Genetic Algorithm-Based Association Rule Mining Approach Towards Rule Generation of Occupational Accidents

Chapter · September 2017

DOI: 10.1007/978-981-10-6430-2_40

Sobhan Sarkar

Indian Institute of Technology Kharagpur

Ankit Lohani

Indian Institute of Technology Kharagpur

Genetic Algorithm-Based Association Rule Mining Approach Towards Rule Generation of Occupational Accidents

Sobhan Sarkar[1(✉)], Ankit Lohani[2], and Jhareswar Maiti[1]

[1] Department of Industrial and Systems Engineering,
Indian Institute of Technology Kharagpur, Kharagpur, India
sobhan.sarkar@gmail.com

[2] Department of Chemical Engineering,
Indian Institute of Technology Kharagpur, Kharagpur, India

Abstract. Occupational accident is a grave issue for any industry. Therefore, proper analysis of accident data should be carried out to find out the accident patterns so that precautionary measures could be undertaken beforehand. Association rule mining (ARM) technique is mostly used in this scenario to find out the association (i.e., rules) causing accidents. But, among the rules generated by ARM, all are not useful. To handle this kind of problem, a new model ARM and genetic algorithm (GA) has been proposed in this study. The model automatically selects the optimal *Support* and *Confidence* value to generate useful rules. Out of 1285 data obtained from a steel industry in India, eleven useful rules are generated using this proposed method. The findings from this study have the potential to help the management take the better decisions to mitigate the occurrence of accidents.

Keywords: Occupational accidents · Steel industry · Association rule mining · Genetic algorithm

1 Introduction

In today's world, advancements in industrialization has led to the establishment of industries and workplaces. In this highly competitive atmosphere, every industry has a number of problems, one of them being industrial accidents. Various circumstances and events lead to occupational accidents in the industry. From prior study, it has been realized that most of these accidents are due to poor safety standards, non—compliance/availability of the standard operating procedures, health of workers and poor machine conditions. Advancements in technology and automations in industries have promoted different types of accidents. Thus, it is required to address the loss of lives and economic resources caused by these occupational accidents. Various studies have been conducted to study the nature of occupational accidents to suggest the ways of preventing

J.K. Mandal et al. (Eds.): CICBA 2017, Part II, CCIS 776, pp. 517–530, 2017.
DOI: 10.1007/978-981-10-6430-2_40

them. With proper analysis of past data, these accidents can be controlled if they can be predicted.

At steel plants, the nature of work pattern, tasks and environment are complex and diverse in nature. Additionally, the transitory nature of factory workforce and workplace makes it more dangerous and prone to accidents. The steel manufacturing process involves the use of high technology and physical labor, making safety management a complicated task. Safety performance has largely been measured and driven by lagging indicators (including injuries, illnesses and fatalities). Different strategies to study the pattern of accidents have been implemented in industries with a view to reduce the number of accidents. This has been very helpful in designing safety standards for the industries. The basis of this analysis is collecting of facts, classifying them, and reporting them precisely in a timely manner. Usually, incident occurs as a culmination of various factors. Some events in the past might also lead to an accident. Hence, various machine learning approach used to identify key association rules between relevant factors is useful to deal with this kind of problem. The association rules can relate various factors which are significant in the occurrence of accidents. One can derive safety measures from this relation to reduce the accidents in future.

In this work, association rule analysis is applied using Genetic Algorithm (GA) to determine a strong level of association among various influencing factors. ARM can give many rules that are irrelevant and needs a manual input of threshold values of *Support* and *Confidence*. But, when implemented with GA, ARM outputs only significant rules with less time complexity and without manually putting any threshold values. The result of this analysis can be used to provide managarial suggestions to the steel plants to take preventive measures from the rules for a safer working environment.

The rest of the paper is organized as follows: Sect. 2 presents a brief literature review on occupational accident analysis using ARM approach, which is followed by research gap and contribution of the present study. In Sect. 3, the methods(i.e. ARM and GA) have been briefly described. A case study from a steel plant has been considered for implementation of the proposed method and is described in Sect. 4. Results and Discussions are illustrated in Sect. 5. Finally, in Sect. 6, conclusion with future scopes is presented.

2 Literature Review

In occupational accident analysis, various methods have been proposed, most of them being parametric models like multivariate models - logistic regression in [1,2]. [3] used the Poisson regression, segmental point process [17], and negative binomial regression techniques in traffic accident analysis. Since there are predetermined assumptions in these models, they give low prediction accuracy. Therefore, researchers developed various non - parametric models like classification and regression tree (CART) model in [4]. A tree-based logistic regression approach for the work zone casualty risk assessment is presented by Cheng et al. [5]. However, these non - parametric models suffer from a great disadvantage of over-fitting. Some studies recently introduced optimization techniques used for parameter tuning of the base algorithm like grid search-based support vector machines (GS-SVM), GA-based SVM [14], GA-CART [15] in occupational accident analysis scenario. Along with these studies, some text mining based approaches have been developed for accident occurrence prediction [16,18].

Association Rule Mining is another famous non-parametric model for safety analysis overcoming the aforementioned problem. Verma et al. worked on finding the accident patterns in a steel plant using association rule mining of incident parameters derieved from investigation reports [6]. Using 10 features it proposed safety actions to be implemented by the company. They pointed out that SOP (Standard Operating Procedure) non-compliance was an important reason for property damage cases. Dehuri et al. in a similar vein, made a similar identification that the root causes of accidents in steel industries are slip/trip/fall (STF), collision/dashing, inadequacy of standard operating procedures (SOPs) and unsafe acts by workers in the workplaces [7].

In case of association rule mining (ARM), there are too many rules generated for a given threshold of parameters. Hence, the analysis becomes cumbersome for management and some major problems are not provided significant importance. Optimization of the model hence becomes a necessity. [8] proposed a rule mining method called multi objective genetic algorithm (MOGA). However, they concentrated only on developing the algorithm and tested it on various datasets. The outcomes of the algorithm were a set of non-dominated solutions. But this method was slow, so, they improved the performance by parallel processing GA (genetic algorithm). In 2008, Dehuri et al., used crossover and mutation methods to modify the solutions using their elitist multi-objective genetic algorithm (EMOGA) [11]. They used Pareto-based rank for fitness evaluation of chromosomes. Some researches were carried out on developing algorithms for rule mining that was based on GA without providing the value of minimum *Support* [9,10]. They used relative *Confidence* as their fitness function and implemented it using frequent pattern (FP) tree. Qodmanan et al. proposed multi objective ARM with GA without specifying threshold values for *Support* and *Confidence* which worked faster than other heuristic approaches [12]. In one of the latest works, Cococcioni et al. presented a semi-supervised learning-aided evolutionary method in [13]. Consistency constraint has te be met in the pair of classifiers designed, between a worker's risk perception with respect to a task and the level of caution of the same worker for the same task.

In all the aforementioned works, ARM technique has been successfully used as a tool in the analysis of accidents. Some of the researchers used variety of optimization techniques to enhance the process. However, to the best of authors' knowledge, similar work in safety analysis has not been done till date. Thus, our present work aims to bridge this gap in occupational accident literature to ensure a safer working environment for the workers.

2.1 Challenge

Based on the review of literature, it has been identified that the research using association rule mining approach on occupational accident domain is very less. Moreover, some studies have also reported the difficulty and have imposed a challenge in handling the huge number of rules generated from ARM approach as this method is more sensitive towards the selection of parameters like *Support* and *Confidence*. Therefore, optimal rule generation is required through proper optimization of parameters of ARM, which is not reported in occupational accident research so far.

2.2 Proposed Work

The present study proposed a new method i.e., GA optimized ARM for the rule generation of incident cases in steel industry. To the best of the authors' knowledge, none of the previous studies have reported this kind of analysis so far in occupational accident domain.

3 Methods Used

In this section, the methods ARM and GA have been described briefly. The proposed methodological flowchart is depicted in Fig. 1. It shows the steps from collection of raw data, pre-processing, feature selection to final analysis. Thereafter, in ARM, which parameters are optimized by GA, has been implemented on the pre-processed data which gives the optimal rules describing the incident outcomes. The entire process of ARM and GA is described below.

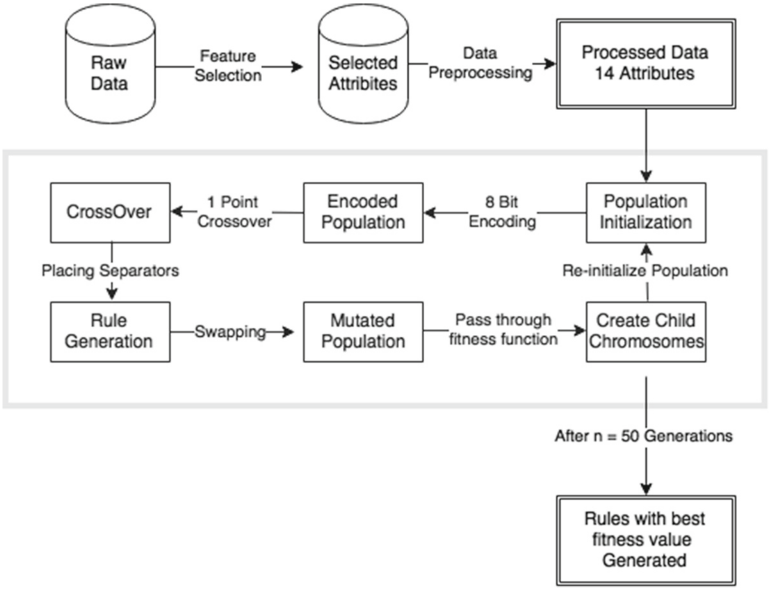

Fig. 1. Proposed methodological flowchart of GA-ARM approach.

3.1 Association Rule Mining

Introduced as an unsupervised approach, Association rule mining (ARM) was used to find patterns in large data sets. The method derived gets its identity from a method where there is a set of transactions from a shop and each transaction is considered a market basket consisting a set of items purchased. A particular item may or may not be a part of the transaction. So, there lies a pattern of rules in the form of a transaction. Apriori algorithm is used to mine these rules from the data set. Let I be a universe of items. A data set D is a set of transactions $T_1, \ldots, T_n$, where each transaction is a set of items from I. An item set X is a subset of I. The frequency of X, freq(X), is the number of transactions T in D for which X is a subset of T, then the *Support* can be calculated from the Eq. (1):

$$Supp(X) = \frac{freq(X)}{|D|} \tag{1}$$

An association rule has two parts- antecedent (denoted by X) and consequent (denoted by Y). Former part is an item found in the data while latter is an item that is found in combination of the earlier one. The intersection of the two parts is a null set. The *Support* of the rule is denoted by supp(X U Y). The *Confidence* of the rule is a ratio of how often it is correct to how often it applies and can be expressed as Eq. (2):

$$Conf(XUY) = \frac{supp(XUY)}{supp(X)} \tag{2}$$

Lift is the ratio of *Confidence* to Expected *Confidence*. It can be expressed as Eq. (3). Among association rules if the value of *Lift* is greater than 1.0, it implies that the relationship between the antecedent and the consequent is more significant than would be expected if the two parts were independent. Higher value of *Lift* denotes higher significance.

$$Lift = \frac{supp(XUY)}{supp(X) * supp(Y)} \tag{3}$$

Association rule mining algorithm is used to find a set of rules which are above a minimum value of *Support* and *Confidence* also known as threshold values. Firstly, all item sets having a *Support* value greater than the threshold value are enlisted as frequent. Then, association rules are generated from these item sets and only those rules are kept that have a *Confidence* value greater than the threshold *Confidence* value. The principle behind apriori is that if an item set is frequent, then all of its subsets must also be frequent. The total of $2^I - 1$ item sets can be generated from I items in the dataset, hence, the step to find the frequent item sets is the more complex than other parts.

3.2 Genetic Algorithm

A genetic algorithm (GA) is a heuristic search algorithm that is inspired from the process of natural selection proposed by Charles Darwin. It is used to generate useful solution to optimization and search problems to find a global optimum in a defined phase space. Before we proceed to apply GA, it is necessary to *encode* the dataset in any format, such as the bit encoding. The three fundamental principles of the algorithm are: **(a) Selection** - In this primary step, individual genomes are chosen from the population for further steps; An *initial population* is generated in the beginning of the procedure. The selection for further generations are done on the basis of their fitness; **(b) Crossover** - This genetic operator is used to generate the next generation of population based of the fitness calculated using the fitness function. This *fitness function* measures the quality of the solution. Since the best chromosomes from the earlier generation is used, the average fitness is greater than the last generation; and **(c) Mutation** - This operator maintains genetic diversity among various generations. In mutation, the solution may entirely change from previous solution. This step overcome the problem if the fitness value at some point is stuck at the local minima.

A combination of these two algorithms is used in this paper for occupations accident analysis in steel industry as described in the next section.

4 Case Study

4.1 Problem Statement and Motivation

The case company is facing a serios issue of occupation accidents occuring at its workplaces. The cost of human live is immense and of utmost importance. Instead of having all precautionary measures, the incidents have been taking place in an uncontrolled way. Cause-effect analyses, though carried out, has not been successful enough to check the occurrences of accidents. Thus, accident patterns in terms of association rules are necessary to be identified so that it can be used to take several precautionary steps. Under these circumstances, application of ARM is realized to be of great value to provide the best managerial suggestions.

4.2 Data Collection and Data Types

The data set for analysis has been collected from a steel plant in India from 24 different divisions in the span of three years (2010–2013). For our analysis, we have used 1285 data points or observations from a particular department of a certain division of the plant.

The data set consists of categorical attributes and each of them has multiple classes. The attributes used in analysis have been enlisted in Table 1 along with the number of classes each feature possesses. They are -

Table 1. Attributes and their corresponding number of classes in the data set used in the study.

Attribute	Number of classes
Injury Type	8
Primary Cause	21
Status	6
Working Condition	3
Machine Condition	3
Observation Type	4
Employee Type	3
Serious Process Incident Score (SPI)	3
Injury Potential Score(IP)	3
Equipment Damage Score(EDS)	3
Safety Standards	18
Incident Category	2
Incident Type	6
Standard Operating Procedure (SOP)	3

(i) **Primary Cause** - This attribute implies the causes behind the accident. There are 21 possible causes of any incident described in the dataset - Crane Dashing, Dashing/Collision, Derailment, Electrical Flash, Energy Isolation, Equipment Machinery Damage, Fire/Explosion, Gas Leakage, Hot Metals, Hydraulic/Pneumatic, Lifting Tools Tackles, Material Handling, Medical Ailment, Occupational Illness, Process Incidents, Rail,Road Incident, Run Over, Skidding, Slip/Trip/Fall, Structural Integrity.

(ii) **Status** - This represents the current state of the investigation of incident case. It may be either close or open.

(iii) **Working Condition** - This attribute implies the condition of work which may be Single Working(SW), Group Working(GW) or Not applicable (Napp).

(iv) **Machine Condition** - Machine failure is a signicant cause of accidents which can be avoided. It represents the machine condition after the incident, i.e. if Idle machine condition(MI), the machine was working while accident occured or Not applicable(Napp) if the accident was not around any machine

(v) **Incident Type** - The accident can occur because of human error or as a reason of an inherent problem with the process. Thus, it can be classified as - Behavorial (Beh); Process (Pro)

(vi) **Employee Type** - There are two types of workers in the workplaces. They have different roles and any accident depends on the work they are performing. On site, Contractor, Employee are the two employee types;

(vii) **Serious Process Incident Score (SPI)** - The incident observed is scored on the basis of severity in three levels, namely - lowspi, mediumspi and highspi

(viii) **Injury Potential Score(IP)** - The injury caused to the causalitites are also categorized in three levels - lowip, mediumip and highip to classify the seriousness of the accident

(ix) **Equipment Damage Score(EDS)** - Many accidents involve damage to machines and equipments too. Based on the level of damage caused to these machines, every accident is associated with an equipment damage score as - loweds, mediumeds and higheds

(x) **Observation Type** - Various safety standards are defined for every activity in these workplaces to maintain a definite protocol of safety. These are categorized as - Unsafe act unsafe condition(UAUC), Unsafe Act(UA), Unsafe condition(UC)

(xi) **Injury Type** - This attribute indicates the type of injury incurred. It has following types - Fatal, First Aid, IOW, No injury, Normal, Injury, Serious Injury.

(xii) **Incident Category** - The incident has been reported as - Property Damage(PrDm), NearMiss(NM) or Injury (Inj) with their meaning as their name depicts;

(xiii) **Standard Operating Procedure (SOP)** - Every process and activity in an industry has a SOP associated with it and are categorized as available but not followed (SANF), available and followed (SF) while in some cases it might mot be required or cannot be proposed like SOP not required (SNR), SOP not available (SNA).

(xiv) **Safety Standards** - Industries follow wide range of safety techniques but still incidents occur. So they form an important criterion in the study to identify how resourceful they have been in preventing accidents. There could be various practices like - barricading, confined space, dismantling, electrical safety, excavation, fire safety, gas cutting & welding, material handling, mines safety, positive isolations, personal protective equipment (PPE), process safety, road safety, wiring, tools & equipment and work permit system.

4.3 Data Pre-processing

In this stage, data cleaning involving missing data handling, outlier detection, and data reduction by feature selection have been done. Due to the limitation of the page, total steps of data pre-processing task have not been included in the scope of the present study. Once data has been pre-processed, it was encoded for application of GA.

4.4 Encoding

First of all, dataset is encoded to initiate the experimentation of GA. This encoding is used in our solution is binary encoding scheme. Any rule, that is treated

Table 2. Schematic arrangement of a chromosome.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14

as chromosome, is encoded in the format as in Table 2, containing 14 genes per chromosome per rule.

A list of 15 items can be used to describe a rule. Out of this 15, 14 genes are classes from the different features enlisted in Table 1 and the last one is a separator that is placed somewhere in between. This separator will separate antecedent from consequent. Any attribute can either be a part of the rule or not. If it is a part of the rule then it can have one out of the multiple classes it has over the whole dataset. Table 1 shows the list of 14 attributes that has been used in the analysis and different classes each attribute can have.

For example, injury potential is one of the features. It has three different classes namely lowip, mediumip, highip. Say, 3 bits are used to encode this feature. Then, we have $2^3 = 8$ combinations of bits. We use three of them to encode the classes of injury potential, namely 000 (lowip), 001 (mediumip), 010 (highip) respectively. In any rule, it can have one of the three values encoded in binary form, or it might be a null value, i.e., it is not a part of the rule. The remaining 5 combination of 3 bits 100, 110, 101, 011, and 111 are assigned a value phi (meaning - null). The class phi signifies that this feature will not be a part of the generated rule during Roulette Wheel selection. The probability of selection of a particular class is depicted in Table 3.

Table 3. Selection Probability of various classes.

Injury Potential Score	Probability of selection for rule
Lowips	1/8
Mediumips	1/8
Highips	1/8
Phi	5/8

Once the data has been encoded, it is ready to implement with the three fundamental steps of GA. Initially, a random set of initial population is chosen from the dataset.

4.5 Initial Population Generation

The 14 genes or attributes are individually encoded using the schematic described in the previous section. We select the size of initial population for our analysis. Hence, our initial population consists of 500 rules/chromosomes. Value of each attribute in every chromosome is selected using Roulette Wheel

Selection; for example, Injury Potential Score depicted in Table 3. Each gene in a chromosome has a significant probability of not being a part of the rule. These probabilities have been decided by the number of bits used for its encoding. If there are more values for an attribute, it has less *Support* and it will be a part of less frequent itemset. For example, primary cause can take 21 different values and its *Support* will be less as compared to an attribute with less diversity. Hence, probability of its existence in the rule will be less as compared to the other one. This population undergoes crossover among itself and a new population or rules are generated.

4.6 Crossover

It is one of the three operators used in genetic algorithm. After generation of initial population, we have 14 genes in each chromosome denoted by respective number of bits as shown in Table 4. There can be two strategies for crossover. We can use *point crossovers* over the whole chromosome. This method has a drawback. Every gene will not have proper crossover and a particular combination of gene is retained and it is simply repositioned in the list. Thus, after few generations, offsprings similar to the parents are generated and the whole point of having diversity in genetic algorithm fails. Hence, we opted for strategy two. In this, we take two chromosomes and operate a 1-point crossover for every gene. This method ensures proper diversity and creation of new rules over every iteration. The two strategies can be explained through Tables 4, 5 and 6.

Table 4. Before crossover of a selected rule.

1001	100010	1100	010	011	000	111	101	110	011	101101	10	1101	110

Table 5. Crossover strategy 1: About middle point of chromosome.

101	110	011	101101	10	1101	110	1001	100010	1100	010	011	000	111

Table 6. Crossover strategy 2: about mid-point of each gene in chromosome.

110	101	011	011110	01	0111	011	0110	100001	0011	001	011	000	111

Placing Separators for Rule Generation: In this step, a separator is placed at one of the random positions in the chromosome generated after crossover. All the attributes which got a value phi after crossover, are dropped at this stage and a rule showing (antecedent → consequent) is left. Thus, rules generated from the previous population are ready. The unfit rules are then segregated from them using the fitness function (see Eq. (4)).

Selection of Population for Next Generation: Association rule mining using genetic algorithm gives many rules. Many of them are insignificant because they have less *Support* for the itemset or less *Confidence* for the rule. We eliminate these weak offsprings by defining a new fitness function that has been described in the paper Qodmanan et Nasiri [12]. The function is

$$fitness = \frac{(1 + supp(XUY))^2}{1 + supp(X)} \tag{4}$$

In this equation, sup(X U Y) is the *Support* of X Y and supp(X) is the *Support* of antecedent part of it. We move toward the third fundamental step of GA - mutation, to prevent the algorithm from stucking at any local optima.

4.7 Mutation

This operator is used to create diversity in the population from the initial population, with the use of this method, such as swapping, we can change the *Confidence* of the rule. We can do the same by swapping the position of separators in the two genes. The algorithm was applied taking an initial population of 1000 chromosomes and rules over 50 generations were generated.

Table 7. Rules generated using GA-ARM approach.

Antecedent	Consequent	Confidence	Lift
PRI, SW, lowspi	NM	1	1.46
SW, UAUC	highip	0.705	0.867
MI, mediumip	Loweds	0.1	19.32
SI	Beh	0.625	0.761
UAUC, contractor	Mediumip	0.33	2.9
Close, SW, MI, mediumip	SANF	0.625	1.6309
UAUC, mediumspi	SNR	0.25	1.98
EI, Pro	SNR	0.14	1.133
Mediumip, loweds	PrDm	1.0	35.69
MI, UA, Contractor	Lowspi, beh	1.0	1.221
Napp	Mediumip	0.117	1.02

5 Results and Discussion

In the present study, GA-based ARM method has been applied in order to generate rules for the occurrence of occupational accidents in steel plant. To start with GA, encoding of the data set has been performed which is followed by the sequential steps i.e., initial population generation, crossover, mutation, and finally ARM. The average of fitness values of all rules for a given generation is calculated and plotted against the iteration number. The average fitness value of all rules in a particular generation has been plotted against the generation number. Figure 2 clearly depicts the output. The best among the 50 generations is selected for analysis. For a particular run, the fifth generation showed the best fitness value for the given generation. The following rules (see Table 7) were deduced.

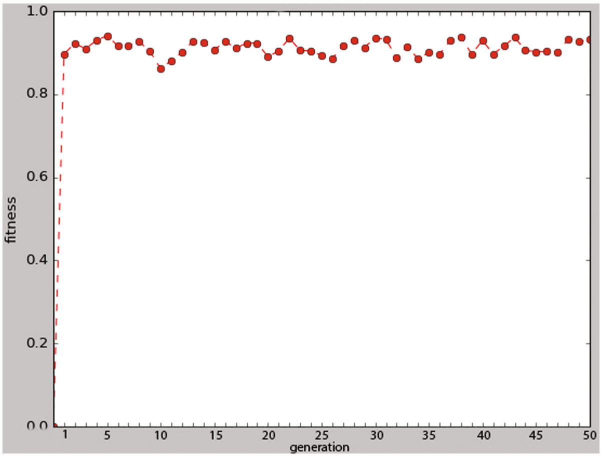

Fig. 2. Graph of generations vs average fitness obtained from GA-ARM approach.

The generated rule [MI, mediumip → loweds] signifies that for accidents that occurred in case of idle machine condition (MI) with average injury potential (mediumip), the equipment damage score (loweds) was also less. From the rule [mediumip, loweds → PrDm], it can be concluded that there are high chances that accidents involving average injury potential (mediumip) and low equipment damage score (loweds) will have high property damage (PrDm). One of the rules, [SI → Beh] depicts the higher risk of having behavioral (beh) accidents with primary cause being structural integrity (SI). Next rule, [Close, SW, MI, mediumip → SANF], tells that in single working conditions (SW) and idle machine (MI) conditions has significant injury potential index (mediumip) even though the cases are closed because of non-availability of SOPs (SANF).

6 Conclusion

From the analysis of rules, it can be inferred that steel industry under study is still not capable of handling hazards successfully to avoid major accidents. The rules that have been derived after this analysis depicts that injury potential of accidents is significant. Cases of unsafe act and unsafe condition (UAUC) are found to be the reason for most of cases.

This work can be further extended to a comprehensive comparision with various other state of art optimization methods, like particle swarm optimization (PSO), ant colony optimization (ACO) etc. The proposed method can be further tuned for a few more values of crossover and mutation probabilities for a intuitive analysis of GA over ARM. Another important future direction can be prediction-based collective class classification rule mining. Structured association map (SAM), another powerful visualization technique, could be used for better visualization of the association rules. Another future direction can be the generation of high-utility item sets from the occupational accident database which could provide potential benefit to the safety manager of any organization.

Acknowledgement. The authors are thankful to the safety personnel for their kind support from the collection of data to the final evaluation phase of the project.

References

1. Bedard, M., Guyatt, G.H., Stones, M.J., Hirdes, J.P.: The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accid. Anal. Prev. **34**(6), 717–727 (2002)
2. Li, Y., Bai, Y.: Development of crash-severity-index models for the measurement of work zone risk levels Accid. Anal. Prev. **40**(5), 1724–1731 (2008)
3. Wong, S.C., Sze, N.N., Li, Y.C.: Contributory factors to traffic crashes at signalized intersections in Hong Kong Accid. Anal. Prev. **39**(6), 1107–1113 (2007)
4. Chang, L., Wang, H.: Analysis of traffic injury severity: an application of nonparametric classification tree techniques. Accid. Anal. Prev. **34**(5), 1019–1027 (2006)
5. Weng, J., Meng, Q., Wang, D.Z.: Tree-Based logistic regression approach for work zone casualty risk assessment. Risk Anal. **33**(3), 493–504 (2013)
6. Verma, A., Khan, S.D., Maiti, J., Krishna, O.B.: Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports. Saf. Sci. **70**, 89–98 (2014)
7. Cheng, C.-W., Yao, H.-Q., Wu, T.-C.: Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. J. Loss Prev. Process Ind. **26**, 1269–1278 (2013)
8. Dehuri, B., Ghosh, A.: Muti objective association rule mining using genetic algorithm. Inf. Sci. **163**, 123–133 (2004)
9. Yan, X., Zhang, C., Zhang, S.: Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support. Expert Syst. Appl. **36**(2), 3066–3076 (2009)
10. Cheng, C.W., Lin, C.C., Leu, S.S.: Use of association rules to explore cause effect relationships in occupational accidents in the Taiwan construction industry. Saf. Sci. **48**(4), 436–444 (2010)
11. Dehuri, S., Patnaik, S., Ghosh, A., Mall, R.: Application of elitist multiobjective genetic algorithm for classification rule generation. Soft Comput. **8**(1), 477–487 (2008)
12. Qodmanan, H.R., Nasiri, M., Minaei-Bidgoli, B.: Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence. Expert Syst. Appl. **38**(1), 288–298 (2011)

13. Cococcioni, M., Lazzerini, B., Pistolesi, F.: A semi-supervised learning-aided evolutionary approach to occupational safety improvement. In: IEEE Congress on Evolutionary Computation (2016)
14. Sarkar, S., Vinay, S., Pateshwari, V., Maiti, J.: Study of optimized SVM for incident prediction of a steel plant in India. In: 2016 IEEE Annual India Conference (INDICON), pp. 1–6. IEEE, December 2016
15. Sarkar, S., Patel, A., Madaan, S., Maiti, J.: Prediction of occupational accidents using decision tree approach. In: 2016 IEEE Annual India Conference (INDICON), pp. 1–6. IEEE, December 2016
16. Sarkar, S., Vinay, S., Maiti, J.: Text mining based safety risk assessment and prediction of occupational accidents in a steel plant. In: 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), pp. 439–444. IEEE, March 2016
17. Gautam, S., Maiti, J., Syamsundar, A., Sarkar, S.: Segmented point process models for work system safety analysis. Saf. Sci. **95**, 15–27 (2017)
18. Brown, D.E.: Text mining the contributors to rail accidents. IEEE Trans. Intell. Transp. Syst. **17**(2), 346–355 (2016)