Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 11 (2013) 225 – 228

www.elsevier.com/locate/procedia

2nd International Through-life Engineering Services Conference

Data Mining and Visualization of Diagnostic Messages for Condition Monitoring

Eckart Uhlmann, Claudio Geisert*, Eckhard Hohwieler, Ian Altmann

Fraunhofer IPK, Pascalstr. 8-9, 10587 Berlin, Germany

* Corresponding author. Tel.: +49 30 39006 133; fax: +49 30 391 1037. *E-mail address:* claudio.geisert@ipk.fraunhofer.de

Abstract

Complex technical systems consist of a huge amount of electromechanical subsystems and components with more or less interdependencies. During the operational phase these systems are subject to wear and tear and other degradation mechanisms. Therefore condition monitoring is a major challenge for operators of such systems. A well-established practice of implementing condition monitoring is to equip the system, or its functionally relevant components with appropriate sensors and measurement technology. Selection of sensors and diagnostic algorithms requires a deep understanding of physical correlations which is usually a core competence of the system provider. Moreover, subsequent installation of a condition monitoring system using additional sensors in most cases is attended by an interference with the system and may affect its functionality if constructional changes of the system for sensor mounting become necessary. A more simple way to access data for technical diagnosis is the use of an on board diagnosis system that collects the system messages that rise while the system is in operation. To control and monitor functionality and interaction between subsystems, each of them generates internal state variables and communicates a subset of these as system messages via task specific bus systems. Here the challenge is to find a suitable way of analyzing the messages. The present paper discusses how to handle system messages for condition monitoring purposes by using data mining and visualization.

Selection and peer-review under responsibility of the International Scientific Committee of the “2nd International Through-life Engineering Services Conference” and the Programme Chair – Ashutosh Tiwari

Keywords: Data Mining; Availability; Diagnosis; Report Generator; Maintenance; Life Cycle

1. Motivation

Capital goods like production systems or systems used in the sectors of public transportation or energy are highly stressed and often operated near full capacity utilization. Therefore strong efforts are made to ensure their technical availability. A well-established practice is the implementation of condition monitoring to be informed at all times about the health state of the system under consideration. Following Bertalanffy "a system may be defined as a set of elements standing in interrelation among themselves and with environment" [1]. The complex interaction of subsystems and components needed for correct overall system behavior makes condition monitoring extremely difficult since deep understanding of underlying physical correlations is a basic prerequisite for adequate diagnosis.

Another way to access the health status of a system is to use already existing internal data like system messages which are generated by the subsystems and are distributed via communication bus. System messages can be divided into so called protocol and diagnosis messages. Protocol messages provide information about the system’s operating state while diagnosis messages comprise information about health status of the overall system or its subsystems (refer to Fig. 1). Both types of messages include timestamps for start and end of the event as well as an ID that is uniquely linked to the event description.

A challenge for using system messages is that they are designed for specific tasks like system control and documentation of events and do not necessarily correlate to the underlying physics. Another disadvantage is that they are listed in chronological order of occurrence without additional

Selection and peer-review under responsibility of the International Scientific Committee of the “2nd International Through-life Engineering Services Conference” and the Programme Chair – Ashutosh Tiwari
doi:10.1016/j.procir.2013.07.045

graphical representation. Nevertheless they are widely used by service technicians to get a first impression of the system's current health status and to analyze previous events that may be associated with the current event.

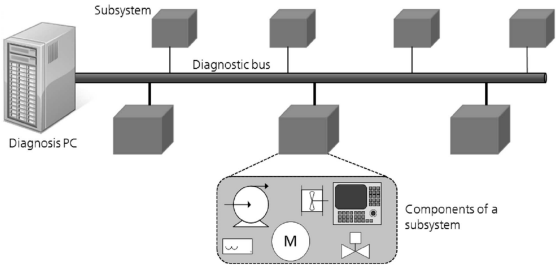

Fig. 1. Principle structure of an on board diagnosis system.

2. Approach

2.1. Hypothesis

Since wear and tear and other degradation mechanisms are subtle processes many serious failures are often announced by smaller intermittent occurring irregularities. These irregularities firstly disturb sub-functions without leading to an immediate breakdown of the overall system. Nevertheless they may build specific patterns becoming apparent within system messages.

2.2. Current Restrictions and Proposal for Solution

As mentioned before, weaknesses of today's possibilities of using logged system messages are:

- Message statement does not necessarily correlate to the underlying physics.
- Data is provided only as lists in tabular format.
- No graphical representation of the data.
- No support for analysis by adequate software tools.

Therefore manual data mining for condition monitoring purposes is very time consuming and extremely depending on specific expert's knowledge.

To close this gap the proposed approach deals with the development of a data mining software application to facilitate manual analysis and to detect (semi-)automatically emerging patterns, e.g. caused by an evolving malfunction of a component, at an early stage. This early detection can finally be used for the introduction of a condition based maintenance strategy.

The proposed solution consists of the following modules:

- On board diagnosis system to collect system messages continuously (not part of the development).
- Database for structured storage of system messages and additional information.
- Graphical user interface (GUI) to enable user interaction.
- Analysis toolbox including algorithms for data mining.
- Export functionality to share results for further analysis.

Fig. 2 shows the interaction between the user and proposed data mining software application in a use case diagram.

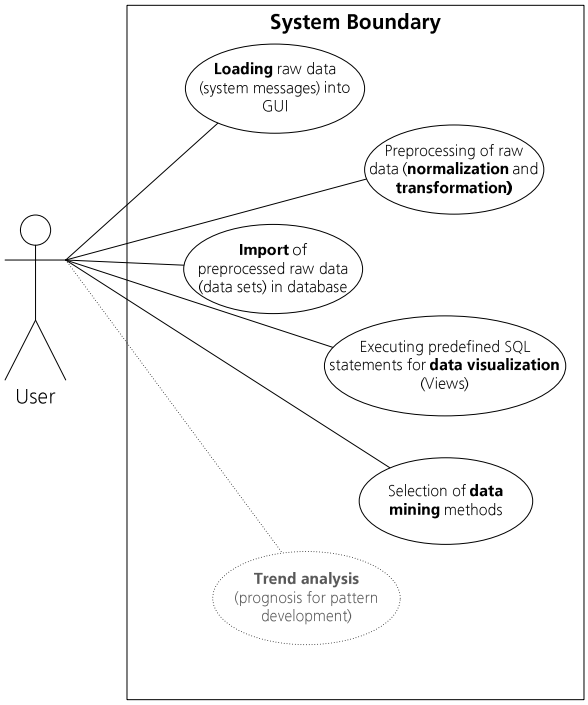

Fig. 2. Data Mining GUI – use case diagram (at the current stage of work trend analysis is not implemented).

3. Data Mining GUI

3.1. IT-Architecture

The basis of the software application is formed by a relational database scheme to store dynamic diagnosis data together with additional information about system configuration as well as static descriptions of message codes (message list) and their causing subsystems. For user interaction a JAVA based graphical user interface (GUI) was realized to meet ergonomic analysis workflows following the interactive data mining process shown in Fig. 6. On the one hand sequential steps like selection of interesting time interval, system population and collection of messages to be used for data mining are presented separately to the user. On the other hand it is possible to go back to the last step to realize needed iterations. A central aspect was the visualization of the dynamic condition data with different graphical diagrams to provide the user with compressed information. Two general kinds of diagrams have been implemented, histogram and Gantt chart. The overall software architecture is illustrated in Fig. 3. The module "generate report" includes the processing of data mining algorithms as well as graphical representation of the results.

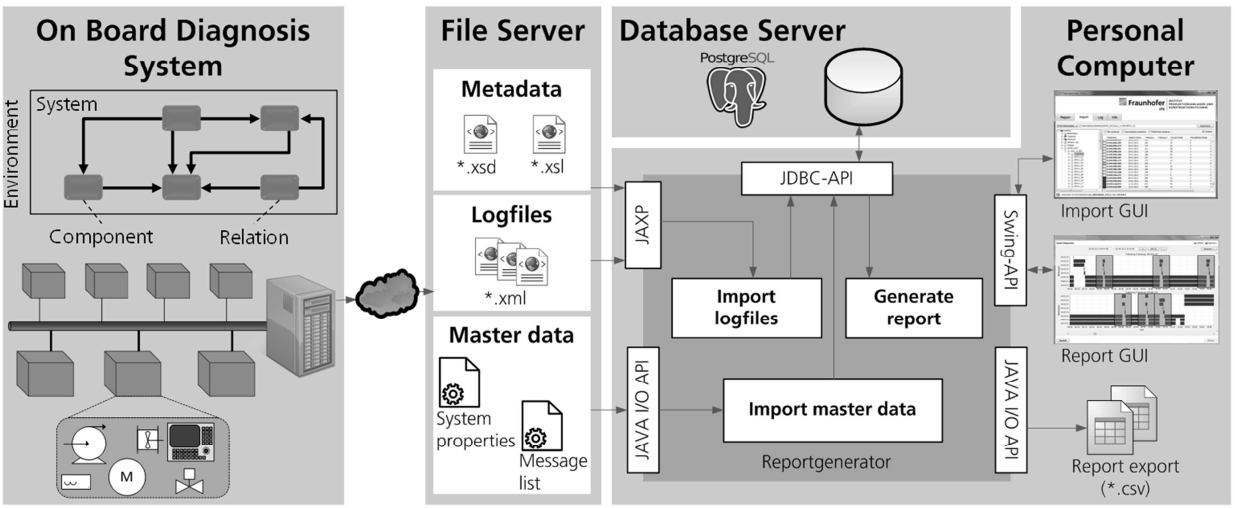

Fig. 3. Data Mining GUI – software architecture.

3.2. Data Mining Methods

Data mining in general means the attempt to disclose hidden correlations between data whereas the term data implies time series of sensor measurements as well as event messages and data without temporal reference. While data mining tools like the open source software RapidMiner [2] offer a wide variety of algorithms for the analysis of time series, methods for data mining within time interval data (event messages) are quite rare [3].

At the current stage two approaches for data mining have been implemented. The first approach finds similar time interval patterns by describing a pattern using Allen's interval algebra [4]. This interval algebra provides 13 relations for setting up a pattern (before, meets, overlaps, starts, during, finishes and respective inverse relations and equal, see Fig. 4). Any two intervals fulfill exactly one of the relations. Disadvantages of robustness due to describing exact time boundaries within the relations were solved by using thresholds and fuzzy extensions.

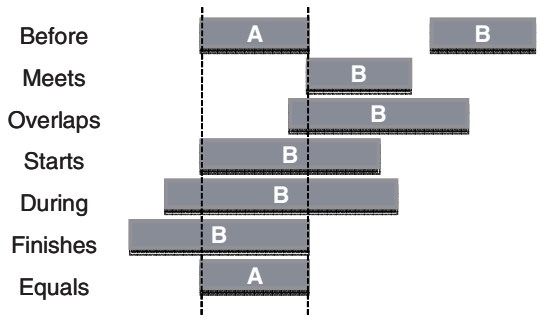

Fig. 4. Relations of Allen's interval algebra (inverse relations are not listed).

The second approach finds frequent time point patterns by implementing the Apriori algorithm [5]. This algorithm is mostly applied in the domain of market basket analysis, where frequent sets of purchased items are mined. Here, it is used to find association rules between given messages of the diagnosis system. The process of interactive data mining is shown in Fig. 6.

3.3. Visualization

As the human ability to recognize complex relationships within graphical representations is very strong, visualization of data and their temporal correlations is an important part of interactive data mining. This means that visualization is not only used for graphically preparation of data mining results. Furthermore, visualization is part of the overall interactive data mining process as depicted in Fig. 6. Besides classic histogram representation, Gantt charts are used to illustrate temporal dependencies between interval data (ref. Fig. 5).

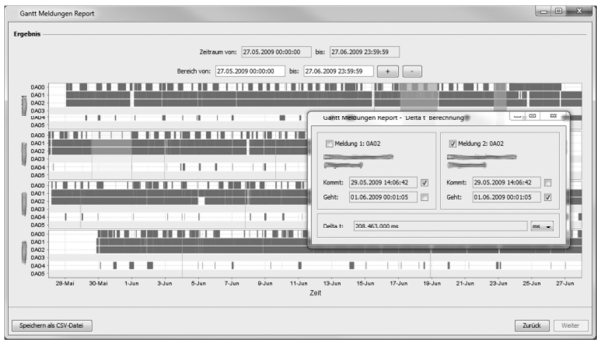

Fig. 5. Gantt chart of selected system messages.

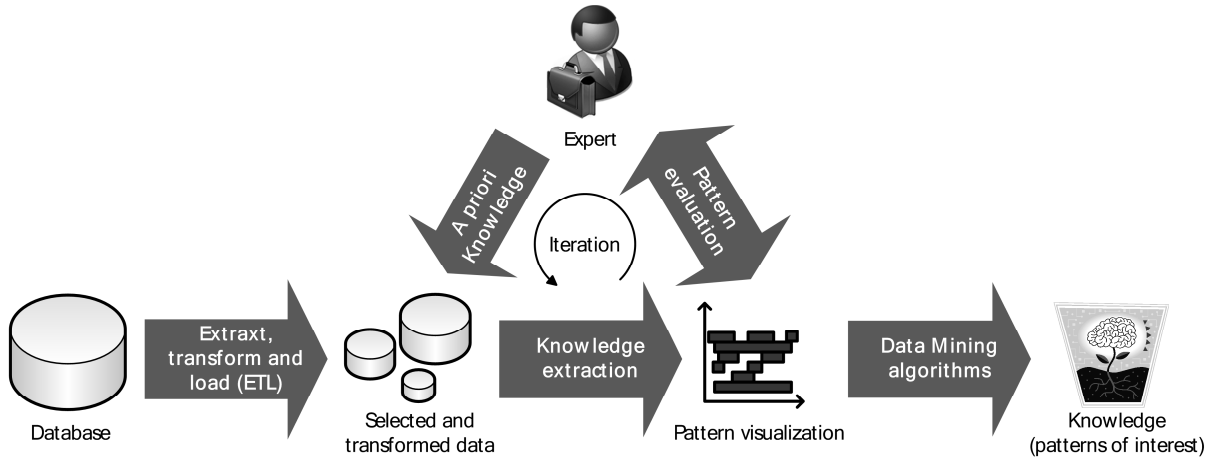

Fig. 6. Interactive data mining process.

4. Summary and Outlook

In the present paper an approach for supporting analysis tasks by graphical representation of condition data and the use of data mining algorithms for finding reoccuring patterns in large data sets was described. The application enables data analysts to recognize evolving technical problems and to launch appropriate counter measures in terms of condition based maintenance. Further research and development has to be done to fully automate the recognition of unknown patterns within large data sets and to forecast their future development in the sense of trend analysis.

Acknowledgements

Parts of the presented work were carried out in the context of the Fraunhofer Innovation Cluster Maintenance, Repair and Overhaul in Energy and Transport.

References

[1] Bertalanffy, L. v.: General System Theory – Foundations, Development, Applications. 14th edition 2003, George Braziller Inc. New York, 1969, p. 252.

[2] NN: RapidMiner – Manual. 2010 by Rapid-I GmbH. Online resource: http://docs.rapid-i.com/files/rapidminer/rapidminer-5.0-manual-english_v1.0.pdf, last visit: June 30, 2013.

[3] Moerchen, F.: Temporal pattern mining in symbolic time point and time interval data. Siemens Corporation, Corporate Research, 2010. Online resource: www.siam.org/meetings/sdm11/moerchen.pdf, last visit: June 30, 2013.

[4] Allen, J. F.: Maintaining knowledge about temporal intervals. In: Communications of the ACM. 26 November 1983, ACM Press, 1983, pp. 832 – 843.

[5] Agrawal, R.; Srikant, R.: Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago, Chile, September 1994, pp. 487 – 499.