2009 Third IEEE International Conference on Secure Software Integration and Reliability Improvement

SSIRI 2009 Fast Abstract
Simplifying Parametrization of Bayesian Networks in Prediction of System Quality

Aida Omerovic[A,B] and Ketil Stølen[A,B]
[A]SINTEF ICT, Pb. 124, 0314 Oslo, Norway, [B]University of Oslo, Department of Informatics, Pb. 1080, 0316 Oslo, Norway
Email: {aida.omerovic, ketil.stolen}@sintef.no

***Abstract*—Bayesian Networks (BNs) are a powerful means for modelling dependencies and predicting impacts of architecture design changes on system quality. The extremely demanding parametrization of BNs is however the main obstacle for their practical application, in spite of the extensive tool support. We have promising experiences from using a tree-structured notation, that we call Dependency Views (DVs), for prediction of impacts of architecture design changes on system quality. Compared to BNs, DVs are far less demanding to parametrize and create. DVs have shown to be sufficiently expressive, comprehensible and feasible. Their weakness is however limited analytical power. Once created, BNs are more adaptable to changes, and more easily refined than DVs. In this paper we argue that DVs are fully compatible with BNs, in spite of different estimation approaches and concepts. A transformation from a DV to a BN preserves traceability and results in a complete BN. By defining a transformation from DVs to BNs, we have enabled reliable parametrization of BNs with significantly reduced effort, and can now exploit the strengths of both the DV and the BN approach.**

I. Introduction

Our recent research (including an extensive industrial case study) indicates the feasibility and reliability of DVs for predicting impacts of architecture design changes on system quality. Promising experiences from the creation and use of DVs argue for their validity in this context. A DV is a structured, directed and parametrized tree showing:

- how and to what degree relevant parts and aspects of a system relate, with respect to a specified quality attribute
- to what degree each system part or aspect fulfills the specified quality attribute (which the DV is dedicated to).

Hence, DVs comprise two notions of parameters:

- EI: Estimated degree of Impact (assigned to an arc pointing to the node being influenced)
- QCF: degree of Quality attribute or Characteristic Fulfillment (assigned to a labelled node).

Figure 1a shows a small extract of a DV dedicated to the quality attribute availability. A quality attribute is defined by the underlying system specific "quality models", which may for example be based on [1]. A QCF value expresses to what degree the node (system part, aspect or similar) is realised so that it, within its own domain, fulfills the quality attribute in question. In the case of the "HW recovery" node of Figure 1a, the QCF value expresses the goodness of hardware recovery with respect to availability. The EI value on an arc expresses the degree of impact of a child

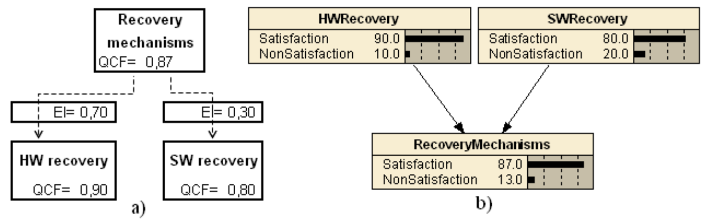

Figure 1. (a) a small extract of a DV. (b) the equivalent BN.

node (which the arc is directed to) on the parent node, or to what degree the parent node depends on the child node. The QCF value of each parent node is recursively (starting from leaf nodes and moving upwards in the tree) calculated by multiplying the QCF and EI value for each closest child and summing these products for all children. For example, with respect to Figure 1a[1]:

$$QCF(R) = QCF(H) * EI(R \rightarrow H) + QCF(S) * EI(R \rightarrow S) \quad (1)$$

The DV based approach constrains the QCF of each node to range between 0 (lowest) and 1 (ideal). The sum of EIs, each between 0 (no impact) and 1 (maximum impact), assigned to the closest children must be 1 (for model completeness purpose). Moreover, all nodes having a common parent have to be orthogonal (independent). The correlating nodes are placed at different levels, when structuring the tree. The leaf nodes of a DV must be observable.

Each DV is created with respect to a quality attribute. The structure of a DV is deduced following a systematic procedure: selecting the relevant parts of the underlying (design and quality) models, and structuring them in a top-down manner. The assigned parameters (EIs and QCFs of leaf nodes) may be based on expert judgment or measurement, thus incorporating objective and subjective input.

A BN [3] is a directed acyclic graph. Observation of known nodes (variables with a probability distribution) allows inferring the probability of others, using probability calculus and Bayes theorem throughout the model (propagation). In the case of three variables (where B and C cause A), Bayes rule states: $P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)}$. From this, we get:

$$P(A) = \frac{P(A|B,C)P(B)P(C|B)}{P(B|A)P(C|A,B)} \quad (2)$$

[1]Denoting: H: HW recovery; S: SW recovery; R: Recovery mechanisms

978-0-7695-3758-0/09 $25.00 © 2009 IEEE
DOI 10.1109/SSIRI.2009.36

According to the Law of Total Probability,

$$P(A) = P(A|B,C)P(B)P(C) + P(A|B,\neg C)P(B)P(\neg C) + P(A|\neg B,C)P(\neg B)P(C) + P(A|\neg B,\neg C)P(\neg B)P(\neg C) \quad (3)$$

While DVs are comprehensible and easy to construct, BNs are far more demanding to parametrize but in turn more adaptable and expressive than DVs. The number of estimates needed for a DV is $(2n-1)$, while traditional parametrization of a BN requires upto 2^n estimates (n: number of nodes). This represents a serious BN scalability issue. Our approach seeks to overcome the weaknesses and combine strengths of both the DV and the BN approach. A correct and complete construction of DVs does not presuppose familiarity with probabilistic reasoning or acquaintance with tools (which is the case with BNs). BNs are, however, more fine grained, more refinable and more expressive than DVs. The tool support of BNs is also more extensive, contributing to the adaptability and effectiveness of BNs. In contrast to DVs, BNs support correlations (between states of the same or several linked nodes), network optimization, self-learning, automatic evidence feeding, diagnostic analysis, notions of utility, cost, decision points, etc. By creating the initial model as a DV and transforming it into a complete BN, the parametrization is made straight forward, and the rest of the prediction analysis can take advantage of the BN approach.

II. Approach

Parametrization of a BN, based on a DV, involves:

- assigning the prior probabilities for each root node of a BN (corresponding to the QCF values on the leaf nodes of a DV)
- assigning the conditional probability tables (CPTs) associated with each non-root node, which quantify the relationship between nodes (corresponding to the EI values on the arcs of a DV).

Figure 1b shows the BN equivalent (from the NETICA [2] BN modelling tool) to the DV from Figure 1a. Leaf nodes of Figure 1a correspond to root nodes of Figure 1b. While arcs of Figure 1a represent direction of dependence, the arcs of Figure 1b represent causes and therefore have opposite direction. The EI values assigned to the DV are embedded in the CPTs of the BN. A node of the DV is assigned one QCF value, while a node of the BN may contain several states, each being assigned a probability of the node's being in the state. Satisfaction and non-satisfaction of the quality attribute are the states used for all BN nodes in this example. By asserting:

- orthogonality of DV nodes with a common parent implies conditional independence and disjoint probabilities of the corresponding BN nodes (having a common child)
- degree of an attribute fulfillment (given by a QCF value of a DV node) corresponds to probability of being in satisfaction state on the corresponding BN node
- EIs from a DV are embedded in the corresponding BN node's CPT, and the missing CPT values are deduced from joint probability distributions
- DV model completeness implies cumulative distribution of adjacent nodes within a BN

we deduce for Figure 1b:

$$P(H) = 0.9; P(\neg H) = 0.1; P(S) = 0.8; P(\neg S) = 0.2; P(R|S,H) = 1; P(R|S,\neg H) = 0.7; P(R|\neg S,H) = 0.3; P(R|\neg S,\neg H) = 0.$$

The marginal distribution on Figure 1b is automatically calculated by NETICA, using Bayes rule. Generally, given a network of three nodes and based on the assertions stated above, the following statements will hold:

$$P(A|B,C) = 1; P(A|B,\neg C) = 1 - P(A|\neg B,C); P(A|\neg B,\neg C) = 0; P(\neg B) = 1 - P(B); P(\neg C) = 1 - P(C); P(C|B) = P(C).$$

Inserting these general statements into the right hand side of Eq.3, results in the expression of the form given by Eq.1, which is used by all DVs. Also, inserting the general statements into the right hand side of Eq.2 results in $P(A)$. Thus, we have analytically argued the equivalence between a three node DV and a BN, provided the general assumptions of DVs are followed during its creation.

We have also empirically confirmed equivalent probability distributions and change impact propagation of a 14-nodes BN (in NETICA), transformed from an arbitrarily chosen, real DV. The empirical part involved three steps:

1) Transforming an existing DV to a BN by transforming QCFs of leaf nodes (of DV) to prior probabilities of root nodes (of BN) and EIs (of DV) to CPTs (of BN).
2) Confirming that node probability distributions of the BN's satisfaction states are equal to QCF values of all the corresponding nodes of the DV. This was the case.
3) Applying a change on both the DV and BN and confirming that the resulting propagations (i.e. QCFs and marginal probability distributions) correspond. This was the case.

III. Conclusions

Our recent research indicates that DVs represent a valid and efficient basis for initial parametrization of BNs. A DV to BN transformation represents a springboard to practical use of BNs and resolves some of the scalability issues related to BNs. Users with minimal statistical knowledge can build large scale realistic DVs for prediction analysis, and extend the analysis with BN facilities. Our approach shows the feasibility of a DV to BN transformation which preserves all the properties of the DV and results in a BN with a complete joint probability distribution. Further analysis based on BNs allows for improved traceability and adaptability of the models.

Acknowledgment

The research for this paper has been funded by the DIGIT (180052/S10) project of the Research Council of Norway.

References

[1] International organisation for standardisation: ISO/IEC 9126 - Software engineering - Product quality. 2004.

[2] Netica application software, http://www.norsys.com/download.html, 2009. Accessed on 11.02.2009.

[3] M. Neil, N. Fenton, and L. Nielsen. Building large-scale Bayesian networks. *Knowl. Eng. Rev.*, 15(3):257–284, 2000.