WILEY | Hindawi

Research Article

Correlation Analysis of External Environment Risk Factors for High-Speed Railway Derailment Based on Unstructured Data

Haixing Wang [ID]**, Yuanlanduo Tian** [ID]**, and Hong Yin** [ID]

School of Traffic & Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Haixing Wang; hxwang@bjtu.edu.cn

In railway operation, unsafe events such as faults may occur, and a large number of unsafe event records are generated in the process of unsafe events' recording and reporting. Unsafe events have been described in unstructured natural language, which often has inconsistent structure and complex sources, involving multiple railway specialties, with multisource, heterogeneous, and unstructured characteristics. In practical application, the efficiency of processing is extremely low, leading to potentially unsafe management utilization. Based on the data on unsafe events, this paper utilizes big data processing technology, conducts association rules mining and association degree analysis, extracts the word segmentation, and obtains the feature vector of unsafe fault event data. At the same time, the unsafe event data analysis model is constructed in combination with regular expression and pattern matching technology. This paper establishes the matching model of high-speed railway derailment-based external environment risk factors and applies it to the occurrence of unsafe events. This model could be utilized to analyze and excavate the link between external environment risk factors and the occurrence of unsafe events and carry out the automatic extraction of characteristic information such as risk possibility and consequence severity; hence, it has potential for identifying, with enhanced accuracy, high-risk factors that may lead to high-speed railway derailment. Based on this study, we could make full use of the unsafe event data, forecast the risk trend, and discover the law of high-speed railway derailment. This study introduces a viable approach to analyzing the unsafe event data, forecasting risk trend, and conceptualizing high-speed railway derailment. It could also enforce the accurate quantification of high-speed railway safety situation, refine the risk index and conduct in-depth analysis combined with the model, and effectively support the digitalization and intellectualization of high-speed railway operation safety.

1. Introduction

The occurrence of high-speed railway derailment accident may result in severe financial and human losses, which have significant disaster characteristics and strong nonlinear characteristics and pose a challenge to high-speed railway safety management. Technological advances have helped to mitigate the internal factors behind railway derailment, but the external factors remain an underexplored area of research.

Advances in computer technology have benefited large-scale numerical calculation by enhancing operation speed, storage capacity, and operation scale. Big data analysis, furthermore, has been building on progress in accuracy, quality, and reliability and has become a major area of academic interest in the context of high-speed railway safety.

To minimize risks that may lead to unsafe events, the railway corporation has accordingly built the reporting system to keep the records, and it has generated a large amount of data thereof. Figure 1 exemplifies a typical record of an unsafe event.

However, the records on unsafe events, which are now described in natural language as shown above, could be incorporated into a digital database, and it could not be established without a consistent description standard. At the same time, the railway infrastructure, rolling stock, and other equipment are diverse and complicated, and the data sources are complex, involving many specialties, such as track maintenance, power supply equipment maintenance, signal and communication equipment maintenance, EMU maintenance, passenger transportation, and external environment. The database covers a wide range of faults' fixed

10:17 on April 18, 2021, at 191 km 188 m on XX line of XX railway corporation the mountain watcher inspected and found that the mountain on the right side of 191 km 201 m was cracked, and about 3 cubic meters rockfall intruded into the clearance (the largest one was 1.1 * 0.9 * 1.2 m). The mountain watcher immediately informed the relevant personnel to block the line. At 10:21, the line was blocked. After being handled by the track maintenance staff, the line was opened at 11:17. The incident affected the train traffic and caused the train no. XXX to change its parking place. The rockfall at 191 km 201 m, and this site is the straight line section of the line, half dike and half cutting section, and the right side is 4 meters away from the track center of the line.

FIGURE 1: Example of unstructured unsafe event data of railway.

equipment faults, mobile equipment faults, perimeter intrusion events, and others. A sizable amount of data have been gathered. In practice, the information needs to be retrieved, read, and updated manually, which costs a huge amount human and material resources, and the processing efficiency is considerably low, resulting in the low utilization rate of railway unsafe event data.

For the requirements of railway safety management and risk analysis for real time and accuracy to be met, it is urgent to carry out railway risk effect factors association analysis based on unstructured data, while taking into account different data structures, different sources, and scattered and independent records of railway unsafe events. Through this research, the heterogeneous data sources could be integrated, the heterogeneity of data can be eliminated, and the accurate association between data and railway risk could be achieved. Therefore, this paper will take the external environment risk of high-speed railway derailment as the research focus and carry out a correlation analysis of external environment risk factors of high-speed railway derailment based on unstructured data. The study seeks to identify the risk factors of high-speed railway derailment and select appropriate models to process unstructured data, including data collection, data cleaning, data dictionary construction, data extraction, data storage, and other steps. Accordingly, the association between data and risk occurrence possibility and consequence severity will be realized. The paper then moves on analyzing the high-speed railway derailment risk factors and effective extraction of unstructured data.

In recent years, with the breakthrough of big data and artificial intelligence technology, a number of investigations have been carried out regarding the railway safety index, unstructured data analysis, and multisource heterogeneous railway safety data identification and extraction. In terms of railway safety index, since 2015, the International Union of Railways (UIC) has started to build the global safety index (GSI) [1]. Based on safety data and accident information, it evaluates the safety level of railway in Europe, some Asian, and Middle Eastern countries and regions and analyzes the statistical data of safety accidents, the impact of accidents, and the safety level and development trend. Zhao et al. [2] established a railway accident index by measuring the occurrence frequency and consequence severity of railway accidents, which is used to evaluate the overall situation of China's railway safety. In the aspect of unstructured data analysis and mining, Zhang [3] built an unstructured data analysis platform based on report documents with Chinese

word segmentation technology, unstructured data extraction method, pattern matching, and other methods. Zhu et al. [4] put forward a new HGD tree index technology and a new partition method, in order to use probability density function to partition data and improve the speed of data access, and gave a solution based on the optimization operation method. Wang et al. [5, 6] analyzed and studied the safety data of dangerous goods transportation based on the data mining method. In the aspect of multisource heterogeneous railway safety data extraction and data analysis, Wang et al. [7] conducted quantitative analysis on railway derailment and the change of accident rate based on American railway safety data. Lin. et al. [8] analyzed the data of American trunk line passenger trains and quantitatively analyzed the causes of passenger train accidents. Liu et al. [9] analyzed the causes of major train derailment and their effect on accident rates. Turla et al. [10] analyzed the freight train collision risk in the United States. Li [11] recognized and extracted the fault features of high-speed railway equipment by establishing the +bilstm and +CRF method for character representation and the +transformer method for word segmentation representation. Zhou and Li [12] established a method of fault data feature recognition and extraction for railway signal equipment based on MCNN.

To sum up, previous studies on high-speed railway risks mostly employ the expert evaluation method, which is arguably based on subjective deliberation and may undermine the research validity. Besides, past investigations mainly focus on feature recognition, extraction for specific structure of safety data, and the processing of unstructured data. As such, there still lacks research on the correlation analysis between safety data and risk. This study proposes a data-driven risk judgment method for analysis to facilitate accurate association between data and railway risk, with implications that the proposed model can contribute to improving the feasibility and accuracy of risk judgment.

2. Analysis of External Environment Risk Associated Factors

To explore the derailment mechanism of high-speed railway, this paper establishes a dynamic derailment-related element model of high-speed railway. As shown in Figure 2, the derailment of high-speed railway is mainly related to EMU subsystem and line subsystem, and the coupling relationship between wheel and rail has an important impact on the derailment of high-speed railway. In addition, the external
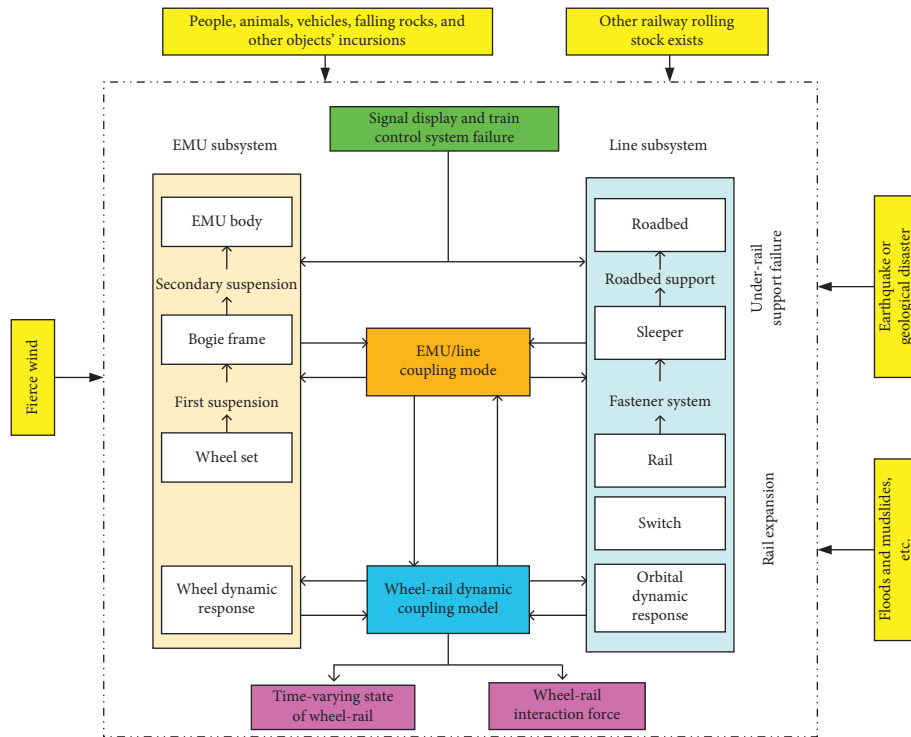
FIGURE 2: Dynamic derailment related element model of high-speed railway.

environmental factors such as natural geology and perimeter intrusion factors also play an important role in the derailment of high-speed railway. The external environmental safety for high-speed railway derailment has been widely investigated and is the focus of the current study [13].

Based on the above analysis, the paper puts forward the framework model of risk associated external environmental factors for high-speed railway derailment, which fall into two categories:: natural and geological factors (natural hazard factors and geological hazard factors) and perimeter intrusion factors (perimeter intrusion of animals, perimeter intrusion of objects and plants, and perimeter intrusion of people), as shown in Figure 3, and the classification is detailed in Table 1 [14–17].

3. Unstructured Data Mining Method

The raw data stem from a database of a railway company that has kept its records of, as many as, 15,000 past unsafe events is obtained.

3.1. General Analysis. At the same time, based on the scientific analysis for the unsafe event data, the study combines regular expression and pattern matching technology and establishes the matching model of external environmental factors for high-speed railway derailment risk associated unsafe events. This paper analyzes and mines the relationship between the external environmental factors of high-speed railway derailment and the unsafe events, automatically, quickly, and accurately extracts the key characteristic information such as the possibility of risk occurrence and the

severity of the consequences, so as to transform unstructured data into structured information. The main data analysis and mining process is shown in Figure 4. The process includes (1) unstructured railway safety data, (2) split and match keywords, and (3) association rules' mining and association degree analysis. Through the process, the accurate association between data and railway risk could be acquired [18–20].

3.2. Keyword Extraction and Matching. If you want to extract keywords in the text, it is relatively simple for English and other languages. Keywords' extraction can be achieved in a number of languages, including English. In the case of English, for example, there are spaces between words as segmentation. In Chinese, however, such expressions are unavailable, so it is necessary to break coherent sentences into keywords. Expressions in Chinese may vary widely, leading to potential ambiguity in the word segmentation. In keyword extraction and matching, the railway safety dictionary is designed, and the algorithms and models such as tire tree, DAG, Viterbi, HMM (hidden Markov model), and keyword matching are comprehensively used. The main processing of keyword extraction and matching is shown in Figure 5 [21–23].

3.3. Association Analysis-Based Apriori Algorithm. This study utilizes the association rules to explore the correlations between data generated by different mode methods, so as to build rules that may inform the decision-making.

The data mining of association rules mainly includes two processes. First, identify the frequent item sets whose
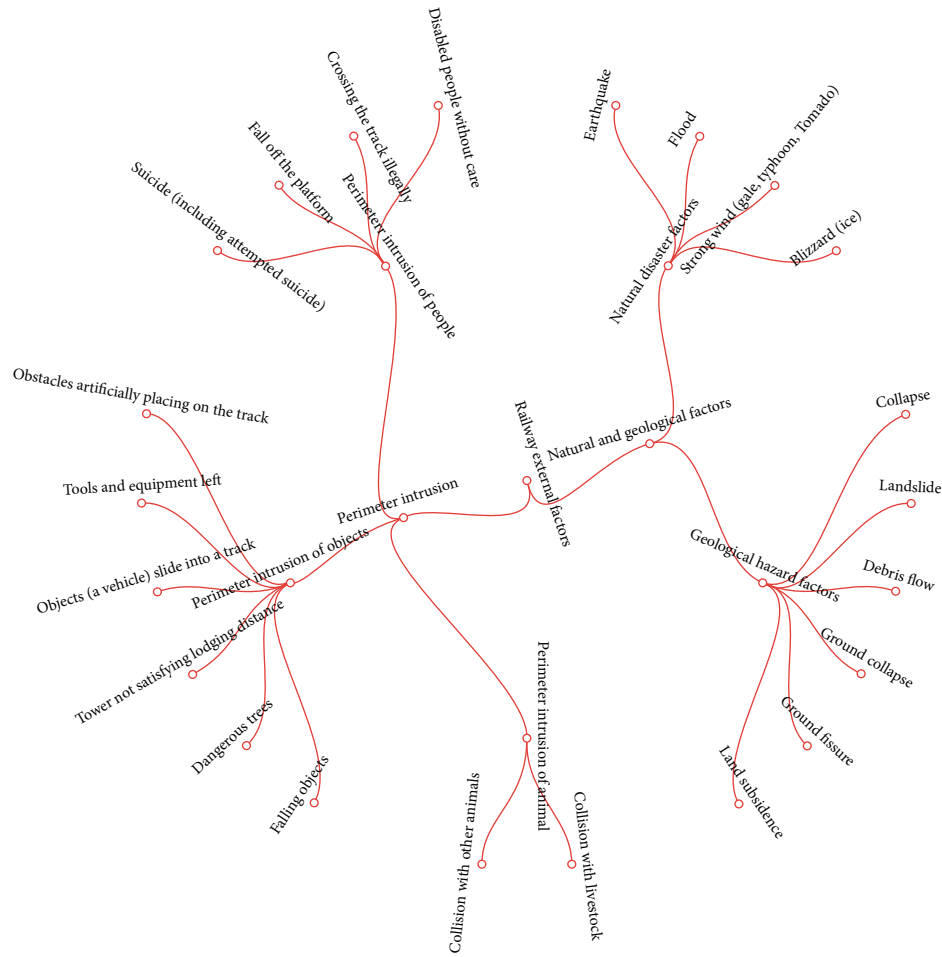
FIGURE 3: Framework model of external environmental factors associated with high-speed railway derailment.

TABLE 1: Main external environmental factors associated with derailment of high-speed railway.

| No.	Serial number	Class I factors	Class II factors	Name of related factors
1	Z11	Natural and geological factors	Natural hazard factors	Earthquake
2	Z12			Flood
3	Z13			Strong wind (gale, typhoon, and tornado)
4	Z14			Blizzard (ice)
5	Z21		Geological hazard factors	Collapse
6	Z22			Landslide
7	Z23			Debris flow
8	Z24			Ground collapse
9	Z25			Ground fissure
10	Z26			Land subsidence
11	Y11	Perimeter intrusion factors	Perimeter intrusion of animals	Collision with livestock
12	Y12			Collision with other animals
13	Y21		Perimeter intrusion of objects and plants	Falling objects
14	Y22			Dangerous trees
15	Y23			Tower not satisfying lodging distance
16	Y24			Objects (a vehicle) slide into a track
17	Y25			Tools and equipment left behind
18	Y26			Obstacles artificially placing on the track
19	Y31		Perimeter intrusion of people	Suicide (including attempted suicide)
20	Y32			Fall off the platform
21	Y33			Crossing the track illegally
22	Y34			Disabled people without care
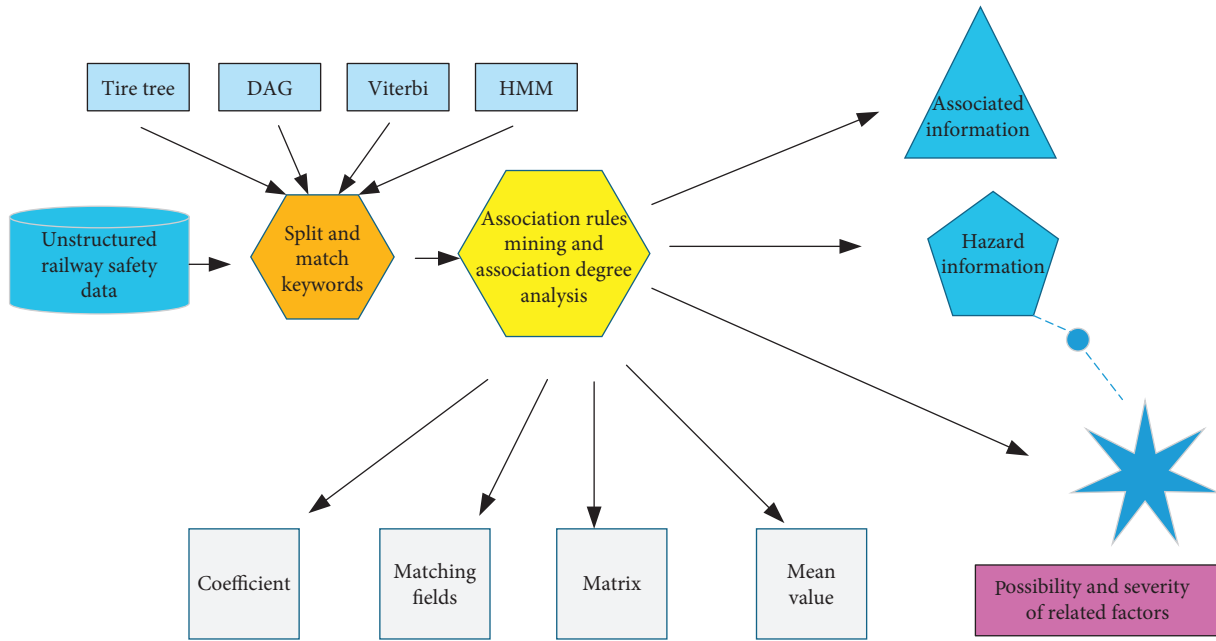
FIGURE 4: Data analysis and mining process.
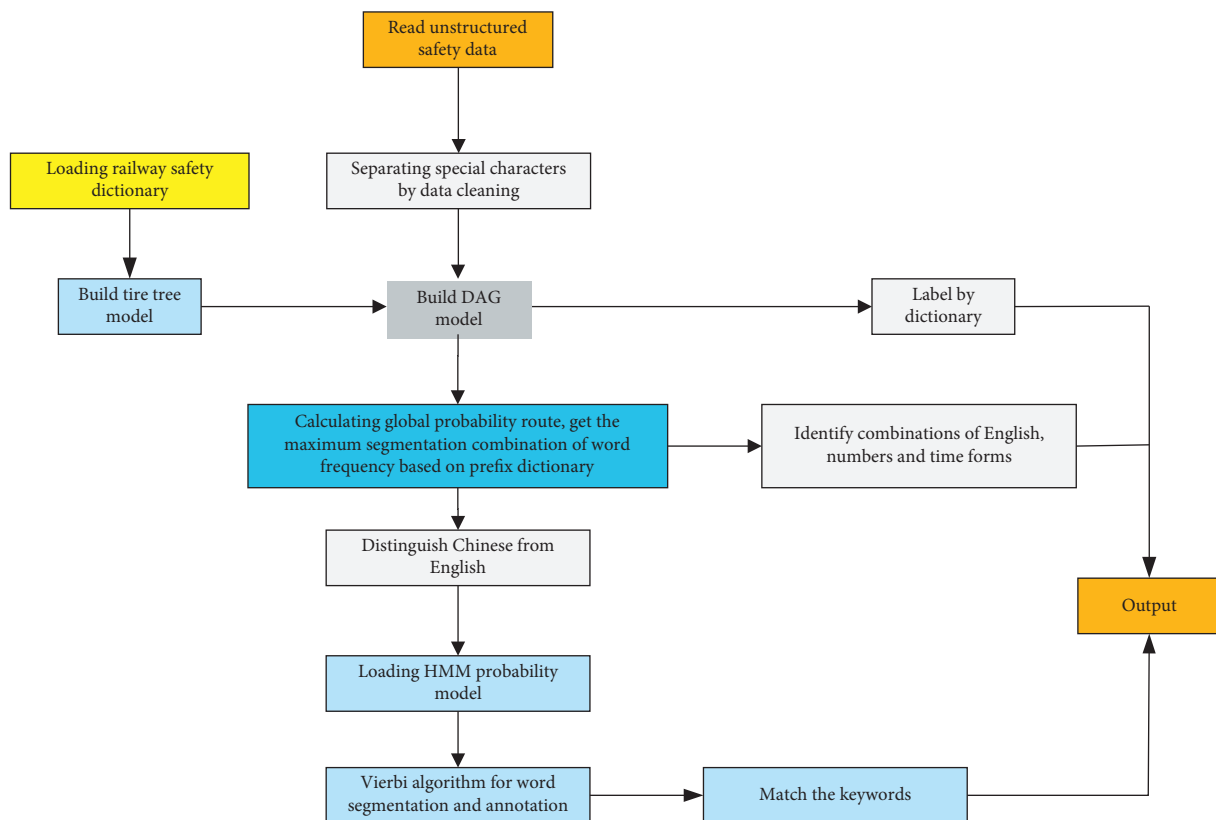
FIGURE 5: Keyword extraction and matching method.

frequency is not less than the minimum support degree of all item sets. Second, conduct mining strong association rules that satisfy the minimum confidence based on the frequent item sets obtained. The overall performance of association

rule data mining is determined by the operation of the previous process [24–28].

Finding frequent item sets is not easy because the data explosion involved in the calculation process may

TABLE 2: Sample data set of Apriori algorithm.

| No.	Key words				
1	Error	Violation	Conflict	Class D accident	
2	Violation	Discipline	Derailment	Class D accident	
3	Violation	Collision	Class D accident		
4	Error	Violation	Train delay	Class D accident	
5	Violation	Dispatch command	Train delay	Class D accident	
6	Equipment	Control	Error	Train delay	Class D accident
7	Vehicle	Construction machinery	Train delay	Class D accident	
8	Error	Loading and unloading	Train delay	Class D accident	
9	Equipment	Fault	Train delay	Class D accident	
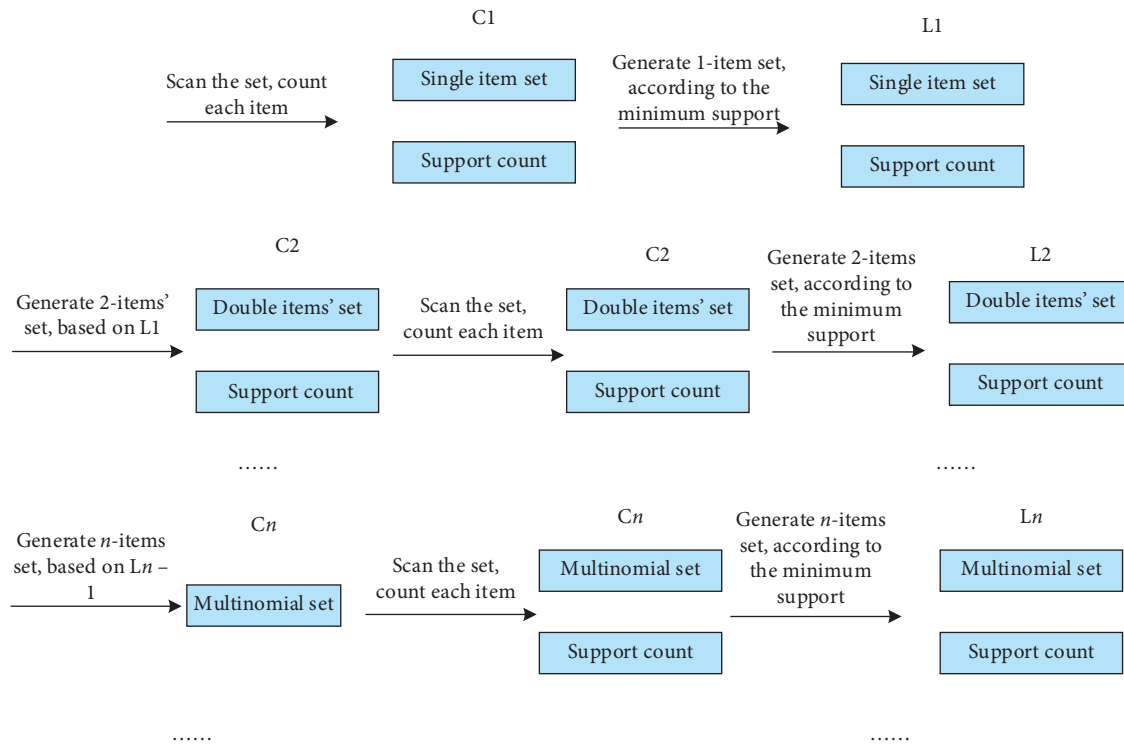
FIGURE 6: Frequent item sets' mining method.

lead to unacceptable computational complexity. However, as long as frequent item sets are obtained, association rules whose confidence is not less than minimum confidence could be explored. The association rules' mining algorithm used in this paper is Apriori algorithm. The following data in Table 2 is an example to illustrate the implementation process of Apriori algorithm [29–32].

Apriori algorithm could be used to mine association rules. The process of mining frequent item sets is shown in Figure 6.

L_k: frequent item sets of length k

C_k: candidate item set of length k

Support_count (k): the support count of k-item sets

It is concluded that all item sets of L1, L2, and Ln are frequent item sets, and then, the confidence of each frequent item set is calculated. When the support threshold is set to

TABLE 3: Calculation results of Apriori algorithm example.

Relationship	Support (%)	Confidence (%)
Train delay-class D accident	66.67	100
Error-class D accident	44.44	100
Class D accident-train delay	66.67	66.67

40% and the confidence threshold is set to 50%, the results shown in Table 3 can be obtained.

3.4. Grey Relation Analysis. Grey relation analysis (GRA) is a multifactor statistical analysis method. The basic method of calculating the correlation degree is to initialize the original data sequence, then calculate the correlation coefficient, get the correlation degree and the correlation matrix through the combination of the correlation coefficient, and finally sort them according to the correlation degree calculation results of each correlation factor sequence [33].
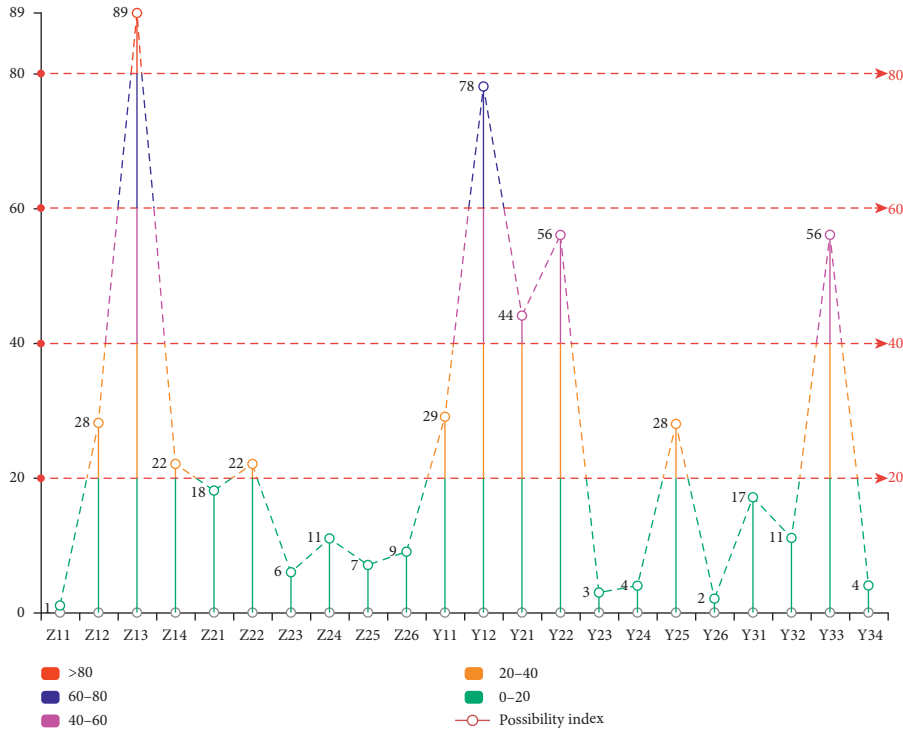
FIGURE 7: Possibility of external environment factors associated with high-speed railway derailment.

The calculation method is as follows.

Let the characteristic behavior sequence of the system be

$$X_i = (x_i(1), x_i(2), \ldots, x_i(n)). \tag{1}$$

Because the units or initial values of each data sequence are different, in order to make them comparable, it is necessary to implement dimensionless processing on the original data so that the data of different dimensions (or magnitudes) could be compared, and the initial value method could be used for calculation:

$$X_i' = \frac{X_i}{x_i(1)} = (x_i'(1), x_i'(2), \ldots, x_i'(n)), \quad i = 0, 1, 2, \ldots, m. \tag{2}$$

Correlation coefficient refers to the degree of correlation based on the geometric shape and development trend of each factor sequence. The expression is as follows:

$$\gamma_{0i}(k) = \frac{\min_i\min_k\left|x_0'(k) - x_i'(k)\right| + \rho\,\max_i\max_k\left|x_0'(k) - x_i'(k)\right|}{\left|x_0'(k) - x_i'(k)\right| + \rho\,\max_i\max_k\left|x_0'(k) - x_i'(k)\right|}, \quad \rho \in (0, 1),\ k = 1, 2, \ldots, n,\ i = 1, 2, \ldots, m. \tag{3}$$

Among them, ρ is called resolution coefficient. The smaller the ρ value is, the greater the resolution is.

The characteristic of association sequence is that it has a huge amount of data. When the information is processed in a centralized way, it is necessary to summarize the association coefficients at different positions of different times into a specific value and calculate their average value. The average value obtained is the correlation degree. The expression is as follows:

$$\gamma_{0i}(k) = \frac{1}{n}\sum_{k=1}^{n}\frac{\min_i\min_k\left|x_0'(k) - x_i'(k)\right| + \rho\,\max_i\max_k\left|x_0'(k) - x_i'(k)\right|}{\left|x_0'(k) - x_i'(k)\right| + \rho\,\max_i\max_k\left|x_0'(k) - x_i'(k)\right|}, \quad \rho \in (0, 1),\ k = 1, 2, \ldots, n,\ i = 1, 2, \ldots, m. \tag{4}$$

4. Risk Analysis Based on Unstructured Data

This paper attempts to analyze the risk of external environment associated with high-speed railway derailment from the possibility and the severity of the consequences and realizes the scientific measurement of the risk by mining the possibility and the severity of the consequences of the unsafe events related to the external environment risk factors associated with high-speed railway derailment. The determination of the occurrence possibility is mainly based on the
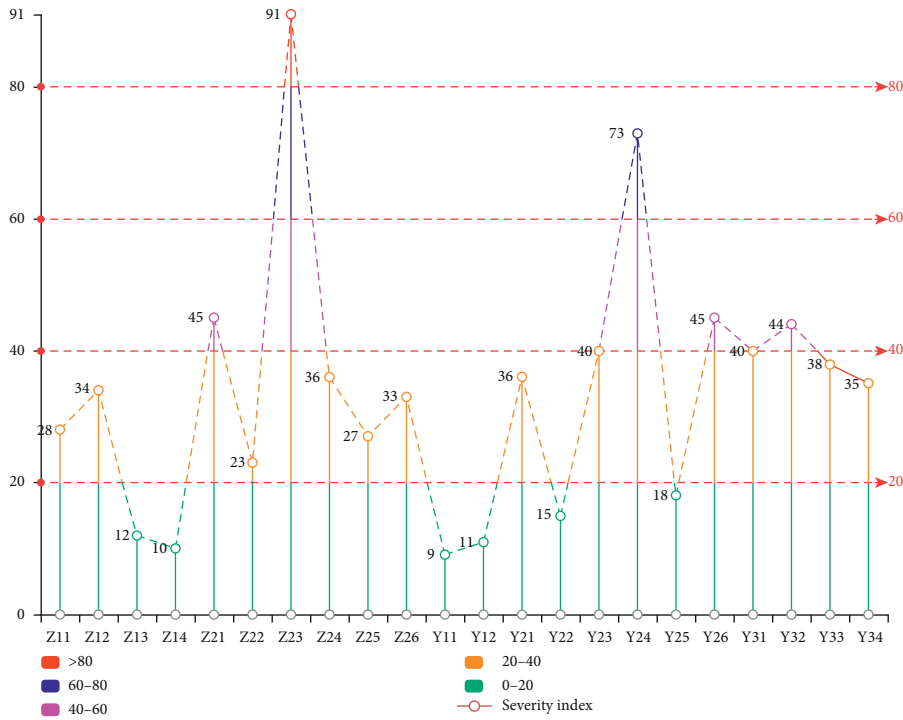
Figure 8: Severity of external environment factors associated with high-speed railway derailment.
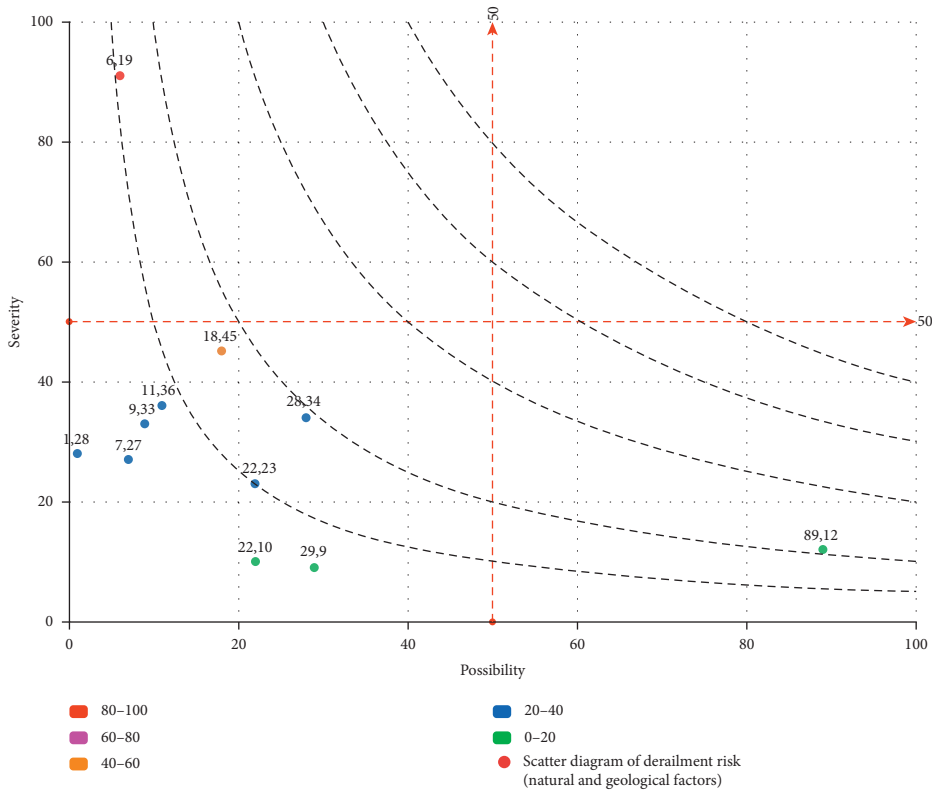

Figure 9: Risk distribution of natural and geological factors associated with high-speed railway derailment.

unstructured safety data mining and then accurately associates the unsafe events with the corresponding risk factors. It is realized by accumulating the occurrence frequency of the unsafe events associated with the risk factors outside the

high-speed railway derailment. Because the basic data involved in the study is mainly unsafe event data, the consequence is mainly the interruption time, so the severity of the consequence is mainly considered to mine the
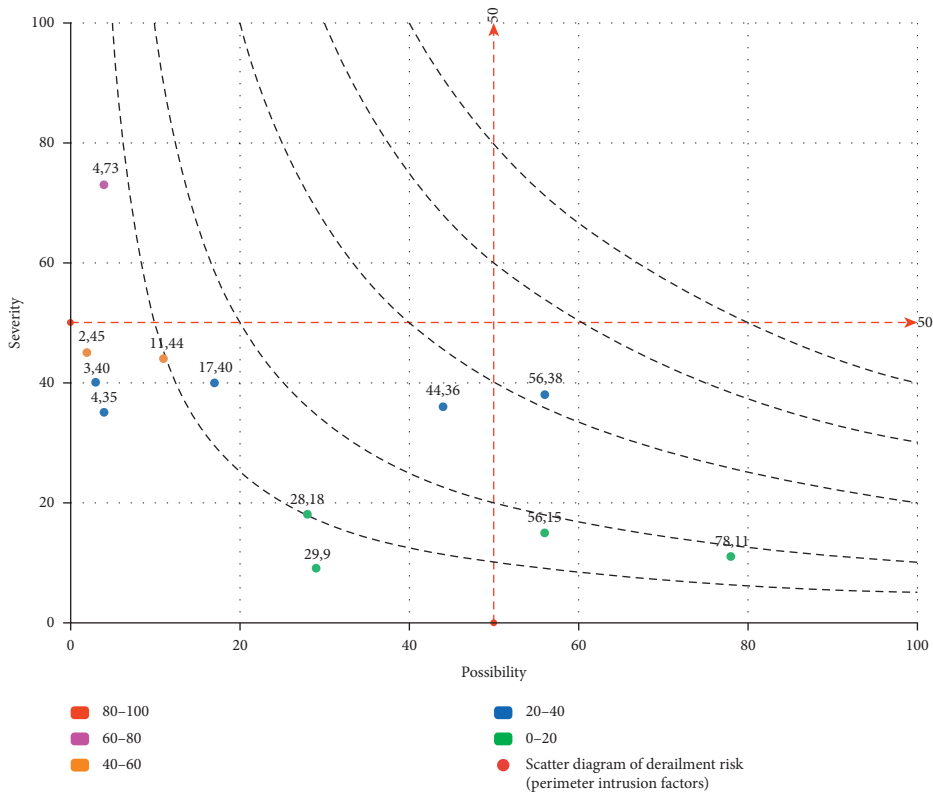
FIGURE 10: Risk distribution of perimeter intrusion factors associated with high-speed railway derailment.
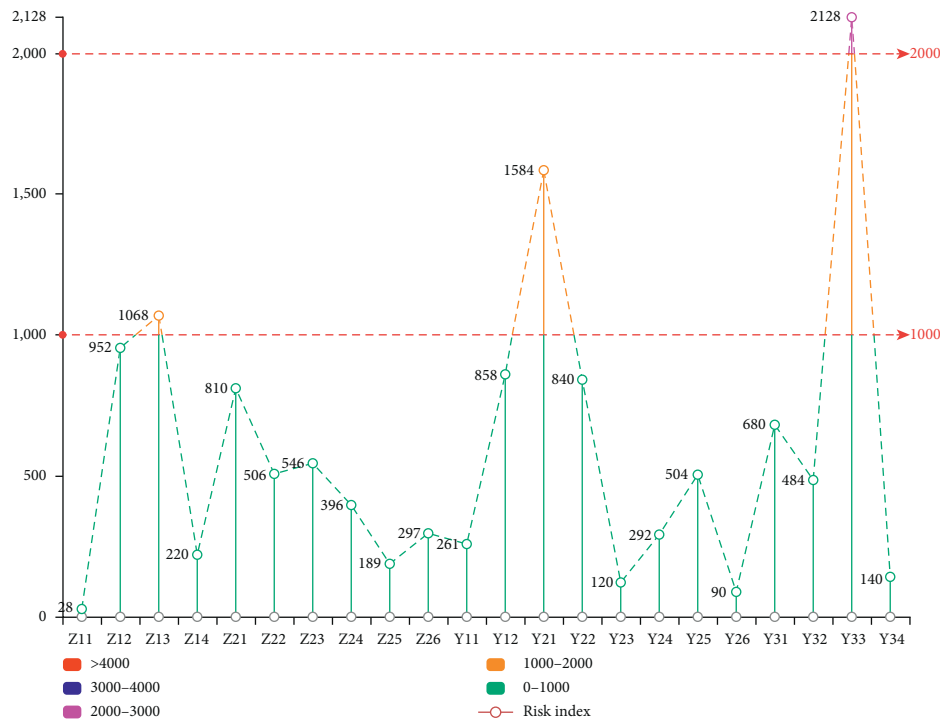

FIGURE 11: Distribution of comprehensive risk index of external environment factors associated with high-speed railway derailment.

interruption time caused by the unsafe events associated with the risk factors of high-speed railway derailment [34–37].

For a small number of factors with low probability of occurrence, they may not be associated with events in the data. In this case, we could consider using some evaluation

methods based on expert experience as a supplement, such as analytical hierarchy process. Based on the above method, after analysis, mining, and unifying the dimensions, the probability of external environment factors associated with high-speed railway derailment is shown in Figure 7, the consequence severity of external environment factors associated with high-speed railway derailment are shown in Figure 8, the risk distribution scatter diagram of natural and geological factors is shown in Figure 9, and the risk distribution scatter diagram of perimeter intrusion factors is shown in Figure 10.

The distribution of comprehensive risk index of external environment factors related to high-speed railway derailment is shown in Figure 11. It could be seen that Z13, Y21, and Y33 are high-risk factors, especially Y33. Risk management procedures should be implemented, and targeted measures should be taken to control them. When implementing control measures, we should pay attention to the actual effect and offer feedback to the implementation to ensure the full implementation of control measures [38–41].

5. Conclusion

Utilizing data on railway fault unsafe events, this paper establishes a matching model that builds correlations between unsafe events and external environment factors in the context of high-speed railway derailment. Operating in an automatic fashion, the model may be employed to analyze and mine the relationship between external environment factors of high-speed railway derailment and unsafe events. The model may also be used, with an enhanced accuracy for identifying high-risk elements, to extract the key feature information such as risk possibility and consequence severity.

The current investigation contributes to the field of research by introducing a statistical method for analyzing unsafe events' data. It seeks to identify high-risk elements of high-speed railway derailment, refines external environment risk index for high-speed railway derailment, and analyzes data in combination with the proposed model. As such, the study achieves a dynamic display of the results arising from external environment risk analysis in the context of high-speed railway derailment. The study is significant in which it seeks to rationalize the methods for analyzing external environment risk and to better visualize the safety laws of high-speed railway derailment. Therefore, the study helps to advance the operation of high-speed railway and its safety arrangements towards a more digitalized and smarter system.

Previous studies on high-speed railway risks mostly employ the expert evaluation method, which is arguably based on subjective deliberation and may undermine the research validity. The current project proposes a data-driven method of risk judgment, which may help to advance the feasibility and accuracy of the analysis. This study contributes to improving the safety level of railway operation by putting forward a method for integrating heterogeneous data sources, minimizing data heterogeneity, and thus, with enhanced accuracy, and building the association between the data and railway risks.

According to the needs of high-speed railway operation safety management, with the continuous accumulation of railway unsafe event data, the external environment risk model related to high-speed railway derailment could be continuously modified and improved, and the correlation matching between risk and unsafe event could be more accurate, which could ensure the continuous improvement of high-speed railway operation safety.

Data Availability

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] UIC, *UIC Safety Report: Significant Accidents 2017 Public Report*, UIC, Paris, France, 2018.

[2] L. Zhao, W. D. Wang, H. M. Xu, and Z. Guo, "Research on construction method of railway accident index in China," *China Railway*, vol. 2020, no. 5, pp. 29–32, 2020, in Chinese.

[3] Y. L. Zhang, *Information Extraction Method of Unstructured Pathological Report Based on Pattern Matching*, Donghua University, Shanghai, China, 2017.

[4] R. Zhu, B. Wang, X. C. Yang, and G. R. Wang, "Research on supporting probabilistic data range query index in big data environment," *Acta Computa Sinica*, vol. 39, no. 10, pp. 1929–1946, 2016.

[5] H. X. Wang and Q. Q. Liang, "Risk analysis and route optimization of dangerous goods transportation based on the empirical path set," *Journal of Advanced Transportation*, vol. 2020, Article ID 8838692, 13 pages, 2020.

[6] H. X. Wang, X. Y. Wang, Z. X. Wang, and X. D. Li, "Clustering analysis of dangerous goods transportation risk driving behavior based on data mining, Transportation System Engineering and Information," vol. 20, no. 1, pp. 183–189, 2020.

[7] B. Z. Wang, C. P. L. Barkan, and M. R. Saat, "Quantitative analysis of changes in freight train derailment causes and rates," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 11, pp. 1–15, 2020.

[8] C. Y. Lin, M. R. Saat, and C. P. L. Barkan, "Quantitative causal analysis of mainline passenger train accidents in the United States," *Journal of Rail and Rapid Transit*, vol. 234, pp. 1–16, 2019.

[9] X. Liu, M. R. Saat, and C. P. L. Barkan, "Analysis of causes of major train derailment and their effect on accident rates," *Transportation Research Record*, vol. 2289, no. 1, pp. 154–163, 2012.

[10] T. Turla, X. Liu, and Z. P. Zhang, "Analysis of freight train collision risk in the United States," *Journal of Rail and Rapid Transit*, vol. 233, no. 9, pp. 817–830, 2019.

[11] X. Q. Li, *Research on Big Data Analysis Method of High Speed Railway Safety Text*, China Academy of Railway Sciences, Beijing, China, 2020.

[12] Q. H. Zhou and X. L. Li, "Research on short text classification method of railway signal equipment fault based on MCNN," *Journal of Railway Science and Engineering*, vol. 16, no. 11, pp. 2859–2865, 2019.

[13] W. G. Xie, J. Liu, T. X. Chen, G. Yang, and Y. J. Liu, "Research on railway safety big data analysis platform," *Railway Computer Application*, vol. 29, no. 9, pp. 21–25, 2020.

[14] W. Dong, "Analysis of factors affecting railway safety and preventive measures," *Science and Technology Innovation Guide*, vol. 16, no. 14, pp. 165-166, 2019.

[15] X. Li, "Analysis of railway safety accidents and counter-measures," *Henan Science and Technology*, vol. 2020, no. 7, pp. 96–98, 2020.

[16] R. X. Wu, *Research on Acoustic Fault Diagnosis Method of Train Bearing Track Based on Two-Stage Learning Model*, Anhui University, Hefei, China, 2020.

[17] Y. H. Ren, "Research on fault diagnosis of railway signal equipment," *Management and Technology of Small and Medium Sized Enterprises*, vol. 2020, no. 8, pp. 162-163, 2020.

[18] H. Yang, Y. J. Yu, G. W. Chen, Y. B. Si, and D. F. Xing, "Research on fault diagnosis algorithm of turnout based on CNN-GRU model," *Acta Sinica Sinica*, vol. 42, no. 7, pp. 102–109, 2020.

[19] A. P. Han, J. Y. Kan, X. Y. Jiang, W. D. Wang, D. J. Fu, and P. Liu, "Construction of China high speed railway electrical safety index system," *China Railway*, vol. 2020, no. 5, pp. 33–38, 2020.

[20] J. F. Guo, Z. T. Ke, W. D. Wang et al., "Calculation method of power supply failure index of high speed railway based on data modeling," *China Railway*, vol. 2020, no. 5, pp. 39–45, 2020.

[21] Y. J. Liu, T. C. Yang, W. Wu, Q. M. Liu, and W. D. Wang, "Design and application of high speed railway operation safety law analysis system based on big data technology," *China Railway*, no. 9, pp. 28–33, 2020.

[22] X. W. Mi and S. Zhao, "Wind speed prediction based on singular spectrum analysis and neural network structural learning," *Energy Conversion and Management*, vol. 216, Article ID 112956, 2020.

[23] X. W. Mi, H. Liu, and Y. Li, "Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine," *Energy Conversion and Management*, vol. 180, pp. 196–205, 2019.

[24] X. Liu, "Analysis of collision risk for freight trains in the United States," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2546, no. 1, pp. 121–128, 2016.

[25] R. T. Anderson and C. P. L. Barkan, "Railroad accident rates for use in transportation risk analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1863, no. 1, pp. 88–98, 2004.

[26] H. X. Wang, G. P. Xiao, and Z. Wei, "Optimizing route for hazardous materials logistics based on hybrid ant colony algorithm," *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 752830, 6 pages, 2013.

[27] T. X. Chen, W. G. Xie, G. Yang, E. C. Su, and Y. J. Liu, "Design scheme of railway safety integrated management system based on safety big data application platform," *Railway Computer Application*, vol. 30, no. 1, pp. 39–43, 2021.

[28] C. F. Lu and D. Wang, "Technical system and safeguard measures for safe operation of high speed railway in China," *Chinese Journal of Safety Sciences*, vol. 29, no. 8, pp. 1–9, 2019.

[29] G. L. Kang, "Research and practice of China's high speed railway security system," *Journal of Railways*, vol. 39, no. 11, pp. 1–7, 2017.

[30] Y. M. Jia, L. Nie, H. X. Wang, and H. F. Liu, "Research on CRSMS architecture," *China Railway*, vol. 2020, no. 1, pp. 43–47, 2020.

[31] Y. X. Fan, G. Fu, Y. W. Zhu, and X. Wang, "Review on the emergence and development of safety management system," *Chinese Journal of Safety Science*, vol. 25, no. 8, pp. 3–9, 2015.

[32] P. Zhang, J. H. Liu, L. V. Feng, and X. Y. Li, "Analysis and enlightenment of European and American railway safety management mode," *Railway Transportation and Economy*, vol. 36, no. 11, pp. 40–44, 2014.

[33] K. L. Wen and C. Chang, "Data preprocessing for grey relational analysis," *The Journal of Grey System*, vol. 11, no. 2, pp. 139–141, 1999.

[34] J. Sun and H. J. Gao, "Analysis on construction and implementation of civil aviation safety management system in China," *China Railway*, vol. 2012, no. 4, pp. 10–14, 2012.

[35] J. Li, "Research on acceptable safety level of civil aviation industry in China," *Science and Technology of Work Safety in China*, vol. 6, no. 5, pp. 137–142, 2010.

[36] H. Liu, X. Mi, and Y. Li, "An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm," *Renewable Energy*, vol. 123, pp. 694–705, 2018.

[37] X. W. Mi, H. Liu, and Y. F. Li, "Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm," *Energy Conversion and Management*, vol. 151, pp. 709–722, 2017.

[38] M. R. Saat and J. Aguilar Serrano, "Multicriteria high-speed rail route selection: application to Malaysia's high-speed rail corridor prioritization," *Transportation Planning and Technology*, vol. 38, no. 2, pp. 200–213, 2015.

[39] M. R. Saat, C. J. Werth, D. Schaeffer, H. Yoon, and C. P. L. Barkan, "Environmental risk analysis of hazardous material rail transportation," *Journal of Hazardous Materials*, vol. 264, pp. 560–569, 2014.

[40] M. R. Saat and C. P. L. Barkan, "Generalized railway tank car safety design optimization for hazardous materials transport: addressing the trade-off between transportation efficiency and safety," *Journal of Hazardous Materials*, vol. 189, no. 1-2, pp. 62–68, 2011.

[41] X. Liu, M. Rapik Saat, and C. P. L. Barkan, "Freight-train derailment rates for railroad safety and risk analysis," *Accident Analysis & Prevention*, vol. 98, no. 1, pp. 1–9, 2017.