1, Fundamentals. *University Computing* 15(2): 58–69.

Berson, A.; Smith, S.; Thearling, K. (2000). *Building Data Mining Aplications for CRM*. USA: McGraw-Hill.

Chu, C.; Proth, J.M.; Wolff, P. (1998). Predictive maintenance: The one-unit replacement model. *Int. J. Production Economics*: 285–295.

Feldmann, C.; Tieso, J.V. (1998). *The Practical Guide to Business Process Reengineering Using Idef0*. New York: Dorset House.

Goebel, M.; Gruenwald, L. (1999). A survey of Data Mining and Knowledge Discovery Software Tools. *SIGKDD* 1(1): 20–33.

Harding, J.; Shahbas, M.; Srinivas, S.; Kusiac, A. (2006). Data Mining in manufacturing: A review. *Journal of Manufacturing Science and Engineering* 128: 969–976.

Hsu, L.; Kuo, S. (1995). Design of optimal maintenance policies based on on-line sampling plans. *European Journal of Operational Research* 86: 345–357.

Michalsky, R.S.; Bratko, I.; Miroslav, K. (1998). *Machine Learning and Data Mining Methods and Aplications*. England: John Wiley & Sons, Inc.

Oppliger, R.; Hauser, R.; Basin, D. (2008). SSL/TLS Session-Aware User Authentication. *IEEE—Xplore—Computer* 41(3): 59–65.

Palmer, R. (1999). Maintenance planning and scheduling handbook. New Jersey, McGraw Hill.

Pinjala, S.K.; Pintelona, L.; Vereecke, A. (2006). An empirical investigation on the relationship between business and maintenance strategies. *Int. J. Production Economics*: 214–229.

Romanowski, C.; Nagi, R. (2001). *Analyzing maintenance data using data mining methods*. MA, USA: Dan Braha Kluwer Academic Publishers Norwell.

Swanson, L. (2001). Linking maintenance strategies to performance. *Int. J. Production Economics*: 237–244.

Sylvain, L.; Fazel, F.; Stan, M. (1999). Data Mining to Predict Aircraft Component Replacement. *IEEE Intelligent Systems* 14(6): 59–65.

Toloo, M.; Sohrabi, B.; Nalchigar, S. (2009). A new method for ranking discovered rules from data mining by DEA. *Expert Systems with Applications* 36: 8503–8508.

Yao, X.; Fernández-Gaucherand, E.; Fu, M.C.; Marcus, S.I. (2004). Optimal Preventive Maintenance Scheduling in Semiconductor Manufacturing. *IEEE Transactions On Semiconductor Manufacturing* 17(3): 345–356.