

## Research Article

# Risk Chain and Key Hazard Management for Urban Rail Transit System Operation Based on Big Data Mining

Yongsheng Tang 

Academic Affairs Office, Shanghai University of Engineering Science, Shanghai 201620, China

Correspondence should be addressed to Yongsheng Tang; tangyongsheng@sues.edu.cn

Received 28 October 2021; Revised 12 November 2021; Accepted 15 November 2021; Published 6 December 2021

Academic Editor: Gengxin Sun

Copyright © 2021 Yongsheng Tang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the promotion of the national transportation power strategy, super large operation networks have become an inevitable trend, and operational safety and risk management and control have become unavoidable problems. Existing safety management methods lack support from actual operational and production data, resulting in a lack of guidance of fault cause modes and risk chains. Large space is available to improve the breadth, depth, and accuracy of hazard source control. By mining multisource heterogeneous operation big data generated from subway operation, this study researches operation risk chain and refined management and control of key hidden dangers. First, it builds a data pool based on the operation status of several cities and then links them into a data lake to form an integrated data warehouse to find coupled and interactive rail transit operation risk chains. Second, it reveals and analyzes the risk correlation mechanisms behind the data and refines the key hazards in the risk chain. Finally, under the guidance of the risk chain, it deeply studies the technologies for refined control and governance of key hidden dangers. The results can truly transform rail transit operation safety from passive response to active defense, improving the special emergency rail transit operation plans, improving the current situation of low utilization of rail transit operation data, but high operation failure rate, and providing a basis for evidence-based formulation and revision of relevant industry standards and specifications.

## 1. Introduction

In recent years, rail transit has achieved rapid development throughout the world. Incidents on such large rail transit networks are highly likely to cause secondary damage and even derivative disasters, which may propagate in a chained manner. Increased line length and networked construction have caused unprecedented pressure and challenges for rail transit operators, making it imperative to improve safety identification and control techniques to minimize the risk of operation for rail transit development.

Accident analysis is the key to safety management. Major potential risks often precede an incident, and incidents are attributed to chained propagation of multiple risks. The risk chain of urban rail transit operation refers an ordered hazard sequence that finally leads to an unexpected incident during operation because hazards fail to be identified and controlled in time and propagate sequentially, achieving a chain effect.

The most effective way to prevent incidents is to identify the risk chain and make timely diagnoses for prevention.

Data mining is used to analyze big data for information with intrinsic value. A new technique involving multiple disciplines including database technology, intelligent algorithms, knowledge engineering, and statistics, data mining can extract a large amount of data from databases as needed. After analysis, the data can be used as a basis for decision-making. All this shows that data mining has a wide range of applications. Applying data mining to rail transit operation is bound to improve its safety management. In 1995, the first ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD & Data Mining Conference) was held in Montreal, Canada. At the conference, the concept of data mining was standardized, turning it into data mining in engineering and knowledge discovery in scientific research [1]. Academic researchers in China and elsewhere have shifted their focus of research to mining methods. Yong et al.

proposed a service-oriented cloud mining framework to design and operate distributed data mining applications [2]. Huang et al. advanced a variable precision rough set theory [3]. Huang applied an evolutionary algorithm to high-dimensional data mining and used a joint evolutionary algorithm to solve the problems of slow and premature convergence arising at the later stages of the algorithm [4]. Common data mining methods include association rule analysis, regression analysis, and cluster analysis. Mazouri et al. collected 6,500 data points on Iranian railway accidents and explored the connections among factors using an association rule algorithm, providing a theoretical reference for railway safety management [5]. Weng et al. used a priori algorithm to explore associations among the factors related to marine accidents and road traffic accidents [6]. Chen et al. improved the association rule algorithm, so that it no longer omitted important factors, and introduced weights into the algorithm, proposing latent class-based association rule mining [7]. The improved algorithm had better mining effectiveness. Regarding accident identification, He developed methods to identify, evaluate, grade, and optimize subway operation hazards, strengthening quantification by the index system of the algorithm, and reducing errors caused by subjectivity [8]. Martello et al. discussed the impact of sea level rise (SLR) and other exogenous flood events on rail transit operations in coastal cities based on a flexible framework model [9]. Li et al. conducted empirical research on the status quo of safety risks of subway operating systems in some cities and the harmful factors in various accident cases and pointed out that it is difficult to find invisible harmful factors [10]. Some researchers have searched for hazards using the accident tree method. Yang et al. identified tramcar operation hazards and found risk factors using the accident tree method, indicating that the method has certain application value [11]. However, this method can only be used for simple causal reasoning and does not apply to complex system analysis. In order to put hazard identification into practice in the information field, Ding conducted an in-depth study on data mining algorithms, built a model for the rail transit dispatching records of a certain supercity to identify the major rain transit hazards, and developed an intelligent hazard identification system based on the data warehouse [12]. The concept of chained risk propagation originates from the fields of project management, trade finance, media, public opinion, culture, etc. but has not yet been extensively studied in the field of rail transit safety management. British professors Chapman and Cooper were the first to propose the idea of risk chains [13]. They then advanced a systematic risk management theory. Kangari studied and verified risk chains from the perspective of supply chains, concluding that risk chains refer to interdependent risks formed by a series of chain effects caused by logical relations [14]. Qiu and Wang et al. handled unstructured accident cases in a structured way and proposed a method to identify possible accident causation chains, achieving targeted prevention and emergency handling [15]. For risk control algorithms, Huang Yuxin et al. designed a dual coal-mine accident prevention information system based on the a priori algorithm, integrating accident hazards

and risk control [16]. Pei and Zhao proposed the idea of “integrative safety management” for information system construction, realizing prompt hazard identification and risk control through information-based management and information sharing [17, 18]. Li combined association rules, complex networks, and Bayesian networks to accurately grasp the dynamics of coupled risks in the construction process, finding the relationships among multiple factors, identifying key risk factors and realizing accident diagnosis, and keeping risks under real-time control [19, 20].

In sum, many scholars have done a large amount of basic research on hazard identification, risk rating, and risk control [21–25]. They have also explored traffic hazards and investigated risk chain propagation mechanisms, but their research still has the following shortcomings. (i) There are still few research findings about hazard identification related to rain transit operation. There are a few arithmetic studies on hazard identification in the field of railway traffic and mining engineering, but most of them focus on the accident tree method and association rule algorithm. These methods can only be used to find the causes of an accident and cannot reveal propagation chains among the hazards, i.e., risk chains. (ii) Much research has been done on incident causation models, whereas there is barely any research on chained coupling mechanisms among traffic incident causation factors, nor on risk chains. Therefore, it is of great practical significance to study risk chains based on major hazards. (iii) Extensive research on big data and data mining algorithms has been conducted in China, but it mainly focuses on commercial purposes or bank loan risk assessment, with very few conclusions about rail transit safety management. It is necessary to build data mining algorithms to provide methodological guidance for hazard identification related to rail transit operation. (iv) Research on risk control algorithms and measures mainly focuses on risk control in industrial engineering and mining engineering. The control measures are mostly applied to information-based management, whereas the preliminary risk assessment and hierarchical control are less refined, making it difficult to avoid careless omissions when an incident occurs.

## 2. Data Fusion and Data Lake Construction for Rail Transit Operation Safety

Rail transit operation safety data types are diverse, such as passenger volume, dispatching record, equipment status, ledger record, and rail damage parameters. The data differ from city to city and from format to format. In this paper, barriers among cities between formats are broken by multisource heterogeneous preprocessing and word cloud merging to build a fused data pool and data lake.

*2.1. Multisource High-Dimensional Heterogeneous Data Preprocessing.* Safety data includes many stop words. A stop word is a word that often appears in textual data but has no practical meaning or always has the same meaning in different data, such as conjunctions and modal particles. These

words usually do not have significance. The purpose of stop word removal is to screen out such words.

### 2.1.1. Word Segmentation and Stop Word Removal

Step 1: create a custom terminological corpus, Fen\_Risk, composed of 804 terms related to urban rail transit hazards and 406 terms related to urban rail transit.

Step 2: with Fen\_Risk, use Python Jieba for word segmentation.

Step 3: remove the stop words from the data, i.e., all words irrelevant to the research, such as punctuation marks, all words that mean “and,” “besides,” and “merely.” Based on Harbin Institute of Technology’s stop word list, create a custom list of 2,301 stop words.

2.1.2. Hazard Word Cloud Merging. Merge the words that are expressed differently but have the same meanings in the text data into technical terms related to rail transit hazards to form a hazard word cloud, and then merge nonstandard expressions in the text into the cloud. Table 1 shows hazard cloud merging and valuation.

2.2. Quantification of Hazard Risks. Risk is the product of probability of hazard occurrence and loss upon occurrence. The higher the hazard occurrence probability, and the more severe the loss, the higher the hazard level. An expression can be established for the risk of rail transit hazards according to the classical risk calculation model.

$$R = P \times C, \quad (1)$$

where  $R$  is the hazard risk,  $P$  is the hazard occurrence probability, and  $C$  is the loss caused by hazard occurrence.

The incident probability can be calculated according to the ratio of incidents caused by a certain type of hazard to those caused by all types of hazards.

$$p = \frac{n_i}{\sum_{i=1}^n n_i}, \quad (2)$$

where  $n_i$  is the frequency of hazard type and  $\sum_{i=1}^n n_i$  is the total frequency of  $n$  hazard types.

Equation (1) can be rewritten as  $R = p \times c = (n_i / \sum_{i=1}^n n_i) \times c_i$ .

If  $R$  is greater than the threshold  $\alpha$ , the hazard is included in the operational hazard database.

After 3.1 and 3.2, the risk data lake is completed, as shown in Table 2.

## 3. Key Hazard Identification, Risk Chain Search, and Algorithm Development

3.1. Optimization of Key Hazards. After processing, the rail transit operation safety data form a dataset that can be used to identify key hazards, but the data need to be classified and reduced. This paper uses SVM to determine the optimal hyperplane for data classification. After training, most of the

training samples can be removed because the final model is related only to the support vector.

The basic idea of SVM is to find the hyperplane with the largest geometric distance where the data can be correctly classified. Hyperplane optimization is actually a quadratic programming problem, as shown in the following equation:

$$\min_{w,b} \frac{1}{2} \|w\|^2. \quad (3)$$

The constraint condition is as follows:  $y_i(w \cdot x_i + b) \geq 1$ ,  $i = 1, 2, 3, \dots, n$ .

The Lagrangian function can be used to solve the above programming problem, as shown in the following equation:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i(w \cdot x_i + b) - 1]. \quad (4)$$

Then the problem is transformed into one of working out the minimum of the Lagrangian function with  $w$  and  $b$  as independent variables. Then, we take the partial derivative of  $w$  and  $b$ , and set it to 0, as shown in the following equations:

$$\frac{\partial L(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0, \quad (5)$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = \sum_{i=1}^l \alpha_i y_i = 0. \quad (6)$$

Equations (5) and (6) are substituted into (3) to solve the dual problem, as shown in the following equation:

$$\begin{aligned} & \max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ & \text{s.t.} \quad \begin{cases} \alpha_i \geq 0, \quad i = 1, 2, 3, \dots, n, \\ \sum_{i=1}^l \alpha_i y_i = 0. \end{cases} \end{aligned} \quad (7)$$

This is an extreme-value problem whose objective function is a quadratic function. Under the constraint of inequality, there exists a unique solution to the problem, so there is also unique solution to the original problem. The solution of the dual problem needs to satisfy the following equation:

$$\alpha_i [y_i(w \cdot x_i + b) - 1] = 0, \quad i = 1, 2, 3, \dots, n. \quad (8)$$

Thus, most Lagrangian coefficients  $\alpha_i$  should equal 0, but some  $\alpha_i$  is not 0. When  $\alpha_i$  is not 0, the corresponding  $y_i(w \cdot x_i + b) - 1 = 0$ ; i.e., hazards are located on the separation boundary. At this point, it corresponds to the support vector.

A set of optimal solutions is obtained, i.e.,  $\alpha^* = (\alpha_1^*, \alpha_2^*, \alpha_3^*, \dots, \alpha_l^*)$ .  $w^*$  and  $b^*$  can be solved through  $\alpha^*$ , as shown in the following equations:

$$w^* = \sum_{i=1}^l \alpha_i^* \cdot y_i x_i, \quad (9)$$

$$b^* = y_i - w^* \cdot x_i. \quad (10)$$

TABLE 1: Hazard cloud merging and valuation.

Hazards	Possible descriptive value	Qty.	Remarks
Incursion by foreign object	A flowerpot fell into the track area; a clothes hanger was hung on the electrification wires; a pedestrian passed by. . .	$n_1$	
Switch splitting	Turnout fault, switch fault, confusion between normal and reverse switch positions. . .	$n_2$	
People jammed into the sliding door	Passengers rushed into the door; someone got jammed in the sliding door due to door failure; the train was packed. . .	$n_3$	
Weak safety awareness	Carelessness, weak sense of responsibility, lack of safety awareness. . .	$n_4$	
Natural disasters	Severe thunderstorm, heavy rain, heavy snow	$n_5$	
.....	.....	$N$	

TABLE 2: Risk big data lake.

SN	Hazards	Operational event description	Qty.	Loss	Remarks
$R_1$	Incursion by foreign object	A flowerpot falls into the track area, forcing the train to stop, and forcing immediate repairs in the track area.	$n_1$	$C_1$	Track maintenance
$R_2$	Switch splitting	.....	$n_2$	$C_2$	Railway operation
$R_3$	People jammed in the sliding door	.....	$n_3$	$C_3$	Station management
$R_4$	Weak safety awareness	.....	$n_4$	$C_4$	Personnel management
$R_5$	Natural disasters	.....	$n_5$	$C_5$	Meteorological management
$R_n$	.....	.....	$N$	$C_n$	

After the above optimization, the original dataset can be reduced and clustered.

### 3.2. Hazard Identification Modeling and Algorithm Design.

From the perspective of SVM, each hazard data point is composed of vectors, while the hazard data is composed of feature words from the perspective of text mining. Each feature word affects the classification results in the form of a vector. Therefore, appropriate feature words can be selected from the external environmental hazard description information and transformed into vectors to improve the accuracy of hazard classification. The existing database includes 78,343 data points. After word segmentation and stop word removal, 43,467 phrases and separate words remain.

By limiting the number of feature vectors to remove some common term vectors, low-frequency words, and words with little information from the categories, we can more accurately pick out the most effective feature words and give them greater weight to improve the classification accuracy. At the same time, term vectors of less importance can be eliminated to effectively reduce the dimensionality of the feature space to improve the calculation efficiency.

Priority process algorithm comprises two main processes: generation of a frequent item set and of association rules. To generate the frequent item set, it finds the item sets that meet the support requirements, while the association rules are a collection of frequent item sets that satisfy the minimum confidence. The formula of support is as follows:  $\text{support}(A \rightarrow B) = P(A \cup B)$ . Support reveals the probability that A and B appear at the same time. If A and B are much less likely to appear at the same time, it shows that A and B are unrelated to each other; if they appear together

frequently, it shows that they are highly related to each other. The confidence formula is as follows:  $\text{confidence}(A \rightarrow B) = P(A|B)$ . Confidence reveals whether or how often B appears when A appears. If the confidence is 100%, A and B can be bundled. If the confidence is low, it means that B does not necessarily appear when A appears, as shown in Algorithm 1.

Table 3 shows the hazards revealed by data mining. The support and confidence of each item can be compared with  $\text{Sup}_{\min}$  and  $\text{Conf}_{\min}$  to reveal key hazards and finally fully identify the key risk chains.

### 3.3. Risk Data Reduction and Risk Chain Construction Based on Key Hazards.

The key hazards can be identified by the calculation in 3.2. The accumulated operation data are big, and the mined data are just some of the core hazards. In order to identify the key hazard risk chains, it is necessary to reduce and extract key hazard life-cycle information. Procedure set (algorithm) 1 is designed for reduction and extraction, as follows:

Step 1: filter and specify the names of key hazards, and build a database table and fields for risk chain search, as shown in Table 3.

Step 2: design a SQL statement processing unit to extract operation event descriptions for the key hazards in the data lake.

Procedure set 1

```
SELECT Distinct Risk FROM Risk_key
Risk_recordset = rs("risk")
```

Procedure set 2

```
SELECT event_description FROM Risk_lake
```

```

Input: hazard dataset  $D$ ; minimum support:  $\text{Sup}_{\min}$ ; minimum confidence:  $\text{Conf}_{\min}$ 
Output: frequent item set  $L$ ; association rules
(1)  $L_1 = \text{find\_frequent\_itemsets}(D)$ ;
(2) For
(3)  $C_k = \text{apriori\_gen}(L_{k-1}, \text{Sup}_{\min})$ ;
(4)  $C_t = \text{subset}(C_k, t)$ ;
(5) End for
(6) for
(7)  $L_k = \{c \in C_k | c.\text{count} \geq \text{sup}_{\min}\}$ 
(8) end
(9) return  $L$ 
(10) for
(11)  $\text{Subitems} = \text{gen\_subitemset}(L_k)$ 
(12)  $\text{Ar\_gen} = \text{Association}(L_k, \text{conf}_{\min})$ 
(13) end.

```

ALGORITHM 1: Priority process algorithm.

TABLE 3: Database field remarks.

Database name	Hazard name	Event description	Hazard no.	Classification
Risk_lake	Risk	Event_description	Risk_ID	Class

WHERE risk IN (SELECT Distinct risk\_id FROM Risk\_key

WHERE dateDiff ("s", "Tc\_time", "Datetime") > 60)

Step 3: segment the hazard words on the Event\_description field of the key hazards, and build a risk chain word unit.

Step 4: build a risk chain and explore risk chain propagation. Focus on the words "cause," "result in," "because," "so," "so that," etc. According to the results of procedure sets 1 and 2, as well as the word segmentation results in Step 3, to identify risk propagation chains based on key hazards, the chains found each time may not be complete, but a whole chain can be formed after many times of mining and concatenation.

Step 5: use the association calculation method presented in Section 4 to reveal the patterns of keyword appearance in the operational time descriptions within the same class name field to form key hazard risk chains. Figure 1 presents key hazard and risk chain results.

Step 6: continue the calculation until a satisfactory chain length is obtained. Otherwise, repeat Steps 4–5.

## 4. Analysis of Examples

### 4.1. Data Acquisition

**4.1.1. Hardware Scheme for Multisource, High-Dimensional, and Heterogeneous Data Acquisition.** Rail transit operation data include passenger traffic video capture, maintenance data, dispatch records, on-board equipment data, station ledgers (paper), data on delays of 15 minutes or more, and AFC passenger traffic data. The specific experimental setup is as follows:

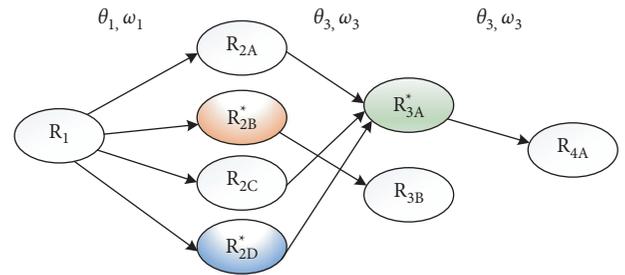


FIGURE 1: Key hazard and risk chain results.

- (i) Experimental site: select 8 representative stations at medium scale or above (all entrances and exits, AFC gate, platform, and escalator) in metropolitan areas, maintenance bases, and compartments under operation
- (ii) Acquisition objects: passenger traffic video, Wi-Fi probe data, track lines, escalators, dispatch records, station ledgers, etc.
- (iii) Main experimental equipment

Wi-Fi probes: fixed or suspended WI-FI probes are used to capture the MAC addresses of intelligent devices carried by passengers.

HD cameras: 8-channel video capture cards, and 4-channel miniature HD cameras, which are 3D HD surveillance cameras with DOF information, used to capture HD passenger traffic video.

**4.1.2. Data Lake-Based Mining Platform Construction and Experiment Design.** Eight medium-sized stations are used as experimental objects. Each station is equipped with the hardware devices shown in Figure 2. The data acquired from each station are sent to the data lake by the data pool

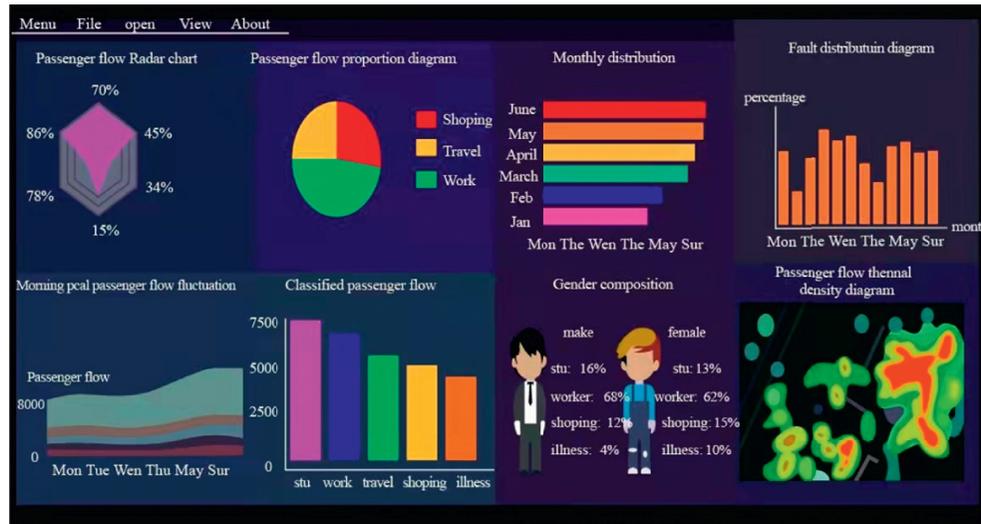


FIGURE 2: Data acquisition architecture and software platform.

through the cloud platform and used as a data warehouse for mining.

Acquisition scheme design: a unified scheme is used to acquire data from eight urban rail transit stations at medium scale or above and rail lines in every city concerned:

- (i) Passenger traffic parameters: in-grid passenger traffic, walking speed, and density
- (ii) Equipment operation status: rail service time, real-time status, degree of fatigue, and load capacity
- (iii) Rail line parameters: foreign objects, turnout status, and rail breakage
- (iv) Dispatch records and station ledgers: fault name, occurrence time, basic description of events, and person in charge

After data acquisition, rough set theory is used to collate the data to delete redundant condition attributes and duplicate records. Then, using deep learning theory, an AI neural network is used to optimize the reduction rules so that it should have self-learning ability to design an operation data lake to be used for intelligent hazard identification.

**4.2. Identification of Key Hazards in Urban Rail Transit Operation.** This paper acquired and processed records of rail transit operation obtained from multiple channels, including the operational status of subway trains under various conditions. Due to space limitation and data confidentiality, only some of the raw data was processed. After word segmentation and stop word removal by Jieba, a database was obtained for identification of key hazards, as shown in Table 4.

**4.3. Risk Chains Based on Key Hazards.** Risk chain propagation: the words “cause,” “result in,” “because,” “so,” “so that,” etc. are mined according to the results obtained

from procedure sets 1 and 2, as well as the word segmentation in Step 3, to identify risk propagation chains based on key hazards. The chains found each time may not be whole chains, but a whole chain can be formed after concatenating many runs. Table 5 displays key hazard risk chains.

Finally, key hazard-based risk chains were obtained, as shown in Figures 3 and 4, providing targeted practical guidance for actual operation and production.

#### 4.4. Management and Control of Key Hazards

**4.4.1. Emergency Planning for Key Hazards.** Many key hazards related to rail transit operation may be identified by data mining. Risk control and governance must be studied, taking incursion by foreign object as an example.

Definition of incursion by foreign object: a foreign object refers to any facility, equipment, structure, or greening device located within or outside the scope of the rail transit. After displacement for any reason, they intrude into the facility or vehicle clearance, thereby affecting safe train operation.

##### (1) Hazard Traceability

- (1) Incursion of internal facilities and equipment, including the evacuation platform, contact passage door, cable holder or signboard, or external objects, including buildings or structures located along the rail line
- (2) Incursion by foreign objects caused by facilities or equipment failure or falling
- (3) Illegal construction in the surrounding area, or deliberate destruction, etc
- (4) Incomplete equipment restoration or clearing after construction
- (5) Inclement weather or natural disasters

TABLE 4: Actual operation data after acquisition and preprocessing.

SN	Hazards	Class 1/2	Operational event description	Remarks
$R_1$	Incursion of foreign object	Maintenance/track maintenance	.....	Track maintenance
$R_4$	Weak safety awareness	Management/personnel management	.....	Personnel management
$R_5$	Natural disasters	Meteorology/meteorological disasters	.....	Meteorological management
.....	.....	.....	.....	.....
$R_{n-1}$	Switch splitting	Maintenance/track maintenance	.....	Maintenance

TABLE 5: Key hazard risk chains.

SN	Hazards	Risk chain	Remarks
$R_1$	Incursion of foreign object	Flowerpot falling-incursion of foreign object-train delay	Track maintenance
$R_1$	Incursion of foreign object	Wire touches the net-incursion of foreign object-train delay-increased passenger traffic at station	Maintenance
$R_4$	Weak safety awareness	Negligent supervision-wishful thinking-people jammed in sliding doors	Personnel management
$R_5$	Natural disasters	Typhoons/rainstorms-train speed limit-increased passenger traffic	Meteorological management
$R_5$	Natural disasters	Freezing-turnstile freezing-switch splitting-increased passenger traffic	.....
.....	.....	.....	.....
$R_{n-1}$	Switch splitting	Weak professional ability-switch splitting-trains withdrawn from schedule	Maintenance

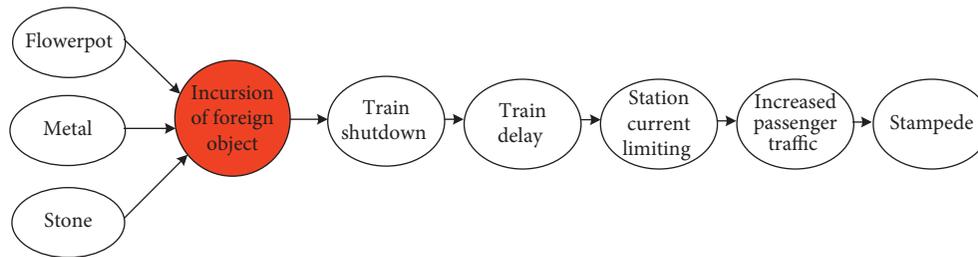


FIGURE 3: Risk chain for incursion by foreign object.

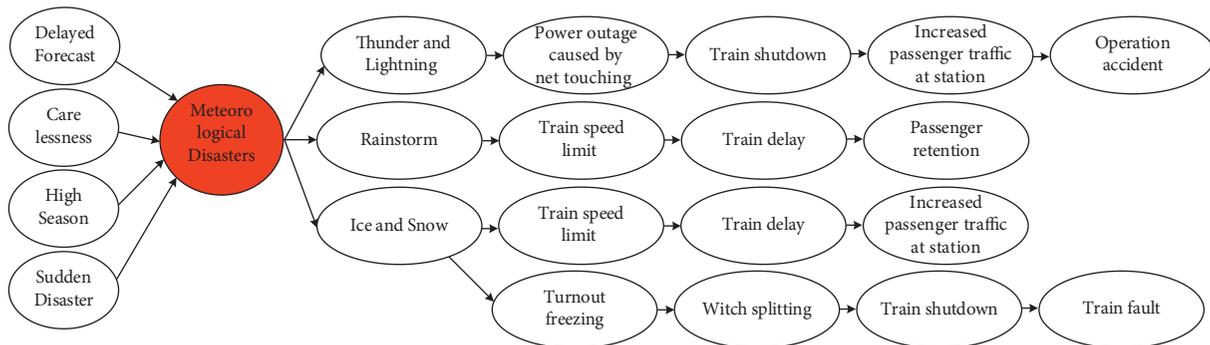


FIGURE 4: Risk chain for meteorological disasters.

(2) Prevention and Control

- (1) The departments concerned should do a good job in maintenance and inspection of facilities and equipment and promptly handle problems.
- (2) Station staff and drivers should strengthen inspection, stop any dubious external facilities or equipment from entering the area, and promptly report to their superiors.

- (3) Train drivers should strengthen distant observation during operation, take immediate measures for any foreign objects identified, and report them to their superiors.
- (4) Construction personnel should operate in strict accordance with the rules and regulations and strictly clear the construction site and may not violate rules or regulations during operation.

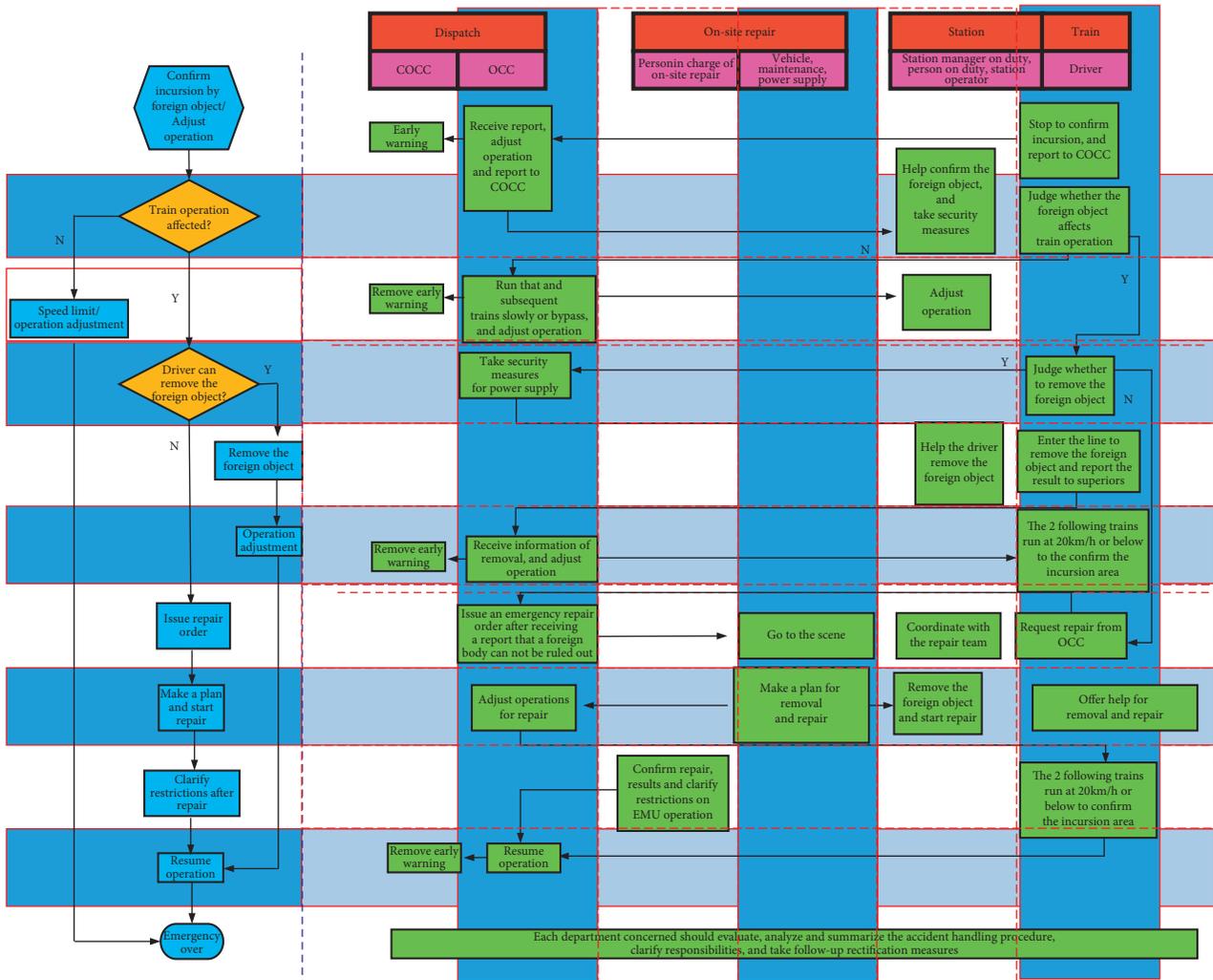


FIGURE 5: Emergency handling flowchart.

(5) The departments concerned should pay attention to early warning information about weather and disasters and take corresponding preventive measures on a timely basis. Construction personnel should strictly implement line clearing procedures to confirm whether the equipment has become normal and meets operational requirements.

(3) Emergency Handling

When the train driver can independently dispose of the foreign object,

- (1) OCC promptly implements safety measures such as third-rail power outages based on rail transit characteristics and on-site conditions, predicts the impact areas and times of foreign object incursions, and adjusts line operation efficiently
- (2) The train driver informs and appeases the passengers based on on-site conditions and takes personal safety measures before entering the line to clear foreign objects

(3) The driver takes any foreign objects beside the rail away from the track area and temporarily puts them in the driver’s cab, and the driver clears any foreign objects beside and on the electric lines with an insulation tool; if the incursion is near a station, OCC can ask the station staff to drive to remove it

(4) After the removal, OCC can ask 2 trains behind to run at a speed of not higher than 20 km/h to observe the condition of the section that had the foreign object

(5) When passing by the section, the train driver strengthens distant observation to ensure driving safety

(4) Posthandlin

- (1) After emergency handling, OCC promptly adjusts the train operation plan to restore normal operation order.
- (2) COCC cancels the early warning as appropriate.

- (3) The departments concerned check and maintain the facility and equipment setup in the area affected by the incursion as needed.
- (4) The departments concerned are responsible for analyzing the causes of the incursion, identifying hazards, and taking corresponding preventive measures.
- (5) All departments evaluate, analyze, and summarize the event handling process, clarify their responsibilities, and take corrective measures. Figure 5 shows emergency handling flowchart for key hazards caused by incursion by foreign objects.

## 5. Conclusion

This paper has established a systematic multisource, high-dimensional, and heterogeneous rail transit operation and production data collection center integrating production and practical data, including operational and maintenance data, forming a data warehouse and risk chain library for intelligent identification of rail transit hazards, helping change the status quo of low utilization of rail transit data, with a high fault rate. From a new perspective, accident causation modes, i.e., the risk chains, were clarified, and key nodes in the chains were found, providing theoretical decision-making support for “chain breakage.” The application of big data on urban rail transit operation, taking safety as the core, and guided by operational requirements, provides data mining methods and application directions, forming a methodological system suitable for intelligent identification of hazards in rail transit operation, thereby standardizing the safety management of rail transit operation and truly realizing refined hazard management. Risk control and governance plans can be developed purposefully for different types of hazards, providing a methodological basis for national, local, and industrial standards and specifications applicable to safe rail transit operation. However, there are still shortcomings in the paper. It does not consider the relationship between the passenger supervisor’s willingness and the rail transit operation risk, which is the focus of future research.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## Acknowledgments

This work was supported by the soft science research project of Shanghai Science and Technology Committee (Grant no. 21692195600), the Key Lab of Information Network Security of Ministry of Public Security (Grant no. C20609), and the Municipal Key Curriculum Construction Project of University in Shanghai (Grant no. S202003002).

## References

- [1] H. Z. Wang and A. Q. Peng, “Research actuality and development trend of data mining,” *Industry and Mine Automation*, vol. 37, no. 2, pp. 29–32, 2011.
- [2] Y. H. Yong, X. J. Xiang, Y. Gao, L. Shang, and Y. B. Yang, “Service-oriented cloud data mining engine,” *Journal of Frontiers of Computer Science & Technology*, vol. 6, no. 01, pp. 46–57, 2012.
- [3] Z. H. Huang, *Design and Implementation of an Internet Information Mining System Based on Semantic Information Extraction*, Zhejiang University, Zhejiang, China, 2010.
- [4] W. J. Huang, *Adaptive Multi-Objective Clustering Ensemble Selection for High-Dimensional Data*, South China University of Technology, Guangzhou, China, 2018.
- [5] F.-Z. E. Mazouri, S. Jabbour, B. Raddaoui, L. Sais, M. C. Abounaima, and K. Zenkour, “Breaking symmetries in association rules,” *Procedia Computer Science*, vol. 148, pp. 283–290, 2019.
- [6] J. Weng and G. Li, “Exploring shipping accident contributory factors using association rules,” *Journal of Transportation Safety & Security*, vol. 12, no. 1, pp. 25–32, 2017.
- [7] Y. J. Chen and A. H. Wu, “Analysis of causes of traffic accidents based on a latent class model and association rules,” *Modern Computer (Professional Edition)*, vol. 15, pp. 8–13, 2018.
- [8] L. He, “Methods for identifying, evaluating, classifying and optimizing subway operation hazards,” *Safety & Security*, vol. 39, no. 8, pp. 8–11, 2018.
- [9] M. V. Martello, A. J. Whittle, J. M. Keenan, and F. P. Salvucci, “Evaluation of climate change resilience for Boston’s rail rapid transit network,” *Transportation Research Part D: Transport and Environment*, vol. 97, no. 01, pp. 1–14, 2021.
- [10] Z. G. Li, P. Xu, and G. L. Wang, “Research on the safety risk identification process and risk factors of subway operation system,” *Journal of Safety Science and Technology*, vol. 15, no. S1, pp. 34–37, 2019.
- [11] N. H. Yang, M. Ye, and A. M. Wu, “Hazard identification and risk assessment for safe operation of modern tramcars,” *Modern Transportation Technology*, vol. 14, no. 2, pp. 75–80, 2017.
- [12] X. Ding, X. Yang, H. Liu, and H. Hu, “The safety management of urban rail transit based on operation fault log,” *Safety Science*, vol. 94, no. 94c, pp. 10–16, 2017.
- [13] C. B. Chapman and D. F. Cooper, *Risk Analysis for Large Project: Models Method and Cases*, pp. 544–549, John Wiley & Sons, New York, NY, USA, 1987.
- [14] T. Kangari, “Risk management perception and trend of U.S. construction,” *Journal of Construction Engineering and Management*, vol. 12, no. 04, pp. 422–429, 1995.
- [15] G. F. Qiu and J. Wang, “Research on knowledge discovery for construction accidents based on concept lattice,” *Journal of Safety and Environment*, vol. 19, no. 5, pp. 1625–1630, 2019.
- [16] Y. X. Huang, Z. G. Yan, J. D. Fan, and C. Li, “Duplex coalmine accident prevention information system based on apriori algorithm,” *Industry and Mine Automation*, vol. 46, no. 10, pp. 92–98, 2020.
- [17] H. Pei and W. Zhao, “On the construction of integrated information system for safety management at Nanjing metro,” *Journal of Safety Science and Technology*, vol. 15, no. S1, pp. 26–33, 2019.
- [18] Z. Liu, H. G. Guo, Y. J. Zhao, and B. Hu, “Optimal pricing decision of composite service offered by network providers in

- e-commerce environment,” *Electronic Commerce Research*, vol. 4, pp. 1–17, 2021.
- [19] C. S. Li, *Research on Dynamic Control of Coupled Risks in Deep Foundation Pits of Metro*, Huazhong University of Science and Technology, Huazhong, China, 2019.
- [20] Z. Liu, L. Lang, L. Li, Y. Zhao, and L. Shi, “Evolutionary game analysis on the recycling strategy of household medical device enterprises under government dynamic rewards and punishments,” *Mathematical Biosciences and Engineering: MBE*, vol. 18, no. 5, pp. 6434–6451, 2021.
- [21] Y. Han, F. Zhang, L. Huang, K. Peng, and X. Wang, “Does industrial upgrading promote eco-efficiency?—a panel space estimation based on Chinese evidence,” *Energy Policy*, vol. 154, Article ID 112286, 2021.
- [22] B. Wu, A. Monfort, C. Jin, and X. Shen, “Substantial response or impression management? Compliance strategies for sustainable development responsibility in family firms,” *Technological Forecasting and Social Change*, vol. 174, Article ID 121214, 2022.
- [23] M. Zhuang, W. Zhu, L. Huang, and W. T. Pan, *Research of Influence Mechanism of Corporate Social Responsibility for Smart Cities on Consumers’ Purchasing Intention*, Library Hi Tech, Bingley, UK, 2021.
- [24] D. Xiang, Y. Zhang, and A. C. Worthington, “Determinants of the use of fintech finance among Chinese small and medium-sized enterprises,” *IEEE Transactions on Engineering Management*, vol. 68, no. 6, pp. 1590–1604, 2021.
- [25] Y. Chen and V. Sivakumar, “Investigation of finance industry on risk awareness model and digital economic growth,” *Annals of Operations Research*, pp. 1–22, 2021.