Journal of Loss Prevention in the Process Industries 32 (2014) 443–453

Contents lists available at ScienceDirect

Journal of Loss Prevention in the Process Industries

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

A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data

Maryam Sadat Hajakbari [a, *], Behrouz Minaei-Bidgoli [b]

[a] Department of Computer Engineering, Qom Branch, Islamic Azad University, Qom, Iran
[b] Department of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 2 June 2014
Received in revised form
21 October 2014
Accepted 22 October 2014
Available online 23 October 2014

Keywords:
Risk assessment
Occupational safety
Data mining

ABSTRACT

The main purpose of reviewing occupational accidents is not only to prevent the accidents but also to minimize the physical and financial damage caused by them. It is obvious that identifying the main risk factors and determining their qualitative and quantitative rates is essential to the implementation of effective safety management. The main purpose of this study is to use the new risk management methods for occupational accidents and improve work safety. Recent research shows that data mining techniques are powerful tools in assessing the risk of occupational accidents; by combining previous methods, such as the coefficients of accident assessment—which are used to assess an organization's safety—and data mining models. This study attempts to present a new risk assessment method for occupational accidents. The results show that the proposed method could identify and evaluate critical points with a high level of accuracy. A practical application of the method is also presented in which the workplaces at most risk in 2010 are identified and the risk level for each category is determined. Then, repeating the procedure for 2011, critical points based on previous data are determined, and a list of workplaces requiring periodic inspections in 2012 is produced.

1. Introduction

Every year, accidents across some industries cause considerable financial, physical, and environmental damage to companies, organizations, and the injured persons and their families. In addition to the direct and obvious damages, there are costs and other consequences, such as losing the head of the family and the psychological trauma suffered by the injured persons, their families, and their relatives. Consequently, it is essential to reduce these risks by examining and analyzing the causes of accidents and identifying risks and hazards in the workplace.

Assessing the risk of occupational accidents is an effective method for improving health and safety, which is vital in industry. According to the International Labor Organization's (ILO) statistics, 250 million accidents happen every year worldwide, and 335,000 people lose their lives (home and car accidents are not considered in these statistics); further, 160 million occupational diseases are seen every year, leading to 1.1 million deaths worldwide. In addition, 4% of the GDP is wasted worldwide in relation to such accidents and diseases. Ignoring work health and safety leads, therefore, to significant economic losses, which has a serious impact on labor productivity. There are three to four times more fatal occupational accidents in developing countries than in developed countries, and these events are usually avoidable; in Iran, about 14,000 occupational accidents occur annually (Dargahi et al., 2012).

Due to the massive volume of data available and the limitations of statistical methods for data analysis, data mining has received a great deal of attention in recent years. Data mining is the process of managing knowledge and extracting it from large databases. Anand et al. (2006) classified the data from the US National Response Center (NRC) database into different groups and, using data mining techniques such as association rules, found interesting patterns in occupational accidents in the American petrochemical industry in terms of the type of equipment involved, the type of chemical released, and the cause of the accident.

Keren et al. (2006) also used the NRC database to try to identify the operational risks in the chemical industries of Harris County, Texas, USA. Again, the association rule data mining tool was used, as

* Corresponding author.
E-mail addresses: m88.hajakbari@gmail.com (M.S. Hajakbari), b_minaei@iust.ac.ir (B. Minaei-Bidgoli).

http://dx.doi.org/10.1016/j.jlp.2014.10.013
0950-4230/

well as lift methodology, to calibrate the failure risk of equipment involved in a chemical process.

Meel et al. (2007), on the other hand, used Bayesian theory to forecast the frequency, relevant causes, equipment involved, and consequences of incidents. This involved the dynamic analysis of incidents in the NRC database in order to design a model for the frequency of incidents in various chemical and petrochemical companies, taking into consideration their cause (e.g., equipment failure, operator error, etc.).

For the period 2000–04, Fabiano et al. (2008) examined the rate of occupational injuries in different Italian industries and compared the working conditions of temporary and permanent employees using injury frequency indices and characteristics of the labor force. The data was derived from the National Organization for Work's injury insurance, which relates to three large manufacturing firms. Then, the results were analyzed using the injury frequency (FI), the characteristics of the labor force, and the accident intensity index (SI). In the most hazardous industrial sectors, the FI and SI for temporary workers were found to be higher than for permanent employees. Evidence was also collected from the responses of injured temporary workers.

Gurbuz et al. (2009) applied data mining methods to incident reports from the Federal Aviation Administration (FAA) database, which records incidents for all categories of aviation in the US between 2000 and 2006. Their aim was to find out the major causes in order to reduce the number of deaths; they used categorization tools to identify relationships and patterns that resulted in deaths. The analysis did reveal the factors that led to deaths, as well as the parameters affecting such incidents, and validity testing was conducted to verify the accuracy and reliability of the results.

A methodology for the development of a model that included the number of lost days due to accidents in a coal mine in Turkey was presented by Sari et al. (2009). In the first step, the frequency and the severity of the accidents were modeled using appropriate statistical distributions; then, the two distributions were combined using the Monte Carlo simulation in order to construct annual relative risk levels; and finally, a simple forecasting model was developed to quantitatively predict the risk levels by using a decomposition technique in a time series analysis.

Hintikka and Saarela (2010) used the European Statistics on Accidents at Work (ESAW) methodology to analyze accidents related to violence, especially those involving women. The database included statistics for Finland's occupational accidents between 2003 and 2006, and indicated that the number of such accidents had increased in recent years (2003–06) due to changes in the labor market and women's working lives.

Applying association rules, Mirabadi and Sharifian (2010) analyzed accident data for the Iran Railway (RAI) and discovered correlations and new patterns. CRISP-DM was the data mining methodology and Clementine 12.0 the software tool used for this study. They selected 6500 records of accidents between 1996 and 2005 from the Iran Railway accidents database (RAI) and examined the conditions and other factors related to each in order to develop and improve preventive regulations and rules: among the most common causes of accidents were human error, wagon, and track.

Papadopoulos et al. (2010) analyzed the factors contributing to the increase in occupational accidents, examining the effect of changes in working hours, years of employment, types of employment contracts, and working conditions on occupational and public health. The results showed that these factors increase worker's fatigue, alter working hour patterns, and affect years of employment. Moreover, job insecurity and occupational stress had a serious impact on workers' health and, as a result, increased occupational accidents. The overall results indicated that the physical, mental, and social health and safety of the labor force must be given more consideration than the retention of their workability.

Peng et al. (2011) designed an effective incident information management system to handle the heterogeneous data that is distributed across multiple databases, which helps decision makers (DMs) using data mining techniques to detect anomalies, extract useful knowledge, and finally assess the risks. This system provides different services to support the requirements of different incident management phases. Validation was undertaken using China's incident data between 1997 and 2001 as a case study, and it demonstrated that the combination of data mining and multiple criteria decision-making (MCDM) methods can fulfill objectives and extensively assess accident risks.

Finally, Conte et al. (2011) designed a general model to recognize and predict accidents in Spain's workplaces. From chronological records, they aimed to identify the real risks, which were then summarized in a contingency table, and the criteria needed for their evaluation and prioritization were calculated by a mathematical-statistical analysis.

As can be seen, several small studies have used data mining techniques to analyze databases of occupational accidents and have proved such techniques can be helpful in revealing important patterns that may not be evident from using other traditional tools (Anand et al., 2006).

Thus, the main purpose of this study is to use new methods for early detection of the level of risks in occupational accidents and therefore, improve workplace safety management. Previous research, particularly in recent years, shows that data mining techniques are strong tools for assessing the risk of occupational accidents, and this study combines data mining models with coefficients of accident assessment that are used in assessing organizational safety. As a result, a new method for assessing the risk of accidents is presented.

The paper is arranged as follows: Section 2 describes the concepts involved; Section 3 explains the steps taken and techniques used in mining the data; Section 4 describes the application of the new scoring system; and Section 5 presents the conclusions as well as recommendations for improving the method in subsequent studies.

2. Basic concepts

2.1. Data mining techniques

In general, data mining techniques can be divided into two categories: directed and undirected, which will now be described.

2.1.1. Directed data mining

In directed data mining there is always a target variable—something to be classified, estimated, or predicted (Berry and Linoff, 2004). The process of building a classifier starts with a predefined set of classes and examples of records that have already been correctly classified. Similarly, the process of building an estimator starts with historical data where the values of the target variable are already known. The modeling task is to find rules that explain the known values of the target variable. In this study, the *k*-means algorithm has been used to group workplaces.

2.1.2. Undirected data mining

In undirected data mining there is no target variable—the data mining task is to find overall patterns that are not tied to any one variable. The most common form of undirected data mining is clustering, which searches for groups of similar records without any instructions about which variables should be considered the most important. Affinity grouping is another example of undirected