Safety Science 104 (2018) 119–134

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Review

Modeling accident scenarios from databases with missing data: A probabilistic approach for safety-related systems design

Reinaldo Squillante Jr[a,*], Diolino J. Santos Fo[a], Newton Maruyama[a], Fabrício Junqueira[a], Lucas A. Moscato[a], Francisco Y. Nakamoto[b], Paulo E. Miyagi[a], Jun Okamoto Jr[a]

[a] Universidade de São Paulo, São Paulo, Brazil
[b] Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, São Paulo, Brazil

ARTICLE INFO

Keywords:
Modeling accident scenarios
Critical events/fault
Databases with missing data
Multivariate imputation by chained equations
Bayesian network learning
Safe-diagnosability property

ABSTRACT

This paper proposes a new probabilistic approach for safety-related systems design based on modeling accident scenarios with databases where missing data is a major concern. The method is based on (1) multivariate imputation by chained equations that addresses the problem of missing data in industrial databases, (2) Bayesian network learning approach that addresses the synthesis of the Fault tree (FT) and the Event tree (ET) diagrams and (3) the *bowtie* diagram that addresses the synthesis of the complete accident scenarios. An experimental application example of a database with missing data related to an accident that occurred in the BP Texas city refinery illustrates the effectiveness of the method. At first, a complete database is submitted to random data extractions and then the proposed method is applied. Experimental results confirm that it is possible to identify the relationship among observed and partially observed critical/undesirable events related to the critical faults even in conditions of missing data. The method might be used for the design of safety-related systems as it is able to support: (1) IEC 61511 and IEC 61508 standards, (2) uncertainty of databases with missing data; and (3) ensuring safe-diagnosability property regarding dynamical aspects of actual systems.

1. Introduction

Safety in the process industries is an important issue that has been receiving attention by the scientific community with increased focus. The reasons are the occurrence of accidents, which require a revision of current engineering practices, such as, process design, process control, risk analysis and control, and risk assessment (Florea & Dobrescu, 2011; Squillante Jr et al., 2015). Accident is the term often used for the occurrence of a single critical event or a sequence of initiator and critical events that causes undesired and/or outcome events. These outcome events may be injury or death, environmental damage and/or property damage (Ferdous et al., 2013; Sklet, 2004). Furthermore, these potential consequences, environmental damage, and financial penalties, are in turn, together with ethical and moral considerations, strong incentives for researches in system safety to prevent such accidents from occurring in the future (Saleh et al., 2010). Therefore, the identification of possible accident scenarios is a key-point in risk analysis and safety-related systems design (Sklet, 2004; Badreddine & Ben Amor, 2010; Squillante Jr et al., 2015).

According to Bakolas & Saleh (2011) when the state of the system is not suitably measured or available, and state estimation techniques are not employed, the system may transit to a hazardous state that remains hidden from the human operators/observers. A system usually assumes a latent hazardous state as a consequence of unobserved events or incidents during system operation, which in conjunction with other factors, can precipitate an accident or aggravate its consequences. These are commonly referred as *pathogenic accident*. Furthermore, several accidents reports identified hidden critical events or unobserved ones as the important contributing factors to the accidents. Actually there is a limitation in the design of safety-related system related to unobserved and/or partially observed critical events that is fundamental to avoid a *pathogenic accident*.

Based on these reasons, this work argues that the design of safety-related systems for process industries must consider the following issues: (*i*) the observability and safety-diagnosability of all observed and unobserved critical and/or undesired events, (*ii*) traceability of all initiator and critical events that were observed and/or unobserved before the occurrence of a topic event (prevention approach); and (*iii*) traceability of all undesired and outcome events arising after the occurrence of a topic event (mitigation approach). In order to address these issues

* Corresponding author.
E-mail addresses: reinaldo.squillante@usp.br (R. Squillante), diolinos@usp.br (D.J. Santos Fo), maruyama@usp.br (N. Maruyama), fabri@usp.br (F. Junqueira), lamoscat@usp.br (L.A. Moscato), nakamoto@ifsp.edu.br (F.Y. Nakamoto), pemiyagi@usp.br (P.E. Miyagi), jokamoto@usp.br (J. Okamoto).

https://doi.org/10.1016/j.ssci.2018.01.001
Received 28 November 2016; Received in revised form 13 December 2017; Accepted 2 January 2018

the safety-related system design from modeling of the accidents scenarios is fundamental.

According to Badreddine & Ben Amor (2013), several techniques have been proposed to modeling the accident scenario of a given risk. A comparison among these techniques is presented in Nivolianitou et al. (2006). The *bowtie* method has become popular in high hazard industries like oil & gas, aviation and mining (Ruijter & Guldenmund, 2016). The *bowtie* diagrams have proven their efficiency in several applications, such as, accident risk assessment, risk management, and safety barrier implementation. Furthermore, Badreddine & Ben Amor (2013) argue the weakness of *bowtie* diagrams is that they are restricted to a graphical representation of different scenarios, exclusively designed by knowledge of experts, that might ignore the dynamic aspect of real systems. They proposed an approach based on Bayesian belief network to construct *bowtie* diagrams that are induced by a complete database. Although these approaches allows the enhancement of the classical *bowtie* diagrams with a numerical estimation for probabilistic decision-making, the question of missing data is not considered and is fundamental from the following points of view: (i) in actual databases, data are unavoidably incomplete (Lakshminarayan et al., 1999), and (ii) given a risk, the determination of relationship among observed and unobserved critical events are fundamental to avoid *pathogenic accidents*. Some papers dealing with missing data or rare events have been proposed. The paper of El-Gheriani et al. (2016) addresses the methodology that uses Hierarchical Bayesian Analysis (HBA) to treat source to source uncertainty and assists in making better estimates of the risk probabilities when dealing with sparse data. Other relevant works dealing with sparse data are referenced in Yan & Haimes (2010), Yang (2013, 2015a, 2015b). Although these papers have a contribution, they are oriented to databases obtained from gathering data from different sources with dissimilar characteristics, such as, different operational conditions, regions, industry sectors or different experts.

Furthermore some papers have also been found in the literature that deal with the use of Bayesian belief networks for the modeling of risks of plants/industrial processes from historical data. The paper of Ale et al. (2013) addresses the development of a dynamic integrated model for risk at the environment process under the point of view of modeling of human activities especially collectively rather than individual human errors. In this context, this approach deals with the use of BBN to Modeling risks based on uncertainties of group of human behavior. Moreover, the paper of Gulijk et al. (2015) introduces the PLATYPUS software that is intended to integrate both technical and human performance for the prevention of process leak and it models the left-hand-side of the *bowtie* diagram (e.g.: FT), and finally the paper of Gulijk et al. (2014) describes the design of a model for calculating the left-hand-side of the *bowtie* diagram for chemical plants. This model is based on Non-Parametric BBN so that uncertainties are automatically included, and it also found that design process for the components of the model is traceable and reproducible.

These approaches have some advantages, for example: (a) they address technical failures and also human error activities in a collective domain such as causes of accidents/disasters, (b) they make use of specially developed PLATYPUS software for the integration of technical and error human as the causes for modeling the left side (e.g.: FT) of the bowtie diagram, (c) they make use of historical data for the modeling process, and (d) they make use of BBN to consider uncertainty in the models generated. However some problems that these methods do not address are described as: (a) they do not address the modeling of the right side (e.g.: ET) of the *bowtie* diagram, that is fundamental for decision making and mitigation actions (e.g.: mitigation barriers) of accidents/disasters, (b) they do not address the existence of missing data in the historical data and this is a problem because most of the industrial historical databases have missing data and this is a unavoidable problem, and (c) given a risk, they do not address the determination of relationship among observed and unobserved critical events that are fundamental to avoid *pathogenic accidents*.

Other relevant works related on application of *bowtie* diagram to BBN for risk analysis perspective can be found in Khakzad et al. (2012), Ferdous et al. (2013), Abimbola et al. (2015), Yang (2015a, 2015b).

In this context, this work initially proposes a new method that deals with the problem of modeling complete accident scenarios from databases that contains missing data and could be use in cases where it is not possible gathering data from different sources with dissimilar characteristics, such as, different operational conditions, regions or industry sectors. These missing data might be related to initiator/critical events and/or undesired/outcome events partially observed and recorded by sensors in the specific plant. In the proposed method the concept of the BBN as a probabilistic directed acyclic graphical (DAG) model, have structural estimated from induced complete data. The induction process is realized via multivariate imputation by chained equations (MICE) and BBN learning techniques. The method shall support designers of safety-related systems in order to construct decision-making systems that can ensure the safe-diagnosability property.[1]

This paper is organized as follows: Section 1 presents arguments for the development of a new method that deals the problem of modeling complete accident scenarios from databases that contain missing data as observed and partially observed initiators/critical and/or undesired outcome. Section 2 presents fundamental concepts, such as, safety-related systems and model-V, *bowtie* diagrams, Bayesian belief Networks and some learning approaches, and imputation of missing data via multivariate imputation by chained equations (MICE) algorithm. Section 3 presents a method for modeling complete accident scenario. Section 4 presents an application example where the BP Texas city refinery accident is utilized to validate the proposed approach. Section 5 presents the analysis of the results. Finally, Section 6 draws some conclusions and discusses possible future developments.

2. Fundamental concepts

2.1. Safety-related systems

According to IEC 61508 (2010) – part 4, safety-related systems implement the required safety functions necessary to achieve or maintain a safe state for plant/process in the critical systems. Furthermore, a safety-related system may: (*a*) be designed to prevent a hazardous event (e.g. if the safety-related systems perform their safety functions then no harmful event arises); (*b*) be designed to mitigate the effects of a harmful event, thereby reducing the risk by reducing the consequences; and (*c*) be designed to achieve a combination of (*a*) and (*b*). Moreover, safety function is the function to be implemented by safety-related system or other risk reduction measure that is intended to achieve or maintain the safe state. Additionally, safety instrumented systems (SIS) is a practical engineering solution to implement a safety-related system based on programmable electronic systems (PES) (e.g. safety programmable logical controllers). The IEC standards recommend layers for risk reduction based on safety-related systems organized hierarchically in order to manage risks, preventing or mitigating critical faults, or bringing the process to a safe state. In this sense, some safety standards for process industries, such as the IEC 61508 (2010), and the IEC 61511(2003a, 2003b), guide different activities related to a SIS, such as design, installation, operation, maintenance, tests, etc. One of the activities related to a SIS, is the development of the program for safety-related control of plant/process. This program is ran by programmable electronic systems (PES), and should be elaborated on the basis of the program life cycle, also called model V and which is illustrated in Fig. 2.1. This program life cycle is composed

[1] A system is said to have the property of safe-diagnosability, if it is diagnosable and it is possible to detect a critical/undesired event or fault to occur before the execution of a given set of prohibitive sequences that violate safety requirements (Paoli & Lafortune, 2005; Squillante Jr et al., 2015).