Expert Systems with Applications 40 (2013) 1034–1045

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Mining association rules for the quality improvement of the production process

Bernard Kamsu-Foguem *, Fabien Rigal, Félix Mauget

Laboratory of Production Engineering (LGP), EA 1905, ENIT, INPT, University of Toulouse, 47 Avenue d'Azereix, BP 1629, 65016 Tarbes Cedex, France

ARTICLE INFO

Keywords:
Data mining
Association rule mining
Knowledge discovery
Continuous improvement
Drilling product manufacturing
Industrial maintenance

ABSTRACT

Academics and practitioners have a common interest in the continuing development of methods and computer applications that support or perform knowledge-intensive engineering tasks. Operations management dysfunctions and lost production time are problems of enormous magnitude that impact the performance and quality of industrial systems as well as their cost of production. Association rule mining is a data mining technique used to find out useful and invaluable information from huge databases. This work develops a better conceptual base for improving the application of association rule mining methods to extract knowledge on operations and information management. The emphasis of the paper is on the improvement of the operations processes. The application example details an industrial experiment in which association rule mining is used to analyze the manufacturing process of a fully integrated provider of drilling products. The study reports some new interesting results with data mining and knowledge discovery techniques applied to a drill production process. Experiment's results on real-life data sets show that the proposed approach is useful in finding effective knowledge associated to dysfunctions causes.

1. Introduction

Engineering applications of artificial intelligence have attracted substantial consideration from industrial practitioners and researchers because of its ability to learn and comprehend facts and principles in order to acquire knowledge and apply it in practice. Continuous improvement refers to both incremental and breakthrough improvement in organizational performance (Linderman, Schroeder, Zaheer, Liedtke, & Choo, 2004). Improvement can result in such things as improved customer value, reduction of defects and errors, improved productivity, improved cycle time performance safety, and motivation (Evans & Lindsay, 2001). This often occurs through the adherence to a stepwise problem solving approach consisting of number of steps for problem contextualization, problem analysis, solution generation, and lessons learned (Kamsu-Foguem, Coudert, Geneste, & Beler, 2008). The problem-solving approach focuses on a characterization of cognitive processes in reasoning tasks and cognitive considerations deal with knowledge capitalization on certain structural and processing regularities that give strength to generalizations (Patel, Arocha, & Kaufman, 2001). Problem-solving methods play a significant role in knowledge acquisition and engineering, since their abstract knowledge level is valuable to achieve goals of tasks by applying domain knowledge with the sequential process of searching for a solution path. They can be applied, among others, to describe the reasoning process in a structured manner, to guide the knowledge acquisition process and to facilitate knowledge sharing and reuse (Benjamins & Fensel, 1998).

Problem-solving research places a greater emphasis on an evolving process (e.g. analyzing with a set of tools) rather than a fixed selection process, by application of deductive reasoning (i.e. a specific conclusion is arrived at from a general principle) and inductive reasoning (i.e. a general conclusion is arrived at by specific examples) (Newell & Simon, 1972). A stepwise problem solving model presents a systematic analysis by ensuring multiple perspectives of a problem are captured and engaged in formulation of an insightful solution (Gibbons, 2000). Such systemic approach helps in problem comprehension, aids identification of its root causes and impacts knowledge creation (Jabrouni, Kamsu-Foguem, Geneste, & et Vaysse, 2011). In general, problem-solving studies are more operational in formalizing latent sources of error as well as describing the root causes of problems or events. Besides, experienced knowledge differs in important respects from intermediate knowledge and has a qualitatively distinct engagement with differential use of reasoning strategies in problem solving: for example, experts are involved in the process of situation assessment with a data-driven reasoning whereas the novices and intermediates are much more proactive in handling solution options and organizing further investigations with a hypothesis-driven reasoning (Patel, Kaufman, & Arocha, 2002).

Plan Do Check Act (PDCA), Lean, and Six Sigma are three of the common stepwise models of problem solving used in industry today. Each has its rationale with relevant features, and each approach, when deployed accurately, can yield some interesting results and sustain improvement (as described in Table 1 (Chiodo,

* Corresponding author. Tel.: +33 6 24 30 23 37; fax: +33 5 62 44 27 08.
E-mail address: Bernard.Kamsu-Foguem@enit.fr (B. Kamsu-Foguem).

0957-4174/$ - see front matter
http://dx.doi.org/10.1016/j.eswa.2012.08.039

Table 1
Common continuous improvement methodologies (Chiodo et al., 2011).

	PDCA	Lean	Six sigma
Definition	Cyclical product and/or process improvement emphasis on control	Elimination of waste, speed, efficiency	Reduction in defects and variation data driven
Objective	Small incremental improvements, repeat process	Relentless pursuit or perfection by Increasing value-adding activities by eliminating waste	Reduce process variation to near perfect (Six Sigma) levels
Methodology	Deming-Shewhart PDCA cycle • Plan • Do • Check • Act	Value stream mapping: 5S: • Sort • Straighten • Scrub • Systematize • Sustain	DMAIC • Define • Measure • Analyze • Improve • Control

PDCA, plan, do, check act; DMAIC, define, measure, analyze, improve, control.

Rosenhauer, & Worsowicz, 2011). The stepwise fashion of problem solving and the associated continuous improvement methodologies can be used at distinct levels of organization, in service and administrative as well as manufacturing processes. Quality management practices that promote monitoring and experience feedback of information and operations management allow learning and knowledge creation (Choo, Linderman, & Schroeder, 2007). Many quality control and improvement activities (e.g. inspection/screening, quality analysis, process control, quality monitoring) that are related to manufacturing problems utilize data analysis methods to mine huge data sets collected through production processes in manufacturing industry. The ideas for improvement provided by such activities are a key element in the experience feedback process to further corrective or preventive actions (Foster, 2008). The generated ideas can be tested through the use of data analysis techniques that link continuous improvement to knowledge creation processes.

Knowledge Discovery in Databases (KDD) has become one of the fastest growing research topics in mathematics and computer science, because the ability to continually change and acquire new understanding is a driving force for its applications (Liao, Chu, & Hsiao, 2012; Washio, 2007). For example data mining have served in the search and retrieval of computer-aided design elements (Liu, McMahon, Ramani, & Schaefer, 2011). The KDD process, specifically data mining techniques, is used to characteristically discover knowledge from data (Zhu & Davidson, 2007). The data mining process extracts knowledge from an existing data set and transforms it into a human-understandable structure for further use (Witten, Frank, & Hall, 2011). Data mining techniques are required to help in identification of model characteristics important to capture and document in an enhancement context of the safety and reliability of complex engineering systems (Saitta, Raphael, & Smith, 2005). Data mining applications are very suitable for quality improvement programs (e.g. Kaizen-PDCA, 9-Steps, 8D, 7-Step, PDCA, Six Sigma-DMAICS) in manufacturing (Köksal, Batmaz, & Testik, 2011), due to advances in data collection systems, analysis tools and interpretation methods (Alzghoul & Löfstrand, 2011). However, there are some factors influencing the adoption of data mining tools (DMTs), primarily the task-oriented dimension (job relevance, output quality, result demonstrability, response time, and format) (Huang, Liu, & Chang, 2012). So, it is decisive to ensure good means of promoting, efficient and effective information access, processing, and use by people and organizations (Detlor, 2010). Data mining involves six common classes of tasks (Kantardzic, 2011) (Ngai, Hu, Wong, Chen, & Sun, 2011):

- *Anomaly detection* (outlier/change/deviation detection) – Anomaly detection is engaged to identify the unusual data records and to detect data objects that are unacceptably different from or inconsistent with the remaining data set. A system protection method can be applied for detecting anomalies in user patterns, with the purpose to provide guidance for facilitating the reconfiguration of collaboration systems (Lee, Ryu, Shin, & Cho, 2012).
- *Association rule mining* (association rule learning) – Association rule learning is employed to discover interesting relations between variables in large databases. This dependency modeling analyses strong rules discovered in databases using different measures of interestingness. The use of association rules mining in frequent patterns captured from industrial processes can provide useful knowledge to explain industrial failures (Martínez-de-Pisón, Sanz, Martínez-de-Pisón, Jiménez, & Conti, 2012).
- *Clustering* – Clustering serves to partition objects into conceptually meaningful groups (clusters), such that similar objects are in the same group, while dissimilar objects are in different groups. Clustering is an unsupervised learning problem where one is only given the unlabeled data and the goal is to learn the underlying structure. A graph clustering algorithm approach for manufacturing cell formation can be used to makes an improvement in the number of intercell moves (Oliveira, Ribeiro, & Seok, 2009).
- *Classification* – Classification is the procedure of assigning labels to objects such that objects' labels within the same categories will match previously labeled objects from a training set, by generalizing known structure. Classification is traditionally a type of supervised learning problem that tries to learn a function from the data in order to predict the categorical labels of unknown objects to differentiate between objects of different classes. Classification procedure can be employed to assist decision makers to classify alternatives into multiple groups, reduce the number of misclassifications and lessen the impact of outliers (Ma, 2012).
- *Regression* – Regression is a statistical methodology for modeling and analyzing several variables and is used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. Most commonly, regression analysis attempts to find a function of the independent variables that models the data with a method of estimation. In the work of Alzghoul and his colleagues, different data-stream-based linear regression prediction methods have been tested and compared within a newly developed fault detection system (Alzghoul, Löfstrand, & Backe, 2012).
- *Summarization* – Summarization is related to the effortlessly understandable presentation of data and to methodology that converts intricate data characteristics into explicit patterns that can make sense to users. It provides a more concise and intelligible representation of the data set, including visualization and report generation in digest form. A visual data mining approach can be suitable for building knowledge base in shop floor control systems of semiconductor wafer fabrication (Shiue, Guh, & Tseng, 2012).