A Data Mining Experiment on Manufacturing Shop Floor Data

J. Jenkole[1], P. Kralj[2], N. Lavrac[2] and A. Sluga[1]

[1]Department of Control and Manufacturing Systems, University of Ljubljana, Slovenia
[2]Department of Knowledge Technologies, Jozef Stefan Institute, Slovenia

Abstract

The paper describes a data mining and visualization experiment performed on a real-world problem and results achieved in the experiment. A subgroup discovery algorithm was used in order to find useful patterns in production data. Workshop and manufacturing work system utilization visualization plots were created in order to gain some useful information. The purpose of this work is to examine the usefulness of data mining and visualization of the shop floor data to support decision making in a manufacturing company.

1. BACKGROUND AND MOTIVATION

Manufacturing systems are becoming more and more complex, distributed, globalised, overwhelmed with data in diverse forms and originated in different sources. In the last decade manufacturing companies have automated and computerized many processes in order to ensure higher productivity, quality of production and minimization of production costs. Such computerized systems produce large volumes of data, whose exploitation is usually neglected even though valuable knowledge might be hidden in it. In the search for competitive advantage the acquisition and management of information and knowledge has become a competitive force [1].

The machinery, inc. manufacturing work systems, plays an important role in the overall performance of a company. Workshop control is likely to become one of the key components of the 21st century manufacturing company [2]. In fact, the machinery utilization is highly correlated with the overall performance of the entire company. Therefore, an efficient control of machinery utilization can be reflected in a better prosperity of the whole manufacturing company.

The integration of automated or semi-automated data acquisition systems, e.g. Supervisory Control and Data Acquisition (SCADA), Manufacturing execution systems (MES), etc., into production systems resulted in enormous data volumes. These systems collect and archive large volumes of raw data, which potentially contain hidden and possibly useful information. These real-time data, recorded to ensure the ability to trace production, can also be used to optimize manufacturing processes [3]. However, the classical approach in data usage, e.g. SQL (Structured Query Language) querying, cannot produce sufficient results in data-rich but information-poor problems [4]. Data mining and visualization enable the elicitation of explicit knowledge from data and represent a suitable approach to support decision making to improve manufacturing operations.

First, the description of the problem is provided, and the methods used and the results of data mining and visualization in a case study from a Slovenian engineer-to-order company, which has recently decided to exploit to a greater extent the datasets collected during production. The company produces complex one-of-a-kind products, e.g. water turbines, hydro mechanical equipment, pumping systems and industrial equipment. The workshop departments are grouped by type of working technology. Manufacturing work systems (MWS) are mostly special, high cost, large-scale, CNC machines working mostly on a 24/7 basis. For example, MWSs are large CNC carousel and universal lathes, horizontal and vertical boring and milling machines, electro-powder welding automats, gear shapers, large grinding machines, and so forth. The company was first of all interested in finding some patterns in break and disruption data in order to make improvements or changes in management, planning and control of workshop operations. Another interesting issue suggested by the company was to visualize the data, in particular the utilization of the workshop and MWSs data in order to understand and to get better insight into what is happening on a shop floor.

2. DATA MINING AND VISUALIZATION EXPERIMENTS

In the experiments described two different sources were used. Firstly, a data mining experiment on the so-called LIMES data was conducted. LIMES is a stand alone manufacturing execution information system [5]. It serves for monitoring, control and management of shop floor operations. LIMES system enables also communication and integration with other information systems over the web. Its monitoring module collects data about events and states in operations related to products, processes, resources and quality. It provides access to the machine, process and work order data and to progress reports and diagnostics of disruptions and malfunctions. This information is available for decision making in the workshop and over the web to other decision-making entities, e.g. project management and plant management [6].

Secondly, an experiment on FactoryLink SCADA data was carried out. FactoryLink is a commercial supervisory control and data acquisition system for real-time machine data acquisition and machine and process monitoring and control.

These intuitively related sources (LIMES and SCADA data) are due to the complexity of the production, very hard to bind. In the next two sections both experiments and results are presented.

Data mining experiment on the LIMES data. From the LIMES system, two of the most significant MWSs were taken into observation, namely NSP and NCV. NSP is a vertical CNC lathe Schiess and NCV is a horizontal CNC Milling centre Schiess. The data underlying the analyses was collected over the period from January till December 2005.

Data mining has been carried out using the Orange data mining tool. Orange is a component based data mining software which includes a range of pre-processing, modelling and data exploration techniques. It is based on C++ components, that are accessed either directly (not very common), through Python scripts (easier and better), or through graphical user interface objects called Orange Widgets. Orange is free software, licensed under the GPL license [7].

The experimental data set consisted of only about 500 records per single MWS per year. The raw dataset were organized in a table with the following relevant attributes: id, id_event, event_type, id_machine, operator, event_time_stamp.

In order to analyze the data, a transformation of dataset attributes has been performed. This task was carried out in order to get as much information as possible from the data. The event_time_stamp attribute was transformed into: month {January, February, March, etc.}, season {winter, spring, summer, autumn}, weekend {yes, no} and work shift {morning, afternoon, night}. Furthermore, disruptions of MWSs were grouped into three logical groups: (1) planned breaks, (2) unplanned disruptions and (3) undefined breakdowns. These groups are represented by numbers 4, 5 and 6, respectively.

A subgroup discovery (SD) algorithm [8] was used in our experiment, as implemented in Orange [7,9]. Formally, the task of the subgroup discovery is defined as follows: given a population of individuals and a specific property of the individuals that we are interested in, find population subgroups that are statistically "most interesting", e.g., are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest.

The result of subgroup discovery is a small rule set of rules of the form:

IF Conditions THEN Class

These rules are short, understandable and easy to interpret. The SD algorithm builds rules in a general-to-specific fashion by beam-search of rules in which attribute-value pairs are added to previously induced rules in order to maximize the generalization quotient:

$$q_g = \frac{TP}{FP + g}$$

Target class examples covered by a rule are also called true positives, TP, while non-target class examples covered by the rule are called false positives, FP. High quality rules cover many target class examples and a low number of non-target examples. The number of tolerated non-target class examples, relative to the number of covered target class examples, is determined by the parameter g. For low g, induced rules will have high specificity since the coverage of every single non-target class example is made relatively very “expensive”. On the other hand, by selecting a high g value, more general rules are generated, covering also non-target class instances.

Despite the fact that input data was relatively information poor, some interesting results came out. As shown in Table 1, (Rule 1) on MWS NSP more unplanned breakdowns occur. This rule suggests that NSP should be maintained and observed more carefully, because unplanned breakdowns are more common. Rules 2 and 3 show that planned breaks last longer than unplanned. Results also indicate that February is the month in which most unplanned breakdowns occur (Rule 4). Furthermore, these breakdowns occur more often on normal working days, rather then on weekends (Rule 5). Based on this, it was suggested that the company should increase maintenance and service staff in this respect in order to decrease this type of disruption.

Table 1 – The rules for MWS NSP

R No.	TPr	FPr	Rule
1	0.68	0.37	MWS=NSP -> Br_group=4
2	0.65	0.27	Br_duration=(68.00, 325.00] -> Br_group=4
3	0.53	0.17	Br_duration=(0.00, 68.00] -> Br_group=5
4	0.13	0.03	Month=feb -> Br_group=5
5	0.12	0.03	Month=feb and WeekEnd=false -> Br_group=5

Visualization experiment on SCADA data. The FactoryLink SCADA system periodically (every two minutes) collects operation states of the MWSs. Every record in the SCADA database represents a vector whose dimension equals the number of wired-to-scada MWSs in the workshops. The vector values are of data type Boolean (values 0 and 1 only): 1 means that MWS is utilized at the time of acquisition, i.e. a cutting process is running, while 0 means that MWS is stopped or not in a working state. The data used in the analysis were collected over the period from January till December 2005. During the observed interval SCADA collected more than 244000 records from 42 MWSs.

First, a novel visualization method of workshop utilization was designed, aimed at eliciting useful information about the overall workshop utilization. The plot in Figure 1 shows that the most efficient utilization of the observed MWS (NSP) is obtained on Tuesdays, Wednesdays, Thursdays and Fridays. The highest utilization (between 50% and 70%) is obtained from Tuesday till Saturday during night shift (between 23.00 and 04.00 hours). This was a surprising discovery for the management while it had always been believed that the night shifts are less efficient. The lowest efficiency is obtained on Saturday and Sunday afternoons and Sunday nights, while these shifts are usually omitted. From the plot one can realize that - on average - morning shifts are less effective. The reasons are machining settings, work preparations and maintenance, etc. which are usually performed during the morning shift. One can also notice the morning break time at around 09:00 and therefore decreased utilization. Utilization degradation is also noticed at the time of shift changes (at 06:00 and 14:00 hours). Afternoon-night shift change does not reflect any utilization changes.

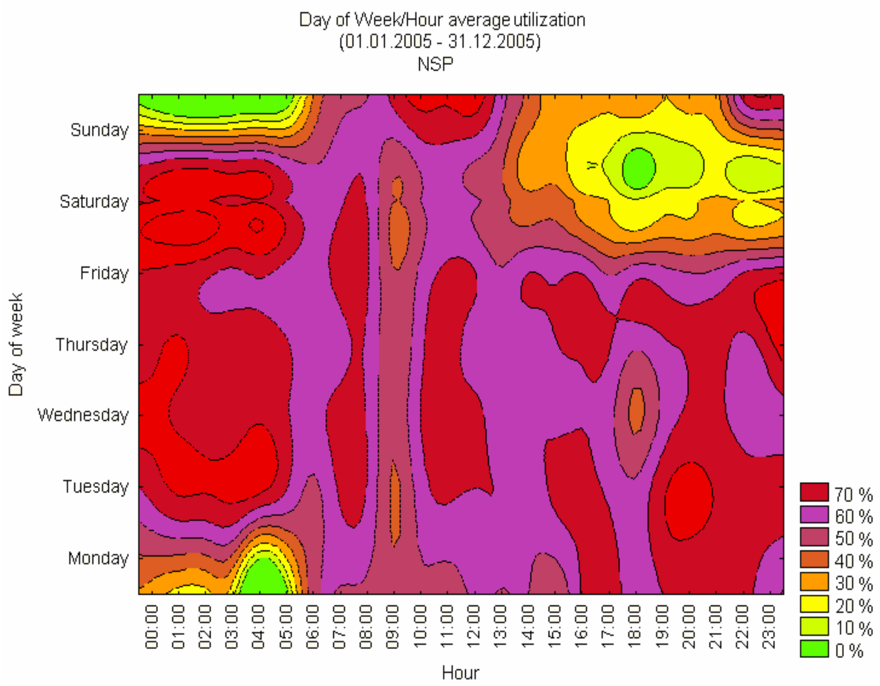

Figure 1. The distribution of the average utilization of the observedMWS NSP for the year 2005.

There were also some trends regarding the regime in which the observed MWS was running. MWSs with 24 hours working regime had plots similar as shown in Figure 1, while a group of MWSs (smaller, not so relevant MWSs) had plots with high utilization in the morning shifts and from Monday till Friday. These MWSs run in a 8-hours (morning shift) regime. The MWSs had the lower left part of the plot similar to the one in Figure 1, while the other part of the plot showed lower utilization.

Figure 1 shows the off-line prototype plot, generated for testing and validating purposes. In order to be more efficient, the next phase of the development will lead to a web-based application with highly settable, on-line, on-the-fly generated plots, available for the management staff. In this way, the workshop managing staff will have on-line information about MWSs utilization, which will help them to make better decisions based on more “objective” information.

4. CONCLUSIONS

Data mining and visualization experiments were conducted on workshop data in the engineer-to-order company. This research has been conducted in order to get some production related patterns, including the information or trends, aimed at gaining some knowledge that could be used to improve operations management.

The data mining experiment, performed on LIMES data, resulted in rules based on which production disruptions can be reduced in order to enhance the utilization of the machinery in the workshop. Visualizations, presented in the paper, gave decision making and planning staff an aggregated view of the workshop and MWSs utilization. Results of the study enabled the managing and planning staff to have a better understanding of the manufacturing processes with respect to machinery conditions and to identify strong and weak points of the production.

Further work will be directed toward creating an on-line, web-based, on-demand and on-the-fly reporting system, which will provide a quick and easy way to retrieve similar plots as presented in the paper. More concern will also be devoted (1) to advanced data acquisition, in terms of improved relation of event driven data collection and periodically collected data, and (2) to a development of a rule based decision system to support more intelligent control of MWS in order to achieve smoother production with near zero downtime.

5 REFERENCES

[1] Reaside R, Walker J, 2001, Knowledge: the key to organizational survival, The TQM Magazine, 13(3): 156-160.

[2] Shin J, Park S, Ju C, Cho H, 2003, CORBA Based Integration Framework for Distributed Shop floor Control, Computers & Industrial Engineering, 45:457-474.

[3] Gertosio C, Dussauchoy A, 2004, Knowledge discovery in industrial databases, Journal of Intelligent Manufacturing, 15: 29-37.

[4] Symeonidis LA, Mitkas AP, 2005, Agent Intelligence through data mining, Springer.

[5] Sluga A, Jenkole J, Butala P, 2005, Web-based shop-floor monitoring and control for new manufacturing structures. CIRP j. manuf. syst. 34(6):589-597.

[6] Husejnagic D, Jenkole J, Butala P, Sluga A, 2005, Performance Assessment of Manufacturing Work Systems Based on Data Mining, Proceedings of the 38th CIRP International seminar on manufacturing systems, Florianopolis, ps. 6.

[7] Demsar J, Zupan B, Leban G, 2004, Orange: From Experimental Machine Learning to Interactive Data Mining, White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana.

[8] Gramberger D, Lavrac N, 2002, Expert-Guided Subgroup Discovery: Methodology and Application, Journal of Artificial Intelligence Research, 17:501-527.

[9] Kralj P, Lavrac N, Zupan B, Gramberger D, 2005, Experimental comparison of three subgroup discovery algorithms: Analysing brain ischemia data, 8th International Multiconference Information Society IS, 220 – 223.