Advances in Reliability, Safety and Security

ESREL 2024
Collection of Extended Abstracts

Using Bayesian Network To Analyze Incident Datasets

Matteo Iaiani, Alessandro Tugnoli, Valerio Cozzani

*LISES – Department of Civil, Chemical, Environmental, and Materials Engineering,
Alma Mater Studiorum – University of Bologna, Bologna, Italy*

Keywords: past incident analysis, Bayesian network, methodology, safety, security, incident dataset, reference scenario

The storage and processing of large quantities of hazardous materials in chemical, process, and Oil&Gas industries can lead to significant incidents such as releases, explosions, and fires, resulting in severe consequences for human life, the environment, and the assets (Mannan, 2012). These events can be triggered by internal system-related factors (safety) or intentional attacks (security), and their proper identification and quantitative assessment is of paramount importance to enhance safety and security and to prevent potential disasters (Iaiani et al., 2023b).

A common approach to gather valuable insights supporting existing qualitative and quantitative risk analysis procedures (e.g., safety QRA studies, security vulnerability assessment methodologies) is the analysis of past incidents (PIA) that occurred in similar facilities (e.g., belonging to the same industrial sector) (Abdolhamidzadeh et al., 2011). In fact, PIA can provide reference scenarios (chain of event from the origin of the risk to the final outcomes suffered by the affected facilities) that can be used by practitioners and authorities as basis to undertake case-specific assessments (Iaiani et al., 2022).

Exploratory Data Analysis (EDA) is conventionally employed in PIA; however, it has limitations in systematically analysing incident datasets and prioritizing relevant incidental chains (Iaiani et al., 2023a). To overcome these limitations, the present study proposes a Bayesian Network modelling-based methodology, alternative to canonical EDA, for the systematic identification of the most relevant incidental chains, that serve as reference scenarios, from safety/security incident datasets.

The flowchart of the proposed 3-steps methodology is shown in Figure 1 (a).

Step 1 involves creating a qualitative BN reflecting the layered scheme of the incidental chain of interest: for example, Figure 1 (b) shows a generic BN corresponding to a 3-layers incidental chain (e.g., in the security domain they may correspond to the attack modes, the physical scenarios generated by the attack modes, and the consequences suffered by targeted facilities). Depending on the structure of the dataset analysed, this process may include an arrangement of the information collected in the entries. Parent-child connections in the BN shall represent the dependencies (e.g., causal relationships) among the classes belonging to the different layers.

Step 2 consists in the quantification of the developed BN by populating Conditional Probability Tables (CPTs) with conditional probabilities derived from statistical analysis of the dataset. The Noisy-OR gate model is suggested to address the issue of data scarcity (Oniśko et al., 2001): the model is used to predict the conditional probabilities referred to combinations of parent nodes occurring at the same time (information typically not contained in the dataset due to the rarity of these events) using as inputs the conditional probabilities referred to each parent node occurring alone (information typically contained in the dataset).

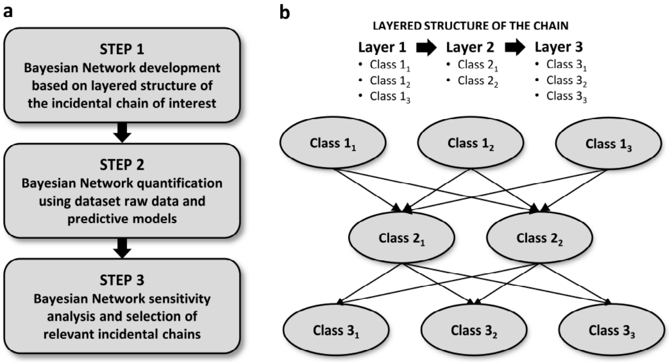
a
STEP 1
Bayesian Network development based on layered structure of the incidental chain of interest
STEP 2
Bayesian Network quantification using dataset raw data and predictive models
STEP 3
Bayesian Network sensitivity analysis and selection of relevant incidental chains
b
LAYERED STRUCTURE OF THE CHAIN
Layer 1
• Class 1₁
• Class 1₂
• Class 1₃
Layer 2
• Class 2₁
• Class 2₂
Layer 3
• Class 3₁
• Class 3₂
• Class 3₃
Class 1₁
Class 1₂
Class 1₃
Class 2₁
Class 2₂
Class 3₁
Class 3₂
Class 3₃

Fig. 1. (a) flowchart of the proposed methodology; (b) example of generic BN referred to a 3-layers incidental chain.

In Step 3, sensitivity analysis of the quantified BN is performed. In particular, the nodes corresponding to the terminal layer of the incidental chain of interest shall be set as the target nodes in the sensitivity analysis (i.e., the three nodes corresponding to layer 3 in Figure 1 (b) that could be the ones corresponding to the consequences in the security domain) and the sensitivity parameter D shall be systematically calculated for each possible incidental chain ending with them. The latter is defined as the first derivative of the posterior probability of interest of a given state of the target node over a specific numerical parameter in the BN which is the combined set of conditional probabilities of the states of the other nodes that are part of the incidental chain under assessment.

Step 3 provides also for the selection of the relevant incidental chains: this is done based on a threshold-based criterion for the sensitivity parameter D. In fact, according to its definition, the higher D, the higher the influence of the upstream part of the incidental chain (i.e., the part of the chain without the terminal layer) on the node set as target (i.e., the node corresponding to the terminal layer), and thus the more relevant the entire chain is.

A systematic step-by-step methodology was developed to identify relevant chains in incidental events using Bayesian Network (BN) analysis of historical datasets. This approach, customizable for safety and security datasets, particularly addresses systematicity issues in chain identification and prioritization, as well as data scarcity issues that are common in small datasets related to rare events. The latter issue is addressed by employing the Noisy-OR gate model in the BN, while the first is addressed by the application of BN sensitivity analysis. The relevant incidental chains are selected based on a derivative-based quantitative parameter, called sensitivity parameter D, which offers a more dynamic insight into the relationship between variables (conditional probabilities in this study) with respect to parameters reliant on simple distribution of variables as those typically adopted in canonical statistical analyses such as EDA.

Future developments are aimed at the application of the proposed methodology to datasets collecting safety/security-related incidents in order to obtain reference scenarios that can support application of existing risk assessment studies.

Acknowledgements

This work was supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union – NextGenerationEU.

References

Abdolhamidzadeh, B., Abbasi, T., Rashtchian, D., Abbasi, S.A. 2011. Domino effect in process-industry accidents – An inventory of past events and identification of some patterns. J Loss Prev Process Ind 24, 575–593.

Iaiani, M., Tugnoli, A., Cozzani, V. 2022. Identification of reference scenarios for security attacks to the process industry. Process Safety and Environmental Protection 161, 334–356.

Iaiani, M., Tugnoli, A., Cozzani, V. 2023a. Identification of cyber-risks for the control and safety instrumented systems: a synergic framework for the process industry. Process Safety and Environmental Protection 172, 69–82.

Iaiani, M., Tugnoli, A., Cozzani, V., Reniers, G., Yang, M. 2023b. A Bayesian-network approach for assessing the probability of success of physical security attacks to offshore Oil&Gas facilities. Ocean Engineering 273, 114010.

Mannan, S. 2012. Lees' Loss Prevention in the Process Industries: Hazard Identification, Assessment and Control, 4th ed. Elsevier, UK: Butterworth-Heinemann.

Oniśko, A., Druzdzel, M.J., Wasyluk, H. 2001. Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. International Journal of Approximate Reasoning 27, 165–182.