MODEL-BASED RECOMMENDATION SYSTEM IN SUPPORT OF CONSTRUCTION EQUIPMENT MANAGEMENT: A CASE STUDY

Hongqin Fan[1], Hyoungkwan Kim[2], and Simaan AbouRizk[3]

ABSTRACT

Major contractors are relying increasingly on computerized equipment management systems to accomplish their objectives in construction equipment management, including maintenance, repair, operations and other managerial tasks. Though the automatic or semi-automatic collection of electronic data for equipment management has been made possible via on-board controls, handheld devices, and electronic fuel data transfer, the output of most current management systems are still delivered through reports that represent facts "as-it-is" and lack a mechanism for automatic knowledge discovery and utilization of the collected data. By means of a case study on real-time work order evaluation, this paper presents our research on embedding data mining modules into current equipment management information systems so that the hidden patterns in the data can be explicitly discovered and represented to the user. This paper will also demonstrate how new cases are predicted using the trained predictive model. The proposed approach is capable of making real-time predictions based on facts rather than experiences; in addition, it explains the logic of reasoning using the transparent data mining models.

KEY WORDS

Construction equipment, maintenance, data analysis, decision making

INTRODUCTION

In 2002, Schexnayder and David [Schexnayder and David 2002] predicted that in the coming years, the most significant changes in equipment for contractors will be in fleet management due to the development of advanced computer tools. Today, most large contractors including the government, rely on a variety of construction management information systems to accomplish their daily tasks in construction equipment management,

1 Graduate Research Assistant, Department of Civil and Envir. Engrg., 3-133 Markin/CNRL Natural Resources Engineering Facility, The University of Alberta, Edmonton, Canada T6G 2W2, Tel: +1 780/492-4235, Fax: +1 780/492-0249, hfan@ualberta.ca

2 Assistant Professor, School of Civil and Environmental Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea, Tel: +82 2/2123-2799, Fax: +82 2/364-5300, hyoungkwan@yonsei.ac.kr

3 Professor, Dept. of Civil and Environmental Engineering, 3-133 Markin/CNRL Natural Resources Engineering Facility, The University of Alberta, Edmonton, Canada T6G 2W2, Tel: +1 780/492-4235, Fax: +1 780/492-0249, abourizk@ualberta.ca

including preventative maintenance, fuel and fluid tracking, inventory management, and work order tracking. To further simplify the process, some automation techniques have been applied to collect operational data, which is then feed into the system. Examples include the on-board control and diagnostics, field data collection with handheld electronic device, real-time recording of physical locations via a Global Positioning System (GPS), and electronic data transfer from fuel suppliers. On the other end of the system, however, the primary output is delivered in various accounting and equipment reports. Due to a shortage of data analysis functions in the current equipment management system, the equipment manager resorts to advanced statistical tools for in-depth data analysis supporting equipment management, which is a technically demanding and time-consuming process. As a result, only a small portion of the data is utilized for analysis in support of decision-making, according to our recent interview with a large road builder in Canada. One solution to this problem is to integrate data mining techniques into the current equipment management system so that both the data analysis and utilization are automated.

Data mining is a hybrid science of many disciplines including statistics, machine learning, and database technology. It is capable of discovering hidden patterns in data, and using discovered patterns or rules for prediction in an automatic or semi-automatic approach. Embedding data mining modules in the current equipment management system provides some unparalleled advantages over the current practice:

- Real-time analysis: the data mining models can be integrated seamlessly with the current system using a live connection to the up-to-date data depository,
- Rules discovered can be explicitly presented in textural or visual format. For example, the decision tree mining model can both identify the prioritized features which lead to the state of the predicted variable and visually present the tree structure containing various decision rules and paths,
- Make predictions in automatic or semi-automatic methods. For example, time-series analysis helps to identify trends in the time-series data in order to make predictions on future events; Bayesian belief network mining models build the network automatically from data except that the directions of cause-effect relations are determined manually.

Statistical analysis is by far the dominant approach currently adopted by construction equipment management. Both the industrial practitioners and researchers use statistical tools for purposes such as diagnostic analysis, life cycle cost analysis, component failure prediction, and time-series forecasting [Gillespie and Hyde 2004; Reid and Bradford 1983]. The major problems with these statistical approaches are their independence from the current equipment management system, their requirements for training in advanced statistical tools, and their inability to reflect the dynamic changes in equipment data. Our survey on the currently available equipment management systems indicates that most of the system outputs are restricted to a handful of customized reports and simple statistical results. The systems that claimed to be intelligent are often merely a close integration among system modules or with an equipment specification/maintenance manual.

Our vision for the intelligent equipment management system is the inclusion of features enabling automatic knowledge discovery and representation, real-time prediction, and forecasting to satisfy the needs of construction fleet management at either managerial level or operational level.

This paper summarizes our proposed framework for building an intelligent equipment management system through a case study on embedding a real-time work order evaluation module in a current equipment management system. A data mining model using an AutoRegressive Tree (ART) technique [Meek et al. 2002] is built and seamlessly integrated with the work order entry module. Whenever the user enters a new work order that includes estimated labour hours, the mining model evaluates the accuracy of the labour hour estimate and makes recommendations based on historical information. Furthermore, the user can visually analyze the derived regression tree structure, identifying the leading causes of inaccuracy as well as browsing through the "most similar" cases in history. Embedding this predictive component helps to improve the accuracy of work order estimation, thus leading to improved scheduling of equipment repair and maintenance as well as resource allocation for large contractors.

BACKGROUND

Discovering hidden knowledge in equipment operational data or making predictions based on historical information, which differs from the general know-how on equipment management and performance, have been studied intensively using advanced statistical tools. Research efforts in this regard include predicting the trend of yearly repair costs per service unit for repair/replace decisions [Gillespie and Hyde 2004] and predicting current residual values of heavy construction equipment [Lucko and Vorster 2002]. Similar research has also been conducted in the areas of agricultural and forestry equipment management [Reid and Bradford 1983; Al-Suhaibani 1999].

Computerized maintenance management is an intensively researched area, which deals with the maintenance operations of facilities, including equipment in power plants, industrial plants, and military facilities. Due to the unexpected failure of major components, the incorporation of intelligent data analysis modules into a Computerized Maintenance Management System (CMMS) is highly expected. The objective is to predict the deterioration of conditions based on monitoring data so as to adjust the maintenance schedule proactively or to repair/replace the components before failure. These intelligent CMMS use artificial intelligence technologies, such as Artificial Neural Network (ANN) [Fu et al. 2004]; Recursive Neural Network (RNN) [Yam et al. 2001]; and Bayesian Network [Zhang et al. 1997].

INTELLIGENT EQUIPMENT MANAGEMENT SYSTEM USING EMBEDDED DATA MINING MODEL

Data mining model has unprecedented advantages as compared with traditional mathematical models and the rule-based system. Firstly, the data-mining model is data-driven, which does not vary with expert opinions (although some prior domain knowledge is important). As a result, the model is based on facts and not influenced by personal subjective judgment. Secondly, the data-mining model may become the only viable solution when the system is too complex to be expressed using mathematical or statistical models. The easy adaptation of the data-mining model also fits well with the dynamic feature of construction equipment management.

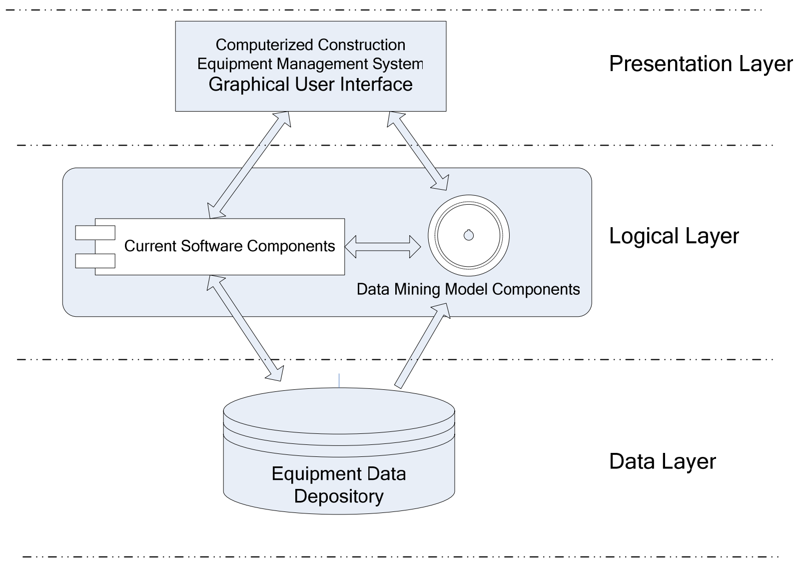

Figure 1. Proposed Intelligent Equipment Management System

A typical construction equipment management information system consists of three layers: the presentation layer, the logical layer, and the data layer. The presentation layer interacts with the user via graphical user interfaces; the logical layer handles the business logic and rules; and the data layer provides data storage and access. The logical layer works in between the other two layers and implements complex logic and rules defined in the system. Recent developments in data-mining engines and data-mining languages such as Analysis Services in Microsoft SQL server 2005 and Data Mining Expressions (DMX) [Tang and MacLennan 2005] allow for a higher level integration between data-mining engines and applications, while providing sufficient flexibility to fine-tune the data-mining algorithms. Therefore, from the system perspective, the integration of data-mining functions into the current system can be realized by adding data-mining modules into the logical layer as shown in Figure 1. Data-mining modules can process data in real-time, take input from and send the mined patterns/rules or prediction results to other modules, or to users in real-time.

A CASE STUDY: WORK ORDER EVALUATION

A road building and maintenance contractor in Alberta, Canada, owns a large construction equipment fleet. While minor repairs and maintenance are usually performed on the spot, the major repair, maintenance, and overhaul of heavy construction equipment are assigned to the distributed shops across Alberta. The general procedures for work order management include:

1. The superintendent makes a preliminary inspection and fills out a work order on the

equipment via the equipment management system; the work order includes the details of estimated labour hours, parts, and the required date.

2. Adjust the maintenance/repair work plan based on the work order estimation, resources, and the priority level of the project,
3. Perform maintenance/repair works according to the schedule, input into the system the actual labour hours and parts,
4. Generate reports on estimated work order, actually incurred items, and costs.

Though the current equipment management system provides a work order entry and manipulation module and generates multiple reports on work order estimation and implementation, the "lessons" learned from the inaccurate estimation of work orders have to be explored by the equipment management team. As a matter of fact, the accuracy of work order estimates is generally not to their satisfaction, and its indirect impact on the project from the unavailability of major equipment is often due to an inappropriate maintenance/repair schedule, which has resulted from inaccurate work order estimation.

The accuracy of estimated labour hours on work order items is a major concern. Many factors are cited as having a potential impact, including equipment division, department, category, class, make, year of manufacture, component, and parts. To improve the accuracy of the work order estimation, the following questions should be investigated: which groups of equipment or equipment of what features have a high accuracy of estimation? Which groups of equipment have underestimated labour hours, and by how much? What are the leading features of impact? These questions can only be answered after a comparative analysis is conducted on both the estimated and actual work orders. On the other hand, the answers to these questions are changing dynamically with time.

AutoRegressive Tree (ART) data-mining technique [Meek et al. 2002] can extract such knowledge from historical data and use the "learned" rules for prediction. The ART model is embedded in the work order entry module after validation; it works behind the user interface to evaluate the data entry on labour estimates according to the other known attribute set. If the user raises any doubts about the "recommendation" from the ART model, he/she can visually analyze the decision tree structure and track those cases that will follow the same decision path.

AUTOREGRESSIVE TREE DATA MINING MODEL

An AutoRegressive Tree is a hybrid data-mining model of decision tree and multivariate linear regression. Although there are different methods for growing the tree, the structure of the ART regression trees is generally the same. Using a large number of training cases, the ART algorithm grows a top-down tree structure for the prediction of a continuous target variable based on known sets of attributes. First of all, the continuous target variable is binned into buckets representing different ranges of values. Then, starting from a root node containing all the cases, the child nodes grow, each node containing purer subsets of cases in accordance to the test results of a most informative attribute. A child node becomes a leaf node if the termination criteria are satisfied; otherwise, it will become a decision node. Recursively split each decision node until all the nodes become leaf nodes. Finally, on each leaf node, a multivariate linear regression model is built.

As a supervised learning algorithm, the ART data-mining model learns by "observation". A large number of historical cases, each of which contains multiple predictor variables and a continuous target variable, must be available for the induction of the mining model structure. The following paragraphs give a brief description of the two underlining component algorithms: decision tree induction and multivariate linear regression.

DECISION TREE INDUCTION

Decision tree induction algorithm C4.5 introduced by Quinlan [Quinlan 1993] is used for the automatic inference of a decision tree structure. The most important feature of C4.5 is the use of an entropy function to measure gains in information after binning a continuous variable or tree splitting. Entropy is a concept in information theory introduced by Claude Shannon. Formula [1] is used to measure how much information (measured in bits) can be conveyed from a dataset in terms of state distribution [Kantardzic 2003].

$$Info(S) = -\sum_{i=1}^{k} \frac{freq(C_i, S)}{|S|} \bullet \log_2\left(\frac{freq(C_i, S)}{|S|}\right) \text{..................} [1]$$

Where S—dataset
K —number of bins
Ci — the i^{th} bin

In this implementation, the decision tree algorithm first renders the continuous variable discrete using entropy function. The values of the continuous target variable are binned into a large number (99, for example) of equal range buckets. Merge the neighbouring buckets if the information of the dataset increases after merging. Repeat the merging process until the buckets of the continuous variable reach an optimum state.

The growing of trees by splitting a decision node is also based on an increase in information. The information obtained after choosing an attribute and attribute test value is calculated as the weighted sum of entropies over subsets. When growing the tree, the algorithm chooses the attribute and its test value, which give the most information after splitting. The procedures for decision tree induction are stated as follows:

1. Start from the root node containing all the training cases
2. Choose the attribute and the test value that produce largest information gain as compared to other alternatives. Split the root node into child nodes based on the attribute test in the decision node. Each child node should contain more homogeneous cases after splitting.
3. Terminate the child node as the leaf node if the number of cases in the node is less than the pre-determined threshold value, or the percentage of support is more than a threshold value based on majority voting. Also terminate the splitting process if the information gain is less than a pre-defined threshold value.
4. Recursively split each decision node as per steps 2 and 3 until all newly grown child

nodes become leaf nodes.

MULTIVARIATE LINEAR REGRESSION

For the cases in the leaf node of the decision tree, the prediction results fall into the same range of values after passing a set of logical tests. To provide a more accurate value prediction, a multiple linear regression model is built on each node with all the continuous predictor attributes as independent variables and the target attribute as a dependent variable. The coefficients of the regression formula are estimated using the least-square approach so that the total summed estimation error is minimized.

The final hybrid data-mining model is an automatically grown decision tree with multiple linear regression models in leaf nodes. Before it is deployed in an application, the induced ART model must pass validation tests. A holdout method is used here for data validation, which requires a random selection of the historical cases for model training; the remaining cases held out for testing. In addition to the tabled comparison of actual versus predicated target values, a scatter plot is usually created to observe visually how many points have their predicted values deviate significantly from their actual values.

WORK ORDER EVALUATION DATA MINING MODEL: BUILDING, VALI DATION AND DEPLOYMENT

The objective of work order evaluation is to predict the accuracy of estimated labour hours in the work order item based on the known factors of impact. Twelve attributes were identified as potential factors of impact: division, department, category, class, unit, manufacturer, equipment age, component, parts, repair type, estimated labour hours, and estimated parts. 1190 cases are collected from the equipment database after the removal of some obvious outliers due to data entry error.

MODEL BUILDING

Using the ART induction algorithm, an ART data-mining model is generated from the training dataset, obtained by randomly selecting 80% of the cases Figure 2 shows part of the model structure, comprised of typical decision trees with regression formulas in the leaf nodes. To make a prediction for a new case, the computer model will answer a series questions by following the decision path from the root node down to the landed leaf node, and then use the regression formula to evaluate the labour hour deviation.

MODEL VALIDATION

The validation of the ART model is conducted by randomly selecting 80% of the cases for induction as an ART model, and by then using the remaining 20% for testing. The scatter plot is used for visually assessing the accuracy of the prediction. As shown in Figure 3, the predicted deviations of the labour hours are plotted against the actual deviations of labour hours, which indicates that only a small percentage of cases are incorrectly predicted. The predicted deviations and actual deviations are tabulated for comparison and it was found that the percentage of cases which are correctly predicted is of acceptable range according to industry practise.

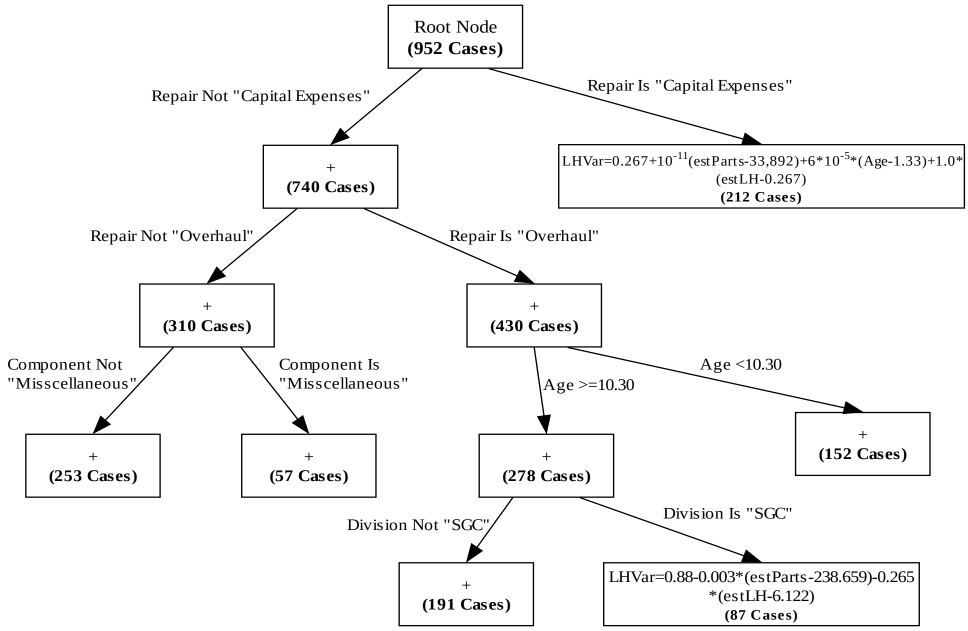

Figure 2. Part of the Induced Work Order Evaluation ART Model
(Decision nodes denoted with an expansible plus sign and leaf nodes with a regression formula)

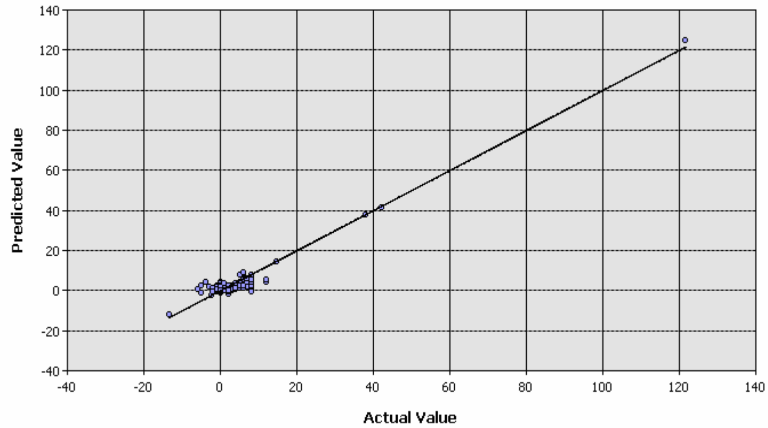

Figure 3. Scatter plot of actual deviations vs. predicted deviations using the test dataset

MODEL DEPLOYMENT

The validated work order of the ART model is hosted using the Analysis Services of Microsoft SQL Server 2005 with a live connection to the equipment database and the current equipment management information system. As shown in Figure 4, when the user enters the estimated labour hours as a work order item, the model is capable of validating the input and making recommendations based on historical information. The data-mining model is updated automatically on a regular basis to reflect recent changes in the data depository.

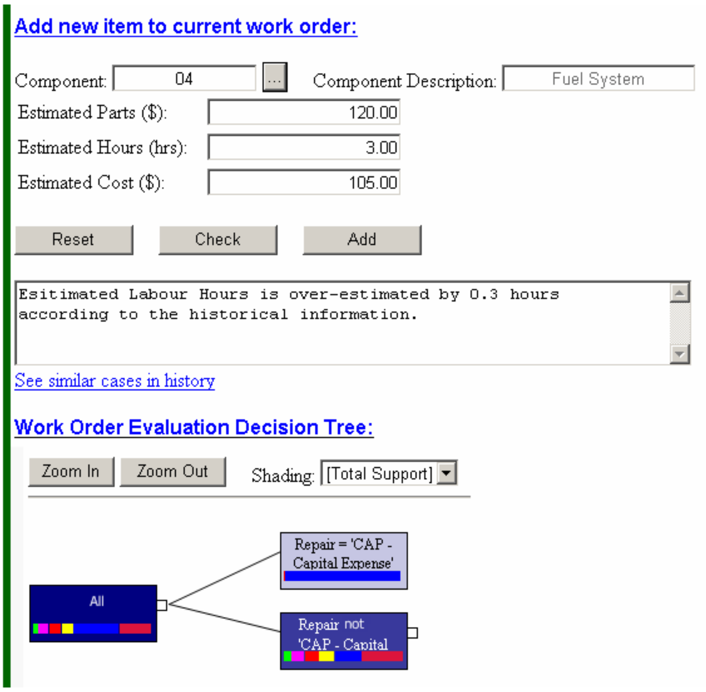

Figure 4. Recommendation on Labour Hour Estimate from Embedded ART Model

CONCLUSIONS AND FUTURE RESEARCH

Embedding data mining models in the computerized equipment management system enables the automatic discovery and representation of hidden knowledge from the equipment operational data, as well as real-time prediction using the induced model. By means of a case study on work order evaluation using an AutoRegressive Tree data-mining model, this paper summarizes our research into the building, validation, and deployment of data-mining models in current equipment management systems, and more importantly, it provides a means for explicitly presenting the discovered knowledge and enables the user to browse through the traversed decision path for fact-based decision support.

Commonly recognized as an automatic or semi-automatic approach for knowledge discovery, data-mining methodologies will not succeed without a good understanding of the domain under study. For example, in the case of work order evaluation, the ART algorithm cannot induce a useable data-mining model when the primary factors of impact on labour-hour estimation are not properly identified.

As compared to traditional expert systems in which prior expert knowledge is built into the information system, the mining model-based approach is completely data-driven. Other data-mining techniques that will be integrated into our current construction equipment management system include outlier detection using nonparametric outlier mining algorithm [Fan et al. 2005], Intelligent Life Cycle Cost Analysis (ILCCA) of equipment using time-series data mining [Meek et al. 2002], and component failure predication using boosted decision trees [Freund 1995]. It is expected that the proposed intelligent equipment management system will serve as a powerful decision support tool in construction industry.

ACKNOWLEDGEMENTS

The authors wish to thank Dale Tillapaugh, the equipment manger of Standard General Inc., for his contributions to this research. The research project is partially funded by the Natural Sciences and Engineering Research Council of Canada under Grant # CRD 226956-99.

REFERENCES

Al-Suhaibani, S.A.M. (1999). "Tractor Repair and Maintenance in Saudi Arabia." ASAE the Power & Machinery Division. 15(6). 591-596 pp.

Fan, H, Kim, H. and AbouRizk S. (2005) "A non-parametric outlier mining algorithm for detecting anomalies in construction equipment database", Proceedings of CSCE Annual Conference, Toronto, June 2005

Freund, Y. (1995). "Boosting a weak learning algorithm by majority." Information and Computation, 121(2), 256–285 pp.

Fu, C., Ye, L., Liu, Y., Yu, R. Iung, B. Cheng, Y. and Zeng, Y. (2004). "Predictive Maintenance in Intelligent-Control-Maintenance-Management System for Hydroelectric Generating Unit." IEEE Transactions on Energy Conversion, 19(1).

Gillespie, J.S. and Hyde, A.S. (2004) *The Replace/Repair Decision For Heavy Equipment*. Virginia Transportation Research Council. Final report: VTRC 05-R8

Kantardzic, M. (2003). Data Mining: *Concepts, Models, Methods, and Algorithms*. John Wiley & Sons, Inc. NJ. USA.

Lucko, G. and Vorster, M.C. (2002) "Predicting the Residual Value of Heavy Construction Equipment." Proceedings of the 4th Joint International Symposium on Information Technology in Civil Engineering 2003, Tennessee, USA.

Meek, C., Chickering, D.M. and Heckerman, D. (2002). "Autoregressive Tree Models for Time-Series Analysis." Proceedings of the 2nd SIAM International Conference on Data Mining, Arlington, VA, USA.

Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan-Kaufmann Publishers, San Mateo, CA. USA.

Reid, D. W. and Bradford, G. L. (1983). "On Optimal Replacement of Farm Tractors." American Journal of Agricultural Economics. 65(2). 326-331 pp.

Schexnayder, C.J. and and David,S.A. (2002). "Past and Future of Construction Equipment—Part IV." J. Constr. Engrg. and Mgmt., ASCE, New York, NY, 128 (4) 279-286 pp.

Tang, Z. and MacLennan, J. (2005). *Data Mining with SQL Server 2005.* Wiley Publishing, Inc. Indianapolis, USA.

Yam, R.C.M., Tsei, P. W., Li, L. and Tu, P. (2001). "Intelligent Predictive Decision Support System for Condition-Based Maintenance." Int. J. Adv. Manuf. Technol, Springer-Verlag London Limited. 17:383–391 pp.

Zhang, J., Tu, Y. and Yeung, E.H.H. (1997) "Intelligent Decision Support System for Equipment Diagnosis and Maintenance Management." Proceedings of Innovation in Technology Management – The Key to Global Leadership, International Conference on Management and Technology, Portland, USA. 733 pp.