DATA MINING FOR NDT INDUSTRY: PREVENTIVE MAINTENANCE, FAILURE PREDICTION, QUALITY CONTROL

1SURENDRAGADVE, 2VIVEK SUNNAPAWAR, 3G. T. THAMPI

1Sardar Patel College of Engineering, Andheri (W), Mumbai, India
2Lokmanya Tilak College of Engineering, Koparkhairane, Navi-Mumbai, India
3Thadomal Shahani College of Engineering, Bandra (W), Mumbai, India
Email: surendragadve@gmail.com, principal@ltcoe.com, gtthampi@yahoo.com

Abstract: NDT enterprises have been collecting and contributing to the cause of commercial data warehouses. The data warehouses 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. This paper provides a brief introduction into knowledge discovery from databases and presents the methodology for data mining in time series. The relevancy of data mining for NDT Industry shall be depicted.

Keywords: data mining, NDT Industry, data warehousing, computer aided design

I. INTRODUCTION

Virtually all NDT enterprises use powerful data acquisition systems to collect, analyze and transfer data from nearly all the processes of the organization. This data may be related to machines, products, processes, maintenance, quality control, failure detection, etc. and is typically stored in databases at various stages. The use of databases and statistical techniques are well established in engineering, because intelligently analyzed data is a valuable resource, since it gains new insights and can provide significant competitive advantage. Traditionally statistical techniques were used in order to find patterns in NDT industry data. But machines with growing complexity and sensor equipment as well as recent proceedings in information technology, data acquisition systems and storage technology lead to an overwhelming amount of data, which is permanently increasing and contains hundreds of attributes. In order to model the system’s behavior accurately, these attributes have to be considered simultaneously. This complexity calls for new techniques and tools for the automated extraction of useful knowledge from huge amounts of raw data. Data Mining is about solving these problems by applying mathematical models to automatically discover patterns in data already present in databases. Data Mining is thereby one step in Knowledge Discovery in Databases, which denotes the entire process of turning low-level data into useful knowledge. This process often includes the following important stages: The first step involves the understanding of the application domain, which is of great importance especially when analyzing NDT Industry data. The goals and tasks of the data mining process have to be determined and all the factors, which might affect the manufacturing process under study, should be revealed and understood. For successful generation of new knowledge a close collaboration of domain experts, data experts and data mining experts is needed. The second step includes selecting the target data set. Since data mining can only uncover patterns already present in the data, the target dataset must be large enough to contain these patterns while remaining concise enough to be mined in an acceptable timeframe. The data sets are often stored in various databases which requires additional integration. Next, the data sets have to be preprocessed, comprising inter alia transformation, handling missing data and removing noise. The fourth step finally is data mining for the extraction of patterns from the data. This involves the selection and application of appropriate (mathematical) data mining algorithms as well as the development of a model, which describes the pattern. This step is often accompanied by ‘traditional’ statistical analysis and data visualization. In a last step the extracted patterns have to be interpreted and verified. Here the evaluation from the domain experts is essential to really transfer the patterns into new knowledge. Usually some of the Knowledge Discovery in Databases steps need to be iterated several times to finally reach this goal. Once useful patterns are found and described, they allow to make (nontrivial) predictions on new data. Depending on the data and the intended outcome of the overall data mining process two main goals can be distinguished: Regression is concerned of computing an expression that predicts a numeric quantity, in classification the outcome to be predicted is a discrete class. A wide range of data mining techniques is available to serve these goals, each with its own advantages and disadvantages. Nowadays NDT industry can store vast amounts of data obtained at a relatively low cost, although these data may fail to provide information. NDT Enterprises are adapting new technologies as Data warehouses, large repositories that integrates data from several sources in an organization for analysis, Online analytical processing (OLAP) systems, which provide fast answers for queries that aggregate large amounts of detail data to find overall trends, Data mining applications, that seek to
discover knowledge by searching semi-automatically for previously unknown patterns and relationships in NDT Enterprises. This paper also document about the ways and means of transforming the NDT workplace also into a research entity. This paper also focuses the techniques for faster inspections with a higher probability of detection. This means reduced downtime, lower costs and higher confidence for even the most difficult inspection problems. The field of Nondestructive Testing (NDT) is a very broad, interdisciplinary field that plays a critical role in assuring that structural components and systems perform their function in a reliable and cost effective fashion. NDT technicians and engineers define and implement tests that locate and characterize material conditions and flaws that might otherwise cause planes to crash, reactors to fail, trains to derail, pipelines to burst, and a variety of less visible, but equally troubling events. These tests are performed in a manner that does not affect the future usefulness of the object or material. In other words, NDT allows parts and materials to be inspected and measured without damaging them. Because it allows inspection without interfering with a product's final use, NDT provides an excellent balance between quality control and cost-effectiveness. Generally speaking, NDT applies to industrial inspections.

II. PREDICTING FUTURE VALUES

In ultrasonic testing, high-frequency sound waves are transmitted into a material to detect imperfections or to locate changes in material properties. The most commonly used ultrasonic testing technique is pulse echo, whereby sound is introduced into a test object and reflections (echoes) from internal imperfections or the part's geometrical surfaces are returned to a receiver. Below is an example of shear wave weld inspection. Notice the indication extending to the upper limits of the screen. This indication is produced by sound reflected from a defect within the weld. Forecasting is of key importance in process and NDT Industry engineering, since the prediction of component malfunctions based on the data collected from sensors attached at machines can save money due to prevented consequential damages. For the prediction of values powerful data mining regression techniques are available.

Figure 1 shows ultrasonic waves spread spectrum measured with a prediction based on the so-called support vector machines method, a famous and mighty data mining technique. Although the prediction is well reproducing the middle of the data, there are deviations at the end of the time series. This leads to the conclusion that regression alone is not sufficient to look into the future, since the model will only work well on parts of the data it has already seen. The reason is a wrong scenario: In reality, i.e. the application in the plant, only the first part of the measurement would be seen and the future should be predicted! In order to develop a more realistic scenario, a technique called windowing can be used, whose basic idea is explained in Figure 2.

The time series is divided into parts (windows) and the values inside the window are used to predict a value outside the window after a given time horizon. This procedure creates the first example of a new data set. By moving the window stepwise and holding the horizon constant, the whole data set to predict the future is created. The new data set with the historical data taken from the windows and the future value (compared to the window) can now be used for regression. The prediction will be a prediction for the future after the horizon has passed. The prediction performance usually decreases with larger prediction horizon.

Figure 3 shows an example for application and explains the advantages of forecasting related to condition monitoring: Sensors, measuring the force on a machine detect an increase. A condition
monitoring system cannot predict the future making it impossible to judge, if the increase is going to be critical and when. This drawback can be overcome by creating a model for the forecasting of the time series using the data mining techniques explained above. It is clearly visible that the force will stabilize and nothing has to be done.

Figure 4: Principle of Data Mining system in NDT Industry

An implementation in the NDT Industry could be realized like sketched in Figure 4. Sensors attached to the machines monitor the process. These sensors may measure all kind of technical time series, e.g. vibration, temperature, force, tension etc. An interface controls the data acquisition and stores the time series in a data base (DB). This is the input data source for the data mining software, which computes the predictions and visualizes them for the user.

III. IDENTIFYING NDT INDUSTRY FAILURE

Beside predicting a numeric value of a time series, there is often the need to classify complete time series in order to perform a forecasting, e.g. to predict a machine failure. Imagine a saw cutting pieces. During each sawing cycle the vibrations of the saw are measured by a sensor and depending on these vibrations the question is: ‘Will the saw break?’ One could simply use the first half of the data (before the crash) and create a model on that. This will, however, not result in a good prediction performance, since the crucial aspect lies in the structure of the series, not in the concrete values. Another big problem when mining in time series is the high dimensionality. Thus feature extraction should be applied to compress the time series, keeping the important information while removing noise and correlations. Not only the algorithm will speed up, it is often the only way to get good prediction results on complex data streams. For feature extraction there is a large amount of methods known from time series handling, including the discrete Fourier Transform (DFT), the Discrete Wavelet Transform (DWT), functional like average, minimum, maximum, windowing and their combinations.

In Figure 5 an example for the case of predicting the crash of a saw is depicted. The time series are measured by a vibrational sensor. The blue-colored one is measured during a normal process, the red one right before a crash. To reduce the dimensionality and to discard noise, some features were extracted using e.g. DFT. In the feature space (right hand side) the two cases crash (red)/no crash (blue) can be classified easily.

IV. SUMMARY & OUTLOOK

Knowledge discovery in Databases (KDD) has a great potential in NDT Industry, since it turns the raw data into useful knowledge. It can give deeper insights into the processes, allows for the prediction of machine failure and for preventive maintenance. Thus KDD can support quality and reduce costs due to damages or loss of production. There is a growing interest and a recent trend to use data mining for NDT Industry. On the other hand side it should be pointed out that data mining in NDT Industry is a demanding task because of a lot of factors like complexity of the processes, (missing) quality of the data and the need to develop precise models with a very high prediction performance.

REFERENCES