

MODIFIED APRIORI ALGORITHM FOR ROAD ACCIDENT ANALYSIS

Lata Sukhadiya¹, Harsh Kumawat²

¹M-Tech Scholar Computer Science, ²Prof. Computer Science, Pratap University Jaipur.

Abstract: Road Accidents are very common in India and taking number of lives which are very precious for their families and even for the nation. In our work we have worked in this segment to find out the causes of the accidents and refining the major causes of accidents. In this paper, we have mentioned the modified apriori algorithm which computes the more refined reasons for road accidents and the comparative analysis is also made with the existing apriori algorithm.

I. INTRODUCTION

India's daily death toll due to road accidents is more than four times the annual death toll from terrorism. As many as 139,671 people lost their lives on India's roads during 2014 – 382 deaths every day. For comparison, the total number of deaths (civilians and security personnel) due to terrorism-related incidents was 83 in all of 2014.

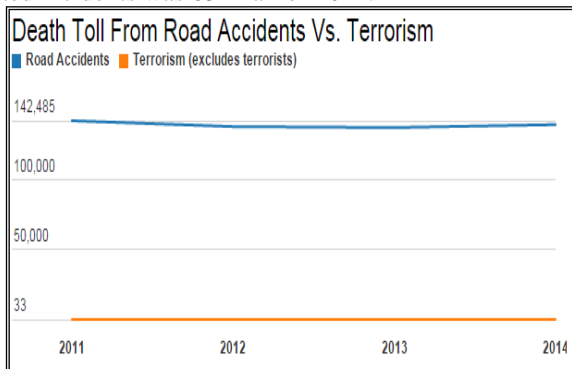


Fig 1. Death ratio in road accidents and terrorism.

Predictably, most of those who die on the roads perish because of preventable causes: speeding, drink driving and overloading.

The large number of deaths among pedestrians and cyclists also indicates that any moves to get people to shun motorized transport in favor of these environmentally-friendly modes are not likely to succeed.

After falling for two years, the number of road accident deaths in India rose again during 2014, according to the latest report from the Ministry of Road Transport & Highways.

The majority of these deaths, nearly three-fourths, have been termed as 'fault of the driver', a catch-all term that includes speeding, drink driving, driving on wrong side of the road and not signaling properly.

- Speeding is the biggest killer on Indian roads. Speeding is the single factor responsible for the maximum number of deaths on Indian roads. During 2014, 57,844 deaths – 41% of the total – were due to accidents caused by speeding.
- Speeding has accounted for a similar share in the

earlier years as well and has consistently accounted for over 50,000 deaths on roads for the past several years.

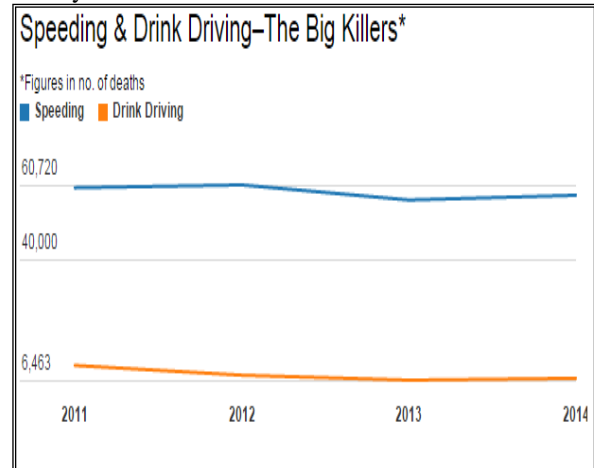


Fig 2. Speeding and Drink Driving Analysis

Speeding is typically the easiest factor to control, and a small reduction in vehicle speed yields disproportionate results in terms of safety. A pedestrian struck by a car driving at 37 km/hr has an average risk of death of 10%, according to a study sponsored by AAA Foundation, a US association dedicated to road safety. This increases exponentially with vehicle speed and rises to 90% for higher speeds.

These numbers suggest some big gains can be made with directed effort.

India's national and state highways, which together account for less than 5% of the road network, accounted for 63% of the total road deaths during 2014. Speeds on highways are typically higher than speeds within city limits and are often violated because of lax enforcement. The Mumbai-Pune Expressway is a prime example; the speed limit of 80 km/hour is observed more in the breach. Strict enforcement of speed limits on highways could save thousands of lives.

II. DATA MINING AND ACCIDENT ANALYSIS

Information mining is a promising zone for managing the expanded, put away information that has been produced in our times. It is the extraction of understood, already obscure and valuable information. In this paper we have examined a percentage of the information mining systems, devices, applications and web indexes for mishap examination and activity investigation. The vast majority of the mishap examination procedures depend on situations of the mischance event and reenactment of mishap circumstance. The expenses of fatalities and wounds because of auto collision greatly affect society.

Architects and analysts in the car business have attempted to outline and assemble more secure vehicles, however auto collisions are unavoidable. As of late, scientists have been using genuine information in examining different parts of car crashes. So measures must be taken to diminish mischances. It is vital that the measures ought to be founded on investigative and target reviews of the reasons for mischance's and seriousness of wounds. Our study highlights different devices, strategies and utilizations of information mining in mischance examination will take out lacks of different systems however covers their focal points. Our fundamental point is to beat the passing rate and the expanded rate of death toll by method for utilizing a few apparatuses, procedures or different calculations in the field of information mining utilizing the movement information bases.

III. LITERATURE IN DATA MINING

"Investigation of Data Mining Tools in Knowledge Discovery Process" By Y. Ramamohan, K. Vasantharao, C. KalyanaChakravarti, A.S.K.Ratnam

Information mining, the extraction of concealed prognostic data from enormous databases, might well be a solid new innovation with pleasant potential to encourage organizations concentrate on most vital data in their data distribution centers. It utilizes machine learning, connected science and picture methods to revelation and blessing data in a structure that basic coherent to people. differed all around enjoyed information mining instruments are get-capable nowadays. information mining apparatuses anticipate future patterns and practices, permitting organizations to make proactive, learning driven decisions. informationmining instruments can answer business questions that generally were excessively time overpowering, making it impossible to determine.

In its least complex kind, information mining robotizes the discovery of significant examples in an extremely data, use plot methodologies and calculations to investigate present and chronicled information which can then be dissected to anticipate future patterns. as a consequences of information reading so as to mine devices anticipate future patterns and practices through databases for concealed examples, they permit associations to make proactive, learning driven picks and answer questions that were precursor too long to resolve. Krutika. K .Jain and Anjali . B. Raut Propose Because of the rapid growth in worldwide information, efficiency of association rules mining (ARM) has beenconcerned for several years. Association rule mining plays vital part in knowledge mining. The difficult task is discovering knowledge or useful rules from the large number of rules generated for reduced support In this paper, based on the Apriori algorithm association rules is based on interestingness measures such as support, confidence and so on. Confidence value is a measure of rule's strength, while support value corresponds to statistical significance. Traditional association rule mining techniques employ predefined support and confidence values. However, specifying minimum support value of the mined rules in advance often leads to either too many or too few rules, which negatively impacts the performance of the overall

System. In this algorithm, we will create association rules depending upon the dataset available in the database. The algorithm majorly works on finding the minimal confidence and so association rules which frequently used and follow the minimum confidence. So the research part of this paper is this by changing the value of minimum confidence, gives different association rules. The value of minimum confidence is high then rules filtered more accurately.[19]

Neelamadhab Padhy¹, Dr.Pragnyaban Mishra ², and Rasmita Panigrahi³In this paper we have focused a variety of techniques, approaches and different areas of the research which are helpful and marked as the important field of data mining Technologies. As we are aware that many MNC's and large organizations are operated in different places of the different countries. Each place of operation may generate large volumes of data. Corporate decision makers require access from all such sources and take strategic decisions .

Akshita Bhandari¹, Ashutosh Gupta², Debasis Das³There are several mining algorithms which have been developed over the years. Apriori Algorithm is one of the most important algorithm which is used to extract frequent itemsets from large database and get the association rule for discovering the knowledge. It basically requires two important things: minimum support and minimum confidence. Firstly, we check whether the items are greater than or equal to the minimum support and we find the frequent itemsets respectively. Secondly, the minimum confidence constraint is used to form association rules. Based on this algorithm, this paper indicates the limitation of the original

Rajni Jindal and Malaya Dutta Borah Educational Data Mining (EDM) is an emerging field exploring data in educational context by applying different Data Mining (DM) techniques/tools. It provides intrinsic knowledge of teaching and learning process for effective education planning. In this survey work focuses on components, research trends (1998 to 2012) of EDM highlighting its related Tools, Techniques and educational Outcomes. It also highlights the Challenges EDM.

Rajni Jindal and Malaya Dutta Borah Educational data mining (EDM) is an emerging discipline that focuses on applying data mining tools and techniques to educationally related data. The discipline focuses on analyzing educational data to develop models for improving learning experiences and improving institutional effectiveness. A literature review on educational data mining follow s, which covers topics such as student retention and attrition, personal recommender systems within education, and how data mining can be use d to analyze course management system data. Gaps in the current literature and opportunities for further research are presented.

Smita¹, Priti Sharma² Data mining is extracts the knowledge/ information from a large amount of data which stores in multiple heterogeneous data base. Knowledge /information are conveying the message through direct or indirect. This paper provides a survey of various data mining techniques. These techniques include association, correlation, clustering and neural network. This research paper also conducts a formal review of the application of

data mining such as the education sector, marketing, fraud detection, manufacturing and telecommunication. This paper discusses the topic based on past research paper and also studies the data mining techniques.

Mohammed Al-Maolegi¹, Bassam Arkok² There are several mining algorithms of association rules. One of the most popular algorithms is Apriori that is used to extract frequent itemsets from large database and getting the association rule for discovering the knowledge. Based on this algorithm, this paper indicates the limitation of the original Apriori algorithm of wasting time for scanning the whole database searching on the frequent itemsets, and presents an improvement on Apriori by reducing that wasted time depending on scanning only some transactions. The paper shows by experimental results with several groups of transactions, and with several values of minimum support that applied on the original Apriori and our implemented improved Apriori that our improved Apriori reduces the time consumed by 67.38% in comparison with the original Apriori, and makes the Apriori algorithm more efficient and less time consuming.

PratibhaMandave, Megha Mane and Prof.SharadaPatil In this paper we have explain one of the useful and efficient algorithms of Association mining named as APRIORI algorithm. Association rule of data mining is used in all real life applications of business and industry. Using this we gets an effective results rather than traditional results. Association rules are the main technique for data mining and APRIORI algorithm is a classical algorithm. Lots of algorithms for mining association rules and their mutation (change/transformation) are proposed on basis of APRIORI algorithm, but traditional algorithms are not efficient. The main intension of this paper is to understand the concept of association rule and how to implement the APRIORI algorithm and improved APRIORI algorithms..

IV. PROPOSED ALGORITHM

The goal of the proposed method is to reduce CPU time which is saved by reducing candidate set size. If candidate set size is less than time required to calculate the support of each candidate is less. We have proposed Method that reduces the number of candidate generated and time required to calculate the support of each candidate. In order to reduce CPU times, we have defined two type of checkpoint in dataset based on support value:

checkpoint1=Totaltransaction-support count+1; \forall min_sup;
 checkpoint2=support count+1; if min_sup>50.

Support property: If support count is n than any item to be frequent it must be appearing in at least n transactions in the dataset.

All new candidates after checkpoint1 cannot be frequent based on support property. At checkpoint1 if min_sup<=50 or checkpoint2 if min_sup>50, scan the candidate set once and check the support value of all candidate. The estimated support value is used to remove infrequent itemsets at checkpoint.

Estimated Support value(Esupport)= support of candidate + Totaltransaction - checkpoint1; if min_sup<=50.

Estimated Support value(Esupport)= support of candidate + Totaltransaction - checkpoint2; if min_sup>50.
 If Estimated Support value < support count then that candidate will be removed from the candidate set.

In this we have suggested the new modified algorithm for data mining which is based on the FApriori algorithm in order to refine the result we have devised the algorithm which works in the following manner.

- Step 1 : Take the Minimum Support count.
- Step 2: Calculate the values of CheckPoint1 , CheckPoint2 , CheckPoint3,CheckPoint4 using the following formulas,

CheckPoint1 =Number of Transactions-Support Count+1

CheckPoint2 =Number of Transactions/2

Step 3: Calculate the Support of DataSet using the Apriori algorithm and then the Esupport in first half is calculated on the basis of Checkpoint1 and checkpoint2 and Esupport on the second half is calculated on the basis of Checkpoint3 and Checkpoint4.

Step4: Find the Combinations on the basis of the Esupport and form the result.

In this work, the author has proposed a system in which we will first manage the data used in the Accident analysis into the various tables and using a GUI software system which summarize the causes of the accident using the Apriori Algorithm.

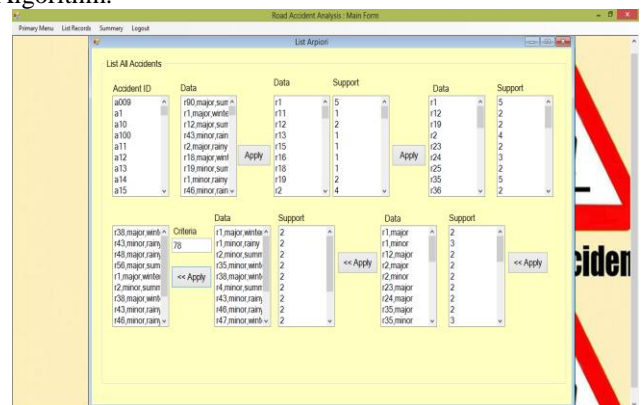


Fig 3. Apriori Algorithm

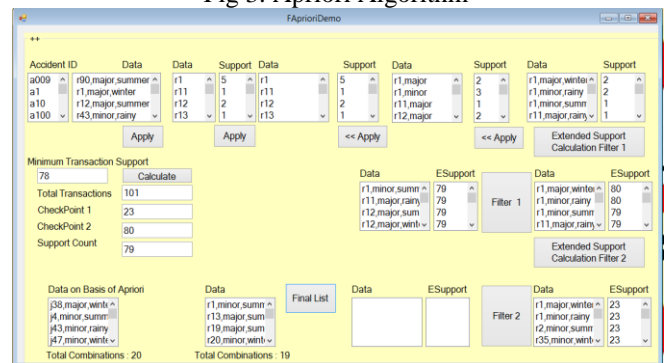


Fig 4. Modified Apriori Algorithm

In the proposed implementation we have taken a dataset in excel file which is used for storing the record information related on the accidents. And this data is collected by surfing the various official sites providing the information related to the road accidents.

accident_id	road_id	severity	accident_type	date	time	weather
1	11	major	22/3/1990	10:0am	winter	
2	12	minor	24/3/1989	3:00pm	summer	
3	13	major	1/3/2001	4:00pm	winter	
4	14	major	24/3/2004	8:00am	winter	
5	15	minor	11/3/1994	8:00pm	summer	
6	16	minor	23/4/1998	7:00pm	summer	
7	17	minor	28/5/1996	8:00pm	rainy	
8	18	minor	10/6/1997	5:00pm	winter	
9	19	minor	25/6/1990	11:00am	summer	
10	20	major	19/6/1991	11:00am	summer	
11	21	major	11/3/1989	7:00pm	rainy	
12	22	major	14/3/2005	8:00am	winter	
13	23	minor	21/6/1999	8:00am	summer	
14	24	minor	24/6/1998	10:00am	rainy	
15	25	minor	29/6/2001	7:00pm	rainy	
16	26	minor	12/6/1991	7:00am	summer	
17	27	minor	24/6/1998	7:00am	winter	
18	28	major	26/6/2001	8:00am	summer	
19	29	minor	30/6/1992	10:00am	winter	
20	30	minor	30/6/1997	7:00pm	rainy	
21	31	minor	29/6/1998	8:00am	summer	
22	32	minor	6/5/2004	winter	winter	
23	33	major	05/6/1994	rainy	rainy	
24	34	major	19/6/2008	winter	winter	

Fig 5. Excel DataSet File

V. CONCLUSION

The objective of the proposed technique is to diminish CPU time which is spared by lessening hopeful set size. On the off chance that competitor set size is not as much as time required to figure the backing of each hopeful is less. We have proposed Method that lessens the number of hopeful produced and time required to figure the backing of every hopeful. Keeping in mind the end goal to decrease CPU times, we have characterized two kind of checkpoint in dataset.

We will further try of implement the other algorithms of the data mining to see the improve results and compare the existing results with the new results. We proposed a strategy in view of bolster esteem that expand the execution of Apriori calculation as portray above. This change is because of the way that the proposed strategy minimizes the quantity of hopeful produced and evacuated applicant at checkpoint which is rare which assistants decreases stockpiling and time required to ascertain support of competitor. At present, there is further work in advancement to develop the proposed calculation in connection to decrease the I/O time.

REFERENCES

[1] J. Han and M. Kamber, "Data Mining Concepts and Techniques, "second Edition, Morgan Kaufmann Publishers, 2006.

[2] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In Proc. 20th Int. Conf. Very Large Data Bases, pp. 487-499, 1994.

[3] J.S. Park, M.S. Chen and P.S. Yu, "An Effective hash-based algorithm for mining association rules," in Proceedings of ACM SIGMOD, pp. 175-186, 1995.

[4] S.A. Ozel, and H.A. Guvenir, "An algorithm for mining association rules using perfect hashing and database pruning," in: Springer 10th Turkish Symposium on Artificial Intelligence and Neural

Networks, Gazimagusa, pp. 257-264, 2001.

[5] M. J. Zaki, "Scalable algorithms for association mining," IEEE Trans. Knowledge and Data Engineering, pp.372-390, 2000.

[6] J. Han, J. Pei and Y. Yin, "Mining frequent patterns without candidate generation", Proceedings of SIGMOD, 2000.

[7] F. Wang and Y. Li, "An Improved Apriori Algorithm Based on the Matrix," 2008 International Seminar on Future BioMedical Information Engineering, pp. 152-155, Dec. 2008.

[8] R. Chang and Z. Liu, "An improved apriori algorithm," Proceedings of 2011 International Conference on Electronics and Optoelectronics, Jul. 2011.

[9] Available:<http://www.philippe-fournier-viger.com/spmf/index.php?link=algorithms.php> [Online].

[10] Available: <http://fimi.ua.ac.be/data/> [Online].

[11] Smita1, Priti Sharma2, "Use of Data Mining in Various Field: A Survey Paper", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. V (May-Jun. 2014), PP 18-21.

[12] Richard A. Huebner and Norwich University, "A survey of educational data-mining research", Research in Higher Education Journal.

[13] Shen Bin, Liu Yuan, Wang Xiaoyi, "Research on Data Mining Models for the Internet of Things",Institute of Electrical and Electronics Engineers,2010

[14] XindongWu,Gong-Qing Wu, and Wei Ding, "Data Mining with Big Data",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, JANUARY 2014

[15] Feng Bao, Xu He, Fengzhi Zhao, "Applying Data Mining to the Geosciences Data",International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE) ,2010

[16] Ms Shweta, Dr. Kanwal Garg "Mining Efficient Association Rules ThroughApriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 6, June 2013.

[17] http://sfb649.wiwi.huberlin.de/fedc_homepage/xplo/re/ebooks/html/csa/node203.html.

[18] http://users.jyu.fi/~samiayr/pdf/mining_road_traffic_accidents.pdf.

[19] Krutika. K .Jain and Anjali . B. Raut, "Review paper on finding Association rule using Apriori Algorithm in Data mining for finding frequent pattern", International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015 ISSN 2091-2730.

[20] Neelamadhab Padhy1, Dr. Pragnyaban Mishra 2, and Rasmita Panigrahi3, "The Survey of Data Mining Applications And Feature Scope",

- International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.3, June 2012.
- [21] Akshita Bhandari¹, Ashutosh Gupta², Debasis Das³, "IMPROVED APRIORI ALGORITHM USING FREQUENT PATTERN TREE FOR REAL TIME APPLICATIONS".
- [22] Rajni Jindal and Malaya Dutta Borah, "A SURVEY ON EDUCATIONAL DATA MINING AND RESEARCH TRENDS", International Journal of Database Management Systems (IJDBMS) Vol.5, No.3, June 2013.
- [23] Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011", S.-H. Liao et al. / Expert Systems with Applications 39 (2012) 11303–11311.
- [24] Nikita Jain¹, Vishal Srivastava², "DDATA MINING TECHNIQUES: A SURVEY PAPER", IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308.
- [25] Y.Elovici¹, A.Kandel², M.Last¹, B.Shapira¹, O. Zaafrany¹, "Using Data Mining Techniques for Detecting Terror-Related Activities on the Web"
- [26] Kalyani M Raval,"Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 10, October 2012 ISSN: 2277 128X
- [27] Mohammed Al-Maolegi¹, Bassam Arkok², "AN IMPROVED APRIORI ALGORITHM FOR ASSOCIATION RULES", International Journal on Natural Language Computing (IJNLC) Vol. 3, No.1, February 2014.
- [28] Jiao Yabing, "Research of an Improved Apriori Algorithm in Data Mining Association Rules", International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013.
- [29] PratibhaMandave, Megha Mane and Prof. SharadaPatil, "Data mining using Association rule based on APRIORI algorithm and improved approach with illustration", International Journal of Latest Trends in Engineering and Technology (IJLTET).