UPDATING BAYESIAN NETWORK FOR DIAGNOSTIC FAILURE ANALYSIS OF CONSTRUCTION EQUIPMENT

*H. Q. Fan

Department of Building and Real Estate

The Hong Kong Polytechnic University

Hung Hom, Kowloon

Hong Kong, P.R. China

*(*Corresponding author: bshfan@polyu.edu.hk)*

UPDATING BAYESIAN NETWORK FOR DIAGNOSTIC FAILURE ANALYSIS OF CONSTRUCTION EQUIPMENT

ABSTRACT

Construction equipment is an important type of resources of heavy construction contractors. Since equipment breakdowns can cause project delays and significant financial losses, the contractors are eager to know those factors causing equipment failures directly or indirectly, related to equipment design, maintenance, and operations. Although Bayesian network can be used for diagnostic analysis of failure events or making predictive analysis, building a Bayesian network for such purpose can be difficult as the cause-effect relations can be subjective and their conditional probabilities change with a wide variety of causal factors. A hybrid approach is proposed in this paper to update the Bayesian diagnostic network structures and parameters using real life data, the conditional probabilities and cause-effect relationships can be dynamically updated with observed failure records to reflect the real life situations of a complex equipment system. A case study is conducted to show the benefits of the hybrid approach in construction equipment diagnostic analysis.

KEYWORDS

Construction equipment maintenance; Bayesian network learning; Failure analysis; Decision support

INTRODUCTION

Every year, expected and unexpected failures of construction equipment cause financial and time losses to equipment-owning contractors and organizations. Although new tools and technologies have emerged to help the equipment management team to conduct maintenance, monitor equipment conditions and uses, and predict impending failures, these unexpected equipment failures will continue to occur in the foreseeable future. Predictive maintenance is a most effective, yet a challenging maintenance strategy striving to "repair before failure". Performing replacement/repairs on major equipment components too early or too late increases construction equipment's operating costs. Currently, contractors are performing repairs before failure or just-in-time repairs, with heavy reliance on condition monitoring and professional judgment.

A large amount of equipment failure data are collected on equipment maintenance history and stored in the contractor's equipment management information system, these equipment data can be analyzed to discover the hidden cause-effect relationships, with the application of data mining in computing science. For example, Bayesian network modeling techniques can be used to discover and describe clearly the inter-dependencies and causalities linking failure events and their factors of impact. Such relationships can help to overcome biased views and judgments in failure analysis.

This research is to study on the identification of hidden cause-effect relationships from equipment failure data for construction of Bayesian Network diagnostic model. While the current practice focuses on a "hypothesis-and-test" process, the proposed approach aims to update knowledge of causality from data currently available. A hybrid approach of combining existing Bayesian network with updating from data is proposed in this research for diagnostic analysis of equipment failures.

RELATED RESEARCH

Because the Bayesian networks can handle well the decision scenarios involving uncertainties and correlated decision variables, Bayesian Network modeling has been applied to construction management for decision support. McCabe et. al. (1998) combined the Bayesian network with simulation models for automatic resource optimization on earth-moving operations; in their research, Bayesian Belief Networks are used to suggest remedial actions that will improve the project performance. Chung et. al. (2006) applied Bayesian technique into a tunneling project for updating the TBM penetrating rate based on accumulated evidence on project performance. Recently Bayraktar and Hastak (2009) applied Bayesian network in decision support system for evaluating different construction strategies based on the Bayesian Network prediction results on a set of project performance indicators. Though these research projects demonstrated the effectiveness of Bayesian network in prediction under uncertain and complex conditions, all the networks were manually constructed and cause-effect relations were identified primarily by domain experts.

In construction plant and equipment management, Edwards et al. (1998) conducted a comprehensive review on the monitoring and predictive analysis techniques in manufacturing industry, the researchers advocates more vigorous use of these techniques and RCFA in construction plant management. Many other researchers and industry leaders have emphasized the important of "repair before failure", prevention, and data analysis for predictive and proactive maintenance of equipment fleet (Steward, 2004; Voster, 2007).

BAYESIAN THEORY AND BAYESIAN NETWORK

Traditional statistical approach for inferring on the occurrence of events with uncertainties relies on a probabilistic distribution based on long term observations. For example, if the probability of a failure event is 5%, it merely indicates the likelihood of an event, and no further conclusion can be drawn regarding its actual occurrence for a particular case. Compared with probabilistic thinking on event occurrence with a probability, Bayesian approach focuses on prior knowledge and observed evidence: for a failure event Y, if events X_1, X_2, X_3… X_n are causal factors of Y, then the likelihood of event Y can be updated based the prior belief on the event occurrence and observed data on these causal factors:

$$\Pr(Y \mid X_1, X_2, \ldots X_n) = \frac{\Pr(Y) \cdot \prod_{i=1}^{n} \Pr(X_i \mid Y)}{\prod_{i=1}^{n} \Pr(X_i)} \qquad \text{(1)}$$

Where Y – Failure event to be predicted,

X_i – Causal factors of Y, i=1,2…n

$\Pr(Y \mid X_1, X_2, \ldots X_n)$ – Probability of event Y given collective evidence of X_i, i=1,2,…n

$\Pr(Y)$ – Probability of event Y

$\Pr(X_i)$ – Probability of the causal factor X_i

$\Pr(X_i \mid Y)$ – Probability of the causal factor X_i given Y

The above theory is called Naive Bayesian theorem which assumes that those causal events (factors) are independent from each other.

The Naive Bayesian theorem can also be used to update the prior knowledge on event outcome. The Bayesian updating method considers the model parameters as changing variants, which can be updated by observed evidence. In real application, a prior distribution of the model parameters is first assumed based on past experiences or subjective judgment; If the factual information is available, the prior distribution is updated by the likelihood that the observed factual data fall into the prior distribution. This updating process draws prior probability distribution closer to its true distribution and hence the posterior probability of model parameters is more accurately obtained. In short, Bayesian theory can be used to describe the occurrence of a probabilistic event with due considerations to its causal factors, or update our current knowledge on an event when additional evidence on its dependents is collected.

Bayesian network is a Directed Acrylic Graphical (DAG) network comprising of nodes and connecting edges. A graphical node denotes causal events or factors, while the arrowed edge connecting two nodes denotes a pair of cause-effect relationship, with edge start (parent) as the cause, and edge end (child) as the effect. Some important features of this prediction model are: (1) parent-child relationships are cause-effect or influence-result with probability values (0-1) depicting the degree of causality; (2) parent-child relationships are based on Bayesian' theorem, i.e. any node (event occurrence) is dependent on the joint probability of its parent nodes; however the parent nodes are independent from each other; and (3) no acrylic edges in the graph.

Bayesian network can be constructed based on the expert opinions (Bayesian Belief network), or statistical evidence (Bayesian probability network). Compared with other decision models such as decision trees, Bayesian network models have some unique advantages in problem modeling and analysis. First of all, the models are easy to interpret because of its graphical representation of a problem domain; secondly, the correlations among these causal factors can be modeled in a Bayesian network, in contrast with most other models where these variables are assumed to be independent and uncorrelated, which is seldom true in real life situations; thirdly, with the rapid development of computing science in AI, the Bayesian network can be updated based on the accumulated knowledge, which is discovered from data automatically or semi-automatically using computer algorithms; expert knowledge on causality can be incorporated into the network prior to learning the Bayesian model from data.

LEARNING BAYESIAN NETWORK MODELS FROM DATA

Bayesian network can be induced automatically from large amounts of observed data on failure events and potential information on their causal events. Given historical observations on failure event Y and conjectured causal variables X_i (i=1,2…n), as well as observed cases represented as a dataset D, computer algorithms can be used to sift through this data case table to learn a Bayesian network model which best fits the relationships hidden in data. Bayesian network learning includes parameters learning and structure learning as explained below:

Parameters Learning

There is a conditional probability table (CPT) attached to each node or variable. In the simplest case, the CPT values can be estimated using Naïve Bayes theory based on training dataset after discretization of continuous variables (Han and Kamber, 2006).

The learning of CPT values is based on the observed data. First, the continuous variables are discretized based on the predefined criteria, data values in each cohort mapped to a label; second, calculate each value in the Conditional Probability Table (CPT) cell based on Naïve Bayesian theory for each node, conditional on the combined conditions of its parent nodes; third, although all the values in CPT can be learned automatically, it is also likely that the computed value is not as trustworthy as "subjective assessment", or prior value due to various reasons. In this case, the decision maker can judge on which one to use, or take a weighted average.

Structure Learning

Structure learning addresses the general problem of determining the existence of statistical dependencies among variables. If variables of influencing factors or events are represented as graphical nodes in the Bayesian network, structure learning identifies the directed edges between nodes with each one indicating a pair of cause (arrow start) and effect (arrow end).

In structure learning, the algorithm searches for an optimum structure in the space of all possible structures for a given set of variables representing the application domain (Luger, 2009). Due to the huge number of possibilities for a problem domain involving a large number of variables, it is inefficient and impractical to search exhaustively on all possible structures by the computer, instead, heuristic search algorithms, including greedy search, greedy search with restarts, best-first search (Heckerman, 1995), are employed. In a typical search algorithm, a scoring criterion is defined for the goodness-of-fit of the Bayesian Model, for example, the coefficient of determination R^2, on the observed data; the algorithm first builds a random tree structure, and evaluates possible changes to the edges. The changes are made if it can increase the overall model score.

The most important step is to learn the cause-effect relations. While the association between variables is easy to discover, their actual casual relations is difficult to identify, as there might be many causes for one results, and the direct causes of an event might be missing or not fully represented in the problem, because the selection of variables and their observations are still subjective. In the Markov blanket structure learning algorithm, a node is considered to be conditionally independent of all other nodes given its Markov blanket, i.e. its parents, children, and spouses (Koski and Noble, 2009). This assumption can greatly simplify the learning process.

For each node, the following nodes can be identified based on calculated information: (1) Father nodes: those nodes that bring more information jointly than alone; (2) Sons: nodes having a direct probabilistic dependence with the target; (3) Spouses: the parents of common children, Those nodes which are marginally independent of the target but become informative when knowing the value of the son (BayesiaLab Version 5.1).

BAYESIAN NETWORK UPDATING BY LEARNING FROM DATA: A HYBRID APPROACH

Many commercial or open-source software packages are available for automatic Bayesian network learning. In practice, the fully automatic learning procedure can be difficult to apply, when there are a large number of variables, or the system is not well understood and the variables are not well defined in the first place. In addition, different learning algorithms can produce different network structures, from the simplest "star" structure (with target in the center and influencing variables outside), to the complex bushy tree-type structure.

A hybrid approach is based on predefined Bayesian network. After obtaining operational data, these factual data can be used to update the parameters and update the structure based on similar learning procedure. During the learning process, we can fix certain known relations or node information and let the learning algorithm to propose changes for further evaluation. The hybrid approach is shown in Figure 1.

CASE STUDY

Hydraulic system is an important sub-system of an excavator; any failure in hydraulic system can lead to a suspension of operations and project stoppage. Hydraulic system is a common type of failure in construction machinery as it delivers power directly to accomplish digging, loading tasks, and this subsystem is affected by a wide range of factors on the jobsite. The contractor is very interested in analyzing the failures of a hydraulic system to: (1) identify the critical factors causing the hydraulic system failures; evaluating their degree of impact; (2) predict if the hydraulic system will fail to function given a set of working conditions; and (3) take preventive measures to effectively mitigate the primary failure sources.

In this case study, a highway construction & maintenance contractor has collected large amount of maintenance data in the equipment management system, all the failure events of the hydraulic system of excavators have been recorded along with the related information. Interpretation of those failure data can help the contractor to describe and better understand the failure mechanism and improve its management practice. Based on the Reactive Maintenance Priority Classifications table, Level III and above are mission critical and can cause imminent impact on project according to Dhillon (2002). Therefore it is important for the contractors to identify failure sources and predict these severe failure events based on the past experience.

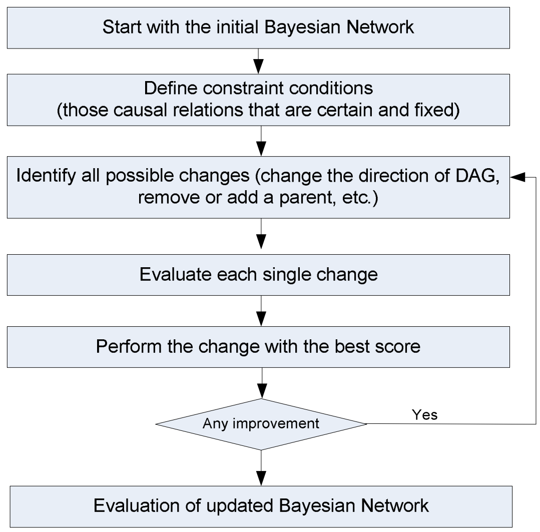

Figure 1 - Flow chart for Bayesian network learning and updating

For diagnostic analysis, Bayesian network can be built directly from expert knowledge. Direct causes of hydraulic system failures include oil leakage, clogged oil filter, failed hydraulic pumps, etc. and the operator can sense the failures from insufficient digging power, alarms on high oil temperature, or a low oil level. In practice, such diagnostic Bayesian network faces some challenges. First, expert judgment on conditional probabilities is subjective and might not be accurate. Second, network topography can be complex and it is difficult to reach consensus among experts, and thirdly, the network is evolving with accumulation of knowledge. The network can change dynamically in both network parameters, and network structure due to changes in environmental conditions, job natures, and operators.

In this case study, a Bayesian network for prediction of failures at Level III and above was built first based on expert knowledge, the Bayesian network is then updated using factual data to reflect more on the real situation based on the following procedures:

(1) Data Preparation

The contractor owns a large fleet of construction equipment for highway construction, among which a total of 7 hydraulic excavators with bucket capacity of 3.0 cubic yard and above are selected for study. Within one year period, there are 52 failure cases which are classified as Level III and above, which caused suspension of excavation works for an extended period of time (3 hours and above). All the data sources related to an equipment failure event are collected and represented in an appropriate format. The target node (Hydraulic system failure event of Class III and above) and all the variables (potential influencing factors) along with the historical cases must be represented in a tabular format, with each column heading representing results and variables, and each row representing failure cases.

(2) Bayesian Network Construction

A Bayesian network is built from expert knowledge and CPT values for each node are estimated through interview with maintenance crew and equipment manager. Originally three direct factors (hydraulic pump failure, oil leakage, clogged oil filter) are identified as the direct sources of hydraulic system failure alone or together. The likely root causes, i.e., the perceived sources of the three categories of problems, are also included in the leaf nodes of the diagnostic network.

(3) Bayesian Network Updating

In the study, BayesiaLab®(BayesiaLab Version 5.1) from Bayesia Ltd. is used for network learning and updating. Although the diagnostic network can be learned completely from data using BayesiaLab, the result is not satisfactory. Different algorithms are used from the simple "association discovery" to sophisticated heuristic search algorithms, with varying results. It is found that a hybrid approach performs best in this experimental study. Based on the original network, the following information on the network is updated: first, the node information (Conditional Probability Table, or CPT) is computed from data and compared with the existing CPT for updating, second, the DAGs in some local areas are modified while keeping the main relations, which are defined as "constraints" in the network. In addition to the updating of node parameters, the directed edge from the node "Operator skills level" to the node "hydraulic pump failure" is removed, the directed edge from the node "Dirty working environment" to "High oil temperature" is added, and "High oil temperature" is moved to be the direct co-parent of the target node. This proves that the previous judgment on these causal relations is not accurate, for example, the leaf node "Operator skills level" is removed as it does not have significant contributions to the "Hydraulic pump failure", and part of its effects is already reflected in the node "Overloaded working conditions", which are primarily caused by the tough working conditions. The node "Operator skills level" does not have statistical significance as an independent contributor to the "Hydraulic pump failure" events according to the equipment operation and maintenance records for this particular contractor. The updated network topography is shown in Figure 2.

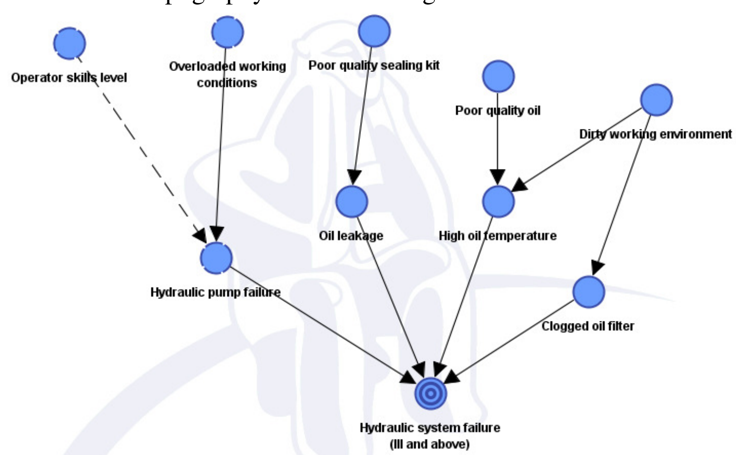

Figure 2 - Bayesian network for hydraulic system diagnostic failures of construction equipment
(Note: directed edge of the dotted line is removed after learning and updating.)

Bayesian Network Validation

There is significant improvement in model quality after updating the Bayesian network through learning from data, compared with the original Bayesian network. For validation analysis of the updated network, the probabilities of system failures are converted into uninterrupted working time of the hydraulic system under 95% reliability, we also measure the time between the previous failure and current failure for each of the 50 failure cases for comparison.

Considering the small dataset in this case study, 10-fold hold-out validation tests are conducted on the Bayesian network updating. The 50 cases are randomly divided into 10 subsets with each containing a subset (5 cases). With one subset is reserved for validation, the other nine subsets (45 cases) are used for Bayesian network updating. The updated Bayesian network is then used to predict the time to failure for the reserved 5 cases, with the results compared with the actual recorded values and the prediction results from the original Bayesian network. After repeating the above procedure for each subset of failure cases, 10 sets of results are obtained with each containing the actual time to failure, predicted time to failure with the original Bayesian network, and predicted time to failure with the updated Bayesian network. According to the experimental results, the average percentage error of prediction with the original Bayesian network is 72.5%, and the average percentage error of prediction with the updated Bayesian network is increased to 86.9%.

Concluding Remarks on the Case Study

Through this case study, the important root causes of the problem are identified and their degree of impact can be quantified through "what-if analysis" by changing root cause node conditions. For example, if a change in oil quality, represented by the node "Poor quality oil", can lead to a significant improvement in the probability of system failure, extra costs in high quality oil can be justified. The network can also be used for life cycle analysis of a system; under the given set of working conditions, the concept of failure probability can be converted into the uninterrupted working time of a system; repair or replacement decisions can be made just ahead of time to avoid sudden and expensive failures on the shift.

DISCUSSIONS

The hybrid approach used in this case is considered a good solution to "learning from data" for diagnostic analysis in an engineering system or a process. A complete acquisition of Bayesian network from data is still difficult although it is theoretically feasible. The difficulty of learning the entire network topography can be explained by: (1) the data collection is not unbiased, rare cases might be equally important and data collected are not sufficient, (2) attribute selection is subjective, and better variables may exist to describe the state of a causal factor, and (3) some failure mechanisms are still difficult to understand in a complex system with disturbance from unpredictable environmental conditions, such as the hydraulic system of construction equipment in this case.

Updating of Bayesian network based on failure data is also important as human beings cannot predict the probability of an event or conditional probability with high accuracy; discrepancies cannot be avoided even with Delphi techniques from a panel of experts. Factual data can be used to learn these prior probabilities and conditional probabilities to reflect the factual information. For a complex system, it is even difficult for the domain experts or system modelers to fully understand the failure mechanism in some cases, hence it is necessary to use factual information to modify or fine-tune the Bayesian network topography using cause-effect knowledge acquired from observational data.

CONCLUSIONS

Heavy construction equipment is an important type of resources for civil engineering contractors. Failures of equipment, especially those mission critical ones, should be controlled and managed properly. Equipment should be fixed in a convenient time and location before failure if such failure can be reasonably predicted. Mitigation of failure factors can also help to reduce the chance of failure if they are properly identified. Considering the scale of outlay incurred in construction equipment repairs and their impact on the project, any improvement in the management goal of "repair just before failure" in equipment operations can have a significant impact on the contractor's economic performance at a project or corporate level. Bayesian networks are a useful tool for diagnostic analysis and decision support and are normally built based on the contractor's expertise or prior knowledge; however both the conditional probability information and causal relations are subjective. In this paper, the research issues on updating Bayesian network based on factual information are discussed. A hybrid approach is used in the case study to update probability information and local cause-effect relations while keeping the known and confirmed causal structures. This research shows that a hybrid approach can be applied to obtain the optimum Bayesian network for failure analysis of construction equipment. The series of causes can be confirmed through the updated network from observational data. The research findings can be applied to any other diagnostic analysis tasks in project management, such as project quality defects analysis, project schedule bottleneck analysis and improvement, structural maintenance strategy analysis. A hybrid approach will best utilize the expert knowledge and the accumulated factual information in constructing a Bayesian Belief network for problem identification and system improvement in construction engineering and management.

ACKNOWLEDGEMENT

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong SAR, P. R. China (Project No. PolyU 5170/09E, General Research Fund).

REFERENCES

BayesiaLab (Version 5.1) [Computer software]. France: Bayesia, Ltd.

Bayraktar, M. E. and Hastak, M. (2009). Bayesian Belief Network Model for Decision Making in Highway Maintenance: Case Studies. *Journal of Construction Engineering and Management*, ASCE, 135(12), 1357-1369.

Chung, T. H., Mohamed, Y., and AbouRizk S. (2006). Bayesian Updating Application into Simulation in the North Edmonton Sanitary Trunk Tunnel Project. *Journal of Construction Engineering and Management*, ASCE, 132(8), 882-894.

Dhillon, B. S. (2002). *Engineering maintenance, a modern approach*. FL: CRC Press.

Edwards, D. J., Holt, G. D., and Harris F. C. (1998). Predictive maintenance techniques and their relevance to construction plant. *Journal of Quality in Maintenance Engineering*, Emerald, Vol. 4 No. 1, 25-37

Han J. and Kamber M. (2006). *Data mining concepts and techniques*, 2^{nd} edition. Singapore: Elsevier Singapore Pte Ltd.

Heckerman, D. (1995). *A Tutorial on Learning with Bayesian Networks*. Technical report, Microsoft Research, Advanced Technology Division, Microsoft Corporation, Seattle, WA. Retrieved from: ftp.research.microsoft.com/pub/tr/tr-95-06.pdf

Koski T. and Noble J. M. (2009). *Bayesian networks, an introduction*. West Sussex, England: John Wiley and Sons, Inc.

Luger, G. F. (2009). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 6^{th} Edition, (pp. 556-557). London, UK.: Pearson Education Inc.

McCabe, B., AbouRizk, S. M., and Goebel R. (1998). Belief Networks for Construction Performance Diagnostics. *Journal of Computing in Civil Engineering*, ASCE, 12 (2), 93-100.

Steward, L. (2004). R for profit: repair before failure. *Construction Equipment: Boston*, ABI/Inform Global, 107(11), 28-32

Vorster, M. (2007). Prevention Is Better Than Cure. *Construction Equipment: Boston*, ABI/Inform Global, 110(2), 65-67