iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: IV Month of publication: April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4207

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Analysing the Train Accident Injuries using Mining Techniques

CH. Nireesha[1], N. Vijay Kumar[2], V. Babu[3]
[1,2,3] *Department of I.T, LBRCE*

Abstract*: Rail accidents describe a significant safety concern for the transportation industry in multiple countries. To better understand the contributors to these extreme accidents, the Federal Railroad Administration has established the railroads involved in accidents to submit reports that contain both fixed function entries and narratives that explain the characteristics of the accident. While a number of studies have noticed the fixed fields, nothing has done an extensive analysis of the narratives. This project describes the use of mining with a combination of techniques to automatically discover accident characteristics that can confess a better understanding of the contributors to the accidents. The results show that predictive accuracy for accident costs significantly improves on the use of features found by mining and predictive accuracy also improves through the use of modern ensemble methods. Importantly, this study furthermore shows through case examples how the findings from mining of the narratives can improve understanding of the contributors to rail accidents in ways not possible on only fixed field analysis of the accident reports.
Keywords: Association rules, Data Mining, Apriori algorithm, Rail safety, Safety engineering.

I. INTRODUCTION

In the 11 years from 2001 to 2012 we collect the information the U.S. had more than 40000 rail accidents with a total cost of $45.9 M. These accidents resulted in some deaths and injuries. Federal Railroad Administration (FRA) has collected the data to understand and find the ways to reduce the number of accidents. Finally the FRA has set “an ultimate goal of zero tolerance of rail-related accidents, injuries, and fatalities”. A review of the data collected from a FRA shows a variety of accidents and most of the accidents are not serious; since, they cause little damage and no injuries. However, there are some that cause over $1M in damages, deaths of crew and passengers, and many injuries. Here the problem is to understand the characteristics of these accidents and also inform the both system design and policies to improve the safety.

This report has number fields that include details about the train like train number, name, and compartment and also include environmental characteristics like temperature and precipitation and operational characteristics like speed and cost estimation and finally primary cause of the accident. Here we mainly focused on the injuries of the accident. After each accident report is completed and submitted to the FRA (Federal Railroad Administration) by the railroad companies involved.

In this paper we applied a prediction algorithm on injuries and predicts the what types injuries are happened more frequently and their combination also. The FRA uses all of these data much as the Federal Aviation Administration uses report on aviation accidents, namely to develop hazard elimination and risk reduction programs that focus on preventing the railroad injuries and accidents. For this we applied an apriori algorithm. Apriori algorithm is a part of risk reduction algorithms. The Apriori algorithm is an influential algorithm for mining frequent item sets for Boolean association rules.

This paper describes to get an idea about the possible predictors or contributors to accidents obtained from “mining” the narrative text in rail accident reports. The purpose of the final analysis is to understand about the accident and about also about the injuries. Finally we collect the how many injuries are occurred and the combination of injuries that frequently occurs. And finally this report is submitted to the FRA (Federal railroad administration) and FRA fined the ways to reduce the number and severity of these accidents.

II. BACKGROUND

In this paper integrates methods for safety analysis with accident report data and text mining to uncover contributors to rail accidents. This section describes related work in rail and, more generally, transportation safety and also introduces the relevant data and text mining techniques. One of the well-studied areas of rail safety concerns rail crossings by roadways. A recent application of fuzzy sets and clustering to guide the selection of rail crossings for active safety systems (e.g., bells, lights, and barriers) is in [1]. Tey *et al.* [2] describe the use of logistic regression and mixed regression to model the behaviour of drivers at railway crossings.

The paper by Akin and Akbas [3] describes the use of neural networks to model intersection crashes and intersection characteristics, such as, lighting, surface materials, etc. Taken together these papers show the use data mining to better understand the factors that can influence and improve safety at rail crossings. Recent work has shown the applicability of data and text mining to broader classes of safety and security problems relevant to transportation. For example, the use of data mining techniques for anomaly detection in road networks is illustrated by the work of [4]. They provide methods to detect anomalies in massive amounts of traffic data and then cluster these detections according to different attributes. Similarly D'Andrea *et al.* mined Twitter and used support vector machines to detect traffic events [5]. Another recent application of text mining is to license plate recognition [3]. These authors use Levenshtein text mining in combination with a Bayesian approach to increase the accuracy of automated license plate matching. Cao *et al*., use data mining in combination with rule-based and machine learning approaches to perform traffic sentiment analysis [3]. Speech processing and message feature extraction have been used for detection of intent in traveller screening [4]. Henzel [6] describes the use of eddy current sensors to provide more precise location of trains for positive control. Parallel control for emergency response is presented in [6]. Meyers *et al.* [7] describe risk assessment methods for evaluating the safety of PTC. They also discuss the many challenges in performing this risk assessment. The work we describe in the subsequent sections of this paper can better inform these risk assessments. In particular, the text mining approach we describe can enable a better understanding of the characteristics of accidents that PTC may prevent and those that it cannot.

III. EXISTING SYSTEM

In the existing system, one of the well-studied areas of rail safety concerns rail crossings by roadways. A recent application of fuzzy sets and clustering to guide the selection of rail crossings for active safety systems (e.g., bells, lights, and barriers) is in this system.
Tey *et al.* [4] describe the use of logistic regression and mixed regression to model the behaviour of drivers at railway crossings. The paper by Akin and Akbas describes the use of neural networks to model intersection crashes and intersection characteristics, such as, lighting, surface materials, etc. Taken together these papers show the use data mining to better understand the factors that can influence and improve safety at rail crossings.

IV. PROPOSED SYSTEM

First this paper describes a broader comparison of techniques than previous studies. Specifically the proposed system gives results for comparisons between no text mining and two contemporary approaches to text mining in combination with three approaches to supervised learning. This three by three design provides a broader range of evaluation than any previous study. Second, this paper focuses on rail accident reports over a longer time span than other studies; namely, 11 years. Third none of the text mining analytics described here have previously been applied to rail accident damage assessment.

A. Proposed Algorithm

Apriori algorithm

Association rule generation is usually in two separate steps

1) First, minimum support is applied to find all frequent item sets in a database.

2) Second, these frequent item sets and for forming the rules we used the minimum confidence constraint.

While the second step is straight forward, the as a matter of choice step needs more attention. Finding for the most part frequent factor sets in a database is difficult since it involves searching for the most part possible item sets (item combinations). The set of possible item sets is the power fit around I and has length 2n − 1 (excluding the empty fit which is not a reliable item set).

B. Apriori Algorithm Pseudo Code

```
Procedure Apriori (T, minSupport)
{
 L1= {frequent items};
For (k= 2; Lk-1! =∅; k++)
{
 Ck= candidates generated from Lk-1
Do
{
```

Volume 5 Issue IV, April 2017
ISSN: 2321-9653

#increment the count of all candidates in Ck that are contained in t Lk = candidates in Ck with minSupport
}
}
}

As is common in association rule mining, given a fit of item sets (for instance, sets of display transactions, each listing individual items purchased), the algorithm attempts to find subsets which are common to at uttermost a minimum number C of the factor sets. Apriori uses a "bottom up" approach.

Apriori uses breadth-first search and a tree process to tell candidate plug in sets efficiently. It generates candidate factor sets of length k from factor sets of term k − 1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate exist contains all frequent k-length factor sets. After that, it scans the transaction database to verify frequent peripheral sets among the candidates.

V. EXPERIMENTAL RESULTS

We have experimentally approved our attacks with an algorithm. The performance analysis graph shows how the level1, level2, level3 graph changes. In this graph x-axis is levels of item sets and the y-axis is how many combinations of the item sets we get in that particular level. It is the dynamic graph it changes based on the minimum support count given by the user. Itemset1 shows the one combination of the candidate generation. Based on the itemset1 data and also taking the minimum support count from user we generate a level1 graph. Itemset2 shows the two combination of the candidate generation. Based on the itemset2 data and also taking the minimum support count from user we generate a level2 graph. Itemset3 shows the three combination of the candidate generation. Based on the itemset3 data and also taking the minimum support count from user we generate a level3 graph.

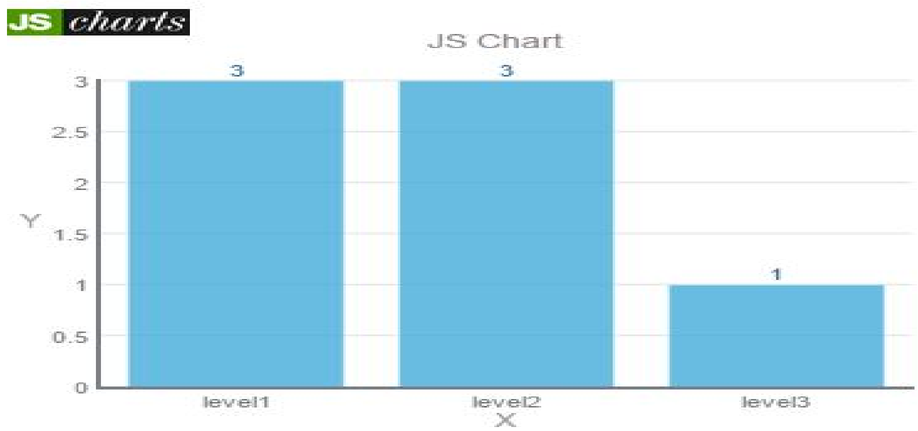

FIG NO:1

And the second graph shows the difference between these paper and previous paper. Here we consider previous paper wroks on primary cause of the accident.

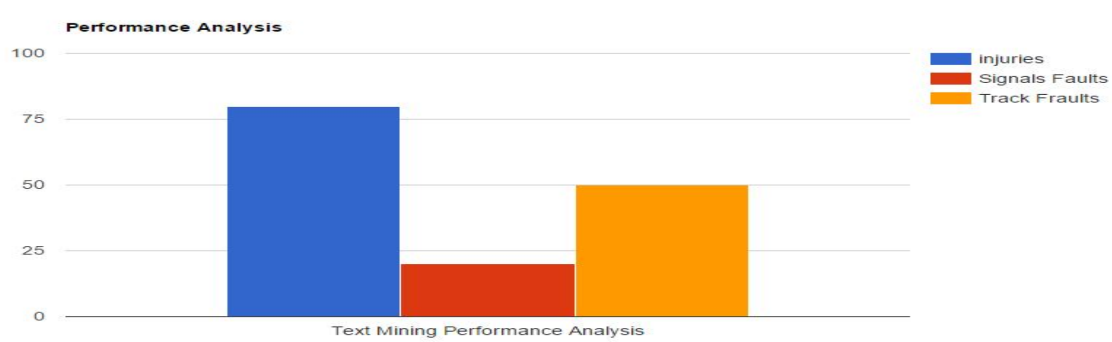

FIG NO:2

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

VI. CONCLUSION

In this Paper, show that the combination of text analysis with ensemble methods can improve the accuracy of models for predicting accident severity and that text analysis can provide insights into accident characteristics. Modern text analysis methods make the narratives in the accident reports almost as accessible for detailed analysis as the fixed fields in the reports. Finally, as described in the work here used standard methods to clean the narratives. And we get combinations of injuries which are more frequently happen and submitted to the FRA (Federal Railroad Administration). For train safety analysis, text mining could benefit from a careful look at ways to extract features from text that takes advantage of language characteristics particular to the rail transport industry.

REFERENCES

[1] "Text mining the contributors to rail accidents" Donald E. Brown, Fellow, IEEE, Feb.2016.
[2] E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, "Real-time detection of traffic from Twitter stream analysis," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 4, pp. 2269–2283, Mar. 2015.
[3] F. Oliveira-Neto, L. Han, and M. K. Jeong, "An online self-learning algorithm for license plate matching," IEEE Trans. Intell. Transp. Syst., vol. 14, no. 4, pp. 1806–1816, Dec. 2013.
[4] J. Burgoon et al., "Detecting concealment of intent in transportation screening: A proof of concept," IEEE Trans. Intell. Transp. Syst., vol. 10, no. 1, pp. 103–112, Mar. 2009.
[5] D. Blei, L. Carin, and D. Dunson, "Probabilistic Topic Models," IEEE Signal Process. Mag., vol. 27, no. 6, pp. 55–65, Nov. 2010.
[6] S. Hensel, C. Hasberg, and C. Stiller, "Probabilistic rail vehicle localization with eddy
[7] current sensors in topological maps," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1525–1536, Dec. 2011.H. Dong et al., "Emergency management of urban rail transportation based on parallel systems," IEEE Trans. Intell. Transp. Syst., vol. 14, no. 2, pp. 627–636, Jun. 2012.