Expert Systems with Applications 40 (2013) 1484–1490

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

# Profiling risk sensibility through association rules

Beatrice Lazzerini *, Francesco Pistolesi

*Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa, Largo Lucio Lazzarino 1, 56122 Pisa, Italy*

ARTICLE INFO

*Keywords:*
Association rules
Frequent patterns
Risk perception
Risk propensity

ABSTRACT

In the last recent years several approaches to risk assessment and risk management have been adopted to reduce the potential for specific risks in working environments. A safety culture has also developed to let workers acquire knowledge and understanding of risks and safety. Notwithstanding, risks still exist in every workplace. One effective way to improve workers' sensibility to risk, i.e., their ability to effectively assess and control the risks they are exposed to, is risk management training. Unfortunately, people may perceive risks in different ways depending on subjective assessment of the characteristics and severity of the considered risks, and may have tendencies to either take or avoid actions that they feel are risky. Therefore, the knowledge of how workers assess each of the risks they may be exposed to in the workplace is a key factor to conceive effective custom risk management training. In this paper we present a novel approach, based on association rules, to workers' profiling with respect to risk perception and risk propensity in order to provide each of them with specific customized risk management training.



## 1. Introduction

Many modern working environments are characterized by a huge quantity of risks the workers are exposed to. When workers do their job, or they simply stay in the workplace, negative events may happen, whose effects can damage their health and safety. Actually, in the last decades safety management has become an important field of study, much emphasis has been placed on the need for a deep understanding of risk management concepts and principles (Aven, 2012), more and more powerful tools have been developed for risk analysis and management (Dubois, 2010; Lazzerini & Mkrtchyan, 2010, 2011; Misra, 2008). At the same time, industrial systems and machineries have been redesigned and provided with efficient safety features, and working environments have evolved into safer places (ISO31000, 2009; Leitch, 2010): despite this, residual risks remain unacceptably high (ISO27001, 2005).

Workers exposed to risks should be aware of the nature of these risks so as to deal with them as well as appropriate. Human subjectivity causes people to behave differently in similar risky situations: the two main variables playing an important role in the interaction between a person and a possibly risky environment are risk perception and risk propensity (Keil, Wallace, Turk, Dixon-Randall, & Nulden, 2000). *Risk perception* is the subjective way with which a person estimates characteristics and gravity of hazardous situations (Bouyer, Bagdassarian, Chaabanne, & Mullet, 2001; Chauvin, Hermand, & Mullet, 2007; Peters & Slovic, 1996; Sjöberg, 2000; Slovic, 1987), while *risk propensity* is a person's tendency to take or avoid risks (Sharma, Alford, Bhuian, & Pelton, 2009; Sitkin & Weingart, 1995; Sueiro, Sánchez-Iglesias, & de Tella, 2011).

Risk perception is influenced by a variety of factors, including past experience and knowledge, past health status, psychological, social, political, and cultural factors, mood and emotions, personal knowledge about the risky condition, trust in risk management institutions, age, sex, locus of control (Horswill & McKenna, 1999; Klein & Helweg-Larsen, 2002), optimism bias (Costa-Font, Mossialos, & Rudisill, 2009; Klein & Helweg-Larsen, 2002), etc. Similarly, risk propensity is influenced by diverse factors as personality and experience, cultural background, mood, feelings, gender, education, job position, age, etc.

Although extensive research has been made about the factors that determine, respectively, risk perception and risk propensity, no general model exists to explain the diverse behavioural strategies in dealing with a given risk. Further, little is known about the interrelations that exist among risk perception, risk propensity, and decisions involving risk (Keil et al., 2000).

For this reason, training policies have been purposely devised and adopted with the aim of improving people's ability of quickly identifying a source of danger and its potential dangerous effects. The main objective of such a training process is increasing the risk sensibility and awareness level of the workers in order to obtain a safer interaction with the risk itself (Nicholson, Fenton-O'Creevy, Soane, & Willman, 2001). Unfortunately, when training is addressed to a heterogeneous group of workers, some workers may obtain a sufficient level of risk awareness, while others may

* Corresponding author. Tel.: +39 050 2217558; fax: +39 050 2217600.
*E-mail addresses:* b.lazzerini@iet.unipi.it (B. Lazzerini), f.pistolesi@iet.unipi.it (F. Pistolesi).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.08.055

maintain an inappropriate way of interaction with the risk. Therefore, the training process should be tailored to the specific risk sensibility profile of the involved worker.

In this paper we describe a way for customizing the training process for risk awareness by adapting it to the worker's risk perception and risk propensity, in order to obtain the best result in terms of learning. In particular, we consider some criticality factors whose correlation with risk perception and risk propensity has been stated by sociology and psychology experts, namely, gender, age, level of education, income, risk knowledge, work control at work site, professional role, injury frequency, effect seriousness, delayed occurrence of effects, role repetitiveness, industrial injuries and diseases, acquired skills, perception of risk control, work gratification, state of health, safety culture in the company, anxiety level, self-esteem, worry level.

Our choice of the criticality factors is essentially based on heuristic considerations as our main objective is to prove the feasibility and effectiveness of the proposed association rule-based approach to risk sensibility profiling.

To achieve our objective, we deduce the risk sensibility profile of a worker from his/her behaviour in dealing with one or more risks, and from the set of criticality factors introduced above. More precisely, given a set of risks to which a group of workers is exposed, we consider, for each risk, a set of actions a worker should or may perform to prevent the risk to occur. Then each worker is interviewed by asking him/her what action (or actions) he/she would perform to prevent the occurrence of each risk to which he/she is exposed. Each worker is also asked to answer some questions related to the criticality factors. After collecting a sufficient number of interviews, through a data mining process we look for association rules (Agrawal, Imieliński, & Swami, 1993a, 1993b) concerning risks, prevention actions and criticality factors. Once the association rules have been generated, each of those associated with a high level of interestingness, expressed in terms of appropriate indexes (Agrawal & Srikant, 1994; Wu, Chen, & Han, 2010), may represent a risk sensibility profile. In this way, a risk sensibility profile consists in a particular configuration of the criticality factors and a specific behaviour towards one or more risks. We consider both single-risk profiles and multi-risk profiles.

This work represents a first step towards significantly improving risk management as we provide the risk manager with direct knowledge of the risk sensibility profiles of the workers that will be the target of training aimed at risk awareness.

The rest of this paper is organized as follows. In Section 2 we introduce the basics of association rule mining; in Section 3 we introduce the concepts of single-risk profile and multi-risk profile, and detail our association rule-based approach to risk sensibility profiling; in Section 4 we give an illustrative example of the application of the proposed risk sensibility profiler. Finally, we provide concluding remarks in Section 5.

## 2. Data mining

*Data mining* is the process of discovering interesting correlations or useful patterns in large data sets. One of the most important data mining techniques is *association rule mining*, first introduced in Agrawal, Imieliński, and Swami (1993b), which aims to discover all significant associations between items in data repositories.

### 2.1. Frequent patterns and association rules

Frequent patterns are itemsets, structures or sequences of items which are recurrent in a dataset, i.e., they appear together with frequency higher than a specified threshold (Han, Cheng, Xin, & Yan, 2007). A frequent pattern is called *frequent itemset*. Finding frequent patterns allows to discover associations and correlations among the items in a dataset.

Let $\mathcal{U} = \{i_1, i_2, \ldots, i_m\}$ be a set of items. A *transaction* $T$ is a subset of $\mathcal{U}$. A set of items $X \subseteq \mathcal{U}$ is called *itemset*; an itemset which contains $k$ items is also named a *k-itemset*. A transaction $T$ contains an itemset $X$ if $X \subseteq T$. Let $A \subset \mathcal{U}$ and $B \subset \mathcal{U}$ be two itemsets such that $A \cap B = \emptyset$. An *association rule* is an implication of the form $A \Rightarrow B$, with $A$ and $B$ called, respectively, *antecedent* and *consequent*. The two most used measures of the importance of an association rule are support and confidence. The *support* expresses the proportion of transactions in the transaction set containing $A \cup B$. The event of finding both sets $A$ and $B$ occurs with a probability $P(\mathcal{A} \cap \mathcal{B})$, where $\mathcal{A}$ and $\mathcal{B}$ are, respectively, the event of finding the itemset $A$ and the event of finding the itemset $B$ in a transaction. Hence the support (abbreviated as supp) of a rule $A \Rightarrow B$ is defined as

$$\mathrm{supp}(A \Rightarrow B) = P(\mathcal{A} \cap \mathcal{B}). \tag{1}$$

The support is also indicated with $\mathrm{supp}(A \cup B)$.

The *confidence* (abbreviated as conf) of the rule $A \Rightarrow B$ is defined as the percentage of transactions containing $A$ that also contain $B$:

$$\mathrm{conf}(A \Rightarrow B) = P(\mathcal{B}|\mathcal{A}) = \frac{\mathrm{supp}(A \cup B)}{\mathrm{supp}(A)}, \tag{2}$$

where $P(\mathcal{B}|\mathcal{A})$ is the conditional probability.

During the rule mining process, only the rules having support and confidence greater than user-defined thresholds are considered: these rules are called *strong rules*.

The association rule generation can be described as a two-step process (Agrawal et al., 1993b):

1. *Large itemset generation*: find all itemsets having support above the chosen minimum support $s_{min}$; these itemsets are called *large* itemsets;
2. *Rule generation*: use the large itemsets to generate the desired rules. Each generated rule has to satisfy also a minimum confidence.

The association rules can be generated starting from the set $\mathcal{L}$ of large itemsets returned by step 1, as follows:

- for each large itemset $l \in \mathcal{L}$, generate all its non-empty subsets;
- for each non-empty subset $s \subset l$, consider the rule $s \Rightarrow (l - s)$ if and only if:

$$\frac{\mathrm{supp}(l)}{\mathrm{supp}(s)} \geqslant c_{min}$$

where $c_{min}$ is the minimum confidence.

Since all the rules are generated from large itemsets, they automatically satisfy the minimum support conditions, so only the minimum confidence level has to be checked.

A lot of algorithms for mining association rules use criteria based on support and confidence. Although these approaches are able to avoid the generation of a high number of uninteresting rules, many of the generated rules may still be not interesting to the user. This may happen, e.g., when a too low support threshold is used. To solve this problem further interestingness measures can be taken into account, based on statistical significance and correlation analysis. Several interestingness measures have been proposed; one of the most popular is, e.g., the *lift*, which is defined as

$$\mathrm{lift}(A \Rightarrow B) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{A})P(\mathcal{B})} = \frac{\mathrm{conf}(A \Rightarrow B)}{\mathrm{supp}(B)}. \tag{3}$$