Supporting the Selection of Prognostic-based Decision Support Methods in Manufacturing

Alexandros Bousdekis[1], Babis Magoutas[1], Dimitris Apostolou[2] and Gregoris Mentzas[1]
[1]School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece
[2]Department of Informatics, University of Piraeus, Piraeus, Greece

Keywords: Condition based Maintenance, Decision Tree Learning, Method Filtering, Decision Support.

Abstract: In manufacturing enterprises, maintenance is a significant contributor to the total company's cost. Condition Based Maintenance (CBM) relies on prognostic models and uses them to support maintenance decisions based on the current and predicted health state of equipment. Although decision support for CBM is not an extensively explored area, there exist methods which have been developed in order to deal with specific challenges such as the need to cope with real-time information, to prognose the health state of equipment and to continually update decision recommendations. We propose an approach for supporting analysts selecting the most suitable combination(s) of methods for prognostic-based maintenance decision support according to the requirements of a given maintenance application. Our approach is based on the ID3 decision tree learning algorithm and is applied in a maintenance scenario in the oil and gas industry.

1 INTRODUCTION

In manufacturing enterprises, high reliability, low environmental impact and safety of operations are important issues for every industry (Peng, Dong, and Zuo, 2010). Maintenance is a significant contributor to the total company's cost, so optimal maintenance policy in terms of cost, equipment downtime and quality should be identified (Garg, and Deshmukh, 2006). Condition Based Maintenance (CBM) is a type of maintenance strategy, which relies on diagnostic and prognostic models and uses them to support decisions about the appropriate maintenance actions based on the current health state of a system through condition monitoring (Jardine, Lin, and Banjevic, 2006). Condition monitoring in manufacturing enterprises is increasingly realised with equipment-installed sensors, which have the capability of measuring with high frequencies a multitude of parameters. This capability leads to storage of a huge amount of data. Generating and storing Big Data has become possible due to recent developments in both hardware and data management software (Zikopoulos, and Eaton, 2011).

Big Data-driven CBM poses challenges to Decision Support Systems. These challenges are not easily addressed within the complex manufacturing environment, especially when dealing with maintenance where several factors should be considered simultaneously such as costs of maintenance actions as a function of time, safety issues and equipment degradation.

Although decision support for CBM is not an extensively explored area, there exist several works focusing on combinations of methods that can be utilised for CBM decision support. Such methods deal with real-time data which are gathered in high frequency, develop prognostic models for the estimation of Remaining Useful Life (RUL) or Remaining Life Distribution (RLD) and provide recommendations for maintenance. In the current paper, we propose a practical approach for supporting analysts to select the most suitable combination(s) of methods for prognostic-based maintenance on the basis of Big Data according to the requirements of the application which they are involved with.

Aiming to support prognostic-based maintenance in various application domains and for a wide range of functional and non-functional application requirements, we follow a practical multistage decision making approach. The basic idea of our hierarchical approach is to break up the problem of selecting the most suitable combination(s) of

Bousdekis A., Magoutas B., Apostolou D. and Mentzas G..
Supporting the Selection of Prognostic-based Decision Support Methods in Manufacturing.
DOI: 10.5220/0005372104870494
In *Proceedings of the 17th International Conference on Enterprise Information Systems* (ICEIS-2015), pages 487-494
ISBN: 978-989-758-096-3

methods for prognosis and prognostic-based maintenance, into a union of several simpler decisions about the suitability of the method combination depending on the functional and non-functional application requirements by developing a Decision Tree (DT).

To formulate the information found in the papers examined in our literature review in the form and structure needed for feeding the DT learning algorithm and building the DT, the following steps were performed. First, generic categories of methods for prognosis and prognostic-based maintenance were defined in our effort to avoid specific method extensions or variations used in the various papers and keep the resulting DT as generic as possible. Second, unique combinations of the previously defined generic methods, which are actually the leaf nodes of the DT, were identified in the papers reviewed. Third, rules for classifying the method combinations using criteria that depend on functional and non-functional application requirements were defined; such rules and criteria are used in the decision nodes of the DT. Fourth, for all method combinations identified, the fulfilment of the criteria used in the DT's decision nodes was assessed. Finally, the appropriate DT learning algorithm was selected and fed with the training dataset.

The rest of the paper is organized as follows: Section 2 presents the literature review; Section 3 outlines the method filtering approach, while Section 4 illustrates its application in a maintenance scenario in the oil and gas industry. Section 6 discusses the method filtering approach and the results and concludes the paper.

2 LITERATURE REVIEW

2.1 Literature Search and Pre-filtering of Results

Several research works have examined and developed maintenance decision support methods, based on historical and real-time data as well as expert knowledge, in order to address different maintenance challenges. Most of the existing research works address maintenance issues for components subjected to condition monitoring in the context of Condition Based Maintenance (CBM). Maintenance decision support is related to reliability, safety and environmental issues as well as associated with equipment downtime costs in cases of breakdowns or malfunctions of machines (Peng, Dong, and Zuo, 2010). First, prognostic methods are applied and then decision methods are developed in order to provide prognostic-based recommendations. Table 1 summarises the prognostic-based decision support methods reviewed, as well as their inputs and outputs. The methods have been separated in two groups: one group supports the Prognostic (P) and the second one the Decision (D) step.

The papers examined were identified by searching Google scholar with the keywords 'CBM, 'recommendations', 'decision support', 'decision making',' manufacturing', 'maintenance' in various combinations among them. We focused on papers dealing with the decision step of CBM. However, we realized that most of them proposed a combination of methods so that they develop a prognostic model based on real-time data and then, based on this, they provide recommendations for maintenance. The focus was on most recent papers, after 2008, with exceptions in cases where an older paper satisfied the keywords and proposed a novel and useful method which has not been extended until now.

2.2 Categorising Methods

The methods found in the literature for prognostic-based decision support can be divided into the following generic categories:

- Bayesian Network (BN), which also include Dynamic Bayesian Network
- Neural Network (NN)
- Statistical Analysis (SA), which include statistical techniques such as Statistical Quality Control (SQC), Support Vector Machine (SVM) and moving average.
- Degradation Modelling (DM), which includes all the mathematical techniques dealing with representing the degradation process.
- Reinforcement Learning (RL), such as State-Action-Reward-State-Action (SARSA) algorithm.
- Markov Chain (MC)
- Mathematical Programming (MP)-Optimisation, which includes operational research methods such as linear, non-linear and stochastic dynamic programming
- Markov Decision Process (MDP), which also includes Semi-Markov Decision Process (SMDP) and Partially Observable Markov Decision Process (POMDP).

- Rules (R), which include rule-based systems such as IF-THEN rules and Event-Condition-Action (ECA) rules.

2.3 Identifying Combinations of Methods

As shown in Table 2, there exist in the reviewed literature ten unique combinations of methods used for providing prognostic-based recommendations.

3 METHOD FILTERING APPROACH

3.1 Identifying Method Filtering Criteria

Following our analysis of existing methods and combinations of methods, we propose criteria for selecting the appropriate ones based on the functional and/or non-functional requirements of specific applications. Selection should be based on desired output that the business analyst expects to get after the implementation of the method combination. Depending on the available input, different combinations of methods can be applicable. Another criterion is whether Domain Knowledge (DK) can be expressed in terms of utility functions. Finally, the existence of degradation knowledge affects the selection of the appropriate prognostic methods and thus the selection of the suitable combination of methods.

3.2 Evaluating Methods

Based on the information given in Table 1, the identified combinations of methods are evaluated according to the four specified criteria. Evaluation of method combinations on the criteria of available input and desired output was done based on the information summarized in Table 1. For evaluating method combinations on the other two criteria, we examined in more detail the information provided in the respective papers. Desired output can be either the optimal time of applying a predefined action (e.g., optimal time of replacement of some part of equipment) or the optimal action and the optimal time of applying it (e.g., lubrication of metal parts accompanied with the optimal time). Available input can include historical data about cause (e.g., vibration, temperature, etc.) and effect (e.g., failure, malfunction, etc.) or prognostic information. Knowledge of the degradation process is a prerequisite for some prognostic methods, while this is not the case for others. Table 3 shows the evaluation of the methods' combinations using the four specified criteria.

3.3 Decision Tree Learning

The data presented in Table 3, were fed into a DT learning algorithm, i.e. a DT classifier, in order to produce the DT for supporting analysts perform prognostic-based maintenance in various application domains and for a wide range of functional and non-functional application requirements. DT classifiers have the ability to handle data which are measured in different scales, they do not require any assumptions about the frequency of data in each class, while they are able to handle non-linear relationships between features and classes. Furthermore, the analyst can comprehend and interpret a decision tree as it is not a 'black box' (Pal, M., and Mather, 2003). The ID3 (Iterative Dichotomiser 3) algorithm, that was used in our case, classifies all training data provided that there are enough attributes to do so.

There are several extensions of the ID3 algorithm, such as J48, C4.5 and C5.0, which, among others, are able to handle continuous attributes, training data with missing attribute values and attributes with differing costs. However, these capabilities are not useful in our case because there are not any related issues to address. Moreover, these additional capabilities provide improvements in terms of speed and memory usage, which are nevertheless not needed in our case because it consists of a small number of combinations of methods and the DT learning is done once. Finally, the aforementioned extensions of ID3 create smaller DTs because the probability of over-classifying the data is much smaller compared to ID3. However, in our case, we want to separate our method dataset as much as possible. The four criteria used for the separation of the combinations of methods and their alternative values were defined in an abstract level. This means that if, for example, two combinations of methods are classified in the same class, they are not necessarily the same and cannot be used interchangeably because e.g., their input may require additional, more specific data or knowledge than those specified in Table 3.

The ID3 decision trees algorithm is based on information theory and tries to minimize the number of comparisons among the data of the training set. The core idea behind the algorithm Is asking

Table 1: Reviewed Research Works on Prognostic-based Decision Support.

Reference		Methods	Input	Output
(Kaiser, and Gebraeel, 2009)	P	BN; DM	Real-time and historical data; Threshold; Replacements	Estimation of RLD
	D	R	RLD; Costs of maintenance; Process knowledge	Compute/update maintenance schedule
(Besnard, and Bertling, 2010)	P	DM; MC (Continuous time);	Degradation condition monitoring; Degradation process	RUL; Failure rate
	D	MC (Continuous time); MP; R	RUL; Failure rate; Maintenance and production knowledge and costs	Optimal maintenance strategy
(Besnard, et al., 2011)	P	-	-	-
	D	MP (Stochastic); R	Wind forecasting ; Failure rate; List of actions; Production and maintenance knowledge and costs	Minimised cost of production losses and transportation
(Castro, et al., 2012)	P	DM	Real-time and historical data; Threshold	Mean Residual Life; Times of replacement
	D	MP (cost minimisation)	Mean Residual Life; Times and costs of maintenance	Minimised maintenance cost; Optimum policy
(Wu, et al., 2007)	P	NN; MP (Non-linear programming); SA (Moving average)	Real-time and historical data; Threshold	Residual Life Percentile Prediction
	D	MP (Non-linear programming)	Predicted Residual Life Percentile; Times of operation; Costs related to maintenance	Minimised cost; Optimal replacement time
(Ivy, and Nembhard, 2005)	P	SA (SQC)	Real-time and historical data; States; Threshold	Transition Matrix; Estimation of observation distribution parameters
	D	MDP (POMDP)	Transition Matrix; Estimation of the observation distribution parameters; Maintenance costs	Minimised expected cost ; Optimal maintenance and monitoring actions
(Aissani, Beldjilali, and Trentesaux, 2009)	P	RL (SARSA algorithm)	Real-time and historical data; Degradation and maintenance knowledge	Solution of SARSA algorithm; Probabilities of events
	D	MDP	Solution of SARSA algorithm; Probabilities of events	Predictive and corrective maintenance tasks
(Elwany, and Gebraeel, 2008)	P	DM	Real-time and historical data; Failure threshold	RLD
	D	MP (replacement model)	RLD; Costs of maintenance; Lead times of spare parts	Optimal replacement and inventory ordering times
(Bouvard, et al., 2011)	P	DM; SA	Real-time monitoring; Degradation process; Threshold	Failure probability; Estimated degradation path; Time-to-failure
	D	MP (maintenance optimization)	Failure probability; Time-to-failure; Maintenance costs	Optimal maintenance cost and planning
(Huynh, Barros, and, Berenguer, 2012)	P	DM	Condition monitoring; Degradation process	Reliability; Probability density function
	D	MP (dynamic replacement model); R	Reliability; Probability density function; Cost function	Replacement Time Estimation; Optimised cost
(Muller, Suhner, and Iung, 2007)	P	BN (DBN)	Real-time and historical data; Process	Degradation process; Prognosis
	D	MC (Discrete time); R	Degradation process; Prognosis; List of actions; Costs	Optimal maintenance policy
(Engel, Etzion, and Feldman, 2012)	P	BN	Real-time monitoring; Historical data of transitions	Probability distribution of an event; Time-to-failure
	D	MDP	Probability distribution of an event; Time-to-failure; States; Actions; Cost function	Optimal action; Optimal time of action

Table 2: Methods and Techniques for Decision Making.

Reference	BN	NN	SA	DM	MC	RL	MP	MDP	R
(Kaiser, and Gebraeel, 2009)	v			v					v
(Besnard, and Bertling, 2010)				v	v		v		v
(Besnard, *et al.*, 2011)							v		v
(Castro, *et al.*, 2012)				v			v		
(Wu, *et al.*, 2007)		v	v				v		
(Ivy, and Nembhard, 2005)			v					v	
(Aissani, Beldjilali, and Trentesaux, 2009)						v		v	
(Elwany, and Gebraeel, 2008)				v			v		
(Bouvard, *et al.*, 2011)				v			v		
(Huynh, Barros, and, Berenguer, 2012)				v			v		v
(Muller, Suhner, and Iung, 2007)	v				v				v
(Engel, Etzion, and Feldman, 2012)	v							v	

Table 3. Methods' combinations evaluation.

Combinations of methods	Desired Output		Available input		DK expressed in utility function		Knowledge of the degradation process	
	Time of action	Action and time	Historical data	Prognosis	Yes	No	Yes	No
NN-SA-MP	v		v			v		v
BN-DM-R		v	v			v	v	
BN-MC-R		v	v		v		v	
RL-MDP		v	v		v			v
DM-MC-MP-R		v	v		v		v	
SA-MDP		v	v		v		v	
DM-MP-R		v	v			v	v	
DM-MP	v		v			v	v	
MP-R		v		v		v		v
BN-MDP		v	v		v		v	

questions the answers of which provide the most information. The splitting criteria are prioritized according to the information gain; splitting criteria with more information gain are used first. The decision tree is constructed by employing a top-down, greedy search through the given sets to test each attribute at every tree node. Information is measured by the entropy which represents the amount of uncertainty of a data set D (Chen, Zhang, and Tong, 2014). Based on the entropy, the information gain can be measured. Information gain is the difference in entropy from before to after the data set D is split on an attribute A or equally, how much uncertainty in the data set was reduced after splitting it on an attribute A (Gaddam, Phoha, and Balagani, 2007).

The DT was built by feeding the combinations of methods identified in the literature as training data to the decision tree learning algorithm and shows the sequence of steps needed to be followed by an analyst in order to decide which combination(s) of methods are the most appropriate ones for a specific problem. The pseudo-code of the application of the ID3 algorithm for the classification of the combinations of methods according to the four criteria is shown below (adapted from (Jin, De-lin, and Fen-xiang 2009)):

```
ID3 (Set of combinations of methods D,
Set of criteria-attributes S, Criteria-
Attributes_values V)
Return Decision Tree.
Begin

Load set of combinations of methods D
first, create decision tree root  node
'rootNode', add learning set D into
root node  as its subset.

For rootNode, we compute
    H(rootNode.subset) first
    If H(rootNode.subset)==0, then
            rootNode.subset consists of
           records all with the same
           value for the  categorical
           criterion-attribute,
           return a leaf  node with
           decision        criterion-
```

```
           attribute:         criterion-
           attribute value;
    If H(rootNode.subset)!=0, then
           compute     IG     for     each
           criterion-attribute       left
           (have   not   been   used   in
           splitting), find attribute S
           with    Maximum(IG(D,    S)).
           Create  child  nodes  of  this
           rootNode and add to rootNode
           in the decision tree.
           For   each   child   of   the
           rootNode,
                apply   ID3(D,   S,   V)
                recursively until reach
                node  that  has  H=0  or
                reach  leaf node.
End ID3
```

Our problem was formulated in the ID3 notation using the RapidMiner machine learning software, while the DT obtained after running ID3 is shown in Figure 1.

4 APPROACH ILLUSTRATION IN A MAINTENANCE SCENARIO

In this section, we outline how our method filtering approach can help in selecting the most appropriate method combination for supporting decision making in a maintenance scenario in the oil and gas industry. In the scenario under consideration, sensors collect data with a very high frequency, and these data accompanied with historical data and domain knowledge are used for detecting the current health state of the equipment examined, estimating RUL and calculating the probability distribution of an undesired event, e.g., breakdown of the gearbox of an oil drilling company's equipment. Historical data show the patterns of the monitored parameters, which are used as indicators of degradation till failure. Domain knowledge can include a list of maintenance actions, failure threshold as well as utility functions considering criteria such as cost, time and safety. Then, the optimal action and the optimal time of applying it are recommended.

The DT flow in the aforementioned scenario is shown in Figure 2. First, as far as the available input is concerned, in the aforementioned scenario there are historical data about causes (sensed parameter) and effects (failure) but not prognostic information (e.g. RUL, probability distributions of the occurrence of failure, etc.), while data are continuously updated with the ones coming from sensors.

The prognostic information is not known in advance, but it will derive from the processing and analysis of data by using the appropriate method.

The output of this method will feed into another method for providing recommendations. Then, the desired output is the optimal action and the optimal time for this action, because the objective is to identify the best maintenance action out of a list of actions as well as the best time to implement it in

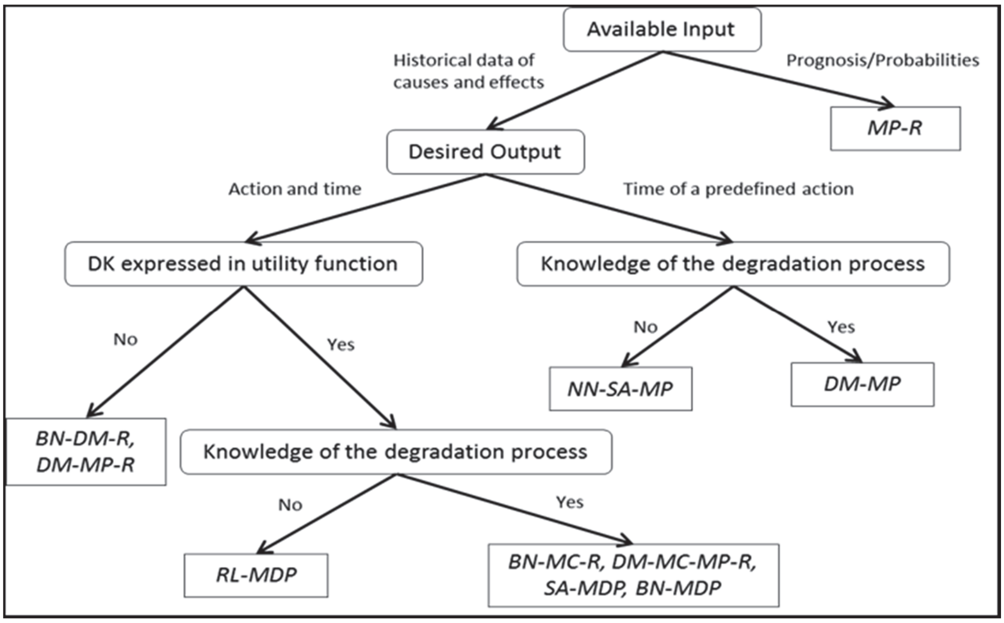

Figure 1: Method filtering process for choosing the appropriate combination of methods.

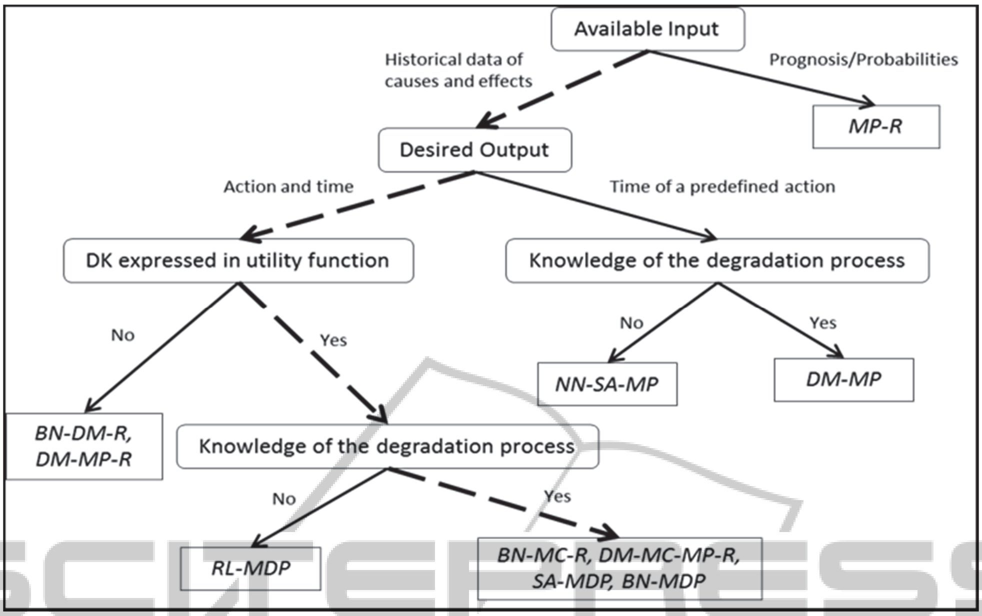

Figure 2: Method filtering process in a maintenance scenario.

terms of cost.

DK can be expressed in utility functions as there is extensive industrial experience on the domain which can be expressed in a systematic way. Issues about cost function, safety, current maintenance policy, etc. can be embedded to a utility function which can be used in optimization techniques in order to provide reliable recommendations. Finally, there is knowledge of the degradation process. Hence, there are four options: BN-MC-R, DM-MC-MP-R, SA-MDP and BN-MDP.

For example, for the BN-MDP combination, BN are used for data-driven estimation of probability distribution of an undesired event (e.g. gearbox breakdown) and MDP for generating recommendations about optimal maintenance actions and optimal time for applying these actions. Moreover, this particular combination can effectively support decision-making in manufacturing enterprise and especially for CBM (Engel, Etzion, and Feldman, 2012). This case shows that MDP and MC are the most suitable methods for extracting this output provided that some domain knowledge exists and probability distribution of an undesired event has been extracted from machine learning or statistical methods.

5 CONCLUSIONS

CBM relies on prognostic models and uses them to support decisions about the appropriate maintenance actions based on the current health state of a system through condition monitoring (e.g. using sensors). To do this, combinations of both machine learning and decision methods are required in a way that they are able to handle real-time data and provide recommendations for maintenance decisions based on predictions about future health state of the equipment.

We examined literature that deals with methods supporting decision making in the context of CBM. The method filtering approach that we propose supports the business analyst to select the most appropriate combination(s) of methods based on the requirements of the specific maintenance scenario. Our method may recommend more than one alternative method or method combinations, which are applicable in specific maintenance scenarios and under specific conditions. In such cases however, methods or combinations of methods that are classified in the same class are not the same and, while our filtering method cannot discern between them, a human expert should be able to do so by taking into account additional data or knowledge in order to select the most appropriate. Although we identified several combinations of methods that are used for prognostic-based maintenance recommendations, most are not able to adequately support proactive decision making by recommending the optimal action and the optimal time of applying it based on predictions. Furthermore, there are limitations regarding the continuous improvement of the recommendations using these combinations of methods.

Our future research will focus on the examination and incorporation of additional machine learning and decision methods specifically targeting proactive decision support. Moreover, we will extend our method filtering approach with a feedback loop, which will support collection of data about the effectiveness of the recommended decisions and will utilize the collected data as a basis for improving the recommendation generation process. Finally, we will test and evaluate our approach in real maintenance scenarios in the oil & gas and automotive industries.

ACKNOWLEDGEMENTS

This work is partly funded by the European Commission project FP7 STREP ProaSense "The Proactive Sensing Enterprise" (612329).

REFERENCES

Aissani, N., Beldjilali, B., Trentesaux, D., 2009. Dynamic scheduling of maintenance tasks in the petroleum industry: A reinforcement approach. *Engineering Applications of Artificial Intelligence*, 22(7), 1089-1103.

Besnard, F., & Bertling, L., 2010. An approach for condition-based maintenance optimization applied to wind turbine blades. *Sustainable Energy, IEEE Transactions on*, 1(2), 77-83.

Besnard, F., Patriksson, M., Stromberg, A., Wojciechowski, A., Fischer, K., Bertling, L., 2011. A stochastic model for opportunistic maintenance planning of offshore wind farms.*In PowerTech, IEEE Trondheim*, pp. 1-8.IEEE.

Bouvard, K., Artus, S., Berenguer, C., Cocquempot, V., 2011. Condition-based dynamic maintenance operations planning & grouping. Application to commercial heavy vehicles. *Reliability Engineering & System Safety*, 96(6), 601-610.

Castro, I. T., Huynh, K. T., Barros, A., Berenguer, C., 2012. A predictive maintenance strategy based on mean residual life for systems subject to competing failures due to degradation and shocks. In *Proc. of the 11th International Probabilistic Safety Assessment and Management Conference & the Annual European Safety and Reliability Conference-PSAM 11/ESREL 2012*, pp. 375-384.

Chen, X. J., Zhang, Z. G., Tong, Y., 2014. An Improved ID3 Decision Tree Algorithm. In *Advanced Materials Research*, Vol. 962, pp. 2842-2847.

Elwany, A. H., Gebraeel, N. Z., 2008. Sensor-driven prognostic models for equipment replacement and spare parts inventory. *IIE Transactions*, 40(7), 629-639.

Engel, Y., Etzion, O., Feldman, Z., 2012. A basic model for proactive event-driven computing. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems*, pp. 107-118.ACM.

Gaddam, S. R., Phoha, V. V., Balagani, K. S., 2007. K-Means+ ID3: A novel method for supervised anomaly detection by cascading K-Means clustering and ID3 decision tree learning methods. *Knowledge and Data Engineering, IEEE Transactions on*, 19(3), 345-354.

Garg, A., Deshmukh, S. G., 2006. Maintenance management: literature review and directions. *Journal of Quality in Maintenance Engineering*, 12(3), 205-238.

Huynh, K. T., Barros, A., Berenguer, C., 2012. Maintenance decision-making for systems operating under indirect condition monitoring: value of online information and impact of measurement uncertainty. *Reliability, IEEE Transactions on*, 61(2), 410-425.

Ivy, J. S., Nembhard, H. B., 2005. A modeling approach to maintenance decisions using statistical quality control and optimization. *Quality and Reliability Engineering International*, 21(4), 355-366.

Jardine, A. K., Lin, D., Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483-1510.

Jin, C., De-lin, L., Fen-xiang, M., 2009. An improved ID3 decision tree algorithm. In *Computer Science & Education. ICCSE'09. 4th International Conference on*, pp. 127-130. IEEE.

Kaiser, K. A., Gebraeel, N. Z., 2009. Predictive maintenance management using sensor-based degradation models. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39(4), 840-849.

Muller, A., Suhner, M. C., Iung, B., 2007. Maintenance alternative integration to prognosis process engineering. *Journal of Quality in Maintenance Engineering*, 13(2), 198-211.

Pal, M., & Mather, P. M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment*, 86(4), 554-565.

Peng, Y., Dong, M., Zuo, M. J., 2010. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4), 297-313.

Wu, S. J., Gebraeel, N., Lawley, M. A., Yih, Y., 2007. A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 37(2), 226-236.

Zikopoulos, P., Eaton, C., 2011. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.