Demichela Micaela
SAfeR- Centro Studi su Sicurezza, Affidabilità e Rischi - Dip. Scienza dei Materiali e Ingegneria Chimica - Politecnico di Torino, Torino, Italy

Murè Salvina
ARIA S.r.l. – Analisi dei Rischi Industriali e Ambientali, Torino, Italy

Cigna Caterina, Monai Laura, Patrucco Mario
Dip. Ingegneria del Territorio, dell'Ambiente e delle Geotecnologie - Politecnico di Torino, Torino, Italy

Occupational accidents data collection and analysis

Keywords

occupational accidents, Fuzzy Logic, Neural networks, expert systems, data collection.

Abstract

Despite of the always growing attention to safety related topics, the enforcement of directives, regulations and technical standards and the improvement of technical solutions aimed to minimize the occupational risks, the number of people dying every day at workplaces is still excessively high. The overall number of injuries is recently decreasing, but both the frequency and the total yearly number of fatalities still remain fundamentally unchanged in the last years. The main problem with accidental data, as officially reported, is that very often, no evaluation is possible in terms of root causes, e.g. standard violations. Since the target of the analysis is the determination of the causal chain of events that lead to the accident to understand how it happened and how to avoid the occurrence of similar situations, the lack of detailed information lead to many difficulties in the definition of the suitable prevention measures. This paper shows three different, but integrated. methods able to collect, manage and analyze the information related to occurred accidents for preventive purposes.

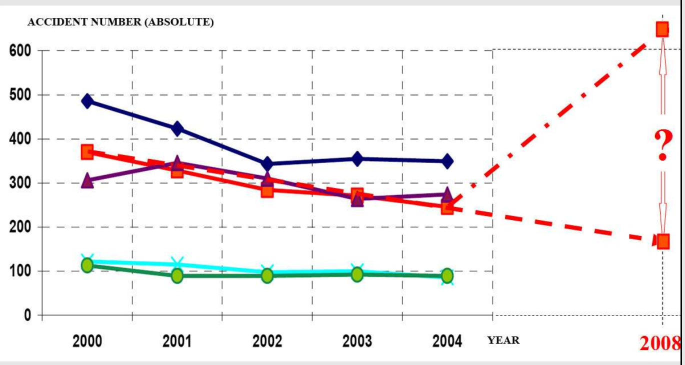
1. Introduction

The Health and Safety of workers at workplaces is still a quite concerning problem worldwide. E.g., , the 2007 Italian data show a rate of 3 fatal accidents / day (INAIL - Istituto Nazionale Assicurazione contro gli Infortuni sul Lavoro - official data for fatal injuries: 1207 fatalities in 2007): it is clear that the only way to reduce such an unacceptable value is to carry out an effective Risk Analysis to preview in detail the critical situations, upon whose results to base strong proactive actions of risk management, and thus correct any safety flaw, keeping in mind that *"one of the best ways to prevent and control occupational injuries, illnesses and fatalities is to 'design out' or minimize hazards and risks early in the design process"* [6], this approach lying at the very base of the Prevention through Design (PtD) way of thinking.

The analysis of a set of non homogeneous or incomplete data can only lead to partial results, as strengthened by many Authors (e.g. [11]), and the problems of a significant consistency of the data base, correct use of the stored data, and difficulties in the evaluation of the main internal and external parameters conditioning each safety situation are often underestimated, this leading to omitted or ineffective risk management actions, poor residual risk management, and inefficient use of the national inspection resources, as e.g. discussed in [1].

Unfortunately, often no official records of necessary detail are available, so that only very general considerations may at present be drawn on this basis (mainly the pointing out of critical industrial sectors), so that, where a specific problem arises, a special and onerous research work is needed: e.g. in [2] and [3] an original approach aimed to point out the criticalities of different types of machines widely used at construction sites is discussed, It is based on a number of fatal accidents case studies, correlation of national and foreign data bases and direct in field data collection.

Possible misuses and misleading of statistical data in not confirmed significant number:
- direct inference of the effectiveness of internal/external safety audits
- forecasting of expectable accident rates



What did it happen?
The disappointing result is three to two possible sets of errors:
1 - reduced statistical basis (region or even province, without due consideration of the local industrial situation);
2 - insufficient analysis of boundary conditioning parameters (e.g. economical situation);
3 – yearly data instead of packaged mean values on 3 or more years.

Some precious info for an effective prevention can however be obtained on special circumstances (e.g. in case of technological modifications).

Figure 1. Possible misuses and misleading of statistical data

Hereafter some discussion is provided on three methodologies and software tools for occupational accidents data collection and analysis developed by the Authors. The research lines that made possible these results were carried on in strict cooperation, and common and different targets carefully identified and discussed, so that in each case, as soon as the context and goal are defined, the selection of the best suitable tool is possible.

2. The integrated methodologies

2.1. Self-organizing maps and clustering algorithms for the analysis of occupational accident databases

It is well known that clustering algorithms are a substantial tool for Data Mining applications ([4], [7], [8]), but they lack of the capability of allowing the analyst to have appropriate insight in the resulting classification. This is mostly important when, like in this case, the data under analysis live in a multidimensional space.
Occupational accidents in available databases are in fact classified accordingly to the multi-parameter ESAW (European Statistics on Accidents at Work) classification, from which, in order to better understand the causes of accidents, only the six qualitative parameters regarding the sequence of events that led the accident were considered and namely:
- the specific activity engaged in during the accident;
- the machinery, the tool or the material involved in the activity;
- the cause that initiated the chain of events leading to the accident;
- the object, substance, or person that generated the source of injury;
- the manner in which the injury was produced or inflicted;
- the object, substance, bodily motion, or exposure that directly produced or inflicted injury.

The advantage of the projection approach is that it allows, the bidimensional visualization of the data set and, therefore, a better understanding of the mutual relationships between the data.
Besides this noteworthy capability, SOM (Self-Organising Maps) performs a first stage of clustering by grouping similar cases in the map prototype vectors, that become in this way a set of protoclusters of the original data.
SOM and clustering algorithms have been originally developed for numerical databases, so they usually employ Euclidean metric. However, Euclidean metric is not suited for categorical data. Thus the ‘Binary-based similarity distances’ were taken into account and among these, the Hamming distance [5] seemed the best choice, because it is a dissimilarity measure, like Euclidean distance for numerical data, and moreover it has proven to give better results for categorical data sets [9].
The normalized Hamming distance, which has range [0, 1] and is defined by equation (1):

$$H(x,y)=\frac{\sum_{i=1}^{n}\delta(x_i,y_i)}{n} \qquad (1)$$

$$\delta(x_i,y_i)=\begin{cases}0 & if \quad x_i=y_i\\ 1 & if \quad x_i\neq y_i\end{cases}$$

where:

- x and y are the two data under comparison;
- n is the number of elements characterizing each data.

The data are shown as a map as in Figure 1, where the colors are linearly changing from blue to red: blue denotes the smallest average distances between each unit and its direct neighbors, and red the largest one.

Therefore, uniform areas in blue indicate neighboring units, red and yellow areas indicate units far from each other. In this way the distances break up the map in zones that divide areas with similar accidental cases.

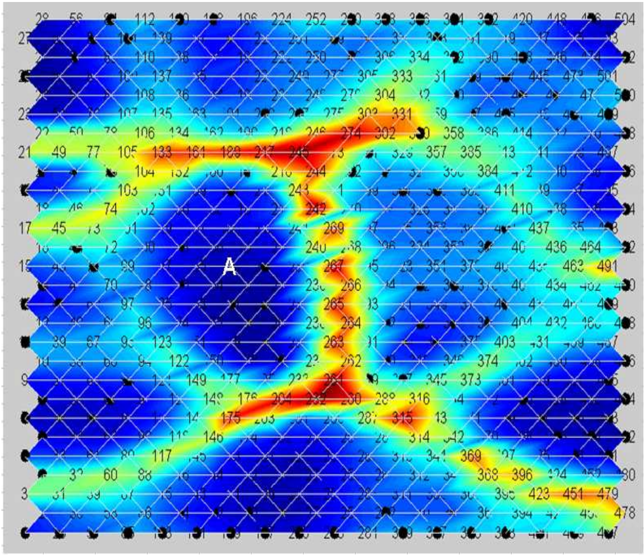


Figure 2. Results from the SOM

Nearly all cells of A region are blue colored, that means that all units classify similar accident data.

The subsequent application of the K-means algorithm to the SOM units splits the data set in crisp partitions. Moreover, this classification is not blind to the analyst, as happens for the direct clustering algorithms, but can be explained in a visually understandable way.

For example, let's consider two accident cases grouped in a single cluster. The binary codes of the accidents are shown in Table 1.

Case 1 describes an accident occurred during a working with a tool: the worker injured because he lost the control of the tool and got in contact with a sharp part of the tool. Case 2 describes instead an accident occurred during the working of a woodwork with a tool: also in this case the worker injured because he lost the control of the woodwork and got in contact with a sharp part of the tool.

The main sequences of events leading to each accident have been identified and grouped in the same classes with an overall accuracy of 92%.

Once clustering has been done and the sequences of events leading to accidents have been identified, a cluster ranking can been performed to evaluate the sequences most critical. For this purpose, a frequency index has been evaluated for each cluster, by dividing the number of cases grouped in the same cluster by the total number of cases in the database under analysis. On the other hand, it has been looked for a seriousness index [12].

In the database analyzed there are no fatal accidents, so it has been looked in each cluster for the number of accidents that have implied losses of limbs.

The original database contained 1700 accident cases occurred in wood processing industry (manufacturing of furniture or building elements) over three years (2002-2004) in northwest Italy. After discarding records with missing values and considering only the secondary wood processing industry (manufacturing of furniture or building elements), 1207 records were retained.

2.2. Fuzzy application procedure

If the neural networks allow analyzing large accident databases, they are not so appropriate for the quantification of risks in the work environment. Given the recognized importance of historical data collection and analysis for prevention purposes, a decisional tool to assess the risk level associated to the exposition of the workers to the generating factors of occupational accidents in the industrial work environment was set up.

Starting from a method based on fuzzy logic approach [13] that allows treating uncertain, qualitative and in some cases contradictory data, as the parameters describing occupational accidents usually are, a procedure was developed able to quantitatively assess the risk of occupational accident in the work environment [10].

The solution of a problem through the fuzzy logic approach can be subdivide in the following operations:

- Definition of the fuzzy set of the input variables and the output ones (fuzzification)
- Definition of the rules that correlate the input variables to the output ones
- Aggregation of the contributions of the rules
- Defuzzification of the results

Table 1. Example of direct inspection of a cluster

	Activity (A)	A Material	Deviation (D)	D Material	Contact (C)	C Material
Case 1	01000000	00010000	00010000	00010000	00000100	00010000
Case 2	01000000	01000000	00010000	01000000	00000100	00010000

Table 2. Fuzzy sets

Frequency	Severity	Contact factor	Protective means	Risk	Improvements
Remote 0-0.3	Very slight 0-0.3	Low 0-0.3	Inadequate 0-0.3	Low 0-0.25	Tolerable risk: no interventions
Low 0.2-0.5	Slight 0.2-0.5	Medium 0.2-0.5	Sufficient 0.2-0.5	Medium 0.26-0.45	Medium risk: Medium interv.
Medium 0.4-0.7	Medium 0.4-0.7	High 0.4-0.7	Good 0.4-0.7	High 0.46-0.75	High risk: heavy interv.
High 0.6-1	Strong 0.6-1	Very high 0.6-1	Very good 0.60-1	Very high 0.76-1	Not tolerable risk: urgent & heavy interv.

The method was validated through several applications performed in a tyre production company, in a chemical plant and in an underground building activity and makes it possible to assess the level of risk the workers are exposed to in each specific sector at present and propose preventive interventions to reduce the risk and assess their efficacy.
In particular, the method allows very different types of parameters (as the frequency of events, the stress factor, the time of exposure, etc.) to be taken into account in risk assessment, highlighting their interactions in originating the accident.
FAP was developed as a support tool in order to obtain a greater efficiency in the analysis and a more user friendly approach. FAP can be used by the only requirement to have an input datasheet, homogeneous, complete and correctly filled, this deriving from company accident databases or other generic databases, given a pre-treatment of the information.
Allows establishing a priority of interventions on the basis of the risk levels assessed.
Being at the present state the risk assessed as a function of:
F– frequency of occurrence of an event,
S – probable damage derived from the event,
E– contact factor,
L – judgment of the analyst on the degree of adequacy of the existing protection measures.
The expression of the risk of accident therefore result to be

$$R = F \times S \times E \times L$$

The variable are managed as Fuzzy sets, described through trapezoidal membership functions, as in table 2. To give as output the risk level, R, thus allowing the decision making, the FAP uses Fuzzy Toolbox of Matlab as a solver.
The model developed therefore allows:

- An assessment to be made of the level of risk of a work phase and/or a work sector
- A verification and quantification to be made of the reduction of the risk after having adopted preventive and/or protective measures
- • A priority of interventions to be established on the basis of the assessed risk levels

The methodology can be considered easy to use for any type of company, with the only prerequisite being to have a record of sufficient and homogeneous number of accidents so as to be able to correctly prepare the software and tune the reference parameters.

2.3. A computer assisted work related accidents analysis technique

The Work Related Accidents Analysis Technique here discussed was developed to make available a method of some help for the analyst to step by step focus the intermediate end very root causes of a work related accident, reducing the possibility of errors due to subjective judgment and hasty evaluations. Basically, it can be considered a combined approach based on the widely used Cause Consequence and Cause-and-Effect techniques from Ishikawa.

The procedure for the analysis, according to the work related accidents analysis approach here discussed, basically contains the following steps:

– selecting an event or type of accident situation to be evaluated (in many cases even if involving more than one victim, the Top event can be assumed to be common, a conjecture to be anyhow carefully verified);
– identifying the safety functions (systems, machinery, operator conduct, etc.) that may have influenced the course of the accident step by step up to the initiating event/s, directly and vs. both the Risk analysis and management company documents and the national safety regulations and standards.

It must be underlined that the accident analysis is basically funded on the comparison between the analyzed situation and a correct work conditions, coherently with the considerations upon which the expected frequency of occurrence level PR was defined. An important advantage of the method developed on this basis is that it is specially targeted to the in deep exam of each single accident, the national/foreign statistical data being only a selectable reference info, so that it is not affected by the afore discussed problems. Moreover, keeping the possibilities typical of a Cause Consequence approach, it is possible to proceed in both directions that is for the analysis of occurred accidents, and to verify the expectable effectiveness of prevention countermeasures in a large number of situations where simple models can be easily displayed. The broad areas of enquiry which shall be treated in a logical process of linking, disjunction, conjunction for the analysis are summarized in Fig.3.

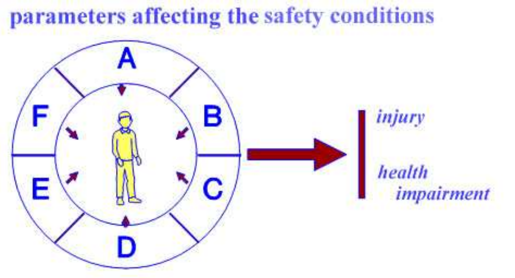
The software was organized in eleven sections, each section consisting of several variables logically linked and with input scrolling facilities, was organized as a sort of user friendly guided tour:

1. General inspection data, Accident code number, Report ID, Event Date, previous inspection activities, etc.;
2. Victim data: ID, job and experience, training, etc.;
3. Company /employer data, activity code, authorizations/permits where required, safety documentation, etc,;
4. General data on the accident: site and operations , direct cause, etc.,
5. Consequences of the accident: seriousness, involved body part, first aid, , etc.;
6. Further available info: work premises characteristics and external/internal criticalities, noise, dust, poor visibility, etc,., presence of witnesses, etc.;
7. Direct causes of the accident, violations and penalties, if any;
8. First step of the data collection/evaluation plan a Cause-and-Effect analysis (involved machinery/plant characteristics and safety/maintenance conditions, standard/special operating situations, etc.);
9. Development of Cause-and-Effect analysis: organizational flaws highlighting and evaluation, if any
10. Accident Investigation Summary;
11. Final report and responsibilities definition, if any, prevention proposals.

The following tools were directly included to support the analysis process:

- a link to the up to date technical standards list, namely the European C type EN standards issued in accordance to the UE Directive on machinery, to be used for a throughout evaluation of the accomplishment to the basic law statement on the comparable or more effective safety level;
- a link to some foreign accident data bases, such as the American DOL-OSHA "Accident Investigation Search", where the accident causes and violation to the safety regulations data can be easily verified, to evaluate the quality of the risk assessment – and deriving Safety Management procedure- in the case under exam, with reference to both a correct design and the evaluation of possible accident involving faults and deviations;
- some examples of safe work procedures directly derived from an affective Risk Analysis, and the basics for workers information, education and training).

The computer assisted approach leads the user to identify the input data useful for the analysis, to realize the accident causes sequence, and to correctly examine the prevention lacks and the non-compliances with the safety regulations and standards, if any. The target being a step by step identification of the relationship between accident consequences and their basic causes in simple or nested lanes, and the combinations of basic causes that can result in the accident sequence, and not for statistical purposes – even if the stored data can be statistically analyzed - the package was developed with Microsoft Visual Basic® 6.0 Professional. To allow an high level data analysis, a remarkable effort has been devoted to make available for many of the involved parameters a large number of pre-defined selection options, to avoid the risk of confusion due to the "fantasy" of each operator: a finite but exhaustive list of parameters (where possible drawn from widely used data bases) removes the possibility that a single parameter, defined with different descriptors, could not be univocally recognized.



If the overall and work load

A + B + C + D + E +... …

is excessive, there is the risk of:

- instantaneous trauma, i.e. accident;
- middle or long term psychophysical decay of the individual, i.e. work related disease.

A Work premises characteristics and external/internal criticalities, machinery and equipment, work organization, etc.;

B Chemical and physical aspects of the working environment: e.g. dust, noise, vibration, lighting, microclimate, etc.;

C Biological aspects: working cycles and management of shifts;

D Physiological aspects: workers age, health, aptitude and experience);

E Ideological aspects: attitude, quality of the management;

F others .....

A and B factors are influenced by the characteristics of the working environment; C, D, E are more directly connected with single workers.

Figure 3. Factors potentially leading to accidents, physical/psychological pathologies and occupational diseases

3. Conclusion

The methodologies and tools here described allow to deal with the procedures related to occupational accident management for preventive purposes.
Data mining applications allows the user to disclose in large accident database the more critical accident sequences to be considered in risk analysis and compared with the local accident situation. Fuzzy Application Procedure allows the risk assessment of occupational accident to be performed and quantified. Last but not least the Work Related Accidents Analysis Technique was developed to make available a method of some help for the analyst to step by step focus the intermediate and very root causes of a work related accident constituting a knowledge and data base of homogeneous data to be fed to FAP and compared to sequences from data mining in order to optimally manage the occupational data information for preventive purposes.

References

[1] Ariano, P.F., Bersano, D., Cigna, C., Patrucco, M., Pession, J.M., Prato, S., Romano, R. & Scioldo, G. (2009). Extractive activities start up and management: a computer assisted specially developed "Prevention through Design" approach. International *Journal of Mining, Reclamation and Environment*.

[2] Camisassi, A., Cigna, C., Nava, S., Patrucco, M. & Savoca, D. (2006). *Load and hauling machinery: an evaluation of the hazard involved as a basis for an effective risk evaluation*. in Cardu M.,Ciccu R.,Lovera E.,Michelotti E. (Eds.) Mine Planning and Equipment Selection (MPES), Torino.

[3] Camisassi, A., Cigna, C. & Patrucco, M. (2004). Sicurezza nei cantieri: analisi di rischio e condizioni di impiego in sicurezza di macchine operatrici e mezzi di sollevamento di materiali. *GEAM-Geoingegneria ambientale e mineraria* XLI, 3, 19-32.

[4] Edelstein, H.A. (1999). *Introduction to Data Mining and Knowledge Discovery*. Two Crows Corporation, third ed.

[5] Hamming, R. (1950). Error Detecting and Error Correcting Codes. *Bell System Technical Journal*, 29, 147-160.

[6] Howard, J. (2008). Prevention through Design: Introduction. *Journal of Safety Research*, 32, 113.

[7] Kohonen, T. (1999). Analysis of processes and large data sets by a self-organizing method. 2nd International Conference on Intelligent Processing and Manufacturing of Materials (IPMM'99), Honolulu, Hawaii, July 10-15, IEEE, vol. 1, pp. 27-36.

[8] Liao, C.W. & Perng, Y.H. (2008). Data mining for occupational injuries in the Taiwan construction industry. *Safety Science* 46, 1091–1102.

[9] Lourenço, F., Lobo, V., Bação, F. (2004). Binary-based similarity measures for categorical data and their application in Self-Organizing Maps. Internal Report, Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa.

[10] Mure`, S., Demichela, M., & Piccinini, N. (2006). *Assessment of the risk of occupational accidents using a FUZZY approach.* Cognition Technology & Work, 8,103–112.

[11] NOHSC - National Occupational Health and Safety Commission. Australian Government, The role of design issues in work related injuries in Australia 1997-2002, 2004.

[12] Palamara, F., Demichela, M. (2007). Occupational accidents risk analysis using clustering algorithms. *Proc. of the European Safety and Reliability Conference, ESREL 2007*.

[13] Zadeh, L.A. (1965). Fuzzy Sets. *Inform Control* 8:338–353.