Safety Science 110 (2018) 11–19

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

From free-text to structured safety management: Introduction of a semi-automated classification method of railway hazard reports to elements on a bow-tie diagram

Peter Hughes[a,*], David Shipp[b], Miguel Figueres-Esteban[a], Coen van Gulijk[a]

[a] University of Huddersfield, Queensgate, Huddersfield HD1 3DH, United Kingdom
[b] Network Rail, The Quadrant, Elder Gate, Central Milton Keynes MK9 1EN, United Kingdom

ARTICLE INFO

Keywords:
Natural language processing
Close calls
Bow-tie diagrams

ABSTRACT

This paper introduces a semi-automated technique for classifying text-based close call reports from the GB railway industry. The classification schema uses natural language processing techniques to classify close call reports in accordance with the threat pathways shown on bow-tie diagrams. The method enables categorisation of a very large number of unstructured text documents where safety-related information has not previously been extracted due to the infeasibility of analysis by human readers. The results demonstrate mixed accuracy in the categorisation of close calls, with the highest accuracy being for the threat pathways that are more frequently reported. This work paves the way to machine-assisted analysis of text-based safety and risk databases, and provides a step forward in the introduction of data analytics in the safety and risk domain. Others working in this area have speculated that approaches such as this could be mandatory for safety management in the future.

1. Introduction

This paper presents a hazard reporting and safety management method that has been developed for the railways in Great Britain (the *GB railways*). The method aligns aggregated data from free-text hazard reports provided by workers with the bow-tie accident causation models that are used by the railways. Since the hazard reports had not previously been used in this way, the new method demonstrates the possibility for valuable new source of information to be used for safety management which, in turn, allows for improved management of safety risks.

Railways around the world invariably have formal reporting mechanisms to record accidents. These records are analysed to gather information about the hazards that led to the accidents and provide recommendations to prevent recurrence of similar events. In addition to such an accident reporting system, the GB railway industry provides a formal mechanism for workers to report safety hazards even without an accident having occurred, so called *close calls*. Since the close call reporting system provides information on the hazards before injury or damage occurs, it provides a valuable complement to an accident reporting system.

The close call reporting mechanism used by the GB railways allows workers to report by sending an email message or by directly entering textual descriptions of hazards into a mobile phone app. This free-text reporting method allows workers to describe hazards that they encounter on the railways without being constrained to pre-existing risk categories, such as systems that use drop-down lists. It also allows workers to report whatever information is relevant to the hazard. Workers are able to describe complex sequences of causation and nuances that would not be feasible with a system that did not allow free-text. Similarly, free-text reporting allows workers to report on emerging hazards that may not be listed in a drop-down list, for example hazards associated with emerging technology that were not known about at the time a drop-down list was developed.

Since implementing the system, the GB railways have obtained a very rich source of information regarding hazards on the railway, with workers reporting approximately 200,000 close calls per year. With such a high volume of data, it is impractical for railway safety managers to obtain an overview of hazards on the railway by manually reading each record. Without an automated process to support the aggregation of data, it is likely that trends in hazards remain undetected. A major impediment to automatically aggregating information from close call records is that the free-text reporting method means that workers can provide many different descriptions of similar types of hazard, which can make it difficult to automatically aggregate hazard information to obtain an overview of trends in hazards. A further impediment is that

* Corresponding author.
E-mail addresses: p.hughes@hud.ac.uk (P. Hughes), david.shipp@networkrail.co.uk (D. Shipp), M.Figueres@hud.ac.uk (M. Figueres-Esteban), C.VanGulijk@hud.ac.uk (C. van Gulijk).

https://doi.org/10.1016/j.ssci.2018.03.011
Received 12 June 2017; Received in revised form 1 February 2018; Accepted 4 March 2018
Available online 13 March 2018

workers using the mobile phone app enter text on a small touch-screen which is vulnerable to mistyping, especially if the workers have cold or gloved hands. Therefore, whilst the data source has the potential to provide valuable safety information, currently no exhaustive learning is being propagated through the railway's management systems to improve safety of the railway.

The GB railways have an extensive safety management system, which includes the use of bow-tie diagrams to model potential accident causation; understand hazards on the railway; assess how the hazards can lead to accidents; and identify suitable controls to manage the risk. Since close calls provide valuable information on hazards on the railway, an ambition of the railway's safety management process is to aggregate information from the close calls in a way that integrates with the existing bow-tie diagrams. In such a system the railway would be able to identify trends in hazards and the risk controls that are breaking down. Furthermore there is the possibility that there may be close calls describing hazards that have not been identified and are therefore not managed within the railway's safety management. Integrating close call information with bow-tie diagrams again provides a valuable opportunity to identify improvements to the existing safety management system. To address this need, this paper introduces a safety management to extract safety-relevant information from close call records so that the records can be categorised according to the structure of an existing bow-tie diagram; no process is know that currently performs this task. The method proposed in this paper uses an iterative process between a safety specialist and automated text analysis techniques

The process presented in this paper seeks use the rich source of data provided by workers in a way that integrates with the railway's existing management systems to improve safety. Since no fully automated process is known that can perform this task, the proposed method uses semi-automated natural language processing (NLP) techniques that rely on an iterative process between a human user and a computer to develop a categorisation schema for records. The method is illustrated by a case study for *worker slip, trip or fall* (STF) events which are common accidents to which all workers on the railway are exposed. A contribution of this paper is that it brings together close call reporting, use of bow-tie diagrams for safety management and natural language processing to inform safety management processes in a way that has not been possible until now.

Section 2 of this paper describes the background of close call reporting, the use of bow-tie diagrams for safety analysis, and NLP. Section 3 sets out the classification method used in the case study. Section 4 presents the results: the hazard trends identified based on the bow-tie diagram. Section 5 discusses the results and potentials for further development; and Section 6 concludes.

2. Theoretical background

The theoretical background to this work is derived from three different areas of research: reporting close calls as a part of a safety management system; use of bow-tie diagrams to understand risks to safety; and natural language processing to extract information from text. The following sections discuss these areas of research. No prior work has been found that merges these three concepts.

2.1. Close call reporting

A close call is "*a hazardous situation where the event sequence could lead to an accident if it had not been interrupted by a planned intervention or by random event*" (Gnoni et al., 2013, Andriulo and Gnoni, 2014). The concept of identifying close calls, near misses or no-injury accidents to improve safety by intervening before an accident is observed goes back to Heinrich (1929). In Lateiner (1958) posited that safety-related incidents could be compared to an iceberg: incidents that result in damage or injuries are only the visible proportion of all incidents; no-injury accidents, whilst larger in number, remain invisible but should nevertheless be actively investigated as a proactive approach to preventing accidents. The concept that no-injury events, near misses and close calls are invaluable information has continued to be upheld by safety researchers who generally agree that action taken to prevent close calls will necessarily result in a reduction in accidents (Bird and Germain, 1966, Petersen, 1971, Wright and Van der Schaaf, 2004, Gnoni and Lettera, 2012). This concept is supported by Gnoni and Lettera (2012) who assert that close calls and serious accidents at major hazard facilities both have the same underlying causes. An inverse relationship between near miss reports and accidents at Norsk Hydro industrial plant between 1985 and 1997 was demonstrated by Jones et al. (1999).

Despite the benefits available from collecting close call information, the process is not without problems. As well as suffering from traditional impediments to reporting, such as fear of disciplinary retaliation, close call reporting can suffer from additional impediments since the reporting process may be disproportionately time consuming and suffer from a lack of feedback to workers reporting hazards. Any close call reporting system needs to be carefully designed system to allow extraction of sensible learning (Cambraia et al., 2010).

Another difficulty is the question of how to assess the seriousness of a close call. Cambraia et al. (2010) and Gnoni and Lettera (2012) have used an elaborate semi-quantitative method to assess the seriousness of close calls based on the risk matrix to establish whether safety interventions would be justified. Disseminating information on near misses and close calls also raises problems; work by Dillon and Tinsley (2008) shows that workers with knowledge of near misses actually undertake in riskier behaviour because they develop a mindset of counterfactuals that suggests hazards do not often materialize into accidents. Bliss et al. (2014) summarize seven problems that impede the development of a sensible near-miss system: development of a theoretical model for close calls; leveraging the use of close calls; management of progression and actions based on close calls; training to optimize benefits from close calls; separating close calls and false alarms; experiencing the gravity of a close call for people not involved in the incident; and quantification of the close call severity.

In general, the available literature on close calls considers only safety management systems that have not benefitted from the widespread introduction of a digital reporting system that is open to all workers in the industry. Rather, the projects described usually have relatively small numbers of close-calls: Gnoni and Lettera (2012) consider just 11 close call reports in an undisclosed period of time; Cambraia et al. consider 122 reports in eight months. The much larger source of data considered by Jones contains approximately 4000 records that were collected over seven years. Given these potential benefits, the work described in this paper seeks to use the railway's close call database, which is currently being populated with approximately 200,000 close calls reports annually, in a way that improves the safety of the railway for all workers. This volume of information provides a benefit and a challenge to safety management: with such a large number of records, there is a rich source of valuable information. Conversely extracting relevant information can be difficult.

2.2. Bow-tie diagrams as safety analysis tools

Bow-tie diagrams illustrate the direct relationships between hazards, tasks, safety controls, risks, and the potential outcomes of accidents. Since their creation by Nielsen (1971), bow-tie diagrams have supported safety management in a wide range of industries but particularly in the process industries (Visser, 1998, Cameron and Raman, 2005, Hudson, 2010). As powerful safety management tools, earlier researchers have described using bow-tie diagrams for modelling hazards related to human factors (Targoutzidis, 2010) and for using bow-tie diagrams to manage multiple, possibly conflicting, objectives such as "*maximizing effectiveness, reliability and availability and minimizing the global cost of all barriers*" (Badreddine et al., 2014). Recently the railways in Britain have integrated bow-tie diagrams into their safety management system (Turner et al., 2015). The railway infrastructure manager has developed a suite of bow-tie diagrams that identify business-critical rules and how they influence operational risks.

Our work uses the railway's existing bow-tie diagrams to provide a framework for classifying information extracted from close call reports. Close calls give an indication of which hazards are most prevalent on a day-to-day basis within the GB railways. Linking information from close call reports to the bow-tie diagrams provides a method of feedback on safety performance that is directly correlated with the process that is used to monitor and review the safety management system. This paper uses a case study based on the railway's bow-tie diagram for worker slip, trip and fall (STF) events.

2.3. Analysing close calls using NLP techniques

Manual analysis of free-text close calls places a considerable demand for resources on an organisation, however the effort required can be reduced by applying semi-automated techniques (Popping, 2000, Xu et al., 2010). Popping describes three approaches for analysing text: thematic, semantic, and network approaches. The thematic approach is the simplest but is nevertheless powerful and effective for safety learning (Hughes et al., 2015). Using this approach records are categorised according to characteristic tokens that occur in the text; a token could be a single word, a consecutive sequence of two words (a *bigram*) such as *level crossing*, or any number of consecutive words (an *n-gram*). The approach requires development of a list of domain-specific tokens that can be used identify key information in the source text. In some cases tokens may be sufficient by themselves to identify the theme of a text record, in general however, a number of tokens are usually required to provide reliable categorisation of a record into a particular theme. A very similar method was employed by Tixier et al. (2016) who noted that "*although this approach is simple, it is very powerful*". An important difference is that the study described in this paper is not looking simply to find what categories of incident can be found in the text records, rather in this study there are pre-defined categories of incidents described in the bow-tie diagram.

Regardless of the approach used to extract meaning from text, there are no fully autonomous methods available that can take only text as an input and provide categorisations, abstractions or any other knowledge as an output (Tixier et al., 2016). Unsupervised methods exist for detecting patterns in text based on the occurrence of words such as *k-means* clustering (Manning and Schütze, 1999) or Latent Dirichlet Allocation (Blei et al., 2003). These methods allow documents to be grouped into themes; but they provide no understanding of what the themes relate to (Noel et al., 2014). By themselves words are merely labels for objects and events that occur in the real world. Whilst software agents can perform mathematical pattern matching on words, software that has not experienced the real world cannot understand the *meaning* of the words; meaning arises from the interpretation of words in the mind of a reader.

Classification of natural language is not a task that can be performed in the absence of an understanding of the purpose of the classification activity (Li, 2015). The corpus of close call records contains a large number of records that could be categorised in many ways. In addition to categorising close call records by the nature of the safety hazard being described, other groupings are also possible such as: length of the record (some close calls are only a few words, some are hundreds); style of writing (some close calls use many abbreviations and railway-specific terms, others use none or hardly any); or by dialect differences (workers submitting close calls in northern Britain use different terms to workers in the south). An unsupervised software agent could not be expected to guess what method of categorisation is required by the human operator without some form of guidance on what information is being sought. Even when supervised machine methods are used to extract information from natural language, such as the use of artificial neural networks, the quality of the results can be varied and in some cases extensive human input is required to obtain acceptable outcomes (Mou et al., 2016). An analogous problem arises with the development of ontologies for categorisation of natural language; Ruiz-Martínez et al. (2011) note the need for manual intervention in the ontology learning process since "*techniques for learning domain ontologies from free natural language text have important drawbacks*". The consensus in the literature is that developing the correct rules for extracting information from text is an incremental process (Sánchez Ruenes, 2007), with no perfect solution classification systems have to be developed that are fit for a specific purpose (van Gulijk et al., 2016).

3. Categorisation of close call records

This paper treats a case study of a worker *slip trip and* fall (STF) bow-tie diagram. To aggregate information from the close call records in a way that supports integration with the railway's safety management system, each close call record was examined to determine how it corresponded with descriptions of hazards, controls or threat pathways on the STF bow-tie diagram. Bow-tie diagrams are in integral part of the

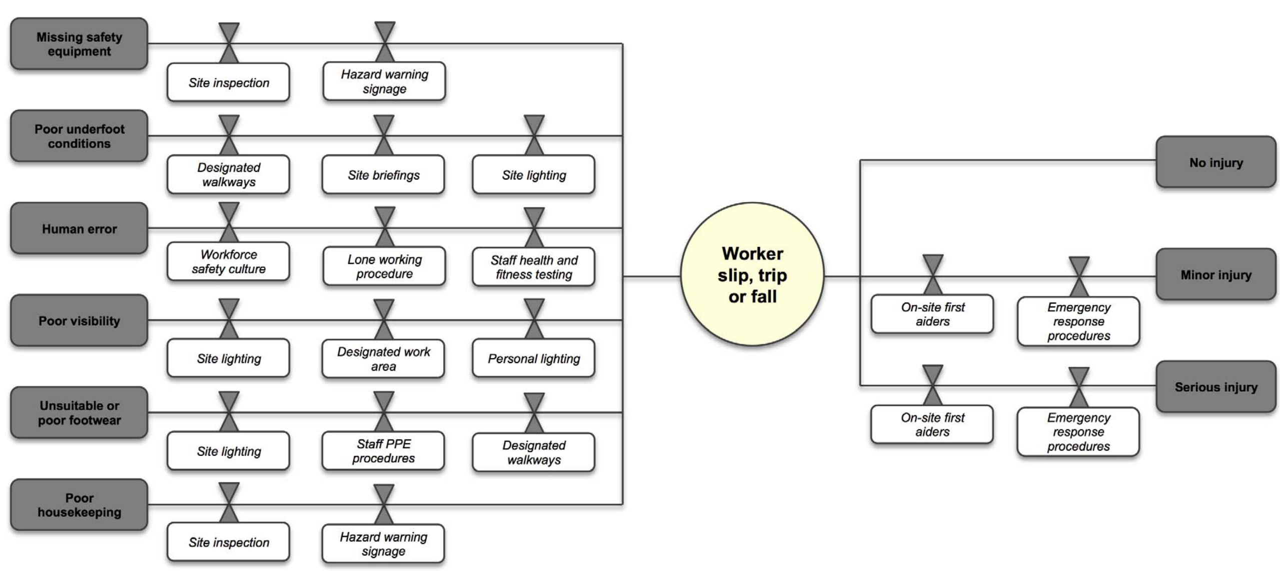

Fig. 1. Overview of STF bow-tie diagram.

railway's safety management system and are a rich source of information on the controls each place for each loss-of-control (LOC) event. Fig. 1 provides an overview of the railway's *slip, trip or fall* bow-tie diagram. At the left-hand side of the diagram are six initiating conditions (hazards): *missing safety equipment*; *poor underfoot conditions*; *human error*; *poor visibility*; *unsuitable or poor footwear*; and *poor housekeeping*. On the threat pathways between the initiating events and the LOC event are a number of barriers which are each represented on the diagram with double triangle. The barriers are the risk control that are intended to prevent the initiating events leading to the LOC event; for each barrier there will be corresponding details in the railway's safety management system describing the control, how it is implemented and monitored.

On the right-hand side of the diagram possible outcomes of the LOC event are shown. In general these outcomes can be any outcome including adverse financial impacts, loss or reputation or legal sanctions subsequent to an accident. For this case study only outcomes directly resulted to injury to people were considered. Again there are barriers that can prevent the LOC event leading to undesired outcomes. Close call events typically report the presence of initiating events, with reference to Fig. 1, a close call may report an instance of a worker on site with inadequate footwear. Alternatively a close call record may report failures of barriers; for example inadequate lighting on site, or the lack of a suitably trained first-aider. In either case, the close call report does not state that an LOC event occurred, nor that an undesired outcome occurred, instead the close call report describes the instances of circumstances where such events may occur.

To perform the analysis, all close call reports from 01 October 2013 to 28 February 2015 were extracted from the database: a total of 219,231 records. Example records are shown in Table 1.

The close call reports generally report hazards that could lead to safety-related incidents, although some reports do describe health hazards, such as exposure to asbestos.

The process was carried out broadly in accordance with the process set out in Hughes et al. (2015), *viz.*:

Step 1: text pre-processing;
Step 2: interactive token selection;
Step 3: record selection; and
Step 4: post-processing.

These steps are described in the following sections and illustrated in Fig. 2.

3.1. Step 1: pre-processing

The close call records imported from the database contain a number of coding artefacts that are not needed to extract information on safety hazards. The example records in Table 1 include the following examples of text: `<br /> <div> </div> <!-- RICH TEXT -->`, and an occurrence of the character in place of an apostrophe ('). These data artefacts are unnecessary since they do not provide information useful for safety analysis and can interfere with NLP processes. As a pre-processing step, Java software was written employing *regular expression* functionality (Thompson, 1968) to cleanse the source data. The software imported the source data, searched for occurrences of unwanted text elements and either removed the unwanted text or, where it was possible to determine with a high degree of confidence what the correct text should be, substituted the text with corrected data. The cleansed text was exported to a new file.

Table 1
Example records from the railway's database of close call records.

Operatives found working on the east platform with insufficient barriers to protect the public from the work site. Members of the public could have strayed into the work area   <!-- RICH TEXT -->
Portmill Station Machine/crane Falkirk TGM no identity badge worn. Operatives/Supervision unclear whom was directing RRV
Reported by A global operative Lap 5-283 - loc 124 (SWB 1084E3) cable damage seen as cable emerge from the trough route. SSL/BCM to rectify asap
Don't walk By Draw s on QJB project have not happened for over 6 months. Reported on Previous DWB s

3.2. Step 2: interactive token selection

To identify tokens the source records were filtered to select only those that had been categorised at the time of entry as relating to worker STFs. From this subset of records, *n-grams* (consecutive sequences of words) were identified in the text in the new file. Unlike the work of Tixier et al. (2016) where tokens were identified through a manual review of records, for this study the Java software was used to extract all 1- to 5-grams (sequences of one to five words) from the corpus of records and sort them by frequency of occurrence. For this work, an point followed by a space (·) was considered to mark the end of a sentence; n-grams were not considered to flow over sentences. Tokens are used to identify commonly occurring terms in the text that are meaningful to safety management; for example n-grams such as *trip hazard* and *pit cover* occur frequently in the text and are instances where two consecutive words (a two-gram) are required to describe a single concept. In performing this task it is necessary to understand the domain of railway operations and safety management in order to identify n-grams that are relevant to identify hazards.

A railway safety specialist manually reviewed and selected the subset of n-grams that would be most useful as tokens for characterising the close call records on each threat pathway on the bow-tie diagram; each threat pathway could be identified by several n-grams. The selection of tokens is a manual process since it requires knowledge of the domain and what information is relevant for safety management. Since natural language contains synonyms it is necessary to identify a number of terms that occur in free-text that can be mapped to a single token; for example two terms that occur commonly in close call reports are: *slippery underfoot conditions* and *icy walking conditions*. Whilst these terms contain different words and different nuances, within the context of managing safety hazards on the railway they have largely the same meaning: namely that the hazard relates to a potential for an STF event due to the condition of the walking surface. Similarly, identification of tokens for analysing free-text must allow for commonly occurring spelling variations. For example within the railway close calls corpus, the words *slippery* and *slippy* are both used commonly to convey the same meaning. Furthermore close call records entered via the mobile phone app contain many spelling variations as workers may be outside or working in cold conditions or with gloved hands. However selection of terms must discriminate carefully where small differences in spelling can have a large impact on meaning, for example within close call reports the term *m.o.p.* is commonly used as an abbreviation for *member of the public*; whilst the similarly spelled *m.o.m.* refers to a *mobile operations manager*. Both terms refer to a type of person, and therefore automatic language processing techniques may identify that the properties of an m.o.p. and an m.o.m. are similar; however in terms of railway safety management the terms represent different concepts and therefore they cannot be considered synonyms in this context.

The process of token selection was performed iteratively: the software was initially run to identify common n-grams from the full corpus of records. The safety specialist identified those n-grams which would be used to select close call records relating to each threat pathway. Having identified candidate n-grams; a subset of close call records was extracted with only those records that contained the n-gram (for example only records containing the term *pit cover*) was extracted, the process of n-gram identification was then repeated on the subset to

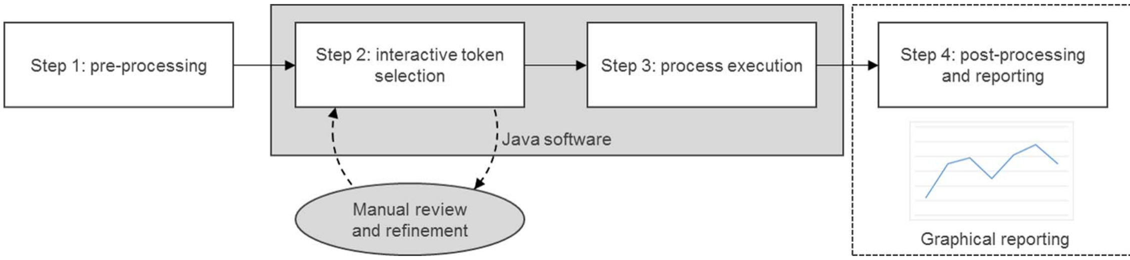

Fig. 2. Overview of the text extraction process.

identify any further n-grams that were considered relevant for the threat pathways. This process allowed the analyst in selecting n-grams when there are large number of frequently occurring terms that are not useful for identifying safety hazards; for example the following two-grams occurring frequently in all records: *to the*, *in the*, *was in*. These example two-grams occur frequently in close call records regardless of what hazard is being described and are not valuable for identifying hazards on any particular threat pathway in the STF bow-tie diagram. During the iterative process, the n-gram set was tested on sample records to determine its effectiveness at correctly identifying records; the results were reviewed to determine whether records were accurately selected using the n-grams. The process of n-gram selection was repeated until an acceptable set of tokens was achieved.

Table 2 shows examples of the n-grams that were selected to identify records belonging to each of the threat pathways. The table shows that there are variations in the terms that are used by workers when describing concepts, for example the terms *over grown*, *overgrown*, and *over-grown* all occur within the source text.

Using this process it is clear that a record could be categorised to more than one threat pathway, for example a record that contained both the tokens footwear and uneven surfaces would be categorised to both the *unsuitable or poor footwear* threat pathway and the *underfoot conditions* threat pathway. Such an outcome is meaningful and is not deficiency in the process. Whilst the bow-tie diagram is structured to consider each threat pathway as distinct a distinct sequence to a worker STF event, in reality this may not be the case. A hazard could occur on the railway as a result of a combination of missing safety equipment, poor underfoot conditions, and human error. Railway workers may use the close call system to report any hazards on the railway and the free-text reporting permitted by the close call system does not constrain workers to report only one hazard per record. A worker may use the system to report multiple hazards that have occurred at different locations and at different times. As such it is meaningful that a close call report could describe hazards in any number of threat pathways.

Table 2
Example tokens used to identify each threat pathway.

Threat pathway	Example n-grams used to classify record	
Missing safety equipment	defect defective missing with no	no cover knocked off not properly secured …
Underfoot conditions	uneven rough surface slippery slippy	rainy rain raining snow …
Human error	forgot forget	not looking …
Poor visibility	overgrown over grown over-grown vegetation	fog foggy visibility …
Unsuitable or poor footwear	footwear foot wear foot-wear boots	shoe shoelace shoe lace
Poor housekeeping	loose untidy messy left on left by	trip over step over was left scrap rubbish …

Since the token identification is based on the structure of the bow-tie diagram, the hazards that are identified from the close call reports are only those that are described on the bow-tie. Close call records that relate to hazards that are not described in an existing bow-tie diagram may not be detected using this method. The case study described in this paper considers the worker STF bow-tie diagram, however a complete implementation of the process could be expected to consider all bow-tie diagrams used by the railway. Records that do not contain tokens relating to any bow-tie diagram could be identified and either reviewed manually, or analysed by software that identified commonly occurring n-grams, in order to identify hazards on the railway that are not described in the bow-tie diagrams and may therefore not have controls within the railway's safety management system.

3.3. Step 3: record selection

The close call records and selected tokens were provided as input to software that selected records and categorised each record to a number of threat pathways based on the presence of tokens in the records. The process of categorisation took approximately 15 s on a standard desktop computer to categorise the full corpus of close call records, or approximately 15,000 records per second. For comparison, a human reader could be expected to read and categorise approximately five close call records per minute, and would therefore take approximately 100 working days. During such a task it cannot be expected that a human reader would categorise records consistently. If the task were shared among a number of readers it can again be expected that there would be variations between the categorisation used by each reader.

3.4. Step 4: post-processing

After the process of categorisation, the number of close call records in each threat pathway was counted for each calendar month and the result was presented graphically in format based on the existing precursor indicator model used by the GB railways (RSSB, 2015).

4. Results

The tokenisation process identified records corresponding to each of the threat pathways. Examples of close call records that were categorised to each threat pathway are shown in Table 3. The results in the table show demonstrate how records can be identified based on the simple token set, for instance considering the results the example tokens in Table 2, it can be seen that the token missing in result 1a

Table 3
Exemplar close calls found for each threat pathway.

Threat pathway		Exemplar close call records matching each threat pathway
1	Missing safety equipment	a. `Handrail missing on stairs to project office` b. `Operatives working around broken pit cover could have fallen and sustained injury` c. `Bearers rotted and broken could have caused slips and falls, RRV could have fall over`
2	Underfoot conditions	a. `Snow and ice on steps to project office, handrail not safe and needs to be repared` b. `Access to the carpark is only possible over uneven and rough ground – one staff got his foot stuck in rabbit hole` c. `Walking route around site has many hazards and need to be cleared up – rubbish and shotgun cartridge found and walking is on slippery ground`
3	Human error	a. `During site briefing the site supervisor forget to mention the icy conditions – could have cause slip trip fall` b. `Cleaning staff omitted to replace nonslip matting at the entrance to sm office risk to staff and public – icy condition underfoot could cause slip fall hazard`
4	Poor visibility	a. `Night work on site with inadequate lighting – safe walk route not visible` b. `The path leading to the worksite was overgrown – vegetetion needs to be cleared before staff are sent to site` c. `Working in fog, operatives were too far away to see lookout, also poor walking conditions could have slipped`
5	Unsuitable or poor footwear	a. `Contractors on site without proper ppe –not wearing approved harnesses or safety boots` b. `on going into the signalbox the signaller was wearing flip flops which is unsuitable footware for any working environment could easily have sutbbed his toe or other` c. `Staff provided with unsuitable ppe boots – cannot be properly laced up leading to potential for slip trip fall`
6	Poor housekeeping	a. `Build up of rubbish and scrap materials at south end of carpark – unsafe access to site and trip hazard` b. `Extension leads have been left trailing across boards causing a trip hazard, operative could have tripped` c. `Gas bottles not stored properly, walking access past crane, storage area needs better organising, tripping hazards (timber, steel rebar, etc.)`

indicates where safety equipment (in this case a handrail) is absent. The results for threat pathway 1 show that the results can identify failures of different types of safety equipment, including a handrail, a pit cover, and a bearer. These results have been obtained because of similarity in the terms used to describe failed safety equipment, rather than similarity in the terms used to described different types of safety equipment. It can also be seen that some of the records may correspond to more than one threat pathway. For example result 3b in Table 3 describes two hazards: a human error where matting had not been replaced, and icy conditions. Using this tokenisation technique the event would be counted as both a *human error* hazard and an *underfoot conditions* hazard.

Fig. 3 shows the numbers of close call records categorised to each threat pathway by the month when the close call was reported. These results show that *poor housekeeping* and *poor underfoot conditions* are the most commonly reported causes of worker STF hazards on the railway. The dip in reporting STF hazards during August and September is consistent with an overall reduction of close calls reported in these months, which in turn corresponds with a popular holiday time when fewer overall hours are worked on the railway. The threat pathway *missing safety equipment* is not expected to necessarily be seasonal and similarly does not show any distinct seasonal variation. The variation in the numbers of close calls reporting poor housekeeping varies in part with the overall numbers of close calls being reported, however it also shows a slowly increasing overall trend, which can also be seen in the total numbers of close calls being reported. The exact causes of the variations in the numbers of close calls being reported are a matter for safety management activities to understand the nature of the underlying hazards and put in place controls to reduce the risk.

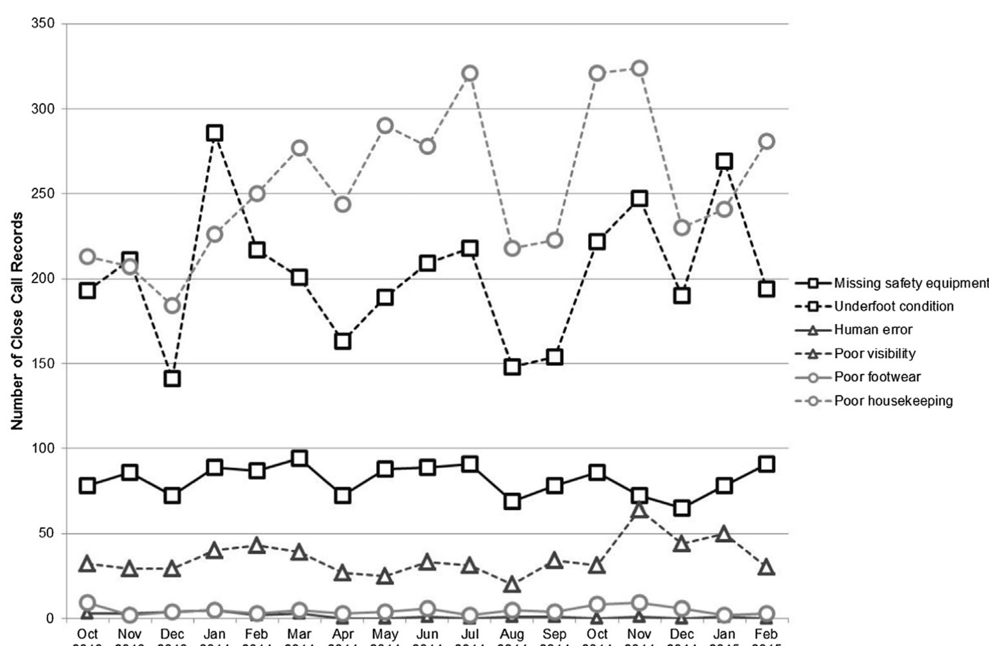

Fig. 3. Count of close calls per month assigned to each threat pathway.

Table 4
Accuracy of the current NLP-based assignment on a sample of 100 records.

Categorisation	Correct	Incorrect
Record does not relate to worker STF event	10	5
Human error	0	3
Missing safety equipment	4	10
Poor housekeeping	31	8
Underfoot conditions	20	9

5. Discussion

5.1. Applicability of the technique

The results show that overall the technique has the ability to categorise close call reports consistently with threat pathways shown on the railway's bow-tie diagrams. In this way, railway safety managers have the ability to understand hazards on the railway in a way that integrates with the existing safety management system. It is possible to identify hazards that occur most frequently and the risk controls that are breaking down. The technique provides insight not only into hazards that are occurring but, as a corollary, where hazards are not occurring. Such insight – provided directly by workers who are using risk controls as part of their day-to-day work – gives information that can strengthen the safety management system. The information from this technique can support other direct sources of information from the railway including site inspections, audits, and incident investigations. Following application of a similar approach in the construction industry, Tixier et al. (2016) consider that "*the use of NLP may soon become mandatory and widespread in construction management to make sense of the ever-growing amount of digital information*". As software for performing NLP becomes increasingly available, safety managers within the GB railways may be similarly become obliged to adopt this technique in order to demonstrate that their safety management systems are compliance with industry best practice.

The case study demonstrated in this paper was limited to a single LOC event: worker STF, however the technique could be applied to every one of the railway's bow-tie diagrams. Such an application would have the benefit of identifying close calls that are not categorised to any threat pathway on any bow-tie. The technique lends itself to iteration on this subset of close calls in accordance with the process described in Section 3.2. Repeating the token identification process on close calls that have not been categorised would identify new tokens that may enhance the tokens sets to improve categorisation to existing bow-tie diagrams. Alternatively the tokens identified from non-categorised records may indicate the presence of hazards that are not considered in the railway's safety management system. Such an application would provide a powerful new source of information for the railway.

5.2. Allocation of close calls to threat pathways

The results show that whilst the process is relatively simple in terms of detecting the occurrence of tokens in close call reports, the process can nevertheless detect a significant number of close calls and categorise them to a corresponding threat pathway. There are two key attributes that are relevant to the categorisation of close calls to bow-tie threat pathways: *precision* and *accuracy* of the results. *Precision* refers to the granularity with which close call records can be categorised to threat pathways in the bow-tie diagram. The process used in this paper considers six threat pathways, which provides more insight for safety management than an alternative method that might categorise close calls to only the bow-tie diagram that they correspond to. In some cases it is possible to categorise a close call record to a single risk control (i.e. an individual node on the bow-tie diagram). For example close call records that report `broken handrail` and `handrail falling off` can be mapped to the individual node that shows handrails as a control to reduce STF risk. In general, however, close calls may report only the presence of hazards and not failures of specific risk controls. Some of the risk controls on the bow-tie diagram relate to organisational factors such as performance monitoring and auditing, whereas close call records generally describe immediate hazards in the workplace which are usually physical conditions. It is therefore considered that the close call records do not lend themselves to greater precision than an individual threat pathway. Greater coverage of the bow-tie might be achieved if additional data sources, such as audit reports, were included in the analysis.

Accuracy describes the degree to which the selection schema assigned records to the correct threat pathway. A validation was performed to test the level of agreement between the categorisation of a human safety analyst and this fairly straightforward NLP-based process. One hundred close call records were randomly selected and categorised by the safety analyst; this categorisation was assumed to be entirely accurate and was the baseline against which the NLP categorisation was evaluated. During this process the safety analyst noted that some of the records did not relate to worker STF hazards. The automated categorisation of the same records was then compared to the human categorisation to determine the number of correct results; the outcome is shown in Table 4.

This small sample shows a significant level of incorrect categorisation: 35% error rate against the human analyst baseline. However this error rate is not constant across the threat pathways: *human error* and *missing safety equipment* show considerable levels of disagreement, whereas *poor housekeeping* and *underfoot conditions* show better correlation. The accuracy rates compares poorly with that reported by Tixier et al. who aimed for 95% accuracy; however there is a fundamental difference between the studies in that Tixier's work sought only to identify categories of incidents from the data that were available. Conversely the study described in this paper contained an important constraint that records must be categorised to events that are described on the bow-tie diagram.

5.3. Limitations and further improvements

These errors in categorisation are likely to reflect the limitations of the thematic approach used, namely that categories are identified only on the basis of the occurrence of specific tokens in the text. The results show that *poor housekeeping* and *underfoot conditions* can be most readily characterised by a relatively simple set of tokens, whereas human error is a complex cause of hazards with a wide context that requires a complex suite of tokens for correct identification. With fewer close call records being reported in these categories there is a smaller pool of records to define a suitable set of tokens. Further refinement to the selection of tokens would be likely to lead to improved accuracy on this sample of results but may result in *overfitting* that reduces the overall accuracy of the results. Overfitting is an effect that occurs when general selection rules have been trained to correctly match a sample of data with high accuracy, but are too specific to achieve high accuracy on a general sample of results. With sufficiently careful selection of tokens, it would be possible to identify token sets that provide 100% accuracy for the tested sample. Such a specific set of tokens, however, would be extremely unlikely to be generally useful for all close calls due to the fact that they had been selected from only a small sample of records.

Another shortcoming of the thematic approach is that it is not able to efficiently detect *negation*. The thematic approach matches records based solely on the presence of tokens, regardless of the context in which the tokens occur. For example the following sentences contain exactly the same words, but have opposite meanings:

- *The road was icy and not in good condition.*
- *The road was in good condition and not icy.*

By dint of having the same words, these two sentences are equivalent in terms of thematic analysis. Moving from a thematic approach to a semantic approach for NLP of close call records could improve the accuracy of the process significantly. As well as identifying negation, a thematic approach may be valuable to synonym resolution, for example to detect where the word *foot* is used to describe a unit of measurement, and where it is used to describe a body part. Such a modification presents a significant complication to the NLP process and is an avenue of research we are currently exploring.

An important factor that influences the accuracy of the results is the degree to which the analyst identified appropriate token sets for each threat pathway, and in particular appropriated synonyms and spelling variations for important terms. This tasks is complicated by the many variations in spelling and use of abbreviations in the free-text provided by railway workers. For example analysis of close call records has found the following variations of the word *palisade*:

- `palasaid`
- `palasaide`
- `palicade`
- `palilsade`
- `palisadade`
- `paliside`
- `palistrade`
- `pallasade`
- `pallasde`
- `pallaside`

Similarly the term *British Transport Police* occurs commonly in close call records but is also abbreviated in a number of ways:

- `B.T.P.`
- `B.T.P`
- `BTP.`
- `BTP`
- `b.t.p`
- `B.t.p`
- `btp`
- `BT.Police`
- `B.T.Police`
- `BT Police`

Incorporating probabilistic techniques such as Latent Dirichlet Allocation (Blei et al., 2003) to support the selection of tokens may result in improved token sets for identification of records to a threat pathway.

The case study used in this work aggregated close calls only by the month in which the close call reported, which is consistent with the railway's existing safety monitoring processes. A further enhancement to the process could be to report geographical trends in close call reports by threat pathway, and to normalise by hours worked on particular tasks.

6. Conclusion

This paper introduces a semi-automated technique for classifying text-based close call reports in accordance with the structure of an existing bow-tie diagram and is currently being adopted by the GB railway industry. Due to the difficulty in aggregating information from the large number of records that are reported, no existing system is in place to provide classification of close calls. The technique therefore unlocks the information from close call records support safety management of the railway in a way that was not previously possible. The information obtained from the technique allows an understanding of the relative occurrence of hazard on the railway and can therefore help direct safety managers to address higher risk. The technique also allows the identification of trends in the numbers of close calls being reported and can therefore help identify where hazards are becoming more prevalent.

The limited accuracy of the results demonstrates that it would be unsafe to commit to safety management activities based solely on the results of this analysis. At this stage it is not yet clear whether the accuracy is a fundamental limit of the nature of close call reporting, however it is expected that greater accuracy can be achieved with more advanced text analysis techniques; semantic text analysis approaches are currently being trialled. Despite the limited accuracy of the results, the technique is able to provide new information that has not previously been possible. Information on trends and the relative occurrences of hazards on the railway can be valuable to safety management of the railway so long as the management processes for committing resources are away of the limitations of the results.

Despite the results being fairly coarse, the case study demonstrates that even fairly straightforward NLP processes can be an asset for detecting some hazard trends, as well as changes in those trends. Where the technique can be applied to reliably detect hazards, the process can be used to complement existing safety management activities. Whilst the paper presents a case study within the railway industry, it is expected that the technique could be generally applied across other industries. Tixer et al. have speculated use of NLP techniques for safety management may become mandatory for the construction industry. With the emergence of new tools for performing NLP, the barriers of application of the approach are being lowered and there may similarly arise an imperative for safety managers in other industries to apply similar techniques in order to demonstrate compliance with best practice.

Acknowledgements

We gratefully acknowledge RSSB for co-funding this research through the memorandum of understanding of 08 August 2013. We are also grateful to Network Rail for sharing their STF bow-tie diagram that supports this work.

References

Andriulo, S., Gnoni, M.G., 2014. Measuring the effectiveness of a near-miss management system: an application in an automotive firm supplier. Reliab. Eng. Syst. Saf. 132, 154–162.

Badreddine, A., Romdhane, T.B., HajKacem, M.A.B., Amor, N.B., 2014. A new multi-objectives approach to implement preventive and protective barriers in bow tie diagram. J. Loss Prev. Process Ind. 32, 238–253.

Bird, F.E., Germain, G.L., 1966. Damage Control: A New Horizon in Accident Prevention and Cost Improvement. American Management Association.

Blei, David M., Ng, Andrew Y., Jordan, Michael I., 2003. Latent dirichlet allocation. J. Mac. Learn. Res. 3 (4–5), 993–1022 (Lafferty, John (Ed.)).

Bliss, J.P., Rice, S., Hunt, G., Geels, K., 2014. What are close calls? A proposed taxonomy to inform risk communication research. Saf. Sci. 61, 21–28.

Cambraia, F.B., Saurin, T.A., Formoso, C.T., 2010. Identification, analysis and dissemination of near misses: a case study in the construction industry. Saf. Sci. 48, 91–99.

Cameron, I.T., Raman, R., 2005. Process Systems Risk Management. Academic Press.

Dillon, R.L., Tinsley, C.H., 2008. How near-misses influence decision making under risk: a missed opportunity for learning. Manage. Sci. 54, 1425–1440.

Gnoni, M.G., Andriulo, S., Maggio, G., Nardone, P., 2013. “Lean occupational” safety: an application for a Near-miss Management System design. Saf. Sci. 53, 96–104.

Gnoni, M.G., Lettera, G., 2012. Near-miss management systems: a methodological comparison. J. Loss Prev. Process Ind. 25, 609–616. http://dx.doi.org/10.1016/j.jlp.2012.01.005.

Heinrich, H.W., 1929. The foundation of a major injury. National Saf. News 19 (1), 9–11 59.

Hudson, P.T., 2010. Integrating organisational culture into incident analyses: extending the bow tie model. SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production. Society of Petroleum Engineers.

Hughes, P., Van Gulijk, C., Figueres-Esteban, M., 2015. Learning from text-based close call data. In: Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015. p. 8.

Jones, S., Kirchsteiger, C., Bjerke, W., 1999. The importance of near miss reporting to further improve safety performance. J. Loss Prev. Process Ind. 12, 59–67.

Lateiner, A., 1958. If we’re to stop accidents, preventing injuries is not enough!. Ind. Supervisor 26 (11), 3–5 14.

Li, J., 201. Domain specific commonsense relation extraction from bag of concepts metadata. In: Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, January. ACM, p. 96.

Manning, C.D., Schütze, H., 1999. Foundations of Statistical Natural Language Processing. MIT Press.

Mou, L., Meng, Z., Yan, R., Li, G., Xu, Y., Zhang, L., Jin, Z., 2016. How Transferable are Neural Networks in NLP Applications?. arXiv preprint arXiv:1603.06111.

Nielsen, D.S., 1971. The cause/consequences diagram method as basis for quantitative accident analysis. Danish Atomic Energy Commission, Risoe. Research Establishment.

Noel, G.E., Peterson, G.L., 2014. Applicability of latent dirichlet allocation to multi-disk search. Digital Invest. 11 (1), 43–56.

Petersen, D.C., 1971. Techniques of Safety Management. McGraw-Hill Book Company, New York, 250pages, 57 references; 1971.

Popping, R., 2000. Computer-Assisted Text Analysis. SAGE Publication.

RSSB, 2015. Annual Safety Performance Report 2015/16 Half-year update.

Ruiz-Martínez, J.M., Valencia-García, R., Fernández-Breis, J.T., García-Sánchez, F., Martínez-Béjar, R., 2011. Ontology learning from biomedical natural language documents using UMLS. Expert Syst. Appl. 38 (10), 12365–12378.

Sánchez Ruenes, D., 2007. Domain Ontology learning from the Web.

Targoutzidis, A., 2010. Incorporating human factors into a simplified “bow-tie” approach for workplace risk assessment. Saf. Sci. 48 (2), 145–156.

Thompson, K., 1968. Programming techniques: Regular expression search algorithm. Commun. ACM 11 (6), 419–422.

Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B., Bowman, D., 2016. Automated content analysis for construction safety: a natural language processing system to extract precursors and outcomes from unstructured injury reports. Autom. Constr. 62, 45–56.

Turner, C., Hamilton, W.I., Bradler, I., 2015. Bowtie diagrams: a user-friendly risk communication tool - contractor management case study. In: Fifth International Rail Human Factors Conference, pp. 735–749.

Van Gulijk, C., Hughes, P., Figueres-Esteban, M., 2016. The potential of ontology for safety and risk analysis. In: Proceedings of ESREL 2016. CRC Press, Chicago.

Visser, J.P., 1998. Developments in HSE management in oil and gas exploration and production. In: Safety Management: The Challenge of Change. pp.43–66.

Wright, L., Van der Schaaf, T., 2004. Accident versus near miss causation: a critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors. J. Hazard. Mater. 111 (1), 105–110.

Xu, H., Stenner, S.P., Doan, S., Johnson, K.B., Waitman, L.R., Denny, J.C., 2010. MedEx: a medication information extraction system for clinical narratives. J. Am. Med. Inform. Assoc. 17 (1), 19–24.