

**2017 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY
SYMPOSIUM
SYSTEMS ENGINEERING (SE) TECHNICAL SESSION
AUGUST 8-10, 2017 - NOVI, MICHIGAN**

**UTILIZING BAYESIAN NETWORKS TO DEVELOP REAL TIME
PROGNOSTIC MODELS FOR GROUND VEHICLES**

Marc Banghart, PhD

Lead Engineer
KBRwyle
Orange Park, FL

David Nelson

Chief Engineer
KBRwyle
Orange Park, FL

Adam Brennan

U.S. Army Tank Automotive
Research Development and
Engineering Center (TARDEC)
Warren, MI

***Disclaimer:** Reference herein to any specific commercial company, product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the Department of the Army (DoA). The opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or DoA, and shall not be used for advertising or product endorsement purposes.*

ABSTRACT

Bayesian networks have been applied to many different domains in order to perform prognostics, reduce risk and ultimately improve decision making. However, these methods have not been applied to military ground vehicle field data sets. The primary objective of this study is to illustrate how Bayesian networks can be applied to a ground vehicle data set in order to predict potential downtime. The study generated a representative field data set, along with tabu search, in order to learn the network structure followed by quantification of link probabilities. The method is illustrated in a case study and future work is described in order to integrate the method into a real-time monitoring system. The study yielded a highly accurate prediction algorithm that can improve decision making, reduce downtime and more efficiently manage resources in the ground vehicle community.

INTRODUCTION

The current and anticipated acquisition climate, along with decreasing budgets are resulting in military ground vehicles remaining in service longer than originally anticipated. These systems must continue to operate safely, effectively and in a cost-effective manner. Thus, these systems must remain supportable over the extended life cycle. An important component of supportability analysis includes identification and anticipation of degraders to readiness. These degraders include high failure items, long lead times (both in terms of

repair and logistics) as well as potential manpower constraints.

Predictive analytics and prognostic models provide powerful decision capability to the Army and Marine Corps. Understanding and anticipating degraders to readiness can be seen as a risk assessment and mitigation activity. A plethora of risk assessment methodologies have been proposed in literature. However, several of these approaches, such as event trees, require a clear understanding of the chain of events (or causal connections) leading to a high risk event. As many accidents have shown, identification and quantification of these

connections (and associated probabilities) is not trivial and is wrought with error and human subjectivity. Equally important is controlling the false alarm rate and manpower requirements of employed methods. Predictive tools with high false positive rates may still have value in certain instances (for example when mitigating high consequence failure modes), however typically should be controlled in order to retain user confidence. Other qualitative methods such as Reliability Centered Maintenance (RCM) typically require teams with sufficient domain expertise sifting through large amounts of data. The teams' primary goal is to identifying both the frequency and severity of failure modes, along with strategies to either prevent failures or mitigate the consequences. This poses a resource challenge which will only be exasperated as the proliferation of Big Data continues. These methods likely also include significant bias, subjectivity and frequently utilize simplistic metrics such as Mean Time Between Failure (MTBF). Probability distributions may also be fit to repair, delay or failure times – however extensive data manipulation is still required. Metrics may also be lagging indicators, which reduce the utility of the model or analysis.

The term machine learning was originally coined in 1959 by Arthur Samuel, predicated on the notion that computers can have the ability to learn without requiring explicit programming. Thus, computers can utilize input data sets and identify patterns or predict outcomes. This field has been applied to a vast array of domains to include radar systems, image processing and signal detection.

Machine learning methods have several benefits to include the capability to learn classification rules automatically, handle large amounts of data in real time, and allow integration of quantitative and qualitative variables within a single model. Additionally, once the model is validated and verified, new data sets can be provided as inputs allowing statistical inference or prediction to be performed.

Bayesian networks are a subclass of probability-based learning methods. Probability-based methods describe causality between features along with associated probabilities. The overall concept is provided in Figure 1. First, a predictive model must be developed which consists of utilizing a training set to first learn the associated Bayesian Network structure, followed by quantification of the various conditional probabilities within the network. The training data set consists of descriptive features and a target feature. The features can both be numerical and categorical. Sensitivity analysis is performed in order to further validate the model during this phase. Once a validated prediction model is established, evidence can be set within the model based on real-time updates.

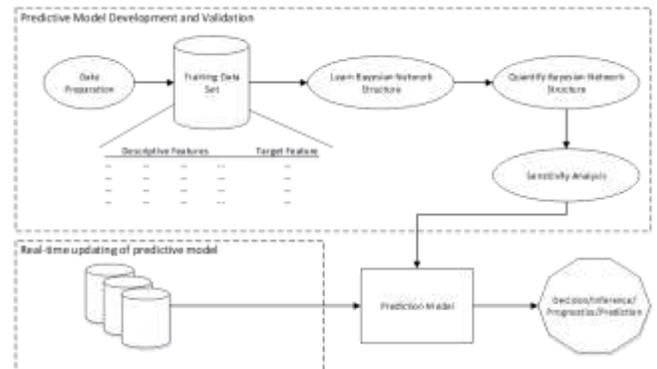


Figure 1: Model Development and Updating Process

PREDICTION AND CAUSAL EXPLANATION

Models are utilized both for causal explanation and prediction. Additionally, in many applications the assumption is made that a model that has high explanatory power is predictive in nature. Explanatory modeling allows “testing causal hypotheses about theoretical constructs” [1], while predictive models are focused on application of statistical or data mining methods to predict new or future observations.

Prediction has gained momentum both in the academic and practical communities. According to

Shmueli [1] prediction has several key attributes to include: (1) large and rich datasets contain complex patterns and relationships that cannot be hypothesized easily, (2) predictive modeling closes the gap between theory and practice, (3) predictive power can be assessed and benchmarking can be employed.

According to Shmueli [1] care must be taken to distinguish between explanatory and predictive models, since the type of uncertainty associated with each differs. The disparity stems from several fundamental philosophical differences between these techniques. First, in explanatory modeling we aim to identify variables that result in a state change of another variable. Exploratory models further aim to develop a function, f that tests an already known set of hypotheses. Predictive modeling does not require causality, but rather considers the association between X and Y . Additionally, prediction develop f with the goal of predicting unknown observations, versus minimizing bias, or the error between predicted and actual observations alone. All models include error, which does not necessarily reduce predictive power. Thus, as stated by Shmueli [1], “wrong” models may sometimes have greater predictive power than correct models.

FIELD DATA UNCERTAINTY

Field data typically contains significant error both in commercial and military applications [2]. Additionally, qualitative estimates within risk management tools such as Failure Modes Effects Analysis (FMEA) has also been shown to contain error [3, 4].

Conducting sensitivity analysis is critical to both assess the potential impact of input errors along with which variables are most important within a constructed model in terms of output metrics. In the case of Bayesian Networks we are interested in how errors within the training data set may influence the constructed model. More specifically, we are interested in how output

decisions change when noise (or error) is purposefully introduced within the training data set.

Previous research by Banghart et al. [5] illustrated that Bayesian Networks are resistant to noise. They demonstrated model credibility by investigating model response utilizing a Design of Experiments (DOE) study. The research indicated that Bayesian networks appear to be robust against noise – however not for all target features. In some high noise cases, the results were drastically impacted. However, the results indicated under low levels of noise the impact was minimal. These results are important and provide credence to utilization of Bayesian Networks in real field data – which will always contain noise or error that is not easily quantified. The researchers are not proposing that “garbage noisy data” can construct Bayesian Networks with high predictive power, or that due diligence should not be performed in terms of design of measurement systems. However, Bayesian Networks may be appropriate even in noisy data sets, where the error cannot be easily quantified – assuming a robust sensitivity analysis is performed [5].

RESEARCH METHOD

Bayesian networks can be defined utilizing several methods. Some of the major methods found within the literature include heuristic search and utilization of expert opinion. Heuristic “score-and-search” techniques dominate the literature and can be used to learn both the underlying topology of the network as well as the link probability, or conditional probability tables (CPTs). In a broad sense these algorithms consider a search space which contains all the feasible solutions or states of the problem, utilize a mechanism to both encode these states and move from state to state within the search space. Finally, a scoring function is utilized to assign a score to a state within the search space. Heuristic methods include K2, genetic algorithms, simulated annealing and tabu search [6]. Based on previous research conducted by Banghart et al. tabu

search has been illustrated as a viable method to predict readiness on Navy aircraft [7].

Tabu Search

The tabu search (TS) algorithm is deterministic in its basic form. TS works on the principle of identifying improved solutions within a specific neighborhood through local search. The algorithm utilizes a tabu list in order to allow “worse” moves. If the move has not been executed before (and thus on the tabu list), the algorithm may evaluate it as a feasible solution. This allows the algorithm to be less susceptible to local optimum traps [8].

Other methods such as the hill climber algorithm can have difficulty in certain solution topographies. For example, if the solution space is fairly “flat”, the hill climber algorithm wouldn’t advance beyond the first local optimum found. The tabu list allows the TS algorithm to escape these local optimums and are thus well suited to complex solution spaces.

TS is employed in order to identify the optimum Bayesian network structure. We define the structure by adding or removing arcs within the network and evaluating an objective function.

When constructing a Bayesian network, we aim to find an optimal network structure. Let S represent a set of moves that lead from one solution to another. These moves can consist of adding, removing or reversing arcs within the Bayesian Network [9]. The complete pseudocode is provided below for reference.

Begin

$t \leftarrow 0$;

Initialize tabu search;

While ($t < t_{max}$) **do**

$t \leftarrow t + 1$;

Search the neighborhood;

Evaluate candidate solutions;

Update tabu list;

End

End

Network quality can be evaluated utilizing several difference score metrics. The reader is referred to Yang and Chang for a detailed discussion of the available metrics [10].

One benefit of tabu search is that it typically proceeds very aggressively (when compared to other techniques) to a local optimum. This is in part because tabu search spends more time/effort in areas of the solution space where solutions are better [11]. Additionally, tabu search can be stopped at any time once a feasible solution is found [12]. Several of these advantages are related to the utilization of memory structures allowing the solution space to be searched more economically. One drawback of tabu search is that the length of the tabu list may not be trivial to determine. Specifically, the length of this list allows a trade-off between the computational burden and algorithm efficiency.

Estimating Link Probability

Once the structure of the network has been defined the CPTs can be calculated. A simple estimator can be used to compute the relative frequencies in the training data set.

Given a training data set of and an associated Bayesian Network we wish to estimate the CPTs for each node (Figure 2).

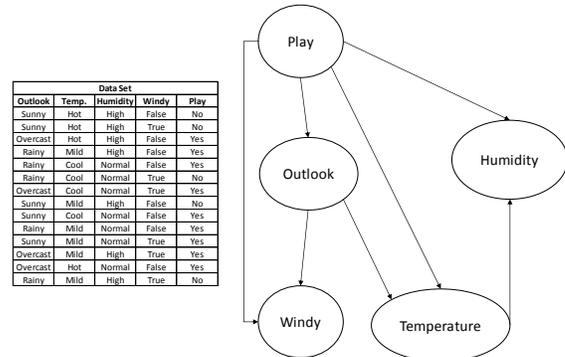


Figure 2: Example Data Set and Model for CPT calculation

Utilizing the temperature node as an example, we note that both play and outlook are connected to

temperature. The conditional probability of $P(\text{temperature}=\text{hot}|\text{play}=\text{yes},\text{outlook}=\text{sunny})$ can be calculated as follows.

First, note in the original data set there are no instances where the temperature=hot, outlook=sunny and play=yes. There were two instances of play=yes and outlook=sunny. In order to avoid the problem of no training samples for a specific class we apply the Laplacian correction. The correction assumes that our training data sample is large and is commonly applied in machine learning [13]. Thus, we calculate the probability:

$$P(\text{temperature} = \text{hot}|\text{play} = \text{yes}, \text{outlook} = \text{sunny}) \\ = (0 + 1)/(2 + 3) = 1/5 = 0.2$$

We repeat this process in order to build the complete conditional probability tables for each feature.

CASE STUDY

In order to illustrate our approach we developed a generic ground vehicle case study. Consider a ground vehicle data set that includes failures from three components (power pack, electronic module and software). Each component exhibits three failure modes with random failure times. Although the data set was generated for this analysis, the approach has been illustrated on actual field data of military systems [7].

Data Set Description

Failure times were generated assuming Weibull distributions with varying slope parameters. Delay and repair times were generated assuming a lognormal distribution. Work centers were assigned based on the respective component. Failure of components were considered equally likely. Finally, failure modes and cost were assigned based on the scheme defined in Table 1.

Table 1: Input Parameters for Derived Data Set

Component	Failure Mode	Failure Mode Ratio	Assumed Cost	Assumed Mean Repair Time
Electronic Module	A	0.5	HIGH	2 hours
	B	0.4	MEDIUM	
	C	0.1	LOW	
Engine	D	0.3	MEDIUM	6 hours
	E	0.3	HIGH	
	F	0.4	MEDIUM	
Software	G	0.9	LOW	3 hours
	H	0.05	LOW	
	I	0.05	HIGH	

In order to visualize the variability and the underlying distributions for each quantitative variable boxplots are provided in Figure 3 by component. Non Mission Capable (NMC) hour values were calculated by the summation of repair time, logistics delay, administrative delay and maintenance delay time. Repair times were typically short, however very long repair times were observed in the data set. The quantitative variables were converted to a nominal scale of LOW, MEDIUM, HIGH and VERY HIGH. The scale was determined by splitting each variable into four equal proportions based on the Interquartile Range.

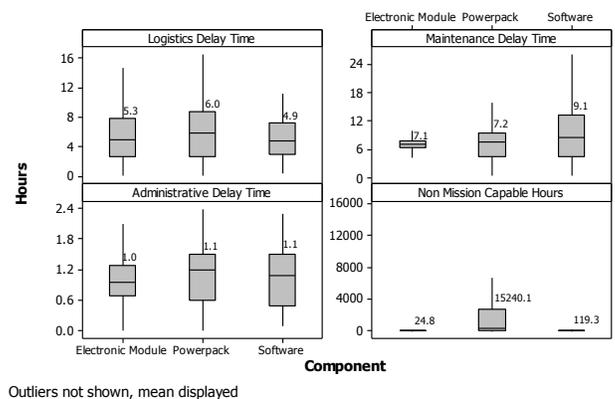


Figure 3: Quantitative Feature Distributions

Thus, the processed data set consisted of 10 features and a sample size of 464 instances.

Visualization of the potential relationships between features is important in predictive modeling. Specifically, if a strong relationship can be observed between variables complex machine learning models may not be needed and the analysis can utilize a regression based model instead. Additionally, features may be redundant if a strong relationship exists and thus be excluded from the predictive model. There did not appear to be a significant linear relationship between the descriptive features.

Results and Discussion

Tabu search yielded a Bayesian Network provided in Figure 4. The network indicated several interesting probabilistic relationships between NMC hours and repair time, failure time, logistics delay, administrative delay as well as the work center. Additionally, relationships were observed between the component and failure mode.

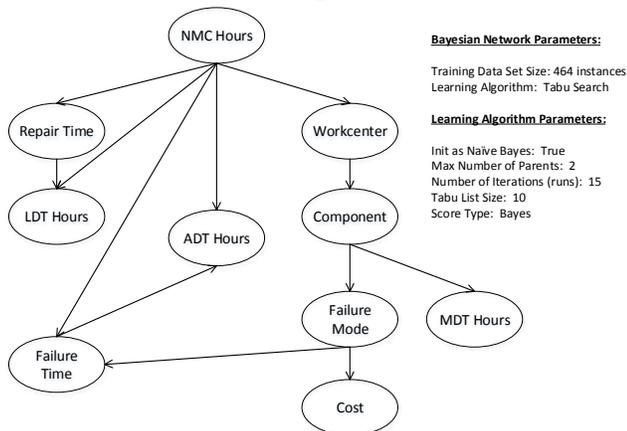


Figure 4: Bayesian Network for a Generic Ground Vehicle

The results indicated that 86.4 percent of instances were correctly classified with a kappa statistics of 0.82 when the target feature was set to NMC hours. The kappa statistics provides a measure indicating the statistical significant of the algorithm. A high kappa indicates the algorithm results are not due to chance alone. The kappa statistic can range from 0 to 1, with a statistic greater than 0.75 considered excellent. The calculated kappa supported that the algorithm had

significant predictive power. True Positive (TP) and False Positive (FP) rates were also calculated. The weighted average true positive rate was 86.4 percent, and associated false positive rate was 4.5 percent. These results further indicated the power of the algorithm when predicting NMC hours.

The target feature was also set to failure mode, since understanding the likelihood of a specific failure mode occurring significantly benefits the decision maker. Specifically, resources can be proactively provided in terms of both manpower, spares and repair supplies in order to reduce the NMC impact. The algorithm achieved an 87.9 percent correctly classified instances with a kappa statistic of 0.85 when the target feature was set to Failure Mode. TP and FP rates were 87.9 and 1.7 percent respectively.

Probability of High NMC Hours

In order to illustrate the practical application of the developed Bayesian Network, consider calculation of the joint probability of NMC hours being either VERY HIGH or LOW given various states of other variables. For example, consider the case of the electronic module. Let's assume that due to several supply chain problems we expect long delays in acquiring this component. We can assess the impact on NMC Hours by setting the states of ADT, LDT and MDT to VERY HIGH and observing the joint probability that NMC hours will be VERY HIGH. We obtain the following joint probabilities:

- NMC hours MEDIUM 62.2 percent
- NMC hours HIGH 23.2 percent
- NMC hours VERY HIGH 5.3 percent

We can consider similar scenarios for other feature states. For example considering the same scenario for power pack we calculate the following joint probabilities:

- NMC hours MEDIUM 12.3 percent
- NMC hours HIGH 17.5 percent
- NMC hours VERY HIGH 68.7 percent

Thus, we conclude that supply chain delays relative to the power pack will impact NMC hours to a much greater extent than the electronic module.

Incorporation of Real-Time Evidence

As illustrated in the previous example the developed Bayesian Network can be utilized to perform predictions, or what-if scenarios. The main output of interest is how the calculated joint probability changes as evidence is set.

In order to incorporate the method into a real time monitoring system two additional challenges must be overcome.

First, an architecture must be designed that allows the decision maker to easily perform aforementioned trade studies. Business rules must be developed and tailored to each program to help facilitate decision making. Additionally, the architecture should automate calculation of the joint probability given evidence and alert the decision maker as appropriate.

Second, mechanisms to set evidence from updated field data must be developed to include consideration of strategies to ensure the Bayesian Network remains validated.

CONCLUSION AND FUTURE WORK

The capability to accurately predict military readiness in complex engineering systems provides an important decision tool. Additionally, quantification of the performance parameters of such a tool, to include false positive and true positive rates, is critical to ensure credibility. Development of these predictive, or prognostic tools is challenging. Two broad categories have been utilized. The first method utilizes system design knowledge in order to understand system operation, define features to measure and apply some type of measurement device. A second method utilizes data already collected, applies advanced algorithms and attempts to predict an outcome based on a known training data set. The research performed utilized machine learning

algorithms (such as Bayesian Networks) and a data set representative of military field data. The research yielded a predictive method that was able to predict the probability of large amounts of downtime given several scenarios. Thus, the analysis provides a framework that can be applied to ground vehicle data sets in order to provide robust decision making tools, reduce risk and improve readiness.

REFERENCES

- [1] G. Shmeuli, "To Explain or to Predict?," *Statistical Science*, pp. 289-310, 2010.
- [2] M. Banghart, C. Comstock and J. Solomon, "Comparison of Military and Commercial Field Data Best Practices for Higher Quality Reliability and Maintainability Data," in *Reliability and Maintainability Symposium*, Orlando, Florida, 2017.
- [3] M. Banghart, L. Bian and K. Babski-Reeves, "Human Induced Variability during Failure Mode Effects Analysis (FMEA)," in *Reliability and Maintainability Symposium*, Tucson, 2016.
- [4] M. Banghart, K. Babski-Reeves, L. Strawderman and L. Bian, "Identification of Human Subjectivity in Failure Modes and Effects Analysis (FMEA)," *International Journal of Quality & Reliability Management*, p. TBD, 2017.
- [5] M. Banghart, A. Tolk and L. Bian, "Assessment of Bayesian Network Credibility under Uncertainty," *Journal of Defense Modeling and Simulation*, p. TBD, 2017.
- [6] L. van der Gaag, S. Renooij and V. Coupe, *Sensitivity Analysis of Probabilistic Networks*, 2007.
- [7] M. Banghart, L. Bian, K. Babski-Reeves and L. Strawderman, "Risk Assessment on the EA-6B Aircraft utilizing Bayesian Networks," *Quality Engineering*, p. TBD, 2017.
- [8] S. Skiena, *The Algorithm Design Manual*, 2nd ed., Berlin: Springer Science and Business Media, 2010.
- [9] R. R. Bouckaert, "Bayesian Network Classifiers in Weka," 2004.
- [10] S. Yang and K. C. Chang, "Comparison of score metrics for Bayesian network learning," *IEEE Transactions on Systems Man and Cybernetics*, vol. 32, no. 3, pp. 419-428, 2002.
- [11] F. Glover, "Tabu Search - Part II," *ORSA Journal on Computing*, vol. 2, no. 1, 1989.

[12] H. Pirim, E. Bayraktar and B. Eksioglu, Tabu Search: A Comparative Study, Vienna, Austria, 2008.

[13] J. Kelleher, B. Namee and A. D'Arcy, Fundamentals of Machine Learning for Predictive Data Analytics, London, England: MIT Press, 2015.