Association Rule mining for Road Accident Prediction Using IARM Algorithm

S. Nagendra Babu, Research Scholar, R & D Center, Bharathiar University, Coimbatore, India.
E-mail:s.nagendrababu@gmail.com
Dr.J. Jebamalar Tamilselvi, Professor, Jaya Engineering College, Thiruninravur, Chennai, India.
E-mail:jjebmalar@gmail.com

_____

Abstract : Lately, the accident rate identified with activity is high. Investigating the accident information and separating valuable data from it can help in taking individual measures to diminish this rate or keep the accident from happening. Related research has been done in the past which included proposing different measures and algorithms to get intriguing accident designs from the accident records. The primary issue is that expansive quantities of examples were delivered and huge number of these examples would be evident or not fascinating. A more profound investigation of the information is required keeping in mind the end goal to get the intriguing examples. With a specific end goal to conquer this circumstance, we have proposed another way to deal with distinguish the most related successive examples in the accident information. We likewise influence utilization of the procedure.The principle objective of this paper is to identify the most related successive examples (MASP) and mine different data examples inside the informational collections produced by MASP utilizing an adjusted FP-growth approach in standard affiliation mining. This paper proposed a new Improved Association Rule Mining(IARM) algorithm for the prediction of road accidents and proposing suggestion to higher authorities for increasing safety measures on highways.

Keywords: Association rule mining;accident data;information mining;rule based mining.

_____

1. Introduction

Investigating, deciphering and making most extreme utilization of the information is problamatic and quality requesting because of the exponential development of numerous business, administrative and logical databases. It is evaluated that the measure of information put away on the database develops at regular intervals at a rate of 100%. This reality demonstrates that we are getting increasingly detonated by information/data but grasping for learning. Information mining accordingly shows up as a valuable instrument to address the requirement for filtering helpful data, for example, concealed examples from databases. In the data set, where the gathering of information is expanding in a disturbing rate, understanding stimulating examples of information is a vital issue to be considered to modify procedures, to make most extreme utilization of it, and find new openings. Associations keeping information on their space region takes each record as an open door in learning actualities.

This paper intends to give recommendations to enhance road security. It diagrams the real reasons for roadaccidents by applying information mining algorithms to information gathered from past accident records of the https://data.gov.in/dataset-group-name/road-accidents, where a huge number of accidents happen each year. After information preprocessing, the cleaned information is examined and mined utilizing the IARM information mining strategy.

Roadaccidents happen under particular movement conditions, including drivers, road clients and different components. Road movement security remains the lasting theme for all transportation experts over the globe. Insights demonstrate that in 2012, there were 46 thousand roadaccidents in India alone, with 11 thousand passings and 50 thousand wounds. Specialists of India have been keeping point by point records on roadway accidents, and assembled a complete auto collision data framework. Be that as it may, the information had just been utilized to lead simple measurable examinations and information mining endeavors, the finishes of which are for the most part examples and insights. The more significant and critical nature of movement causation investigation, attributes and relationships thinks about between various causational factors have frequently been neglected.

The proposed IARM algorithm defines the association rules for the prediction of road accidents. After the successful data clustering the IARM algorithm effectively designs the association rules which predicts better than Fp-Growth algorithm.

1. Literature Survey

As indicated by the World Health Organization (WHO) [8], 1.24 million individuals bite the dust every year on the world's roads, and upwards of 50 million are harmed. Moreover, the Centers for Disease Control and Prevention (CDCP) have declared that roadaccidents cost 100 billion in medicinal care each year.

As a survey of the writing appears, numerous information mining methods have been proposed to dissect roadaccidents. In this specific circumstance, H. Schneider et al.[1] utilized CART and MARS to dissect an epidemiological case-control investigation of wounds coming about because of engine vehicle accidents. They additionally recognized potential regions of hazard, to a great extent caused by the driver circumstance [10]. B. Brorsson et al. [2] utilized strategic relapse models to investigate the elements engaged with accidents, and found that shopping regions were more risky than town destinations. K.W.Y. Kelvin et al. [3] utilized the three information mining methods of choice trees, neural systems and strategic relapse to find huge components influencing the seriousness of road activity.

S. Reed et.al [4] considered driver duty utilizing the ID3, J48, and multilayer perceptron (MLP) algorithms to find the related factors, and found that numerous components directly affect the seriousness of accidents, for example, permit grades and the driver's age and experience. T. Beshah et al. [5] utilized CART and multinomial calculated relapse (MLR) to investigate the parts played by the attributes of drivers, and found that the CART strategy gave generally exact outcomes.

Yun. Li et.al [6] investigated how road car accidents seriousness and driving condition factors are identified with each other utilizing a few systems and algorithms. The specialists connected neural system, choice tree classifiers, classifier combination in light of the Dempster– Shafer algorithm, the Bayesian technique, strategic model, and grouping utilizing k-mean algorithm. Their test comes about demonstrated that bunching system was superior to different procedures [14].

R. Marukatatet.al [7] connected distinctive procedure approaches and contrasted them with find accident seriousness factors. The creators initially utilized an arrangement of hypotheses to examine car accident information to see if the information had finish data about the conditions related with the event of accidents, and afterward thought about these subordinate conditions. Their outcomes demonstrated that deadly accidents came about because of a blend of elements. Besides, rules with high or low help demonstrated distinctive highlights [22].

J. Mikeet.al. [8] concentrated the connection between loss frequencies and the separation of the accidents from the zones of habitation. As might have been foreseen, the setback frequencies were higher closer to the zones of living arrangement, conceivably because of higher introduction. The investigation uncovered that the setback rates among inhabitants from territories delegated moderately denied were fundamentally higher than those from generally wealthy zones.
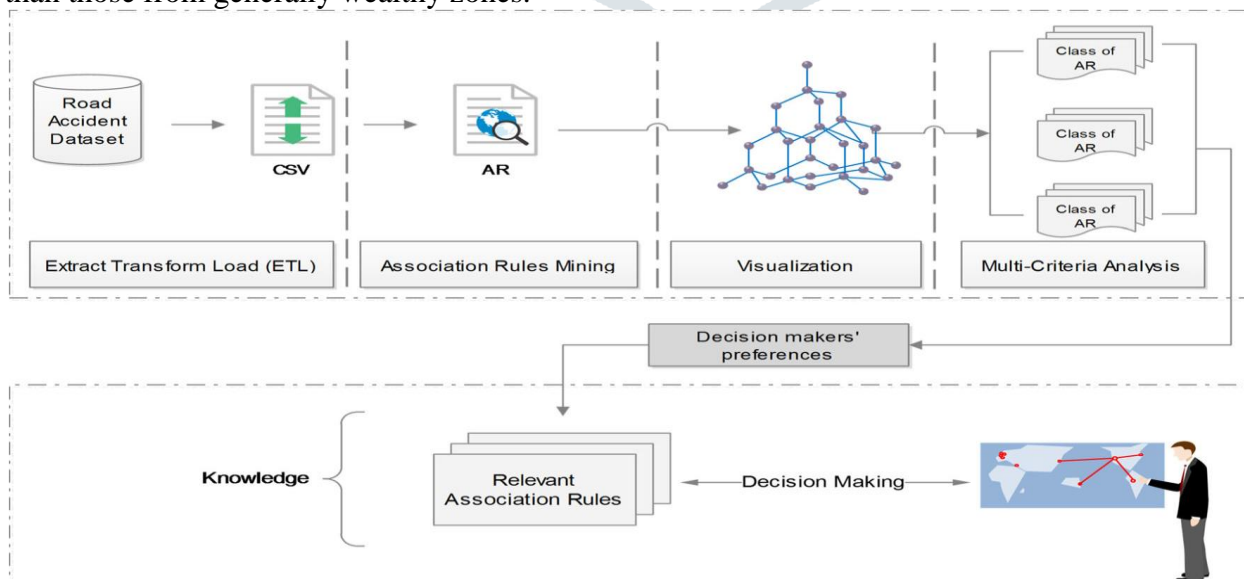

Fig-1 Association rule mining architecture

3. Existing Methods

The AHP-Apriori examination strategy

Diagnostic Hierarchy Process (AHP) is a multi-criteria basic leadership strategy, and is displayed by American specialist Satty T. L., teacher of tasks inquire about. AHP is the subjective judgment of chiefs to evaluate the quantity of types of articulation and handling, through the quantitative information as a mix of subjective and quantitative investigation to enable leaders to decide. Systematic progression process (AHP), be that as it may, just the significance degrees of each factor level are analyzed, and the quantitative connection between different variables is ignored.

Relationship govern is the investigation of connection between's various occasions. It is an information mining strategy. In straightforward terms, relationship lead can deliver a sort of "A=> B" impact, in particular when A happens, B will happen therefore. Affiliation investigation is with the objective of mining the concealed connection between information, searching for a similar occasion in the relationship of various things. So keeping in mind the end goal to compensate for the deficiency of these two techniques, this paper advances the AHP - Apriori strategy to examine the normal for road auto collision information. In the first place to clear the elements with littler impacts on the accident, the primary impact factors and the aftereffect of the accident ought to be held. So as to lessen the measure of accident record, the primary variables are related to decide properties identified with the principle factors and enhance the productivity of relationship. On this premise, through following the current condition to push the reason for the accident, thinking about the elements of intellectual qualities of drivers and driving conduct, joining target information and sane thinking, to a wide range of elements influencing driving wellbeing firmly connected with the car accident.

FP-Growth algorithm

The Frequent Pattern (FP)- Growth strategy is utilized with databases and not with streams. The algorithm extricates visit thing sets that can be utilized to remove affiliation rules. This is finished utilizing the help of a thing set. The wording, that is utilized for this algorithm is portrayed in part IV-A1.

The primary thought of the algorithm is to utilize a separation and vanquish procedure:

Pack the database which gives the regular sets; at that point separate this compacted database into an arrangement of contingent databases, each related with a continuous set and apply information mining on every database.

To pack the information source, an uncommon information structure called the FP-Tree is required [6]. The tree is utilized for the information mining part.

At long last the algorithm works in two stages:
- Development of the FP-Tree
- Concentrate visit thing sets.

4. Proposed Method

Here another Improved ARM algorithm is characterized for associationrulemining for roadaccident information expectation. The proposed algorithm is quite enhanced in association rule mining when contrasted with FP-Growth algorithm.This algorithm characterizes how standards can be set to characterize new techniques for mining regular designs.

```
//Generate vehicle ids, accident types, their accident IDs

Algorithm IARM()
{
 V1 = find_frequent_1_itemsets (P);
For (k = 2; Lk-1 ≠Φ; k++)
{
 Ck = candidates generated from Lk-1;
x = Get _item_min_sup(Ck, V1);
 Tgt = get_accident_ID(x);
 For each accident t in Tgt Do
Increment the count of all vehicles in Ck that are found in Tgt;
Lk=vehicles in Ck ≥ defined_Tgt;
 End;
 }
```

The performance of the proposed association rule mining algorithm is compared with FP-Growth algorithm. The performance levels are compared in seconds.

Type	FP-Growth	IARM
DATA SET<1000 records	3.66 s	3.03 s
DATA SET<50000 records	8.87 s	3.25 s
DATA SET<100000 records	34 m	5.07 s
DATA SET >2 Lack records	4+ hours (Never finished, crashed)	8.82 s

Table-1 Association rule mining performance levels.

5. Results

Our examination depends on the Real-life car accidents records gathered from National Highway authority Department,India from 2010 to 2017. In the wake of applying stratified examining system on the gathered information, 17632 records were chosen in view of the most entire records [11]. Be that as it may, we just centered in this exploration around finding the standards identified with three kinds of classes; Death, Severe, and Moderate. Affiliation rules algorithms are connected in parallel on the chose dataset.

Examination of Rules for Death Class

The best principles created for Death class utilizing Fp-Growthalgorithm and the arrangement of outlined last standards are appeared in the accompanying areas.

Rules Generated Using IARM Algorithm

Table 2 introduces the best standards produced for death class utilizing the proposed IARM algorithm and fig-2 illustrates the performance levels of the proposed and existing methods.

Rule	Best Rule
1	If Driv_gen=M 47 ==> Class=Death 47conf:(1)
2	If Road_surface=Dry 47 ==> Class=Death 47conf:(1)

3	If Weather_cond=Clear 46 ==> Class=Death46 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=ClearRoad_surface=Dry 45 ==> Class=Death 45 conf:(1)
8	If Driv_drink=Not_checked 42 ==>Class=Death 42 conf:(1)
9	If N_of_p_injured=1 Driv_gen=M 31 ==>Class=Death 31 conf:(1)
10	If Driv_gen=M andDriv_drink=Not_checked andWeather_cond=Clear and Road_surface=Dry 40 ==> Class=Death 40 conf:(1)

Table-2 Association rules for road accident prediction.

The performance of the proposed algorithm is compared with the Fp-Growth algorithm and the results show that the proposed IARM algorithm yields better results when compared to the existing methodology.
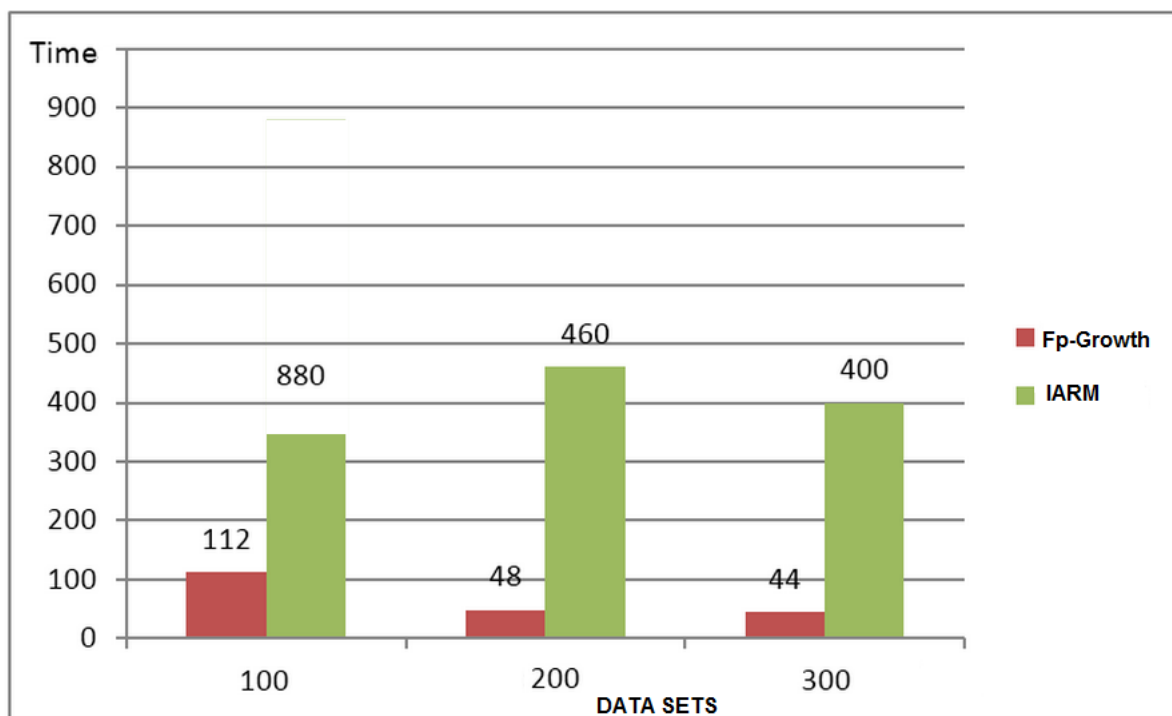

Fig-2 Performance level analysis

6. Conclusion

This paper conducts examine on the causation investigation of accidents from two viewpoints, the conclusions are: in antagonistic driving conditions, where principle factors specifically impact accidents; then again, where fundamental components affect accidents, as indicated by the aftereffects of examinations, recommends the proposed display is material. Setting up the FP-Growthalgorithm voids the imperfections of either model or algorithm, This is utilized to decide the principle impacting factors, sifting through components with less commitments to the accident, which incredibly decreases the measure of count required. Through connection investigation of the primary impact factors, it can be inferred that there is solid relationship between's natural elements and the accident write.

The proposed IARM algorithm is used to build association rules for the prediction of the road accidents and the performance levels of the IARM algorithm is compared with the Fp-Growth algorithm and the results show that the proposed IARM is yielding better results.These outcomes may help the highway authorities to take activities in view of some concealed examples from the information. As future work, other information mining methods and algorithms can be connected to the dataset, for example, grouping and worldly information mining procedures keeping in mind the end goal to think about in various ways which accident components can influence the event of accidents. Likewise, area driven information mining methodology can likewise be connected to produce more significant guidelines.

References

[1]. H. Schneider, Louisiana Traffic Records …0000Data Report 2013, Louisiana State University,Baton Rouge, Louisiana, 2014.

[2]. B. Brorsson, H. Rydgren, J. Ifver, Single-vehicle accidents in Sweden: a comparativestudy of risk and risk factors by age, J. Saf. Res. 24 (1993).

[3]. K.W.Y. Kelvin, Risk factors affecting the severity of single vehicle traffic accidents inHong Kong, Accid. Anal. Prev. 36 (2004) 333–340.

[4]. S. Reed, A. Morris, Characteristics of fatal single-vehicle accidentes in Europe, Int. J.Accidentworthiness 17 (6) (2012) 665-664.

[5]. T. Beshah, Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in Ethiopia, AAAI SpringSymposium-Technical Report, 10 (01)(2010) 14-19.

[6]. Yun. Li , Studies and Applications of Association Rule Mining Methods in Data Mining, Xi An University, 2007.

[7]. R. Marukatat, Structure-Based Rule Selection Framework for Association Rule Mining of Traffic Accident Data, International Conferenceon Computational Intelligence and Security, 1(2007)781-784.

[8]. J. Mike, National transit Bus accident data collection and analysis, Florida: University of South Florida, 2002.

[9]. Y. S. Murat, Modelling traffic accident data by cluster analysis approach, Technical Journal of Turkish Chamber of Civil Engineers, 20(3)(2009)4759-4777.

[10]. C.Y. Qiu ,Research on Multi-objective Particle Swarm Optimization and Its Application in Accident Severity Analysis, Beijing University ofPosts and Telecommunications, 2013.

[11]. K. B. Sudhir, Data mining applications in Transportion Engineering, Transport, 5(2003)216-223.

[12]. J.B. Zhao, W. Deng, J. Wang, Bayesian network-based urban road traffic accidents analysis, Southeast University,41(6)(2011)1300-1306.

[13]. Bedard, M., Guyatt, G. H., Stones, M. J., &Hireds, J. P., "TheIndependent Contribution of Driver, Accident, and VehicleCharacteristics to Driver Fatalities, Accident analysis and Prevention,Vol. 34, pp. 717-727, 2002.

[14]. Miao M. Chong, Ajith Abraham, MarcinPaprzycki, "Traffic AccidentAnalysis Using Decision Trees and Neural Networks", IADISInternational Conference on Applied Computing, Portugal, IADISPress, Pedro Isaias et al. (Eds.), Volume 2, pp. 39-42 2004.

[15]. Chang, L. and W. Chen "Data mining of treebased models to analyzefreeway accident frequency", Journal of Safety Research 36: 365- 375,2005.

[16]. T. Beshah, "Application of data mining technology to support RTAseverity analysis at Addis Ababa traffic office", Addis Ababa, AddisAbaba University, 2005.

[17]. Chang, L. and H. Wang, "Analysis of traffic injury severity: Anapplication of non-parametric classification tree techniques Accidentanalysis and prevention", Accident analysis and prevention 38(5):1019-1027, 2006.

[18]. Srisuriyachai, S., "Analysis of road traffic accidents in NakhonPathomprovince of Bangkok using data mining", Graduate Studies, Bangkok,Mahidol University, 2007.

[19]. Wong, J. and Y. Chung,"Comparison of Methodology Approach toIdentify Causal Factors of Accident Severity." Transportation ResearchRecord 2083: 190-198, 2008.
    Zelalem, R., "Determining the degree of driver's respon000sibility for caraccident: the