Reffat, R., Gero, J. and Peng, W. (2004) **Using data mining on building maintenance during the building life cycle**, *Proceedings of the 38th Australian & New Zealand Architectural Science Association (ANZASCA) Conference*, School of Architecture, University of Tasmania, Tasmania, Australia, pp. 91-97.

Using data mining on building maintenance during the building life cycle

Rabee M. Reffat[1], John S. Gero[2] and Wei Peng[2]

[1]Department of Architecture, Assiut University, Assiut, Egypt
[2]Key Centre of Design Computing and Cognition, University of Sydney, NSW 2006 Australia

ABSTRACT: The data generated within the construction industry has become increasingly overwhelming. Data mining technology presents an opportunity to increase significantly the rate at which the volumes of data generated through the maintenance process can be turned into useful information. This can be done using classification algorithms to discover patterns and correlations within a large volume of data. This paper investigates the potentials of applying data mining techniques on maintenance data of buildings to identify the impediments to better performance of building assets. It demonstrates what sorts of knowledge can be found in maintenance records. The benefits to the construction industry lie in turning passive data in databases into knowledge that can improve the efficiency of the maintenance process and of future designs that incorporate that maintenance knowledge.

Keywords: Building Maintenance, Building Life Cycle and Data Mining

INTRODUCTION

The growth of many business, government, and scientific databases has begun to far outpace human's ability to interpret and digest the data. Such volumes of data clearly overwhelm the traditional methods of data analysis such as spreadsheets and ad-hoc database queries. The traditional methods can create informative reports from data, but cannot analyse the contents of those reports. A significant need exists for a new generation of techniques and tools with the ability to automatically assist humans in analysing the mountains of data for useful knowledge. The increasing use of databases to store information about building facilities, their use, and their maintenance provides the background and platform for the use of data mining techniques for future projections. The current technology for facility maintenance uses databases to keep track of information and for notification of maintenance schedules.
Knowledge Discovery in Databases (KDD) and Data Mining (DM) are tools that allow identification of valid, useful, and previously unknown patterns so that the facility manager may analyse the large amount of project data. These technologies combine techniques from machine learning, artificial intelligence, pattern recognition, statistics, databases, and visualization and can help to automatically extract concepts, interrelationships, and patterns of interest from large databases. These techniques find patterns in data that can assist in planning. The need to feed back the patterns of maintenance and other facilities management activities in existing facilities to improve future maintenance and the design of new facilities can be addressed by applying data mining techniques to the records of existing facilities.

1. BUILDING MAINTENANCE

The primary objective of building maintenance is that building systems and components should be always functioning to support building operations. The challenge is to allow the building operating managers and the owner to share information about the current status of building systems and components and the business situation to plan maintenance to meet this time-varying objective. Supporting this objective is difficult. It is difficult to assess the amount of risk posed by an observed non-critical problem to future production. There are multiple goals (e.g., high long-term availability, minimal short-term cost); goals change (e.g., between availability and cost concerns); goals conflict; indicator data are almost never completely reliable or adequate. The problem has multiple aspects, including interpretation of observed data, diagnosis of problems, repair and maintenance planning, and business evaluation of the value-added of different repair and maintenance options. Finally, significant judgment is needed to interpret both available engineering and business data, and clear business policy is needed to define the "value" of maintenance.

1.1 Corrective and Preventative Maintenance

Considering the example of a boiler that has tripped due to a drop off of the primary airflow signal, the technician goes out to the unit and, seeing a differential pressure of zero, he changes out the transmitter. Repair work such as this, performed following breakdown, is called Corrective Maintenance (CM), also known as reactive maintenance. This paradigm can be described as “fix it when it breaks”. The technician wants to prevent this from becoming a problem again and schedules a check on the boiler transmitter every six months. Maintenance performed periodically based on utilisation metrics such as hours of operation or calendar time in order to prevent failures is called Preventative Maintenance (PM), also known as scheduled maintenance. This paradigm can be described as “fix it at regular intervals”. While preventative maintenance may avoid the unscheduled down time and costly repairs associated with reactive maintenance, it may be scheduled more often than is necessary. Preventative maintenance is not needed most of the time it is performed, thus introducing costs that are to some degree unwarranted and therefore should be minimised without sacrificing plant performance.

1.2 Predictive and Proactive Maintenance

Continuing with the boiler example, once the transmitter has been replaced, the boiler is started up again, but after a short period of time it again trips due to a zero primary airflow signal. Now looking at the situation more closely, the technician finds that the impulse lines from the transmitter to the orifice plate had become plugged and were the true source of the problem. The lines are cleaned out and the unit is brought back up. The technical team now decides to add on-line instrumentation that measures the boiler’s airflow and sends the signal to a processor that monitors this and other signals and uses a model of the system in order to anticipate and diagnose failures before they occur; this is called Predictive Maintenance (PdM). By isolating faulty components and calculating the best time to repair or replace them, this approach minimises both maintenance labour and the risk of an unscheduled outage and can be characterised as “fix it just before it breaks”.

In traditional approaches, the interpretation of a measurement had no correlation to previous tests. As long as an asset’s parameters were found to be within specification, it was considered to be fine and no action was taken. In current approaches, when a Computerised Maintenance Management System (CMMS) issues a work order, it identifies the specific test, instrument, and settings to use for the piece of equipment in question. A technician performs the test with the same settings and procedures used the time before. After the test, the technician can download data to a system that aggregates data from previous tests and builds trend reports. For more than a decade, PdM has enabled facilities technicians to identify and solve problems before they have a chance to damage equipment. In Proactive Maintenance, the general time frame that a component will fail is determined before the failure occurs or is about to occur. Typically, a model of the system is used to anticipate failure. This approach minimises the risk of failure by eliminating root causes. Using data mining can allow proactive maintenance to occur based on the knowledge and patterns extracted from past maintenance records, thus allowing for generalisation from experience of similar building systems and components as well as pinpointed information specifically relevant to a particular system or component.

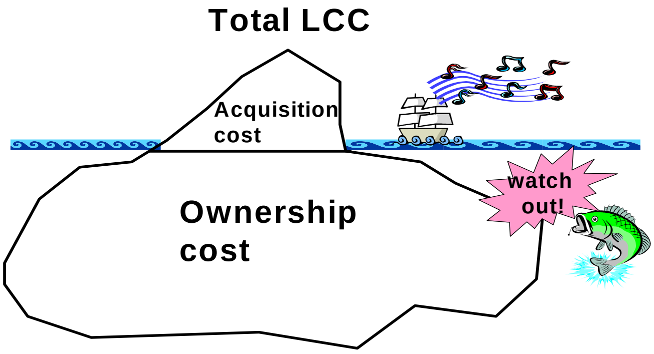
2. LIFE CYCLE OF BUILDINGS

2.1. Life cycle cost

Life cycle costs (LCC) are summations of cost estimates from inception to disposal for both equipment and projects as determined by an analytical study and estimate of total costs experienced during their life. LCC not only comprises initial acquisition cost, it also contains other cost like ‘ownership cost” – operation costs, maintenance costs, logistics costs, etc, which is usually higher than the original acquisition cost. The major objective of LCC analysis is to choose the most cost effective approach from a series of alternatives so the least long term cost of ownership is achieved (Barringer and Weber 1996).It is believed based on Life Cycle Cost analysis (Kawauchi and Rausand 1999) that a typical range of the ownership costs is 60 percent to 80 percent of the total LCC. An interesting view (the iceberg model) that may provide a general understanding of those underlying costs that seems overlooked is shown in Fig. 1.

2.2. Life Cycle Modelling of Buildings

The life cycle cost concept is addressed in the British Standards as ‘Terotechnology’ which is defined as a combination of management, financial, engineering, building and other practices applied to physical assets in pursuit of economic life-cycle costs. Life cycle cost modelling (LCM) contributes to competitiveness of the company by providing strategic planning on rehabilitation and enhanced information for decision making. LCM helps facility manager in evaluating alternative equipment and process selection based on total costs rather than the initial purchase price. The multidimensional information that LCM presents is merged from hybrid project domains such as management, engineering, as well as finance. LCM may be applied in a wide range of critical functions including: (a) evaluation and comparison of alternative design; (b) assessment of economic viability of projects and products; (c) identification of cost drivers and cost effective improvements; (d) evaluation and comparison of alternative strategies for product use, operation, test, inspection, maintenance, etc.; (e) evaluation and comparison of different approaches for replacement, rehabilitation/life extension or disposal of aging facilities; (f) optimal allocation of available funds to activities in a process for product development; (g) assessment of product assurance criteria through verification tests and their trade-offs; and (h) long-term financial planning.



Source: (Dangel 1969)

Figure 1: An interesting view – the iceberg model – of life cycle cost

With the booming of information technology at the end of last century, the increase of information availability has become a dilemma due to inefficiency in processing the information for decision-making. This problem becomes critical in the building industry when the high degree of complexity of work flows involved in and the accompanying uncertainty for decision making in the lifetime of a building are considered. Thus, efficiently dealing with information from different stages of a building's life cycle to improve profitability, productivity as well as strategic resource planning has been turning out to be the business drive for life cycle modelling.

However, the design of new buildings and facilities tends to focus on short-term cost and the immediate needs of the owner for a building that meets various business and functional requirements. Current technologies such as Computer-Aided Design (CAD) have focussed on the needs of designers to develop designs that meet design briefs that do not include life cycle design. Very little attention has been given to modelling of the life cycle costs of buildings at the design and management stages to forecast and achieve the most economical life cycle cost.

There are several life cycle models available for buildings as a whole and for their component systems; although there is no one model that has been accepted as a standard, there are some areas of commonality. Life cycle cost models form predictions based on several parameters, some of which include a degree of uncertainty, such as the reliability of a part. These inputs can range from the cost of installation to the cost associated with carrying spare parts in inventory (Siewiorek and Swarz 1982). By accurately predicting failure rates and repair costs, it is also possible to compute the optimal schedule of preventative maintenance for each asset. What can be predicted and the accuracy of those predictions depends of course on the availability and accuracy of the maintenance data that is available. Furthermore, current life cycle modelling systems fail to provide a seamless integration of hybrid information that provides users access to previously unreachable knowledge.

3. EXISTING COMPUTATIONAL SYSTEMS APPLIED ON BUILDING MAINTENANCE

3.1. SERFIN project

The SERFIN Project is developed at the KBS-Media Lab, Lund University (Christiansson 1998). SERFIN provides facilities management knowledge handling that aims to: (a) identify and capture problems that arise in connection with technical maintenance of buildings; (b) make the problem solution process arising during maintenance of buildings more effective; and (c) make experiences from technical maintenance easily available and accessible in time and space. SERFIN is currently based on web pages on the Internet together with text query. The Demonstrator provides advice on how to fix an item in need of repair. SERFIN is used for searching building maintenance information with the use of hierarchical menus to choose: Building Part, Material, Environment, Problem and Action. The result of search includes similar documents are presented in order of relevance.

3.2. Intelligent Real-Time Maintenance Management System

The Intelligent Real-Time Maintenance Management (IRTMM) System is a project of the Centre for Integrated Facility Engineering (CIFE) at Stanford University. The objective of the system is to perform *value-based* plant maintenance as needed (Kunz et al 1995). The IRTMM system provides integrated subsystems for the following aspects of the maintenance and repair planning problem:

Situation Assessment (SA): interprets observed data as being normal or abnormal and diagnoses causes and effects of plant equipment problems. It analyses system performance to identify indications for condition-based maintenance. Given specialised input data from instrumentation (e.g., pressures, flow rates) and expert diagnostic systems (e.g., vibration analysis), the SA focuses on systems diagnosis. It identifies root causes and effects of component problems, where some of those causes and effects are in the component with a problem, and others are in subcomponents or connected systems. The SA uses a combination of methods for assessment: (a) model-based diagnosis to identify details of possible problems; (b) heuristic classification to identify idiosyncratic problems; and (c) case-based reasoning to compare observed data with previous cases.

Planning: Given a set of problems to repair, provided by the SA or by a user who is considering maintenance for any reason, the planner builds plans of the activities needed to repair a diagnosed problem. The planner can also merge related plans for the same or different components that can be performed during the same plant outage. Specifically, it identifies required activities to perform particular repairs, suggests alternative start times, identifies plans that can be merged, and identifies dependencies among plans that imply ordering of the sequence of plans.

Value Analysis (VA): identifies the dollar costs and predicted benefits of performing every selected repair plan at different times. The user can consider the repair for any reason. Using predicted power demand and plant operating and maintenance cost data, the VA assesses the net unit operating costs associated with performing plans at different possible future times. The VA uses a decision-analysis procedure: it considers possible choices of repair actions and the chance outcomes that can arise given any choice. The value of a choice depends both on the choices and chances and on the probabilities and costs of each occurrence. Since it is usually impossible to get good probabilities, the system identifies "break-even" probabilities of failure such that an owner is indifferent between two options. The user then judges whether actual probability of failure exceeds the break-even.

The IRTMM system provides interactive analyses to facilitate engineering decision-making, not automate it. Given some data from a data acquisition system, the system identifies candidate causes of problems and predicted effects. The user selects one or more components to analyse in more detail. After reviewing the system-generated plans and value analysis, the user selects the one or more components for repair, selecting both the desired repair activity and the planned repair time, after considering the system-generated options. The system is designed to reside on a computer network. It can receive component status information from an on-line data acquisition system and any available diagnostic expert systems, staff and equipment availability information from a computerized maintenance management system (CMMS), and projected product demand, cost and selling price data from a business database. Recommended work could be logged in the CMMS.

4. APPLICATION OF DATA MINING ON BUILDING MAINTENANCE

Past experience often plays a very important role in enhancing the building maintenance. “How often will a building system or a component need repair?” or “How much time is this repair going to take?” are the types of questions that project managers face daily in their planning activities. Failure or success in developing good schedules, budgets and other project management tasks depends on the project manager's ability to obtain reliable information in order to be able to answer these types of questions. Students and young practitioners tend to rely on information that is a regional average provided by various publishing companies. This is in contrast to experienced project managers who tend to rely heavily on their personal experience. Another aspect of building maintenance is that facility managers seek to improve the available scheduling algorithms, estimating spreadsheets and other project management tools. Such a “micro-scale” level of research is important in providing the required tools for the project manager's tasks. However, even with the best such tools, low quality input information will produce inaccurate schedules and budgets as output. Thus, it is also important to have a broad approach of research in a "macro-scale" level.

The advancements of data collection, the introduction of bar codes for almost all commercial products, and computerisation in the Architectural, Engineering, and Construction (AEC) industry have generated a flood of data. Substantial development of data storage technology, such as faster, higher capacity, and cheaper storage devices , better database management systems, and data warehousing technology, have allowed the transformation of this enormous amount of data into computerized database systems. As the AEC industry is adapting to new computer technologies in terms of hardware and software, computerised building data is becoming more and more available. However, in most cases, this data may not be used, or even properly stored for several reasons (Soibelman and Kim 2002): (a) project managers do not have sufficient time to analyse the computerised data; (b) the complexity of the data analysis process is beyond the capabilities of the relatively simple building maintenance systems commonly used; and (c) there has been no well defined automated mechanism to extract, pre-process and analyse the data and summarise the results so that the site managers can use it.

However, there is a great deal of valuable knowledge that can be obtained from an appropriate use of this data; there is a need to analyse this increasing amount of available data and Data Mining can be applied as a powerful tool to extract relevant and useful information from this sea of data. Data Mining (also known as Knowledge Discovery in Databases, or KDD) has been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" (Frawley et al 1992). It uses machine learning, statistical and visualisation techniques to discover and present knowledge in a form that is easily comprehensible to humans. Mining data will enable us to understand how systems that were once thought to be completely chaotic actually have predictable patterns (Peitgen et al 1992). Through KDD, the patterns and causal relationships behind apparently random data in AEC projects can be found. By applying KDD to identify novel patterns, project managers will be able to build knowledge models that may be used for the recurrent activities of on-going construction projects, as well as for a future project activities, and avoid unanticipated consequences (Soibelman and Kim 2002). It presents a significant potential for addressing the problem of transforming knowledge implicit in data into explicit knowledge for decision makers.

In contrast to traditional methods of statistical data analysis, KDD is an automated process that discovers new trends and patterns without the need for human intervention. KDD takes input variables whose relevance may not be obvious to a designer but which becomes evident as result of this process. In addition, KDD makes no prior assumption about the probability distribution of the input variables (Gaussian, Poisson, etc.), as is required in statistics, and is therefore more robust and general. However, like other methods, the process of transforming the data to be in a format suitable for knowledge discovery is not automated and has a large impact on the results obtained. The pre-processing of data, shown schematically in Fig. 2 is an important step in a successful application of data mining.

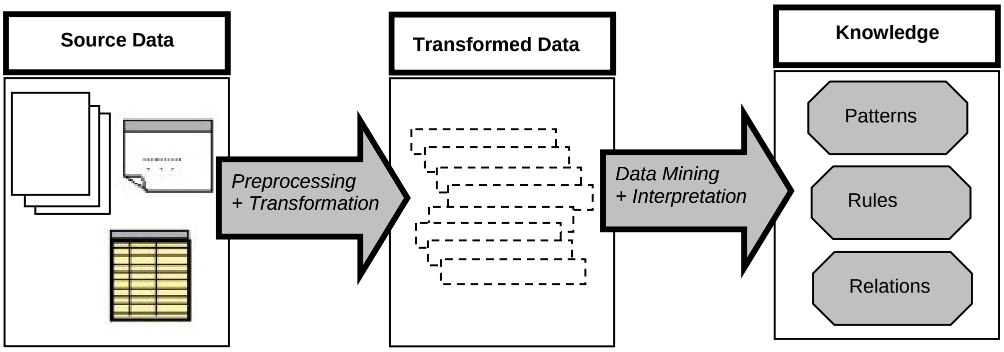


Figure 2: Pre-processing and transformation of data into a format suitable for data mining

5. A DATA MINING APPROACH ON BUILDING MAINTENANCE DURING BUILDING LIFE CYCLE

The approach presented in this paper is based on a comprehensive view of the building management problem. It views the process of building design, maintenance, and replacement as a process generating an enormous amount of information. While current practice addresses parts of this information generation and management, our approach attempts to account for the life cycle flow of this information. The costs of designing and building structures are much smaller than the costs of operating a building or other structure over the course of its life span. The knowledge that becomes available through data mining enables a building owner to make important decisions about life cycle costs in advance, thereby significantly affecting and improve design decisions. The rich set of building data that is created during the design and documentation phase of the building remains relevant even after the building is constructed, and the data only becomes richer as maintenance data is added. Architects, interiors designers and engineers, as well as contractors, marketing and sales personnel, and building managers and owners can extract information from the databases for the building's renovation, maintenance, and operation. Fig. 3 outlines the proposed model of the flow of information in building design and maintenance. The bold arrows depict the functionality provided in this approach while the dashed arrows describe the scope of present approaches to building information management.

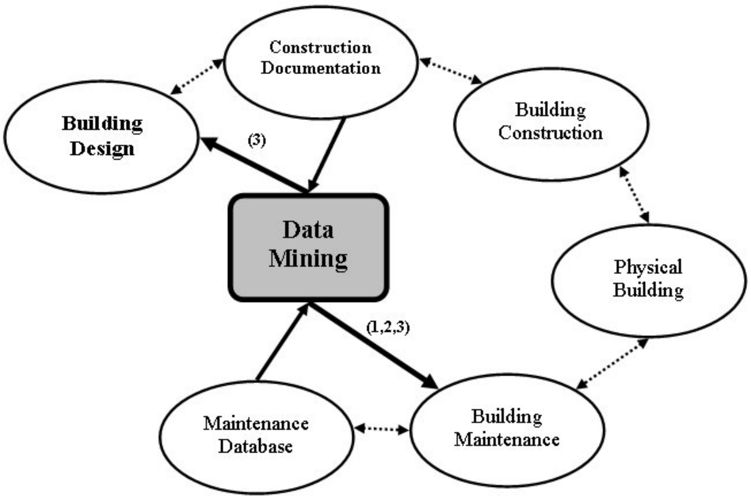


Figure 3: The role of Data Mining in the building life cycle

Data mining techniques can be used effectively on data stored in a Building Maintenance System (BMS) by creating knowledge that can be used in future management and design decision making. Knowledge that implicitly resides in BMS databases includes information about: (1) components that frequently need maintenance and therefore need to be inspected carefully; (2) historical consequences of maintenance decisions that may inform future decisions; and (3) components of buildings that significantly determine maintenance cost and therefore may inform future building designs, as well as refurbishment of the building in question. This information can be extracted using Data Mining techniques and used to improve all phases of the building life cycle, both for current and future buildings, as indicated by the numbers on the arrows in Fig. 3 that correspond to the three types of knowledge listed above.

The quality problems in the AEC industry are very costly; it has been shown that major factors contributing to construction quality problems include inadequate information and poor communication (Arditi and Gunaydin 1998). The detection of previously undiscovered patterns in BMS data can be used to determine factors such as the cost effectiveness and expected failure rate of assorted building materials or equipment in varying environments and circumstances. These factors are important throughout the life cycle of a building, and such information could be used in the design, construction, refurbishment, and maintenance of a building site, representing a substantial decrease in cost and increase in reliability. The Data Mining process can be used to identify the causes of problems such as cost overrun and quality control/assurance. Predictable patterns may be revealed in construction data that were previously thought to be chaotic.

5.1 Analysis of data mining software

There are currently hundreds of mining tool vendors. A large number of reviews of data mining software is also available at (Goebel and Gruenwald 1999; Elder and Abbott 1998). The six leading data mining tools have been reviewed and illustrated in Table 1. Table 2 shows comparisons between these six tools on classification methods. It is noted that commercial data mining vendors in many cases do not disclose the nature of the algorithms. Therefore, our review of these tools from the point of the algorithms they are using is somehow limited. Also in the last three years, a significant convergence emerged between the major tools that not only look very similar from user point of view but also the performance appears very similar as well. Based on this analysis, WEKA (Witten and Frank 2002) has been chosen since it is better in scalability and functionality than the other packages and WEKA is freely available. Furthermore, the WEKA package is an open source software built in Java which can easily embedded into a pilot decision support system.

Table 1: Data mining tools evaluated

Product	Company	URL
Knowledge STUDIO	Angoss	www.angoss.com
Clementine	SPSS Inc.	www.spss.com/spssbi/clementine/
Enterprise Miner	SAS Institute	www.sas.com/products/miner
Intelligent Miner	IBM	www.software.ibm.com/data/intellimine/
Oracle 9i Data Mining	Oracle Corporation	www.oracle.com/ip/deploy/database/oracle9i/
S-Plus	Insightful Corporation	www.insightful.com
WEKA	University of Waikato	www.cs.waikato.ac.nz/~ml/weka

Table 2: Classification methods comparison

Product	Regression	Decision Tree	Neural Nets	Bayes Class	Meta Learn	K-nearest neighbour	Explain based
Knowledge Studio	X	X	X	X	X		X
Clementine	X		X			X	
Enterprise Miner	X	X	X				X
Intelligent Miner	X	X	X			X	
Oracle 9i Data Mining	X	X	X	X			X
S-Plus	X						
WEKA	X	X	X	X	X		

5.2 Demonstration of Using data mining techniques on building maintenance

Data mining techniques applied in this demonstration includes visual data analysis and data mining algorithms of decision trees and association rule. A histogram is used in the visual data analysis wherein data is collected and sorted into categories. Histograms focus on the frequencies and distributions of one particular attribute. WEKA incorporates a stacked histogram which allows three judgments: (a) the trends on the total height of the columns, (b) the proportion of each category within each column; and (c) the trends in the lowest category (Dix and Ellis 1998). This interactive stacked histogram solves the problem of cross comparison of standard histogram by allowing different trends to be analysed using the same graph. The decision tree algorithm is a tree-based knowledge representation methodology used to present classification rules. The leaf nodes present class labels. Various classification algorithms offered by WEKA have been applied and it was found that several algorithms were not able to deal with the maintenance data sets available due to some limitation in processing certain data types. For instance, some algorithms were not able to accommodate numeric values while others failed to accommodate nominal variables. The C4.5 algorithm (built on the top of ID3 proposed by Quinlan (1993), was selected because of its capability to deal with numeric and nominal variables, and to handle missing values and pruning. The latter can be done by replacing a whole sub-tree by a leaf node if the expected error rate in the sub-tree of a rule obtained is greater than it in the single leaf. The C4.5 algorithm generates a classification-decision tree for a given data-set by recursive partitioning of data. Once the tree is constructed, rules can be generated by traversing each branch of the tree and collecting the conditions at each branch of the decision tree. The association rule technique involves finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories (Han and Kamber 2001). The association rule algorithm adopted in WEKA is “Apriori” which is developed by IBM’s Quest project team. Apriori finds all associations that satisfy a set of criteria with minimum support and minimum confidence. Support (also called coverage) refers to the number of instances predicted correctly. Confidence (also called accuracy) is the proportion of the number of instances that a rule is correctly applied to them (Witten and Frank 2000). Rules with high support are of interest and some rules are pruned out due to their low coverage. The above data mining techniques were applied on a selected data set of battery chargers at Building No. 10, Royal Prince Alfred Hospital, Central Sydney Area Health Service. The evaluation of the results obtained from mining the maintenance data and their potential benefits to facility managers on building maintenance are shown in Tables 3, 4 and 5 for battery chargers, air handling units and thermostatic mixing valves respectively.

Table 3: Some results of applying data mining techniques on maintenance data of battery chargers

Data Mining Technique	**Data Mining Results**	**Potential benefits to facility managers on building maintenance**
Visual Analysis	• All outstanding works took place around December 2002.	Direct the allocation of maintenance resources at the appropriate time of the year to achieve better planning and scheduling of maintenance work.
	• Asset “EPG0101” belongs to cost centre “1000” while the cost centre for other assets were not available;	Devise an appropriate way of billing to monitor the life cycle cost for each asset (system or a component).
	• There is fee charge with asset “EPG0101” while no charge for “EDG1000-01”.	
Decision Tree Algorithm (C4.5)	• For all tasks with work order No > 66195, and some tasks with work order No. between 48002 and 66195, completions did not meet the expectation of completion date.	A special attention should be directed to certain tasks in the building wherein maintenance work is required more often.
Associative Rules Algorithm	• 57% of workorders were completed within the expected completion date.	Only 57% of maintenance works completed as planed might reflect substantial problems of maintenance procedures carried out with battery chargers.

Table 4: Some results of applying data mining techniques on maintenance data of air handling units

Data Mining Technique	**Data Mining Results**	**Potential benefits to facility managers on building maintenance**
Decision Tree AlgorithmC4.5,	Department 26462 only reports A/C malfunction. (all 18 cases)	A special attention should be directed to certain places in the building wherein maintenance work is required more often.
	96% jobs for cost_centre = 0 is CM (corrective maintenance).	
Association Rule Algorithm	For floors 5, 6 and 7, the workOrder_Status was always completed.	Benefiting from successful maintenance practices including both equipments and labour is useful to achieve a high level of an overall maintenance performance.

Table 5: Some results of applying data mining techniques on maintenance data of thermostatic mixing valves

Data Mining Technique	**Data Mining Results**	**Potential benefits to facility managers on building maintenance**
Decision Tree Algorithm (C4.5)	All monthly high priority works are carried out in the later part of the year – July to November.	Distribution of priority of maintenance work is important in planning and scheduling maintenance work and resources.
Association Rule Algorithm	There is an incremental relationship between the work priority, the estimated time to complete the work and associated budget.	A better planning and scheduling will help to advance this pattern of relationship.

6. DISCUSSION

Data mining is a capability consisting of the hardware, software, "warmware" (skilled labour) and data to support the recognition of previously unknown but potentially useful relationships. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data and using them to make valid predictions. It supports the transformation of data to information, knowledge and wisdom. It is important to note that data mining is not software alone, as some vendors would have clients believe. While software may play an important role, it is only in the context of clear objectives and careful thinking from management point of view along with the skill of the analyst that data mining ultimately leads to a success story rather than an embarrassing and costly failure. Data mining can be used for anything from asset and inventory management to predictive maintenance to banking and finance. Knowing when a building system or a component will break before it breaks, in plenty of time for repairs to be conveniently and cost-effectively scheduled and executed, is an exciting application of this technology that can add dollars to the bottom line. A successful application of data mining requires commitment to financial and technical resources, highly qualified analytical talent, the cleanup and maintenance of clean data, and the willingness to learn and act upon the results. Therefore, data mining can provide a powerful competitive advantage (Labovitz 2003).

Facility managers and building owners are more concerned with highlighting areas of existing or potential maintenance problems in order to be able to improve the building performance, satisfy occupants and minimise the operational cost of maintenance. Applying data mining techniques on the available industrial maintenance data has helped to discover useful rules that allowed locating some critical issues that will have substantial impact on improving the management of building life cycle. Before using rules to change operations, it is important to examine the rules. Rules produced by associations and sequencing show correlation, not cause. For this a domain expert is required who must then determine. Unexpected rules that don’t make sense may also signal other, more nefarious, activity, such as pilfering.

Data mining of building maintenance can help to discover: procedures that reduce future failures; repairs or maintenance operations that are being done improperly; ways to improve repairs that reduces subsequent down time; undocumented methods being used by experienced personnel that result in reduced down time; advance notice of likely failures before failures occur. Such discoveries can be used to modify building maintenance and repair procedures thereby reducing downtime, increasing uptime, and significantly reducing the costs of maintenance and repair.

ACKNOWLEDGEMENT

The research presented in this paper is funded by CRC-Construction Innovation, Project No.: 2001-002-B “Life Cycle Modelling and Design Knowledge Development in Virtual Environment”.

REFERENCES

Arditi, D., and Gunaydin, M. (1998) *Factors that affect process quality in the life cycle of building projects*, Journal of Construction Engineering and Management, ASCE, **124** (3): 194-203

Barringer, H. and Weber, D. (1996) *Life Cycle Cost Tutorial*, Fifth International Conference on Process Plant Reliability, Gulf Publishing Company: Houston, TX

Christiansson, P. (1998) *Using knowledge nodes for knowledge discovery and data mining*, in Smith, I. (ed), Information Technology for Design, Collaboration, Maintenance, and Monitoring, Springer-Verlag: Berlin, pp 48-59

Dangel, R. (1969) *Integrated logistic support (ILS) implementation in the Naval Ship System Command*, ASE 6th annual technical symposium, pp. 1-25

Dix, A. and Ellis, G. (1998) Starting Simple - adding value to static visualisation through simple interaction, in Catarci, T., Costabile, T., Santucci, M. and Tarantino, L. (eds.), Proceedings of Advanced Visual Interfaces AVI98, ACM Press: L'Aquila, Italy, pp. 124-134

Elder, J., and Abbott, D. (1998) *A Comparison of Leading Data Mining Tools,* tutorial presented at KDD-98: New York

Frawley, W., Piatetsky-Shapiro, G. and Matheus, C. (1992) *Knowledge discovery in databases: An overview*, AI Magazine, Vol. **13**: 57-70

Goebel, M. and Gruenwald, L. (1999) *A survey of data mining software tools*, SIGKDD Explorations, 1(1)

Han, J. and Kamber, M. (2001) *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers: San Francisco

Kawauchi, Y. and Rausand, M. (1999) *Life Cycle Cost (LCC) Analysis in Oil and Chemical Process Industries,* adopted from http://www.ipk.ntnu.no/rams/notater/lcc.doc

Kunz, J., Jin, Y., Levitt, R. Lin, S.D. and Teicholz, P. (1995) *The intelligent real-time maintenance management (IRTMM) system: Support for integrated value-based maintenance planning*, CIFE Technical Report 100: Stanford, CA

Labovitz, M. (2003) *What is data mining and what are its uses?* Darwin Magazine, October 2003, http://www.darwinmag.com/read/100103/mining.html.

Peitgen, H. O., Jurgens, H. and Saupe, D. (1992) *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag: New York

Quinlan, R. (1993) *C 4.5 Programs for Machine Learning*, Morgan Kaufmann: San Mateo, Calif

Siewiorek, D. P. and Swarz, Robert S. (1982) *The Theory and Practice of Reliable System Design*, Digital Press: Bedford, MA

Soibelman, L. and Kim, H. (2002) *Data preparation process for construction knowledge generation through knowledge discovery in databases*, Journal of Computing in Civil Engineering, **16** (1), 39-48

Witten, I. and Frank, E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufman: San Diego