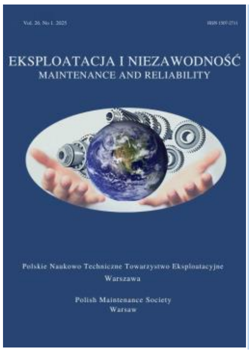
EKSPLOATACJA I NIEZAWODNOŚĆ
MAINTENANCE AND RELIABILITY
Polskie Naukowo Techniczne Towarzystwo Eksploatacyjne
Warszawa
Polish Maintenance Society
Warsaw

Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 27 (2025), Issue 1

journal homepage: http://www.ein.org.pl

Article citation info:
Zhang H, Wang Q, Risk identification model of aviation system based on text mining and risk propagation, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2025: 27(1) http://doi.org/10.17531/ein/192767

Risk identification model of aviation system based on text mining and risk propagation

Indexed by:

Han Zhang[a,b,*], **Qiang Wang**[a]

[a] Air Force Engineering University, China
[b] Xian University of Finance and Economics, China

Highlights

- A risk feature vector space model is proposed based on textual data mining
- The correlation network of risk features is proposed
- The coupling mechanism of risk features is captured
- Procedure, wind, and visibility are critical risk features.

Abstract

Due to the aviation accident is rarely predictable and often irreversible, how to ensure aviation safety is of uttermost importance. Textual aviation accident reports contain the cause and process of the accident which could help people understand incidents. However, the cause of the accident always is summarized by the expert and the accident report would be incomplete, the identification of aviation safety accident risk is not timely and accurate. In this paper, a safety risk identification model is proposed, aiming to identify the correlation between aviation safety accident risk factors by textual data mining from textual aviation accident reports. In detail, the feature of aviation accidents is extracted and classified by text mining technology, on this basis, the correlation coefficient matrix between different features is established. Finally, the correlation network of aviation safety risk is proposed, and the risk propagation process of accidents is developed based on the network to identify aviation safety accident risk.

Keywords
aviation safety, aviation accident reports, complex network, risk propagation, risk identification

1. Introduction

Due to the seriousness of the consequences of aviation accidents, the demand for safety requirements in aviation systems is increasing, continuous airworthiness risk management is the research focus [1]. Continuous airworthiness risk management is a risk management technology for civil aircraft to ensure that the safety level of aircraft operation always meets the basic safety standards or airworthiness level [2]. In general, the way to ensure the safety of aviation system has two aspects. One of these aspects is a reactive, incident-based approach to ensure the safety of aviation system. The other is a proactive, systems-based approach which means the operator has taken appropriate measures to prevent aviation accidents when an unsafe event occurs[3].

The study on the first aspect is to summarize the cause of the accident that could find the risk factors. At first, most of the research focused on the evolution of the number of accidents based on flight data [4]. Next, some researchers focused on classic aviation accidents to find critical risk features[5]. Aim at runway incursion accidents, Stroeve et al. presented a framework for the evaluation of runway incursions, which

(*) Corresponding author.
E-mail addresses: H. Zhang, xicai_zh1020@163.com, Q. Wang 3198501761@qq.com,

could provide feedback to managers about structure causes and risk implications[6]. Aiming at aircraft loss-of-control (LOC) accidents, Ancel et. al. proposed a generic framework of LOC accident risk identification that contains risk factors from the domain of human factors, aircraft system malfunction factors, and environmental conditions[7]. Based on the object-oriented Bayesian network, Ancel et. al. presented an in-flight loss-of-control accident framework model for the large and complex aviation accident model that could identify the most sensitive causal factors for the accident[8]. Aim at controlled flight into terrain (CFIT) accidents, Kelly et. al. analyzed 50 CFIT accidents from 24 counties over 10 years and found human factors, such as distraction, complacency, and fatigue, represent the main cause of the accident[9]. Due to the key role of human factors in risk identification, Gautam et al. analyzed the relationship between aviation safety attitude, flight experience, perceived stress, and hazardous event involvement among aviators based on the data from 360 aviators by using the aviation safety attitude scale, hazardous event scale, and perceived stress scale[10]. For aircraft system malfunction factors, Lee et al. used agent-based modeling, stochastically and dynamically colored Petri net to assess the safety and efficiency of aircraft maintenance strategy which could identify the risks in aircraft maintenance[11]. Zhou et al. proposed a novel data-driven hybrid-learning algorithm that could identify the riskiest sub-systems of the civil aircraft engine to improve the efficient execution of the maintenance strategy and reduce the risk of aviation systems[12].

However, the aviation system is a complex system, which contains risk factors such as human, mechanical engineering, environment, and policy [13]. Identifying the linkages among the risk factors may be an effective way to interrupt risk propagation, which is beneficial for the avoidance of aviation accidents[14]. To better understand the coupling mechanism of various risk factors and present the risk propagation process, aviation accidents must be investigated in more detail to address the question “how did this accident happen?” Therefore, considering this, developing the second way which could detect and predict the relations between risk factors to prevent accidents or predict the process of the accident is necessary. With the development of data technology, natural language processing (NLP) techniques could automatically recognize the cause of aviation accidents, and have become a research hotspot to identify the risk.[15]. The Aviation Safety Reporting System (ASRS) was created in 1976 by the Federal Aviation Administration and the National Aeronautics and Space Administration to receive, process, and analyze voluntarily submitted aviation safety reports [16]. These reports cover a broad scope of safety-related topics, ranging from flight operations, airport ground, and ramp operations, avionics, air traffic control flight, crew communication, general aviation, flight training, meteorology, and weather, to human factors [17]. However, the data are unstructured and high-dimensional, how to identify and classify risk factors from ASRS is a challenge [18]. Zhang et. al. formulated a four-step procedure to construct a Bayesian network, which could realize visualization of the escalation of initiating events into aviation accidents and capture the causal and dependent relationships between the contributory factors and the aviation accident[19]. Zhou et. al. proposed a risk identification and prediction model based on a support vector machine optimized by particle swarm optimization and long short-term memory neural networks, and the model could effectively identify risk factors and accurately predict the trend of parameters to improve the safety of aircraft[20]. Zhang et. al. developed classification models by Word Embedding and the Long Short-term Memory neural network that could predict adverse events like accidents, aircraft damage, or fatalities based on the sequences of events[21]. Miyamoto et al. used natural language processing tools, K Means clustering, and dimensionality reduction by t-distributed Stochastic Neighbor Embedding to categorize and visualize narratives, and found maintenance is the main cause of delays[22]. Zhou et al. proposed a model fusion strategy for aircraft risk identification based on a convolutional neural network and a bidirectional long short-term memory neural network with an attention mechanism[2]. The proposed model with a fusion strategy could realize the stable identification of imbalanced data, which can effectively improve the reliability of aircraft risk identification in the field of civil aviation.

As mentioned above, the risk identification model is a reactive, incident-based approach or a proactive, predictive and systems-based approach, the data is the foundation of these models. ASRS data is composed of heterogeneous data, and the cause of the accident always is summarized by the expert. The accident report would be incomplete, and the identification of aviation safety accident risk is not timely and accurate. Although machine learning tools were used to detect potential links between different reports or risk features[23, 24], the above research ignores the characteristics of influence and propagation among risk features. In this paper, a safety risk identification model is proposed, aiming to identify the correlation between aviation safety accident risk factors by textual data mining from textual aviation accident reports. Firstly, based on textual data mining the dictionary is obtained from ASRS data, and a risk feature vector space model is proposed to get risk features. Secondly, based on the risk propagation model, a risk network model is developed, and a risk propagation process based on the SIS model is proposed to capture the coupling mechanism of risk features. Finally, the risk rate of each risk feature which considers the propagation dynamics of each risk feature is derived for risk identification. The model in the paper could identify the risks in the aviation system and could guide managers, such as air traffic controllers should enhance situational awareness of the airport and aircraft by surveillance equipment such as scene surveillance radar when the visibility is poor.

2. Methodology

The overview of the proposed aviation safety risk identification model includes three phases: data collection, risk features classification and risk identification.

2.1. Data collection

The data in this paper is from the Aviation Safety Reporting System (ASRS), which contains the report submitted by pilots, crew, maintenance personnel, and other relevant staff after the end of each flight mission, according to the problems encountered in the course of flight. We choose assessments and narratives in the report to analyze features of aviation accidents. The assessments represent evaluation results by the expert, and the narrative is the process of the detailed account of the entire airline incident.

2.2. Risk features classification

In this section, we briefly describe the data cleaning process and then delve into how we do the risk feature vector space model. Then we describe risk features that are used to identify the risk.

2.2.1. Date Analysis

Before identifying the aviation safety risk, we need to analyze the textual data and format them as inputs to our models in the form of document feature matrices as shown in Figure 1. The analysis process of the textual data is as follows:

Step 1: For a given sentence, we tokenize the text such that each word is separated. Step 2: We set the stop word such as ‘the’, ‘in’, and ‘by’, and remove them from the sentence.

Step 3: The step of Stemming and Lemmatization is that words such as ‘maintained’ and ‘maintenance’ would be all standardized to ‘maintain’.

Step 4: Counting the processed words and their frequency, creating the document feature matrix.

Step 5: Repeating Step 1 to Step 4 based on ASRS text data, taking 1000 words with the highest frequency as the dictionary. Using the cleaned text, we create a $m \times n$ document, with m number of frequencies and n number of words (or terms), to record the frequency of words in ASRS.

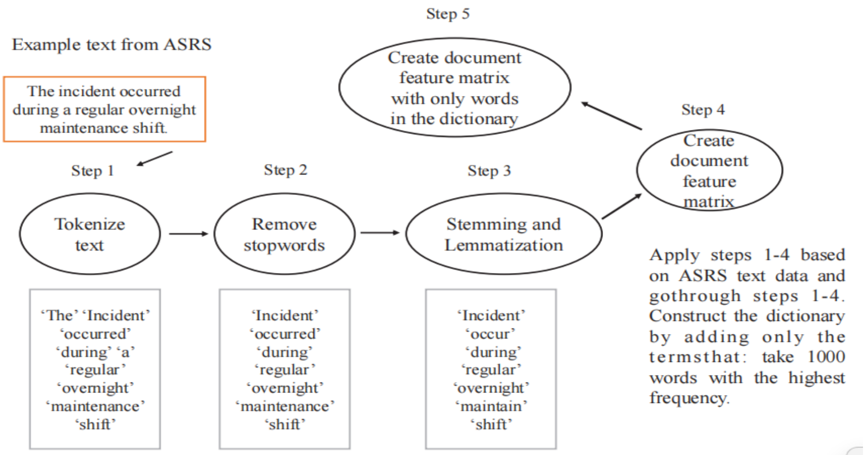
Step 5
Example text from ASRS
Create document feature matrix with only words in the dictionary
The incident occurred during a regular overnight maintenance shift.
Step 4
Create document feature matrix
Step 1
Step 2
Step 3
Tokenize text
Remove stopwords
Stemming and Lemmatization
‘The’ ‘Incident’ ‘occurred’ ‘during’ ‘a’ ‘regular’ ‘overnight’ ‘maintenance’ ‘shift’
‘Incident’ ‘occurred’ ‘during’ ‘regular’ ‘overnight’ ‘maintenance’ ‘shift’
‘Incident’ ‘occur’ ‘during’ ‘regular’ ‘overnight’ ‘maintain’ ‘shift’
Apply steps 1-4 based on ASRS text data and gothrough steps 1-4. Construct the dictionary by adding only the termsthat: take 1000 words with the highest frequency.

Figure 1. Illustration of the process of cleaning the textual data, demonstrating the process of a sentence from the narrative of an accident report in ASRS.

Figure 1: Illustration of the process of cleaning the textual data, demonstrating the process of a sentence from the narrative of an accident report in ASRS.

Due to the dimensions of the dictionary being high and containing a lot of useless information, we need to reduce the dimension of the dictionary to get the risk feature. Here we use the Chi-square statistical method, which shows obvious advantages on recall rate and precision rate, to reduce the dimension. Through this step, we could get the risk feature. The risk feature represents the occurrence of an aviation accident in relation to it. This relationship refers to the accident caused by the risk feature or the accident that happened in the risk feature.

2.2.2. Risk feature vector space model

To get the importance of each risk feature, which represents the ability of the risk feature to cause the occurrence of aviation abnormal events. Term frequency-inverse document frequency (TF-IDF) is a statistical method used to assess the importance of a word to a document in the data set[25]. The importance of a word increases with the number of times it appears in a document but decreases with the frequency of its appearance in the data set. TF-IDF consists of two parts: term frequency (TF) and inverse document frequency (IDF) which represents the rarity of the word. TF represents the local importance of the term, and IDF represents the global importance of the term. Here we use TF-IDF to calculate the weight of each risk feature, the equation is

$$tf - idf = tf_{i,j} \times idf_i \tag{1}$$

where $tf_{i,j}$ can be calculated from the ratios between the appearance frequency of risk features in report j and the total frequency of all risk features in all reports, idf_i can be calculated through logarithm the ratios between the number of reports and the number of the report containing risk feature i.

To avoid the occurrence of rare words and the situation of some words such as 'I' and 'when' appearing high frequently that may lead the IDF of the risk feature to be 0, we improve the inverse document frequency (IDF) in Eq.(1) as

$$idf_i = log \frac{N+1}{|\{j: t_i \in d_j\}|+1} + 1 \tag{2}$$

where N is the total number of reports, $|\{j: t_i \in d_j\}|$ is the number of reports which include word t_i.

2.3. Risk identification

In this section, we establish the relation between risk features to get the risk network, and then describe the risk propagation process and derive the risk index to identify the aviation safety risk.

2.3.1. Risk network model

To get the risk network, the key is to determine the nodes and links between nodes. Here, these risk features are represented as nodes in the risk network. If risk feature i and risk feature j appear in a report at the same time, we consider there is a link between risk feature *i* and risk feature *j*. The correlation between these features would be different. For example, there is a noticeable correlation between the collision accident and the vision factor. It is necessary to consider the weight between risk features, we use the frequency of simultaneous occurrence to be the weight of links. The parameters in the risk network G are defined as follows:

- *i* represents the risk feature in Sec. 2.2 and it also is the node in the risk network.
- w_{ij} the weight between risk feature *i* and risk feature *j*.
- A_{ij} is the network adjacency matrix. If $A_{ij} > 0$ represents risk feature i and risk feature j have connections, otherwise, $A_{ij} = 0$.

Obviously, the risk network is a directed weighted network. Some topological properties are needed to analyze risk features. The out-degree and in-degree are classical indexes in directed weighted networks. The number of edges from *i* to other nodes is used to determine the out-degree of node *i*, it can be calculated through $k_{out} = \sum_{j=G} A_{ji}$. The number of edges from other nodes to *i* is used to determine the in-degree of node *i*, it can be calculated through $k_{in} = \sum_{j=G} A_{ij}$. The sum of the in-degree and out-degree of node *i* is defined as the total degree of node *i*. These indicators which consider static properties of the risk network can be used to evaluate risk.

2.3.2. Risk identification model

Risk is sometimes described as an extensive assessment of the likelihood of mishaps or failures and the seriousness of the results [14]. According to the risk network, risks can spread between risk features through correlation, which can be efficiently represented by infectious illness models [26]. There are other infectious disease models, including SI, SIR, SIS, and SEIRS, according to Gani [27]. The SI model is quick and efficient in simulating the overall risk propagation process. However, when using the SI model, particularly when considering the risk associated with the aviation system, the positive features of human-related activities are not given adequate consideration. In reality, when Flight 3U8633 was cruising in 2018, the front windshield of the right seat of the cockpit broke and fell off, and the crew took emergency measures to make a safe diversion in time, successfully transforming situation from dangerous to normal, that is, the process in which the infected node did not infect other nodes in the system. In 2010, due to pilot fatigue, the pilot ignored the warning when the system issued a warning, and did not fully estimate the situation, which eventually led to the accident of plane crash. In other words, infected nodes infected other nodes and finally triggered the occurrence of aviation accidents. In other words, there are two states of risk features in the aviation system, one is the susceptible state and the other is the infected state. The susceptible risk features would be infected with the susceptible risk features if there is a link between these risk features. The state of infected risk features would turn into a susceptible state through some positive measures. As a result, risk propagation is subsequently prevented spontaneously, analogous to the healing process in the infectious disease model. Therefore, it is more appropriate to replicate the risk propagation associated with aviation accidents using the infectious disease model of SIS[28]. The SIS model also has the advantage of being able to dynamically analyze the impact of risk feature severity on node criticality. As opposed to the analysis methodology suggested in Section 2.3.1, which is static and assumes that the hazardous event's severity is constant.

Here we consider the dynamical models with the SIS model that can be written as

$$\frac{dx_i}{dt} = -Bx_i + \sum_{j=1}^{N} A_{ij} R(1 - x_i) x_j \tag{3}$$

Within the framework of SIS, all nodes in the network are labeled with one of two states: S-susceptible, and I-infected. The dynamics includes a susceptible model in which a node is infected by one of its nearest neighbors at rate R, and an infected node is recovered at rate $B.x_i(t)$ represents the infection rate of node i at time t. In this paper, the infection rate of node i is used to assess risk, indicating the ability of risk feature i to cause adjacent events to become hazardous events, whereas the recovery rate is used to characterize the node's capacity to defend against risk, indicating the ability of risk feature to recover from an abnormal state to a normal state. According to Eq.(3), the infection rate of each node is valued within [0, 1].

To get the infection rate of each node, we consider the steady state of Eq.(3), the equation can be written as

$$-Bx_i^* + \sum_{j=1}^{N} A_{ij} R(1 - x_i^*) x_j^* = 0, \tag{4}$$

where x_i^* represents the infection rate of each node in the steady state.

According to the method in [29], Eq.(4) can be derived as

$$-Bx_{eff} + R\beta_{eff}\left(1 - x_{eff}\right)x_{eff} = 0, \tag{5}$$

where $\beta_{eff} = \frac{\langle k_{in} \cdot k_{out} \rangle}{\langle k \rangle}$, $\langle k \rangle$ is the average degree of the network, $\langle k_{in} \cdot k_{out} \rangle$ is the vector product between in-degree and out-degree, x_{eff} is the weighted nearest neighbor infection rate in the steady state.

To get the infection rate of each node in the steady state, we just use x_{eff} to replacex_j^* in Eq. (4), Eq. (5) becomes

$$-Bx_i^* + Rk_i(1 - x_i^*)x_{eff} = 0, \tag{6}$$

it allows us to capture the steady-state infection rate of each node if the in-degree of node i is known and x_{eff} which can be got from Eq. (5). The steady-state infection rate of node i is

$$x_i^* = \frac{Rk_i x_{eff}}{B + Rk_i x_{eff}}, \tag{7}$$

Through Eq.(7), we could get the steady-state infection rate of each node to assess the risk level of each node. By ranking the risk level of each node, we could identify the critical risk feature. Meanwhile, when some incidents happen during a flight, the manager could take different contingency plans based on the risk network which can prevent further propagation.

3. Experimental results

The data in this paper is from the ASRS database. The aviation accidents from 2021 to 2022 that contain 22064 reports are used to analyze. The full report can be found on the website https://akama.arc.nasa.gov.

3.1. Risk features classification

Through the method from Sec. 2.2, the dictionary of aviation safety reports from the Aviation Safety Reporting System (ASRS) can be extracted. The dictionary includes 1000 words, but it contains a lot of useless information that could affect the computational efficiency. Here we use the chi-square statistics method to reduce dimension to capture risk features. Fig.2 shows the risk features of the aviation system. The horizontal axis represents the frequency of each risk feature, and the vertical axis represents risk features. The word cloud is also shown in Fig.2. The results show that the frequency of risk features, such as aircraft weather, environment, procedure, engine, airport, and wind, is high and their frequency is over 100.

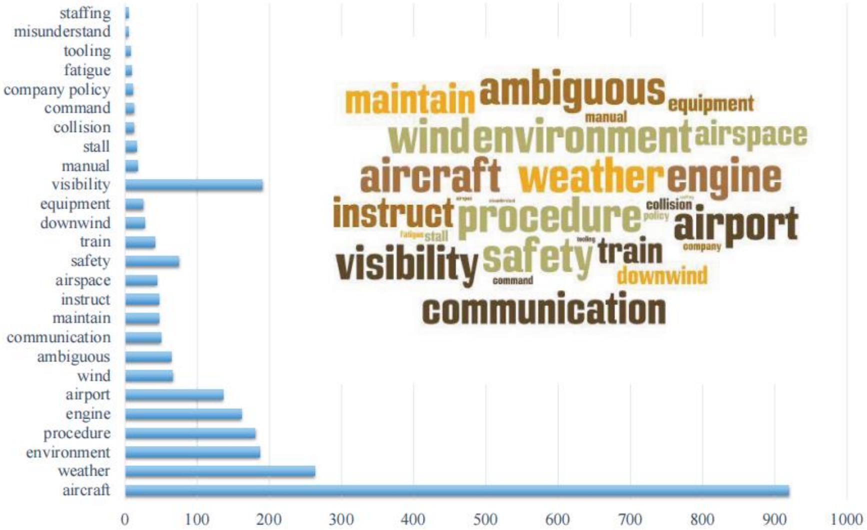

Figure 2: The risk features of aviation safety system

To transform the aviation safety incident report into vector form, we use Eq. (1) to get the weight of each risk feature in different reports. Fig.3 shows the value of TF-IDF for each risk feature. The horizontal axis represents risk features, and the vertical axis represents the frequency of each risk feature. Through comparison of Fig.2 and Fig.3, we find the frequency of aircraft is highest in Fig.2, but it is low in Fig.3. The reason is that the results in Fig.3 are calculated by Eq. (1), the importance of the risk feature is determined by its frequency and rarity. Only considering the frequency, the result could show the topic content of the document, but can not accurately show the ability of the risk feature to cause the occurrence of an aviation abnormal event. Therefore, risk features, such as ‘visibility’, ‘’weather’, ‘environment’, ‘procedure’, ‘engine’, ‘airport’, ‘wind’, ‘ambiguous’, ‘communication’, ‘maintain’, ‘instruct’, ‘airspace’, ‘safety’, ‘train’, ‘misunderstand’, ‘equipment’, are major risk features.

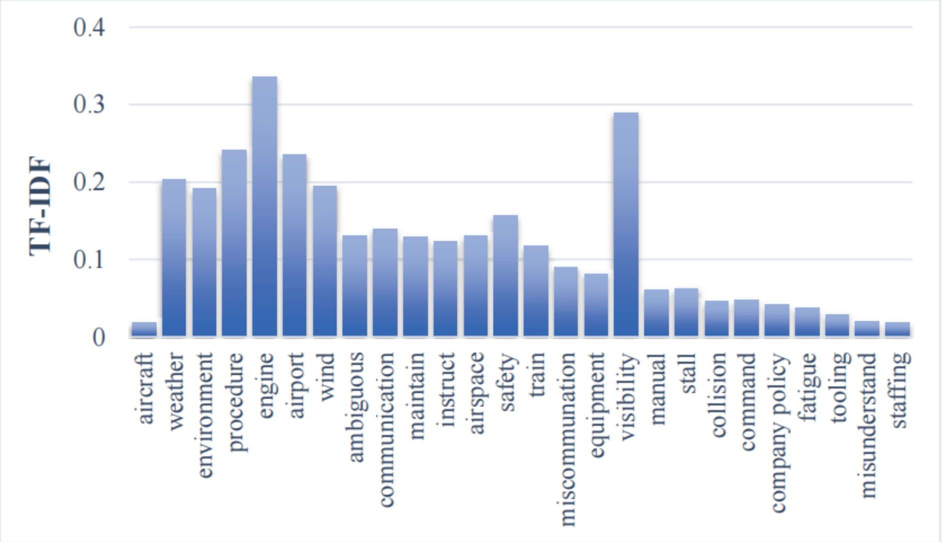

Figure 3. The risk features of aviation safety systems based on TF-IDF.

3.2. Risk network

Based on the method in Sec. 2.3.1, we construct the risk network which includes 26 nodes and 137 links. The network is weighted and directed. Fig. 4a shows the topological structure of the risk network, and Fig. 4b shows the relation between risk features. The results show that the risk feature communication has a high correlation with airspace, aircraft, and maintenance. The risk feature collision has a high correlation with aircraft and visual. The high correlation means the occurrence of the risk feature would cause the occurrence of the associated risk feature. The topological structure feature is also analyzed in Table 1. The out-degree represents the ability of the risk spillover for the risk feature. The in-degree represents the ability of risk tolerance for the risk feature. The total degree could represent the comprehensive ability of the risk tolerance for the risk feature. By calculating the degree of each risk feature in Table 1, we find 'aircraft', 'airport', 'engine', and 'weather' take more risks, and 'maintain', 'staff', and 'fatigue' have great ability of the risk spillover.

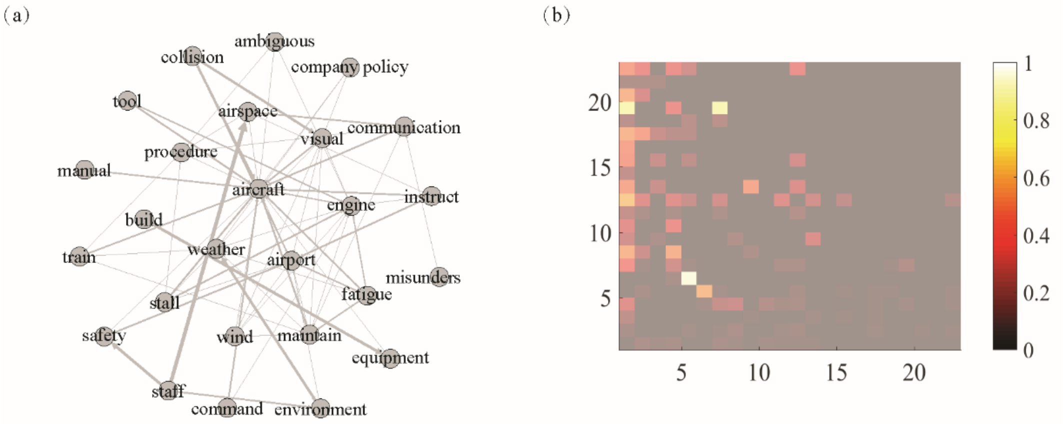

Figure 4. The risk network. (a) The topological structure of the risk network, and (b)the correlation between risk features.

Table 1. The degree of the risk network.

Number	Risk features	In-degree	Out-degree	Total degree
1	aircraft	7.18	0.46	7.64
2	engine	2.17	0.39	2.56
3	procedure	0.76	0.25	1.01
4	airport	2.30	1	3.3
5	weather	1.83	0.76	2.59
6	environment	1.07	1	2.07
7	visual	1.54	0.95	2.49
8	instruct	0.41	1.3	1.71
9	airspace	1.32	0.73	2.05
10	safe	1.14	0.68	1.82
11	wind	0.46	0.49	0.95
12	maintain	1.22	2.4	3.62
13	communication	0.79	1.4	2.19
14	equip	1.0	0.4	1.4
15	train	0.22	1	1.22
16	manual	0.01	0.44	0.45
17	stall	0.1	1.4	1.5
18	company policy	0.04	0.4	0.44
19	collision	0.15	2	2.15
20	command	0.02	0.75	0.77
21	ambiguous	0.01	0.27	0.28
22	fatigue	0.16	1.7	1.86
23	build	0.24	1	1.24
24	tool	0.01	0.83	0.84
25	misunderstand	0.03	0.25	0.28
26	staff	0.16	2.2	2.36

3.3. Risk identification

Before identifying the risk, the dimension method needs to be verified. Here we use a fourth-order Runge-Kutta to get the numerical solution of Eq.(3). The initial value of each risk feature is set to 1. The risk feature would reach the steady state by the fourth order Runge-Kutta. Fig.5a shows the process of each risk feature from the initial state to the steady state. Fig.5b shows the comparison between our proposed method and simulation results. Our method has high precision and predicts the risk rate of each risk feature. Compared to the degree of each risk feature which just considers static property, ’procedure’, ’maintain’, and ’communication’ would also attract the attention of the manager.

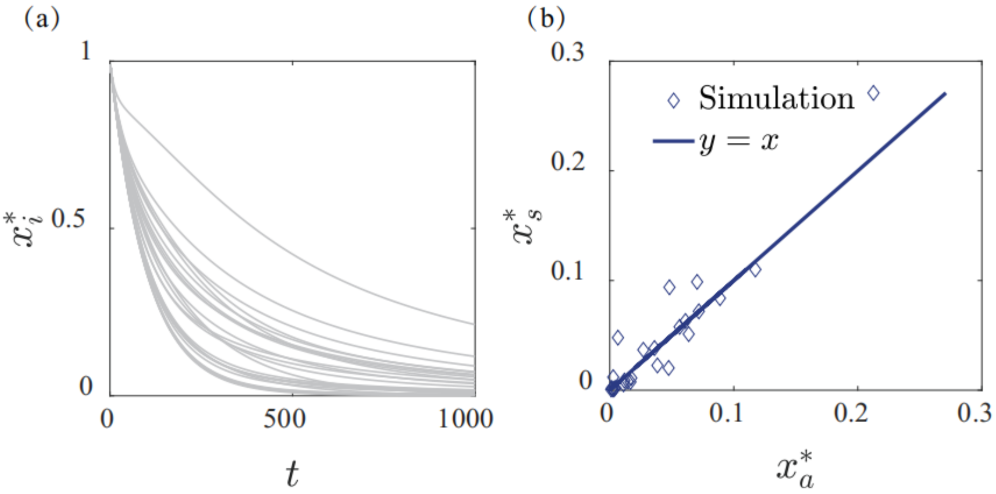

Figure 5. The risk level of each risk feature. (a) The process of the risk level from the initial state to the steady state, (b)The comparison between analytical solution and simulation results.

To analyze the ranking of risks under different indicators, we list the top 10 risk features in Table 2. It shows that risk features such as procedure, wind, weather, communication, and maintenance are critical risk factors.

Table 2. The top 10 risk features under three indicators

TF-IDF	Total degree	Risk rate
engine	aircraft	aircraft
visibility	procedure	procedure
procedure	airspace	weather
airport	staffing	communication
weather	weather	engine
wind	communication	wind
environment	engine	maintain
safety	wind	airspace
communication	safety	airport
ambiguous	maintain	train

4. Conclusions and Discussion

To reduce the consequence severity of unsafe events, managers should identify risks and take appropriate actions for unsafe events. Benefiting from the massive safety incident data and textual data mining techniques, we first analyzed the textual data from ASRS to get the dictionary and get 26 risk features through the risk feature vector space model. The result shows that procedure, visibility, wind, communication, and ambiguity are important risk features. Then we develop a weighted directed risk network that considers the frequency of simultaneous occurrence as the weight of links. The network shows the relation between different risk features and the static characteristics of each risk feature. In reality, accidents always are due to some risk feature accumulation. Therefore, considering the characteristics of influence and propagation among risk features, we propose a risk propagation process based on the SIS model. The risk rate of each risk feature is derived which considers the propagation dynamics of each risk feature. Through the risk rate of each node, we could identify the critical risk feature. Meanwhile, when some incidents happen during a flight, the manager could take different contingency plans based on the risk network which can prevent further propagation. At last, we list the top 10 risk features under different indicators and find procedure, wind, weather, communication, and maintenance are critical risk factors.

Furthermore, understanding the textual data from ASRS is beneficial to understand the many effects of aviation accidents. The proposed in this paper can help the manager determine where a particular aviation accident is in the risk network, and then the appropriate measure is taken. However, the results are limited by the database and would be incomplete. In the future, the method in the paper could be optimized from different data sources and applied in real aviation management which is meaningful.

Reference

1. W. K. Lee, S. J. Kim, Roles of safety management system (sms) in aircraft development, International journal of aeronautical and space sciences 16 (2015) 451–462, https://doi.org/10.5139/IJASS.2015.16.3.451.
2. D. Zhou, X. Zhuang, H. Zuo, J. Cai, X. Zhao, J. Xiang, A model fusion strategy for identifying aircraft risk using cnn and att-bilstm, Reliability Engineering & System Safety 228 (2022) 108750, https://doi.org/10.1016/j.ress.2022.108750.
3. G. Walker, Redefining the incidents to learn from: Safety science insights acquired on the journey from black boxes to flight data monitoring, Safety Science 99 (2017) 14–22, https://doi.org/10.1016/j.ssci.2017.05.010.
4. M. Janic, An assessment of risk and safety in civil aviation, Journal of Air Transport Management 6 (2000) 43–50, https://doi.org/10.1016/s0969-6997(99)00021-6.
5. X Li, F I Romli, S Azrad, M Amzari. An Overview of Civil Aviation Accidents and Risk Analysis."Proceedings of Aerospace Society Malaysia 1.1 (2023): 53-62, https: //www.aerosmalaysia.my/aeros_proceedings/index.php/journal/article/view/24
6. S. H. Stroeve, P. Som, B. A. van Doorn, G. B. Bakker, Strengthening air traffic safety management by moving from outcome-based towards risk-based evaluation of runway incursions, Reliability Engineering & System Safety 147 (2016) 93–108, https://doi.org/10.1016/j.ress.2015.11.003.
7. E. Ancel, A. T. Shih. The analysis of the contribution of human factors to the in-flight loss of control accidents. 12th aiaa aviation technology, integration, and operations (atio) conference and 14th aiaa/issmo multidisciplinary analysis and optimization conference. 2012. https://doi.org/10.2514/6.2012-5548
8. E. Ancel, A. T. Shih, S. Jones, M. S. Reveley, J. Luxhøj, J. K. Evans. Predictive safety analytics: inferring aviation accident shaping factors and causation. Journal of Risk Research 18.4 (2015): 428-451. https: //doi.org/10.1080/13669877.2014.896402
9. D Kelly, E Marina. An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017. Journal of safety research 69 (2019): 155-165, https://doi.org/10.1016/j.jsr.2019.03.009.
10. A. Gautam, N. Garg, Impact of perceived stress safety attitude and flight experience on hazardous event involvement of aviators, Defence Life Science Journal 6 (2021) 235–241, https://doi.org/10.14429/dlsj.6.16800.
11. J. Lee, M. Mitici, An integrated assessment of safety and efficiency of aircraft maintenance strategies using agent-based modelling and stochastic petri nets, Reliability Engineering & System Safety 202 (2020) 107052, https://doi.org/10.1016/j.ress.2020.107052.
12. H. Zhou, T. A. L. Genez, A. Brintrup, A. K. Parlikad, A hybrid-learning decomposition algorithm for competing risk identification within fleets of complex engineering systems, Reliability Engineering & System Safety 217 (2022) 107992, https://doi.org/10.1016/j.ress.2021.107992.
13. C. V. Oster Jr, J. S. Strong, C. K. Zorn, Analyzing aviation safety: Problems, challenges, opportunities, Research in transportation economics 43 (2013) 148–164, https://doi.org/10.1016/j.retrec.2012.12.001.
14. X. Ma, W. Deng, W. Qiao, H. Lan, A methodology to quantify the risk propagation of hazardous events for ship grounding accidents based on directed cn, Reliability Engineering & System Safety 221 (2022) 108334, https://doi.org/10.1016/j.ress.2022.108334.
15. M. Rey, D. Aloise, F. Soumis, R. Pieugueu, A data-driven model for safety risk identification from flight data analysis, Transportation Engineering 5 (2021) 100087, https://doi.org/10.1016/j.treng.2021.100087.
16. L. J. Connell, Aviation safety incident reporting: Nasa's aviation safety reporting system, in: Transportation Research Board Conference Proceedings, 22, 2000, https://doi.org/10.1136/bmj.39071.441609.80.
17. Y. Gao, Y. Hao, S. Wang, H. Wu, The dynamics between voluntary safety reporting and commercial aviation accidents, Safety Science 141 (2021) 105351, https://doi.org/10.1016/j.ssci.2021.105351.
18. O. Sjo¨blom, Data mining in promoting aviation safety management, in: Safe and Secure Cities: 5th International Conference on Well-Being in the Information Society, WIS 2014, Turku, Finland, August 18-20, 2014. Proceedings 5, Springer, 2014, pp. 186–193, https://doi.org/10.1007/978-3-319-10211-5_19.
19. X. Zhang, S. Mahadevan, Bayesian network modeling of accident investigation reports for aviation safety assessment, Reliability Engineering & System Safety 209 (2020), https://doi.org/10.1016/j.ress.2020.107371.
20. Zhou, Di, et al. Deep learning-based approach for civil aircraft hazard identification and prediction. IEEE Access 8 (2020): 103665-103683, https://10.1109/ACCESS.2020.2997371.

21. X. Zhang, P. Srinivasan, S. Mahadevan, Sequential deep learning from ntsb reports for aviation safety prognosis, Safety Science (2021), https://doi.org/10.1016/j.trip.2021.100502.
22. A. Miyamoto, M. V. Bendarkar, D. N. Mavris, Natural language processing of aviation safety reports to identify inefficient operational patterns, Aerospace 9 (2022) 450, https://doi.org/10.3390/aerospace9080450.
23. T. Dong, Q. Yang, N. Ebadi, X. R. Luo, P. Rad, Identifying incident causal factors to improve aviation transportation safety: Proposing a deep learning approach, Journal of advanced transportation 2021 (2021) 1–15, https://doi.org/10.1155/2021/5540046.
24. A. O. Alkhamisi, R. Mehmood, An ensemble machine and deep learning model for risk prediction in aviation systems, in: 2020 6th Conference on Data Science and Machine Learning Applications (CDMA), IEEE, 2020, 54–59, https://doi.org/10.1109/CDMA47397.2020.00015.
25. Aizawa, Akiko. "An information-theoretic perspective of tf–idf measures. Information Processing & Management 39.1 (2003): 45-65. https://doi.org/10.1016/S0306-4573(02)00021-3
26. M. Wu, W. Dai, Z. Lu, Y. Zhao, M. Wang, The method for risk evaluation in assembly process based on the discrete-time sirs epidemic model and information entropy, Entropy 21 (2019) 1029, https://doi.org/10.3390/e21111029.
27. M. J. Keeling, K. T. Eames, Networks and epidemic models, Journal of the royal society interface 2 (2005) 295–307, https://doi.org/10.1098- /rsif.2005.0051.
28. B Barzel, and A L Barabási. Universality in network dynamics. Nature physics 9.10 (2013): 673-681. https://doi.org/10.1038/nphys2741
29. C. Lv, Z. Yuan, S. Si, D. Duan, S. Yao, Cascading failure in networks with dynamical behavior against multi-node removal, Chaos, Solitons & Fractals 160 (2022) 112270. https://doi.org/10.1016/j.chaos.2022.112270.