A data-driven narratives skeleton pattern recognition from accident reports dataset for human-and-organizational-factors analysis

Journal of Loss Prevention in the Process Industries 83 (2023) 105047

Contents lists available at ScienceDirect

Journal of Loss Prevention in the Process Industries

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

A data-driven narratives skeleton pattern recognition from accident reports dataset for human-and-organizational-factors analysis

Shuo Yang [*], Micaela Demichela

Applied Science and Technology Department Politecnico di Torino, Turin, Italy

ARTICLE INFO

Keywords:
Human and organizational factors
Human reliability analysis
Natural language processing

ABSTRACT

Accidents in the process industry involve several interacting factors, including human and organizational factors (HOFs). A long-standing obstacle to HOFs analysis is lack of data. Accident reports are an essential data source to learn from the past and contain HOFs-related data, but they are usually unstructured text in a not standardized format. Some studies have explored the extraction of information automatically from accident reports based on Natural Language Processing (NLP) techniques. However, they were not dedicated to HOFs. Risk communication is considered an essential pillar in safety and risk science. This research develops a HOFs-focused risk communication framework based on the NLP techniques that can support risk assessment and mitigation. The proposed approach automatically extracts the target groups oriented "Who, When, Where, Why" (4Ws) information from accident reports.

This framework was applied to explore the eMARS database. The results show that the "4Ws" skeleton of narratives has appreciated performance in pattern recognition and holistic information analysis. The graphical representation interfaces are designed to display the features of HOFs-related accidents, which can better be communicated to the sharp-end operators and decision-makers.

1. Introduction

Potentially involving large amounts of hazardous materials, accidents in the process industry can lead to severe social, economic, and environmental consequences. A process plant can be regarded as a complex social-technical system. Thus, accidents in the process industry involve multidimensional interacting factors, including technical, human and organizational factors (Hollnagel, 1998). Previous studies of major accidents and disasters in chemical process industries confirm that more than 80% of accidents have been caused by human errors (Zarei et al., 2021). Data analysis of the eMARS database showed that HOFs contributed to about 47% of hazardous material-related accidents in the process industry domain (Yang et al., 2022). With the development of technology, the reliability of technical equipment and components has been significantly raised. In contrast, the human and organizational factors (HOFs), despite the opportunities offered by technological development as the increased monitoring capabilities given by the industry 4.0 domain, require further work (Olivier Fontaine et al., 2016).

It is widely recognized that HOFs have a vital contribution to adverse events and influence the performance of socio-technical systems (Accou and Carpinelli, 2022). The knowledge about HOFs started within the field of ergonomics (Skogdalen and Vinnem, 2011). Human and organizational factors sometimes are named human factors for short. The HSE definition of HOFs/human factors is "environmental, organizational, and job factors, and human and individual characteristics which influence behaviour at work in a way which can affect health and safety" (Reducing error and influencing behaviour, 1999). At the same time, the organization is characterized by the division of tasks, the design of job positions, including selection and training and cultural indoctrination, and their coordination to accomplish the activities(Bellamy et al., 2008).

Various analysis frameworks have been proposed to analyze the human factors related causality of accidents: the Human Error Assessment and Reduction Technique (HEART) (Kirwan, 1994), Analytical Hierarchy Process (AHP) (Bhushan and Rai, 2004), Fault Tree Analysis (FTA) and Human Factor Analysis and Classification System (HFACS) (Khan et al., 2022). But a long-existing obstacle on the way of HOFs-related research is lack of data to test and validate. The primary way to get relevant data should be by analyzing previously occurred events through survey questionnaires or retrieving them directly from

* Corresponding author.
E-mail address: shuo.yang@polito.it (S. Yang).

https://doi.org/10.1016/j.jlp.2023.105047
Received 30 December 2022; Received in revised form 10 March 2023; Accepted 30 March 2023
Available online 8 April 2023
0950-4230/

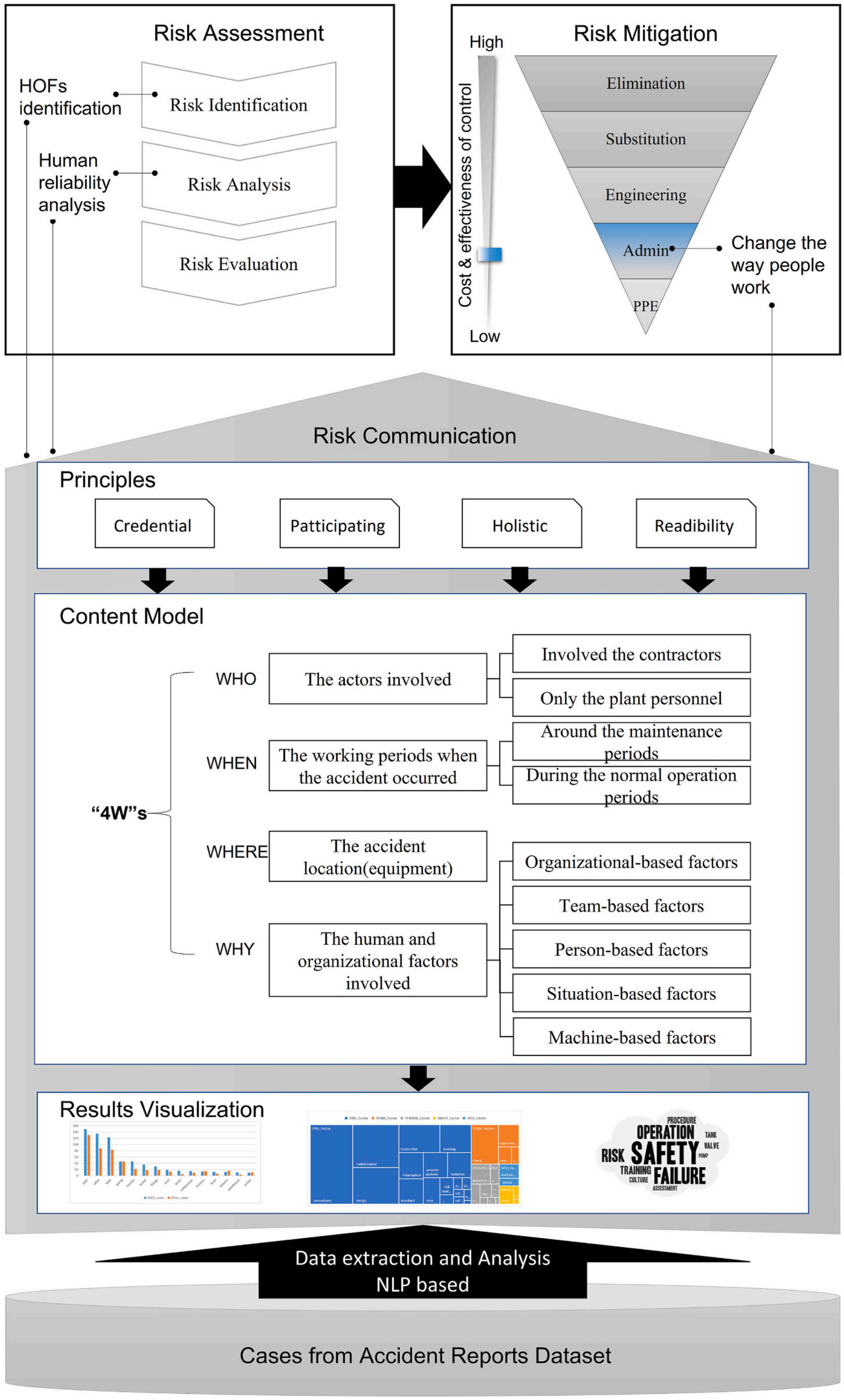

Fig. 1. The framework for the data-driven narratives skeleton pattern recognition.

the operators, which may introduce subjective bias both in the tools design and data collection phase. This study, in particular, will adopt accident reports to mine relevant data.

Accident reports are the primary data sources for high-hazard organizations to learn from themselves and others' experiences instead of managing safety on a trial-and-error basis. Many international and regional organizations have built accident report databases. The process industry shares data via Major Accident Reporting System (eMARS), Analyze, Recherche et Information sur les Accidents (ARIA), ZEMA, and Chemical and Safety Hazard Investigation Board (CSB). This information can provide collaborative learning opportunities across industrial companies, subcontractors, labour representatives, regulators and inspectors, legislators and interested public members. After more than twenty years of accumulation, many accident reports data are available for analysis. But the accident descriptions, although they contain detailed information and are often unstructured text.

The traditional way to extract critical information from these descriptions is by manually coding. This process could be time-consuming and error-prone, requiring knowledge of both process industry and human factors. Therefore, some studies have been done to develop tools

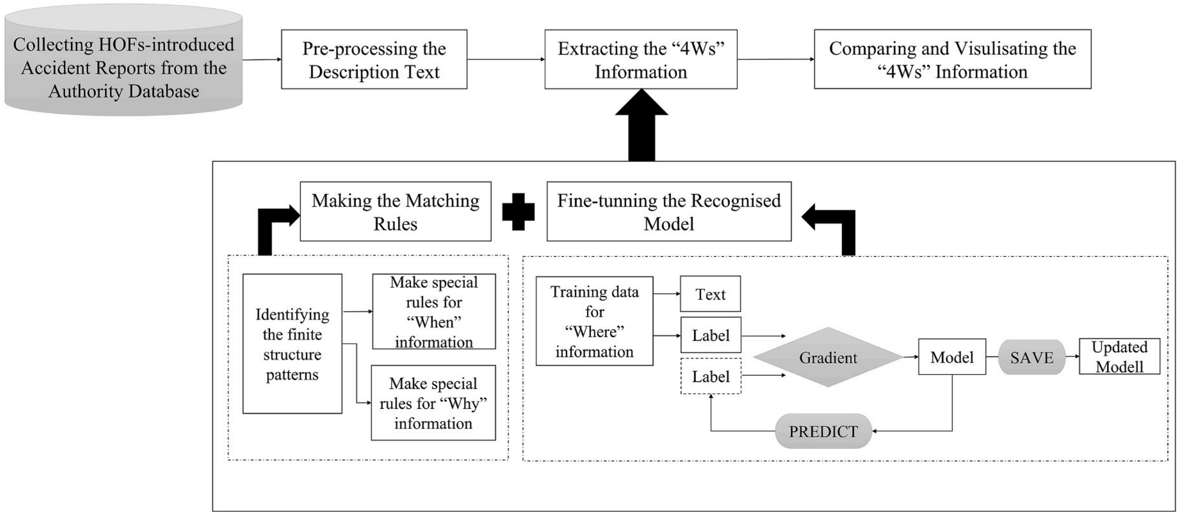

Fig. 2. The methodology of HOFs-focused data analysis framework.

based on Natural Language Processing (NLP) to extract this information automatically. Suh (2021) used rule-based text mining and latent Dirichlet allocation algorithms to identify the sectoral patterns of the accident process. Perboli (2021) developed the Software Hardware Environment Liveware accident causality model, utilizing machine learning techniques. Single et al. (2020) developed a custom tag-based pattern recognition to construct an ontology structure to populate the knowledge automatically. These studies mainly utilize the tag-based or rule-based matching approaches. Difficulties arise when an attempt is made to implement these approaches in more generic scenarios.

A data-driven methodology was proposed for empirical study from the accident report dataset to address the gaps mentioned above. A demonstrative study was given to illustrate the effectiveness of the proposed approach. The research questions guiding this study were ：

RQ1.What is the skeleton of narratives of an accident description?

RQ2.What are the unique features of HOFs-related accidents?

RQ3.How to deliver friendly risk communication to sharp-end operators and decision-makers?

Section 2 presents a theoretical background for the data-driven narratives skeleton pattern recognition from accident reports dataset for HOFs; Section 3 describes the methodology for applying the selective NLP technologies according to the different natures of HOFs-related accident features. Section 4 shows the results of applying this methodology to the eMARS databases for validation. The eMARS database is chosen because it is open access, also the database has explicitly addressed HOFs among the causes' classifiers, which make it possible to better select the target cases. Section 5 discusses the key findings, and section 6 leads to conclusions.

2. Theoretical background

Fig. 1 shows the framework of the data-driven narratives skeleton pattern recognition. Accident reports could be regarded as many narratives, and this research seeks to extract its skeleton. The communication domain has a "5Ws and 1H" frame, including Who, What, When, Where, Why and How in news writing. This structure could orient audiences toward their communal environment and help link audiences with the environment transcending their limited sensory experiences and ensuring that information meets the audience's needs (Pan and Kosicki, 1993). The target group of HOFs-focused risk communication are sharp-end operators and decision-makers. According to Mallam et al. (2022), Sharp-end operators tend to have more interest in human factors-related issues than the technical relevant information because human factors content is relatable to them and their professional identity and culture. To extract the relevant information from accident records, the "4Ws" HOFs information model is built as the generic story tell frame with "Who, When, Where, Why". The two unique features of HOFs-related accident in this research is based on two empirical hypotheses that can be listed as follows: Contractors need special attention, for their interface with the owner is crucial for process safety (Tamim et al., 2017). Maintenance operations, emergency operations, and control room operations are critical operations in hazardous process systems where human interference has enormous potential for human error (Zarei et al., 2021). For the "Who" information, the concern is whether the accident was related to contract or internal personnel. For the "When" information, the concern is whether the accident was around the maintenance periods. The concern for the "where" information is the relevant equipment(locations) the accident occurred.

For the most critical part, "Why" the human reliability analysis (HRA) methods such as THERP(Swain and Guttmann, 1983), CREAM (Hollnagel, 1998), SPAR-H(Gertman et al., 2004), and HEART were considered. They use performance shaping factors (PSFs) to represent the aspects of the human-system interaction. There is still no standard PSFs set. Groth and Mosleh (2012) introduced a hierarchical set of performance influencing factors (PIFs) that have clearly defined units of analysis, such as organizational-based, team-based, personnel-based, situation-based, and machine-based, with a corresponding set of behaviours and metrics that are visible indicators of invisible PIFs. This set of PIFs is selected because it is clearly defined, data-informed and orthogonal. Also, all the PIFs in this set had been tested in accident datasets.

3. Methodology

3.1. Proposed approach

This research begins by collecting the HOFs-related accident reports from the relevant dataset. The logical steps are described, as shown in Fig. 2. At first, NLP techniques are used to pre-process the description text to tokens through the SpaCy package (Honnibal & Montani., 2017) under the Python programming platform. The tokens in the doc can be selected, transformed, and analyzed. After that, the "4Ws" information extraction through rules or fine-tuned models was detailed. Finally, the analysis results are compared and visualized.

3.2. Text mining algorithm

In this research two types of NLP techniques are utilized for text

Table 1
Organization-based factors terms.

Single Pattern Tag		Double Pattern Tag				Triple Pattern Tag					
Text	tag	text		Tag		text			tag		
Training	TRA	Corrective	Action	COR	ACT	Management	of	Change	MANAG	OF	CHAN
Culture	CUL	Workplace	Adequacy	WOP	ADEQ						
Staffing	STA	Outsourcing	Management	OUT	MANAG						
Scheduling	SCHE	Permit	Management	PERM	MANAG						
Procedures	PROD	Process	Analysis	PROS	ANS						
Tools	TOOL	Risk	Analysis	RIS	ANS						
Information	INFO										
Design	DESN										

Table 2
Team-based factors terms.

Single Pattern Tag		Double Pattern Tag			
text	tag	text		tag	
Communication	COM	Role	Awareness	ROL	AWAR
Coordination	COO				
Cohesion	COH				
Supervision	SUP				

Table 3
Person-based factors terms.

Single Pattern Tag		Double Pattern Tag			
text	tag	text		tag	
Attention	ATTE	Sensory	Limits	SENS	LIM
Alertness	ALE				
Fatigue	FATI				
Impairment	IMPA				
Knowledge	KNOW				
Experience	EXPE				
Skills	SKIL				
Bias	BIA				
Morale	MOR				
Motivation	MOTI				
Attitude	ATTI				
Familiarity	FAM				

mining, a fine-tuned model training method, and rule-based matching method. The methods are chosen according to the expression attributes of process industry accident information. Training a fine-tuned model is helpful for the "Where" information because of the wide potentially involved set of equipment. While Rule-based systems are a good choice for the "When" and "Why" information since there is a finite number of examples in these categories with a straightforward, structured pattern that can be expressed with token rules or regular expressions for the "Who" information. The actors are divided into employed operators and contract operators, and this information can be extracted from the dataset directly.

3.2.1. Matching rules making

According to the process operating practice, the working periods are divided into two categories "operational periods and maintenance periods". So, for the "When" information, the critical step is to summarize the critical expression identifying a maintenance operation as: shutting down time, cleaning/purging time, repair time, solder/welding time and isolation time". To identify all the maintenance periods, like "during, before, and after" maintenance periods, the inclusive match patterns are built using token patterns with one dictionary describing one token (list), e.g.:

{'LEMMA': {'IN': ['maintenance', 'clean', 'repair', 'shut', 'weld', 'solder', 'hot']}},{'LOWER':{'IN':['routine', 'out']},' OP':'?'}, {'LOWER':{'IN':['work', 'operation']},' OP':'?'}

Table 4
Situation/Stress-based factors terms.

Single Pattern Tag		Double Pattern Tag			
text	tag	text		tag	
Environment	ENV	Condition	Events	CON	EVE
Stress	STRE	Task	Load	TS	LOD
		Time	Load	TI	LOD
		Task	Complexity	TS	COMP
		Perceived	Situation	PERC	SITU
		Perceived	Decision	PERC	DECI

Table 5
Machine-based factors terms.

Triple Pattern Tag					
text			tag		
Human	Machine	Interface	H	M	I
Human	System	Interface	H	S	I

For the "Why" information. This research selects the set of PIFs developed by Katrina M. Groth and Ali Mosleh (M. Groth & Mosleh, 2012). Based on this set of PIFs, the management activities were extended to outsourcing management, permit management, management of change, process analysis, and risk analysis. Then an extended set of terms categories is developed, as shown in Tables 1–5.

3.3. Fine-tuned model raining

For the "Where" information, the equipment(location) terms can be very wide. This means that using a rule-based method will lose some useful data. Employed the annotation tool Prodigy, this research employs the custom fine-tuned model to mix the rule-based and statistic models. First, words like "pipe, tank, pump" are used as seeds to generate a worklist pattern. Then the texts of the "Cause of accident" column from the database have been used to teach and correct the automatic annotation through Prodigy tools. In the end, the pre-trained English Language model in SpaCy has been combined with our fine-tuned one to form a new name entities recognition model. Finally, 514 entities are used as training data, and 236 entities (30% split) are used as evaluation data. The precision, recall, and F1-score are employed to evaluate NER model, as discussed in the following section.

4. Results

4.1. Collecting HOFs-related accident reports from the eMARS database

The cases from the chosen dataset are initially filtered with the causal factors to differentiate the HOFs-related accidents or not. First, the raw accident cases from eMARS are filtered with "cause of the accident". Only the HOFs-related cases have been collected. Meanwhile, this research has not considered the malicious and no/too simple cause

Table 6
The selection of the case from the eMARS database.

Total	Human error cases	Organizational factor cases	Redundant cases	No/too simple cause description cases	Other cases	Final HOFs cases
1128	209	464	142	42	639	489

Table 7
The SpaCy Tokenized doc.

TEXT	LEMMA	POS	TAG	DEP	SHAPE	ALPHA	STOP
A	a	DET	DT	det	X	TRUE	TRUE
leakage	leakage	NOUN	NN	nsubj	xxxx	TRUE	FALSE
in	in	ADP	IN	prep	xx	TRUE	TRUE
a	a	DET	DT	det	x	TRUE	TRUE
pipeline	pipeline	NOUN	NN	pobj	xxxx	TRUE	FALSE
caused	cause	VERB	VBD	ROOT	xxxx	TRUE	FALSE
the	the	DET	DT	det	xxx	TRUE	TRUE
release	release	NOUN	NN	dobj	xxxx	TRUE	FALSE
of	of	ADP	IN	prep	xx	TRUE	TRUE
chlorine	chlorine	NOUN	NN	pobj	xxxx	TRUE	FALSE
.	.	PUNCT	.	punct	.	FALSE	FALSE

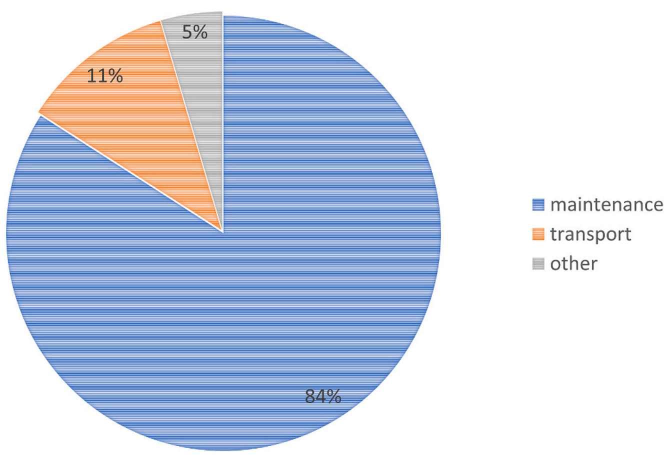

Fig. 3. The accident involved contractors.

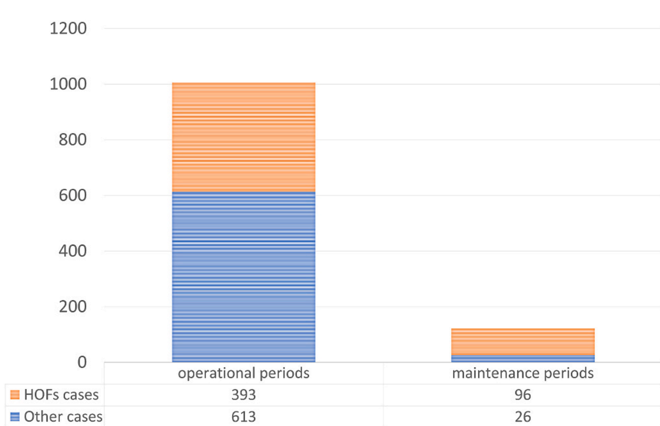

Fig. 4. The accident involved working periods.

description cases. The primary statistic description of the database has shown in Table 6. Then from the 73 columns of the dataset, the raw texts from "accident description", "causes of the accident", and "lesson learned" are selected.

4.2. Pre-process

For the pre-processing part, the raw texts are tokenized in SpaCy's built-in pipelines and become 'docs', e.g., the sentence "A leakage in a pipeline caused the release of chlorine." the 'doc' result, as Table 7 shows. Then using the "Lemmatizer component" to remove the stop word, to get the cleaned words data as "leakage pipeline caused release chlorine".

4.3. "4Ws" information extraction and visualization

As for the "Who" information, 44 cases involved contract operators (see Fig. 3). In 37 cases, contract operators performed maintenance, including hot work, cleaning, repair, and replacement. In 5 cases they carried out transport services. Moreover, 2 cases worked in routine operations.

With respect to the "When" information, nearly three times HOFs cases happened during the maintenance periods compared to other cases. About 20% of the HOFs case, while only 7% of other cases occurred during maintenance periods, as shown in Fig. 4. This result validates human errors that often occur around maintenance activities. Moreover, many accidents happen because of inadequate procedures and instructions about maintenance work. Also related to rare or unusual activities (Baldissone et al., 2019). The decision-maker should pay more attention to the maintenance working periods when making risk reduction policies and allocating safety control resources.

As for the "Where" information, more than 1/2 of the HOFs-related cases happened around the reactor compared to other cases. More than 1/3 of the HOFs-related cases happened around the valve compared to other cases. They were followed by "tank" and "pipe" locations, as shown in Fig. 5.

As for the "Why" information, after extracting and analyzing, 48 cases have three factors identified, 91 cases have two factors identified, 146 cases have one factor identified, and 167 cases have no factors identified. More than 2/3 of the identified factors are organizational factors. The occurrence frequencies of the factors are expressed as the area of the rectangular, and the color indicate the different types of factors, then the hierarchical data of "Why" factors is visualized as a tree map is shown in Fig. 6.

4.4. Performance evaluation

As introduced above, the performance of the model for the "where" identification was accessed through the F1-score. The precision is calculated using the number of true positive results divided by the

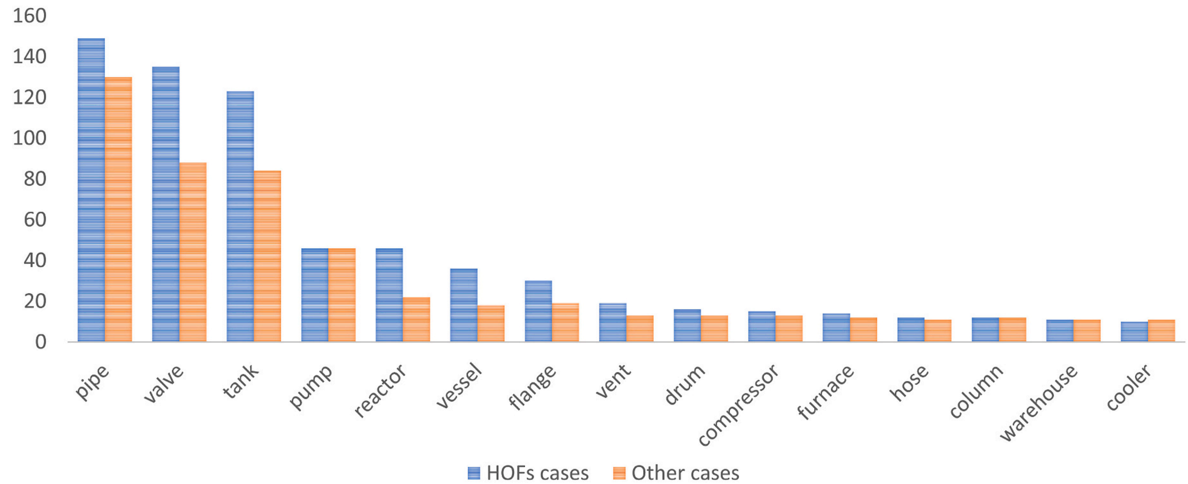

Fig. 5. Identified equipment(locations) frequencies.

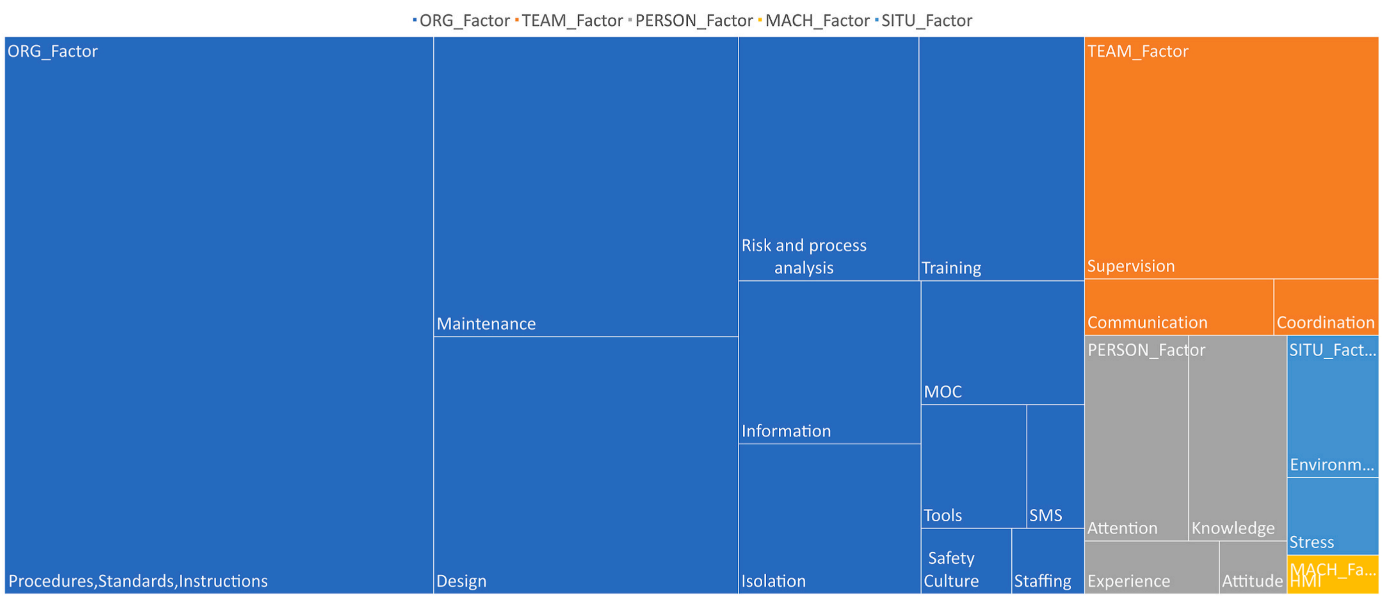

Fig. 6. Identified HOFs tree map.

number of all positive results, including those not identified correctly. The recall is calculated using the number of true positive results divided by the number of all samples that should have been identified as positive. The F1-score of a classification model is combined the precision and recall of a classifier into a metric(Van Rijsbergen, 1974), calculated as follows:

$$\frac{2(P*R)}{P+R} \quad \text{Eq (1)}$$

where, P is the precision, R is the recall of the classification.

The performance evaluation of the "Where" information identified model shows good results, with a precision of 93.00, a recall of 93.94, and an F1-score of 93.47.

5. Discussion

With the development of NLP technology, several scholars have contributed to applying it to extract information from accident reports, especially single at.(2020) developed an orthogonal knowledge pattern to extract accident information. However, few studies use model-based NLP technologies to deal with the item with infinite expressions, such as chemical equipment. The current study integrated the risk communication framework with HOFs theories and designed the "4Ws" skeleton of narratives to lead the pattern recognition. The methodology was then applied to explore the eMARS database.

The storytelling style has been proven to affect risk communication positively. However, this process may lose some authority at the same time. This research initially finds a way to balance the benefits and deficiencies of narratives to employ the simple version of storytelling. Therefore, the "4Ws" information model, modified based on the "5Ws1H" theory as the standard pattern of the accident descriptions from accident reports. This pattern works well for skeleton extraction for both HOFs-related accidents and other accidents. With the help of this pattern, information is categorized smoothly, and the graphical representation visualization tools work efficiently.

The results show significant differences for the "Who" and "When" information, when dealing with HOFs related cases, with respect to purely technical related ones. Tamim et al. (2017) discussed the critical roles of engineering contractors in managing risk throughout a plant life cycle based on the investigation of nine major process safety events. Zarei et al. (2021) summed up the three critical operations for process safety: maintenance, control room, and emergency operations. In this

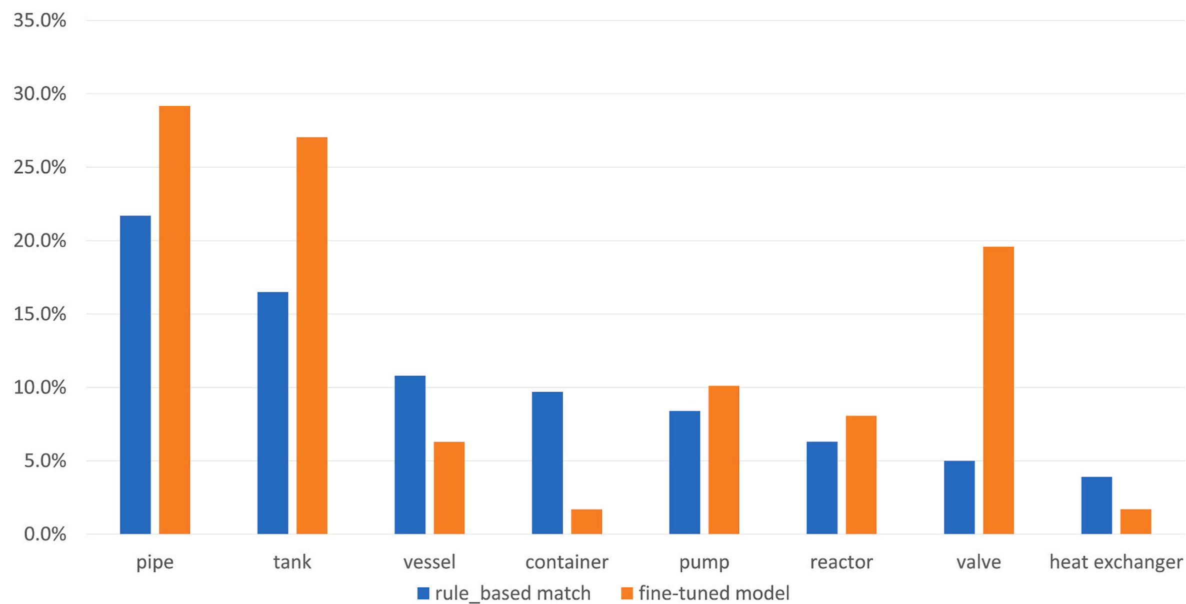

Fig. 7. Identified equipment(locations) frequencies comparison between two methods.

research, HOFs-related accidents are numerically statistically proven to be more related to contractors and tend to be around the maintenance periods, whereas other scenario shows fewer occurrences. These evidence-based quantified features can be conveyed to the sharp-end operator and decision-maker more effectively to relocate the safety resource and the safety management design.

For the "Where" information extraction, Single et al. (2020) conducted a customed tags-based NLP match analysis of the main categories of the eMARS database equipment. From 889 cases, 77 locations were identified, only 8.7%. To find a more generic matching method, we trained a fine-tuned model to recognize the critical equipment for the accident automatically. Compared to the result of research of Single at. (2020), the proposed method recognized much more equipment information. Because many cases involve more than one kind of equipment (locations), a total of 1728 equipment entities are identified from 1128 cases. Moreover, Fig. 7 compares the results of identified equipment (locations) frequencies between the two studies. The result validated that the fine-tuned model can better identify various equipment.

As for the "Why" information, over 2/3 of the identified factors are organizational factors. On the one hand, this result proves the dominant position of organizational factors, which could be the root cause of human factors and technical factors. On the other hand, this implies that NLP technologies may not work well for recognizing human character and psychology-related factors, which are hard to express in accident descriptions or less care for the hardness of validating evidence during the accident investigation.

Many HRA studies stop on the way to exploring the causality of HOFs, but the intervention and loss prevention parts also need attention. This research aims to promote HOFs-focused risk communication to introduce behavioral change. Despite this would be a future development based on the actual work, some consideration can already be given. As a form of communication tools graphs and diagrams have often played an essential role in understanding complex phenomena and discovering laws and explanations (Friendly & Wainer, 2021). Thus, to deliver friendly risk communication both to the sharp-end operators and decision-makers, the critical features of HOFs-related accidents have to be shared through quantitative graphical representation. This work will be a prime foundation for taking advantage of the edge technology of data visualization. This research designed graphical results display interface, including the word clouds and bar charts to enhance the risk communication phase. That will be tailored to different receivers and extend in its effectiveness in future work.

6. Conclusions

Previous studies have shown that applying NLP technologies to extract information from accident reports may be an effective strategy to improve the shortcomings of traditional manual coding. This study demonstrated the usefulness of a "4Ws" narrative skeleton to combine the fine-tuned model with rule-based matching technologies as a tool for accident report information extraction. The proposed skeleton narratives pattern allows an automatic NLP algorithm transition from the unstructured text to category data, which can be a quantitative representation of accident descriptions. Based on this, abundant machine learning and data analysis can dive deep. The results can be communicated in graphical forms to support sharp-end behaviour change and identify priorities to support decisions. Overall, this study contributes to the safety management field as follows:

- The proposed framework utilizes the "4Ws" information model as an information skeleton to obtain underlying patterns among the accident reports data. Such a model-based and data-driven manner enables automatic and efficient information identification without human intervention, reducing human efforts.
- This research brings insights into attributes of HOFs that contributed to accidents, especially the quantity data about HOFs factors, which can be a foundation for future work on HRA.
- The "4Ws" skeleton works well in pattern recognition tasks. It shows potential to be a foundation for the machine learning data analysis innovative technology framework to the intelligent analysis accident report.
- The proposed framework is generic and can be applied to the newest data from the eMARS and other accident report databases.

Although this study compensates for drawbacks in the tag-based and rule-based NLP technologies and applies the methodology to explore the HOFs-related accident features. Some limitations can be further explored. Focusing on the data extraction and results visualization parts, this study did not explore the evaluation module in the proposed framework. Future work can be based on the provided principles to conduct the relevant feedback survey. Along with the human reliability research line, further research can be done based on the extracted data to test and compare the HRA models or explore the data-driven dependencies of HOFs on PSFs to improve the estimation of the failure rate of human activities to aid decision-making.

Author contribution statement

Shuo Yang: Conceptualization, Writing-original draft, Methodology, Project administration. **Micaela Demichela:** Writing-review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shuo yang reports financial support was provided by China Scholarship Council. Shuo yang reports a relationship with China Scholarship Council that includes: funding grants.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the China Scholarship Council (grant number No.202009110127).

References

Accou, B., Carpinelli, F., 2022. Systematically investigating human and organisational factors in complex socio-technical systems by using the "SAfety FRactal ANalysis" method. Appl. Ergon. 100, 103662 https://doi.org/10.1016/j.apergo.2021.103662.

Baldissone, G., Comberti, L., Bosca, S., Murè, S., 2019. The analysis and management of unsafe acts and unsafe conditions. Data collection and analysis. Saf. Sci. 119, 240–251. https://doi.org/10.1016/j.ssci.2018.10.006.

Bellamy, L.J., Geyer, T.A.W., Wilkinson, J., 2008. Development of a functional model which integrates human factors, safety management systems and wider organisational issues. Saf. Sci. 46 (3), 461–492. https://doi.org/10.1016/j.ssci.2006.08.019.

Bhushan, N., Rai, K., 2004. The analytic hierarchy process. In: Bhushan, N., Rai, K. (Eds.), Strategic Decision Making: Applying the Analytic Hierarchy Process. Springer, pp. 11–21. https://doi.org/10.1007/978-1-85233-864-0_2.

Fontaine, Olivier, Martinetti, Alberto, Sofia Michaelides Mateou, 2016. Remote pilot aircraft system (rpas): just culture, human factors and learnt lessons. Chemical Engineering Transactions 53, 205–210. https://doi.org/10.3303/CET1653035.

Gertman, D., Blackman, H., Marble, J., Smith, C., Boring, R., 2004. The SPAR-H Human Reliability Analysis Method.

Groth, K.M., Mosleh, A., 2012. A data-informed PIF hierarchy for model-based Human Reliability Analysis. Reliab. Eng. Syst. Saf. 108, 154–174. https://doi.org/10.1016/j.ress.2012.08.006.

Hollnagel, E., 1998. Cognitive Reliability and Error Analysis Method: CREAM, first ed. Elsevier.

Khan, R.U., Yin, J., Mustafa, F.S., Ahmad Farea, A.O., 2022. A data centered human factor analysis approach for hazardous cargo accidents in a port environment. J. Loss Prev. Process. Ind. 75 https://doi.org/10.1016/j.jlp.2021.104711. Scopus.

Kirwan, B., 1994. A Guide to Practical Human Reliability Assessment. CRC Press. https://doi.org/10.1201/9781315136349.

Mallam, S.C., Marit Wahl, A., Aas, J., 2022. Reintroducing the sharp-end operator to organizational learning: How accident reports are used by maritime officers. Saf. Sci. 147, 105632 https://doi.org/10.1016/j.ssci.2021.105632.

Pan, Z., Kosicki, G., 1993. Framing analysis: an approach to news discourse. Polit. Commun. 10 (1), 55–75. https://doi.org/10.1080/10584609.1993.9962963.

Perboli, G., 2021. Natural Language Processing for the identification of Human factors in aviation accidents causes: an application to the SHEL methodology. Expert Syst. Appl. 7.

Reducing Error and Influencing Behaviour, 1999. HSE Books.

Single, J.I., Schmidt, J., Denecke, J., 2020. Knowledge acquisition from chemical accident databases using an ontology-based method and natural language processing. Saf. Sci. 129, 104747 https://doi.org/10.1016/j.ssci.2020.104747.

Skogdalen, J.E., Vinnem, J.E., 2011. Quantitative risk analysis offshoreHuman and organizational factors. Reliab. Eng. Syst. Saf. 96 (4), 468–479.

Suh, Y., 2021. Sectoral patterns of accident process for occupational safety using narrative texts of OSHA database. Saf. Sci. 142, 105363 https://doi.org/10.1016/j.ssci.2021.105363.

Swain, A.D., Guttmann, H.E., 1983. Handbook of human-reliability analysis with emphasis on nuclear power plant applications. *Final report* (NUREG/CR-1278. https://doi.org/10.2172/5752058. SAND-80-0200, 5752058; p. NUREG/CR-1278, SAND-80-0200, 5752058).

Tamim, N., Scott, S., Zhu, W., Koirala, Y., Mannan, M.S., 2017. Roles of contractors in process safety. J. Loss Prev. Process. Ind. 48, 358–366. https://doi.org/10.1016/j.jlp.2017.04.023. Scopus.

Van Rijsbergen, C.J., 1974. Foundation of evaluation. J. Doc. 30 (4), 365–373. https://doi.org/10.1108/eb026584.

Yang, Shuo, Micaela, Demichela, Geng, Jie, Tao, Hongjie, 2022. Analysis of human and organization factors related accident reports based on natural language processing. Chem. Eng. Trans. 90, 745–750. https://doi.org/10.3303/CET2290125.

Zarei, E., Khan, F., Abbassi, R., 2021. Importance of human reliability in process operation: a critical analysis. Reliab. Eng. Syst. Saf. 211, 107607 https://doi.org/10.1016/j.ress.2021.107607.