Computers and Chemical Engineering 93 (2016) 128–142

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Multivariate probabilistic safety analysis of process facilities using the Copula Bayesian Network model

Seyed Javad Hashemi, Faisal Khan*, Salim Ahmed

Centre for Risk, Integrity and Safety Engineering (C-RISE), Faculty of Engineering & Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada

ARTICLE INFO

Article history:
Received 18 February 2016
Received in revised form 20 May 2016
Accepted 18 June 2016
Available online 23 June 2016

Keywords:
Correlation
Dependence structure
Multivariate probabilistic model
Akaike's information criterion

ABSTRACT

Integrated safety analysis of hazardous process facilities calls for an understanding of both stochastic and topological dependencies, going beyond traditional Bayesian Network (BN) analysis to study cause-effect relationships among major risk factors. This paper presents a novel model based on the Copula Bayesian Network (CBN) for multivariate safety analysis of process systems. The innovation of the proposed CBN model is in integrating the advantage of copula functions in modelling complex dependence structures with the cause-effect relationship reasoning of process variables using BNs. This offers a great flexibility in probabilistic analysis of individual risk factors while considering their uncertainty and stochastic dependence. Methods based on maximum likelihood evaluation and information theory are presented to learn the structure of CBN models. The superior performance of the CBN model and its advantages compared to traditional BN models are demonstrated by application to an offshore managed pressure drilling case study.

1. Introduction

Process safety and risk assessment are often multidimensional and hence require the study of several potentially correlated random variables from different risk sources. Consequently, risk practitioners usually deal with complex process systems with multiple correlated variables rather than considering independent risk factors. Looking for relationships among variables is an essential part of process safety analysis to understand the system, identify the cause(s) of process symptoms, predict abnormal conditions and protect systems from catastrophic events. There are some recent works extracting and analyzing interrelationships among random variables in the context of process facilities (Hashemi et al., 2015a; Mohseni Ahooyi et al., 2014a; Yu et al., 2015a). However, development of a tool to simultaneously capture different aspects of variables' interrelationships (including causality and dependency) in complex systems with high dimensionality is still a formidable challenge.

In recent years, Bayesian Network (BN) analysis has been used in process safety analysis mainly through multivariate probabilistic analysis (Mohseni Ahooyi et al., 2014a) and probability updating (Oktem et al., 2013). However, BN analysis has some restrictions from the multivariate analysis perspective, which are mainly lack of control of the marginal distribution of variables and inability to capture the non-linear dependence structure (Mohseni Ahooyi et al., 2014a).

To address the limiting properties of BN analysis, Elidan (2010) proposed the Copula Bayesian Network (CBN) that fuses the frameworks of the statistical copula and BNs. Copulas allow the modelling of complex real-valued distributions by separating the choice of the univariate marginal distributions and the dependence function that "couples" them into a coherent joint distribution (Nelsen, 2006). In contrast to BN analysis that uses conditional probability distributions to define a joint density, in the CBN model a collection of local copula functions is used to capture the direct dependence among system variables (Elidan, 2010).

The objective of this work is to address the limitations of traditional BN analysis by adopting the concept of the CBN model for application in multivariate probabilistic analysis of abnormal conditions in process facilities. The contribution of this work is twofold. Firstly, by using the language of probabilistic graphical models, this work applies copula functions to extend the BN applications in the context of process facilities' safety analysis of higher dimensions. Secondly, a learning mechanism based on a combination of maximum likelihood evaluation and information theory is introduced to address the issue of structure learning for CBN models.

* Corresponding author.
E-mail address: fikhan@mun.ca (F. Khan).

http://dx.doi.org/10.1016/j.compchemeng.2016.06.011
0098-1354/

The rest of the paper is organized as follows. In Section 2 a comparison of the complementing properties of BN analysis and copula functions is provided, followed by a discussion of the interesting synergy that can be achieved by combining copulas and BNs. The proposed CBN model for process safety analysis is provided in Section 3. The practical application of the proposed model is demonstrated using a case study in Section 4, followed by some concluding remarks.

2. Preliminaries

2.1. Inter-relationships of process variables

Multivariate probabilistic process safety analysis requires identification of the inter-relationships among process variables. Connectivity, causality, and correlation are three important attributes which are used to describe such inter-relationships (Yang et al., 2014). The illustrative example in Fig. 1, adopted from Yang et al. (2014), is provided to facilitate better explanation of the physical interpretation and practical use of these different concepts. As shown in Fig. 1, a liquid hydrocarbon feed stream is fed into a distillation column. Following the principle of fluid dynamics, the feed flow rate (F_1) influences the liquid level in the column (L) and L influences the bottom product flow rate (F_3). In terms of the information flow path, the signal line is connected to valve V_2 to transmit the level signal L to the control valve. This connectivity is shown in Fig. 1b. However, F_3 also influences L, which is different from connectivity. In fact, valve V_2 controls the flow rate F_3 based on the signals transmitted from the level meter to the control valve. The same causality relationship also exists for the overhead flow rate (F_2) and the column top pressure P (Fig. 1b). Thus, causality does not exist without connectivity.

To describe the concept of correlation, consider the flow rate F_1 which affects both flow rates F_2 and F_3. Intuitively, F_2 and F_3 are correlated, which can be shown by investigating the process data. However, there is no causality between F_2 and F_3, since by ruling out their common cause, F_1, F_2 and F_3 are both independent.

From the illustration above it can be concluded that correlation is a necessary (not sufficient) condition for causality. Beside correlation, some additional conditions are required to imply a causal relationship, such as connectivity, responsiveness, and the direction of the relationship between two process variables (Yang et al., 2014). Therefore, different tools are required to capture causality and dependency among variables.

Process knowledge can be used to capture causality using common methods such as structural equation models (SEM), graphical models, and rule–based models (Yang et al., 2014). However, as reliable process knowledge is not always available, it is also important to explore capturing causality from process data. Lag-based methods, such as Granger causality and transfer entropy, and conditional independence methods, such as BNs, are widely used approaches to capture causality from process data. Linear relationships among process variables and stationary data time series are restrictive assumptions of Granger causality and transfer entropy, respectively (Yang et al., 2014). Application of BN analysis is of main interest in this work to capture causality due to its ability to represent intuitive cause-effect relationships among process variables, as well as several other modelling advantages as discussed in Section 2.2.

To measure the dependency, the common approach in process facilities literature has been the application of correlation coefficients. The Pearson ρ for linear relationships and rank correlation coefficients (such as Spearman's ρ_s and Kendall's τ) for nonlinear relationships are the frequently used correlation coefficients (Hashemi et al., 2015e). However, given the complexity of relationships among process variables, dependence can be quantified in more sophisticated ways than merely through these numeric coefficients. Copula functions, sometimes referred to as "dependency functions", contain all of the dependence information among random variables (Klaus, 2012). Using copulas, the dependence pattern of the random variables and their individual behaviours (more precisely, their marginal probability distributions) can be studied separately.

Beyond modelling dependency and causality, copulas and BNs are both widely used in the literature to provide a framework for modelling multivariate distributions. In the subsequent subsections, a brief review of advantages and shortcomings of both approaches is provided first. Then, the potential synergy from the integration of copulas and BNs to allow simultaneous modeling of stochastic and topological dependencies among process variables is discussed.

2.2. Bayesian networks (BNs)

BN analysis offers a general framework for analyzing causal influences and constructing multivariate distributions. Basically, BNs are probabilistic networks which rely on Bayes' theorem to draw inferences based on prior evidence (Pearl, 1986). A BN can be defined as a directed acyclic graph (DAG) associated with a joint probability distribution (Mittnik and Starobinskaya, 2010). BNs' main application in process safety analysis is as an inference engine for updating the prior occurrence probability of events given new information (Bauer, 2013; Pariyani et al., 2012). This advantage addresses one of the main shortcomings of the traditional fault tree, event tree, and bowtie safety analysis methods. However, despite the broad scope of applicability, the following shortcomings are identified for BN applications:

i Deterministic point-based probability values are used in most BN applications, ignoring the uncertainty associated with probability estimations.
ii To tackle the above shortcoming, Gaussian distribution has been used as the marginal distribution in some applications. However, there is no doubt that the assumption of joint normality fails to yield suitable models in many applications. Aside from the case of the normal distribution, application of other probability distributions for marginal distributions is not practical due to the limitations of the BN structure (Elidan, 2010).
iii Constructing conditional probability tables (CPTs) to describe the strength relationships quickly becomes very complex and difficult to compute as the number of parents and states increases (Mohseni Ahooyi et al., 2014a).
iv Furthermore, representation of the dependence structure is simply limited to the definition of nodes' relationships using CPTs. Therefore, BN models fail to model complex non-linear dependencies.

There have been some recent developments to improve the practical application of BNs, such as the application of multinomial likelihood functions (Khakzad et al., 2014) and nonlinear Gaussian belief networks (Yu et al., 2015b) to model non-linear interactions, and the application of object-oriented BN (Kjaerullf and Madsen, 2008) and Noisy-OR technique (Bobbio et al., 2001) to simplify the analysis of complex networks. Although such developments have enhanced the practical implementation of the BNs, the limitations mentioned above still exist.

2.3. Copulas

An alternate and markedly different approach for constructing multivariate distributions is the application of copula functions to link univariate marginal distributions. Let $\boldsymbol{U} = (U_i)$, $i \in \{1, \ldots, d\}$ and