Safety Science 70 (2014) 89–98

Contents lists available at ScienceDirect

Safety Science

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

Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports

Abhishek Verma [a], Sudha Das Khan [a], J. Maiti [a,*], O.B. Krishna [b]

[a] Department of Industrial & Systems Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, India
[b] Corporate Safety, Safety, Health and Environment, Tata Steel, Jamshedpur, India

ARTICLE INFO

Article history:
Received 10 November 2013
Received in revised form 13 April 2014
Accepted 13 May 2014
Available online 6 June 2014

Keywords:
Incident investigation
Association rules
Incident patterns
Steel plant

ABSTRACT

The aim of this paper is to find out the patterns of incidents in a steel plant in India. Occupational incidents occur in steel plant mainly in form of injury, near miss, and property damage or in combination. Different factors are responsible for such incidents to occur. An incident investigation scheme is proposed. Association rule mining approach is used to discover cause-and-affect patterns (rules) using 843 incidents. Thirty-five meaningful association rules are extracted using three criteria, support (S), confidence (C) and lift (L). For example, the results show that unsafe acts done by others are more frequent in injury cases ($S = 4.86\%$, $C = 78.8\%$, $L = 2.3$). Similarly, one of the SOP (standard operating procedures) related rule: 'SOP required, available, adequate but not complied' led to property damage ($S = 11.03\%$, $C = 49.2\%$, $L = 1.525$). Another useful rule 'SOP required, available but inadequate, followed' led to near miss ($S = 1.66\%$, $C = 38.89\%$, $L = 1.163$). It is also found that for slip, trip and fall incidents, workers working alone ($S = 3.91\%$, $C = 76.74\%$, $L = 2.239$) or in a group ($S = 3.20\%$, $C = 75.00\%$, $L = 2.188$) does not make much difference. The findings pinpoint the areas of improvement such as inadequate SOPs, non-compliance of SOPs, training, and slip, trip and fall prevention to minimize incidents.

1. Introduction

Steel manufacturing is one of the most hazardous industries because of its complex socio-technical system. In one hand, it uses high technology and on the other hand, it is labor intensive. It has all components of process safety and personal safety. Managing safety in steel industry is a daunting task. According to domino theory proposed by Heinrich (1959), maximum industrial incidents occur due to unsafe acts, unsafe conditions, or both; which are controllable. Reason's Swiss Cheese model depicts that an hazard becomes an accident, when a series of events in between align together thus creating a path from hazard to accident (Reason, 1990). Identification and quantification of such paths are utmost important for preventing accidents to occur.

The analysis of accident data is as old as the development of Poisson distribution. Fatal accident being a rare event fits Poisson distribution very well. Over the years, several statistical models have been used to explore factors causing incidents. Khanzode et al. (2012) have done a review work on accident causation theories generation wise; like 1st generation theories on accident proneness (Greenwood and Woods, 1919), 2nd generation theory about Domino theories (Heinrich, 1959), 3rd generation injury epidemiology theories (Haddon et al., 1964) and 4th generation theories on system approach (Trist and Bamforth, 1951). Some pioneering works on accident data analysis includes (Cooper, 2000; Shankar and Mannering, 1995; Maher and Summersgill, 1996).

In recent years, the analysis of incident/accident data using data mining techniques and algorithms has been gaining much attention among researchers (Pande and Abdel-Aty, 2009; Liao and Perng, 2008; Cheng et al., 2013; Arunraj et al., 2013). Recently, Maiti et al. (2014) proposed a methodology to determine safety rules for derailments in a steel plant using corresponding analysis. Most of industrial incidents occur due to lack of knowledge, lack of standard operating procedure and its compliance, insufficient training, etc. Researchers are trying to discover the causes of incidents which can be used to improve safety management system in industry. Many data mining techniques such as support vector machines, classification and regression trees and Bayesian networks have been used to identify hidden patterns and structures in a large amount of data consisting of various factors associated with incidents or accidents.

The association rule mining (Agrawal et al., 1993) has been used to analyze incident data to get rules for incident patterns. For example, in road incident, crash data analysis was done to find

* Corresponding author. Tel.: +91 3222283750; fax: +91 3222255303.
E-mail addresses: abhishekverma.cs@gmail.com (A. Verma), essd.iit@gmail.com (S.D. Khan), jhareswar.maiti@gmail.com (J. Maiti), obkrishna@gmail.com (O.B. Krishna).

http://dx.doi.org/10.1016/j.ssci.2014.05.007
0925-7535/

out the contributory factors and their interdependency (Pande and Abdel-Aty, 2009; Montella, 2011; Montella et al., 2012); in railway, accidents data analysis was done to discover the factors and relationship between them (Mirabadi and Sharifian, 2010); and in construction industry, incident data was analyzed to find out cause and effect relationships of occupational incidents (Liao and Perng, 2008; Cheng et al., 2010). One of the well-received algorithm for association rule mining is apriori algorithm (Agrawal and Ramakrishnan, 1994). Apart from apriori algorithm there are different algorithms, techniques and approaches that have been used for mining association rules are, FP-Growth algorithm (Han et al., 2000), Eclat algorithm (Zaki, 2000), Maxclique algorithm (Zaki et al., 1997), partitioning technique (Savasere et al., 1995), sampling approach (Toivonen, 1996), continuous association rule mining algorithm (CARMA) (Hidber, 1999), vertical itemset partitioning for efficient rule extraction algorithm (VIPER) (Shenoy et al., 2000), aprioriTid and aprioriHybrid algorithms (Agrawal and Ramakrishnan, 1994). Although there are several studies on association rule mining conducted, but in safety data analysis, literature is scanty.

Most of the leading organizations across the world have their own incident or accident investigation and reporting system. The primary task of incident investigation is to identify "what", "where", "when", "why" and "how" incidents have happened. It also aims to find out the root causes behind a particular incident to take corrective actions to avoid its reoccurring. The key information regarding incidents is recorded in investigation reporting system. Appropriate factors are used in incident investigation reporting according to industry type (Brazier, 1994). Incident investigation reporting has been used by various industries to find the causal factors and near misses to improve safety performance (Ji and Zhang, 2012; Okstad et al., 2012; Nesmith et al., 2013; Jones et al., 1999). Basso et al. (2004) proposed an incident investigation database system to record factors causing incidents as well as corrective actions that allow monitoring safety performance. Oktem et al. (2010) also proposed a model to design near miss management system by a framework from event identification to solution implementation. Gnoni et al. (2013) proposed similar type of model taking benefit of lean thinking. While the incident reporting system is somehow well managed but it is still flawed from two counts: (i) lack of process approach incorporating workflow information among key stake holders like supervisor, department head, safety professional, etc. and (ii) data gathered is hardly analyzed in the way it is purposed. In this study, we attempt to provide a scheme for incident investigation followed by incident data analysis.

This study starts with describing the incident data and explaining the proposed incident investigation process (Section 2.1). It comprises the most prominent combination of factors leading to one of three types of incidents: injury, near miss, and property damage. Association rule mining using apriori algorithm is employed to extract incident patterns in Section 2.2. Data codification and analysis is given in Section 2.3. The results obtained and its practical implications are given in Section 2.4. Inference of the study to take managerial decisions has been discussed in Section 3. Finally, conclusions of the study with future scope of research are given in Section 4.

2. Methods

2.1. Incident investigation process

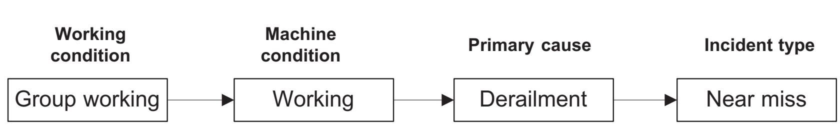
The study was conducted in coke, sinter and iron (CSI) division of a steel manufacturing company in India. To ensure safety, the company uses online safety management system (SMS) which is available on local network inside the organization. This system provides worksite observations, fatality risk control programs, incident investigations, and job cycle checks. For this study, incident data recorded in the SMS is considered. The SMS generates excel report of the incidents logged on. In our study we have proposed the workflow for incident investigation as shown in Fig. 1. All incidents are investigated, and incident investigation reports are created. On the basis of investigation, actions are taken to improve safety (both unsafe acts and/or conditions). Any employee involved or witnessing an incident can report the same to the corresponding supervisor of the department. Supervisor then log on the incident investigation module of the SMS and fills the information fields. Then the severity of the incident is assessed considering all the hazardous elements and conditions prevailed during the incident. Depending upon the severity, risk score is given to that particular incident. If risk score is higher than a threshold limit, a pre-specified value determined by the organization, it is considered as high priority incident and is then sent to the head of department (HOD) for further consideration. Low priority incident cases are taken care at the supervisory level. To handle high priority incident cases, HOD forms an investigation team of specialists to further investigate the incident scenario and explore the causal factors. After the investigation, recommendations are released by the team for implementation. Whether recommendations are correctly implemented or not is verified by safety professional. Presently, 15 information fields are generated. For our study ten out of fifteen factors (fields) have been considered as per the discussion with safety expert for in-depth analysis. Proper description for those ten factors with related information is given in Table 1.

2.2. Association rule mining

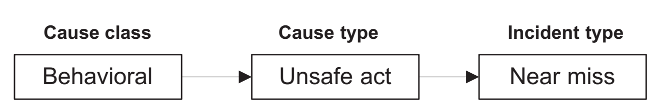
2.2.1. Illustrative examples

Example 1: A group of workers tries to couple three loaded boxes with a loco, which already is attached with four empty boxes. One of the loaded boxes is derailed because of hard pushing. No one is injured, i.e., near miss happens. But this might be lead to a major accident. While analyzing the incident, the following rules may be generated:

Rule 1:



Rule 2:



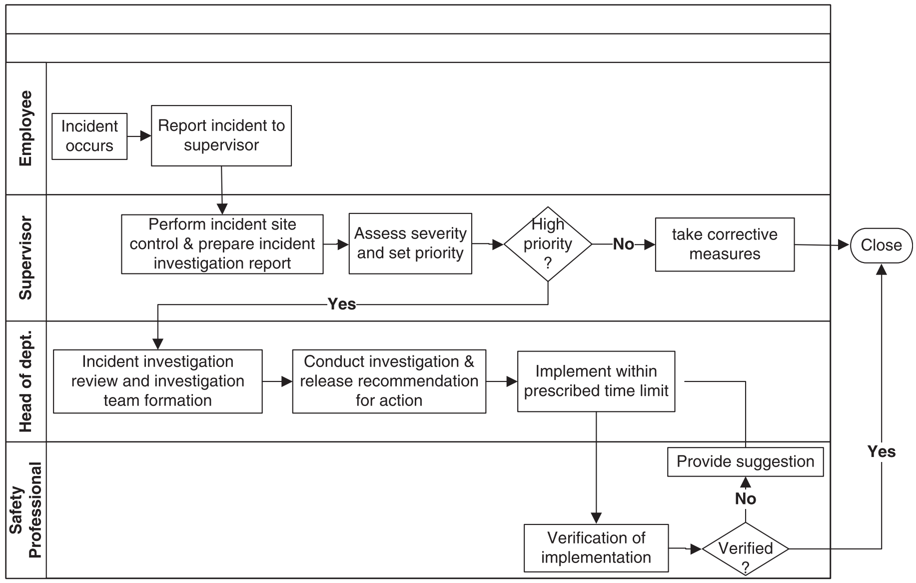
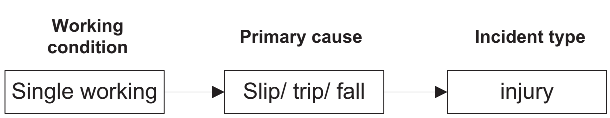


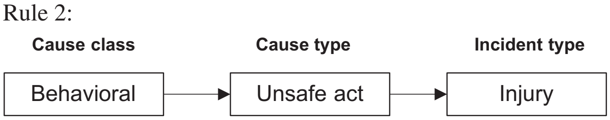
Fig. 1. Proposed work flow of incident investigation module.

Example 2: One of the very common incident causation in every industry is slip/trip/fall. This is a repetitive phenomenon. As an example, a worker is engaged with a dumper driver for coke breeze removing activity at breeze pond of a battery. He observes that the hook of the door of the dumper does not close properly due to the deposition of coke breeze. Resting his bare left hand on the cargo space of the dumper, he starts cleaning the deposited coke breeze. Suddenly his hand slips and he falls down. This happens because of not wearing hand gloves and not using proper tools. While analyzing the incident, the following rules may be generated:

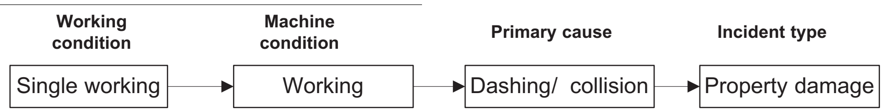
Rule 1:



Rule 2:



Example 3: Dashing/collision are very common accident causation. An example of property damage may be as follows: a dumper truck is carrying lime heads towards a lime crushing plant. On its way, the hydraulic jack is suddenly released unintentionally and the cargo space lifts up and hits the barrier gate of the crossing. This results either by a mechanical failure of the jack release switch of the truck or by a careless conduct of the driver who operates the switch. While analyzing the incident, the following rule may be generated:



The above examples with proposed rules provide avenue for learning from past incidents in terms of rules. These rules are created considering different factors that are associated with such incidents. However, from one or two examples we cannot make general rules. We require large amount of incident data that support such rules through frequent recurrences. But this creates complexity in analyzing large data. We require relying on some logic/reasoning that ultimately help us defining such rules from the large data set. Rules generated while considering association of several incident related factors are termed as association rules in safety context and one of the techniques to generate such rules is association rule mining which is used in this study. The following section details the association algorithm (apriori algorithm) to extract association rules.

2.2.2. Apriori algorithm

Association rule identification is the discovery of item sets, which occur mutually in a given data set (Han and Kamber, 2001). The rules are based on the frequency number of an itemset, which occurs alone, or in combination with other sets in a database. Association rule is mostly used to identify 'interesting' hidden relationships among attributes of huge database. Generally, a standard association rule is expressed in $X \rightarrow Y$ form, where X is the antecedent and Y is the consequent, which signifies that X will occur with Y for the same instance in a database with a minimum level of significance. Note that each rule can have multiple items, i.e., a set of items, as antecedent and consequent. It is worth to

Table 1
Description of each factor and information fields for corresponding factor.

	Factors	Description	Information fields (Items)
1.	Incident type	Categories of incident occurred	Injury, near miss and property damage
2.	Primary cause	Describes the primary reason at the time of incident occurrence	Dashing/collision, derailment, energy isolation, road incident, slip/trip/fall, process incidents, occupational illness, equipment/machinery damage, material handling, lifting tools tackles, structural integrity, working at height
3.	Cause class	Describes whether a behavior or a process is responsible for the incident occurrence	Behavior, process
4.	Cause type	Describes whose fault was to happen an incident	Unsafe act, unsafe condition, unsafe act by other, unsafe act and unsafe condition
5.	Working condition	What was the working condition at the instance of incident occurrence	Single, group, no working
6.	Machine condition	What was the machine condition at the instance of incident occurrence	Idle, working, not applicable
7.	SOP requirement	Whether SOP is needed for the job	Required, not required, not applicable
8.	SOP availability	Whether it is applicable or not for that job given SOP required	Available, not available, not applicable
9.	SOP adequacy	Whether it is adequate or not for that job given SOP required and available	Adequate, not adequate, not applicable
10.	SOP compliance	Whether it is followed or not for that job given SOP required, available and adequate	Followed, not followed, not applicable

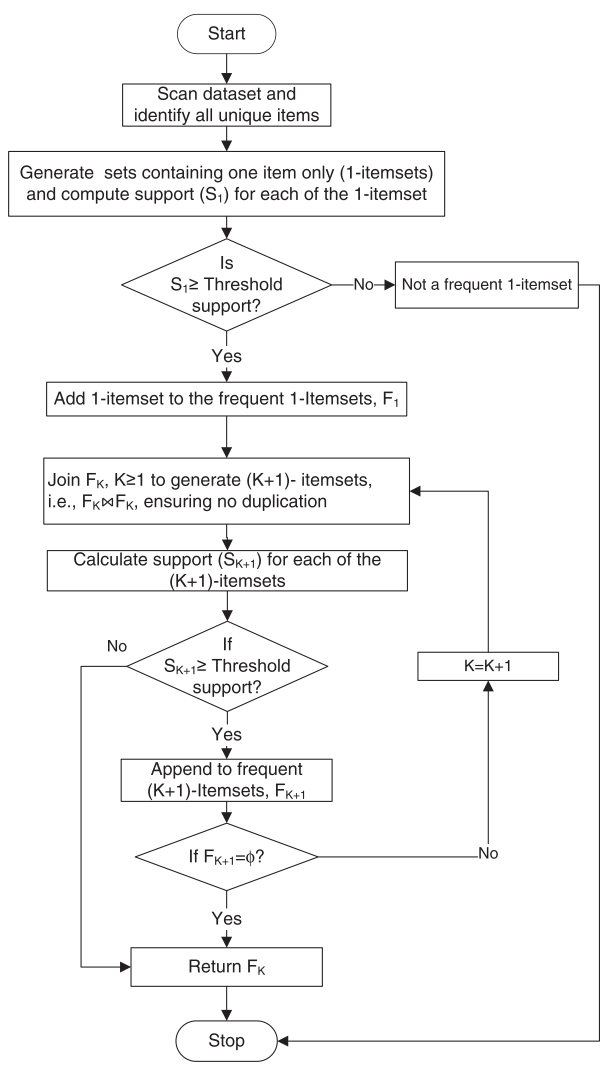


Fig. 2. Flow chart for frequent itemsets generation using apriori algorithm.

mention that association rules only associate the itemsets, not inferred the direct causation (SAS Institute, 2001).

Apriori algorithm, proposed by Agrawal et al. (1993), has been used to find out strong association rules among itemsets of the incident data collected. The apriori algorithm to find out frequent itemsets is shown in Fig. 2 and generation of association rule from frequent item set is shown in Fig. 3. Let $I = i_1, i_2, \ldots i_n$ defines the itemsets and N be the total number of itemsets. Let $X \subseteq I$ and $Y \subseteq I$ where X and Y are two distinct subsets of I, and X and Y are not void, i.e., $X = \varnothing$, $Y = \varnothing$. The main two measurements of rule effectiveness are support and confidence, which reflect the usefulness and certainty of discovered rules, respectively (Han and Kamber, 2001). Support (S) is the measurement of the proportion occurrence of any itemset or combination of itemsets (e.g., X and Y) in a database (D). Support is expressed as follows:

$$S(X) = \frac{N(X)}{N}$$

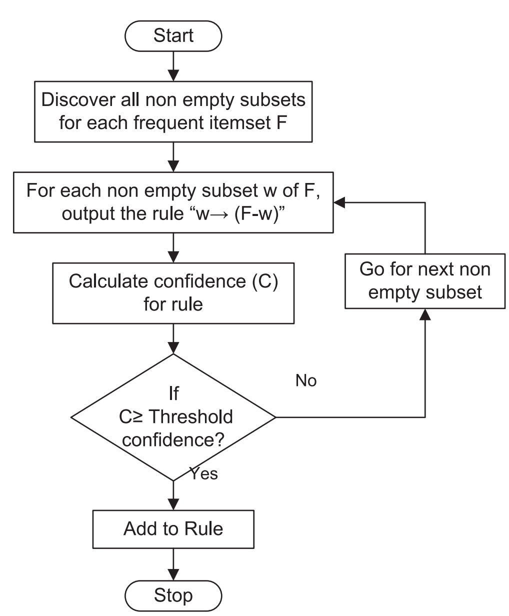


Fig. 3. Flow chart for association rule generation form frequent itemsets.

$$S(X \to Y) = P(X \cup Y) = \frac{N(X \cup Y)}{N} \tag{1}$$

Confidence (C) is defined as the conditional probability (P) of occurrence of the consequent of an itemset given that the antecedent of that itemset has occurred. Confidence is expressed as follows:

$$C(X \to Y) = \frac{P(X \cup Y)}{P(X)} = \frac{S(X \cup Y; D)}{X(X; D)} \tag{2}$$

For example, a rule for a group of people trying to attach boxes in a loco and one box is derailed causing a near miss incident can be generated as follows. Out of N incident data, let $N(X)$ of them occur under "machine working + group working + derailment" and $N(Y)$ of them are near-miss incidents. Out of $N(Y)$ near-miss incident type, $N(XY)$ happens under the situation "machine working + group working + derailment". Here, "machine working + group working + derailment" is the antecedent and "near miss" is the consequent. So, the rule for the incident ($X \to Y$) is "machine working + group working + derailment" → "near miss". The support values are:

$$S(antecedent) = \frac{N(machine\ working + group\ working + derailment)}{total\ incidents} = \frac{N(X)}{N}$$

$$S(consequent) = \frac{N(near\ miss\ incidents)}{total\ incidents} = \frac{N(Y)}{N}$$

$$S(rule) = \frac{N(\text{“}machine\ working\ condition + group\ working + derailment\text{“} \cup \text{”}neat\ miss\text{”})}{total\ incident} = \frac{N(X \cup Y)}{N}$$

The confidence for the rules is

$$C(rule) = \frac{S(rule)}{S(antecedent)}$$

Both support and confidence of a rule should satisfy corresponding threshold values. The threshold value for support as well as confidence is determined based on sample size and minimum frequency of required factors. It is user specific. Lower the threshold values, the larger the number of rules and vice versa (Fig. 4). Unnecessary large numbers of rules are confusing while pruning many rules (using higher threshold) may discard some important rules. Hence, a tradeoff is needed. Agrawal et al. (1993) considered only support and confidence for generating association rules. However, these two parameters are inadequate to extract interesting association rules from all possible rules. In order to improve the rule generation procedure, a third evaluation parameter 'lift' or 'interest' has been proposed by Brin et al. (1997) and Zhang and Zhang (2002). Montella (2011) defined lift (L) as the frequency of co-occurrence of the antecedent and the consequent to the expected frequency of co-occurrence under the assumption of conditional independence. Actually, lift is the measurement of correlation between X and Y Lift is expressed as follows:

$$L(X \to Y) = \frac{P(X \cup Y)}{P(X)P(Y)} = \frac{X(X \cup Y)}{S(X) \times S(Y)} \tag{3}$$

According to Lee et al. (2012)

- $L = 1$ indicates no correlation between antecedent and consequent.
- $L > 1$ indicates positive correlation between antecedent and consequent.
- $L < 1$ indicates negative correlation between antecedent and consequent.

The negative lift between antecedent and consequent indicates that occurrence of one hardly leads to the occurrence of other. For the rule described above,

$$L(rule) = \frac{S(rule)}{S(antecedent) \times S(consequent)}$$

In addition, following Argyris and Schon (1978) single and double loop learning concept, we adopt multiple loop rules generation process (equivalence to rule learning). Figs. 5 and 6 depict this. As it has been seen, association rule mining can help us in obtaining single to multiple loop learning. What will be the degree of loop (1, 2 or more) will depend on the purpose to be served. In safety

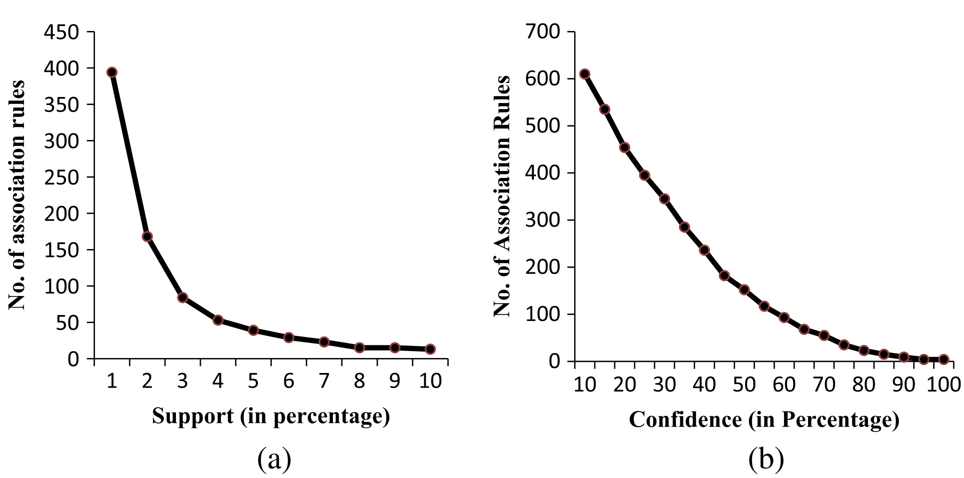


Fig. 4. (a) Variation of number of association rules with support values keeping confidence value constant. (b) Variation of number of association rules with confidence values keeping support value constant.

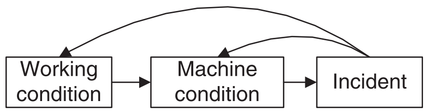


Fig. 5. Double (3-item) loop learning.

studies, the essence of any data analysis is to identify the causes of incidents as well as the causal chain. Therefore, higher the loop, better the rule, provided the rule is meaningful from accident causation and prevention point of view. Based on the discussion with the safety professional and process/work system experts of the plant studied double (3-items), triple (4-items), and quadruple (5-items) rules are preferred.

2.3. Data and analysis

For this study incident data for a period from March 2010 to July 2013 is considered. In total 843 incidents were recorded. The codification and frequency for information fields for corresponding factors are shown in Table 2.

As described earlier, in order to discover the rules, minimum thresholds for support and confidence need to be specified. Numbers of association rules generated are inversely proportional to the threshold support and threshold confidence. It depends on the user to fix the threshold values for pruning large number of association rules as per requirement. There is no established criterion for selecting threshold values for support and confidence. Different studies considered different threshold support and confidence values as per the availability of number of data points and achievement of strong rules (Montella et al., 2012; Pande and Abdel-Aty, 2009; Montella, 2011). In this study, the threshold support and confidence values have been considered 1% and 34% respectively for all the three-incident types, i.e., injury, near miss, and property damage. It means that no rules with support <minimum support and/or confidence <minimum confidence would be considered along with lift value greater than one. All generated

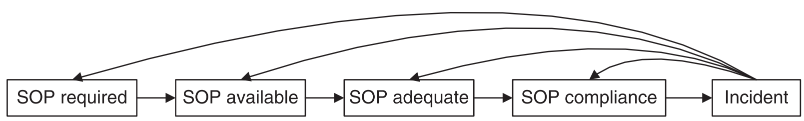


Fig. 6. Quadruple (5-item) loop learning.

Table 2
Codification and frequency distribution of each of the factors considered.

Sr.	Factors	Information fields (Items)	Code	Frequency
1.	Incident type	Injury	It1	289
		Near miss	It2	282
		Property damage	It3	272
2.	Primary cause	Dashing/collision	PrC1	103
		Derailment	PrC2	149
		Energy isolation	PrC3	14
		Road incident	PrC4	142
		Slip/trip/fall	PrC5	118
		Process incidents	PrC6	209
		Occupational illness	PrC7	12
		Equipment machinery damage	PrC8	23
		Material handling	PrC9	42
		Lifting tools tackles	PrC10	11
		Structural integrity	PrC11	15
		Working at height	PrC12	5
3.	Cause class	Behavior	CC1	620
		Process	CC2	223
4.	Cause type	Unsafe act	Ct1	343
		Unsafe act & unsafe condition	Ct2	88
		Unsafe act by other	Ct3	68
		Unsafe condition	Ct4	344
5.	Working condition	Group working	WC1	454
		Single working	WC2	279
		Not a working place	WC3	110
6.	Machine condition	M/C idle	Mc1	42
		M/C working	Mc2	185
		Not applicable	Mc3	616
7.	SOP requirement	SOP required	SR1	418
		SOP not required	SR2	57
		Not known	SR3	368
8.	SOP availability	SOP available	SA1	374
		SOP not available	SA2	101
		Not known	SA3	368
9.	SOP adequacy	SOP adequate	SAd1	297
		SOP not adequate	SAd2	76
		Not known	SAd3	470
10.	SOP compliance	SOP followed	SC1	144
		SOP not followed	SC2	288
		Not known	SC3	471

rules for injury, near miss, and property damage are shown in Figs. 7–9 respectively. The figures include the following parameters:

1. Lift
2. Support (%)
3. Confidence (%)

In Figs. 7–9, the three-item, four-item, and five-item rules are separately shown, for three different incident types separately along with their support, confidence, and lift values.

2.4. Result and discussion

Rules consisting of higher lift (greater than one) values are more strong and interesting. There are three three-item rules, seven four-item rules and two five-item rules for injury incident type that have satisfied the minimum support (1%) and minimum confidence (34%) along with lift value greater than one. The first most useful three-item rule having $L = 2.3$ for injury type incident is ‘cause class-CC1 and cause type-Ct3 → incident type-injury ($S = 4.86\%$, $C = 78.8\%$)’ which signifies that the ***injury is most likely to have happened due to behavioral related problem and unsafe act by others***. The finding supports the studies of Brown et al. (2000) in steel industry and Paul and Maiti (2007) in mining industry. An useful four-item rule having $L = 2.24$ is ‘M/C condition-Mc3 and working condition-WC1 and primary cause-PrC5 → incident type- injury ($S = 3.91\%$, $C = 76.74\%$)’ indicates that slip/trip/fall is also a reason for creating injury incident while people working in group. In case of five-item rules, the most useful one having $L = 2.15$ is ‘SOP requirement-SR2 and availability-SA2 and adequacy-SAd3 and compliance-SC2 ($S = 4.98\%$, $C = 73.68\%$)’ which signifies that injury incident happened for those jobs where SOP was not available because they were thought not to be required. For example, after discussion with safety expert, it was found that this condition prevails in case of road incidents. So, there must be SOP for every job to be performed.

There are five three-item rules, six four-item rules, two five-item rules that have satisfied the minimum support (1%) , minimum confidence (34%) and lift (greater than one) in case of near miss incident. The first most useful three-item rule having $L = 1.87$ for near miss is ‘cause class-CC2 and cause

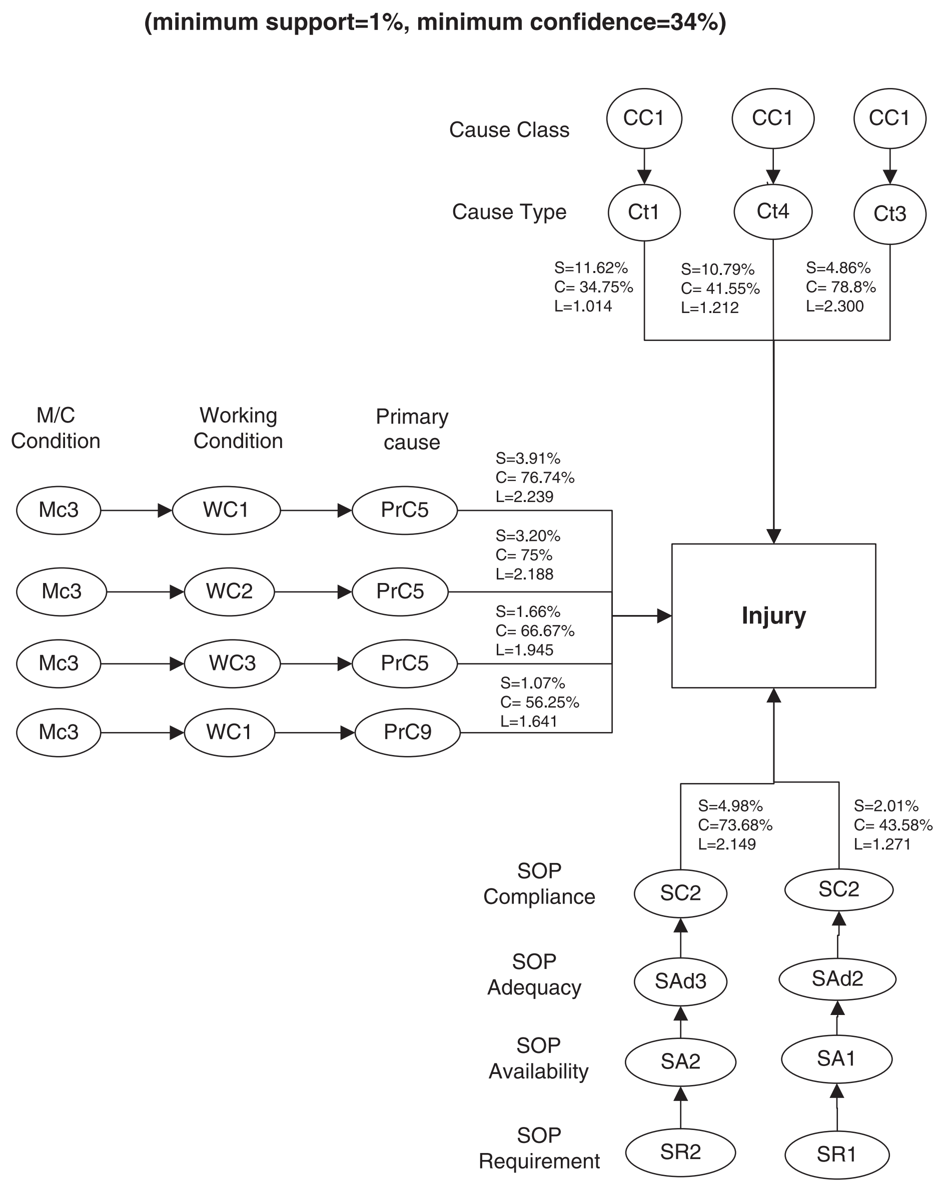


Fig. 7. Association rule for injury type incident.

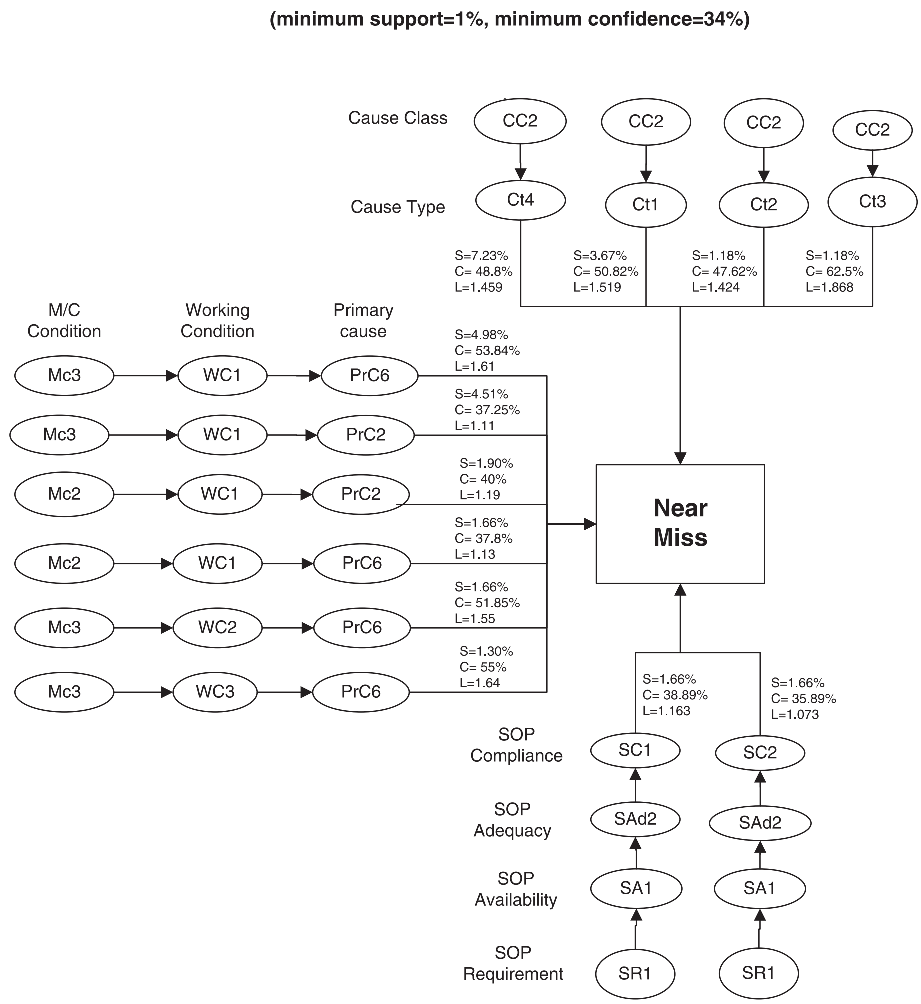


Fig. 8. Association rules for near miss type incident.

type-Ct3 → incident type-near miss ($S = 1.18\%$, $C = 62.5\%$)' which signifies that ***near miss is most likely to have happened due to process related problem and unsafe act by others***. The first most useful four-item rule having $L = 1.64$ for near miss is 'M/C condition-Mc3 and working condition-WC3 and primary cause-PrC6 → incident type-near miss ($S = 1.3\%$, $C = 55\%$)' which signifies that process incidents are most likely to be the reason for near miss where machine and working condition are not applicable. In case of five item rules for near miss, the most useful one is 'SOP requirement-SR1 and availability-SA1 and adequacy-SAd2 and compliance-SC1 → incident type- near miss ($L = 1.163$, $S = 1.66\%$, $C = 38.89\%$)'. This rule indicates that ***near misses are most likely to be happened because of available inadequate SOPs***.

There are four three-item rules, seven four-item rules, three five-item rules that have satisfied the minimum support (1%), minimum confidence (34%) and lift value (greater than one) in case of property damage incident. The most useful three-item rule having $L = 1.33$ is 'cause class-CC2 and cause type-Ct2 → incident type-property damage ($S = 1.06\%$, $C = 42.85\%$)' that implies that ***property damage incidents are most likely due to process related problems and unsafe act & unsafe condition***. The first most useful four-item rule having $L = 2.38$ for property damage is 'M/C condition-Mc2 and working condition-WC1 and primary cause-PrC1 → incident type-property damage ($S = 1.19\%$, $C = 76.9\%$)'. This rule signifies that ***dashing/collisions are most likely to be the reason for property damage while people are working in a group and machine is in a working condition***. In case of five item rules for property damage, the most useful one is 'SOP requirement-SR1 and availability-SA1 and adequacy-SAd1 and compliance-SC2 → incident type- property damage ($L = 1.525$, $S = 11.03\%$, $C = 49.2\%$)' which implies that ***property damage is most likely to be happened due to not following the available adequate SOP***. Proper monitoring of SOP compliance can reduce the amount of property damage.

Most of the rules developed above are plant specific. Unfortunately such studies are rare and hence, cannot be backed up by previous studies. The authors therefore argue that these rules can be considered as preliminary hypotheses for future studies. The limitation of the proposed work flow based incident investigation model is that it requires full commitment from the employees and management to make it effective (Gnoni and Lettera, 2012). It requires training and motivating employees so that reliable information can be collected to perform data analysis.

3. Managerial implication

It is also observed that slip/trip/fall occur mainly during group working conditions, implicating lack of alertness and awareness. This can be improved by spreading awareness of safety culture among workers. Training should be provided mainly to newly joined employees and temporary workers who do not have adequate experience. They are not familiar with new working environment. It is found that near misses are frequent with process incidents. Interestingly, available SOPs are found to be inadequate which is one of the significant management problems. So adequate

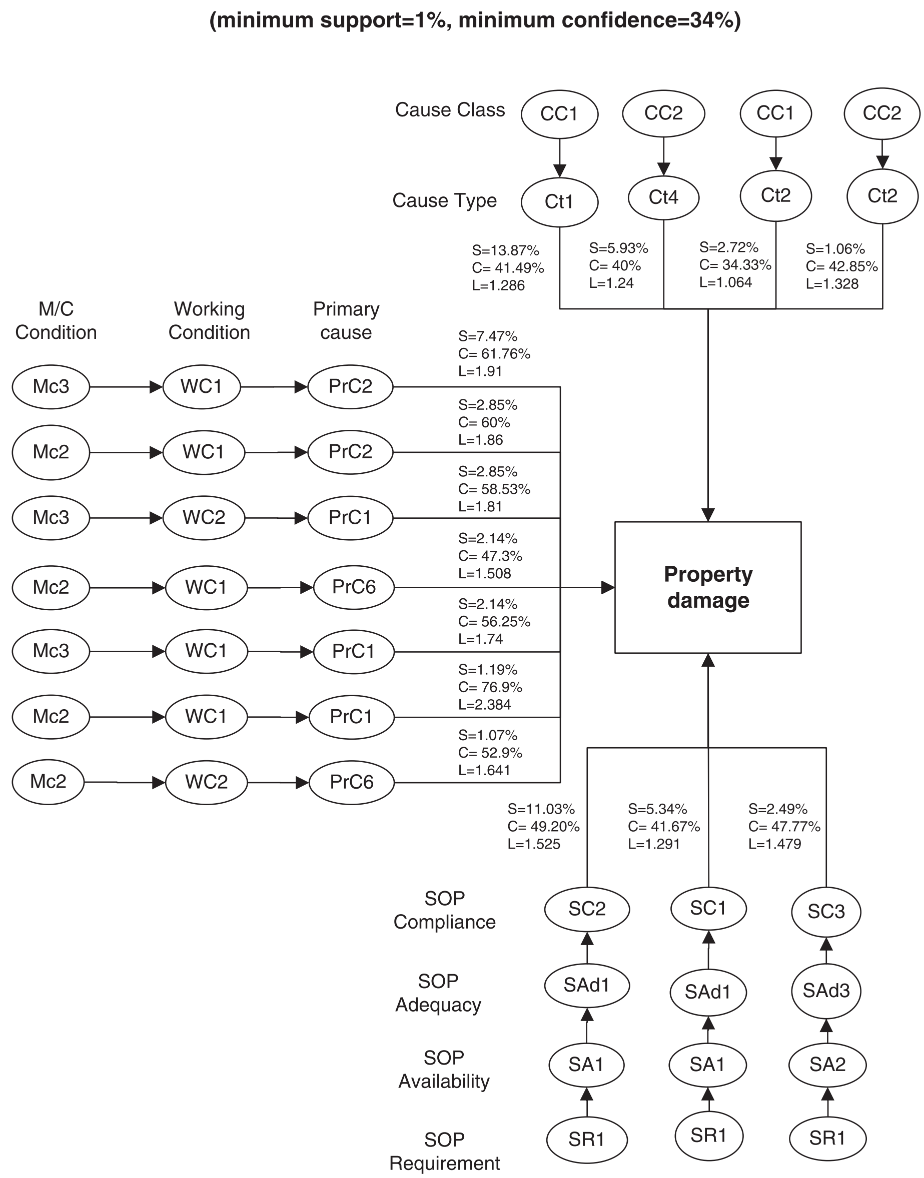


Fig. 9. Association rules for property damage incident type.

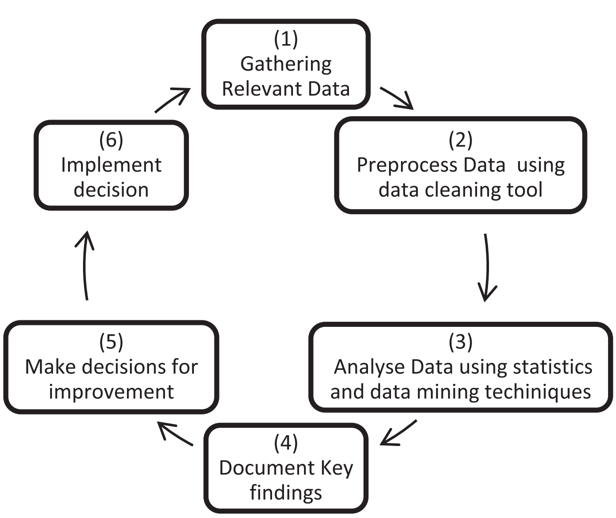


Fig. 10. Cyclic process for relevant and reliable data collection.

SOPs must be provided by management and compliance should be monitored. The results have shown that occurrences of property damage incidents primarily happened due to dashing/collision and derailment bring about by human behavior because it is found that even SOPs are required, available and adequate but not followed.

So as per above discussion it is found that association rule mining successfully extract some rule which can be used by management to take action and make policy to improve safety culture in organization. Association rule mining can also improve the data collection process because to take effective decision out of data and predicting future incident correctly it is important to get quality data. Fig. 10 shows the general process that how the statistics and data mining techniques jointly can help the management to take decision. After implementation of decision taken, data collection process will get improved as a result of lesson learned and statistical analysis. It will also reveal the factors that are more crucial and should be collected next time which would have been not included previously.

4. Conclusions

The purpose of this study was to investigate and to identify the operating factors and generating association rules for incidents occurring in a particular steel industry in India and also to suggest preventive and safety measures to prevent or minimize incidents. A scheme for incident investigation is explained that is being carried out in the organization studied. From the results, it is identified that the main factors contributing root causes are slip/trip/fall, dashing/collision, inadequate SOP's and unsafe acts by co-workers. The findings also indicate that behavior related problems, such as 'unsafe acts done by others', are found to be considerable in number in injury cases and SOP non-compliance is notable in property damage cases. After discussion with safety experts it was found that root causes behind these unsafe behavior related problems include work stress, production pressure, overconfidence, lack of concentration, lack of training and supervision for new workers. In addition, seasoned workers rely more on experience rather than SOP. So these association rules act as guidelines for the management to make policy such that safety performance improves. Limitations of this study can be attributed to the quality of the data used. The data quality mainly depends on the understanding and expertise of supervisors who are documenting the incidents in the database. Fallacy among supervisors about the cause of the same incident has been observed in filling the information fields. However this problem was overcome after taking expert suggestions and missing data problem was overcome after discussions with the supervisors and safety professional. For the future work, other variant of apriori algorithm or other algorithms that we discussed in the introduction section can be used to find association rules for safety related incidents and their comparison with apriori algorithm can be done. Nevertheless, the present study provides avenue for learning from past and their management implications.

Acknowledgements

The authors are thankful to the learned reviewers for their valuable suggestions in enriching the quality of the paper. The authors also acknowledge the support and cooperation provided by the management of the plant studied.

References

Agrawal, R., Ramakrishnan, S., 1994. Fast algorithms for mining association rules. In: International Conference on Very Large Data Bases, pp. 1–32.

Agrawal, R., Road, H., Jose, S., 1993. Mining Association Rules between Sets of Items in Large Databases, pp. 1–10.

Argyris, C., Schon, D.A., 1978. Organizational Learning: A Theory of Action Perspective. Addison – Wesley Series on Organization Development.

Arunraj, N.S., Mandal, S., Maiti, J., 2013. Modeling uncertainty in risk assessment: an integrated approach with fuzzy set theory and Monte Carlo simulation. Accid. Anal. Prev. 55, 242–255.

Basso, B., Carpegna, C., Dibitonto, C., Gaido, G., Robotto, A., Zonato, C., 2004. Reviewing the safety management system by incident investigation and performance indicators. J. Loss Prevent. Process Ind. 17 (3), 225–231.

Brazier, A.J., 1994. A summary of incident reporting in the process industry. J. Loss Prev. Process Ind. 7 (3), 243–248.

Brin, S., Motwani, R., Ullman, J.D., Tsur, S., 1997. Dynamic itemset counting and implication for market basket data. In: International Conference on Management of Data SIGMOD 97, pp. 255–264.

Brown, K.A, Willis, P.G., Prussia, G.E., 2000. Predicting safe employee behavior in the steel industry: development and test of a sociotechnical model. J. Oper. Manage. 18 (4), 445–465.

Cheng, C.-W., Lin, C.-C., Leu, S.-S., 2010. Use of association rules to explore cause–effect relationships in occupational accidents in the Taiwan construction industry. Saf. Sci. 48 (4), 436–444.

Cheng, C.-W., Yao, H.-Q., Wu, T.-C., 2013. Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. J. Loss Prevent. Process Ind.

Cooper, M.D., 2000. Towards a model of safety culture. Saf. Sci. 36, 111–136.

Gnoni, Maria Grazia., Lettera, G., 2012. Near-miss management systems: a methodological comparison. J. Loss Prev. Process Ind. 25 (3), 609–616.

Gnoni, M.G., Andriulo, S., Maggio, G., Nardone, P., 2013. "Lean occupational" safety: AN application for a near-miss management system design. Saf. Sci. 53, 96–104.

Greenwood, M., Woods, H., 1919. The incidence of industrial accidents upon individuals with special reference to multiple accidents. doi: Tech. rep., Industrial Fatigue Research Board.

Haddon, W., Suchman, E., Klein, D., 1964. Accident Research: Methods and Approaches. Harper & ROW, New York.

Han, J., Kamber, M., 2001. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco.

Han, J., Pei, J., Yin, Y., 2000. Frequent pattern tree: design and construction. In: SIGMOD '00 Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, pp. 1–12.

Heinrich, H., 1959. Industrial Accident Prevention. McGraw-Hill, New York.

Hidber, C., 1999. Online association rule mining. In: SIFMOD. ACM, Philadelphia, PA, pp. 145–156.

Institute, S., 2001. SAS Enterprise Miner 4. 3. In: Release 4.1.SAS Institute, Cary, NC.

Ji, C., Zhang, H., 2012. Accident investigation and root cause analysis method. In: 2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, IEEE, pp. 297–302.

Jones, S., Kirchsteiger, C., Bjerke, W., 1999. The importance of near miss reporting to further improve safety performance. J. Loss Prev. Process Ind. 12 (1), 59–67.

Khanzode, V.V., Maiti, J., Ray, P.K., 2012. Occupational injury and accident research: a comprehensive review. Saf. Sci. 50 (5), 1355–1367.

Lee, C., Song, B., Park, Y., 2012. Design of convergent product concepts based on functionality: an association rule mining and decision tree approach. Expert Syst. Appl. 39 (10), 9534–9542.

Liao, C.-W., Perng, Y.-H., 2008. Data mining for occupational injuries in the Taiwan construction industry. Saf. Sci. 46 (7), 1091–1102.

Maher, M.J., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accident models. Accid. Anal. Prev. 28 (3), 281–296.

Maiti, J., Singh, A.K., Mandal, S., Verma, A., 2014. Mining safety rules for derailments in a steel plant using correspondence analysis. Saf. Sci. 68, 24–33.

Mirabadi, A., Sharifian, S., 2010. Application of association rules in Iranian Railways (RAI) accident data analysis. Saf. Sci. 48 (10), 1427–1435.

Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid. Anal. Prev. 43 (4), 1451–1463.

Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2012. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accid. Anal. Prev. 49, 58–72.

Nesmith, G., Keating, J.T., Zacharias, L.A., 2013. Investigating process safety near misses to improve performance. Am. Inst. Chem. Eng. 32 (2).

Okstad, E., Jersin, E., Tinmannsvik, R.K., 2012. Accident investigation in the Norwegian petroleum industry – Common features and future challenges. Saf. Sci. 50 (6), 1408–1414.

Oktem, U.G., Wong, R., Oktem, C., 2010. Near-Miss Management : Managing the Bottom of the Risk Pyramid. Risk & Regulation-Magazine of the ESRC Centre for Analysis of Risk and Regulation, July, 12–13.

Pande, A., Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Saf. Sci. 47 (1), 145–154.

Paul, P.S., Maiti, J., 2007. The role of behavioral factors on safety management in underground mines. Saf. Sci. 45 (4), 449–471.

Reason, J., 1990. Human error. Bmj, New York: Cambridge University Press, vol. 320, pp. 768–770.

Savasere, A., Edward, O., Navathe, S., 1995. An efficient algorithm for mining databases association rules in large. In: 21st VLDB Conference. Zurich, pp. 432–444.

Shankar, W., Mannering, F., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.

Shenoy, P., Haritsa, J.R., Sudarshan, S., Bhalotia, G., Bawa, M., Shah, D., 2000. Turbo-charging vertical mining of large databases. In: MOD 2000. ACM, Dallas, TX, USA, pp. 22–33.

Toivonen, H., 1996. Sampling large database for association rules. In: 22nd VLDB Conference, Mumbai, India, pp. 1–12.

Trist, E.L., Bamforth, K.W., 1951. Some social and psychological consequences of the longxwall method of coal-getting. Human Relat. 4 (1), 3–38.

Zaki, M.J., 2000. Scalable algorithms for association mining. IEEE Trans. Knowl. Data Eng. 12 (3), 372–390.

Zaki, Mohammed J., Parthasarthy, S., Ogihara, M., 1997. Parallel algorithms for discovery of association rules. Data Min. Knowledge Discov. 373, 343–373.

Zhang, C., Zhang, S., 2002. Association Rule Mining: Models and Algorithms. Springer, New York.