Improving the Management of Building Life Cycle: A Data Mining Approach

Clients Driving Innovation
International Conference

Rabee Reffat, John Gero & Wei Peng
john, wpeng@arch.usyd.edu.au

Motivations of the Project

Integrating data mining within the life cycle of buildings

Facility Maintenance and Management

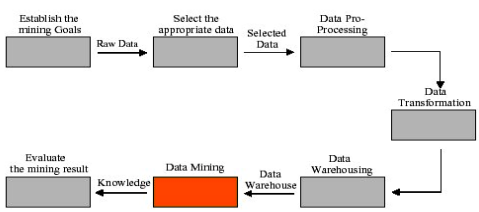
Designing + managing for facilities maintenance = major task …for both the designer and building manager.

- No system that provides users with feedback on the kind of design and management knowledge required to improve building facilities

How to develop such a system for designers and managers to improve building design and maintenance management?

- By adding value to information contained in existing building facilities databases through identification of significant patterns.
- Results may then be utilised for improved management, operation and enhance building life cycle.
- Variety of system benefits for: designers, facilities managers, construction industry and the environment.

Data Mining Processes



Stages of data mining process

Interactive Data Mining Scenario

Data requirements for database:

- Pre-processing: what is task relevant data? What kind of knowledge do I need?

Data mining process:

- What are the appropriate mining approaches? Can these approaches be generalised to other data sets?

Evaluation of the results:

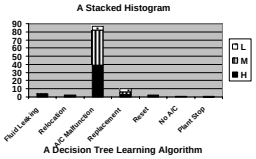
- What background knowledge could be useful?
- How do I present the results?

Friendly interface:

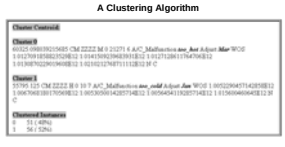
- Allow interactions;
- Navigations, usability concerns, etc.

Data Mining Approaches

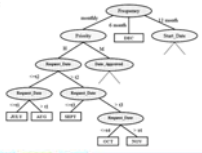
A Stacked Histogram



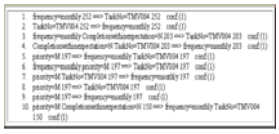
A Clustering Algorithm



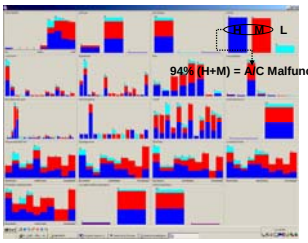
A Decision Tree Learning Algorithm



Associative Rule Algorithm



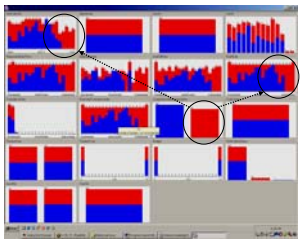
Visual Analysis on Stacked Histogram



For attribute "priority" and its correlations:

- 94% medium and high priority works belong to A/C malfunction
- Stacked histogram allows cross comparing
- Stacked histogram allows different trends analysis
- Perform poor in Date, continuous numeric attribute

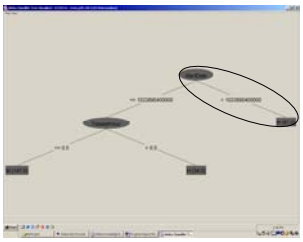
Visual Analysis on Stacked Histogram



For attribute "work completed on time" and its correlations:

- At the end part of work order list, there is a trend of not able to meet completion deadlines;
- This is illustrated in diagram with red shades at the end of workorderNo and startdate, etc.)

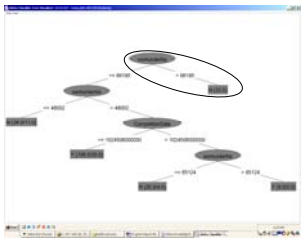
Decision Tree Learning



Decision Tree C4.5 on "priority" for thermostatic valves:

- Start Date after the date must belong to "High" priority work
- A trend of recently emphasizing of thermostatic mix valves maintenance?

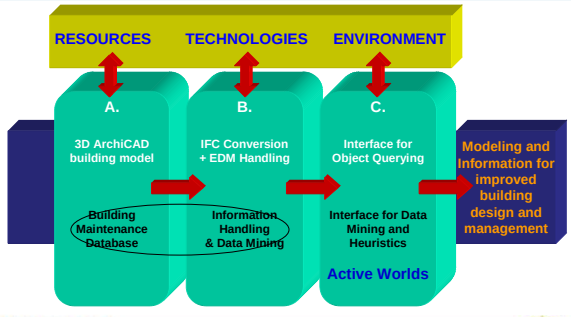
Decision Tree Learning



Decision Tree C4.5 on "work completed on time" for battery units:

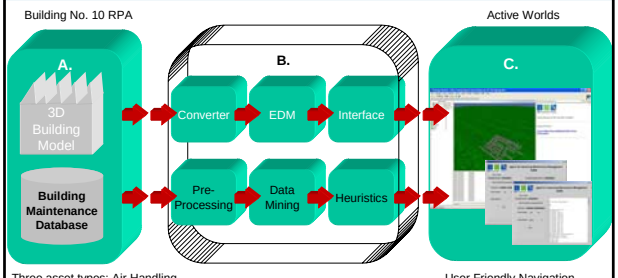
- **All** tasks with work order No > 66195, **some** tasks with work order No between 48002 and 66195 -- not complete on time
- A trend that with higher work order number, the less able to meet deadlines?

Approaches and Techniques



AIMM System Architecture

Agents for Improved Maintenance Management



Three asset types: Air Handling Unit, Thermostatic Mixing Valve & Battery Charger

User Friendly Navigation and Data Mining Interface

Results and Implications

Patterns	Rules and Implications	Suggestions
floor=7 25 = = > causeofrepair = A/C_Malfunction 25	▪ All A/C maintenance works in floor 7 belong to A/C malfunction; ▪ Failures concentrate on a particular floor;	Investigate the possibility of poor design or maintenance of A/C in 7th floor due to its high demand of corrective maintenance

Data Mining Technique	Rules Obtained	Potential Impact on Facility Maintenance and Design
Visual Analysis	Approximately all "A/C malfunction" belongs to high and medium priority.	"A/C malfunction" is of a major concern in guiding the allocation of maintenance resources.
Visual Analysis + Decision Tree Algorithm (C4.5)	For all monthly high priority works, all the works in July complete within expectation and those in August fail.	There are possible failures in relation to seasons or resulting weather and humidity; This also may show that changing services or personnel in August, who are not efficient; Suggestion: Compare maintenance records between July and August and analyse discrepancies.
	All 7th floor jobs were of high and medium priority and the cause of repairing was "A/C malfunction".	Suggestion: Investigate the possibility of poor design or maintenance of air conditioning function in 7th floor. A special attention in the design should be given to a specific floor due to its high demand of corrective or preventive maintenance or special design of A/C.
Decision Tree Algorithm (C4.5)	Department 26462 only reports A/C malfunction. (all 18 cases)	Failure abnormally concentrate on a particular department. Suggestion: Investigate the possibility of poor design or maintenance of air conditioning function in a particular department. 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.

Potential Benefits

Useful results can be obtained from the maintenance data by applying the appropriate data mining scenarios, i.e. classification, clustering and associative rules in conjunction with filtering agents.

Assist facility managers in identifying critical cost issues.

- Decrease operational costs by reducing high levels of unplanned maintenance.

Constructing predictive plans based on correlations obtained from the application of data mining.

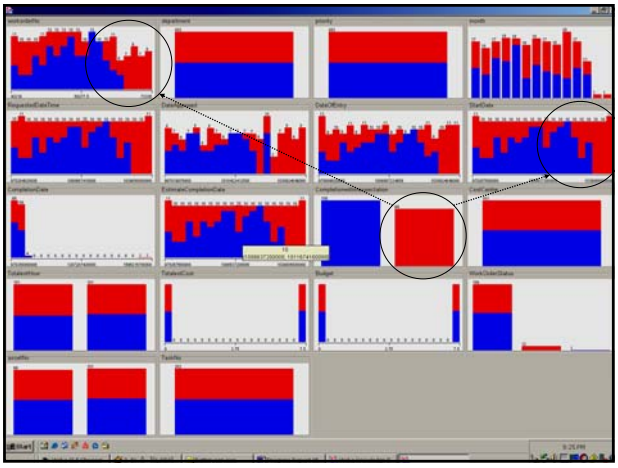
- Ability to guide the allocation of maintenance resources through potential correlations between seasons and malfunction rates.
- Investigate abnormal phenomenon discovered from the maintenance data set such as "provide example here".

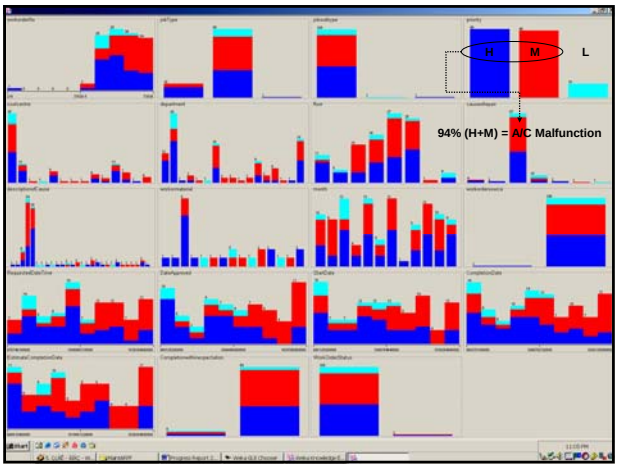
Discussion

Requirements of data gathering and knowledge:

- What kind of knowledge do I want to mine?
  Mean time between failures, critical maintenance issue, etc.
- What is missing and where does it come from?
  Cost-related data, contract information, etc.
- What background knowledge could be useful?
  Need a knowledge base to hold mined results?

The End
Thanks!




H
M
L
94% (H+M) = A/C Malfunction