PROGNOSTIC MODELLING WITH DYNAMIC BAYESIAN NETWORKS

Ken McNaught and Adam Zagorecki
{K.R.McNaught, A.Zagorecki}@cranfield.ac.uk
Decision Analysis and Risk Modelling Laboratory
Department of Engineering Systems and Management
Cranfield University
Defence Academy of the UK, Shrivenham, Swindon SN6 8LA.

ABSTRACT

In this paper, we review the application of dynamic Bayesian networks to prognostic modelling. An example is provided for illustration. With this example, we show how the equipment's reliability decays over time in the situation where repair is not possible and then how a simple change to the model allows us to represent different maintenance policies for repairable equipment.

1. INTRODUCTION

Prognostics is concerned with predicting the residual life of an item or the time until the next failure is expected. If this could be predicted with any degree of accuracy, it would allow necessary maintenance activities to be timed in an optimal manner as well as eliminating some unnecessary maintenance altogether. This in turn would provide many benefits, including reduced maintenance costs, reduced downtime and improved equipment availability.

Prognostics is becoming an increasingly important aspect of predictive maintenance, and is closely related to condition-based monitoring (CBM)[14]. While CBM provides information about the past and present condition of a system, however, prognostics is also concerned with its future condition. To be able to predict this also requires a reliable forecast of future equipment use. Although often mentioned in the same breath, prognostics and diagnostics are quite different. Fault diagnosis is a post-event activity, aiming to assist recovery from a failure while prognostics is a pre-event activity, aiming to avoid failure.

There are many approaches to prognostic modelling, including Physics of Failure models and traditional statistically based models. Of most relevance to this conference, however, are those approaches which make use of AI. These include artificial neural networks and dynamic Bayesian networks (DBNs). It is this latter approach which we concentrate on in this paper.

2. BAYESIAN NETWORKS

Bayesian networks (BNs) belong to the family of probabilistic graphical models and have been very successfully employed in fault diagnosis. In this section, we will provide a short description and explanation of BNs. We will also briefly review their applications in the fields of reliability and maintenance, concentrating in particular on fault diagnosis.

Since their introduction in the mid-1980s [17], Bayesian networks have become increasingly popular as a framework for reasoning under uncertainty. While much of the initial growth stemmed from their popularity within the AI community, the last decade or so has brought the attentions of the wider world of science and engineering, including the field of reliability or, more generally, dependability modelling. A recent review of BNs in reliability modelling is provided by Langseth and Portinale [15].

Much of their appeal can be attributed to the flexibility which the modelling framework provides. For example, the same Bayesian network might be used to predict the probability of failure of some equipment and then to diagnose its cause once a failure has occurred. Furthermore, hard statistical data can be combined with softer expert opinion, perhaps regarding environmental factors or design considerations. Extensions of the basic BN further enhance the capability of the approach to not just match but exceed the power of traditional reliability modelling techniques such as fault trees, reliability block diagrams and state-space methods. Demonstrating the power of probabilistic representations, BNs have also been shown to offer numerous advantages in fault diagnosis and detection over alternative AI approaches such as rule-based systems, neural networks and case-based reasoning.

The qualitative structure of a BN is represented by a directed acyclic graph (DAG), portraying probabilistic dependencies and independencies within the domain. This contains a great deal of information, even before we consider any probability distributions. A fully specified BN, however, requires the construction of conditional probability tables (CPTs) for each node. For nodes with no arcs entering them, i.e. with no parent nodes, only a single prior distribution has to be specified. For nodes with a single parent, a conditional probability distribution will have to be specified for each possible state of the parent variable. Finally, for chance nodes with more than one parent, a conditional probability distribution is usually required for every possible combination of parent states. While at first sight, this may appear rather burdensome, there are often special cases where this requirement can be relaxed, e.g. Noisy-OR gates. This amounts to making various independence assumptions.

Consider the BN shown in Figure 1. Prior marginal distributions are required for $P(A), P(B)$ and $P(E)$, while conditional distributions are required for $P(C \mid A), P(D \mid B), P(F \mid C, D)$ and $P(G \mid D, E)$. The joint probability distribution over the full set of variables $U = (V_1, V_2, ...)$ is then given by $P(U) = \prod_i P(V_i \mid pa(V_i))$, where $pa(V_i)$ denotes the set of parents of variable V_i. A major advantage of the BN approach, however, is that the full joint distribution does not usually have to be computed in order to obtain posterior probabilities when new evidence arrives.

The ability of BNs to provide a flexible and powerful probabilistic modelling framework makes them suitable for applications in the field of reliability and maintenance. The earliest and still the most popular application is to fault diagnosis. Sanseverino and Cascio [22] describe one such application to automotive electronic

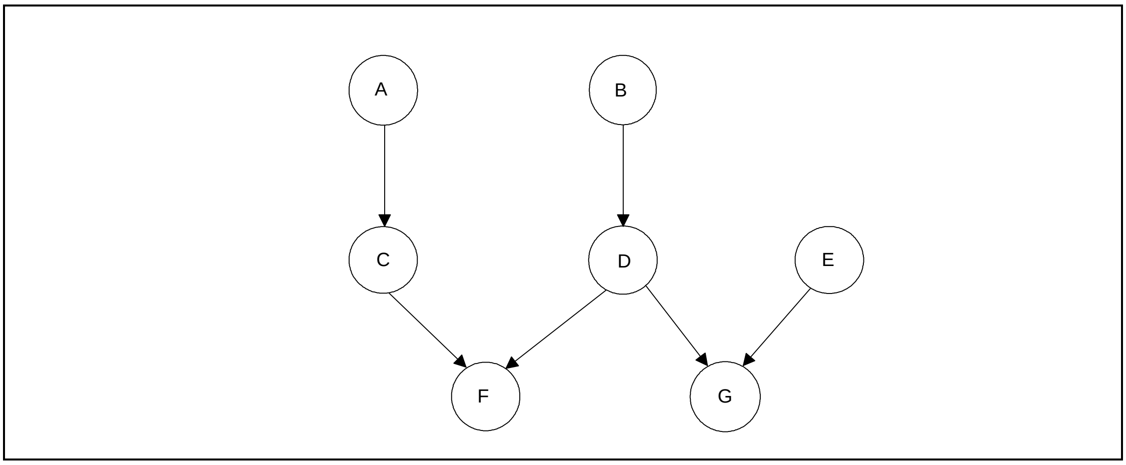

Figure 1. Example DAG for a BN.

sub-systems which has been implemented in hundreds of Fiat repair centres across Italy. Romessis and Mathioudakis [21] developed a BN for fault diagnosis of gas turbine performance in the jet engine domain. Chan and McNaught [6] describe how BNs play a key role in a decision support system designed to provide advice on fault diagnosis and correction during the final system testing of mobile telephone base stations at Motorola. Another application in the mobile telephone industry involves the use of a BN to help resolve problems in Nokia networks [2].

An example of how BNs can be used for more traditional system reliability modelling is given by Langseth and Portinale [15]. As well as discussing the logical basis of the approach and practical aspects of model building, they provide an example of fault-tree-like analysis and show how features such as multi-state variables, parameter uncertainty and components sharing a common environment can be conveniently handled.

A more detailed description of how a fault tree can be translated into a BN, together with a discussion of the attendant advantages, is given by Portinale and Bobbio [18] and by Bobbio et al.[3]. Similarly, Torres-Toledano and Sucar [23] describe how reliability block diagrams can be represented as BNs. One advantage of the BN representation is the ease with which the traditional fault tree's deterministic AND and OR gates are generalized to probabilistic relationships. Furthermore, while this permits the representation of relatively complex relationships, there are also intermediate, more parsimonious options available to the BN modeller, such as the use of so-called Noisy-OR and Noisy-AND gates. In a Noisy-OR gate, for example, the presence of at least one of the lower-level causes does not automatically lead to the presence of the associated higher-level effect. It is instead assumed that each cause can be inhibited with some probability and that the inhibition mechanism of each cause is independent.

Bouissou and Pourret [5] show how a BN can be used to assess the capacity of a production system represented as a series-parallel system in which each component has a capacity to process and pass on a given fraction of the total system capacity and where some components might initially be in a cold standby state. They demonstrate the use of the system for troubleshooting as well as for system evaluation.

3. DYNAMIC BAYESIAN NETWORKS

Whereas an ordinary BN is a static model, representing a joint probability distribution at a fixed point or interval of time, a DBN can represent the evolution of a system over time. In particular, it permits variables to be represented at multiple points in time within the same network structure.

There are a number of ways of representing the passage of time within BNs but the most popular to date is the method proposed by Dean and Kanazawa [9]. In this approach, time is modelled discretely as in a discrete Markov chain. Each variable, X_t, has a time index subscript to indicate which time slice it belongs to. As well as containing the static or within-slice dependencies which ordinary BNs represent, additional temporal dependencies are represented in DBNs by arcs between the time slices. In many cases, it is only necessary to consider first order time dependencies in which case a two-slice network is sufficient to display all of the relationships. However, the actual number of slices over which inference is performed depends, of course, on the problem situation. In essence, the compact representation can be regarded as being 'unrolled' over the number of time slices needed to solve the problem at hand. An example is provided in Section 5.

DBNs provide a general probabilistic graphical modelling framework which encompasses many well-known special cases such as Hidden Markov Models and Kalman filters. The advantage of DBNs is the modelling flexibility they provide and the efficient computational mechanisms which they offer. However, alternative methods of temporal modelling with BNs have also been proposed, including an event-based approach by Arroyo-Figueroa and Sucar [1]. This approach has been applied to reliability modelling by Boudali and Dugan [4].

4. APPLICATIONS OF DBNs TO PROGNOSTIC MODELLING

Roemer et al [20] identify five main approaches to prognostic modelling dependent on the nature of the information available to conduct the prognosis. These are experience-based, evolutionary, feature progression and AI-based, state estimator and physics-based. The first of these is closest to traditional reliability estimation, perhaps making predictions based on a Weibull distribution and only making amendments when new failure data become available. Data-driven approaches such as neural networks have proven to be effective in processing certain data types such as vibrational data, usually after the raw data has been pre-processed to extract a number of relevant features. Radial basis function networks and probabilistic neural networks appear to be particularly popular for this task. Unlike neural networks, DBNs can compensate for limited data availability by making more use of expert knowledge.

Model-based reasoning along with knowledge of past failure rates and other relevant characteristics is increasingly being used alongside condition monitoring information, either sampled periodically or measured continuously as part of a sensor suite or health monitoring system.

For machinery containing moving parts, vibration analysis is a popular prognostic indicator. So too is particle analysis from lubricants, although there can be a significant time delay in obtaining results from laboratory analysis. Acoustic signals can be used to check structural integrity in ships and aircraft. The combination of such disparate information is not always straightforward and is one reason why prognostic modelling has attracted the interest of data fusion practitioners.

Just as Bayesian network approaches have been becoming increasingly popular in data and information fusion applications, so we believe that they may also be well suited to prognostic modelling. This is precisely because their framework naturally permits the combination of disparate information streams, including historical data, expert knowledge and opinion, and measured, noisy observations. Interestingly, most applications of BNs, including DBNs, to prognosis to date appear to be in the medical domain [11]. However, this is perhaps not surprising as the medical domain was also one of the first to prove the value of BNs in diagnostic reasoning. After this, industrial applications grew quickly in number and scope, and we speculate that the same pattern will be repeated with prognostics.

Muller et al. [16] describe a prognostic maintenance model employing a DBN with a manufacturing example involving metal bobbins. Dynamic variables are used to track a number of degradation mechanisms and the impact of different maintenance policies can then be evaluated.

Weber and Jouffe [25] advocate an object oriented approach to DBN modelling in the case of complex systems. They make use of SADT to obtain a functional decomposition of the system, FMEA to identify failure modes of the various functions and construct the DBN based on this decomposition. They provide an example concerning an immersion water heater, showing how the various functional and system reliabilities vary with time. The effect of a maintenance intervention on one of the components at a fixed time is incorporated within this example. They do not consider condition-based maintenance, however

Much of the current research in maintenance modelling is concerned with condition-based monitoring and much of this was influenced by Christer's work on delay time models [7,8]. Wang [24] proposes the use of Hidden Markov Models (HMMs) as part of a two-stage prognostic modelling process in which the length of the second stage corresponds to the delay time. HMMs represent the evolution of a system in which the underlying condition of the system follows a Markov chain but the true state variables are hidden and only imperfect, noisy measurements of them are observable. In fact, HMMs are a special case of DBNs and so the latter offer a potentially richer representation.

Work by Boudali and Dugan [4] has sought to apply timed BNs to system level reliability estimation. The term 'timed BN' is used to make a distinction from DBN since their approach is different but still involves a temporal representation. They demonstrate how their approach can be used to convert dynamic fault trees (DFTs) into timed BNs and discuss the advantages this brings. Each basic component and gate of the DFT is associated with a node in the timed BN. The state space of each such node represents a set of mutually exclusive and exhaustive discrete time intervals during which the component or gate will fail. Of most interest is the probability of the system level gate failing in the final time interval which is usually taken to be the interval from the mission time of interest until infinity.

A discussion of possible metrics related to prognosis is given by Roemer et al. [19]. This is aimed at providing standard metrics to compare alternative prognostic algorithms. It relates to validation and verification efforts applied to PHM tools being developed for the Joint Strike Fighter. Kacprzynski et al. [12] also discuss metrics for prognostics, primarily focusing on the role which probabilistic sensitivity analysis can play in this.

Extending the FMECA to consider aspects related to prognostics is an idea proposed by Kacprzynski et al. [13]. They suggest that the enhanced FMECA could, for example, identify what environmental parameters might influence the item under study, what sensors are capable of capturing this information and where they should be placed.

5. EXAMPLE

The DBN shown in Figure 2 is concerned with an equipment for which two condition monitoring indicators are available, CM1 and CM2. These provide imperfect indicators of the equipment’s True Condition. True Condition is represented here as having six discrete states – Good, Wear 1, Wear 2, Wear 3, Failure Mode 1 and Failure Mode 2. In the first four states, the equipment is functioning but with increasing levels of wear and in the final two states, it is failed but in different ways.
The True Condition in the next time slice is dependent on the True Condition in the current time slice and on the Load exerted on the equipment during the current time slice. Here, this is simply represented by two states – Normal and Abnormal. The presence of an abnormal load increases the transition probabilities associated with greater wear and failure. Over time and in the absence of any maintenance, the equipment gradually progresses through all of the wear states unless it enters one of the absorbing states associated with failure. In this example, Failure Mode 1 can be entered from the states Wear 1, Wear 2 and Wear 3 (with increasing likelihood) while Failure Mode 2 can only be entered from the state Wear 3.

CM1 is highly correlated with True Condition and provides a good indication of the level of wear suffered by the equipment. CM2 is intended to represent the severity of vibration exhibited by the equipment and has three states – Low, Medium and High. Here, this is more weakly correlated with wear than the CM1 indicator is. However, the vibrational state for Failure Mode 1 is most likely to be Medium while the vibrational state for Failure Mode 2 is most likely to be High.

The arcs in Figure 2 identify direct probabilistic dependencies between the variables in the system. When there is no number on an arc, the relationship is within the same time slice. However, a number appearing on an arc indicates a dependence across time slices and the number itself denotes the order of the dependence. Hence, the Loads exerted on the equipment at time slices t and t+1 are probabilistically dependent, as we might expect. In addition, a maintenance action carried out in time slice t influences the True Condition of the equipment at t+1. There are first and second order dependencies shown for the variable True Condition on itself, i.e. True Condition at t is influenced by True Condition at t-1 and t-2. The reason for this is the nature of the maintenance actions which we permit and model later. The allowable actions are None, Reset and Replace. Taking action None has no effect on True

Condition. Taking action Reset will get the equipment out of Failure Mode 1 and return it to the wear state that it entered that failed state from – hence the need for the second order dependence. With only a first order dependence we would have to assume that the equipment was always returned to a fixed condition, possibly Good or Wear 1. Action Replace, however, does always return the equipment to the state Good. An extract from the CPT for True Condition is shown in Table 1.

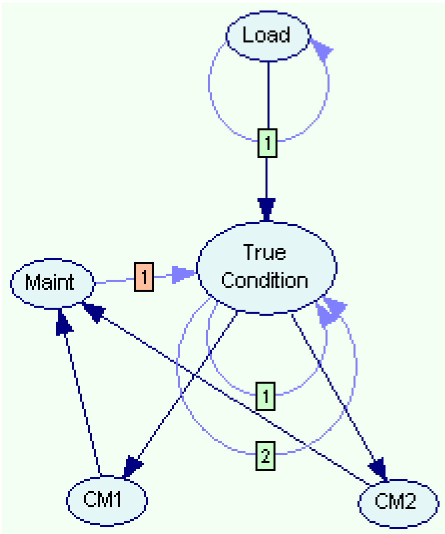

Figure 2. The DBN for an equipment with two condition monitoring indicators.

We have implemented the above model in the BN modelling software Genie*[10]. The results from 1,000 time slices are displayed in Figure 4. This shows how the distribution of True Condition changes over time when no maintenance is performed. At t=111, the reliability drops below 0.9 and by t = 277, it has dropped below 0.5.

Table 1. Extract from the CPT for the True Condition variable.

Load (t)	Normal		
Self (t-1)	Wear 2		
Self (t-2)	Wear 2		
Maint(t-1)	None	Reset	Replace
Good	0	0	1
Wear 1	0	0	0
Wear 2	0.988	1	0
Wear 3	0.01	0	0
FM 1	0.002	0	0
FM 2	0	0	0

Next, we permit condition-based maintenance activities to either reset the equipment (returning it to its previous wear state before it entered Failure Mode 1) or replace the equipment which is the only way to recover from Failure Mode 2 (therefore returning True Condition to the state Good). Here, maintenance is modelled as a deterministic variable, effectively a fixed policy based only on the states of its parent nodes, CM1 and CM2 which provide condition-based information on the equipment.

*The models described in this paper were created using the GeNIe modeling environment developed by the Decision Systems Laboratory of the University of Pittsburgh (http://dsl.sis.pitt.edu).

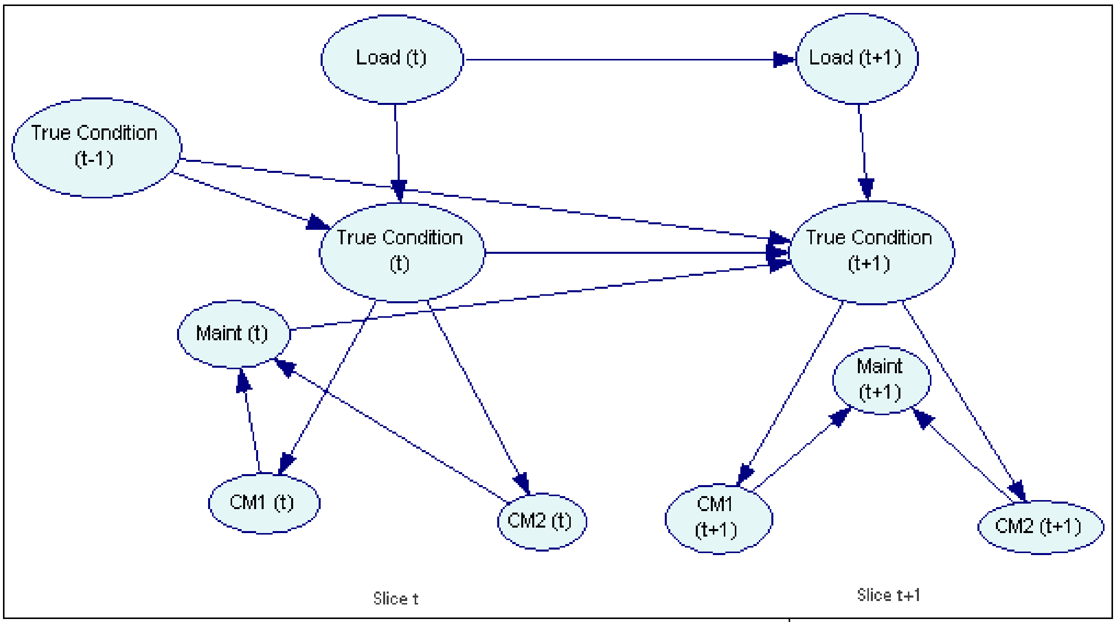

Figure 3. DBN unrolled over two full time slices and also showing the 2nd order dependence of True Condition on itself.

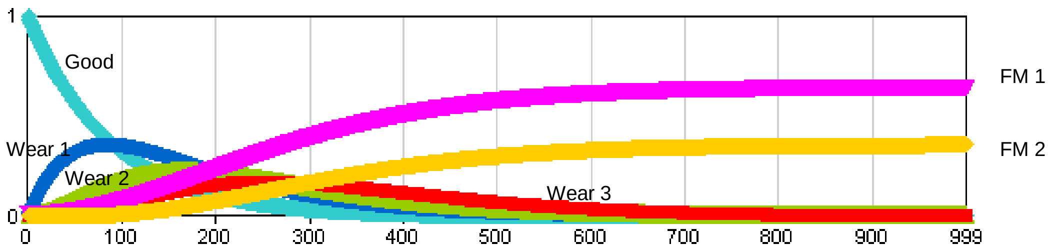

Figure 4. Probability distribution of True Condition over time given no maintenance.

The policy adopted for this example is as follows. If CM1 is Good or Wear 1 and CM2 reports High Vibration, then the equipment is Reset. If CM1 is Wear 2 and CM2 reports Medium or High Vibration, then the equipment is Reset. Finally, if CM1 is Wear 3, the equipment is Reset if CM2 reports Medium Vibration but is Replaced if CM2 reports High Vibration. This is captured in the CPT for the maintenance node.

Figure 5 shows the results associated with the variable True Condition for this case. We can see that the system approaches a steady state and that the probabilities of being in either failed state are very low, particularly for the second failure mode.

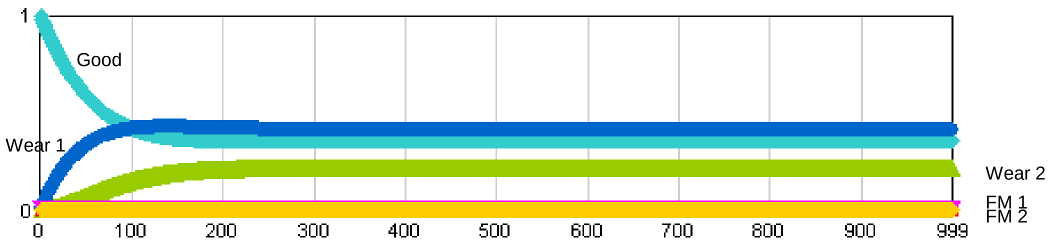

Figure 5. Probability distribution of True Condition over time with maintenance.

Note that it would also be possible to model maintenance as a chance variable, to represent, for example, situations where given the same information, different maintenance engineers might act differently. Yet another possibility is to extend our DBN representation to that of a dynamic decision network and then to model maintenance as a decision. This will be explored in future work.

6. DISCUSSION AND CONCLUSIONS

It is interesting to observe that the objectives associated with reliability estimation and prognostic modelling seem virtually identical – to estimate the probability distribution of time until failure. (In this context, failure might also equate to reaching a threshold level of degradation at which some form of maintenance is required.) Given this, there seems to be less overlap between these topics in the literature than might be expected. Perhaps this is attributable to the fact that recent interest in prognostics has mainly come from those with a background in condition-based monitoring, maintenance, sensors and data fusion rather than in traditional reliability engineering. Another difference may be that prognostics is currently more associated with component level predictions while reliability engineering is traditionally more associated with system level predictions. Nonetheless, prognostics would seem to be an area which promises to bring these various communities together.

Increasing pressure to reduce maintenance costs and to improve equipment availability is likely to keep prognostic modelling an active research area over the coming years. DBNs have a useful role to play in this, particularly as they offer a unifying framework which generalizes a number of seemingly distinct approaches, including HMMs and Kalman filters. Specific research areas related to applying DBNs to prognostic modelling which require further work include learning models from databases, handling missing data and the use of approximations such as particle filtering.

7. REFERENCES

[1] Arroyo-Figueroa G and Sucar LE (1999). A temporal bayesian network for diagnosis and prediction. In: Laskey KB and Prade H (eds), Proc. 15th Conference on Uncertainty in Artificial Intelligence, Stockholm, 1999, 13-20.
[2] Barco R, Wille V and Diez L (2005). System for automated diagnosis in cellular networks based on performance indicators. European Transactions on Telecommunications 16, 399-409.
[3] Bobbio A, Portinale L, Minichino M and Ciancamerla E (2001). Improving the analysis of dependable systems by mapping fault trees into bayesian networks. Reliability Engineering and System Safety 71(3): 249-260.
[4] Boudali H and Dugan JB (2005). A discrete-time Bayesian network reliability modelling and analysis framework. Reliability Engineering and System Safety 87, 337-349.
[5] Bouissou M and Pourret O (2003). A Bayesian belief network based method for performance evaluation and troubleshooting of multistate systems. International Journal of Reliability, Quality and Safety Engineering 10(4), 407-416.
[6] Chan, A and McNaught, KR (2008). Using bayesian networks to improve fault diagnosis during manufacturing tests of mobile telephone infrastructure. Journal of the Operational Research Society 59, 423-430.
[7] Christer AH (1976). Innovative decision making. In: Brown KC and White DJ (eds), Proc. NATO Conference on Role and Effectiveness of Theory of Decision

Practice, pp 368-377.
[8] Christer AH (1999). Developments in delay time analysis for modelling plant maintenance. Journal of the Operational Research Society 50, 1120-1137.
[9] Dean T and Kanazawa K (1989). A model for reasoning about persistence and causation. *Computational Intelligence* 5: 142-150.
[10] Genie. Decision Systems Laboratory, University of Pittsburgh. http://genie.sis.pitt.edu/
[11] van Gerven MAJ, Taal BG and Lucas PJF (2008). Dynamic Bayesian networks as prognostic models for clinical patient management. Journal of Biomedical Informatics 41, 515-529.
[12] Kacprzynski GJ, Liberson A, Palladino A, Roemer MJ, Hess AJ and Begin M (2004). Metrics and development tools for prognostic algorithms. . Proc. IEEE Aerospace Conference, 3809-3815.
[13] Kacprzynski GJ, Roemer MJ, Hess AJ and Bladen KR (2001). Extending FMECA – health management design optimization for aerospace applications. Proc. IEEE Aerospace Conference Vol. 6, 3105-3111.
[14] Kothamasu R, Huang SH and VerDuin WH (2006). System health monitoring and prognostics – a review of current paradigms and practices. International Journal of Advanced Manufacturing Technology 28, 1012-1024.
[15] Langseth H and Portinale L (2007). Bayesian networks in reliability. Reliability Engineering and System Safety 92, 92-108.
[16] Muller A, Suhner M-C and Iung B (2008). Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system. Reliability Engineering and System Safety 93, 234-253.
[17] Pearl J (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA.
[18] Portinale L and Bobbio A (1999). Bayesian networks for dependability analysis: an application to digital control reliability. In: Laskey KB and Prade H(eds), Proc. 15th Conference on Uncertainty in Artificial Intelligence, Stockholm, 1999.
[19] Roemer MJ, Dzakowic J, Orsagh RF and Byington CS (2005). Validation and verification of prognostic and health management technologies. Proc. IEEE Aerospace Conference, 3941-3947.
[20] Roemer MJ, Emmanuel ON and Bloor G (2001). Development of diagnostic and prognostic technologies for aerospace health management applications. Proc. IEEE Aerospace Conference Vol. 6, 3139-3147.
[21] Romessis C and Mathioudakis K (2006). Bayesian network approach for gas path fault diagnosis. Journal of Engineering for Gas Turbines and Power Transactions of the ASME 128, 64-72.
[22] Sanseverino M and Cascio F (1997). Model-based diagnosis for automotive repair. IEEE Expert 12(6), 33-37.
[23] Torres-Toledano JG and Sucar LE (1998). Bayesian networks for reliability analysis of complex systems. Lecture Notes in Artificial Intelligence 1484, 195-206.
[24] Wang W (2007). A two-stage prognosis model in condition based maintenance. European Journal of Operational Research 182, 1177-1187.
[25] Weber P and Jouffe L (2006). Complex system reliability modelling with dynamic object oriented Bayesian networks (DOOBN). Reliability Engineering and System Safety 91, 149-162.

Prognostic Modelling with Dynamic Bayesian Networks

McNaught, K.

2009-11-04T00:00:00Z