R-Language Based Analytics System For Monitoring Railway Disasters

G S Prasada Reddy\textsuperscript{1}, P Rama Rao\textsuperscript{2}, Vemula Harish\textsuperscript{3} Nirmala Teegala\textsuperscript{4}

\textsuperscript{1}Vardhman College of Engineering, Hyderabad, India
\textsuperscript{2}G Pullaiah College of Engineering and Technology
\textsuperscript{3}Vardhman College of Engineering, Hyderabad
\textsuperscript{4}Department of Computer Science and Engineering MLR Institute of Technology, Hyderabad

*Corresponding author E-mail: gsprasad.reddy56@gmail.com

Abstract

The everyday workplace of railroads is backbreaking and frequently includes high-hazard activities. These operations have its impact on railways crew and public people who lives along the town and cities. To avoid the employees and public's exposure to risk, railway companies uses safety training programs\cite{1}, advanced technology concepts like data analytics and machine learning to analyze about frequent accident occurring zones and safety precautions to be taken in advance. So in this paper we are proposing an visualization model for identifying the frequent occurrence of railway accident zones so as to take enough precautions to avoid them. Here we will be collecting the data on the frequently occurring railway accidents and the causes for it. After the data is been collected, we will be analyzing the data and presenting the information about most frequent and less frequent accident prone areas using visualization tools which provides the clear picture of incidents so as certain precautions or counter measures will be taken at proper times to minimize railway accidents. The results of the analysis done can be used to make and communicate decisions about accidents so as to minimize them.

Keywords: railways, data analytics, visualization, machine learning, R-language

1. Introduction

The Indian Railways(IR)\cite{2} with network size 121407 km and total track 67368km is fourth largest network. Indian railway is running more than 12500 passenger trains covering long distances over country. The Indian railways also in the job of carrying goods across the country. The IR serving huge number of passengers as well as carrying goods. The railway network in India\cite{15} is one of the biggest and busiest railway arranges on the planet, dealing with gigantic quantities of travelers and amounts of products every day. Railroads are the most well known methods for long-separate transportation in India; consequently the railway is frequently portrayed as the foundation of this present country's economy.

For example, factors, for example, the movement between real urban communities surpassing the arranged limit and over-used railroad tracks are bringing about trains traveling at lessened speeds and convey lesser measures of cargo, in this way expanding the cost and time of transportation. Likewise, Indian Railways have been developed way back by British, and have created during the time to its present state. Since the railroad mastermind outlines the establishment of system in India, the progression of the rail route can fill in as a marker of the money related improvement of the country also. Accordingly, the improvement of railroad has also been inspected in this proposition.

1.1 Recent railway accident history:

Despite of the many measures taken by Indian railways, more number of accidents taken place in the from 2017 onwards to till date. Here we collected railway accidents data from 2017 January onwards. There occurred a spate of railway accidents in the year 2017 resulting in loss of large number of human lives and damage to the nation property, disturbances to normal day to day life. There are nearly 9 major railway mishaps\cite{2} and overall 35 above railway accidents in the year 2017 to till date resulting in loss of hundreds of human lives and injuring people. The below figure gives a glance on railway accidents taken place in 2017 year.

<table>
<thead>
<tr>
<th>SNO</th>
<th>Accident type</th>
<th>Accident date</th>
<th>Accident area</th>
<th>Killed people</th>
<th>Injured people</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Derailed</td>
<td>2017</td>
<td>Andhra Pradesh</td>
<td>41 and 68</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Bomb explosion</td>
<td>2017</td>
<td>Madhya Pradesh</td>
<td>0 and 10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>derailed</td>
<td>2017</td>
<td>Uttar Pradesh</td>
<td>0 and 52</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>derailed</td>
<td>2017</td>
<td>Rampur</td>
<td>0 and 24</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>derailed</td>
<td>2017</td>
<td>Uttar Pradesh</td>
<td>23 and 97</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>derailed</td>
<td>2017</td>
<td>Uttar Pradesh</td>
<td>0 and 100</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>derailed</td>
<td>2017</td>
<td>Uttar Pradesh</td>
<td>3 and 9</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Hit</td>
<td>2017</td>
<td>Uttar Pradesh</td>
<td>5 and 2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: List of railway accidents in year 2017

The railways are facing some of the challenging problems. one of them is railway accidents which damages railway and public properties, huge loss of human life and employees of railways etc. There were many reasons for the railway disasters occurring in the daily operations of railways. Some of the most common railway accident types are:

- derailed
- bridge collapse
These kind of railway accidents causing heavy losses in terms of property and life. Hence to reduce these kind of losses at least to small extent, a railway accident analytics system is required to monitor the railway disaster incidents.

Present scenario:
- Data collection
- Precautions are implemented
- But no exact analysis

2. Proposed System:
Here we propose a analytics system which involves
- Collection of Data
- Analysis of Data using analytics package
- Frame Exact precautions depending on location and accident type
- Analysis on more and Less frequent accident prone areas

We can build the analytics system using analytics tool like R with the support of database.

3. Proposed Analytics System:

Work Flow of proposed analytics system

The proposed analytics system involves the following steps:

Collection of Data: The data about different railway accidents occurred in the span of last 1-2 years is collected from various sources like railways reports etc. The data collected will have attributes like accident type, year of accident, location, number of people injured or died etc.
Chart to Histograms or else dot curves. The following are the user roles in our project,

- Request for the Data
- View the Charts

**Admin Module:** The Admin modules are designed to help in managing the server. The admin module in the proposed system is used to do the following tasks,

- Maintain Server to fulfill all the requirements
- Maintain the updating of the data

**Server Module:** The job of server is to take requests from client and respond accordingly. Security is the key concern for running servers on a dedicated computer. The server module in our project maintains the following data, such as

- Database of the datasets inserted by the admin
- Respond to the client on request

5. Design diagrams:

5.1 Visual Representation of functionality:

The business functionality can be represented by using use case diagrams. The identification of use case starts with listing the discrete business functions in software requirements document. The business functions identified can be declared as a potential use case.

![Use case diagram](image)

**Figure 5:** Use case diagram

5.2 Class Structure:

![Class diagram](image)

**Figure 6:** Class diagram

5.3 Interactions between classes:

![Sequence diagram](image)

**Figure 7:** Sequence diagram
5.4 Workflow of System:

![Workflow Diagram](image)

Figure 8: Work flow diagram

6. Introduction to R:

Statistical computing and visual graphics can be best implemented or developed by using R language which is open source. The different numeric statistical and visual graphics techniques supported by R-language are modelling(linear and non-linear), classification and clustering etc. By using R-language, clear and quality visual plots can be produced for any given data.

6.1 The R environment

R is an integrated package of software facilities for the data manipulation, calculation and graphical display.

It includes:
- efficient handling of data and storage function,
- a set of operators for matrix calculations
- a huge set of intermediate tools for analysis of data
- visual graph representation facilitating data analysis and a robust programming language which includes Decision making constructs, loops, user-defined functions, recursion facility and I/O facilities.
- can be dynamically linked.

7. Sample data set of railway accidents in India:

<table>
<thead>
<tr>
<th>S.NO</th>
<th>Accident type</th>
<th>Accident date</th>
<th>Accident area</th>
<th>Killed people</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Derailed</td>
<td>1950</td>
<td>Bihar</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>bridge</td>
<td>1950</td>
<td>Bihar</td>
<td>81</td>
</tr>
<tr>
<td>3</td>
<td>collision</td>
<td>1960</td>
<td>Bihar</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Collision</td>
<td>1962</td>
<td>Bihar</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Bomb</td>
<td>1966</td>
<td>Bihar</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>Bomb</td>
<td>1986</td>
<td>Bihar</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>Collision</td>
<td>1979</td>
<td>Bihar</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>river</td>
<td>1981</td>
<td>Bihar</td>
<td>800</td>
</tr>
<tr>
<td>9</td>
<td>Collision</td>
<td>1986</td>
<td>Bihar</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>Fire</td>
<td>2013</td>
<td>Bihar</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>Collision</td>
<td>1990</td>
<td>Bihar</td>
<td>60</td>
</tr>
<tr>
<td>12</td>
<td>Bump</td>
<td>2014</td>
<td>Bihar</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>Signal</td>
<td>2013</td>
<td>Bihar</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>fire</td>
<td>1990</td>
<td>Bihar</td>
<td>70</td>
</tr>
</tbody>
</table>

The above data set is visualized as shown below by using tools like weka or R-tool.

![Analysis using WEKA](image)

Figure 9: Analysis using WEKA

Using Weka tool, The above given data set is analysed based on attribute accident type:

![clustering Analysis using WEKA](image)

Figure 10: clustering Analysis using WEKA
8. Motion Charts in R-Language:

Motion Chart which is possible in R-platform by using Google Vis brings your data to life. The basic requirements for generating motion charts are minimum four variables. One of the four variables represents the variable we want to plot, one related to time variable and remaining at least two numerical variables.

8.1 Commands in R to build analytics by using dataset given in Table 1:

- install.packages()
- library('googleVis')
- data1 <- read.csv(file.choose(), header=T)
- visualization1 <- gvisMotionChart(data1, idvar='Train.per.million', timevar='Year')
- plot(visualization1)

8.2 Motion Charts Generated using R-packages:

Motion charts are useful in visualization of data where data variables can be mapped on to two dimensional coordinate axis which helps us in understanding of large amounts of data.

9. Conclusion

With data set taken, we can draw some conclusions like which type of railway accident happened more times and at which locations, reason for accident etc. Hence we are proposing an analytical system based on R-language which can be used for monitoring railway disasters. The railway disasters can be analyzed region wise across country and can identify areas with less, more railway accidents prone areas based on motion charts generated by considering past history of accidents. This conclusions drawn from analytics system can be given to railway people so as they can take measures to avoid railway disasters saving human life and property losses.

10. Future Scope

Extensive Investigation to be done to find other incident types which are not included in this project. The missing data in this project can also be introduced into the database in future so that the results will be more accurate and the precautions can be taken more precisely. Till now, with this analytics system, we are able to collect data and identify different attributes set so as to draw necessary conclusions.

References: