Active Safety Methodologies for Rail Transportation

Book · October 2020

CITATIONS
0

READS
51

2 authors:

Li-min Jia
Beijing Jiaotong University
585 PUBLICATIONS **2,119** CITATIONS
SEE PROFILE

Yong Qin
Beijing Jiaotong University
463 PUBLICATIONS **1,710** CITATIONS
SEE PROFILE

Advances in High-speed Rail Technology

Yong Qin · Limin Jia

Active Safety Methodologies of Rail Transportation

Springer

Advances in High-speed Rail Technology

More information about this series at http://www.springer.com/series/13506

Yong Qin • Limin Jia

Active Safety Methodologies of Rail Transportation

Springer

Yong Qin
Beijing Jiaotong University
Beijing, China

Limin Jia
Beijing Jiaotong University
Beijing, China

ISSN 2363-5010 ISSN 2363-5029 (electronic)
Advances in High-speed Rail Technology
ISBN 978-981-13-2259-4 ISBN 978-981-13-2260-0 (eBook)
https://doi.org/10.1007/978-981-13-2260-0

Library of Congress Control Number: 2018954656

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Railway is one of the most sustainable ground transportation modes with advantages of safety, reliability, punctuality, high efficiency, and environmental protection. It is the backbone of the comprehensive transportation system in Europe, America, Japan, and Korea, especially in China. Safety is the core competitiveness and permanent goal of rail transportation. The lack of technical safety assurance would be a devastating strike to the railway industry, and leads to serious social issues. To ensure the rail transportation safety, some scientific problems should be solved in theory to satisfy urgent industry demands, including assessing the real-time risk quantitatively, classifying the service state of equipment in high risk level and predicting its growth accurately, conducting the risk control in system level, and building a perception and pre-warning-based active safety assurance system. In addition, the indigenous innovations on core technical equipment and software should be developed.

The main purpose of rail transportation system safety is to describe the real-time risk and its evolution rule quantitatively and formally, analyze the system risk profiles and their relationship, as well as research the accident control method in high risk level. Traditional risk analysis methods could not meet above demands. Because these methods are based on the mechanism or experience and they are qualitative or semi-quantitative. This book proposes the data-driven-based active safety analysis methodologies – safety region based safety analysis theory. The content of this book includes three aspects. The first one is the system of safety analysis methodology based on the safety region. The second one is the method of fault diagnosis and prognosis for rail transit trains. The third one is the dynamic assessing method for the rail transport network operation. In addition to the above contents, the authors also research on the traffic operation risk analysis model based on safety region and the traffic crash risk evaluation model based on reliability theory. Sufficient field examples are provided to verify the proposed methodology.

The authors have constantly studied in this area for nearly 20 years and have received much support from national projects on basic theory and key technology

researches, such as the natural science foundation of China (NSFC) [Grant: 60332020], national 863 plans projects [Grant: 2011AA110501], national technology support projects [Grant: 2011BAG01B02], the doctoral fund of higher education program [Grant: 20120009110035], and so on. Research achievements include one second class prize of national prize for progress in science and technology and ten first or second class prizes at ministerial and provincial levels. The authors also have published 60 papers of SCI or EI and 3 Chinese books. Patents include 2 American patents and 21 Chinese patents. Besides two national or industry standards have been formulated. Thirty Ph.D. or master's students have graduated during this research process. The research team has developed the technology and equipment for the trains, including online safety monitoring and warning, safety and reliability assessment, and operation and maintenance optimization. They have been applied to CRH380A/AL high-speed trains and urban trains. The authors also developed safety monitoring and emergency command systems for the rail transit network. They are applied to the operation and command center of rail transit network in Beijing and Guangzhou and so on.

This book will be pretty useful to many individuals, including reliability and safety professionals working in the transportation industry, transportation system administrators, transportation engineering undergraduate and graduate students, researchers and instructors in the area of transportation, and engineers at large. The Ph.D. candidates Xuejun Zhao and Linlin Kou mainly assisted in writing this book. During this process, the authors also received the great help from Yuan Zhang, Yangfang Yang, Yong Fu, Mingming Wang, Shan Yu, Zhenyu Zhang, Ting Yun, Wantong Li, Dandan Wang, and so on. The authors are really grateful for that.

In this book, many valuable references have also been cited. The authors tried to keep a style of clear definition, with best effort, so as to make all kinds of readers have a clear understanding of the rail transportation safety. Due to the author's knowledge level and the depth and breadth of the study, the views, methods, and theories mentioned in the book certainly have some deficiencies. Do not hesitate to connect the authors to provide your valuable advice.

Beijing, China

Yong Qin
Limin Jia

Contents

Chapter 1
Fundamental of Rail Transportation Active Safety

1.1 Research Paradigm of Rail Transportation Active Safety

1.1.1 Concepts and Methodologies

Active safety assurance plays a key role in complex large-scale engineering system, and it is one of ways to keep system working properly (the other is the passive safety). Based on the system safety theory, active safety emphasizes the ability of perception of system state and making output under control, so as to reduce the system risk and avoid accident. Essentially, active safety is to model, analyze and control the activities of complex system. Recently, active safety becomes an interdisciplinary research area across safety science, control science, information science, intelligent system and other disciplines, and has already have a more and more essential role in the transportation, power system, Internet, military system, nuclear industry, aerospace system and so on.

Safety region (SR) is a quantitative model used for the description of safe and stable running region of the whole system. Currently, the theory and approaches of safety region has been studied deeply in complex power systems [1, 2] and also introduced to rail transportation field [3, 4]. This theory provides a new method for the development of online monitoring, early warning and risk assessment of rail transportation system.

Safety region is defined as a state or feature space to describe the system dynamic behavior. Let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of characteristic(state or feature) variables, in which n is the number of the space dimension. The characteristic variables may contain both discrete variables and continuous variables. Define space E as safety region: within the boundary of E is safe space; otherwise is.

Accident or unsafe space $\bar{E}$. The boundary is determined by the threshold of system safe state, i.e. the accepted risk level that can ensure system safety.

Y. Qin, L. Jia, *Active Safety Methodologies of Rail Transportation*, Advances in High-speed Rail Technology, https://doi.org/10.1007/978-981-13-2260-0_1

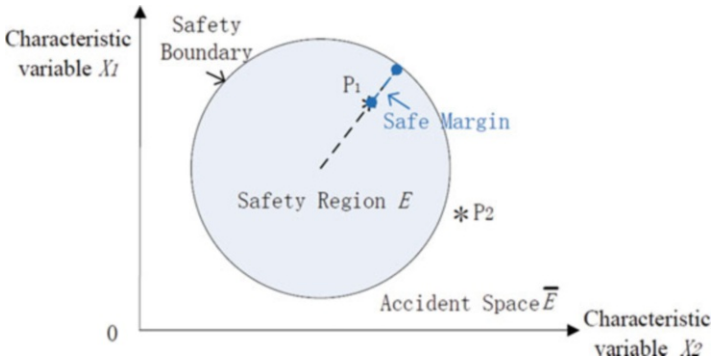


Fig. 1.1 Schematic diagram of two-dimension safety region

The boundary of the safety region is only determined specifically to a certain system. Usually, the state of the system located in safety region is called the balanced state or safe state. If the character point falls in the safe space, then the system is confirmed to be safe, with the distance between the point and boundary, called safe margin, to assess the safety level of the system. Otherwise, the point falls in the accident space when it breaks through the safety boundary, indicating that the risk reaches an unacceptable level and then causes the accident.

Figure 1.1 shows a safety region consists of 2 dimension variables, in which represents respectively system running safely and accident taking place. Obviously, the crucial task to use safety region to denote system safety is to obtain the safety boundary – a decision function returning a safe threshold that differentiate the state of safety and accident.

Off-line space division and boundary estimation have to be conducted before online risk evaluation, as shown in Fig. 1.2. This book provided two ways to divide safety space for two typical data classification situations. One is the directed acyclic graph SVM and the interval fuzzy set based IT2FCM methods which mainly focus on the dataset with complete samples (fault and normal), and those samples are nonlinear signals embedded in strong noise. The other is the kNN and one-class SVM based method for single class classification with big data in normal condition. System state determination and its safety margin calculation in real time are provided after off line computation, as shown in Fig. 1.3.

1.1.2 Research Architecture

A data driven enclosure circle for active safety assurance model for railway transport system was provided as seen in Fig. 1.4. This model has four steps: condition monitoring, risk assessment, risk control and emergency management. It has the following characteristics: (1) Data-driven based: It is an automatic and intelligent knowledge process based on real-time monitoring data. (2) Integrated treatment: It is

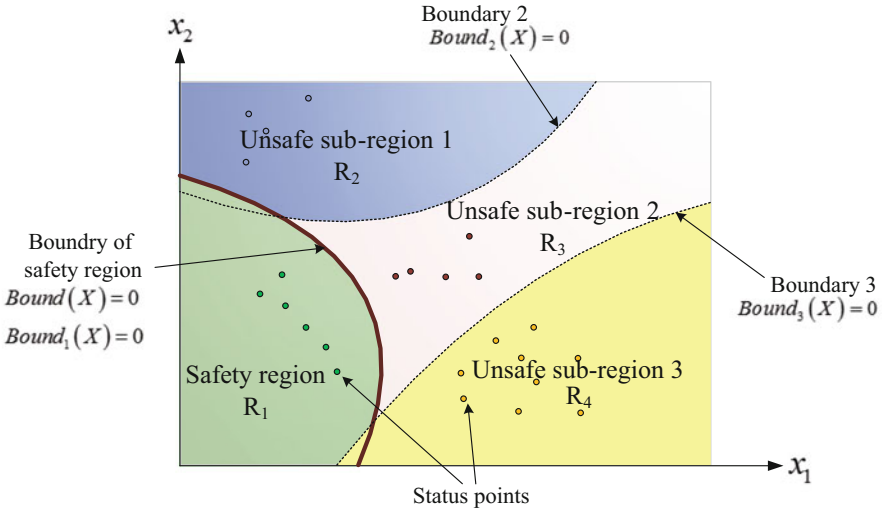


Fig. 1.2 Space classification of safety region

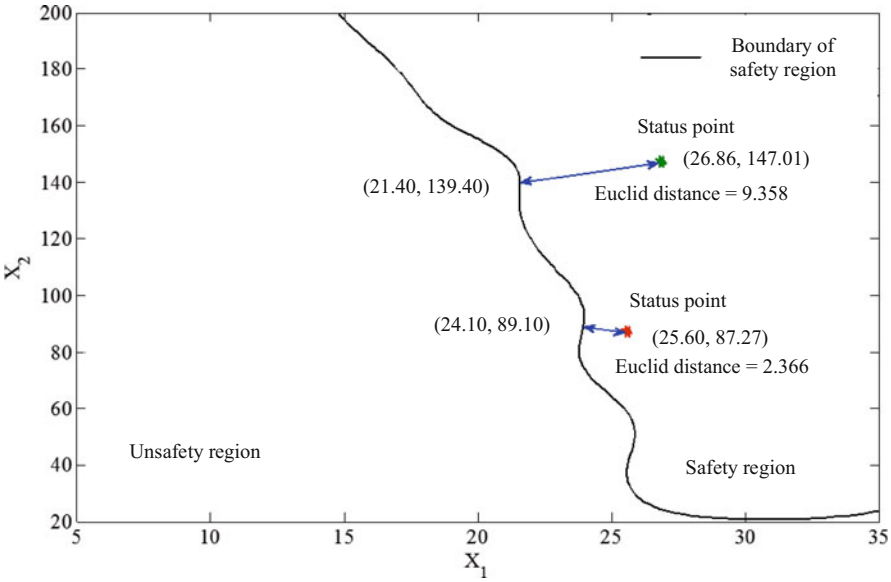


Fig. 1.3 Real time risk evaluation

an entire process optimization from risk identification to risk control. (3) Advanced information processing technologies: It took full use of advanced technologies, such as Internet of Things, sensor networks, cloud computing, knowledge processing and so on.

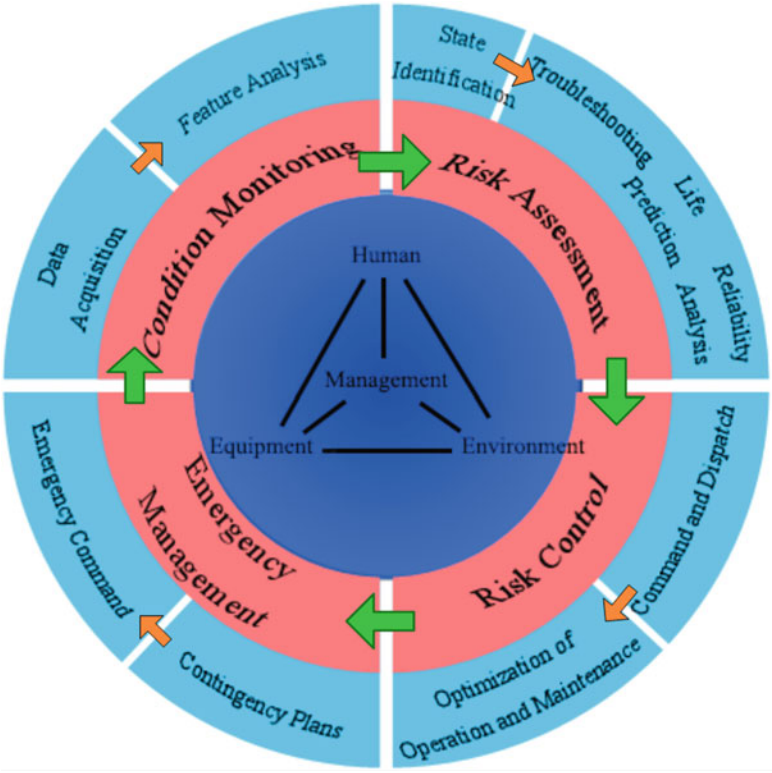


Fig. 1.4 Closed-loop model of the active safety system

From an implementation view, the active safety assurance system has a hierarchical structure, including perception layer, intelligent layer and system layer, as shown in Fig. 1.5. Real time monitoring and information fusion with advanced sensor and network techniques are conducted in perception layer. Intelligent layer is the core layer mainly for fault diagnosis, risk evaluation, reliability calculation and remaining useful life prediction, based on feature extraction and state identification. System layer provides user interface. It integrates the perception layer and intelligent layer to invoke their functions for user requirements, and supplies a systematic service.

1.2 Literature Review

1.2.1 Safety Region Estimation Theory and Methods

This book is based on the basic principles of regional division, putting forward the idea of using state-based safety domain estimation to identify. So it is necessary to

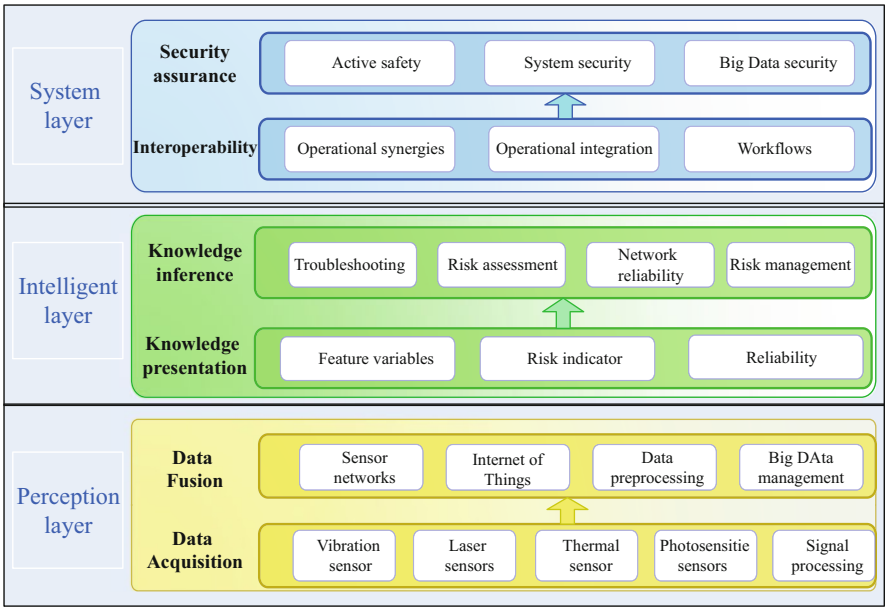


Fig. 1.5 Hierarchical structure of active safety system

introduce the state-based identification of regional division of the state related research.

In the fields of power system, automotive vehicle, aerospace, mechanical electron, computer and so on, some scholars have studied the relevant aspects such as state identification, fault diagnosis and pattern recognition based on regional division. Zheng Tao et al. [5] focused on multi-power open-loop operation of distribution network, combining the idea of regional division, by judging the substation transformer low-voltage side of the switch fault information to improve the fault location method. Tian et al. [6] studied the boundary search of vehicle operating characteristic parameters and proposed a hybrid search algorithm which combined adaptive genetic algorithm and floating search algorithm, and selected the optimal characteristic parameter of vehicle operating state subset. In order to overcome the shortcomings of DV-Hop algorithm process, Xia Shaobo et al. [7] proposed a DV-Hop improvement algorithm based on hop count region partitioning, introducing RSSI ranging technology and limit hopping mechanism to optimize the combination of beacon nodes. Then with multiple trilateration methods, he used the centroid method to determine the unknown node coordinates. Safizadeh M S et al. [8] aimed at the state detection of rolling bearings, extracted the state features from vibration acceleration signals and load signals respectively, and clustered the two-dimensional state feature points to obtain the distribution regions of different state points. Yan Zhiyong [9] studied the classifier based on decision-making boundary from the perspective of dividing data space and studied the classifier

using decision-making boundary as tools. The theoretical framework of classifier was studied based on decision-making boundary from the perspective of dividing data space. Based on this theoretical framework, the classifier is improved.

In the field of information safety, a safety region refers to different logical subnets or networks within the same system that are classified according to the nature of information, use of the main body, safety goals and strategies. Each logical area has the same safety protection requirement and the same safety access control and border control policies, and have trust relationship between regions. And the same network safety domains share the same safety policy [10]. Intuitively interpreted as a safety region is to protect different safety needs of information and information carriers, the system with the same safety requirements of the trusted or untrusted part is divided into different safety zones which are established safety connection by trusted way [11]. The research and application of safety domain based on this concept have been extended to network control [12, 13], road traffic [14, 15], e-government and so on [16].

In the field of power system safety, as early as the 1980s, some scholars in the United States proposed the safety region approach to the stability and safety of the power system [17]. In the early 1990s, some scholars in our country also researched the power safety based on the dynamic safety region to study the power system safety [18]. Recently, many scholars at home and abroad have carried out extensive and in-depth research on the practical safety region of complex power system. Among them, Yu Yixin's researches are the most in-depth study. The United States Department of Energy [19] conducted a methodological study of the wide area safety region of power systems and developed practical applications that take into account various constraints such as thermal, voltage, voltage stability, transients, and potential oscillation stability limits. The safety region is described in the form of hyper planes. The validity of the safety region estimation method is verified by numerical simulation of the system model of the Western Electricity Coordination Committee. The Saudi scholar Mohamed A. El-Kady et al. [20, 21] gave a framework of the identification method of power system operation safety zone. The method of quantifying the safety level by Euclidean distance was given. Based on this, the safety assessment under different operating states is carried out. Yu Yixin et al. [22] proposed that the study of a large real power system shows that under some important predictions, the practical dynamic safety region boundary which can be surrounded by a plane perpendicular to the coordinate axis and a hyper plane that describes the critical point of transient stability by injecting upper and lower limits into each power injection space in the injection power space that guarantees the stability of transient power angle. And with numerical simulation method searching for a large number of critical points and the practical dynamic safety domain boundaries are obtained by least squares fitting. In addition, based on the theoretical results of safety do-main estimation, Yu Yixin and his team also studied the dimensionality reduction visualization method of practical dynamic safety domain, and proposed a method to reduce the practical dynamic safety domain with high dimension to ensure transient stability into three dimensions which can allow

dispatchers to see the stability margin at different points in the injection space for the current operating point so that pre-decisions and even emergency controls can be quickly and accurately implemented [23]. Based on the probabilistic stability model of the safe region, Wang et al. [24] used the theory of dynamic systems to determine the linear boundary of the dynamic safety region near the dominant unstable equilibrium point of power system transient stability. And the joint probability distribution weighted by random variables is obtained by Edge worth series expansion based on semi-invariant, so as to construct the transient stability probability model of power system.

In the field of rail transportation, after 2010, some scholars conducted field-related researches on the safety analysis of rail transportation vehicles from different angles and depths. Jin Xuesong et al. [25] established the derailment model of train based on vehicle-track coupling dynamics in complex environment, and used different derailment criterion and dynamic simulation results to get the safe operation limit of high-speed train under complex environment and defined it as operating safety region. Zhang et al. [26] studied the influence of slip angle and bridge height on the aerodynamic load characteristics of high-speed trains. The aerodynamic loads are applied to the dynamic model of high-speed train as external loads. The operational safety of high-speed trains on the bridge is analyzed and the operational safety zone of high-speed trains on the bridges is given. Qin Yong, Jia Limin and Zhang Yuan et al. took the lead in independently and systematically proposed a method for estimating the safety margin of a rail transportation system. For the first time, the definition, formalization and processing steps for the safety of a rail transportation system [27, 28]. And apply it to the analysis of the safe operation of the key equipment of rail vehicles and the influence of rail lines on the running safety of trains [29]. On the basis of this, a technology system of active safety guarantee for rail transportation based on safety domain is gradually established.

1.2.2 State Identification and Predication of Train Equipment

Rail transportation trains mainly include running system, power system, braking system, communication signals, auxiliary system and other subsystems. Running system is used to guide the vehicle along the track, and the weight of the vehicle passed to the rail. The power system is mainly used to supply power for trains and its main function is to transfer electrical energy from the traction power supply network to the mechanical energy to drive the vehicle. Braking system is mainly used to achieve the train speed control (deceleration, no acceleration or stop), its functions include the implementation of braking and braking relief. Auxiliary system is mainly used to achieve the train lighting, ventilation, air conditioning, emergency power supply and other ancillary functions.

1.2.2.1 Safety State Analysis of Equipment

1. Running System

The research on the safety of running-critical equipment mainly focuses on the objects of wheel sets, suspensions, axle boxes, gears, motors and frames. Abroad, research methods made use of signal processing, statistical limit checking and based on empirical methods. Arash et al. [30] used time-spectrum kurtosis method to analyze the high-frequency sound data of axle box bearings collected along the railway to infer the fault condition. Goo [31] analyzed the stability of the bogie using the finite element analysis method. British scholar Stefano Bruni et al. [32] summarized the research and application of the condition monitoring methods of suspension system and bogie based on sensor information and mechatronics technology. Gharavian et al. [33] based on Fisher discriminant analysis and principal component analysis of the two methods to extract the fault characteristics of the gear box, with good fault classification separation performance. Polach et al. used non-linear parameter estimation method to analyze wheel-rail contact force to estimate wheel wear.

In our country, the research methods adopted are similar to those abroad. The research mostly focuses on the fault diagnosis based on vibration signals and the design of condition monitoring system. Xiao et al. [34] established the three-dimensional wheel-rail instantaneous rolling contact elastic-plastic finite element model with the train speed of 300 km/h, by using the nonlinear finite element software ABAQUS to analyze the wheel rail contact plaque lateral creep, longitudinal creep and creep force distribution. And the result is used as the input of the stability diagram and the damage function to analyze the influence of different friction coefficient on the contact patch fatigue index and the wheel damage distribution. Zhao Rong et al. [35] proposed a high-order spectral feature extraction of rail vibration signals and a particle swarm optimization (PSO-SVM) algorithm for vehicle wheel flaw recognition. By establishing the vertical coupling model of vehicle track and the wheel abrasion model, the rail vibration response of normal wheel and abrasion wheel is calculated. In order to study the law of tread wear on high-speed train wheels, Han Peng et al. [36] conducted follow-up tests on a high-speed EMU serving a line and recorded the amount of tread wear during the rounding cycle. And, based on the two-time fitting of the wear statistics, a wheel-wear prediction model is proposed. Cao Qingsong et al. [37] proposed a vertical coupling dynamic model of vehicle body-frame-suspension-rolling bearing-wheel and rail to solve the problem of the looseness of the inner race and journal of the high-speed train rolling bearing. In addition, the nonlinear fourth-order Runge-Kutta numerical integration and test method are used to study the nonlinear dynamic characteristics of high-speed train rolling bearings supported loose system under different clearances and traveling speeds. Wang Jing et al. [38] studied the fault vibration characteristics of traction gears, locomotive axle box bearings and motor bearings, the mechanism of frequency band failure, as well as the monitoring methods and fault diagnosis signal processing algorithms. Based on the traditional

particle filter and Kalman filter parameter estimation algorithm, the introduction of uniform sampling strategy, Fang Yu et al. [39] established a rail vehicle suspension system condition monitoring method. According to the established vertical dynamic model of railway vehicle system and the vertical state space model, the parameters of suspension system of railway vehicle are simulated and estimated. In addition, some scholars have studied the state of the body based on optical fiber sensing technology [40].

2. Power System

The study of the safety state of the key components of the operation safety system of the power system mainly focuses on the objects of the bow power supply equipment, traction motors and traction converters.

Aydin et al. [41] proposed an optimized kernel parameter tracking method to detect the abnormalities in the bow network system. Karakose [42] based on the real-time acquisition of arch network current data, using low-pass filter and fuzzy S-transform method to process the bow network current signal to obtain the characteristic data, in order to achieve the bow network system state monitoring. Marco Bocciolone et al. [43] proposed the use of FBG optical sensor monitoring network system, this method can monitor bow network contact for online monitoring. Bucca et al. [44] evaluated the wear condition of the bow web by plotting the wear curves of the bow system with the contact and wear models of the flow receiver and catenary. Switzerland ELAG company [45] has developed a non-contact inspection system based on a 2D laser sensor for catenary and pantograph. Mohamed et al. [46] used the fault diagnosis method based on neural network to classify the fault types of power transformers. D. Grillo et al. [47] pro-posed a monitoring system for an electrified railway traction system that can measure and transmit the relevant electrical parameters online. H. Kamijo et al. [48] designed a superconducting floor-mounted locomotive vehicle traction converter system that can measure the electrical parameters and test the dynamic characteristics of the traction converter based on the dynamic simulation of rolling stock. Mermet-Guyennet, M. et al. [49] studied the reliability of railway traction systems and discussed the design of traction converters.

In China, bow net monitoring, traction motor and traction converter fault diagnosis research and the results are also more abundant. Qin Yong et al. [50] established the dynamic model of bow mesh and analyzed the relationship between the catenary irregularity and contact force, and established the bow network system state prediction model based on catenary irregularity. Liu Kai et al. [51] analyzed the bow network operation by detecting the contact pressure and pull-out value of the pantograph and the catenary, and independently developed a bow-net monitoring system based on the embedded system. Peng Wei et al. [52] proposed a mathematical model to solve the contact line vibration based on triangulation method for the pull-out value and the contact line height fault caused by bow vibration, and adopted the method of video surveillance and fault image diagnosis based on 3G technology, which realized online monitoring and fault diagnosis of bow network. Zhang et al. [53] studied and designed a comprehensive system for fault diagnosis of traction system of metro vehicles integrated with vehicle-mounted, ground and monitoring

center. Chen et al. [54] pro-posed a fault diagnosis model based on space-time fusion of information and successfully applied it to the fault diagnosis of electric locomotive traction motors. Liu Ling et al. [55] studied the fault identification method of traction converter of CRH5 EMU based on BP neural network and simulated the open circuit fault diagnosis of converter inverter. Wang Yi et al. [56] proposed a method of fault diagnosis based on wavelet analysis and decision tree based on the change of output voltage of main converter under different failure mode of main converter of Shaoshan 8 electric locomotive.

3. Braking System

The research on the state assessment of the key equipment of brake system at home and abroad is more common in Europe and Asia and more related to the control of brake parameters and the reliability analysis.

Foreign studies mostly focus on the target parameters of the braking system and fault detection. Niu et al. [57] used the method of combining linear categorical transformation model and data-driven to monitor the abnormal state of braking system online. Zhuan [58] studied the fault detection and isolation of the braking system through the difference between the steady-state speed fault and the intact brake system.

Domestic research emphasizes reliability analysis of key equipment in braking system. Tu Jiliang et al. [59] proceeding from the design principle of high-speed EMU braking system, it elaborates the design concept and implementation method of fault diagnosis and safety measures of EMU braking system. From the design of intrinsic safety proceed and taking into account the various possible faults and their possible consequences of the severity and designing the corresponding identification and control methods, the system automatically or prompts the crew to handle faults or isolate faulty equipment, speed-limit operation or automatic parking, thus ensuring the safety of the train. Liu Jie et al. [60] established support vector machine frame including feature selection, feature vector selection, model construction and decision boundary to monitor the braking system faults of high-speed trains. Jia Limin [61] conducted a reliability analysis on the brake system of subway vehicles. After the qualitative and quantitative parameters were estimated by the method of fault tree analysis, the fault tree model was derived by Monte-Carlo method based on probability and statistical analysis of basic event lifetimes. Ding Jianbo et al. [62] summarized the common fault characteristics by analyzing the pneumatic transmission principle and structural characteristics of the 120 type freight train brake engine, established the fault tree and fault knowledge base of the brake system by using the fault tree analysis method, and developed a fault diagnosis system. Wu Mengling et al. [63] developed the domestic urban rail transportation vehicle unit brake and its reliability characteristics, described the method of durability test. Based on the test results, the reliability of the unit brake was evaluated. Huang Zhiwu et al. [64] studied MAS-based fault diagnosis theory and technology, and developed an online diagnosis system for synchronous braking system of heavy-duty combined trains. The system was applied to HXDl heavy-haul combined trains on Daqin Railway. Tian Chun et al. [65] evaluated the reliability of the relay valve used in the brake

system of rail transportation vehicles through the durability test. The results show that the relay valve failure obeys the Weibull distribution with shape parameter m = 3.43, and the failure rate increases with the increase of test cycles. The main failure modes are fatigue failure of the V-ring and return spring.

4. Auxiliary System

The research results on the evaluation of key equipment state in the auxiliary system mainly focus on the storage battery, auxiliary reverse and air-conditioning devices, among which the research on the battery has been carried out as early as the 1980s.

The French scholar Khadija El Kadri et al. [66] proposed a model and control algorithm for heavy-load electric drive system simulation, in which the battery was modeled and the shape behavior of the battery was analyzed. The research results can provide support for the study of decision-making and fault-tolerant behavior in the case of fault. Wu Canpei et al. [67] studied the remote monitoring and control system based on Web for the emergency power supply of high-speed railway trains. The real-time display technology based on Ajax and SVG remotely monitors the emergency power supply on the ASP.NET platform. Berger et al. [68] proposed a simulation model of on-line inverter current protection for trains, which has good effect on the protection of inverter inverter of trains. Wu et al. [69] based on the sound signal collected from the air conditioning system, the use of wavelet packet transform method in the neural network combined method of air conditioning system fault diagnosis.

Jia Limin et al. [70] aiming at the fault types of auxiliary inverter system of urban rail transportation, proposed a fault diagnosis method based on wavelet packet and neural network. Liu Gang [71] studied the basic method of fuzzy logic for state inspection of battery technology, established the mathematical model of battery failure, and proposed the basic idea of classifying the failure state of battery as the premise of whether it can be repaired without disassembly. The technical staff of Shanghai metro operation company analyzed the overvoltage monitoring failure and start-up failure of the auxiliary inverter's three-phase AC output, and put forward the corresponding solutions [72, 73]. Guangzhou metro corporation head technicians on the Guangzhou metro Line 5 emergency lighting frequent failures, combined with the train lighting system control principles, to identify the cause of the malfunction and put forward reasonable optimization and improvement measures [74]. Chen Huanxin et al. [75] developed TAFDES, a bus air conditioner fault diagnosis expert system. The system can perform fault diagnosis on the performance test of air conditioner in passenger car and can indicate the cause of the malfunction and the method of repair and adjustment.

1.2.2.2 State Monitoring and Evaluation of Wheel Rolling Bearings

The research objects and application examples of this book are mainly concerned with the rolling bearing or axle box system in the traveling system of trains.

Therefore, the following is a detailed analysis of the research state at home and abroad for the on-line condition monitoring and safety evaluation of train rolling bearings.

The condition monitoring methods of rolling bearings for rolling stock vary with the amount of inspection, mainly including vibration, ultrasonic and acoustic emission, thermal infrared, oil sample analysis and temperature monitoring [76], Among them, vibration and sound-based monitoring methods are the hot topics in the field of rolling bearing monitoring and diagnosis in recent years. Although the method based on sound signal has the characteristics of rich information and non-contact measurement, the sensors required are expensive and susceptible to various external noise and interfere with low signal-noise ratio and high technical difficulty. The vibration monitoring method is widely researched and applied in bearing condition monitoring because it does not need disassembly equipment, vibration signal acquisition is easy and it contains abundant equipment state information and signal processing methods are flexible. Nowadays, most of the bearing diagnosis instruments and systems used in locomotive depots, locomotive factories and some depots in our country, as well as the vast majority of bearing diagnostic instruments on the market, all use vibration analysis methods [77].

Vibration signal-based monitoring methods mainly involve signal feature extraction and feature-based state identification.

1. Signal Feature Extraction

Vibration signal feature extraction technology is the key to rolling bearing monitoring and diagnosis, there are time domain analysis, frequency domain analysis and time-frequency analysis.

Time domain analysis is mainly used for statistical and analysis of signal time-domain features include statistics, model methods and time-domain signal processing [78]. Among them, the statistical method mainly calculates the time-domain index of mean, root mean square, peak, skewness and kurtosis. The model method is to process the vibration signal as a time series, fit the time-serialized parameter model, and take the model parameters as signal features. Time-domain signal processing methods include traditional filtering, convolution, correlation and other digital signal processing, DSP technology and chaotic theory based feature extraction and other methods.

Frequency domain analysis is the commonly used method at present, which can realize fault analysis, that is, identify the fault location compared with time domain analysis [78]. Frequency domain analysis includes spectrum analysis, envelope analysis, cepstrum analysis and high-order spectral analysis. The spectrum or power spectrum is mainly obtained by Fourier transform and Fast Fourier transform, taking the spectrum at the entire spectrum or feature frequency as the signal feature. Envelope analysis, also known as amplitude demodulation or high-frequency resonance technology, includes two steps: band-pass filtering and envelope estimation. It can detect early faults of bearings and is therefore maturely applied in the condition monitoring of rolling bearings [79]. Cepstral analysis is the logarithm of the power spectrum of the signal, which is used to monitor the frequency spectrum. Higher-

order spectra usually refer to bispectrum and tri-spectrum, which are the Fourier transforms of the third and fourth order statistics of the signal.

Time-frequency analysis is the same time in the time domain and frequency domain signal analysis, and more for the analysis of non-stationary signal. Commonly used method are short-time Fourier transform, Wigner-Ville distribution, wavelet transform, Hilbert-Huang transform and other methods. Short-time Fourier transform, also known as window Fourier transform, the signal within the sliding window is Fourier transformed to obtain different resolutions in time and frequency. Wigner-Ville distribution is a bilinear transform to get high-precision FM signal frequency distribution. The wavelet transform is built using translation and scaling invariance to provide analysis of different resolutions on time and frequency scales. Hilbert-Huang transform is to decompose the signal into a number of mode functions, and extract the signal features based on this.

2. Feature-Based State Recognition

Based on the extracted signal characteristics, a variety of identification and diagnostic methods can be used to identify and evaluate states. State identification methods are divided into two major categories, including the traditional one and intelligent one.

Traditional methods of identification include distance classifier, clustering analysis, Bayesian classification and other classification methods, but these methods have some limitations. Therefore, more and more researchers have paid attention to the identification methods of intelligent and hybrid intelligence recently. These methods mainly include neural network, fuzzy reasoning, support vector machine, intelligent group algorithm, rough set theory and so on [80]. Neural networks have good non-linear learning ability and distributed parallel information processing capabilities, and are easy to combine with other intelligent computing methods, so they are widely used in pattern recognition and other fields. Fuzzy reasoning can gather the prior knowledge of experts and is widely used in the processing of uncertainty data. Support vector machine is a novel machine learning algorithm based on statistical learning theory, which can adapt well to small sample and nonlinear data environment. Intelligent group algorithm is a new evolutionary computing technology, including particle swarm and ant colony algorithm, which can be used to solve discrete optimization problems. Rough set theory is a new mathematical tool dealing with ambiguity and uncertainty. It does not need prior knowledge and mathematical models, simplifies information and is very suitable for mechanical fault diagnosis.

In the field of train rolling bearing condition monitoring and fault diagnosis, many scholars at home and abroad have done extensive and in-depth work on the combination and application of the above various signal feature extraction and state identification methods. He Shenghan et al. [81] designed a wireless collector based on low-power wireless communication module and single-chip microcomputer to acquire the vibration signal of high-speed train's axle box, which can be used to detect and evaluate the axle box online. Shang Wanfeng et al. [82] proposed to apply high-order cumulant adaptive filtering algorithm to ace fault diagnosis and

monitoring. The algorithm extracts the characteristics of the monitoring signal to separate the normal bearing signal from the fault signal. He Ping et al. [83] proposed an acoustic sensor-based race fault acoustic signal acquisition system to evaluate the operating state based on the analysis of race fault acoustic signals. Ding Fuyan et al. [84] proposed a relatively complete locomotive bearing condition detection and quality control system solution, including three subsystems: vehicle monitoring system, ground detection and diagnosis system and locomotive bearing state information system. Nabiyev N.K [85] proposed fault diagnosis of axle box bearing based on identification measurement. The method is based on the measurement method, the variability of the vibration signal and the characteristics of the signal it-self. The current automatic monitoring system used in vehicle maintenance company's technical service and maintenance. Yang Jianwei et al. [86] proposed improved wavelet packet and BP neural network for race fault detection. Piezoelectric acceleration sensors are used to collect vibration signals of potentially faulty bearings and perform noise elimination with wavelet. Then, the improved wavelet packet is used to analyze the data and train the improved BP neural network as a fault sample to realize the fault diagnosis. Wei He et al. [87] proposed the use of wavelet SOFM network for rolling race fault diagnosis, Based on the frequency domain features of the vibration signals, the method performs race fault diagnosis through time-frequency domain analysis of neural networks and vibration signals. Yang Jiangtian et al. [88] proposed a locomotive bearing diagnosis system based on Lacplace wavelet analysis and envelope spectrum analysis to extract the fault characteristic frequency. The impulse response of the fault bearing was composed of a series of unilateral attenuated oscillatory signals. The characteristic of race fault characteristic frequency contains less energy and is disturbed by noise. Laplace wavelet is introduced into bearing vibration signal analysis.

The existing methods have good theoretical analysis and practical application effects on typical and significant serious faults. However, there is a great deficiency both in theory and in practice for the early fault diagnosis and life prediction under complicated conditions. And integration and operation and maintenance plan optimization methods have yet to be further in-depth study. At the same time, with the development of detection data toward massive big data, the method of state identification and prediction based on big data is also an important research direction in the future.

1.2.3 Train Safety and Reliability Evaluation

The reliability of a train system is the ability of the system to fulfill the prescribed functions within the stipulated time and under specified conditions. The higher the reliability of the system, the less likely it is to fail, and the greater the probability of completing the prescribed function. Train system safety refers to the ability of the system to avoid accidents. Safety is usually a condition. In many cases, unreliable train systems can lead to system in safety. In the event of a system failure, not only

does it affect the functioning of the system, it can sometimes lead to accidents, resulting in death or property damage. Therefore, to take measures to improve system reliability, both to ensure that the system functions, but also can improve system safety. However, the reliability of a train system is not exactly the same as safety. Their focus is different: reliability focuses on maintaining system function and achieving system goals, safety focuses on preventing accidents and avoiding personal injury and property damage. Reliability studies the state of the system up to the point where the failure occurred before failure occurred, safety focuses on the impact of the failure on the system after the failure. Because of the close relationship between system reliability and system safety, the study of train system safety should be based on the research of train system reliability.

At present, the railway system in developed countries has formed a relatively complete system of safety assessment and safety management and has formulated a series of practical technical standards for safety assessment. For example, the IEC61508 standard for electronic system safety management was officially released in 2000. Based on the IEC61508 standard, CENELEC has successively launched a series of safety standards for different applications of rail transportation: EN50126, EN50128, EN50129 and EN50159, which are adopted by the IEC organization and are currently used by European national railways. Taiwan's high-speed railway in our country is also independently verified and confirmed according to the EN series of standards. So far, the Chinese railway has introduced the European standard IEC61508 in terms of safety management and has formulated the Chinese national standard GB T20438. In the aspect of rail transportation system RAMS management, China transferred the rail transportation industry RAMS standard IEC62278: 2002 developed by the International Electro technical Commission to China's national standard GB T21562–2008 'Reliability, Usability, Maintainability and Safety of Rail Transportation in 2008' norms and examples. ".

At present, some experts and scholars have studied the safety and reliability of high-speed train systems from the perspective of system safety and reliability. For example, based on the reliability theory, Yu Mengge et al. [89] established the multi-body dynamics model of high-speed train and deduced the extremes of system reliability sensitivity to determine the safety curve of high-speed train. Based on the correlation between system reliability and safety, Li Chao et al. [90] analyzed the process of system safety change from three aspects of entity danger event, coupling danger event and system danger event, and put forward a method of stratified coupling analysis of equipment system safety with reliability and improvement of impulse process. Yu Zhuo-min et al. [91] proceeded with the structural sys-tem safety and reliability and applied the product life-cycle management theory. According to the four stages of designing, manufacturing, using, repairing and scrapping of rolling stock, the paper proposed to establish the train of thought, main content and system framework of the whole life cycle structure safety management system of rolling stock. Su Hongsheng [92] analyzed the structure and function of CTCS-3 train control system and combined with the fault data of train control system equipment. From the perspective of system engineering, the reliability of CTCS-3 train control system was analyzed and evaluated by using FTA

method and BN technology, considering the characteristics of multi-factor fault modes such as maintainability, common cause failure and polymorphism. By determining the potential risks of the train control system, the causes of the risks and the possible consequences of the risks, BN is used to establish a risk assessment model to evaluate the safety risks of the train protection alert system.

At present, the analysis of the safety and reliability of trains, especially high-speed train systems, lacks pertinent analysis methods, complete methodologies and assessment processes, and has not yet formulated corresponding industry and national standards. Therefore, there is an urgent need to establish a safety and reliability analysis and evaluation method system suitable for the design and operation of high-speed trains in our country, and to support the design of safety and reliability of high-speed train systems and the healthy operation and maintenance work.

1.3 Research Work of Authors' Group

Railway is one of the most sustainable ground transportation mode with advantages of safety, reliability, punctuality, high efficiency and environmental protection. It is the backbone of the comprehensive transportation system in Europe, America, Japan, and Korea, especially in China. Much progress has been made in China that 20,688 electric multiple units (EMU), about 2586 high speed trains, over 22,000 km of high-speed railway and 4153 km urban railway are in operation, as well as many pioneer achievements around the world such as the longest Beijing-Guangzhou high speed railway, Harbin-Dalian high speed railway in the highest latitude areas, Lanzhou-Sinkiang high speed railway in desert with strong wind. China has taken the first place in operation mileage and equipment manufacture scale of high speed railway. In accordance with the national Medium and Long term Railway Network Plan (revised in 2008), the total length of China's high-speed railways will reach 30,000 km and cover 80% large cities by 2020, and the speed will reach 350 km/h among megalopolis, while the total railway operating mileage is expected to reach 150,000 km.

Safety is the core competitiveness and permanent goal of railway. The lack of technical safety assurance would be devastating blow to the railway industry, and lead to serious social issues. The 7.23 Yong-Wen line major transportation accident in 2011, led to great progress stagnation and negative impact on Chinese high speed rail industry. The operational risk analysis and control of rail vehicle is the foundation to rail safety assurance, as train is the direct carrier of railway transportation. And real time monitoring, diagnosis and prediction of component in high risk level should be conducted after that to reduce the total system risk. However, rail train is an extremely complex electro-mechanical system, with more than 20,000 coupled components. The long service time, extended locomotive routing, heavy-duty, severe environment, impact in high speed condition have brought unprecedented challenges to long term service. Over 40% rail accidents were caused by failed

equipment on train according to statistics. In the meantime, train equipment failure is also the overriding factor of accident in urban rail transportation. Serious accidents caused by defect of rolling stock happened occasionally. In 1998, the wheel set failure of high speed rail train caused 101 deaths in German. On 27th, August, 2007, a dangerous accident of derailment took place on the No. 35203 freight train because of the heat cutting-axle fracture at the 16th car.

Some scientific problems should be solved in theory to satisfy the urgent industry demand: How to assess the real-time risk quantitatively, how to classify the service state of equipment in high risk level and predict its growth accurately, how to conduct the risk control in system level and build a perception and pre-warning based active safety assurance system. In addition, the indigenous innovations on core technical equipment and software should be made.

The authors have worked on railway transport in safe and efficient operation over 20 years, and insisted on an idea that theoretical innovation should be driven by industry requirements, be the foundation of real expertise, and in turn to support industrial development. The author and his team fully took part in the construction, operation and important technical creation of the national train speeding project, Qinghai-Tibet plateau railway project, high speed and urban rail network project, high speed and urban rail train project and so on. They have made significant success in the basic theory,key technologies and application systems of rail system safety and dispatching optimization(shown in Fig. 1.6).

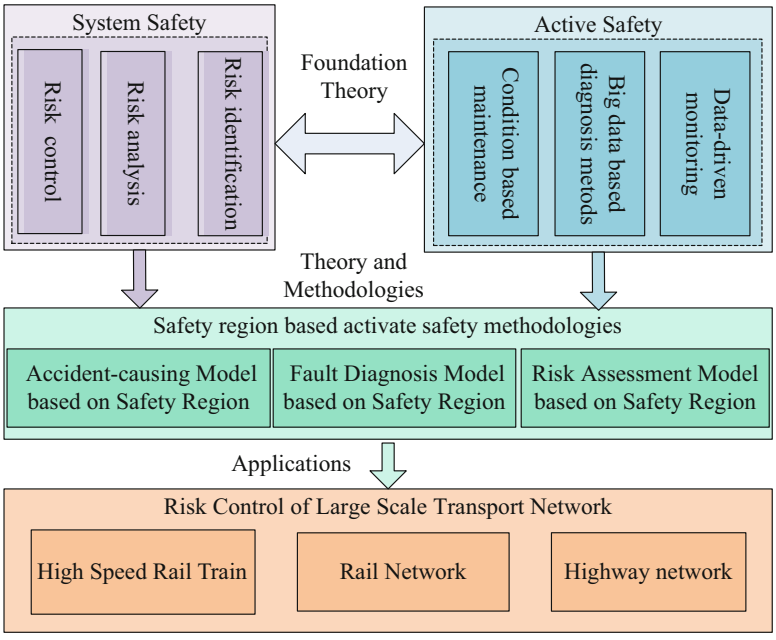


Fig. 1.6 Safety region based active safety technical framework

These innovations have made great contributions to safe operation of the largest and highest-speed railway network, highway network, plateau railway (the Qinghai-Tibet Railway), urban railway network in megalopolises and the key equipment on high speed railway vehicle. They also helped getting more than 11 awards, that the National Science & Technology Progress Award (second class), safe production and scientific achievement (first class) from the State administration of Work Safety, and second prize from the Education Ministry and China Railway Society etc. Despite that, more than 85 patents has been applied, and about 43 patents was granted, one of them received a golden medal for invention in Nuremberg, Germany. Moreover, these innovations have gotten 50 software copyrights. 155 SCI/EI papers have been published and one of them was selected into the global top 1% highly cited papers. The authors edit 4 international academic proceedings, and some sections of an international monograph.

The authors have been constantly studied in this area for many years, and have got much support of national projects on basic theory researches, such as the Natural Science Foundation of China (NSFC) [Research on intelligent transport comprehensive information system and technical of High speed railway, No. 60332020], the New Century Talent Supporting Project by education ministry [Research on complex dynamic system safety theory of high speed rail train group operation, No. NCET-08-0719], the Doctoral fund of higher education program [online safety assessment of key equipment on railway vehicle based safety region theory, No. 20120009110035], IBM International Cooperation Program (Reliability center maintenance) and so on. Much engineering practice and pilot application had been carried out with the support of national 863 plans projects [Research on operational hidden risk mining and evaluation with pre-warning techniques and system (2011AA110501), Research on reliability and safety macro model of high speed rail vehicle genealogy (2012AA112001–07)] and the National Science technology Support Plan Projects [Holography Testing and fault diagnosis technology and equipment development of urban railway operation project (2011BAG02B13), Safety, reliability and availability evaluation method of the next urban rail vehicle system (2015BAG12B01–06)]. Especially during the 12th Five-Year-Plan period, as the technical executive director in chief of the only program in urban rail vehicle operational safety assurance (2011AA110500, total fund is 2.3 billion Yuan, with government grants of 7 3.71 million Yuan), the authors worked 3 years, coordinated more than 10 companies, including the Guangzhou Metro, CRRC, Tongji university etc. They worked more than 6575 person-months and developed the hidden risk mining and evaluation technique, operational fault diagnosis of key equipment on rail vehicle, sensor network of rail vehicle, comprehensive maintenance decision supporting system and some other techniques. The first operational fault diagnosis detection and pre-warning systematic equipment for urban rail vehicle with Chinese own intellectual property rights has been inducted and engineering validated, which was installed on 15 trains and worked on 2 lines between three stations, two sections. All the programs and projects had been finished on August, 2014. Achievements from them were reported to the Ministry of Science and Technology and played a leading role in this area.

These researches not only deepen and widen the system safety theory and its application in transport system, but also improve the safety assurance mode changing from single component to system, negative safety to active safety, small sample analysis to big data analysis. It is original innovation support of system safety related theory to high speed railway, and a new solution of operational system safety for transport network in large scale.

References

1. A.E. Mohamed, A.A. Essam, Framework for identification of power system operating safety regions. The Third International Conference on Net-work and System Safety, Queensland, 2009, pp. 415–419
2. Y.U. Yixin, Review of study on methodology of safety regions of power system. J. Tianjin Univ. **41**(6), 635–646 (Ch) (2008)
3. Y. Zhang, Y. Qin, L. Jia, et al., Research on method framework of safety region estimation in rail transportation system operation safety assessment. J. Syst. Simul. Technol. Appl. **13**, 1018–1022 (2011)
4. Y. Zhang, Y. Qin, L. Jia, Research on methodology of safety region estimation of railway system operation safety assessment. Proceedings of World Congress on Engineering and Technology 2011, 06, pp. 803–807
5. T. Zheng, Y. Pan, K. Guo, et al., Study on fault location method of distribution network based on immune algorithm. Power. Syst. Protect. Control. **42**(1), 77–83 (2014)
6. T. Yi, Z. Xin, Z. Xin, Z. Liang, Study on the recognition method of automobile operation state (I) – selection of characteristic parameters. China. Mech. Eng. **24**(9), 1258–1263 (2013)
7. X. Shaobo, Z. Jianmei, X. Zhu, et al., An improved DV-Hop algorithm based on the hop count region. Chin. J. Sens. Actuarors. **27**(7), 964–969 (2014)
8. M.S. Safizadeh, S.K. Latifi, Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. Inf. Fusion. **18**, 1–8 (2014)
9. Y. Zhiyong, Classifier Based on Decision-Making Boundaries in the Perspective of Partitioning Data Space. Zhejiang University, Hangzhou, 2011
10. Z. Xiang, H.-H. Shen, C. Bing, A cross-safety access control method based on bidirectional defense. Netinfo. Saf. **10**, 19–21 (2009)
11. Z. Zhijie, Safety Domain Division of the Key Theory and Application of [D]. Kunming University of Science and Technology, Kunming, 2006, pp. 5–15
12. N. Wensheng, L. Yahui, Z. Yadi, Research on access control mechanism of embedded systems based on safety domain isolation. Comput. Sci. **40**(06A), 320–322 (2013)
13. N. Wang, Z. Ying-Jian, Z. Jian-Hui, et al., An identity-based routing protocol for secure domain access. J. Softw. **20**(12), 3223–3239 (2009)
14. D. Xiang, Y. Wei, Research on vehicle image recognition and longitudinal safety domain control in front of vehicle. Sci. Technol. Eng. **1**, 100–104 (2014)
15. M. Barzegar, N. Mozayani, M. Fathy, Secure safety messages broadcasting in vehicular network. International Conference on Advanced Information Networking and Applications Workshops, Bradford, 2009, pp. 1055–1060
16. W. Miao, L. Jie, H. Yanjun, Research and application of safety domain division technology in e-government. Comput. Eng. Sci. **32**(8), 52–55 (2010)
17. R.J. Kaye, F.F. Wu, Dynamic safety regions of power systems. IEEE Trans. Circ. Syst **CAS-29** (9), 612–623 (1982)
18. Y. Yu, F. Fei, Active power steady-state safety region of power system. Sci. China. Ser. A **33** (121), 1488–1500 (1990)

19. Y. Makarov, P. Du, S. Lu, T. B. Nguyen, Wide area safety region final report [EB/OL]. http://www.pnl.gov/main/publications/external/technical_reports/PNNL-19331.pdf, March 2010
20. A. E. Mohamed, A. A. Essam, Framework for identification of power system operating safety regions. The Third International Conference on Network and System Safety, Queensland, 2009, pp. 415–419
21. A. Essam, A. Mohamed, Application of operating safety regions in power systems. IEEE PES Transmission and Distribution Conference and Exposition. New Orleans, LA, USA, 2010
22. Z. Yuan, F. Jichao, Y. Yu, et al., A practical dynamic safety domain for power system. Autom. Electr. Power. Syst. **1**, 6–10 (2001)
23. Y.Y. Yanbin, Z. Yuan, J. Hongjie, N. Ben, H. Nanqiang, T. Zhiyu, Z. Yiming, F. Hongjun, Visual dimension visualization using dynamic secure domain. Autom. Electr. Power. Syst. **29** (12), 44–48 (2005)
24. L.-j. Wang, G. Wang, Performance evaluation of transient stability of power systems based on dynamic safety and edgeworth series. Proc. CSEE **31**(1), 52–58 (2011)
25. J. Xue-song, L. Liang, X. Xin-biao, et al., Numerical simulation of dynamic behaviors of high-speed trains in complex environment and analysis of operational safety domains. Comput. Aided. Eng. **03**, 29–41 (2011). 59
26. M. Yu, Z. Jiye, Z. Weihua, Safety of transverse wind traffics on high speed trains on bridges. Chin. J. Mech. Eng. **48**(18), 104–111 (2012)
27. Z. Yuan, Q. Yong, J. Limin, et al., Study on safety domain estimation method for operational safety assessment of rail transportation system. Syst. Simul. Technol. Appl. **13**, 1018–1022 (2011)
28. Y. Zhang, Y. Qin, Z. Xing, et al., Roller bearing safety region estimation and state identification based on LMD–PCA–LSSVM. Measurement **46**(3), 1315–1324 (2013)
29. Z. Yuan, Q. Yong, J. Limin, A estimation of uneven-peak or peak-safe-area based on distribution of risk Points-SVM. J. Cent. South. Univ. Sci. Technol. **43**(11), 4533–4541 (2012)
30. A. Amini et al., Wayside detection of faults in railway axle bearings using time spectral kurtosis analysis on high-frequency acoustic emission signals. Adv. Mech. Eng., **8**(11) (2016)
31. J.S. Goo, J.S. Kim, K.B. Shin, Evaluation of structural integrity after ballast-flying impact damage of a GFRP lightweight bogie frame for railway vehicles. J. Mech. Sci. Technol. **29**(6), 2349–2356 (2015)
32. S. Bruni, R. Goodall, T.X. Mei, H. Tsunashima, Control and monitoring for railway vehicle dynamics. Veh. Syst. Dyn. Int. J. Veh. Mech. Mob. **45**(7–8), 743–779 (2007)
33. M.H. Gharavian, F.A. Ganj, A.R. Ohadi, et al., Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes. Neurocomputing **121**, 150–159 (2013)
34. X. Qian, F. Jun, L. Wang, Effect of friction coefficient on instantaneous rolling contact fatigue of high-speed train wheels. China. Railw. Sci. **37**(3), 68–74 (2016)
35. Z. Rong, S. Hongmei, Research on wheel galling identification algorithm of high speed trains based on high order spectral feature extraction. Chin. J. Mech. Eng. **53**(6), 102–109 (2017)
36. H. Peng, Z. Weihua, Statistic rule and prediction model of wheel pair wear of high speed train. Chin. J. Mech. Eng. **52**(2), 144–149 (2016)
37. C. Qingsong, X. Guo, X. Guoliang, et al., Study on dynamic characteristics of high speed train bearing loose bearing. Chin. J. Mech. Eng. **21**, 87–95 (2016)
38. W. Jing, Study on Vibration Characteristics of Train Wheels and Key Techniques of Diagnosis. Central South University, Changsha, 2012, pp. 5–10
39. F. Yu, C. Long, S. Zheng, et al., Monitoring method of rail vehicle suspension system based on parameter estimation. J. China. Railw. Soc. **35**(5), 15–20 (2013)
40. Science and Technology Department of Liaoning Province, Application of optical fiber sensing technology in condition monitoring of train body and railway facilities [EB/OL]. http://www.lninfo.gov.cn/kjzx/show.phpitemid=423266, 2010, pp. 11–12
41. I. Aydin, M. Karakose, E. Akin, Anomaly detection using a modified kernel-based tracking in the pantograph–catenary system. Expert Syst. Appl. **42**(2), 938–948 (2015)

42. E. Karakose, M.T. Gencoglu, M. Karakose, et al., A new arc detection method based on fuzzy logic using S-transform for pantograph–catenary systems. J. Intell. Manuf., 1–18 (2015)
43. M. Bocciolone, G. Bucca, A. Collina, L. Comolli, An approach to monitor railway pantograph-catenary interaction with fiber optic sensors. Proceedings of the SPIE-The International Society for Optical Engineering, 2010, 7653:76533Q (4 pp.)
44. G. Bucca, A. Collina, A procedure for the wear prediction of collector strip and contact wire in pantograph–catenary system. Wear **266**(1), 46–59 (2009)
45. E.A. Mohamed, A.Y. Abdelaziz, A.S. Mostafa, A neural network-based scheme for fault diagnosis of power transformers. Electr. Power Syst. Res. **75**(1), 29–39 (2005)
46. S. Hedayati Kia, H. Henao, G.A. Capolino, Mechanical health assessment of a railway traction system. The 14th IEEE Mediterranean Electrotechnical Conference, Ajaccio, France, 2008, pp. 453–458
47. D. Grillo, C. Landi), M. Luiso, N. Pasquino, An on-board monitoring system for electrical railway traction systems. IMTC 2006 – Instrumentation and Measurement Technology Conference, Sorrento, Italy, 24–27 April 2006, pp. 2306–2311
48. H. Kamijo, H. Hata, H. Fujimoto, A. Inoue, K. Nagashima, K. Ikeda, A. Iwakuma, K. Funaki, Y. Sanuki, A. Tomioka, H. Yamada, K. Uwamori, S. Yoshida, Tests of superconducting traction transformer for railway rolling stock. IEEE Trans. Appl. Supercond. **17**(2), 1927–1930 (2007)
49. M. Mermet-Guyennet, M. Piton, Railway traction reliability. The 6th International Conference on Integrated Power Electronics Systems (CIPS), Germany, 16–18 March 2010, pp. 1–6
50. Z. Yuan, Q. Yong, X. Cheng, P. Xuemiao, X. Zongyi, Relationship analysis between contact surface unevenness and bow mesh contact force based on improved NARX Neural Network. China. Railw. Sci. **33**(3), 84–91 (2012)
51. L. Kai, F. Yao-ping, L. Ying-long, Design and implementation of bow network monitoring system. Comput. Meas. Contr. **14**(5), 600–602 (2006)
52. P. Wei, D. He, M. Jian, X. Yang, Study on condition monitoring and fault diagnosis of pantograph. J. Guangxi Univ. Nat. Sci. Ed. **36**(5), 718–722 (2011)
53. Z. Yanyan, Metro Vehicle Traction System Fault Diagnosis Technology and System. Beijing Jiaotong University, Beijing, 2009, pp. 11–25
54. C. Xiao-xuan, L. Yong-yi, S. Yong-teng, Fault diagnosis of locomotive traction motor based on neural network information fusion. Comput. Meas. Contr. **15**(5), 563–565, 573 (2007)
55. L Ling, Research on Fault Diagnosis of Traction Converter. Southwest Jiaotong University, Chengdu 2010, pp. 5–35
56. W Yi, Locomotive Traction Converter Fault Diagnosis Based on Data Mining. Southwest Jiaotong University, Chengdu, 2005, pp. 15–26
57. G. Niu, Y. Zhao, M. Defoort, et al., Fault diagnosis of locomotive electro-pneumatic brake through uncertain bond graph modeling and robust online monitoring. Mech. Syst. Signal Process. **50**, 676–691 (2015)
58. X. Zhuan, X. Xia, Fault-tolerant control of heavy-haul trains. Veh. Syst. Dyn. Int. J. Veh. Mech. Mob. **48**(6), 705–735 (2010)
59. L. Wan-xin, Z. Yang, R.-w. Lin, et al., Fault diagnosis and safety measures for brake system of harmony No. EMU. Railw. Locomot. Car. **31**(5), 39–42 (2011)
60. J. Liu, Y.F. Li, E. Zio, A SVM framework for fault detection of the braking system in a high speed train. Mech. Syst. Signal Process. **87**, 401–409 (2017)
61. C. Guoqiang, J. Yang, Z. Liming, J. Limin, Reliability analysis of passenger brake system of metro vehicles based on FTA. Int. Conf. Model. Simul. Optim. Beijing, 323–328 (2009)
62. D. Jianbo, C. Hangfeng, The design of fault diagnosis system of 120-type freight train brake. 2011 International Conference on Electric Information and Control Engineering (ICEICE), Wuhan, 2011, pp. 5463–5465
63. X. Wang, M. Wu, Research on unit brake reliability of urban rail transportation vehicles. Urban. Mass. Transport. **11**, 52–53 (2010)

64. L. Xinghua, On-line diagnosis of synchronous braking system for heavy haul combined train based on MAS. Changsha. Cent. South Univ., 2–15 (2009)
65. M. Wu, X.-y. Wang, T. Chun, Reliability of relay valve used in braking system of rail transportation vehicles. J. SouthWest JiaoTong Univ. **44**(3), 365–369 (2009)
66. El K. Kadri, A. Berthon, Simulation of a dual hybrid generator for heavy vehicle application. The 32nd Annual Conference on IEEE Industrial Electronics (IECON), Paris, France, 2006, pp. 2642–2647
67. W. Canpei, S. Xianhai, H. Shunhao, Web based remote monitoring and control system for emergency power supply of highspeed rail train. International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), Changchun, China, 2011, pp. 694–697
68. M. Berger, C. Lavertu, I. Kocar, et al., Proposal of a time-domain platform for short-circuit protection analysis in rapid transportation train DC auxiliary systems. IEEE Trans. Ind. Appl. **52** (6), 5295–5304 (2016)
69. J.D. Wu, S.Y. Liao, Fault diagnosis of an automotive air-conditioner blower using noise emission signal. Expert Syst. Appl. **37**(2), 1438–1445 (2010)
70. Y. Dichen, J. Limin, Q. Yong, P. Wei, Y. Yang, Fault diagnosis of auxiliary inverter system of urban rail transportation based on wavelet BP neural network. Chin. J. Constr. Mach. **11**(6), 542–546 (2013)
71. L. Gang, Fuzzy logic for battery state of technology. 1995 China Intelligent automation conference and intelligent automation professional committee proceedings proceedings (second). Tianjin, China, 1995, pp. 936–941
72. G. Yajun, Fault analysis of three-phase ac overvoltage monitoring for auxiliary inverter of AC02 train. Science and Technology Conference of Shanghai Metro Operation Co., Ltd., Shanghai, 2006, pp. 261–264
73. Y. Qiang, C. Ang-long, W. Li, T. Sheng-gui, Starting Failure Analysis and Countermeasure of Auxiliary Inverter for DCO l Type Electric Train. Shanghai Subway Operation Co., 2006, pp. 191–193
74. C. Xiaoliang, Analysis of emergency lighting failures of Guangzhou metro line 5 train. Urban. Mass. Transport. **14**(7), 70–71, 75 (2011)
75. C. Huanxin, Z. Jun, W. Shanzhe, Fault diagnosis expert system for passenger car air conditioning units (2002) 1:69–72
76. M Chuan. Fault Feature Extraction and Application of Rolling Bearing. Dalian University of Technology, 2009, pp. 2–5
77. H. Weiguo, Condition Monitoring and Fault Diagnosis of Rotating Machinery Based on Feature Extraction and Expression of Vibration Signals. University of Science and Technology of China, Hefei, 2010, pp. 4–10
78. S. Wentao, Roller Bearing Surface Damage Fault Feature Extraction and Diagnostic Methods. Jinan, Shandong University, (2011), pp. 10–30
79. T. Deyao, *Generalized Resonance, Resonance Demodulation Fault Diagnosis and Safety Engineering: Railway* (China Railway Publishing House, Beijing, 2006), pp. 8–105
80. L. Yongbin, *Study on Condition Monitoring and Diagnosis of Rolling Bearing Based on Nonlinear Signal Analysis* (University of Science and Technology of China, Hefei, 2011), pp. 3–15
81. S. He, J. Lin, Zhang Bing. high speed train shaft vibrating harvester. Eng. Test. **50**(1), 54–57 (2010)
82. S. Wanfeng, Z. Shengtang, H. Jie, Fault diagnosis of train bearing based on high order cumulant adaptive algorithm. J. Vib. Eng. **19**(2), 234–237 (2006)
83. P. He, L. Pan, S. Huiqi, S. Nanxiang, Design of acoustic signal acquisition system for train bearace fault. Autom. Instrum. **10**, 8–11 (2011)
84. D. Fuyan, Research on general scheme of locomotive bearing monitoring and diagnosis. Diesel. Locomot. **8**, 15–17 (2006)
85. N.K. Nabiyev, Diagnostics of axle boxes bearings based on identification measuring method. Trans. Univ. Karaganda. State. Tech. Univ. **1**, 77–79 (2010)

86. J. Yang, C. Guoqiang, Y. Dechen, H. Qiang, L. Jie, Fault diagnosis method for the rolling bearing of railway vehicle based on wavelet packet transform and BP neural network. China Railw. Sci. **31**(6), 68–73 (2010)
87. W. He, X. Zhou, Application of the wavelet-SOFM network in roll bearing defect diagnosis. 2009 WRI Global Congress on Intelligent Systems (GCIS 2009), 2009, pp. 8–12
88. J. Yang, Z. Mingyuan, Robot bearing diagnosis system based on vehicle bus and Laplace wavelet. J. China. Railw. Soc. **33**(8), 23–27 (2011)
89. M.-g. Yu, Z. Ji-ye, Z. Wei-hua, Analysis of crosswind safety of high-speed trains based on reliability. J. Vib. Shock. **32**(20), 90–96 (2013)
90. Li Chao, Wang Ying. Hierarchical coupling analysis of equipment system safety based on reliability and IPPM. China. Saf. Sci. J., 2013, 8 (08)
91. Z.-m. Yu, Z. Hong-lun, Life-cycle safety management system of rolling stock structure based on reliability. China. Railw. Sci. **26**(6), 1–5 (2005)
92. S. Hongsheng, C. Yulong, Z. Youpeng, Reliability evaluation of on-board subsystem of CTCS-3 train control system based on bayesian network. China. Railw. Sci. **05**, 96–104 (2014)

Chapter 2
Safety Region Based Active Safety Methods

2.1 Safety Region Analysis Model

The safety region based state identification theory and method was discussed in this section. A boundary estimation algorithm was proposed based on formalized description of some basic definitions like safety region estimation, multi-domain division.

2.1.1 Basic Concepts

Safety region is a quantitative model that studies the system safety and stability problem from the regional perspective. Relative relationship between safety region boundary and system operating point can provide quantitative safety margin and optimal control information under various conditions. For a certain study object, safety region is proposed to describe safe operation of studied object in the space decided by safety related variables.

Broadly defined, if the object is working in safety region, it is considered to be healthy, or at least at a low risk level. Otherwise it is considered to be in two states, except safety region, as described below. One is that when the object was failed, and the failure would lead to unsafe event (e.g. accident), at this time it is considered to be at high risk level. The other one is the object was abnormal or unhealthy, but would not cause any adverse events directly. At this point, the object was within moderate risk. We discussed a narrow safety region definition here, only the state that would cause accident was divided into unsafe region.

$$g(X) = g(x_1, x_2, \ldots, x_n) = c \tag{2.1.1}$$

Y. Qin, L. Jia, *Active Safety Methodologies of Rail Transportation*, Advances in High-speed Rail Technology, https://doi.org/10.1007/978-981-13-2260-0_2

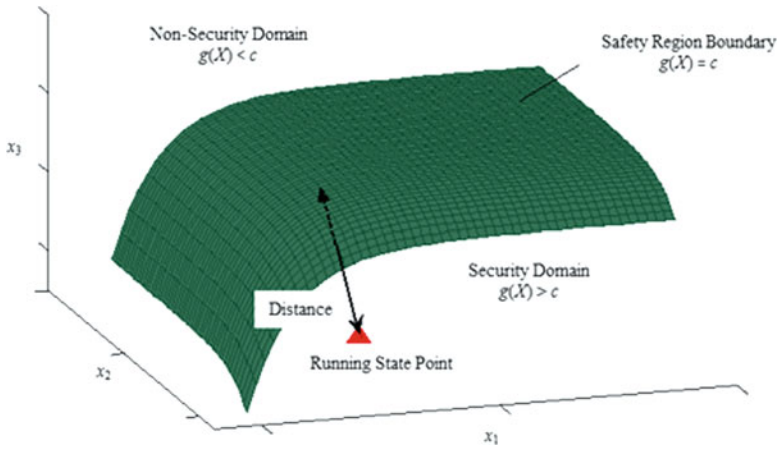


Fig. 2.1 Visualize description of safety region

Mathematically, safety region boundary can be described by a safety region boundary formula, where x_1, x_2,. . ., x_n denote safety related variable, n is variable numbers, $g(X)$ is output variable characterizing safety state of running, and c is a constant of safety threshold. If $n = 3$, safety region boundary is curved surface in three-dimensional space; if $n = 2$, safety region boundary is a curve in two-dimensional space; if $n = 1$, the safety region boundary formula will be univariate threshold. The region of $g(X) < c$ (assume the region below curved surface determined by a safety region boundary formula) is defined as safety region. Conversely, the region of $g(X) > c$ is defined as non-safety region. According to whether the object's current operating state point lies within safety region. The object's safety state can be judged. If the point is within safety region, the object is safe. If the point is outside the safety region boundary, there may be risks.

In addition, as shown in Fig. 2.1, the safety degree can be described by the distance between the object's real-time operating state point and the safety region boundary. Furthermore, quantitative safety margin can be given to take the optimal prevention and control measures and ensure further running safety.

In view of complicated rail vehicles equipped with a set of various types of equipment, safety region research can be divided into three layers in accordance with the structure of objects in Fig. 2.2. The top is system layer namely the whole rail vehicles. Then unit layer includes several larger units of rail vehicles, such as running system, power system and brake system and so on. The bottom is device layer, includes key equipment of unit layer, wheel sets in running system, for example. Different objects of various layer can be studied separately, meanwhile, safety region estimation about the top layer can combine the lower's results.

For some study object safety assessment of service state based on safety region includes two steps as follows.

1. Determine safety related variables, estimate safety region boundary and divide safety region and non-safety region of service state.

2. Judge whether the object's current operating state point lies within safety region based on service state data and safety region boundary formula. If yes, then the distance between the object's real-time operating state point and the safety region boundary should be calculated to gain safety margin. If not, alarm information should be given. Finally, give quantitative evaluation results.

The safety region of railway systems is a domain for evaluating the system safety state within a space as determined by system's safety-related variables (such as running speed, track irregularity, safe interval, pantograph catenary voltage/current, cross wind, etc.). As shown in Fig. 2.3, if the system is operating in the safety region (as shown in green), the system operating state is considered to be safe, otherwise it is considered to be unsafe. If the railway system performs a transportation from the safety region to the warning region (as shown in yellow) due to the equipment failure or other external disturbances, i.e., system operation at risk, prevention and control measures should be adopted timely to change the system state back to safety from a risk (as shown by dashed arrow). Without appropriate prevention and control

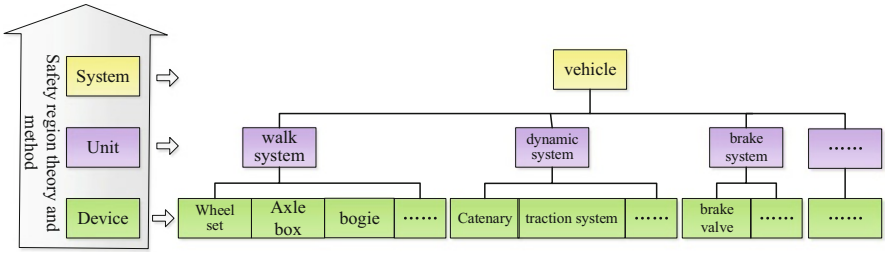


Fig. 2.2 Safety region objects of railway vehicles

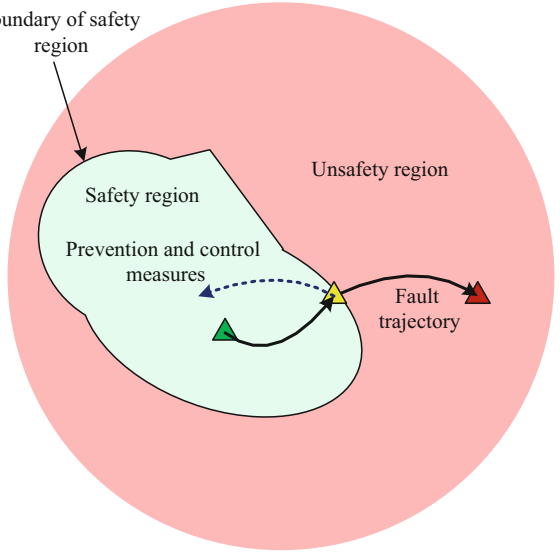


Fig. 2.3 Schematic diagram of safety region

measures when system is running in warning region, the system state will deteriorate further and will be evolved to the emergency region (the area shown in red), which eventually leads to the accident (as shown by failure trace). In other words, the behavior of railway system operation is changeable with the system state, and the evolved trajectories are different due to the different interaction imposed to the system. For the railway system as a complex dynamic autonomous system, a direct mapping exists between the safety region and the stability region of its safety-related model. Based on this, the fundamental conception and geometric interpretation of safety region are proposed as follows.

Now, consider the research object as an autonomous nonlinear dynamical system,

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \dot{X} = f[X, t] \tag{2.1.2}$$

where $X = (x_1, x_2, \ldots, x_n) \in R^n$ is safety-related state vector; $f(X) = (f_1(X, t), f_2(X, t), \ldots, f_n(X, t))^T$ and f is locally Lipschitz on R^n.

So, SR of RTSOSA is defined as follows.

Definition 2.1 (Safety Region) Let set $A \subset R^n$, with the initial condition X_0, $D_s(A) = \cup[D_s(\varepsilon, A): \varepsilon \in R^+]$ is safety region of (1) on set A if there exists a neighborhood of A $D_s(\varepsilon, A) \in R^n$ satisfy $\rho[X(t; X_0), A] < \varepsilon$, $t > 0$ (ρ is the distance between $X(t; X_0)$ and A) when $X_0 \in D_s(\varepsilon, A)$ for $\forall \varepsilon \in \mathrm{R}^+$.

A more intuitive explanation of safety region: in the railway system, a point X in safety-related variable space starts movement from an initial point X_0, and if X gradually approaches the equilibrium point X_e over time and each point of its trajectory makes the system operating safely at the same time, then X is called safety point. The body of safety points regarded as the safety region. The geometric interpretation of a two-dimensional safety region is shown in Fig. 2.4.

The relationship between region division and state classification is discussed below. As mentioned before, safety region and unsafety region correspond to the normal and fault states of device respectively. Meanwhile, unsafety region can be divided into several sub regions and each one represents one fault type (shown in Fig. 2.5). Take race fault as an example, unsafety regions are the ball fault, inner race fault, outer race fault and so on (Fig. 2.6).

For two class identification, the decision function is

$$f(X) = \mathrm{sign}[\mathrm{Bound}(X)] \tag{2.1.3}$$

Where: $X = (x_1, x_2, \cdots, x_n) \in R^n$ is the feature, n is the dimension of feature variable, and $x_1, x_2, \cdots, x_n$ are the values of feature in every dimension; Bound(X) is the boundary between two regions. Bound(X) = 0 is the boundary function to divide two regions.

As for multi-class identification, a rule function, named multi-class discrimination function should be determined to locate different states into their region. For a state point X in certain feature space, the classification function is

$$\text{Class}(X) = \underset{i=1,2,\ldots k}{\text{multisign}}\left[\text{Bound}_i(X)\right] = \{1, 2, \ldots, m\} \tag{2.1.4}$$

where, Class(X) is the state discriminant decision function, $X = (x_1, x_2, \cdots, x_n) \in R^n$ is the variable in feature space, n is the dimension, and $x_1, x_2, \cdots, x_n$ are the values of feature in every dimension; Function multisign() is the multi-value sign function. Its value ranges in $\{1, 2, \ldots, m\}$, and represent Class 1, Class 2, …, Class m, where m is

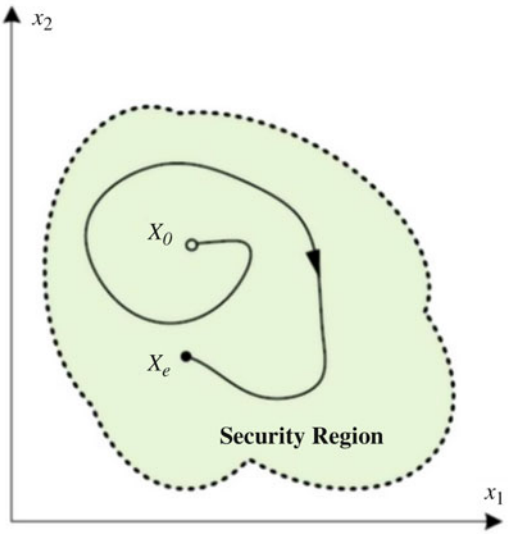


Fig. 2.4 Geometric interpretation of safety region

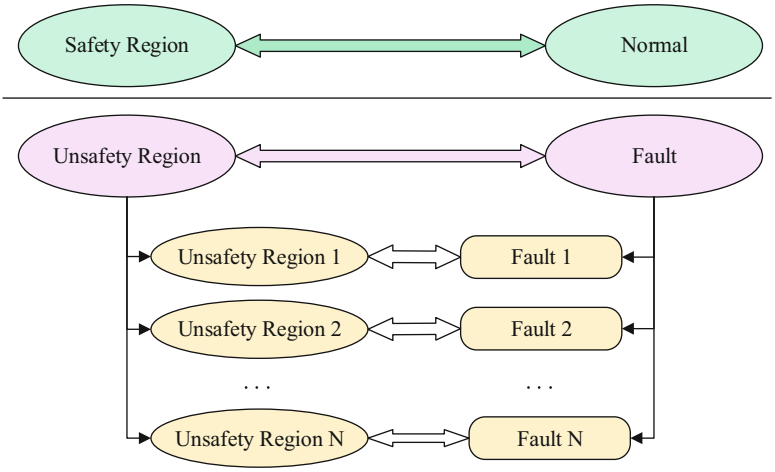


Fig. 2.5 Correspondence between the region and the state

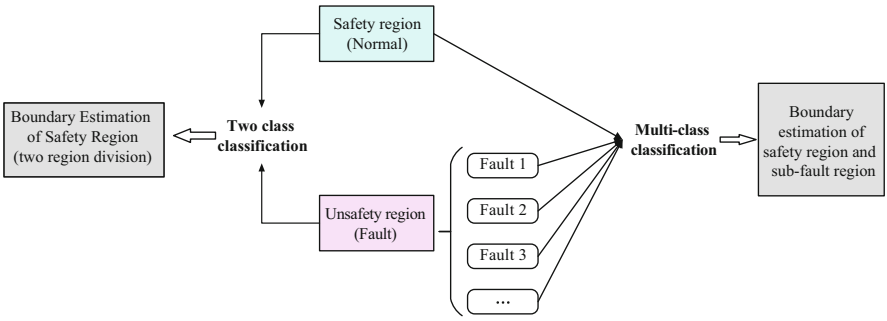


Fig. 2.6 State identifications under different requirements

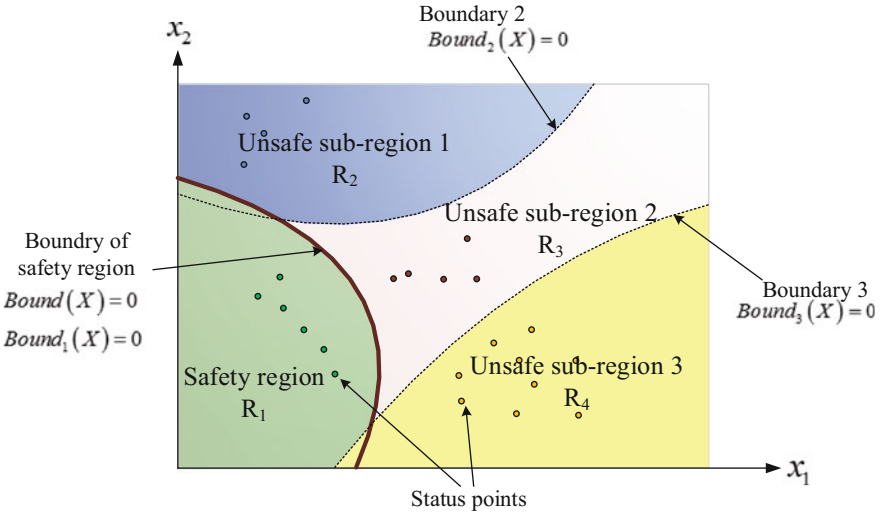


Fig. 2.7 Schematic diagram of multi-state identification based on regional division

the number of regions in state space; $\text{Bound}_i(X)$ is the ith boundary function, $i = 1, 2, \ldots, k$, k is the number of boundaries.

Figure 2.7 shows the region division of two dimensional feature variables of x_1 and x_2. Bolded line is the boundary between safety and unsafety regions, where $\text{Bound}(X) = 0$; Dotted borders are the boundaries among sub unsafety regions, where $\text{Bound}_1(X) = 0$, $\text{Bound}_2(X) = 0$, $\text{Bound}_3(X) = 0$. The space represent by x_1 and x_2 has been divided into four regions, R_1, R_2, R_3, R_4, which are safety region, unsafety region 1, unsafety region 2, and unsafety region 3 corresponding to normal, fault 1, fault 2 and fault 3 states.

2.1.2 Processing Procedures

For a specific object, it generally takes the following five steps to complete the state identification based on safety region theory.

Step 1: For a specific object, analyze its characteristics based on its mechanism and working condition. For example, whether it is convenient to modeling or whether it can obtain the monitoring data.
Step 2: Study the express of the state characteristic, fully consider the working environment and object characteristics, and chose the feature which can sensitively reflect the change of the working state.
Step 3: Extract features and calculate their values based on the chosen state variables. Then the point in feature space has been obtained. Noise and the environment uncertainty should be taken into consideration if necessary.
Step 4: Estimate the boundary of safety region, and the feature space would be divided into several sub regions.
Step 5: Corresponds the state point to a certain region based on the completed boundary evaluation.
Step 6: Safety margin will be calculated later if the state point locates in safety region, otherwise, find out which unsafety region it belongs to.
Step 7: Prognose the remaining useful life and future operating information of the target device or system when its state point lying in safety region.

A brief demonstration of safety margin is provided in Fig. 2.8. There are two state variables, which mean the state space is in two dimensions. Assuming boundary between safety and unsafety region has already been obtained and represented by the bold line in Fig. 2.8, and the green dot is the state point in this 2D space. Calculate the smallest distance between the state point and safety region boundary, which is called safety margin. For example, a state point is (26.86, 147.01) in the given space, and its smallest Euclidean distance to safety region boundary is 9.358. Then the safety margin of state point (26.86, 147.01) is 9.358. Furthermore, if the state point changed to the red dot of (25.60,87.27), the margin is 2.366 after evaluation, which means state of the target device is getting worse.

Furthermore, as the description of the Step 6, the quantitative state identification result can be obtained by calculating the safety margin if the state point is in the safety region. For the key equipment of the train, especially the electro mechanic equipment, the change or transfer of the working state point is very slow. Therefore, the safety margin can be described by the minimum distance between the state point and the boundary of the safety region. When the safety margin is reduced and the state point is approaching the boundary of the safety region, it is necessary to

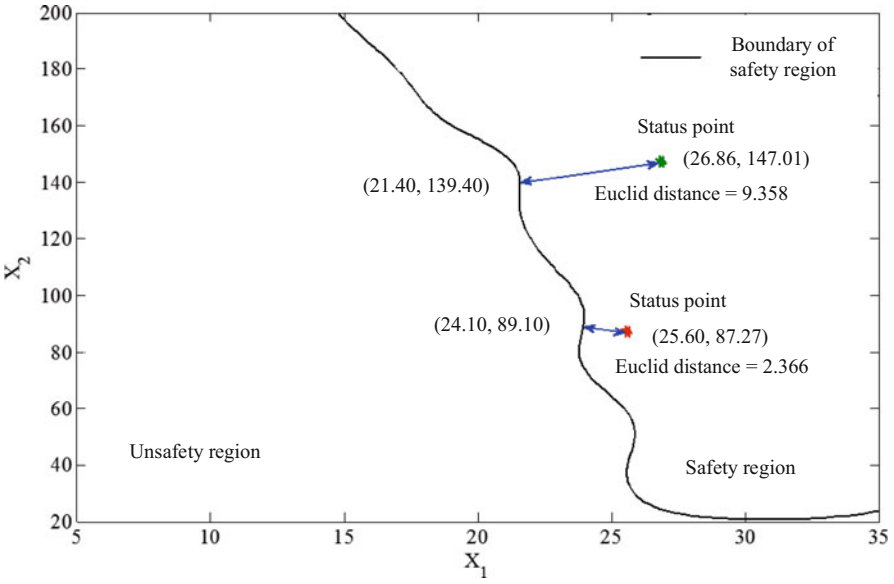


Fig. 2.8 Safety margin calculation

warning for reminding the relevant personnel to take appropriate preventive and maintenance measures, ensuring safe working state of the equipment.

The basic procedure of state identification based on safety region theory is shown is Fig. 2.9.

2.1.3 Computation Methods

The methodology of SRE of RTSOSA is proposed which includes two different research methods.

1. The method based on stability region estimation of safety-related state-space model

The safety-related state-space model of railway system need to be established; and the different parameter matrix of the state-space model used to represent different state of the system which may run either properly or in multifarious fault conditions. The different boundary of SR (also known as the different SR_s) can be obtained by estimating stability region of the state-space model with different parameter matrix. This approach is indirect as object model should be built first; therefore, it is called the indirect method.

2. The method based on intelligent analysis of safety-related data

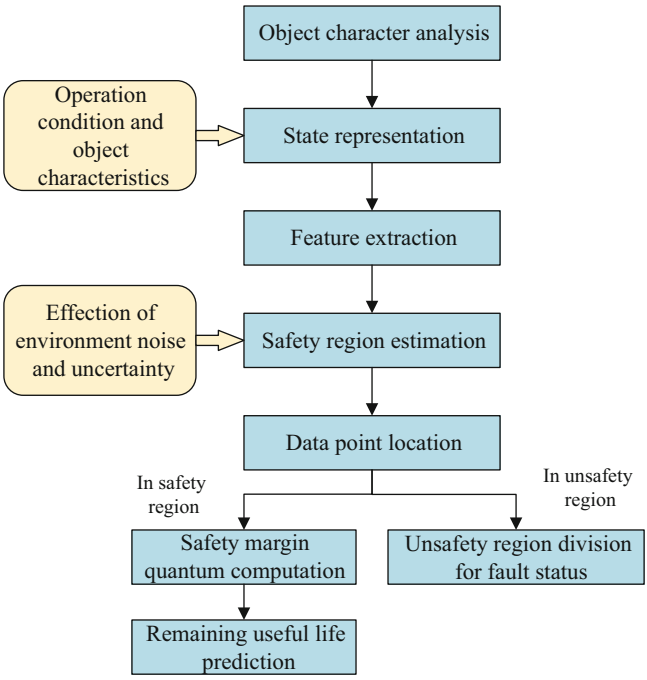


Fig. 2.9 Procedure of state identification based on safety region theory

As the system evolves from the safety state to the risk state, state information and data information of the whole process are required to collect when safety-related variables take on diverse values. According to the whole process information, the corresponding data of safety-related variables can be divided into two categories which can respectively indicate safe or unsafe state of system using some intelligent classifiers. The separating surface that achieves the best classification is the boundary of SR to estimate. In this case, no mathematical model of the system needs to be created and estimation of the SR can be achieved directly, and hence this method is referred to as the direct method.

Details of the implementation of the two methods are explained as below.

1. Implementation of the Indirect Method

    1. Implementation Steps

As described above, to apply the indirect method, four issues should be taken into consideration, which are to select safety-related variables, to build the safety-related state-space model, to identify different parameter matrixes of the model under various conditions and to estimate the stability region.

Step 1: Select safety-related variables

With reference to the standards of safety and comfort and other relevant literatures of domestic and international railway system, and full consideration to feasibility of establishing the state-space model and the actual situation of railway system operation, select accessible and representative safety-related variables using expert experience, statistical data, correlation analysis and other methods.

Step 2: Establish the safety-related state-space model

Reference to a wealth of existing dynamic models of railway system and safety-related testing data, with appropriate simplification, reasoning and conversion methods and advanced state-space model identification techniques, establish the state-space model $dX/dt = f(X)$ in the safety-related variables space, where $X = (x_1, x_2, x_3 \ldots x_n)$ is an n-dimensional vector and x_i is a certain safety-related variable.

Step 3: Identify parameter matrixes

Base on the state-space model established, refer to the normal values of the basic parameters in the existing railway system models, and use of expectation-maximization (EM) algorithm, hierarchical identification and other methods to determine and identify the parameter matrix P_1 for intact system and P_2, P_3, …, P_p for diversified system troubles.

Step 4: Estimate stability region

Utilize the existing stability region estimation methods of complex system to achieve the stability regions SR_1, SR_2, $SR_3 \ldots SR_p$ which respectively correspond to parameter matrix P_1, P_2, $P_3 \ldots$, P_p. These estimated stability regions are the required safety regions for multiple system states.

2. Stability Region Estimation

Now, many computational methods including Lyapunov function method [1–3], flow estimation method [4], Monte Carlo method, inverse iteration method [5] and energy function method, as well as some optimization methods like linear matrix inequality(LMI) [6] and genetic algorithm [7] are applied to estimate stability region of complex objects. The following is a brief introduction of several popular approaches.

(a) Lyapunov function method

This method is primarily to construct a suitable Lyapunov function. If the Lyapunov function constructed on a stable fixed point is convex at some other points and the trajectories which start from these points converge to the stable fixed point, the convex boundary of the Lyapunov function is the boundary of stability region. The method is applicable to continuous or discrete dynamical systems in any dimension, however, no general means or rules to construct Lyapunov function can be followed. Moreover, the results obtained are mostly conservative.

(b) Flow estimation method

The boundaries of the stability region are contained in the stable manifold of an unstable set on the border. For a mapping system, the stable flow of the saddle fixed point (or unstable set) can be obtained by numerical method according to the Center Manifold Theorem, then boundary of stability region can be gained. This approach can be convenient to get all stability regions of mapping system, but for many actual dynamics systems, it is difficult to find an efficient way to solve the stable manifold of the system.

(c) Monte Carlo method

For getting the stability region, a defined work domain in the variables space is divide into a finite number of initial points (scan pixels), and use numerical method to scan these initial points and mark their motion state trajectory. This method is practicable for systems in any dimension and the calculation error can be adjusted by changing the number of the initial points. During the time-domain simulation, very huge workload and the low efficiency are the deficiencies of this approach. In addition, constant test is needed to determine the work domain.

(d) Inverse iteration method

The idea is to determine a small stability region in advance in system's state space as a target set. Then compute the reachable set with inverse time searching namely the stability region. This method is generally used for evolution system, and it can get the final outcome of the evolution. Nevertheless, the effective definition of the stability region can be ensured only on the basis that the inverse mapping of system exists and can be effectively solved.

2. Implementation of the Direct Method

1. Implementation Steps

As mentioned above, implementation of the direct method needs to address four major issues: the selection of safety-related variables, the establishment of simulation models, the collection and analysis of the whole process data, and intelligent classification of dataset.

Step 1: Select safety-related variables

This step is basically the same with the first step in the indirect method.

Step 2: Build the simulation models

Under full consideration of actual operating conditions of railway system, using advanced multi-body dynamics simulation software (Simpack or ADAMS/Rail), identify and summarize common risks and failures of railway system, and build multiple simulation models of the system or some subsystems in the case of system intact or system failure.

Step 3: Collect and process the whole process data

Based on numerical simulation methods, determine the input data corresponding to safety-related variables, input the input data to the multiple simulation models, and collect the simulation models' output data which can characterize the whole process of system operating state from safety to risk (therefore, the input data and output data called the whole process data), and then process the source data using data cleaning, data transformation, data reduction and other methods.

Step 4: Classify the data intelligently

Based on multiple sets of the whole process data, train intelligent classifiers using related intelligent optimization algorithms to classify the data of each set into two classes which are labeled as 'safe' and 'unsafe', and obtain the formulas $SR_i(X) = 0$, $i = 1 \sim p$ of the best separation surfaces. The formulas are the boundaries of SRs under various system states.

2. Intelligent Classification Based on Support Vector Machine

Support Vector Machine (SVM) is a creative machine learning method based on the foundation of the statistical learning theory and the optimization theory, which is proposed by Vapnik and his copartners in 1995 [8], and the basic idea is to transfer a classification problem in a low dimensional input space to a high dimensional space by using nonlinear transformation defined in the kernel function, and find the generalized optimal separating surfaces in this space. The training algorithms based on SVM have already been used in many fields aiming to find the decision boundary that separates the dataset into a discrete predefined number of classes in a fashion consistent with the training examples.

With the rigorous theoretical foundation and structural risk minimization principle, SVM is particularly appealing in small-sample and nonlinear cases of classification [9]. And the excellent learning and generalization performance of SVM exceeds the neural network and some other artificial intelligence methods in high dimensional dataset classification.

Furthermore, SVM can be applied to fulfill the data classification better by combining advantages of other intelligent optimization theory and algorithms, including the fuzzy theory, genetic algorithm, the rough set theory, the hidden Markov model and the DS evidence theory, etc. Respectively, the combination with the fuzzy theory can effectively solve the noise problem in the training sample [10]; the combination with genetic algorithms can solve the parameter selection problem within the SVM and its kernel function [11]; the combination with rough set theory can improve the capacity to deal with imprecise and incomplete data [12]; the combination with Hidden Markov model allows improved robustness for data classification and pattern recognition; and the SVM combined with DS evidence theory has advantages in terms of multi-source information fusion [13] (Fig. 2.10).

3. Framework of SRE Method

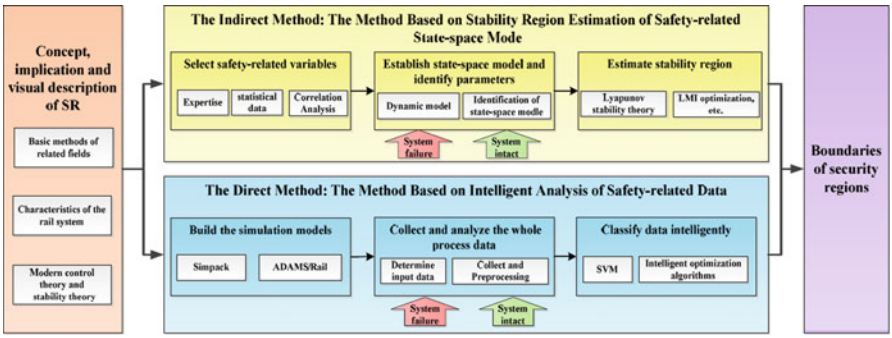


Fig. 2.10 The framework of SRE method

2.2 Safety Region Based Accident-Causing Model

Accident-causing theory mainly studies why accident happens and the mechanism of its process [14]. In order to prevent future accidents, the relationship of the causation found out in each part of procedure is established by disclosing the interaction of the components in the system. Traditional accident-causing theory, like Domino theory proposed by Heinrich in the 1940s [15], takes single element such as human, equipment or other causes separately into consideration as a chain or sequence of events [16], which explains well accidents caused by physical components and relatively simple systems [17]. Whilst systems we build today are increasingly complex and linear model is no longer adequate to capture the interactions and coupling within the system; thus it requires us to analyze the accident causation systematically as a whole. To catch up with the complexity, the accident theories developed via previous linear causation theories to present-day systematic theories, such as: system theory, perturbation accident-causing theory, energy transfer theory and information theory [18].

The system approach addresses the notion that safety is an emergent property, which arises from non-linear interactions between multiple components across complex system and the relationship of behaviors implicated in operation [19]. In systemic safety models, the accident process is described as a complex and interconnected network of events to model the dynamics of complex systems [20]. Rasmussen's hierarchical framework [21] and Leveson's system theoretic accident modeling and processes [17] are two notable approaches. Even though these accident models considered the joint effect of multi-factors in an accident with their dynamic interactions, the descriptions of them (human, equipment, environment and etc.) are mainly qualitative, and the outcome of those interactions of system components are described respectively without an uniform expression. On the one hand, these models are sufficient to help us learn from accidents that have already happened, and thereby preventing hazards from the similar kind. On the

other, as they hardly reveal the course of the outcome of system change, they are inadequate to guide real-time emergency response to prevent accident when the system is disturbed and prone to accident. This is mainly because the consideration of system state as a whole is lacked in these models. And the challenges we meet today to achieve safety is going beyond accident analysis to the extent of resilience engineering [22]. Hereby, the accident analysis should also be able to implement in the real-time field work to prevent accident not only after but during its process, by enhancing its resilience against disturbance.

To achieve this goal, the conception of safety region, which depicts the safe state affected by different factors in a unified way, is introduced with the combination of perturbation accident-causing theory to establish the perturbation-safety region (P-SR) accident-causing theory. In this theory, in addition to analysis causality systemically, the safe state of the system after perturbation is described quantitatively with the changing course of it in P-SR model. And then by exploiting the safe state as risk assessment, the monitoring and evaluation of system safe state as well as the corresponding control measures are brought into the model to enable its practicability in safety management of production activities.

2.2.1 Concepts and Procedures

Inevitable as perturbation is in production activities, Amalberti [23] argued that these 'noises' (e.g. equipment malfunction or human errors) jeopardize operation safety; conceptually they should be symmetrically assessed and then calculate the associated risks. With new safety methods and perspectives that keep up with the continuously increasing complexity of industry, accident models aiming at explaining events and guide risk assessment need to match this complexity [24]. Specific to the complex system, the P-SR model promotes a quantitative description of the safe state and risk boundary of the system, which will better instruct safety monitoring and relative control measures. The concept, perspectives and processes are defined and described in this section.

2.2.1.1 Definition of Safety Region

Safety region analysis have been applied to monitor the safety and stability of power system [25]. The concept of region quantitatively describes the safety boundary of a system so that it could dynamically and consecutively monitor the system state with its changing process, and evaluate the safe state to provide warning information.

On the basis of the object studied in accident models, the safety region is defined as a changing space to describe the multifactor. Let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of

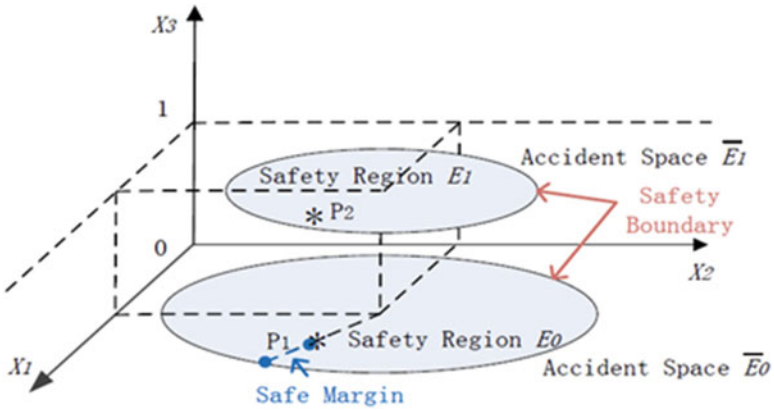


Fig. 2.11 The change of system safety region

characteristic variables representing the characteristic state of the system, in which n is the number of the critical subsystem. The characteristic variables, derived from multifactor of human, equipment, environment, management or other factors, contain both discrete variables and continuous variables. Define space E as safety region: within the boundary of E is safe space; otherwise is accident space $\bar{E}$. The boundary is determined by the threshold of system safe state, i.e. the accepted risk level that can ensure system safety.

The safety region is determined as a n dimension space by the number of the characteristic variables n, in which the lower dimension spatial scope may vary with high dimension variables. Figure 2.11 gives an example of a 3-dimension safety region composed of $X = \{x_1, x_2, \ldots, x_n\}$, in which x_3 is a discrete variable, representing two types of system state at this dimension: when $x_3 = 0$, the safety region is E_0; when $x_3 = 1$, it changes to E_1.

The boundary of the safety region is only determined specifically to a certain system. Usually, the state of the system located in safety region is called the balanced state. If the character point falls in the safe space, then the system is confirmed to be safe, with the distance between the point and boundary, called safe margin, to assess the safety level of the system. Otherwise, the point falls in the accident space when it breaks through the safety boundary, indicating that the safe state reaches an unacceptable level and then causes the accident.

In production activities, the system state continually deviates from safe space under the influence of perturbation. As it reaches a certain extent that beyond the safety boundary, the system enters the accident space. Figure 2.12 show a safety region consists of 2 dimension variables, in which represent respectively system running safely and accident taking place. Obviously, the crucial task to use safety region to denote system safety is to obtain the safety boundary, a decision function returning a safe threshold that differentiates the state of safety and accident [26].

Fig. 2.12 A schematic diagram of two-dimension safety region

2.2.1.2 Analysis of Accident-Causing Model

The P-SR accident-causing model consists of four critical parts: the risk resource part, the perturbation part, the alarm and system change part, and the accident part, shown as Fig. 2.13.

To study the nature of accidents, in the first part, the risk resource is prominently analyzed in the perspective of energy carrier, followed by the analysis of the direct cause of perturbation. The moving device, electrified equipment, and containers loaded of hazardous chemicals constitute the energy carrier in the system, which is the material basis of an accident. And the severity of the accident is related to the types, quantity, property, state, and energy storage method of the energy carrier. Normally, the system maintains safety by effectively taking control of the energy. Only when the unsafe multifactor disturbs the system will it result in failure of energy control mainly because of the unsafe state and unsafe behavior:

1. Unsafe state includes environment change and the defect of the equipment itself. Firstly, natural disasters and extreme weather, e.g. lightning, earthquake, typhoon, debris flow and blizzard, are uncontrollable stochastic factors, which will influence the equipment and energy transmission in the system by causing the perturbation to the balanced state and further the accidental release of energy. Secondly, the equipment has problems of wear, deformation, and metal fatigue due to the long time use, thereby increasing the probability of mechanical fault. And the device itself may also have design flaws. Meanwhile, with the increasing complexity of the system, the dynamic interaction of each part is more complicated that the fault of single equipment may affect the whole system. Thus, the system is vulnerable to the unsafe state.

2. Unsafe behavior mainly refers to the unsafe operation and management of human. The role people play in the system mainly includes: design personnel, operation staff, maintenance staff and management personnel. They together determine the reliability, stability and safety of a system. Yet each person is an

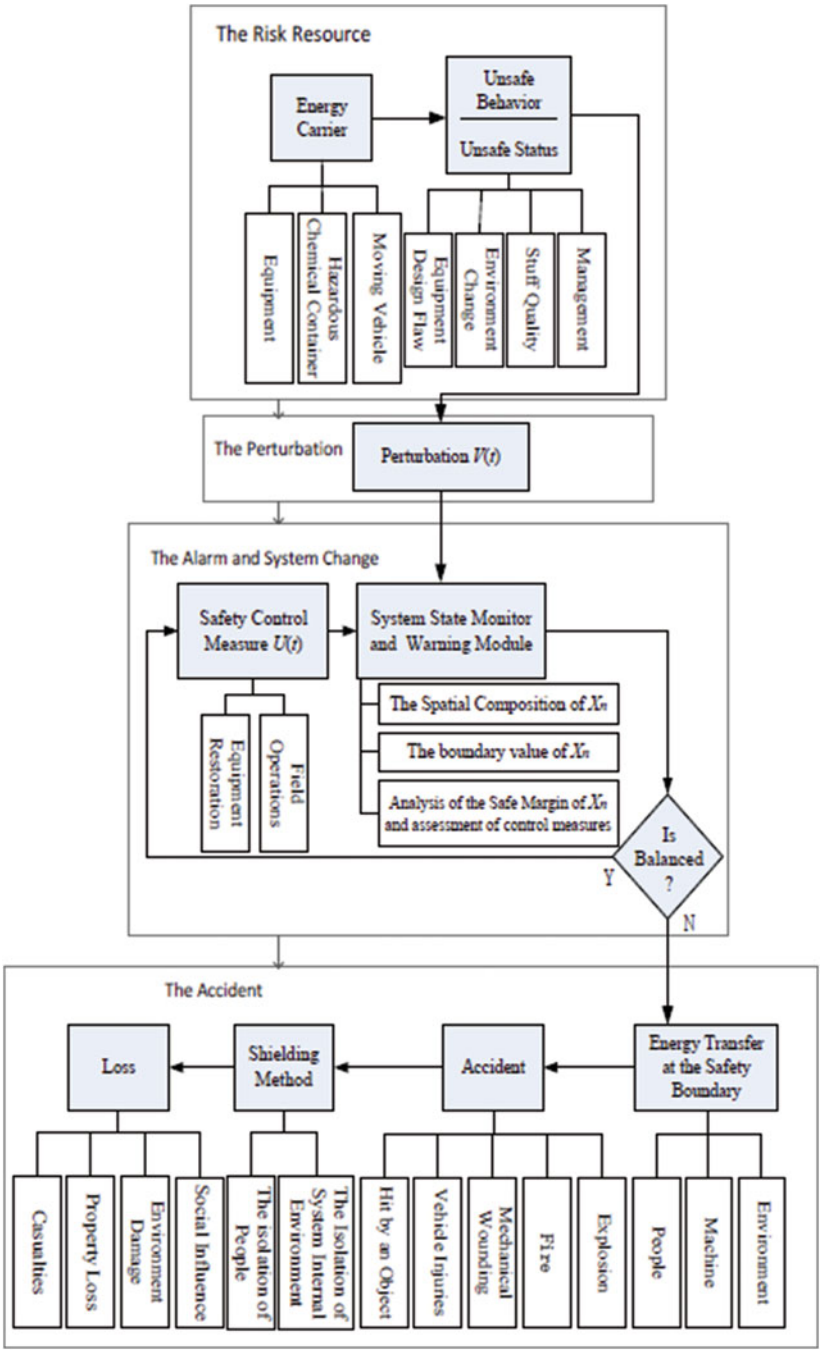


Fig. 2.13 Perturbation-safety region accident-causing model

individual with different quality, characteristic, education and etc. In the process of production, man's operation ability, management level and experience are closely related to system safety. Unsafe behaviors such as sneaking off in work, illegal operation, the decision-making mistakes, and loose management are the possible causes of an accident.

The effect of the unsafe state and behavior engenders the perturbation $V(t)$, shown in the perturbation part of the model in Fig. 2.13, which is the direct cause that deviates the safe state from balanced state. The perturbation should be further analyzed in term of the specific system and situations.

As the controllers or decision makers are highly dependent on feedbacks to take action after perturbation, the necessary information about the actual state of the process is crucial to avoid accidents [27]. The question then arises about how we express and present the actual safe state. In the next stage, the alarm and system change, the concept of safety region we introduced is the solution to this problem. At the beginning, the initial balanced state is expressed as $X(t) = \{x_1(t), x_2(t), \ldots, x_n(t) \mid x \varepsilon E\}$. After the perturbation, it changes to $X(t + 1) = AX(t) + V(t), x \varepsilon E$, in which A is the system parameter. In order to ensure the system to still be in balanced after the disturbance, the changes of state in safety region need to be monitored so that the safe margin can be calculated. Then, according to the safe margin, corresponding prevention and control measures should be taken to rebalance the system. If the adopted measures are inadequate, the system will break the safety boundary and into the accident space. Herein, a system state monitoring and warning module based on safety region is included in this part. As $X(t + 1)$ moves to the safety boundary, the safe margin decreases. Then the warning system generates alarm information; based on the alarm information, safety control measure $U(t)$ should be applied on the system, which is expressed as $X(t + 1) = AX(t) + B\ U(t) + V(t), x \varepsilon E$, ($B$ is the safety control parameter). If the system restores balance, it continues to monitor the change of safe margin and assess the control measures, so that the safety control measures module responds appropriately; if the system state broke the balanced state, it means undesired energy transfer has occurred and resulted in an accident.

Figure 2.14 depicts the rebalance or accident procedure after perturbation under the action of system state monitor and early warning module (the arrows are the state locus, and the blue lines show the safe margin at each time).

The system is in balanced state before t_1. At t_1 the safe state begins to move towards the safety boundary under the effect of perturbation $V(t)$. Then the warning module detects the reduction of the safe margin and raises alarm. Afterwards, the countermeasure $U(t)$ is applied at t_2 to slow down the decrease of safe margin. Later, the safe margin decreases slower at t_3, indicating that the system tends to restore the balanced state. Still, appropriate safety measures continue to be implemented at t_3. Finally the safe margin begins to move toward the internal safe space at t_4, which means the system state has been effectively controlled, thereby avoiding the accident.

Another trace in Fig. 2.14 shows an opposite situation where the safety measure $U(t)$ fails to work. The difference is that the countermeasure taken at t_2 is far enough to slow down the decreasing speed of the safe margin. Thus, at t_3 the system state is

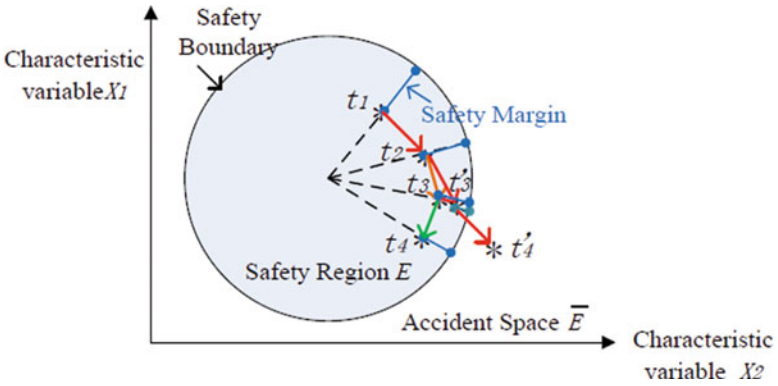


Fig. 2.14 The state transportation after perturbation

already close to the safety boundary and keeps approaching it. Ultimately, the system state breaks through the boundary, with the energy (chemical energy, mechanical energy, kinetic energy, or electric energy) transferring to people, equipment, and environment.

According to the previous analysis, the safety control measure based on the monitoring and warning module is critical to restore system safety after perturbation, as it decides the trend as well as the speed of the system state change. Therefore, in the accident prevention and control procedure, we should establish corresponding emergency plans specific to the object; and strengthen its disturbance control measures to reduce the probability of accidents, eventually avoiding the accidents.

Nevertheless, when the accident happens, there's still shielding method-the isolation of people, environment and energy carrier which we can take to control the damage degree of the energy releases. If the shielding measure fails or not timely, the accident may cause severe direct loss like casualties and property loss, as well as the indirect loss such as damage of the environment, the social influence and the production stagnation, which is described in the accident part in Fig. 2.13.

To sum up, the key of P-SR model is to extract the characteristic state variables of safety critical subsystem to build the safety region; and then determine the safety threshold to establish the safety boundary. That's when the system state can be quantitatively calculated as safe margin.

2.2.2 Case Study

As China's railway transportation system thrives, the train speed is increasingly faster, train numbers are much denser, power supply capacity is bigger, and the multi-factors coupling is higher. With a lot of risk sources, the railway system is both an ultra-safe system and a typical complex system, confronted with enormous

challenges of accident prevention and control. The P-SR model herein provides a solution to solve these problems as the following one accident analysis example and one emergency control example confirm.

2.2.2.1 The Wenzhou Train Collision Accident Analysis

According to the accident investigation report established by State Council of China [27], the P-SR model is employed to analyze and reconstruct the Wenzhou train collision process so as to provide decision support for future accident prevention and the improvement of safety measures.

On 23 July 2011, high speed train D301 from Beijing to Fuzhou collided with the high speed train D3115 from Hangzhou to Fuzhou on Yongwen railway line, Wenzhou, Zhejiang province, China. The analysis of the accident based on P-SR model is established in Table 2.1.

As the relative speed and position of a train with adjacent trains is the essence to control safety, this system safety-critical state space is defined as three-dimension: train running control mode, train speed and train interval. So the safety region is also three-dimension, in which the train running control mode is discrete variable with the value of automatic block control or manual control; the train speed is continuous variable ranging from 0 to 350 km/h; the train interval is discrete variable indicating the number of blocks between two trains running on the same rail at the same direction. To facilitate the graphical display of the safety region, the traffic control mode is set as a third dimension, thus we can describe the changing of the system's safe state in two-dimension space.

Previously we introduced that the special extent of the safety region in dimensionality reduction space is possible to vary with the value of high-dimension variables. In this example, along with the change of train running control mode, the boundary of the two-dimensional safety region made up by the train speed and train interval changes as well, as seen in Fig. 2.15. In automatic block control mode, also the normal operation mode, the safety space is in a large range as shown in area E0; while in manual control mode, the spatial extent of safety region reduces to E1, as automatic train protection (ATP) requires the speed to be lower than 20 km/h and the train interval is required to be as the distance between adjacent stations.

The system safety region composes of the velocity v (km/h) of the first train running into the section and the interval of the subsequent train n (the number of the blocks between two successive trains). In automatic train control mode, the safety boundary is made up of the safety threshold, in which the train running speed is 250 km/h and the minimum safe interval of 2 blocks, as $E_0(v, n) = \{v \leq 250, n \geq 2\}$. In the manual mode, the safety threshold of the speed changes to 20 km/h and the minimum safety interval increases to 3 blocks, as $E_1(v, n) = \{v \leq 20, n \geq 3\}$, for sufficiently stopping the train before any collision.

The safety region is E_0 at E_1, when D3115 set off from Yongjia station at a normal speed into the section under automatic train control mode. However, the control mode changed into manual mode at t_2, with the safety region narrowed down to E_1.

Table 2.1 The accident-causing analysis of Wenzhou train collision

Items		Content	
Energy			
Carrier		Moving motor train unit	
Trigger factors	*Unsafe state*	1. The lightning activity unusually intensive alone Wenzhou-Yongjia and Wenzhou-Ouhai railway line;	
		2. The host in the control center only transfer the fault message received the form track circuit to the monitor and maintenance terminal, while continuing outputting the signal control message according to the occupancy of track at the last moment before malfunction (the track was free so the control center authorized green signal).	
		3. The integrated wireless communication devices in D3115 lost its signal, so the driver couldn't connect to train dispatcher in time.	
	Unsafe behavior	Management	1. The equipment design company had severe defects in the design process and quality control of control center equipment;
			`2. The project director ministry had a series of management failures on equipment bidding, technical examination and inspection for service for newly developed signaling equipment;
			3. The project director ministry had a series of management failures on equipment bidding, technical examination and inspection for service for newly developed signaling equipment;
		Operation	1. The field stuff didn't perform joint interaction control of train running and track occupancy under manual mode;
			2. The D315 was authorized onto the section at automatic control mode without confirmation that the D3115 had arrived at the next station or the equipment had restored to work normally.
The perturbation	1. The lightning struck a trackside signal assembly, burning out its fuses F2, while the transmitter in track circuit 5829AG lost connection with the control center;		
	2. The control center gave an incorrect indication, based on the state before the fault when the track was free, that the track section containing train D3115 was occupied, thereby allowing the signal instruction staying green;		
	3. Due to the communication error between 5829AG track circuit and control center, 5829AG track circuit began to send messy code, causing the computer in-terloacking system in Wenzhou south station displayed red bond on the corresponding section;		
	4. As D3115 run into the malfunctioned track 5829AG the messy code transmitted to the train triggered automatic braking of ATP, so that D3115 came to a halt with 3 times failure to override the system into visual driving mode.		
The monitor and warning	1. The computer interlocking system in Wenzhou south station appeared 'red band';		
	2. The frequency shift track circuit terminal at mechanical room in Wenzhou south station displayed red alarm light;		

(continued)

Table 2.1 (continued)

Items	Content
Energy	
Carrier	Moving motor train unit
	3. The last two communication boards in the track circuit interface unit in Wenzhou south station indicated red warning light;
	4. The computer interlocking system in Wenzhou south station appeared 'red band', while the Centralized Traffic Control System (CTC) in dispatching station didn't.
Safety measures	1. The track maintenance workers walked alone the Wenzhou-Ouhai and Yongjia-Wenzhou railway line to check the occupancy of track;
	2. The railway electricity workers attempted to restore the faulted equipment;
	3. The control mode was change from automatic into manual control mode in Yongjia station, Wenzhou south station and Ouhai station;
	4. The dispatcher instructed the driver of D3115 driving under visual mode at a speed lower than 20 km/h, when encountering red light in the section.
Accident space	
Energy transfer	Train D301 ran at 99 km/h crashed into the rear-end of the D3115 run at 16 km/h.
Accident	The 15th and 16th coached at rear of D3115 and the front five coaches of D301 were derailed.
Shielding	The driver of D301 pulled on emergency brake at the sight of D3115.
Loss	40 people were killed and 172 injured;
	7 motor train set vehicles was scrapped, 2 broken heavily, 5 broken at medium, 15 broken slightly; the network of overhead Contact System in accident section collapsed; the railway line at accident section shut down for 32 h and 35 min.

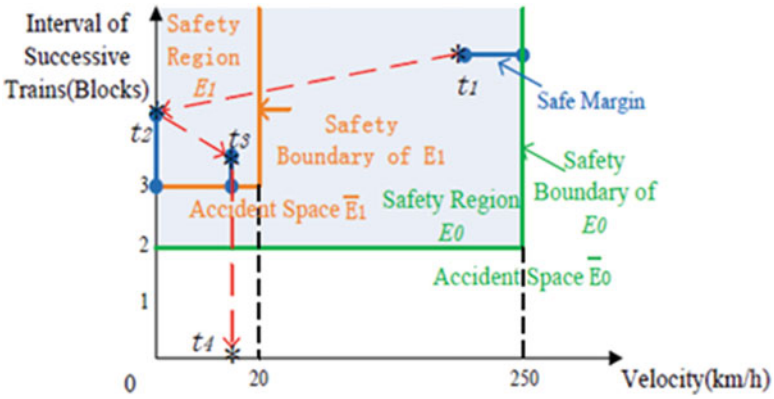


Fig. 2.15 The evolution of system state in Wenzhou train collision based on safety region

Table 2.2 The monitor and warning information Wenzhou collision and corresponding evaluation

Time	The monitor and warning of equilibrium state	Safety measures	Safety regions	Safety margin	Evaluation of safety measures
t_1	None	None	E_0	Equilibrium state	–
t_2	The inconformity of the display in CTC and train control center	The train control mode was change to manual control mode in Yongjia station, Wenzhou south station.	E_1	Increasing	Effective
	Track circuit sent messy code	D3115 was stopped by the Automatic Train Protection (ATP)	E_1	Increasing	Failed
	None	The driver of train D3115 overrode the ATP and drove at visual mode.	E_1	Decreasing	Dangerous
t_3	None	The following train D301 approached onto the section of track where D3115 had been stopped at automatic mode.	E_1	Decreasing dramatically	Slight
t_4	None	Emergency brake of D301	E_1	Enter accident space	

Soon after, D3115 was stopped by the ATP when running onto the track 5829AG with faulted track circuit. At the time of t_3, D301 entered the same section occupied by D3115 as a way of the automatic mode, which it shouldn't. Two minutes later, D3115 finally overrode the ATP to start the visual driving mode. Nonetheless, the interval between these two trains decreased sharply at this time. As there was no effective warning, no imperative safety measure was taken.

Thus the safe margin diminished dramatically. Eventually, D301 collided with D3115 at t_4 that the system state broke through the safety boundary, with energy transfer, causing the accident. The course of the accident is shown in Fig. 2.15 as red arrow lines. The warning and monitor information with relative safety measures at each time is evaluated according to safe margin in Table 2.2.

According to the analysis of the P-SR accident-causing model, it is the joint efforts and the interaction between multiple factors that put the system at risk of accident. However, it is the control measures that finally decide whether an accident will happen or not. In Wenzhou train collision accident, the safety measures adopted according to the early warning has somewhat maintained system safe margin. But when the system neither obtained the early warning information in the field, nor did any imperative human or equipment safety control measures are taken, the system safe margin began to drop dramatically until the accident happened.

2.2.2.2 Specified Application of the Safety Control Measures

This section focuses on the system safety control measures to restore the order of the system. Specific to the railway system, train dispatching and rescheduling is the imperative method to ensure both the operation safety and transportation capability of the whole system, as essentially they avoid the time and space conflicts between different trains, which is the decisive factor to the range of safety region. Therefore, a train rescheduling method is specially proposed in this part.

1. The principle and strategy of train rescheduling

When the railway system is in unbalanced state, strategies to restore the system need to follow certain principles.

(A) Principles of train rescheduling

- Schedule the train in the original path and avoid detour and outage to the greatest extent.
- When detour is necessary, confirm the train and the line and choose the shortest one.
- Higher grade trains can't be overtaken by lower ones.
- Passenger trains can't be overtaken by freight trains.
- The punctual trains have a higher priority.
- Passenger trains can arrive in advance but can't departure in advance.

(B) Strategies for train rescheduling

- Detour, outage, reconnection and turn-back can be adopted when necessary.
- Change the section running time.
- Change the dwelling time in station.

2. Rescheduling method

- Change the overtaking station or time.

Paper [28] summarizes 3 rules for the events dispatching. A first-to-start dispatcher selects the next train to be moved based on the earliest start time. A first-to-finish dispatcher selects the next train to be moved based on the earliest finish time on its next segment. Other possible dispatchers can be created by setting the dispatching decision time for train i as $t_i = (1 - \delta)u_i + \delta v_i$, where u_i is the start time for train i and v_i is its expected finish time on its next immediate segment and $\delta \in [0,1]$.

While the trains' priority is not considered in the dispatching rules mentioned before. As the priorities are different between the neighboring trains, there will be 3 situations: the neighboring trains have the same priority (Fig. 2.16a), higher priorities train run after the lower priority train (Fig. 2.16b), and lower priority train run after the higher priority train (Fig. 2.16c).

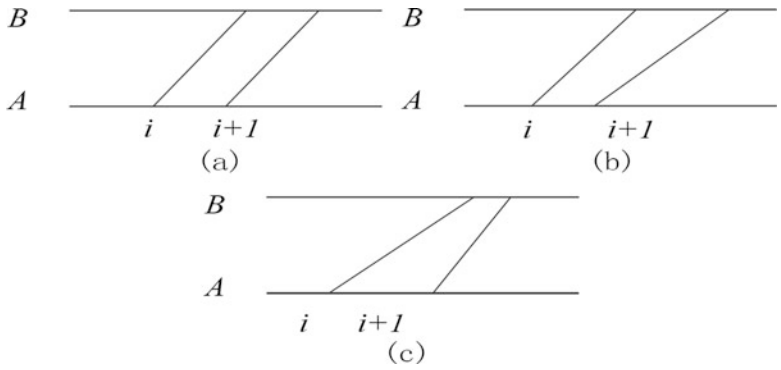


Fig. 2.16 Different tracking form of different train degree

Table 2.3 Formula as for the actual start time

(a)	If $S''_{i+1} - S_i \geq I$	Then $S_{i+1} = S''_{i+1}$, $S_i = S_i$
	If $S''_{i+1} - S_i < I$	Then $S_{i+1} = S_i + I$, $S_i = S_i$
(b)	If $S''_{i+1} - S_i \geq I$	Then $S_{i+1} = S''_{i+1}$, $S_i = S_i$
	If $S''_{i+1} - S_i < I$	Then $S_{i+1} = S_i + I$, $S_i = S_i$
(c)	If $S''_{i+1} - S_i \geq I + t_i - t_{i+1}$	Then $S_{i+1} = S''_{i+1}$, $S_i = S_i$
	If $S''_{i+1} - S_i < I + t_i - t_{i+1}$	Then $S_{i+1} = S''_{i+1}$, $S_i = S_{i+1} + I$

When the actual start time of trains (AST) in each section is obtained, the timetable is got too. So the calculation of AST is the key of the problem. In this paper, AST is calculated by the formulas in Table 2.3.

In Table 2.3, s_i stands for AST, while s_i'' stands for the earliest start time (EST). The two concepts can be distinguished that AST is EST considering constrains between trains. EST can be got by two factors, (a) the reckoning time according to AST and section running time in last section and the operation time in last station, (b) the start time in the original timetable. We choose the bigger one as the result. It can be seen in (2.2.1).

$$s''_k = \max\left(s_{k-1} + t_{k-1} + t_j, s^*_k\right) \tag{2.2.1}$$

Where, s''_k stands for EST in section k, s_{k-1} stands for AST in section $k-1$, t_{k-1} stands for the running time in section $k-1$, and t_j stands for the operation time in station j.

On account of factors such as weather, track condition, equipment condition and etc., the velocity of trains is not constant. So we consider the pure running time as a variable number. The section running is depicted in (2.2.2).

$$t_p = \alpha_{p-1}\tau_q + \alpha_{p+1}\tau_t + \delta \tag{2.2.2}$$

Where, α is a $0-1$variable representing whether train stops in station or not, τ_q and τ_t stand for the addition time of start and stop, t_p stands for the pure running time, δ is a stochastic number.

The variation of section running time enriches the problem space, and we can find a better solution. The value of δ is vital to the quality of the result. R. Albrecht [29] made many experiment to obtain a more proper value in his doctoral dissertation, finding that when distributed normally (that is $\delta \in N(0, \alpha_T)$ and $\alpha_T = m/2$ [28], the result could be better. m stands for the section running time. The conclusion is still applied in this paper.

The algorithm is depicted in the following.

Step1: Choose all the events in section I.
Step2: Calculate the earliest start time of section i according to formula (2.2.1).
Step3: Calculate the actual start time of the section event according to Table 2.3.
Step4: Do $I = I + 1$ until the last section.
Step5: Repeat step 1 to step 5 N times (N is determined by decision maker, it can be 100 or another), thus we have N feasible schemes, and find the best solution according to the object function among the N feasible schemes.
Step6: Draw the adjusted train diagram.

3. An experimental example of the method

The results of using the method before are discussed here for a representative example on Jin-qin passenger railway (approximately 260 km with 9 stations), China. The case is based on real data and a scene with disorder is assumed.

The assumed scene: the section Junliangcheng north station to Binhai station suffered heavy rainfall during the period 13:00–17:00. And the allowed speed of the trains passed by then is 100 km/h. Because of the bad weather, 11 trains are late. So a quick adjustment of train timetable is needed.

We take the minimum deviation between the original timetable and the adjusted timetable as objective, then carry out the algorithm before with related data and the output objective distribution is shown in Fig. 2.17. The results are normal distributed. The result with minimum deviation time is an ideal scheme and the rescheduled timetable is shown in Fig. 2.18.

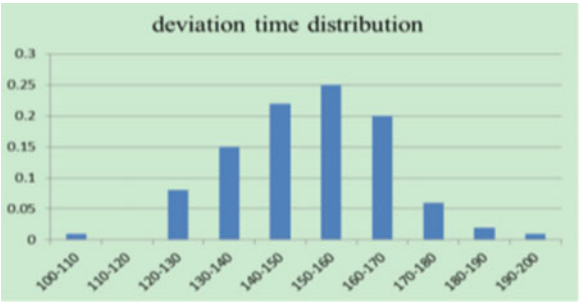


Fig. 2.17 Objective distribution with running time

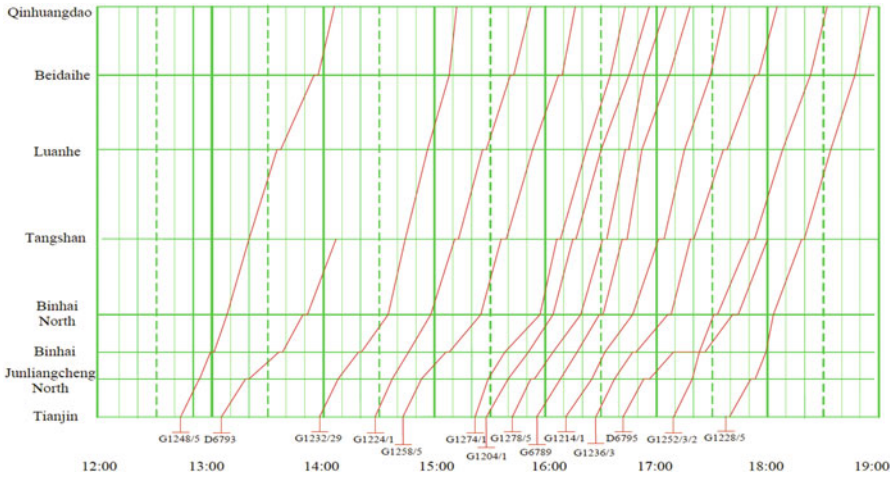


Fig. 2.18 Train timetable with running time

We can see that in Fig. 2.18, D6795 is a train with lower priority compared to others and in order to cause larger deviation it is overtaken by train G1253 in Binhai station only. In this case the objective number is 109.3672 and the result can be got in an acceptable time. The method has also been applied with success to a range of test problems with various network sizes, number of trains and works well.

2.2.2.3 Corresponding Prevention Measures

On the basis of the theory and analytical method of the P-SR accident model, we can further conclude the following preventive measures against accidents.

1. Strengthen the implement of technical engineering in the system changes and control measures parts. As the external disturbance is almost inevitable, to maintain system balanced state is the critical process to prevent an accident.
2. Strengthen the monitoring of the system running state and quantitative analysis of safety region, so as to timely reflect the safe state of the system. And then offer the safe state analysis and early warning information to provide basis for adopting corresponding control measures;
3. Take comprehensive and effective safety control measures based on safe state and early warning information, and at the same time constantly monitor the system state to assess the effectiveness of safety measures to adjust inappropriate control measures in time.
4. Strengthen the construction of emergency management and human emergency response. As human bears huge psychological pressure when the system works out of order after disturbance, they are likely to make inappropriate decisions or take unsuitable actions that may aggravate the reduction of system safety margin.

References

1. A. Vannelli, M. Vidyasagar, Maximal Lyapunov functions and domains of attraction for autonomous nonlinear systems. Automatica **21**, 69–80 (1985)
2. G. Chesi, Estimating the domain of attraction via union of continuous families of Lyapunov estimates. Syst. Control Lett. **56**, 326–333 (2007)
3. F.M.A. Ghali, SA. El Motelb. Stability region estimation of hybrid multi-machine power system, in *Proceedings of the Robot and Human Interactive Communication* (2000), pp. 49–154
4. X. Lin, The waveform relaxation method for estimating the asymptotic stability region of nonlinear diferential dynamic systems. Chin. J. Eng. Math. **27**(3), 479–486 (2010)
5. Y. Lin, Z. Cai, Determination of power system stability region using hamilton-jacobi formula. Proc. Chin. Soc. Electr. Eng. **27**(28), 19–23 (2007)
6. H. Xin, J. Tu, J. Xie, et al., LMI-based stability region estimation for dynamical systems with saturation nonlinearities and a short time-delay. Control Theory Appl. **26**(9), 970–976 (2009)
7. P.L. Benjamin, S.D. Sudhoff, S.H. Zak, et al., Estimating regions of asymptotic stability of power electronics systems using genetic algorithms. IEEE Trans. Control Syst. Technol. **18**(5), 1011–1022 (2010)
8. C. Cortes, V.N. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
9. V.N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1999)
10. C.F. Lin, S.D. Wang, Fuzzy support vector machines with automatic membership setting. Stud Fuzz **177**, 233–254 (2005)
11. X. Shi, Z. Guo, Multi-classification method of GA-VM on identifying grade of expansive soils. J. Civil Archit. Environ. Eng. **31**(4), 44–48., 59 (2009)
12. X. Li, A. Li, X. Bai, Robust face recognition using HMM and SVM. Opto-Electron. Eng. **37**(6), 103–107 (2010)
13. H. Che, F. Lu, Z. Xiang, Defects identification by SVM-DS fusion decision-making with multiple features. J. Mech. Eng. **46**(6), 101–105 (2010)
14. A. Kuhlmann, *An Introduction to Safety Science* (Germany, 1981)
15. H.W. Heinrich, D. Petersen, N. Roos, *Industrial Accident Prevention: A Safety Management Approach*, 5th edn. (Mcgraw-Hill, New York, 1980)
16. H. Xueqiu, *Safety Engineering* (China University of Mining and Technology, China, 2000)
17. N. Leveson, A new accident model for engineering safer systems. Saf. Sci. **42**(4), 237–270 (2004)
18. China higher education committee in safety engineering guidancs, Safety System Engineering (China Coal Industry, China, 2002)
19. P.M. Salmon, N. Goode, F. Archer, C. Spencer, D. McArdle, R.J. McClure, A system approach to examining disaster response: using accimap to describe the factors influencing bushfire response. Saf. Sci. **70**, 114–122 (2014)
20. Y.X. Fan, Z. Li, J.J. Pei, H. Li, J. Sun, Applying system thinking approach to accident analysis in China: case study of "7.23" Yong-Tia-Wen High-speed train accident. Saf. Sci. **76**, 190–201 (2015)
21. J. Rasmussen, Risk management in a dynamic society: a modelling problem. Saf. Sci. **27**, 183–213 (1997)
22. E. Hollnagel, Investigation as an impediment to learning, in *Remaining Sensitive to the Possibility of Failure, Resilience Engineering Series*, ed. by E. Hollnagel, C. Nemeth, S. Dekker (Ashgate, Aldershot, 2008)
23. R. Amalberti, The paradoxes of almost totally safe transportation systems. Saf. Sci. **37**, 109–126 (2001)
24. R. Woltjer, E. Pinska-Chauvin, T. Laursen, B. Josefsson, Towards understanding work-as-done in air traffic management safety assessment and design. Reliability Engineering & System Safety. (2015 March). [Online]. Elsevier. http://www.sciencedirect.com

25. X. Ancheng, F. Wu, F.L. Qiang, M. Shengwei, Power system dynamic safety region and its approximations. IEEE Trans. Circuits Syst. Regul. Pap. **53**, 2849–2859 (2006)
26. Q. Yong, S. Jingxuan, Z. Yuan, Z. Shengzhi, J. Limin, Online safety assessment of rail vehicles in service state based on safety region estimation. J. Cent. South Univ. Sci. Technol. **44**, 195–200 (2013)
27. The State Council of China, *Yongwen railway line major traffic accident investigation report*. China. (2011 December)
28. A.R. Albrecht, D.M. Panton, D.H. Lee, Rescheduling rail networks with maintenance disruptions using problem space search. Comput. Oper. Res. **40**, 703–712 (2013)
29. A.R. Albrecht, *Integrating railway track maintenance and train timetables*. University of South Australia (2009)

Chapter 3
Train Equipment Fault Diagnosis and Prognosis

3.1 Fault Diagnosis of Rolling Bearings Based on Safety Region

3.1.1 The Configuration and Faults of Rolling Bearings

Train rolling bearings are the most commonly used mechanical components in the rail transportation. They are important components of the running part of the train. According to the statistical report, about ten percentage to twenty percentage of the rolling bearings can reach to the designed life longevity. Many kinds of fault may happen during the train operation, including wearing, erosion, breakage, and so on.

Currently, two-column cone element bearings are mostly used on trains. In addition to the domestic production, those bearings are imported from abroad such as SKF and FAG. The configuration of the train rolling bearings is shown in Fig. 3.1, which is made up of inner race, outer race, rolling elements, cage, and space ring.

According to the fault location of the rolling bearing, the faults can be divided into three kinds, including the outer race fault, inner race fault, and the rolling element fault, shown in Fig. 3.2.

3.1.2 Rolling Bearings Vibration Mechanism

According to the difference of the feature during the operation of the train rolling bearings, the vibration signal can be divided into two categories, including wearing and the surface damage. Wearing will not lead to the rolling bearing damage immediately. Its harm is far less than the surface damage. Therefore, we mainly discuss the surface damage. When the surface damage occurs, as the rolling elements strike a local fault on the outer or inner race, a shock is introduced that excites high-frequency resonances of the whole structure between the bearing and the response

Y. Qin, L. Jia, *Active Safety Methodologies of Rail Transportation*, Advances in High-speed Rail Technology, https://doi.org/10.1007/978-981-13-2260-0_3

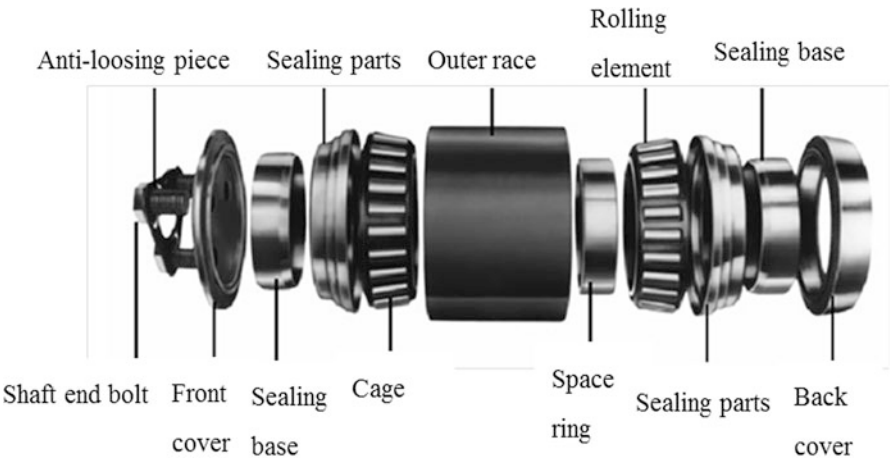


Fig. 3.1 The configuration of the train rolling bearing

Fig. 3.2 Train rolling elements faults

transducer. The same happens when a fault on a rolling element strikes either the inner or outer race.

3.1.3 Procedure of the Safety Region Identification of Rolling Bearings

The safety region identification of the train rolling bearings is composed of two stages, including the state feature extraction and the boundary division of the safety region. The first stage is to finish the signal decomposition of bearings vibration signals and calculate the state feature index. The second stage is to finish the classification of the different faults based on the feature index, named safety boundary classification.

Specific steps of the safety region identification are as follows:

Step 1: Collect the vibration data under the normal state and the fault state, respectively.
Step 2: Considering the data length of collected vibration data, the collected data are divided into several segments: the number of the segment of the decided by the data sampling frequency and experiment condition.
Step 3: Apply the local mean decomposition to the segmented data, and get components of the corresponding data.
Step 4: To ensure the same dimension of every state feature vector, calculate the minimum number of the segment of corresponding data, and take the number as the dimension number of the corresponding data.
Step 5: Choose the state feature index, and calculate every index value of the corresponding signal segment; thus, the state feature vector can be obtained.
Step 6: The state feature index are marked as normal state vector and fault state according to the data condition from the Step 1. Fault states are numbered from 1 to N if this is a multi-classification problem.
Step 7: Apply the LSSVM to the classification problem, and decide the safety region boundary.
Step 8: If this is a multi-classification problem, apply the DAGSVM to the classification of the problem, and obtain the multi-classification model.

Steps of safety region identification of the train rolling bearings are also shown in the Fig. 3.3.

3.1.4 LMD of the Vibration Signal of Rolling Bearings

The local mean decomposition (LMD) was put forward by Jonathan S. Smith and has been used to analyze electroencephalogram signal [1]. LMD can self-adaptively decompose a complicated multicomponent signal into a set of product functions (PFs), each of which is the product of an envelope signal from which instantaneous amplitude of the PF can be got and a purely frequency-modulated signal from which a well-defined instantaneous frequency could be calculated. Therefore, each resulting PF component is, in fact, a mono-component amplitude-modulated and frequency-modulated (AM-FM) signal. Furthermore, the complete time-frequency distribution of the original signal could be obtained by assembling the instantaneous amplitude and instantaneous frequency of all PF components.

Since a multicomponent AM-FM signal could be decomposed into a set of mono-component AM-FM signals by LMD, LMD is suitable for processing of the multicomponent AM-FM signal. When gear or roller race fault occurs in the rotating machinery, it is generally the case that the vibration signals measured by sensor present AM-FM feature. For this kind of signal, the demodulation analysis is the most common method [2]. Therefore it is possible to apply LMD to the feature extraction of gear and roller race fault vibration signals because the decomposition process of LMD is exactly the process of demodulation.

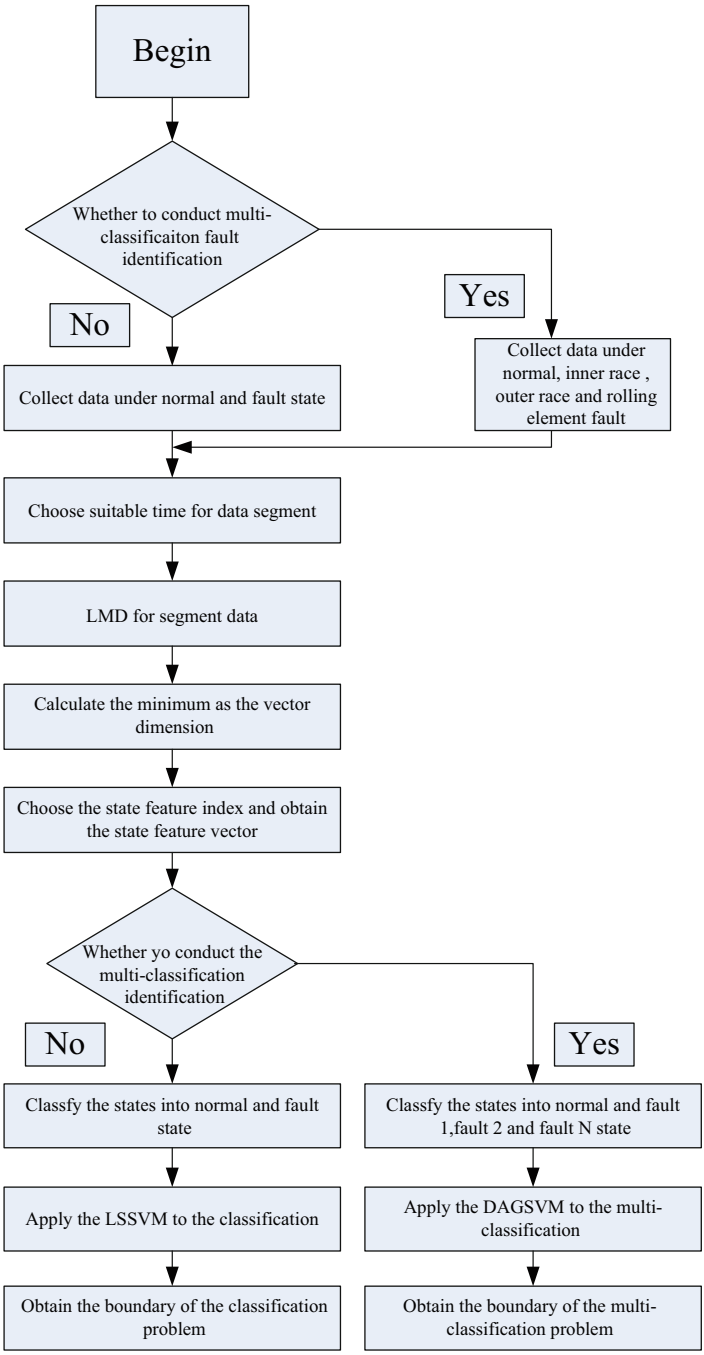


Fig. 3.3 The flow chart of the safety region identification

Furthermore the modulation feature could be extracted effectively by applying spectra analysis to instantaneous amplitude of each PF. Until now, lots of research results have been proposed based on the LMD, such as the order tracking [3], Fourier transform [4], energy demodulation [5], and envelope analysis [6].

1. Product Function

The frequency-modulated signal can be written as $a(t)$ and $f(t)$ which are the instantaneous amplitude and the instantaneous phase, respectively. The formula suits the physical expression of the signal component signal, so the instantaneous frequency has physical meaning.

$$x(t) = a(t)\cos\left[2\pi \int f(t)dt\right] \tag{3.1.1}$$

Thus, the product function (PF) is defined; it can be represented by the production of an envelope signal and a pure frequency-modulated signal whose amplitude is one, shown as (3.1.2)

$$PF(t) = a(t)\mathrm{s}(t) \tag{3.1.2}$$

Firstly, the PF component is a modulated signal which includes the amplitude modulation and frequency modulation. It contains some certain important features which can be obtained easily form those components. Besides, the PF component is signal component signal, which only represents one vibration model. Therefore, combining all the PF components can obtain the time-domain distribution of the original signal.

2. LMD Analysis Method

The nature of LMD is to demodulate AM-FM signals. By using LMD a complicated signal can be decomposed into a set of product functions, each of which is the product of an envelope signal and a purely frequency-modulated signal. Furthermore, the completed time-frequency distribution of the original signal can be derived [7].

We apply the LMD based on the cubic spline function, and the steps are shown as follows [8]:

Step 1: Determine all local extreme n_i of the original signal x(t), and then the mean value of two successive extreme n_i and n_{i+1} can be calculated by

$$m_i(t) = \frac{n_i + n_{i+1}}{2} \tag{3.1.3}$$

All mean value mi of two successive extreme are connected by straight lines, and then local mean function $m_{11}(t)$ can be formed by using moving averaging to smooth the local means $m_i(t)$.

Step 2: A corresponding envelope estimate c_i is given by

$$a_i(t) = \frac{|n_i - n_{i+1}|}{2} \tag{3.1.4}$$

Step 3: The local mean function $m_{11}(t)$ is subtracted from the original signal x(t), and the resulting signal $h_{11}(t)$ is given by

$$h_{11}(t) = x(t) - m_{11}(t) \tag{3.1.5}$$

Step 4: $h_{11}(t)$ can be amplitude demodulated by dividing it by envelope function $c_{11}(t)$

$$s_{11}(t) = \frac{h_{11}(t)}{c_{11}(t)} \tag{3.1.6}$$

Step 5: Ideally, $s_{11}(t)$ is a purely frequency-modulated signal, namely, the envelope function $c_{12}(t)$ of $s_{11}(t)$ should satisfy $c_{12}(t) = 1$. If $c_{12}(t) = 1$, then $s_{11}(t)$ is regarded as the original signal, and the above procedure needs to be repeated until a purely frequency-modulated signal. Therefore,

$$\begin{cases} h_{11}(t) = x(t) - m_{11}(t) \\ h_{12}(t) = s_{11}(t) - m_{12}(t) \\ \vdots \\ h_{1n}(t) = s_{1(n-1)}(t) - m_{1n}(t) \end{cases} \tag{3.1.7}$$

$$\begin{cases} s_{11}(t) = \dfrac{h_{11}(t)}{c_{11}(t)} \\ s_{12}(t) = \dfrac{h_{12}(t)}{c_{12}(t)} \\ \vdots \\ s_{1n}(t) = \dfrac{h_{1n}(t)}{c_{1n}(t)} \end{cases} \tag{3.1.8}$$

Step 6: Envelope signal $c_1(t)$, namely, instantaneous amplitude function, can be derived by multiplying together the successive envelope estimate functions that are acquired during the iterative process described above

$$c_1(t) = c_{11}(t) \cdot c_{12}(t) \cdots c_{1n}(t) = \prod_{i=1}^{n} c_{1i}(t) \tag{3.1.9}$$

Step 7: Multiplying envelope signal $c_1(t)$, by the purely frequency-modulated signal $s_{1n}(t)$, the first product function PF_1 of the original signal can be obtained.

$$PF_1(t) = c_1(t) \cdot s_{1n}(t) \tag{3.1.10}$$

Step 8: Subtract the first PF component $PF_1(t)$ from the original signal x(t), and we have a new signal $r_1(t)$, which becomes the new original signal, and the whole of the above procedure is repeated, up to k times, until r_k becomes monotonic function.

$$\begin{cases} r_1(t) = x(t) - PF_1(t) \\ r_2(t) = r_1(t) - PF_2(t) \\ \vdots \\ r_k(t) = r_{k-1}(t) - PF_k(t) \end{cases} \tag{3.1.11}$$

Thus, the original signal x(t) was decomposed into k product and a monotonic function r_k

$$x(t) = \sum_{v=1}^{k} PF_v(t) + r_k(t) \tag{3.1.12}$$

Furthermore, the corresponding complete time-frequency distribution could be obtained by assembling the instantaneous amplitude and instantaneous frequency of all PF components, shown in Fig. 3.4.

3. LMD Characteristics

According to the elaboration of the LMD algorithm, the characteristic of the LMD can be summarized as follows [2].

1. Self-Adaptation

The time scale is an important factor to describe the signal feature, and it has close correlation to the local extreme points. Generally, the time range between two local extremes is chosen as the time scale of the signal, and it can reflect the vibration wave of the signal. The practical vibration signal usually contains multiple vibration waves, and every vibration wave contains a kind of signal feature. The small time-scale parameter corresponds to the high-frequency vibration wave and reverse either. Based on the time-scale parameter, different vibration scale can be separated out from the original signal. Therefore, LMD is a kind of self-decomposition without prior knowledge. Besides, different signals contain different local extremes and different vibration wave component. These components have different frequency range and central frequency. But the LMD can obtain different decomposition

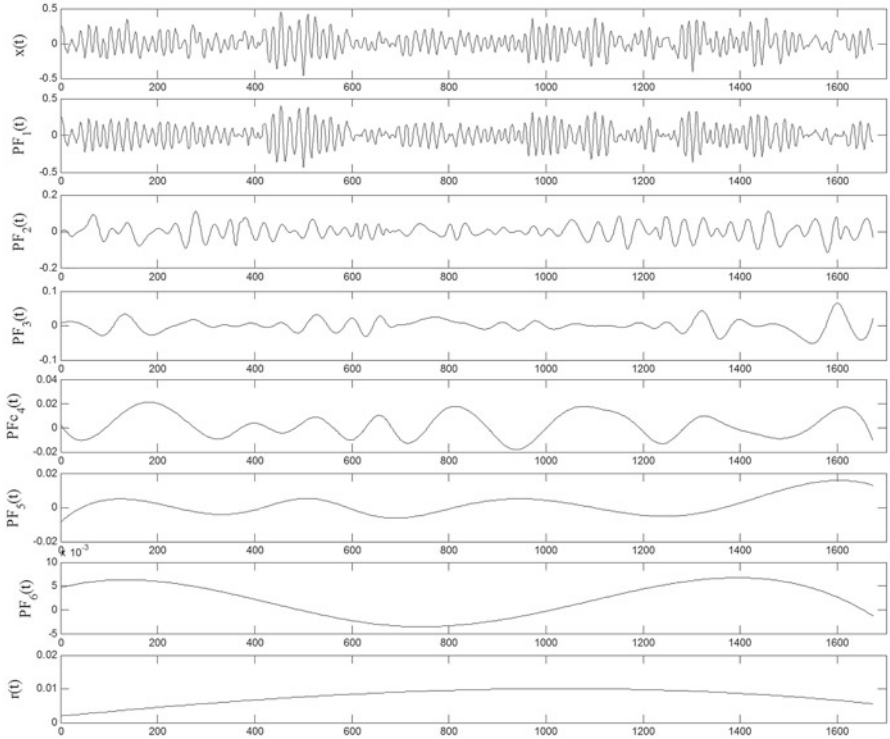


Fig. 3.4 Vibration signal and corresponding PFs

results toward different signals to suit the signal-wave components. Therefore, this shows the self-adaptation of the LMD.

2. Independence

LMD decomposes the original into the pure frequency modulation and the envelope signal. Then the production of the pure frequency modulation and the envelope signal can be taken as the PF component. For the same PF component, the instantaneous amplitude and the instantaneous frequency are independent. This independence can bring about two merits. The first merit is that the PF component can hold more local feature which will not disappear after the production. Secondly, the instantaneous frequency is positive according to the calculation; thus the result is physically meaningful.

3. Orthogonality

Every PF component decomposed by the LMD is a one-time scale of the signal. Therefore, the PF components are orthogonal to each other, shown as (3.1.13)

$$\sum_{i,j=1}^{k} PF_i(t)PF_j(t) = 0, \quad i \neq j \tag{3.1.13}$$

It is noted that the orthogonality is a local orthogonality, because some certain signal may have the same frequency when the signal is decomposed into two adjacent PF components. This phenomenon will turn out to be more obvious as the signal becomes longer.

4. Completeness

The completeness of the LMD represents that the sum of the all those PF components equal to the original signal.

Above all, compared to other methods, LMD has a better performance when dealing with the endpoint effect, reducing the calculation time and holding the integrity of the original signal.

3.1.5 Safety Region Feature Extraction of Rolling Bearings

After the decomposition of the vibration signal, the feature value can be calculated based on that. The following content will introduce the feature extraction method.

The first class of method relates the direct calculation of vibration data so as to obtain some certain time-domain feature parameters. This kind of method doesn't need any time-frequency transformation, and the calculation volume is small. But the time-domain feature can represent little useful information, and the performance is not convincing enough.

The second class of method relates the decomposition or transformation of vibration signal. Different from the first class of method, this method doesn't calculate the time-domain parameters directly. Pretreatment methods, such as Fourier transform, wavelet decomposition, and empirical mode decomposition, should be carried out to analyze the signal. After that, the feature will be extracted. Compared to the first class of method, those methods usually have a better performance, but the calculation volume is large.

The first method has a long history, and the time-domain parameter has closed to intactness. With the development of the machinery industry, the complexity of the signal has increased a lot, and the first class of method cannot meet the demand of analysis accuracy. Therefore, recent researches pay attention to the second class of method. And feature index based on the energy and entropy has developed a lot.

Assume the collected data are $x = \{x_1, x_2, \ldots, x_N\} = \{x_i\}$, $i = 1,2,\ldots, N$, where N represents the sample numbers. The following will introduce the direct time-domain feature and features based on energy and entropy.

1. Time-Domain Feature

   1. Root mean square (RMS)—the root mean square represents root mean square the vibration amplitude

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \tag{3.1.14}$$

where x is the mean value of the whole sample. The RMS increases with the fault develops and it can reflect the vibration energy. Thus it is sensitive to the race fault and insensitive to the peeling or scratching fault [9].

2. Peak value, which reflects the maximum difference value of the vibration amplitude

$$Peak = \frac{1}{2}(\max(x_i) - \min(x_i)) \tag{3.1.15}$$

3. Crest factor

$$Crest\ factor = \frac{Peak}{RMS} \tag{3.1.16}$$

The crest factor reflects the signal intensity, which is suitable for the surface corrosion and damage. It is also sensitive to the instantaneous impulse of the rolling elements or the cage [9].

4. Square root amplitude

$$x_R = \left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2 \tag{3.1.17}$$

5. Absolute mean value

$$x_R = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{3.1.18}$$

6. Skewness, which measures the asymmetry of the signal probability distribution

$$Skewness = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3 \tag{3.1.19}$$

7. Skewness factor, which is a dimensionless factor relates to the skewness

$$Skewness\ factor = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3}{RMS^3} \tag{3.1.20}$$

8. Kurtosis

$$Kurtosis = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^4 \tag{3.1.21}$$

9. Kurtosis factor, which is a dimensionless factor relates to the kurtosis

$$Kurtosis\ factor = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^4}{RMS^4} \tag{3.1.22}$$

Kurtosis and kurtosis factor are used to reflect the rule vibration amplitude; when the fault occurs, the kurtosis will increase and it is sensitive to the impulse. So those two features are suitable for the early fault diagnosis.

10. Shape factor

$$Shape\ factor = \frac{RMS}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \tag{3.1.23}$$

11. Impulse factor

$$Crestf\ actor = \frac{Peak}{|\bar{x}|} \tag{3.1.24}$$

12. K factor

$$Kfactor = RMS \times Peak \tag{3.1.25}$$

2. Feature Index Based on the Energy and Entropy

Besides the direct time-domain parameters, feature indexes based on the energy and entropy are introduced.

1. Energy value

$$Energy = \sum_{i=1}^{N} |x_i|^2 \tag{3.1.26}$$

The energy value is widely used in fault diagnosis. When the fault occurs, the machinery tends to vibrate violently. Thus the energy value of the amplitude can be used to diagnose the state of the machinery.

2. Energy moment

$$Energy\ moment = \sum_{i=1}^{N} (i \cdot \Delta t)|x_i|^2 \tag{3.1.27}$$

where Δt is the sample period. This feature not only considers the energy value of the vibration amplitude; the amplitude distribution is also put into account so as to uncover the fault.

3. Shannon entropy

$$Shanon\ entropy = -\sum_{i=1}^{N} p(x_i)\log p(x_i) \tag{3.1.28}$$

where p(xi) represents the probability of x = xi, and $\sum_{i=1}^{N} p(x_i) = 0$. Shannon entropy is used for the measuring the uncertainty of the signal. The fault signal has more uncertainty.

4. Renyi entropy

$$Renyi\ entropy = \frac{1}{1-a}\sum_{i=1}^{N}\log[p(x_i)^a] \tag{3.1.29}$$

where a represents the order of the Renyi. Renyi entropy equals to Shannon entropy when a = 1.

5. Energy entropy

$$Energy\ entropy = -\sum_{j=1}^{M} p_j \log p_j = -\sum_{j=1}^{M}\frac{E_j}{E_A}\log\left(\frac{E_j}{E_A}\right) \tag{3.1.30}$$

where p_j is the ratio of the energy D_j to the whole energy

E_j is the j_{th} energy of the signal, $E_j = \sum_{i=1}^{N}\left|D_j\right|^2$

E_A is the sum of signal, $E_A = \sum_{j=1}^{M} E_j$

3.1.6 The Safety Region Identification of Rolling Bearings Based on SVM

1. *Support Vector Machine*

After the analysis of the above sectors, the classification of different feature indices is critical for the state identification based on the safety region. Support vector machine (SVM) is powerful classification tool for this purpose, and it can figure out the boundary function directly. Later in this book, support vector data description is also applied in the degradation performance of rolling bearings. So the basic concept of SVM is introduced firstly.

SVM is proposed by Vapnik in 1995 [10]. SVM is a creative machine learning technology, and it has become one of the most standard tools for data science. SVM is developed based on the statistical learning theory (SLT) and the structure risk minimization (SRM) [11]. SRM is more advantageous than traditional empirical.

Risk minimization (ERM) so as to make the SVM spreads widely [12]. SVM has strict math and theory base without local minimum value, which makes it widely used in the area of pattern recognition and control field [13–15].

In this sector, the basic concept of SVM and machine learning technology based on SVM is introduced and discussed.

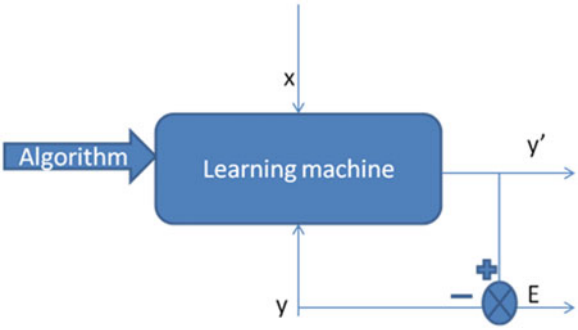


Fig. 3.5 Machine learning model

1. SVM theory

A. Basic theory

SVM theory involves the machine learning theory, ERM, and SRM. All those theory need to be introduced.

(a) Machine learning theory

Machine learning theory is to design some algorithms which can make the algorithm learn automatically. Classical definition of SVM is that a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Assume the input and output variables are x and y, respectively. The variable x and the variable y have corresponding relationship. The problem solved by the machine learning is that training a model to make this model can simulate the relationship between x and y. The model can output a result which approaches the variable y as much as possible. Thus, when the model is obtained, the predicted value of y can be regarded variable y itself. The machine learning model is shown in Fig. 3.5.

(b) Empirical risk minimization

As shown above, machine learning can be treated as the relationship between x and y. This relationship meets the union probability distribution $F(x, y)$. The task of the machine learning is to find the optimal function to predict y according to n-independent identical distribution samples, such as (x_1, y_1), (x_2, y_3), …, (x_n, y_n). The predictive result has to decrease the expectation risk as much as possible

$$R(w) = \int L(y, f(x, w))dF(x, y) \tag{3.1.31}$$

where $\{f(x, w)\}$ is the predictive function set, w is the parameter, and $L(y, f(x, w))$ is the loss function [16].

In general, the risk R(w) cannot be computed because the distribution F(x, y) is unknown to the learning algorithm (this situation is referred to as agnostic learning). However, we can compute an approximation, called empirical risk, by averaging the loss function on the training set, named empirical risk minimization (ERM) which is shown as Eq. 3.1.32.

$$R_{emp}(w) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, f(x_i, w)) \tag{3.1.32}$$

Empirical risk minimization for a classification problem with 0–1 loss function is known to be an NP-hard problem even for such relatively simple class of functions as linear classifiers, though it can be solved efficiently when minimal empirical risk is zero, i.e., data is linearly separable.

In practice, machine learning algorithms cope with that either by employing a convex approximation to 0–1 loss function (like hinge loss for SVM), which is easier to optimize, or by posing assumptions on the distribution $F(x, y)$ (and thus stop being agnostic learning algorithms to which the above result applies).

(c) VC dimension

Vapnik-Chervonenkis dimension (VC dimension) has been proposed and improved by Russian mathematicians Vapnik and Chervonenkis from 1960 to 1990. In Vapnik-Chervonenkis theory, the VC dimension is a measure of the capacity (complexity, expressive power, richness, or flexibility) of a space of functions that can be learned by a statistical classification algorithm. It is defined as the cardinality of the largest set of points that the algorithm can shatter.

Formally, the capacity of a classification model is related to how complicated it can be. For example, consider the thresholding of a high-degree polynomial: if the polynomial evaluates above zero, that point is classified as positive, otherwise as negative. A high-degree polynomial can be wiggly, so it can fit a given set of training points well. But one can expect that the classifier will make errors on other points, because it is too wiggly. Such a polynomial has a high capacity. A much simpler alternative is to threshold a linear function. This function may not fit the training set well, because it has a low capacity. This notion of capacity is made rigorous below.

A classification model f with some parameter vector theta is said to shatter a set of data points (x_{1}, x_{2},. . ., x_{n}) if, for all assignments of labels to those points, there exists a theta such that the model f makes no errors when evaluating that set of data points [13].

The VC dimension of a model f is the maximum number of points that can be arranged so that f shatters them. More formally, it is the maximum cardinal D such that some data point set of cardinality D can be shattered by f.

(d) Probabilistic upper bound

The VC dimension can predict a probabilistic upper bound on the test error of a classification model. Vapnik proved that the probability of the test error distancing

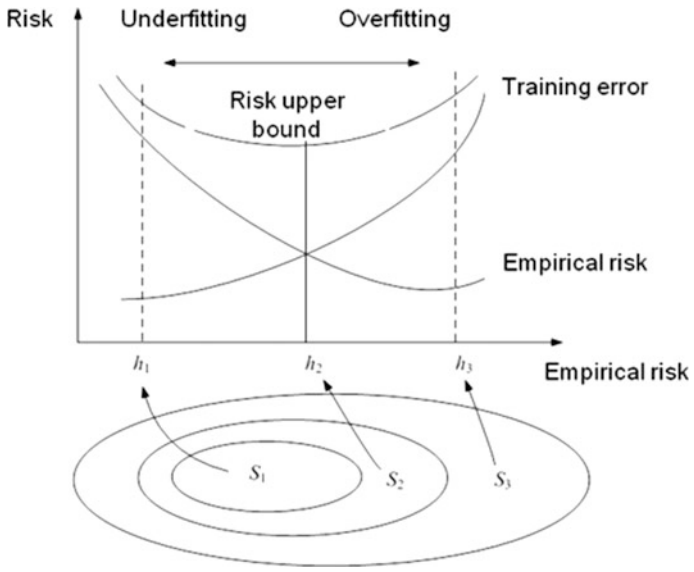


Fig. 3.6 Structural risk minimization

from an upper bound (on data that is drawn i.i.d. from the same distribution as the training set) is given by Eq. 3.1.33

$$R(w) \leq R_{emp}(w) + \varphi\left(\frac{h}{n}\right) \tag{3.1.33}$$

$$R(w) \leq R_{emp}(w) + \sqrt{\frac{h\left(\ln\left(\frac{2n}{h}\right)+1\right) - \ln\left(\frac{\eta}{4}\right)}{n}} \tag{3.1.34}$$

The Eq. (3.1.33) shows that the risk of machine is made up of two components; one is the empirical risk and the other is the training error. Replace the training error with the VC dimension and training samples. Then the Eq. (3.1.33) turns into Eq. (3.1.34).

When the training sample number is fixed, the training error will increase with the VC dimension, leading to the over-fitting phenomenon. Therefore, when designing the machine learning algorithm, the VC dimension should be kept as low as possible to obtain the small risk.

(e) Structural risk minimization

Structural risk minimization (SRM) is an inductive principle for model selection used for learning from finite training datasets. It describes a general model of capacity control and provides a trade-off between hypothesis space complexity (the VC dimension of approximating functions) and the quality of fitting the training data (empirical error), shown in Fig. 3.6. The procedure is outlined below.

1. Using a priori knowledge of the domain, choose a class of functions, such as polynomials of degree n, neural networks having n hidden layer neurons, a set of splines with n nodes, or fuzzy logic models having n rules.
2. Divide the class of functions into a hierarchy of nested subsets in order of increasing complexity, for example, polynomials of increasing degree.
3. Perform empirical risk minimization on each subset (this is essentially parameter selection).
4. Select the model in the series whose sum of empirical risk, and VC confidence is minimal.

B. Classification theory

The purpose of the SVM classification is to develop an efficient method to draw a hyperplane in the high-dimensional space.

(a) SVM characteristic

SVM method is based on the VC dimension and SRM principle. According to the limited sample information, balance the learning accuracy and the generalization ability so as to obtain a better algorithm. SVM has several following merits [17].

① SVM focuses on the limited sample number; its purpose is to obtain the optimal solution of with current information instead of the solution under an infinite sample number.

② SVM can find the global optimal solution which avoids the local optimal solution of some certain method such as neutral network.

③ SVM transform the nonlinear problem in low-dimensional space into the linear problem in high-dimensional space, which keeps the generalization ability of the algorithm.

Assume a binary classification problem; the sample number in dataset LS is n, LS = {(xi, yi), i = 1,2,. . ., n}, xi∈Rl. If the xi belongs to the first class, then the label is yi =1. If the xi belongs to the second class, then the label is yi = −1. The following will discuss the linear condition and the nonlinear condition.

(b) Linear condition

If the hyperplane is existed

$$\langle w \cdot x \rangle + b = 0 \tag{3.1.35}$$

which satisfies the Eq. 3.1.35 when $i = 1,2,\ldots, n$

$$\begin{cases} \langle w \cdot x_i \rangle + b \geq 1 & y_i = 1 \\ \langle w \cdot x_i \rangle + b \leq 1 & y_i = -1 \end{cases} \tag{3.1.36}$$

Then the dataset is linearly separable. $\langle w \cdot x_i \rangle$ in Eqs.(3.1.35) and (3.1.36) is inner production of the weight vector, where $w \in R^l$, $b \in R$ have been normalized to make the sample points satisfy the Eq.(3.1.37).

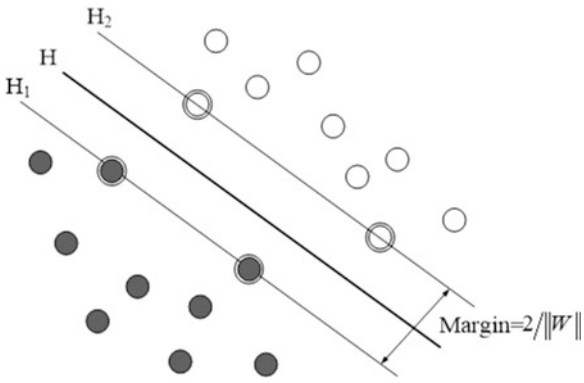


Fig. 3.7 The optimal classification boundary in two dimension space

$$y_i(\langle w \cdot x_i \rangle + b) \geq 1 \quad i = 1, 2, \ldots, n \tag{3.1.37}$$

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier, shown in Fig. 3.7. H represents the best classification line. The distance between H1 and H2 are taken as the margin.

The discriminant function of optimal hyperplane is shown as Eq. 3.1.38:

$$f(x) = \text{sgn}(\langle w \cdot x \rangle + b) \tag{3.1.38}$$

where sgn() is symbolic function.

The linear discriminant function is normally defined as $g(x) = \langle w \cdot x \rangle + b$. Compared to the geometric interval, the output interval will change, and the function output is called function interval. Therefore, the geometric interval should be optimized and minimize the norm of the weight vector, which means the function interval is set as 1. Assume w is weight vector, and the geometric interval can be calculated as follows:

$$\begin{cases} \langle w \cdot x^+ \rangle + b = 1 \\ \langle w \cdot x^- \rangle + b = -1 \end{cases} \tag{3.1.39}$$

At the same time, the weight w should be normalized in order to calculate the geometric interval. The geometric interval is the function interval of the classification machine:

$$\begin{aligned} margin &= \frac{1}{2}\left(\left\langle \frac{w}{\|w\|_2}\cdot x^{+}\right\rangle - \left\langle \frac{w}{\|w\|_2}\cdot x^{-}\right\rangle\right) \\ &= \frac{1}{2\|w\|_2^2}(\langle w\cdot x^{+}\rangle - \langle w\cdot x^{-}\rangle) = \frac{1}{\|w\|_2^2} \end{aligned} \tag{3.1.40}$$

Therefore, the geometric interval is $1/\|w\|_2^2$, where $\|w\|_2$ is geometric norm, marked as $\|w\|$.

Normalize the discriminant function, and make all samples of the labels satisfy the formula $|g(x)| \geq 1$. That's to make sample which is closest to the classification plane satisfy the formula $|g(x)| = 1$. Therefore, the maximum interval can be regarded as the minimum $\|w\|$, and then it should satisfy the Eq.(3.1.40) to make the classification right.

$$y_i(\langle w\cdot x_i\rangle + b) - 1 \geq 0 \quad i = 1, 2, \ldots, n \tag{3.1.41}$$

Therefore, the optimal classification plane is to make the Eq. (3.1.40) equality holds. Those samples are called as support vectors. The optimal hyperplane problem can be expressed as Eq. (3.1.42):

$$\begin{cases} \min\limits_{w,b,\xi} \dfrac{1}{2}\|w\|^2 \\ s.t. \quad y_i(\langle w\cdot x_i\rangle + b) \geq 1 \quad i = 1, 2, \ldots, n \end{cases} \tag{3.1.42}$$

This is a typical quadratic programming problem. The problem maximum interval classification is that it always generates a result without training error. When the dataset can't be totally separated, the maximum interval is negative. To solve this problem, the relaxation variable ξ_i, $i = 1, 2, \ldots, n$, is introduced. Then the Eq. (3.1.42) is transformed into Eq. (3.1.43)

$$\begin{cases} \min\limits_{w,b,\xi} \dfrac{1}{2}\|w\|^2 + \gamma\sum\limits_{i=1}^{n}\xi_i \\ s.t. \quad y_i(\langle w\cdot x_i\rangle + b) \geq 1 - \xi_i \\ \quad \xi_i \geq 0, \quad i = 1, 2, \ldots, n \end{cases} \tag{3.1.43}$$

where γ is punishment factor. Its value represents the punishment level. The Lagrange multiplier method is used to solve this problem.

$$\begin{cases} \max\limits_{\alpha,\beta}\min\limits_{w,b,\xi}\left\{L_P = \dfrac{1}{2}\|w\|^2 + \gamma\sum\limits_{i=1}^{n}\xi_i - \sum\limits_{i=1}^{n}\alpha_i[y_i(\langle w\cdot x_i\rangle + b) - 1 + \xi_i] - \sum\limits_{i=1}^{n}\beta_i\xi_i\right\} \\ s.t. \quad \alpha_i \geq 0, \quad \beta_i \geq 0 \end{cases} \tag{3.1.44}$$

where α_i and β_i are Lagrange multipliers. The dual form is shown as Eq. (3.1.45).

$$\begin{cases} \dfrac{\partial L_P}{\partial w} = w - \sum\limits_{i=1}^{n} y_i \alpha_i x_i = 0 \\ \dfrac{\partial L_P}{\partial \xi} = \gamma - \alpha_i - \beta_i = 0 \\ \dfrac{\partial L_P}{\partial b} = \sum\limits_{i=1}^{n} y_i \alpha_i = 0 \end{cases} \tag{3.1.45}$$

The dual optimal can be obtained by bringing the Eq. (3.1.45) into Eq. (3.1.44), shown as Eq. (3.1.46):

$$\begin{cases} \max\limits_{\alpha} \left[L_D = \sum\limits_{i=1}^{n} \alpha_i - \dfrac{1}{2} \sum\limits_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \langle x_i \cdot x_j \rangle \right] \\ s.t. \quad 0 \le \alpha_i \le \gamma, \quad \sum\limits_{i=1}^{n} y_i \alpha_i = 0 \end{cases} \tag{3.1.46}$$

According to the Karush-Kuhn-Tucker (KKT) condition, at the optimal point, the Lagrange multiplier is 0

$$\begin{cases} \alpha_i [y_i (\langle w \cdot x_i \rangle + b) - 1 + \xi_i] = 0 \\ \beta_i \xi_i = 0 \end{cases} \quad i = 1, 2, \ldots, n \tag{3.1.47}$$

For any standard support vector, because $0 < \alpha_i < C$, $\beta_i > 0$ according to the Eq. (3.1.48). Then for any standard support vector Xi, satisfy the Eq. (3.1.48)

$$y_i (\langle w \cdot x_i \rangle + b) = 1 \tag{3.1.48}$$

Then the parameter b can be obtained as follows:

$$b = y_i - \langle w \cdot x_i \rangle = y_i - \sum_{x_j \in SV} \alpha_j y_j \langle x_j \cdot x_i \rangle \quad x_i \in NSV \tag{3.1.49}$$

To make the calculation reliable, the parameter b is taken the mean value, which is shown as follows:

$$b = \frac{1}{N_{NSV}} \sum_{x_i \in NSV} \left[y_i - \sum_{x_j \in SV} \alpha_j y_j \langle x_j \cdot x_i \rangle \right] \tag{3.1.50}$$

where NNSV is the number of the support vectors.

(c) Nonlinear condition

When the training dataset is nonlinear, the nonlinear function is used to map the nonlinear data into high-dimensional linear space. Thus to solve the nonlinear problem, the classification hyperplane is constructed as Eq. (3.1.51).

$$\langle w \cdot \phi(x) \rangle + b = 0 \tag{3.1.51}$$

The discriminant function is Eq. (3.1.52).

$$y(x) = \text{sgn}(\langle w \cdot \phi(x) \rangle + b) \tag{3.1.52}$$

The optimal hyperplane problem can be described as Eq. (3.1.53):

$$\begin{cases} \min\limits_{w,b,\xi} \dfrac{1}{2}\|w\|^2 + \gamma \sum\limits_{i=1}^{n} \xi_i \\ s.t. \quad y_i(\langle w \cdot \phi(x_i) \rangle + b) \geq 1 - \xi_i \\ \quad \xi_i \geq 0, \quad i = 1, 2, \ldots, n \end{cases} \tag{3.1.53}$$

The obtained dual problem is shown as Eq. (3.1.54):

$$\begin{cases} \max\limits_{\alpha} \left[\begin{aligned} L_D &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i.j=1}^{n} y_i y_j \alpha_i \alpha_j \langle \phi(x_i) \cdot \phi(x_j) \rangle \\ &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i.j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \end{aligned} \right] \\ s.t. \quad 0 \leq \alpha_i \leq \gamma, \quad \sum\limits_{i=1}^{n} \alpha_i y_i = 0 \end{cases} \tag{3.1.54}$$

where $K(x_i, x_j)$ is the kernel function. The discriminant function is shown as Eq. (3.1.55):

$$y(x) = \text{sgn}\left[\sum_{X_i \in SV} y_i \alpha_i K(x_i, x) + b \right] \tag{3.1.55}$$

where the threshold value b is shown as Eq. (3.1.56):

$$b = \frac{1}{N_{NSV}} \sum_{x_i \in NSV} \left[y_i - \sum_{x_j \in SV} \alpha_j y_j K(x_j, x_i) \right] \tag{3.1.56}$$

Form the Eq. (3.1.54), (3.1.55), and (3.1.56), we can conclude that while calculating the optimal problem and the discriminant function, only the kernel function is

Table 3.1 Several common kernel functions

Kernel function	Expression
Linear kernel function	$K(x, x_i) = x^T x_i$
Polynomial kernel function of D order	$K(x, x_i) = (x^T x_i + 1)^d$
Gauss radial basis kernel function	$K(x, x_i) = \exp\left[\frac{-x(x-x_i)^2}{\sigma^2}\right]$
Multilayer perceptron kernel function	$K(x, x_i) = \tanh(\kappa x^T x_i + \theta)$
B-spline kernel function	$K(x, x_i) = B_{2n+1}(x - x_i)$
Sheet-spline kernel function	$K(x, x_i) = \|x - x_i\|^{2n+1}$
Multiple quadric kernel functions	$K(x, x_i) = \sqrt{\|x - x_i\|^2 + c^2}$
Inverse multiple quadric kernel function	$K(x, x_i) = \frac{1}{\sqrt{\|x-x_i\|^2 + c^2}}$
Trigonometric polynomial kernel function of D order	$K(x, x_i) = \frac{\sin\left(d+\frac{1}{2}\right)(x-x_i)}{\sin\left(\frac{(x-x_i)}{2}\right)}$

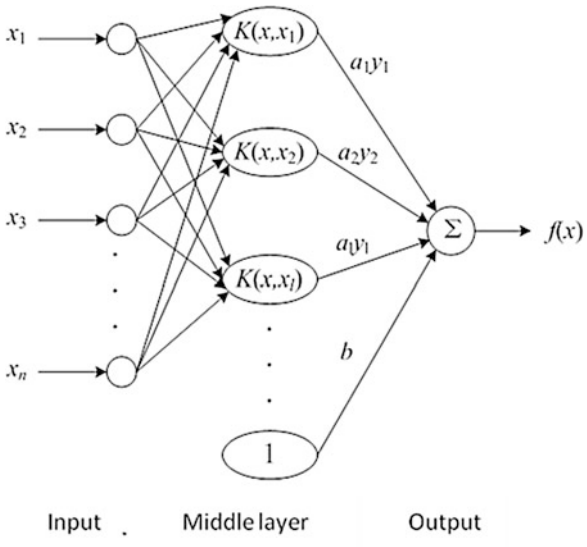


Fig. 3.8 SVM classification

needed to be figured out. Therefore, the dimension disaster is avoided. Several common kernel functions are shown in Table 3.1.

Above all, SVM is similar to neutral network; the output is the linear combination of the middle layer points. Every point in the middle layer corresponds to a support vector, shown in Fig. 3.8.

2. Least square SVM

Although the SVM can adapt the nonlinear, high-dimension, and small-number samples, it also has problems such as high complexity, large scale, and so on. Therefore, Suykens proposed the least squares support vector machine to deal with this problem [18].

LSSVMs are least squares versions of SVMs, which is a set of related supervised learning methods that analyze data and recognize patterns and which are used for classification and regression analysis. In this version one finds the solution by solving a set of linear formulas instead of a convex quadratic programming (QP) problem for classical SVMs. Least squares SVM classifiers were proposed by Suykens and Vandewalle [19]. LSSVMs are a class of kernel-based learning methods.

Same as the above binary classification, LSSVM can describe the optimal problem as Eq. (3.1.57) [20]:

$$\begin{cases} \min\limits_{w,b,\xi} J(w,b,\xi) = \dfrac{1}{2}w^{\mathrm{T}}w + \dfrac{1}{2}\gamma\sum\limits_{i=1}^{n}\xi_i^2 & i = 1,\ 2,\cdots,n \\ s.t.y_i[w^{\mathrm{T}}\phi(x_i) + b] = 1 - \xi_i \end{cases} \tag{3.1.57}$$

where J is the objective function.

W is the weight vector.
B is the threshold value.
ξ is the relaxation variable.
γ is the punishment factor.
$\phi(\cdot)$ is the nonlinear mapping.

The corresponding Lagrange function is shown as Eq.(3.1.58):

$$L(w,b,\xi,\alpha) = J(w,b,\xi) - \sum_{l=1}^{N}\alpha_i\left\{y_i\left[w^T\phi(x_i) + b\right] + \xi_i - 1\right\} \tag{3.1.58}$$

where α_i is Lagrange multiplier; combing the KKT condition (3.1.59), the Eq. (3.1.60) can be obtained:

$$\left.\begin{array}{c} \dfrac{\partial L}{\partial w} = 0 \rightarrow w = \sum\limits_{i=1}^{n}\alpha_i\phi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum\limits_{i=1}^{n}\alpha_i = 0 \\ \dfrac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma\xi_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i\left[w^{\mathrm{T}}\phi(x_i) + b\right] + \xi_i - 1 = 0 \end{array}\right\} \quad i = 1,2,\cdots,n \tag{3.1.59}$$

$$\begin{bmatrix} \mathbf{I} & 0 & 0 & -\mathbf{Z}^T \\ 0 & 0 & 0 & -\mathbf{y}^T \\ 0 & 0 & \gamma\mathbf{I} & -\mathbf{I} \\ \mathbf{Z} & \mathbf{y} & \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ \overline{\xi} \\ \overline{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1_n \end{bmatrix} \tag{3.1.60}$$

The Eq. (3.1.60) can also be written as Eq. (3.1.61):

$$\begin{bmatrix} 0 & \mathbf{y}^T \\ y & \mathbf{Z}\mathbf{Z}^T + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \overline{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ 1_n \end{bmatrix} \tag{3.1.61}$$

$$\text{where } \mathbf{Z} = [\phi(x_1), \phi(x_2), \ldots, \phi(x_n)]^T$$

$$\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$$

$$1_n = [1, 1, \ldots, 1]_{1\times n}^T$$

$$\overline{\xi} = [\xi_1, \xi_2, \ldots, \xi_n]^T$$

$$\overline{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T$$

The inner product of the nonlinear function can be replaced by the kernel function K(xi, xj) under the condition Eq. (3.1.62):

$$\Omega = \mathbf{Z}\mathbf{Z}^T \tag{3.1.62}$$

$$\Omega_{ij} = y_i y_j \phi(x_i)^T \phi(x_j) = y_i y_j K(x_i, x_j) \tag{3.1.63}$$

Then the LSSVM classification can be expressed as Eq.(3.1.64):

$$f(x) = \text{sgn}\left[\sum_{i=1}^{n} \alpha_i K(x_i, x) + b\right] \tag{3.1.64}$$

Normally, the kernel function of the LSSVM is Gauss radial basis kernel function.

3. Multi-classification support vector machine

SVM is designed for binary classification problem, but there are a lot of multi-classification problems in practice. To make SVM deal with multi-classification problems, a lot of researches have been done.

The dominant approach for doing so is to reduce the single multiclass problem into multiple binary classification problems. Common methods for such reduction include:

① Building binary classifiers which distinguish between one of the labels and the rest (one-versus-all) or between every pair of classes (one-versus-one) Classification of new instances for the one-versus-all case is done by a winner-takes-all strategy, in

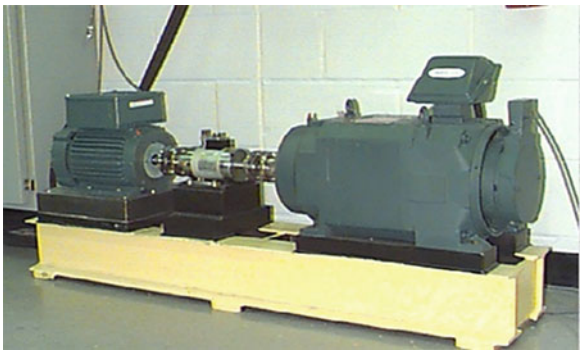


Fig. 3.9 Rolling bearing test rig

which the classifier with the highest output function assigns the class (it is important that the output functions be calibrated to produce comparable scores). For the one-versus-one approach, classification is done by a max-wins voting strategy, in which every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by one vote, and finally the class with the most votes determines the instance classification.

② Directed acyclic graph SVM (DAGSVM)

A directed acyclic graph (DAG) is a graph whose edges have an orientation and no cycles. A rooted DAG has a unique node such that it is the only node which has no arcs pointing into it. A rooted binary DAG has nodes which have either 0 or 2 arcs leaving them.

SVM will face the problem of wrong classification during the one against one classification. The training phase of DAGSVM is the same as the one-against-one method by solving binary SVMs. However, in the testing phase, it uses a rooted binary-directed acyclic graph which has internal nodes and leaves. Each node is a binary SVM of th and th classes. Given a test sample, starting at the root node, the binary decision function is evaluated. Then it moves to either left or right depending on the output value. Therefore, we go through a path before reaching a leaf node which indicates the predicted class. An advantage of using a DAG is that some analysis of generalization can be established. There are still no similar theoretical results for one-against-all and one-against-one methods yet. In addition, its testing time is less than the one-against-one method (Fig. 3.9).

2. Experiment and Analysis

1. Data Acquisition

A. Laboratory Data Acquisition

The vibration data in this book are provided by the Case Western Reserve University [12]; the test rig is shown in the Fig. 3.5. The test stand consists of a 2 hp. motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The test bearings support the motor shaft. Single-point faults were introduced to the test bearings using electro-discharge machining

with fault diameters of 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils (1 mil = 0.001 inches). SKF bearings were used for the 7, 14, and 21 mils diameter faults, and NTN equivalent bearings were used for the 28 mil and 40 mil faults. Drive-end and fan-end bearing specifications, including bearing geometry and defect frequencies, are listed in the bearing specifications. Vibration data were collected using accelerometers, which were attached to the housing with magnetic bases. Accelerometers were placed at the 12 o'clock position at both the drive end and fan end of the motor housing. During some experiments, an accelerometer was attached to the motor supporting base plate as well. Vibration signals were collected using a 16 channel DAT recorder and were post processed in a Matlab environment. All data files are in Matlab (*.mat) format. Digital data was collected at 12,000 samples per second, and data was also collected at 48,000 samples per second for drive-end race faults. Speed and horsepower data were collected using the torque transducer/encoder and were recorded by hand.

Outer raceway faults are stationary faults; therefore, placement of the fault relative to the load zone of the bearing has a direct impact on the vibration response of the motor/bearing system. In order to quantify this effect, experiments were conducted for both fan- and drive-end bearings with outer raceway faults located at 3 o'clock (directly in the load zone), at 6 o'clock (orthogonal to the load zone), and at 12 o'clock.

Data in this book come from the 205-2RS JEM SKF rolling bearing. The motor load is 3 hp. and the rotating speed is 1730 rpm.

B. Operation Condition Simulation

(a) Environment noise analysis

The main research object of this book is rolling bearing of rail transportation train. So the noise should be added to simulate the real working condition. Train rolling bearings are mainly affected by two kinds of noise including the white noise and the impulsive noise.

The white noise mainly comes from the three following source [21].

① Vibration noise coming from the traction motor, bogie and gears, and so on, which are named as background noise

② Vibration noise coming from the poor lubrication, improper assemble, and poor material of rolling bearings

③ Vibration noise coming from the crash between the wheel and the track during the train operation

The impulsive noise mainly comes from three following source [22].

① The impulsive noise generated by the train when passing through switches

② The impulsive noise generated by the damage of the wheel, such as the wheel flats impact

③ The impulsive noise generated by the electromagnetic wave such as the pantograph electromagnetic wave [23]

Based on the above analysis, white noise and impulsive noise were added into the collected data to simulate the practical operation condition.

(b) Noise signal simulation:

① White noise signal

Based on the above analysis of the noise environment, each of the three sources of white noise is composed of multiple sources of interference, so the whole white noise signal is also the sum of multiple interference sources, and the number of interference sources is quite large. By the Chebyshev large number theorem, the noise data after the superposition of an infinite number of interference sources is bound to conform to the Gauss distribution. Therefore, in order to simulate the white noise signal of the actual working condition to the maximum extent, we also select the high-intensity Gauss white noise signal with the amplitude of the original normal vibration signal amplitude 100% as the interference signal to fully excavate the algorithm performance.

② Impulsive noise simulation

Among the three impulsive noise sources, the third one which is electromagnetic interference to the sensor electrical signal. This interference will not have influence on the device itself. Therefore, only the first two sources are considered. Aiming at simulating the practical operation conditions of rolling bearings, assuming the wheel diameter is 840 mm and the operation speed is 50 km/h, the impulsive noise occurs during every circle. Thus the interval frequency between contiguous impulsive noise is 5.62 Hz. Considering the weld of the rail track, the interval frequency between contiguous impulsive noise under the assumption that the rail is welded every 25 meters. Considering the random distribution of the impulsive noise, the impulsive is set randomly without certain frequency interval. The number of the intervals is controlled between 55 and 60 in 10 seconds.

For the signal impulsive noise, the symmetric α stable distribution is always used for model construction [24, 25]; the feature function is shown as Eq. (3.1.65):

$$\varphi(t) = e^{-\gamma|t|^{\alpha}} \tag{3.1.65}$$

where α is feature index, $0 < \alpha < 2$ and

γ is dispersion coefficient, $\gamma > 0$.

The smaller α is, the larger dispersion trailing is, and the impulsive feature will be more obvious, when the feature index $\alpha = 2$ which is Gaussian distribution; if feature index $\alpha = 1$, which is Cauchy distribution; and if $0 < \alpha < 2$ and the dispersion coefficient $\gamma = 1$, which is symmetric distribution. According to the narrow-band feature of the impulsive, the feature index α is valued between 0.2 and 0.4, in which the noise amplitude is ten times larger than the normal vibration signal.

③ Compound noise signal

In order to simulate the noise interference environment as close to the practical working condition as possible, the white noise signal and the shock noise signal are superimposed to get the compound noise signal. The signal obtained by the superposition of the original signal and the compound noise signal in the laboratory environment is used to simulate the vibration data collected in the actual working environment.

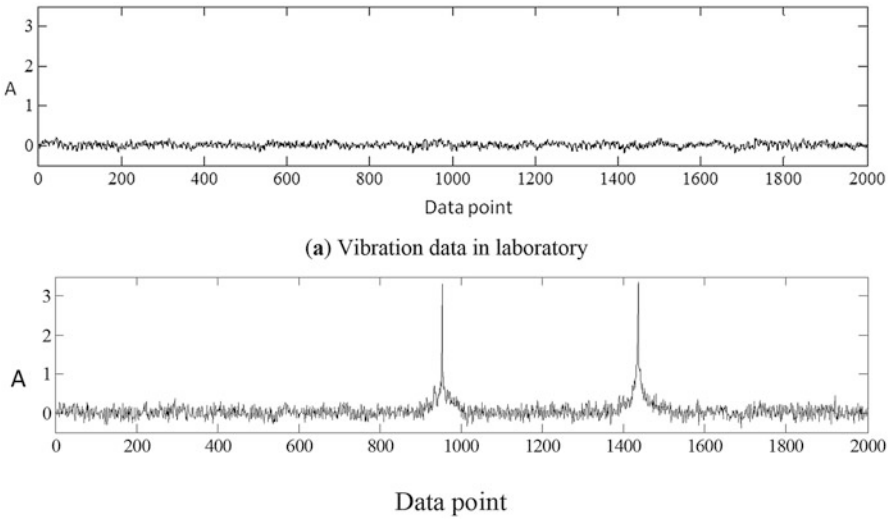


Fig. 3.10 Vibration data under certain conditions. (**a**) Vibration data in laboratory. (**b**) Vibration data after noise simulation

Through field investigation, it is found that in city rail train vibration sensor of rolling bearings, acceleration sensor sampling frequency is about 10 k Hz. The sampling point is set at the load end, so laboratory data with 12 k Hz sampling

frequency are taken as the original vibration data in the simulation of practical working condition. Figure 3.10 shows the original vibration data in the laboratory environment and the simulated working condition data after the composite noise is superimposed.

2. Experiment Preparation

A. Experiment Grouping

To verify the performance of the proposed method, the experiment data include the laboratory data and the simulated operation condition data.

The experiment is divided into Group 1 and Group 2. Every group has different test according to the vibration data. Besides, to verify the performance, early fault data are chosen for the experiment.

Group 1: Laboratory environment

Test 1.1: Sample frequency (Fs) 12 k Hz, load-end data, safety region estimation of normal condition and fault conditions

Test 1.2: Fs-48 k Hz, drive-end data, safety region estimation of normal condition and fault conditions

Test 1.3: Fs-12 k Hz, load-end data, safety region estimation of normal condition and multi-fault conditions

Test 1.4: Fs-48 k Hz, drive-end data, safety region estimation of normal condition and multi-fault conditions

Group 2: Simulated practical working condition

Test 2.1: Fs-12 k Hz, load-end data, safety region estimation of normal condition and fault conditions

Test 2.2: Fs-12 k Hz, load-end data, safety region estimation of normal condition and multi-fault conditions

B. Parameter Selection

Main parameter selections include the following.

① Vibration data segment interval

Every round of the rolling bearing is taken as a data segment. The rolling bearing rotates 288 rounds. There are 416 points when the Fs is 12 k Hz and 1666 points when the Fs is 48 k Hz.

② Feature index for the state identification

RMS, energy, Shannon entropy, and energy moment are chosen for the simulation experiment.

③ LSSVM kernel function: Gauss radial basis kernel function

DAGSVM classification rule

④ The ratio of the training data number to the test data number

C. Performance Assessment Index

Detection rate, false-alarm rate classification rate, and Fleiss' kappa statistic are chosen for the performance assessment.

① Detection rate (DR)

DR is defined as the ratio of the correctly detected sample number to the total number of the certain sample Eq. (3.1.66):

$$DR = \frac{\text{Sample number of detected certain classification}}{\text{The total sample number of the certain classification}} \tag{3.1.66}$$

② False alarm rate (FAR)

FAR is defined as the ratio the incorrectly detected sample number to the total number of the certain sample (3.1.67):

$$FAR = \frac{\text{Sample number of certain classification not belonging this classification}}{\text{Total sample number not belonging this classification}} \tag{3.1.67}$$

③ Classification rate (CR)

CR is defined as correctly detected sample number to the total number of the whole sample (3.1.68).

Table 3.2 Safety region identification result

	12 k Hz load-end data				48 k Hz drive-end data			
	RMS	Energy	Shannon entropy	Energy moment	RMS	Energy	Shannon entropy	Energy moment
DR_{normal}	0.9113	0.9610	0.9103	0.9615	0.8965	0.9439	0.9201	0.9589
DR_{fault}	0.9431	0.9499	0.9501	0.9487	0.9385	0.9405	0.9651	0.9459
CR	0.9399	0.9508	0.9400	0.9502	0.9398	0.9401	0.9499	0.9489
FK	0.8858	0.8969	0.8898	0.8950	0.8861	0.8801	0.8997	0.8987

$$CR = \frac{\text{Number of corrected classified sample}}{\text{Total sample number}} \tag{3.1.68}$$

④ Fleiss's kappa statistic (FK)

FK statistic is used to evaluate the coherence between the predictive output and the label. When FK statistic is more than 0.8, the predictive output and the label have high coherence degree.

In addition, to compare the result between the laboratory environment and the simulated operation conditions, the float percentage is also used, shown as Eq. (3.1.69):

$$I_{float} = \frac{I_{ND} - I_{RD}}{I_{RD}} \times 100\% \tag{3.1.69}$$

where I_{float} is the float percentage of some certain index,

I_{RD} is the whole index of the data in laboratory environment

I_{ND} is the whole index of the data-simulated operation condition environment.

When the float percentage is more than zero, the index value increases. When the float percentage is less than zero, the index value decreases. It is noted that the change of FAR can be shown by the change of the DR.

3. Result analysis

A. Result under laboratory environment:

① Test 1.1and Test 1.2 safety region identification results are shown in Table 3.2.

In the result of the 12 k Hz load-end data, classification result based on energy features is the best, with correct rate 0.9508, and the corresponding FK value is 0.8969, followed by the energy moment feature, with correct rate 0.9502, and the corresponding FK value is 0.8950. The third one is the result based on the Shannon entropy, with correct rate 0.9400 and the corresponding FK value 0.8898. The worst one is based on the RMS, with correct rate 0.9399, and the corresponding FK is 0.8858.

The classification of energy moment feature based on the correct rate of 0.9502, the FK value is 0.8950; again for classification Shannon based on entropy, the correct classification rate is 0.9400, and FK value is 0.8898; the worst performance

is the classification based on RMS feature, the correct classification rate is 0.9399, and the corresponding FK value is 0.8858. It can be seen that the feature extraction method based on energy is the best performance, and the performance of RMS feature extraction is the worst.

Seen from the detection rate of the "normal" and "fault," classification results show that features based on the RMS and Shannon entropy perform better when detecting the fault state. Classification results show that features based on the RMS and Shannon entropy perform equally when detecting those two states. Meanwhile, under the condition that the number of two samples is largely different, features based on the energy and energy moment show its adaptation and superiority.

In the result of the 48 k Hz drive-end data, classification result based on Shannon entropy has the best performance, with the value 0.9499, and the corresponding FK value is 0.8997, followed by the energy moment, whose value is 0.9489, and the corresponding FK value is 0.8987. Result based on the RMS is not satisfying with 0.9398 and the corresponding FK value 0.886.

Seen from the detection rate of the "normal" and "fault," classification results show that features based on the RMS and Shannon entropy perform better when detecting the fault state. Classification results show that features based on the Shannon moment perform equally when detecting those two states. Meanwhile, under the condition that the number of two samples is largely different, features based on the energy and energy moment show its adaptation and superiority.

No matter what sample frequency is chosen, the difference of four-feature index is small. The maximum difference of classification correction rate under 12 k Hz sampling frequency is 0.0103. The maximum difference of classification correction rate under 48 k Hz sampling frequency is 0.0101.

All in all, from the perspective of the classification correction rate, mean value of four-feature index under 12 k Hz sampling frequency is 0.9452. Mean value of four-feature index under 48 k Hz sampling frequency is 0.9447. The difference is small.

Besides, to show the optimal classification boundary, named safety region boundary, several figures are given. It should be noted that the common classification plane is high dimension. Figure 3.11 and Fig. 3.12 are the classification plane under Test 1.1and Test 1.2, respectively.

② Test 1.3 of multistate identification result is shown in Table 3.3 and Table 3.4, respectively.

The classification correctness and FK values of the subclassifiers in the DAGSVM of the 12 k Hz load-end data of the Test 3.1.3 are shown in Table 3.3. Seen from Table 3.3, in each of the two subclassifiers classification, better classification results can be obtained from three subclassifiers under normal or other three faults. CR values are higher than 0.9 no matter which kind of feature extraction method is used. CR values are higher than 0.9; FK value is very close to 0.9. The worst subclassifier is "roller fault VS outer race fault" subclassifiers. The highest CR value was only 0.8142 based on the four features, followed by 0.8103, 0.7903, and 0.7504, while the FK value was not higher than 0.7 of the four features. The performance of "inner race fault VS outer race fault" subclassifier is slightly better than "roller fault VS outer race fault" subclassifiers, but the results are not satisfying.

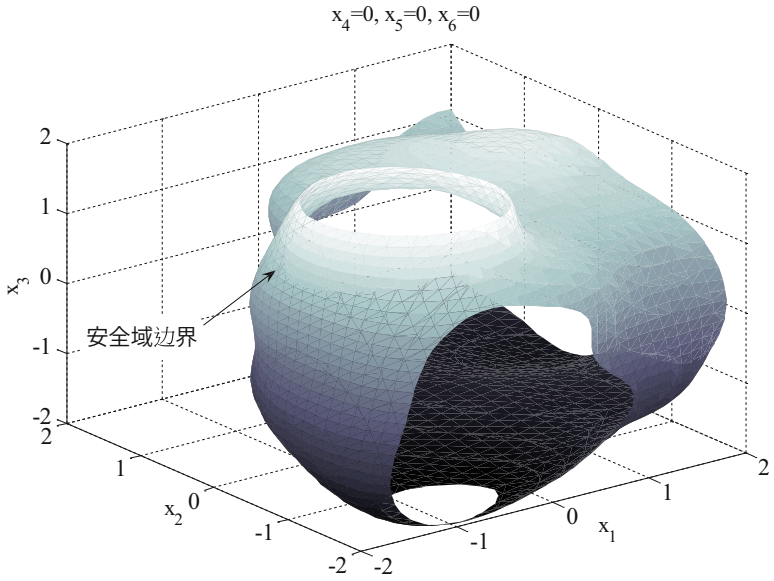


Fig. 3.11 Test 1.1 Safety region boundary based on the energy feature

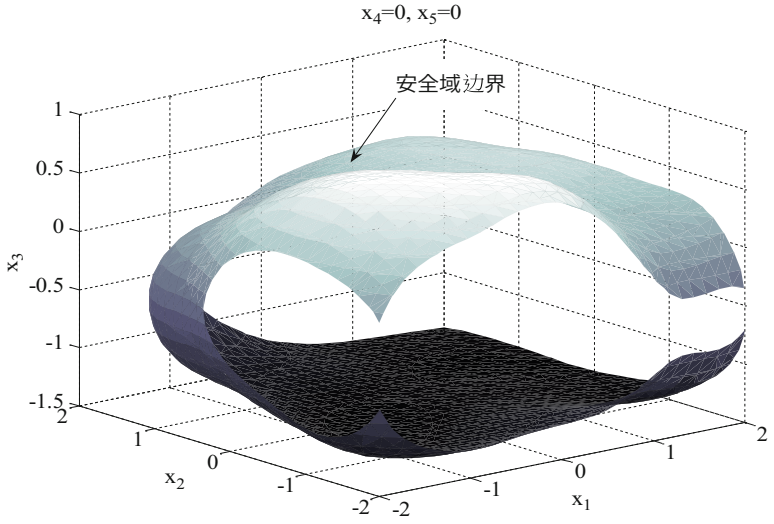


Fig. 3.12 Test 1.2 Safety region boundary based on the energy feature

Only one CR value is larger than 0.85, and its FK value was lower than 0.8. The "roller fault VS inner fault" subclassifier is not as good as three other classifiers. The CR value is between 0.85 and 0.9; FK value is about 0.8.

Table 3.3 Test 1.3 DAGSVM multistate identification result

		12 k Hz drive-end data			
		RMS	Energy	Shannon entropy	Energy moment
Normal VS outer race fault	CR	0.9140	0.9260	0.9258	0.9260
	FK	0.8981	0.9220	0.9215	0.9220
Roller fault VS outer race fault	CR	0.7903	0.8103	0.7504	0.8142
	FK	0.6499	0.6903	0.5695	0.6984
Inner race fault VS outer race fault	CR	0.8382	0.8582	0.7863	0.8342
	FK	0.7459	0.7855	0.6420	0.7365
Normal VS roller fault	CR	0.9059	0.9059	0.9180	0.9059
	FK	0.8819	0.8819	0.9059	0.8819
Normal VS inner race fault	CR	0.9360	0.9260	0.9220	0.9260
	FK	0.9160	0.9220	0.9140	0.9220
Roller fault VS inner race fault	CR	0.8739	0.8939	0.8538	0.8578
	FK	0.8175	0.8577	0.7776	0.7854

Table 3.4 Test 1.3 Four kinds of state identification result

	12 k Hz load-end data			
	RMS	Energy	Shannon entropy	Energy moment
DR_{normal}	0.9200	0.9200	0.9028	0.8993
FAR_{normal}	0.0857	0.0846	0.0801	0.0802
$DR_{Roller\ fault}$	0.8203	0.8513	0.8513	0.8616
$FAR_{Roller\ fault}$	0.1040	0.0960	0.1246	0.1177
$DR_{Inner\ race\ fault}$	0.8926	0.9166	0.8549	0.8755
$FAR_{Inner\ race\ fault}$	0.1109	0.1075	0.1178	0.1120
$DR_{outer\ race\ fault}$	0.7835	0.8005	0.7459	0.8005
$FAR_{outer\ race\ fault}$	0.1075	0.0960	0.1063	0.0915
CR	0.8540	0.8620	0.8385	0.8591
FK	0.8319	0.8459	0.8113	0.8388

Compared with the classification results in Table 3.3 based on each feature, we can see that the CR and FK values of the classifier based on energy and energy moments feature are better than those of RMS and Shannon entropy features, and the performance indexes of each subclass based on energy feature are optimal. Take "normal VS outer race fault" subclassifier as an example; the CR and FK were 0.9260 and 0.9220 based on the energy feature. CR and FK values were 0.9140 and 0.8981 based on RMS. The worst subclassifier is "roller fault VS outer race fault." CR and FK value based on energy are 0.8103 and 0.6903. CR and FK value based on Shannon entropy is 0.7504 and 0.5695.

The comprehensive results of four-state identification of the normal, roller, inner, and outer race faults of Test 1.3 are shown in Table 3.4. It is the first prior to pay attention to the detection rate and the error rate index of the four states. From

Table 3.4, we can see that no matter which feature extraction method is applied, the DR value of "normal" state is the highest among the four states, and the highest DR value of four different characteristics is 0.9200; the lowest one is 0.9028, followed by the "inner race fault" and "roller fault." The same as Test 2.1, the DR value of the "outer race fault" state is the lowest, the highest DR value of the four different feature is 0.8005, and the lowest one is 0.7459. This result shows that the sample points in the "outer race fault" are not properly classified to the number of the sample points of the class. From the FAR index, the four-classification result, "roller fault" and "inner race fault" of the FAR value, is relatively large. Among four-feature extraction methods, "roller fault" is the highest with FAR value 0.1246; the lowest is 0.0960, which is close to 0.1; and FAR value of four inner race fault is greater than 0.1. This indicates that the number of two states of the sample points of other states is misclassified as the "roller fault" and "inner race fault." Secondly, the classification result of "outer race fault," of which two of its four FAR values, is more than 0.1, and the other two are more than 0.09. The FAR value of the "normal" classification is the lowest, all over 0.09. Then we focus on the overall performance indicators CR and FK values of the four states. It can be seen that in the four classification results based on different features, the largest CR value is 0.8620, the smallest is 0.8385, the largest FK value is 0.8459, the smallest is 0.8113, and the overall classification accuracy is not more than 0.9.

Comparing classification results in Table 3.3, the CR value is the highest based on energy feature, the same with the FK value, followed by the Shannon moment, whose CR value and FK value are 0.8591 and 0.8388, respectively. The performance is worst when based on the Shannon entropy. The CR and FK values are 0.8385 and 0.8113.

Figures 3.13, 3.14, 3.15, and 3.16 show the identification results of multistate data based on the four features of RMS, energy, Shannon entropy, and energy moments after the decomposition of Test 1.3 LMD, respectively. In the four figures, from the

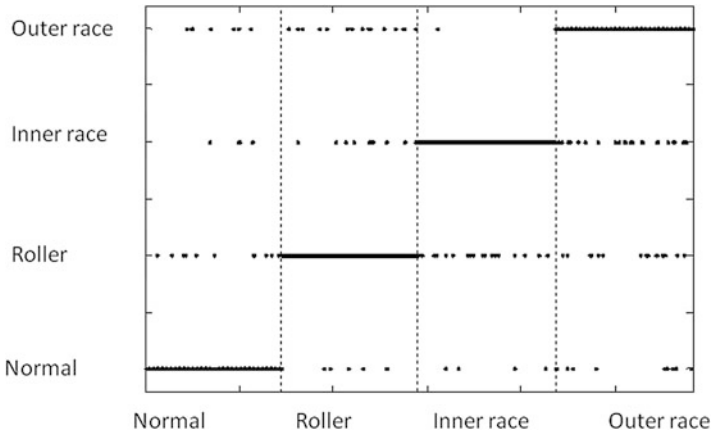


Fig. 3.13 Test 1.3 Multistate identification result based on RMS

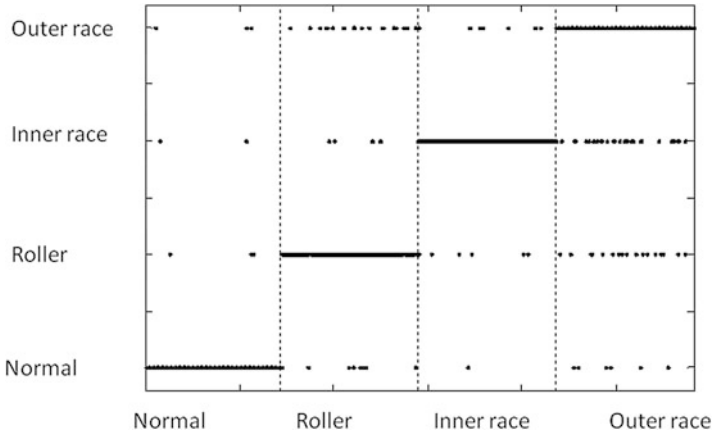


Fig. 3.14 Test 1.3 Multistate identification result based on energy

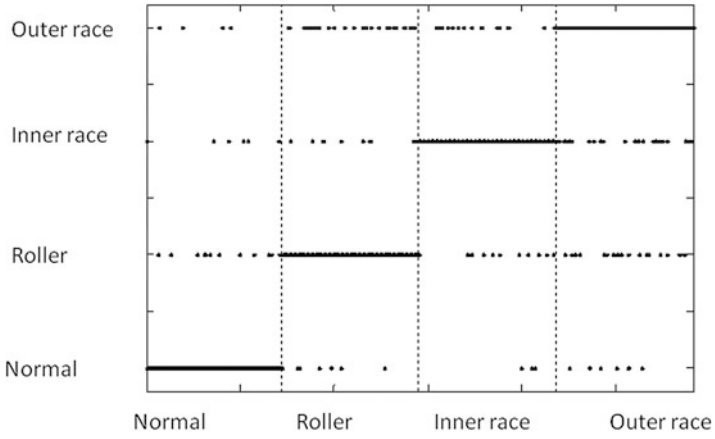


Fig. 3.15 Test 1.3 Multistate identification result based on Shannon entropy

vertical view, no matter what kind of feature extraction method is used, the "outer fault" is mostly misclassified as other classes, which have low detection rate, corresponding to low "outer race fault" DR value, followed by the "roller fault" and "the inner race fault." But the "normal" class is least misclassified sample. From the lateral view, no matter what kind of feature extraction method is used, more samples are misclassified to the "inner race fault" and "roller fault," followed by the "outer race fault."

By comparing four figures, the corresponding results of all kinds of sample point in Fig. 3.14 and Fig. 3.16 are slightly better than those in Fig. 3.15. That is, the performance of multi-classifiers based on energy and energy moments is better than that of multiple classifiers based on Shannon entropy. The results shown in the figures are consistent with the results in Tables 3.6 and 3.7.

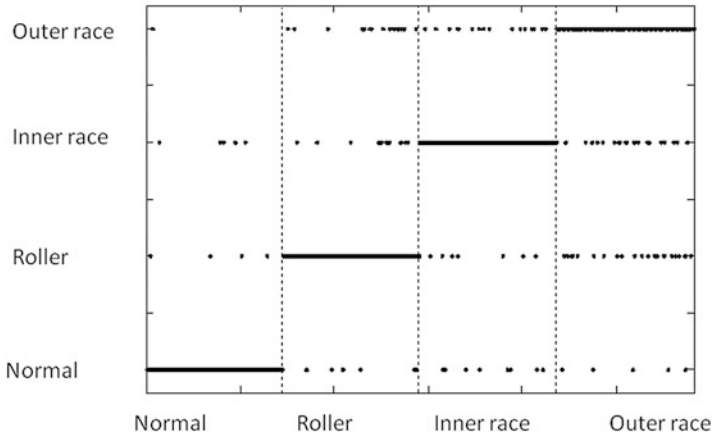


Fig. 3.16 Test 1.3 Multistate identification result based on energy moment

Table 3.5 Test 3.1.4 DAGSVM multistate identification result

		48 k Hz drive-end data			
		RMS	Energy	Shannon entropy	Energy moment
Normal VS outer race fault	CR	0.8536	0.8610	0.8610	0.8800
	FK	0.8492	0.8421	0.8421	0.8741
Roller fault VS outer race fault	CR	0.8686	0.8686	0.8724	0.8648
	FK	0.8572	0.8572	0.8648	0.8496
Inner race fault VS outer race fault	CR	0.8648	0.8724	0.8762	0.8724
	FK	0.8496	0.8648	0.8724	0.8648
Normal VS roller fault	CR	0.8486	0.8686	0.8572	0.8648
	FK	0.8402	0.8572	0.8345	0.8496
Normal VS inner race fault	CR	0.8724	0.8801	0.8648	0.8801
	FK	0.8648	0.8740	0.8496	0.8736
Roller fault VS inner race fault	CR	0.8762	0.8760	0.8762	0.8724
	FK	0.8724	0.8714	0.8724	0.8648

③ Test 1.4 multistate identification result is shown in Table 3.5 and Table 3.6.

The classification correctness and FK values of the subclassifiers in the DAGSVM of the 48 k Hz driving-end data of the Test 1.4 are shown in Table 3.5. Among six binary classifiers, the classification results are similar. No matter what feature extraction method is used, each subclassifier CR value of the nearly distributed between 0.86 and 0.90. Only "normal VS roller fault" subclassification characteristics of RMS based on the CR value is 0.8486. The FK values were both greater than 0.8, which are between 0.84 and 0.88. The results show that the performance of the six subclassifiers is good and balanced.

Table 3.6 Test 1.4 Four kinds of state identification result

	48 k Hz drive-end data			
	RMS	Energy	Shannon entropy	Energy moment
DR_{normal}	0.8543	0.8509	0.8431	0.8612
FAR_{normal}	0.0629	0.0599	0.0624	0.0599
$DR_{Roller\ fault}$	0.8613	0.8682	0.8682	0.8647
$FAR_{Roller\ fault}$	0.0633	0.0613	0.0619	0.0606
$DR_{Inner\ race\ fault}$	0.8750	0.8748	0.8716	0.8740
$FAR_{Inner\ race\ fault}$	0.0687	0.0653	0.0650	0.0707
$DR_{outer\ race\ fault}$	0.8613	0.8681	0.8750	0.8681
$FAR_{outer\ race\ fault}$	0.0648	0.0615	0.0592	0.0593
CR	0.8730	0.8755	0.8735	0.8773
FK	0.8599	0.8654	0.8653	0.8697

The comprehensive results of four-state identification of the normal, roller, inner, and outer race faults of Test 1.4 are shown in Table 3.6. From Table 3.6, we can see that no matter which feature extraction method is applied, the DR value of the "inner race fault" state is the highest in the four features, and the DR highest value is 0.8750 in the four different features; the lowest one is 0.8716, which are greater than 0.87. Followed by the "outer race fault," the DR highest value is 0.8750 in the four different features, and the lowest one is 0.8613, which are greater than 0.86. The followed results are "roller fault" and "normal," among which the DR values of four different features of roller fault are all greater than 0.86, while DR value of the "normal" state are all greater than 0.84, and the difference between them is not large. Seen from the FAR index, the difference between the "inner race fault" and "outer race fault" is large. The FAR of the "inner race fault" is largest with value 0.0707. Followed by the "outer race fault," the highest FAR value is 0.65.

The results show that the number of two states of the sample points in other states is misclassified for "inner race fault" and "outer race fault." The FAR values of the two classifications of "normal" and "roller faults" are relatively low, of which the maximum value of "normal" FAR is 0.0629 and the minimum value is 0.0599. The maximum value of FAR of "outer race fault" is 0.0648, and the minimum value is 0.0592. Generally speaking, the classification effect of the four-state points is very small, and the classification accuracy is high.

Comparing classification results with different features based on Table 3.6, the high CR and FK value is based on the energy moment, which are 0.8773 and 0.8697, respectively, followed by the energy features. The worst is the RMS value, with the CR value and FK value 0.8730 and 0.8599, respectively.

Figure 3.17–3.20 show the identification results of multistate data based on the four features of RMS, energy, Shannon entropy, and energy moments after the LMD decomposition of Test 2.4, respectively.

In the four pictures, from the vertical view, no matter what kind of feature extraction method is used, the "normal" class in the sample is mostly misclassified as other class. The detection rate is the lowest, corresponding to the "normal" low

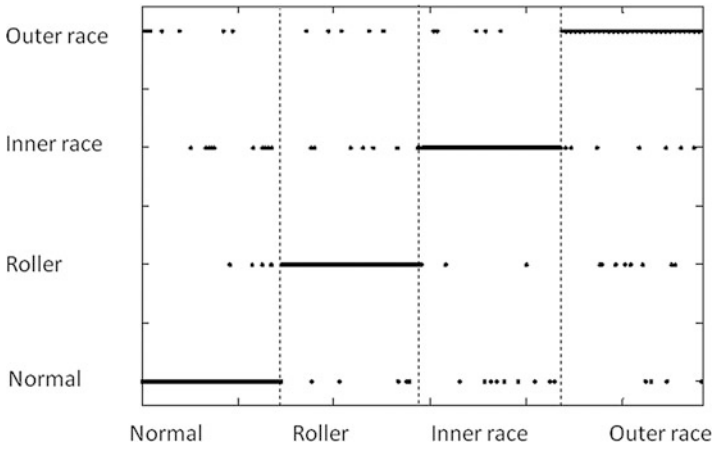


Fig. 3.17 Test 1.4 Multistate identification result based on RMS

DR value. Followed by the "roller fault," the "inner race fault" and the "outer race fault" class in the sample are least misclassified samples. Seen from the lateral view, no matter what kind of feature extraction method is used, samples which are misclassified into the "inner race fault" are the most, which is followed by the "outer race fault." The number of samples which are classified into the "roller fault" and "normal" is least, corresponding to low FAR value. Comparing four figures, the corresponding results of all kinds of sample points in Figs. 3.18, 3.19 and 3.20 are better than those in Fig. 3.17. That is, the performance of multiple classifiers based on energy and energy moments is better than that of RMS-based multiple classifiers. The results shown in the diagram are consistent with the results in Tables 3.5 and Table 3.6.

④ Based on results from the pure vibration data, we come the following conclusions.

No matter for the binary identification estimated by the safety region or for multi-fault identification of multiple fault types, the identification methods based on real-time state feature all have the identification accuracy of more than 0.85, which can effectively accomplish the identification work.

The size of the data sampling frequency has little effect on the accuracy of the identification method based on the real-time state feature, and the method can adapt to the data of different sampling frequencies.

In the binary identification, the identification accuracy differences of energy, entropy, energy moment, and Shannon features are small. The identification accuracy of RMS is in low accuracy. Features based on the energy, energy moment has a good ability to overcome the unbalanced data problem.

In multistate identification, real-time state features based on energy and energy moments are more capable of distinguishing different states and improving the accuracy of state identification. However, real-time state features based on RMS

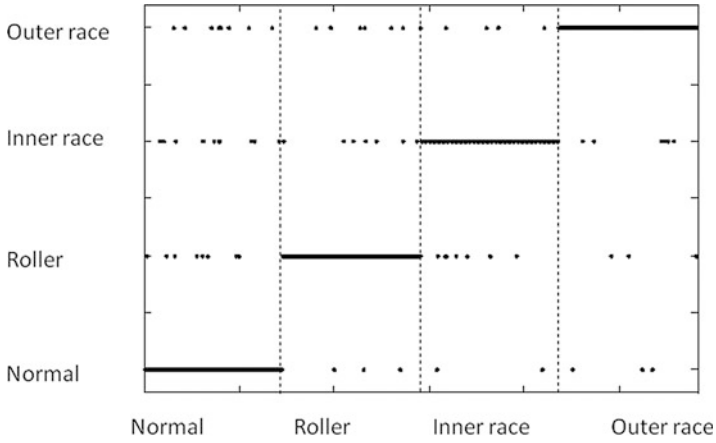


Fig. 3.18 Test 1.4 Multistate identification result based on energy

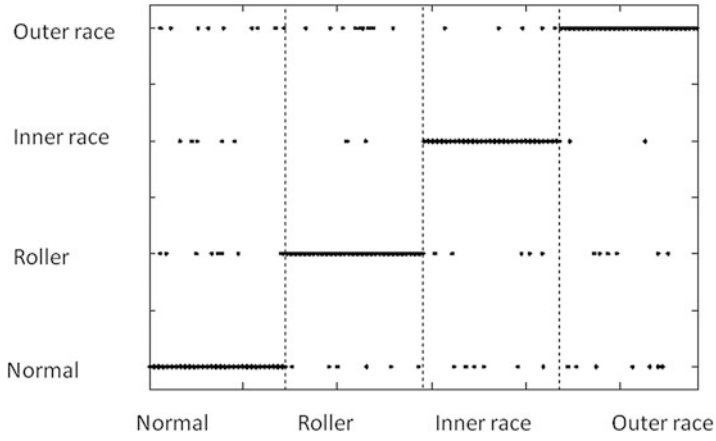


Fig. 3.19 Test 1.4 Multistate identification result based on Shannon entropy

and Shannon entropy do not perform well in subclassifier training and multistate identification.

Considering the classification accuracy and adaptability, four kinds of feature extraction methods can be ranked as energy moment—energy-Shannon entropy—RMS according to their performance.

B. Results of Simulated Operation Condition

Experimental results of the above section, the accuracy difference of identification of four kinds of real-time feature is not very large. To simplify the experiment, the feature with high identification accuracy is chosen for experiments under simulated operation condition. In addition, the identification accuracy of the data

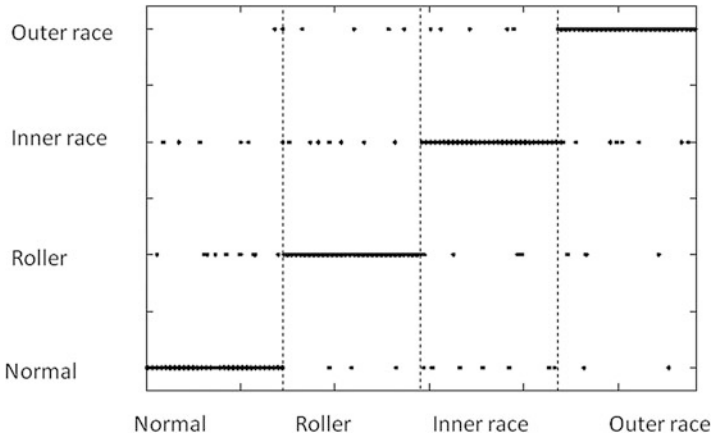


Fig. 3.20 Test 1.4 Multistate identification result based on energy moment

sampling frequency is not sensitive in above section, so the choice of 12 k Hz load data as the original vibration data to simulate the actual working environment to verify the algorithm, the results can represent the identification method based on the characteristics of real-time state simulation performance in actual working environment.

① Test 2.1safety region binary identification result is shown in Table 3.7.

Table 3.7 shows the Test 2.1 simulation of the normal, and fault state identification results in the actual operating environment. In order to compare with the experimental results in the laboratory environment, the floating percentage of DR normal, DR failure, CR and FK index is given in the table.

Seen from the Table 3.7, simulation of the practical working environment, real-time feature extraction method showed that the normal and fault state of the two detection rates were 0.7787 and 0.9433, the normal detection rate is far lower than

the fault rate, the CR and the FK values of all samples were 0.9095 and 0.7826, the correct classification rate is still higher than the 90%, and the FK value is close to 0.8. Further, the performance degradation can be obtained from the float percentage. The float percentage of two states are −18.97% and 0.69%, respectively. The drop of the CR and FK value is 4.34 and 12.74, respectively. The DR of the normal state drops largely.

The result shows that the identification method based on the features of the real-time state of actual working environment in the high-strength composite noise still can accurately complete the normal and fault condition identification.

② Test 2.2 multistate identification result can be seen from Table 3.7 and Table 3.8.

Table 3.8 shows the multistate identification result. Table 3.9 shows the identification result of normal, roller fault, inner race fault, and outer race fault.

Seen from Table 3.8, classifiers related to the outer race fault, such as “roller fault VS outer race fault,” are in poor performance. The Cr value is lower than 0.6,

Table 3.7 Safety region estimation of simulated operation condition

	Index value	Float percentage
DRNormal	0.7787	−18.97
DR fault	0.9433	−0.69
CR	0.9095	−4.34
FK	0.7826	−12.74

Table 3.8 DAGSVM identification result under simulated operation condition

		Index value
Normal VS outer race fault	CR	0.8861
	FK	0.8421
Roller fault VS outer race fault	CR	0.5908
	FK	0.2506
Inner race fault VS outer race fault	CR	0.5948
	FK	0.2588
Normal VS roller fault	CR	0.8819
	FK	0.8336
Normal VS inner race fault	CR	0.8939
	FK	0.8577
Roller fault VS inner race fault	CR	0.7496
	FK	0.5687

Table 3.9 Identification results of four-fault-type under simulated operation conditions

	Index value	Float percentage
DR_{normal}	0.8439	−8.28
FAR_{normal}	0.1272	——
$DR_{Roller\ fault}$	0.6848	−19.56
$FAR_{Roller\ fault}$	0.4554	——
$DR_{Inner\ race\ fault}$	0.7654	−16.15
$FAR_{Inner\ race\ fault}$	0.3802	——
$DR_{outer\ race\ fault}$	0.6237	−22.09
$FAR_{outer\ race\ fault}$	0.4811	——
CR	0.7206	−16.40
FK	0.6577	−22.25

followed by the "roller fault VS inner race fault," whose CR value is 0.7496. Another three classifiers are relatively good, with CR value 0.8.

From Table 3.9, we can see that under the simulated practical operation conditions, the classification effect of normal and inner race fault states is better, and the DR are 0.8439 and 0.7654, respectively, followed by roller fault, with value 0.6848. The detection rate of the sample points in the outer race fault state is the lowest, only 0.6237. As for the overall multistate identification accuracy, CR and FK values are 0.7206 and 0.6577, respectively. The classification accuracy is slightly higher than 70%, and the FK value is also low.

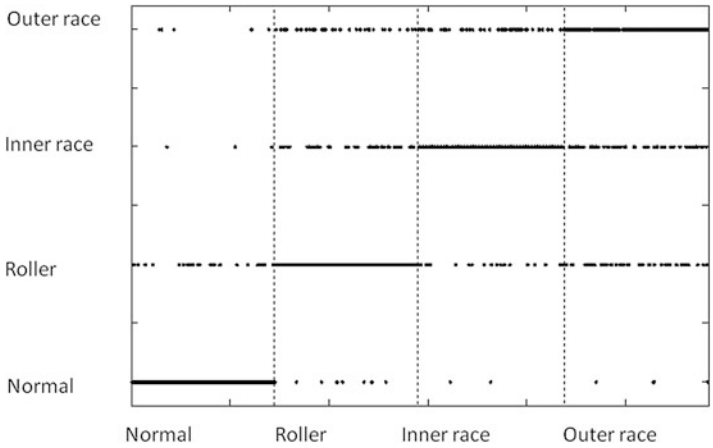


Fig. 3.21 Simulated results of multistate under simulated operation condition

Further, seeing the changes from the perspective of the value of each index from Table 3.9, outer race fault state of the sample point detection rate dropped by 22.09%, while the DR floating percentage roller and the inner race fault are 19.56% and 16.15%, respectively. The drop rate is more than 15%. The DR floating rate of the normal state is nearly 10%. From the point of view of the overall recognition accuracy, the method of CR floating percentage is relatively larger, with value 16.40%. The corresponding FK value is −22.25%, which indicates a poor performance.

To simulate multistate identification results based on real-time feature extraction of state of practical operation condition more clearly display. Figure 3.21 shows multi-classification results of simulated practical operation condition. Different misclassified situations and the results are consistent in Table 3.9.

The experimental results show that identification accuracy will drop a lot under practical operation conditions. The overall identification accuracy is less than 80%, and the robustness is poor.

③ According to the analysis of the above experimental results, the following conclusions can be obtained in view of the vibration data containing complex noise in the simulated practical operation condition.

In the binary identification, the identification accuracy based on the real-time feature is more than 90%, which can meet the actual engineering needs.

In the multistate identification, the identification accuracy based on the real-time feature is seriously affected by the noise. The accuracy of identification is greatly reduced and the robustness is general.

Table 3.10 Running time of algorithm based on the real-time feature

	12 k Hz sample frequency		48 k Hz sample frequency	
	Binary identification	Multistate identification	Binary identification	Multistate identification
Running time	0.0206	0.0366	0.0398	0.0491

B. Efficiency Verification of Algorithm

In the field of application, the online identification is required. So the real-time verification is required. In order to investigate the efficiency and execution speed of the algorithm, the execution time of the algorithm is tested.

The computer hardware environment of this book algorithm real-time verification test is Intel (R) Core (TM) 2 Duo CPU E7500 @ 2.93GHz, 2G RAM. All the algorithms involved are executed in the Matlab environment. As mentioned in experiment preparation, the load data of 12 k Hz are tested in this section, and the energy feature index is chosen as the representative for real-time analysis and comparison of the algorithm.

In the safety state identification of binary identification and multistate identification, the running time refers to the finishing the online running state identification after the boundary is set by the off-line data, including reading from the original data to obtaining the whole process of the state identification.

Table 3.10 gives the algorithm execution time for binary identification and multistate identification. Given the load data of 12 k Hz sampling frequency, completion time of a binary state identification is 0.0206 seconds and 0.0366 seconds for a multistate identification. Given the load data of 48 k Hz, completion time of a binary state identification is 0.0398 seconds and 0.0491 seconds for a multistate identification.

All in all, the running time of the calculation time is less than 0.0.5 s no matter what sample frequency is used. The state identification method based on real-time state feature has a high efficiency. Moreover, with the increase of data sampling frequency, that is, the increase of data points, the algorithm time is increased, but the growth rate is not large. Therefore, the state identification method based on real-time state features has high computing efficiency, and it should be able to meet the requirements of high real-time field application.

3.2 Degradation Assessment of Rolling Bearings Based on SVDD

3.2.1 Support Vector Data Description

Support vector data description was originally proposed by Tax and Duin. Given a target object set $x_i \in R^d$ i = 1,. . ., N, the basic idea of SVDD is to find a minimum-

volume hyper sphere in high-dimensional space with center a_F and radius R to enclose most of the objects, as shown in Eq.(3.1.70).

$$\text{Minimize } O_p(R, a_F, \xi) = R^2 + c\sum_{i=1}^{N}\xi_i \tag{3.1.70}$$

Subject to $\|\phi(x_i - a_F)\|^2 \leq R^2 + \zeta_i$

$$\zeta_i \geq 0 \quad \forall \mathrm{i} = 1, \ldots \mathrm{N}$$

where c is the penalty weight which gives the trade-off between the volume of the hyper sphere and the number of errors. ζ_i is a slack variable which allows a probability that some of the training samples can be wrongly classified. ϕ is a nonlinear mapping which maps the input object into a high-dimensional feature space F.

The dual problem of (3.1.70) is as (3.1.71), where $K(x_i, x_j)$ is the kernel function.

$$\text{Maximize } O_d(\alpha) = 1 - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K\left(x_i, x_j\right) \tag{3.1.71}$$

Subject to $\sum_{i=1}^{N}\alpha_i \quad 0 \leq \alpha_i \leq C \; \mathrm{i} = 1, \ldots, \mathrm{N} \quad \mathrm{C} \in [1/\mathrm{N}, 1]$.

In this study, the Gaussian kernel, $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = \exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)$ is selected. It is because Gaussian kernel has only one free parameter to be turned and is shown to yield tighter boundaries than other kernel choices, where α is Lagrange multiplier. According to the Kuhn-Tucker conditions, the objects can be classified into three categories: the object with $\alpha_i = 0$ are inside of the hyper sphere; the objects whose $0 < \alpha_i < C$ are on the hyper sphere boundary; and the objects whose $\alpha_i = C$ fall outside the hyper sphere and have nonzero ξ_i.The objects with $\alpha_i > 0$ are the support vectors. Objectors lying on the hyper sphere boundary ($0 < \alpha_i < C$) are also called unbounded support vectors. Objects lying outside the hyper sphere ($\alpha_i = C$) are also called bounded support vectors. The center can be expressed as Eq. (3.1.72). And its radius R can be determined by utilizing the distance between a_F and any support vector x on the ball boundary (unbounded support vectors), as (3.1.73). Finally, for the test object x, the output can be obtained by comparing its distance to the center a_F with radius D in space F. The SVDD decision function is as (3.1.74):

$$a_F = \sum_{i=1}^{N_s}\alpha_i\phi(x_i) \tag{3.1.72}$$

$$R = \left(1 - 2\sum_{x_i \in SV_s} \alpha_i K(x_i, x_k) + \sum_{x_i \in SV} \sum_{sx\, j \in SVs} \alpha_i \alpha_j K(x_i, x_k)\right)^{\frac{1}{2}} \tag{3.1.73}$$

$$D(x) = \|\phi(x_i - a_F)\|^2 - R^2 = c - 2\sum_{i=1}^{N_s} \alpha_i K(x, x_i) \tag{3.1.74}$$

where $c = \left(1 - R^2\right) + \sum_{i=1}^{N_s} \alpha_i \alpha_j K\left(x_i, x_j\right)$ is a constant. For the rolling bearing fault detection, the real-time monitoring data x are accepted as target objects if $D(x) \leq D$, which indicates the rolling bearing is normal. Otherwise, it is rejected as an outlier, which indicates the rolling bearing is abnormal. There are two parameters needed to be tuned, C and q. C controls the trade-off between the volume of the hyper sphere and the classification error of the model. It can be tuned to archive the determined confidence level of the fault detection process. By changing the value of the width parameter $q = 1/2\sigma^2$ in the Gaussian kernel, the description transforms from a solid hyper sphere to a Parzen density estimator.

The above inference involves the inner product of the vector. According to the theory proposed by V. Vapnik, the kernel function can be used for the calculation of the inner product of vectors. So the nonlinear problem in low dimension can be converted in the linear problem in the high dimension. The following will introduce the kernel function.

1. Polynomial kernel function

$$K\left(x_i, x_j\right) = \left[\left(x_i \cdot x_j\right) + 1\right]^d \tag{3.1.75}$$

where d represents the order of the kernel function. The difference between the polynomial kernel function and other functions is the redundant vectors from the order 2 to order n.

When the dimension and the order is small, such as dimension equals to 2 an order equals to 3, the mapping can be represented as Eq.(3.1.76):

$$\phi(x) = \left(1, x_1, x_2, x_3, x_1 \cdot x_2, x_2 \cdot x_3, x_1 \cdot x_3, x_1^2, {}^1x_2^2, x_3^2\right) \tag{3.1.76}$$

2. Gaussian radial basis kernel function

$$K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{3.1.77}$$

3. Multi-quadric kernel function

$$K(x_i, x_j) = \sqrt{\|x_i - x_j\|^2 + c^2} \tag{3.1.78}$$

4. Perceptual kernel function

$$K(x_i, x_j) = \tanh(\beta x_i \cdot x_j + b) \tag{3.1.79}$$

5. B-spline kernel function

$$K(x_i, x_j) = B_{2n+1}(x_i - x_j) \tag{3.1.80}$$

3.2.2 *Particle Swarm Algorithm Based on Dynamic Weight Adjustment*

In 1995, James Kennedy and Russell Eherhart propose the particle swarm algorithm [13]. This is an evolving algorithm which seeks the optimal solution in the space. According to the transmission of the information among particle swarm, the whole swarm moves to the optimal solution. This algorithm has been used widely for its good performance, but it tends to fall into the local optimal solution. Therefore, a lot of researches pay attention to its improvement.

1. Particle swarm algorithm

In the particle swarm optimization (PSO) algorithm, the optimal solution searching process will turn into the particle-searching process, and object function of every location of the will be evaluated. Every particle will decide the next move based on the current location and global optimal location. Like the searching food process of birds, particles will have influence on each other. And this influence will drive the particle to move toward the global optimal solution. Therefore, the particle swarm algorithm can be described as follows [26]:

In a continuous space with D dimensions, a particle swarm composed of m particles fly in a certain speed. In the searching process for every particle, considering the optimal point of a certain particle and the global optimal point of the whole

swarm, the location of the swarm is changed. Suppose the *i*th particle is composed of three D dimension vectors. Their state can be expressed as follows:

$$\text{Current location} : x_i = (x_{i1}, x_{i2}, \ldots, x_{iD}).$$
$$\text{History optimal location} : p_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$$
$$\text{Speed} : v_i = (v_{i1}, v_{i2}, \ldots, v_{i3})$$

where $i = 1, 2, \ldots, n$.

The current location of the particle can be described by coordinates, and the object location will be assessed after each iteration. The optimal location of the whole swarm is marked as $p_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$. The speed and location of each particle can be updated as Eq. (3.1.81) and Eq. (3.1.82), respectively:

$$v_{id} = v_{id} + c_1 \cdot rand() \cdot (p_{id} - x_{id}) + c_2 \cdot rand() \cdot \left(p_{gd} - x_{id}\right) \tag{3.1.81}$$

$$x_{id} = x_{id} + v_{id} \tag{3.1.82}$$

where the acceleration constant $c_1 \geq 0$, $c_2 \geq 0$ and two constants show the intelligence of particles. Rand() is a random function in the [0,1]. Vmax is the maximum speed set by users, and the speed of particles can be controlled between $[-V_{\max}, V_{\max}]$.

To further improve the performance of PSO, Y. Shi and R. Eberhart introduce the weight into the algorithm. The weight can decide the influence of the historical speed on the current speed, and a new speed update formula can be expressed as Eq. (3.1.83):

$$v_{id} = \omega \cdot v_{id} + c_1 \cdot rand() \cdot (p_{id} - x_{id}) + c_2 \cdot rand() \cdot \left(p_{gd} - x_{id}\right) \tag{3.1.83}$$

Y. Shi and R. Eberhart found that the algorithm can perform well when the weight is between 0.9 and 1.2. With the iteration of the algorithm, the weight can decrease linearly. Therefore, maximum weight $\omega_{\max}$, minimum weight $\omega_{\min}$, and the largest iteration time $t_{\max}$ are introduced. The weight can be adjusted as Eq. (3.1.84):

$$\omega = \omega_{\max} - \frac{t}{t_{\max}}(\omega_{\max} - \omega_{\min}) \tag{3.1.84}$$

Therefore, the improved PSO can be summarized as follows.

Step 1: Initialize the speed and the location of each particle in the D dimension.
Step 2: Assess the adaptation value of the optimization function in D dimension.
Step3: Compare the current value and the historical value, and update the location and speed according to the comparison result. The updated rule is based on the Eq. (3.1.81) and Eq. (3.1.82).
Step 5: Stop the iteration if the result is satisfying, or return to the step either.

2. Particle swarm algorithm based on dynamic weight adjustment

According to the 4.3.1, if the whole particle swarm converge at a certain particle p_g, the iteration will stop. The optimal particle is local particle p_g and the optimal value cannot be obtained. To keep the diversity of the whole swarm, which means the particle I can change randomly when the above situation happens, the similarity $s(i, g)$ of particles is introduced.

Definition: The similarity degree of two particles has to satisfy the following rules:

1. $s(i, i) = 1$
2. When $d(i, j) \to \infty$,$s(i, j) \to 0$.
3. For any particle such as i and j, $s(i, j) \in [0, 1]$.

Based on the rule, the similarity between i and j is calculated through Eq. (3.1.85):

$$s(i,j) = \begin{cases} 1, d(i,j) \le d_{\min} \\ 1 - \left[\frac{d(i,j)}{s_{max}}\, max\right]^{\alpha}, d_{\min} \le d(i,j) \le d_{\max} \\ 0, d(i,j) \ge d_{\max} \end{cases} \tag{3.1.85}$$

where $d(i, j)$ is the Euclidian distance between particle i and particle j. $d_{\max}$ and $d_{\min}$ are constants.

Set iteration times t of $d_{\max}$ and $d_{\min}$, respectively, and calculate similarity degree $s(i, j)$.When the similarity is zero, the inertia weight of the particle is the largest one representing by $\omega_{\max}$.When the similarity is one, the inertia weight of the particle is the lowest one representing by $\omega_{\min}$.When the similarity is between zero and one, similarity decreases accordingly. The calculated formulas of inertia weight are show as follows:

$$\omega_i =^{i} \omega_{\max} - s(i,g)(\omega_{\max} - \omega_{\min}) \tag{3.1.86}$$

$$\omega_i = \omega_{\min} - (\omega_i - \omega_{\min}) \cdot \frac{t_{\max} - t}{t_{\max}} \tag{3.1.87}$$

Finally, the kernel parameter and penalty weight selection method based on DPSO is made up of six steps:

Step 1: Generate n locations and initial speeds of kernel parameter or penalty weight.
Step 2: Evaluate the adaptation of every kernel parameter or penalty weight.
Step 3: Confirm the best location of every particle and the global best location.
Step 4: Calculate particle and global similarity of every penalty or penalty weight according to Eq. (3.1.85), and calculate particle and global weight according to Eq. (3.1.86) and Eq. (3.1.87).

Step 5: Update the location and position of kernel parameter or penalty weight.
Step 6: If the result satisfies the stopping condition, output the result. Otherwise, turn to Step2.

3.2.3 Research on the Self-Adaptation Warning

The degradation state of rolling bearings can be assessed by the calculation of the distance between the test value and the center of the SVDD hyper sphere. However, the SVDD distance is a dimensional index which means the distance value will be a lot more different. Thus the warning threshold is hard to set. However, the SVDD distance is a continuous value; thus, the SVDD distance should subject to the same probability distribution [27]. Therefore, the abnormal value detection method such as Pauta method and Chauvet Nat method can be used for the warning. The Pauta method is used for the application. The Pauta method is also called as 3S method which assumes the value is abnormal when the expectation value difference between the current data point and the whole set, shown as Eq. (3.1.88):

$$|x - \bar{x}_i| > 3S \tag{3.1.88}$$

The warning variable used in the research is SVDD distance, because the SVDD distance is stable when rolling bearings operate normally. When rolling bearings enter into an early degradation stage, the SVDD value will increase in an impulsive manner. Then, the SVDD distance keeps in a stable value until rolling bearings enter into a deep degradation stage.

Inspired by the above findings, the data smoothing method is introduced into the Pauta method to set the warning value. N continuous data are chosen for the mean value calculation so as to obtain the proper expectation value difference, shown as Eq. (3.1.89).

$$\left| \frac{(x_{i1} + x_{i2} + \ldots + x_{iN})}{N} - M \right| > 3S \tag{3.1.89}$$

The self-adaptation method can be described as Fig. 3.22.

3.2.4 Case Study

1. Data Acquisition

In this section, the rolling bearing life vibration data are from the University of Cincinnati intelligent maintenance system (IMS) center. The test apparatus is shown in Fig.3.6. Four ZA-2115 Rexnord rolling bearings are mounted on the same output

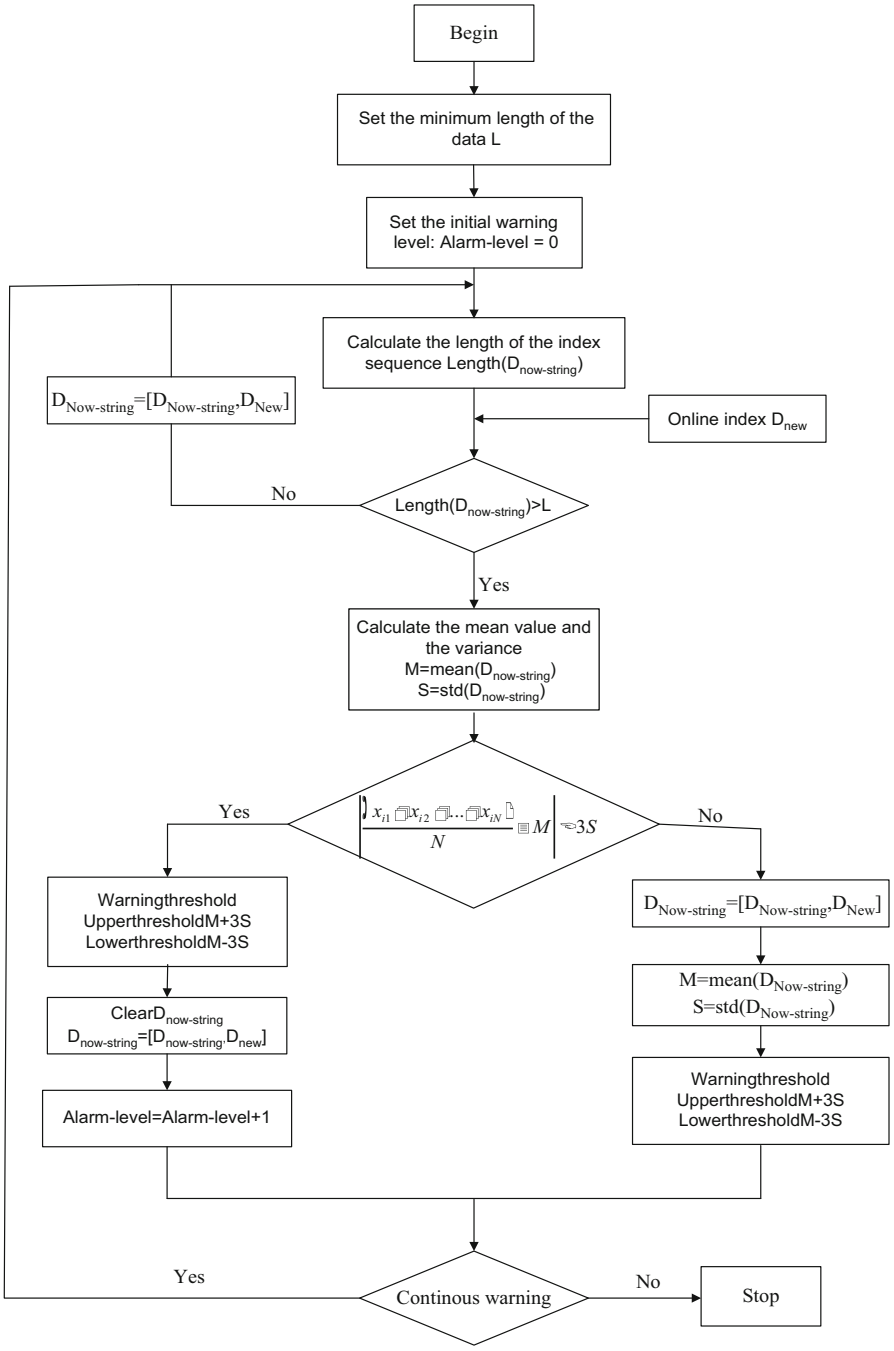


Fig. 3.22 Main steps of adaptive alarm method

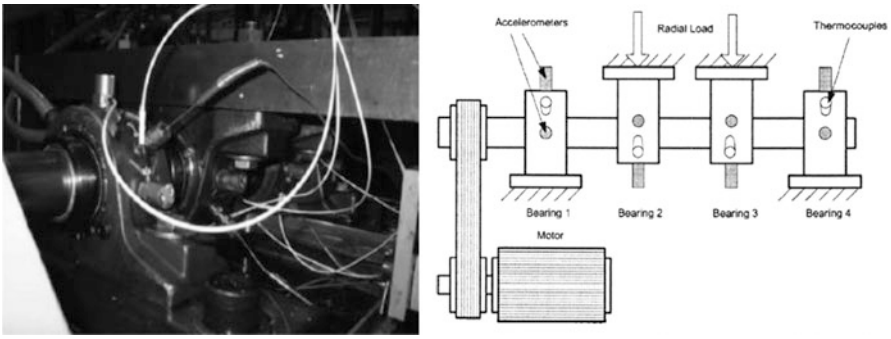


Fig. 3.23 Rolling bearing test rig

shaft with different positions. The speed of the output shaft is 2000 rpm. In the shaft and bearing, the radial load of 6000 lb. is exerted by the spring mechanism. Collect the data by data collection card NI DAQ6062EJ, and the sample rate is 20 kHz. Collect the data every 10 minutes and the collection time is 1 second. The data length is 20,480 (Fig. 3.23).

2. DPSO Algorithm Verification

To research the optimization ability of the DPSO, several commonly used functions are applied to the performance examination, including Rosenbrock, Rastrigin, Griewank, and Ackley, shown in Fig. 3.24. The optimization ability between PSO and DPSO is simulated via four examination functions. The optimization process between the proposed algorithm and the PSO is shown in Fig. 3.25. Concluding from the Fig. 3.25, both DPSO and PSO converge quickly at the earlier stage. However, Fig. 3.25(d) shows that PSO stops declining from the beginning of the iteration, indicating PSO is easily affected by the partial optimum solution. Therefore, the DPSO has a better simulation result.

3. Degradation Assessment of the Train Rolling Bearing

The degradation assessment of the rolling bearing is carried out with the feature extracted via PCA. This paper mainly discusses the incomplete data and takes the 50 groups normal data out of 2000 groups for training. After standardizing process, the data are input into the SVDD model for finding the center and radius of the SVDD hyper sphere. Then data to be tested are input into the existing SVDD model, and the degradation degree is obtained by calculating the distance between the data and the hyper sphere center.

The SVDD distance is shown in Fig. 3.24, and it is divided into three stages. The first stage covers the data between 0 and 735, which means the rolling bearing is normal stage and the SVDD distance is very small and stable. The second stage covers the data between 736 and 1638, which means SVDD distance increases and fluctuates in a larger scale indicating that the rolling bearing enters into an initial stage. The third stage covers the data between 1639 and 2000, and the SVDD

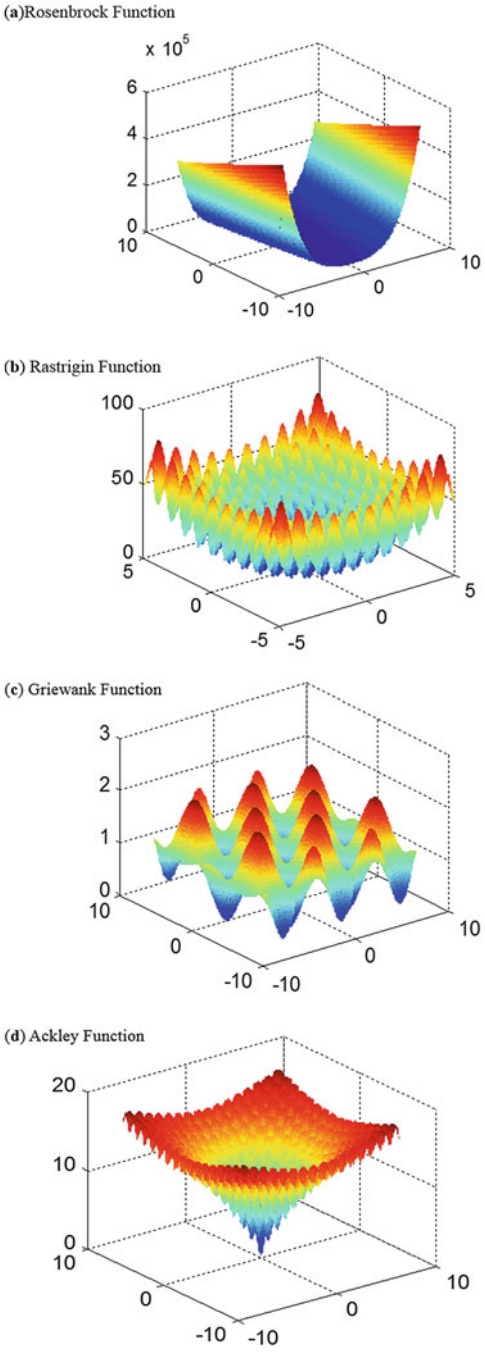


Fig. 3.24 Figure for test function

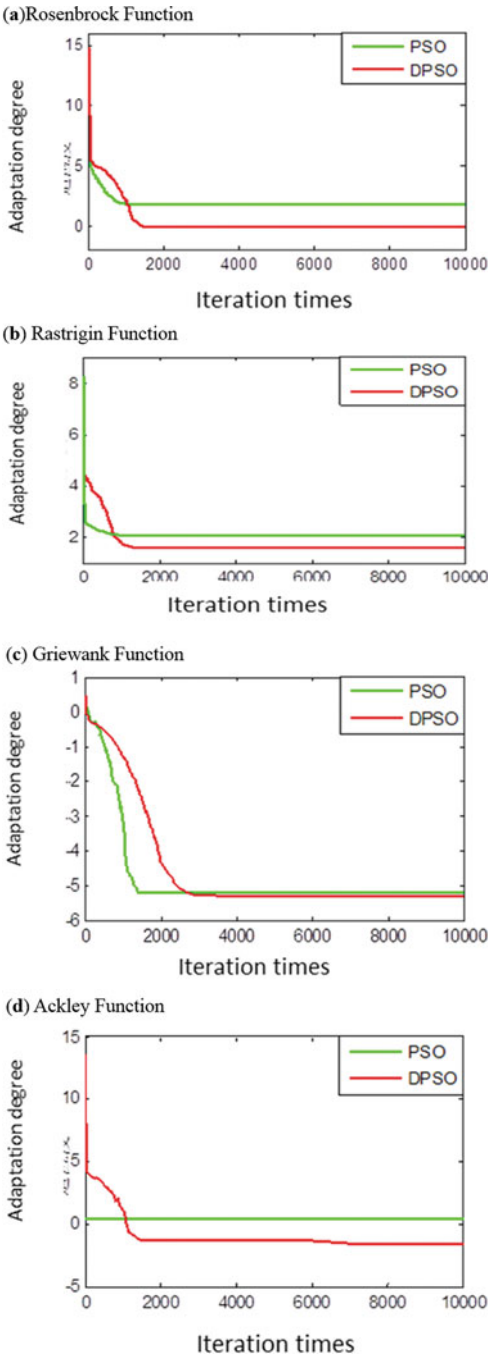


Fig. 3.25 Convergence performance comparison of two methods

distance increases rapidly up to another scale indicating that the rolling bearing enters into a deep degradation stage (Fig. 3.26).

3.3 Fault Diagnosis of Door System Based on the Extended Petri Net

3.3.1 Subway Train Door: Open Process Analysis

The door-opening process model of door system model can be parted into three layers: electrical control layer, EDCU control layer, and mechanical action layer. Electrical control layer includes relays, electrical loops, power supply, and electrical machines, accomplishing giving out signals of opening the door and driving the motor. EDCU control layer consists of EDCU units and their peripheral circuits, giving out signals of door movement velocity interactively controlled by EDCU and the electrical machine. And the mechanical action layer includes the door and the connected mechanical parts (masts, screw rods, guide rails, and so on), with the function of accomplishing the movements of mechanical parts of the door. The three subnets correlate and interact with each other, controlling the movement process of door system jointly.

3.3.2 Subway Train Door System Fault Diagnosis Theory and Method

1. Extended Time Petri Net

If time intervals $tm_i = [a, b]$ and $tm_j = [c, d]$, a and c are the lower bounds of tm_i and tm_j, respectively, while b and d are the upper bounds of tm_i and tm_j, respectively. Σ_F represents a FDES. $e \in E$ is a fuzzy event, state identification $M_i[e > M_j$. State transportation can be done in time interval tm_{ij}, with time constrain, which is denoted by $M_i[e \lhd tm_{ij} > M_j$.

Compound fault casual chain model based on temporal casual relations is established by extended time petri net (ETPN) [28–32]. ETPN is an eight tuple $\sum_{ET} = (S', T'; F', E', I', \delta', \tau', M_0')$, in which:

1. $(S', T'; F')$ is a prototype Petri net, with $\forall s \in S : \{s\} \cap s^{\bullet} \overrightarrow{=} \varnothing, \forall t \in T' : |{}^{\bullet}t| \leq 1$.
2. E' is the finite set of events.
3. $I' \subset F'$ is the finite set of inhibitor arcs.
4. $\delta' : E' \to 2^{|T'|}$ is event map in transportation subset.
5. $\tau' : T' \to R_0 \times (R_0 \cup \{\infty\})$, $\tau'(t_i)$ is time delay interval related to transportation t_i.
6. M_0' is the original state identification of Σ_{ET} and $M'(s_i) \in R_0 \times R_0$ is the token of s_i.

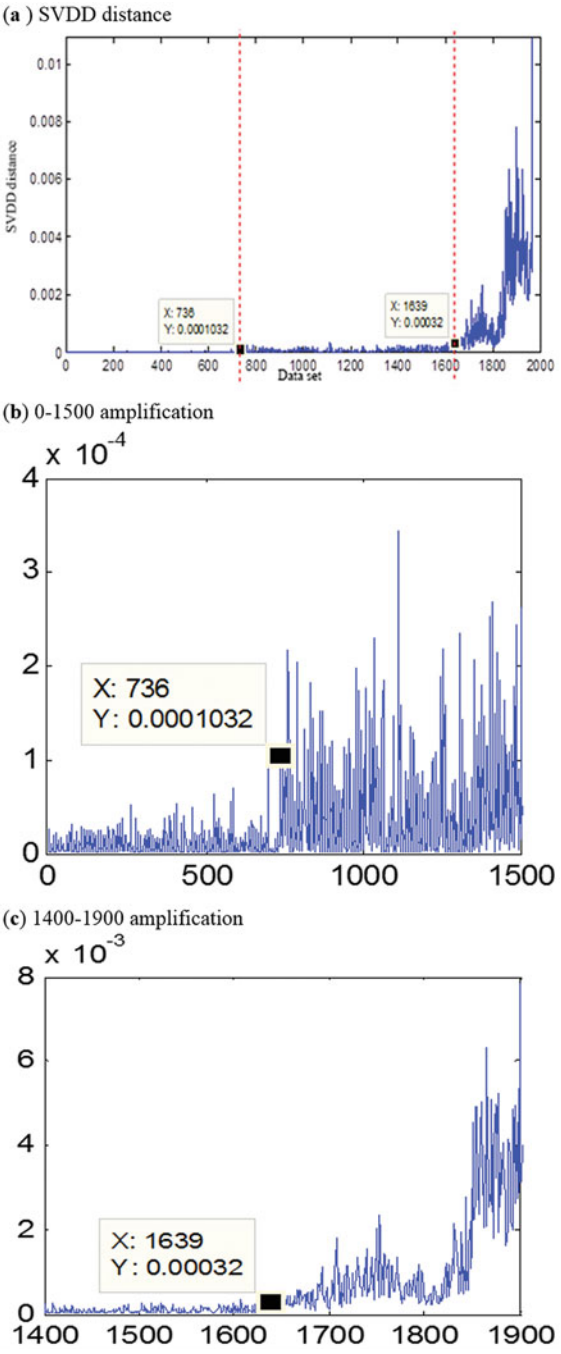


Fig. 3.26 The curve of performance degradation

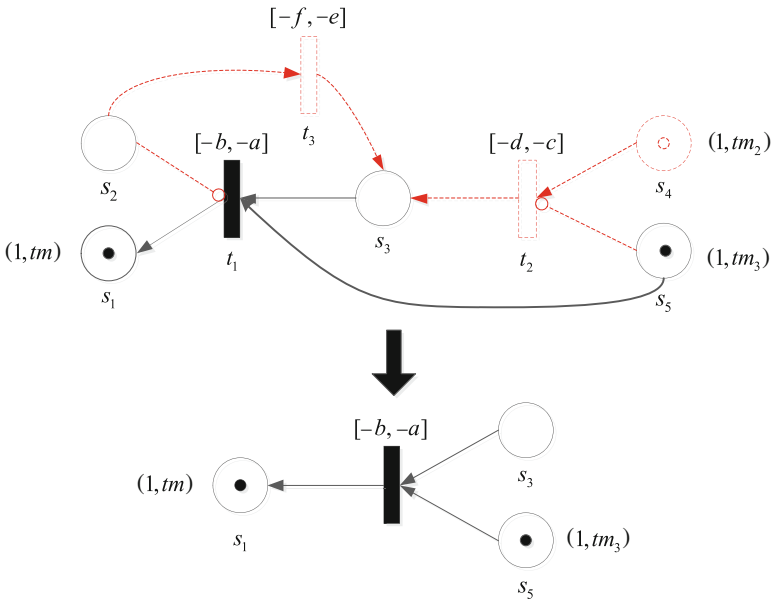


Fig. 3.27 TC-PPN reduction

2. Possibility Petri Net with Time Constraints

If Σ_{TP} is possibility petri net with time constraints (TC-PPN), fuzzy state identification of Σ_{TP} is $M = [M(s_1), M(s_2), \cdots, M(s_n)]^T$, in which $n = \mid S|$.

Through ETPN model inversion, TC-PPN can be got for correct decoupling, which can eliminate the impacts of virtual event and timing error. Based on inhibitor arcs and original state, model reduction can be done in TC-PPN. We will get TC-PPN after reduction.

Fig. 3.27 is the reduction process of TC-PPN, in which the dotted line elements can be deleted in the original net.

In Fig. 3.27, inhibitor arcs are represented by lines with circles in the tail.

3. Decoupling Net Construction

If Σ_{TP} is a FDES, time constraint of Σ_{TP} is defined as a three tuple $TG(\Sigma_{TP}) = (V, A_{rc}, H)$, in which:

1. $V = \{V_i | i = 1, 2, \ldots, n\}, V_i = \{(a, b) | a, b \in Ra \leq b\}$ together with $s_i \in S$ are time constraint, which is called vertex set $TG(\Sigma_{TP})$.
2. $A_{rc} \subseteq F$is the arc set of $TG(\Sigma_{TP})$.
3. $H : A_{rc} \rightarrow 2^{|T|}$ is the mark of arc.

The time constraint graph of TC-PPN is established, which is used to calculate ODDT with time constraint.

If Σ_{TP} is a TC-PPN and $TG(\Sigma_{TP}) = (V, A_{rc}, H)$ is the time constraint graph, ODDT of event $e \in E$ is

$$d_e = \left(d_e^o, d_e^{uo}\right) \tag{3.1.90}$$

$$d_e^o = \frac{\sum\limits_{\mu \in \delta(e)} \mu}{\mid \delta(e) \mid - \mid \varepsilon \mid} \tag{3.1.91}$$

$$d_e^o + d_e^{uo} = 1 \tag{3.1.92}$$

in which $|\varepsilon|$ represents the number of $\mu = \varepsilon$ in $t \in \delta(e)$. The calculation method of μ is:

1. If $s_i \in {}^\bullet t$, $s_j \in t^\bullet : \mid V_i \mid > 1 \vee \mid V_j \mid > 1$, then,

$$\mu = \varphi\left(V_i, V_j, \tau(t)\right) \tag{3.1.93}$$

$$\varphi = \frac{\sum\limits_{v_i \in V_i,\, v_j \in V_j,\, v_i + \tau(t) \neq v_j} \kappa\left(v_i, v_j, \tau(t)\right)}{\mid V_i \mid \cdot \mid V_j \mid - \mid v_i + \tau(t) \equiv v_j \mid} \tag{3.1.94}$$

in which $|v_i + \tau(t) \equiv v_j|$ is time interval number of satisfying $v_i + \tau(t) = v_j$ and κ is calculated by Eq.3.1.90. φ is non-self-voting function.

2. If $s_i \in {}^\bullet t$, $s_j \in t^\bullet : V_i \notin V \wedge V_j \notin V$, then $\mu = 0$.
3. If $s_i \in {}^\bullet t$, $s_j \in t^\bullet : \mid V_i \mid = 1 \wedge \mid V_j \mid = 1$, then $\mu = \varepsilon$, which means unknown state out of d_e^o calculation.

Set $\sum_{TP} = (S, T; F, E, I, \delta, \tau, M_0)$ as a FDES with multiple casual coupling. Then $D^T = \left\{d_e^T | e \in E\right\}$ is an ODDT set of all events in Σ_{TP}. With ODDT, multiple casual decoupling algorithm of FDES is established and used to do decoupling. The result is type I decoupling net in the first step.

Search leakage library in type I decoupling net. If the leakage library is not the source library of original ETPN, type II decoupling net needs to be constructed.

Σ_{I} is given as type I decoupling net of Σ_{TP}, and the reachable state mark set is R_{I}. Σ_{I} only has one state mark $M_l = (M_{\sigma\_l}, M_{\upsilon\_l})$, which satisfies $\forall M_\upsilon = (M_\sigma, M_\upsilon)$, $M_\upsilon \in R_{\mathrm{I}} : M_{\sigma\_l}(s) = \max\{M_\sigma(s)\} \wedge M_\upsilon(s) \subseteq M_{\upsilon\_l}(s)$. $\dot{\delta}(e_7) = \{t_7, t_9\}$ is the final state of Σ_{I}.

Σ_{I} is given as type I decoupling net of Σ_{TP}. R and R_{I} are the reachable state mark set of Σ_{TP} and Σ_{I}, respectively. Σ_{TP} projection in Σ_{I} is $P(R)$. And then $P(R) = R_{\mathrm{I}}$.

Σ_{I} is given as type I decoupling net of Σ_{TP}, $\sum_{TP} = (S, T; F, E, I, \delta, \tau, M_0)$, $\sum_{\mathrm{I}} = \left(S_{\mathrm{I}}, T_{\mathrm{I}}; F_{\mathrm{I}}, E_{\mathrm{I}}, I_{\mathrm{I}}, \delta, \tau, M_0^{\mathrm{I}}\right)$. M_l and M_l^{I} are given as the final state of Σ_{TP} and Σ_{I}, $S_{\mathrm{I}\_in} = \{s | s \in S_{\mathrm{I}} \wedge s^\bullet = \varnothing\}$, $S_{in} = \{s | s \in S \wedge s^\bullet = \varnothing\}$. If $S_{\mathrm{int}} = S_{in} \cap S_{\mathrm{I}\_in} \wedge S_{\mathrm{int}} \neq \varnothing$, as to $\forall s \in S_{\mathrm{int}}$, $M_l(s) = M_l^{\mathrm{I}}(s)$.

Set $\sum_{TP} = (S, T; F, E, I, \delta, \tau, M_0)$, $\sum_{\mathrm{I}} = \left(S_{\mathrm{I}}, T_{\mathrm{I}}; F_{\mathrm{I}}, E_{\mathrm{I}}, I_{\mathrm{I}}, \delta, \tau, M_0^{\mathrm{I}}\right)$, in which Σ_{II} is given as type II decoupling net of Σ_{TP}. R_{II} is the reachable state mark set of Σ_{II}. Σ_{II} has only one state mark $M_l = (M_{\sigma\_l}, M_{\upsilon\_l})$, satisfying $\forall M_\upsilon = (M_\sigma, M_\upsilon)$, $M_\upsilon \in R_{\mathrm{I}} : M_{\sigma\_l}(s) = \max\{M_\sigma(s)\} \wedge M_\upsilon(s) \subseteq M_{\upsilon\_l}(s)$. M_l is the final state of Σ_{II}.

Σ_{II} is the reduction result of Σ_{I}, which means all the elements in Σ_{II} are subsets of Σ_{I} elements. Thereby,

$$\sum_{TP} = (S, T; F, E, I, \delta, \tau, M_0) \tag{3.1.95}$$

$$\Sigma_{\mathrm{II}} = \left(S_{\mathrm{II}}, T_{\mathrm{II}}; F_{\mathrm{II}}, I_{\mathrm{II}}, E_{\mathrm{II}}, \delta, \tau, M_0^{\mathrm{II}}\right) \tag{3.1.96}$$

M_l and M_l^{II} are given as the final state of Σ_{TP} and Σ_{II}, respectively, $S_{\mathrm{II}\_in} = \{s | s \in S_{\mathrm{II}} \wedge s^\bullet = \varnothing\}$, $S_{in} = \{s | s \in S \wedge s^\bullet = \varnothing\}$. If $S_{\mathrm{int}} = S_{in} \cap S_{\mathrm{II}\_in} \wedge S_{\mathrm{int}} \neq \varnothing$, as to $\forall s \in S_{\mathrm{int}}$, $M_l(s) = M_l^{\mathrm{II}}(s)$.

This inference illustrates the final state in accordance with the original model can be obtained by recalling the reduction model based on time interval and conditional probability. In the information system with complex causalities, such as backward inference, fault analysis, and diagnosis, analysis and reduction with this method will lead to a result consistent with original model at a low computing cost.

3.3.3 *Case Study*

In order to test whether this method is effective or not, a decoupling deduce is made. This simulation experiment follows the opening-door process of rail transportation vehicles. According to the recorded frequency of door fault, the input of compound fault during the opening-door process is set as the faults of screw-nut and screw-rod unlocking, masts, electrical machines, brakes, and popping of backing pins (Table 3.11).

In the opening-door process carried out by the door system model in Fig. 3.28, the monitoring system will capture the following information (the unit is ms): $O(S_{38}) = 2951$, $O(S_{35}) = 2892$, $O(S_{28}) = 2243$, $O(S_{25}) = 1145$, and $O(S_{19}) = 1077$, among which $O(x)$ refers to the time when x is captured for the first time. It aims to get the possible faulted running equipment and their triggered event chain.

Probability data of some nodes used in experiment are as in the following table.

Based on the above parameters and analysis, along with the causal chain of running equipment, an ETPN model is as in the following Fig. 3.28.

$\sum_{ET} = (S', T'; F', E', I', \delta', \tau', M_0')$ is given as ETPN in Fig. 3.28, among them (Table 3.12).

Table 3.11 Probability data of some nodes

Node	Probability data	Node	Probability data	Node	Probability data
1	0.0005	11	0.0015	21	0.0034
2	0.0010	12	Y0.0033	22	0.0028
3	0.0001	13	0.0010	23	0.0002
4	0.0002	14	(Y)0.0143	24	0.7793
5	Y0.0011	15	0.0001	25	0.2207
6	Y0.0042	16	(Y)0.0001	26	0.0010
7	0.0064	17	0.0001	27	0.0001
8	(Y)0.0220	18	0.0001	28	0.0001
9	0.0167	19	Y0.0046	29	(Y)0.0723
10	Y0.0029	20	0.0004	30	0.1187

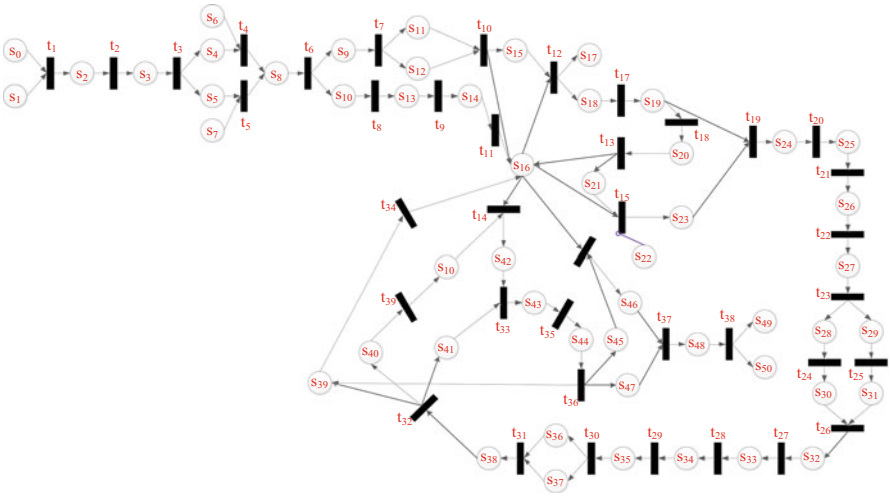


Fig. 3.28 ETPN model of the door system in the opening-door process

1. $S^{'}$ is a collection of libraries. Implications of each library can be found in Table 3.12. $T^{'} = \{t_1, \cdots t_{39}\}$, $F^{'}$ is shown as in the figure:
2. $E^{'} = \{e_1, \cdots e_{39}\}$;
3. $I^{'} = \{(S_{22}, t_{15})\}$;
4. $\delta^{'}(e_1) = \{t_1\}$, $\delta^{'}(e_2) = \{t_2\}$, $\delta^{'}(e_3) = \{t_3\}$, …, $\delta^{'}(e_{39}) = \{t_{39}\}$;
5. $\tau(t_{15}) = \tau(t_{17}) = \tau(t_{18}) = \tau(t_{19}) = \tau(t_{22}) = \tau(t_{26}) = \tau(t_{28}) = \tau(t_{30}) = [10, 40]$,
   $\tau(t_4) = \tau(t_9) = \tau(t_{11}) = \tau(t_{12}) = \tau(t_{14}) = \tau(t_{32}) = \tau(t_{34}) = \tau(t_{37}) = [310, 340]$,
   $\tau(t_5) = \tau(t_6) = \tau(t_{21}) = \tau(t_{23}) = \tau(t_{24}) = \tau(t_{25}) = \tau(t_{39}) = [510, 540]$,
   $\tau(t_1) = \tau(t_3) = \tau(t_7) = \tau(t_{16}) = \tau(t_{33}) = \tau(t_{35}) = \tau(t_{39}) = [110, 140]$,
   $\tau(t_2) = \tau(t_8) = \tau(t_{10}) = \tau(t_{13}) = \tau(t_{20}) = \tau(t_{27}) = \tau(t_{29}) = \tau(t_{31}) = \tau(t_{36}) = [20, 40]$;
6. $M_0(S_{38}) = [2951, 2951]$, $M_0(S_{35}) = [2892, 2892]$, $M_0(S_{28}) = [2243, 2243]$, $M_0(S_{28}) = [2243, 2243]$, $M_0(S_{19}) = [1077, 1077]$, and $M_0^{'}(else) = \varepsilon$.

Table 3.12 Implications of each library in the ETPN model

Marking	Implication	Marking	Implication	Marking	Implication
S_0	Zero-speed signal	S_{18}	Driving signals of electrical machines	S_{36}	Upper guide rail
S_1	Stop-in-place signal	S_{19}	Electrical machines work	S_{37}	Lower guide rail
S_2	ATC	S_{20}	Locking devices turn on	S_{38}	Door opens to 85% of the maximum
S_3	Door enabling signal	S_{21}	Minor control library 2	S_{39}	Position sensor of the door
S_4	HMI	S_{22}	S_2 moves	S_{40}	Backing pins on top of the door work
S_5	ATO	S_{23}	Door movements allowed	S_{41}	Minor control library 3
S_6	ATO switch not on automatic catch	S_{24}	Unlocking of electrical machines and screw-nut pairs	S_{42}	Minor control library 7
S_7	ATO switch on automatic catch	S_{25}	Screw-nut locking devices quit the LS locking segment	S_{43}	Current controlled by electrical machines decline
S_8	Opening-door signal	S_{26}	S_1 switch off	S_{44}	Door opens to the maximum
S_9	VCU	S_{27}	S_4 switch off	S_{45}	Minor control 4
S_{10}	Relays in door system get electricity	S_{28}	Backing pin popping	S_{46}	Minor control 5
S_{11}	Zero-speed signal	S_{29}	Racks slide on the short guide pillar, with the long guide pillar moves laterally	S_{47}	Minor control 6
S_{12}	Opening-door signal	S_{30}	Lower suspension arm swings	S_{48}	Mechanical machines shut down
S_{13}	Normally open contacts of EDCU power circuits close	S_{31}	Idler wheel slides in the cross slide way	S_{49}	Yellow lights on
S_{14}	EDCU gets electricity	S_{32}	Idler wheel enters the lower slide way	S_{50}	Door stop

(continued)

Table 3.12 (continued)

Marking	Implication	Marking	Implication	Marking	Implication
S_{15}	Minor control library 1	S_{33}	Screw rod	S_{51}	Motor current increases instantly
S_{16}	EDCU	S_{34}	Screw nut		
S_{17}	Yellow lights wink	S_{35}	Masts drive the door moves longitudinally		

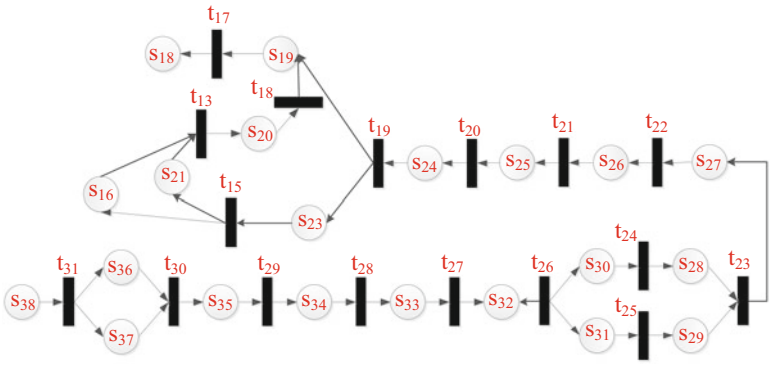


Fig. 3.29 Schematic diagram of TC-PPN of the door system during the opening-door process

Inverse Σ_{ET} and $\Sigma_{TP}{}' = (S, T; F, E, I, \delta, \tau, M_0)$ is drawn. Moreover, $\Sigma_{TP} = (S, T; F, E, I, \delta, \tau, M_0)$ is got when $\Sigma_{TP}{}'$ is reduced, shown in Fig. 3.29.

Making the extension degree of time interval $\lambda = 0.5$, the ODDT of each event is calculated by the time constraint derived from TG: $d^o_{e31} = \varepsilon$, $d^o_{e30} = 1$, $d^o_{e_{29}} = 0$, $d^o_{e_{28}} = 0$, $d^o_{e_{27}} = 0$, $d^o_{e_{26}} = 0$, $d^o_{e24} = 0.935$, $d^o_{e22} = 0$, $d^o_{e21} = 0.841$, $d^o_{e20} = 0$, $d^o_{e19} = 0.725$, $d^o_{e_{15}} = 0$, $d^o_{e_{13}} = 0$, $d^o_{e_{18}} = 0.5$, $d^o_{e_{17}} = 0$, $d^o_{e_{25}} = 0$, and $d^o_{e_{23}} = 0$. Create type I decoupling net of Σ_{TP} based on ODDT of events is as follows (Fig. 3.30):

Construction of type II decoupling net of Σ_{TP}, Σ_{II}, based on type I decoupling net Σ_{I} is shown in the following Fig. 3.31.

From 3.28 and 3.29, if the final condition of Σ_{II} is got, the conclusion that it is in accordance with Σ_{TP} under the condition of ODDT may be drawn. Thus, in this example, possible faulted running equipment and their triggered event chains can be got based on analysis of Σ_{II}.

Create a whole working model functioning in door opening created in the system emulator. In this model, if inhibitor arcs are added to nodes where faults can be tested, possible faults are expressed. When several faults occur at the same time, the most likely casual chain can be decoupled based on faults tested. The fault chain deduced by simulation environment is as in Fig. 3.32.

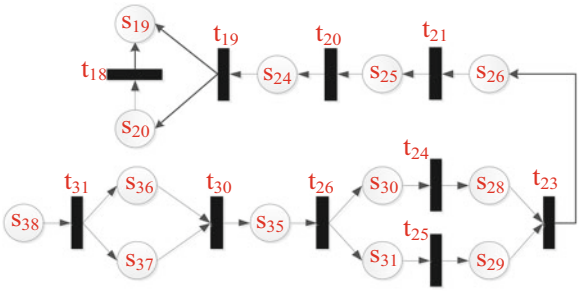


Fig. 3.30 Type I decoupling net of the door system during the opening-door process

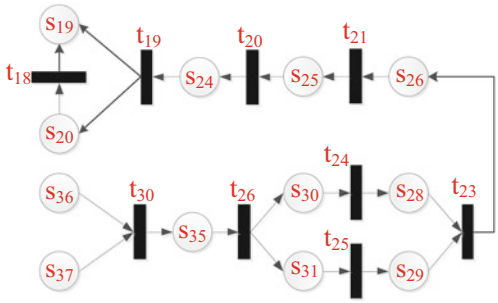


Fig. 3.31 Type II decoupling net of the door system during the opening-door process

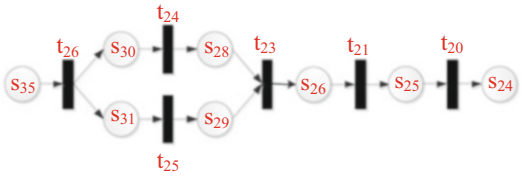


Fig. 3.32 Fault causal chain analyzed by decoupling

References

1. J.S. Smith, The local mean decomposition and its application to EEG perception data[J]. J. R. Soc. Interface **2**(5), 443–454 (2005)
2. Z. Kang. *Local mean mean decomposition method and its application in fault diagnosis of rotating machinery* [D]. Hunan, Changsha: Hunan University, (2012)
3. J. Cheng, K. Zhang, Y. Yu, An order tracking technique for the gear fault diagnosis using local mean decomposition method [J]. Mech. Mach. Theory **55**, 67–76 (2012)
4. Y. Wang, Z. He, Y. Zi, A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis[J]. J. Vib. Acoust. **132**(2), 021010 (2010)
5. J. Cheng, Y. Yang, Y. Yang, A rotating machinery fault diagnosis method based on local mean decomposition[J]. Digital Signal Process **22**(2), 356–366 (2012)
6. H. Li. Local mean decomposition based bearace fault detection [J] Adv. Mater. Res.. 2012, Mechatronics and Intelligent Materials II: 360–364

7. C. Park, D. Looney, M.M. VanHulle, et al., The complex local mean decomposition [J]. Neurocomputing **74**(6), 867–875 (2011)
8. M.D. Wang, L.B. Zhang, W. Liang, et al., Local mean decomposition method based on B-spline interpolation[J]. J. Vib. Shock **29**(11), 73–77 (2010)
9. C.L. Nikias, M.R. Raghuveer, Bispectrum estimation: a digital signal processing framework [J]. Proc. IEEE **75**(7), 869–891 (1987)
10. C. Cortes, V. Vapnik, Support vector networks[J]. Mach. Learn. **20**, 273–295 (1995)
11. V.N. Vapnik, *The Nature of Statistical Learning Theory [M]* (Springer-Verlag, New York, 1999)
12. R. Chen, Y.D. Sun, Modeling and precision influencing factors of engine support vector machine [J]. J. Cent. South Univ. **41**(4), 1391–1397 (2010)
13. X.G. Zhang, On statistical learning theory and support vector machines [J]. J. Autom. **26**(1), 32–42 (2000)
14. HU Tangao, Pan Yaozhong, Zhang Jinshui, etc. Integration of soft and hard classifications using linear spectral mixture model and support vector machines[J]. Spectrosc. Spectr. Anal., 2011, 31(2):508–511
15. Y.H. Wang, J.Y. Gao, Nonlinear predictive control technology based on support vector machine [J]. Inf. Control. **33**(2), 133–136 (2004)
16. V.N. Vapnik, *Estimation of Dependencies Based on Empirical Data[M]* (Springer Verlag, Berlin, 1982)
17. Y.H. Jiang, *Machine Learning Method [M]* (Electronic Industry Press, Beijing, 2009)
18. J.A.K. Suykens, T. Van Gestel, J. De Brabanter, et al., *Least Squares Support Vector Machines [M]* (World Scientific Publishing Co Pte, Singapore, 2002)
19. J.A.K. Suykens. Nonlinear Modeling and Support Vector Machines[C]. IEEE Instrumentation and Measurement Technology Conference, Budapest, (2001), pp. 287–294
20. H.-S. Tang, S.-T. Xue, R. Chen, et al., Online weighted LS-SVM for hysteretic structural system identification [J]. Eng. Struct. **28**(12), 1728–1735 (2006)
21. J. Weston, C. Watkins. Support vector machines for multi-class pattern recognition. In: Proc European symposium on artificial neural networks, Bruges, Belgium, vol. 4(6) (1999), pp. 219–224
22. K. Crammer, Y. Singer, On the learnability and design of output codes for multiclass problems. Mach. Learn. **47**(2), 201–233 (2002)
23. T.G. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes. J. Artif. Intell. Res. **2**, 263–286 (1995)
24. W.H. Chih, J.L. Chih, A comparison of methods for multiclass support vector machines [J]. IEEE Trans. Neural Netw. **13**(2), 415–425 (2002)
25. J.C. Platt, N. Cristianini, J. Shawe-Taylor, Large margin DAG's for multiclass classification [J]. Adv. Neural Inf. Proces. Syst. **12**, 547–553 (2000)
26. Y. Shi, R. Eberhart, A modified particle swarm optimizer[C]. In:IEEE World Congress on Computational Intelligence, Ahchorage, AK, USA, (1998), pp. 69–73
27. C. Sanchez-Hernandez, D.S. Boyd, G.M. Foody, One-class classification for mapping a specific land-cover class: SVDD classification of fenland[J]. IEEE Trans. Geosci. Remote Sens. **45**(4), 1061–1073 (2007)
28. J. Zhou, J. WANG, Reliability and safety research for passenger compartment door of shanghai metro vehicles [J]. Electric Locomotives Mass Transportation Veh. **4**, 002 (2006)
29. T. Denton *Advanced Automotive Fault Diagnosis: Automotive Technology: Vehicle Maintenance and Repair[M]* (Routledge, 2016)
30. V.R. Vuchic, *Urban transportation systems and technology[M]* (Wiley, Hoboken, 2007)
31. E.J. Joung, G.D. Kim. A Study on development of modularized electric plug—in door[A], International Conference on Electrical Machines and Systems (ICEMS), Incheon, South Korea, 2010. IEEE, 2010:4pp
32. E.J. Joung. Sliding sliding step for the disabled in the railway vehicle[A]

Chapter 4
Train Reliability and Safety Analysis

4.1 Introduction

With the rapid development of China's high speed railway industry and the rapid increase of EMU demand, the safety and reliability of high speed train system has attracted more and more attention. At present, the railway system in developed countries has formed a relatively perfect safety assessment and management system, and has developed a series of feasible technical standards for safety assessment. The international standardized management system, IEC 61508 standard, released in 2000. Later, a series of safety standards: EN 50126, EN 50128, EN 50129 and EN 50159 were launched for different railway transportation applications by European Committee for Electromechanical Standardization, which were qualified by the IEC organization and applied by European railways. Besides, the rules and standards of series of EN were used for reference in Taiwan area of China.

Up to now, the standard IEC 61508 has been introduced into the safety management of China's railways, and the national standard GB/T 20438 has been worked out. However, there is still lack of targeted analysis method and corresponding industry and national standards for the safety and reliability analysis of high speed train system, there is a pressing need to establish a system of operational and designed safety and reliability analysis of traffic train to support the safety and reliability design and operation of traffic train.

4.1.1 Reliability and Safety Standards of European Railway System

Compared with the present domestic situation of the safety and reliability evaluation system research, the safety and reliability evaluation system has been developed a relatively perfect safety assessment and management system abroad. The above EN

Y. Qin, L. Jia, *Active Safety Methodologies of Rail Transportation*, Advances in High-speed Rail Technology, https://doi.org/10.1007/978-981-13-2260-0_4

standard, EN 50126, EN 50128, EN 50129 and EN 50159 which advocate the concept that safety to a certain degree can be measured by reliability indices, that is, the basic idea is functional safety, safety integrity and safety ensured by technology.

Take the standard EN 50126 as an example, it is the specification and description of RAMS in the railway applications. RAMS is short of reliability, availability, maintainability and safety, which defines the requirements of RAMS in each phase of the system's safety lifecycle. RAMS is an important feature of system service quality, which can be obtained through the system's safety lifecycle by the designed idea and technical method. In order to achieve the required RAMS, it is necessary to make some effective control of the influence factors of RAMS in the whole system, that is, the random failure and system failure. As for the standard EN 50126, the system life cycle can be divided into 14 stages, and each stage has its own work, covering the whole safety lifecycle from the initial design to the waste of the system, represented by the 'V' shape. The branch from top to bottom can be often called development, which is a gradual refinement process, starting from the system concept until the manufacturing of the system components. The branch from bottom to top represents the assembly, installation, acceptance and operation of the whole system. The standard EN 50126 has been applied in China and revised the national standard TB/T 3133 in China.

4.1.2 System of Train Operational Reliability and Safety Analysis

The system of operational safety and reliability analysis of traffic train can be regarded as a system engineering methodology system for traffic train's system safety, which covers the general steps, specific stage and knowledge scope of operational safety and reliability analysis for traffic train. Traffic train is a complex electromechanical system, whose safety evaluation system is not only related to the system operational maintenance, but also refers to the relationship between system reliability and safety. Therefore, the system of operational safety and reliability analysis of traffic train can be constructed as a three-dimensional structure, which can be shown in Fig. 4.1.

4.1.2.1 Data

As for the system safety assessment of traffic train, the corresponding data are needed. Data acquisition usually takes a certain period of time, and the structural designed data are prepared for the structural reliability analysis of traffic train system. The data of operation, fault and maintenance can be collected during the

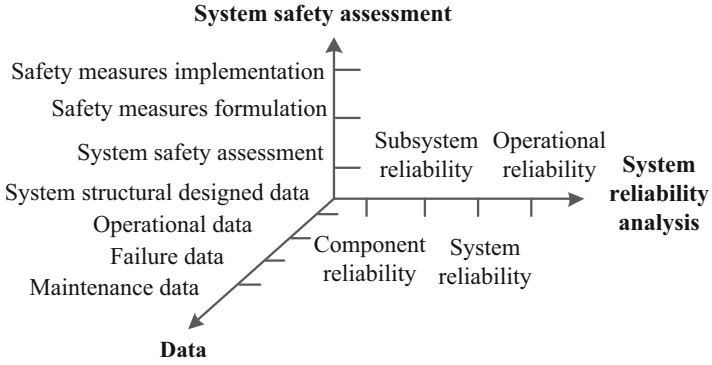


Fig. 4.1 Three-dimensional structure of operational safety and reliability analysis

running of traffic train, which can be also prepared for the system safety evaluation of traffic train.

4.1.2.2 System Reliability Assessment

The system reliability assessment is based on the operational failure data after the acquisition and processing, including the reliability of the component, subsystem, system of traffic train and the operational reliability of traffic train.

4.1.2.3 System Safety Assessment

The system safety assessment is mainly based on the safety affected data and system reliability assessment, the general steps of which includes system safety assessment, safety measures formulation and implementation.

4.1.3 Procedure of Train Operational Reliability and Safety Assessment

The procedure of operational safety and reliability assessment of traffic train is a complete closed loop process, which includes the failure description and recording of traffic train, the safety and reliability analysis of the component, subsystem, system of traffic train and the operational reliability of traffic train, and application of traffic train's safety and reliability, which can be shown in Fig. 4.2.

The first step in the operational safety and reliability assessment of traffic train is the failure data analysis, which needs to standardize the description and record and

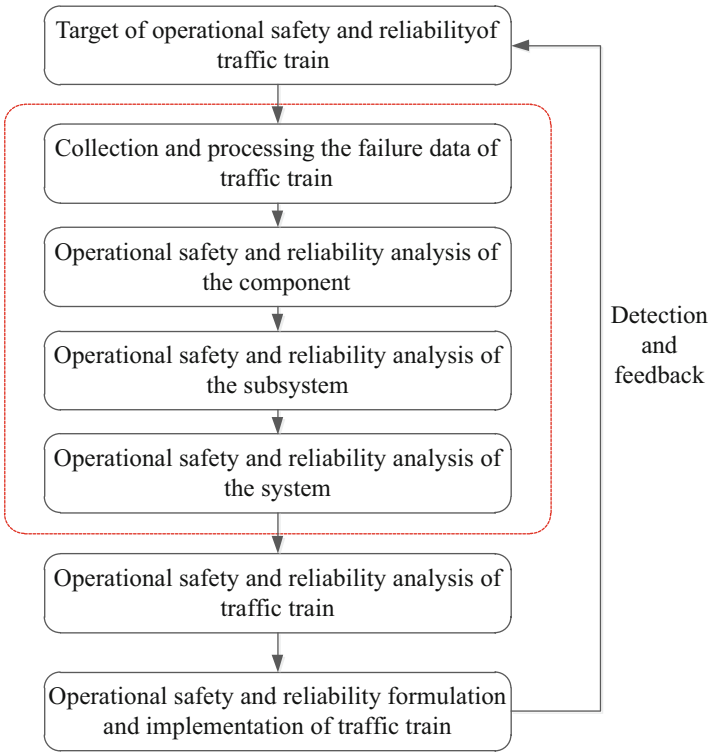


Fig. 4.2 The procedure of operational safety and reliability assessment of traffic train

analyze the failure mode of traffic train. The purpose of the reliability analysis of the component is to explore the design life and failure time interval of the related component, which can provide the basis for the reliability analysis of the component and system reliability analysis of traffic train. Besides, the reliability analysis of the component is the process to look for, record and describe the component failure and reliability, which is the premise for the system reliability analysis based on the coupling of each component, as well as the active maintenance of traffic train. As for the system reliability analysis, it is to identify the dangerous source in the traffic train system in the operation period, and reduce the risk to an acceptable level, which can avoid major casualties and property losses.

The operational safety and reliability assessment of traffic train is to compare the results of assessment and the safety and reliability standard set in advance, or compare the risk degree among each component, so as to determine the safety level of each component and the whole system. The safety measures formulation and implementation of traffic train is to take measures based on the assessment results, which can be affected by many factors. Therefore, relevant evaluation information needs to be collected and updated timely.

4.2 Reliability Analysis and Prediction of Bogie Frame

Bogie frame is the main load bearing component, the installation requirement of other components. Bogie frame not only supports the vehicle body, but also passes the vertical and longitudinal forces between the vehicle body and the wheel, and the reliability of which directly affects the performance and safety of the locomotive.

The research of bogie frame's reliability mainly focuses on the fatigue life analysis and sensitivity analysis [1–3], the lack of uncertainty in life data was considered. The key point of the research of bogie frame failure rate prediction lies on how to use artificial intelligence accounting to set up a prediction model with high accuracy. Typical prediction methods include time order prediction, neutral network prediction. BP is of many advantages such as being fast in convergence rate, small in absolute error, accurate in the prediction of developing tendency in failure rate [4]. However, a problem that we are blind in selecting the training parameters will appear if we only use BP as prediction model.

Therefore, survival analysis is used for reliability analysis of bogie frame to solve uncertain life time problem. A prediction model of failure rate composed of PSO-BP is came up with in this paper, so BP prediction model get developed, furthermore it can be used to predict the failure rate with high accuracy.

4.2.1 Reliability Analysis of Bogie Frame Based on Survival Analysis

4.2.1.1 Survival Analysis Theory

Survival analysis theory is developed as a new branch of mathematical statistics in the past three decades, which focuses on statistical analysis of randomly censored data [5]. Survival analysis not only can be used in biological and medical fields, but also can be used in engineering sciences, such as reliability engineering. Survival analysis is used for reliability analysis of bogie frame to solve uncertain life time problem, resulting in a more reasonable assessment results.

Survival analysis focuses on non-negative random variable T, count according to the object observed. Four types of values are obtained by observing each of the individual's life.

- Complete data. The exact value of individual life is observed.
- Right censored data. The exact value of individual life is not observed; only know that is greater than a specified number, denoted t^{+}.
- Left censored data. The exact value of individual life is not observed; only know that is less than a specified number, denoted t^{-}.

- Interval censored data. The exact value of individual life is between the two numbers Probability density function $p(t_j)$, survival function $S(t_j)$, failure rate function are

$$p(t_j) = p(T = t_j) = D_j/N, 1 \le j \le n \tag{4.2.1}$$

$$s(t_j) = R(t_j) = P(T > t_j) = \sum_{t_j > t} p(t_j) \tag{4.2.2}$$

$$\lambda(t_j) = \frac{\lim\limits_{\Delta x \to 0} P(t_j < T \le t_j + \Delta t | T t_j)}{\Delta t} = -\frac{\mathrm{d}}{\mathrm{d}x} \log(R(t_j)) \tag{4.2.3}$$

The failure of bogie frame cannot be found during train operation, only can be found through maintenance. Therefore, the failure time data of bogie frame can be expressed as

$$(L_i, R_i] \tag{4.2.4}$$

Where, L_t is time of last maintenance of i-th bogie frame; R_i is time of this maintenance of i-th bogie frame.

4.2.1.2 Maximum Likelihood Estimation

Assuming that, the number of life time data is N, n_1 of which are right censored, n_2 of which are left censored, and n_3 of which are interval censored. The likelihood function is

$$L(\lambda) = \prod_{i=1}^{n_1} [1 - F(L_i, \lambda)] \cdot \prod_{i=n_1+1}^{n_1+n_2} F(R_i, \lambda) \cdot \prod_{i=n_1+n_2+1}^{N} \times \ln [F(R_i, \lambda) - F(L_i, \lambda)] \tag{4.2.5}$$

Take the logarithm of both sides of Eq. (4.2.5)

$$\ln L(\lambda) = \sum_{i=1}^{n_1} \ln [1 - F(L_i, \lambda)] + \sum_{i=n_1+1}^{n_1+n_2} \ln F(R_i, \lambda) + \sum_{i=n_1+n_2+1}^{N} \times \ln [F(R_i, \lambda) - F(L_i, \lambda)] \tag{4.2.6}$$

Take the logarithm of both sides of Eq. (4.2.6) and assume the result of the formula is Eq. (4.2.7).

$$\frac{\mathrm{d}\ln L(\lambda)}{\mathrm{d}\lambda} = 0 \tag{4.2.7}$$

The maximum likelihood estimates of the parameters are obtained.

4.2.1.3 Goodness of Fit Test

Anderson–Darling (A–D) test method in Minitab software focuses on degree of data subject to a particular distribution. A–D statistic is smaller, the degree of data subjects to the distribution is better. The formula of A–D statistic [6] is shown as follow.

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(x_i) + \ln(1 - F(x_{n+1-i}))] \tag{4.2.8}$$

Where, $F(x_i) = \Phi\left(\frac{x_i - \bar{x}}{\sigma}\right)$ is cumulative distribution function subject to normal distribution.

4.2.2 Failure Rate Prediction of Bogie Frame Based on BP and PSO-BP Methods

4.2.2.1 BP Neural Network

BP neural network have hierarchical feed forward network architecture, is suitable for nonlinear prediction. In the classical structure of BP neural network, the output of each layer is sent directly to each neuron in the layer above. While there can be many layers, but the process can be done with a minimum of three layers: one layer that receives and distributes the input pattern, one middle or hidden layer that captures the nonlinearities of the input/output relationship, and one layer that produces the output pattern.

For a group of contiguous sequence $x(i)$, the $x(i)$, $x(i + 1)$, $x(i + 2)$, $\ldots x(k + i - 1)$ *as* an input vector, $x(k + i)$ as the output value, in order to establish the training sample, and BP prediction model is described as

$$x(k+i) = f(x(i), x(i+1), \ldots, x(k+i-1)) \tag{4.2.9}$$

4.2.2.2 Basic Principles of PSO

PSO was proposed by Dr. Eberhart and Dr. Kennedy, originated from the study on the behavior of birds foraging. Assuming a D-dimensional search space, there are N particles forming a population in which the position of each particle is represented as a D-dimensional vector with $X_i = (x_{i1}, x_{i2}, \ldots x_{iD})$ representation; particles' flight speed recorded as $V_i = (v_{i1}, v_{i2}, \ldots v_{iD})$; particle i have been found as the optimal position so far, denoted P_{best};the best position of whole swarm have been searched so far, denoted by g_{best}; after finding the two positions, the particles update their speed and position according to Eqs. (4.2.10 and 4.2.11).

$$v_{id}^{k+1} = w^* v_{id}^k + c_1 r_1 \left(p_{id}^k - x_{id}^k\right) + \mathrm{c}_2 r_2 \left(p_{gd}^k - x_{id}^k\right) \tag{4.2.10}$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \tag{4.2.11}$$

Where, v_{id}^{k+1} is the flight speed of i-th particle in $(k + 1)$ generations; k_{id}^{k+1} is the position of i-th particle in $(k + 1)$ generations; p_{id}^k is the best position of i-th particle to k-generation; p_{gd}^k is the best position populations to k-generation; $p_{id}^k - x_{id}^k$ is individual cognition; $p_{gd}^k - x_{id}^k$ is the population cognition; w is the inertia weight; v_{id}^k is the velocity of the particle; c_1 and c_2 are learning factors; r_1 and r_2 is the random number of uniformly distribution in [0,1]; $i = 1,2,\ldots, N$.

4.2.2.3 PSO-BP Prediction Model

PSO-BP prediction model is divided into three modules: (1) data preprocessing module is that experimental data are normalized to obtain the required model for training and testing data sets. (2) PSO parameter optimization module is the use of the PSO algorithm to optimize the BP parameters, then the optimal parameters pass to the BP. (3) BP prediction module using the training data set has been training to obtain prediction model, and use the test data set for prediction. The structure of prediction model has been shown in Fig. 4.3.

In PSO parameter optimization module, the fitness function should be determined. PSO algorithm used fitness value to evaluate the merits of an individual or population in the search process of evolution, and as the basis of the particle velocity and position changes, it gradually evolved to the optimal solution. The root mean square error (RMSE) is the fitness function in this paper.

$$\mathrm{RMSE}(y, y_m) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y(i) - y_m(i))^2} \tag{4.2.12}$$

Where y is the actual value of the training sample, y_m is the predicted value of the model; N is the number of data samples. The RMSE is smaller, which means the higher prediction accuracy, the predicted value closer to the target value.

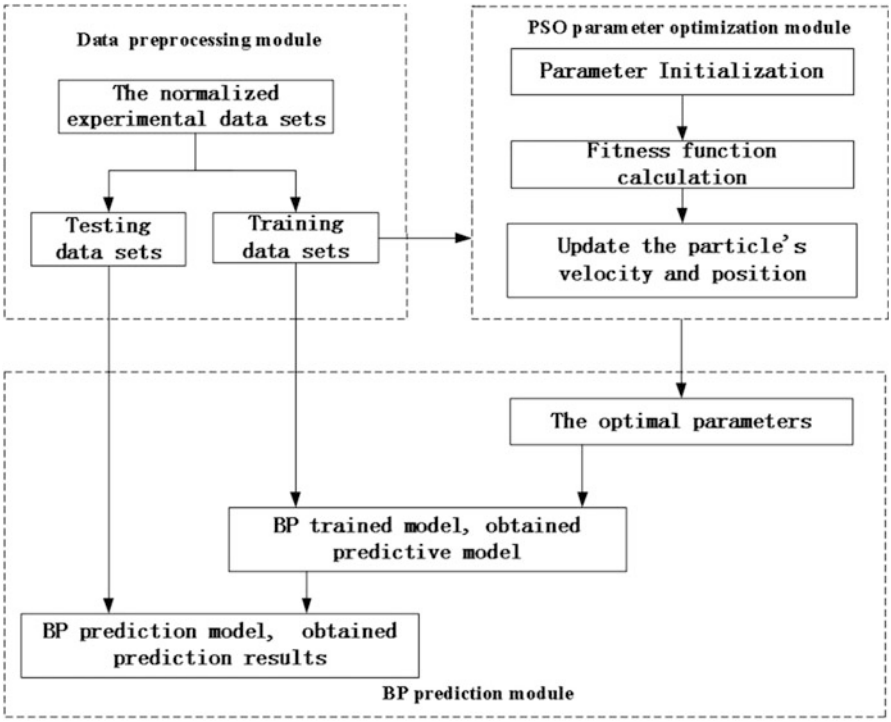


Fig. 4.3 The structure of PSO-BP prediction model

Based on the above prediction model, failure rate prediction algorithm steps based on PSO-BP will be shown in Fig. 4.4.

4.2.3 Case Study

4.2.3.1 Analysis of Bogie Frame

The basic idea of bogie frame reliability frame is as follows: Firstly, censored data of life time was counted, and then the maximum likelihood estimation method was used to estimate the parameters of the life time, and finally find the best distribution model according to the A–D statistic.

The life time of bogie frame includes two types of censored data, interval censored data and right censored data, life time of censored data is shown in Table. 4.1.

The censored life time data is inputted into Minitab software. Maximum likelihood estimation method is used for the exponential, lognormal, two-parameter Weibull distribution and three-parameter Weibull distribution parameter estimation

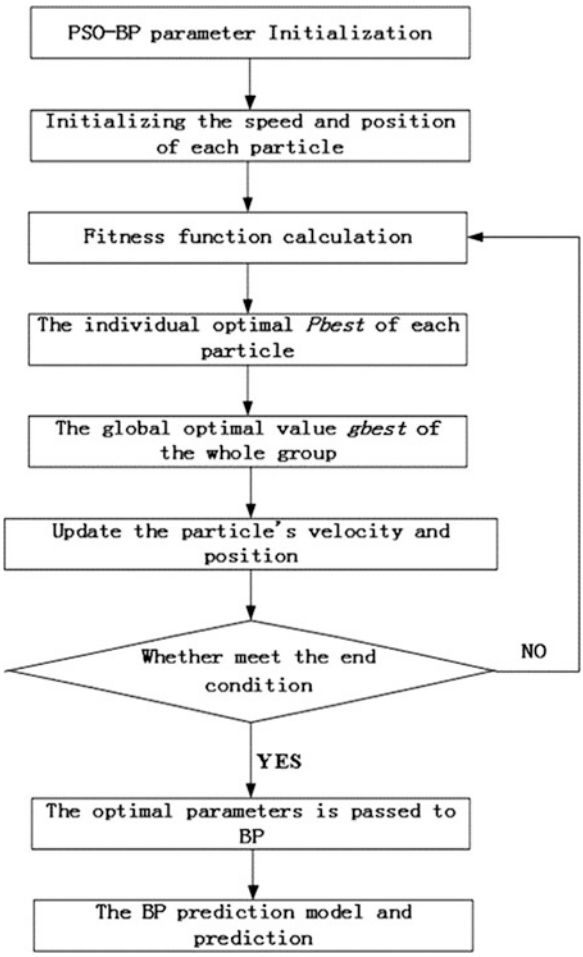


Fig. 4.4 Failure rate prediction algorithm steps based on PSO-BP

Table 4.1 Censored life time data

No.	Life time
1	(4,]
2	(5,6]
. . .	. . .
59	(13,]

and goodness of fit test. The result is shown in Fig. 4.5, the worst of the fit is exponential distribution. The A–D statistic values of exponential, lognormal, two-parameter Weibull distribution and three-parameter Weibull distribution are 0.530, 0.500, 0.820, 0.484. Therefore, the best distribution of bogie frame life time is three-parameter Weibull distribution.

The life time of bogie frame subjects to three-parameter Weibull distribution, shape parameter β is 1.11034, scale parameter η is 17.3206, threshold γ is 0.576497.

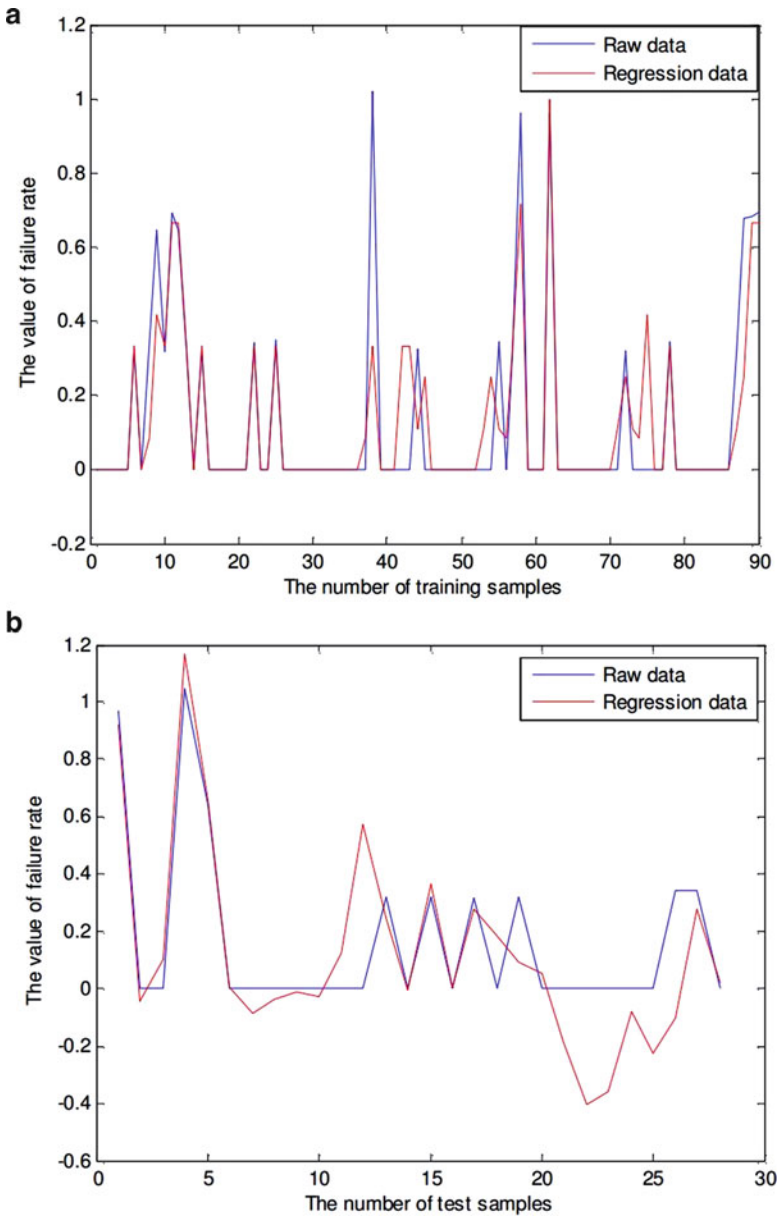


Fig. 4.5 The comparison between the predicted and actual value. (**a**) The training set predictive data comparison. (**b**) The testing set predictive data comparison

The probability density function, survival function (or reliability function), failure rate function are shown as Eqs. (4.2.13, 4.2.14, and 4.2.15).

$$f(t) = \frac{1.11034}{17.3206^{1.11034}}(t - 0.576497)^{0.11034} e^{-\left(\frac{t-0.576497}{17.3206}\right)^{1.11034}} \tag{4.2.13}$$

$$R(t) = e^{-\left(\frac{t-0.576497}{17.3206}\right)^{1.11034}} \tag{4.2.14}$$

$$\lambda(t) = \frac{1.11034}{17.3206}\left(\frac{t - 0.576497}{17.3206}\right)^{0.11034} \tag{4.2.15}$$

4.2.3.2 Prediction of Failure Rate

To verify the validity of PSO-BP bogie frame failure rate prediction model, Line 2 of one metro corporation, for example, the data of year 2010–2012 was selected as the original experimental data sample. Using Eq. (4.2.15) on failure rate value data are normalized, taking the dimension input vector is 10, building a experimental sample set whose capacity is 118 groups, where the data of the first 90 groups as the training sample set, the data of the last 28 groups as the test sample set.

$$x_i^{\text{scal}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{4.2.16}$$

In order to measure the accuracy of the prediction model, this section selected root mean square error as predictive accuracy of fitness function evaluation model. However, due to the small root mean square error value, it is difficult to visually give the difference between the actual value of bogie frame failure rate and the predicted value of PSO-BP failure rate prediction model, and analysis the correlation through linear regression analysis. Therefore, using the correlation coefficient R to further measure the curve fitting and linear regression analysis between the predicted value and the actual value of the prediction model.

$$R(y, y_m) = \frac{\sum_{i=1}^{N}(y(i) - \bar{y})(y_m(i) - \bar{y}_m)}{\sqrt{\sum_{i=1}^{N}(y(i) - \bar{y})^2 \cdot \sum_{i=1}^{N}(y_m(i) - \bar{y}_m)^2}} \tag{4.2.17}$$

Where, $\bar{y}$ is the average of the test sample, $\bar{y}_m$ is the average of the predicted value. R-value is closer to i, which means that the model prediction has higher accuracy, the closer to the actual system.

Set PSO parameters are: population size is 30; inertia weight initial $w_{\max} = 0.9$; inertia weight final value $w_{\min} = 0.4$; learning factor $c_1 = c_2 = 2$; maximum velocity $V_{\max} = 5$; maximum speed and position limits scaling factor $k = 0.6$; maximum evolution generation $T_{\max} = 20$. Training process adaptation curve is shown in Fig. 4.6.

Survival analysis is used for reliability analysis of bogie frame to solve uncertain life time problem. It turns out that the best life distribution model of is three-

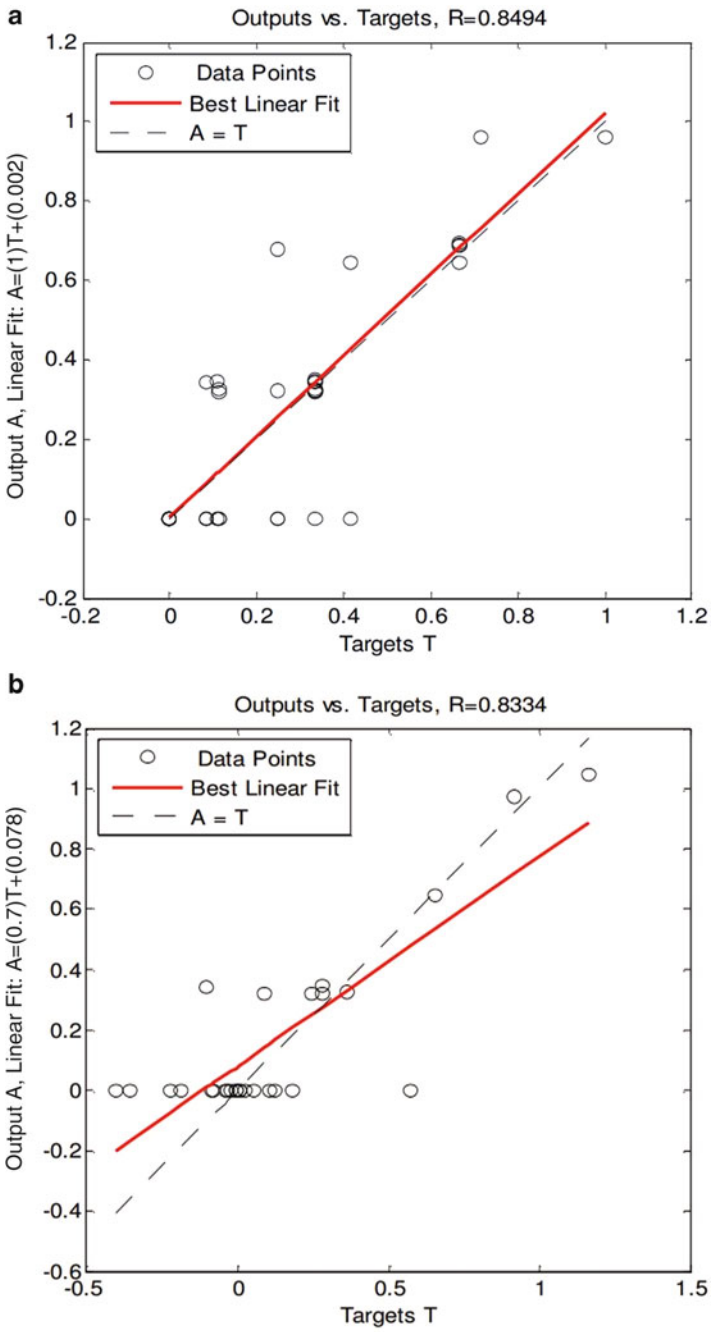


Fig. 4.6 The correlation curve of predicted value and the actual value. (**a**)Correlation of the training set. (**b**) Correlation of the testing set

parameter Weibull distribution. A prediction model of failure rate composed of PSO-BP is come up with to predict the failure rate of bogie frame accurately.

4.3 Residual Life Prediction of Rolling Bearings Based on GA-BP

Rolling bearing, as a critical component of metro trains, is widely applied to the frequent transmission of heavy load in the mechanical systems of trains. The state of rolling bearing can directly affect the train's operational safety. Serious accidents such as train derailment, overturn and operational collision will be brought about as a result of a little problem from rolling bearings. As for the bearing failure, most maintenance measures at present that the rolling bearings need to be serviced at the depot regularly, which makes it difficult to grasp the change of the operational state. Due to the untimely and inaccurate maintenance measures, a large number of rolling bearings operated in the field can work with problems or be replaced far from the life expectancy, which can bring about the hidden danger and a waste of resources for the safety operation of the train. Therefore, accurate and effective residual life prediction based on condition monitoring for the rolling bearings of train is urgent in demand for the sake of reducing train accidents, improving operational safety and reducing maintenance costs.

Currently, some sophisticated and formulaic life prediction algorithms have been applied for general rolling bearings. But the algorithm parameters are often fixed, these algorithms cannot adapt to the situation of frequent change of state variables (such as load, speed, temperature, noise, etc.). And for the traditional life prediction methods based on reliability theory, only event statistical data can be used. For a specific rolling bearing, prediction accuracy of this type of approaches based on reliability may not satisfy the site requirements because the operating state information is ignored. Therefore, the residual life prediction method based on GA-BP is put forward in this paper.

4.3.1 Residual Life Prediction Model of Rolling Bearings Based on GA-BP

With the harsh operating conditions, complex structure and sophisticated mechanism, the rolling bearings of rail vehicles are generally designed to be non-standard forms. And the types of rolling bearings are different for different installation positions. Generally, according to the shape of rolling element, rolling bearings can be divided into cylinder rolling bearings, tapered rolling bearings, and spherical rolling bearings. Usually, spherical rolling bearings and cylinder rolling bearings are often used to support the motor rotor for traction motor. For the bogie system, the rail

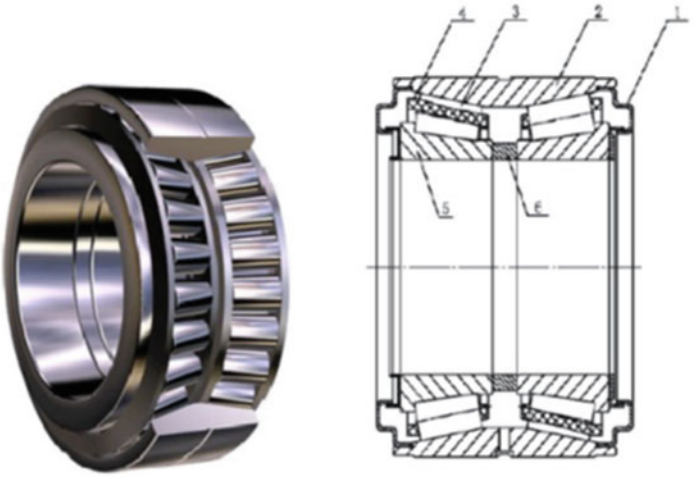


Fig. 4.7 The basic configuration of the rolling bearings. *1 Sealing ring; 2 The outer race; 3 The rolling elements; 4 The cage; 5 the inner race; 6 The middle spacer race*

vehicles operating conditions with high load, high speed operation, frequent starting should be considered, so cylinder rolling bearings and tapered rolling bearings will be selected to withstand and transmit various loads between wheel and the truck parts.

Figure 4.7 shows the basic composition of a rolling bearing (tapered rolling elements) of rail vehicles. Rolling bearing is mainly composed of outer race, inner race, rolling elements, and cage. Usually, the inner race assembles on the shaft neck and rotates together with the axis. The outer race is fitted within the axle boxes or bearing block and plays a supportive role. The rolling elements located between the inner and outer race. And when the shaft neck and the inner race rotate together and outer race does not rotate, the rolling elements not only can rotate around its axis, but also can scroll around the inner and outer raceway. The size and number of rolling elements determine the carrying capacity of rolling bearing. The cage function is to make each of the rolling elements distributed evenly, and prevent the collision and friction, and keep the rolling elements roll well.

As for the traditional BP neural networks, the weights and thresholds of the initial values are randomly selected. It can fall into the local minimum with the slow convergence of the model if these parameters are not appropriate. As mentioned in the Sect. 4.2.2, Genetic Algorithms (GA) is a global optimization algorithm based on the principle of natural selection and natural genetic mechanism, which can simulate the life evolution mechanism and achieve the optimization of specific target in artificial system. The essence of GA is to get the global optimal solution based on group search technology and the principle of survival of the fittest [6]. Therefore, GA algorithm is applied to get a good initial weight distribution, and BP neural network model based on LM training algorithm is used to adjust weights reasonably to find the global optimal solution in the chapter. The basic procedure of prediction model

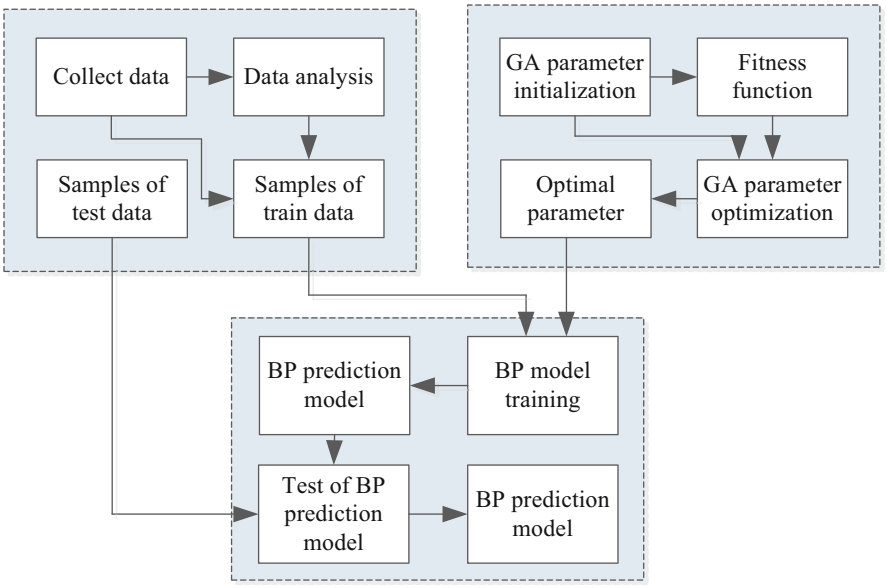


Fig. 4.8 Process of GA-BP model

construction based on GA-BP model can be expressed as follows, which can be shown in the Fig. 4.8.

- Collect and select sample data. Preprocess the above data and divide training samples and test samples.
- Determine the GA algebra, initial population size, fitness function and so on. Determine the related coefficients and parameters of BP neural network model optimized by GA.
- Input the optimal parameters obtained by GA and data preprocessed into BP neural network model so as to get BP neural network trained in the process.
- Test the prediction accuracy of the trained model by the pretreated test data and construct the final residual life prediction model.

4.3.2 Case Study

The experimental data of rolling bearing full life cycle was proposed by Center for Intelligent Maintenance Systems (IMS) [7]. The bearing used in the experiment is Rexnord ZA-2115 double row rolling bearing, and the operational speed is 2000 rpm/min. The field bench of life data collection can be shown in Fig. 4.9. To collect the vibration data, 353B33 high sensitivity acceleration sensor produced by PCB is applied in the axle box housing. The 6062E data acquisition card of National Instruments Corporation is selected to collect data and the sampling frequency was 20 k Hz. The full life cycle vibration data is composed of data sets collected at three

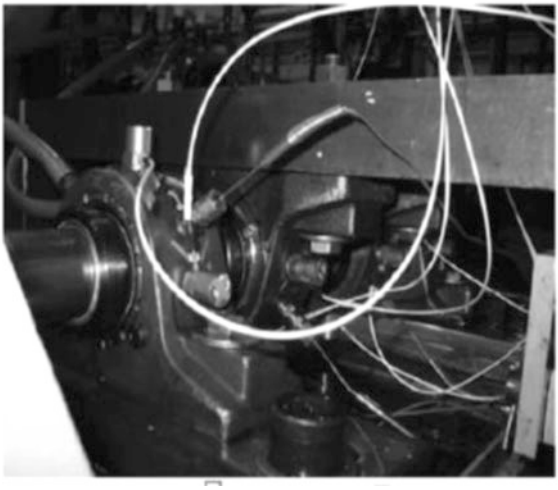


Fig. 4.9 Life data collection of rolling bearings

different time intervals with 20 K Hz sampling frequency, and the data was sampled every 10 min, and each sampling time is 1 s.

This chapter applies vibration data from No. 3 bearing of the third data sets. The process of the data acquisition began from the bearing installed until the bearing failure takes about 740 h, nearly for a month (from the beginning of March 4, 2004 to the end of April 4, 2004). To reduce the computational burden of model, a segment vibration data is selected every 24 segments (about 4 h) from total 4448 segments vibration data, and 186 segments vibration data has been selected, where each segment contains 20,480 vibration acceleration data points.

After full life vibration data are collected and selected, the five dimensional features of 186 segments vibration data can be extracted based on statistical model method:

- T1: standard deviation of normal distribution.
- T2: log likelihood value of normal distribution.
- T3: scale of logistic distribution.
- T4: log likelihood value of logistic normal distribution.
- T5: log likelihood value pf Weibull distribution.

Figure 4.10 shows the variation curve diagrams of the extracted five dimensional features with the bearing running time respectively. The features of the Fig. 4.10a, c change lowly before the bearing operates to 400 h, while changes after 400 h are obvious. The feature trends of the Fig. 4.10b, d are consistent with the Fig. 4.10c. In summary, the variation of five dimensional features extracted based on statistical model is sensitive, which can effectively reflect the residual life of the bearing. Therefore, the five dimensional features extracted based on statistical model are applied as inputs in the GA-BP neural network model for the residual life of the bearing.

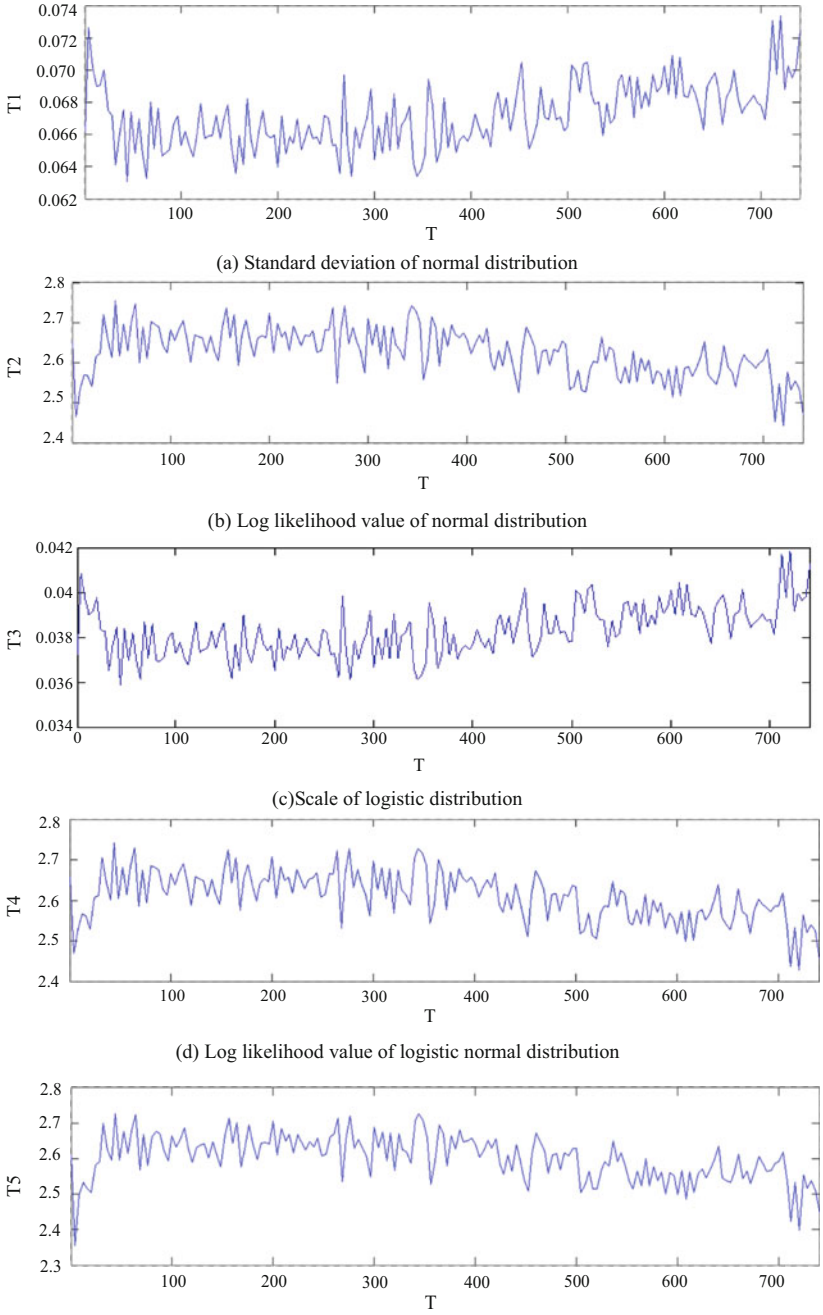


Fig. 4.10 Feature trends of the five dimensional features. (**a**) Standard deviation of normal distribution. (**b**) Log likelihood value of normal distribution. (**c**) Scale of logistic distribution. (**d**) Log likelihood value of logistic normal distribution. (**e**) Log likelihood value Weibull distribution

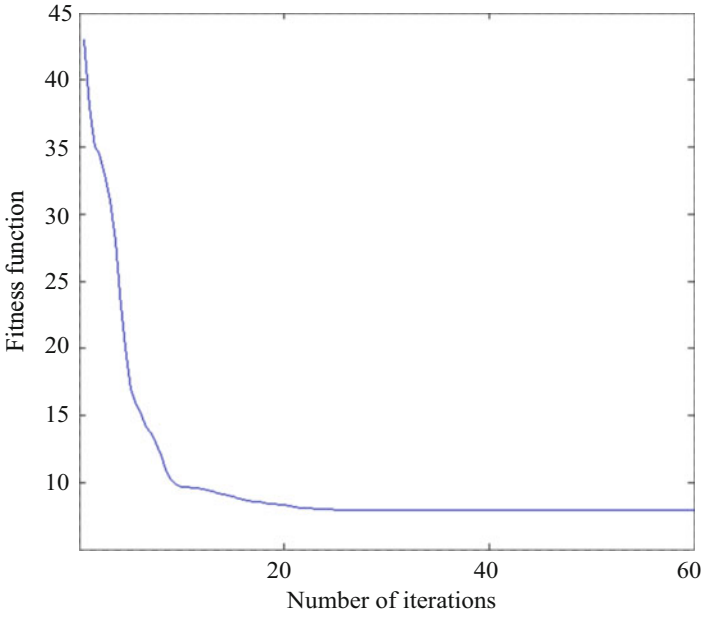
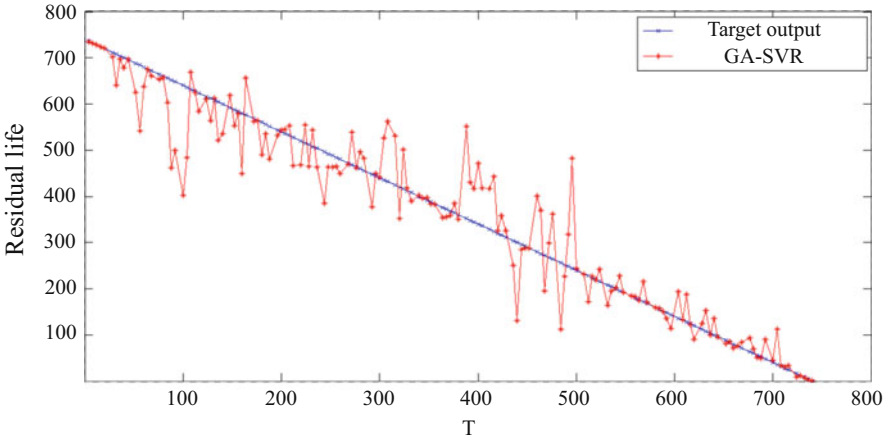


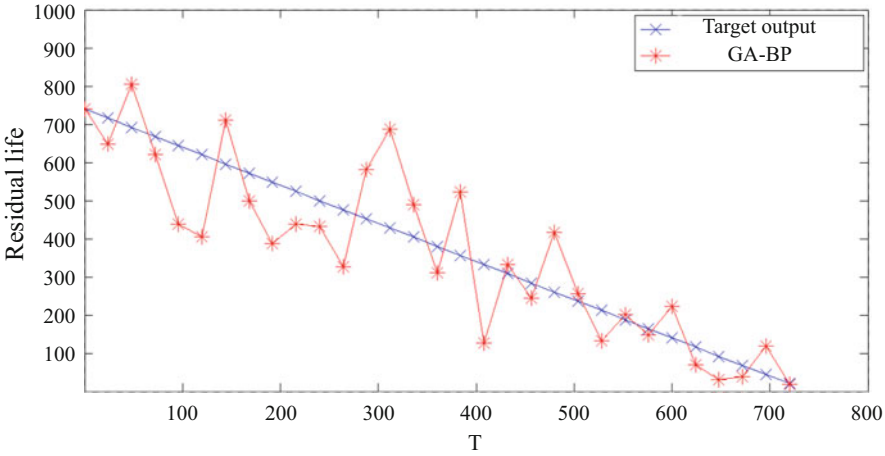
Fig. 4.11 Training process of GA

In order to overcome the shortcomings of the traditional BP neural network, this chapter applies GA to optimize the connection weights of the BP neural network. In the process of residual life prediction model establishing, the coding method, initial population size, fitness function, individual selection method, individual crossover mode and mutation probability of GA need to be determined. This chapter takes entity coding method, where the individual value is in the range of [0,1], the individual coding length l = number of hidden units × input node number + number of hidden units × output node number + number of hidden units + output node number. The initial population size is 40, and the maximum generation is 60. The fitness function is the RMS value of the actual output compared with the target output. Roulette method is applied to select the excellent individual. The individual crossover mode is arithmetic crossover, where the crossover probability is 0.7 and the mutation probability is 0.05. The training process of GA is shown in Fig. 4.11, and the result can be stable when the number of iterations is over 25.

In order to facilitate the analysis of the residual life prediction effect based on GA-BP neural network model, 155 training samples and 31 test samples are applied in the experiment. Figure 4.12 shows the comparison between the target output of training samples and testing samples and GA-BP neural network respective, which indicates that the output of GA-BP neural network model drops followed by the target output during the whole life cycle (0 ~ 740 h), and varies in a slight float around the target output, which can better reflect the actual situation of bearing residual life. Figure 4.13 shows the difference between the target output and GA-BP



(a) Training sample

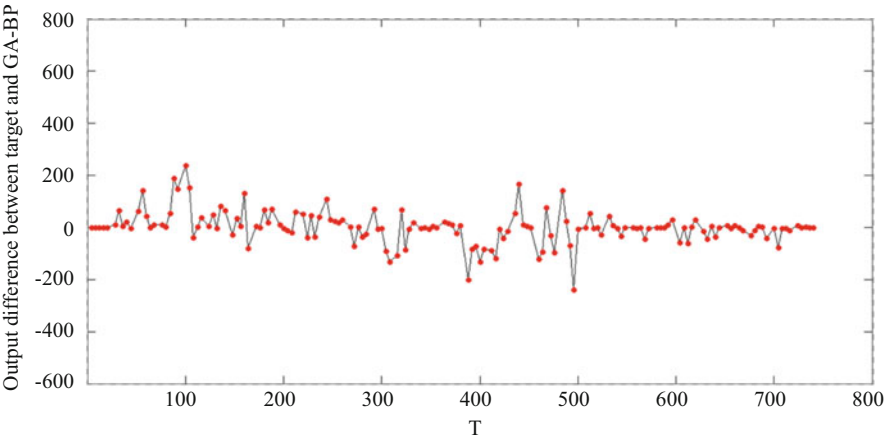


(b) Test sample

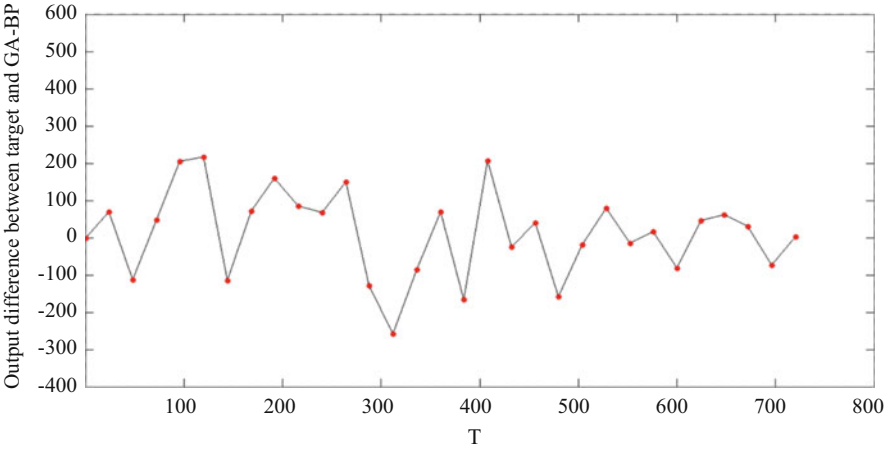
Fig. 4.12 The comparison between target output and GA-BP. (**a**) Training sample. (**b**) Test sample

of each training sample and test sample, which indicates that the change of difference is around 0 with slightly floating.

Similarly, in order to further evaluate the prediction accuracy and model performance of residual life based on GA-BP neural network, the RMSE and correlation coefficient R are presented in Table 4.2, and the corresponding results of correlation analysis can be seen in Fig. 4.14. The results show that the RMSE value of training samples and testing samples is 7.8873 and 9.6969 respectively, which is less than 10 and has been improved compared with the BP neural network model. The correlation coefficient of the training samples is 0.9500, which is over than that of BP neural network. The correlation coefficient of test sample number is 0.8662, which is over than that of BP neural network.



(a) Training sample



(b) Test sample

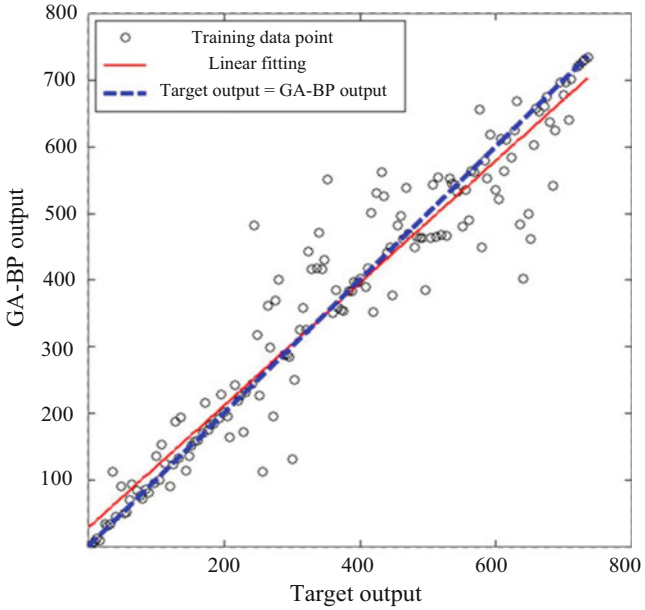
Fig. 4.13 Difference between target output and GA-BP. (**a**) Training sample. (**b**) Test sample

Table 4.2 The evaluation index of output based on GA-BP neural network model

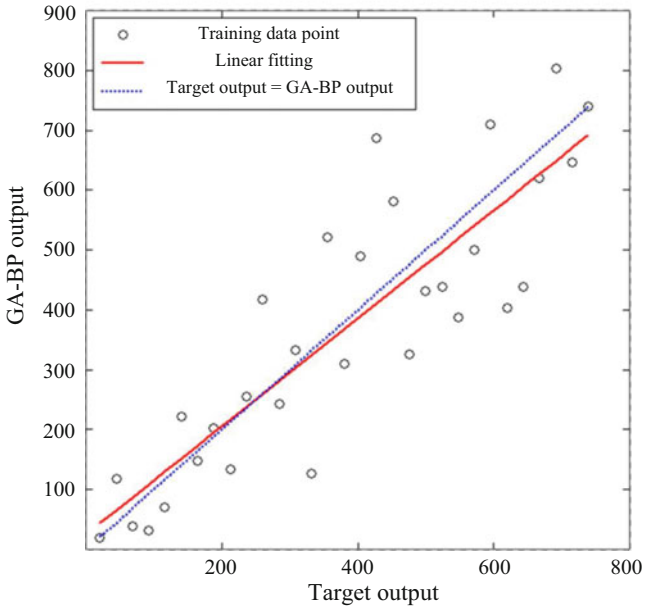
	Training sample	Test sample
RMSE	7.8873	9.6969
R	0.9500	0.8662

4.4 Operational Risk Assessment of High Speed Train

Nowadays, the flourish of economy and trade in China drives the booming of high speed railway to satisfy the great demand of transportation. High speed train provides services for a large number of passengers, which plays a crucial role in its safety operation. Great loss like train derailment, overturn and operation collision



(a) Training sample



(b) Test sample

Fig. 4.14 Correlation analysis of target output and GA-BP. (**a**) Training sample. (**b**) Test sample

will be brought about as a result of a little problem from the train. Hence, advanced study and analysis must be implemented to the significant issue for the risk assessment of high speed train in order to identify and control the risk.

The risk of high speed train refers to the likely outcome of probability of occurrence and the severity of consequence, which can be influenced by the interaction of various factors in the operation environment. Staff [8], environment [9], infrastructure [10] and train itself can be the major influence factors which can be divided into two main categories of internal and external influence factors. Operational risk assessment of high speed train is to identify the risk factors of the train, thus to get the risk ranking of each component in the system, for the risk of the train can be controlled and involved train maintenance plans can be prepared and updated periodically. During the process of risk assessment, some researchers had made many contributions in risk assessment of high speed train. Many traditional approaches such as fault tree analysis (FTA) [11], event tree analysis (ETA) [12] and Bayesian analysis [13], failure mode and effect analysis (FMEA) [14] and analytic hierarchy process (AHP) based on expert judgments [15] are used to get the risk state of trains. However, FTA, ETA and Bayesian analysis can be better aimed at the analysis of a particular accident rather than a whole equipment system. Besides, FMEA and AHP can be often carried out by qualitative analysis. However, high speed train is a complex electromechanical system of more than 40,000 interconnected components under different condition, which makes it more difficult for the above approaches to meet the technical demands of high speed train operational risk assessment with system analysis, quantitative analysis and period analysis. Thus, an efficient modified risk assessment approach for high speed train is urgent in demand.

Due to the various risk state and complexity of high speed train, multiple risk indices with qualitative and quantitative information based on the staff, environment, infrastructure and train itself may be a good choice to evaluate the risk of high speed train. However, multiple risk indices with qualitative and quantitative information always require to be transformed into the same type in most researches. The process of transformation may bring about information distortion or loss, which can bring about the calculation uncertainty and cause some impact on the results. In addition, the situations where the experts and engineers are not able to express their preference in the risk assessment can also bring about the uncertainty for the assessment results. Therefore, VIKOR approach [16], which is directed against the calculation uncertainty, expert preference and information indeterminacy has been put forward to deal with this problem. VIKOR approach, developed for multi-criteria evaluation on the basis of TOPSIS approach and overmatched it in the algorithm, can be well applied to the problem of high speed train risk assessment, since it is directed against concentrating the mixed quantitative and qualitative information, which can better deal with the various risk state and complexity of high speed train. Besides, VIKOR approach can well deal with the calculation uncertainty and expert preference by a maximum group utility of the 'majority' and a minimum individual regret of the 'opponent' [17]. Jahan et al. [18] applied VIKOR approach to evaluate the material selection by the hybrid information. Kavita [19] applied VIKOR approach to cope

with multiple criteria decision making problems and the result illustrated the effective of the method. Mohsen [20] adopted fuzzy VIKOR approach to rank and prioritize the failure modes in the FMEA, which has been proved to improve the applicability of the conventional FMEA approach. VIKOR approach provides a more practical way to solve the hybrid risk assessment with qualitative and quantitative information on the basis of various risk state and system complexity, which can be well applied to the problem of high speed train operational risk assessment in this paper.

As for the qualitative risk information which cannot be expressed distinctly by quantitative data in the risk assessment of high speed train, it can be expressed by the professional judgment of experts applied with the fuzzy sets. Fuzzy set theory was first proposed by Zadeh in 1965 [21], which was applied to describe inaccurate and qualitative information. In 1986, Atanassov [22] filled up the deficiency of fuzzy set in depicting fuzzy relationship and raised the theory of intuitionistic fuzzy set (IFS). IFS can concentrate its two aspects of information from membership and non-membership, which makes it more flexible in describing the fuzzy problems. Despite the successful study and application of IFS in risk analysis and assessment [23–25], the membership degree and the non-membership degree are difficult to be described by numerical values because of the complexity and fuzziness of the qualitative risk information on high speed train. Thus, type-2 intuitionistic fuzzy set (type-2 IFS) was proposed to deal with these problems in a better way. Type-2 IFS [26] possesses many advantages over type-1 IFS, as their membership functions and non-membership functions are themselves fuzzy, making it possible to model and minimize the effects of indeterminacy in fuzzy matters. Since interval number IFS [26, 27] was raised as an application of type-2 IFS to cope with the fuzzy problems, sometimes the extreme points value of the interval require to be too large or too small in order to cover the entire range of interval, which can further enlarge the interval range and affect the final result [28]. Therefore, triangular fuzzy number between 0 and 1 was applied in the IFS as another application of type-2 IFS to overcome this problem [29]. Triangular fuzzy number intuitionistic fuzzy set (TFNIFS) has been applied in some fields of science and technology for the fuzzy problems [30], there is few applications to the high speed train risk assessment yet. Accordingly, TFNIFS can be commendably applied to the problem of high speed train operational risk assessment in this paper.

During the process of the risk assessment, most of researches [20, 23–25, 28, 30] have been applied the static indices without involving the influence of time on the assessment, which may not fully reflect the facts of the risk state to a certain extent. The risk state of high speed train system and related risk factors with staff, environment and infrastructure can change constantly with the time, showing a dynamic risk feature. In addition, the recognition and understanding about high speed train of the relevant experts and decision makers can be more accurate and clear as time goes on. With the combination of time, it can be more systematic and effective to evaluate the mixed risk information. Therefore, a new ranking approach of dynamic VIKOR

based on the different periods is proposed for high speed train operational risk assessment in this paper.

As discussed above, we are devoted to studying the operational risk assessment of high speed train based on triangular fuzzy number intuitionistic fuzzy set and dynamic VIKOR approach in this paper, which is put forward to cope with the uncertainty and complexity of the risk information and rank the risk of system components. The assessment period can be chosen according to the maintenance plan, and the operational risk assessment index system of high speed train can be proposed based on the factors of staff, environment, infrastructure and train itself. The field test data of a specific high speed train is implemented and can provide technique support for the high speed train operational risk assessment with significant practice.

The rest of this paper can be organized as follows. In Sect. 4.2, basic problem of high speed train operational risk assessment are explained. In Sect. 4.3, application of dynamic VIKOR in high speed train operational risk assessment based on constant different periods is explained. In Sect. 4.4, a numerical study based on the field test data of a specific high speed train is implemented. In Section 5, some conclusions are made of the paper.

4.4.1 Basic Challenges of High Speed Train Operational Risk Assessment

The risk of high speed train refers to the likely outcome of probability of occurrence and the severity of consequence, which can be influenced by the interaction of various factors in the operation environment. The operational risk assessment is to identify the risk factors of high speed train and get the risk ranking of system components in order to control the risk, prepare and update the maintenance plan for the train. The maintenance schedule of a specific high speed train in China can be divided into five classes [31]: (1) the primary class operated every 2 days for visual inspection and functional test, (2) the second class operated around every year for component performance detection, (3) the third class operated around every one and a half years for bogie system disassembling inspection, (4) the forth class operated every 3 years for disassembling inspection of each system and (5) the fifth class operated every 6 years for the whole train maintenance. Due to the different schedule classes for high speed train maintenance, the component performance detection of the second class and the system disassembling inspection of the forth class need to be taken into account in the operational risk assessment of high speed train. Therefore, the dynamic operational risk assessment of high speed train can be carried out comprehensively for three periods according to the second and forth class maintenance schedule.

4.4.1.1 Construction of High Speed Train Operational Risk Assessment Index System

High speed train is a complicated mechatronics system, and the risk of the train can be influenced by the interaction of various factors which can be divided into two main categories of internal and external influence factors from staff [8], environment [9], infrastructure [10] and train itself.

Staff factor may be one of the most critical and pivotal factors among the various elements of high speed train. As for the operation of high speed train in a relatively closed environment, the activity range of passenger is only in the train carriage. However, train staffs like train driver and maintenance personnel are closely related to the train operation. It may cause large damage or even accident if the train driver or maintenance personnel is inexperienced or fatigue working with continuous long periods. Therefore, the staffs' operation skills as well as mental state can be applied to risk elements as staff factor. The operational skills of the staff can be comprehensively affected by the age, education and technical title. The mental state of the staff can be influenced by constant working period.

As one of the critical risk elements, environment factor has a crucial influence on the safety state of high speed train. During the operation of high speed train, the wind, heavy rain, snow, dust storms and other bad weather provide enormous challenges for the train system, which can result in the performance degradation and function failure. Thus, environment factor of high speed train synthetically based on the extreme weather, temperature and humidity should be taken into account.

Infrastructure in the operation of the train also has a significant impact on the risk state of the high speed train. Infrastructure factor involves various elements such as tracks, bridges, electrical facilities, signal facilities and so on. Above the mentioned factors, the tracks have directly relationship with the trains operation. The quality of the tracks can be referred to the track quality, which can affect the normal operation and safety state of high speed train as well as the comfortable trip of passengers. Therefore, track quality index (TQI) [10] can be applied to evaluate the risk of high speed train.

As for the internal risk factors of train itself, mean distance between failure (MDBF), mean time to restoration (MTTR), maintenance cost, fault detectability and the risk effect on system, people and environment are the inherent risk factors of high speed train, which can play a great important role in the safe operation of high speed train. MDBF is another form of mean time between failure (MTBF), as time is replaced by running distance. MDBF describes the probability of risk occurrence. In addition, MTTR, maintenance cost and the risk effect on system, people and environment of the train describe the severity of risk consequence. MTTR is the variable to measure the average time cost to repair the component. Maintenance cost focuses on the complexity and significance of the system. Risk effect on system,

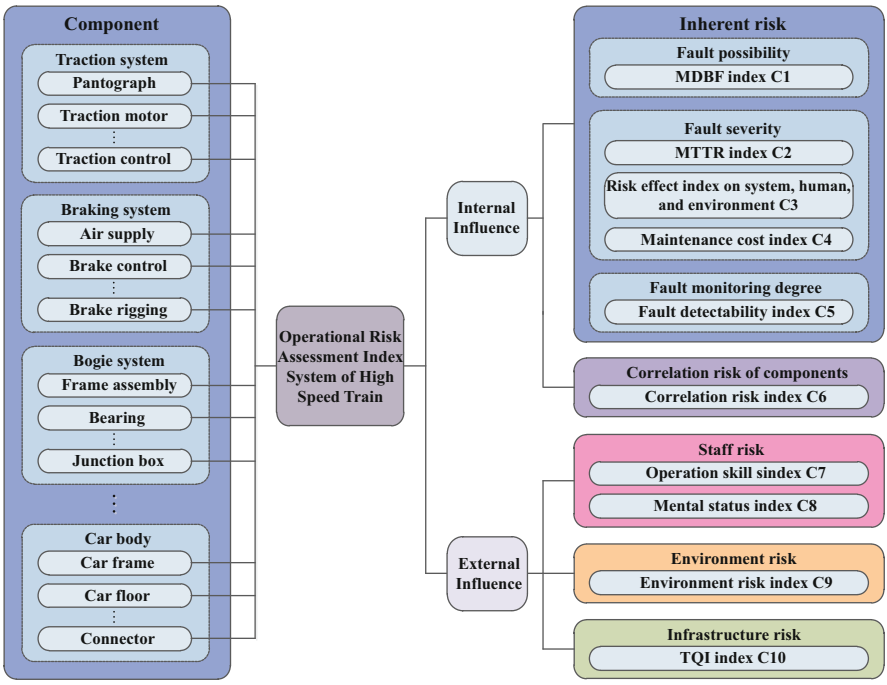


Fig. 4.15 Operational risk assessment index system of high speed train

people and environment reflects the consequence severity of the failure components. Besides the inherent risk factors above, high speed train is a complicated mechatronics system consisting of various parts and subsystems with different relationships and interactions. Hence, correlation risk index integrated with system structural risk and components failure is proposed in this paper.

As discussed above, the operational risk assessment index system of high speed train for each component can be set up based on the above factors, which can be shown in Fig. 4.15. The above factors consist of quantitative and qualitative information, dividing the assessment index into numerical, interval and fuzzy type. MDBF index, MTTR index, correlation risk index, operation skills index, mental state index and TQI index can be calculated by some specific formulas [32], the result of which can be presented as numerical type with accurate number. However, risk effect index on system, people and environment, fault detectability index and environment risk index cannot be presented and calculated by the accurate number, which need fuzzy set theory to capture the subjective judgments of experts and engineers to represent the risk state. Besides, the above factors can be also divided into benefit type and cost type according to their result effect. The operational risk assessment index information can be seen in Table 4.3.

Table 4.3 Operational risk assessment index information

No.	Assessment index	Data type	Index type	Calculation formula
C_1	MDBF	Numerical	Benefit	MDBF $= \sum L/N_f$
C_2	MTTR	Numerical	Cost	MTTR $= \sum_{i=1}^{n} t_i/n$
C_3	Risk effect on system, people and environment	Fuzzy	Cost	–
C_4	Maintenance cost	Interval	Cost	$R_c = \bar{C}/T$
C_5	Fault detectability	Fuzzy	Benefit	–
C_6	Correlation risk	Numerical	Cost	CR $= \sum_{j=1}^{5} w_j C_j$
C_7	Operation skill	Numerical	Benefit	$S_{rc} = \sum_{i=1}^{3} n_{1i}\varepsilon_i/N + \sum_{i=1}^{5} n_{2i}\lambda_i/N + \sum_{i=1}^{5} n_{3i}\eta_i/N$
C_8	Mental state	Numerical	Benefit	$Y_i = 1 + \sin(2\pi X/23)$
C_9	Environment risk	Fuzzy	Cost	–
C_{10}	TQI	Numerical	Cost	TQI $= \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)^2/n}$

4.4.1.2 Applications of TFNIFS in the Risk Assessment Index System

In the risk assessment index system of high speed train, risk effect index on system, people and environment, fault detectability index and environment risk index cannot be presented and calculated by the accurate number, which need fuzzy set theory to capture the subjective judgments of experts and engineers to represent the risk state. TFNIFS has been applied as an application of type-2 IFS in this paper to illustrate the value of the qualitative risk information for the as assessment. TFNIFS is proposed by Liu and Yuan [29] to cope with the fuzzy problems in the way that definite value of the membership and non-membership of IFS have been adjusted to the triangular fuzzy numbers, which can better express the indeterminacy of qualitative evaluation information.

Definition 1 An intuitionistic fuzzy set $\tilde{A}$ in $X = \{x\}$ can be defined as follows [22].

$$\tilde{A} = \left\{ < x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x) > | x \in X \right\} \tag{4.4.1}$$

Where the function membership and non-membership $\mu_{\tilde{A}} : X \to [0, 1]$, $\nu_{\tilde{A}} : X \to [0, 1]$ and $0 \leq \mu_{\tilde{A}}(\mathrm{x}) + \nu_{\tilde{A}}(x) \leq 1$.

Definition 2 The value of $\pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(\mathrm{x}) - \nu_{\tilde{A}}(x)$ can be called the degree of hesitancy of x to $\tilde{A}$.

In the description of intuitionistic fuzzy sets, $\tilde{A}$ can concentrate its two aspects of information from the degree of membership and non-membership of x to $\tilde{A}$, which makes it more flexible in describing the fuzzy problems. However, the membership degree and the non-membership degree are difficult to be described by numerical values because of the complexity and fuzziness of the qualitative risk information on high speed train. Therefore, TFNIFS integrated with triangular fuzzy number and IFS has been applied to overcome this problem in this paper.

Definition 3 A triangular fuzzy number $a = (a_l, a_m, a_r)$ in $X = \{x\}$ is a special fuzzy number, its membership function $\mu_a : X \to [0, 1]$ can be defined as follows [29].

$$\mu_a(x) = \begin{cases} (x - a_l)/(a_m - a_l) & a_l \leq x \leq a_m \\ (x - a_r)/(a_m - a_r) & a_m \leq x \leq a_r \\ 0 & \text{otherwise} \end{cases} \tag{4.4.2}$$

Where the membership function $\mu_a(x) \in [0, 1]$, $0 \leq a_l \leq a_m \leq a_r \leq 1$, a_m is the barycenter of fuzzy number a.

Definition 4 A triangular fuzzy number intuitionistic fuzzy set $\vec{A}$ in $X = \{x\}$ can be defined as follows [29].

Table 4.4 Linguistic variable of fault detectability C5

TFNIFS	Linguistic variable
((0.0,0.0,0.3),(0.6,0.8,1.0))	Very Difficult (*VD*)
((0.1,0.3,0.5),(0.4,0.6,0.8))	Difficult (*D*)
((0.3,0.5,0.7),(0.2,0.4,0.6))	Subcritical (*SU*)
((0.5,0.7,0.9),(0.0,0.2,0.4))	Easy (*E*)
((0.7,1.0,1.0),(0.0,0.0,0.2))	Very Easy (*VE*)

Table 4.5 Linguistic variable of risk effect on system, people and environment C3

TFNIFS	Linguistic variable
((0.0,0.0,0.3),(0.6,0.8,1.0))	Slight (*S*)
((0.1,0.3,0.5),(0.4,0.6,0.8))	Light (*L*)
((0.3,0.5,0.7),(0.2,0.4,0.6))	Subcritical (*SU*)
((0.5,0.7,0.9),(0.0,0.2,0.4))	Fatal (*F*)
((0.7,1.0,1.0),(0.0,0.0,0.2))	Disastrous (*D*)

Table 4.6 Linguistic variable of environmental risk C9

TFNIFS	Linguistic variable
((0.0,0.0,0.3),(0.6,0.8,1.0))	Very Safe (*VS*)
((0.1,0.3,0.5),(0.4,0.6,0.8))	Safe (*S*)
((0.3,0.5,0.7),(0.2,0.4,0.6))	Subcritical (*SU*)
((0.5,0.7,0.9),(0.0,0.2,0.4))	Fatal (*H*)
((0.7,1.0,1.0),(0.0,0.0,0.2))	Very Hazardous (*VH*)

$$\vec{A} = \left\{ \left(x, < \mu_{\vec{A}}(x), \nu_{\vec{A}}(x) > \right) | x \in X \right\} \tag{4.4.3}$$

Where $\mu_{\vec{A}}(x) = \left(\mu\vec{A}^{1}(x), \mu\vec{A}^{2}(x), \mu\vec{A}^{3}(x) \right)$ and $\nu_{\vec{A}}(x) = \left(\nu\vec{A}^{1}(x), \nu\vec{A}^{2}(x), \nu\vec{A}^{3}(x) \right)$ are triangular fuzzy numbers of $X = [0, 1]$, which express the membership and non-membership degree of x in X.

As for risk effect on system, people and environment, fault detectability and environment risk of high speed train, the index linguistic variable described as TFNIFS type can be seen in Tables 4.4, 4.5 and 4.6. Take the environmental risk as an example, Fig. 4.16 shows TFNIFS with "Very Safe (VS)" – ((0.0, 0.0, 0.3), (0.6, 0.8, 1.0)) in green, "Safe (S)" – ((0.1, 0.3, 0.5), (0.4, 0.6, 0.8)) in blue, "Subcritical (SU)" – ((0.3, 0.5, 0.7), (0.2, 0.4, 0.6)) in purple, "Fatal (H)" – ((0.5, 0.7, 0.9), (0.0, 0.2, 0.4)) in pink and "Very Hazardous (VH)" – ((0.7, 1.0, 1.0), (0.0, 0.0, 0.2)) in red, respectively. The solid lines describe the membership functions of TFNIFS, while the dotted lines describe non-membership functions of TFNIFS.

TFNIFS possesses many advantages as its membership and non-membership functions are themselves fuzzy, which can make it possible to minimize the defect and error of the subjective judgments of experts and engineers. Therefore, TFNIFS can be more appropriate to describe the risk state of high speed train.

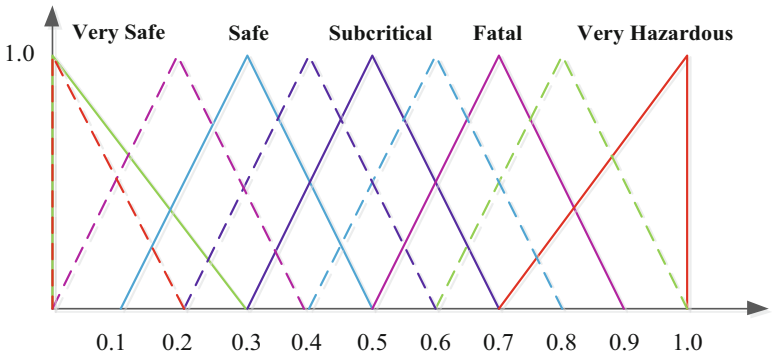


Fig. 4.16 Linguistic variable of environmental risk

4.4.1.3 Correlation Risk Index

High speed train is a complex electromechanical system consisting of more than 40,000 components, with various degradation and complex connection properties. Once a small failure of components occurs, it may lead to a loss of production and investments in relevant components or casualties and damage of the whole train system. There can be some defects in the internal risk factors analysis which only consider the components properties in an independent way and ignoring the relationship between the components. Thus, correlation risk index integrated with system structural risk and components failure, which is aimed at the interactions of components and the state importance in the train system, is proposed in this paper.

In recent years, complex network theory based on the system structure, which focuses on the structural properties such as node degree and clustering coefficient, has been utilized to the risk analysis of the relationships and interactions in the complex mechanical system [33–36] However, these researches are just conducted through a single-layer network, which cannot meet the need to represent the risk and complexity of high speed train with different mechanical, electrical and information interactions. Thus, multiplex network [37–40], which is coupled with various subnets of nodes and edges, is proposed to analyze the risk and complexity of high speed train.

Components in the high speed train system contact and interact with each other through their respective functions. However, a component may have various functions, which bring about different interactions. The loss or incomplete of a function may not affect the relationships between the relative components, as the different interactions between them. Therefore, the interactions of the components need to be divided according to the actual function of high speed train. The same function is placed on the same level, forming a mechanical, electrical and information interaction layer of the train system, and a multiplex network of high speed train is established, which can be seen in Fig. 4.17.

In the multiplex network of high speed train, components and their interactions can be described as nodes and edges respectively. The multiplex network of high

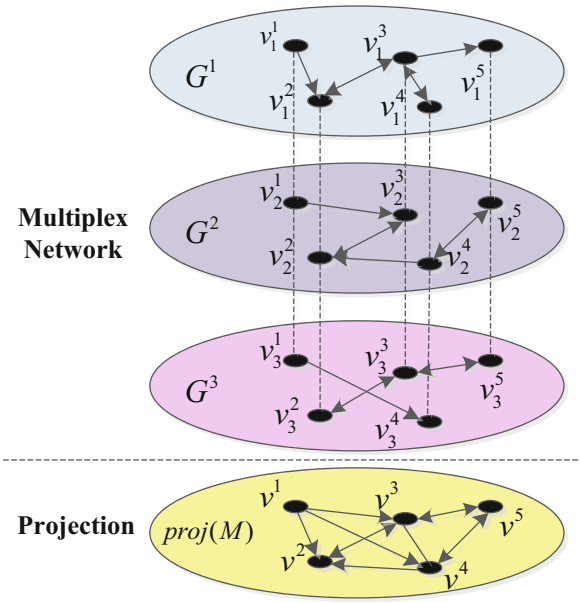


Fig. 4.17 High speed train multiplex network model

speed train can be constructed by the mechanical interaction layer G^1, electrical interaction layer G^2 and information interaction layer G^3 according to the actual function of high speed train, where $proj(M)$ is the projection of the multiplex network, which integrates all the interactions. In the mechanical interaction layer, components connect with each other by bolting and riveting, and the mechanical force can be transferred through the connections. Components in the electrical interaction layer are connected by cables and wires of different conductors, which can transfer the electric energy. Components in the information interaction layer are connected through the transmission medium, which can transfer the command and state information of each component. In these layers, the same component can integrate the different mechanical, electrical and information function through the layer connections, representing the relationship between the functions of the same component. The relationship between different components can be represented by a directed edge. One-way edge represents the one-way transmission, while two-way edge represents two-way transmission.

Multiplex network model, which can be applied to the analysis of correlation risk for high speed train, can be always combined with the structural properties such as node degree, clustering coefficient, closeness centrality, eigenvector centrality, etc. However, each component of high speed train system has its practical functional behaviors, which can affect the risk state of the whole train system in the form of failure, and present dynamic properties varying with time. Therefore, dynamic

Table 4.7 Dynamic properties with failure of high speed train multiplex network

Properties	Calculation formula
$K(i)$	$K(i) = \overline{\lambda_i} k_i = \sum_{j=1}^{m} \overline{\lambda_i} a_{ij}$
$C(i)$	$C(i) = 2\sum_{j=1}^{m} \overline{e_{ij}}/k_i(k_i - 1)$
$C_D(i)$	$C_D(i) = (m-1)/\sum_{j=1}^{m} \overline{d_{ij}}$
$C_E(i)$	$C_E(i) = \sum_{j=1}^{m} \overline{\lambda_i} a_{ij} x_j/\lambda$
$I(i)$	$I(i) = 1 - E_i/E_0$

properties with failure characteristic of the multiplex network based on the structural properties and the failure state of the train components are proposed to analyze the correlation risk in this paper, which can be seen in Table 4.7.

$K(i)$ is the failure degree of v_i, which is the result value based on the topological degree and node failure synthetically. $K(i)$ represents the affected average range of neighbor nodes as the node fails. $C(i)$ is the failure clustering coefficient of v_i, which measures the cluster degree after the node fails. $\bar{e}_{ij}$ is the number of the normal connections of v_i after the node fails. $\bar{d}_{ij}$ is the length of failure path for v_i. $C_D(i)$ is the failure closeness centrality of v_i, which describes the centrality of v_i as the node fails. $C_E(i)$ is the failure eigenvector centrality of v_i, which represents the importance of v_i with the feedback of neighbor nodes, λ is the eigenvalues of its neighbor matrix. $I(i)$ is the value at risk of failure of v_i, which can reflect the network efficiency after the node fails. E_0 and E_i are the network efficiency at the period of normal and failure operation respectively.

These five properties of the train multiplex network model combined with failure and its failure rate present the dynamic feature of high speed train system, which can be better to describe the risk and complexity of high speed train. Correlation risk index can be an amalgamation of these five properties based on the PCA approach [41]. Therefore, the weight of each property can be defined as

$$w_j = \sum_{j=1}^{n} a_j p_{ij} / \sum_{j=1}^{n} \sum_{i=1}^{m} a_j p_{ij} \tag{4.4.4}$$

Where λ_i is the eigenvalue of covariance matrix p with $n*m$ dimension, f_i is the principal component of the property C_j, and the correlation risk index of each component can be defined as

$$\mathrm{CR} = \sum_{j=1}^{5} w_j C_j \tag{4.4.5}$$

To summarize, correlation risk index of high speed train is mainly aimed at the interactions and relationships of components as well as the state importance in the train system, which can consummate the operational risk assessment.

4.4.2 Dynamic VIKOR Method for High Speed Train Operational Risk Assessment

The operational risk assessment of high speed train with qualitative and quantitative information brings about calculation uncertainty in the evaluation process, which can cause some impact on the results. Therefore, VIKOR approach [16], directed against the calculation uncertainty and concentrating the mixed data without information distortion or loss, has been applied in this paper.

VIKOR approach, developed for multi-criteria evaluation can be well applied to the problem of high speed train risk assessment, since it can reckon with the uncertainty particularly in situations where the experts and engineers are not able to express their preference in the risk assessment. VIKOR approach can determine the compromise ranking solution by a maximum group utility of the 'majority' and a minimum individual regret of the 'opponent', which can be the basis for negotiation, involving the preference of experts and engineers by indices weights [17]. The basic idea of VIKOR approach is the development from the Lp metric [17]:

$$L_{p,i} = \left\{ \sum_{j=1}^{n} \left[w_j \left(\left| r_j^* - r_{ij} \right| \right) / \left(\left| r_j^* - r_j^- \right| \right) \right]^p \right\}^{1/p} \tag{4.4.6}$$

Where w_j is criterion weight of the judgment, r_{ij} is the score of alternative A_j from the jth criterion. r_j^* and r_j^- is the ideal and negative ideal solutions of each alternative.

Nevertheless, the risk assessment of high speed train is carried out for three periods in this paper according to the second and forth class maintenance schedule of a specific high speed train in China. Relevant risk information and the expert cognition may experience constant and extreme change. Besides, the former stage would affect the latter stage, which would take on a dynamic feature about the process. Therefore, an extend VIKOR approach based on dynamic time is proposed for high speed train operational risk assessment in this paper.

As for the risk indices, the qualitative indices can be calculated according to the Table 4.3, while the quantitative indices should be illustrated as TFNIFS by the judgments from experts and engineers including maintenance personnel D_1, design manufacturer D_2 and high speed train driver D_3, under the three different annual check periods. $((a_{l1}, a_{m1}, a_{r1}), (b_{l1}, b_{m1}, b_{r1}))$ is a TFNIFS number $\vec{A}$, which is used as the value of expert evaluation. (a_{l1}, a_{m1}, a_{r1}) and (b_{l1}, b_{m1}, b_{r1}) is the membership and non-membership of $\vec{A}$, respectively, which represents the positive and negative degree of the risk index. $[aL\ 1, aU\ 1]$ can be always applied to express interval data set $\bar{\bar{A}}$, $aL\ 1$ and $aU\ 1$ is the upper and lower bound of $\bar{\bar{A}}$. A_i $(A = \{A_1, A_2, \ldots, A_n\})$ is the evaluation alternative (component), C_j $(C = \{C_1, C_2, \ldots, C_m\})$ is the evaluation criteria (risk index), D_s $(D = \{D_1, D_2, \ldots, D_d\})$ is the experts and $uk\ ij$ is the comprehensive evaluation value under the period K_k $(K = \{K_1, K_2, \ldots, K_v\})$, $Wk\ j$ is the weight of criteria C_j. The mixed fuzzy decision matrix U^k can be represented as follows.

$$U^k = \begin{array}{c} \\ A_1 \\ A_2 \\ \vdots \\ A_n \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_m \end{array} \\ \begin{bmatrix} u_{11}^k & u_{12}^k & \cdots & u_{1m}^k \\ u_{21}^k & u_{22}^k & \cdots & u_{2m}^k \\ \vdots & \vdots & \vdots & \vdots \\ u_{n1}^k & u_{n2}^k & \cdots & u_{nm}^k \end{bmatrix} \end{array} \tag{4.4.7}$$

$$W_j^k = \left[w_1^k, w_2^k, \cdots, w_m^k\right]^T \tag{4.4.8}$$

Usually, the mixed fuzzy decision matrix needs to be standardized by

$$r_{ij}^k = u_{ij}^k / \max u_{ij}^k \text{ or } r_{ij}^k = \left[u_{ij}^{kL} / \max u_{ij}^{kU}, u_{ij}^{kU} / \max u_{ij}^{kU}\right] \tag{4.4.9}$$

$$r_{ij}^k = \min u_{ij}^k / u_{ij}^k \text{ or } r_{ij}^k = \left[\min u_{ij}^{kL} / u_{ij}^{kU}, \min u_{ij}^{kU} / u_{ij}^{kU}\right] \tag{4.4.10}$$

Equations (4.4.9) and (4.4.10) are applied to the criteria of cost and benefit type respectively. Obviously, the *rk ij* after standardization can be between 0 and 1. It is easy to know that the ideal and negative ideal solution of the mixed fuzzy decision matrix can be expressed as.

$$r_j^{k*} = \max_i r_{ij}^k, \quad r_j^{k-} = \min_i r_{ij}^k \text{ for benefit type } C_j \tag{4.4.11}$$

$$r_j^{k*} = \min_i r_{ij}^k, \quad r_j^{k-} = \max_i r_{ij}^k \text{ for cost type } C_j \tag{4.4.12}$$

Later the weight of expert can be determined by Eqs. (4.4.11 and 4.4.12) based on expert credibility [22], and the weight of evaluation criteria can be determined by Eqs. (4.4.5, 4.4.6, 4.4.7, and 4.4.8) based on entropy weight [20].

$$B_s^k(\pi) = -\left(\left(\sum_{i=1}^n \sum_{j=1}^m \pi_{sij}^k\right) \ln\left(\sum_{i=1}^n \sum_{j=1}^m \pi_{sij}^k\right)\right)^{-1} \tag{4.4.13}$$

$$\lambda_s^k = \frac{B_s^k(\pi)}{\sum_{s=1}^d B_s^k(\pi)} \tag{4.4.14}$$

$\tilde{r}_j^k$ is the aggregation operator of criteria C_j, Eqs. (4.4.3, 4.4.4, and 4.4.5) are applied to the value of numeric, interval and TFNIFS type respectively.

$$\tilde{r}_j^k = \left(r_{1j}^k + r_{2j}^k + \cdots + r_{nj}^k\right)/n \tag{4.4.15}$$

$$\tilde{r}_j^k = \left[\left(\sum_{i=1}^n r_{ij}^{kL}\right)/n, \left(\sum_{i=1}^n r_{ij}^{Uk}\right)/n\right] \tag{4.4.16}$$

$$\tilde{r}_j^k = \left(\left(\prod_{i=1}^{n}\left(a_{lij}^k\right)^{1/n}, \prod_{i=1}^{n}\left(a_{mij}^k\right)^{1/n}, \prod_{i=1}^{n}\left(a_{rij}^k\right)^{1/n}\right), \left(\left(1-\prod_{i=1}^{n}\left(1-b_{lij}^k\right)^{1/n}\right), \left(1-\prod_{i=1}^{n}\left(1-b_{mij}^k\right)^{1/n}\right), \left(1-\prod_{i=1}^{n}\left(1-b_{rij}^k\right)^{1/n}\right)\right)\right) \tag{4.4.17}$$

The entropy of criteria C_j is $ek\ j$, which can be calculated as

$$e_j^k = -\left[\left(\sum\nolimits_{i=1}^{n} d\left(r_{ij}^k, \tilde{r}_j^k\right) / \sum_{i=1}^{n} d\left(r_{ij}^k, \tilde{r}_j^k\right)\right) \ln\left(d\left(r_{ij}^k, \tilde{r}_j^k\right) / \sum_{i=1}^{n} d\left(r_{ij}^k, \tilde{r}_j^k\right)\right)\right] / \ln(n) \tag{4.4.18}$$

Therefore, the weight of evaluation criteria can be calculated as

$$\omega_j^k = \left(1-e_j^k\right) / \sum\nolimits_{j=1}^{m}\left(1-e_j^k\right) \tag{4.4.19}$$

The weight of each period k can be calculated under entropy weight approach in accordance with Eqs. (4.4.17, 4.4.18, and 4.4.19).

$$r_i^k = \sum\nolimits_{j=1}^{m}\left(r_{ij}^k/m\right), \quad \bar{r}^k = E\left(r_i^k\right) \tag{4.4.20}$$

$$e^k = -\sum\nolimits_{i=1}^{n}\left[d\left(\tilde{r}_j^k, \bar{r}^k\right) / \sum_{i=1}^{n} d\left(\tilde{r}_j^k, \bar{r}^k\right)\right]\left[\ln\left(d\left(\tilde{r}_j^k, \bar{r}^k\right) / \sum_{i=1}^{n} d\left(\tilde{r}_j^k, \bar{r}^k\right)\right)\right] / \mathrm{In}(n) \tag{4.4.21}$$

$$\eta^k = \left(1-e^k\right) / \sum\nolimits_{K=1}^{K}\left(1-e^k\right) \tag{4.4.22}$$

The maximum group utility, the minimum of individual regret and the comprehensive risk value can be calculated in accordance with Eqs. (4.4.20, 4.4.21, and 4.4.22).

$$S_i = \sum\nolimits_{k=1}^{K}\sum\nolimits_{j=1}^{m}\eta^k\omega_j\left(d\left(r_j^{k*}, r_{ij}^k\right) / d\left(r_j^{k*}, r_j^{k-}\right)\right) \tag{4.4.23}$$

$$R_i = \max_j\left[\omega_j\eta^k\left(d\left(r_j^{k*}, r_{ij}^k\right) / d\left(r_j^{k*}, r_j^{k-}\right)\right)\right] \tag{4.4.24}$$

$$Q_i = v(S_i - S^*)/(S^- - S^*) + (1-v)(R_i - R^*)/(R^- - R^*) \tag{4.4.25}$$

The proposed dynamic VIKOR approach has the ability to reckon with the various risk state and complexity of high speed train operational risk assessment under the different periods with dynamic features. The procedure of dynamic VIKOR approach can be explained as follows, which can be seen in Fig. 4.18.

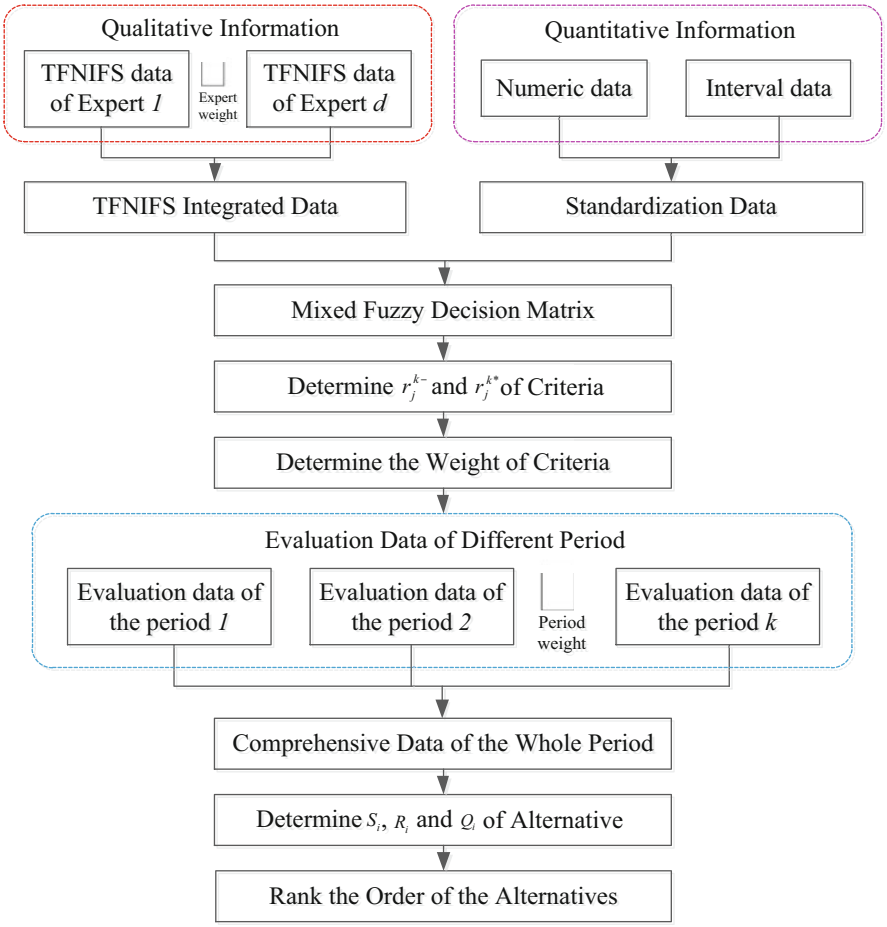


Fig. 4.18 Procedure of dynamic VIKOR approach for high speed train operational risk assessment

Step 1. Compute each risk evaluation index based on the assessment index system under the period *k*.
Step 2. Determine the weight of expert for the qualitative index under the period *k*.
Step 3. Standardize the mixed fuzzy decision matrix under the period *k*.
Step 4. Determine *rk* j*and *rk- j*of the mixed fuzzy decision matrix under the period *k*.
Step 5. Determine the weight of evaluation criteria for the mixed fuzzy decision matrix under the period *k*.
Step 6. Determine the weight of each period *k* for the comprehensive assessment.
Step 7. Determine the maximum group utility, the minimum of individual regret and the comprehensive risk value and determine the ranking order of the alternatives based on the comprehensive risk value.

4.4.3 Case Study

In this paper, the operational risk assessment based on TFNIFS and dynamic VIKOR approach is validated by taking a specific high speed train bogie system as an example. The components of bogie system have been applied to establishing the multiplex network model of high speed train bogie system, which can be shown in Fig. 4.19. Nodes of each sub network can be shown in Table 4.8.

Based on the multiplex network model of high speed train bogie system, the correlation risk index can be calculated with the fusions of failure degree, failure

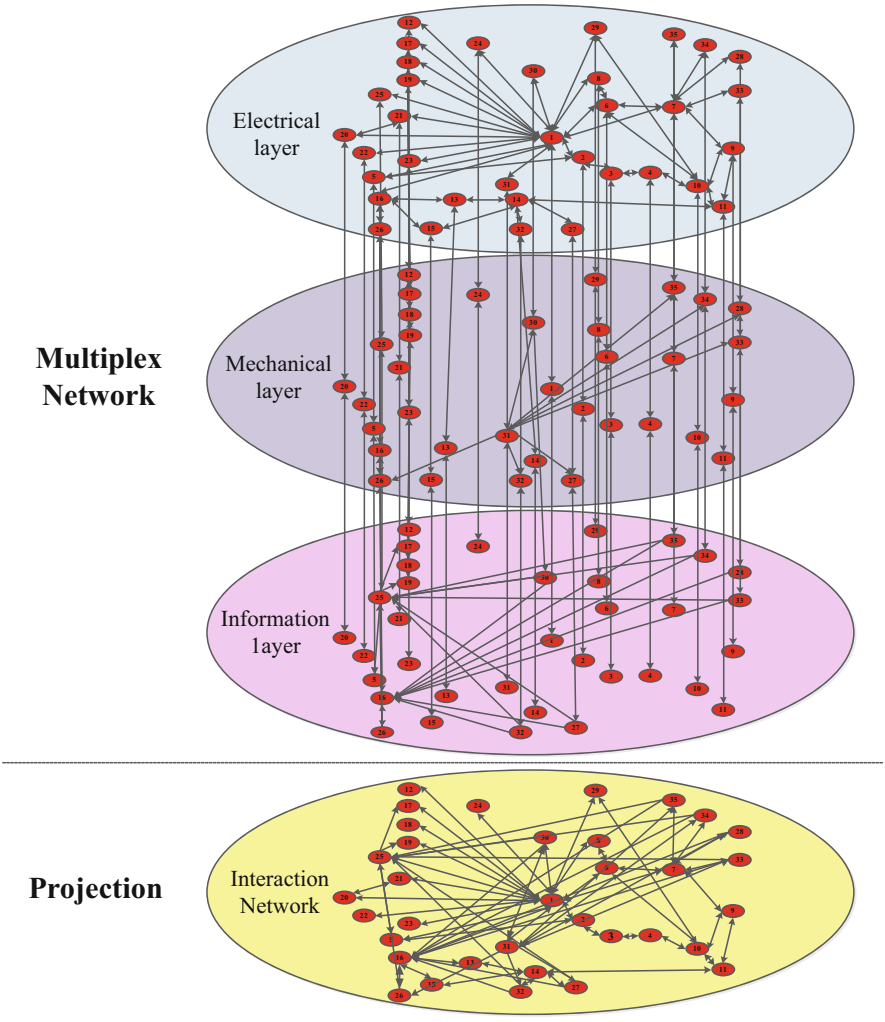


Fig. 4.19 Multiplex network model of high speed train bogie system

Table 4.8 Components in a Specific High Speed Train Bogie System

No.	Component	No.	Component	No.	Component
1	Frame assembly	13	Coupling	25	Main duct and solenoid valve
2	Brake clamp	14	Gearbox assembly	26	Velocity sensor1
3	Brake lining	15	Ground device	27	Velocity sensor2
4	Wheel-mounted disc brake	16	Traction motor	28	Velocity sensor3 LKJ2000
5	Pressure cylinder	17	Height adjustment device	29	Surface cleaning device
6	Spring assembly	18	Coil resistance shock absorber	30	Acceleration sensor
7	Axle box body	19	Air spring	31	Junction box
8	Primary vertical shock absorbers	20	Center for traction pin	32	Gear box bearing temperature sensor
9	Bearing	21	Traction rod	33	Journal temperature sensor
10	Wheel	22	Transverse shock absorber	34	Velocity sensor4 AG37
11	Axle	23	Lateral stop	35	Velocity sensor5 AG43
12	Vertical shock absorber – 2	24	Anti-roll bar		

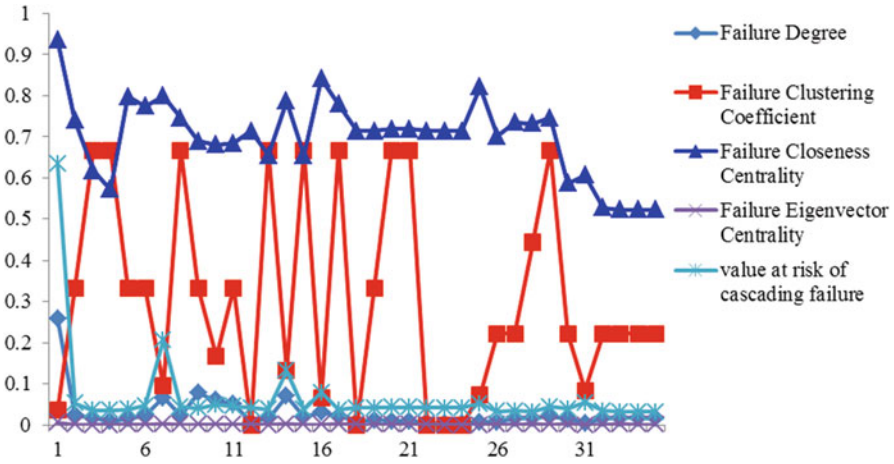


Fig. 4.20 Properties of high speed train bogie system under Period 1

clustering coefficient, failure closeness centrality, failure eigenvector centrality and the value at risk of failure of each component based on the PCA approach according to the Eqs. (4.4.9 and 4.4.10), Fig. 4.20 gives an example of the above properties of the multiplex network model under the period 1. Correlation risk index of these three constant periods can be comprehensively integrated and illustrated in the Fig. 4.21.

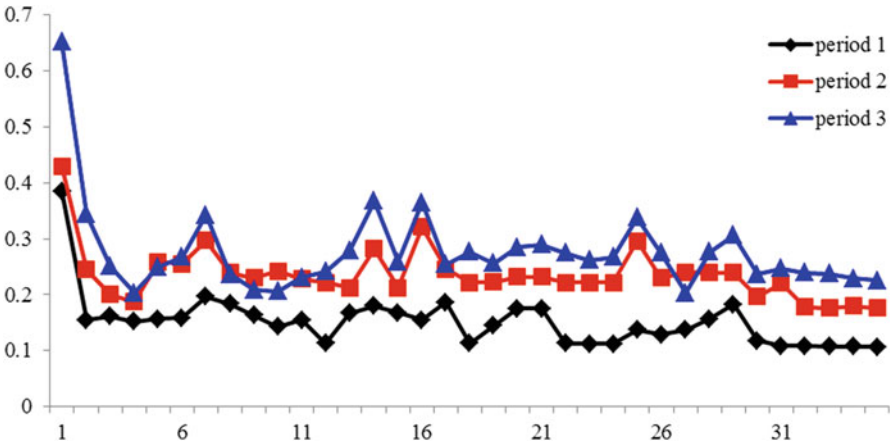


Fig. 4.21 Correlation risk index of three Periods

Table 4.9 Risk effect value on system, people and environment for first 10 components from D1

No.	Period 1	Period 2	Period 3
1	((0.6,0.8,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))
2	((0.5,0.7,0.9),(0.0,0.0,0.2))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.7,0.9,1.0),(0.0,0.0,0.2))
3	((0.6,0.8,1.0),(0.0,0.2,0.4))	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.2,0.4,0.6),(0.2,0.4,0.6))
4	((0.1,0.3,0.5),(0.1,0.3,0.5))	((0.0,0.0,0.2),(0.4,0.6,0.8))	((0.2,0.4,0.6),(0.2,0.4,0.6))
5	((0.2,0.4,0.6),(0.0,0.0,0.2))	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.2,0.4,0.6),(0.3,0.5,0.7))
6	((0.3,0.5,0.7),(0.0,0.2,0.4))	((0.3,0.5,0.7),(0.1,0.3,0.5))	((0.3,0.5,0.7),(0.1,0.3,0.5))
7	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.6,0.8,1.0),(0.0,0.0,0.2))
8	((0.1,0.3,0.5),(0.1,0.3,0.5))	((0.2,0.4,0.6),(0.2,0.4,0.6))	((0.0,0.2,0.4),(0.2,0.4,0.6))
9	((0.2,0.4,0.6),(0.1,0.3,0.5))	((0.2,0.4,0.6),(0.4,0.6,0.8))	((0.4,0.6,0.8),(0.0,0.2,0.4))
10	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.6,0.8,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))

As it is shown in the Fig. 4.21, the correlation risk index of each component is in the interval of 0.1 and 0.7, where the risk degree of the dynamic failure properties takes on ascend trend with the time and the maximum correlation risk index comes to the frame assembly.

Three experts and engineers including maintenance personnel D_1, design manufacturer D_2 and high speed train driver D_3 have been invited to assess the risk effect on system, people and environment, fault detectability and environment risk index, which are expressed as TFNIFS, under the three different periods. Tables 4.8, 4.9 and 4.10 give a view of the assessment index value of the risk effect on system, people and environment from maintenance personnel D_1, design manufacturer D_2 and high speed train driver D_3 which are just listed the number of first 10 in the alternative of components.

For instance, ((0.6, 0.8, 1.0), (0.0, 0.0, 0.2)) is the value of the risk effect on system, people and environment from the maintenance personnel D1, which expressed as TFNIFS. (0.6, 0.8, 1.0) is a triangle fuzzy number, which is defined

Table 4.10 Risk effect value on system, people and environment for first 10 components from D2

No.	Period 1	Period 2	Period 3
1	((0.8,1.0,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))
2	((0.7,0.9,1.0),(0.0,0.0,0.2))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.8,1.0,1.0),(0.0,0.0,0.2))
3	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.3,0.5,0.7),(0.0,0.2,0.4))	((0.4,0.6,0.8),(0.0,0.2,0.4))
4	((0.3,0.5,0.7),(0.1,0.3,0.5))	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.4,0.6,0.8),(0.0,0.2,0.4))
5	((0.2,0.4,0.6),(0.0,0.0,0.2))	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.2,0.4,0.6),(0.3,0.5,0.7))
6	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.3,0.5,0.7),(0.1,0.3,0.5))	((0.3,0.5,0.7),(0.1,0.3,0.5))
7	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.4,0.6,0.8),(0.0,0.2,0.4))
8	((0.2,0.4,0.6),(0.1,0.3,0.5))	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.0,0.2,0.4),(0.2,0.4,0.6))
9	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.2,0.4,0.6),(0.4,0.6,0.8))	((0.4,0.6,0.8),(0.0,0.2,0.4))
10	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.5,0.7,0.9),(0.0,0.2,0.4))

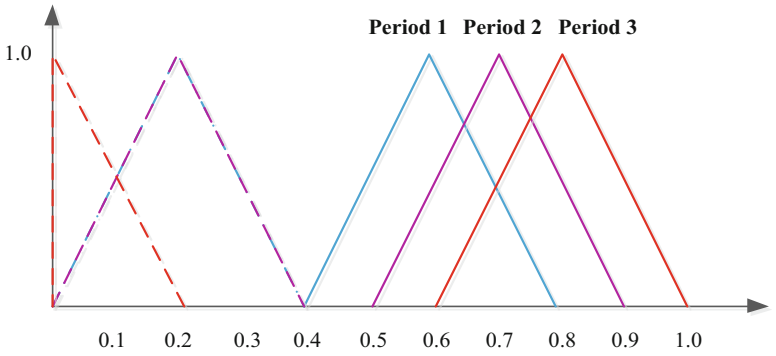


Fig. 4.22 Trend of risk effect index on system, people and environment over time for axle box body

as the membership of TFNIFS and indicates the positive degree for the risk effect index on system, people and environment. (0.0, 0.0, 0.2) is the other triangle fuzzy number defined as the non-membership of TFNIFS, which indicates the negative degree for the risk effect index on system, people and environment. It can be easy to see that the positive degree is much higher than the negative degree, which expresses the expert idea that the risk effect on system, people and environment is more likely to be higher and affects bogie system.

Figure 4.22 gives an example of the trend of the risk effect index value on system, people and environment over time for the axle box body (component 7). ((0.4, 0.6, 0.8), (0.0, 0.2, 0.4)) in blue, ((0.5, 0.7, 0.9), (0.0, 0.2, 0.4)) in pink and ((0.6, 0.8, 1.0), (0.0, 0.0, 0.2)) represents the risk effect index value on system, people and environment under the Period 1, Period 2 and Period 3, respectively. It can be easy to see that the trend is getting higher, which expresses the expert idea that the risk effect on system, people and environment is getting higher over time.

Later, the qualitative assessment value of the index, which consists of the risk effect on system, people and environment, fault detectability and environment risk, can be compromised into the comprehensive assessment value of each index

Table 4.11 Risk effect value on system, people and environment for first 10 components from D3

No.	Period 1	Period 2	Period 3
1	((0.6,0.8,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))
2	((0.5,0.7,0.9),(0.0,0.0,0.2))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.7,0.9,1.0),(0.0,0.0,0.2))
3	((0.6,0.8,1.0),(0.0,0.2,0.4))	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.2,0.4,0.6),(0.2,0.4,0.6))
4	((0.1,0.3,0.5),(0.1,0.3,0.5))	((0.0,0.0,0.2),(0.4,0.6,0.8))	((0.2,0.4,0.6),(0.2,0.4,0.6))
5	((0.2,0.4,0.6),(0.0,0.0,0.2))	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.2,0.4,0.6),(0.3,0.5,0.7))
6	((0.3,0.5,0.7),(0.0,0.2,0.4))	((0.3,0.5,0.7),(0.1,0.3,0.5))	((0.3,0.5,0.7),(0.1,0.3,0.5))
7	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.6,0.8,1.0),(0.0,0.0,0.2))
8	((0.1,0.3,0.5),(0.1,0.3,0.5))	((0.2,0.4,0.6),(0.2,0.4,0.6))	((0.0,0.2,0.4),(0.2,0.4,0.6))
9	((0.2,0.4,0.6),(0.1,0.3,0.5))	((0.2,0.4,0.6),(0.4,0.6,0.8))	((0.4,0.6,0.8),(0.0,0.2,0.4))
10	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.6,0.8,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))

according to the Eqs. (4.4.11 and 4.4.12) based on expert credibility [8]. Table 4.11 gives a view of the comprehensive value of the index for the risk effect on system, people and environment based on these 3 experts and engineers. The weight of D1, D2 and D3 under the different 3 periods can be determined as follows.

- $\lambda 1\ 1 = 0.43$, $\lambda 1\ 2 = 0.36$, $\lambda 1\ 3 = 0.21$;
- $\lambda 2\ 1 = 0.33$, $\lambda 2\ 2 = 0.35$, $\lambda 2\ 3 = 0.32$;
- $\lambda 3\ 1 = 0.34$, $\lambda 3\ 2 = 0.32$, $\lambda 3\ 3 = 0.34$.

The weight of experts and engineers in one period is disciplines distinct from the other, which can account for the reasons that their jobs and working environment differ a lot. Therefore, the cognition and preference can make a large change to the risk matters with bogie system. In addition, the information obtained from the former period can have a great influence on the assessment of the latter period. The information quantity and quality obtained from different period can also differ a lot with each other. As a result, the weight of each expert in the different period can have a great difference.

With the advance of the proposed approach, the mixed fuzzy decision matrix under different period can be set up based on the risk factors with MDBF, MTTR, risk effect on system, people and environment, maintenance cost, fault detectability, and correlation risk, operation skills, mental state, environment risk and TQI index. Table 4.12 illustrates a mixed fuzzy decision matrix under the period 2, which is just listed the number of first 10 in the alternative of components.

In the mixed fuzzy decision matrix under the period 2, the external elements for the operation skills, mental state and TQI index are calculated as the overall indicators, which can affect the whole high speed train. Therefore, the index value for these three risk factors can be illustrated as a same numerical value. Apart from these, maintenance cost is illustrated as interval numbers, which can represent the expense as a range of numbers with the definitive extreme points of the intervals (Table 4.13).

After the mixed fuzzy decision matrix is standardized according to the Eqs. (4.4.13 and 4.4.14), rk^* $j$and r_j^{k-} of the mixed standardized fuzzy decision

Table 4.12 Comprehensive risk effect value on system, people and environment for First 10 components

No.	Period 1	Period 2	Period 3
1	((0.6,0.8,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))
2	((0.6,0.8,1.0),(0.0,0.0,0.2))	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.8,1.0,1.0),(0.0,0.0,0.2))
3	((0.5,0.7,0.9),(0.0,0.2,0.4))	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.3,0.5,0.7),(0.1,0.3,0.5))
4	((0.2,0.4,0.6),(0.1,0.3,0.5))	((0.1,0.3,0.5),(0.4,0.6,0.8))	((0.3,0.6,0.7),(0.1,0.3,0.5))
5	((0.2,0.4,0.6),(0.0,0.0,0.2))	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.2,0.4,0.6),(0.3,0.5,0.7))
6	((0.2,0.4,0.6),(0.0,0.4,0.4))	((0.3,0.5,0.7),(0.1,0.3,0.5))	((0.3,0.5,0.7),(0.1,0.3,0.5))
7	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.5,0.7,0.9),(0.1,0.3,0.5))	((0.5,0.7,0.9),(0.0,0.0,0.2))
8	((0.1,0.3,0.5),(0.1,0.3,0.5))	((0.1,0.3,0.5),(0.3,0.5,0.7))	((0.0,0.2,0.4),(0.2,0.4,0.6))
9	((0.1,0.3,0.5),(0.2,0.4,0.6))	((0.2,0.4,0.6),(0.4,0.6,0.8))	((0.4,0.6,0.8),(0.0,0.2,0.4))
10	((0.4,0.6,0.8),(0.0,0.2,0.4))	((0.5,0.7,0.9),(0.0,0.0,0.2))	((0.8,1.0,1.0),(0.0,0.0,0.2))

matrix under the different period can be determined to provide the basic preparation for the dynamic VIKOR approach. Later, the weight of each index can be calculated and the result can be shown as follows.

- w1 = [0.03 0.05 0.05 0.02 0.04 0.03 0.25 0.25 0.03 0.25];
- w2 = [0.03 0.02 0.04 0.03 0.02 0.03 0.27 0.27 0.02 0.27];
- w3 = [0.03 0.04 0.04 0.03 0.02 0.03 0.26 0.26 0.03 0.26].

Then, the weight of each period for the comprehensive assessment can be determined according to the Eqs. (4.4.18, 4.4.19, and 4.4.20), which can be obtained as

- $\eta^1 = 0.24$, $\eta^2 = 0.37$, $\eta^3 = 0.39$.

In the end, the maximum group utility S_i, the minimum of individual regret R_i and the comprehensive risk value Q_i can be calculated according to the Eqs. (4.4.18, 4.4.19, and 4.4.20). Figure 4.23 illustrates the results of Q_i, from which the ranking order of the alternatives can be obtained. Furthermore, Fig. 4.23 also compares the results between static [32] and dynamic assessment. From the results comparison between static and dynamic assessment of high speed train bogie system in the Fig. 4.23 frame assembly is still the highest risk component. The risk degree of brake clamp, wheels and axle box body exceed over the state of static assessment, which need more attention to take care of.

In addition, these two different results make it clear that the relevant risk information and the expert cognition can experience constant and extreme change over time. The operational risk assessment of high speed train is carried out for three periods in this paper according to the second and forth class maintenance schedule of a specific high speed train in China. It can thank to the assessment of the former period in the dynamic process, which makes it more reliable to assess the risk state of high speed train. What is more, relevant information can be more systematic and effective, for the failure accumulates as the time goes by. Therefore, operational risk

Table 4.13 Mixed fuzzy decision matrix of Period 2

No.	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
1	11.9	30	((0.8,1.0,1.0), (0.0,0.0,0.2))	[4.0,5.0]	((0.4,0.6,0.8), (0.0,0.2,0.4))	0.6	8	6.5	((0.5,0.7,0.9), (0.0,0.2,0.4))	7.3
2	5.2	20	((0.5,0.7,0.9), (0.0,0.2,0.4))	[2.0,2.5]	((0.4,0.6,0.8), (0.0,0.2,0.4))	0.3	8	6.5	((0.4,0.6,0.8), (0.1,0.3,0.5))	7.3
3	28.4	25	((0.4,0.6,0.8), (0.0,0.2,0.4))	[2.0.2.5]	((0.4,0.6,0.8), (0.0,0.2,0.4))	0.3	8	6.5	((0.4,0.6,0.8), (0.1,0.3,0.5))	7.3
4	3.8	20	((0.1,0.3,0.5), (0.4,0.6,0.8))	[2.5,3.5]	((0.4,0.6,0.8), (0.1,0.3,0.5))	0.2	8	6.5	((0.3,0.5,0.7), (0.2,0.4,0.6))	7.3
5	10.6	15	((0.1,0.3,0.5), (0.3,0.5,0.7))	[2.0,2.5]	((0.3,0.5,0.7), (0.1,0.3,0.5))	0.3	8	6.5	((0.4,0.6,0.8), (0.2,0.4,0.6))	7.3
6	16.3	20	((0.3,0.5,0.7), (0.1,0.3,0.5))	[2.5,3.0]	((0.2,0.4,0.6), (0.3,0.5,0.7))	0.3	8	6.5	((0.4,0.6,0.8), (0.1,0.3,0.5))	7.3
7	6.1	20	((0.5,0.7,0.9), (0.1,0.3,0.5))	[2.5,3.0]	((0.5,0.7,0.9), (0.0,0.2,0.4))	0.3	8	6.5	((0.3,0.5,0.7), (0.1,0.3,0.5))	7.3
8	46.7	20	((0.1,0.3,0.5), (0.3,0.5,0.7))	[2.5,3.0]	((0.4,0.6,0.8), (0.1,0.3,0.5))	0.2	8	6.5	((0.6,0.8,1.0), (0.0,0.2,0.4))	7.3
9	17.9	10	((0.2,0.4,0.6), (0.4,0.6,0.8))	[2.0,2.5]	((0.6,0.8,1.0), (0.0,0.2,0.4))	0.3	8	6.5	((0.3,0.5,0.8), (0.1,0.3,0.5))	7.3
10	11.9	25	((0.5,0.7,0.9), (0.0,0.0,0.2))	[2.5,3.5]	((0.5,0.7,0.9), (0.0,0.2,0.4))	0.3	8	6.5	((0.4,0.6,0.8), (0.0,0.2,0.4))	7.3

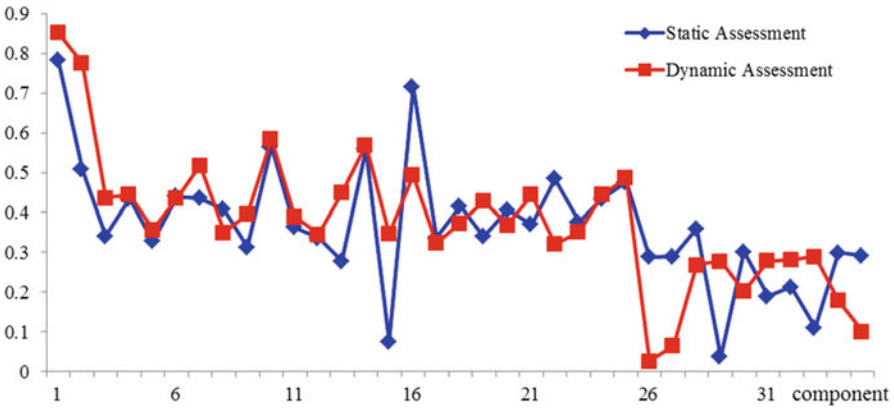


Fig. 4.23 Results comparison between static and dynamic assessment

assessment based on TFNIFS and dynamic VIKOR approach can provide technique support for high speed train safety operation.

This chapter works on the study of the operational risk assessment of high speed train based on TFNIFS and dynamic VIKOR approach under constant periods, a comprehensive operational high speed train risk assessment index system is established associated with the risk factors of staff, environment, infrastructure and the train itself. As for the calculation uncertainty, expert preference and information indeterminacy in the assessment, VIKOR approach can be well applied to cope with this problem. Due to the qualitative risk information which cannot be expressed distinctly by quantitative data, TFNIFS as an application of type-2 IFS can reckon with this problem as its membership and non-membership functions are themselves fuzzy, which can make it possible to minimize the defect and error of the subjective judgments of experts and engineers. Since the risk assessment of high speed train is carried out for three periods in this paper according to the second and forth class maintenance schedule of a specific high speed train in China, and the comprehensive risk assessment is integrated synthetically based on the result of three periods. Relevant risk information and the expert cognition may experience constant and extreme change. In addition, the former stage would affect the latter stage, which would take on a dynamic feature about the process. Therefore, an extend VIKOR approach based on the different time periods is proposed with TFNIFS for high speed train operational risk assessment in this paper. Finally, a specific example of high speed train bogie system is implemented to validate the proposed approach. The result is also compared with the static operational risk assessment result, which the authors once proposed. Between the comparisons, it can be found out that dynamic VIKOR approach and TFNIFS can be better utilized in the operational risk assessment for high speed train.

References

1. S.H. Baek, S.S. Cho, W.S. Joo, Fatigue life prediction based on the rainflow cycle counting method for the end beam of a freight car bogie. Int. J. Automot. Technol. **9**(1), 95–101 (2008)
2. Y. Lu et al., Reliability and parametric sensitivity analysis of railway vehicle bogie frame based on monte-carlo numerical simulation (2010)
3. S.G. Zhang, Study on testing and establishment method for the load spectrum of bogie frame for high-speed trains. Sci. China **51**(12), 2142–2151 (2008)
4. X. Wang, X. Li, F. Li, Analysis on oscillation in electro-hydraulic regulating system of steam turbine and fault diagnosis based on PSOBP. Expert Syst. Appl. **37**(5), 3887–3892 (2010)
5. B. Yazici, S. Yolacan, A comparison of various tests of normality. J. Stat. Comput. Simul. **77** (2), 175–183 (2007)
6. J.F. Lawless, *Statistical Models and Methods for Lifetime Data* (Wiley-Interscience, New York, 2003), pp. 264–265
7. A. Barabadi, J. Barabady, T. Markeset, Maintainability analysis considering time-dependent and time-independent covariates. Reliab. Eng. Syst. Saf. **96**(1), 210–217 (2011)
8. M. Guo et al., The impact of personality on driving safety among Chinese high-speed railway drivers. Accid. Anal. Prev. **92**, 9–14 (2016)
9. G. Gou et al., Effect of humidity on porosity, microstructure, and fatigue strength of A7N01S-T5 aluminum alloy welded joints in high-speed trains. Mater. Des. **85**, 309–317 (2015)
10. C.F. Hung, W.L. Hsu, Influence of long-wavelength track irregularities on the motion of a high-speed train. Veh. Syst. Dyn. **12**, 1–18 (2017)
11. E. Jafarian, M.A. Rezvani, Application of fuzzy fault tree analysis for evaluation of railway safety risks: An evaluation of root causes for passenger train derailment. Proc. Inst. Mech. Eng. F J. Rail Rapid Transp. **226**(1), 14–25 (2012)
12. G. Bearfield, W. Marsh, Generalising event trees using bayesian networks with a case study of train derailment. Lect. Notes Comput. Sci **3688**, 52–66 (2005)
13. P. Antonio, R. Fabrizio, A. Raffaele, Bayesian analysis and prediction of failures in underground trains. Qual. Reliab. Eng. Int. **19**(4), 327–336 (2010)
14. W.J. Zhang, N. Lan, Research on the reliability growth management techniques of high-speed train for whole life cycle (2013)
15. Y. Wang et al., Research on design evaluation of high-speed train auxiliary power supply system based on the AHP, in *Transportation Electrification Asia-Pacific* (2014)
16. S. Opricovic, G.H. Tzeng, Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. Eur. J. Oper. Res. **156**(2), 445–455 (2004)
17. S. Opricovic, G.H. Tzeng, Extended VIKOR method in comparison with outranking methods. Eur. J. Oper. Res. **178**(2), 514–529 (2007)
18. A. Jahan, K.L. Edwards, VIKOR method for material selection problems with interval numbers and target-based criteria. Mater. Des. **47**(47), 759–765 (2013)
19. K. Devi, Extension of VIKOR method in intuitionistic fuzzy environment for robot selection. Expert Syst. Appl. **38**(11), 14163–14168 (2011)
20. O. Mohsen, N. Fereshteh, An extended VIKOR method based on entropy measure for the failure modes risk assessment – A case study of the geothermal power plant (GPP). Saf. Sci. **92**, 160–172 (2017)
21. L.A. Zadeh, Fuzzy sets, information and control. Inf. Control. **8**(3), 338–353 (1965)
22. K.T. Atanassov, Remarks on the intuitionistic fuzzy sets. Fuzzy Sets Syst. **33**(1), 37–45 (1989)
23. F. Shen et al., An extended intuitionistic fuzzy TOPSIS method based on a new distance measure with an application to credit risk evaluation. Inf. Sci. (2017)
24. L.E. Wang, H.C. Liu, M.Y. Quan, *Evaluating the risk of failure modes with a hybrid MCDM model under interval-valued intuitionistic fuzzy environments* (Pergamon Press, 2016), pp. 175–185
25. S.P. Wan, F. Wang, J.Y. Dong, *A novel risk attitudinal ranking method for intuitionistic fuzzy values and application to MADM* (Elsevier Science Publishers B. V., 2016), pp. 98–112

26. C. Yue, A geometric approach for ranking interval-valued intuitionistic fuzzy numbers with an application to group decision-making. Comput. Ind. Eng. **102**, 233–245 (2016)
27. G. Kumar, R.K. Bajaj, N. Gandotra, Algorithm for shortest path problem in a network with interval-valued intuitionistic trapezoidal fuzzy number. Procedia Comput. Sci. **70**, 123–129 (2015)
28. S. Guo, W. Yin, Multiple attribute decision making method based on 2-type intuitionistic fuzzy information. Fuzzy Syst. Math. **27**(3), 128–133 (2013)
29. F. Liu, X.H. Yuan, Fuzzy number intuitionistic fuzzy set. Fuzzy Syst. Math. **21**(1), 88–91 (2007)
30. X.U. Danqing, X. Chen, D.O. Mathematics, multi-attribute decision-making method based on hesitant intuitionistic fuzzy linguistic set. J, in *Huaibei Normal Univ*, (2016)
31. B. Zhou, M. Xie, W.U. Keming, Analysis and prediction on the current situation of the repair class and repair system of electric multiple units(EMU). Electric Drive for Locomotives (2017)
32. Y. Fu, et al., Operation safety assessment of high-speed train with fuzzy group decision making method and empirical research, in *International Conference on Cloud Computing and Internet of Things* (2017)
33. Y. Yang, Y. Liu, M. Zhou, F. Li, C. Sun, Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing Subway. Saf. Sci. **79**, 149–162 (2015)
34. A.M. Sarhan, Reliability estimations of components from masked system life data. Reliab. Eng. Syst. Saf. **74**(1), 107–113 (2001)
35. M. Macchi, M. Garetti, D. Centrone, L. Fumagalli, G.P. Pavirani, Maintenance management of railway infrastructures based on reliability analysis. Reliab. Eng. Syst. Saf. **104**, 71–83 (2012)
36. J. Lin, J. Pulido, M. Asplund, Reliability analysis for preventive maintenance based on classical and Bayesian semi-parametric degradation approaches using locomotive wheel-sets as a case study. Reliab. Eng. Syst. Saf. **134**, 143–156 (2015)
37. M. Kurant, P. Thiran, Layered complex networks. Phys. Rev. Lett. **96**(13) (2006)
38. V.Y. Guleva, M.V. Skvorcova, A.V. Boukhanovsky, Using multiplex networks for banking systems dynamics modelling. Procedia Comput. Sci. **66**, 257–266 (2015)
39. R. Mittal, M.P.S. Bhatia, Anomaly detection in multiplex networks. Procedia Comput. Sci. **125**, 609–616 (2018)
40. S. Opricovic, G.-H. Tzeng, Extended VIKOR method in comparison with outranking methods. Eur. J. Oper. Res. **178**(2), 514–529 (2007)
41. L. Zhang, W. Dong, D. Zhang, G. Shi, Two-stage image denoising by principal component analysis with local pixel grouping. Patt. Rec. **43**(4), 1531–1549 (2010)

Chapter 5
Operational Risk Analysis of Rail Transportation Network

5.1 Operational Risk Assessment Model

Plenty of operational risk analysis of urban railway transportation network have been carried out. Diversified characteristics, which develop toward intelligent, systematic and in-depth trend, and focus on the analysis of risk factors of 'man-machine-environment-management' and formulation of relevant standards [1]. Brian put forward measures to improve safety of railway transportation from the people, vehicles, lines, laws and other operational risk factors of urban railway transportation [2]. Shi and others made a depth analysis of the main problems of operational safety of railway transportation from technical equipment, network transport capacity, operation organization, emergency and other aspects [3]. Lv [4] explored the factors that influenced the operational safety of urban railway transportation, which could be divided into internal and external factors. The internal factors included the equipment state, design reasons and personnel quality. Besides, vehicle system, maintenance system, signal system, communication system, power supply system and other system factors were among the equipment state. The external factors included personnel interference, construction annoyance, criminal activities, terrorist activities, natural climate and other factors. As for the above operational risk factors, some prevention, control measures and rescue measures were put forward. Ye [5] established four evaluation index systems by 'man-machine-environment-management', aiming at the complex system of urban railway transportation, such as the safety evaluation index systems of operational equipment, personnel, external environment and government of urban railway transportation.

The operational safety changes of urban railway transportation network can be brought about by the interaction and interrelationship of man, machine, environment and management [6]. In order to make the evaluation index system fully integrated with field operation, plenty of investigation and analysis have been made in the urban railway transportation, and the main factors of operational safety of urban railway transportation have been summarized in this chapter. The evaluation index

Y. Qin, L. Jia, *Active Safety Methodologies of Rail Transportation*, Advances in High-speed Rail Technology, https://doi.org/10.1007/978-981-13-2260-0_5

system based on the 'microcosmic, middle and macroscopic' level has been constructed above the main factors, and the operational safety evaluation of 'station – line road network' has been accomplished as well.

5.1.1 Operational Safety Assessment Index System of Metro Station

The metro station is a huge and complex system. The change of metro station's risk is the result of interaction and interrelationship of the four elements: people, equipment, environment and management. Based on the survey of metro station, the number of accidents, casualties and economic loss can be found to affect the risk of the metro station. As a summary, the main factors which affect the risk of metro station are passenger flow, equipment, environment, management and accident.

People are the most critical and flexible elements in the system, the performance of people directly affect the operation of the metro station. In order to reflect the risk state of the metro station, some features of passenger flow are selected. The load degree of AFC (Auto Fare Collection) at the exit and entrance reflects the use of AFC. The value of the load degree is higher, the speed of passengers is slower, which leads to passenger gathering and has a stampede risk. The congestion degree of platform, stairs, passageway and escalator reflects the intensity of platform, stairs, passageway and escalator. The value of the congestion degree is higher, means the passenger flow is larger and the overcrowding and stampede events are more likely to occur.

As one of the factors, the equipment are the basis for the safe operation of the metro station. The operating state of the equipment must be highly concerned about in the risk assessment. The equipment in the metro station consist of escalator, drainage system, fire alarm system (FAS), screen door system, lighting system, air conditioning system.

Environment has a significant impact on safe operation of the metro station. It affects the safety in a subtle way. Its influence may be a positive effect, but sometimes also be a negative effect. The temperature, humidity, PM2.5, PM10 and CO2 are taken into account.

Management plays a central role in the metro station system. It penetrates into every aspect to prompt various elements into a whole. Good management can strengthen the positive effect and impair the negative effect, which provide good conditions for safe operation of the metro station.

In addition, the number of accidents, the casualties and economic loss during the accidents are taken into account.

In summary, the safety evaluation index system is built up from the five indices: passenger flow, equipment, environment, management and accident. The safety evaluation index system of metro station can be shown in Fig. 5.1.

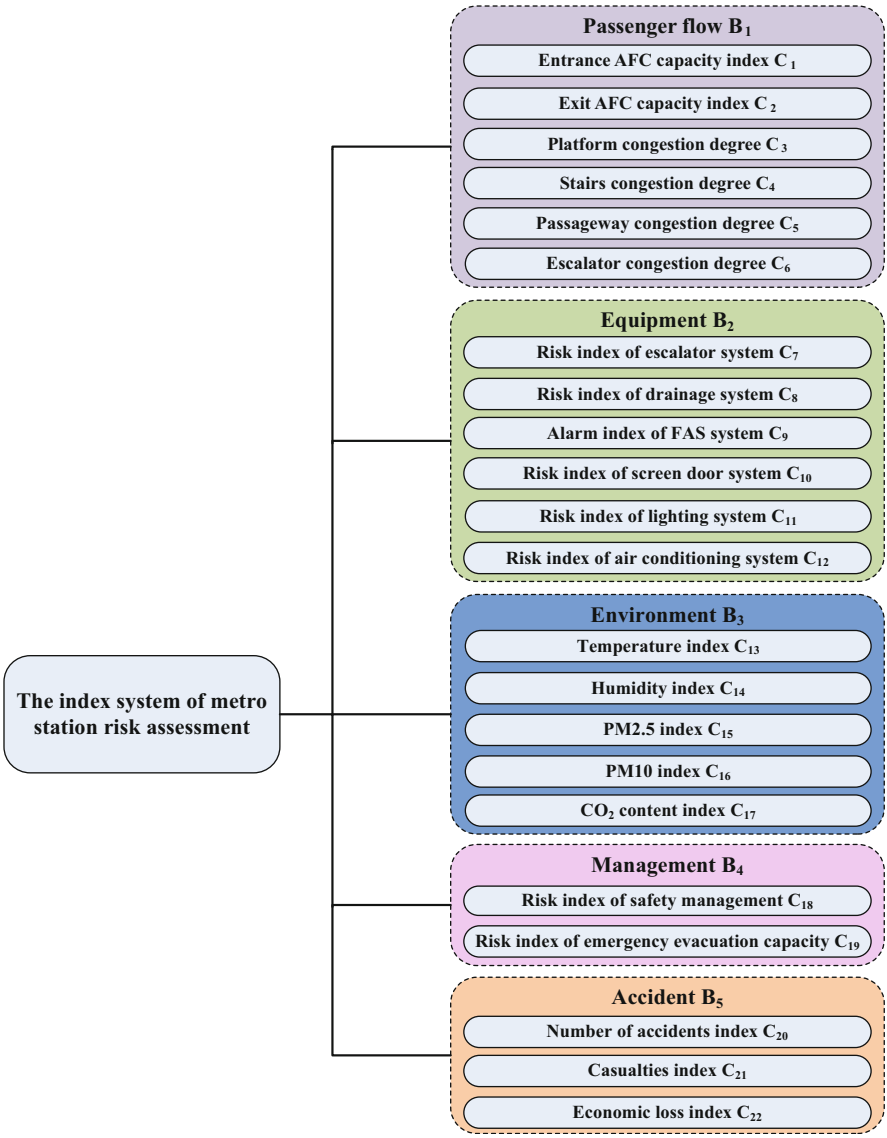


Fig. 5.1 The index system of metro station risk assessment

This chapter analyzes the influence of indicators on the station's safety situation, and on the basis of field investigation and literature analysis, deduced the calculation formula of each index, as shown in Table 5.1. Quantitative values of each index calculated can objectively reflect the station's risk state and ensure the accuracy of the result on metro station risk assessment.

Table 5.1 The calculation formula of each index

The meaning of indicators	Calculation formula & parameters
$C_1/C_2/C_4/C_5/C_6$ is the load degree of entrance AFC/exit AFC/stairway/passageway/escalator during the peak times of passenger flow of station	$\begin{cases} C = \sum_{i=1}^{n} \lambda_i \cdot S_i \\ S_i = \frac{Q_i}{C_i} \end{cases}$ S_i—The load degrees of the i-th entrance AFC/exit AFC/stairway/passageway/escalator; λ_i—The weight of the i-th entrance AFC/exit AFC/stairway/passageway/escalator; Q_i—The actual passenger traffic of the i-th entrance AFC/exit AFC/stairway/passageway/escalator; C_i—The traffic capacity of the i-th entrance AFC/exit AFC/stairway/passageway/escalator; n—The number of entrance AFC/exit AFC/stairway/passageway/escalator
C_3 is the ratio of the actual assembling on the station platform and the actual area of the station platform during the peak times of passenger flow of station	$C_3 = \frac{(Q_1 \cdot T_1 + Q_2 \cdot T_2)}{S} \phi$ Q_1—The pitted people of per second;Q_2—The outbound people of per second; S—Actual area of the station platform, m^2;T_1—Train arrived in time interval,s;T_2—The longest travel time from the platform to the station hall, s; Φ—The platform uneven coefficient of passenger flow density
$C_7/C_8/C_9/C_{10}/C_{11}/C_{12}$ is the ratio of escalator/drainage/FAS/shielding door/lighting/air conditioning system failure numbers and system numbers of station	$C = \frac{M}{N}$ M—Escalator/drainage/FAS/shielding door/lighting/air conditioning system failure numbers; N—Escalator/drainage/FAS/shielding door/lighting/air conditioning system numbers
$C_{13}/C_{14}/C_{15}/C_{16}/C_{17}$ is the ratio of the measurement value of temperature/humidity/particles smaller than 2.5 micrometers/particles smaller than 10 micrometers/CO_2 and the standard value of station	$C = \frac{c}{S}$ c—The measurement values; S—The standard values
C_{18} is the comprehensive evaluation for the safety management of the station of station	$C_{18} = 1 - \frac{c}{1000}$ c—The score given by the experts
C_{19} is the accident evacuation time of the platform layer during the peak time of passenger flow. Of station	$C_{19} = \frac{Q_1 + Q_2}{0.9[A_1(N-1) + A_2 B]}$ Q_1—Train passenger numbers; Q_2—The total number of waiting passengers and staff on the station platform; A_1—Through capacity of escalator, people/min·m; A_2—Through capacity of stairway, people /min·m; N—The number of escalator; B—The total width of the stairway, m; 0.9 is reduction factor
C_{20} is the number of accidents during of station	$C_{20} = \sum_{i=1}^{5} w_i s_i$ w_i—The weight of i-th accident; S_i—The number of i-th accident; accidents include

(continued)

Table 5.1 (continued)

The meaning of indicators	Calculation formula & parameters
	special major accidents, major accidents, accidents, accident insurance and general accident.
C_{21} is the casualty rate of the station during statistical period of station	$C_{21} = \frac{n}{N}$ n—The number of casualties during statistical period; N—The passenger flow in the station
C_{22} is the economic loss caused by accidents of station	$C_{22} = \sum_{i=1}^{5} w_i s_i$ w_i—The weight of i-th accident; S_i—The economic loss of i-th accident

5.1.2 Operational Safety Assessment Index System of Traffic Line

The traffic line, which is composed of various metro stations, is an important part of urban railway transportation network, and the change of traffic line's risk is also the result of interaction and interrelationship of the five elements: people, equipment, environment, management and accident.

The factors of people can be the risk of stampede, when total scale of passenger flows exceeds the maximum transmission capacity of the line. Therefore, the maximum full load rate and the mean of full load rate of each section at peak hour at both upside and downside direction can be referred to the risk formula.

As for the equipment, vehicle system, as the carrier of passenger transport, plays an important role in the operation of urban railway transportation. Signal system is one of the key facilities to ensure the safety of operation and improve the operation efficiency of urban railway transportation. Besides, the interlocking equipment which can monitor and record the working state itself contributes a lot to the safety of urban railway transportation. Power supply system is also an important link to the safety of urban railway transportation, whose power supply mode can be divided into unilateral, bilateral and beyond area supply modes. Communication system consisted of transmission system, private line system, closed circuit telephone system, broadcast system and wireless system can make a great impact on the safety of urban railway transportation. Electromechanical system aimed at aeration system at interval tunnel links in the safety operation chain. The civil engineering system mainly considers the design and the structure of underground line, overhead line, station building, vehicle base and operation center. The failure of platform screen door system can directly affect the number of normal thoroughfare for vehicles. The line system mainly considers the line and its subsidiary system. The rail damage is a prominent problem in the line system which has a close relationship with driving

safety and can severely affect the safety operation of railway train. AFC and safety system is an important link to ensure the safe arrival of passengers, which can also influence the safety state of traffic line. Apart from the above equipment factors, other factors such as external environment can also have an effect upon the safety of urban railway transportation.

Analogy to the safety management at station, the safety management at traffic line can also be the safety operation standardization evaluation index of urban railway transportation. Different from the risk of accident at station, the risk of accident at traffic line need combine operation mileage with the number of accidents. Besides, the influence of stations at traffic line can also impact the safety of urban railway transportation. Therefore, the passenger and environment comprehensive index of line and station can be proposed to measure the safety state of traffic line.

In summary, safety evaluation index system of traffic line basically designed with reference to the standards of the metro station's safety evaluation index system, shown in Fig. 5.2. Based on the field investigation and literature analysis, the deduced calculation formula of each index for traffic line can be shown in Table 5.2.

5.1.3 Operational Safety Assessment Index System of Traffic Network

The traffic network of urban railway transportation is made up of different metro stations and traffic lines, which makes it more complex to analyze the safety state of urban railway transportation. The change of traffic network's risk is also the result of interaction and interrelationship of the five elements: people, equipment, environment, management and accident.

As for the factors of people, the capacity matching of lines risk index can be presented by the full load rate of each section, which reflects the transfer matching between the lines. Concerning about the equipment, the impaction of vehicle, signal, power supply, communication, civil engineering, platform screen doors, AFC, safety system and other factors of network can be measured by the weighted mean of each system's operational risk index. With regard to the environment for the network, the comprehensive risk can be measured by the comprehensive index of environment at each line based on Weighted Algorithm. Analogy to the factor of accidents at line, the risk of accident at traffic network also need combine operation mileage with the number of accidents.

In short, safety evaluation index system of traffic network basically designed with reference to the standards of the metro line's safety evaluation index system, which can be shown in Fig. 5.3. Based on the field investigation and literature analysis, the deduced calculation formula of each index for traffic line can be shown in Table 5.3.

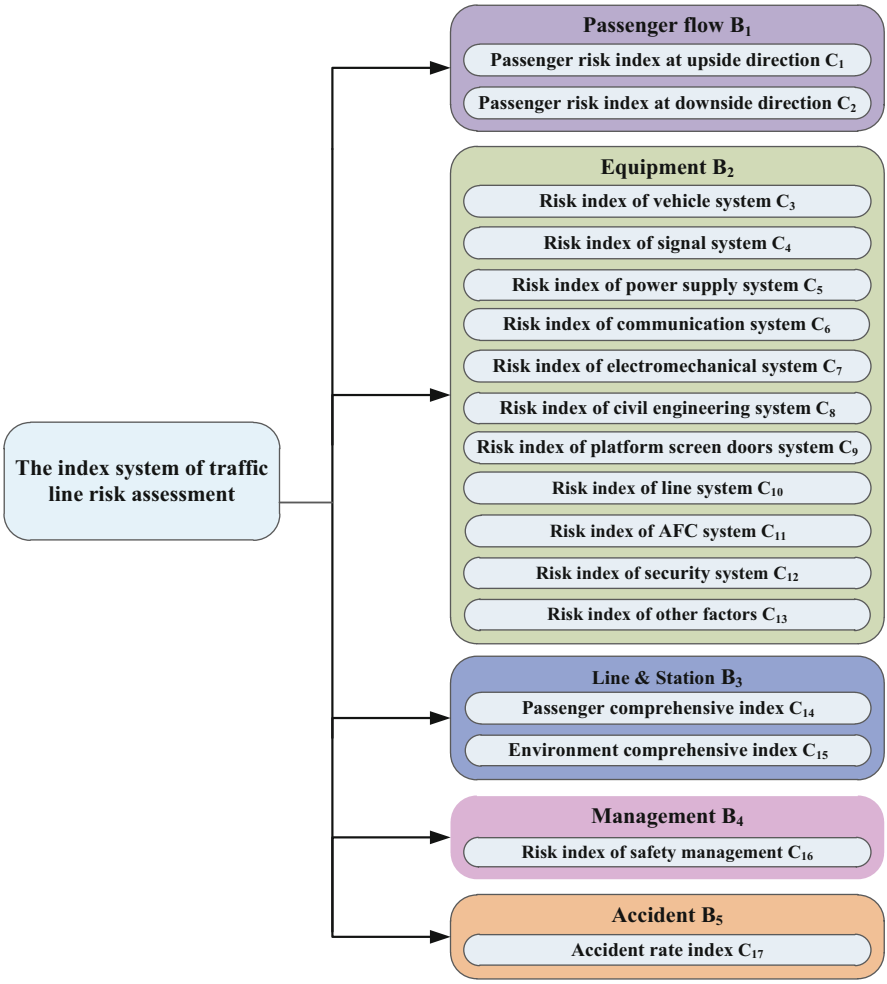


Fig. 5.2 The index system of traffic line risk assessment

5.2 Operational Risk Prediction Model

Railway transportation network, as the backbone of urban public transport system, is a place for passengers to wait, ride, drop and transfer. Railway transportation operation system has the characteristics of highly specialized operation, complicated infrastructure and large passenger flow. Once an accident occurs, it will cause a lot of damage to the economy and society. Therefore, it is of great significance to guarantee the safe and reliable operation of railway transportation system.

Table 5.2 The calculation formula of each index

The meaning of indicators	Calculation formula & parameters
C_1/C_2 is the risk index of passenger upside/downside direction of line	$c = w_1\frac{P_{\max}}{P} + w_2\frac{n_a}{N} + w_3\frac{n_t}{T}$ $P_{\max}$—The maximum full load rate of section at peak hour at upside/downside direction; P—The mean of full load rate of section at upside/downside direction; n_a—The number of section whose full load rate is over 70% at upside/downside direction; n_t—The number of time interval whose full load rate is 100% at upside/downside direction; N—The number of sections; T—The time of operation; w_i—Weight coefficient
$C_3/C_4/C_5/C_6/C_7/C_8/C_9/C_{10}/C_{11}/C_{12}/C_{13}$ is the operational risk index of vehicle/signal/power supply/communication/electromechanical/civil engineering/platform screen doors/AFC/safety system/other factors of line	$c = w_1\frac{t}{T} + w_2\frac{d}{D} + w_3\frac{n}{N}$ t—The failure time of vehicle/signal/power supply/communication/electromechanical/civil engineering/platform screen doors/AFC/safety system/other factors; T—The time of operation; d—The affected operation mileage of vehicle/signal/power supply/communication/electromechanical/civil engineering/platform screen doors/AFC/safety system/other factors; D—The operation mileage of main line; n—The failure times of vehicle/signal/power supply/communication/electromechanical/civil engineering/platform screen doors/AFC/safety system/other factors; N—The operation mileage; w_i—Weight coefficient
C_{14}/C_{15} is the passenger/environment comprehensive index of line & station	$c = \sum_{i=1}^{m} w_i Q_i$ Q_i—The comprehensive index of passenger/environment at station; w_i—Weight coefficient; m—The number of stations
C_{16} is the risk index of safety management of line	$c_{16} = 1 - \frac{c}{1000}$ c—The score given by the experts
C_{17} is the accident rate index of line	$c_{17} = \sum_{i=1}^{n} \frac{w_i s_i}{l}$ w_i—The weight of i-th accident; S_i—The economic loss of i-th accident; l—The operation mileage; n—The number of accidents

In the safety state of urban railway transportation network, the safety region refers to the area where the safety state variables are determined in the traffic network system, and are applied to evaluate whether the traffic network system is safe or not. If the safety state of railway transportation network in the boundary near the critical value cannot be controlled well in the effective time, it will be worsening deeper in the future, in a state of uncontrollable, and eventually evolved into the accident. Therefore, in order to prevent the state in the boundary near the critical value from

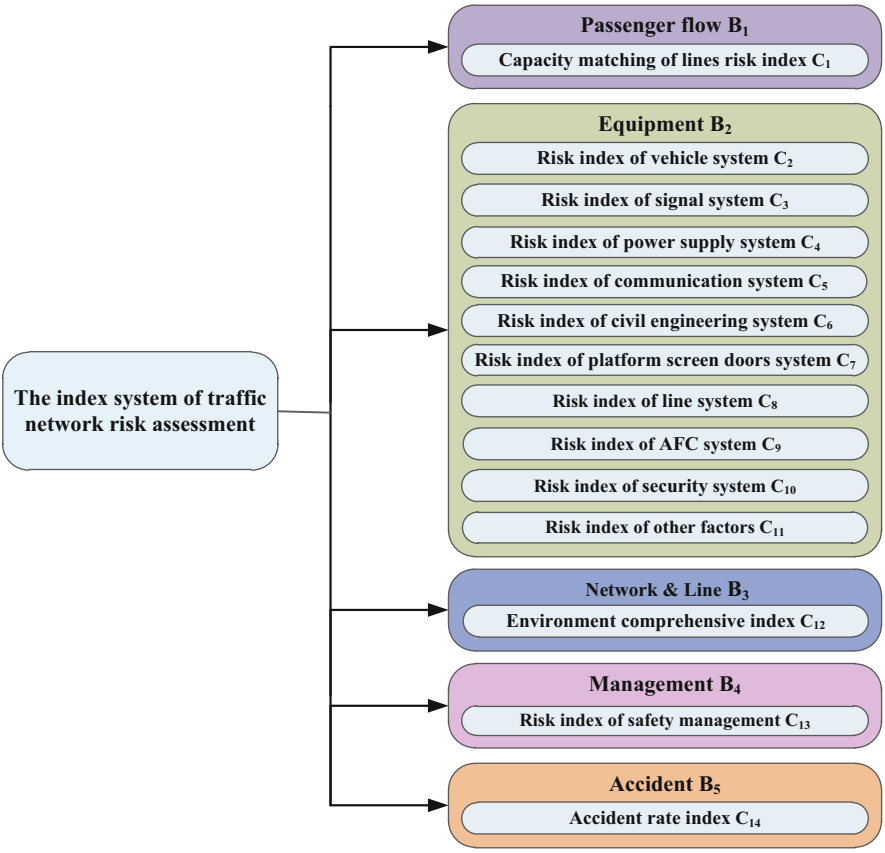


Fig. 5.3 The index system of traffic network risk assessment

deteriorating, the accurate prediction of the safety state of the railway transportation network is an inevitable trend for the realization of active safety research.

Accurate prediction of the safety state of the railway transportation network can reduce the risk factor of the network, to avoid accidents caused by the threat to the safety of the risk factor, bring about the realization from "passive safety" to "active safety", and it has an important significance to protect the safety operation of the traffic network for preventing accidents and reducing the loss of life and property of passengers.

At present, the methods of predicting the safety state of railway transportation network include the prediction of time series, neural network, gray model and support vector machine (SVM) [7, 8]. The key point of safety state prediction research is that the high precision prediction model. The data of safety state changes over time, which can be viewed as a group of time series arranged chronologically. Auto-regressive and moving average model (ARMA model) has better prediction accuracy in predicting such data [9]. Support vector regression (SVR) has the

Table 5.3 The calculation formula of each index

The meaning of indicators	Calculation formula & parameters
C_1 is the capacity matching of lines risk index of network	$C_1 = \sum k \cdot (w_1 \cdot p_a + w_2 \cdot (p_a - p_b))$ k—The ratio of transfer number of each direction to total transfer number of lines.; P_a—The full load rate of a section; P_b—The full load rate of the former section; w_i—weight coefficient
$C_2/C_3/C_4/C_5/C_6/C_7/C_8/C_9/C_{10}/C_{11}$ is the operational risk index of vehicle/signal/power supply/communication/civil engineering/platform screen doors/AFC/safety system/other factors of network	$C = \sum w_i x_i$ x_i—The operational risk index of vehicle/signal/power supply/communication/civil engineering/platform screen doors/AFC/safety system/other factors of line; w_i—weight coefficient
C_{12} is the environment comprehensive index of network & line	$C_{12} = \sum w_i Q_i$ Q_i—The comprehensive index of environment at line; w_i—weight coefficient
C_{13} is the risk index of safety management of network	$c_{13} = 1 - \frac{c}{1000}$ c—The score given by the experts
C_{14} is the accident rate index of network	$c_{14} = \sum \frac{w_i s_i}{l}$ w_i—The weight of i-th accident; S_i—The economic loss of i-th accident; l—The operation mileage

advantages of fast convergence, small absolute error, strong fitting ability and high accuracy in the prediction [10].

This chapter selects ARMA and GA-SVR method to build railway transportation network safety state prediction model and find a high precision prediction model through the comparative analysis, so as to realize the high precise prediction of railway transportation network safety state.

5.2.1 Safety State Prediction Based on ARMA Model

ARMA model is one of the most common models to describe stationary random sequence features, which has been structured and standardized, and it is more convenient to realize the model by the existing statistical software. The basic idea of safety state prediction of railway transportation network based on ARMA model is to regard the safety state value varying with time as time series. In this series, the safety state at time of n can be affected not only the disturbance of safety state at the former time from 1 to n-1, but also the safety state itself at the former time from 1 to n-1, so the safety state prediction model can be constructed. Safety state prediction of railway transportation network based on ARMA model can be viewed, shown as Eq. (5.2.1).

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{5.2.1}$$

Where $\{X_t\}$ represents value by zero mean processing.$\{\varepsilon_t\}$ is white noise with independent and identically distributed as $\{X_t\}$, and $E(\varepsilon_t) = 0$ while $Var(\varepsilon_t) = \sigma^2 > 0$. $\phi_1, \phi_2, \ldots, \phi_p$ and $\theta_1, \theta_2, \ldots, \theta_q$ represent the autoregressive coefficients and moving average coefficients of the model respectively, which can be expressed as ARMA(p, q).

The procedure of the safety state prediction of railway transportation network based on ARMA model can be expressed as follows.

(1) Collecting and preprocessing data.

Firstly, the safety state time series can be expressed as $\{x_1, x_2, \ldots, x_t\}$. Then, state curve can be plotted according to the safety state time series, and determine whether the curve changes periodically, if the periodic change exists, the safety state time series need to be differentiated. The new time series can be formulated according to $\{X_t\} = \{x_{t+i} - x_t\}$ with i as cycle length.

After that, the autocorrelation coefficients and partial autocorrelation coefficients of the new time series can be calculated according to Eqs. (5.2.2) and (5.2.3), and the autocorrelation analysis is performed.

$$\widehat{\rho}_k = \frac{\sum_{i=1}^{n-k} X_t^{'} \cdot X_{t+k}^{'}}{\sum_{i=1}^{n} X_t^{'} 2} \tag{5.2.2}$$

$$\widehat{\phi}_{kk} = \begin{cases} \widehat{\rho}_1, k = 1 \\ \dfrac{\widehat{\rho}_k - \sum_{j=1}^{k-1} \widehat{\phi}_{k-1,j} \cdot \widehat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \widehat{\phi}_{j,j} \cdot \widehat{\rho}_j}, k = 2, 3, \ldots \end{cases} \tag{5.2.3}$$

Let $\widehat{\phi}_{k,j} = \widehat{\phi}_{k-1,j} - \widehat{\phi}_{kk} \cdot \widehat{\phi}_{k-1,k-j}$.

Then, the self-correlation of the safety state time series can be analyzed according to the confidence interval $(-2/\sqrt{n}, 2/\sqrt{n})$ [11], and the stationarity and randomness of the new time series after zero mean processing need to be tested. As $k > 3$, the autocorrelation coefficients of the new time series tends to be 0, and are in the scope of the confidence interval, which can show that the new time series is relatively stable. While the autocorrelation coefficients of the new time series fall in the scope of the confidence interval, it can show that the new time series is random. As the new time series has both stationarity and randomness, the new time series after zero mean processing can be analyzed based on ARMA model for the next stage of prediction analysis.

(2) Model building and parameter estimation.

After the self-correlation analysis, the ARMA (p, q) model need to be chosen reasonably. The coefficient p and q need to be tested by the autocorrelation coefficient and partial autocorrelation coefficient of the new time series.

- If the autocorrelation coefficient is censored at q, $p = 0$, and the model is MA (q).
- If the partial autocorrelation coefficient is sensor at p, $q = 0$, and the model is MR (p).
- If both the autocorrelation coefficient and partial autocorrelation are trailed, the model is ARMA (p, q).

After the model is determined, the model order should be selected. At present, the time series method is more common in economy analysis, and the statistical software SPSS and SAS are more convenient to deal with such problems. These statistical software can directly output R square value and BIC value under different combinations. The larger the R value and the smaller the BIC value is, the higher the prediction accuracy of the model order will be. Then, the model order can be used as the final model order.

As for the selection of the model order in the statistical software, it can also output the estimation of parameters. Thus, the safety state prediction of railway transportation network based on ARMA model can be expressed as follows

$$X_t^{'} = \phi_1 X_{t-1}^{'} + \ldots + \phi_p X_{t-p}^{'} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \tag{5.2.4}$$

(3) Safety state prediction.

Finally, the safety state of the next stage can be predicted by the ARMA model In general, the more abundant the data is, the higher the prediction accuracy of the model will be.

5.2.2 Safety State Prediction Based on GA–SVR Model

(1) Support vector regression (SVR) model.

In 1995, Corinna Cortes firstly proposed the SVR model, which is superior to neural network model in dealing with the generalization problems. To a certain extent, SVR model can make up for the deficiencies in the structural risk minimization of neural network, and has been widely used in function approximation, pattern recognition and state prediction research. SVR model is based on SVM. Therefore, SVM can be introduced firstly.

The main idea of SVM can be introduced as: the main way in the regression or classification process in the Euclidean space is to determine the real function $g(x)$ of

R^n, and output variable y with the mapping relation to the variable x by decision-making function $f(x) = sgn[g(x)]$. The solution of $g(x)$ is to construct nonlinear mapping $\phi(\cdot)$ to express its duality nonlinear programming problem.

Kernel function $K(x_i, x) = \phi(x_i)^T\phi(x)$is the main way to complete the mapping of high dimensional space, and needs to meet the requirements of Mercer theory [12], that is, kernel function is expanded with positive coefficients αm, which can be shown as follows

$$K(u, v) = \sum_{m=1}^{\infty} \alpha_m \psi(u)\psi(v) \tag{5.2.5}$$

If the Eq.(5.2.5) is workable, the Eqs. (5.2.6) and (5.2.7) must be satisfied as well.

$$\iint K(u, v)g(u)g(v)\mathrm{d}u\mathrm{d}v > 0 \tag{5.2.6}$$

$$\int g^2(u)\mathrm{d}u < \infty \tag{5.2.7}$$

Kernel function plays a key role in SVM, common kernel function can be shown as:

- Sigmoid kernel function

$$K\left(x_i, x_j\right) = \tanh\left[v\left(x_i \cdot x_j\right) + c\right] \tag{5.2.8}$$

Sigmoid kernel function can only satisfy the Mercer theory in particular case of v and c, slightly different from the following Gauss radial basis and polynomial.

- Gauss radial basis kernel function

$$K\left(x_i, x_j\right) = \exp\left[-\left|x_i - x_j\right|^2 / \left(2\sigma^2\right)\right] \tag{5.2.9}$$

Gauss radial basis kernel function can be short for Gauss RBF kernel function, which has the highest frequency in SVM algorithm, in which the kernel parameter is expressed as σ.

- Polynomial kernel function

$$K\left(x_i, x_j\right) = \exp\left[-\left|x_i - x_j\right|^2 / \left(2\sigma^2\right)\right] \tag{5.2.10}$$

Where q is the order of the kernel function.

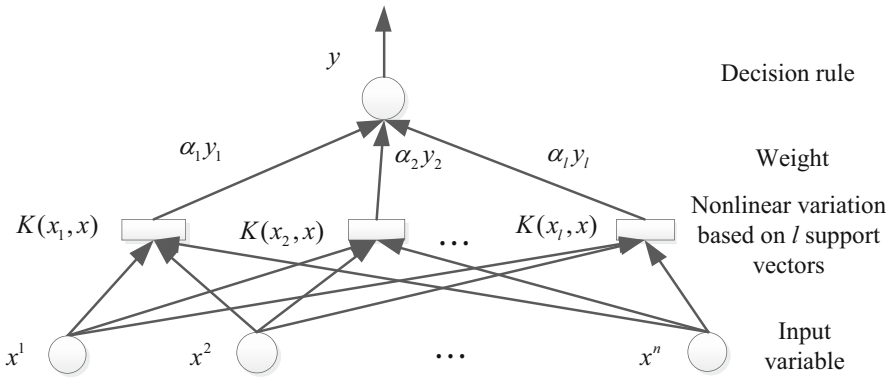


Fig. 5.4 Structure of support vector machine

According to the theory of kernel function, the decision function of Rn can be expressed as

$$f(x) = \text{sgn}\left[\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right] \tag{5.2.11}$$

Where l is the number of SVM, α_i and γ_i are the weight, b is the threshold.

The output of SVM is a linear combination of intermediate nodes, the structure of which is similar to the neural network, as shown in Fig. 5.4.

The main idea of the support vector regression, as a derivative of SVM, can be expressed as: determine the SVR input and output data (x_1,y_1), (x_2,y_2),..., (x_l,y_l), $x_t \in R^n$, $y_t \in R$, $t = 1, 2, \ldots, N$, t expresses the number of samples, $f(x)$ can be established by the training, and ensure that the difference between the target sample value and output value within a certain threshold, and that function is smooth. The model can be expressed as

$$f(x_i) = w \cdot \phi(x_i) + b \tag{5.2.12}$$

Where w and $\phi(\cdot)$ are respectively the mapping coefficients and nonlinear mappings in the high dimensional spaces, $b \in R$.

The SVR model can be considered as an optimization problem [12], and the model is expressed as follows.

$$\min \frac{\|w\|^2}{2} + C\sum_{i=1}^{t}\left(\xi_i + \xi_i{}^*\right) \tag{5.2.13}$$

$$\text{s.t.} \begin{cases} y_i - (w \cdot \phi(x_i)) - b \le \varepsilon + \xi_i \\ (w \cdot \phi(x_i)) + b - y_i \ge \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{5.2.14}$$

Where the insensitive loss function can be expressed as Vapnik ε, penalty factor $C > 0$, relaxation variable can be expressed asξ_i and ξ_i^*. If $|y_i - (\phi(x_i)^T w + b)| > \varepsilon$, the coefficient of insensitivity will be$|y_i - (\phi(x_i)^T w + b)| - \varepsilon$. If $|y_i - (\phi(x_i)^T w + b)| \le \varepsilon$, the coefficient of insensitivity will be 0.

The Lagrange function is introduced to solve the above problems, which can be expressed as

$$\begin{aligned} L = & \frac{\|w\|^2}{2} + C \sum_{i=1}^{t} (\xi_i + \xi_i^*) - \sum_{i=1}^{t} \alpha_i (y_i - (w \cdot \phi(x_i)) - b + \varepsilon + \xi_i) \\ & - \sum_{i=1}^{t} \alpha_i^* (y_i - (w \cdot \phi(x_i)) - b + \varepsilon + \xi_i^*) - \sum_{i=1}^{t} (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \tag{5.2.15}$$

Where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \ge 0$. Take the derivative of Eq. (5.2.16):

$$\left.\begin{aligned} \frac{\partial L}{\partial b} &= \sum_{i=1}^{t} (\alpha_i - \alpha_i^*) = 0 \\ \frac{\partial L}{\partial w} &= w - \sum_{i=1}^{t} (\alpha_i - \alpha_i^*) x_i = 0 \\ \frac{\partial L}{\partial \xi_i^{(*)}} &= C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \end{aligned}\right\} \; t = 1, 2, \cdots, N \tag{5.2.16}$$

The dual variables can be obtained by substituting the Lagrange function with the Eq. (5.2.17):

$$\begin{aligned} W(\alpha, \alpha^*) = & -\frac{1}{2} \sum_{i,j=1}^{t} (\alpha_i - \alpha_i^*) \left(\alpha_j - \alpha_j^*\right) K(x_i, x_j) \\ & + \sum_{i=1}^{t} (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^{t} (\alpha_i + \alpha_i^*) \varepsilon \end{aligned} \tag{5.2.17}$$

When $\sum_{i=1}^{t} (\alpha_i - \alpha_i^*) = 0$, α_i, α_i^* can be obtained by maximizing $W(\alpha, \alpha^*)$, and can be brought it into regression function, the result can be expressed as

$$f(x) = w \cdot \phi(x) + b = \sum_{i=1}^{t} (\alpha_i - \alpha_i^*) \phi(x_i) \cdot \phi(x) + b$$
$$= \sum_{i=1}^{t} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{5.2.18}$$

(2) Safety state prediction based on GA-SVR model.

Genetic Algorithms (GA) is a global optimization algorithm based on the principle of natural selection and natural genetic mechanism, which can simulate the life evolution mechanism and achieve the optimization of specific target in artificial system. The essence of GA is to get the global optimal solution based on group search technology and the principle of survival of the fittest [1]. The basic procedure of the algorithm includes coding, selection, crossover, mutation, fitness function and selection of control parameters, as shown in Fig. 5.5.

SVR mainly applies nonlinear mapping to map input to high-dimensional state space, so as to solve nonlinear regression problem based on the linear regression function in high dimensional space. Nonlinear mapping $\phi(\cdot)$ is usually composed of kernel functions, which have great influence on generalization and learning ability of SVR as well as the parameter selection. The common kernel functions mentioned above include Sigmoid function, Gauss radial basis function, polynomial kernel function, etc. In the absence of prior knowledge, RBF kernel function is better suited to deal with this problem in comparison with other functions [2]. Safety state prediction of railway transportation network is lack of prior knowledge. Therefore, RBF kernel function is suitable for dealing with the safety state prediction of railway transportation network, which can be expressed as the Eq. (5.2.17).

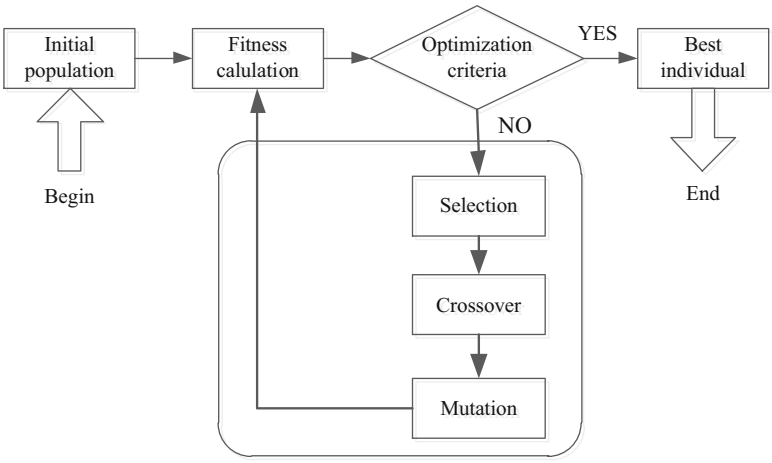


Fig. 5.5 Optimization of GA

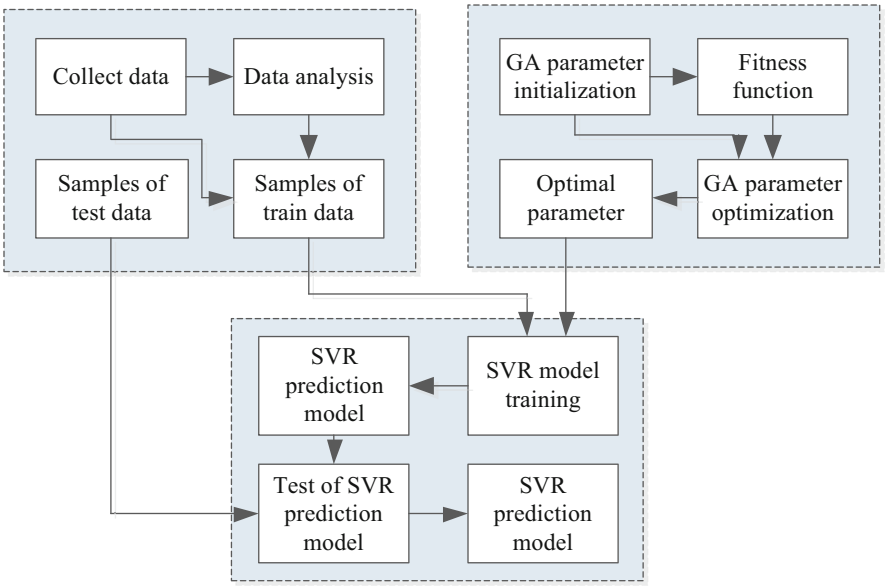


Fig. 5.6 Process of GA-SVR model

The determination of the kernel function is very important, and the selection of the kernel parameter σ, the insensitivity coefficient ε, and the penalty parameter C must be chosen reasonably. σ expresses the nuclear width, which mainly reflects the distribution of training samples. The value of σ can affect the size of the RBF function fitting and generalization ability. The better fit ability is, the smaller the kernel width will be. However, the value cannot be too small for the generalization ability. The value of ε can affect the number of SVM. The smaller the value is, the weaker the generalization ability of the model will be, the more the SVM will be and the higher the complexity and accuracy of the model will be C is a parameter to compromise the generalization ability and complexity of SVR. The smaller the value is, the stronger the generalization ability of model will be and the more the VC (Vapnik-Chervonenkis) dimensional weight will be. However, when the value of C is small to a certain extent, it can also cause the failure of sensitive coefficient and the poor training result in the end [12]. The optimization process of GA is mainly the optimal selection process of parameters such as kernel parameter, insensitivity coefficient and penalty factor. The basic procedure of prediction model construction based on GA-SVR can be expressed as follows, which can be shown in the Fig. 5.6.

- Collect and select sample data. Preprocess the above data and divide training samples and test samples.
- Determine the GA algebra, initial population size, fitness function and so on. Determine the nuclear parameters, insensitivity coefficient and penalty parameter of SVR model by GA.

- Input the optimal parameters obtained by GA and data preprocessed into SVR model.
- Test the prediction accuracy of the trained model by the pretreated test data and construct the final prediction model.

In this paper, the root mean square error *RMSS* and correlation coefficient *R* are used as the indices to evaluate the predictive performance of ARMA and SVR models.

$$\mathrm{RMSE}(y, y_m) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y(i), y_m(i))^2} \tag{5.2.19}$$

$$R(y, y_m) = \frac{\sum_{i=1}^{N}(y(i) - \bar{y})(y_m(i) - \bar{y}_m)}{\sqrt{\sum_{i=1}^{N}(y(i) - \bar{y})^2 \cdot (y_m(i) - \bar{y}_m)^2}} \tag{5.2.20}$$

Where N is the number of samples, y_m, y, $\bar{y}$, $\overline{y_m}$ are the model predictive values, the safety state value, the average value of the safety state, the average value of the model prediction respectively. The closer the *R* value is to 1, the better the prediction fitting effect will be. The more the RMSS value is, the lower the prediction accuracy will be.

5.3 Case Study

5.3.1 Case Study on ARMA Model

This section applies the safety state data of Beijing railway transportation network from January to September in 2013, which can be divided into two groups: training data and test data. The first 243 groups are training data, and the latter 30 groups are test data.

By analyzing the results of safety state assessment, a more obviously cyclical change rule cannot be found among the data. Therefore, the data need to be processed by zero mean and autocorrelation analysis by SPSS software, which can be shown in the Fig. 5.7.

The autocorrelation analysis shows that both the autocorrelation coefficient and the partial autocorrelation coefficient begin to fall into the confidence interval as $k = 14$, presenting an increasing trend without obvious convergence, which indicates that the trailing property can be found both in the autocorrelation coefficient and the partial autocorrelation coefficient. Therefore, the ARMA model can be judged, the correlation test results can be shown in Table 5.4.

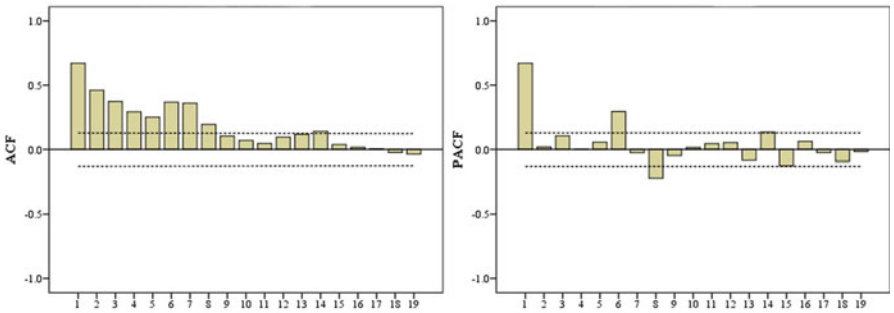


Fig. 5.7 Auto correlation function and partial correlation function of safety state

Table 5.4 Test result of model

Model	R square	RMSE	BIC
ARMA(14,14)	0.593	0.105	−3.843
ARMA(14,15)	0.591	0.106	−3.809
ARMA(15,14)	0.592	0.106	−3.813

Table 5.4 shows that ARMA (14, 14) has the largest R square test and the value is 0.593, the root mean square error of RMSE and BIC are 0.105 and − 3.843 respectively. In contrast to the three models, the mean of ARMA (14, 14) model is the smallest, which indicates that ARMA (14, 14) model has the best prediction accuracy. Therefore, the best model can be ARMA (14, 14), and the model parameters can be estimated as shown in Table 5.5.

The result predicted by the ARMA model can be shown in Fig. 5.8, the blue line represents the actual state of the safety value, and the red line represents the fitting value of the ARMA model. In the first 243 sets data, the red line and the blue line have a higher degree of anastomosis, and the fitting effect is better. However, the latter 30 sets data maintain a downward trend.

5.3.2 Case Study on GA–SVR Model

The selection of kernel function parameters, penalty parameter and insensitivity coefficient are closely related to the accuracy of SVR prediction model. This chapter applies RBF as kernel function and applies root mean square error as fitness function to select kernel parameter, insensitivity coefficient and penalty parameter, which can be calculated as reference Eq. (5.2.14). The coding method of genetic algorithm adopts entity coding, and the initial population size is 20, the maximum evolution algebra is 100, and the excellent individuals are selected by roulette mode. The crossover way of individuals is arithmetic crossover, the cross probability is 0.7, and the mutation probability is 0.05. The value range of the parameters C, ε and σ are [0.1, 1000], [0, 1] and [0.001, 100] respectively. In the process of genetic algorithm

Table 5.5 Parameter estimation of ARMA model

Parameter	Estimated value	Parameter	Estimated value	Parameter	Estimated value	Parameter	Estimated value
Φ_1	0.231	Φ_8	−0.158	Θ_1	−0.460	Θ_1	0.165
Φ_2	0.343	Φ_9	−0.093	Θ_2	0.080	Θ_2	0.282
Φ_3	0.319	Φ_{10}	0.066	Θ_3	0.343	Θ_3	0.351
Φ_4	−0.366	Φ_{11}	0.004	Θ_4	−0.243	Θ_1	0.101
Φ_5	−0.075	Φ_{12}	0.161	Θ_5	−0.009	Θ_1	0.183
Φ_6	0.433	Φ_{13}	−0.296	Θ_1	0.204	Θ_1	−0.033
Φ_7	0.435	Φ_{14}	−0.092	Θ_1	0.329	Θ_1	−0.293
Constant	−0.007						

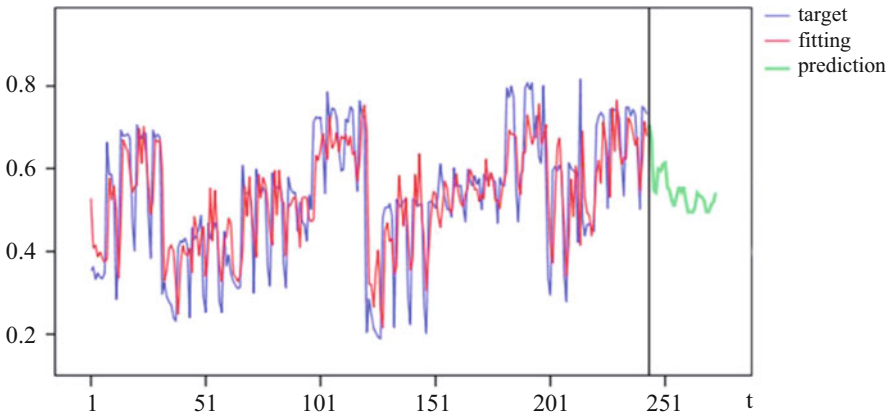


Fig. 5.8 Safety state prediction result of ARMA model

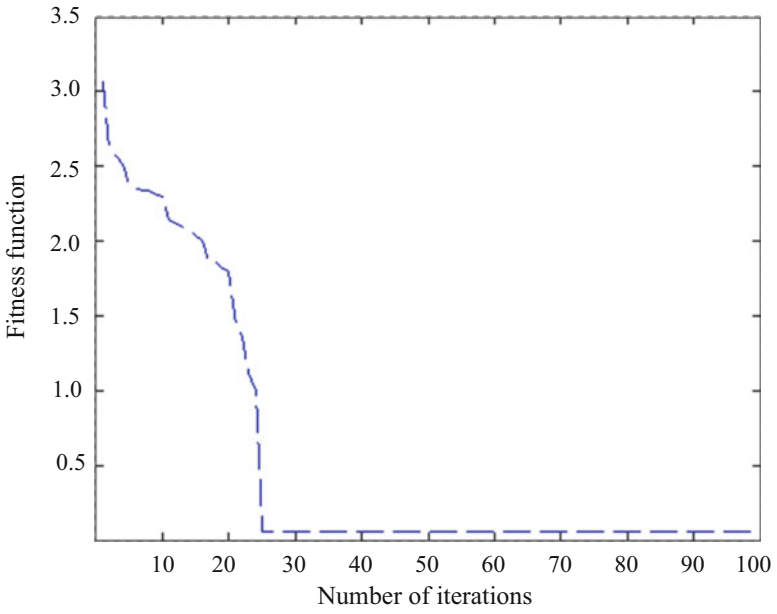


Fig. 5.9 Fitness curve of safety state

optimization, the changing process of fitness function of safety state can be shown in Fig. 5.9. The model achieves steady state in the 25 generation of iteration. At this time, the best parameter can be obtained by genetic algorithm, and the best parameters C, ε and σ are 11.3618, 0.1022 and 0.0020 respectively.

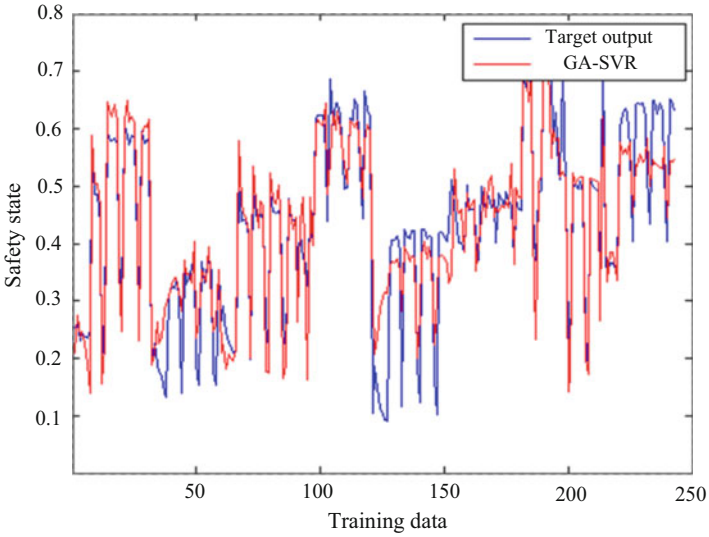
In order to facilitate the analysis and comparison of the safety state prediction effect based on GA-SVR model and ARMA model, the same 243 training samples and 30 test samples were used for the experiment.

Training samples and testing samples of the target output and GA-SVR output of the model can be shown in the Fig. 5.10, which can contrasted in the figure. The model can accurately track the target output value in the training process, with a smaller error. The safety state output of GA-SVR model can be in good agreement with the target, which indicates that GA-SVR model can be well used to predict the actual safety state of situation.

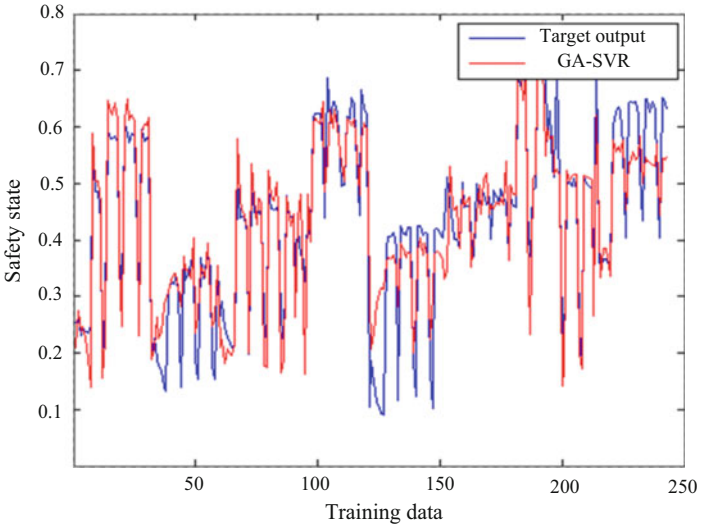
In order to make a further evaluation of the prediction accuracy and performance of safety state based on GA-SVR model, the correlation between the training samples and testing samples can be shown in the Fig. 5.11. Correlation coefficients of the training and testing safety state are 0.9106 and 0.9238 respectively, which are both higher than 0.8, indicating the good prediction effect.

At the same time, the root mean square error RMSE of the training samples and testing samples based on GA-SVR model is 0.0645 and 0.0831 respectively, which shows that the prediction results have a small error. As a whole, GA-SVR prediction model can accurately predict the safety state of the railway transportation network.

The safety state predicted by the ARMA and GA-SVR models can be compared in the Fig. 5.12 and the Table 5.6. The safety state based on GA-SVR model maintains a downward trend in the last 30 samples, while the real safety state value fluctuates up and down. The safety state curve predicted by the GA-SVR model is more consistent with the real safety state curve, which also shows that the accuracy of GA-SVR model is higher than that of ARMA in this prediction. Based on the comparative analysis of two models in the Table 5.6, the root mean square error of training and testing of GA-SVR model are 0.0645 and 0.0831 respectively, which are both less than 0.1, indicating the better prediction effect than that of ARMA model's 0.1049 and 0.2042. R, the correlation coefficient of the training and testing samples are 0.9106 and 0.9238 respectively, better than that of ARMA model's 0.7407 and 0.0261. Besides, the correlation coefficient of the training and testing samples are both higher than 0.9, which can meet the requirements to be higher than 0.8. Therefore, GA-SVR model can be selected as the safety state prediction model.

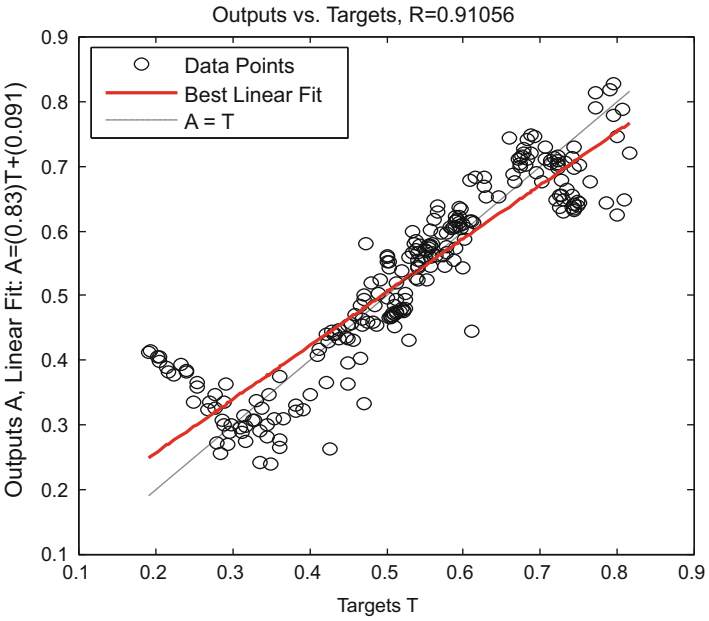


(a)Training sample (safety state)

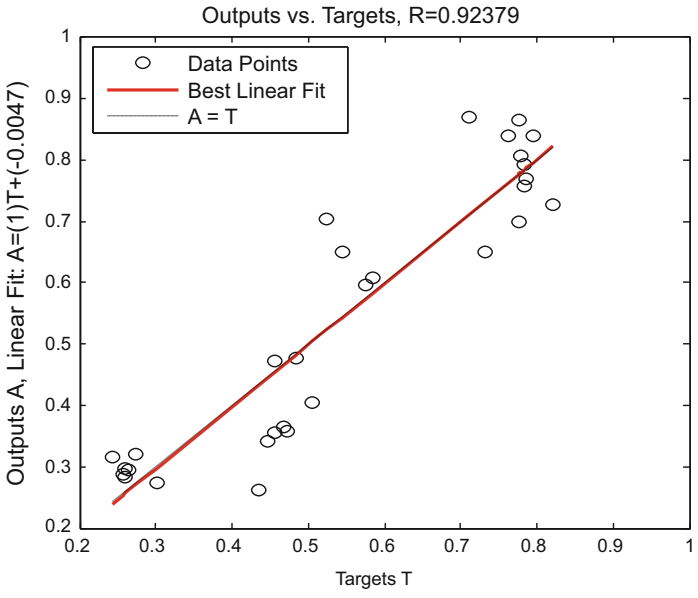


(b) Testing sample (safety state)

Fig. 5.10 Comparison of targets and GA-SVR model predicted values (**a**)Training sample (safety state) (**b**) Testing sample (safety state)



(a) Training data (safety state)



(b) Testing data (safety state)

Fig. 5.11 Correlation coefficient of GA-SVR predicted values and actual results (**a**) Training data (safety state) (**b**) Testing data (safety state)

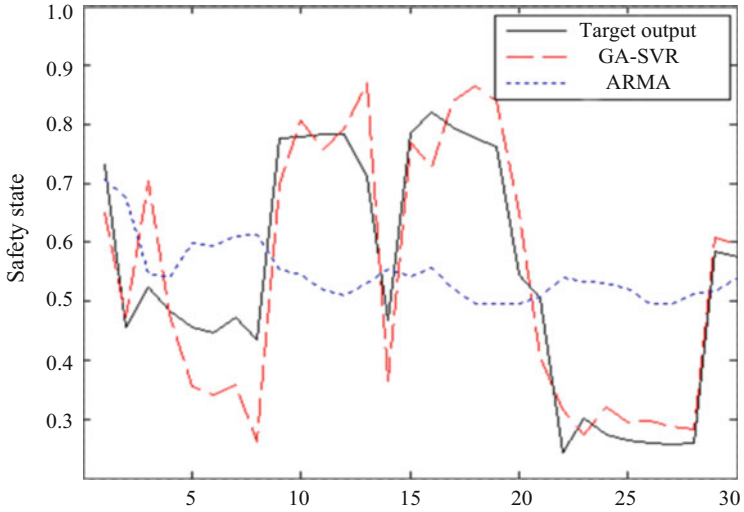


Fig. 5.12 Safety state prediction results of ARMA and GA-SVR

Table 5.6 Safety state prediction results of ARMA and GA-SVR

Model	ARMA	GA-SVR
R of the training samples	0.7407	0.9106
R of the testing samples	0.0261	0.9238
RMSE of the training samples	0.1049	0.0645
RMSE of the testing samples	0.2042	0.0831

References

1. Y.K. Huang, Research of the assessment method of the city rail transportation network. Beijing Jiaotong University (2014)
2. B.L. Mishara, Railway and metro suicides: Understanding the problem and prevention potential. J. Crisis Interv. Suicide Prev. **28**(S1), 36–43 (2007)
3. Y.F. Shi, C. Yang, L.T. Sun, The management analysis of the city rail traffic. Res. City Rail Transp. **6**(2), 26–28 (2003)
4. C.J. Lv, The operation safety analysis of the city rail traffic. Shanghai Railw. Technol. **3**, 52–53 (2006)
5. Pw.W. Ye, The efficiency assessment research of the city rail transportation based on the envelope analysis. Beijing Jiaotong University (2009)
6. M. Li, The comprehensive assessment model research of the city rail traffic. Beijing Jiaotong University (2012)
7. X.I. Rong-Rong et al., Research survey of network safety situation awareness. J. Comput. Appl. **32**(1), 1–133 (2012)
8. K. Gao et al., A hybrid safety situation prediction model for information network based on support vector machine and particle swarm optimization. Power Syst. Technol. **35**(4), 176–182 (2011)
9. R.Y. Li, R. Kang, The research of the fault rate based on the ARMA model. Syst. Eng. Electron. Technol. **30**(8), 1588–1591 (2008)

10. M.R. Endsley, R. Sollenberger, E. Stein, Situation awareness: A comparison of measures (2000)
11. M.A. Abdel-Aty, R. Pemmanaboina, Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data. IEEE Trans. Intell. Transp. Syst. **7**(2), 167–174 (2006)
12. D. Wang et al., Prediction of total viable counts on chilled pork using an electronic noise combined with support vector machine. Meat Sci. **90**(2), 373–377 (2012)

Chapter 6
Safety Prognostic Analysis in Traffic System

6.1 Traffic Operation Risk Analysis Model Based on Safety Region

6.1.1 Sequential Forward Selection and Principal Components Analysis

The quality of observed traffic variables (e.g., speed, volume, occupancy) can influence the effectiveness of data mining/machine learning algorithms in safety region estimation. If the observed traffic variables contain irrelevant or redundant features, the knowledge discovery process becomes noisy and unreliable. In this paper, a state variable extraction method, combining sequential forward selection and principal components analysis (SFS-PCA), is considered to construct the state space of the traffic system.

Supposing vector V is the input vector of the SFS-PCA method. V is composed of the observed traffic variable vector X and the corresponding class label vector Y. $Y = \{Y_1, Y_2, \ldots, Y_N\}^T$, $Y_l \in \{1,-1\}$, where $l = 1,2,\ldots, N$, $Y_l = 1$ corresponds to crash case and $Y_l = -1$ corresponds to crash case and $Y_l = -1$ corresponds to non-crash case.

The V can be denoted as:

$$V = \{(X_l, Y_l)|1 = 1, 2, \ldots, N\} = \{(x_{l1}, x_{l2}, \ldots, x_{lm}, Y_l)|1 = 1, 2, \ldots., N\} \quad (6.1.1)$$

where X_l is the lth sample in X, each X_l contains m observed traffic variables.

The goal of SFS-PCA method is to find a minimal set of state variables $F = \{f_1, f_2, \ldots, f_k\}$ ($k \leq m$.) to represent the observed traffic variables in a lower dimensional state space. The SFS-PCA method can be described as follows:

The best possible subset S (i.e.,$S \subset X$) of the observed traffic variables is selected by SFS firstly. SFS starts from an empty set, and then iteratively updates S by including the observed traffic variable X_i ($i = 1, 2, \ldots, m$) which results in maximal

Y. Qin, L. Jia, *Active Safety Methodologies of Rail Transportation*, Advances in High-speed Rail Technology, https://doi.org/10.1007/978-981-13-2260-0_6

score $G(S, X, M)$ in each step [1]. Thus, the size of S, denoted by d ($d \leq m$), is given by

$$S_d = S_{d-1} \cup \arg\max_{X_i} G(S_{d-1} \cup X_i, X, M) \tag{6.1.2}$$

where M denotes the k-nearest neighbor model, which is used as a classification model to evaluate $G(S, X, M)$.

After the SFS procedure, the final state variable set F is extracted from S by PCA. PCA decomposes S into two subspaces (a lower dimensional feature subspace composed of principle components and a residual subspace) by multiple projections. Two statistic indicators, $T^2$and squared prediction error (SPE), are calculated in the two subspaces respectively [2]. T^2 reflects the change of the principle component model in feature subspace and SPE measures the interference and noise in the residual subspace.

T^2 and SPE can be calculated by using the following formulas respectively:

$$T_l^2 = s_l P_b \lambda^{-1} P_b^T s_l^T l = 1, 2, .., N \tag{6.1.3}$$

$$\mathrm{SPE}_l = s_l \left(I - P_b P_b^T\right) s_l^T l = 1, 2, .., N \tag{6.1.4}$$

where s_l is the lth sample in subset S, P_b is the matrix of the b loading vectors, which could be calculated by PCA, I is the identity matrix.

6.1.2 Computation Procedure

The implementation procedures of traffic operation risk analysis based on safety region are shown as follows

Step 1. Collect crash data and non-crash data as the training data for traffic risk evaluation method. Crash data include crash information (time, location) and the matched traffic flow data collected from the traffic surveillance system (speed, volume, occupancy). Non-crash data are traffic flow data in the given times interval when the traffic states are under the safe condition.

Step 2. Extract state variables through SFS-PCA. First of all, obtain subset S from the observed traffic variable set by using SFS. Secondly, process subset S of the observed traffic variables by PCA, and calculate statistics T^2 and SPE. The two statistics forms a two-dimensional statistical state vector for each sample, and the vectors would be the final state variable set F.

Step 3. Use the two-dimensional statistical state vector as the input data for LSSVM. Classify the traffic states into safe state and unsafe state and obtain the best classified line which is the boundary of the traffic safety region.

Step 4. With the validation data, distinguish the state points in the safety region from the state points in the unsafe region. If a state point is in the unsafe region, it

means the corresponding traffic sate is under the unsafe condition. Otherwise, if the state point is in the safety region, it means the traffic state is under the safe condition and the corresponding safety margin is calculated.

6.1.3 Case Study

6.1.3.1 Data Description

In this study, field data were obtained from a 35-mile freeway section on the I-880 freeway in Alameda in the United States and the studied segment started from milepost 10.55 and ended at milepost 45.42. A total of 70 loop detector stations, which spaced at approximate 0.5 mile, located along the selected freeway segment in the northbound direction. Crash data and the paired real-time traffic flow data were collected from January 1, 2011 to December 31, 2012. All the data were provided by the Highway Performance Measurement System (PeMS) [3], which is maintained by the California Department of Transportation (Caltrans).

Caltrans PeMS database provided raw loop data, i.e., speed, volume and occupancy, for each lane at 30-s intervals. The raw data were prepared by first aggregating into 5-min intervals. For example, time interval 0:00 denoted the measuring period from 0:00:00 to 0:04:59, and time interval 23:55 denoted the measuring period from 3:55:00 to 23:59:59. And then each crash was assigned to the nearest loop detector (as shown in Fig. 6.1). The corresponding traffic flow data 5–10 min prior the crash time and the traffic flow data in the crash time were selected to represent the traffic condition. At the same time, the traffic data of upstream and downstream were also extracted. For example, if a crash happened at 13:32, at the milepost 15.46. Traffic condition of nearest loop detector at milepost 15.54 in time intervals 13:20, 13:25 and 13:30 is the corresponding traffic state for this crash. To eliminate the geometric characteristics' influences on crash risk evaluation, matched case-control structure was used to extract non-crash data [4]. For each specific crash case, two non-crash cases, one week before and one week after the crash time, were identified and matched. For example, a crash

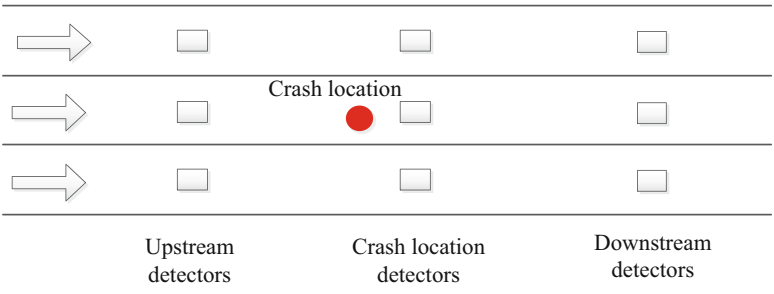


Fig. 6.1 Illustration of field data collection

happened on April 26, 2011, the corresponding non-crash cases (April 19, 2011 and May 3, 2011) at the location of crash occurrence were selected.

For each sample, average and standard deviation values of the speed, occupancy, and volume for the three detectors ($2 \times 3 \times 3 = 18$) constituted the observed traffic variable set. In this study, a total of 417 crashes and 837 non-crash cases were identified and used for further data analysis.

6.1.3.2 State Variable Extraction

Variables important scores are calculated via SFS and the subset S is determined based on the scores. A total of 8 observed traffic variables are selected, i.e., downstream standard deviation of speed (DDS), crash location average occupancy (CAO), upstream standard deviation of speed (UDS), crash location standard deviation of occupancy (CDO), downstream average speed (DAS), upstream standard deviation of occupancy (UDO), crash location standard deviation of speed (CDS) and downstream average occupancy (DAO). Furthermore, multi-collinearity test for the 8 selected traffic variables has been carried using SPSS and the correlation coefficients between two variables in the subset are calculated, as listed in Table 6.1. The results imply that some of variables exist highly correlated relations, e.g., the correlation between DAO and DAS is 0.825, approximating to 1, which suggests that a further analysis should be conducted on the selected traffic variables before being used in the following classification models. In order to eliminate the high correlation among the selected traffic variables, the PCA is applied to the observed traffic variable subset, see Fig. 6.2. Cumulative percentage of total variation 80% rule is used to determine the number of components. Finally, three components are chosen. Figure 6.2 shows the cumulative proportion for the first 3 components. The variances of first 3 components are 0.475, 0.223 and 0.137, respectively. The corresponding two statistics T2 and SPE are calculated by using Eqs. 6.1.3 and 6.1.4, which would be the final input state variable set to the LSSVM method.

Table 6.1 Correlation matrix for selected observed traffic variables

	DDS	CAO	UDS	CDO	DAS	UDO	CDS	DAO
DDS	1	0.069	−0.014	0.057	0.171	0	−0.413	0.042
CAO	0.069	1	0.096	−0.357	0.041	−0.202	0.153	−0.348
UDS	−0.014	0.096	1	0.237	0.045	**−0.729**	−0.422	0
CDO	0.057	−0.357	0.237	1	0.097	−0.312	**−0.699**	0.121
DAS	0.171	0.041	0.045	0.097	1	0.016	−0.115	**0.825**
UDO	0	−0.202	**−0.729**	−0.312	0.016	1	0.283	0.053
CDS	−0.413	0.153	−0.422	**−0.699**	−0.115	0.283	1	−0.096
DAO	0.042	−0.348	0	0.121	**0.825**	0.053	−0.096	1

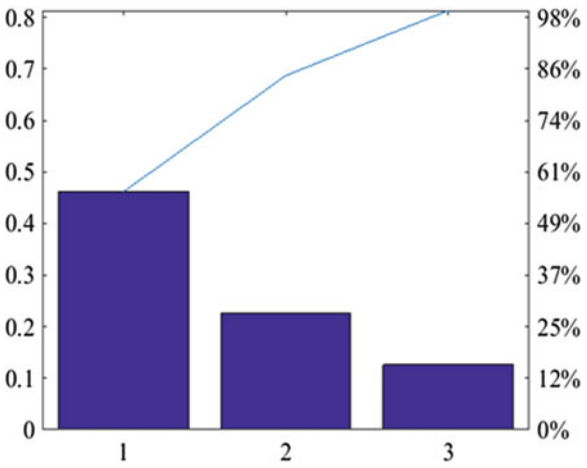


Fig. 6.2 The cumulative variance proportion of the first 3 components

6.1.3.3 Traffic Safety State Identification

The k-fold cross-validation is employed during the classification experiments. The state variable set is divided into k subsets, and the classification method is repeated k times. Each time, one of the k subsets is used as the validation dataset and the other k-1 subsets are put together to form the training set [5]. In this paper, the four-fold cross-validation is taken, which implies that 75% of the whole data set is used as the training dataset and the rest 25% of the whole data set is used to form the validation dataset. The corresponding classification results of training dataset and validation dataset are shown in Fig. 6.3. In Fig. 6.3a, the safety region boundary, which tries to classify the crash state data and non-crash state data into two regions, is estimated by using training dataset. Figure 6.3b plots the testing data points on the classified region. It is intuitive that the SFS–PCA-LSSVM method works efficiently in classifying the traffic safety states. However it is necessary to evaluate these results in a quantitative way.

$$\mathrm{CR} = \frac{\text{The number of samples correctly classified for all classes}}{\text{The total number of samples}} \times 100\% \tag{6.1.5}$$

Classification accuracy is used to evaluate the classification performance. The classification accuracy for the dataset is measured by correct rate (CR).

The classification performance of the proposed method is compared with the SFS. LSSVM and PCA-LSSVM methods. According to Eq. 6.1.5, the CR of the SFS-PCA-LSSVM is 88.84%, the CR of the SFS-LSSVM is 77.19% and the CR of the PCA-LSSVM is 68.82%, as listed in the first row of the Table 6.2. Furthermore, in

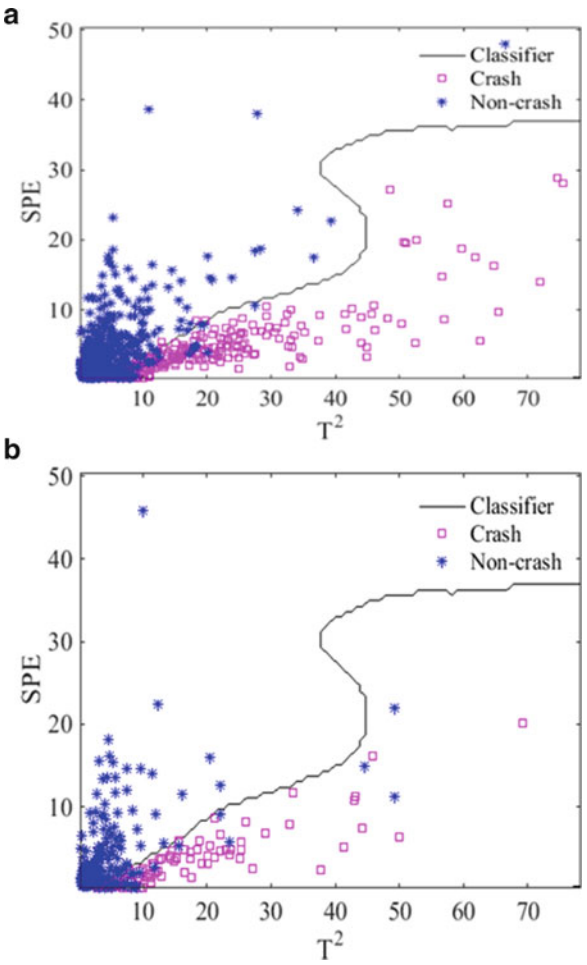


Fig. 6.3 Fold classification results: (**a**) training dataset and (**b**) test dataset

Table 6.2 CR and AUC Values for three methods in different training dataset

Criteria	Training dataset	SFS-LSSVM	PCA-LSSVM	SFS-PCA-LSSVM
CR	4-fold	77.19%	68.82%	88.84%
	3-fold	76.12%	68.34%	88.28%
	2-fold	75.54%	67.70%	88.16%
AUC	4-fold	0.6983	0.5495	0.8894
	3-fold	0.6854	0.5456	0.8850
	2-fold	0.6887	0.5456	0.8806

order to demonstrate the classification performance, the classification results of three methods are plotted. As shown in Fig. 6.4, the horizontal axis represents the number of state points. The 1st ~ 417th points are crash state samples, the 418th ~ 1254th are non-crash state points. The whole area is divided into two subareas by pink dotted

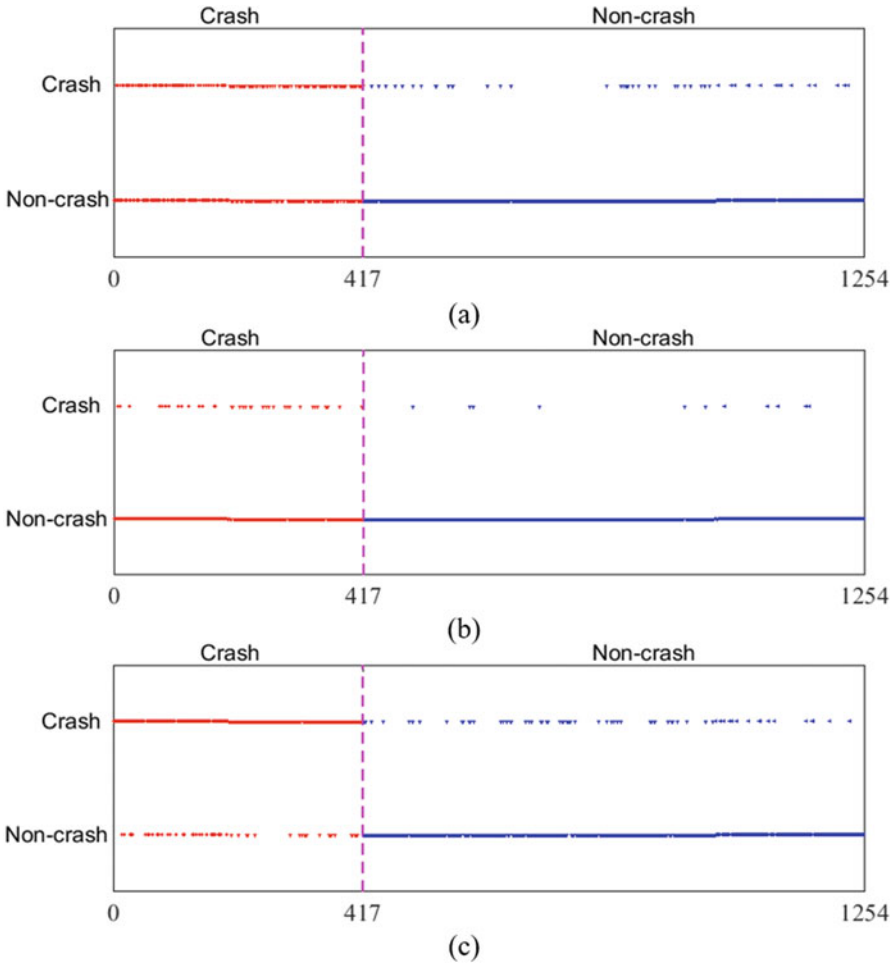


Fig. 6.4 The classification results of three methods: (**a**) SFS-LSSVM (**b**) PCA-LSSVM (**c**) SFS-PCA-LSSVM

lines. Vertical axis from bottom to top represents the classifier results for the two classes of data samples. In Fig. 6.4a and 6.4b, it can be seen that the scatter points of the non-crash classifier results are denser than that of the crash classifier results, which means that the first two models misclassify most of the crash state point as the non-crash points. However, the proposed model identifies most of the samples correctly, especially in the crash sample set. It can conclude that the proposed method performs better than other two methods.

As the safety region estimation is dependent on the size of training dataset. In this paper, two additional cross-validation experiments, i.e., three-fold cross-validation and two-fold cross-validation, are conducted. The CR and AUC values for the three models mentioned above are calculated. As listed in Table 6.2, both the CR and

AUC values of SFS-PCA-LSSVM are higher than that of other two models. While the size of training dataset becomes bigger, the corresponding CR and AUC values increase. It can be concluded that for different size of training dataset, sufficient training data may improve the classification results. Moreover, state variables extraction procedure is needed prior to the data mining/machine learning algorithms. In this paper, experiments with different variable extraction methods are conducted. It can be concluded that hybrid intelligent methods perform better than single intelligent methods.

According to the obtained safety region boundary, the traffic states are classified into two classes: safe condition and unsafe condition. When the traffic state points are in the safety region, the corresponding safety margin would be calculated. In this paper, subdivision algorithm is applied to calculate the safety margin. An example of the safety margin calculation is given in Fig. 6.5a. The state point P (39.3533, 22.7842) represents the traffic condition of the non-crash case at milepost 13.14 in time interval 16:25, June 21, 2012. With calculation, a point (41.705, 26.6418) on the boundary is found to own the shortest Euclidean distance, 4.5167, which is the value of the corresponding safety margin to this state point.

Safety margin has been applied into the safety risk prediction of the rail system's key equipment. In highway traffic system, safety margin also can be used to estimate the traffic crash risk. According to the non-crash case mentioned above, the corresponding crash case happened at milepost 13.14 in time interval 16:25, June 28, 2012. The corresponding state point is denoted by P_3. Two time intervals, i.e., 16:20 and 16:15, prior to the crash time interval also are analyzed, which are denoted by P_2 and P_1 respectively. Three state points are calculated and plotted in Fig. 6.5b. The safety margin for P_1 is 5.4525, for P_2 is 3.1780. Then in the next time interval, traffic crash happens. When the state points get close to the boundary of safety region, the crash risk may increase. The results allude that safety margin could be used to grade the traffic crash risk and predict the traffic crash risk in the future works.

6.2 Traffic Crash Risk Evaluation Model Based on Reliability Theory

6.2.1 Structural Reliability Analysis Theory

The structural reliability analysis deals with the calculation of the failure probability under a defined limit state condition [6]. The failure probability of a structural can be formally calculated from

$$P_f = \int_{G(\mathbf{X}) \leq 0} f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x} \tag{6.2.1}$$

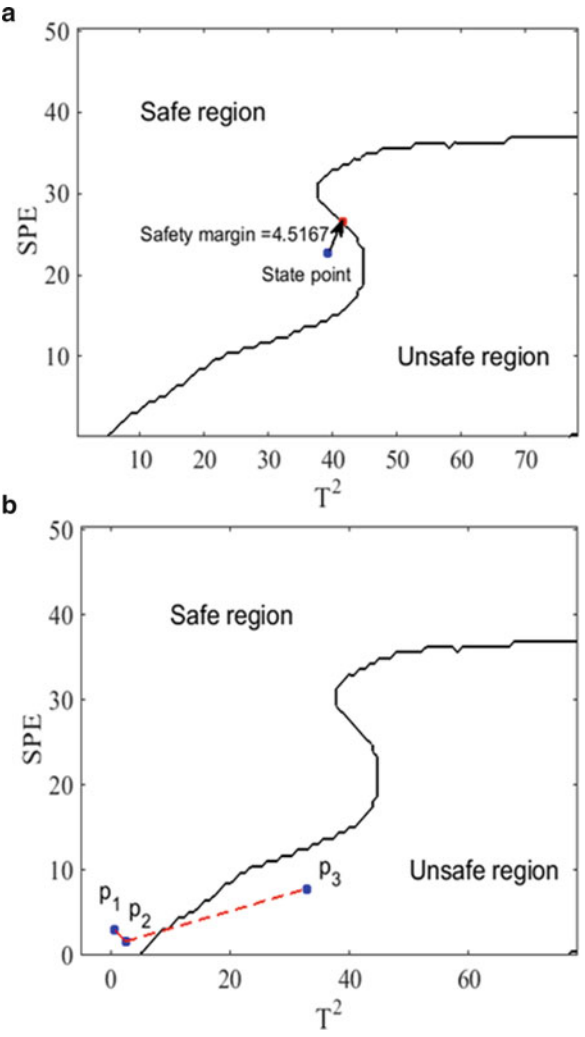


Fig. 6.5 Safety margin: (**a**) state point P (39.3533, 22.7842) and (**b**) values versus time

where x represents the vector of basic random variables, $y = G(x)$ is the limit state function (LSF), $G(x) = 0$ is the limit state surface separating the unsafe region $G(x) \leq 0$ from the safe region $G(x) > 0$ and $f_x(x)$ is the joint probability density function of the random variables x1, x2, . . ., xn.

However in most realistic scenarios, the joint probability density function $f_x(x)$ is always difficult to calculate, the approximate methods, such as the First Order Reliability Methods (FORM) [7] and the Second Order Reliability Methods (SORM) [8] are adopted to evaluate the failure probability. The main ideas of the approximate methods are all trying to calculate the shortest distance β between the origin and the limit state surface in standard normal space. The shortest distance β is

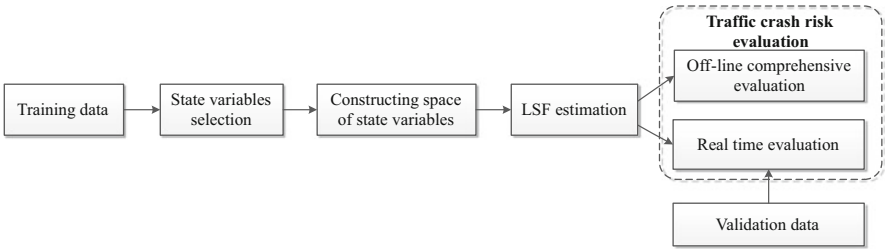


Fig. 6.6 Modeling procedure in this study

also termed reliability index. Then the failure probability could be evaluated by $Pf = \Phi(-\beta)$, where Φ is the standard normal cumulative distribution function.

The Hasofer-Lind index β [9] is the most popularly used in structural reliability analysis. Its matrix formulation could be expressed as follows

$$\beta = \min_{x \in F} \sqrt{(\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m})} \tag{6.2.2}$$

where x is a vector representing the set of random variables, m is the mean values, C is the covariance matrix, and F is the failure domain. In the original space of the random variables, the solution of Eq.(6.2.2) is equivalent to finding the smallest ellipsoid tangent to the limit state surface [10]. And the tangent point, termed design point, is the most probable failure point. The formula of the ellipsoid is given by quadratic form as follows.

$$(\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}) = \beta^2 \tag{6.2.3}$$

6.2.2 Analysis Procedure

This study adopted reliability analysis model to evaluate highway traffic crash risk. Figure 6.6 presents the flowchart of the main modeling procedure for this study. The reliability analysis procedure can be summarized as follows:

Step 1. The data are split into training and validation datasets. The training dataset is utilized to estimate the model and the validation dataset is meant to test the prediction performance of the reliability model.

Step 2. Random variables are selected as the state variables. The state variables would construct the state space of the highway traffic system. In this paper, the classification and regression tree (CART) is utilized to select the most significant contributing variables from the observed variables.

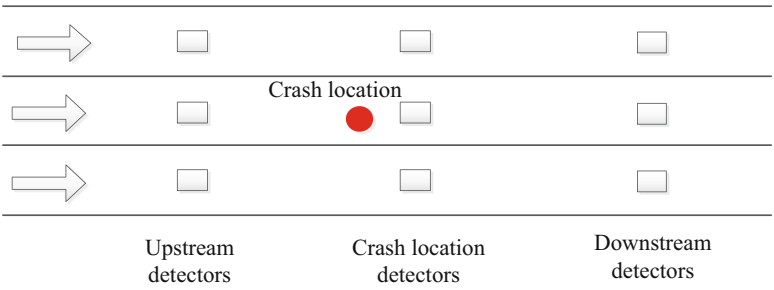


Fig. 6.7 Illustration of field data collection

Step 3. Base on the chosen state variables, the distribution of each state variable is estimated. Mathematical procedure is provided to calculate the joint probability density function.

Step 4. The support vector machines (SVM) model is adopted to approximate the LSF. The limit state surface separated the state space into two regions, i.e., safe region and unsafe region.

Step 5. On the basis of the given LSF, highway traffic risk is evaluated in two ways. One way is evaluating comprehensively off line, including the reliability index β, the probability of traffic crash Pf, and the design point; the other way is estimating the real time traffic risk with the validation dataset.

6.2.3 Case Study

In this section, we only focus on the traffic flow state of point of disruption occurred. That is means, the research data sets just be selected form the nearest loop detectors at the point of disruption occurred (as shown in Fig. 6.7). Furthermore, the other data sets clearing, selected and variables calculation methods are mentioned in Sect. 6.1.3. There are a total of 455 crashes and 1039 non-crash cases identified and used for further data analysis.

6.2.3.1 State Variable Selection

The classification and regression Tree (CART) have been adopted to select the significant variables from the 6 observed variables mentioned above. In this study, CART procedure is conducted in SAS Enterprise Miner with the following setting in the program: Splitting Criterion: Gini; Maximum Depth: 10; Leaf Size: 10; Split Size: 20; and Number of Surrogate: 3. Crash locations standard deviation of occupancy (CDO) and logarithm of crash locations average volume (Log CAV) are selected as state variable of traffic system. Moreover, the variable importance is

calculated based on the number of times a variable appeared and its relative position in the tree. Final variable selection results are presented in Table 6.3.

6.2.3.2 State Variable of Traffic System

In order to construct the state space for traffic system, distribution fitting should be done for each selected state variable and the joint distribution would be set up. Five candidate distributions, i.e., Normal, Gamma, Exponential, Lognormal, and Weibull distributions, are prepared to fitting the distribution for each state variable. The procedure is developed by MATLAB. During the procedure, the maximum likelihood method [11] is used to estimate the parameters of the distributions. Moreover, the likelihood-based statistics are supplied to indicate the data fitting of the estimated distributions. Among the likelihood-based statistics the Bayesian information criterion (BIC) is selected to identify the most appropriate distribution for the variables. The smaller the BIC value is, the better the distribution fits the data. Among the five candidate distributions, the Normal distribution provides the best fit for the CDO and Log CAV according by BIC, as shown in Table 6.4.

According to the results, the state variables CDO and log CAV satisfy the distribution of Normal Gamma, and the normal distribution is written as:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{6.2.4}$$

where μ and σ are the mean and standard deviation of the variables, respectively.

Table 6.3 Variable selection results by CART

Variables	Description	Mean	SD	Importance
CDO	Crash locations standard Deviation of occupancy	1.6727	1.2351	1
Log CAV	Logarithm of crash locations Average volume	2.5432	0.2626	0.785

Table 6.4 Distribution fitting results for state variables

	CDO			Log CAV		
Distribution	Converged	BIC	Selected	Converged	BIC	Selected
Normal	Yes	2415.90	Yes	Yes	31.24	Yes
Exponential	Yes	2433.10	No	Yes	3845.10	No
Lognormal	No	–	No	Yes	308.27	No
Gamma	No	–	No	Yes	210.18	No
Weibull	No	–	No	Yes	483.78	No

The process of gaining the joint probability density function of state space is the following.

Supposing x1 represents CDO, and x2 represents Log CAV. According to the distribution fitting results, x1 and x2 both follow Normal distribution. Transforming x1 and x2 into standard normal distributions, which have the formulations:

$$\mathbf{y}_1 = \frac{\mathbf{x}_1 - \mu_1}{\sigma_1}, \mathbf{y}_2 = \frac{\mathbf{x}_2 - \mu_2}{\sigma_2} \tag{6.2.5}$$

Supposing ρ is the correlation coefficient between y1 and y2. According to Cholesky decomposition [12], the covariance matrix C could be decomposed by C = LLT, where L= $\begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$.

Assuming $\mathbf{u}_1$, $\mathbf{u}_2$ are independent and follows standard normal distribution. $\mathbf{y}_1$ and $\mathbf{y}_2$ should satisfy $\mathbf{y}_1 = g_1(\mathbf{u}_1,\mathbf{u}_2) = \mathbf{u}_1; \mathbf{y}_2 = g_2(\mathbf{u}_1, \mathbf{u}_2) = \rho\mathbf{u}_1 + \sqrt{1-\rho^2}\mathbf{u}_2;$.

The joint probability density function of the state variables is set up as follows:

$$f_{\mathbf{y}_1\mathbf{y}_2} = \frac{f_{\mathbf{u}_1\mathbf{u}_2}}{|\, J(\mathbf{u}_1, \mathbf{u}_2)\,|}\Bigg|_{\substack{\mathbf{u}_1 = g_1^{-1}(\mathbf{y}_1, \mathbf{y}_2) \\ \mathbf{u}_2 = g_2^{-1}(\mathbf{y}_1, \mathbf{y}_2)}} = \frac{f_{\mathbf{u}_1} f_{\mathbf{u}_2}}{\sqrt{1-\rho^2}}\Bigg|_{\substack{\mathbf{u}_1 = g_1^{-1}(\mathbf{y}_1, \mathbf{y}_2) \\ \mathbf{u}_2 = g_2^{-1}(\mathbf{y}_1, \mathbf{y}_2)}} \tag{6.2.6}$$

where $J(\mathbf{u}_1, \mathbf{u}_2)$ is Jacobi matrix, and $|\, J(\mathbf{u}_1, \mathbf{u}_2)\,| = \begin{vmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{vmatrix} = \sqrt{1-\rho^2}$. By substituting Eqs. (6.2.4), (6.2.5) and (6.2.6), the joint probability density function can be formulated as follows:

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\mathbf{y}_1^2 - 2\rho\mathbf{y}_1\mathbf{y}_2 + \mathbf{y}_2^2\right]\right\}$$

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{\mathbf{x}_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\mathbf{x}_1 - \mu_1}{\sigma_1}\right)\left(\frac{\mathbf{x}_2 - \mu_2}{\sigma_2}\right) + \left(\frac{\mathbf{x}_2 - \mu_2}{\sigma_2}\right)^2\right]\right\} \tag{6.2.7}$$

From Eq. 6.2.7, it could conclude that the joint probability in the state space follows a bivariate normal distribution. As shown in Fig. 6.8, the joint probability density function is plotted in a 3- dimensional perspective.

6.2.3.3 Limited State Function Estimation

Support vector machine (SVM) is a statistical classification algorithm that classifies data by separating two classes with the help of a hyper plane. In structural reliability analysis, LSF acts in a similar manner to the hyper plane in SVM. In this study, SVM model is adopted to approximate the target limit state function $G(\mathrm{x}) = 0$.

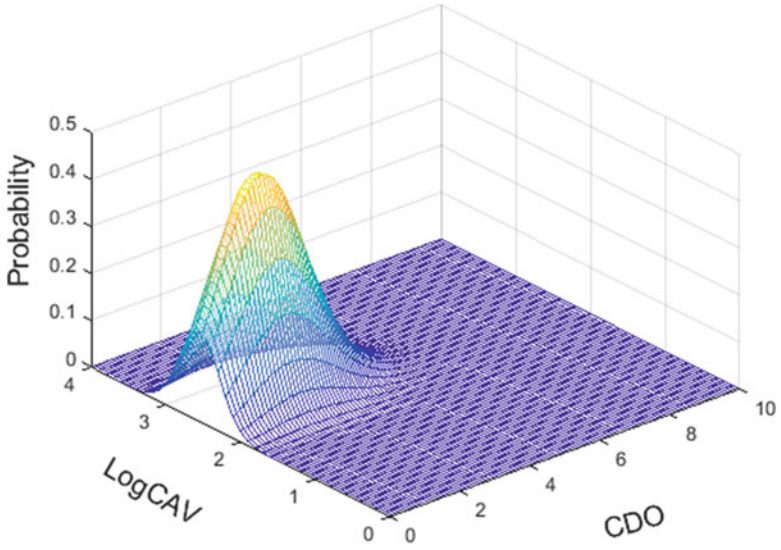


Fig. 6.8 The joint probability density function in state space

Let it be known that a set of N training set (X1, l1), (X2, l2), . . ., (XN, lN), where X = [x1, x2] T represents the matrix for variables selected by CART model, Xi = (xi1,xi2), $i = 1, 2, \ldots, N$, represents the ith sample of the training set, and li $\{-1, 1\}$, li $= -1$ responds to crash case and li $= 1$ responds to non-crash case.

According to the description of hyper plane function, the LSF can be defined as

$$G(\mathbf{x}) = \mathbf{w} \cdot h(\mathbf{x}) + b \tag{6.2.8}$$

where w is the normal to the limit state surface, h(x) is the mapping function and b is the bias value.

The LSF is regarded as the optimal separating hyper plane with maximum margin. By introducing the Lagrange multiplier, the optimization problem has the dual quadratic programming form:

$$\begin{cases} \min\limits_{\boldsymbol{\alpha}} \dfrac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} l_i l_j \alpha_i \alpha_j \big(h(\mathbf{X}_i) \cdot h\big(\mathbf{X}_j\big)\big) - \sum\limits_{i=1}^{N} \alpha_i \\ \text{s.t.} \quad \sum\limits_{i=1}^{N} l_i \alpha_i = 0 \\ \alpha_i \geq 0, \quad i = 1, 2, \ldots, N \end{cases} \tag{6.2.9}$$

The dual problem can be solved by using [12] and the α_i ($i = 1, 2, \ldots, N$) is calculated finally. Then the parameters of the LSF can be estimated as follows

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i l_i h(\mathbf{X}_i), b = \frac{1}{n_{\mathrm{SV}}} \left[\sum_{i=1}^{N} l_i - \sum_{i=1}^{N} \sum_{j=1}^{N} l_j \alpha_j \left(h(\mathbf{X}_i) \cdot h\left(\mathbf{X}_j\right) \right) \right. \tag{6.2.10}$$

where nSV is the number of support vectors. Support vectors are the vectors on the margin. And the LSF has the form

$$G(\mathrm{x}) = \sum_{i=1}^{N} \alpha_i l_i h(\mathbf{X}_i) h(\mathbf{X}) + b = \sum_{i=1}^{N} \alpha_i l_i K(\mathbf{X}_i, \mathbf{X}) + b \tag{6.2.11}$$

where $\boldsymbol{K}(\mathbf{X}_i, \mathbf{X})$ is the kernel function. In this paper, linear kernel is considered

$$K(\mathbf{X}_i, \mathbf{X}) = \mathbf{X}_i^T \mathbf{X} \tag{6.2.12}$$

In this section, the training data set, 70% of the total prepared data, is used for developing LSF and calculating the off-line comprehensive evaluation results. According to Eqs. (6.2.10) and (6.2.11), the LSF is set up as

$$G(\mathbf{x}) = -0.6617\,\mathbf{x}_1 - 1.6551\mathbf{x}_2 + 6.5076 \tag{6.2.13}$$

Figure 6.9 plots the LSF in 3-dimensional and 2-dimensional perspectives, respectively. In Fig. 6.9a, it is intuitive that the LSF $G(\mathbf{x}) = 0$ also satisfies a normal distribution. In Fig. 6.9b, a top view is provided. The limit state surface separated the state space into two regions: safe region and unsafe region, which is the fundament of calculating reliability index β, design point, and predicting the traffic crash risk.

6.2.3.4 Test Results Analysis

The probability of traffic crash Pf in a historical period is used as a comprehensive traffic crash risk evaluation criteria. According to the definition of reliability index β in structural reliability theory, β is closely related to the probability Pf. The bigger the β value is, the smaller the Pf. value would be. So the problem of estimating Pf can be transformed into calculating β. The ellipsoid approach via spreadsheet is adopted to calculate the reliability index β. In Fig. 6.10, the bigger ellipse, termed 1– σ ellipse, is corresponding to the quadratic formula, (x – m)T C-1 (x – m) = 1. The correlation coefficient ρ of state variables is 0.1004. The result suggests that the state variables are not highly correlated. So the semi-major axis of ellipse approximately parallels to the x coordinate axis shown as Fig. 6.11. The critical ellipse that is tangent to the limit state surface is β times the size of the 1– σ ellipse. The tangent point is the design point, which is the point of maximum traffic crash likelihood. In this paper, the value of β is 0.2912, and the design point is (3.4, 2.5726). The value of

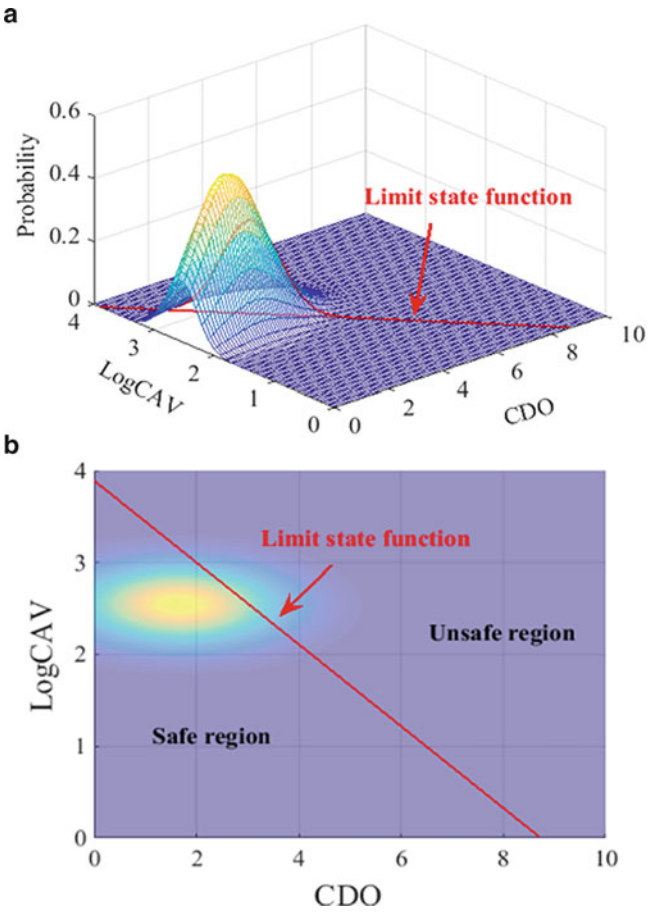


Fig. 6.9 The limit state surface in state space: (**a**) side view and (**b**) top view

design point means when the CDO approximates to 3.4, and Log CAV approximates to 2.5726, the crash occurrence of the highway traffic system is in a high probability. The probability of traffic crash is $Pf = \Phi(-\beta) = 0.3854$, which is the total probability of free highway traffic crash occurred. The reliability index β provides a general traffic crash risk estimation criteria, which could provide a critical index for classifying traffic safety grades, for example, most prone to traffic accidents, prone to traffic accidents, and not prone to traffic accidents.

According to calculation of reliability index β, we can evaluate the conditions of traffic system operation risk comprehensively. However, it also needs to do real-time traffic crash risk evaluation. With the validation data set (30% of the total prepared data), real-time traffic crash risk evaluation accuracy of traffic reliability model is tested in this section. Based on the foundation of state space (divided into safe region and unsafe region), when $G(x) \leq 0$, the traffic state is in an unsafe region, and It

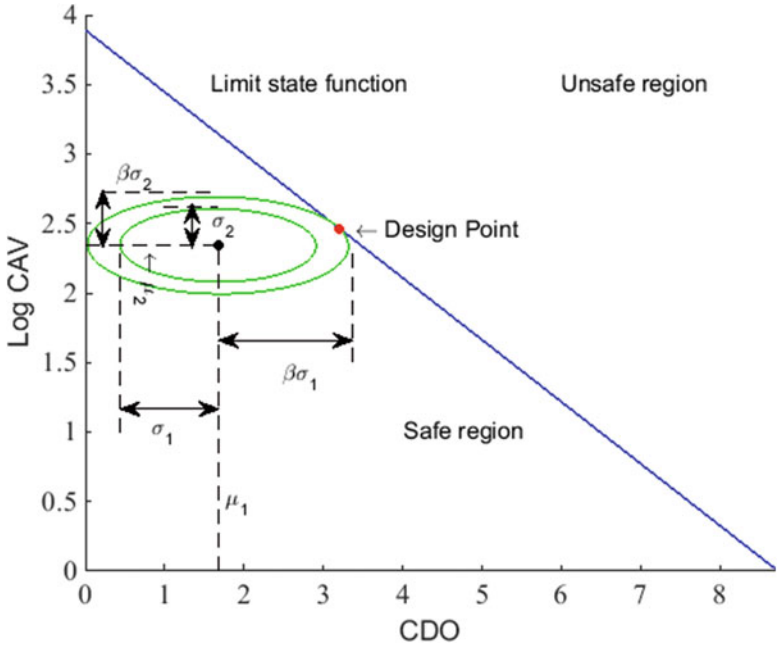


Fig. 6.10 1– σ ellipse, critical ellipse and design point in state space

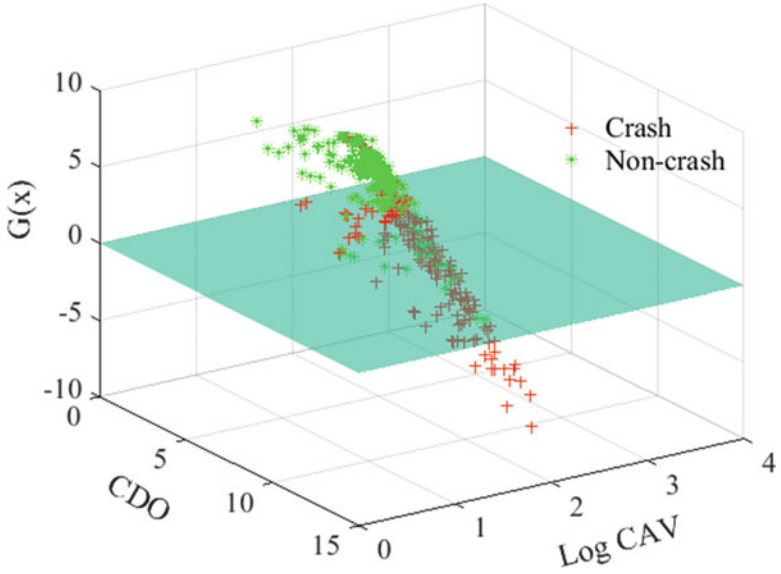


Fig. 6.11 Illustration of the traffic crash risk prediction results

Table 6.5 The traffic crash risk evaluation accuracy for validation data

Field data	Classification results		Total
	Crash	Non-crash	
Crash	104	28	132
Non-crash	42	275	317
Accuracy	78.79%	86.75%	84.41%

means that the traffic crash is likely to happen in a probability. Otherwise, $G(x) > 0$, the traffic state is in a safe region. Table 6.5 lists classification numeral results. It could been seen that the traffic reliability model could classify the validation data set with an overall accuracy rate of 84.41%, for crash cases with an accuracy rate of 78.79%, and for non-crash cases with an accuracy rate of 86.75%, respectively. The results, on one hand, show that utilizing reliability model to estimate highway traffic crash risk is feasible and the accuracy is in an acceptable range, on the other hand, real-time traffic crash risk evaluation provides the foundation for the traffic crash risk prediction.

References

1. J. Pohjalainen, O. Räsänen, S. Kadioglu, Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. Comput. Speech Lang. **29**, 145–171 (2015)
2. K. Pearson, On lines and planes of closest fit to systems in space. Phios. Mag. **2**, 559–573 (1901)
3. C. Xu, P. Liu, W. Wang, Z. Li, Evaluation of the impacts of traffic states on crash risk on freeways. Accid. Anal. Prev. **47**, 162–171 (2012)
4. R. Yu, M. Abdel-Aty, Utilizing support vector machine in real-time crash risk evaluation. Accid. Anal. Prev. **51**, 252–259 (2013)
5. K. Polat, S. Güneş, A novel approach to estimation of E. coli promoter gene sequences: Combining feature selection and least square support vector machine (FS_LSSVM). Appl. Math. Comput. **190**, 1574–1582 (2007)
6. H.B. Basaga, A. Bayraktar, I. Kaymaz, An improved response surface method for reliability analysis of structures. Struct. Eng. Mech. **42**(2), 175–189 (2012)
7. A.M. Hasofer, N.C. Lind, Exact and invariant second-moment code format. J. Eng. Mech. **100** (1), 111–121 (1974)
8. A.D. Kiureghian, H.Z. Lin, S.J. Hwang, Second-order reliability approximations. J. Eng. Mech. **113**(8), 1208–1225 (1987)
9. B.K. Low, W.H. Tang, Reliability analysis using object-oriented constrained optimization. Struct. Saf. **26**, 69–89 (2004)
10. R. Yu, Q. Shi, M. Abdel-Aty. Feasibility of incorporating reliability analysis in traffic safety investigation, Transportation Research Record: Journal of Transportation Research Board, No. 2386, Transportation Research Board of the National Academies, Washington, DC., (2013), pp. 35–41
11. H. William, A.S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University England EPress, New Delhi, 1992)
12. U. Alibrandi, A.M. Alani, G. Ricciardi, A new sampling strategy for SVM-based response surface for structural reliability analysis. Probab. Eng. Mech. **41**, 1–12 (2015)