Four Checkpoint Modified Apriori Algorithm for Data Mining in Accidents Analysis

Shivangi Dheer & Miss. Priyanka Punjabi
Gyan Vihar University

Abstract: Mischances specifically to discuss the railroad mishaps happen consistently in India yet then can be lessened on the off chance that we dissect the reason for mischances. In our study we are thinking about some elements like Road which are associated with the Junctions where the mishaps occurred, climate in which the mischance occurred, day time or evening time when the mischance happens and more variables. And afterward we attempt to discover the mixes utilizing the Modified Checkpoint based Apriori Algorithm the mixes of the variables which causes the most mishaps and attempt to mull over them.

Introduction

Indian Railways is get ready for taking off quick voyager advantages even as it gives off an impression of being set to record its most exceedingly awful execution in three years on accidents. Fresh data on wellbeing exhibits complete number of accidents per million train kilometers has found the center estimation of 0.14 in the present cash related year so far - higher than the numbers recorded in each of the last three years.[1]

The extent of number of accident's that bounced out at every million kilometer secured by means of trains has declined dependably from 0.29 in 2004-05 to 0.10 in 2013-14. Regardless, as showed by railways' latest capability parameter report, the extent stayed at 0.14 amidst April and September 2014. In case kept up at the present level, the current money related would check the foremost yearly rising in the extent in the past decade.[1]

Data mining is a promising zone for dealing with the extended, set away data that has been delivered in our times .It is the extraction of saw, effectively dark and profitable data. In this paper we have inspected a rate of the data mining frameworks, gadgets, applications and web lists for accident examination and action examination. Most by far of the setback examination methodology rely on upon circumstances of the accident occasion and reenactment of incident situation. The costs of fatalities and wounds due to car accident enormously influence society.

Draftsmen and examiners in the auto business have endeavored to diagram and gather more secure vehicles, however car accidents are unavoidable. Starting late, researchers have been utilizing real data as a part of looking at changed parts of auto collisions. So measures must be taken to decrease incidents. It is key that the measures should be established on investigative and target surveys of the purposes behind incident's and reality of wounds. Our study highlights diverse gadgets, procedures and usages of data mining in setback examination will take out absence of various frameworks however covers their central focuses. Our major point is to beat the passing rate and the extended rate of loss of life by technique for using a couple mechanical assemblies, methodology or diverse computations in the field of data mining using the development data bases [2].

1. Importance And Relevance Of The Study

Shen Bin 1, Liu Yuan 1, Wang Xiaoyi 1 propose four data burrowing models for the Internet of Things, which are multi-layer data mining model, circled data mining model, Grid based data mining model and data mining model from multi-advancement joining perspective. Among them, multi-layer model consolidates four layers: 1) data gathering layer, 2) data organization layer, 3) event planning layer, and 4) data mining organization layer. Passed on data mining model can deal with issues from putting away data at different destinations. Cross section based data mining model grants Grid structure to comprehend the components of data mining. Data mining model from multi-development compromise perspective depicts the relating structure for the future Internet. [1]

Xindong Wu 2,Gong-Qing Wu 2, and Wei Ding 2 exhibits a HACE hypothesis that describes the components of the Big Data transformation, and proposes a Big Data handling model, from the information mining point of view. This information driven model includes request driven total of data sources, mining and examination, client enthusiasm demonstrating, and security and protection contemplations. They break down the testing issues in the information driven model furthermore in the

Big Data insurgency. [2]

Feng Bao 3, Xu He 3, Fengzhi Zhao 3, addresses the components of the petro physical information, logging information, seismic information and land information in view of the ideas of the information mining. The mining thoughts with respect to the petro physical and logging information, seismic information and topographical information are made in light of their elements. They utilizes diverse mining approaches to handle the comparing information, and portrays the outcome from the point of view of the elements of information mining.

By data mining frameworks, the petro physical data are associated with find the relations and guess storehouse the logging data will be used to survey the fleecy stores and see the capable supplies in confounded geological conditions; the space mining outcome of the 3D seismic data; the diagrams and substance mining eventual outcomes of the land data. [3]

Ms Shweta 4, Dr. Kanwal Garg 4 considers data (bank data) and tries to gain the result using Weka a data mining instrument. Association rule figurings are used to find the best blend of different qualities in any data. In this paper maker uses Apriori to find connection rule. Here maker consider three connection rule counts: Apriori Association Rule, Predictive Apriori Association Rule and Tertius Association Rule. Ms Shweta, Dr. Kanwal Garg investigates the eventual outcome of these three counts and exhibits the result. By result procured using data mining instrument maker find that Apriori Association figuring performs better than the Predictive Apriori Association Rule and Tertius Association Rule computations. [5]

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

Data Mining, the extraction of concealed prognostic data from huge databases, may well be a solid new innovation with pleasant potential to encourage organizations concentrate on most vital data in their information stockrooms. It utilizes machine learning, connected arithmetic and picture procedures to revelation and blessing data in a structure that basic clear to people. Differed very much preferred Data mining devices are get-capable nowadays. Data mining devices foresee future patterns and practices, permitting organizations to make proactive, learning driven decisions. Data Mining devices can answer business questions that customarily were excessively time overpowering, making it impossible to determine.

In its most straightforward kind, Data Mining robotizes the location of pertinent examples in an extremely data, use diagram methodologies and calculations to investigate present and verifiable information which can then be broke down to foresee future patterns. as an aftereffects of Data Mining apparatuses foresee future patterns and practices by perusing through databases for shrouded designs, they permit associations to make proactive, learning driven picks and answer inquiries that were predecessor too long to resolve.[5]

Another paper is "Mining Big Data in Real Time" By Albert Bife.

Spilling data examination progressively is changing into the fastest and most antiquated on account of get supportive data from what is going on presently, allowing associations to respond rapidly once issues appear or to see new patterns serving to upgrade their execution. Developing information streams square measure contributory to the development of learning made in the course of the most recent couple of years. we tend to square quantify making the same measure of data every 2 days, as we tend to made from the beginning of time up till 2003.Evolving information streams procedures are getting a modest, unpracticed system for timeframe on-line forecast and examination. we tend to talk about this and future patterns of mining advancing learning streams, and the difficulties that the held can got the opportunity to overcome all through continuous years.

These days, the quantity of data that is made every 2 days is measurable to be five Exabyte's. This amount of learning is like the amount of data made from the beginning of your time up till 2003. Besides, it completely was measurable that 2007 was the rest of inside which it totally was unattainable to store all the data that we tend to are assembling. This expansive amount of data opens new troublesome disclosure errands. information stream ongoing examination square measure required to deal with the learning without further ado created, at A regularly expanding rate, from such applications as: gadget systems, estimations in system recognition and track administration, log records or snap streams in web investigating, delivering forms, choice point of interest records, email, blogging, twitter post sand others. Indeed, all data produced are regularly considered as gushing learning or as a photograph of spilling data, since it is gotten from an interim of your time. In the learning stream model, data achieve fast, and calculations that strategy them ought to do hence underneath horrendously strict requirements of house and time. Thusly, information streams make numerous difficulties for information preparing

algorithmic project style. To begin with, calculations ought to make utilization of limited assets (time and memory). Second, they need to handle with data whose nature or conveyance changes overtime.[6]

Proposed Concept

The objective of the proposed technique is to decrease CPU time which is spared by lessening competitor set size. In the event that applicant set size is not as much as time required to compute the backing of every hopeful is less. We have proposed Strategy that lessens the quantity of applicant produced and time required to ascertain the backing of every competitor. 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
CheckPoint3 =Number of Transactions/2 +1
CheckPoint4 =Support Count+1

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 creator has proposed a framework in which we will first deal with the information utilized as a part of the Mishap examination into the different tables and utilizing a GUI programming framework which compresses the reasons for the mischance utilizing the Changed Checkpoint based Apriori Calculation.

Proposed System has the following DATABASE SCHEMA

- Person Table
- Junction Table
- Roads Table
- Road_Check
- Accident Table

Person Table

This table is used for storing the information regarding the drivers driving the trains at the time of the accident.

Junction Table

This table is used for storing the details regarding the junctions.

Roads Table

This table is used for storing the information regarding the roads which are connecting roads or roads at the crossing to the nearest junctions.

Accident Table

This table is used for storing the details of accidents .

Road_check table

This table is used for categorization of the road.

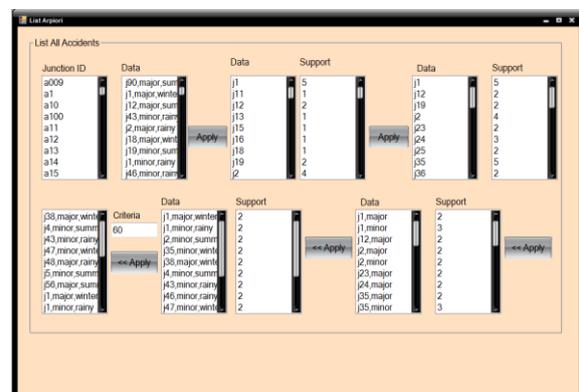In the figure 1 , we have implemented the Aprori based implementation on the sample data.

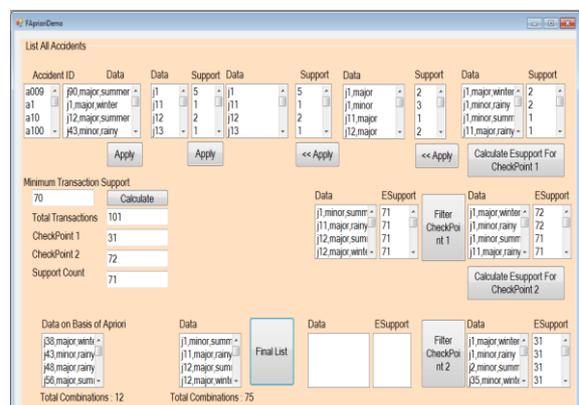**Figure 1: Apriori Algorithm Implementation On The Sample Data.**
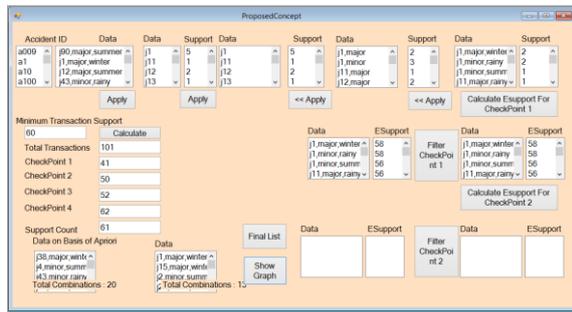
Figure 2 FAprori Algorithms

Figure 3 Proposed Concept

In the figure 4 , we have shown the implementation of the proposed four checkpoint based apriori algorithm.

.

Figure 4 Example of the sample data which we have used in the analysis process.

Figure 4 shows the data which we have used in the proposed algorithm for the simulation of the algorithm.
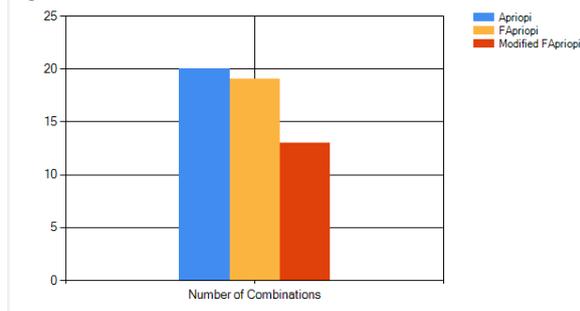

Figure 5 Graphical Comparison of 3 algorithms

Conclusion & Future Scope

This study is revolved around how to deal with the successful issues of Apriori count and propose another changed 4 check point based data mining apriori estimation. It focus on component finding for train incidents examination and still data mining is an incredibly unfathomable field so constantly a degree to improve and develop new thoughts for upgrading the eventual outcomes of the data mining operations.

References

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

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

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

[4] Ms Shweta, Dr. Kanwal Garg "Mining Efficient Association Rules Through Apriori 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.

[5]http://sfb649.wiwi.huberlin.de/fedc_homepage/xplore/ebooks/html/csa/node203.html

[6]

http://users.jyu.fi/~samiayr/pdf/mining_road_traffic_accidents.pdf

[7]

http://www.thearling.com/text/dmtechniques/dmtechniques.htm