WORKING PAPER

ITLS-WP-23-13

RecoMap - a semi-automated tool for analysing railway accident recommendations across jurisdictions and over time

By
Wei-Ting Hong[a], Geoffrey Clifton[a] and John D Nelson[a]

[a] Institute of Transport and Logistics Studies (ITLS), The University of Sydney, Australia

July 2023

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-23-13

TITLE: **RecoMap - a semi-automated tool for analysing railway accident recommendations across jurisdictions and over time**

ABSTRACT: To maintain a safer railway operational environment, recommendations are proposed by independent investigators after accidents. Despite a considerable number of (sometimes similar) recommendations made across jurisdictions, practitioners suffer from a lack of synthesised recommendations made across jurisdictions and time due to the high complexity of analysing textual data. To fulfil the gap, an auto mated tool for the analysis of accident report recommendations is developed, allowing the railway industry to learn from other countries. The Structural Topic Model (STM) is used to extract critical insights from recommendations to depict how independent railway accident investigators mitigate risks observed. Empirical data is retrieved from official railway accident reports published by Rail Accident Investigation Branch (RAIB), Australian Transport Safety Bureau (ATSB), National Transportation Safety Board (NTSB) and Transportation Safety Board of Canada (TSB). The resulting *RecoMap* is developed as a framework to help practitioners learn across jurisdictions and time. The study also identifies a transition from making interfering recommendations addressing operational issues to making supportive recommendations addressing organisational issues in the railway industry across countries. Additionally, the concept of triple-loop learning is insufficient in the railway industry of the investigated jurisdictions, implying that current practices might result in railway accidents that could have been prevented by learning from other jurisdictions and implementing corresponding mitigation measures in advance.

KEY WORDS: ***Railway safety, accident investigation, recommendation analysis, Structural Topic Model, learning behaviours***

AUTHORS: **Hong, Clifton, Nelson**

ACKNOWLEDGEMENTS: N/A

CONTACT: INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H04)

The Australian Key Centre in Transport and Logistics Management
The University of Sydney NSW 2006 Australia
Telephone: +612 9114 1813
E-mail: business.itlsinfo@sydney.edu.au
Internet: http://sydney.edu.au/business/itls

DATE: July 2023

1 Introduction

Railway accidents significantly disrupt the transportation network and cause catastrophic impacts on society, such as fatalities, injuries and economic loss. Each railway accident should be fully investigated to understand immediate causes and underlying factors and address hazards identified. As a critical part of the investigation, recommendations and remedial actions can be considered as crucial components to ensure that lessons are learnt and future strategies for addressing risks are implemented to increase railway safety. Several studies have discussed the challenge of making recommendations and ensuring their further implementation (Akel et al., 2022; Cedergren, 2013; Lundberg et al., 2012), implying the importance of the role that recommendations play in the progress of advancing railway safety.

Despite the consensus that railway safety can be improved through analysing recommendations from accident reports, investigating this issue becomes more complicated and time-consuming due to the difficulty of handling a huge volume of textual data, whilst the body of railway accident reports and recommendations increases over time. Nevertheless, most works in this context concentrate on single jurisdictions and discuss the inter-organisational challenges (Cedergren, 2013) or the learning behaviour while implementing recommendations (Stemn et al., 2018). Limited attention is given in the literature to comparing recommendations made by individual railway accident investigation bodies across jurisdictions, hindering practitioners from advancing railway safety by learning from other countries' experience. Although railway systems operated in different jurisdictions differ from one another in terms of the infrastructure design, signal systems and management approaches, valuable insights can still be retrieved and applied to reinforce the awareness of railway safety through learning from accidents in other jurisdictions. Additionally, recommendations that result from each investigation are extremely precious given that considerable effort is devoted to mitigating risks identified. Similar accidents can be prevented from occurring once these recommendations proposed in other jurisdictions are learnt and implemented in advance elsewhere.

To reduce the difficulty of analysing textual data, a considerable number of works put emphasis on leveraging the power of Natural Language Processing (NLP) and machine learning to streamline the analysis of crowdsourced textual data. NLP is a technology developed for processing human languages and addressing the interface between text and programming (Collobert et al., 2011; Luo et al., 2020; Syeda et al., 2017). Machine learning is the technique for identifying the relationship and patterns within the data and providing potential trends and features to the NLP model for understanding human languages. Several empirical works have proven that NLP can help to reduce manual effort and increase the accuracy of developed models (Dong et al., 2022; Kume &Kozaki, 2021; Single et al., 2020). Despite a broad discussion about the application of NLP in practice, researchers suffer from the absence of a consistent and systematic framework for interpreting the result. Nevertheless, some studies adopt (semi-)supervised learning approaches, demanding a large amount of annotated data that requires heavy human effort whilst training the NLP model (Augenstein et al., 2016; Yang et al., 2021). These obstacles can create a significant barrier for practitioners and researchers to utilising these state-of-the-art technologies effectively.

Thus, this paper aims to overcome the difficulties mentioned above by developing a consistent analysis framework based on the NLP and unsupervised-based machine learning approaches. The proposed model is implemented to analyse railway accident report recommendations. *RecoMap* is a practice-oriented and data-driven model for helping the railway industry learn how recommendations are made across jurisdictions and time. The *RecoMap* also reveals the potential transition in the style of making recommendations in each jurisdiction, which is beneficial for understanding the learning behaviours in the railway industry.

This paper is organised as follows. Firstly, the existing literature body related to recommendations analysis is briefly reviewed (Section 2). Secondly, the framework of semi-automated analysis of textual data is elaborated (Section 3). Next, details related to the development of *RecoMap* based on the application to the railway industry are provided, including information on processing data and interpretation (Section 4). Additional findings from the *RecoMap* related to the transition in the style of making recommendations in the railway industry are highlighted (Section 5). Lastly, several conclusions, limitations and suggestions for future work are proposed (Section 6).

2 Literature context

The relatively limited literature in the field of railway accident recommendation analysis primarily treats recommendations as a proxy of the behaviour of investigators. It concentrates on the role they play in a railway accident. For instance, it has been revealed that investigators can have difficulty determining the scope of recommendations and allocating responsibility for the tasks derived from recommendations (Cedergren, 2013). Additionally, previous studies also find that the potential underlying factors making similar railway accidents occur include the lack of organisational learning and understanding even though recommendations are proposed and implemented immediately after the railway accident (Drupsteen & Hasle, 2014; Wrigstad et al., 2014). Effective organisational learning has been proven to significantly improve the safety of work environment by learning from historical accidents (Fahlbruch &Schöbel, 2011) and adapting lessons learnt (such as recommendations in railway accident reports) for further implementation and operations (Choularton, 2001). However, the gap and delay between the recommendations made by investigators or academic researchers and implementation by the industry is also revealed (Brath, 2020; Underwood &Waterson, 2013). This implies a need to improve the connection between the recommendations proposed and learning behaviour in the railway industry.

Several factors might hinder the railway industry from organisational learning, such as technological barriers, the lack of a practical learning framework and legislative restrictions. Some prior works have discussed potential barriers to learning in the railway industry. For example, Blackwood &Renaud (2022) identify that the railway industry might suffer from barriers to learning due to internal concerns such as the lack of evidence, time constraints and limited cost-benefit analysis. Elliott et al. (2000) reveal significant barriers to organisational learning and argue that the railway industry repeats mistakes and fails to learn from the aftermath of accidents due to the lack of understanding of the role of organisational learning. This argument is also supported by later studies although cross-country analysis is still absent in the literature (Gray, 2008; Johnsen et al., 2006; Nolan-McSweeney et al., 2022). Lack of organisational learning might result in the failure to learn from accidents and repeat mistakes.

On the other hand, several definitions exist in the literature for expressing organisational learning more generically. One commonly used definition is the concept proposed by Georges &vanWitteloostuijn (1999), dividing the learning level into single-loop, double-loop, and triple-loop. Single-loop learning refers to taking corrective action after the mistake is identified, whereas double-loop learning includes considering underlying factors at the organisational level and modifying norms, objectives and policies (Georges &vanWitteloostuijn, 1999; Størseth &Tinmannsvik, 2012). Finally, triple-loop learning manifests structural opportunities, promotes people’s participation in making well-informed decisions, and extends the deepness and fullness of diverse issues (McClory et al., 2017; Wang &Ahmed, 2003).

Many studies have shown the transition from single-loop to double-loop learning and the benefit to railway safety in the railway industry (Pilbeam et al., 2016; Rosness, 2013; Steiro et al., 2004). In recent years, there has been growing attention to the shift to triple-loop learning and the potential importance for railway safety. For instance, the lack of incentives to report experience at the management level has postponed the triple-loop learning in the railway industry (Rydstedt Nyman, 2019). Additional evidence also suggests that triple-loop learning can improve railway safety by achieving dynamic maintenance regulation (Granström et al., 2022). However, most studies emphasise the potential benefits of triple-loop learning or the present performance in learning processes. The understanding of how the triple loop learning is driven in the railway industry remains unclear, hindering practitioners from adopting the learning concept at the organisational level. Given that recommendations made by independent railway accident investigators play a critical role in leading the improvement of railway safety (Cedergren, 2013; Watson, 2004), more evidence related to the recommendations should be collected and analysed to understand the learning culture and promotion in the railway industry.

3 Semi-automated analysis of textual data

The topic modelling approach can be used to extract critical insights from recommendations made in the rail industry and to understand the focus each jurisdiction puts on specific recommendations. Topic modelling is a practical application in information retrieval and Natural Language Processing (NLP) to categorise text into domain topics and rank documents against topics (Dornick et al., 2021; Roque et al., 2019; Yang and Anwar, 2016). A topic model reveals the relationship between topics and documents by different features, such as the probability of occurrence of words and high dimensional word embeddings. The model assumes that a document contains a collection of underlying themes, and the distribution of words in the document over the whole corpus might derive topics representing these underlying themes. A set of keywords is identified to reflect underlying topics and their trend and contribute to informative statistics for further methodological and practical applications (Blei and Mcauliffe, 2007).

Several NLP models have been developed for the topic modelling task (Angelov, 2020; Grootendorst, 2022; Han and Eisenstein, 2019; Lata et al., 2022; Ly et al., 2020). These approaches can be roughly divided into word-embedding-based methods and bag-of-words methods based on the mechanism of understanding the natural language. The word-embedding-based approaches, such as the BERTopic model, identify the meaning of one word by considering the words in the same document. The higher dimensionality of the word-embedding-based approaches allows the model to store the characteristics of each word from different dimensions. On the other hand, the bag-of-words-based approaches, such as the Structural Topic Model (STM), use the dimensionality equal to the volume of vocabulary used in the data. Each word is treated individually and uniquely regardless of words with similar meanings. For instance, “rail” and “track” are identified as two independent words in the STM even though they share a part of the concept that supports wheels to roll upon.

The recommendations in the railway industry accident reports show a strong semantic homogeneity of descriptions in terms of words used. For example, consider the following recommendations from the RAIB recommendations dataset:

- *It* ***is expected*** *that* ***Network Rail*** *will take account of principles identified by recent research when modifying* ***crossings****.* (RAIB, 2017)
- ***Network Rail should*** *review the design of long hoods that can be fitted at* ***level crossings*** *and implement any necessary changes identified to make them more effective.* (RAIB, 2009)
- *When addressing risks identified by the implementation of the revised process,* ***Network Rail should*** *prioritise the implementation of required mitigation measures to* ***level crossings*** *where consequences of operator error are severe and not protected by engineered safeguards.* (RAIB, 2014)

The recommendations above are assigned to the same topic by the word-embedding-based method because the semantic meaning of Network Rail's obligation on level crossing risks is detected by words in bold, including "expected", "should", "Network Rail", and "level crossings". However, the topic of interest in this study is "how" recommendations address the risk. Keywords with underlines including "review", "implementation" and "principles" should be identified and assigned to topics. In this case, the bag-of-words-based approach is more applicable because the occurrence of words is more meaningful than the semantic context information.

There are several bag-of-words-based approaches popularly used in empirical analysis, such as the Latent Dirichlet Allocation (LDA), the Correlated Topic Model (CTM), and the Structural Topic Model (STM). In recent years, the STM has shown many advantages over other approaches, such as allowing co-variance analysis and temporal analysis and high flexibility on small and sparse documents (Bai et al., 2021; Kuhn, 2018; Kwayu et al., 2021). Thus, the STM is utilised for analysing recommendations in the railway industry in this study.

3.1 Structural Topic Model (STM)

The STM is an unsupervised learning-based probabilistic topic modelling method derived from the Latent Dirichlet Allocation (LDA). The LDA is a generative statistical model that classifies documents based on the observations of each individual word collected in the documents and assumes that the topic of each document is derived from the aggregation of the words in that document. Suppose a word is a basic item from a set of vocabulary indexed by $\{1, 2,\ldots, V\}$, a document (w) is a sequence of N words noted by $w = (w_1, w_2, \ldots, w_N)$, and a corpus is a set of M documents noted by $D = \{w_1, w_2, \ldots, w_M\}$. Assume documents ($D$) are created by a random combination of latent topics, characterised by a specific distribution over words (N) and follow a generative probabilistic model (Blei et al., 2003). The generation of each document (d_i) in a corpus D follows the consecutive theorems:

1. The number of words N is chosen by a Poisson(ζ) distribution.
2. A random parameter θ drawn from a Dirichlet (α) distribution is chosen to represent the proportions of topics in one document.
3. For each word w_n in N words within one document, a random topic z_n is assigned to w_n drawn from a Multinominal(θ) distribution.
4. For each topic z_n, proportions of each word are drawn from another Multinomial distribution $p(w_n|z_n, \beta)$, where β is a parameter representing the proportions of words in one topic.

The basic LDA model is commonly applied and virtualised for an alternative approach to be explained by

Blei (2012) and shown in Figure 1. Assuming that the dimensionality of the Dirichlet distribution is a fixed and known value k representing the number of topics, the β can be parameterised as a $k \times V$ matrix for mapping the probabilities of words based on the bag-of-words approach. Blei (2012) also notes that N is an independent variable, and its randomness is ignored during the development of the LDA model. Thus, the probability density of the proportions of topics in one document retrieved from the Dirichlet (α) distribution can be illustrated as:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \times \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1} \qquad \text{Equation 1,}$$

where $\Gamma(\mathrm{x})$ is the Gamma function and α_i is a k-vector mapping the distribution of topics. $p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \times \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$ Equation 1 reflects the nature of two plates in Figure 1, representing documents (M) and the recurring choice of words and topics in one document. The outer plate represents the association between all documents (M) and the random parameter θ drawn from a Dirichlet (α) distribution; whereas the inner plate illustrates the association between all words (N) in one document and random topics z_n drawn from a Multinominal (θ) distribution. Therefore, the link between topics and words appearing in each document is built. On the other hand, another parameter β is estimated to identify the link between the proportions of words in one topic. The joint distribution of θ for a set of words w and topics z can be expressed as:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \times \prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n, \beta) \qquad \text{Equation 2.}$$

In this case, key parameters α and β can be inferred with a Bayesian approach by estimating the posterior distribution of known variables from the original corpus (Kuhn, 2018).

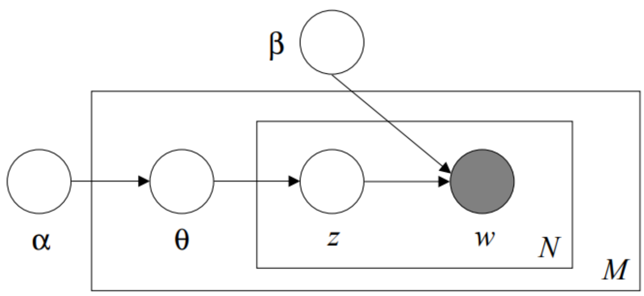

Figure 1, a generic concept of the LDA model illustrated as a plate diagram (Blei et al., 2003)

The LDA has been widely improved and implemented in accordance with the context of interests. For example, Li et al. (2018) advance the structure of the LDA model by training a Word2Vec embedding on the dataset. The journal articles dataset is partitioned into summary, method, and conclusion according to the cosine similarity of embeddings. A weighted topic embedding is created to improve the accuracy of the clustering result. Another example is that Guo et al. (2019) who improve the accuracy of LDA by partitioning the documents into paragraphs and applying weighted summation to obtain the predicted topics. Despite the convenience of retrieving document-level information delivered by the LDA, the assumption that the probability of the occurrence of one word within one document is fixed after the LDA model is developed restricts the flexibility of analysis. For instance, estimated parameters α and β are not allowed to be sensible to temporal factors or other potential covariates (Kuhn, 2018).

The STM is developed on the same statistical basis as the LDA in addition to allowing correlations of external factors among topics. The main difference lies in the pre-generalised linear models derived from the nature of the data used while estimating parameters. In doing so, the parameter θ is not applied to all documents equally drawn from the Dirichlet (α) but from the logistic-normal distribution to estimate the topical prevalence on document-level data. Furthermore, the assumption is that fixed parameter β (distribution of topics over words) should be released by replacing the multinomial distribution with a multinomial logit model for estimation. Mathematically, the parameter ($\beta_{d,k,v}$) for an individual word v in document d within the topic k should be as the following equation for capturing the influence of covariate data (Roberts et al., 2013):

$$\beta_{d,k,v} \propto \exp\left(m_v + \varphi_v^{\cdot,k} + \varphi_v^{y,\cdot} + \varphi_v^{y,k}\right) \qquad \text{Equation 3,}$$

where m_v is the baseline occurrence of word v, the $\varphi_v^{\cdot,k}$ is the effect of topic k, the $\varphi_v^{y,\cdot}$ is the effect of covariate y, and the $\varphi_v^{y,k}$ is the mixed effect among topic k and covariate y. Thus, the plate diagram (Figure 1) can be further extended in Figure 2. The main distinction lies in the prior estimation of parameter θ during topic prevalence analysis and additional consideration of covariate variables in topical content. More mathematical details and theorems can be found in Roberts et al. (2013). The STM is more suitable than the LDA for analysis of railway accident reports because critical covariates are usually disclosed and discussed in reports, such as the occurrence of time and the involved mode of rail transport and organisations. These critical covariates can offer valuable insights for better understanding the nature and prevalence of railway accidents across time. For instance, the STM may reveal how the platform-train interface incidents occur on the light rail system and other modes of rail transport system. The trend of how it happens may also be revealed by supplementing the occurrence of time as an additional covariate in STM temporal analysis.

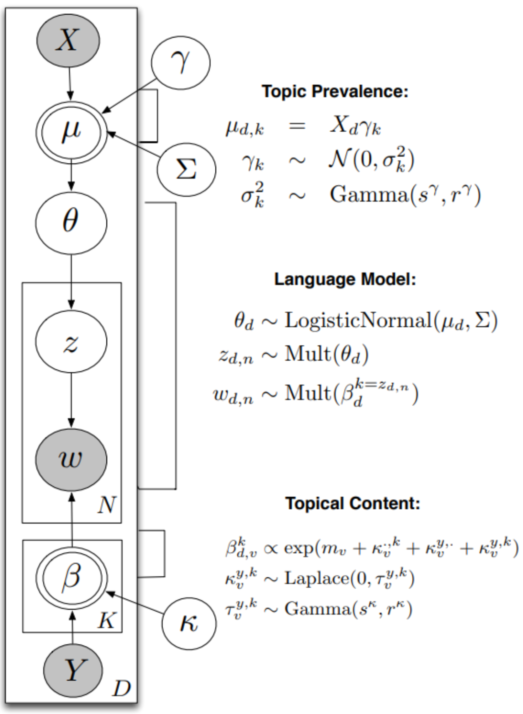

Figure 2, the concept of the Structure Topic Model illustrated as a plate diagram (Roberts et al., 2013)

3.2 Model selection and validation

To ensure the performance of candidate models during training, two metrics are introduced as indicators: Semantic Coherence (SC) and Exclusivity. The SC is a measurement determining the occurrence of individual words and the co-occurrence of the pairs of distinctive words. For instance, we have terms "freight train" and "passenger train" in our dataset with the same word "train", and a distinct topic should be able to detect this and assign these two words to different topics. On the other hand, exclusivity means the extent to which the model is able to assign one critical keyword to one topic with a high level of possibility of appearing and ensure the possibility of appearing is low in other topics. A higher SC usually leads to lower exclusivity and vice versa. A model with a lower number of topics (*k*) would have higher SC because a limited number of topics and words are used for estimation. However, it would lead to lower exclusivity as well because the option of critical keywords is limited. When the *k* increases, the SC will decrease, whereas the Exclusivity increases because more topics are available for assigning distinct keywords. Once *k* is equal to the number of words in the vocabulary (*V*), the exclusivity will become almost infinite, and the result will not offer any valuable insight. To reach a balance between SC and exclusivity by determining a suitable number of topic *k*, $Performance_i = \left(\frac{SC_i - SC_{max}}{SC_{min} - SC_{max}}\right) \times \left(\frac{\text{Exclusivity}_i - \text{Exclusivity}_{max}}{\text{Exclusivity}_{min} - \text{Exclusivity}_{max}}\right)$ Equation 4, covered by both metrics, is designed in this study to estimate the balanced performance of the STM developed.

$$Performance_i = \left(\frac{SC_i - SC_{max}}{SC_{min} - SC_{max}}\right) \times \left(\frac{\text{Exclusivity}_i - \text{Exclusivity}_{max}}{\text{Exclusivity}_{min} - \text{Exclusivity}_{max}}\right) \qquad \text{Equation 4,}$$

where SC_i represents the Semantic Coherence of the *i* model, SC_{max} and SC_{min} is the highest and lowest Semantic Coherence value in all models respectively, Exclusivity_i represents the exclusivity of the *i* model, Exclusivity_{max} and Exclusivity_{min} is the highest and lowest Exclusivity value in all models respectively. This indicator allows us to select the model with the lowest marginal effect on SC and Exclusivity for a range of *k*.

4 The development of RecoMap and application to the railway industry

4.1 The datasets

In case of inconsistency in the analysis and bias resulting from the different language used in recording information in the railway industry, this study only considers recommendations included in railway accident reports published by independent railway accident investigation bodies in countries with the following conditions to secure the model's performance: (1.) having and maintaining a comprehensive documentation system to ensure the consistency of data processing, (2.) being granted independent authority to investigate railway accidents, (3.) making recommendations that the railway industry are required to take actions, (4.) being in the English-speaking countries with reports written in consistent English language regardless of investigation engagement, time and types of accident to reduce the complexity of analysis, and (5.) have published over 100 reports.

Thus, the recommendation section in railway accident reports published by Rail Accident Investigation Branch (RAIB), Australian Transport Safety Bureau (ATSB), National Transportation Safety Board (NTSB) and Transportation Safety Board of Canada (TSB) are selected based on the requirements mentioned above. A variety of time periods is covered by datasets provided by investigators. All railway accident reports are retrievable from the official websites of selected investigators. Note that we have removed scanned documents because of technical difficulties.

Table 1 illustrates the overview of the processed recommendations dataset at the sentence level. The number of recommendations made by the TSB is limited because the TSB only publishes recommendations at the highest level of accidents with severe consequences. Similar circumstances can be found in the ATSB dataset given that only identified risks are highlighted without publishing recommendations directly, offering the railway industry the flexibility to propose strategies for managing risk factors. Another note is that the NTSB provides the independent recommendations dataset ranging from 1966 to 2020 and stored in an editable way. Therefore, all recommendations are retrieved to understand the composition of recommendations across time.

Table 1, the overview of the processed recommendations datasets

	No. of sentences	**Period covered**	**Note**
RAIB	4,807	2005-2019	All reports are linked to corresponding recommendations.
ATSB	1,074	1999-2021	Only a limited number of reports lead to recommendations.
NTSB	3,185	1966-2020	Reports earlier than 1996 are scanned files, but the recommendations dataset is independent, editable and retrievable from 1966 to 2020.
TSB	76	1991-2021	Only a limited number of reports lead to recommendations.

4.2 Data pre-processing

The data pre-processing for the STM consists of the following steps: lowercasing, digital number removing, punctuation removing, and stemming. The R package *textProcessor* is implemented in the STM. The metadata is associated with processed text by the *quanteda* package, converting data into a sentence-term matrix and holding covariates at the sentence-level (Benoit et al., 2018). Other libraries under *quanteda* also provide a wide range of functions, such as reading data in multiple forms. The output can be directly fit into STM functions.

4.3 Selecting the number of topics

For the STM model, the number of topics influences the performance of the model and needs to be estimated carefully. An iterative analysis sets the number of topics from 5 to 50 and each model's performance is recorded using Equation 4. The result suggests that using 26, 21, 12, and 5 topics for the RAIB, ATSB, NTSB and TSB recommendation datasets respectively results in the best performance and balance between semantic coherence and exclusivity.

4.4 An overview of results

Figure 3 to Figure 6 show the extracted keywords of the RAIB, ATSB, NTSB and TSB datasets respectively with the highest occurrence probability and the assigned name of each topic. The interpretation is completed by reviewing keywords and representative sentences from the perspective of recommendations for railway accidents rather than the nature of railway accidents. For instance, although the keywords “cross”, “user”, and “level” in Topic 22 of the RAIB dataset (Figure 3) might refer to the mechanism of level crossing accidents, the label “review of consideration of design and standard for level crossing safety” is assigned to highlight the representation of other keywords and the real meaning of sentences sorted to this topic.

For the RAIB recommendation dataset (Figure 3), several identified topics of recommendations have been widely discussed in the railway accident studies and recommendation analysis in the literature such as removal of the hazard (assessment and measurement), enhancement of design, enhancement of design assurance and approvals, steps to address safety culture (attitudes and behaviours), management process, enhancement of procedures, and training and competency (Braut et al., 2014; Cedergren &Petersen, 2010; Hulme et al., 2019; Tretten &Candell, 2021; Zhan &Zheng, 2016). However, other topics including standardisation of process and operation, cooperation, lesson learnt processes, and documentation are seldom discussed.

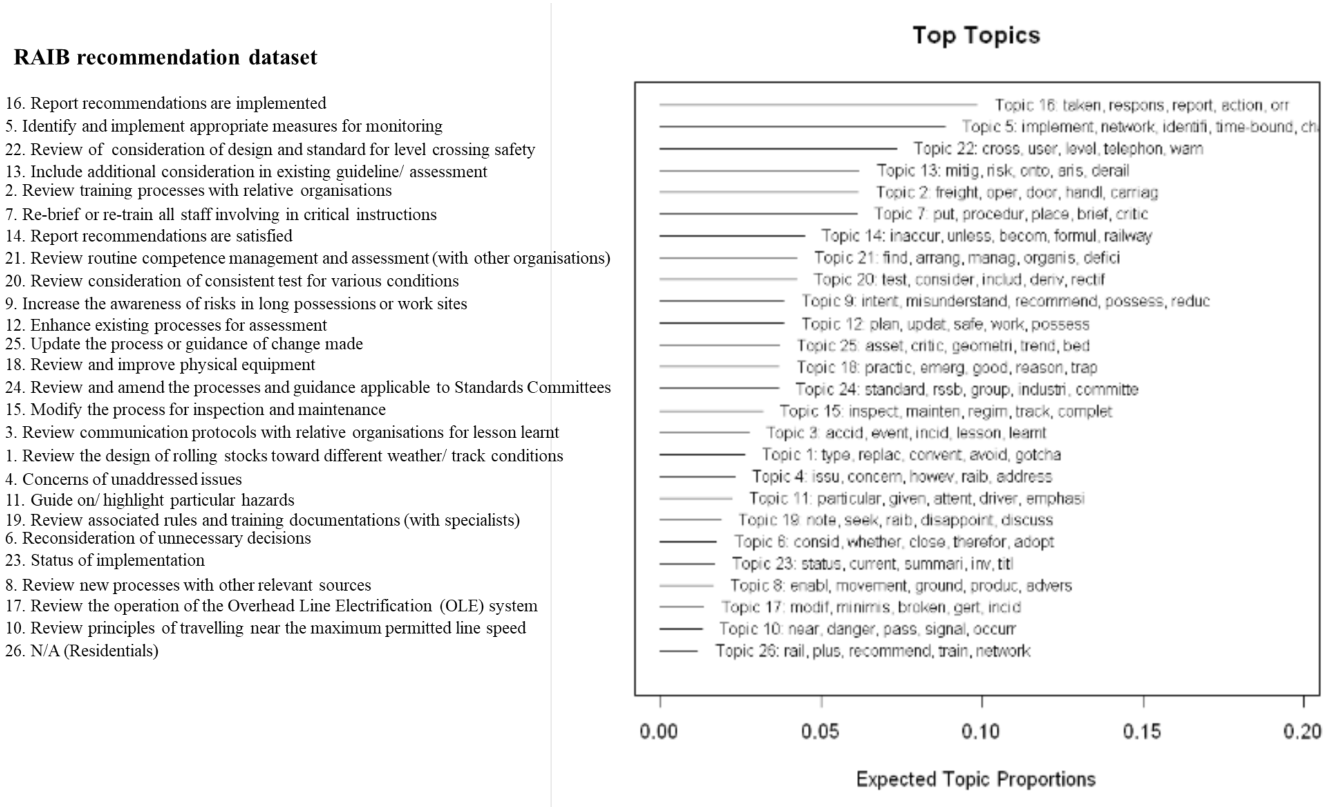

Figure 3, the extracted topics and keywords of the RAIB recommendation dataset from the STM

On the other hand, recommendations frequently proposed by the ATSB are reviewing communication technology, exchanging knowledge with other organisations, and undertaking risk mitigation strategies (topic 9 in Figure 4). A notable finding is that a more significant proportion of the sentences is sorted to the topic “request to take action to address identified safety issue” (topic 4), implying that some parts of the recommendations are made to remind the reader about compliance of existing rules or procedures.

ATSB recommendation dataset

4. Request to take action to address identified safety issues
9. Review communication technologies and cabin design
6. Exchange knowledge with other organisations to reduce level crossing risk
13. Undertake a review for risks mitigation strategies (lesson learnt process)
17. Introduce and update the regulations of freight loading manual
11. Introduce signal protection protocol and review permitted speed
15. Review the effectiveness of testing equipment/ process
5. Develop documented process or specification for safety management system
19. Assess the effectiveness of measurement of monitoring process
16. Develop rules and procedures for governing coordination between organisations
3. Review requirements of track maintenance and monitoring standards
8. Consider appropriate modifications to the assessment and analysis interpretations
18. Assess the effectiveness of maintenance practices and data recording
20. Develop and validate procedures for monitoring management
7. Revise the testing and exanimating process/ maintenance schedules
14. Introduce framework/ exercises for prioritising tasks during emergency
1. Offer suggestions for instructions of maintenance/ operation process
2. Review process/ regulations of lookout working and worksite safety management
10. Advise on monitoring and reporting shift lengths and break between shifts
21. Ensure compliance to existing regulations on Safeworking arrangements
12. Introduce/ modify the track stability management plan

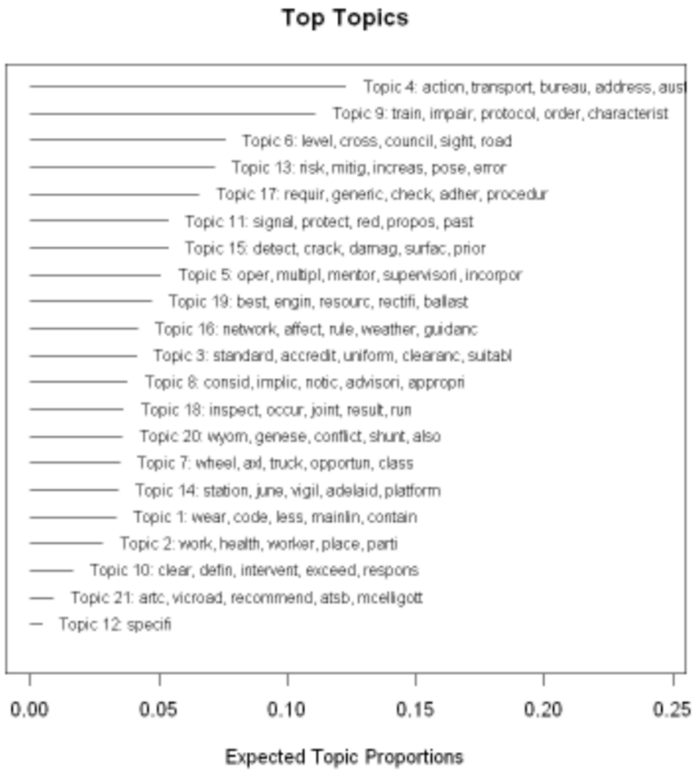

Figure 4, the extracted topics and keywords of the ATSB recommendation dataset from the STM

According to Figure 5, a considerable amount of focus is put by the NTSB on cooperation with organisations within the railway industry, implying less intervention and restrictions on the approach operators apply to address identified hazards. Furthermore, assisting research and programs is also mentioned with high frequency, which might indicate the promotion of cooperating with third parties and producing a comprehensive solution. Another note is that assigning specific methods to address identified hazards is rarely found in the NTSB recommendations. Most recommendations are supportive and offer high flexibility for the railway industry to implement improvements.

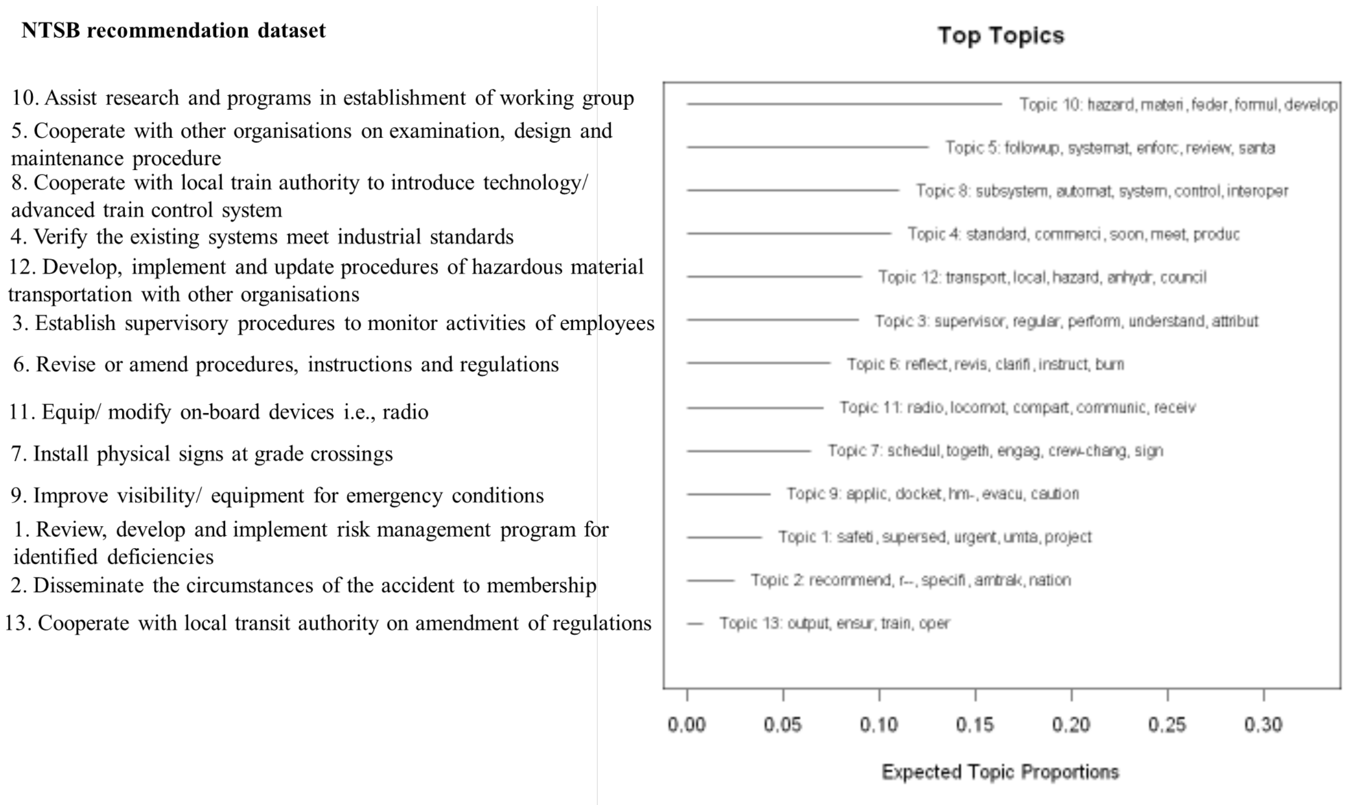

Figure 5, the extracted topics and keywords of the NTSB recommendation dataset from the STM

Figure 6 illustrates the extracted topics and keywords of the TSB recommendation dataset from the STM. The TSB dataset comprises recommendations requesting the examination and reassessment of current procedures rather than developing new rules or processes. On the other hand, limited suggestions are given to cooperate with organisations within the railway industry. Furthermore, most recommendations are directed to individual railway companies, assigning an objective to resolve identified hazards. Lastly, interfering recommendations such as assigning specific instructions to organisations involved are not identified, implying the TSB tends to propose supportive advice primarily and remains a large degree of flexibility for the railway industry.

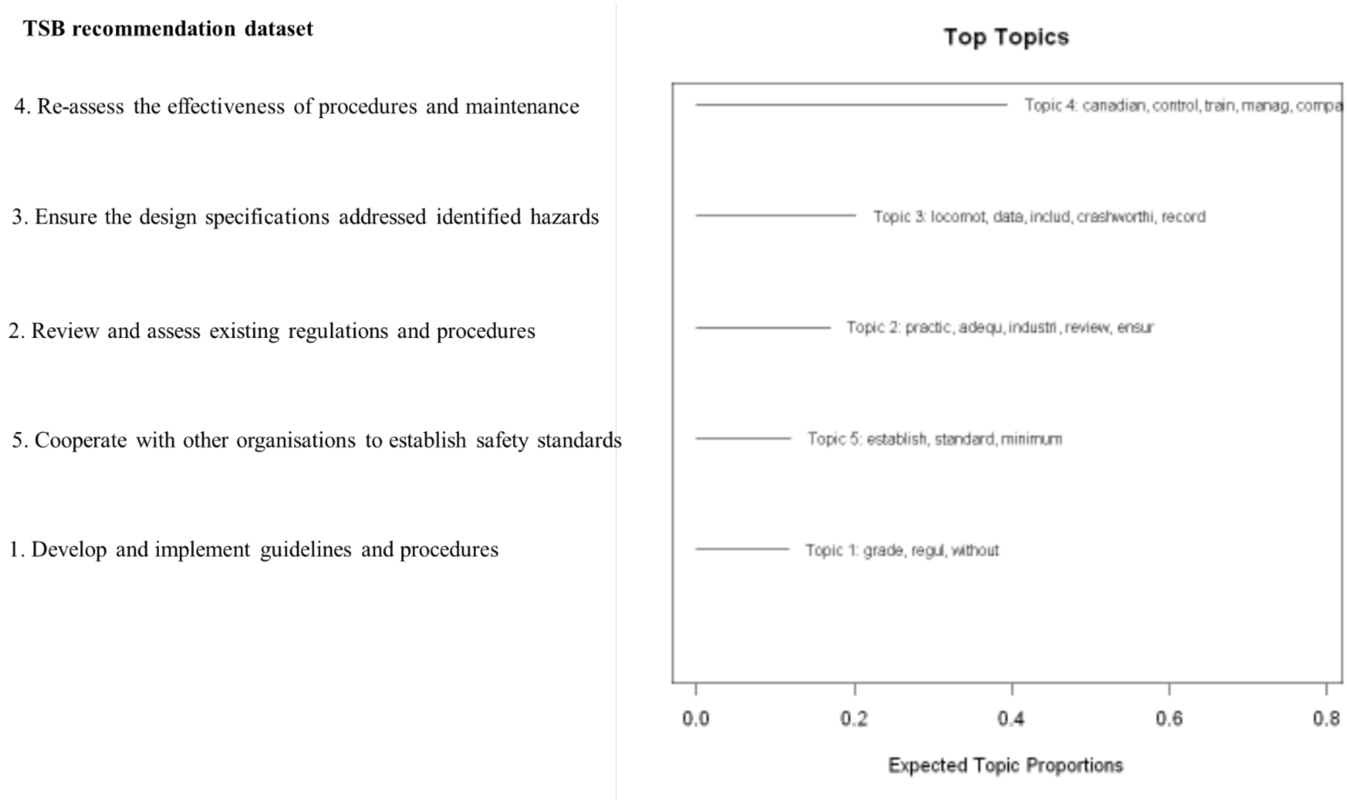

Figure 6, the extracted topics and keywords of the TSB recommendation dataset from the STM

4.5 A systematic perspective of topics extracted

In previous sections, we discuss the outcomes of STM and the distribution of each topic over different countries. However, these interpretations are limited to revealing underlying topics without thoroughly understanding how issues identified in the railway accident reports are intended to be addressed and what perspective investigators consider in investigating the issues. Furthermore, initial outcomes shown in previous sections cannot discover the transition in the style of making recommendations in the railway industry, hindering stakeholders from managing systematic risks in the railway recommendation system. Therefore, a systematic view of the outcomes of STM is required to clarify the relationship between recommendations made and the socio-technical system in the railway industry.

For the systematic interpretation, topics extracted are further connected to existing risk management theory to understand the behaviour of railway accident investigation bodies whilst proposing recommendations. The argument about modelling risk management by considering it to be a control problem and addressing issues from the perspective of a control structure inclusive of the society for each type of hazard begins with the work done by Rasmussen (1997). Subsequently, several studies have applied such a concept to real-world cases and proposed frameworks for interpretation (Arenius &Sträter, 2014; Grant et al., 2016; Yuyua et al., 2021). AcciMap is one of the commonly used frameworks representing the interactions between hazardous elements from different systems in a structured way, such as technology, human factors and environment (Stanton et al., 2019; Thatcher et al., 2019; Underwood &Waterson, 2013; Wheway &Jun, 2021). The AcciMap also elaborates on the decision flow, including consequences and reactions from the top to the bottom of the system. However, this framework might not be applicable to the railway recommendations data because the shape of one recommendation consists of several system levels in the socio-technical system. For example, a recommendation for a railway accident suggesting learning across organisations might involve system levels at regulatory bodies, local governments and the railway industry. This might be difficult to implement into the existing AcciMap. Furthermore, the AcciMap cannot neither visualise multiple types of recommendations nor describe the trend of decisions made by one organisation over time, thus hindering users from having a holistic map of all recommendations made by different countries.

The taxonomy of recommendations proposed by Karanikas (2016) is used to discriminate the recommendation type based on the extent to which the railway has the flexibility to address hazards identified by investigators. There are three types of recommendation proposed by Karanikas, namely assignment, action and reminder (Table 2Table 2). The assignment type of recommendation offers a distinct objective for organisations to come up with solutions and implementations and is considered a supportive recommendation. In contrast, the action type of recommendation might contain specific approaches assigned by the investigator to address hazards, limiting the flexibility of organisations to adopt solutions; as such it is categorised as the interfering type of recommendation. Lastly, the remainder type of recommendation is another supportive recommendation, providing enormous flexibility to organisations in modifying the existing rules and procedures of the operation.

Table 2, Types of made recommendations (based on Karanikas, 2016)

Recommendation type	Description	Example	Role
Assignment	Assign an objective for organisations to resolve identified hazards	Network Rail should identify and implement suitable measures to mitigate the risk of a runaway [train].	Supportive
Action	Assign specific methods to address identified hazards	Network Rail should amend its National Hazard Directory to include the access point alongside South Hampstead station;	Interfering
Reminder	Remind the compliance of existing rules or procedures	Federal Railroad Administration should increase monitoring of their employees for compliance with existing applicable rules and procedures	Supportive

There is no one-size-fits-all solution for addressing hazards in the railway industry given that each country uses a wide range of systems and has developed an inherent railway safety culture. In addition, recommendations also need to reflect the nature of the investigated railway accident and should be balanced between each type to ensure moderate flexibility in implementing solutions.

Next, this study slightly modifies the method of describing each recommendation's role (Table 2) in the railway system and proposes the customised model, referred to as *RecoMap*, to address the issues of being unable to consider systematic factors mentioned above. Instead of showing the decision flow, the *RecoMap* enables a variety of recommendations to be positioned at multiple system levels and depict the trend of the occurrence of recommendation’s types made over countries and time. Figure 7 shows the proposed *RecoMap* applied to the outcomes of the STM. The extracted topics are compressed based on their similarity and placed in the *RecoMap* in accordance with the covered systems, and the number of occurrences is labelled as well. The depth of the colour represents the time that recommendations sorted to the topic are proposed. Given that each dataset covers different periods, each colour is divided into three levels of depth representing one third of period of time covered. The lightest colour refers to the recommendations made in the first one third of period of time covered. The railway system is divided into the organisational and operational levels, representing how the socio-technical system works in the railway industry. Therefore, the *RecoMap* addresses the concern of aggregating findings and insights obtained from railway accidents in other jurisdictions. Furthermore, practitioners can review what role each recommendation plays in the socio-technical framework and understand how the legislative framework and regulations influence the railway safety in each jurisdiction through the implementation of recommendations.

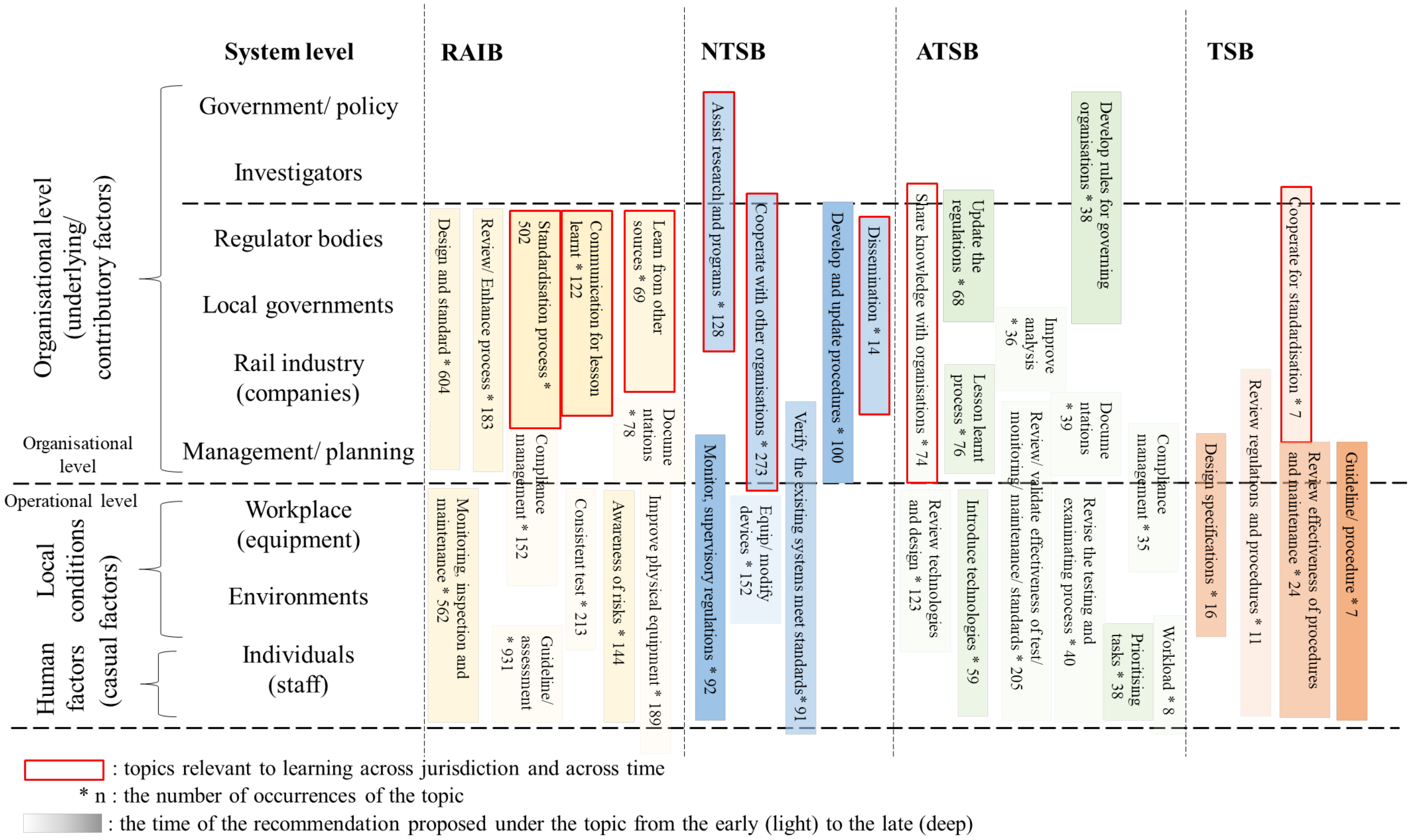

Figure 7, the proposed RecoMap applied to the outcomes of the STM

Overall, *RecoMap* maps out how investigators in different countries address identified hazards and provides the possibility for the railway industry of each jurisdiction to improve railway safety by learning across jurisdictions and time. Examples of common recommendations at the operational level are procedures of maintenance and inspection, consistency of testing processes, introducing state-of-the-art equipment, and reviewing existing designs and technologies. On the other hand, recommendations at the organisational level popularly proposed are process standardisation, cooperation with other organisations and dissemination of railway safety knowledge.

5 The transition in the style of making recommendations in the railway industry

A growing shift from addressing hazards at the operational level to the organisational level is found in ATSB and RAIB recommendations, implying the railway industry gradually adopts system theory and control theory to improve railway safety and addresses risks from the perspective of an integrated whole. Such a trend might provide useful predictive capabilities to make the railway system adapt to the dynamic environment. On the other hand, NTSB consistently offers recommendations at the organisational level, in contrast to recommendations made by TSB.

Several recommendations made by ATSB indicate detailed instructions at the operational level, such as the prioritisation of tasks, the management of workload and validation of the effectiveness of existing standards. This might imply that ATSB tends to propose interfering recommendations (the action type in Table 2) to address identified issues. However, most interfering recommendations are coloured in light, indicating a potential transition from making interfering recommendations addressing operational issues to making supportive recommendations addressing organisational issues. A similar shift can also be observed in RAIB, transferring from interfering recommendations such as improving physical equipment and assessments of individuals to design and standardisation of the system.

Furthermore, NTSB proposes many recommendations related to cooperation between organisations and assisting research and programs, implying a solid promotion of learning across jurisdictions and sharing knowledge with other research organisations. The trend continued recently along with recommendations relating to disseminating railway safety knowledge. It is also observed that NTSB consistently tends to propose precise but interfering recommendations, such as verifying existing systems and assisting research and programs.

Lastly, the number of recommendations made by TSB is extremely limited because only investigating major railway accidents results in recommendations. In addition, most of them address hazards from the operational perspective and recommendations at the organisational level are proposed only rarely, hindering how the railway industry deals with hazards as an integrated whole.

Different combinations of the style and system level of recommendations might be feasible for different roles in the railway industry. Therefore, the role that local railway regulators and national railway accident investigators play is suggested to be clearly defined under the legislative framework. For local regulators, operational recommendations might be appropriate to be proposed given the high homogeneity of railway systems and operation. In addition, local regulators have more experience and understanding of railway systems under their jurisdictions. The level of cooperation is higher than national railway accident investigators, indicating that interfering recommendations might provide more efficiency to the performance of railway safety improvement.

On the other hand, national railway accident investigators are eligible to instruct the whole railway industry, including local railway regulators. Therefore, the emphasis should be put on proposing a positive railway safety culture, disseminating railway safety knowledge and ensuring lessons are fully learnt and applied to all relevant railway organisations across the country. In doing so, the recommendations made by national railway accident investigators need to be supportive, offering organisations the best flexibility for local railway regulators and railway organisations to modify the day-to-day operation and gradually adopt new approaches to manage potential impacts. Furthermore, recommendations at the organisational level are suggested to be proposed by national railway accident investigators to enhance the communication and safety culture of the whole railway industry. Promoting learning behaviours and eliminating the obstacle of the interface between organisations by supporting cross-section engagement are also critical objectives to be achieved. Thus, proposing supportive recommendations to address hazards and manage risks from the organisational perspective might be the most beneficial for railway safety.

Table 3 shows the comparison matrix for investigated countries between the style and system level of recommendations. Cells from the top left (dark grey) to the bottom right (light grey) represent the implied combination of the style and system level of recommendations adopted by the investigator at the lower system level (i.e., local railway regulators) to the higher system level (i.e., national railway accident investigators). Each investigated country has been divided into two stages: the early stage and the current stage. Overall, the style of proposing recommendations of all countries at the early stage tends to be interfering at the operational level. Such a trend has gradually shifted to making supportive recommendations at the organisational level. However, the majority of recommendations made by NTSB are still interfering, and TSB thus far proposes most recommendations at the operational level. Therefore, it is suggested that investigators at different levels consider the role they play before coming up with recommendations.

Table 3, the comparison matrix for investigated countries between the style and system level of recommendations

	Interfering	Neutral	Supportive
Operational	ATSB (early years)	RAIB (early years) TSB (early years)	TSB (current)
Neutral	NTSB (early years)	RAIB (current)	ATSB (current)
Organisational	NTSB (current)		

Apart from the style of making recommendations, learning behaviour also plays an essential role in advancing railway safety (Paul et al., 2018; Zhan &Zheng, 2016). Topics related to learning across jurisdictions and time are highlighted with red outlines (Figure 7), including lessons learnt, communication, dissemination, and cooperation. The result suggests that investigators gradually put emphasis on exchanging knowledge and learning across organisations within the jurisdiction in recent railway accident reports, indicating that the adjustment to correct mistakes (single loop learning) and the identification of underlying factors (double loop learning) have been fully implemented in the railway industry. However, the participation of people in making well-informed decisions for addressing complicated and dynamic risks (triple loop learning) is not yet found in made recommendations. For instance, recommendations are seldom found to review cultural dimensions. The idea of learning across organisations has been proposed, but investigators rarely remind the rail industry to understand the value of making these decisions which might result in a passive attitude toward railway safety.

To sum up, the proposed *RecoMap* provides a holistic view of how different accident investigation bodies countries make recommendations, enabling the railway industry in other jurisdictions to learn potential approaches to address similar risks from these countries. The style of making recommendations is discussed, and the result suggests that the most appropriate type for each organisation might vary based on its role in the railway system. A shift from making interfering recommendations at the operational level to making supportive recommendations at the organisational level is also identified in this study. Lastly, the learning behaviours are also observed, and the analysis suggests that the behaviour of triple loop learning is still insufficient in the railway industry of the investigated countries. Learning from recommendations might not be the only way to improve railway safety but understanding recommendations can help the railway industry understand how similar issues are addressed in other jurisdictions.

6 Conclusions and suggestions

This work analyses over 9,000 sentences in the recommendation section of railway accident reports published by RAIB, ATSB, NTSB and TSB. The STM has been applied to explore latent topics within each dataset, enabling us to understand the emphasis investigators put on mitigating hazards identified. The performance metric for the STM is designed to ensure that models established reach the proper balance between SC and Exclusivity. The initial result shows distinct approaches that each investigator applied. For instance, NTSB concentrates on promoting cooperation and sharing knowledge between organisations in the railway industry, whereas ATSB makes recommendations as a reminder instead of requesting immediate actions of modifications.

To advance the interpretation of the result, this study introduces the taxonomy of recommendations, system theory and control theory to extend the analysis. The developed model *RecoMap* is proposed to describe the distribution of recommendations made from the organisational perspective over different countries, providing an alternative approach for interpreting the outcomes of topic modelling which prior works have struggled with. Additionally, a shift from making interfering recommendations at the operational level to supportive recommendations at the organisational level is observed. A growing trend of promoting learning across jurisdictions and knowledge sharing is also found across investigators. However, railway safety is proactively led by accidents and driven by authorities. Recommendations have not included people's participation in making well-informed decisions for addressing dynamic risks. In other words, railway operators know the decision should be made but might not understand why this should be done (Huang et al., 2019; Tappura et al., 2022). This might imply an insufficient understanding of triple-loop learning and result in a potential passive attitude due to underestimating the value of railway safety (Li et al., 2020).

Although the *RecoMap* allows the railway industry to learn across jurisdictions and time by offering a systematic view on recommendations made in different jurisdictions, several limitations remained unsolved and would be worth investigating. Firstly, incentives and barriers making the railway industry follow safety-related instructions are critical for decision-makers to understand the behaviour of practitioners but have not yet been revealed by *RecoMap*. Secondly, the performance and effectiveness of recommendations made by different countries cannot be evaluated and compared by the *RecoMap* although they have played an important role in cost-benefit analysis. Lastly, the interpretation of topics extracted from the topic model still requires adequate manual effort. Further work should develop advanced NLP models to overcome such difficulties.

Acknowledgments

The authors gratefully acknowledge the funding provided in the form of a Taiwan Ministry of Education - University of Sydney Scholarship (funding code SC3261).

Reference

Akel, A. J. N., Gravio, G.Di, Fedele, L., &Patriarca, R. (2022). Learning from Incidents in Socio-Technical Systems: A Systems-Theoretic Analysis in the Railway Sector. *Infrastructures 2022, Vol. 7, Page 90*, *7*(7), 90. https://doi.org/10.3390/INFRASTRUCTURES7070090

Angelov, D. (2020). *Top2Vec: Distributed Representations of Topics*. https://doi.org/10.48550/arxiv.2008.09470

Arenius, M., &Sträter, O. (2014). Resilience engineering in railways-results from a systemic accident and event analysis in German railways. *Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, ESREL 2013*, 423–428. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84899995697&partnerID=40&md5=14b941368beba5d99620ea271931e223

Augenstein, I., Maynard, D., &Ciravegna, F. (2016). Distantly supervised Web relation extraction for knowledge base population. *Semantic Web*, *7*(4), 335–349. https://doi.org/10.3233/SW-150180

Bai, X., Zhang, X., Li, K. X., Zhou, Y., &Yuen, K. F. (2021). Research topics and trends in the maritime transport: A structural topic model. *Transport Policy*, *102*, 11–24. https://doi.org/10.1016/J.TRANPOL.2020.12.013

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., &Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/JOSS.00774

Blackwood, L., &Renaud, F. G. (2022). Barriers and tools for implementing Nature-based solutions for rail climate change adaptation. *Transportation Research Part D: Transport and Environment*, *113*, 103529. https://doi.org/10.1016/J.TRD.2022.103529

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., &Mcauliffe, J. D. (2007). Supervised Topic Models. *Advances in Neural Information Processing Systems*, *20*. www.digg.com

Blei, D. M., Ng, A. Y., &Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. In *Journal of Machine Learning Research* (Vol. 3).

Brath, O. B. (2020). *Learning from incidents: Evaluating safety interventions after incident investigations* [University of Stavanger, Norway]. https://uis.brage.unit.no/uis-xmlui/handle/11250/2681460

Braut, G. S., Solberg, O., Nja, O., Solberg, T., &Njå, O. (2014). Organizational effects of experience from accidents. Learning in the aftermath of the Tretten and Asta train accidents. *Transportation Research Part A: Policy and Practice*, *69*, 354–366. https://doi.org/10.1016/j.tra.2014.08.013

Cedergren, A. (2013). Implementing recommendations from accident investigations: A case study of inter-organisational challenges. *Accident Analysis and Prevention*, *53*, 133–141. https://doi.org/10.1016/j.aap.2013.01.010

Cedergren, A., &Petersen, K. (2010). Learning from accident investigations - A cross-country comparison. *10th International Conference on Probabilistic Safety Assessment and Management 2010, PSAM 2010*, *2*, 1693–1703. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84873598544&partnerID=40&md5=ecb5bf683237e9cf0275420e5142cac5

Choularton, R. (2001). Complex learning: organizational learning from disasters. *Safety Science*, *39*(1–2), 61–70. https://doi.org/10.1016/S0925-7535(01)00026-1

Collobert, R., Weston, J., Com, J., Karlen, M., Kavukcuoglu, K., &Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. In *Journal of Machine Learning Research* (Vol. 12, Issue 76). http://jmlr.org/papers/v12/collobert11a.html

Dong, K., Romanov, I., McLellan, C., &Esen, A. F. (2022). Recent text-based research and applications in railways: A critical review and future trends. *Engineering Applications of Artificial Intelligence*, *116*, 105435. https://doi.org/10.1016/J.ENGAPPAI.2022.105435

Dornick, C., Kumar, A., Seidenberger, S., Seidle, E., &Mukherjee, P. (2021). Analysis of Patterns and Trends in COVID-19 Research. *Big Data, Iot, and Ai for a Smarter Future*, *185*, 302–310. https://doi.org/10.1016/J.PROCS.2021.05.032

Drupsteen, L., &Hasle, P. (2014). Why do organizations not learn from incidents? Bottlenecks, causes and conditions for a failure to effectively learn. *Accident Analysis and Prevention*, *72*, 351–358. https://doi.org/10.1016/j.aap.2014.07.027

Elliott, D., Smith, D., &McGuinness, M. (2000). Exploring the Failure To Learn: Crises and the Barriers to Learning. *Review of Business*, *21*(3), 17–17. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00346454&v=2.1&it=r&id=GALE%7CA73183463&sid=googleScholar&linkaccess=fulltext

Fahlbruch, B., &Schöbel, M. (2011). SOL – Safety through organizational learning: A method for event analysis. *Safety Science*, *49*(1), 27–31. https://doi.org/10.1016/J.SSCI.2010.05.004

Georges, L. R. A., &vanWitteloostuijn, A. (1999). Circular organizing and triple loop learning. *Journal of Organizational Change Management*, *12*(5), 439–454. https://doi.org/10.1108/09534819910289110

Granström, R., Söderholm, P., &Eriksson, S. (2022). Rail View, Sky View and Maintenance Go – Digitalisation Within Railway Maintenance. *Lecture Notes in Mechanical Engineering*, 224–239. https://doi.org/10.1007/978-3-030-93639-6_19/FIGURES/9

Grant, E., Salmon, P. M., Stevens, N. J., Goode, N., &Read, G. J. (2016). Exposing resistant problems in complex systems: A review of accident causation tenets. *51st Annual Conference of the Human Factors and Ergonomics Society of Australia 2016: Healthy, Safe and Productive By Design*, 35–49. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018425920&partnerID=40&md5=41fa56ac2a9b20ccb422c0cb0fe44ba7

Gray, I. (2008). What Can Railway Organisations Learn from Railway Cultural Traditions? *Historic Environment*, *21*(2), 58–67. https://researchoutput.csu.edu.au/en/publications/what-can-railway-organisations-learn-from-railway-cultural-tradit

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. https://doi.org/10.48550/arxiv.2203.05794

Guo, C., Lu, M., &Wei, W. (2019). An Improved LDA Topic Modeling Method Based on Partition for Medium and Long Texts. *Annals of Data Science 2019 8:2*, *8*(2), 331–344. https://doi.org/10.1007/S40745-019-00218-3

Han, X., &Eisenstein, J. (2019). Unsupervised Domain Adaptation of Contextualized Embeddings for Sequence Labeling. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 4238–4248. https://doi.org/10.18653/V1/D19-1433

Huang, W., Shuai, B., Zuo, B., Xu, Y., &Antwi, E. (2019). A systematic railway dangerous goods transportation system risk analysis approach: The 24 model. *Journal of Loss Prevention in the Process Industries*, *61*, 94–103. https://doi.org/10.1016/J.JLP.2019.05.021

Hulme, A., Stanton, N. A., Walker, G. H., Waterson, P., &Salmon, P. M. (2019). What do applications of systems thinking accident analysis methods tell us about accident causation? A systematic review of applications between 1990 and 2018. *Safety Science*, *117*, 164–183. https://doi.org/10.1016/J.SSCI.2019.04.016

Johnsen, S. O., Vatn, J., Rosness, R., &Herrera, I. A. (2006). Cross border railway operations: Improving safety at cultural interfaces. *Cognition, Technology and Work*, *8*(1), 76–88. https://doi.org/10.1007/s10111-005-0022-8

Karanikas, N. (2016). Exploiting data from safety investigations and processes to assess performance of safety management aspects. *Http://Dx.Doi.Org/10.1080/14773996.2016.1255444*, *14*(2), 115–127. https://doi.org/10.1080/14773996.2016.1255444

Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, *87*, 105–122. https://doi.org/10.1016/J.TRC.2017.12.018

Kume, S., &Kozaki, K. (2021). *Extracting Domain-specific Concepts from Large-scale Linked Open Data; Extracting Domain-specific Concepts from Large-scale Linked Open Data*. https://doi.org/10.1145/3502223.3502227

Kwayu, K. M., Kwigizile, V., Lee, K., &Oh, J. S. (2021). Discovering latent themes in traffic fatal crash narratives using text mining analytics and network topology. *Accident Analysis & Prevention*, *150*, 105899. https://doi.org/10.1016/J.AAP.2020.105899

Lata, K., Singh, P., &Dutta, K. (2022). Mention detection in coreference resolution: survey. *Applied Intelligence*, *52*(9), 9816–9860. https://doi.org/10.1007/S10489-021-02878-2/TABLES/17

Li, C., Lu, Y., Wu, J., Zhang, Y., Xia, Z., Wang, T., Yu, D., Chen, X., Liu, P., Guo, J., Li, C., Lu, Y., Wu, J., Zhang, Y., Xia, Z., Wang, T., Yu, D., Chen, X., &Liu, P. (2018). LDA Meets Word2Vec: A Novel Model for Academic Abstract Clustering. *WWW '18: Companion Proceedings of the The Web Conference 2018*, 1699–1706. https://doi.org/10.1145/3184558.3191629

Li, M., Zhai, H., Zhang, J., &Meng, X. (2020). Research on the Relationship Between Safety Leadership, Safety Attitude and Safety Citizenship Behavior of Railway Employees. *International Journal of Environmental Research and Public Health 2020, Vol. 17, Page 1864*, *17*(6), 1864. https://doi.org/10.3390/IJERPH17061864

Lundberg, J., Rollenhagen, C., Hollnagel, E., &Rankin, A. (2012). Strategies for dealing with resistance to recommendations from accident investigations. *Accident Analysis & Prevention*, *45*, 455–467. https://doi.org/10.1016/J.AAP.2011.08.014

Luo, M., Yu, L., &Yao, Y. (2020). Analysis on RAMS information for metro vehicles using natural language processing algorithm: Evidence from China. *ASME International Mechanical Engineering Congress and Exposition, Proceedings (IMECE)*, *14*. https://doi.org/10.1115/IMECE2020-24363

Ly, A., Uthayasooriyar, B., &Wang, T. (2020). A survey on natural language processing (nlp) and applications in insurance. *ArXiv*. http://arxiv.org/abs/2010.00462

McClory, S., Read, M., &Labib, A. (2017). Conceptualising the lessons-learned process in project management: Towards a triple-loop learning framework. *International Journal of Project Management*, *35*(7), 1322–1335. https://doi.org/10.1016/J.IJPROMAN.2017.05.006

Nolan-McSweeney, M., Ryan, B., &Cobb, S. (2022). Interviews with Rail Industry Leaders About the Management of Organisational Change. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4243482

Paul, S., Marsden, E., Verschueren, F., Tulonen, T., Ferjencik, M., Dien, Y., Simola, K., &Kopustinskas, V. (2018, October). *Accident Investigation and Learning to Improve Safety Management in Complex System: Remaining Challenges Proceedings of the 55th ESReDA Seminar Hosted by Romanian Railway Investigation Agency AGIFER*. https://doi.org/10.2760/242130

Pilbeam, C., Davidson, R., Doherty, N., &Denyer, D. (2016). What learning happens? Using audio diaries to capture learning in response to safety-related events within retail and logistics organizations. *Safety Science*, *81*, 59–67. https://doi.org/10.1016/J.SSCI.2015.07.034

RAIB. (2009). *Fatal accident at Halkirk level crossing*. https://www.gov.uk/raib-reports/fatal-accident-at-halkirk-level-crossing-caithness

RAIB. (2014). *Fatal accident at Moreton-on-Lugg*. https://www.gov.uk/raib-reports/fatal-accident-at-moreton-on-lugg-near-hereford

RAIB. (2017). *Fatal accident at Alice Holt footpath crossing, Hampshire*. https://www.gov.uk/raib-reports/fatal-accident-at-alice-holt-footpath-crossing-hampshire

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science*, *27*(2–3), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0

Roberts, M. E., Stewart, B. M., Tingley, D., &Airoldi, E. M. (2013). The structural topic model and applied social science. *Adv. Neural Inf. Process. Syst. Workshop Top. Models: Comput. Appl. Eval.*, 1–20.

Roque, C., Lourenço Cardoso, J., Connell, T., Schermers, G., &Weber, R. (2019). Topic analysis of Road safety inspections using latent dirichlet allocation: A case study of roadside safety in Irish main roads. *Accident Analysis & Prevention*, *131*, 336–349. https://doi.org/10.1016/J.AAP.2019.07.021

Rosness, R. (2013). The proceduralization of traffic safety and safety management in the Norwegian rail administration: A comparative case study. In *Trapping Safety into Rules* (pp. 173–189). Ashgate Publishing Ltd. https://doi.org/10.1201/9781315549774-11/PROCEDURALIZATION-TRAFFIC-SAFETY-SAFETY-MANAGEMENT-NORWEGIAN-RAIL-ADMINISTRATION-COMPARATIVE-CASE-STUDY-RAGNAR-ROSNESS

Rydstedt Nyman, M. (2019). Collective Learning in Organizations—Opportunities and Constraints: Case Study of an Avalanche Blocking a Railway Line. *Risk, Hazards & Crisis in Public Policy*, *10*(3), 332–359. https://doi.org/10.1002/RHC3.12159

Single, J. I., Schmidt, J., &Denecke, J. (2020). Knowledge acquisition from chemical accident databases using an ontology-based method and natural language processing. *Safety Science*, *129*(February), 104747. https://doi.org/10.1016/j.ssci.2020.104747

Stanton, N. A., Salmon, P. M., Walker, G. H., &Stanton, M. (2019). Models and methods for collision analysis: A comparison study based on the Uber collision with a pedestrian. *SAFETY SCIENCE*, *120*, 117–128. https://doi.org/10.1016/j.ssci.2019.06.008

Steiro, T., Johnsen, S. O., Bergersen, C. E. B., &Tungland, M. (2004). Assessment and Improvement of Safety Culture across Borders in European Railways. *Probabilistic Safety Assessment and Management*, 1071–1076. https://doi.org/10.1007/978-0-85729-410-4_173

Stemn, E., Bofinger, C., Cliff, D., &Hassall, M. E. (2018). Failure to learn from safety incidents: Status, challenges and opportunities. *Safety Science*, *101*, 313–325. https://doi.org/10.1016/J.SSCI.2017.09.018

Størseth, F., &Tinmannsvik, R. K. (2012). The critical re-action: Learning from accidents. *Safety Science*, *50*(10), 1977–1982. https://doi.org/10.1016/J.SSCI.2011.11.003

Syeda, K. N., Shirazi, S. N., Asad, S., Naqvi, A., &Parkinson, H. J. (2017). Exploiting Natural Language Processing for Analysing Railway Incident Reports. *IGI Global Publishing*, 1–18. http://www.digitalrail.co.uk/wp-content/uploads/2018/01/Exploiting-Natural-Language-Processing-for-Analysing-Railway-Incident-Reports.pdf

Tappura, S., Jääskeläinen, A., &Pirhonen, J. (2022). Creation of satisfactory safety culture by developing its key dimensions. *Safety Science*, *154*, 105849. https://doi.org/10.1016/J.SSCI.2022.105849

Thatcher, A., Nayak, R., &Waterson, P. (2019). Human factors and ergonomics systems-based tools for understanding and addressing global problems of the twenty-first century. *Ergonomics*, *63*(3), 367–387. https://doi.org/10.1080/00140139.2019.1646925

Tretten, P., &Candell, O. (2021). AHFE (7) - Digitalization of Railway Maintenance: A Situation Awareness Perspective. In *Advances in Human Factors and System Interactions: Vol. NA* (Issue NA). https://doi.org/10.1007/978-3-030-79816-1_25

Underwood, P., &Waterson, P. (2013). Systemic accident analysis: Examining the gap between research and practice. *Accident Analysis and Prevention*, *55*, 154–164. https://doi.org/10.1016/j.aap.2013.02.041

Wang, C. L., &Ahmed, P. K. (2003). Organisational learning: A critical review. *The Learning Organization*, *10*(1), 8–17. https://doi.org/10.1108/09696470310457469/FULL/PDF

Watson, S. (2004). Training rail accident investigators in UK. *Journal of Hazardous Materials*, *111*(1–3), 123–129. https://doi.org/10.1016/J.JHAZMAT.2004.02.026

Wheway, J. L., &Jun, G. T. (2021). Adopting systems models for multiple incident analysis: utility and usability. *International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care*, *33*(4), NA-NA. https://doi.org/10.1093/intqhc/mzab135

Wrigstad, J., Bergström, J., &Gustafson, P. (2014). Mind the gap between recommendation and implementation - Principles and lessons in the aftermath of incident investigations: A semi-quantitative and qualitative study of factors leading to the successful implementation of recommendations. *BMJ Open*, *4*(5), 5326. https://doi.org/10.1136/bmjopen-2014-005326

Yang, J., &Anwar, A. M. (2016). Social Media Analysis on Evaluating Organisational Performance: A Railway Service Management Context. *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*, 835–841. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.143

Yang, J., Han, S. C., &Poon, J. (2021). *A Survey on Extraction of Causal Relations from Natural Language Text*. https://arxiv.org/abs/2101.06426v1

Yuyua, G., Xu, K., Li, J., Yao, X.-W., Wu, C.-Y., Li, S.-Y., Yan, F., Zhang, J., Xu, Q.-W., &Ji Zhang. (2021). Interaction Theory of Hazard-Target System. In *NA: Vol. NA* (Issue NA, p. NA-NA). https://doi.org/NA

Zhan, Q., &Zheng, W. (2016). An in-depth learn from incidents (LFI) approach based on HFACS-RAs. *2016 IEEE International Conference on Intelligent Rail Transportation, ICIRT 2016*, 177–185. https://doi.org/10.1109/ICIRT.2016.7588729