Analyzing Road Accident Data using Association Rule Mining

K. Shanmuga Srivalli
MCA III Year
Department of C.S.E
SVU CM&CS, Tirupati

Dr. E. Kesavulu Reddy
Asst. Professor
Department of C.S.E.
Svu.Cm&Cs ,Tirupati

Abstract: **Road accident is one of the crucial areas of research in India. A variety of research has been done on information collected through police records covering a limited portion of highways. The strong rules hidden in these frequent item sets often uncover the association between influencing factors of accidents, which can be utilized to reduce the event of accidents by breaking them. The standards can likewise be utilized to investigation usual scenes of accidents, and some corresponding security improvement measures can be taken to prevent the accidents, and ultimately improve the city's traffic safety level. The study first designed a method to calculate lowest Support price of training parameters, and further put forward a way to extract strong rules automatically. Therefore, an automatic modeling algorithm exploitation association rules was finally established to promote the effective application of association rule mining on intelligent transportation system.**

Keywords: Traffic accident, Association rules, Automatic modeling algorithm

I. INTRODUCTION

Data mining is searching for covered up, substantial, and feasibly valuable designs in tremendous data sets. Data mining is all with respect to finding unsuspected/ beforehand incomprehensible connections among the data. It is a multi-disciplinary expertise that utilizations AI, measurements, and AI and database innovation. The bits of knowledge inferred by means of information mining can

be utilized for promoting, misrepresentation identification, and logical revelation, and so forth. Data mining is moreover called as information disclosure, data extraction, information/design investigation, data gathering, etc. Road and car crashes are questionable also, eccentric episodes and their investigation requires the information of the variables influencing them. Road and traffic accidents are defined by a set of variables which are mostly of discrete nature. The most important downside within the analysis of accident information is its heterogeneous nature. So no uniformity should be thought of throughout analysis of the information otherwise, some relationship between the information could stay hidden. Although, researchers used segmentation of the information to cut back this heterogeneousness using some measures such as expert information, but there is no guarantee that this will lead to an optimal segmentation which consists of homogeneous groups of road accidents. Therefore, cluster analysis will assist the segmentation of road accidents. The existing association rule mining generally determined the model parameters (such because the minimum Support, etc.) by repeated experiments. For the massive results excavated, it is necessary for the experts to screen useful rules according to personal experience manually. The method is inefficient and the subjective screening method cannot be translated into an objective algorithm, so it hinders the direct use of association rule mining in intelligent transportation system. We proposed a method to calculate the minimum Support in the modeling parameters, and put forward a way to extract strong rules from the massive rules by a clustering method, or automatically filtering out the weak rules based on an expert expertise related method.
Finally, we built up an automated modeling algorithm using association rules which would better promote the practical application of association rule mining in existing intelligent transportation system.

II. RELATIVE STUDY:

A. Investigation of work zone crash setback designs utilizing association rules. Accident; investigation and avoidance

Investigation of the bad luck crash attributes and contributory elements is one of the high-need issues in rush hour congestion wellbeing investigation. We propose a strategy dependent on association rules to break down the attributes and contributory components of work zone crash bad luck. A contextual investigation is led utilizing the Michigan M-94/I-94/I-94BL/I-94BR work zone crash information from 2004 to 2008. The got affiliation rules are partitioned into two segments just as rules with high-lift, and rules with high-support for the any examination. The outcomes show that practically all the high-lift rules contain either ecological or occupier attributes. Most of affiliation rules are focused on explicit attributes, for example, drinking, driving, the superhighway with multiple paths, and speed-limit over 40mph and not utilization of traffic control devices. It must to be ***recognized that some more grounded related guidelines were found in the high-bolster part. With the system perception, the association***

rule strategy can give progressively reasonable outcomes to investigative the examples of work zone crash setbacks.

B. A unique Bayesian system model for ongoing accident forecast utilizing traffic speed conditions information

Car accidents happening on interstates/roads are considered to relate near past traffic conditions, which are time-shifting. Then, most examinations use volume/inhabitance/speed parameters to forestall the probability of accidents, which are invalid for streets where the traffic conditions are measured utilizing speed information separated from inspected coasting vehicles or advanced mobile phones. In this manner, a unique Bayesian system (DBN) model of time succession traffic information has been proposed to research the connection between crash event and dynamic speed condition information. Besides, the traffic conditions close to the accident site were distinguished as a few state combinations as indicated by the degree of block and remembered for the DBN model. In light of 551 crashes and comparing speed data gathered on freeways in Shanghai, China, DBN models were worked with time arrangement speed condition information and diverse state mixes. A near investigation of the DBN model utilizing stream indicator information and a static Bayesian system model was likewise directed. The outcomes show that, with just speed condition information and nine traffic state mergers, the DBN model can accomplish an accident forecast exactness of 76.4% with a bogus caution pace of 23.7%. What's more, the aftereffects of transferability testing suggest that the DBN models are relevant to other comparative roads with 67.0% accident expectation exactness.

C. Examples of Single-Vehicle Crashes on two-path Rural Highways in Granada Province, Spain: In Depth Analysis Through choice Rules

In Spain, 74 of damage crashes happen on provincial two-path superhighways. In this way, one of the dynamic needs point by point in interstate wellbeing plans is the particular investigation of these parkways. This investigation planned for exploring crash designs and contributory factors on provincial two-path expressways in order to propose explicit street security countermeasures. The investigation technique comprised in recognizing choice principles removed from choice trees (DTs). As the normal strategy for rule extraction is constrained by a DT's structure, some significant connections between factors may not be recognized. For this issue to be continued, an inside and out technique for extricating rules from confusion tremens was utilized. Since the usage of any corrective street security measure is compelled by the accessible assets, the most grounded examples that portray the street wellbeing issue must be removed. For recognizable proof of the most grounded standards, another basis, lift increment foundation, was characterized. Single-vehicle crashes on two-path rural parkways in the region of urban focus in Spain were broke down. Crash information were comparative with the 7-year time frame 2003 to 2009. Rules were gotten by utilizing both increase data and the data gain proportion as parting basis. The standards got by use of the two criteria were reliable and correlative; hence, the inventors prescribe the utilization of the two techniques to manufacture insanity tremens. Consequences of the examination featured a few examples adding to serious accidents and possibly powerful countermeasures. Principle designs were person on foot crashes, run-off-the-street crashes, including controlled bikes, crashes including fueled bikes, and crashes around evening time without instruction.

III. EXISTING SYSTEM

The existing association rule mining generally determined the model parameters (such as the minimum Support, etc.) by repeated experiments. For the massive results excavated, it is necessary for the experts to screen useful rules according to personal expertise manually. We built up an automated modeling algorithm using association rules which would better promote the practical utilization of association rule mining in existing smart transportation framework. The remainder of this paper is sorted out as follows. Designing an automatic modeling algorithm using association rules, is done which can promote the application of association rule mining in the existing intelligent transportation system. Finally, based on these methods, an automatic modeling algorithm using association rules is proposed, which helps to promote the application of association rule mining in the existing intelligent transportation systems.

A. Proposed System

There are also studies on the impact of road conditions on traffic accidents, they proposed a point that high and steep roadbed will undermine the traffic safety. We proposed a method to calculate the minimum Support in the modeling parameters, and put forward a way to extract strong rules from the massive rules by a clustering method, or automatically filtering out the weak rules based on an expert experience related method. In addition to the direct use of association rules, some scholars have proposed methods such as multi-dimensional association rules model, association rules in spatial and temporal Data. A method of computing minimum Support based on interested frequent item sets was proposed, and two powerful methods of automatic acquiring strong rules were proposed.

B. Algorithm: Automated Modeling Algorithm Using Association Rules

Normally, the parameters of association rule mining are determined by repeated experiments, and confronted with the massive rules of excavation, it mainly relies on manually screening, which greatly limits the direct application of association rule mining in an intelligent transportation system. Based on the above methods, we designed an automated modeling algorithm using association rules and applied it to intelligent traffic systems. As long as the user selects the sample data, gives some necessary information interactively, and then clicks

the Mining button, the system can dig out the effective strong rules and visualize them automatically.

Step1: User specified sample data (user interaction)

Step2: Show the qualities of the example information the user selects the interested frequent items to obtain interested frequent item set (user interaction)

Step3: The system calculates the minimum Support based on interested frequency item set which is described in part A of section IV, and specifies the constant minimum Confidence, and display them (system execution)

Step4: The user confirms the modeling parameters, including minimum Support and minimum Confidence, and the user can manually modify these parameters (user interaction)

Step5: The user chooses the strong rules selection method, including method described in part C of section IV or method described in part D of section IV, and appoints the number of rules (sorted by Lift in descending order) which will be returned by system (user interaction)

Step6: Call the association rule mining (we used Apriority function of the A rules R package), to automatically mine the sample data (system execution)

Step7: Filter the mining result using strong rules automatic screening method, return the specified number of rules and visualize them (system execution)

IV. CONCLUSION

We utilized the affiliation rules to examine the connection between the affecting elements of auto collisions gathered from Shanghai urban freeway, and proposed a insignificant Support count technique dependent on intrigue visit thing set. Simultaneously, so as to successfully translate the mining results, two in number standards screening strategies are proposed, one is the solid principles programmed screening technique in light of bunching and the other is the frail principles programmed separating technique dependent on master understanding. These two strategies can likewise be utilized in blend. Through the first strategy, we unearthed the circle finder information and mishap information of urban road in Shanghai, and indicated some solid governs in Table II. During that time strategy, we unearthed the conventional car crash information (counting traffic attributes, individuals, vehicles, streets, condition, and so forth.) and shown some solid principles in Table IV. The consequences of the tests show that the two strategies are powerful. At last, in light of these techniques, a programmed demonstrating calculation utilizing affiliation rules is proposed, which advances the utilization of affiliation rule mining in the current smart transportation frameworks. Through the solid guidelines which contain the affiliation among affecting elements of auto collisions, we can break such relationship to bring down the event of mishaps. With respect to it implies that it is progressively inclined to mishaps on the street with no speed limit signs in blustery night, and we can break the relationship by including quite far signs or improving the road lighting conditions to decrease the accidents.

REFERENCES

[1] Worrall R D, Bullen A G R. AN EMPIRICAL ANALYSIS OF LANE CHANGING ON MULTILANE HIGHWAYS[J]. Highway Research Record, 1970(303).
[2] Xu H, Cheng G, Pei Y. Study on effect of lane-changing behavioral characteristic to capacity[J]. Sciencepaper Online, 2010.
[3] Yulong P, Ji M, Research on countermeasures for road condition causes of traffic accidents [J]. China Journal of Highway and Transport, 2003, 16(4):77-82.
[4] Das S, Sun X. Investigating the Pattern of Traffic Crashes Under Rainy Weather by Association Rules in Data Mining[J]. Transportation Research Board Annual Meeting, 2014.
[5] He S B, Wang Y J, Sun Y K, et al. The Research of Multidimensional Association Rule in Traffic Accidents[C]// International Conference on Wireless Communications, NETWORKING and Mobile Computing. IEEE, 2008:1-4.
[6] Mennis J, Liu J W. Mining Association Rules in Spatio-Temporal Data[C]// 2003:699–700.
[7] Kumar S, Toshniwal D. A data mining framework to analyze road accident data[J]. Journal of Big Data, 2015, 2(1):26.
[8] López G, Abellán J, Montella A, et al. Patterns of Single-Vehicle Crashes on Two-Lane Rural Highways in Granada Province, Spain: InDepth Analysis Through Decision Rules[J]. Transportation Research Record Journal of the Transportation Research Board, 2014, 2432(1):133-141.
[9] Weng J, Zhu J Z, Yan X, et al. Investigation of work zone crash casualty patterns using association rules[J]. Accident; analysis and prevention, 2016, 92:43.
[10] Sun J, Sun J. A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data[J]. Transportation Research Part C Emerging Technologies, 2015, 54:176-186.
[11] Apiletti D, Baralis E, Cerquitelli T, et al. Characterizing network traffic by means of the N et M ine, framework[J]. Computer Networks, 2009, 53(6):774-789.
[12] Pande A, Abdel-Aty M. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool[J]. Safety Science, 2009, 47(1):145-154.