Hybrid Bayesian Networks for the Reliability Analysis of Systems with Continuous Variables

Luigi Portinale
Computer Science Institute, DiSIT
Research Center on Artificial Intelligence AI@UPO
Università del Piemonte Orientale
Alessandria, Italy
luigi.portinale@uniupo.it

Abstract

The standard way of dealing with continuous variables into reliability models is to discretize them, resulting in discrete state models. The present paper proposes an approach where continuous system variables can be directly exploited by resorting to Hybrid Bayesian Networks (HBN), where both continuous and discrete variables can be mixed in a general way. This allows one to: model the inter-dependencies between discrete state components or subsystems, model the inter-dependencies between continuous system variables, model the influence of contextual information on system variables and components, model the definition of specific system events or conditions given specific values of the system variables. We will show how the above issues can be captured in a principled way by the HBN formalism, by making the final analyses more grounded on the actual values of every system variable. We finally present a case study where the model of a granule storage tank system of a petrochemical plant is considered, and we present the results of specific analyses implemented as inference on the HBN model.

Introduction

Reliability analysis of critical systems can be carried out using a plethora of different formalisms (Trivedi and Bobbio 2017). Probabilistic Graphical Models (PGM) can be adopted profitably in order to properly model and analyze a critical system for dependability (Codetta-Raiteri and Portinale 2015). In several real-world situations, dependencies among the different parts of the system are present both in terms of basic system variables and sub-system components, and cannot be neglected without compromising the final analysis results. Some modeling choices also depends on specific contextual situations (e.g, an operational or a stand-by system). PGMs, and in particular Bayesian Networks (BN) are really well suited to deal with such issues (Bobbio et al. 2001; Langseth and Portinale 2007; Portinale, Codetta-Raiteri, and Montani 2010; Khakzad, Khan, and Amyotte 2011; Lakehal, Nahal, and Harouz 2019) In practical situations, several systems variables are of continuous nature; the standard way of introducing such system variables into the dependability model is to discretize them. In Fault Tree Analysis (FTA) (Ruijters and Stoelinga 2015), the modeling process departs from the fact that a suitable boolean abstraction of the system variables is provided, in such a way that events are modeling the working/failure dichotomy of components and subsystems. The discrete nature of the underlying variables is assumed, even if the formalism is extended to deal with multi-state components (Veeraraghavan and Trivedi 1994). This way forward affects consequently the PGM models that can be produced from the analysis: the resulting model is usually a BN with only discrete (often binary) nodes. In this paper, we claim that continuous system variables can be directly exploited into reliability analysis by exploiting the main features of Hybrid Bayesian Networks (HBN) (Salmerón et al. 2018). The mixing of both discrete and continuous nodes in the HBN formalism allows us to: (1) model the inter-dependencies between discrete variables (discrete state components or subsystems); (2) model the inter-dependencies between continuous system variables; (3) model the influence of contextual information on system variables and components; (4) model the definition of specific system events or conditions given specific values of the system variables. We will show how the above issues can be captured in a principled way by the HBN formalism, by making the final analyses more grounded on the actual variables values. of every system variable. As a case study, we consider the model of a granule storage tank system of a petrochemical plant (Yazdi, Nikfar, and Nasrabadi 2017), and we present the results of specific analyses implemented as inference on the HBN.

Hybrid Bayesian Networks

Definition 1 *A Bayesian Network (BN) is a pair $\langle G, P\rangle$ where: $G = \langle V, E\rangle$ is a DAG whose vertices $V = \{X_1, ...X_n\}$ represent (discrete) random variables and any edge $(X_i \to X_j) \in E$ represents a direct influence of X_i over X_j; P is a probability distribution over the variables represented by V, such that;*

$$P(X_1, ...X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i))$$

where $pa(X_i)$ is the set of parents nodes of X_i.

Definition 2 *A Linear Gaussian Bayesian network (LGBN) is a pair $\langle G, P\rangle$ where: $G = \langle V, E\rangle$ is a DAG whose vertices*

$V = \{X_1, ...X_n\}$ *represent continuous random variables with Gaussian distribution and any edge* $(X_i \rightarrow X_j) \in E$ *represents a linear dependency of* X_i *over* X_j*;* P *is a probability distribution over the variables represented by* V.

A root node X follows a normal probability distribution $\mathcal{N}(\mu, \sigma)$ with mean μ and standard deviation σ. Let Y be a child node with parents $X = \{X_1, ...X_k\}$: the conditional distribution of Y given X is given by the following expression with parameters: b: offset, $W^T = [w_1, ..., w_k]$: vector of weights, σ: standard deviation.

$$(Y|X = x) \sim \mathcal{N}(b + W^T x; \sigma) \quad (1)$$

The mean of Y is then given by $\mu_y = \sum_{i=1}^{k} w_i x_i + b$

When discrete variables are allowed as parents of continuous variables, but no discrete variable is a child of a continuous one, the model is called Conditional Linear Gaussian Bayesian Network (CLGBN). The parameterization is the same as in the case of standard BN and LGBN, except that the distribution of a continuous node is actually a set of distributions as in eq. (1), one for each configuration of the discrete parents. Finally, if also the restriction concerning the presence of discrete children for continuous nodes is removed, we get the most general class of models mixing discrete and continuous variables, properly called **Hybrid Bayesian Networks** (HBN). In this case, a discrete child having only discrete parents is specified as in a standard BN; in case continuous parents are present, a soft threshold model like probit or logit (Aldrich and Nelson 1984) can be adopted (e.g., a soft-max function). Of course, if discrete nodes are present as parents of D, a different logistic model has to be provided, for each configuration of discrete parents.

Model design principles

We propose some high-level design principles for modeling physical systems characterized by different components and subsystems, as well as specific system variables that may be either discrete or continuous. In particular, the status or the behavior of the system may be influenced by contextual situations such as an operational or a stand-by system, some external conditions such as the working temperature, etc... We then consider the set of contextual conditions as influencing factors for system variables. System variables may be characterized by either discrete states or continuous values (double edge ovals) which may depend on contextual factors, but that may also influence each other. Finally specific (sub)-system events (or conditions) may be defined considering the status of the system variables; moreover, these events may influence each other (e.g., the faulty status of a set of components that defines the faulty status of a whole sub-system as in a series or parallel system). Given the above characterization we can build a model based on the HBN formalism: (1) for each contextual condition create a discrete state random variable node; (2) for each discrete (respectively continuous) system variable, create a discrete state (respectively continuous) random variable node; (3) for each specific system event of interest create a binary state random variable node; (4) connect every node X with every node Y such that Y depends on X; (5) quantify the resulting HBN model.

Case Study

As a case study, we consider tank corrosion problems arising in a petrochmical plant as described in (Yazdi, Nikfar, and Nasrabadi 2017). Storage tank corrosion has been identified as the main threat to tank safety worldwide and corrosion detection becomes the main approach to ensure the safety of storage tanks(Feng, Yang, and Huang 2019). We identify the entities of interest for the case study as follows:

Season (S): context; a discrete variable with possible states `dry, wet`.

Budget (B): continuous; the monthly amount for maintenance measured in thousands money units, considering a range $r_B = [10, 150]$, a mean of $\mu_B = 80$ and a standard deviation $\sigma_B = 25$.

Internal Coating (ICoa): continuous; the width of internal coating measured in micron ($1\mu m = 10^{-6} m$) with a normal range $r_{Icoa} = [50, 200]$; it depends from B and MP, thus the mean value is assumed to be a linear function of the available maintenance budget, conditioned by the presence/absence of a material problem. In particular, we consider a situation where, the presence of a material problem reduces the width of a 10% factor; in the considered case study, we fitted this situation with the following linear regression models (w is the internal coating width and b is the budget amount): $w = 0.883b + 23.7$ when there is no material problem, and $w = 0.79b + 21.3$ when a material (pH) problem occurs. Standard deviation has been set to 31.2 for a dry season and to 28.1 for a wet season.

Granule (G): continuous; the measure of the pH of granules. It influences the width of internal coating (acidic or alkaline granule will reduce the internal coating width), by producing a material problem. We assume a mean value $\mu_G = 7$ and a standard deviation $\sigma_G = 2.5$.

Atmospheric Condition (AC): continuous; the percentage of humidity in the air; it depends on the type of season. We consider an average humidity $\mu_{AC} = 50\%$ with standard deviation $\sigma_{AC} = 3$ during a dry season, and $\mu_{AC} = 80\%, \sigma_{AC} = 2.8$ during a wet season. It may cause the occurrence of a weather problem.

External Coating (ECoa): continuous; the total width of external coating measured in micron with usual range $r_{Ecoa} = [300, 500]$; it is influenced by the available maintenance budget and the atmospheric conditions, and may cause an external coating problem. In the case study we fitted the following regression models on the available data (b is the amount of budget, ac is the air humidity and ec the external coating width):

$ec = 0.03b - 1.19ac + 454.8$ Standard deviation has been set to 50.2.

Tensile Stress (TS): continuous; stretching force per unit area measured in *megapascal* (MPa); in the current case study we consider a steel with a cracking value of 841 MPa, with critical values starting from 785 MPa. We assume an a-priori mean of 770 MPa, and a standard deviation of 6.5 MPa. A large tensile stress is likely to augment corrosion penetration speed.

Stress Corrosion Cracking (SCC): continuous; the corrosion penetration speed measured as micron per year ($\mu m/y$), with critical values being larger than 2.6 $\mu m/y$; it depends on the tensile stress and in the case study, we fitted the following linear regression model (s is the penetration speed and ts the tensile stress): $s = 0.0124ts + 6.8$. Standard deviation has been set to 0.58.
Internal Coating Problem (ICoaP): event; the occurrence of a internal coating deterioration in the tank depending from values of $ICoa$; the smaller the internal coating width, the larger the event's probability. In the case study we fitted the following logistic model (ip is the occurrence of the event and w the internal coating width): $P(ip) = 1 - \frac{1}{1+e^{-2w+90}}$ (with $w < 44\mu m$ the occurrence probability tends to one while with $w > 47\mu m$ tends to zero)
Acidic: event; presence of acidic material and depending from values of pH of G; the fitted logistic model is (ac is the event occurrence and g the granule's Ph):
$P(ac) = 1 - \frac{1}{1+e^{-3g+15}}$ (for $g < 4.5$ the events tends to be true and for $g > 6.5$ it tends to be false)
Alkaline (Alk): event; presence of alkaline material and depending from values of pH of G; the fitted model is (alk is the event occurrence and g the granule's Ph):
$P(alk) = \frac{1}{1+e^{-3g+27}}$ (for $g < 8$ the events tends to be false and for $g > 10$ it tends to be true)
Material Problem (MP): event; the occurrence of a material deterioration in the tank; it occurs when the granule is either acidic or alkaline (i.e., a logical `or` between Alk and $Acidic$).
Weather Problem (WP): event; the occurrence of a weather problem and depending from the air's humidity (AC); the fitted model is (wp is the event occurrence and ac the air's humidity): $P(wp) = \frac{1}{1+e^{-1.5ac+130}}$ (for $ac < 84\%$ the events tends to be false and for $ac > 89\%$ it tends to be true)
External Coating Problem (ECoaP): event; the occurrence of an external coating deterioration on the tank and depending from values of $ECoa$ (width of external coating); the fitted model is (ecp is the event occurrence and ec the external coating width): $P(ecp) = 1 - \frac{1}{1+e^{-2ec+680}}$ (for $ec < 338\mu m$ the events tends to be true and for $ec > 342\mu m$ it tends to be false)
SCC Problem (SCCP): event; the occurrence of a deterioration due to the corrosion penetration speed (SCC); the fitted model is ($sccp$ is the event occurrence and scc the corrosion penetration speed):
$P(sccp) = \frac{1}{1+e^{-3scc+763}}$ (for $scc < 252.5\mu m/y$ the events tends to be false and for $scc > 255.5\mu m/y$ tends to be true)
Internal Corrosion (IC): event; the occurrence of internal tank corrosion: the occurrence of either an internal corrosion problem ($IcoaP$) or a material problem (MP).
External Corrosion (EC): event; the occurrence of external tank corrosion: the occurrence of either an external corrosion problem ($EcoaP$) or a weather problem (WP).
Corrosion (C): event; the catastrophic event representing the occurrence of corrosion in the tank; it is defined by the occurrence of either internal corrosion (IC) or external corrosion (EC) or tensile stress related corrosion ($SCCP$).

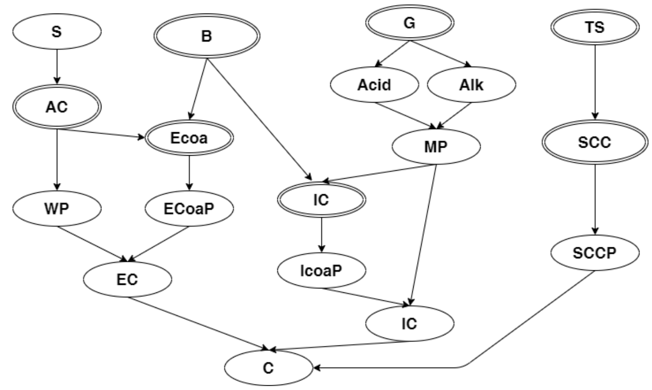

Figure 1: Structure of the HBN for the case study

Figure 1 represents the structure of the HBN resulting from the characterization of the system (double edge nodes represent continuous variables and single edge nodes represent discrete variables). Quantification is given by the above described fitted models: a (conditional) linear regression model for the mean of continuous variables, a logistic model for discrete variables having continuous parents, and a deterministic function (conditional probability distribution modeling a deterministic logical function) for discrete variables with only discrete parents (e.g. a logical `or`). We consider a time horizon of 180 days (a 6 months season); a monitoring of the system every 10 days (i.e., we performed a posterior probability query on the HBN every 10 days); a fixed *Tensile Stress* during the 6 months period with $TS = 720MPa$; an initial budget $B = b_0 = 150K$ and a maintenance plan contemplating a monthly reduction by 3%; finally the granule has an initial neutral pH, but it is subject to an acidification process reducing the Ph by 5% each month. We consider the above scenario under the alternative contextual situations of wet and dry season. In both situations, during the monitored period we expect an increase of corrosion due to several factors: the maintenance budget reduction, impacting on both internal and external corrosion (the coating width cannot be kept completely under control); the acidification of the granule, impacting on the internal corrosion process. Moreover, we expect more corrosion during wet months since air humidity impacts on external corrosion. Analysis are performed by computing the posterior probability on the nodes of interest, by setting the evidence on nodes B, TS, G and S at the considered time instant. Figure 2 shows the tank corrosion probability (top event); as expected, corrosion probability is larger during wet season and shows an increasing pattern. The material deterioration due to the acidification of the granules is evident in figure 3 where the probability of the detection of a material problem is also increasing during time. In this case the seasonal context has no influence on the problem (it is a problem related to the internal layer of the tank) and the pattern is the same for both seasons. It is worth noting that the internal corrosion event IC shows the same trend, with very similar probability values; we do not report here this plot, but figure 4 reports the trend in width reduction for the internal coating (the graph shows the error bars

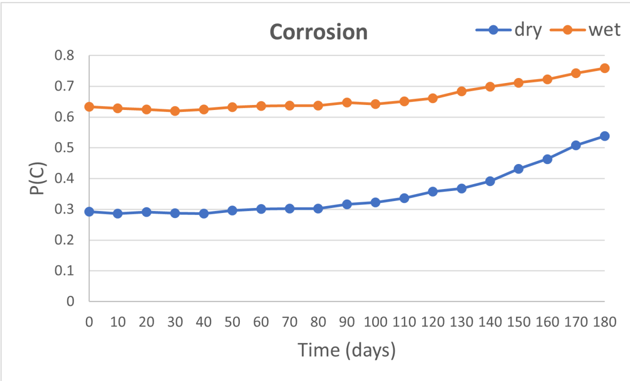

Figure 2: Tank corrosion probability

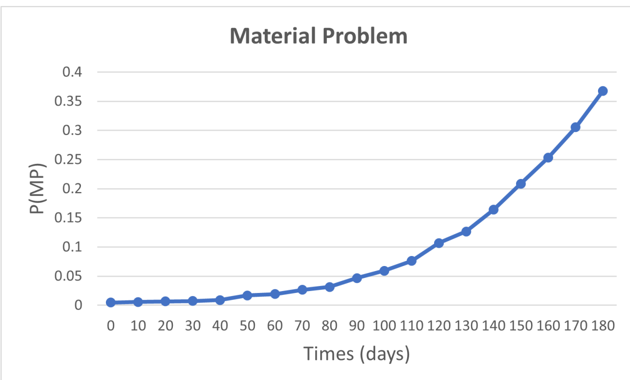

Figure 3: Material Problem probability (acidification)

with mean value and standard deviation). This means that the internal corrosion process is mainly due to the material problem related to acidic granule. In other words, the impact of the budget reduction is limited, and this is also evidenced by the analysis of external corrosion (figure 5) where we can notice that the probability of external corrosion depends on season, but is is almost stable during the monitored period (budget reduction has no practical impact on that).

Finally, another important property of PGM is that the system reliability analysis is not restricted to monitoring and prediction (as usually done in standard FTA for example), but it can be performed in a more general way, for instance to perform diagnostic reasoning. Consider the following situation: we are in dry season ($S = \texttt{dry}$), we measure an air's humidity $AC = 60\%$, we have a fixed maintenance budget $B = 150K$ and we know that the granule has neutral pH (measured pH $G = 6.85$). Let us also suppose that a corrosion process is detected by inspection (i.e., $C = \texttt{true}$). We can investigate potential causes related to both coating and stressing problems by setting the above nodes as evidence, and by querying variables $Icoa, Ecoa$ (to check internal and external coating width respectively, TS and SCC (to check for dangerous stretching and penetration speed). We get the following posterior distributions: $ICOA \sim \mathcal{N}(156.28; 31.44)$, $ECOA \sim \mathcal{N}(377.03; 54.5)$, $TS \sim \mathcal{N}(780.39; 6.49)$, $SCC \sim \mathcal{N}(2.86; 0.53)$ It is likely

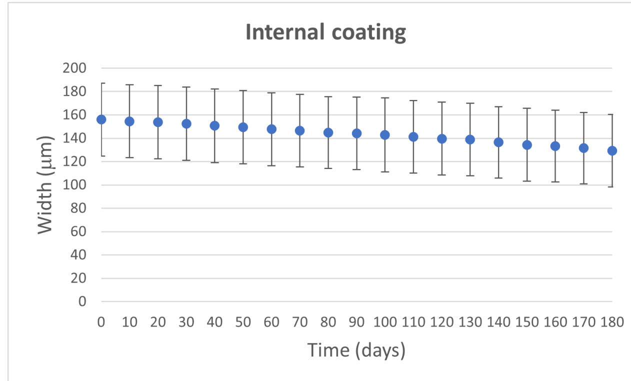

Figure 4: Internal coating (width reduction)

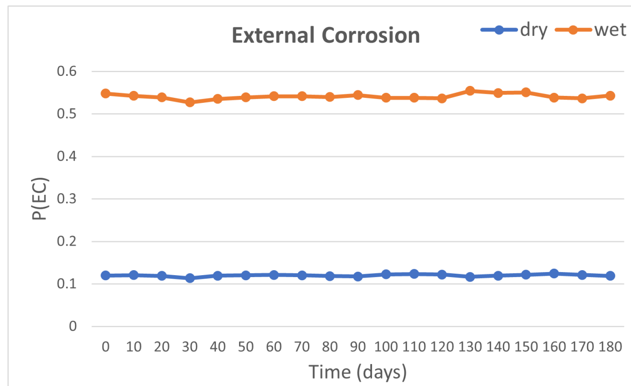

Figure 5: External corrosion probability

that corrosion is due to stressing problems, and in particular because of a high penetration speed; this is confirmed by querying the $SCCP$ node, since $P(SCCP = \texttt{true}|evidence) = 0.846$.

Conclusions

This paper proposes to exploit in reliability analysis the HBN formalism which is able to mix both discrete and continuous random variables in a general way. We have shown the potential offered by this approach, by considering the storage tank subsystem of a petrochemical plant, where reliability analysis (monitoring, prediction and diagnosis) is performed as posterior probability computation on the HBN, The case study has been implemented in MATLAB by exploiting the BNT toolbox (Murphy 2014); since exact inference is possible only when continuous nodes with discrete children are observed, we resorted to approximate inference, and in particular to likelihood weighting algorithm (Schachter and Peot 1990) setting a quite large number of samples ($30K$), and with results that are computed in a few seconds.

References

Aldrich, J., and Nelson, F. 1984. *Linear Probability, Logit and Probit Model*. Sage Publ.

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 & System Safety* 71(3):249–260.

Codetta-Raiteri, D., and Portinale, L. 2015. *Modeling and Analysis of Dependable Systems: a Probabilistic Graphical Model Perspective*. World Scientific Publ.

Feng, Y.; Yang, Y.; and Huang, B. 2019. Corrosion analysis and remaining useful life prediction for storage tank bottom. *Int. Journal of Advanced Robotic System* 1–9.

Khakzad, N.; Khan, F.; and Amyotte, P. 2011. Safety analysis in process facilities: Comparison of fault tree and bayesian network approaches. *Reliability Engineering & System Safety* 96(8):925–932.

Lakehal, A.; Nahal, M.; and Harouz, R. 2019. Development and application of a decision making tool for fault diagnosis of turbocompressor based on Bayesian network and fault tree. *Management and Production Eng. Review* 10:16–24.

Langseth, H., and Portinale, L. 2007. Bayesian networks in reliability. *Reliability Engineering & System Safety* 92(1):92–108.

Murphy, K. 2014. Bayes Net Toolbox for Matlab. https://github.com/bayesnet/bnt.

Portinale, L.; Codetta-Raiteri, D.; and Montani, S. 2010. Supporting reliability engineers in exploiting the power of Dynamic Bayesian Networks. *International Journal of Approximate Reasoning* 51(2):179–195.

Ruijters, E., and Stoelinga, M. 2015. Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools. *Computer Science Review* 15-16:29 – 62.

Salmerón, A.; Rumí, R.; Langseth, H.; Nielsen, T.; and Madsen, A. 2018. A review of inference algorithms for hybrid bayesian networks. *JAIR* 62:799–828.

Schachter, R., and Peot, M. 1990. Simulation approaches to general probabilistic inference on belief networks. In *Proc. 6th Intern. Conf. on Uncertainty in AI (UAI 90)*, 221–230.

Trivedi, K. S., and Bobbio, A. 2017. *Reliability and Availability Engineering: Modeling, Analysis, and Applications*. Cambridge University Press.

Veeraraghavan, M., and Trivedi, K. 1994. A combinatorial algorithm for performance and reliability analysis using multistate models. *IEEE Trans. on Computers* 43(2):229–234.

Yazdi, M.; Nikfar, F.; and Nasrabadi, M. 2017. Failure probability analysis by employing fuzzy fault tree analysis. *Int. Journal of System Assurance Eng. and Management* 8(2):1177–1193.