Accident Analysis and Prevention 70 (2014) 155–166

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Extracting recurrent scenarios from narrative texts using a Bayesian network: Application to serious occupational accidents with movement disturbance

F. Abdat [a], S. Leclercq [a,*], X. Cuny [b], C. Tissot [c]

[a] INRS – Working Life Department, 1 rue de Morvan, 54500 Vandoeuvre les Nancy, France
[b] CNAM – Honorary Professor of Occupational Hygiene and Safety, 292 rue Saint-Martin, 75003 Paris, France
[c] INRS – Library & Literature Watch Division, 65 boulevard Richard Lenoir, 75011 Paris, France

ARTICLE INFO

Article history:
Received 21 May 2013
Received in revised form 2 April 2014
Accepted 7 April 2014
Available online 25 April 2014

Keywords:
Bayesian network
Recurrent scenarios
Narrative text
Occupational accident with movement disturbance

ABSTRACT

A probabilistic approach has been developed to extract recurrent serious Occupational Accident with Movement Disturbance (OAMD) scenarios from narrative texts within a prevention framework. Relevant data extracted from 143 accounts was initially coded as logical combinations of generic accident factors. A Bayesian Network (BN)-based model was then built for OAMDs using these data and expert knowledge. A data clustering process was subsequently performed to group the OAMDs into similar classes from generic factor occurrence and pattern standpoints. Finally, the Most Probable Explanation (MPE) was evaluated and identified as the associated recurrent scenario for each class. Using this approach, 8 scenarios were extracted to describe 143 OAMDs in the construction and metallurgy sectors. Their recurrent nature is discussed.

Probable generic factor combinations provide a fair representation of particularly serious OAMDs, as described in narrative texts. This work represents a real contribution to raising company awareness of the variety of circumstances, in which these accidents occur, to progressing in the prevention of such accidents and to developing an analysis framework dedicated to this kind of accident.

1. Introduction

Prevention of trips, collisions, slips and other movement disturbances in the workplace represents an undeniable human and financial challenge. Bureau of Labor Statistics (BLS, 2012) data show that, in the USA, these accidents represented about 30% of the 1,181,290 non-fatal occupational accidents (OA) with days lost in 2011. In its 2008 document on the "causes and circumstances of accidents at work in the European Union", the EC states that, among the 3,983,881 non-fatal accidents causing more than 3 lost work days in 2005, 19% were slips, trips, missteps, stumbles without a fall or with a fall on the level (CE, 2008). At companies operating under the French general social security system, slips, trips and other movement disturbances in work situations (excluding working at height) represented 32% of accidents with days lost (213,940 accidents); 34% of accidents with permanent partial disability (13,759 accidents); 35% of lost work days due to temporary disability (13,591,652 days) and 5% of fatal accidents (25 accidents) in 2011 (CNAMTS, 2012).

Analysis of this kind of accident is often limited to factors close to the injury in the accident genesis. However, accident analysis has also revealed explanatory factors distant from the injury, such as equipment usage (Kines, 2003), access system configuration (Leclercq et al., 2007), work system design (Derosier et al., 2008), work organization or safety management (Bentley and Haslam, 2001). Each factor revealed by analyzing an accident is required for its occurrence, irrespective of its position in the accident genesis. Investigating the accident genesis as far upstream of the injury as possible therefore assists prevention in terms of highlighting a maximum number and a variety of levers for action. A diversity of OAMD occurrence circumstances for different activity sectors (Leclercq and Tissot, 2004) has also been observed within a single company (Leclercq and Thouy, 2004). Combinations of factors common to several slips, collisions and other movement disturbances have been empirically identified in all the accidents subject to in-depth analysis at a regional power distribution facility (Leclercq and Thouy, 2004) and at a railroad company (Leclercq et al., 2007). The authors termed each of these combinations a "recurrent scenario".

* Corresponding author. Tel.: +33 383 509 866; fax: +33 383 502 185.
E-mail address: sylvie.leclercq@inrs.fr (S. Leclercq).

http://dx.doi.org/10.1016/j.aap.2014.04.004
0001-4575/

Haslam and Bentley (1999) had already observed that a combination of slippery conditions, use of footwear with worn treads and time-saving behavior was encountered in 50% of slip, trip and fall accidents among postal delivery workers.

Representing accidents by combinations of factors, rather than by isolated factors, allows us to characterize more closely accident-causing situations since an isolated accident factor (congested floor, person running, etc.) is more representative of a usual occupational situation than of an accident. Prevention is thus more a question of controlling factors, whose combination can be harmful, rather than trying to eliminate every risk factor (Monteau, 1997), which would indeed appear illusory in the case of OAMDs. Furthermore, a fact that has contributed to accident occurrence can sometimes only be considered an accident factor within the context of its occurrence. It may, in fact, be a safety-related factor in another context. For example, knowledge of a location is a safety-related factor, when a person anticipates a step (abrupt change in level) at a location where it is unusual (e.g. midway along a corridor). This same knowledge can be an unsafeness-related factor, when there is an unfamiliar obstruction and a person, trusting his/her knowledge of the location, does not notice it. Characterizing an accident by a combination of factors, such as an accident scenario, rather than by an isolated factor therefore allows us to consider contextual information reflecting the accident-causing nature of certain identified factors.

The purpose of this research is to develop serious recurrent OAMD scenarios, which go beyond the exclusively empirical stage of this development process adopted by Leclercq et al. (2007). Our work falls within the framework of a systemic accident model (Hollnagel, 2004), which has proved beyond any doubt its value to OAMD prevention (Bentley, 2009).

Bayesian Network (BN)-based approaches appear better suited to answering this kind of issue. They provide an adequate representation of our pre-processed data, being a set of accident factors combinations built by experts. Each combination is composed of qualitative knowledge (accident factors) and logical links (links between factors in each logical combination). BNs are well adapted to model such data, bridging the gap between different types of knowledge and unifying all available knowledge into a single type of representation. They are capable of apprehending qualitative knowledge, in terms of accident factors, and links through BN structure. They can also apprehend quantitative knowledge, in terms of frequency of accident factor occurrence among data set, through BN parameters, allowing recurrent scenarios extraction. Unlike other methods such as neural network models, regression methods etc., all the parameters in Bayesian networks have an understandable semantic interpretation. This method can therefore combine expert knowledge with data, to build the model. This is particularly useful when the amount of data is small. Moreover, if machine learning techniques are used (with or without expert knowledge) to build the model from a data set, it can be explained in terms that are understandable by domain experts.

BN-based occupational safety studies have been conducted by several authors in recent years. Using coded data, they have analyzed the effect of task performance-related factors in situations involving risks of falling from ladders or equipment such as scaffolding (Martín et al., 2009), the effect of safety climate- and individual experience-related factors on human behavior (Zhou et al., 2008) or the effect of accident factors (Zhao et al., 2012) or working conditions (García-Herrero et al., 2012) on accident occurrence. In the field of road accidents, BNs are increasingly used, e.g. to model and classify accidents according to their injury severity (Simoncic, 2004; Oña et al., 2011) or to predict the number of accidents of different severity (Deublein et al., 2013) or crash in real time (Hossain and Muromachi, 2012). To our knowledge, no BN-based research has investigated a methodology for determining recurrent scenarios as a diagnostic step toward improving occupational safety. This aim requires in-depth analysis of a set of accidents, which can be found in a database whose richest information is contained in narrative texts. Indeed, Lincoln et al. (2004) have shown that narrative text analysis is a useful supplement to traditional epidemiological analyses because it provides qualitative data, usually based on the accident/injury process, which offers a deeper understanding of the underlying accident process. Fatality investigation reports, in particular, contain data elements not routinely analyzed with coded occupational injury surveillance data (Bunn et al., 2008). The issue now is, "Is it possible to extract recurrent scenarios from a set of serious OAMD narrative texts?"

Further studies aimed at understanding accidents based on narrative text have been conducted in recent years. McKenzie et al. (2010) describe recent advances in using this kind of text in injury surveillance research. Narrative texts need to be pre-processed, unlike coded data which can be directly applied within the scope of BN-based approaches. Automatic methods, such as text mining, have been developed to extract clusters of words with a high probability of target category association (Brooks, 2008). However, these methods do not allow accurate identification of accident factors from a narrative text, i.e. facts that make sense in terms of the accident progression. Similarities or identities can effectively be expressed in words with different meanings or, conversely, a similar meaning can be expressed in differently spelt words (McKenzie et al., 2010). These facts can only be extracted, if the whole narrative is considered. To date, most OA analyses based on narrative text have implemented, for example a 'reconstruction template' (Lincoln et al., 2004), a priori-defined generic accident factors (Shibuya et al., 2010) or Haddon's matrix (Bunn et al., 2008) to process information manually. Analysis of these processed data is usually based on the occurrence and co-occurrence of factors or a number of related keywords.

Our aim is to extract combinations of factors common to several accidents, so we need to identify, from narrative texts, both accident generic factors, which have contributed to injury occurrence and how these factors have combined to cause injury. A BN-based approach has been developed to extract such combinations or recurrent scenarios.

2. Method

2.1. Data pre-processing

2.1.1. Data

The OAMD data used in this study were taken from the France's anonymous EPICEA database consolidating more than 18,000 OA cases that have occurred, since 1990, at companies operating within the French general social security system (EPICEA, 2011). EPICEA lists nearly all fatal occupational accidents and some accidents that were serious or significant for prevention. Identifying OAMDs contained in the database is not automatic. In particular, it requires analysis of the narrative text wording and reading of each account prior to its inclusion in the corpus data. Our study concentrated on the construction and metallurgical industries because these are the industrial sectors most affected by occupational accidents. These industries are dynamic and hazardous due to the diverse and complex nature of their work tasks, trades and environments, as well as the temporary and transitory nature of the workplaces and workforces (Kines, 2002). 143 accidents were ultimately extracted from EPICEA database, 79 cases from the construction sector and 64 cases from the metallurgical industries. However, this set is not representative of all OAMDs, so results could not be extrapolated to OAMDs occurring within the French general social security system. The construction and metallurgical industry set does allow us to develop the recurrent scenario extraction methodology and synthesize a set