International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

Rail Accidents Analyzing By Text Mining

KADARI VARALAXMI **ENUMULA.KIRAN**

PG Scholar, Department of CSE, Vaagdevi College of Engineering, Autonomous, Bollikunta, Warangal Telangana, Mail.varalaxmikadari@gmail.com

Associate Professor Department of CSE,Vaagdevi College of Engineering, Autonomous Bollikunta, Warangal Telangana, Mail id: kiran.enu@gmail.com

Abstract

Death costs and injuries caused by rail accidents have a significant impact on society. As part of our effort to understand the characteristics of past rail accidents, this article presents an analysis of major rail accidents worldwide from 2000 to 2015. The theory of gross sets and related rules are applied to analyze data . collected. The results show that, although most derived rules are unique, some rules deserve to be highlighted. Collision accidents generally result in lower accidents than derailment accidents, and the most common cause of accidents is human error. In addition, most train accidents occur during the summer. These findings can provide train leaders with lessons and resolutions from past accidents, facilitating the creation of a safer rail operating environment around the world. Accident investigation and analysis are essential to strengthen and improve rail safety. Many railway accidents have been caused by degraded human performance and human error, and the tasks of train drivers and markers have remained essentially the same. Although new technologies and equipment have progressively reduced railway operational accidents, no research has been conducted to determine whether the factors influencing rail performance (PRW) attributed to deteriorating human performance have changed. or have remained constant. The results show that the predictive accuracy of accident costs is significantly improved by using features found in text mining and that predictive accuracy is further improved by the use of modern sets. It is important to note that this study also shows through case examples how the results of textual exploration of narratives can improve the understanding of railway accident contributors in an impossible way through a fixed field analysis of accident reports.

1. INTRODUCTION

For a long time, the Ministry of Railways in the world is a relative monopoly organization. When an accident occurs, the Ministry of Railways forms a research team to investigate; the team would finally publish the crash investigation report. But the report of the investigation could only be seen by the railway staff, basically there was no way to know the outside world. After the reform of the Ministry of Railways in 2013, this situation has changed. During this year, the world railway system publicly published annual statistics on deaths by rail traffic, 1336 people lost their lives due to rail accidents. By 2014, this figure reached 1232. Overall, the number of deaths declined in 2014, but the prospects for rail safety in the world remain bleak. The rail system is a major component of the economy of most countries. passengers, as well as millions of dollars in goods from origin to destination (1). Therefore, the relevant operational, regulatory and governmental bodies in each country with a railway network aim at a safe, reliable and high-quality railway system (2). In the UK in particular, railroads have played an important role in the everyday life and economy of society since the 1820s. The annual safety report (3) recently published by Rail Safety and Standards Board Ltd.. (RSSB) reports on its crucial role: for 2013 to 2014, approximately 1.59 billion passengers, 60.1 billion passenger-kilometers and 48.5 million passenger-kilometers This article describes a survey to understand possible predictors or contributors to accidents resulting from the "exploitation" of narrative text in railway accident reports. To do this, the approach integrates a combination of analytical methods to first identify the relevant accidents and then look for relationships in structured and unstructured data that may suggest contributors to accidents. This available study evaluates the effectiveness of the characteristics found in the mining of texts when using models that contain these characteristics to predict the costs of extreme accidents. In conducting this assessment, the study also considers the usefulness of modern comprehensive approaches that integrate these features of text to predict accident costs. Finally, the study leaves aside the characteristics of text mining, whose importance is confirmed by predictive accuracy, for its understanding of taxpayers to rail accidents. The purpose of this final analysis is to understand railway safety information that text mining can provide, excluding fixed field ratios. These

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

studies have shown interesting results, however, they are not able to adequately analyze the cognitive aspects of the causes of accidents. They often choose to omit important qualitative and textual information from datasets because it is difficult to create meaningful observations. The consequence of textual ignorance results in a limited analysis leading to less substantial conclusions. Text mining methods attempt to fill this void. Text Mining is the discovery of new unknown information, which is automatically extracted from different written resources (text). Text mining methods can extract important concepts and emerging themes from the collection of text fonts. Used in a practical situation, the possibilities of discovering knowledge through the use of text mining are immense. As far as we know, there are few or no reliable studies that have used text mining in this data domain, however, previous field studies indicate a real need for textual exploration to better understand relationships. contextual data.

2. Literature review

Rail accidents cause many casualties each year, attracting the interest of many researchers and analysts. Existing research on rail accidents is empirically and methodologically diverse. From an empirical point of view, most studies have attempted to identify the risk factors that influence the severity of train accidents. Many of these studies report that train accidents often result from a chain or sequence of events rather than a single cause.3,5,6 For example, Reinach and Viale7 analyzed six train accidents and developed a model for analyzing and classifying human errors with 36 likely contributing factors. The results demonstrate that each accident was associated with multiple contributing factors. Related studies in other countries have reported similar results. Although these models can be used to assign responsibility for accidents, they are generally ineffective in preventing future accidents. In addition, many studies have focused on the factors influencing the severity of accidents. Ilkjær and Lind8 analyzed a railway accident in Denmark in 1994 and concluded that the inside of the wagon (the seating arrangement, whether there is an armrest on the seat, the trunk or not, etc.) has a influence injuries. Mirabadi and Sharifian3 investigated accident data from the Iranian Railways from 1996 to 2005 and concluded that human errors, rail cars and tracks most often contribute to an increase in the

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

severity of accidents. Evans9 analyzed fatal rail accidents on major European railway lines from 1980 to 2009. The results show that the most common immediate causes of serious accidents at level crossings are errors or violations by road users. of iron. Other studies have focused on the development of computer-controlled systems to prevent rail accidents.1,10,11 From a methodological point of view, researchers have applied various statistical approaches to analyze train accident rates. and trends in various countries. Evans9 estimated that the overall trend in the number of fatal collisions and derailments per train-kilometer was 6.3% per year between 1990 and 2009 in Europe, with a 95% confidence interval. However, there are statistically significant differences between the different European countries in terms of the rate of fatal rail accidents. Evans10 analyzed data from almost all fatal rail accidents in the United Kingdom from 1967 to 2003, and observed a downward trend in all major categories of train-driven accidents over the 27 years prior to 1993. , non-parametric statistical methods are widely used. in the analysis of train accidents. Chong et al.12 applied artificial neural networks and decision trees to model the severity of injuries resulting from train accidents. In all cases, decision trees outperformed neural networks. Yan et al.13 applied a hierarchical tree regression to explore the prediction and analysis of train-vehicle collisions at passive highway crossings and concluded that the installation of stop signs can effectively improve safety at these crossings. at the level. Liu et al.14 developed two regression models that provide a better understanding of the severity distribution of the train derailment. Knowledge discovery and data mining techniques have also been applied in many recent explorations of this topic. Wong and Chung15 explored accidental presence in Taiwan using raw sets theory, a technique that was used to make decisions in the presence of uncertainties and imprecision applied rule techniques to analyze crash data on accidents. the Iranian railways. the data. Models extracted through these methods can be used to develop regulations and rules to prevent the occurrence of similar accidents in the future. 2.1 Current limitations The current design The analysis of this study is limited to the extraction of the causal text relating to accidents "Groundings", "Collisions", "Machinery breakdowns" and "Fire". The scope of the study was also limited by focusing only on classification of models and connection methods to extract causal relationships in order to keep the

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

study at a reasonable size. Investigation reports on an accident pose many challenges. Reports are written in natural language without a standard template. Spelling errors and abbreviations are often found. Composite word detection such as "safety culture", "state of mind", etc. is difficult because the order of importance is unknown. The contextual meaning of the words "security" and "culture" differs considerably, but the word "safety culture" has a completely different meaning. Therefore, context and semantics play an important role in text mining. To date, they have not reported large-scale story analysis for information that can inform security policy and design. They focused on recovery, not prediction.

3 PROPOSED SYSTEM

3.1 Generate Accident Report This article incorporates security investigation techniques with mischance reports and content mining information to find citizens for rail mishaps. This area depicts the related work in the field of rail transport and, all the more for the most part, transport wellbeing, and furthermore presents pertinent systems of information mining and content. Accessible on the web: Accident report includes This report incorporates various fields including train or prepare qualities, prepare working staff (eg speed at the season of the mischance, most extreme speed before the mishap and weight) and the fundamental driver of the mishap. This territory has turned out to be progressively critical in view of the huge measure of information accessible in records, news articles, look into papers and mishap reports. Put away in databases Textual databases are semi-organized on the grounds that, notwithstanding free content, they additionally contain organized fields containing titles, creators, dates, and other metadata. The mishap reports utilized as a part of this archive are semi-organized.

IV CONCLUSION

In this paper, show that combining text analysis with ensemble methods can improve model accuracy in predicting the severity of the crash and that text analysis can provide insight into the characteristics of the accident. accident. Modern methods of text analysis make narratives in accident reports almost as accessible for detailed analysis as fixed fields in reports. More important than the illustrated examples, text extraction of stories can provide a much richer amount of information than is possible in fixed fields. Finally, as described

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

in the work here used standard methods to clean up stories. However, stories about train accidents use jargon common to the rail industry, and traditional elimination and the elimination of stop words do not necessarily characterize the words used in this industry. For train safety analysis, text mining could benefit from a careful look at ways to extract text characteristics that benefit from the linguistic characteristics of the rail transport sector. A basic research need is to characterize the variation and uncertainty inherent in text mining techniques In this study, the use of LDA and PLS did not yield consistent results with different training and test selections. These differences need to be formally characterized and, ideally, described with a probabilistic model that further improves the understanding of accident contributors. Finally, as described in Section V, the work here used standard methods to clean up stories. However, stories about train accidents use jargon common to the rail industry, and traditional elimination and the elimination of stop words do not necessarily characterize the words used in this industry. For train safety analysis, text mining could benefit from careful consideration of how to extract features from a text that takes advantage of the language characteristics of the rail industry. The references

[1] "Rail Safety Statistics - Annual Report 2009 - Final", Federal Railroad Admin., Washington, DC, United States, April 2011. [Online]. Available: http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Publications.aspx

[2] "Bureau of Safety Analysis", Federal Railroad Administration, Washington, DC, United States, October 2009. [Online]. Available: http://safetydata.fra.dot.gov/officeofsafety/

[3] G. Cirovic and D. Pamucar, "Decision Support Model for the Prioritization of Railway Grade Crossings to Improve Safety: Application of the Adaptive Neuro-Fuzzy System", System Expert. Appl., Vol. 40, pp. 2208-2223, 2013. [4] L.-S. Tey, G. Wallis, S. Cloete and L. Ferreira, "Modeling Driver Behavior for Novel Warning Devices at Crossings", Neural Comput. Appl., Vol. 51, pp. 104-111, March 2013.

[5] D. Akin and B. Akbas, "A Neural Network Model (NN) for Predicting Intersection Crashes Based on Surface Characteristics of the Driver, Vehicle and Roadway", Sci. Res. Essays, vol. 5, pp. 2837-2847, 2010.

[6] H. Gonzalez, J. Han, Y. Ouyang and S. Seith, "Multidimensional exploration of traffic anomalies on large-scale road networks", Transp. Res. Rec., Vol. 2215, pp. 75-84, 2011.

[7] E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, "Real-time Detection of Twitter Stream Analysis Traffic," IEEE Trans. Intell. Transp. Syst., Vol. 16, no. 4, pp. 2269-2283, March 2015.

[8] F. Oliveira-Neto, L. Han and M. K. Jeong, "An online self-learning algorithm for register plate matching", IEEE Trans. Intell. Transp. Syst., Vol. 14, no. 4, pp. 1806-1816, dec. 2013. Available online: http://internationaljournalofresearch.org 2097 International Research Review p-ISSN: 2 3 4 8 - 6848 e -IS SN: 2 3 4 8 - 795X Available at https://edupediapublications.org/journals Vo lu me 0 4 Is su e 0 2 February 2017

[9] J. Cao et al., "Web traffic sentiment analysis: methods and applications", IEEE Trans. Intell. Transp. Syst., Vol. 15, no. 2, pp. 844-853, Apr 2014

[10] J. Burgoon et al., "Detecting intention concealment in transportation screening: a proof of concept," IEEE Trans. Intell. Transp. Syst., Vol. 10, no. 1, pp. 103-112, March 2009.

[11] Y. Zhao, T. H. Xu, and W. Hai-feng, "Mining-Based Fault Diagnosis of Vehicle On-Board Equipment for High-Speed Railroad", in Proc. IEEE 17th Int. Conf. CSTI, Oct. 2014, pp. 900-905. [12] T. Hofmann, "Probabilistic latent s emantic indexing," in Proc. 22nd Annu.Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 1999, pp. 50–57.

[13] R. Nayak, N. Piyatrapoomi, J. W. R. Nayak, N. Piyatrapoomi, and J. Weligamage, "Application of text mining in analysing road crashes for road asset management," in Proc. 4th World Congr. Eng. Asset Manage., Athens, Greece, Sep. 2009, pp. 49–58.

[14] "Leximancer Pty Ltd." [Online]. Available: http://info.leximancer.com/ academic

[15] A. E. Smith and M. S. Humphreys, "Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping," Behav. Res. Methods, vol. 38, no. 2, pp. 262–279, 2006.

[16] U.S. Grant, The Personal Memoirs of U.S. Grant., 1885. [Online]. Available: http://www.gutenberg.org/files/4367/4367-pdf/4367- pdf.pdf

[17] W. Jin, R. K. Srihari, H. H. Ho, and X. Wu, "Improving knowledge discovery in document collections through combining text retrieval and link analysis techniques," in Proc. 7th IEEE Int. Conf. Data Mining, Omaha, NE, USA, Oct. 2007, pp. 193–202.

[18] D. Delen et al., Practical Text Mining and Statistical Analysis for Non-StructuredText DataApplications. Waltham,MA, USA:Academic, 2012.

[19] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees. Belmont, CA, USA: Wadsworth, 1984.

[20] T. Hastie, R. Tibshirani, and J. H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: SpringerVerlag, 2009.

[21] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.

[22] H.Wold, "Estimation of principal components and related models by iterative least squares," in Multivariate Anal., P.Krishnaiaah, Ed. NewYork, NY, USA: Academic, 1966, pp. 391–420.

[23] L. Li, R. D. Cook, and C. Tsai, "Partial inverse regression," Biometrika, vol. 94, no. 3, pp. 615–625, Aug. 2007.

[24] M. Taddy, "Multinomial inverse regression for text analysis," J. Amer. Statist. Assoc., vol. 108, no. 503, 2012. [Online]. Available: http://dx.doi. org/10.1080/01621459.2012.734168

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

[25] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J.Mach. Learn. Res., vol. 3, pp. 993– 1022, Mar. 2003.

[26] M. Steyvers and T. Griffiths, "Probabilistic topic models," in Handbook of Latent Semantic Analysis, vol. 427. Hillsdale, NJ, USA: Erlbaum, 2007.

[27] D. Blei, L. Carin, and D. Dunson, "Probabilistic Topic Models," IEEE Signal Process. Mag., vol. 27, no. 6, pp. 55–65, Nov. 2010.

AUTHOR'SPROFILE:

KADARI VARALAXMI,PG Scholar, Department of CSE, Vaagdevi College of Engineering, Autonomous, Bollikunta, Warangal Telangana, Mail.varalaxmikadari@gmail.com

ENUMULA.KIRAN,Associate Professor Department of CSE,Vaagdevi College of Engineering, Autonomous

Bollikunta, Warangal Telangana.

Mail id: kiran.enu@gmail.com