



Article Establishment of a Key Hidden Danger Factor System for Electric Power Personal Casualty Accidents Based on Text Mining

Dan Lu¹, Changqing Xu¹, Chuanmin Mi^{2,*}, Yijing Wang², Xiangmin Xu³ and Chufan Zhao²

- ¹ Economic and Technological Research Institute, State Grid Henan Electric Power Company, Zhengzhou 450000, China; ludan1@ha.sgcc.com.cn (D.L.); xuchangqing@ha.sgcc.com.cn (C.X.)
- ² College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; wangyijing0302@163.com (Y.W.); zhaochufan@nuaa.edu.cn (C.Z.)
- ³ Suzhou China International Engineering Consulting Corporation, Suzhou 215000, China; xuxiangmin@ciecc.com.cn
- * Correspondence: cmmi@nuaa.edu.cn

Abstract: Based on actual safety management difficulties and needs, this paper aims to screen and extract the key accident potential factors of personal injuries and deaths within the electric power industry to provide a reference for electric power companies' accident prevention effort. First, this document sorts out and analyzes all of the causes and influencing elements that may lead to the occurrence of electric personal injuries and deaths, based on which rough accident potential factors are initially identified and combined with the definition of accident potentials. Second, this paper mines and analyzes relevant accident report texts using text-mining technologies such as term count, word cloud, and term frequency–inverse document frequency (TF-IDF), and thus a system of key accident potential factors for personal injuries and deaths within the electric power industry, including three key factors (human, material, and management), is finally constructed. Workers' habitual violation behavior, in particular, has a larger risk than other key accident potential components, implying that additional steps should be made to eradicate this type of critical accident potential in time.

Keywords: text mining; word cloud; TF-IDF; safety management; personal injuries and deaths; potential factors

1. Introduction

Power generating safety is becoming increasingly crucial as China's power system develops and becomes more complicated, as well as the increased power demand brought on by China's economic development. An accident is defined as an incident that causes a personal casualty or direct economic loss in production and operation activities under China's Regulations on the Reporting, Investigation, and Handling of Work Safety Accidents [1]. Personal accidents, as a subset of accidents, are accidents that result in a personal casualty, excluding property damage and equipment damage. As a result, an electric power personal casualty accident is defined as an event that causes a personal casualty in the electric power industry's production, engineering construction, marketing, and industrial fields, typically involving electric shock, falling accidents, being struck by objects, collapsing, and burial [2]. According to statistics from the National Energy Administration's Electric Power Safety Supervision Department, 44 electric power personal casualty accidents occurred in China in 2019, resulting in 54 casualties, endangering the lives of various personnel and affecting the development of electric power enterprises and the national economy. As a result, it is urgent to determine the reasons for these events and prevent them from occurring again.

According to research, hazard sources and hidden threats are commonly regarded as the causes of accidents. There are three types of danger sources in the theory of accident causes. The first type of danger source is the energy carrier or hazardous substance itself,



Citation: Lu, D.; Xu, C.; Mi, C.; Wang, Y.; Xu, X.; Zhao, C. Establishment of a Key Hidden Danger Factor System for Electric Power Personal Casualty Accidents Based on Text Mining. *Information* **2021**, *12*, 243. https://doi.org/ 10.3390/info12060243

Academic Editors: Gennady Agre and Diego Reforgiato Recupero

Received: 9 April 2021 Accepted: 31 May 2021 Published: 10 June 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/). which is the material source of the disaster and cannot be eliminated. The second class relates to harmful conditions and unsafe human behavior, while the third class relates to organizational issues that do not correspond to safety standards. The first category includes dangers that cannot be avoided through management, whereas the second and third categories include dangers that may be eliminated and must be addressed by power companies. A class I hazard source is the material root of an accident that cannot be eliminated [3], whereas a hidden danger is defined as the hazardous state of an object as well as people's risky behaviors and management defects [4] that may cause an accident, including class II and III hazard sources. Addressing these issues is an important way for safety management to play its part and lower the probability of an accident. As a result, it is required to objectively determine the influencing variables of electric power personal injury accidents from the standpoint of standardizing language and focusing on the latter two kinds of danger sources, build a more comprehensive and effective system for the key hidden risk causes of electric power personal injury accidents, and offer a reference for electric power companies to improve the level of production and management safety.

Most scholars and practitioners develop a highly subjective hidden danger factor system using literature study, expert interviews, and questionnaire surveys. However, in recent years, the power industry has developed a reasonably uniform method and system for accident investigation and recording, and a large number of accident reports have been developed, in which many examples of effective accident law and information are hidden. Text information, on the other hand, is less structured, and human processing is time-consuming and labor-intensive. If we can apply text mining technology to mine accident report text, we will be able to eliminate subjectivity and enhance the utilization rate of information.

The following is how this document is structured. Section 2 mostly introduces the research's current status. Section 3 identifies the hidden danger factors of electric power personal injury accidents and provides the factor index method. Section 4 proposes a method for extracting critical hidden hazard indicators from document mining, which involves document pre-processing, word cloud creation, and term frequency counting. Section 5 includes an empirical investigation to validate the scientific nature and effective-ness of the suggested method.

Section 6 then establishes the primary hidden danger factor system for electric power personal casualty accidents based on empirical analysis. Section 7 summarizes the report by discussing the implications and limitations of our findings as well as future research proposals. In summary, the following are the research's main contributions:

- (1) The hidden danger factors of electric power personal injury accidents are preliminarily determined through literature research and based on the definitions of the three types of hazard sources, by removing the factors that cannot be eliminated by management means.
- (2) This study develops a text mining approach for the development of a hidden danger factor system for electric power personal injury accidents, and it picks the essential hidden danger factors based on absolute and relative word frequency. We employ the average maximum TF-IDF value of hidden danger factor proper phrases associated with hidden danger factors as one of the assessment indices during the process.

2. Related Works

2.1. Research on the Extraction of Influencing Factors of Power Accidents

In the study of factors influencing power accidents, both domestic and international experts have produced a wealth of research results. Williamson and Feyer [5] looked at the causes of electrical death at work from four perspectives: the environment, unsafe state of objects, human physiological state, and human unsafe behavior, and then divided the causes of accidents into eight categories: environmental factors, environmental factors, equipment factors, work habits factors, regulatory factors, training factors, work errors factors, physical condition and other factors (including alcohol/drug abuse, etc.). Zhao [6] used a literature review and expert consultation method to create an influencing factor

system with 15 factors to analyze the causes of electrical fatalities in the construction industry in the United States of America, including operation time, task type, safety training, operation space, and voltage level. Chi [7] and others investigated the causes of 255 electrical fatal accidents in Taiwan's construction industry using personal factors (age, gender, and experience), task factors (tasks performed), environmental factors (humid areas and confined spaces), and management factors (company size), as well as accident mode classification. Wei Xiuning [8] and others employed safety behavior, psychology, and accident cause theory to separate the elements impacting human safety in power supply operations into five categories: people, equipment, materials, procedures, and environment. Sun Huiling [9] argued that there are more than 20 risk factors that affect the safety production of power enterprises, and she classified them into five categories, including technical risk, environmental risk, management risk, and so on. The first three types of risk variables are the most important risk variables affecting power plant safety.

Based on the current research findings on the causes and influencing factors of power accidents, it can be concluded that: first, most literatures expound their own viewpoints and select terms at random to describe the causes and influencing factors of power accidents, without distinguishing the specific connotation of these terms, resulting in a lack of unified expression standards in the relevant research. Second, most studies just list and summarize all conceivable accident causes and influencing factors without considering whether or not they can be removed or managed, resulting in an influencing factors system that lacks a practical reference value for safety management practice. Finally, the majority of the literature relies on prior experience, various safety theories, or the use of expert consultation and other subjective qualitative approaches to determine the causes and influencing aspects of the accident. They do not extract information from the accident report text, which contains a wealth of useful data, and they do not use big data mining technology effectively.

2.2. Research on Identifying Accident-Influencing Factors Based on Text Mining

In the process of power production, electric power firms have amassed a huge number of accident text data, which is an essential source for companies to research accident rules and summarize work experience. Due to its unstructured or semi-structured characteristics, text information is more difficult to fully utilize its data value than digital information. As a result, text mining should be used to identify the contributing aspects more scientifically and efficiently.

In many areas of accident investigation, domestic and international researchers have made significant progress in identifying influencing elements utilizing text mining technology. Li Jie and Wang Jianping [10] chose 151 subway construction safety risk accident reports as text data from 2002 to 2015. They performed text preprocessing using the R language and text mining methods, such as word segmentation, feature extraction, spatial vector model development, and so on. The primary risk elements and general risk factors of subway construction safety were mined, and the results of text mining were visualized using visualization methods including word clouds and network structure diagrams, providing useful reference information for future subway construction safety. In 2018, Wu Chen and Jiang Fucai [11] used the R language and text mining algorithm to obtain a set of high-dimensional feature space vectors, then successfully reduced the dimension using statistical methods, and successfully constructed a Bayesian network structure of a "ship environment management" system. The difficulty of being unable to recognize professional jargon is solved by enhancing the TF-IDF algorithm, and the confidence rate of risk factors for ship collision accidents is enhanced. Human factors are thought to be the leading cause of ship collision accidents. Zhang Lu et al. [12] analyzed 41,791 hidden danger records of a coal mining enterprise, using a word cloud and TF to extract coal mine safety risks. From 156 accident reports, Li et al. [13] developed a lexicon and used document frequency (i.e., DF value) to identify 15 high-occurring safety risk variables and three participants.

After reviewing prior research, it was discovered that text mining technology has not yet been fully used in the electricity industry, particularly in China. Moreover, the existing research has not taken into account the factors that can help to eliminate the hidden hazard. As a result, we proposed a three-step strategy for the construction of a major hidden danger factors system applicable to China's electric power personal injury occurrences. The first stage is to identify hidden danger elements that can be eliminated through management by conducting literature research and integrating it with the three-class theory of danger sources. The second step is to design the text mining process, which includes document preprocessing, word cloud building, and term frequency counting, as well as to screen the most important hidden danger factors using the parameters of hidden danger factor proper terms in the text that are related to the hidden danger factors obtained in the first step. The system of the main hidden danger factors of power personal injury accidents is developed in the last step, using the results of the previous literature research and text mining.

3. Preliminary Establishment of Hidden Danger Factor Index System of Electric Power Personal Casualty Accidents

3.1. Analysis on Influencing Factors of Electric Power Personal Casualty Accidents

The relevant literature is first analyzed in this paper to summarize the most common causes and influencing factors of electric power personal casualty accidents. There are 24 related works (1992 to 2019) that have been collected and analyzed, including 14 Chinese and 10 English papers. The study and statistical results of the causes and influencing factors of accidents can be found in Table 1.

Table 1. Analysis and Statistical Results of Causes and Influencing Factors of Electric Power Personal Casualty Accidents in the Related Literature.

Factor Category	actor Category No. Specific Causes and Influencing Mentioned Paper Factors of Accidents		Quantity of Mentioned Paper	
	1	Gender	Literature [6,7,14–16]	5
	2	Nationality	Literature [14]	1
	3	Age	Literature [6-8,14-20]	10
	4	Education background/education level/intelligence level	Literature [15,19,21,22]	4
	5	Working years/work experience level/familiarity/qualifications	Literature [7,8,14,15,17–21,23]	10
	6	Time for current work	Literature [20]	1
	7	Contract type	Literature [15]	1
	8	Work skill level/knowledge level/ability and quality	Literature [7–9,15,17,19,21–28]	15
Personnel factor	9	Working habit	Literature [5–9,15,17,19,21,22,26–30]	16
	10	Absenteeism	Literature [19]	1
	11	Handedness	Literature [19]	1
	12	Personality trait/character trait	Literature [20,24-26,28]	5
	13	Physical health/physical condition	Literature [5,8,15,17,19–27,29,31]	15
	14	Lifestyle	Literature [5,16,19,28]	4
	15	Mental state	Literature [8,15,17,19–27]	12
	16	Work mood/psychological condition/psychological quality	Literature [8,17,19,21–27,29,31]	12
	17	Communication level/team cooperation/interpersonal relationship/coordination degree	Literature [9,15,19,21,23,24,27,29]	8

Factor Category	No.	Specific Causes and Influencing Factors of Accidents Mentioned Paper		Quantity of Mentioned Paper		
	1	Equipment design and quality/equipment state	Literature [5,7,9,15,17,18,20, 21,26,27,29,30]	12		
Equipment factor	2	Equipment nameplate quality/safety sign	Literature [15,21,30]	3		
2quipinent incloi	3	Equipment location/site layout	Literature [21,27,29,30]	4		
	4	Control system state	Literature [29]	1		
Tool factor	1	Instrument configuration	Literature [5–9,15,17,18,20,21,26,27,30]	13		
Toornactor	2	Personal protective equipment configuration	Literature [5,7– 9,15,17,18,20,21,27,29,30]	12		
	1	Working height type	Literature [7,8,17,19,20,27]	6		
	2	Work power outage type	Literature [7,8,17,19,27]	5		
	3	Working potential type	Literature [7,8,17,19,27]	5		
	4	Working voltage class	Literature [6]	1		
	5	Working equipment type	Literature [6-9,17,19,20,27]	7		
	6	Working nature/technical environment/technical nature	Literature [6,8,9,14,15,17–20,22,27]	12		
Method factor	7	Working personnel allocation/personnel scale/personnel arrangement/personnel assignment	Literature [8,17,21,27]	6		
	8	Working organization form	Literature [8,17,21,27]	4		
	9	Working resource allocation/resource management condition	Literature [19,22]	2		
	10	Working difficulty	Literature [21,27]	2		
	11	Working duration/workload/working intensity/working pressure	Literature [8,17,19–21,27]	6		
	12	Working time	Literature [5,6,8,16–19,24]	8		
	13	Working instruction basis	Literature [5,6,8,9,15,17,20–24,27,29,30]	14		
	1	Working weather	Literature [8,9,15– 17,19,20,26,27,29,32,33]	11		
	2	Working temperature	Literature [15,19–21,23,26,27,29,32]	9		
	3	Working air force	Literature [15,19,20,26,29,32]	6		
	4	Working humidity	Literature [5,7,15,19,20,23,29,32]	8		
	5	Working air quality/dust	Literature [15,19,20,26,29]	5		
	6	Working ventilation	Literature [15,19,20,30]	4		
Environmental factor	7	Working noise/interference	Literature [15,19–21,23,26,29]	7		
	8	Working radiation	Literature [15,19,20,26,29]	5		
	9	Working illumination/lighting	Literature [5,15,19–21,27,29,30]	8		
	10	Working area/working site	Literature [5,7–9,15,17–20,27,29]	11		
	11	Working space	Literature [5–9,15,18–20,23,27]	11		
	12	Working atmosphere/safety culture atmosphere/organizational atmosphere/organizational safety commitment	Literature [9,19,21,22,24,29]	6		
	13	Company size	Literature [7,18]	1		
	14	Social environment/policy environment	Literature [30]	1		

Table 1. Cont.

Factor Category	ctor Category No. Specific Causes and Influencing Mentioned Paper Factors of Accidents		Quantity of Mentioned Paper		
	1	Investigation before working/analysis of hazardous points in working/safety inspection/site investigation	Literature [21,23,27,29]	4	
	2	Arrival of work leaders/assignment of safety officers and guardians	Literature [5,6,8,9,15,17– 19,21–23,27,29,30]	14	
	3	Implementation of site safety control measures/effective supervision by work leaders	Literature [8,9,15,19,23]	5	
Management factor	4	Training selection/safety education training	Literature [5,6,15,21,26,29,30]	7	
0	5	Accident handling ability	Literature [29]	1	
	6	Reward and punishment mechanism state/investigation degree of accountability for accidents	Literature [29]	1	
	7	Management consulting service mode	Literature [18]	1	
	8	Project jurisdiction ownership	Literature [15]	1	
	9	Project contract amount	Literature [15]	1	

Table 1. Cont.

All influencing elements may be grouped into six groups, as shown in Table 1, after integrating the study results of associated literatures, which are human factors, equipment factors, tool factors, method factors, environmental factors, and management factors, with 59 unique components. Personnel factors account for the majority of the factors (17 out of 59), with work skill level/knowledge level/ability and quality, working habits, and physical health/physical condition appearing the most frequently; in most of the literature, the arrival of work leaders/assignment of safety officers and guardians in method factors and the arrival of work leaders/assignment of safety officers and guardians in management factors are also considered key aspects.

3.2. Preliminary Identification of Hidden Danger Factors of Electric Power Personal Casualty Accidents

Many of the 59 causes and influencing factors of electric power personal casualty accidents listed in Table 1 are class I hazards that cannot be eliminated by safety management measures, such as working height type and working power outage type, which are determined by the work task and cannot be changed (high-altitude work, for example, necessitates climbing and cannot be accomplished by not climbing or lowering the climbing height.) As a result, the influencing variables of a class I hazard source are not hidden risk elements, according to the definition. The hidden danger elements of electric power personal injury accidents are preliminarily identified and categorized from four aspects of persons, object, environment, and management [4] by analyzing and judging the nature of each factor; see Table 2 for the results.

Table 2 demonstrates that there are 42 hidden danger factors in electric power personalcasualty incidents, which may be split into four groups, with people and management hidden danger elements accounting for the most, totaling 13 each. AP_i means the *i*-th hidden danger factor group and AP_{ij} means the *j*-th hidden danger factor belonging to the *i*-th group. Management refers to the activity process in which managers in a particular organization coordinate other people's activities by implementing the planning, organizing, leading, coordinating, and controlling functions, so that others can achieve the established goals alongside them. In the final analysis, management is carried out by people, so the hidden dangers of personnel and management are essentially those of human factors, with the difference being that the hidden dangers of personnel focus on the unsafe state and behavior of workers, whereas hidden dangers of management focus on the unsafe state and behavior of workers, whereas hidden dangers of management focus on the unsafe state and behavior of workers, whereas hidden dangers of management focus on the unsafe state and behavior of managers and the unsafe systems they established. Furthermore, the hidden danger factors of the object are the least, primarily in four aspects: working equipment, control system, working instruments, and personal protective equipment; the hidden danger factors of the environment comprise natural and humanistic working environment components.

Table 2. Preliminary Identification Results of Hidden Danger Factors of Electric Power Personal Casualty Accidents.

Hidden Danger Factor Category	Specific Hidden Danger Factor		
$AP_i \ (i = 1, 2, \dots, 4)$	AP_{ij} (<i>i</i> = 1,2,,4; <i>j</i> = 1,2,,13)		
	Education background/education level/intelligence level AP_{11}		
	Working years/work experience level/familiarity/qualifications $\overline{AP_{12}}$		
	Time for current work AP_{13}		
	Contract type AP_{14}		
	Work skill level/knowledge level/ability and quality <i>AP</i> ₁₅		
	Working habit AP_{16}		
Hidden danger of personnel <i>AP</i> ₁	Absenteeism AP_{17}		
	Personality trait/character trait AP_{18}		
	Physical health/physical condition AP_{19}		
	Lifestyle $AP_{1\times 10}$		
	Mental state $AP_{1 \times 11}$		
	Work mood/psychological condition/psychological quality $AP_{1\times 12}$		
	Communication level/team cooperation/interpersonal relationship/coordination degree		
	$AP_{1\times 13}$		
	Equipment design and quality/equipment state AP_{21}		
	Equipment nameplate quality/safety sign AP_{22}		
Hidden danger of object AP	Equipment location/site layout AP_{23}		
The definition of object 711 2	Control system state AP_{24}		
	Instrument configuration AP_{25}		
	Personal protective equipment configuration AP_{26}		
	Working weather AP_{31}		
	Working temperature AP_{32}		
	Working air force AP_{33}		
	Working humidity AP ₃₄		
	Working air quality/dust <i>AP</i> ₃₅		
Hidden danger of environment AP_3	Working ventilation AP_{36}		
0	Working noise/interference AP_{37}		
	Working radiation <i>AP</i> ₃₈		
	Working illumination/lighting AP_{39}		
	Working atmosphere/safety culture atmosphere/organizational		
	atmosphere/organizational safety commitment $AP_{3\times 10}$		
	Investigation before working/analysis of hazardous points in working/safety		
	inspection/site investigation AP_{41}		
	Working instruction basis AP_{42}		
	Working organization form AP_{43}		
	Working personnel allocation/personnel scale/personnel arrangement/personnel		
	assignment AP_{44}		
	Working resource allocation/resource management condition AP_{45}		
Hiddon danger of management AP.	Working time AP_{46}		
i nouen uanger of management AP ₄	Working duration/workload/working intensity/working pressure AP_{47}		
	Arrival of work leaders/assignment of safety officers and guardians AP_{48}		
	Implementation of site safety control measures/effective supervision by work leaders		
	AP_{49}		
	Training selection/safety education training $AP_{4\times 10}$		
	Accident handling ability $AP_{4\times 11}$		
	Keward and punishment mechanism state/investigation degree of accountability for		
	accidents $AP_{4\times 12}$		
	Management consulting service mode $AP_{4\times 13}$		

Note: Although factors such as gender, nationality, and age can be controlled through management means (e.g., recruitment conditions), they will not be considered during the development of the hidden danger factor system that can be eliminated through various safety management means in this paper because they are linked to the sensitive social issue of employment discrimination.

4. Extraction Method of Key Hidden Danger Factors of Electric Power Personal Casualty Accidents Based on Document Mining

4.1. Experimental Scheme of Document File

Document information mining has become an important direction of data mining with the application of the Internet and social networks, and it is frequently employed in disciplines such as emotion analysis [34] and public opinion recognition [35].

Figure 1 depicts the experimental scheme for this paper by document mining, which consists of six steps: document collection, document storage, document preprocessing, document mining, analysis and comparison of experimental results, and final establishment of key hidden danger factor system of electric power personal casualty accidents. This paper collects 225 electric power casualty accident reports from the Compilation of National Electric Power Accidents and Electric Power Safety Events (2014 to 2018) [2,36–39] provided by the National Energy Administration's Electric Power Safety Supervision Department; document mining experiments are carried out for the accident brief and cause in each accident report.





The cases are all in Chinese and come from the State Grid Corporation of China. The following are the reasons why we chose the Chinese case: (1) Chinese and English have distinct text characteristics that cannot be mixed during text mining. (2) The safety management models and cultures of power firms varies among countries. Case texts in Chinese are being studied in depth in order to better target the characteristics of China's power safety management.

4.2. Document Preprocessing

1. Due to the redundancy, incompleteness, and complexity of a huge quantity of document data in the electric power personal casualty accident report, it is required to pre-process the content first to increase the effect of document mining [40], which mostly includes Stopword's filtration. In the original accident report document, there are numerous punctuation marks such as ",", ".", and "...", as well as meaningless phrases for experimental purposes such as "of", "very", "limited company", and "power supply station" that should be removed. As a result, in order to filter the document content, a stop list must first be specified.

- 2. Segmentation of Chinese words: some specific words hidden in the whole language in the electric power personal casualty accident report express the precise connotation of the important concealed danger aspects of electric power personal casualty accident. For example, "XiaoX Li, an untrained masonry worker, has weak self-protection awareness and inadequate safety knowledge". It reflects that the accident is related to the work skill level/knowledge level/ability and quality in personnel hidden threats. As a result, word segmentation for the document content using Chinese word segmentation is required, as well as word segmentation for each concealed risk factor's connected terms, in order to prepare for its following characteristic calculation and reduction.
- 3. Create your own dictionary: Chinese sentences are made up of meaningful genuine words and function words that help create the sentence grammar; because there is no space between words, Chinese sentences cannot be subdivided directly by spaces and punctuation between words as English sentences can. As a result, a dictionary or an algorithm is often needed to finish the word segmentation [40]. Jieba, a Python-based word segmentation tool, is utilized in this work to finish the word segmentation of an electric power personal casualty accident report document. A custom dictionary was imported to realize the fixed collocation and combination of words related to various hidden danger factors in order to ensure the accuracy and completeness of the word segmentation results related to various hidden danger factors and avoid the segmentation of proper nouns such as "..... weak/self/protection/consciousness", "self-protection awareness", and "safety awareness" to guarantee that the follow-up trial runs well.

4.3. Word Cloud Building

Word Cloud, also known as word cloud analysis, is a technology that highlights the "keywords" that appear frequently in document data and converts them into cloud-like color pictures, allowing visitors to appreciate or understand the main idea conveyed by the document at a glance, primarily using document mining and visualization technology. Based on the term count result of words directly reflecting the connotation of hidden danger factors, this paper draws a word cloud picture for the accident cause document in the electric power personal casualty accident report and shows the general situation of the important terms related to hidden danger factors; see Figure 2 for the experimental process.



Figure 2. Word cloud of important words related to hidden danger factors of electric power personal casualty accidents.

4.4. Term Frequency Counting

The word frequency count refers to the statistics of how many times a term appears in the document. In general, the absolute term frequency (i.e., term count) counts the number of terms that appear in a document. However, this creates a problem: a term may appear more frequently in a long document than in a short one, but this does not imply that the term is more essential or has more weight. As a result, the normalized term frequency, also known as relative term frequency, is frequently used as the weight value of the term. Word frequency-inverse document frequency (TF-IDF), a popular weighting method for information retrieval and data mining, is based on the basic principle of relative term frequency) refers to the frequency of a certain term appearing in the document and normalizes the number of terms to avoid bias towards large documents, whereas IDF (inverse document frequency) is a measure of the universal relevance of terms [41,42].

Due to the long time span, large number, and numerous writers of the 225 electric power casualty accident reports chosen for this paper, the language style is difficult to unify, the description methods are diverse, and the number of words is quite different, resulting in a disadvantage in the simple term count experiment. Some high-frequency keywords directly expressing the meaning of concealed danger elements, for example, are overstated because they appear frequently in many exceedingly long and inconsequential accident cause documents. This is because their description documents are too short, and some other keywords directly representing the meaning of hidden danger factors may be overestimated, impacting the final screening result of important hidden danger factors. As a result, in this paper, the word count experiment is carried out first on the terms related to hidden danger factors, and then the relevant important terms are screened based on the term count results; additionally, the TF-IDF method is introduced to count the relevant important words with the highest TF-IDF value in each accident cause document and their TF-IDF values. After classification and merging, the comprehensive average maximum TF-IDF value of the relevant important words of each hidden danger factor in 225 electric power personal casualty accident reports can be obtained, reflecting the importance and influence of each hidden danger factor on an electric power personal casualty accident.

The accident cause document set in 225 electric power personal casualty accident reports is considered as set $D = \{D_1, D_2, ..., D_m\}$, and $D_r(r = 1, 2, ..., m)$ means the *r*-th accident cause document in the document set. The term set appearing in the accident cause document set is considered as $W = \{W_1, W_2, ..., W_n\}$, and $W_s(s = 1, 2, ..., n)$ means the *s*-th term in the term set. The proper term set of hidden danger factors appearing in the accident cause document set is considered as $R = \{R_1, R_2, ..., R_l\}$, and $R_t(t = 1, 2, ..., l)$ means the *t*-th proper term of hidden danger factors.

The influence of specific hidden danger factors is calculated and screened using the following steps:

Step (1): calculate the term frequency of the proper term of hidden danger factors R_t :

$$TC_t = \sum_{r=1}^m x_{tr}, t = 1, 2, \dots, l$$
 (1)

 TC_t means the term frequency of R_t in the document set D; x_{tr} means the frequency of the occurrence of R_t in D_r .

Step (2): set a threshold of term frequency to screen the important proper term of hidden danger factors:

The proper terms of hidden danger factors whose term frequencies is greater than threshold (α) will form set $U = \{U_1, U_2, ..., U_p\}$, and $U_k (k = 1, 2, ..., p, p < l)$ means the *k*-th important proper terms of hidden danger factors. The screening discriminant is as follows:

$$\Phi = \begin{cases} R_t \in U, \ TC_t \ge \alpha \\ R_t \notin U, \ TC_t < \alpha \end{cases}, \ t = 1, 2, \dots, l$$
(2)

The threshold is generally empirically set according to the text analysis field. If TC_t exceeds the threshold, R_t will be considered as an important proper term of hidden danger factors and be classified into set U; otherwise, it will be removed.

Step (3): determine whether there is a logical correlation between the important proper term of hidden danger factors (U_k) and each specific hidden danger factor (AP_{ij}) in Table 2, the determination formula is:

 $A_{ijk} = \begin{cases} 1, \text{ there is a logical correlation between } U_k \text{ and } AP_{ij} \\ 0, \text{ there is not a logical correlation between } U_k \text{ and } AP_{ij} \end{cases}, k = 1, 2, \dots, p; i = 1, 2, \dots, 4; j = 1, 2, \dots, 13$ (3)

Step (4): calculate the term frequency of important proper terms of hidden danger factors related to each specific hidden danger factor (AP_{ij}) :

$$C_{ijk} = \sum_{k=1}^{p} A_{ijk}, i = 1, 2, \dots, 4; \ j = 1, 2, \dots, 13$$
(4)

 C_{ijk} means the number of important proper terms of hidden danger factors that have a logical correlation with AP_{ij} .

$$Q_{ijk} = \sum_{k=1}^{p} TC_k \cdot A_{ijk}, \ i = 1, 2, \dots, 4; \ j = 1, 2, \dots, 13$$
(5)

 Q_{ijk} means the term frequency of important proper terms of hidden danger factors related to AP_{ij} ; TC_k means the term frequency of U_k in the document set D.

For a specific hidden danger factor AP_{ij} , the larger Q_{ijk} is, the more influence the specific hidden danger factor AP_{ij} will have on electric power personal casualty accidents.

Step (5): calculate the *TF-IDF* value of each important proper term of hidden danger factors (U_k) in each accident cause text (D_r) :

$$TF_{kr} = \frac{x_{kr}}{\sum_{s=1}^{n} x_{sr}}, \ k = 1, \ 2, \ \dots, \ p; \ r = 1, \ 2, \ \dots, \ m$$
(6)

 TF_{kr} means the *TF* value of the *k*-th important proper term of hidden danger factors (U_k) in the *r*-th accident cause document (D_r) ; x_{kr} means the frequency of the occurrence of U_k in the *r*-th accident cause document; $\sum_{s=1}^{n} x_{sr}$ means the number of terms in the *r*-th accident document.

$$IDF_{k} = \log_{10} \frac{m}{|\{r: U_{k} \in D_{r}\}|}, \ k = 1, \ 2, \ \dots, \ p$$
(7)

 IDF_k means the IDF value of the *k*-th important proper term of hidden danger factors (U_k) , *m* means the number of documents in the accident cause document set, and $|\{r : U_k \in D_r\}|$ means the number of accident cause documents containing the *k*-th important proper term of hidden danger factors (U_k) .

$$TF - IDF_{kr} = TF_{kr} \times IDF_k \tag{8}$$

TF- IDF_{kr} means the TF-IDF value of the k-th important proper term of hidden danger factors (U_k) in the r-th accident cause document.

Step (6): determine whether the *TF-IDF* value of each important proper term of hidden danger factors (U_k) in each accident cause text D_r is the largest,

$$B_{kr} = \begin{cases} 1, \ TF - IDF_{kr} \ge TF - IDF_{qr}, q \in \{1, 2, \dots, p\} \\ 0, \ \text{otherwise} \end{cases}, \ k = 1, 2, \dots, p; \ r = 1, 2, \dots, m \qquad (9)$$

Step (7): calculate the average highest *TF-IDF* value of every important proper term of hidden danger factors (U_k) that have correlation with each specific hidden danger factor AP_{ij} in the accident cause text set,

$$\overline{TF-IDF_{ij}^{max}} = \frac{\sum_{r=1}^{m} \sum_{k=1}^{p} TF-IDF_{kr} \cdot B_{kr} \cdot A_{ijk}}{\sum_{r=1}^{m} \sum_{k=1}^{p} B_{kr} \cdot A_{ijk}}, \ i = 1, 2, \dots, 4; j = 1, 2, \dots, 13$$
(10)

 $\overline{TF-IDF_{ij}^{max}}$ means the comprehensive average maximum *TF-IDF* value of important proper terms of hidden danger factors.

It has been found that if an important term related to a hidden danger factor appears more frequently in a specific accident cause document and less frequently in the entire set of accident cause

documents, the hidden danger factor has a greater importance and influence on the occurrence of the specific accident.

5. Analysis on Document Mining Results of Electric Power Personal Casualty Accident Report

5.1. Data Sources

As stated in Section 4.1, the accident text used in this study is a compilation of 225 electric power personal casualty accident reports from the national compilation of electric power accidents and electric power safety incidents (2014 to 2018) compiled by the national energy administration's electric power safety supervision department. Each article is divided into three sections: a brief account of the accident, the causes of the event, and ideas on how to rectify it. As a result of the unified subject compiling the compilation set and the state having relevant provisions on the terms and format of power accident report writing, the 225 texts may better adapt to the text mining method and process of this study. This study begins with a simple word count of accident type phrases in the accident description section to determine the number distribution of various sorts of accidents. The text of the accident's cause is then mined using word cloud and TF-IDF technologies.

5.2. Document Mining of Accident Brief

The term count experiment on the accident brief in 225 electric power personal casualty accident reports (2014 to 2018) was carried out using the document mining method; see Figure 3 for the results.



Figure 3. Statistical results of the main types of electric power personal casualty accidents (2014 to 2018).

As illustrated in Figure 3, the six most common types of accident from 2014 to 2018 were falling, electric shock, mechanical injury, collapse, object striking, and derrick falling down, accounting for 87 percent of all accident kinds. As the frequency of falling and electric shock accidents is significantly higher than that of other accident kinds, how to effectively limit the occurrence of these two categories of accidents is the emphasis and challenge of accident prevention work in power enterprises.

5.3. Document Mining of Accident Causes

5.3.1. Word Cloud of Accident Hidden Danger Factors

The relevant phrases connected to the concealed hazard variables as accident causes in 225 electric power personal injury accident reports (2014 to 2018) are graphically represented using the word cloud analysis approach described in Section 4.3; see Figure 4 for the drawn word cloud. Terms such as "absence", "work leader", "management", "safety management", "supervisor", "inspection", and "violation" appear most frequently and are most frequently used in accident cause documents to explain the accident cause, and their frequent occurrence also represents the high possibility of related hidden danger factors in common accidents.

operation regulations illegal command audits safety management agemen' operation regulations in management index water to be a second and the second and illegal work pre technology accident hidden danger strict enforcement range performance Supervison egulations education low security awareness ndard regulation ct No unauthorized belt conveyor oduct safe operation regulations flaw safety hidden danger temporary were and the find out in the set of the plan and the set of the set OSENCE adventure plan scaffold Р safety management supervisor construction plan safety measures supervision et safety education and training hidden wire wet safety consciousness sat safety belt serious violation of regulations operation

Figure 4. Term cloud of important terms related to hidden danger factors of electric power personal casualty accidents (2014 to 2018).

5.3.2. Screening Results and Analysis of Key Hidden Danger Factors Based on Term Count and TF-IDF

In order to address the inherent defects in the collected 225 electric power personal casualty accident reports (2014 to 2018), the approach of term count combination with TF-IDF developed in Section 4.4 was used: first, the term count threshold for hidden risk factors ws set to 5, and 389 essential terms linked to hidden danger factors were screened from α them; second, three characteristic values of the key keywords associated with each concealed hazard factor were calculated: the count of the important terms related to each hidden danger factor (C_{ijk}), the comprehensive cumulative frequency of the important terms related to each hidden danger factor (Q_{ijk}), and the comprehensive average maximum *TF-IDF* value of the important terms related to each hidden danger factor ($\overline{TF-IDF_{ijk}^{max}}$); finally, based on the overall distribution of the experimental results, the threshold was set as follows: for a specific hidden danger factor AP_{ij} (i = 1, 2, ..., 4; j = 1, 2, ..., 13), if $C_{ijk} \ge 10$, $Q_{ijk} \ge 200$ and $\overline{TF-IDF_{ijk}^{max}} \ge 0.1$, it will be considered as the key hidden danger factor of electric power personal casualty accident. See Table 3 for the screening results.

i	j	AP_{ij}	C_{ijk}	Q_{ijk}	TF-IDF ^{max} _{ijk}
$1 $ $\frac{1}{2}$	1	Work skill level/knowledge level/ability and quality	105	1486	0.2156
	2	Working habit	179	2958	0.2354
2	1	Equipment design and quality/equipment state	47	727	0.3602
	2	Equipment nameplate quality/safety sign	21	344	0.2948
	3	Instruments configuration	13	211	0.2666
	4	Personal protective equipment configuration	35	493	0.2102
3	1	Investigation before working/analysis of hazardous points in working/safety inspection/site investigation	82	1621	0.2319
	2	Working instruction basis	109	2231	0.2238
	3	Arrival of work leaders/assignment of safety officers and guardians	93	1925	0.2119
	4	Implementation of site safety control measures/effective supervision by work leaders	72	1767	0.1829
	5	Training selection/safety education training	37	989	0.1793
		Average	72	1341	0.2375

Table 3. Screening Results of Key Hidden Danger Factors of Electric Power Personal Casualty Accidents Based on TermFrequency Count.

In Table 3, for a specific hidden danger factor AP_{ij} (i = 1, 2, 3; j = 1, 2, ..., 5), if Q_{ijk} is greater with the increasing of C_{ijk} , it shows that that hidden danger factor appears frequently in general electric power personal casualty accidents; if $\overline{TF-IDF_{ijk}^{max}}$ is greater, it demonstrates that such a hidden danger element has a stronger impact on a personal injury catastrophe using electric power. The primary hidden hazard variables of electric power personal casualty incidents filtered out in this paper include two features based on the assumption of setting a threshold: first, hidden danger factors are present in the majority of electric power personal casualty accidents, emphasizing their prevalence; second, the appearance of the hidden danger factor has a significant impact on the occurrence of an electric power personal casualty accident, emphasizing its harmfulness. The risk concept is the combination of these two characteristics; based on the formula (risk = possibility influence), it can be seen that the occurrence possibility and harmfulness of the working habit (a key hidden danger factor) in the personnel hidden danger are above average, which belongs to the high-risk key hidden danger factor and is the "top priority" in the key hidden danger factor system.

6. Establishment of Key Hidden Danger Factor System of Electric Power Personal Casualty Accidents Based on Empirical Analysis

In this paper, the key hidden danger factor system of electric power personal casualty accidents is ultimately formed, the names of key hidden danger factors are defined and standardized, and their specific connotations are reinforced; see Table 4 for specific contents.

Table 4. Key Hidden Danger Factor System of Electric Power Personal Casualty Accident.

Hidden Danger Factor Category	Key Hidden Danger Factors	Specific Connotation	
Hidden danger of personnel	Safety quality and ability of workers	Means workers' safety awareness ability and safety production skill level (including the work skills, skills mastering working safety devices and facilities, and knowledge and skills of proper handling and emergency rescue in case of emergencies)	
rinden amger of personaer	Executive ability of rules and regulations of workers	Means whether the workers have good working habits, whether they follow numerous power sector safety production rules and regulations while working, and whether they have habitual violations	
	Equipment design and quality	Means whether the design parameters of the equipment fulfill industry and national standards, whether the commissioning time of the equipment exceeds the maximum working life, and whether the equipment has component damage and function loss	
Hidden danger of object	Equipment nameplate configuration and quality	Means whether the equipment nameplate is missir damaged or unclear	
	Configuration and quality status of operation instruments	Means whether the instruments are missing, unqualit and damaged during working	
	Personal protective equipment configuration and quality	Means whether the personal protective equipment is missing, unqualified, damaged, or exceeds the shelf life during working	
	Site investigation of managers	Means whether the managers have conducted site investigation before working, and whether the site investigation is accurate	
	Formulation of working instruction basis	Means whether the managers have developed the necessary working instruction base (containing three measures and one scheme, labor discipline, working drawings, various rules and regulations, and so on), and whether the formulation is accurate.	
Hidden danger of management	Arrival of managers	Means whether the managers arrive and whether the perform guardianship duties on the worksite	
	Implementation of site safety control measures by managers	Means whether the managers have implemented safety control measures during working and whether the implementation is accurate	
	Safety education training of enterprise	Means whether electric power firms have provided safety education training to their personnel as part of their daily safety management job, and whether the effect has been positive	

Table 4 shows that, in contrast to other industrial fields where production accidents are frequently attributed to personnel, object, environment, and management, the key hidden danger factors of electric power personal casualty accidents extracted based on document mining in this paper exclude the hidden danger factors of the environment and only focus on hidden dangers of personnel, object, and management. This is exactly consistent with the actual situation in which many electric power enterprises adjust working arrangements based on environmental weather conditions (e.g., not working on hot days or heavy rain and typhoon days). Therefore, the hidden danger factors of the environment can be ignored.

7. Conclusions

The significant information in the accident report has not been effectively exploited in the safety management of China's power system. The first step toward realizing the digitalization, automation, and intelligence of power grid safety management is to extract important information from power grid accident reports using machine learning and text mining. This study builds an index system for hidden danger factors based on text mining of power grid safety accident reports based on the current state of China's power grid safety management, employing safety management theory and text mining methods. This method not only reduces the subjectivity of identifying and judging potential safety threats, but it also lays the groundwork for the future implementation of digital and intelligent safety management.

The following are the limitations of this study. As the majority of electric power personal injury accidents are minor or minor-level accidents, and there are few significant accidents, the 225 electric power personal injury accidents collected were minor or minor-level accidents (level 4). Based on text information in Chinese, we only focused on power grid safety management in Chinese.

Author Contributions: Conceptualization, D.L. and C.M.; methodology, C.M. and Y.W.; software, C.Z.; validation, X.X.; formal analysis, Y.W.; resources, C.X.; data curation, D.L.; writing, Y.W. and C.Z.; supervision, C.M.; project administration, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the science and technology project of the headquarters of the State Grid Corporation of China (52170018000T).

Conflicts of Interest: The authors declare no conflict of interest.

References

- State Council of the PRC. Report on Production Safety Accident and Regulations of Investigation and Treatment. 28 March 2007. Available online: http://www.gov.cn/zwgk/2007-04/19/content_588577.htm (accessed on 4 June 2020).
- 2. Electric Power Safety Supervision Department of National Energy Administration. *Compilation of National Power Accidents and Power Security Incidents (2014)*; China Power Media Group Co., Ltd.: Beijing, China, 2015.
- 3. Tian, S.C. Study on Identification and Control of the Third Type of Hazard; Beijing Institute of Technology: Beijing, China, 2001.
- 4. State Administration of Work Safety. Interim Provisions on the Investigation and Control of Safety Accidents. 28 December 2007. Available online: https://www.mem.gov.cn/fw/flfgbz/gz/200801/t20080110_233408.shtml (accessed on 4 June 2020).
- 5. Williamson, A.; Feyer, A.M. The causes of electrical fatalities at work. J. Saf. Res. 1998, 29, 187–196. [CrossRef]
- 6. Zhao, D.; Thabet, W.; McCoy, A.; Kleiner, B. Electrical deaths in the US construction: An analysis of fatality investigations. *Int. J. Inj. Control. Saf. Promot.* **2014**, *21*, 278–288. [CrossRef]
- Chi, C.F.; Yang, C.C.; Chen, Z.L. In-depth accident analysis of electrical fatalities in the construction industry. *Int. J. Ind. Ergon.* 2009, 39, 635–644. [CrossRef]
- 8. Wei, X.N.; Liu, Y.G.; Huang, Y.L.; Chen, G.S.; Xiao, J.H. Prediction on human casualty hazard in power-supplying operation based on multiple linear regression. *J. Saf. Sci. Technol.* **2015**, *11*, 178–184.
- 9. Sun, H.L. Study on Risk Management of the Electric Power Enterprise Production Safety; Qingdao University of Technology: Shandong, China, 2018.
- 10. Li, J.; Wang, J.P.; Xu, N.; Zhou, Z. Analysis of safety risk factors for metro construction based on text mining method. *Tunn. Constr.* **2017**, *2*, 33–42.
- 11. Wu, J.; Jiang, F.C. An analysis and risk forecasting of inland ship collision based on text mining. J. Transp. Inf. Saf. 2018, 3, 154–161.
- 12. Tan, Z.L.; Chen, X.; Song, Q.Z.; Chen, X.C. Analysis for the potential hazardous risks of the coal mines based on the so-called text mining. *J. Saf. Environ.* 2017, 4, 1262–1266.
- 13. Feng, J.; Gong, C.; Li, X.; Lau, R.Y. Automatic approach of sentiment lexicon generation for mobile shopping reviews. *Wirel. Commun. Mob. Comput.* **2018**, *10*, 232–240. [CrossRef]
- 14. Castillo-rosa, J.; Suarez-cebador, M.; Rubio-romero, J.C.; Aguado, J.A. Personal factors and consequences of electrical occupational accidents in the primary, secondary and tertiary sectors. *Saf. Sci.* **2017**, *91*, 286–297. [CrossRef]
- 15. Cheng, C.W.; Leu, S.S.; Lin, C.C.; Fan, C. Characteristic analysis of occupational accidents at small construction enterprises. *Saf. Sci.* **2010**, *48*, 698–707. [CrossRef]
- 16. Dokov, W. Assessment of risk factors for death in electrical injury. Burns 2009, 35, 114–117. [CrossRef]
- 17. Yang, Y.Q. Evaluation Model of Power Operation of Personal Risk Quantification; South China University of Technology: Guangdong, China, 2015.
- Suarez-cebador, M.; Rubio-romero, J.C.; Lopez-arquillos, A. Severity of electrical accidents in the construction industry in Spain. J. Saf. Res. 2014, 48, 63–70. [CrossRef] [PubMed]
- 19. Khanzode, V.V.; Maiti, J.; Ray, P.K. Occupational injury and accident research: A comprehensive review. *Saf. Sci.* **2012**, *50*, 1355–1367. [CrossRef]

- 20. Paivinen, M. Electricians' perception of work-related risks in cold climate when working on high places. *Int. J. Ind. Ergon.* **2006**, *36*, *661–670*. [CrossRef]
- 21. Tang, J.X. Research on Human Reliability Analysis and Accident Pre-control of Power Grid; Zhejiang University: Zhejiang, China, 2015.
- 22. Zhang, J.W. Research on Human Error of Power Accidents Based on HFACS; Soochow University: Jiangsu, China, 2016.
- 23. Ren, C.Y. Application on Safety Management of Human Factors Engineering Theory in Electric Power Company; Tianjin University: Tianjin, China, 2016.
- 24. Wu, S.S. Research and Application on safety Human Factors in Electricity Enterprises; Beijing Jiaotong University: Beijing, China, 2013.
- 25. Wu, S.P. Study on the Influencing Factors and Management of Human Errors in the Production of Electricity Enterprises; Beijing Jiaotong University: Beijing, China, 2009.
- 26. Yang, Z.L. *Electric Power Production Safety Risk Causes and Pre-Control Method Research;* North China Electric Power University: Beijing, China, 2016.
- 27. Ye, J.G. Safety Risk Identification and Control. System of Power Supply Enterprise; North China Electric Power University: Beijing, China, 2014.
- 28. Mellen, P.F.; Weedn, V.W.; Kao, G. Electrocution: A review of 155 cases with emphasis on human factors. *J. Forensic Sci.* **1992**, 37, 1016–1022. [CrossRef]
- 29. Zhu, L.N. Research on the Early Warning of Production Safety Accidents in Power Generation Enterprises Based on Bayesian Network; North China Electric Power University: Beijing, China, 2019.
- 30. Editorial board of "Construction and Practice of Intrinsic Safety Capability of Power Grid Enterprises". *Construction and Practice of Intrinsic Safety Capability of Power Grid Enterprises*; China Electric Power Press: Beijing, China, 2018.
- Shen, Y.L.; Yang, Z.X.; Li, H.L.; Zhang, K.; Liu, Y.Z. A review of the safety psychology and physiology related research in electric power human accident. *Telecom World* 2019, 26, 368–369.
- 32. Li, Y.B.; Han, Y.; Zhang, R.; Li, Y. Research on impact model of meteorological factors on the power accidents. *Power Syst. Technol.* **2013**, *37*, 1683–1687.
- 33. Yu, X.; Xu, C.; Lu, D.; Zhu, Z.; Zhou, Z.; Ye, N.; Mi, C. Design and application of a case analysis system for handling power grid operational accidents based on case-based reasoning. *Information* **2020**, *11*, 91. [CrossRef]
- 34. Yang, Y.J.; Yuan, H.H.; Wang, Y.L. Sentiment analysis method for comment text. J. Nanjing Univ. Sci. Technol. 2019, 43, 280–285.
- 35. Tan, J.Y.; Ma, S.C. Research on network microblogging lyrics topic recognition and tracking technology based on clustering. *J. Chongqing Univ. Technol.* **2019**, *33*, 176–181.
- 36. Electric Power Safety Supervision Department of National Energy Administration. *Compilation of National Power Accidents and Power Security Incidents (2015)*; China Power Media Group Co., Ltd.: Beijing, China, 2016.
- 37. Electric Power Safety Supervision Department of National Energy Administration. *Compilation of National Power Accidents and Power Security Incidents (2016)*; China Power Media Group Co., Ltd.: Beijing, China, 2017.
- 38. Electric Power Safety Supervision Department of National Energy Administration. *Compilation of National Power Accidents and Power Security Incidents (2017)*; China Power Media Group Co., Ltd.: Beijing, China, 2018.
- 39. Electric Power Safety Supervision Department of National Energy Administration. *Compilation of National Power Accidents and Power Security Incidents (2019)*; China Power Media Group Co., Ltd.: Beijing, China, 2019.
- 40. Huang, C.N.; Zhao, H. Chinese word segmentation: A decade review. J. Chin. Inf. Process. 2007, 3, 8–19.
- 41. Yang, B.; Han, Q.W.; Lei, M.; Zhang, Y.P.; Liu, X.G.; Yang, Y.Q.; Ma, X.F. Short text classification algorithm based on improved TF-IDF weight. *J. Chongqing Univ. Technol.* **2016**, *30*, 108–113.
- 42. Zhang, J. A Method of intelligence key words extraction based on improved TF-IDF. J. Intell. 2014, 33, 153–155.