Analysis research of industry accident law based on barrier theory and text semantic recognition technology

Hongqiao Yan\textsuperscript{1}, Kun Tian\textsuperscript{1}, Zhiqiang Fan\textsuperscript{1} and Yaming Mao\textsuperscript{1}

\textsuperscript{1}CNPC Research Institute of Safety and Environmental Technology 102206, Building 1, Yard 1, Yellow River Street, Changping District, Beijing, China

E-mail: yanhongqiao@cnpc.com.cn

Abstract. One of the important reasons for the repeated occurrence of the same type of accidents in the group company over the years is the insufficient utilization of accident resources. This paper introduces barrier theory and establishes an accident analysis method system based on barrier theory. Barrier theory changes the traditional accident analysis mode, and adjusts the subject of accident analysis from consequence to barrier, which provides a new perspective for the analysis of accident law. This paper develops a HSE text semantic recognition model which can automatically extract the production operation link, production stage and type of failed barriers of the accidents from the accident investigation report, finds out the cause factors leading to the failures of these barriers and establishes a direct relationship with the HSE(Health, Safety and Environment) management system, in a way to find out some common laws of the accidents and reveal the defects and shortcomings in business management.

1. Introduction

At present, with the vigorous promotion of the HSE management system in recent years, enterprises have gradually improved their understanding on accident events. However, in the management of accident events, the same problem is repeated and similar accident events occur repeatedly. One of the important reasons is the lack of understanding of the resources of accident events and the insufficient utilization of their value. Therefore, it is necessary to establish a scientific concept of accident event resources and use scientific and systematic analysis methods to study the mechanism and law of the occurrence of accident events, find out the root causes of enterprises in institutional system and management process, effectively control risks, improve early warning management capabilities and reduce the accident rate, in order to help the group company keep improving its safe production situation.

The accident investigation results of the Norwegian Petroleum Safety Authority clearly show that the weakness or failure of safety barriers is a common factor for safety accidents. Barrier analysis and barrier management are now key factors in critical hazard management. Especially after Macondo and Montara, the industry and regulatory authorities are increasingly focusing on barriers to ensure they are identified, in place and operated on demand. The experience of the major oil industry accidents at home and abroad in recent years shows that many accidents occur because of the failure or weakening of barriers [1]. Based on Swiss cheese model and barrier theory, this paper takes some fatal accidents amongst industrial production safety accidents as the data basis. Through description of text data (i.e. accident process, direct cause, indirect cause and management...
cause) and adoption of barrier theory to identify relevant human barriers and hardware barriers and their defects, this paper uses data to further identify the causes of these barrier defects and the management factors causing these defects, as well as analyzes the law of occurrence of the fatal accidents and typical similar fatal accidents on this basis. The research of this paper is of great significance in revealing the inherent law of accidents, finding out management defects and preventing the occurrence of major accidents.

2. Barrier theory

2.1. Swiss cheese model

Accident-causing theory is the basis and guideline for modern HSE management [2]. According to literature research and the general practice of the international oil and natural gas industry, Swiss cheese model, which is the current main accident-causing theory, explicitly proposes the accident-causing theory in Figure 1.

Hazard prevention and control theory is the basis and guideline for modern HSE management [2]. According to literature research and the general practice of the international oil and natural gas industry, Swiss cheese model, which is the current main accident-causing theory, explicitly proposes the accident-causing theory in Figure 1.

Figure 1. Accident-causing theory based on barriers.

Hazards are objective and related to energy and production or operations. Defences, also known as barriers, can be flawed at any time [3]. An accident or event will occur when a series of events occurs in a specific order after passing through all barriers (or defences) in turn. Losses, accidents or events: To find out the law and reveal the defects of HSE risk management through systematic analysis of accidents and events, and in turn, to increase barriers, make up for voids and improve hazard identification, thus systematically improving HSE management.

Under the guidance of the cheese model [4]: 1) systematically identify all hazards, 2) systematically find out all barriers in the organization, and 3) understand, explore and enhance the failure mechanism of every barrier, forming the key point for modern HSE risk management [5].

2.2. Barrier theory

A barrier is a measure that prevents a hazard from further evolving into an accident at an early stage of the accident. It can limit the harm and consequences of the hazard [6]. The barrier itself should be able to stop certain scenarios, causes or consequences without the help of other barriers, equipment or tasks. A scientific accident prevention and control model is based on the objective law of accident prevention and control, and sets corresponding barriers according to the characteristics of hazards to carry out targeted prevention and control [7].

Figure 2 shows the basic concept of the barrier model. The left side shows hazards (including the energy that can cause accidents), followed by hardware and human barriers to prevent hazards from developing into an undesired or out-of-control event. Various hardware and human barriers would mitigate the ultimate consequences [8]. Every barrier, if intact, should be able to completely prevent further development of the event. Meanwhile, human and hardware barriers are affected by various HSE management system elements. At the bottom of the management system are leadership and safety culture which are the common cause that affects all management factors.
3. **Barrier-based accident data analysis method**

To analyze the accident investigation report by using a barrier-based method and to identify the causes and characteristics of the accidents in facilities, departments or organizations by understanding the failure data of hardware barriers and human barriers will help us fully understand the causes of the accidents and find out the weak links in safety management. This paper mainly adopts the following four steps, as shown in Figure 3:

**Step 1:** Data collection and examination. Collect the accident investigation report and HSE information system accident data of the group. The data type is unstructured text data, including but not limited to accident process description data and cause description.

**Step 2:** Develop a common barrier analysis knowledge graph, including barrier type, cause factors for barrier failure, and corresponding management factors.

**Step 3:** Annotate and label every investigation report or every accident record of the HSE information system. Make certain the accident process and determine the cause factors and management factors for each barrier failure.

**Step 4:** Develop a HSE text semantic recognition model, and use the model to automatically read the key information in the text data of the accident investigation report, including production operation link, accident type and failure barrier information, etc.

**Step 5:** After extracting the key information, carry out statistics and classification for the failed barriers, cause factors and management factors to determine the mode of safety barrier failure.

The unstructured data of all the accidents are labelled under the guidance of barrier theory, and the text data of the fatal accidents are transformed into structured data. There is a difference between the structure of the current accident investigation report information and the accident investigation based on Swiss cheese model and barrier theory. The collection and extraction of barrier information is therefore insufficient. This part will be further discussed later.

**Figure 2.** Basic concept of barriers (human barriers, hardware barriers and management factors).

**Figure 3.** Technical route for barrier-based accident data analysis.
4. Barrier analysis knowledge graph

4.1. Barrier type

4.1.1. Hardware barriers. In order to effectively realize the function of each barrier, IOGP (The International Association of Oil and Gas Producers) divides barriers into hardware barriers and human barriers [9]. Hardware barriers refer to the equipment and system part that realizes the function of safety barriers [10]. In the upstream business field of the petroleum industry, hardware barriers are classified into 8 types from the perspective of safety function. These hardware barriers are also applicable to the midstream and downstream businesses of the petroleum industry.

Type 1: System related to structural integrity
Type 2: System related to process integrity
Type 3: Ignition source control system
Type 4: Detection system
Type 5: Shutdown system, including isolation system for production wells and well control system for drilling
Type 6: Protection system, including spray and fire protection systems
Type 7: Emergency response system
Type 8: Rescue and escape system

According to the 8 types of hardware barriers mentioned above, the project team analyzed the text data of the fatal accidents and searched for the shortcomings of hardware barriers from the text data on the basis of these types.

4.1.2. Human barriers. Human barriers refer to the human behaviors that realize the function of safety barriers, such as monitoring and evaluation of bottom hole flow and pressure during drilling, manual operation of equipment, responding to alarms, and the like. In reference to the IOGP544 report and the classification of human barriers by the UK Health and Safety Executive [11], and in combination with the typical activities of human in the human-machine system, human barriers are divided into the following 5 types:

Type 1: Human operation - operate in accordance with the requirements of work procedures
Type 2: Human monitoring - including patrol inspection and inspection and verification activities at various sites
Type 3: Human command - including approval, verification, on-site command, approval startup and other activities
Type 4: Human disposal - disposal of process alarms, production anomalies, and operational anomalies
Type 5: Human response - response to emergencies, mainly human activities in emergencies

According to the 5 types of human barriers mentioned above, the project team analyzed the text data of the fatal accidents and searched for the shortcomings of human barriers from the text data based on these 5 types.

4.2. Barrier failure cause

This paper adopts the following barrier failure causes, of which the causes for hardware barrier failure are divided into 6 types:

- Device/equipment design defect
- Device/equipment manufacturing defect
- Device/equipment installation defect
- Device/equipment operation control defect
- Poor working environment
- Poor natural environment

The causes for human barrier failure are divided into 7 types:

- Poor physiology
- Poor mentality
- Lack of awareness/ negligence/ omission
- Insufficient knowledge/ skills
- Poor communication
- Poor working environment
- Poor natural environment

4.3. Management factors
In the IOGP 544 report [12], the management factors affecting human and equipment barriers include the following 10 factors: 1) Commitments and responsibilities; 2) Policies, standards and targets; 3) Organizations, resources and capabilities; 4) Stakeholders and customers; 5) Risk assessment and control; 6) Asset design and integrity; 7) Plans and procedures; 8) Operational implementation; 9) Monitoring, reporting and learning; 10) Support, review and promotion.

These 10 factors are in fact part of the HSE management system. CNPC’s HSE management system consists of 7 primary elements and 27 secondary elements. The 27 secondary elements need to be merged and simplified as a dimension for management factor data analysis. The management factors adopted in this paper are divided into 13 types and synthesize system elements and CNPC’s management habits. Therefore, this paper intends to use the classification of the 13 types of management factors to analyze the text data of the fatal accidents.

- Compliance management
- Organization and responsibilities
- Systems and procedures
- Educational training
- Risk prevention and control and hidden danger investigation and governance
- Production organization
- Work permit
- Equipment management
- Change management
- Supply management
- Contractor management
- Accident and emergency management
- Leadership and organizational culture

5. Data annotation and labelling
Annotation and labeling are the key steps and difficulties in the analysis of accidents based on barrier theory. In this process, it is necessary to extract barrier failure information and determine the barrier types and the cause factors and management factors of barrier failures according to the text data of the accident investigation report. One accident is taken as an example to illustrate the process of labeling, as shown in Table 1.

The case of the accident investigation report is as follows:

“During the period, the pressure gauge of the pump in the pump room leaked out liquid methanol which was vaporized. There was a strong smell in the room at that time. The two persons did not take ventilation and safeguard measures in a timely manner and did not treat the leakage point. Their self-protection consciousness was poor, hence they inhaled methanol gas by different degrees and one of them had a mild methanol inhalation reaction.

Analysis of the major causes: The employees violated the operating procedures or labor disciplines, and the training was insufficient. Secondly, the technical control was not good enough in the purchasing process, the material of the pressure gauge failed to meet the requirement, and the medium corrosion caused the leakage.”
Table 1. Example of data analysis.

| Copied words of barrier failure text | Barrier types | Copied words of cause factor text | Cause factors | Copied words of management factor text | Management factors |
|--------------------------------------|--------------|----------------------------------|--------------|---------------------------------------|--------------------|
| The pressure gauge of the pump leaked out liquid methanol. | Process integrity system barrier | The material of the pressure gauge failed to meet the requirements, and the medium corrosion caused the leakage. | Device/equipment design defect | The technical control was not good enough in the procurement process. | Supply management |
| The employees violated the operating procedures or labor disciplines. | Human operation barrier | Poor self-protection consciousness and safety awareness. | Lack of awareness/negligence/omission | The training was insufficient. | Educational training |

Note: The copied words of barrier failure text, the copied words of cause factor text and the copied words of management factor text are part of the words or sentences from the original text of the accident event.

6. HSE text semantic recognition model
According to the data labeling and annotation method in Table 1 of Chapter 5, the accident data are manually labeled one by one to form a training data set.

According to the training data set, this paper creates a customized model that conforms to the characteristics of HSE information. As shown in Figure 4, the model includes two algorithms, i.e. syntactic analysis and semantic analysis. Syntactic analysis is mainly obtained from simple nouns, verbs, etc. For example, according to “falling from the derrick”, the accident type can be extracted as falling from a high place. Semantic analysis needs to get the overall meaning of the text from sentences and paragraphs. For example, according to “unable to provide protection during climbing under the existing process conditions”, the cause factor of barrier failure can be deemed as a device/equipment design defect.

Figure 4. HSE text semantic recognition model diagram.
Syntactic analysis involves four algorithms, including Fuzzy String Matching (FSM), word embedding similarity, Stanford NER model, and Google BERT model. Among them, FSM and word embedding similarity achieve text classification mainly by making type matching between text and HSE dictionary, having higher accuracy for “accident type” extraction. BERT (Bidirectional Encoder Representations from Transformers) is a new method of pre-training word embeddings open sourced by Google. It has achieved the best results in various downstream natural language processing tasks, being of a higher recognition rate for the “production operation link”, “production stage”, “risk barrier”, “cause factor” and “management factor” of accident occurrence. Stanford NER model itself can recognize the organization name, facility name, and accident type in the text.

Semantic analysis mainly includes unsupervised feature creation SVM classification model and Facebook FastText N-gram bag-of-words classification model. The former combines SVM (Support Vector Machine) and LDA (Latent Dirichlet Allocation) theme models for text classification. The SVM theme model is more efficient in classifying or regressing multi-dimensional and non-linear data, that is, using the same amount of data to obtain higher accuracy, or getting the same accuracy with less data. The LDA theme model can find out the theme to which a document belongs, and generate a document theme distribution vector for each document. Firstly, LDA is used to generate a theme distribution vector for each document, and the said vector is used to represent the characteristics of the document, and then this portion of data is used to train SVM. The Facebook FastText N-gram bag-of-words classification model is a text classifier open sourced by Facebook. Fasttext does not use a deep learning architecture, hence it has a high training speed. Meanwhile, it has the performance comparable to deep learning text classification, that is, it has the accuracy and recall comparable to deep learning. By adopting the idea of ensemble learning, the final classification results are achieved by comprehensively considering the results of the unsupervised feature creation SVM classification model and the Facebook FastText N-gram bag-of-words classification model.

Through machine learning and natural language, the data team structured the unstructured event data, and then performed HSE analysis for these structured event data and the structured data of the original accident/event. The comprehensive model prediction accuracy was not less than 62%.

7. Analysis results
This paper analyzes the fatal accidents from 2000 to 2018 of the group company. It reveals the failure distribution of hardware barriers and human barriers, and associates barrier types with their failure causes and corresponding management factors to find out the defects and shortcomings of HSE management and to provide recommendations and measures for continuous improvement of HSE management.

(1) Failure analysis of hardware barriers
As shown in Figure 5, the defects of the structural integrity barrier rank first, affecting 22% of the fatal accidents. The defects of the protection system barrier rank second, affecting a total of 15% of the fatal accidents.

![Figure 5. Influence ratio of hardware arriers to the fatal accidents.](image)
(2) Failure analysis of human barriers
As shown in Figure 6, the human operation barrier ranks first. The defects of this type of barrier affect a total of 59% of the fatal accidents. The human monitoring barrier ranks second. The defects of this type of barrier affect a total of 38% of the fatal accidents.

![Figure 6. Influence ratio of human barriers to the fatal accidents.](image)

(3) Analysis of the causes of barrier failure
As shown in Figure 7, the major causes of the failure of the structural integrity barrier are device/equipment design, installation and operational control. The major causes of the failure of the human operation and monitoring barriers are lack of awareness/negligence/omission, and insufficient knowledge/skills.

![Figure 7. Correlation analysis of the barrier failure causes of the fatal accidents.](image)

(4) Analysis of the management factors corresponding to barrier failure
As shown in Figure 8, risk prevention and control and hidden danger investigation and governance is correlated with 35% of the fatal accidents, and educational training is correlated with 33% of the fatal accidents. These two factors are the most common management factors for the occurrence of the fatal accidents.
Figure 8. Distribution of the influence of management factors on the fatal accidents.

Other management factors may not be fully revealed by the accident investigation report, especially leadership and organizational culture which actually affects all the other 12 management factors.

Figure 9. Correlation between barrier defects and management factors of the fatal accidents.
As shown in Figure 9, the defects of the structural integrity barrier are mainly correlated with the following two management factors: risk prevention and control and hidden danger investigation and governance; equipment management. The defects of the human operation barrier are mainly correlated with the following five management factors: capabilities and training; risk prevention and control and hidden danger investigation and governance; contractor management; production organization; systems and procedures.

8. Conclusions and recommendation
This paper attempts to use a new theory - barrier theory to analyze the data of the accident investigation report. It establishes a HSE text semantic recognition model to extract the barrier type in the accident event text, and reveals the inherent law of accidents and the defects of HSE management from the perspective of barriers. The following conclusions are achieved:

The number of the barrier defects of structural integrity in hardware barriers is the most and should be concerned in accident prevention, while the number of the barrier failures of emergency response system and detection system are the least. The failures of hardware barriers are the causes of accidents. Any barrier function may avoid the happening of accidents. It is recommended to pay attention to mitigation and emergency measures when enterprises are in danger.

The defects of the human operation barrier in human barriers affect the most accidents, and the number of the accidents affected by the human disposal barrier is the least. The defects of the human operation barrier should be concerned in accident prevention.

The three major management factors affecting accidents are: risk prevention and control and hidden danger investigation and governance, followed by educational training and contractor management. These three management factors are the shortcomings of business management that should be concerned continuously.

In the future, it is planned to supplement the sample data of accident events, iteratively optimize models, continuously improve the accuracy of the HSE text semantic recognition model, and extract reliable and accurate laws from accident events, so as to provide a basis for enterprise safety management decision-making.

References
[1] O. Lauridsen, E. Lootz, Husebo. Function Barrier Model: Barrier Management and the Interaction between Technical, Operational, and Organizational Barrier Elements: A Selection of Essays from SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production[C], Beijing, Safety and Environmental Protection Department of CNPC, 2016.

[2] Reason. Managing the Risks of Organizational Accidents [M]. England: Ashgate Publishing, 1997.

[3] Energy Institute, High level framework for process safety management[R]. London: 2010.

[4] Petroleum safety authority Norway. Principles for barrier management in the petroleum industry barrier memorandum[R]. Norway: 2017.

[5] Hu Yueting. Safety Risk Prevention and Control [M]. Beijing: Unity Press, 2017:197

[6] International Association of Oil & Gas Producers. Standardization of barrier definitions[R]. IOGP Process Safety Subcommittee, 2016.

[7] Energy Institute, Human Factor Performance Indicators for the Energy and related process industries[R], London: 2010.
[8] Marijana Blesic, Valentina Dichiarante, Roberto Milani, Markus Linder and Pierangelo Metrangolo. Evaluating the potential of natural surfactants in the petroleum industry: the case of hydrophobins [J]. Pure and Applied Chemistry, 2018, 90 212-3.

[9] Xingwei Zhen, Jan Erik Vinnem and Xue Yang. Quantitative risk modelling in the offshore petroleum industry: integration of human and organizational factors [J]. Ships and Offshore Structures, 2020, 15 812.

[10] Vibeke Milch, Karin Laumann. The influence of interorganizational factors on offshore incidents in the Norwegian petroleum industry: challenges and future directions [J]. Reliability Engineering and System Safety, 2018, 11 132-4.

[11] Jan Erik Vinnem, Jon Andreas Hestad and Jan Terje Kvaløy. Analysis of root causes of major hazard precursors (hydrocarbon leaks) in the Norwegian offshore petroleum industry [J]. Reliability Engineering and System Safety, 2010, 95 1142-53.

[12] Shih-Tzung Chen. An approach of identifying the common human and organisational factors (HOFs) among a group of marine accidents using GRA and HFACS-MA [J]. Journal of Transportation Safety & Security, 2020, 12 142-3.