Applying Association Rules Mining Algorithms for Traffic Accidents in Dubai

Amira A. El Tayeb, Vikas Pareek, Abdelaziz Araar

Abstract*— *Association rule mining algorithms are widely used to find all rules in the database satisfying some minimum support and minimum confidence constraints. In order to decrease the number of generated rules, the adaptation of the association rule mining algorithm to mine only a particular subset of association rules where the classification class attribute is assigned to the right-hand-side was investigated in past research. In this research, a dataset about traffic accidents was collected from Dubai Traffic Department, UAE. After data preprocessing, Apriori and Predictive Apriori association rules algorithms were applied to the dataset in order to explore the link between recorded accidents' factors to accident severity in Dubai. Two sets of class association rules were generated using the two algorithms and summarized to get the most interesting rules using technical measures. Empirical results showed that the class association rules generated by Apriori algorithm were more effective than those generated by Predictive Apriori algorithm. More associations between accident factors and accident severity level were explored when applying Apriori algorithm.

Index Terms*— *Association Rule Mining, Apriori, Predictive Apriori, Dubai Traffic Accidents

I. INTRODUCTION

Data mining uses various techniques and algorithms to discover knowledge from huge amounts of data and identify understandable patterns from data. It is considered as one of the most important trends in information technology in the preceding decade [1]. Association rule mining algorithms are used to find all rules in the database satisfying some minimum support and minimum confidence constraints [2]. Classification rule mining aims to find a small number of rules in the database to form an accurate classifier [3]. A small set of rules are generated by standard classification algorithms to form a classifier, but these algorithms use domain independent biases and heuristics [3]. The Apriori Algorithm is a powerful algorithm for mining frequent itemsets for boolean association rules [4]. Two separate steps are followed in this algorithm to generate association rules; applying minimum support to find all frequent itemsets in a database, and using these frequent itemsets and the minimum confidence constraint to generate rules [4]. On the other hand, support and confidence constraints are combined into a single measure known as accuracy which is used to generate the Apriori association rules in Predictive Apriori algorithm [5]. In WEKA, these algorithms are used to generate "n" best association rules based on the number of rules "n" determined by the user [5].

Revised Version Manuscript Received on July 08, 2015.

Eng. Amira A. El Tayeb, PhD. Scholar, Department of Computer Science, Banasthali University, Rajasthan, India

Dr. Vikas Pareek, Assoc. Prof., Department of Computer Science, Banasthali University, Rajasthan State, India.

Dr. Abdelaziz Araar, Assoc. Prof., College of Information Technology, Ajman University, Ajman, UAE.

In order to decrease the number of generated rules, the adaptation of the association rule mining algorithm to mine only a particular subset of association rules known as the class association rules (CARs) was investigated in past research [3]. This technique is integration between two data mining techniques; association rule mining and classification rule mining [6]. The integration is applied to focus on a special subset of association rules where the classification class attribute is assigned to the right-hand-side [3]. It consists of two parts; a rule generator using Apriori algorithm for discovering association rules, and a classifier builder [7]. Empirical results showed that big savings and more accurate classifiers could be built this way than those generated by the classification algorithm C4.5 [3]. In addition, this integration helped in solving a number of problems in the classification algorithms [3]. Data mining has widely been implemented in many real-life domains. Many researchers highlighted the significance of applying data mining techniques in the traffic accident field in order to extract patterns that can reduce the severity of the accidents. According to a traffic report generated by the World Health Organization (WHO), millions of road traffic accidents take place worldwide annually, and the fatalities due to these traffic accidents are also in the millions [8]. More than 1.2 million people are killed every year due to road traffic accidents, and around fifty million get injuries or disabilities [8]. According to the centers for Disease Control and Prevention (CDC), road traffic accidents cost $100 billion in medical care every year [9]. A car accident victim is treated in an emergency room suffering from accident injuries every 10 seconds and almost 40,000 deaths are killed annually due to traffic accidents [9]. With the increased number of vehicles accompanied by the rapid expanding road construction executed programs, UAE experiences increasing number of traffic accidents with injuries and fatalities causing a serious public health problem [10]. In UAE, there are about 600 people killed in car accidents each year [11]. Road traffic accidents are the second major cause of deaths in the United Arab Emirates [11]. The costs of fatalities and injuries due to road traffic accidents have a great impact on the society. Previous studies showed that casualty and fatality rates in UAE and in other Gulf countries are much higher than in the developing countries with comparable vehicle ownership levels [10]. Dubai in particular suffered a loss of Dh4.7 billion due to road traffic accidents in the last 10 years. Traffic accidents are not only resulting in loss of lives and injuries but are also badly affecting the emirate's economy. Traffic accidents in 2007 caused an economic loss of some Dh720 million which is around one percent of the gross domestic product (GDP) of Dubai [11]. Road traffic accidents problem needs more research to discover associated accidents risk factors and

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

identify new methods to reduce the large number of accidents and fatalities. Applying data mining into traffic accidents domain still needs further research. This research work can assist in discovering interesting rules from a set of generated rules using both association rules algorithms. Empirical results can assist police decision makers in the formulation of new policies and traffic rules from some hidden patterns. Conclusion and future work are generated at the end.

II. RELATED WORK

Various studies and many pieces of research have focused on predicting the factors in road traffic accident severity using data mining techniques. Previous research has been reviewed for the period between 2001 and 2015. Ossenbruggen and Pendharkar et al. (2001) applied a logistic regression model for discovering the factors of accidents and accident-related injuries. The models used to make a risk assessment of a specific region included attributes such as design of roadside, land use activity, traffic control devices usage, and exposure to traffic. Their study concluded that village sites were less dangerous than residential or shopping sites [12]. Sohn and Hyungwon (2001) conducted research on road traffic accidents (RTA) severity in Korea. The researchers applied three techniques of data mining; decision tree, neural network and logistic regression, investigated a number of significant factors, generated classification models for accident severity, and compared the accuracy of their three classifications [13]. Sohn and Lee (2002) analyzed how road traffic accidents severity and driving environment factors are related to each other using several techniques and algorithms. The researchers applied neural network, decision tree classifiers, classifier fusion based on the Dempster–Shafer algorithm, the Bayesian procedure, logistic model, and clustering using k-mean algorithm. Their experimental results showed that clustering technique was better than other techniques [14]. Ng, Hung and Wong (2002) combined regression analysis, geographical information system (GIS), and cluster analysis techniques for grouping homogeneous accident data, predicting traffic accidents number, and assessing road traffic accidents risk in Hong Kong. Their algorithm performed efficiently for fatal accidents and pedestrian-related accidents. Their proposed algorithm could help authorities identify high risk areas and could help town planners to plan more safe roads [15]. Bedard et al. (2002) used a multivariate logistic regression technique to determine how driver, crash, and vehicle characteristics could independently contribute to drivers' fatality risk. It was concluded that increasing the use of seatbelt, reducing speed and severe impacts of driver might avoid fatalities [16]. Miao M. Chong, Ajith Abraham, Marcin Paprzycki (2004) studied the severity of traffic accidents injuries using decision trees and neural networks. Their research showed that the usage of driver's seat belt, driver's alcohol, and light condition of the roadway are the most three significant factors leading to fatal injury [17]. Chang and Chen (2005) built tree-based models to analyze freeway accident frequency using accident data of National Freeway 1 covering the years 2001-2002 in Taiwan. The authors generated classification and regression tree (CART) and negative binomial regression models to investigate experimentally how traffic accidents, traffic characteristics, highway geometric variables, and environmental factors were related to each other. The authors found out that rain variables and the average volume of daily traffic and were the key factors of freeway accident frequency [18]. Tibebe (2005) analyzed historical road traffic accident records at the Addis Ababa Traffic Office to explore the suitability of applying data mining technology to accident severity analysis in Addis Ababa, Ethiopia. The developed model classified accident severity into fatal injury, serious injury, slight injury, and property-damage using the Knowledge SEEKER algorithm of the Knowledge STUDIO data mining tool and the decision tree technique. The basic factors for the level of injury severity were accident cause, accident type, road condition, vehicle type, light condition, road surface type, and driver age [19]. Chang and Wang (2006) analyzed accident data covering the year 2001 for Taipei, Taiwan by applying non parametric classification tree techniques. The relationships between driver/vehicle characteristics and injury severity, accident variables and highway/environment variables were investigated by developing a CART model. The most significant factor associated with crash severity was vehicle type [20]. Srisuriyachai (2007) used Simple K-Means clustering technique and J48, NaïveBayes, and One R classification algorithms in WEKA to analyze traffic accidents in the Nakhon Pathom province of Bangkok. The resultant traffic accident profiles could be used as a useful tool for analyzing road accidents in Nakhon Pathom [21]. Wong and Chung (2008) applied different methodology approaches and compared them to discover accident severity factors. The authors first used a set of theories to analyze traffic accident data to find out whether the data had complete information about the circumstances associated with the occurrence of accidents, and then compared these derivative circumstances. Their results showed that fatal accidents resulted from a combination of factors. Furthermore, rules with high or low support showed different features [22]. Zelalem (2009) studied how to classify levels of driver's responsibility in road accidents in Addis Ababa. The significant factors influencing the driver's responsibility level were investigated. Applying ID3, J48, and MLP algorithms, models were built using WEKA. Rules generated from the decision tree showed major relationships between driver's level of responsibility variables such as age, driving experience, license grade, education, , and some environmental factors [23]. Getnet (2009) studied how to develop models using data mining to identify and predict main risk factors related to driver and vehicle that cause road traffic accidents. J48 decision tree algorithm and PART rules induction algorithm techniques were applied in WEKA. Experience, vehicle type, and service year were the most significant variables for predicting severity of accident [24]. Sami Ayramo, Pasi Pirtala, Janne Kauttonen, Kashif Naveed, Tommi Karkkainen (2009) analyzed road traffic accidents on the Finnish roads using data mining techniques. The dataset consisted of 1203 fatal accidents. The authors investigated the usability of association rules, clustering, and visualization methods to analyze road traffic accident. Results showed that selected data mining techniques could produce comprehensible patterns from the data. Nearly all fatal accidents occurred where the speed limit was between 80-100km/h on single roadway on main roads outside

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

urbanized areas. Although middle-aged drivers caused most fatal accidents, young drivers caused the highest number of non-fatal traffic accidents [25]. Tibebe Beshah and Shawndra Hill (2010) investigated the link between recorded road factors to accident severity using data mining techniques in Ethiopia. Different classification models were developed using NaïveBayes, k-nearest neighbor, and decision tree classifiers in WEKA. Generated rules showed that the severity of accident varied with various combinations of road-related variables [26]. Galvão ND and de Fátima Marin H. (2010) studied 139 real pairs of victims of traffic accidents data at the city of Cuiabá-MT using data mining technology. Apriori algorithm in WEKA data mining tool was applied. Among the 10 best rules generated, six of them were so useful and understandable for characterizing the victims of road accidents in Cuiabá. The results highlighted the need of prevention measures for accidents caused by males [27]. Amirhossein Ehsaei and Dr. Harry Evdorides (2011) used the Bayesian classifiers in WEKA and generated rules to describe the interrelationship between the different techniques of data mining. Findings indicated that more accidents occurred in daylight than in darkness, the probability of fatal accidents to occur in darkness was double the possibility of serious injury accidents to occur, and the possibility of serious and fatal accidents to occur on wet surfaces was declined over the five-year period being considered [28]. S.Krishnaveni and Dr.M.Hemalatha (2011) classified the type of injury severity of road traffic accidents by comparing different classifiers such as AdaBoostM1 Meta, Naive Bayes, PART Rule, J48 Decision Tree, and Random Forest Tree. They focused on Hong Kong's road traffic accident dataset of 2008 for drivers' records. The attributes selected were of three categories; accident, vehicle, and casualty. The final results showed that the Random Forest performed better than other four algorithms [29]. S.Krishnaveni and Dr.M.Hemalatha (2011) also applied various classification algorithms to predict the patterns related to vehicle collision occurred in training accident dataset. C4.5, RndTree, CS-MC4, C-RT, Decision List, Naïve Bayes and ID3 classification algorithms were applied to predict vehicle collision patterns. Experimental results indicated that RndTree classification algorithm outperformed the other algorithms in classifying collision manner, and the Feature Ranking method improved the accuracy of classifiers [30]. Beshah, T., Ejigu, D., Abraham, A., Snasel, V., Kromer, P. (2011) focused on exploring and predicting the role of road users on injury risks. Road accident data were collected from Addis Ababa Traffic Office. Classification and Adaptive Regression Trees (CART) and RandomForest approaches were applied to identify related patterns and demonstrate the performance of these techniques for the road safety domain. Results showed that the models could classify traffic accidents with reasonable accuracy [31]. Vandana Munde, Sachin Deshpande, and S.K.Shinde (2012) analyzed database incidents in Maharashtra and spatial data obtained from the Maharashtra police using clustering and visualization using Google earth considering road conditions and weather conditions. Their research could effectively help local agencies allocate resources to improve safety measures in sites with high accidents frequency causing fatalities or injuries. It could also provide information for highway engineers and transportation planners to design safer roads [32]. Abdelaziz Araar and Amira A. El Tayeb (2013) analyzed road traffic accidents data in the emirate of Dubai, UAE. The authors collected a dataset covering accidents between 2008 and 2010. Four classification methods (Decision trees, Rules induction, BayesNet, and MultilayerPerceptron) were applied and compared. Experimental results showed that accidents could be classified with reasonable accuracy using and the neural networks classifier (MutilayerPerceptron algorithm) was the best classifier for all classes [11]. Olutayo V.A and Eludire A.A (2014) applied decision trees and artificial neural networks techniques to find out knowledge about accidents in one of the busiest roads in Nigeria so that to reduce fatalities on highways. They collected continuous and categorical data from Nigeria Road Safety Corporations. Artificial Neural Networks technique was applied to analyze continuous data and Decision Trees technique was applied to analyze categorical data. Experimental results showed that decision tree approach outperformed the artificial neural network with a higher accuracy rate and a lower error rate, and the three most important causes of accident were over speeding Tyre burst, and loss of control [33]. Rajdeep Kaur Aulakh (2015) proposed an algorithm in his research to reduce the number of generated rules by Apripri algorithm. After applying Apriori algorithm to generate frequent item sets, he used the frequent item sets to generate the association rules. This approach succeeded to reduce the number of generated association rules [34]. Amit Mittal, Ashutosh Nagar, Kartik Gupta and Rishi Nahar (2015) reviewed frequent mining algorithms such as Apriori, FP, and DIC. The authors described each algorithm and compared different frequent mining techniques based on different important parameters. Experimental results showed that FP-growth algorithm was the best one among the three algorithms in terms of speed [35]. Dr. S. Vijayarani and Ms. R. Prasannalakshmi (2015) conducted a research on the use of the traditional algorithms for generating association rules in data streams. They applied Frequent Item sets, Assoc Outliers, and Supervised Association Rule using Tanagra data mining tool. The major performance factors were the number of generated rules and execution time. Empirical results showed that Frequent Item set was most efficient algorithm [36].

III. SIGNIFICANCE OF THE RESEARCH

The significance of this research is in proposing an approach using association rules mining algorithms to deliver interesting rules from a large set of discovered rules extracted from Dubai traffic accidents real data. Figure1 depicts the methodology used in the research.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

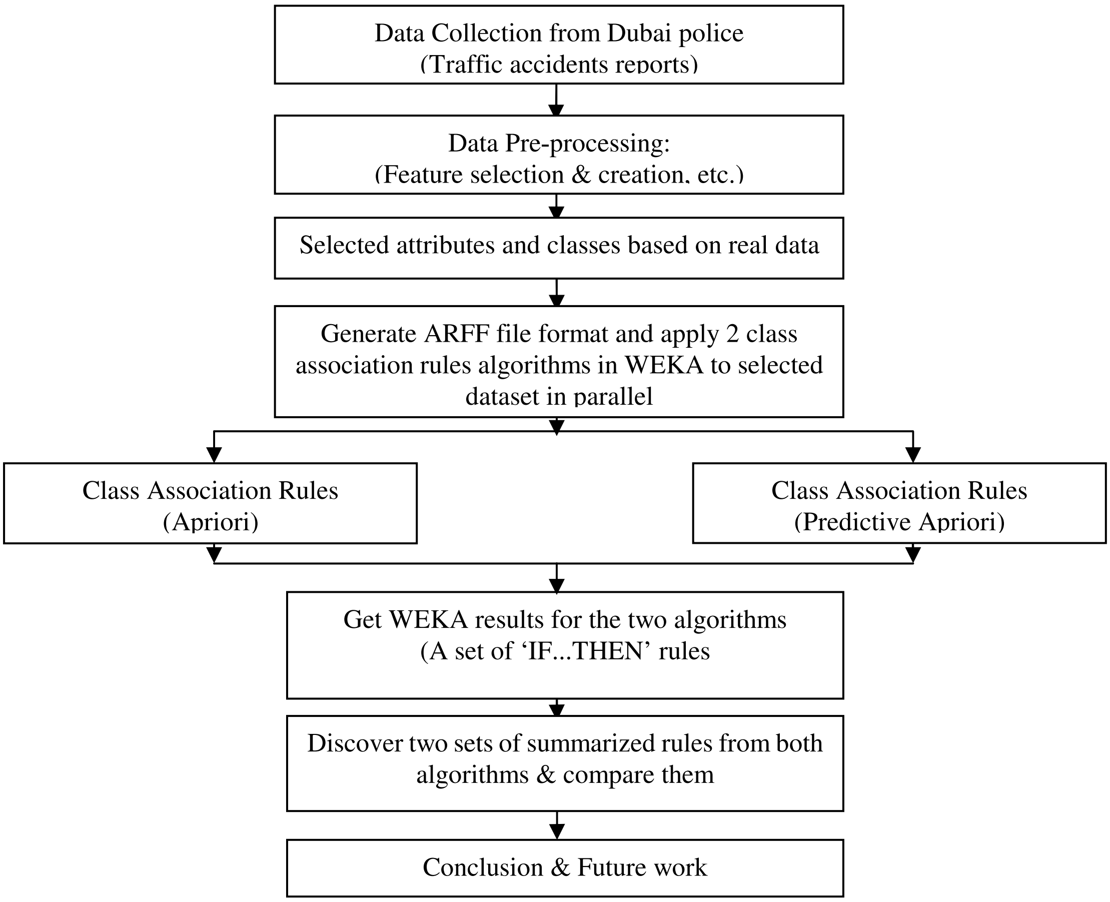

Figure 1. Summary of research methodology

As shown above, data related to traffic accidents were collected from the Dubai police authority. A number of interviews with expert domain were conducted. Collected data were then cleaned and preprocessed involving many tasks such as data aggregation, sampling, feature selection and creation, dimensionality reduction, and handling of missing values [30]. The data mining tool, WEKA, was used to apply association rules data mining techniques on the dataset. The cleaned dataset contained a number of attributes of three types, driver-related attributes, road-related attributes, and accident-related attributes. Attributes and class labels are created based on the real data. The class label ('Accident Severity') had three nominal values: ('Death', 'Severe', and 'Moderate') [11].

IV. ASSOCIATION RULES DATA MINING

Association rules are used to find the relationship between data items in a transactional database [37]. Association rules data mining algorithms are used to discover frequent association or correlation in database by generating association rules of the form: If antecedent then consequent where antecedent and consequent are itemsets of one or more items [37]. To generate association rules, two steps are involved. First, minimum support is used to find all frequent itemsets in a database [37]. Second, these frequent itemsets and the minimum confidence measure are used to generate association rules [37]. There are many association rule algorithms, but we focused on two association rule algorithms; Apriori and Predictive Apriori in our research. Both association rule algorithms are used to discover the best combination of accident factors that links to the accident severity level.

A. Apriori Algorithm

Apriori Association rule algorithm is used for mining the frequent patterns in database where support and confidence are the two measures used to measure association rule quality [38]. Support is the percentage of transaction in the database that contains XUY for the association rule X->Y [38]. Confidence is the ratio of the number of transaction that contains XUY to the number of transaction that contains X [38].

B. Predictive Apriori Algorithm

Predictive APRIORI association rule algorithm is also used for mining patterns in database. It differs than Apriori algorithm in that both support and confidence measures are combined into a single measure called predictive accuracy [38].

V. EXPERIMENTAL SETUP

Our research is based on the Real-life traffic accidents records collected from Dubai Traffic Department from 2008 to 2010. We collected 1887 records from Dubai Police. After applying stratified sampling technique on the collected data, 600 records were selected based on the most complete records [11]. However, we only focused in this research on discovering the rules related to three types of classes; Death, Severe, and Moderate. Association rules algorithms are applied in parallel on the selected dataset. Class association rules are generated based on technical interestingness and summarized to find the final set of interesting rules from

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

both algorithms for each class. Data mining approach is used to link recorded accidents factors to accident severity in Dubai using the data mining software, WEKA [11].

A. Data Preprocessing

After collecting 1887 traffic accident records from the traffic department in Dubai Police Authority, data were cleaned and preprocessed [11]. The problem of having many missing values in many records was solved by ignoring all those records with missing or incomplete values. Stratification method was also applied to the dataset to reduce the number of records used for analysis. Stratification is a process of splitting the population into equal groups by taking a sample of each group (stratum) into consideration during the research [39].

B. Selection of Attributes and Records

After pre-processing, 17 attributes covering three types of factors (Accident, Driver, and Road) were developed. Our focus on this research was on three types of severity classes of accidents; death, severe, and moderate. The following tables 1, 2 and 3 illustrate the attributes and class labels selected with their description and selected traffic accidents records [11].

Table 1. Attributes with their Description

Attribute Name	Description
Accid_day	The day in which the accident occurred.
Accid_month	The month in which the accident occurred.
Accid_year	The year in which the accident occurred.
N_of_p_injured	Number of injured persons
Accid_cause	The cause of the accident
Accid_type	The type of the accident
Driv_gen	Driver's gender
Driv_nation	Driver's nationality
Driv_age	Driver's age
Driv_vec_type	Driver's vehicle type causing the accident
Driv_drink	Whether the driver was drunk or not.
Driv_belt	Whether the driver's seat belt was fastened or not
Road_accid_place	The place where the accident occurred
Road_accid_type	Type of the accident road
Road_light_cond	The light condition of the road
Weather_cond	The weather condition when the accident occurred
Road_surface	Whether the surface of the road was dry, wet, sandy, etc.

Table 2. Class Labels

Class Label	Class Description
Death	An accident where one or more persons dies within 30 days of the accident.
Severe	An accident where a person is severely injured and needs intensive care.
Moderate	An accident where one or more persons is injured and kept in hospital for more than 12 hours.

Table 3. Selected Records for 2008-2010 Dataset

Class	Year		
	2008	**2009**	**2010**
Death	17	16	15
Severe	5	17	16
Moderate	29	47	41
Total	51	80	72

C. Summarizing Generated Class Association Rules

Rule Covers method is a technique used to summarize the number of best rules obtained. It is applied to compute a cover for a set of rules with the same consequent [40]. After generating a set of rules has the same consequent. This is already done by generating class association rules based on the class attribute. The greedy algorithm is then being applied to find a cover [40]. For a set of rules R = {r1: A1-> B, r2= A2-> B, rd= Ad-> B}, the number or rules covered by ri is computed. If Ai contains or equal to Aj, then ri covers rj. If a rule ri has maximum cover(ri) where it is the number of rules covered by ri. If ri has maximum cover(ri), then it is added to the rule cover R new after removing and discarding all its covered rules. By applying this algorithm, the rule cover should definitely have at most the original number of rules and every rule is covered. Although this algorithm is not optimal to the number of rules selected, it is fast to classify and summarize the rules in a post-processing stage [40]. In this research, we could summarize the rules generated by Apriori algorithm using Rule Covers method. Applying this method for the rules generated by Predictive Apriori algorithm was not working. No combination of variables was generated within the rules. Therefore, we used another method to summarize the rules generated by Predictive Apriori algorithm using OR to combine the different rules. The final accuracy of the summarized rule can be calculated as follows: accuracy of summarized rule = Sum (accuracy of each combined rule)/ total number of combined rules.

VI. EXPERIMENTAL RESULTS

There are four subsections to focus on when discussing the results related to the analysis; the generated rules discovered from both association rules algorithms for each of the three classes with summary of these rules and the comparison of the rules extracted for each class from each algorithm to find out which algorithm is more suitable for traffic accident analysis. The rules for the years 2008 to 2010 are generated for each of the three classes; death, severe, and moderate using Apriori and Predictive Apriori algorithms. Best twenty rules were first generated for death class, severe class, and moderate class by WEKA. The summarized final sets of rules for each class were then generated using rule covers method. A rule covers method is applied to summarize the best rules generated by Apriori algorithm. It summarizes the rules by removing shorter rules covered in bigger rules. For the rules generated by Predictive Apriori algorithm, we combined the rules and calculated the accuracy for the each summarized rule.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

A. *Analysis of Rules for Death Class*

The best rules generated for Death class using Apriori algorithm and Predictive Apriori algorithms and the set of summarized final rules are shown in the following sections.

A.1 Rules Generated Using Apriori Algorithm

Table 4 presents the best rules generated for death class using Apriori algorithm and Table 5 shows the final set of summarized rules after applying rule covers method.

Table 4: Best Rules for Death Class Using Apriori Algorithm

Rule#	Best Rule
1	If Driv_gen=M 47 ==> Class=Death 47 conf:(1)
2	If Road_surface=Dry 47 ==> Class=Death 47 conf:(1)
3	If Weather_cond=Clear 46 ==> Class=Death 46 conf:(1)
4	If Driv_gen=M Road_surface=Dry 46 ==> Class=Death 46 conf:(1)
5	If Weather_cond=Clear Road_surface=Dry 46 ==> Class=Death 46 conf:(1)
6	If Driv_gen=M Weather_cond=Clear 45 ==> Class=Death 45 conf:(1)
7	If Driv_gen=M Weather_cond=Clear Road_surface=Dry 45 ==> Class=Death 45 conf:(1)
8	If Driv_drink=Not_checked 42 ==> Class=Death 42 conf:(1)
9	If Driv_drink=Not_checked Road_surface=Dry 42 ==> Class=Death 42 conf:(1)
10	If Driv_gen=M Driv_drink=Not_checked 41 ==> Class=Death 41 conf:(1)
11	If Driv_drink=Not_checked Weather_cond=Clear 41 ==> Class=Death 41 conf:(1)
12	If Driv_gen=M Driv_drink=Not_checked Road_surface=Dry 41 ==> Class=Death 41 conf:(1)
13	If Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 41 ==> Class=Death 41 conf:(1)
14	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear 40 ==> Class=Death 40 conf:(1)
15	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 40 ==> Class=Death 40 conf:(1)
16	If N_of_p_injured=1 31 ==> Class=Death 31 conf:(1)
17	If N_of_p_injured=1 Driv_gen=M 31 ==> Class=Death 31 conf:(1)
18	If Driv_vec_type=Private 30 ==> Class=Death 30 conf:(1)
19	If N_of_p_injured=1 Road_surface=Dry 30 ==> Class=Death 30 conf:(1)
20	If Driv_vec_type=Private Weather_cond=Clear 30 ==> Class=Death 30 conf:(1)

Table 5: Summarized rules for Death Class Using Apriori Algorithm

Rule#	Summarized Rule
15	If Driv_gen=M and Driv_drink=Not_checked and Weather_cond=Clear and Road_surface=Dry 40 ==> Class=Death 40 conf:(1)
17	If N_of_p_injured=1 Driv_gen=M 31 ==> Class=Death 31 conf:(1)
19	If N_of_p_injured=1 Road_surface=Dry 30 ==> Class=Death 30 conf:(1)
20	If Driv_vec_type=Private Weather_cond=Clear 30 ==> Class=Death 30 conf:(1)

As stated above, death accidents occurred mostly by male drivers driving private vehicles at clear weather condition and on dry roads. It is not clear whether the driver was drunk or not. One injured person was recorded at the time of accident. When applying rule covers method on the best rules for death class generated using Apriori algorithm, we can observe that amongst the twenty best rules, four rules only appear after eliminating shorter rules covered in longer ones.

A.2 Rules Generated Using Predictive Apriori Algorithm

Table 6 presents the best rules generated for death class using Predictive Apriori algorithm. Table 7 shows the final set of summarized rules after combining the rules.

Table 6: Best Rules for Death Class Using Predictive Apriori Algorithm

Rule#	Best Rule
1	If Driv_gen=M 47 ==> Class=Death 47 acc:(0.98307)
2	If Road_surface=Dry 47 ==> Class=Death 47 acc:(0.98307)
3	If Weather_cond=Clear 46 ==> Class=Death 46 acc:(0.9827)
4	If Driv_drink=Not_checked 42 ==> Class=Death 42 acc:(0.98106)
5	If N_of_p_injured=1 31 ==> Class=Death 31 acc:(0.97427)
6	If Driv_vec_type=Private 30 ==> Class=Death 30 acc:(0.9734)
7	If Road_light_cond=Dark_lighted 28 ==> Class=Death 28 acc:(0.97147)
8	If Driv_belt=Fastened 21 ==> Class=Death 21 acc:(0.9619)
9	If Driv_belt=Unknown 19 ==> Class=Death 19 acc:(0.95793)
10	If Accid_month=Dec 18 ==> Class=Death 18 acc:(0.95564)
11	If Accid_year=2008 17 ==> Class=Death 17 acc:(0.9531)
12	If Accid_year=2009 16 ==> Class=Death 16 acc:(0.95026)
13	If Accid_type=Run_over 16 ==> Class=Death 16 acc:(0.95026)
14	If Accid_year=2010 15 ==> Class=Death 15 acc:(0.94708)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

15	If Driv_nation=India 15 ==> Class=Death 15 acc:(0.94708)
16	If Accid_cause=Lack_of_respect_for_rd_users 13 ==> Class=Death 13 acc:(0.9394)
17	If Accid_cause=Others 12 ==> Class=Death 12 acc:(0.93472)
18	If Accid_day=Tu 9 ==> Class=Death 9 acc:(0.91545)
19	If Accid_day=Th 9 ==> Class=Death 9 acc:(0.91545)
20	If Accid_day=Fr 9 ==> Class=Death 9 acc:(0.91545)

Table 7: Summarized rules for Death Class Using Predictive Apriori Algorithm

Rule#	Summarized Rule
1-7, 10, 13, 15	If Driv_gen=M 47 OR Road_surface=Dry 47 OR Weather_cond=Clear 46 OR Driv_drink=Not_checked 42 OR N_of_p_injured=1 31OR Driv_vec_type=Private 30 OR Road_light_cond=Dark_lighted 28 OR Accid_month=Dec 18 OR Accid_type=Run_over 16 OR Driv_nation=India 15 ==> Class=Death acc:(0.97017)
8,9	If Driv_belt=Fastened 21 OR Driv_belt=Unknown 19 ==> Class=Death acc:(0.95991)
11,12,14	If Accid_year=2008 17 OR Accid_year=2009 16 OR Accid_year=2010 15==> Class=Death acc:(0.95014)
16,17	If Accid_cause=Lack_of_respect_for_rd_users 13 OR Accid_cause=Others 12==> Class=Death acc:(0.93706)
18,19, 20	If Accid_day=Tu 9 OR Accid_day=Th 9 OR Accid_day=Fr 9==> Class=Death acc:(0.91545)

As stated above, death accidents occurred mostly during the three years of study (2008-2010) on Tuesdays or Thursdays or Fridays in December where seat belt is fastened or unknown, and the accident cause was either lack of respect for road users or others. In addition, death accidents mostly occur by male Indian divers driving private vehicles at clear weather condition and dry dark lighted roads. When applying rule covers method on the best rules for death class generated using Predictive Apriori algorithm, we can observe that no rules are eliminated. We can only aggregate them together. We combined the rules by using OR and calculated the accuracy for each summarized rule using this equation; accuracy of summarized rule = Sum (accuracy of each combined rule)/ total number of combined rules.

B. Analysis of Rules for Severe Class

The best rules generated for severe class using Apriori and Predictive Apriori algorithms and the set of summarized final rules are shown in the following sections.

B.1 Rules Generated Using Apriori Algorithm

Tables 8 presents the best rules generated for severe class using Apriori algorithm. Table 9 shows the final set of summarized rules after applying rule covers method.

Table 8: Best Rules for Severe class using Apriori Algorithm

Rule#	Best Rule
1	If Weather_cond=Clear 38 ==> Class=Severe 38 conf:(1)
2	If Road_surface=Dry 38 ==> Class=Severe 38 conf:(1)
3	If Weather_cond=Clear Road_surface=Dry 38 ==> Class=Severe 38 conf:(1)
4	If Driv_gen=M 37 ==> Class=Severe 37 conf:(1)
5	If Driv_gen=M Weather_cond=Clear 37 ==> Class=Severe 37 conf:(1)
6	If Driv_gen=M Road_surface=Dry 37 ==> Class=Severe 37 conf:(1)
7	If Driv_gen=M Weather_cond=Clear Road_surface=Dry 37 ==> Class=Severe 37 conf:(1)
8	If Driv_drink=Not_checked 34 ==> Class=Severe 34 conf:(1)
9	If Driv_drink=Not_checked Weather_cond=Clear 34 ==> Class=Severe 34 conf:(1)
10	If Driv_drink=Not_checked Road_surface=Dry 34 ==> Class=Severe 34 conf:(1)
11	If Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 34 ==> Class=Severe 34 conf:(1)
12	If Driv_gen=M Driv_drink=Not_checked 33 ==> Class=Severe 33 conf:(1)
13	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear 33 ==> Class=Severe 33 conf:(1)
14	If Driv_gen=M Driv_drink=Not_checked Road_surface=Dry 33 ==> Class=Severe 33 conf:(1)
15	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 33 ==> Class=Severe 33 conf:(1)
16	If N_of_p_injured=1 30 ==> Class=Severe 30 conf:(1)
17	If N_of_p_injured=1 Weather_cond=Clear 30 ==> Class=Severe 30 conf:(1)
18	If N_of_p_injured=1 Road_surface=Dry 30 ==> Class=Severe 30 conf:(1)
19	If N_of_p_injured=1 Weather_cond=Clear Road_surface=Dry 30 ==> Class=Severe 30 conf:(1)
20	If N_of_p_injured=1 Driv_gen=M 29 ==> Class=Severe 29 conf:(1)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Table 9: Summarized rules for Severe Class Using Apriori Algorithm

Rule#	Summarized Rule
15	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 33 ==> Class=Severe 33 conf:(1)
19	If N_of_p_injured=1 Weather_cond=Clear Road_surface=Dry 30 ==> Class=Severe 30 conf:(1)
20	If N_of_p_injured=1 Driv_gen=M 29 ==> Class=Severe 29 conf:(1)

As stated above, severe accidents mostly caused by male drivers at clear weather condition and on dry roads. It wasn't checked whether the driver was drunk or not. One injured person was recorded during the accident. When applying rule covers method on the best rules for severe class generated using Apriori algorithm, we can observe that amongst the twenty best rules, three rules only appear after eliminating shorter rules covered in longer ones.

B. 2 Rules Generated Using Predictive Apriori Algorithm

Table 10 presents the best rules generated for severe class using Predictive Apriori algorithm. Table 11shows the final set of summarized rules after combining the rules.

Table 10: Best rules for Severe class Using Predictive Apriori

Rule #	Best Rule
1	If Weather_cond=Clear 38 ==> Class=Severe 38 acc:(0.9766)
2	If Road_surface=Dry 38 ==> Class=Severe 38 acc:(0.9766)
3	If Driv_gen=M 37 ==> Class=Severe 37 acc:(0.97598)
4	If Driv_drink=Not_checked 34 ==> Class=Severe 34 acc:(0.97394)
5	If N_of_p_injured=1 30 ==> Class=Severe 30 acc:(0.97059)
6	If Driv_vec_type=Private 25 ==> Class=Severe 25 acc:(0.96497)
7	If Road_light_cond=Daylight 22 ==> Class=Severe 22 acc:(0.96043)
8	If Accid_year=2009 17 ==> Class=Severe 17 acc:(0.94963)
9	If Accid_cause=Lack_of_respect_for_rd_users 17 ==> Class=Severe 17 acc:(0.94963)
10	If Accid_type=Run_over 17 ==> Class=Severe 17 acc:(0.94963)
11	If Driv_belt=Fastened 17 ==> Class=Severe 17 acc:(0.94963)
12	If Accid_year=2010 16 ==> Class=Severe 16 acc:(0.94674)
13	If Driv_belt=Unknown 14 ==> Class=Severe 14 acc:(0.93985)
14	If Road_light_cond=Dark_lighted 14 ==> Class=Severe 14 acc:(0.93985)
15	If Driv_nation=UAE 11 ==> Class=Severe 11 acc:(0.92551)
16	If Driv_nation=Pakistan 11 ==> Class=Severe 11 acc:(0.92551)
17	If Accid_cause=Others 8 ==> Class=Severe 8 acc:(0.90249)
18	If Accid_day=Th 7 ==> Class=Severe 7 acc:(0.89139)
19	If Accid_day=Fr 7 ==> Class=Severe 7 acc:(0.89139)
20	If Accid_day=Sa 6 ==> Class=Severe 6 acc:(0.8775)

Table 11: Summarized rules for Severe Class Using Predictive Apriori

Rule#	Summarized Rule
1-6, 10	If Weather_cond=Clear 38 OR Road_surface=Dry 38 or Driv_gen=M 37 OR Driv_drink=Not_checked 34 OR N_of_p_injured=1 30 OR Driv_vec_type=Private 25 OR Accid_type=Run_over 17==> Class=Severe acc:(0.96975)
7,14	If Road_light_cond=Daylight 22 OR Road_light_cond=Dark_lighted 14==> Class=Severe acc:(0.95014)
8,12	If Accid_year=2009 17 OR Accid_year=2010 16==> Class=Severe acc:(0.94818)
11,13	If Driv_belt=Fastened 17 OR Driv_belt=Unknown 14==> Class=Severe acc:(0.94474)
15,16	If Driv_nation=UAE 11 OR Driv_nation=Pakistan 11 ==> Class=Severe acc:(0.92551)
9,17	If Accid_cause=Lack_of_respect_for_rd_users 17 or Accid_cause=Others 8==> Class=Severe acc:(0.92551)
18,19, 20	If Accid_day=Th 7 or Accid_day=Fr 7 or Accid_day=Sa 6==> Class=Severe acc:(0.88676)

As stated above, severe accidents occurred mostly during the years of study (2009-2010) on Thursdays or Fridays or Saturdays where seat belt is fastened or unknown, and the accident cause was either lack of respect for road users or others. In addition, severe accidents mostly occur by Emarati or Pakistani male drivers driving private vehicles at clear weather condition and daylight or dark lighted dry roads. One injured person was recorded during the accident. When applying rule covers method on the best rules for death class generated using Predictive Apriori algorithm, we can observe that no rules are eliminated. We can only aggregate them together using this equation; accuracy of summarized rule = Sum (accuracy of each combined rule)/ total number of combined rules.

C. Analysis of Rules for Moderate Class

The best rules generated for moderate class using Apriori algorithm and Predictive Apriori algorithms and the set of summarized final rules are shown in the following sections.

C.1 Rules Generated Using Apriori Algorithm

Tables 12 presents the best rules generated for moderate class using Apriori algorithm. Table 13 shows the final set of summarized rules after applying rule covers method.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Table 12: Best Rules for Moderate Class Using Apriori Algorithm

Rule#	Best Rule
1	If Weather_cond=Clear 117 ==> Class=Moderate 117 conf:(1)
2	If Road_surface=Dry 114 ==> Class=Moderate 114 conf:(1)
3	If Weather_cond=Clear Road_surface=Dry 114 ==> Class=Moderate 114 conf:(1)
4	If Driv_drink=Not_checked 112 ==> Class=Moderate 112 conf:(1)
5	If Driv_drink=Not_checked Weather_cond=Clear 112 ==> Class=Moderate 112 conf:(1)
6	If Driv_drink=Not_checked Road_surface=Dry 109 ==> Class=Moderate 109 conf:(1)
7	If Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 109 ==> Class=Moderate 109 conf:(1)
8	If Driv_gen=M 106 ==> Class=Moderate 106 conf:(1)
9	If Driv_gen=M Weather_cond=Clear 106 ==> Class=Moderate 106 conf:(1)
10	If Driv_gen=M Road_surface=Dry 103 ==> Class=Moderate 103 conf:(1)
11	If Driv_gen=M Weather_cond=Clear Road_surface=Dry 103 ==> Class=Moderate 103 conf:(1)
12	If Driv_gen=M Driv_drink=Not_checked 102 ==> Class=Moderate 102 conf:(1)
13	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear 102 ==> Class=Moderate 102 conf:(1)
14	If Driv_gen=M Driv_drink=Not_checked Road_surface=Dry 99 ==> Class=Moderate 99 conf:(1)
15	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 99 ==> Class=Moderate 99 conf:(1) yyyy
16	If N_of_p_injured=1 80 ==> Class=Moderate 80 conf:(1)
17	If N_of_p_injured=1 Weather_cond=Clear 80 ==> Class=Moderate 80 conf:(1)
18	If N_of_p_injured=1 Driv_drink=Not_checked 77 ==> Class=Moderate 77 conf:(1)
19	If N_of_p_injured=1 Road_surface=Dry 77 ==> Class=Moderate 77 conf:(1)
20	If N_of_p_injured=1 Driv_drink=Not_checked Weather_cond=Clear 77 ==> Class=Moderate 77 conf:(1)

Table 13: Summarized Rules for Moderate Class Using Apriori Algorithm

Rule#	Summarized Rule
15	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 99 ==> Class=Moderate 99 conf:(1)
19	If N_of_p_injured=1 Road_surface=Dry 77 ==> Class=Moderate 77 conf:(1)
20	If N_of_p_injured=1 Driv_drink=Not_checked Weather_cond=Clear 77 ==> Class=Moderate 77 conf:(1)

As stated above, moderate accidents mostly caused by male drivers at clear weather condition and on dry roads. It was not checked whether the driver was drunk or not. One injured person was recorded at the time of accident. When applying rule covers method on the best rules for death class generated using Apriori algorithm, we can observe that amongst the twenty best rules, three rules only appear after eliminating shorter rules covered in longer ones.

C.2 Rules Generated Using Predictive Apriori Algorithm

Tables 14 presents the best rules generated for moderate class using Predictive Apriori algorithm. Table 15 shows the final set of summarized rules after combing the rules.

Table 14: Best Rules for Moderate Class Using Predictive Apriori Algorithm

Rule#	Best Rule
1	If Weather_cond=Clear 117 ==> Class=Moderate 117 acc:(0.99332)
2	If Road_surface=Dry 114 ==> Class=Moderate 114 acc:(0.99323)
3	If Driv_drink=Not_checked 112 ==> Class=Moderate 112 acc:(0.99317)
4	If Driv_gen=M 106 ==> Class=Moderate 106 acc:(0.99296)
5	If N_of_p_injured=1 80 ==> Class=Moderate 80 acc:(0.99155)
6	If Driv_vec_type=Private 76 ==> Class=Moderate 76 acc:(0.99122)
7	If Road_light_cond=Daylight 74 ==> Class=Moderate 74 acc:(0.99104)
8	If Driv_belt=Unknown 60 ==> Class=Moderate 60 acc:(0.98933)
9	If Driv_belt=Fastened 48 ==> Class=Moderate 48 acc:(0.98686)
10	If Accid_year=2009 47 ==> Class=Moderate 47 acc:(0.98658)
11	If Road_light_cond=Dark_lighted 42 ==> Class=Moderate 42 acc:(0.98498)
12	If Accid_year=2010 41 ==> Class=Moderate 41 acc:(0.9846)
13	If Accid_type=Run_over 35 ==> Class=Moderate 35 acc:(0.98184)
14	If Accid_type=Side_collision 32 ==> Class=Moderate 32 acc:(0.98001)
15	If Driv_nation=Pakistan 30 ==> Class=Moderate 30 acc:(0.97857)
16	If Accid_year=2008 29 ==> Class=Moderate 29 acc:(0.97777)
17	If Accid_cause=Lack_of_respect_for_rd_users 29 ==> Class=Moderate 29 acc:(0.97777)
18	If Accid_month=Sep 23 ==> Class=Moderate 23 acc:(0.97139)
19	If Accid_day=Tu 22 ==> Class=Moderate 22 acc:(0.96998)
20	If Accid_day=Th 21 ==> Class=Moderate 21 acc:(0.96842)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Table 15: Summarized Rules for Moderate Class Using Predictive Apriori

Rule#	Summarized Rule
1-6, 15,17,18	If Weather_cond=Clear 117 OR Road_surface=Dry 114 OR Driv_drink=Not_checked 112 OR Driv_gen=M 106 OR N_of_p_injured=1 80 OR Driv_vec_type=Private 76 OR Driv_nation=Pakistan 30 OR Accid_cause=Lack_of_respect_for_rd_users 29 OR Accid_month=Sep 23==> Class=Moderate acc:(0.98702)
8,9	If Driv_belt=Unknown 60 OR Driv_belt= Fastened 48==> Class=Moderate acc:(0.98809)
7,11	If Road_light_cond=Daylight 74 OR Road_light_cond=Dark_lighted 42==> Class=Moderate acc:(0.98801)
10,12,16	If Accid_year=2009 47 OR Accid_year=2010 41 OR Accid_year=2008 29 ==> Class=Moderate acc:(0.98298)
13,14	If Accid_type=Run_over 35 OR Accid_type=Side_collision 32==> Class=Moderate acc:(0.98092)
19,20	If Accid_day=Tu 22 OR Accid_day=th 21==> Class=Moderate acc:(0.9692)

As stated above, moderate accidents occurred mostly during the years (2008, 2010) on Tuesdays or Thursdays where seat belt is fastened or unknown, and the accident cause was lack of respect for road users. In addition, moderate accidents mostly occured by Pakistani male drivers driving private vehicles at clear weather condition and daylight or dark lighted dry roads. The accident type was run over or slide collision. It wasn't checked whether the driver was drunk or not. One injured person was recorded during the accident. When applying rule covers method on the best rules for death class generated using Predictive Apriori algorithm, we can observe that no rules are eliminated. We can only aggregate them together by using the equation; accuracy of summarized rule = Sum (accuracy of each combined rule)/ total number of combined rules.

D. Comparison of Association Rules

Following are some remarks observed when comparing the association rules extracted for each class by Apriori and Predictive Apriori algorithms:

- When applying rule covers method on the best rules for the three classes generated using Apriori algorithm, we can observe that amongst the twenty best rules, few rules appear after eliminating shorter rules covered in longer ones.
- When applying rule covers method on the best rules for the three classes generated using Predictive Apriori algorithm, we can observe that no rules are eliminated. We can only aggregate them together and calculate the new accuracy for each summarized rule.

Table 16 and fig. 1 show the number of summarized rules by each algorithm when applying rule covers method.

Table 16: Number of Summarized Rules

Algorithm	Number of Summarized Rules for Each Class		
	Death	Severe	Moderate
Apriori	4	3	3
Predictive Apriori	5	7	6

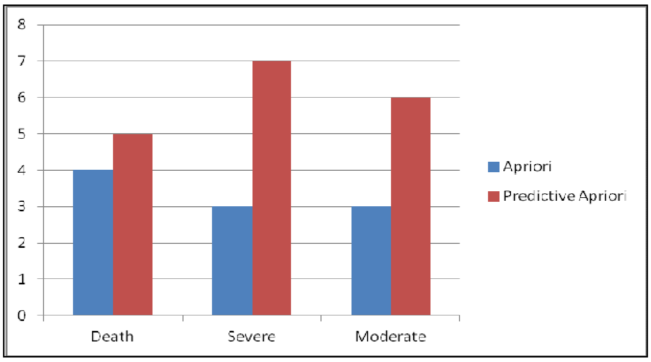

Figure 1: Number of Summarized Rules

VII. Discussions and Recommendations

Based on this research, we can observe that when applying rule covers method on the best rules for the three classes generated (death, severe, and moderate) using Apriori algorithm, few summarized rules were obtained after eliminating shorter rules covered in longer ones. On the other hand, when we applied rule covers method on the best rules for the three classes using Predictive Apriori algorithm, no rules were eliminated. We could only aggregate them together. In addition, the best twenty rules generated by Apriori algorithm contained combinations of various accidents' factors unlike the best twenty rules generated by Predictive Apriori algorithm where each rule contained a single factor at a time and lacked the associations among the different accident factors. Therefore, empirical results showed that class association rules generated by Apriori algorithm were more effective than those generated by Predictive Apriori algorithm if associations between the different factors are of high significance. More number of rules could be eliminated and more associations between accident factors and accident severity were explored when applying Apriori algorithm.

VIII. Conclusion and Future Work

Our empirical experiments showed that when applying rule covers method on the generated class association rules using Apriori and Predictive Apriori algorithms, many class association rules generated by Apriori algorithm were eliminated and more effective rules than those generated by Predictive Apriori algorithm were obtained. In addition, more associations between accident factors and accident severity were explored when applying Apriori algorithm. On the other hand, Predictive Apriori algorithm could derive more number of rules that could be useful when studying the effect of each individual factor to accident severity. The adaptation of the association rule mining algorithm to mine only a particular subset of association rules where the classification class attribute is assigned to the right-hand-side could successfully generate more effective rules

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

covering all three classes. It solved the problems that could occur when traditional classification algorithms applied including no interesting rules could be discovered if the population of a particular class was so large. Apriori algorithm outperformed Predictive Apriori algorithm when applying rule covers method for summarizing the rules by giving less number of significant interesting rules. These results may help the decision makers in the traffic accident department to take actions based on some hidden patterns from the data. As future work, other data mining techniques and algorithms can be applied to the dataset such as clustering and temporal data mining techniques in order to study in different ways which accident factors can affect the occurrence of accidents. In addition, domain-driven data mining approach can also be applied to generate more actionable rules.

ACKNOWLEDGMENT

We would like to thank the Traffic Accident Department staff for their cooperation and providing us with traffic accidents data covering the years 2008 to 2010.

REFERENCES

[1] Adeyemi Adejuwon, Amir Mosavi, “Domain Driven Data Mining-Application to Business”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No.2, July 2010, pp. 41-44.
[2] Qiankun Zhao, Sourav S. Bhowmick, “Association Rule Mining: A Survey”, Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116, 2003.
[3] Bing Liu, Wynne Hsu, Yiming Ma, “Integrating Classification and Association Rule Mining”, KDD-98, New York, Aug 27-31, 1998.
[4] Paresh Tanna, Dr. Yogesh Ghodasara, “Using Apriori with WEKA for Frequent Pattern Mining”, International Journal of Engineering Trends and Technology (IJETT), Volume 12 Number 3, Jun 2014, pp. 127-131.
[5] Divya Bansal, Lekha Bhambhu, “Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September 2013, pp. 54-62.
[6] Zhou Aoying, Wei Li, YU Fang, “Effective Discovery of Exception Class Association Rules”, Journal of Computer Science and Technology, Volume 17, Issue 3, May 2002, pp. 304-313.
[7] Saddys Segrera, Maria N. Moreno, “Classification Based on Association Rules for Adaptive Web Systems”, Advances in Soft Computing, Volume 44, 2007, pp. 446-453.
[8] David Halbert (October 20, 2008), “The World’s Worst Drivers – Car Accident Statistics From Around the World”, [Online]. Available at URL http://www.articlesbase.com/cars-articles/the-worlds-worst-drivers-car-accident-statistics-from-around-the-world-609862.html.
[9] Ronald V Miller (September 8, 2010), “Car Accident Statistics from the CDC”, [online]. Available at URL: http://www.marylandinjurylawyerblog.com/2010/09/car_accident_statistics_from_t.html.
[10] A. Bener, D. Crundall, “Road traffic accidents in the United Arab Emirates compared to Western countries”, Advances in Transportation Studies an international Journal Section A6, 2005, pp. 5-12.
[11] Abdelaziz Araar, Amira A. El Tayeb,“Mining Road Traffic Accident Data to Improve Safety in Dubai”, Journal of Theoretical and Applied Information Technology, Vol. 47 No.3, 31st January 2013, pp. 911-925.
[12] Ossenbruggen, P. J., J. Pendharkar, et al. "Roadway safety in rural and small urbanized areas." Accidents Analysis and Prevention 33(4): 485-498, 2001.
[13] Sohn, S. and S. Hyungwon. "Pattern recognition for a road traffic accident severity in Korea", Ergonomics 44(1): 101-117, 2001.
[14] Sohn, S. and S. Lee, "Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea." Safety Science 41(1): 1-14, 2002.
[15] Ng, K. S., W. T. Hung, et al., "An algorithm for assessing the risk of traffic accidents." Journal of Safety Research 33: 387-410, 2002.
[16] Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., “The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities, Accident analysis and Prevention, Vol. 34, pp. 717-727, 2002.
[17] Miao M. Chong, Ajith Abraham, Marcin Paprzycki, “Traffic Accident Analysis Using Decision Trees and Neural Networks”, IADIS International Conference on Applied Computing, Portugal, IADIS Press, Pedro Isaias et al. (Eds.), Volume 2, pp. 39-42 2004.
[18] Chang, L. and W. Chen "Data mining of treebased models to analyze freeway accident frequency", Journal of Safety Research 36: 365- 375, 2005.
[19] T. Beshah, “Application of data mining technology to support RTA severity analysis at Addis Ababa traffic office”, Addis Ababa, Addis Ababa University, 2005.
[20] Chang, L. and H. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques Accident analysis and prevention", Accident analysis and prevention 38(5): 1019-1027, 2006.
[21] Srisuriyachai, S., “Analysis of road traffic accidents in Nakhon Pathom province of Bangkok using data mining”, Graduate Studies, Bangkok, Mahidol University, 2007.
[22] Wong, J. and Y. Chung,"Comparison of Methodology Approach to Identify Causal Factors of Accident Severity." Transportation Research Record 2083: 190-198, 2008.
[23] Zelalem, R., “Determining the degree of driver’s responsibility for car accident: the case of Addis Ababa traffic office”, Addis Ababa, Addis Ababa University, 2009.
[24] Getnet, M., “ Applying data mining with decision tree and rule induction techniques to identify determinant factors of drivers and vehicles in support of reducing and controlling road traffic accidents: the case of Addis Ababa city, ”Addis Ababa, Addis Ababa University, 2009.
[25] Sami Ayramo, Pasi Pirtala, Janne Kauttonen, Kashif Naveed, Tommi Karkkainen, “Mining road traffic accidents”, University of Jyvaskyla, Finland, 2009.
[26] T. Beshah and S. Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia", Proceedings of AAAI Artificial Intelligence for Development (AI-D'10), 2010.
[27] Galvão ND, de Fátima Marin H, “Traffic accident in Cuiabá-MT: an analysis through the data mining technology”, Federal University of Mato Grosso-UFMT, Brazil, 2010.
[28] Amirhossein Ehsaei, Harry Evdorides,“Temporal Variation of Road Accident Data caused by Road Infrastructure”, 3rd International Conference of Road Safety and Simulation, September 14-16, Indianapolis, USA, 2011.
[29] S.Krishnaveni, Dr.M.Hemalatha, “A Perspective Analysis of Traffic Accident using Data Mining Techniques”, International Journal of Computer Applications,Volum 23- No. 7, pp. 40-48, June 2011.
[30] S.Krishnaveni,Dr.M.Hemalatha,“Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms”, International Journal of Computer Applications, Volume 35– No.12, December 2011, pp. 30-37.
[31] Beshah, T.; Ejigu, D.; Abraham, A.; Snasel, V.; Kromer, P., “Pattern recognition and knowledge discovery from road traffic accident data in Ethiopia: Implications for improving road safety”, World Congress on Information and Communication Technologies (WICT), December 2011, pp. 1241 - 1246.
[32] Vandana Munde, Sachin Deshpande, S.K.Shinde,“Data Mining for Traffic Accident Analysis”, International Conference on Advances in Computing and Management, 2012.
[33] Olutayo V.A, Eludire A.A, “Traffic Accident Analysis Using Decision Trees and Neural Networks”, I.J. Information Technology and Computer Science, 2014, 02, 22-28.
[34] Rajdeep Kaur Aulakh, “Association Rules Mining Using Effective Algorithm: A Review”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 3, March 2015, pp. 831-835.
[35] Amit Mittal, Ashutosh Nagar, Kartik Gupta, Rishi Nahar, “Comparative Study of Various Frequent Pattern Mining Algorithms”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 4, April 2015, pp. 550-553.
[36] Dr. S. Vijayarani, Ms. R. Prasannalakshmi, “Comparative Analysis of Association Rule Generation Algorithms in Data Streams”, International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 1, February 2015, pp. 15-25.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

[37] Sunita B Aher, Mr. LOBO L.M.R.J., "Data Mining in Educational System using WEKA", International Conference on Emerging Technology Trends (ICETT), 2011, pp. 20-25.

[38] Sunita B Aher and Lobo L.M.R.J, "A Comparative Study of Association Rule Algorithms for Course Recommender System in E-Learning", International Journal of Computer Applications, Volume 39 – No. 1, February 2012, pp. 48-52.

[39] "Stratified Random Sampling", [online]. Available at URL: http://www.stat.ualberta.ca/~prasad/361/STRATIFIED%20RANDOM%20SAMPLING.pdf.

[40] Carlos Ordonez, Norberto Ezquerra, Cesar A. Santana, "Constraining and Summarizing Association Rules in Medical Data", Knowledge and Information Systems, Volume 9, Issue 3, 2006, pp.. 259 - 283.

Eng. Amira A. El Tayeb been working as a senior lecturer for eleven years at Preston University-Ajman, University of Bedfordshire,-UK, Preston campus, and Jumeira University, Dubai. She obtained her B.Sc in Computer Engineering in 2002 and completed her M.Sc in Computer Engineering course requirements in 2004 from the Arab Academy for Science and Technology, Egypt. She also got M.Sc in Information Systems from Ajman University of Science & Technology, UAE, 2012. She is currently a PhD. scholar in Information Systems at Banasthali University, India. Her research interests include: Databases, Data Mining, and Software Engineering.

Dr. Vikas Pareek is currently an Associate Professor in the Department of Computer Science at Banasthali University, India. He is actively engaged in research on various aspects of Information security ranging from cryptographic algorithms to random number generation and cyber crimes. He has many publications to his credit in information security, e-learning, mobile computing and related areas.

Dr. Abdelaziz Araar is an associate professor at the college of Information Technology at Ajman University of Science and Technology, UAE. He received his B.Sc. in computer science from University of Annaba in 1983, his M.Sc. in 1986, and then his PhD in 1991 from Case Western Reserve University, Cleveland, Ohio, USA. His area of interest is mainly on advanced simulation models for networking, applied artificial intelligence, data mining and NFC technology.