

DYN-RISK- Design and Development of a Dynamic Risk Assessment Tool

Tomaso Vairo^{a*}, Paolo Bragatto^b, Maria F. Milazzo^c, Margherita Pettinato^a, Bruno Fabiano^a

^aDICCA – Civil, Chemical and Environmental Engineering Department - University of Genoa, Via Opera Pia, 15 – 16145 Genoa, Italy

^bDipartimento Innovazioni Tecnologiche INAIL Via Fontana Candida, 1 00078 Monteporzio Catone (RM), Italy

^cDipartimento Ingegneria Università di Messina C.da di Dio, Villaggio Sant'Agata, 98158 Messina, Italy

tomaso.vairo@edu.unige.it

This paper concerns the design and development of a dynamic risk assessment tool, and the determination of a dynamic indicator of the safety level. The DYN-RISK system is intended to be a predictive decision-making tool, which provides a constantly updated risk, based on the ongoing operations, for the whole plant and all the components. The aim of the DYN-RISK project is to overcome the limitations of the classical approaches to QRA, mainly related to their intrinsically static nature, which reflects on an inability to adequately catch interdependencies and likelihoods in fault trees and event trees. In accordance with the resilience engineering practice, the objective of this project is to enhance the operational control, focusing on the overall performance of the complex system constituted by plant and environment. The DYN-RISK system relies on a robust Machine Learning architecture providing the system the ability to automatically learn and improve from experience, without being explicitly programmed. The overall model includes Deep Neural Networks, for predicting the critical process variables values, Hidden Markov Models for inferring the most probable sequence of states for the system, and Hierarchical Bayesian Networks, for combining Boolean events, such malfunctions and failures of critical equipment, and system states predictions. The final outcome is a real time dynamic risk indicator. The DYN-RISK system is validated on a Chemical and Petrochemical storage terminal in northern Italy.

1. Introduction

The need for a more dynamic and affordable risk assessment process is underlined by several authors (Leveson, 2020; Jain et al., 2018; Vairo et al., 2019 and 2021) and most of the proposed strategies are about a deep change in the inferential process. Inferences are the steps of reasoning. Inferential statistics is the procedure according to which observations from a data sample are used to make inferences about the largest population from which the sample was drawn. It stands to reason that the considered sample needs to accurately and significantly reflect the population. To be more arbitrary as possible, it is recommended to use a random sampling method. When using the inferential statistics, there will always be an error between the properties of the global population and the sample's ones, so that the uncertainty is always included in the results and an interval of confidence is outlined.

The traditional Quantitative Risk Assessment (QRA) is based on a frequentist approach that calibrates the plausibility of propositions by considering repeated sampling of a population distribution. The evaluation of frequentist properties of a statistical proposition provides fixed values, that's why the traditional QRA is static. For designing a dynamic risk assessment model, the branch of inferential statistics has to be explored, i.e., the abductive inference. The definition of abduction indicates it as "a syllogism" composed of an evident major premise and a probable minor premise, leading to an only probable conclusion. Essentially, a conclusion is drawn from the information that is known. Abductive reasoning typically begins with an incomplete set of observations and proceeds to the likeliest possible explanation for the set (Thagard et al., 1997). In the abductive approach, the conclusion is always a best guess that is continuously updated by new evidences, thus making

the abductive approach an intrinsically dynamic one. The framework developed in the present work relies on Bayesian inference, which is the heart of the implemented model. As amply known, BN model is conceived as a directed acyclic graph including nodes and arcs (i.e., variables and direct causal relationships). The derived Bayesian statistics rely on probability to quantify uncertainty, or degree of belief, so that probability distributions add a quantitative dimension and can be used to represent the states of belief, attaining predictive models on the dynamic propagation of an hazardous event (Yang et al., 2013; Vairo et al., 2019).

2. Methods

Abduction derives the best explanations for observations. Statistical abduction attempts to define a probability distribution over explanations and to evaluate them by their probabilities. The framework of statistical abduction is general since many well-known probabilistic models, i.e., Bayesian Networks (BN), Hidden Markov Models (HMMs) and Tree-based probabilistic models, are formulated as statistical abduction (Sato and Kameya, 2001). In a logical framework, abduction, is usually defined as follows:

- Given: Background knowledge B and observations O, both represented as sets of formulae in first-order logic, where O is typically restricted to a conjunction of ground literals.
- Find: A hypothesis H, also a set of logical formulae, such that $B \cup H \not\models \perp$ and $B \cup H \models O$.

Where \models means logical entailment and \perp means false, i.e., find a set of assumptions that is consistent with the background theory and explains the observations. There are generally many hypotheses H that explain a particular set of observations O. Several researchers have applied logical abduction to tasks like plan recognition and diagnosis (Ng et al., 1992).

The logical steps for the DYN-RISK model are:

- prediction of critical variables values;
- prediction of system states sequence;
- real time update of the risk parameters, which define a dynamic risk indicator.

2.1 Prediction of critical variables values

The model component in duty of predicting the critical variable values is a series of Deep Neural Networks (DNN) ad hoc designed (Vairo et al., 2021), according to the following characteristics:

- hidden layers: 10;
- neurons in each hidden layer: 24;
- learning rate: 10^{-7} ;
- step max: 10^8 ;
- activation function: Tanh;
- error function: SSE;
- prediction time interval: 15 min.

The activation function performs data analysis and processing; when the sum of weighted inputs and biases exceeds a precise activation threshold, the activation function considers the argument valid and processable. Weight parameter quantifies the inputs importance. Weights and bias are corrected by learning algorithm for adapting the DNN to the input dataset. The data processing scheme is depicted in Figure 1.

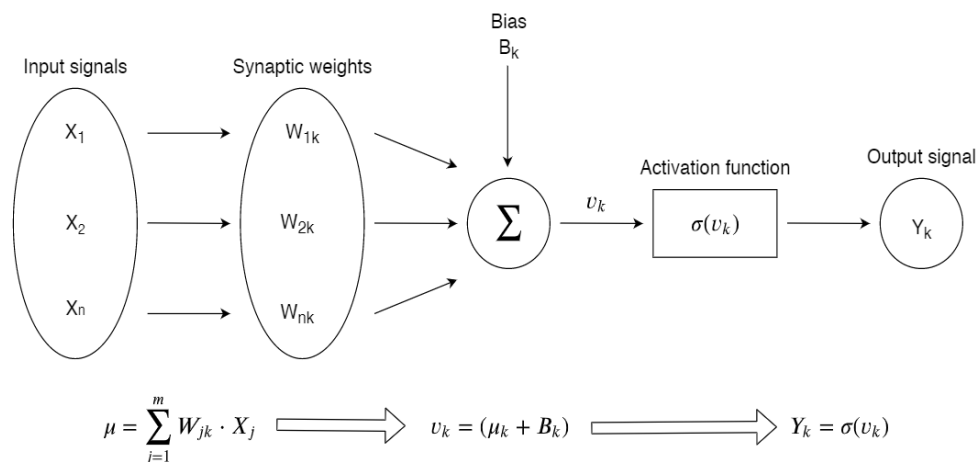


Figure 1: Data processing in Neural Networks

2.2 Prediction of system states sequence

As commented by Sarkar et al. (2019), accidents do not happen chaotically, then underlying patterns and trends exist and can be caught. As already mentioned, the Bayesian approach has been proven to be a robust probability reasoning method under uncertainty, providing a tool for incorporating evidence during operations. HMMs seems to be one of the most promising and reliable approaches (Galagedarage et al., 2019).

An HMM is a generative probabilistic model, in which a sequence of observable X variables is generated by a sequence of internal hidden states Z. The hidden states are not observed directly. The transitions between hidden states are assumed to have the form of a first order Markov chain. They can be specified by the start probability vector π and a transition probability matrix A. The emission probability of an observable can be any distribution with parameters θ conditioned on the current hidden state. The HMM is completely determined by π , A and θ . In the present work, the hidden states are the states between a regular performance and a failure of a sub-system. The only known states are the first one (the component is performing well) and the last one (the component fails), while the hidden states in between may cautiously represent the precursors of accidental events. The resulting outputs of the system are the process variable values.

The abductive inference in the developed HMM relies on the Metropolis Hastings (MH) sampling algorithm. MH performs forward and backward inference by computing the distribution space of the model parameters and determine the most likely outcome.

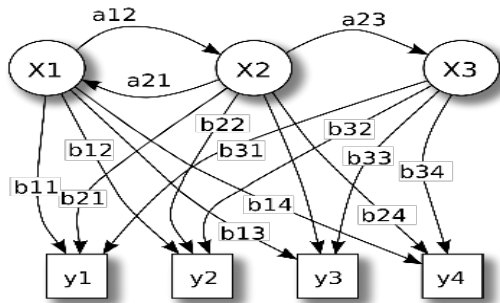


Figure 2: Conceptual Representation of a Hidden Markov Model

The samples are then analysed by the HMM with the *Viterbi* algorithm (Lember et al., 2019). Given a HMM with state space S, initial probabilities π_i of being in state I, transition probabilities $a_{i,j}$ of transitioning from state I to state j, and a set of observations y_1, \dots, y_T , according to the *Viterbi* algorithm the most likely state sequence x_1, \dots, x_T , that produces the observations is given by the recurrence relations according to Eqs(1) and (2).

$$V_{1,k} = P(y_1 | k) \pi_i \quad (1)$$

$$V_{t,k} = \max_x (P(y_t | k) a_{x,k} V_{t-1,x}) \text{ with } x \in S \quad (2)$$

Where $V_{t,k}$ is the probability of the most probable state sequence $P(x_1, \dots, x_t, y_1, \dots, y_t)$ responsible for the first observations that have k as its final state. The *Viterbi* path V can be retrieved by saving back pointers that remember which state x was used in the second equation.

2.3 Dynamic Risk Indicator

The Boolean elements of the risk analysis and the predictions of system states sequence are integrated in a Hierarchical Bayesian Network (HBN). The HBN is obtained by remapping the Fault Trees (FT) (Zurheide et al., 2021). The quantitative analysis of a BN may proceed along two lines. A forward (or predictive) analysis, in which the probability of occurrence of any node of the network is calculated based on the prior probabilities of the root nodes and the conditional dependence of each node. A more standard backward (diagnostic) analysis that concerns the computation of the posterior probability of any given set of variables given some observations (the evidence), represented as instantiation of some of the variables to one of their admissible values.

In the DYN-RISK model, the prior probabilities of the nodes are represented by the FT failure rates. The parameters are updated, from one side, by the process variable values predictions, and, from another side, by

the prediction of the state sequences, which defines the parameters of the probability distributions for each network node.

The general architecture of the DYN-RISK model is represented in Figure 3.

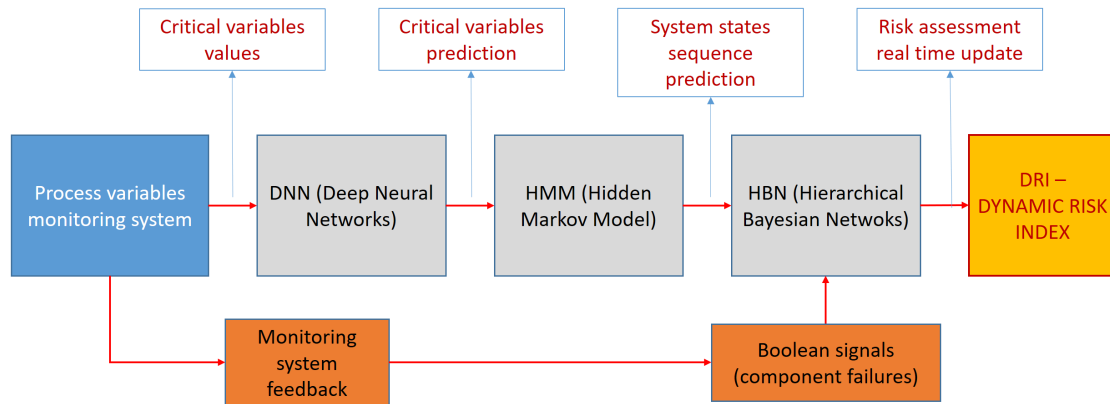


Figure 3: Conceptual Architecture of the DYN-RISK model

3. Case Study validation

The DYN-RISK framework was tested in a petroleum products distribution terminal, focusing on fire and explosion hazards connected to the Vapour Recovery Unit (VRU). Figure 4 provides as an example the simplified Fault Tree for a leakage in the VRU, possibly originating a source term for accident scenario evolution.

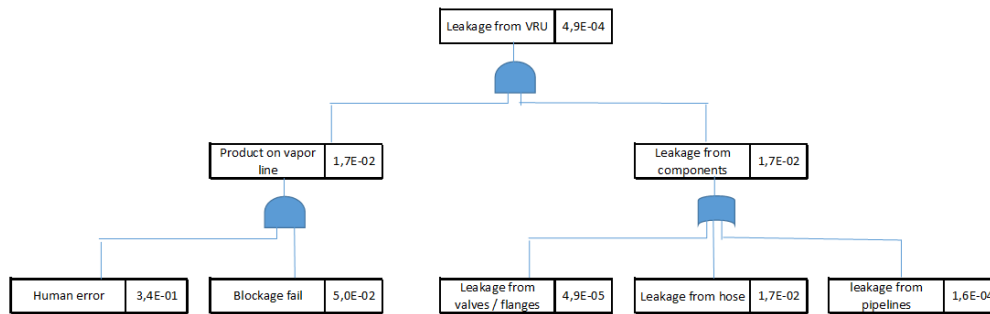


Figure 4: Fault Tree Analysis of the VRU section

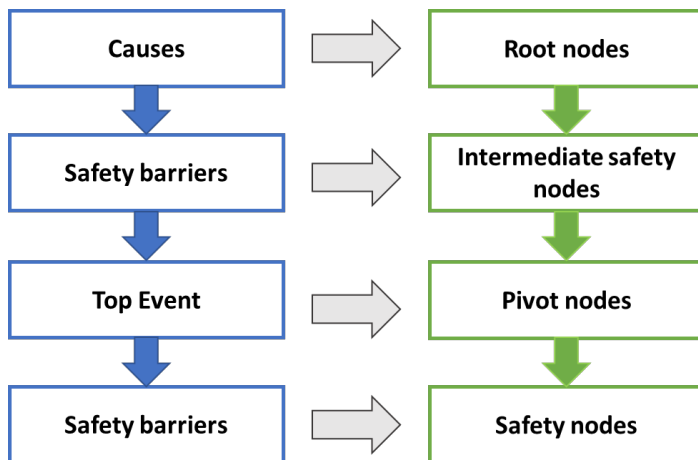


Figure 5: Flowchart of the approach adopted in this study and relevant mapping steps

The transposition of the FT in a HBN follows the flowchart outlined in Figure 5, developed following the reasoning by Zurheide et al. (2021). According to the developed approach, the resulting structure is depicted in Figure 6.

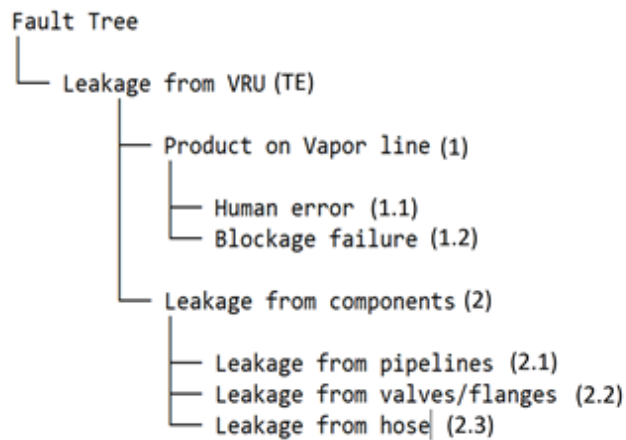


Figure 6: Transposing FT into HBN

The risk parameters of the tree are updated, according to the designed model (Figure 3), as shown in Table 1.

Table 1: Risk assessment dynamic parameters

Hierarchical level	Node	State	Expected probability
2	1	works	0.999998
2	1	fails	0.000002
2	2	works	0.999999
2	2	fails	0.000001
3	1.1	works	0.999987
3	1.1	fails	0.000013
3	1.2	works	0.999982
3	1.2	fails	0.000018
3	2.1	works	0.999992
3	2.1	fails	0.000008
3	2.2	works	0.999988
3	2.2	fails	0.000012
3	2.3	works	0.999991
3	2.3	fails	0.000009

At last, the dynamic update of the above represented parameters, defines the probability distribution of the Top Event (TE), as summarized in Table 2.

By combining all the instant values of Phi for all the identified Top Events, the Dynamic Risk Indicator is defined.

Table 2: Top Event real time expected probability

Hierarchical level	Node	State	Phi (TE)
1	TE	works	0.9999999
1	TE	fails	0.0000001

4. Conclusions

In the near future (Pasman et al., 2021), under fast development of process digitalization, fully integration of sensors and data streams with digital twins, machine learning and artificial intelligence-based method can make hidden abnormal features fully detectable with continuous updating, while transient operations become more accessible for risk analysis. In this research line, the presented approach relies on a hybrid model incorporating different data driven models in a complete logical and interconnected model (DNN – HMM – HBN). The former exhibits a robust predictive capability on the process variables, the second, which is the real predictive model,

performs abductive inference on the system states sequence and the latter explores the interdependencies among the system components and their modification alongside process variables fluctuation. The combined model outcomes are subsequently used to generate a dynamic risk indicator connected with the process variable predictions. The preliminary industrial validation evidenced the framework capability for attaining reliable results, in terms of accuracy and test error performance, allowing a refinement of the system risk picture and a decision-making support based on field evidences. This possibility is currently constrained by the need of filling the gap between process data sources and the frontline to provide dynamic, real-time risk insights.

Acknowledgment

This research was partially funded by INAIL within the framework of the call BRIC/2019/ID2 (Project DYN-RISK).

References

- Galagedarage D.M., Khan F., 2019, Process fault prognosis using Hidden Markov Model–Bayesian Networks hybrid model, *Industrial & Engineering Chemistry Research*, 58, 12041–12053.
- Jain P., Rogers W.J., Pasman H.J., Mannan M.S., 2018, A Resilience-based integrated process systems analysis: Part II management system layer, *Process Safety and Environmental Protection*, 118, 115-124.
- Lember J., Gasbarra D., Koloydenko A., Kuljus K., 2019, Estimation of Viterbi path in Bayesian Hidden Markov models, *Metron*, 77, 137-169.
- Leveson N., 2020, Safety III: A Systems Approach to Safety and Resilience, MIT Engineering Systems Lab <sunnyday.mit.edu/safety-3.pdf> accessed 16.11.2021.
- Ng H.T., Mooney R.J., 1992, Abductive plan recognition and diagnosis: A comprehensive empirical evaluation, *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, 499-508.
- Pasman, H.J., Fabiano, B. 2021. The Delft 1974 and 2019 European Loss Prevention Symposia: Highlights and an impression of process safety evolutionary changes from the 1st to the 16th LPS. *Process Safety and Environmental Protection* 147, 80-91.
- Sarkar S., Vinay S., Raj R., Maiti J., Mitra P., 2019, Application of optimized Machine Learning techniques for prediction of occupational accidents, *Computers and Operations Research*, 106, 210–224.
- Sato T., Kameya Y., 2001, Parameter learning of logic programs for symbolic-statistical, Modeling, *Journal of Artificial Intelligence Research*, 15, 391-454.
- Thagard P., Shelley C., 1997, Abductive reasoning: Logic, visual thinking, and coherence, Chapter In: Della Chiara M. L., Doets K., Mundici D., van Benthem J. (Eds), *Logic and Scientific Methods*, Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science), Vol 259, Springer, Dordrecht, Netherlands, 413-427.
- Vairo T., Gualeni P., Reverberi A.P., Fabiano B., 2021, Resilience dynamic assessment based on precursor events: application to ship LNG bunkering operations, *Sustainability*, 13, 12, 6836.
- Vairo T., Milazzo M.F., Bragatto P., Fabiano B., 2019, A dynamic approach to fault tree analysis based on Bayesian Beliefs Networks, *Chemical Engineering Transactions*, 77, 829-834.
- Yang M., Kahn F., Lye L., 2013, Precursor-based hierarchical Bayesian approach for rare event estimation: a case of oil spill accident, *Process Safety and Environmental Protection*, 91, 333-342.
- Zurheide F.T., Hermann E., Lampesberger H., 2021, pyBNBowTie: Python library for Bow-Tie analysis based on Bayesian Networks, *Procedia Computer Science*, 180, 344–351.