Safety Science 50 (2012) 1049–1055

Contents lists available at SciVerse ScienceDirect

Safety Science

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

Analysis of factors that influence hazardous material transportation accidents based on Bayesian networks: A case study in China

Laijun Zhao [a], Xulei Wang [a,b,*], Ying Qian [a]

[a] Shanghai University, School of Management, Shanghai 200444, PR China

[b] Qingdao Agriculture University, College of Economics and Management, Qingdao 266109, PR China

ARTICLE INFO

Article history:
Received 24 September 2010
Received in revised form 3 November 2011
Accepted 4 December 2011
Available online 27 December 2011

Keywords:
Hazardous materials (Hazmat)
Transportation
Bayesian networks
Dempster–Shafer evidence theory
Expectation–maximization algorithm

ABSTRACT

In this study, we applied Bayesian networks to prioritize the factors that influence hazardous material (Hazmat) transportation accidents. The Bayesian network structure was built based on expert knowledge using Dempster–Shafer evidence theory, and the structure was modified based on a test for conditional independence. We collected and analyzed 94 cases of Chinese Hazmat transportation accidents to compute the posterior probability of each factor using the expectation–maximization learning algorithm. We found that the three most influential factors in Hazmat transportation accidents were human factors, the transport vehicle and facilities, and packing and loading of the Hazmat. These findings provide an empirically supported theoretical basis for Hazmat transportation corporations to take corrective and preventative measures to reduce the risk of accidents.

1. Introduction

In recent years, many accidents caused by hazardous materials (Hazmat) transportation have resulted in catastrophic losses of human life and damage to the environment in China. Such accidents have attracted increasing attention from the general public and the government, and a growing amount of research about accident prevention has been conducted (Wang et al., 2005; Zhao et al., 2009; Yang et al., 2010). Analysis of the factors that influence Hazmat transportation accidents is an important issue in accident prevention because it can provide transportation corporations with actionable information on the causes of accidents. As a result, corrective and preventative measures can be implemented to exercise greater control over these factors.

The most common methods that are used to analyze the factors that influence Hazmat transportation accidents, such as statistical methods and fault tree analysis, consider all factors to be independent rather than related (Wang et al., 2005; Oggero et al., 2006; Samuel et al., 2009; Trépanier et al., 2009; Zhao et al., 2009; Yang et al., 2010). In reality, multiple factors contribute to a Hazmat transportation accident, and these factors are often interrelated (Bird and Germain, 1990). Thus, it is necessary to consider the interplay of factors when analyzing such accidents. As a tool for studying uncertainty, Bayesian networks combine graph theory with probability theory, and can represent both the uncertainty and the interplay among the variables (Pearl, 1988, 2003). In this approach, probabilistic inference can be conducted to predict the values of some variables based on the observed values of other variables, and the predicted value is referred to as the "posterior probability". The posterior probability serves as a universally sufficient statistic for detection applications because it captures the relationships and interplay among the variables that describe a situation (Pearl, 2003). If we apply Bayesian networks to analyze Hazmat transportation accidents, we can assign the probability of the accident to be 1 (i.e., 100% probability because the accident has already occurred), and the posterior probability of each factor can then represent its influence on the accident. Based on this analysis, we can identify the most important factors that contributed to the accident and find relationships among these factors. The results of this analysis will help transportation corporations take the measures required to reduce the risk of an accident.

A Bayesian network can be constructed manually, (semi-)automatically from the data, or by a combination of a manual and data-driven processes (Kjaerulff and Madsen, 2008). The last approach first develops the structure of the Bayesian network by taking advantage of the knowledge of domain experts, and then learns the parameters of the Bayesian network from a database using a learning algorithm such as the maximum-likelihood estimation algorithm or the expectation–maximization (EM) algorithm. This approach is sufficiently easy that it can be used in practice. Moreover, the network structure can be easily understood and the

* Corresponding author at: Shanghai University, School of Management, 99 Shangda Road, BaoShan District, Shanghai 200444, PR China. Tel.: +86 21 6613 7925; fax: +86 21 6613 4284.

E-mail address: stonewangxu@gmail.com (X. Wang).

0925-7535/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2011.12.003

expert knowledge can be easily incorporated in the model. However, this method has two limitations: expert knowledge is subjective and limited, which results in poor reliability when building the network's structure, and conditional independence among the variables is ignored when the experts confirm the causal relationship among variables. Overcoming these limitations is the key issue when constructing Bayesian networks using this approach.

The remainder of this paper is organized as follows: Section 2 presents a literature review on the analysis of factors that influence Hazmat transportation accidents and Bayesian networks. Section 3 identifies key factors involved in Hazmat transportation accidents and builds a Bayesian network structure based on expert knowledge using Dempster–Shafer (D–S) evidence theory. Moreover, the structure of Bayesian network is modified according to the conditional independence test. Section 4 computes the posterior probability of the factors using the EM algorithm to find the influence of the different factors on transportation accidents. Section 5 analyzes the results and describes their implications for China's Hazmat transportation corporations. Section 6 provides our conclusions.

2. Previous related work

Most previous studies that analyzed the factors responsible for Hazmat transportation accidents used a statistical method. The researchers largely confined themselves to the collection, analysis, and interpretation of data derived from accident reports or an accident database. For example, Oggero et al. (2006) studied 1932 accidents that occurred during the transport of Hazmat by road and rail from the beginning of the 20th century to July 2004, and concluded that the following major factors were responsible for the accidents: external factors such as the weather, human factors such as operator error, and mechanical failures. Wang et al. (2005) studied Hazmat transportation accidents in China and concluded that six groups of factors were responsible for these accidents: human, vehicle, packing, transportation facilities, road conditions, and environmental conditions. Zhao et al. (2009) studied 650 Hazmat transportation accidents from 2005 to 2008 to analyze the influencing factors and found results similar to those of Wang et al. (2005). Although statistical methods can analyze the relationships between an accident and the factors that influence it, they cannot account for the interplay among different factors and fail to reflect the fact that an accident is usually the result of more than one factor (Bird and Germain, 1990).

To avoid this problem, tree-based methods such as fault-tree analysis, event-tree analysis, and cause-consequence analysis are often used to identify the influencing factors (Roland and Moriarty, 1990; Hamada et al., 2004). However, tree-based methods are based on three assumptions: (1) accidents are binary events, (2) events are statistically independent, and (3) the relationship among events can be represented by means of "logical gates" (Singer, 1990; Andrews and Moss, 2002; Rao et al., 2009). These assumptions restrict the use of these methods to static, logic-based modeling. The tree-based methods are therefore not suitable to describe influencing factors with more than two potential states and make it difficult to represent the relationships among factors. For Hazmat transportation accidents, some factors may have more than two states; for example, weather conditions and road conditions have a wide range of possible states. In addition, the relationships among the factors responsible for an accident cannot be easily represented by means of logical gates. Thus, tree-based methods are not suitable for analyzing the factors that influence Hazmat transportation accidents.

A more promising approach involves the use of Bayesian networks, which have previously been applied in accident analysis. For example, Trucco et al. (2008) developed a Bayesian belief network to model the maritime transport system. Their model represented various factors and their mutual influences by means of a set of dependent variables. Marsh and Bearfield (2004) described a method of modeling the organizational causes of accidents using Bayesian networks, and demonstrated how such a model can be used for risk assessment based on examples from a model of the causes of Signals Passed at Danger (SPAD) incidents in the UK railway system. Maglogiannis et al. (2006) developed a Bayesian network model to concisely represent all the interactions among undesirable events in a risk analysis for health information systems, and the Bayesian network model identified and prioritized the most critical events. The structures of the abovementioned Bayesian networks were often developed based on the causal relationships determined by domain experts. However, due to the subjectivity of the opinions of domain experts, the Bayesian network's structure may be inconsistent with the actual situation. Moreover, the domain experts in these studies did not consider the possibility of conditional independence relationships between nodes, which can lead to an incorrect description of the relationships between nodes in the network structure.

3. A Bayesian network structure for Hazmat transportation accidents

For detailed information about Bayesian networks, refer to the work of Pearl (1988, 2003). Here, we present only a basic mathematical description.

Given a directed acyclic graph, $\zeta = (V, E)$, where V denotes a set of nodes and E denotes a set of directed edges, we can describe a joint probability distribution, P, over the set of variables $X = \{X_1, X_2, \ldots, X_n\}$, which can be factorized as follows:

$$P\{X_1, X_2, \ldots X_n\} = \Pi_i P(X_i|\pi[X_i]) \quad (1)$$

where $\pi[X_i]$ denotes the set of parent variables of variable X_i for each node $v \in V$. The nodes in V are in one-to-one correspondence with the variables X_i. For a node without any parent nodes, the conditional probability is the same as the prior probability.

3.1. Define the variables and their structures

Based on a review of the literature (Oggero et al., 2006; Samuel et al., 2009; Trépanier et al., 2009; Zhao et al., 2009) and a survey of transportation corporations, we identified 11 direct and indirect factors related to Hazmat transportation accidents (Tables 1 and 2). According to the theory of accident causation (Bird and Germain, 1990), the root causes of accidents can be grouped as "immediate" or "contributing". The immediate causes are unsafe acts by a worker and unsafe working conditions. The contributing causes include management-related factors, environmental conditions, and the physical or mental condition of the worker. Here, we have defined human factors (H), Hazmat packing and loading (C), and transport vehicles and facilities (T) as direct factors responsible for accidents. Human factors include skill level ($H1$), health ($H2$), and safety awareness ($H3$). Hazmat packing and loading includes packing ($C1$), and loading and unloading ($C2$). Transport vehicles and facilities (T) include transport vehicles ($T1$), protective equipment ($T2$), and maintenance and monitoring ($T3$). H, C, and T are parent direct factors, and $H1$, $H2$, $H3$, $C1$, $C2$, $T1$, $T2$, and $T3$ are child direct factors. The indirect factors are road conditions (R), weather conditions (W), and management (M). Road conditions and weather might increase the probability of accidents through their impact on humans, transport vehicles, and the hazardous material itself. Both humans and materials are subject to management impacts. Therefore, road conditions, weather, and management (which have