J. of Comp. and I.T. Vol. 3(1&2), 13-18 (2012).

Data Mining: A Tool for Knowledge Management and Eco- Friendly Sustainable Development

NEERAJ BHARGAVA[1], RAJEEV ARYA[2], ANIL RAJPUT[3] and ATIF AZIZ[4*]

[1]Department of Computer Science, School of Engineering & System Sciences, M.D.S. University, Ajmer, Rajasthan (INDIA)
[2]Department of Mechanical Engineering Truba college of Science &Technology Bhopal (INDIA)
[3]Department of Mathematics and Computer Science Government college Sehore (INDIA)
[4]Department of Mechanical Engineering Truba Institute of Engineering and Information Technology Bhopal (INDIA)

(Acceptance Date 6th January, 2012)

Abstract

The paper reviews the scope of data mining which leads to knowledge management in engineering and science field. A vast change can be brought in the field of engineering design, manufacturing Systems, decision Support Systems, Shop Floor Control and Layout Customer Relationship Management**,** banking, finance, retail, marketing, insurance, fraud detection, medical sciences, semiconductor industry aerospace industry etc for sustainable development at low cost using artificial intelligence. This can be by implementing the most appropriate data mining and knowledge management software in the organization. This would give competitive advantage to the business over its competitors and also reduce the overall cost of research and development of the organization. Nowadays the market is flooded with enormous number of Data Mining tools (software) and Knowledge Management software. There is need to study and compare the data mining tools. The study would be of helpful recommending a particular type of tool according to the requirement of the organization.

Introduction

With the rapid development of computer and information technology in the last several decades, an enormous amount of data in science and engineering has been and will continuously be generated in massive scale, either being stored in gigantic storage devices

*Address for Correspondence: HIG D-18 Kahkashan Complex phase 2 BDA colony Kohefiza Bhopal India 462001 Email:atifaziz82@gmail.com mobile:+91-9425605537

and out of the system in the form of data streams. Such tremendous amount of data, in the order of tera- to peta-bytes, has fundamentally changed science and engineering, transforming many disciplines from data-poor to increasingly data-rich, and calling for new, data-intensive methods to conduct research in science and engineering.[1] Data mining techniques have been used to discover various biological, drug discovery, and patient care knowledge and patterns using selected statistical analyses, machine learning, and neural networks methods. Biomedical entities such as drug names, proteins, genes, and diseases can be automatically extracted from published documents and used to construct gene pathways or to provide mapping into existing medical researchers.[2] Sustaining a competitive advantage in firms demands a combination of three unique prerequisites. These unique elements are skilled and capable people, an organizational culture focused on learning, and the use of leading-edge information technology tools for effective knowledge management.[9]

It is important to examine the challenges on data mining posed in data-intensive science and engineering and explore how to further develop the technology to facilitate new discoveries and advances in science and engineering. Data mining is a blend of concepts and algorithms from machine learning, statistics, artificial intelligence, and data management. With the emergence of data mining, researchers and practitioners began applying this technology on data from different areas such as banking, finance, retail, marketing, insurance, fraud detection, science, engineering, etc., to discover any hidden relationships or patterns. Data mining is therefore a rapidly expanding field with growing interests. *Manufacturing is an application area where it can provide significant competitive advantage*. Areas such as manufacturing operations, fault detection, design engineering, and decision support systems have gained the attention of the research community, although there is still enormous potential for research in these areas. Other areas like maintenance, layout design, resource planning, and shop floor control require even greater attention and further exploration[4] Data mining uses a combination of an explicit knowledge base, sophisticated analytical skills, and domain knowledge to uncover hidden trends and patterns. These trends and patterns form the basis of predictive models that enable analysts to produce new observations from existing data. Data mining is the search for relationships and global patterns that exist in large databases, but are "hidden" among the vast amount of data. These relationships represent valuable knowledge about the database and objects in it. Data Mining tends to produce understanding the relationship between information which in turn is used in understanding patterns out of information and these patterns become knowledge, this helps in formulating principles by Knowledge Management which goes to make proper judgment. Knowledge Management (KM) is a emerging field which can contribute a lot to engineering field. Knowledge originates in the head of an individual (the mental state of having ideas, facts, concepts, data and techniques, as recorded in an individual's memory) and builds on information that is transformed and enriched by personal experience, beliefs and values with

decision and action-relevant meaning.[3]. Knowledge exists in all business functions, including purchasing, marketing, design, production, maintenance and distribution, but knowledge can be notoriously difficult to identify, capture, and manage. Knowledge management tools are defined as tools which support the performance of applications, activities, and actions such as knowledge generation knowledge codification, and knowledge transfer, they also promote and enable the knowledge process in order to improve decision making.[7] Knowledge should be managed as an asset or resource, just like land, capital, and labour. This is a shift away from how to simply obtain knowledge to how to use it productively. Knowledge is to be seen as an activity as well as an object; it is a product and a process.[8] The model organizes knowledge flows into four primary activity areas: knowledge creation, retention, transfer and utilization[6]

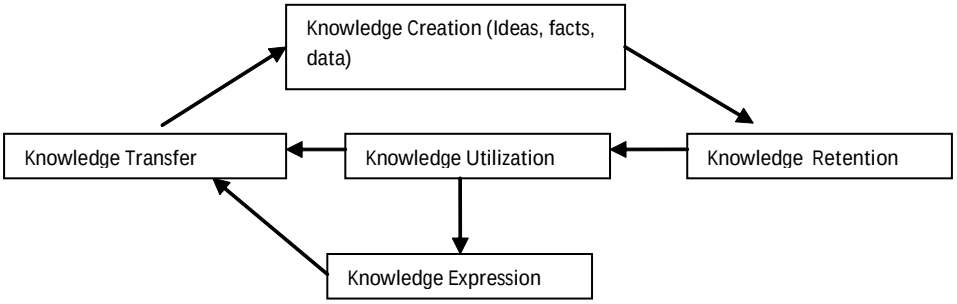

Fig. 1. Knowledge flow

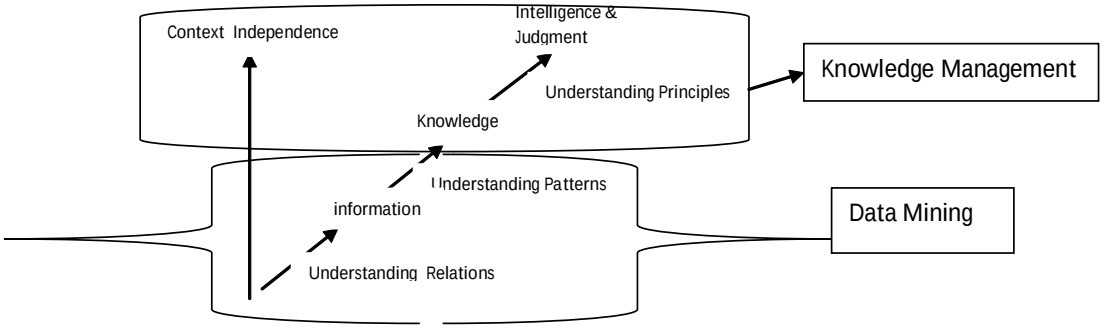

Fig. 2. Data Mining and Knowledge Management

Data Mining Applications :

Engineering Design: Engineering design is a multidisciplinary, multidimensional, and non-linear decision-making process where parameters, actions, and components are selected. This selection is often based on historical data, information, and knowledge. It is therefore a prime area for data mining applications and although as yet only a few papers have reported

applications of data mining in engineering design[4].

Manufacturing Systems: Data collection in manufacturing is common but its use tends to be limited to rather few applications. Machine learning and computational intelligence tools provide excellent potential for better control of manufacturing systems.[4]

Decision Support Systems: Knowledge is the most valuable asset of an organization. Decisions are made based on a combination of judgement and knowledge from various domains. Decision support, knowledge management, and processing are interdependent activities in many organizations. Ideally, all relevant knowledge should be available before making a decision. The knowledge extracted from databases can be integrated with existing expert systems.[4]

Shop Floor Control and Layout: The shop floor control and layout problems are concerned with the efficient and effective utilization of resources, at the lowest level of control in manufacturing. A vast amount of data is recorded during the operation of a shop floor, often to ensure that parts and production steps can be traced. These data can also be used to optimize the process itself, since the knowledge generated from mining historical work-in process data helps in characterizing process uncertainty and parameter estimation of the system concerned.[4]

Fault Detection and Quality Improvement: Fault diagnosis is an area that has seen some of the earliest applications of data mining.

Maintenance: Preventive maintenance is of key importance in process and manufacturing engineering. Databases containing the events of failure of the machines and the behavior of the relevant equipment at the time of the failure can be used in the design of the maintenance management systems[5].

Customer Relationship Management: The marketing model has shifted from product-focused to customer-focused. Customer Relationship Management (CRM) is concerned with increasing the value of interaction with customers and maximizing the profit. CRM is as important as producing high quality and low cost products and is complementary to demand management which may be defined as a set of practices aimed at managing and coordinating a demand chain, starting from the end customer and working backwards to raw material and suppliers. Similarly, in service industries, data from customers is the only source of knowledge about their satisfaction with the product.[4] An exponential growth of data mining applications in the semiconductor industry has been observed. The reasons for this may be that large volumes of data are generated during manufacture and that small improvements can have a significant impact in this industry. No other sector of manufacturing industry reports such large increases of data mining applications. This is rather surprising as other industries such as aerospace routinely collect huge quantities of data during product manufacture and hence are good potential environments for data mining studies. Many reported applications are related to the causes of malfunctioning of different

types of manufacturing systems or processes and hence the discovered knowledge leads toward the better functioning of the manufacturing enterprise. Developments in data mining are generally directed at the refinement of algorithms and their application in manufacturing, their integration with existing systems, standardization, the use of common methods and tools, and the definition of repeatable projects. Future research effort is therefore also needed to enhance the expressiveness of the knowledge. The development and application of data mining algorithms requires the use of powerful software tools. As the number of available tools continues to grow, the choice of one special tool becomes increasingly difficult for each potential user. This decision making process can be supported by criteria for the categorization of data mining tools.[10] These criteria can be based on user groups, data structures, data mining tasks and methods, import and export options, and license models. Many advanced tools for data mining are available either as open-source or commercial software. Recent tools are able to handle large datasets with single features, time series, and even unstructured data-like texts.

Conclusion

The tools of data mining should be able to detect the necessary information from the available data .To achieve this, a more generic process for data cleaning is essential to enable the growth of data mining in industry. The data-mining research often does not consider the quality of the rules or knowledge discovered. The knowledge generated is sometimes cumbersome and the relationships obtained are too complex to understand. Future research effort is therefore also needed to enhance the expressiveness of the knowledge. Data mining tools should have some characteristic like: User friendly environment, Efficiency and effectiveness of the tool in finding patterns, Basic task of knowledge management should be accomplished, Low cost of implementation, low maintenance cost. There is a lack of powerful and generalized mining tools for multidimensional datasets such as images and videos. There is a need to study the various data mining tools available in the market so that the most appropriate tool can be selected for the organization as per the requirement.

References

1. Jiawei Han, and Jing Gao, Research Challenges for Data Mining in Science and Engineering Chapter 1 1-18
2. Hsinchun Chen, Sherrilynne S. Fuller, Carol Friedman, and William Hersh, Knowledge Management, data Mining, and Text Mining in Medical Informatics chapter 1 3-5
3. Apurva Anand, M.D. Singh Understanding Knowledge Management: a literature review *International Journal of Engineering Science and Technology* Vol. *3(2),* 926-939 (2011).
4. Harding, J., A., Shahbaz, M., Srinivas, Kusiak, A., Data Mining in Manufacturing: A Review *Journal of Manufacturing Science and Engineering* Vol. *128,* 969-976 (2006).
5. Jayanthi Ranjan, A Review of Data Mining Tools In Customer Relationship Management *Journal of Knowledge Management*

Practice, Vol. 9(1) (2008).
6. Brian (Bo) Newman, Kurt W. Conrad, A Framework for Characterizing Knowledge Management Methods, Practices, and Technologies (1999).
7. Peter Tyndale, A taxatomy of knowledge Management software tools: origins and applications *Evaluation and program planning* Vol. *25,* 183-190 (2002).
8. Syed Raiyan Ghani, Knowledge Management: Tools and Techniques *Journal of Library & Information Technology,* Vol. *29(6),* 33-38 (2009).
9. John H. Heinrichs, Jeen-Su Lim: Integrating web-based data mining tools with business models for knowledge management Decision Support Systems Elsevier Science B.V. Vol. *35*, 103-112 (2003).
10. Ralf Mikut, Markus Reischl: Data mining tools advanced review John Wiley & Sons, Inc. Vol. *1,* 431-443(2011).