Safety Science 83 (2016) 12–22

Contents lists available at ScienceDirect

Safety Science

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

Prediction of occupational risk in the shipbuilding industry using multivariable linear regression and genetic algorithm analysis

V.D. Tsoukalas *, N.G. Fragiadakis

Merchant Marine Academy of Athens, Engineering Department, Paralia Aspropyrgou, 19300 Aspropyrgos, Greece

ARTICLE INFO

Article history:
Received 5 June 2015
Received in revised form 11 October 2015
Accepted 7 November 2015
Available online 26 November 2015

Keywords:
Risk prediction
Occupational accident
Shipbuilding industry
Soft computing techniques
Genetic algorithm

ABSTRACT

In this research, an effective approach based on Multivariable Linear Regression (MVLR) and Genetic Algorithm (GA) methods has been applied to study the effect of working conditions on occupational injury, using data of occupational accidents accumulated by ship repair yards. The work aims at the development of a calculating model that will use soft computing techniques to assess the occupational risk in the working place of shipyards using occupational accidents data. For each accident the following parameters have been considered as the model's input features: day and time, individual's specialty, type of incident, dangerous situation and dangerous actions involved. Reported accident data were used as the training data for the MVLR model to map the relationship between the working conditions and occupational risk. With the fitness function based on this model, genetic algorithms were used for the prediction of occupational risk taking into consideration the severity and the frequency of occupational accidents data accumulated by ship repair yards. The working parameters' values for minimum occupational risk were obtained using GAs. By comparing the predicted values with the reported data, it was demonstrated that the proposed model is a useful and efficient method for predicting the risk of occupational injury.

1. Introduction

Shipbuilding and repair are highly technical and complex processes requiring a number of skilled trades and expertise. Ship construction and repair are among the most hazardous industries in the world, which means that quite complicated tasks have to be performed in parallel. Carelessness of the workers, insufficient safety training and education, unawareness of costs of accidents, erroneous series of human operations and inadequate work site environment remain the key risk factors for occupational accidents. The occupational accidents are followed by costs; namely, injuries, fatalities, material and/or environmental damages. Common causes of occupational accidents are high elevation, toxic, flammable and explosive materials, fire, moving machinery, dangerous gases, work on/close to haphazard established heavy structures, misuse or failure of equipment, poor ergonomics, untidiness, poor illumination, exposure to general hazards including electricity, and inadequate protective clothing (Krstev et al., 2006; Barlas, 2012).

Various descriptive and analytical parametric modeling procedures such as descriptive statistical methods and regression analysis have been applied for the study and classification of occupational injuries. A great number of studies describe the distribution of injuries (numbers, rates, frequency index) usually in terms of person, place and workplace characteristics, which are useful for identifying hazardous industries, occupations and work situations (Yorio and Wachter, 2014; Salminen, 2005; Chi et al., 2013; Trontin and Bejean, 2004). Researchers have often preferred logistic regression when analyzing data from a case-control study since they can easily quantify the results in the form of odds ratio (OR) and associated confidence intervals. Regression analysis has been used to evaluate the relationship between the injury frequency index and one or more covariates or predictor variables (Ciarapica and Giacchetta, 2008). Roudsari and Ghodsi (2005) adopted logistic regression, while Fabiano et al. (2004), Siu et al. (2004) and Smith et al. (2005) used the statistical correlation and significance study with the evaluation of *P*-value, odds ratio and likelihood ratio test to support the results. Poisson distribution is widely employed as an analytical tool in safety analysis and reliability engineering and it is also frequently used in epidemiological medical studies. Examples of its application in occupational safety studies may be found in Jensen and Laursen (2014), Wang et al. (2003) and Yau et al. (2004).

Since the logistic regression techniques may not yield high prediction accuracy, many non-parametric models like artificial neural

* Corresponding author.
E-mail addresses: vdtsoukalas@yahoo.com (V.D. Tsoukalas), nikolasf2002@yahoo.gr (N.G. Fragiadakis).

http://dx.doi.org/10.1016/j.ssci.2015.11.010
0925-7535/

network models, fuzzy logic and adaptive neuro-fuzzy inference system have been employed for occupational risk analysis. Artificial neural networks (ANN) which don't require any pre-defined underlying relationship between the dependent and independent variables have proved to be a powerful tool, especially when dealing with the problems of classification and prediction. There have been relatively few applications of artificial neural networks (ANN) for analyzing safety issues (Ciarapica and Giacchetta, 2008) and some of them have been specifically addressed at road accidents (Abdel-Aty and Pande, 2005; Chang, 2005). Fuzzy logic is a powerful tool for facing with uncertainty and solving problems where there are no sharp boundaries and precise values. This method has been used to solve different aspects of risk problems (Jamshidi et al., 2013; Bajpai et al., 2010). Adaptive neuro-fuzzy inference system (ANFIS) is a fuzzy inference system implemented in the framework of adaptive networks. By using a hybrid learning procedure, ANFIS can construct an input–output mapping based on both human knowledge (in the form of fuzzy if-then rules) and stipulated input–output data pairs. Reports by Ciarapica and Giacchetta (2008), and Fragiadakis et al. (2014) show the supremacy of the fuzzy neural network over the conventional artificial neural network.

Recently, the Genetic Algorithm (GA) approach has been applied to solve safety design problems in transportation research (Carter et al., 2005; Khosravi et al., 2011) or to assess the work zone casualty risk (Meng and Weng, 2011). It should be pointed out that the GA approach has been widely used in the fields of business, medicine and manufacture (Khosravi et al., 2011; Ren et al., 2014; Tsoukalas, 2008). In these works genetic algorithms were used to sweep a region of interest and select the optimal (or near optimal) settings to a process. The GA is a global optimization algorithm, and the objective function does not need to be differentiable. This allows the algorithm to be used in solving difficult problems, such as multimodal, discontinuous, or noisy systems. The GA approach has been shown to outperform logistic regression analysis, classification tree models, fuzzy logic and artificial neural network models in terms of the prediction accuracy (Dehuri and Mall, 2006).

However, although the GA provides higher prediction accuracy it fails to determine the marginal impacts of the working parameters on the occupational risk. For this reason, in this study the GA technique has been integrated with regression analysis to predict more accurately the occupational risk in the shipbuilding industry.

The separate evaluation of occurrence frequency and severity of an occupational accident cannot completely reflect an occupational risk, rendered by a broad range of accidents from frequent minor to rare major. Therefore, the focus of this work is to develop a calculating model that will use MVLR–GA methodology to predict the overall safety level in the shipbuilding industry. The proposed methodology includes the following steps: (i) Definition of four injury severity levels, (ii) calculation of an injury frequency index for the statistically selected working conditions, defined as the number of times the same level of injury has occurred under the same incident parameter, (iii) calculation of a Normalized Risk Index from the injury frequency values and the severity levels for each parameter, (iv) development of the Multivariable Linear Regression (MVLR) and Genetic Algorithm (GA) model for the calculation of occupational risk in the workplace of shipyards using as input values the NRIes for each parameter. The proposed methodology can learn from historical data in order to predict the occupational risk in any working place. Genetic Algorithm (GA) efficiently exploits useful information contained in a population of solutions to generate new solutions with better performance.

2. Parameters affecting occupational injury

Occupational risk is characterized by two indices, one that indicates the frequency of the injuries and another that refers to the severity of the injury. These two indices have been taken into account in the study of 294 Reports of Occupational Accidents collected by the Greek Labour Inspectorate Agency and included in its Records. These records correspond to accidents that have occurred in four ship repair yards (Perama, Keratsini, Salamina, Drapetsona) and have been reported within two decades, i.e. from 1989 to 2008. The typical form of these data reports include parameters such as year, place, worker's specialty, age, nationality, day and time of the incident, short description of the incident, kind of injury (literature description) as well as six coded parameters. These six parameters include type of accident, material factor involved, type of injury, part of the body been injured, dangerous situation involved in the incident and dangerous actions involved in the incident. A cause and effect diagram (Fig. 1) was constructed to identify the relation between these parameters and occupational risk in ship repair yards.

Since there were no data concerning the degree of permanent disability or workdays lost until complete clinical recovery, nor data concerning the social and economical results of the incident, a mind procedure was used to define the injury severity levels on the basis of experts' opinions and with the aim to take into account the severity attributed to the injuries. This procedure took into consideration the recorded parameters of the small description of the incident, the kind of injury (literature description), the type of injury and the part of the body been injured. According to this procedure four injury severity levels were established, namely negligible wounding, slight wounding, severe wounding and death (see Table 1).

The data, as retrieved from the records, have been statistically processed using SPSS (PASW Statistics ver. 21) in order to define the parameters for the classification of the injury cases and the criteria for calculating the frequency and severity indices. The injury frequency index is produced considering day and time, specialty, type of incident, dangerous situation and dangerous actions involved as the incident parameters. Tables 3 and 4 present the codification list for the incident parameters. These parameters have been defined applying Pearson's correlation criterion (*P*-values) to the four levels of injury severity.

In this study the frequency has been calculated as the number of times the same level of injury has occurred under the same incident parameters. The above-mentioned parameters have been codified according to Records of Occupational Accidents used in the Greek Labour Inspectorate Agency Records. A Normalized Risk Index (NRI) resulted from the injury frequency values and the severity levels for each parameter (see Tables 2–4) and used as input value in order to calculate Risk. A normalization formula was used to estimate the Normalized Risk Index (NRI) of each category concerning the parameters "day and time, specialty, type of accident, dangerous situation involved in the incident and dangerous activity involved in the incident":

$$\mathrm{NRI} = \frac{\sum_i x_i y_i}{3 * \sum_i x_i}, \quad i = 1\text{–}4, \tag{1}$$

where x_i is the percentage for each resulting occupational injury and y_i is the respective severity level according to Table 1. Thus the final value of NRI is scaled from 0 to 1.

The resulting values for each parameter are presented in Tables 2–4. These values of NRI were finally used as input values in the MVLR–GA model. In order to estimate occupational risk using the MVLR–GA model, the total number of 294 data sets were divided into two groups, one for training and the other for testing (Chang