A Bayesian Approach for Estimating Complex System Reliability

Ozge DOGUC
School of Systems and Enterprises, Stevens Institute of Technology
Hoboken, NJ 07030, USA
odoguc@stevens.edu

Jose Emmanuel RAMIREZ-MARQUEZ
School of Systems and Enterprises, Stevens Institute of Technology
Hoboken, NJ 07030, USA
Jose.Ramirez-Marquez@stevens.edu

Abstract. Although a number of recent studies on using BN for system reliability estimation have been proposed, these studies are based on the assumption that a pre-built BN was designed to represent the system. In these studies, the task of building the BN is typically left to a group of specialists who are BN and domain experts. However the process of building a system-specific BN is generally very time consuming and may lead to incorrect deductions. As there are no existing studies to eliminate the need for a human expert in the process of system reliability estimation, this paper introduces a holistic method that uses historical data about the system to be modeled as a BN and provides efficient techniques for automated construction of the BN model and estimation of the system reliability. Moreover, very limited human intervention is sufficient for the process of BN construction and reliability estimation.

Introduction

System reliability can be defined as the probability that a system will perform its intended function during a specified period of time under stated conditions (Gran and Helminen 2001). Traditionally, engineers estimate reliability by understanding how the different components in a system interact for system success. Based on this understanding, typically a graphical model (usually in the form of a fault tree, a reliability block diagram or a network graph) is constructed to represent component interactions. Using the graphical model, different analysis methods such as minimal cut sets, minimal path sets, Boolean truth tables, etc (Coyle, Arno, and Hale 2002; Fenton, Krause, and Neil 2002; Gopal, Kuolung, and Nader 2001) are used to represent system reliability quantitatively. At the end, the reliability characteristics of the components in the system are introduced into the mathematical representation in order to obtain a system level reliability estimate. This approach is valid whenever the system success or failure behavior is well understood. However, for complex systems (i.e., systems with large numbers of components and/or complex component interactions), understanding component interactions, which usually requires intervention of a domain expert, may prove to be a challenging problem.

Bayesian networks (BN) have been proposed as an alternative to traditional reliability estimation approaches (Amasaki et al. 2003; Boudali and Dugan 2006; Gran and Helminen 2001). BN have significant advantages over traditional frameworks, partly because they are easy to use in interaction with domain experts in the reliability field (Sigurdsson, Walls, and Quigley 2001). Current approaches for reliability analysis via a BN (Amasaki et al. 2003; Bobbio et al. 2001; Sigurdsson, Walls, and Quigley 2001) use specialized networks, each of which is designed for a specific system. In these studies, the BN structure to be used for estimating system reliability should be known *a priori*. This assumption presupposes that the BN should be built by an expert who has "adequate" knowledge about the system behavior. However, finding such an expert may not be possible at all times for every system under consideration. Moreover, the number of such experts is limited and finding one is usually difficult and costly (Lagnseth and Portinale 2005). Also, human intervention is always open to unintentional mistakes, which could cause discrepancy in the results. These issues are particularly true in complex systems, where the number of components and interactions are larger and thus, the likelihood of miscalculations can be substantial.

To address these issues, this study introduces a holistic method for estimating system reliability by linking BN construction from raw component and system data, association rule mining and evaluation of conditional probabilities. Based on our literature review, this is the first study that incorporates these methods for estimating system reliability to reduce the need for human intervention. The proposed method automates the process of BN construction by using the *K2* algorithm (a commonly used association rule mining algorithm), which has been proven to be efficient and accurate for finding associations (Cooper and Herskovits 1992) from a dataset of historical data about the system. Moreover, unlike previous approaches, the proposed solution is not system specific, it can be applied to systems following any kind of configuration (two terminal, k-terminal, all terminal, etc…) and behavior (binary, capacitated and multi-state). In essence, our approach can build a BN and estimate reliability for any system when observed system data is available (Doguc and Ramirez-Marquez 2007).

Literature Survey

Estimating systems reliability using BN dates back as early as 1988, when it was first defined by Barlow (Barlow 1988). The concept of BN has been discussed in several earlier studies (Cowell et al. 1999; Jensen 2001; Pearl 1988); the idea of using BN in systems reliability has gained acceptance within the last decade because of the simplicity it allows to represent systems and the efficiency for evaluating component associations. More recently, BN have found applications in software reliability (Fenton, Krause, and Neil 2002; Gran et al. 2000), fault finding systems (Jensen 2001), and general reliability modeling (Bobbio et al. 2001). In recent studies, predefined BN are used for reliability estimation for specific systems. For example, Gran and Helminen (Gran and Helminen 2001) study on building BN for nuclear power plants and introduce a hybrid method for estimating the reliability of the plant. In their study, they considered the nuclear plant as two subsystems; a software system and the plant hardware. Therefore they combined two BN that were being used for corresponding systems: 1) The Halden Project (HRP) (Dahll and Gran 2000) uses a BN for risk assessment based on disparate evidences. 2) The VTT Automation (Helminen 2000) focuses on the reliability of software-based systems using BN. Additionally they discuss another challenge; each BN uses a different modeling and simulation environment.

In another study Helminen and Pulkkinen present a BN-based method for reliability estimation of computer-based motor protection relay (Helminen and Pulkkinen 2003). In their study, Helminen and Pulkkinen assume existence of a BN that models the system and introduce methods for estimating prior probabilities and assessing the system reliability accordingly.

In addition to these, Amasaki *et al.* (Amasaki et al. 2003) use BN for software quality assessment. They modeled the phases of a software system as a BN, and by using this model they simulated the faults that may occur in their system. After this step, they used the actual data and performed sensitivity analysis of the BN model that they constructed. In addition to these, Boudali and Dugan (Boudali and Dugan 2006) introduce a method for reliability assessment in dynamic systems by using temporal BN; where the system components change states at different time intervals. Moreover, Singh *et al.* (Singh et al. 2001) presents their work on reliability estimation in component based systems. They classify the component based system reliability estimation methods into three as state based models, path based models and additive models.

Although all of the studies introduced in this section use BN for reliability estimation, they require human domain experts to evaluate the prior probabilities and understand the structure of the BN. In the next section, we introduce a methodology that automates the process of BN construction and reduces the need for a human expert for system reliability estimation.

Bayesian Networks

As discussed in the previous sections, BN have been used in various studies for estimating system reliability. In this section we first provide definitions of BN and Bayes' theorem. Then we discuss the *K2* algorithm that we used to create BN in this study.

Using Bayesian Networks for System Reliability. One could summarize the BN as an approach that represents the interactions between the variables from a probabilistic perspective. This representation is modeled as a directed acyclic graph, where the nodes represent the variables and the links between each pair of nodes represent the causal relationships between the variables. In general, a fundamental assumption for the construction of a BN is that, the strength of the interaction/influence among the graph nodes is uncertain and thus, this uncertainty is represented by assigning a probability of existence to each of the links between nodes.

From systems engineering perspective, the variables of a BN are defined as the components in the system while the links represent the interactions of the components leading to system "success" or "failure". Under a reliability analysis perspective, a variable *A* in BN constitutes the success of a specific system component and therefore, p(*A*) represents the probability of success for such a component. For non-trivial systems -systems not following a series, parallel or any combination of these configurations- the failure/success probability of a system is usually dependent on the failure/success of a non-evident collection of components. Strictly speaking, the probability of success of a component is conditional on the available evidence from other components. In a BN this dependency is represented as a directed link between two components, forming a *child* and *parent* relationship, so that the dependent component is called as the *child* of the other. Therefore, the success probability of a child node is *conditional* on the success probabilities associated with each of its parents (Fenton, Krause, and Neil 2002). The *conditional probabilities* of the child nodes are calculated by using the Bayes' theorem via the probability values assigned to the parent nodes. Also, absence of a link between any two nodes of a BN indicates that these components do not directly interact for system failure/success thus, they are considered *independent* of each other and their probabilities are calculated separately.

To illustrate these concepts, the BN shown in Figure 1 presents how five components of a system interact. In this BN the child-parent relationships of the components can be observed, where on the quantitative side the *degrees* of these relationships (associations) are expressed as probabilities (Lagnseth and Portinale 2005).

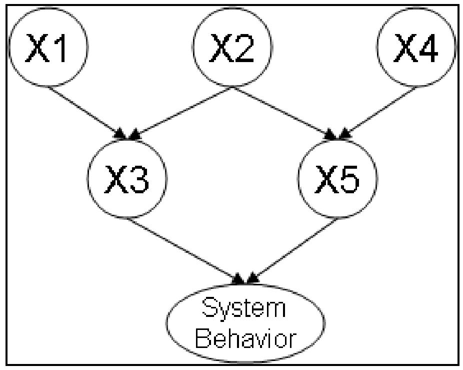

Figure 1. A sample Bayesian network

In Figure 1 the topmost nodes (X_1, X_2 and X_4, representing components 1, 2, and 4 respectively) do not have any incoming edges, therefore they are conditionally independent of the rest of the components in the system. The *prior probabilities* that are assigned to these nodes should be known beforehand -with the help of a domain expert or using historical data about the system. Based on these prior probabilities, the conditional probability table (CPT) that belong to a *dependent* node, such as X_3, can be calculated using Bayes' theorem as illustrated by equation (1):

$$p(X_3 \mid X_1, X_2) = \frac{p(X_1, X_2 \mid X_3)\, p(X_3)}{p(X_1, X_2)} \qquad (1)$$

Equation (1) shows that the probability for the node X_3 is *independent* of nodes other than X_1 and X_2 in the system. Similar to prior probabilities, CPT can be computed by using historical system and component data. However, an important question on how to discover the associations among the system components still remains. As an alternative to using a domain expert for this purpose, an unsupervised BN construction algorithm, *K2* is used in this paper.

The *K2* Algorithm. The *K2* algorithm, for construction of a BN, was first defined by Cooper and Herskovits (Cooper and Herskovits 1992) as a greedy heuristic search method. This algorithm searches for the parent set for a node that has the maximum association with it. The *K2* algorithm is composed of two main factors: a scoring function *f* to quantify the associations and rank the parent sets according to their scores, and a heuristic to reduce the search space to find the parent set with highest degree of association. The *K2* algorithm would need to examine all possible parent sets, i.e. starting from the empty set, it should consider all subsets of set of possible parents without the heuristic. With the help of the heuristic, the *K2* algorithm does not need to consider the whole search space; it starts with the assumption that the node has no parents and adds incrementally that parent whose addition most increases the scoring function. The *K2* algorithm stops adding parents to the node when addition of no single parent can increase the score.

Illustration of Our Methodology

This section provides a step-by-step explanation of the BN construction framework and system reliability estimation method discussed in the previous section. Table 1 presents an example historical dataset that contains observations on the sample system shown in Figure 1 with five components labeled X_1 to X_5. Each row in Table 1 shows the states of the system components at an instance of time t_i; when the observation was done. For the sake of simplicity and without loss of generality in the proposed method, component failure data exhibits binary behavior.

That is, for each component X_i, the value of 0 represents failure while the value of 1 represents full functionality for the corresponding observation. Also, in Table 1, information about the overall *System Behavior* is provided in last column.

Table 1: Dataset for the illustrative example

Observation	X_1	X_2	X_3	X_4	X_5	*System Behavior*
1	1	1	0	0	0	1
2	0	1	1	0	0	0
3	1	0	1	1	1	1
4	0	0	0	0	0	0
5	1	1	1	0	1	1
6	0	1	1	1	0	0
7	1	0	0	1	0	1
8	0	0	1	1	1	1
9	1	1	1	0	0	0
10	0	1	0	1	1	1

Our proposed method uses a dataset such as displayed in Table 1; finds associations between the columns (system components); calculates the degrees of these associations; builds the associated BN and finally uses it to estimate overall system reliability. In the first step of our method the *K2* algorithm starts with the first component in the dataset, X_1. Since X_1 does not have any succeeding components (i.e. possible candidate parents), the *K2* algorithm skips it and picks the second component in the dataset, which is X_2.

For X_2, there are two alternative parent sets: the empty set ϕ, or X_1. Therefore, the *K2* algorithm computes the scoring function *f* for each of these alternative parent sets and compares the results. Then, the set of candidate parents with highest *f* score is chosen as the parent set for X_2. At the end of this iteration the values $\frac{1}{2310}$ and $\frac{1}{3600}$ are calculated and then compared; and the former, representing the score of the empty set $\{\phi\}$, picked as the parent. So the *K2* algorithm decides that X_2 has no parents, which means that there is no association between X_1 and X_2.

In the next iterations of the *K2* algorithm, the number of possible candidate parent sets to be considered and the amount of computations for *f* score calculation increases. Skipping the details, *f* scores of the candidate parent sets for the X_3 component are given in Table 2. Because the *K2* algorithm iterates on the components according to their ordering in dataset, components X_4 and X_5 are not taken into account as candidate parents for X_3. The *K2* algorithm selects the set $\{X_1, X_2\}$ as parent set of X_3, because it has the highest *f* score. The number of computations grows with the order of the component in the system, and when the *K2* algorithm finishes with the last column (*System Behavior* in Table 1), it outputs the BN structure displayed in Figure 1.

Table 2: *f* scores for all possible candidate parent sets for X_3

Parent Set	*f* score
ϕ	$\frac{1}{2310}$
$\{X_1\}$	$\frac{1}{2772}$
$\{X_2\}$	$\frac{1}{3600}$

$\{X_1, X_2\}$	$\frac{1}{288}$

The next step of the proposed method estimates system reliability using the BN that was constructed by the *K2* algorithm. Besides the associations that were discovered by the *K2* algorithm in the previous step, the inference rules should be used to calculate the conditional probabilities between the nodes in the BN. The conditional probabilities are essential in calculating the overall reliability of the system, as they represent the degrees of associations between components of a system. Each component with a non-empty parent set in the network is associated with a CPT. In this step, with the help of CPT and the prior probabilities that X_1 and X_2 have, the success probability value for X_3 can be calculated. According to the BN structure in Figure 1, components X_1 and X_2 are independent of others; therefore their success probabilities can be directly inferred from the observations dataset in Table 1. From Table 1 it can be evaluated that $p(X_1=1)=0.5$ and $p(X_2=1)=0.6$ and the probability of success for component X_3 can be calculated as 0.57 using Bayes' rule provided in Equation (1). Extending the computations for the other components in the network, success probabilities for the rest of the components in the sample system can be evaluated; such that $p(X_4=1)=0.4$ and $p(X_5=1)=0.6$. In the last step, the system reliability can be calculated by using these probability values and the CPT of the "System Behavior" node in the BN structure given in Figure 1. The success probability for the *System Behavior* node is calculated as 0.72 or 72%; which is the reliability of the sample system presented in this section. The proposed method for estimating system reliability using observations dataset is superior to previously defined methods due to its unsupervised nature; almost all steps of the required computations can be carried out without any human intervention.

Experimental Analysis

In this section, experimental analysis on the performance of our proposed method for system reliability estimation is provided. In order to give a better perception of analysis, performances of the two phases of the proposed method (BN construction and reliability estimation) are examined separately. First, performance and correctness of the *K2* BN construction algorithm is analyzed using historical data (obtained via Monte Carlo simulation) for the following BN:

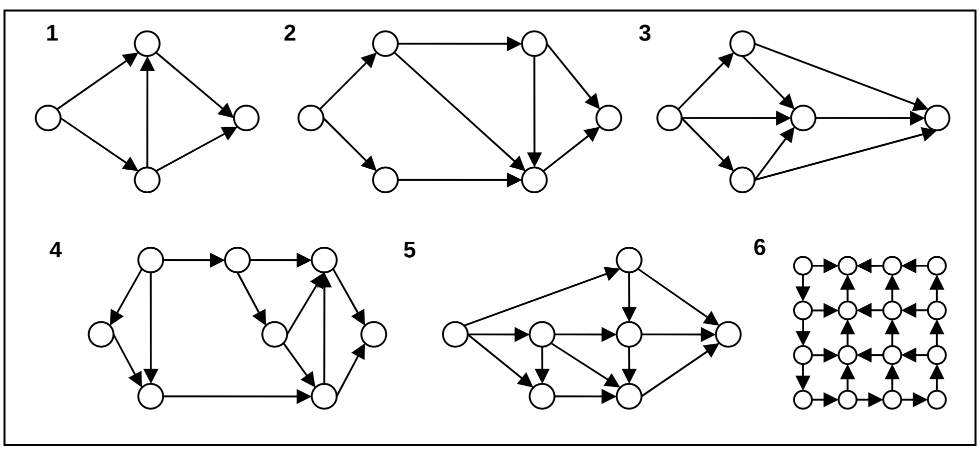

Figure 2. Case BN tested on the *K2* algorithm.

BN displayed in Figure 2 represent different systems with various components. For our experimental analysis, separate data sets -similar to Table 1- are used for each example BN. As it was explained in the previous section, the *K2* algorithm uses historical system data as input.

Therefore running time of this algorithm is highly dependent on the size of the input data set, i.e. number of nodes (*n*) and number of observations (*t*). Each of the case BN shown in Figure 2 has different number of nodes ranging from 5 to 16, and the performance of the *K2* algorithm on each BN is analyzed using different input data sets. Figure 3 shows the experimental results on the performance of the *K2* algorithm. It can be observed from Figure 3 that the running time of the *K2* algorithm is quadratic ($O(n^2)$) with the number of nodes and linear with the number of observations. This is an expected result, since the *K2* algorithm reduces the time-complexity of finding associations from exponential (2^n) to quadratic (n^2). This brings the conclusion that for even substantially large systems ($n > 100$) the *K2* algorithm will be efficient to use (i.e., doing 10,000 iterations instead of 2^{100}).

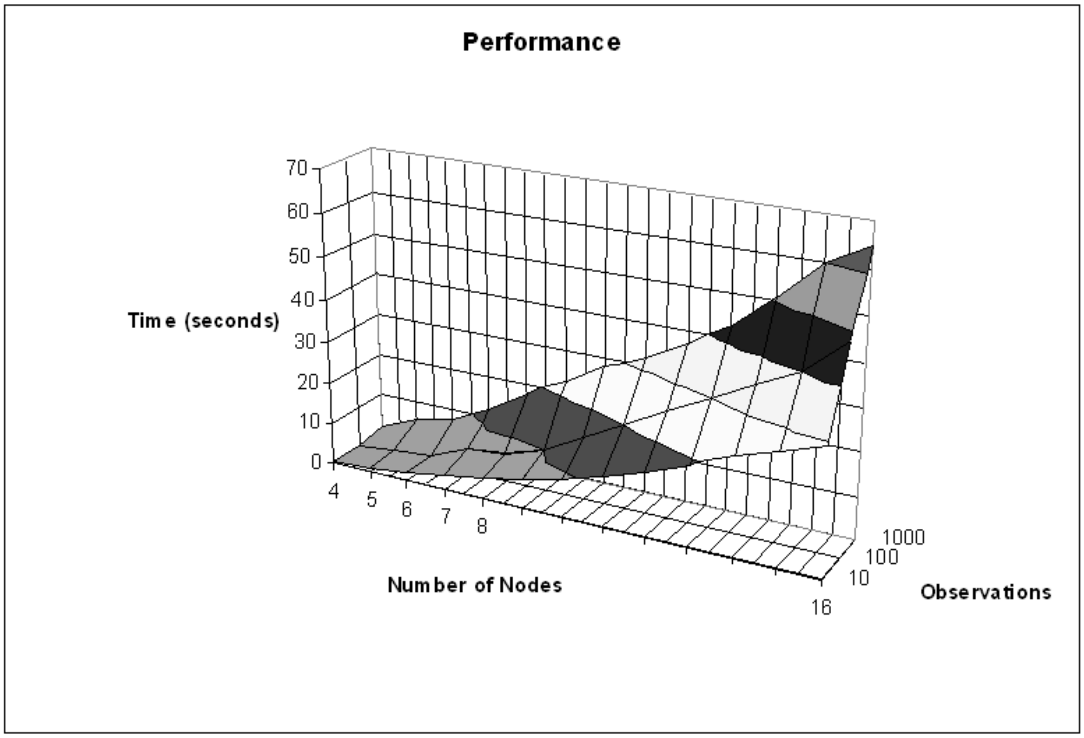

Figure 3. Running time of the *K2* algorithm

The number of observations that are used for discovering the associations between system components is an important measure for both efficiency of the *K2* algorithm and correctness of the constructed BN. Also accuracy of the reliability estimation is highly dependent on correctness of the underlying BN model. Errors in the *K2* algorithm would lead to incorrect assignments of associations in the BN; which will end up with inaccuracies in the reliability values. Once the BN is correctly constructed, estimating the system reliability is simple and straightforward as discussed in the previous section; therefore in this section we evaluate correctness of the BN constructed by the *K2* algorithm. The *K2* algorithm can be expected to find out associations more accurately when more observations used as input (Cooper and Herskovits 1992). Using the constructed BN, error rate (ρ) of the *K2* algorithm can be calculated by using Equation (2):

$$\rho = \frac{A_M}{A_T} = \frac{A_{FP} + A_{FN}}{A_T} \qquad (2)$$

In Equation (2), a *false positive* (A_{FP}) is defined as an association decided by the *K2* algorithm; however does not exist in the actual BN given in Figure 2. Conversely, a *false negative* (A_{FN}) is defined as an existing association in the actual BN that is missed by the *K2* algorithm. Both should be taken into account while calculating accuracy (where accuracy = 1-ρ) of the *K2* algorithm with the constructed BN. In this study, correctness of the constructed BN models are evaluated by using data sets with 10, 100 and 1000 observations for different case networks. General analysis results on the correctness and accuracy are provided in Figure 4.

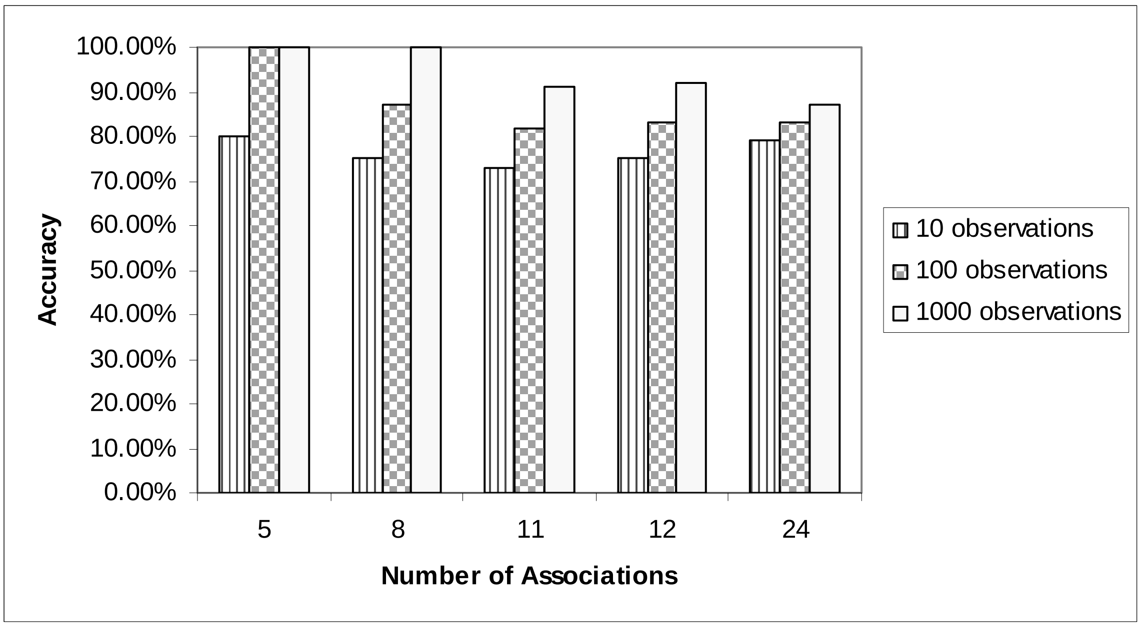

Figure 4. Accuracy of Results in BN Construction

According to Figure 4, regardless of the size of the constructed BN, accuracy of the BN model increases as more observations are used. This is an expected result, since associations between the components are decided by using the observations and these associations can be figured out more precisely when more observations available.

We used the case networks displayed in Figure 2 in our experiments. Our experimental results for the accuracy of BN construction as well as the CPU times for BN construction and reliability estimation are presented in Table 6. For our experiments we used a computer equipped with an Intel Centrino 2Ghz CPU and 2GB RAM. Moreover we implemented our proposed method in Matlab 7.0.

Table 6: Compilation of results for case BN

Case Network	Number of Nodes	Number of associations	*K2* algorithm CPU time (seconds)	Accuracy of the constructed BN (w/1000 observations)	Reliability Estimation CPU time (seconds)
1	4	5	1.839	100.00%	0.465
2	6	8	10.471	100.00%	0.981
3	5	8	5.567	100.00%	0.830
4	8	11	37.898	90.91%	1.004

5	7	12	26.012	91.66%	0.991
6	16	24	148.762	87.50%	1.587

Conclusions

Estimating system reliability using BN is a very popular practice and has been widely studied recently. There are numerous methods in the literature defined for estimating system reliability, which are mainly focused on doing it for specific systems, such as nuclear plants. However none of these studies dealt with the problem of requiring a human expert to construct the BN. This is the first study that introduces a methodology for efficient construction of BN models and estimating system reliability, with limited very human expert requirement. The proposed method uses historical data about the system to be modeled and constructs the BN model automatically. The *K2* algorithm is used for this purpose, which is a popular and efficient association rule mining method.

Next it was shown that the system reliability can efficiently be estimated by using the BN model. According to the experimental results, reducing the running time of finding associations from $O(2^n)$ to $O(n^2)$, the proposed methodology can work efficiently even with substantially large systems. Moreover, the BN models constructed by the *K2* algorithm are shown to be accurate, especially when more historical data about the system is available. As expected, the experimental results show that when 1,000 historical observations on the system are available, the constructed BN are more than 90% accurate. Accuracy of the constructed BN is highly influential on the correctness of the system reliability values, as incorrect associations in the BN would lead to biased calculations while estimating system reliability. In conclusion, the methodology introduced in this study will help systems engineers as it minimizes human interaction and provides efficient ways of automatically building a BN model and estimating system reliability.

References

Amasaki, S., Y. Takagi, O. Mizuno, and T. Kikuno. 2003. A Bayesian Belief Network for Assessing the Likelihood of Fault Content. In Proceedings of the 14th International Symposium on Software Reliability Engineering.

Barlow, R. E. 1988. Using influence diagrams. *Accelerated life testing and experts' opinions in Reliability*:145–150.

Bobbio, A., L. Portinale, M. Minichino, and E. Ciancamerla. 2001. Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliability Engineering and System Safety* 71 (3):249–260.

Boudali, H., and J.B. Dugan. 2006. A Continuous-Time Bayesian Network Reliability Modeling, and Analysis Framework. *IEEE Transaction on Reliability* 55 (1):86-97.

Cooper, G. F., and E. Herskovits. 1992. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9 (4):309-347.

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. 1999. *Probabilistic networks and expert systems*. New York, NY: Springer-Verlag.

Coyle, T., R.G. Arno, and P.S. Hale. 2002. Application of the minimal cut set reliability analysis methodology to the gold book standard network. In Proceedings of the Industrial and Commerical Power Systems Technical Conference.

Dahll, G., and B. A. Gran. 2000. The Use of Bayesian Belief Nets in Safety Assessment of Software Based Systems. *Special Issues of International Journal on Intelligent Information Systems* 24 (2):205-229.

Doguc, O., and J.E. Ramirez-Marquez. 2007. A Generic Method for Estimating System Reliability Using Bayesian Networks. *Journal of Reliability Engineering and System Safety (under revision)*.

Fenton, N., P. Krause, and M. Neil. 2002. Software Measurement: Uncertainty and Causal Modeling. *IEEE Software* 10 (4):116-122.

Gopal, C., H. Kuolung, and A. Nader. 2001. A New Approach to System Reliability. *IEEE Transactions on Reliability* 50 (1):75-84.

Gran, B. A., G. Dahll, S. Eisinger, E. J. Lund, J. G. Norstrøm, P. Strocka, and B. J. Ystanes. 2000. Estimating Dependability of Programmable Systems Using BBNs. In Proceedings of the Safecomp 2000.

Gran, B. A., and A. Helminen. 2001. A Bayesian Belief Network for Reliability Assessment. *Safecomp 2001* 2187:35–45.

Helminen, A. 2000. Reliability Estimation of Software-based Digital Systems Using Bayesian Networks. Espoo: Helsinki University of Technology.

Helminen, A., and U. Pulkkinen. 2003. Quantitative Reliability Estimation of a Computer-Based Motor Protection Relay Using Bayesian Networks. *Safecomp 2003* 2788:92–102.

Jensen, F. V. 2001. *Bayesian networks and decision graphs*. New York, NY: Springer Verlag.

Lagnseth, H., and L. Portinale. 2005. Bayesian Networks in Reliability.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA: Morgan Kaufmann.

Sigurdsson, J. H., L. A. Walls, and J. L. Quigley. 2001. Bayesian belief nets for managing expert judgment and modeling reliability. *Quality and Reliability Engineering International* 17:181–190.

Singh, H., V. Cortellessa, B. Cukic, E. Gunel, and B. Bharadwaj. 2001. A Bayesian Approach to Reliability Prediction and Assessment of Component Based Systems. In Proceedings of the 12th International Symposium on Software Reliability Engineering (ISSRE'01).

Biography

Ozge Doguc is a Business Analyst at Asset Control Inc. in New York, NY. Prior to that, she had worked as a Research Assistant in the Systems Engineering Department at Stevens Institute of Technology between 2004 and 2006. She holds a M.E. degree in Systems Engineering from Stevens Institute of Technology. She is graduated from Beykent University in Istanbul, Turkey obtaining B.S. degrees in Mathematics and Computer Sciences, and Management Information Systems. She is currently a PhD candidate in Systems Engineering and her research interests and areas include Reliability Estimation in Complex Systems, Grid Systems and System Operational Effectiveness.

Jose E. Ramirez-Marquez is an Assistant Professor at the Stevens Institute of Technology in the Department of Systems Engineering. His research interests include system reliability and quality assurance, uncertainty modeling, advanced heuristics for system reliability analysis, applied probability and statistical models and, applied operations research. He has authored more than 12 articles in leading technical journals on these topics. He obtained his Ph.D. at the Rutgers University in Industrial and Systems Engineering. He received his B.S. degree in Actuarial Science from the UNAM in Mexico City in 1998 and M.S. degrees in Industrial Engineering and Statistics from the Rutgers University.