



Article

Correlation in Causality: A Progressive Study of Hierarchical Relations within Human and Organizational Factors in Coal Mine Accidents

Ziwei Fa , Xinchun Li *, Quanlong Liu , Zunxiang Qiu and Zhengyuan Zhai

School of Management, China University of Mining & Technology, Xuzhou 221116, China; TB20070001B3LD@cumt.edu.cn (Z.F.); qll2016@cumt.edu.cn (Q.L.); TS20070117A31@cumt.edu.cn (Z.Q.); cumtzzy@cumt.edu.cn (Z.Z.)

* Correspondence: TS20070078A31LD@cumt.edu.cn; Tel.: +86-15366762329

Abstract: It has been revealed in numerous investigation reports that human and organizational factors (HOFs) are the fundamental causes of coal mine accidents. However, with various kinds of accident-causing factors in coal mines, the lack of systematic analysis of causality within specific HOFs could lead to defective accident precautions. Therefore, this study centered on the data-driven concept and selected 883 coal mine accident reports from 2011 to 2020 as the original data to discover the influencing paths of specific HOFs. First, 55 manifestations with the characteristics of the coal mine accidents were extracted by text segmentation. Second, according to their own attributes, all manifestations were mapped into the Human Factors Analysis and Classification System (HFACS), forming a modified HFACS-CM framework in China's coal-mining industry with 5 categories, 19 subcategories and 42 unsafe factors. Finally, the Apriori association algorithm was applied to discover the causal association rules among external influences, organizational influences, unsafe supervision, preconditions for unsafe acts and direct unsafe acts layer by layer, exposing four clear accident-causing "trajectories" in HAFCS-CM. This study contributes to the establishment of a systematic causation model for analyzing the causes of coal mine accidents and helps form corresponding risk prevention measures directly and objectively.

Keywords: coal mine accidents; HFACS framework; data-driven; text mining; association rules



Citation: Fa, Z.; Li, X.; Liu, Q.; Qiu, Z.; Zhai, Z. Correlation in Causality: A Progressive Study of Hierarchical Relations within Human and Organizational Factors in Coal Mine Accidents. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5020. <https://doi.org/10.3390/ijerph18095020>

Academic Editors: Roger Jensen and David P. Gilkey

Received: 25 March 2021

Accepted: 7 May 2021

Published: 10 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the resource with the highest proportion in China's energy consumption, the rapid growth of coal mine production guarantees the steady development of its national economy [1]. In recent years, the overall safety situation of China's coal-mining industry has been progressively improving, with the mortality rate per million tons, the number of total deaths and the number of accidents showing a downward trend annually. In 2020, the mortality rate per million tons was only 0.058, which was the lowest in the safety records of China's coal-mining industry [2]. However, serious accidents still occur from time to time, and there is still a huge gap compared with developed countries [3,4]. Of the 122 coal mine accidents that occurred in 2020 alone, ten were major accidents (death toll between 3 and 10) and three more were serious accidents (death toll between 10 and 30). Among them, on 27 September 2020, a fire accident occurred in Chongqing City, China, causing 16 deaths and 42 injuries, with a direct economic loss of 3.84 million dollars; on 4 December 2020, another fire accident occurred in Chongqing, resulting in 23 deaths and 1 serious injury, with a direct economic loss of 3.63 million dollars; and on 29 November 2020, a major flood accident occurred in Hunan Province, China, leading to 13 deaths and direct economic losses of 5.36 million dollars [2].

For the coal-mining industry, the specific causes of each accident are different. However, according to the analysis of accident investigation reports, the causes can be roughly

divided into direct causes and indirect causes [5]. The direct causes refer to the unsafe behaviors or state of the operators who cause the accidents, which can be regarded as the first source of causes, while indirect causes are defects in internal management and external supervision, which can be taken as the second source. Thus, it can be inferred that coal mine accidents can mainly be attributed to human and organizational factors (HOFs). Under these circumstances, in order to better explain the mechanism of accidents from the perspective of human and management, accident-causing models, as important analysis tools, have been proposed and improved to adapt them to the development of sociotechnical systems.

The domino model was the first analytical model put forward to explicitly point out human error in industrial accidents [6]. The collapse of the first domino will knock down the subsequent dominoes one after another until injury occurs. Since the domino model oversimplifies the process of control over human behaviors in accidents, Bird, Adams and Weaver extended this model, and incorporated managerial decisions into accident causes [7–10]. Nevertheless, simple linear accident models of this kind have generally been considered too straightforward, with limited applicability in increasingly complex sociotechnical systems. Owing to the need for a more reasonable method and a stronger model for understanding accidents, simple linear models were gradually replaced by epidemiological models in the 1980s, which compared the events that lead to accidents with the spread of diseases, and held that accidents were the results of a variety of factors accidentally existing in both space and time. The most famous epidemiological model is the Swiss cheese model proposed by Reason, who considered that an accident involves four categories of factors (four pieces of cheese), namely, organizational influences, unsafe supervision, preconditions for unsafe acts, and direct unsafe acts [11]. Each piece of cheese represents a layer of the defense system, while the holes in the cheese represent loopholes or defects in the corresponding defense system. Since the positions and sizes of these holes are constantly changing, and when the holes in each piece of cheese are arranged in a straight line in a given instant, an accident-causing “trajectory” is formed, where the risks pass through the loopholes in all of the defense systems and eventually lead to accidents. However, this model does not present exact definitions of “holes” in each piece of cheese, thus limiting its practical application [12]. To facilitate the investigation and analysis of accidents, Shappell and Wiegmann, based on the Swiss cheese model, established the Human Factors Analysis and Classification System (HFACS) by analyzing hundreds of aviation accidents, clearly defining the “holes” in the cheese, and this has been widely used in various fields [13].

Since the development of this model, Baysari et al. identified the errors that frequently resulted in the occurrence of rail accidents under the guidance of HFACS framework, finding that slips of attention were the most common unsafe acts committed by drivers [14]. Rashid et al. performed a statistical analysis of 58 helicopter maintenance-induced safety occurrences, put forward the HFACS-ME model in maintenance extension, and studied the survivability and severity distribution of such occurrences [15]. Chauvin et al. applied the HFACS system to maritime collision accidents to analyze HOFs in the shipping industry, and found that most accidents were the result of decision errors [16]. Cohen et al., through the investigation of the causes in surgical near-miss events, discovered that most issues involved the preconditions for unsafe acts, followed by unsafe acts, organizational influences and unsafe supervision [17]. Chen put forward the HFACS-CI framework in the construction industry, and verified its effectiveness through concrete accident analysis [18]. Wang et al. revised the HFACS system for small- and medium-sized enterprises in the chemical industry, and considered that it could effectually identify and distinguish the causes in chemical accidents [19]. In the coal-mining industry, Patterson and Shappell analyzed 508 coal mine accidents in Australia, and the results indicated that skill-based errors in the HFACS model were the most common unsafe acts in miners [20]. Liu et al. collected over 300 coal mine accidents in China, established the HFACS-CM framework of China’s coal-mining industry, and calculated the weight of each causation factor [21].

However, previous scholars have mainly used research methods based on grounded theory, such as literature summary, on-the-spot investigation, in-depth interview, or manual analysis to identify unsafe factors in accidents. Under these circumstances, some important accident-causing factors or hidden variables are not recognized, and consequently, are not considered in the combinations of variables. Thus, the adaptability between assumptions and data as well as explanatory ability in traditional models will be weak. In addition, although HFACS framework, as a hierarchical model, has apparent causal relationships within layers, since there are many elements in the system, the correlation of unsafe elements between layers is often ignored, and the causality of specific elements that have an important role in restraining accidents is not recognized, that is, the transmission paths of the accident-causing “trajectory” are still unclear, which can limit the effectiveness of precautions in the coal mine risk management process. Based on this, by adopting the data-driven research paradigm, this study uses text mining technology to identify manifestations of accident-causing factors that restrict safe production in the coal-mining industry and maps them into the HFACS framework, forming a modified HFACS-CM in the coal-mining industry. Moreover, under the guidance of the inherent causal relationships in the system, frequent patterns of specific unsafe factors between layers are further discovered, forming a causal chain with clear unsafe elements in coal mine accidents, so as to upgrade the relevant accident precautions.

The remainder of this paper is organized as follows. Section 2 introduces the data sources, the basic concepts of each method, and the research framework. In Section 3, manifestations of the accident causes are extracted through text mining, and hierarchical association rules are mined using the Apriori algorithm under the guidance of the modified HFACS-CM framework. Finally, some concluding remarks are provided in Section 4.

2. Preliminaries and Research Framework

2.1. Data Set

Since accident investigation reports can reveal the causes of accidents in detail, they can also reveal mistakes within the process of risk management and the defects of various constructive documents in coal-mining enterprises. Therefore, accident investigation reports were used in this paper as the original data for text mining. In the selection of coal mine accident reports, some samples came from the State Administration of Work Safety, China’s Emergency Management Department, and the official website of the provincial and municipal safety supervision administration, while other samples were obtained through field investigation of China’s coal-mining enterprises. After data cleaning, a total of 883 accidents in the coal-mining industry in China from 2011 to 2020 were identified, including eight accident types: gas, electro-mechanical, transportation, flood, fire, blasting, roof, and other accidents. Specific information on the various types of accidents is shown in Figure 1.

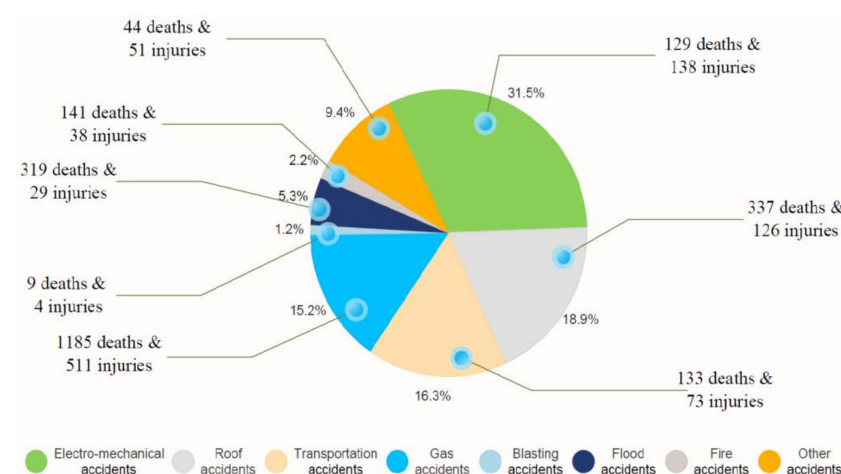


Figure 1. Statistics of 883 coal mine accidents.

2.2. HFACS Framework

The original HFACS framework contained 4 categories and 17 subcategories. However, considering the influences of environment on unsafe behaviors, Wiegmann and Shappell added “physical environment” and “technical environment” into the category of “preconditions for unsafe acts” (Figure 2) [22]. After that, Patterson and Shappell, when analyzing coal mine accidents, introduced the fifth category “outside factors”, based on the original HFACS framework, and divided it into subcategories including “regulatory factors” and “other factors” [20]. Liu et al. refined “outside factors” into “management factors”, “policy factors”, “economic factors”, and “historical factors” [21].

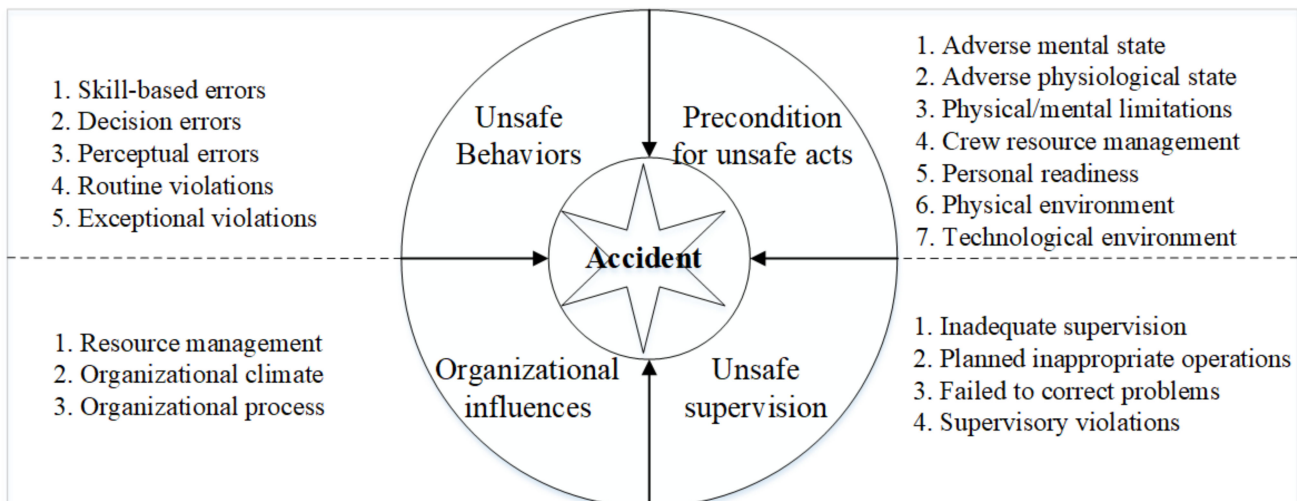


Figure 2. The original HFACS framework.

Since the coal mine industry is a sociotechnical system with the interaction of technology, society and organizational influences, the increasing complexity of technology and the rapid development of society mean that the system is not only affected by internal fluctuations, but also by the external environment, including market competition, economic and political pressure. Therefore, on the basis of the original HFACS framework, this paper took “external influences” as the fifth category in this model.

2.3. Text Mining

Text mining refers to the computer processing technology used to extract valuable information from text data, which can also be called knowledge discovery in a database [23]. In recent years, there has been a rising trend in the achievements related to text mining, and it is already becoming one of the most effective methods for studying the relationships between elements in various disciplines [24–27]. This paper mainly uses the function of text segmentation in text mining to extract the manifestations with the characteristics of 883 coal mine accidents. The process is illustrated in Figure 3.

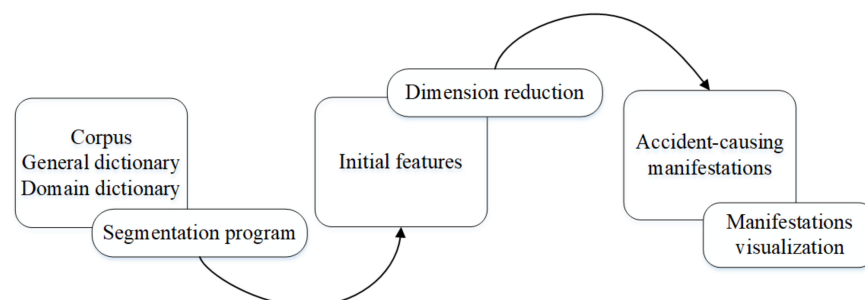


Figure 3. Process of manifestation extraction.

2.4. Association Rules

Traditional models determine the combination of variables or constructs based on observation, theoretical deduction and empirical refinement, thus establishing theoretical assumptions, which are then tested through data demonstration. However, in the context of multidimensional data, numerous combinations of variables need to be examined, making difficult to build traditional models and test them one by one. At this time, it is necessary to use the association rules of data-driven paradigms to determine the correlation between variables in order to reduce the space and combination scale of the variables [28,29]. Nevertheless, correlations among large-scale data are not always accessible, since the traditional association algorithms are usually not perfect, that is, the rules generated usually contain some meaningless or even incorrect rules, referred to as weak rules and negative rules. Moreover, the general rules often fail to reflect the useful relationships among data, and thus cannot fully express deep knowledge in practical applications. In this context, this paper aims to discover the association patterns of unsafe factors under the guidance of the HFACS framework, which has its own causal and logical characteristics, giving the antecedents and consequents in the association rules a certain causality, such that the rules mined are strong and contain more valuable knowledge than traditional models.

At present, the commonly used association algorithms are the Apriori algorithm, the DIC algorithm, and the FP-Growth algorithm [30]. In this paper, to extract the frequent patterns of unsafe factors, the Apriori algorithm was used to find meaningful connections in multidimensional data sets and to improve inspection efficiency in coal-mining enterprises. In this algorithm, the correlation between elements is mainly reflected by three indicators: “support”, “confidence” and “lift” [31–33].

Let $I = \{i_1, i_2, i_3 \dots, i_n\}$ be defined as the set of items in all events, and let $T = \{t_1, t_2, t_3 \dots, t_m\}$ be the set of all events. Support count $\sigma(X) = |\{X_{t_i}, t_i \in T\}|$ represents the number of occurrences of a given item set in all transactions, and the support between two indicators can be expressed as follows:

$$\text{support}(X \& Y) = p(X \& Y) = \frac{\sigma(X \cup Y)}{N} \quad (1)$$

The confidence between indicators can be expressed as follows:

$$\text{confidence}(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2)$$

The value of lift between indicators can be expressed as follows:

$$\text{lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{support}(Y)} \quad (3)$$

2.5. Research Framework

As shown in Figure 4, the research process of this paper includes two stages and four steps. The first stage contains the mining of manifestations of causes in coal mine accidents (Step 1) and the revision of HFACS framework in the coal-mining industry (Step 2). The second stage is the extraction of causal association rules among unsafe factors in different layers (Step 3) and the formulation of specific accident-causing trajectories (Step 4).

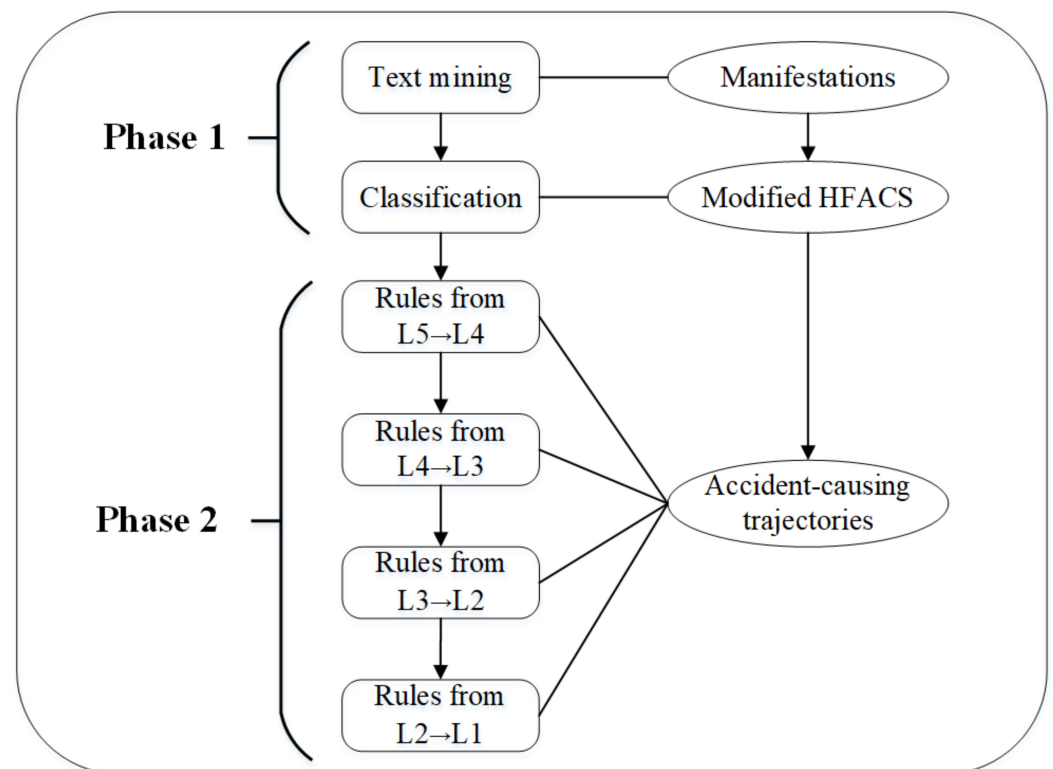


Figure 4. Research framework.

3. Results and Analysis

3.1. Text Mining Results

After initial text segmentation, there were as many as 3000 original items containing a lot of useless information that could seriously interfere with the subsequent analysis, so it was necessary to reduce the dimensions of the featured items. In this paper, the chi-squared statistic was used to extract useful features [34]. In addition, the calculation formula was:

$$\chi^2(t, c_i) = \frac{n \times (a \times b - c \times d)^2}{(a + c)(b + d)(a + b)(c + d)} \tag{4}$$

In Equation (4), n represents the total amount of text; c_i refers to the category of text; a is text frequency that belongs to c_i and contains item t ; b is text frequency that does not belong to c_i but contains item t ; c stands for the text frequency that belongs to c_i but does not contain item t ; d indicates the text frequency that does not belong to c_i and does not contain item t either.

$$\chi^2_{max}(t) = \max_{i=1}^m \{ \chi^2(t, c_i) \} \tag{5}$$

Most meaningless items can be removed by using the m value in Equation (5), so as to reduce the dimensions of characteristics. A total of 55 features representing different causes in coal mine accidents were finally obtained, as shown in Table 1, where F represents frequency.

Table 1. Fifty-five manifestations extracted from coal mine accidents.

Item	Manifestations	F	Item	Manifestations	F
1	Inadequate control of superior companies	492	29	Incomplete safety monitoring system	37
2	Failure to provide guidance	423	30	Defects in ventilation system	35
3	Weak safety awareness of employees	366	31	Inadequate safety supervision	28
4	Distempered technical specifications	329	32	Distraction at work	27
5	Failure to inspect and fix hidden risks	289	33	Unreasonable labor organization	26
6	Act against regulations	193	34	Inadequate risk assessment	24
7	Lack of safety training	173	35	Failed to learn from the past	21
8	Inadequate hazards identification	168	36	Lack of personal protective equipment	20
9	Failure to enforce rules	156	37	Falsified data and documents	16
10	Lack of government supervision	152	38	Fluke mind	14
11	Inadequate coal-mining administration control	140	39	Unreasonable working face layout	12
12	Illegal production	128	40	Defects in transportation management	12
13	Improper procedure	87	41	Production excess capability	12
14	Insufficient staffing	74	42	Lack of funding	7
15	Lack of preshift meetings	73	43	Lack of guard lines	5
16	Lack of safety confirmation	70	44	Emphasis on production rather than safety	4
17	Inadequate working ability	69	45	Aging equipment	4
18	Operation at risk	65	46	Defects in equipment management	3
19	Imperfect aggregate regulations	64	47	Defects in ventilation management	3
20	Lack of self and mutual protection awareness	59	48	Physical fatigue	3
21	Complicated geological structure	56	49	Lack of equipment	2
22	Imperfect management organization	52	50	Disorganized workplace	1
23	Authorized operators without certificate	49	51	Unqualified machinery	1
24	Defects in roof management	46	52	Equipment against regulations	1
25	Misjudgment of information	40	53	Defects in vehicle management	1
26	Inadequate facility maintenance	40	54	Lack of specified equipment	1
27	Failure to intervene in unsafe acts	38	55	Inadequate performance assessment	1
28	Lack of communication in shift change	38			

3.2. The Modified HFACS-CM Framework

As the 55 manifestations cover different aspects of coal mine safety production, it is necessary to classify them with a suitable theoretical framework. As for human factor modeling and assessment, there are various techniques available, such as the Success Likelihood Index Method (SLIM), the Swiss Cheese Model, the Human Factors Analysis and Classification System (HFACS) and Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H). Of all of these models, HFACS describes accident causes in more detail than the others and can help analyze multiple cases, enabling us to systematically excavate and categorize the direct causes of security failure with respect to coal mine accidents and the indirect causes behind them [11,12,35,36]. The process of mapping manifestations into HFACS was completed by the following three steps:

Step 1: Roughly categorize the manifestations. According to their nature, all manifestations were sequentially classified into the five categories of HFACS.

Step 2: Accurately map manifestations into HFACS in accordance with definitions of subcategories defined in the original HFACS framework. During this process, the identified manifestations were respectively mapped into the subcategories in the model. Since there are no manifestations in the 55 featured above that fit “physical/mental limitations”, this subcategory was removed. In addition, the descriptions of employees’ behavioral violations were vague in selected accident reports. For example, in one of the roof accidents, when analyzing the direct causes, it was disclosed in the accident report that “Some worker stood on coal gangue, operating against rules, caused the coal gangue to fall and was buried eventually”, making it hard to distinguish whether it was an “exceptional violation” or a “routine violation”, so the two types of violations were unified as “violations”. Additionally, since there is no category of “external influences” in the original HFACS model, subcategories were put forward according to the connotations of all of the manifestations above and the classification of previous studies, so the category of “external influences” was further divided into “policy factors” and “management factors” [20,21]. Consequently, 55 manifestations were split into 19 subcategories.

Step 3: Merge comparable manifestations. Under the same subcategory, if there are multiple causal factors expressing a similar concept, they can be grouped together. For example, factors such as “defects in roof management”, “defects in transportation management”, “defects in vehicle management”, “defects in equipment management” and “defects in ventilation management” are all caused by defective regulations applicable to a specific management department, so they were merged as “unsound sectional regulations”. Likewise, “inadequate performance assessment” and “imperfect aggregate regulations” refer to defective norms applicable to all employees and departments in coal mines, and were consolidated as “unsound integral regulations”. In addition, if there are inclusion relationships among manifestations, they should be merged as one. Therefore, “equipment against regulations” was merged into “unqualified machinery”; “lack of guard lines” together with “unreasonable working face layout” were merged to “disorganized workplace”; “lack of safety confirmation” was incorporated into “operation at risk”; “lack of self and mutual protection awareness” was merged into “weak safety awareness of employees”; “lack of specified equipment” was incorporated into “lack of equipment”; “insufficient staffing” and “inadequate safety supervision” were incorporated into “unreasonable labor organization” and “failure to intervene in unsafe acts”, respectively. Finally, some manifestations were re-expressed to reveal the essential causes of coal mine accidents more clearly. “Lack of preshift meetings” was modified to “unseasonable work arrangement” and “lack of communication in shift change” was changed to “miscommunication”.

After identifying and mapping all manifestations, the HFACS-CM framework was formed, containing 42 accident-causing factors. In this framework, organizational influences contained 10 factors. Unsafe supervision contained 10 factors. Preconditions for unsafe acts contained 15 factors. Unsafe acts of operators contained 4 manifestations, and external influences contained 3 factors. The modified HFACS-CM framework is illustrated in Figure 5.

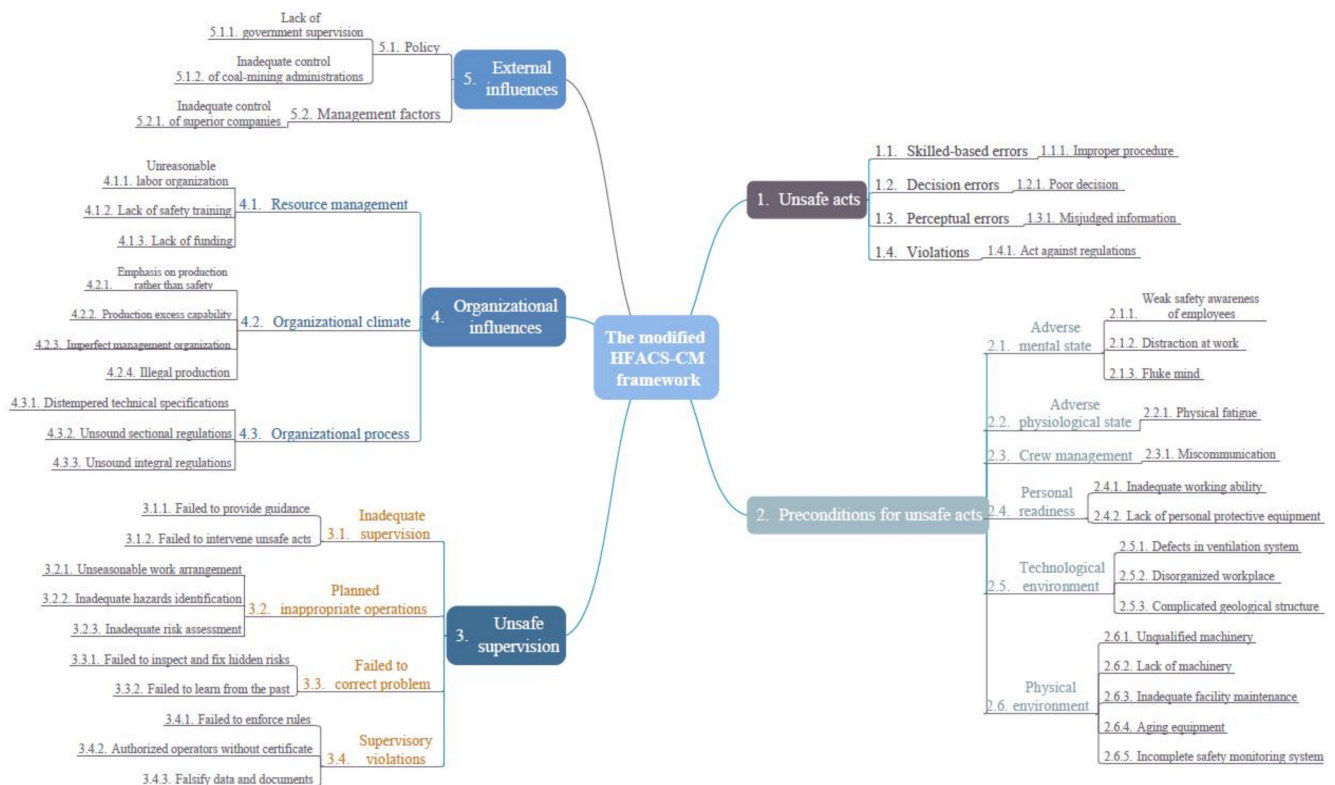


Figure 5. The modified HFACS-CM framework.

3.3. Causal Association Rules

Text segmentation can extract common factors that cause various accidents, but cannot discover the intrinsic relations. However, accidents are the result of multiple unsafe factors spreading in the accident chain, so determining the progressive modes of unsafe factors in HFACS-CM framework can systematically reveal patterns of factors involved in past coal mine accidents. To further clarify interactions among different contributing factors, this paper used the Apriori algorithm to mine the rules and trace the accident-causing paths from external influences to employees' unsafe acts layer by layer. When preprocessing data, 0–1 coding was used to convert the causative factors in text form into digital representation. If an accident corresponds to a specific factor, it is recorded as 1, otherwise as 0. Therefore, the accident database was converted into a matrix consisting of 0's and 1's.

3.3.1. Causal Rules from External Influences to Organizational Influences (L5→L4)

Based on a minimum support degree of 0.01, a minimum confidence degree of 0.5 and a minimum lift degree of 1.0, the antecedents of the rules were set as factors in external influences, and the consequents were factors in organizational influences. Relationships of 13 unsafe elements in two layers were carried out, and the causal rules are shown in Table 2. The visualization of six strong association rules is presented in Figure 6, where bubbles indicate the existence of association rules, the size of bubbles represents the value of confidence, and the gradation of color stands for lift value, revealing the degree of connection between factors.

Table 2. Causal rules between external influences and organizational influences.

Number	Rules	Support	Confidence	Lift
1	{Inadequate control of coal-mining administrations} => {Lack of safety training}	0.12	0.74	1.34
2	{Inadequate control of superior companies} => {Lack of safety training}	0.11	0.69	1.26
3	{Lack of government supervision} => {Lack of safety training}	0.07	0.70	1.26
4	{Lack of government supervision} => {Illegal production}	0.06	0.65	6.31
5	{Lack of government supervision} => {Distempered technical specifications}	0.05	0.54	1.88
6	{Lack of government supervision} => {Unreasonable labor organization}	0.05	0.52	2.93

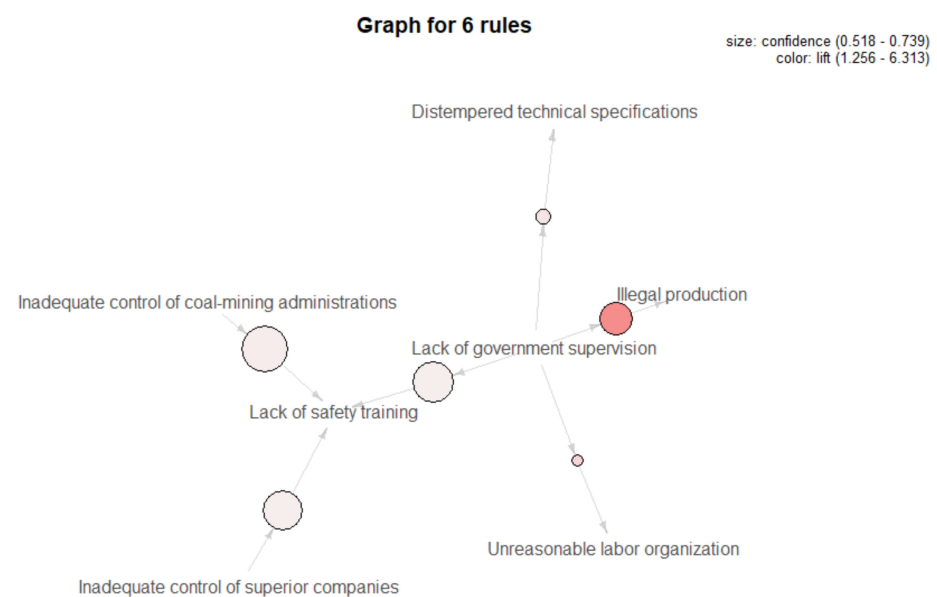


Figure 6. Visualization of causal rules between L5 and L4.

Specifically, it is apparent that most of the rules are linked to government supervision (Rules 3–6). If the government fails to supervise safe production, it can lead to a poor quality of safety education and training, illegal production, technical specifications that do

not meet the requirements, and unreasonable labor organization in coal-mining enterprises, among which government supervision and enterprises' illegal production have the highest lift degree (6.31), indicating the strongest correlation between them. In addition, apart from government supervision, the poor control of coal-mining administrations and parent companies allows enterprises to conduct inadequate safety training since 12% and 11% cases in all accidents were caused by them respectively.

3.3.2. Causal Rules from Organizational Influences to Unsafe Supervision (L4→L3)

Under the same working environment, the association patterns of 23 unsafe factors under the layers of unsafe supervision and organizational influences were mined, with the results shown in Table 3 and Figure 7. It can be seen that the co-occurrence frequency of inadequate safety training and failure to provide guidance in Rule 1 is as high as 29%, and by the confidence value (0.53), it can be concluded that when the safety training for employees is not in place, there is over a 50% chance that supervisors will deliver commands against rules, while when the organizational climate in companies emphasizes production over safety, the possibility of illegal commands increases to 68%, and the lift degree changes from 1.19 to 1.51. Moreover, Rule 2 and Rule 4 indicate that supervisors failing to inspect and fix hidden risks are likely to be triggered by unreasonable labor organization or lack of funding, which has a stronger influence on the quality of hidden risks investigation in coal-mining enterprises.

Table 3. Causal rules between organizational influences and unsafe supervision.

Number	Rules	Support	Confidence	Lift
1	{Lack of safety training} => {Failed to provide guidance}	0.29	0.53	1.19
2	{Unreasonable labor organization} => {Failed to inspect and fix hidden risks}	0.09	0.52	1.29
3	{Emphasis on production rather than safety} => {Failed to provide guidance}	0.03	0.68	1.51
4	{Lack of funding} => {Failed to inspect and fix hidden risks}	0.01	0.67	1.66

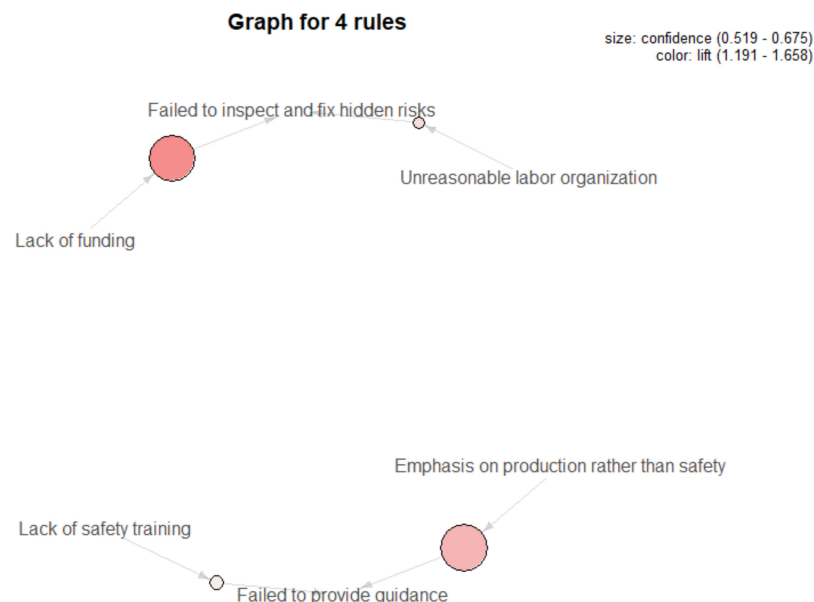


Figure 7. Visualization of causal rules between L4 and L3.

3.3.3. Causal Rules from Unsafe Supervision to Unsafe Preconditions (L3→L2)

Continuing to progressively dig down into the rules between unsafe supervision and unsafe preconditions, as shown in Table 4 and Figure 8, all causal rules concern employees' ideological states. Among them, failing to provide guidance or intervene in unsafe acts, inadequate investigation and rectification of hidden risks, unreasonable work arrangement

at preshift meetings, failure of leaders to strictly implement relevant regulations, permission of supervisors for unqualified crew to operate, and inadequate hazard identification will all lead to unsafe mental states among employees such as insufficient awareness of self-protection and mutual protection or a weak sense of responsibility. Combined with the visual map of the rules, it can be seen that although the co-occurrence of supervisors’ illegal commands and employees’ weak safety awareness is the highest, when leaders fail to strictly implement relevant regulations, there is a more than 80% possibility that the safety awareness of front-line operators will be weak, and the correlation between them is also the strongest.

Table 4. Causal rules between unsafe supervision and unsafe preconditions.

Number	Rules	Support	Confidence	Lift
1	{Failed to provide guidance} => {Weak safety awareness of employees}	0.31	0.70	1.16
2	{Failed to inspect and fix hidden risks} => {Weak safety awareness of employees}	0.28	0.68	1.12
3	{Failed to intervene unsafe acts} => {Weak safety awareness of employees}	0.18	0.70	1.14
4	{Unseasonable work arrangement} => {Weak safety awareness of employees}	0.17	0.71	1.16
5	{Failed to enforce rules} => {Weak safety awareness of employees}	0.10	0.84	1.38
6	{Authorized operators without certificate} => {Weak safety awareness of employees}	0.07	0.64	1.05
7	{Inadequate risk assessment} => {Weak safety awareness of employees}	0.03	0.73	1.20

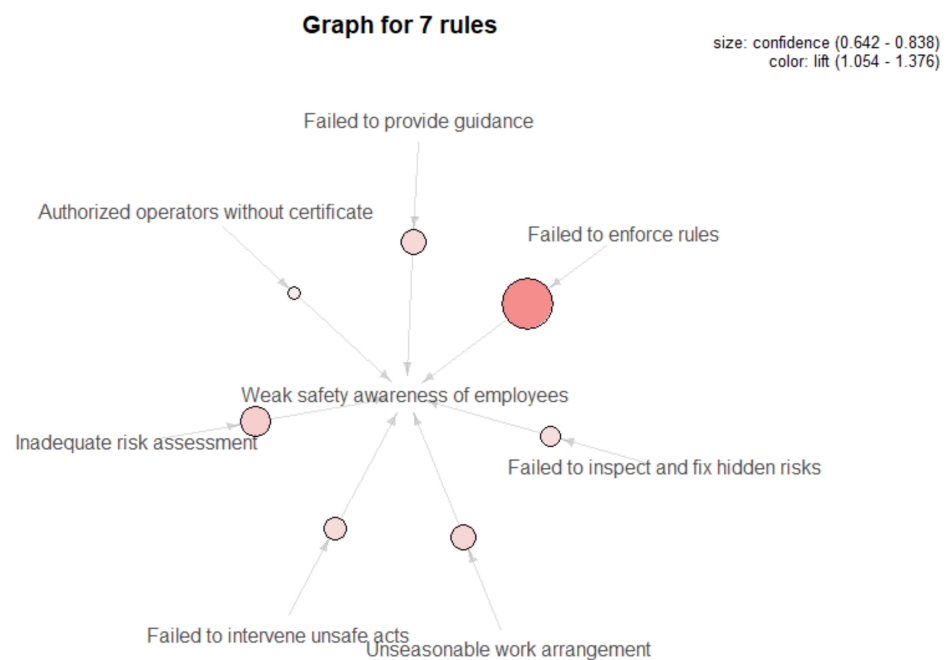


Figure 8. Visualization of causal rules between L3 and L2.

3.3.4. Causal Rules from Unsafe Preconditions to Unsafe Acts (L2→L1)

Finally, the association algorithm was used to mine useful knowledge under unsafe preconditions and direct unsafe acts, and the results are presented in Table 5 and Figure 9. Through analysis, it is found that the majority of unsafe factors in unsafe preconditions have a direct influence on operators’ unsafe acts. In particular, front-line operators’ weak safety awareness, fluke mind, not wearing personal protective equipment or poor communication during shift change can all result in violations during operation. In addition, unsafe state of technological and physical environment can also bring about employees acting against rules, such as lack of equipment and defective safety monitoring systems and ventilation systems, and when the design of ventilation systems in coal-mining enterprises is unreasonable, the possibility of employees acting illegally goes up to 72%, with the highest correlation, while skill-based errors like improper procedures are mainly caused by workers’ inadequate working ability or experience and complex geological structures.

Table 5. Causal rules between unsafe preconditions and unsafe acts.

Number	Rules	Support	Confidence	Lift
1	{Weak safety awareness of employees} => {Act against regulations}	0.32	0.53	1.21
2	{Inadequate working ability} => {Improper procedure}	0.11	0.55	1.30
3	{Lack of personal protective equipment} => {Act against regulations}	0.08	0.58	1.32
4	{Complicated geological structure} => {Improper procedure}	0.07	0.52	1.23
5	{Lack of machinery} => {Act against regulations}	0.06	0.57	1.29
6	{Miscommunication} => {Act against regulations}	0.05	0.58	1.32
7	{Incomplete safety monitoring system} => {Act against regulations}	0.04	0.56	1.28
8	{Defects in ventilation system} => {Act against regulations}	0.02	0.72	1.65
9	{Fluke mind} => {Act against regulations}	0.02	0.53	1.21

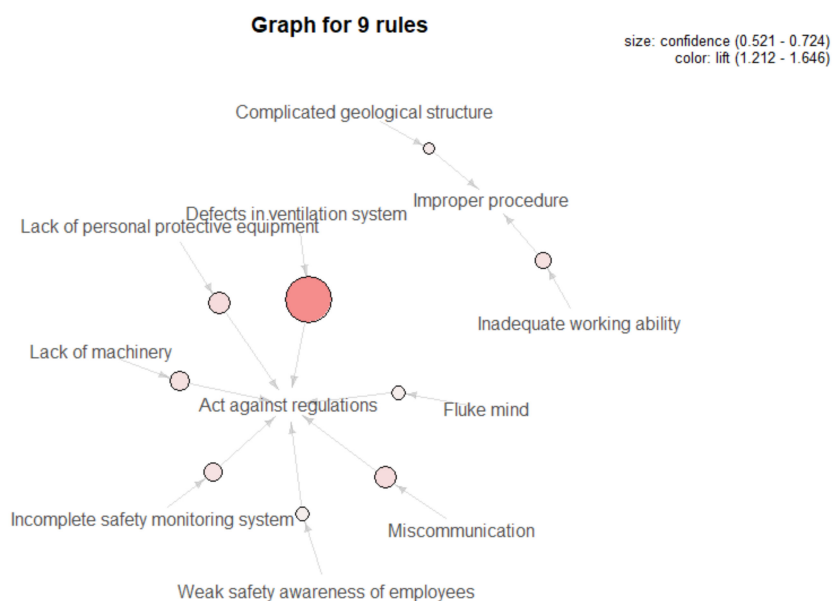


Figure 9. Visualization of causal rules between L2 and L1.

3.3.5. Accident-Causing Trajectories of Key Factors in the HFACS-CM Framework

According to the unsafe factors in the causal association rules above, the adapted fishbone diagram was used to list the key accident-causing factors in HFACS-CM framework, as shown in Figure 10, where the “eye” of the fish represents the occurrence of accidents, and the main bone, which is located at the diagram axis, consists of five arrows with graduated colors in accordance with the five layers in the HFACS-CM framework. For the fishbone diagram, key causation factors of coal mine accidents were added to fill in the bones of the fish. Accordingly, the accident-causing trajectories composed of specific factors were formed (Figure 11). It can be clearly seen that there are four main paths for the formation of coal mine accidents. In the first place, ineffective supervision by superior companies, coal mine administrations and the government can initially affect the quality of safety education and training in coal-mining enterprises, which will further lead to illegal commands by supervisors, where the safety awareness of front-line operators will gradually fade, and eventually illegal operations being conducted, directly causing accidents. Beyond that, the government’s ineffective supervision can also lead to a company’s unreasonable labor organization, such as insufficient staffing and unclear distribution of responsibilities, which can directly influence the quality of investigation and rectification towards hidden risks, gradually bringing about a long-term decline in employees’ vigilance against hazards or ability to assess risks, and conducting acts against rules.

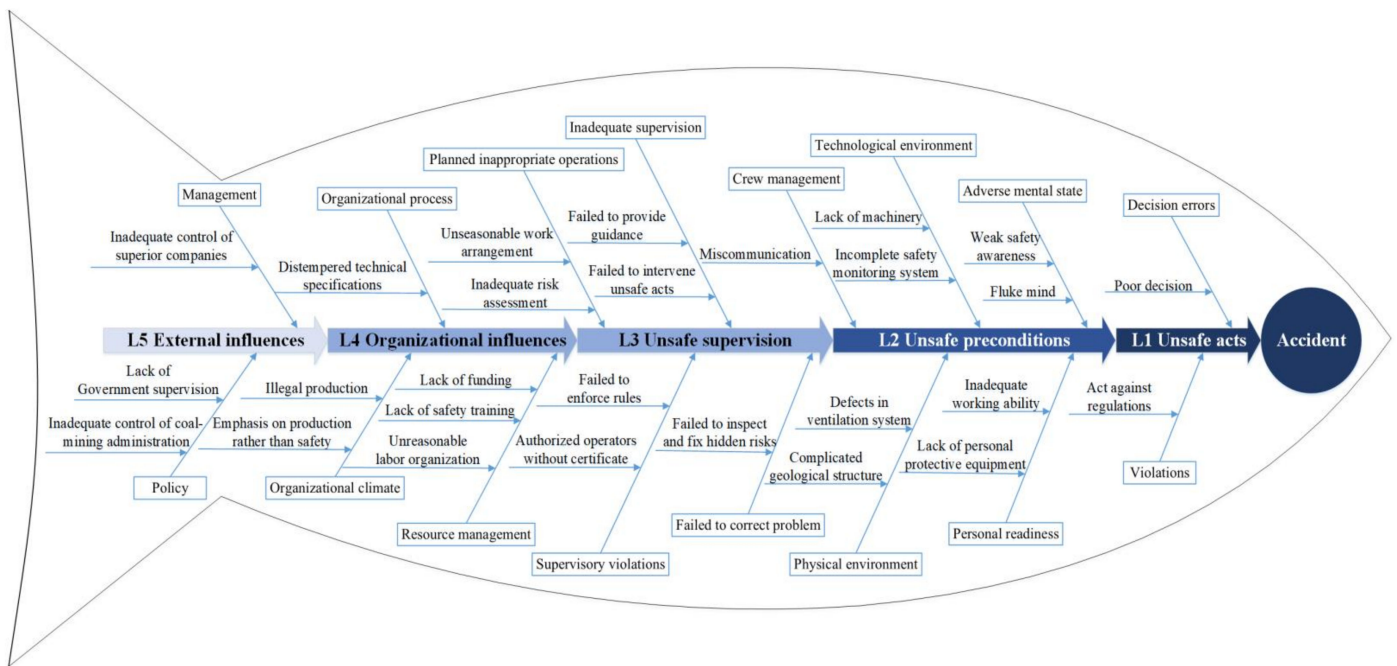


Figure 10. The adapted fishbone diagram about the key contributing factors in coal mines.

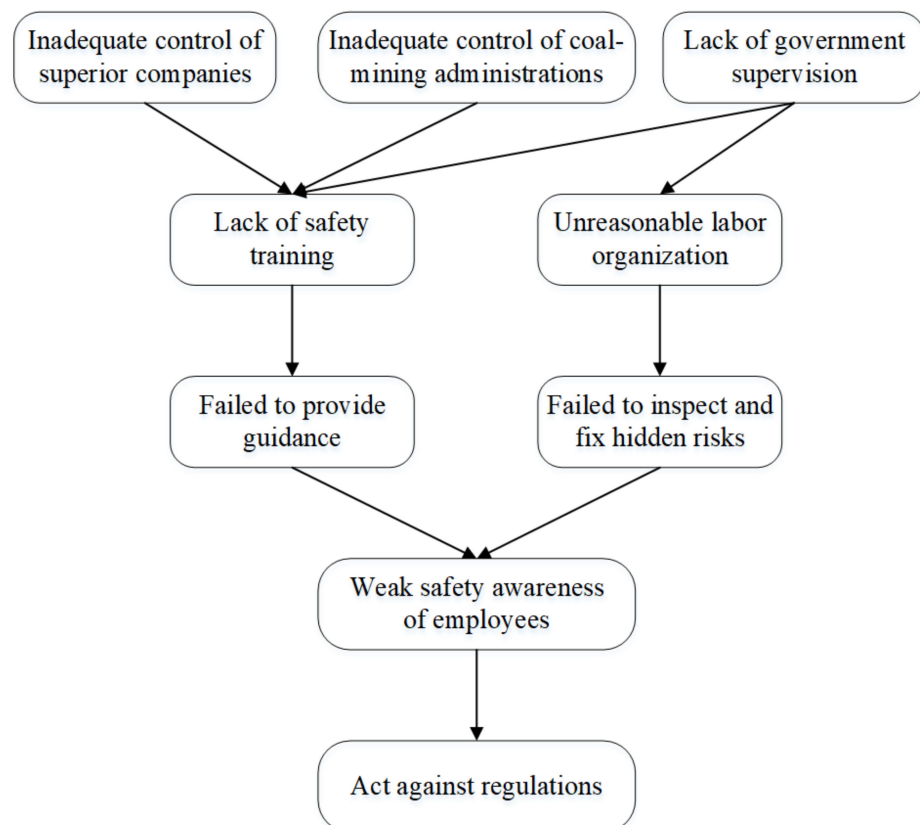


Figure 11. Accident-causing trajectories of specific factors in HFACS-CM framework.

In view of the accident routes in the process of coal mine safety management, this paper puts forward the following suggestions.

- (1) First, the government, coal-mining administrations, and superior companies need to strengthen the guidance and supervision with respect to the quality of enterprises' safety training and education, as well as using science-based approaches to ensure that

- accidents are effectively managed and positive learning is achieved [37]. Apart from that, it is also crucial for government departments to formulate censorship regulations so as to regularly check whether the enterprise's labor organization is appropriate.
- (2) Second, when organizing safety education and training, coal-mining enterprises should increase attention devoted to the examination of mastery of safety knowledge among supervisors, improving their work efficiency, putting an end to illegal commands, and ensuring that they can make the correct decisions in coal mine production. In addition, a department for investigating and rectifying hidden risks should be set up with adequate staff, so as to ensure regular inspection and timely correction of hidden risks can be implemented throughout the whole process of production.
 - (3) Finally, for ordinary workers, a safety education and assessment system should be established to help them conduct regular examinations of their awareness of self-protection and mutual protection, legal production awareness, and their ability to identify hazards, so as to expose the weaknesses of employees' safety awareness and develop targeted solutions for different situations.

4. Conclusions and Discussion

The aim of this study was to reveal the accident-causing trajectories in China's coal mine industry by using data-driven methods instead of traditional grounded-theory-based approaches. Theoretically, grounded theory starts with the actual observation, emphasizing the generalization of experience from the original data to a systematic theory. However, this qualitative method has its shortcomings. Firstly, since the data collection is based on the personal observation of researchers, the objectivity of the conclusions is limited. Secondly, when the data are massive and complicated, the burden for "human brains" to analyze is heavy and the accuracy and reliability will accordingly be reduced. Thus, this paper adopted a data-driven approach, which, through the statistical analysis of the whole body of data with the help of different algorithms, aimed to find correlations among the data objectively and to use them to guide decision making directly. Overall, through data-driven methods, unsafe human factors and management factors that restrict safe production in coal mines were found, and the causal association rules among unsafe factors were extracted layer by layer to reveal the causal paths of specific factors that can lead to coal mine accidents. The following conclusions are drawn.

- (1) With text segmentation technology to extract feature words with accident-causing characteristics of the coal mine accidents, a total of 55 manifestations were obtained. Then, according to the nature of each manifestation, they were mapped into the HFACS framework, forming a revised HFACS-CM model composed of 5 categories, 19 subcategories and 42 specific factors for coal-mining industry.
- (2) In the process of mining progressive modes among external influences, organizational influences, unsafe supervision, unsafe preconditions and unsafe acts, it was found that the ineffective supervision of the government and other external supervisory departments could bring about the unsafe state of internal management such as insufficient safety training and unreasonable labor organizations, which will continually cause supervisors to issue illegal commands and to fail to investigate or rectify hidden risks, thus leading to weak safety awareness among employees and behavioral violations.

This study contributes to establishing a systematic causation model for analyzing the causes of the failure of security to prevent coal mine accidents in China. It introduces the HFACS-CM framework for systematically and quantitatively analyzing the underlying human, organizational, environmental and technical factors. Such a model can facilitate more comprehensive accident investigation and form targeted precautions directly. This study also enriches existing risk analysis methods in the field of safety management by extending the application of HFACS and text mining along with association algorithm to expose effective causation rules among unsafe factors instead of general and weak rules.

The limitations of the current study should also be acknowledged, and these suggest directions for further research. The limitations mainly lie in the setting of the HFACS model.

Firstly, with the development of accident analysis tools, there exist some nonlinear accident causation models that are more advanced for the tightly coupled and nontractable socio-technical systems compared with complex-linear causation model such as HFACS [38]. In addition, for unsafe factors that do qualify for the accident-causing trajectories presented, when accidents are seen as complex phenomena, there may not be an obvious relationship between the behavior of parts in the system and system-level outcomes, which is the weakness of the HFACS model [39]. Under these circumstances, complexity theory should be taken into account for further study. Secondly, if there are real links between accident-causing factors that do in reality have a negative impact, but their correlation does not align with the HFACS framework, then some real meaningful rules will be left out. Thirdly, due to the nature of this study, there will be great variability between accident descriptions and bias in data processing [40]. The method of grouping factors together when modifying the HFACS-CM framework will inevitably influence the taxonomy, which is the central mechanism in HFACS. For example, if factors are merged with a less detailed level of granularity, the association rules generated may not contain too much useful information. Overall, these limitations suggest the need to think beyond the HFACS model itself, considering the complexity of accident occurrence mechanism as well as the combination of data-driven results with actual facts.

Author Contributions: Conceptualization, Z.F.; methodology, Z.F. and Z.Q.; software, Z.Q.; validation, X.L.; formal analysis, Z.F.; investigation, Z.Z.; resources, Z.F.; data curation, Z.Z.; writing—original draft preparation, Z.F.; writing—review and editing, Q.L.; visualization, Q.L.; supervision, X.L.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by State Key Program of National Social Science Foundation of China, grant number 19AGL030.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Conflicts of Interest: We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. Chen, L.J.; Li, P.C.; Liu, G.M.; Cheng, W.M.; Liu, Z.X. Development of cement dust suppression technology during shotcrete in mine of China—A review. *J. Loss Prev. Process Ind.* **2018**, *55*, 232–242. [CrossRef]
2. The State Administration of Coal Mine Safety. 2020. Available online: <https://www.mem.gov.cn/xw/zhsqxx/> (accessed on 12 January 2021).
3. Liu, Q.; Li, X.; Meng, X. Effectiveness Research on the Multi-player Evolutionary Game of Coal-Mine Safety Regulation in China Based on System Dynamics. *Saf. Sci.* **2019**, *111*, 224–233. [CrossRef]
4. Liu, Q.; Meng, X.; Li, X.; Luo, X. Risk precontrol continuum and risk gradient control in underground coal mining. *Process Saf. Environ. Protect.* **2019**, *129*, 210–219. [CrossRef]
5. Meng, X.; Liu, Q.; Luo, X.; Zhou, X. Risk assessment of the unsafe behaviours of humans in fatal gas explosion accidents in China's underground coal mines. *J. Clean. Prod.* **2019**, *210*, 970–976. [CrossRef]
6. Heinrich, H.W.; Petersen, D.; Roos, N.R. *Industrial Accident Prevention: A Safety Management Approach*, 5th ed.; McGraw-Hill: New York, NY, USA, 1980.
7. Stranks, J.W. *Health and Safety at Work: Key Terms*; Butterworth-Heinemann: Oxford, UK, 2012.
8. Ridley, J.; Channing, J. *Safety at Work*, 7th ed.; Butterworth-Heinemann: Oxford, UK, 2012.
9. Abdelhamid, T.S.; Everett, J.G. Identifying root causes of construction accidents. *J. Constr. Eng. Manag.* **2000**, *126*, 52–60. [CrossRef]
10. Taylor, G.; Easter, K.; Hegney, R. *Enhancing Occupational Safety and Health*; Butterworth-Heinemann: Oxford, UK, 2004.
11. Reason, J. *Human Error*; Cambridge University Press: Cambridge, UK, 1990.
12. Shappell, S.A.; Wiegmann, D.A. The human factors analysis and classification system-HFACS. *Am. Libr.* **2000**, *1*, 20–46.
13. Wiegmann, D.A.; Shappell, S.A. Human error analysis of commercial aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Aviat. Space Environ. Med.* **2001**, *72*, 1006–1016.

14. Baysari, M.T.; Caponecchia, C.; Mcintosh, A.S. Classification of errors contributing to rail incidents and accidents: A comparison of two human error identification techniques. *Saf. Sci.* **2009**, *47*, 948–957. [CrossRef]
15. Rashid, H.S.J.; Place, C.S.; Braithwaite, G.R. Helicopter maintenance error analysis: Beyond the third order of the HFACS-ME. *Int. J. Ind. Ergon.* **2010**, *40*, 636–647. [CrossRef]
16. Chauvin, C.; Lardjane, S.; Morel, G. Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accid. Anal. Prev.* **2013**, *59*, 26–37. [CrossRef] [PubMed]
17. Cohen, T.N.; Francis, S.E.; Wiegmann, D.A. Using HFACS-Healthcare to identify systemic vulnerabilities during surgery. *Am. J. Med. Qual.* **2018**, *33*, 614–622. [CrossRef]
18. Chen, Y. The development and validation of a human factors analysis and classification system for the construction industry. *Int. J. Occup. Saf. Ergon.* **2020**. [CrossRef] [PubMed]
19. Wang, J.; Fan, Y.; Gao, Y. Revising HFACS for SMEs in the chemical industry: HFACS-CSMEs. *J. Loss Prev. Process Ind.* **2020**. [CrossRef]
20. Patterson, J.M.; Shappell, S.A. Operator error and system deficiencies: Analysis of 508 mining incidents and accidents from Queensland, Australia using HFACS. *Accid. Anal. Prev.* **2010**, *42*, 1379–1385. [CrossRef]
21. Liu, R.; Cheng, W.; Yu, Y.; Xu, Q. Human factors analysis of major coal mine accidents in China based on the HFACS-CM model and AHP method. *Int. J. Ind. Ergon.* **2018**, *68*, 270–279. [CrossRef]
22. Wiegmann, D.A.; Shappell, S.A. *A Human Error Analysis of Commercial Aviation Accidents: Using the Human Factors Analysis and Classification System*; Ashgate: Burlington, VT, USA, 2003.
23. Mitchell, T. *Machine Learning*; McGraw-Hill: New York, NY, USA, 1997.
24. Liang, D.; Yi, B. Two-stage three-way enhanced technique for ensemble learning in inclusive policy text classification. *Inf. Sci.* **2021**, *547*, 271–288. [CrossRef]
25. Soares, V.H.A.; Campello, R.J.G.B.; Nourashrafeddin, S.; Milios, E.; Naldi, M.C. Combining semantic and term frequency similarities for text clustering. *Knowl. Inf. Syst.* **2019**, *61*, 1485–1516. [CrossRef]
26. Chen, P.; Chai, J.; Zhang, L.; Wang, D. Development and application of a Chinese webpage suicide information mining system (Sims). *J. Med. Syst.* **2014**, *38*, 88. [CrossRef] [PubMed]
27. Kyebambe, M.N.; Cheng, G.; Huang, Y.Q.; He, C.H.; Zhang, Z.Y. Forecasting emerging technologies: A supervised learning approach through patent analysis. *Technol. Forecast. Soc. Chang.* **2017**, *125*, 236–244. [CrossRef]
28. Kuhn, T. *The Structure of Scientific Revolution*; The University of Chicago Press: Chicago, IL, USA, 1962.
29. Hey, T.; Tansley, S.; Tolle, K. *The Fourth Paradigm: Data-Intensive Scientific Discovery*; Microsoft Research: Redmond, WA, USA, 2009.
30. Han, J.; Kamrr, M. *Data Mining Concepts and Techniques*; Morgan Kaufmann Publishers: San Francisco, CA, USA, 2000.
31. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. *ACM SIGMOD Rec.* **1993**, *22*, 207–216. [CrossRef]
32. Harun, N.; Makhtar, M.; Abd Aziz, A.; Zakaria, Z.; Abdullah, F.; Jusoh, J. The application of apriori algorithm in predicting flood areas. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2017**, *7*, 763–769. [CrossRef]
33. Toivonen, H. Sampling Large Databases for Association Rules. Available online: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.9385&rep=rep1&type=pdf> (accessed on 10 May 2021).
34. Zia, T.; Akhter, M.P.; Abbas, Q. Comparative study of feature selection approaches for Urdu text categorization. *Malays. J. Comput. Sci.* **2015**, *23*, 93–109.
35. Embrey, D.; Humphreys, P.; Rosa, E.; Kirwan, B.; Rea, K. *SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment. Volume I. Detailed Analysis of the Technical Issues*; Brookhaven National Lab: Upton, NY, USA, 1984.
36. Gertman, D.; Blackman, H.; Marble, J.; Byers, J.; Smith, C. The SPAR-H Human Reliability Analysis Method. *US Nucl. Regul. Comm.* **2005**, *230*, 35.
37. Heraghty, D.; Rae, A.J.; Dekker, S.W.A. Managing accidents using retributive justice mechanisms: When the just culture policy gets done to you. *Saf. Sci.* **2020**. [CrossRef]
38. Hulmea, A.; Stantonb, A.N.; Walkerc, G.H.; Watersond, P.; Salmona, P.M. What do applications of systems thinking accident analysis methods tell us about accident causation? A systematic review of applications between 1990 and 2018. *Saf. Sci.* **2019**, *117*, 164–183. [CrossRef]
39. Dekker, S.; Cilliers, P.; Hofmeyr, J. The complexity of failure: Implications of complexity theory for safety investigations. *Saf. Sci.* **2011**, *49*, 939–945. [CrossRef]
40. Lundberg, J.; Rollenhagen, C.; Hollnagel, E. What-You-Look-For-Is-What-You-Find—The consequences of underlying accident models in eight accident investigation manuals. *Saf. Sci.* **2009**, *47*, 1297–1311. [CrossRef]