Applied Intelligence (2022) 52:16555–16571
https://doi.org/10.1007/s10489-021-03114-7

Intelligent identification for vertical track irregularity based on multi-level evidential reasoning rule model

Zhenjie Zhang[1,2] · Xiaobin Xu[1,2] · Xuelin Zhang[2] · Xiaojian Xu[1,2] · Zifa Ye[2] · Guodong Wang[3] · Schahram Dustdar[4]

Accepted: 13 December 2021 / Published online: 24 March 2022

Abstract

Vertical track irregularity is one of the most significant indicators to evaluate track health. Accurate identification of vertical track irregularity is beneficial to achieve precise maintenance of the track and thus avoid accidents. However, the continuous variation of the track irregularity and the imbalance of the abnormal/normal data samples make it difficult to guarantee the accuracy of identification models. Therefore, by considering the interaction between train and track, a multi-level evidential reasoning (M-ER) rule model is proposed to build the nonlinear causal relationship of vibration signals and vertical track irregularity. In the modeling process of M-ER, the referential evidence matrix (REM) and fusion parameters (i.e., reliability factors and importance weights) are determined and optimized. In the model, the reliability factor of evidence is determined through trend analysis, while the importance weights of evidence and REM are optimized by sequential quadratic programming (SQP). In the inference process of M-ER, sample expansion strategy and two-level evidence fusion mechanism are designed. Specifically, in the first level, samples on each vibration signal are fused with their nearest neighboring historical samples obtained by *K*-Nearest Neighbor(*K*-NN) method. In the second-level, the results generated in the first-level are integrated by ER rule. We evaluate the M-ER rule model with an actual data set from China railway. The experimental results show that the model can identify the vertical track irregularity more accurately compared with the single-level ER rule model and other typical machine learning based models.

Keywords Vertical track irregularity · Evidential reasoning rule · Multi-level · Sample imbalance · Identification model

1 Introduction

Railway track is the most important infrastructure to carry trains in the railway transportation network. Due to frequent actions of trains with heavy loads, differential sub-grade settlement and harsh environment, geometric deformation of tracks often occurs [1]. It has become one of the most significant potential risks which influence the normal operation of trains. Vertical track irregularity is an indicator to measure the vertical concave and convex degree between the extending direction of track surface and the datum plane, reflecting the health condition of tracks [2]. It can cause abnormal vibration of trains and reduce ride quality [1]. What's worse, it may destroy the track structure and train parts, and thus induce accidents. To improve running safety and ride comfort, vertical track irregularity should be identified with proper condition monitoring methods [3, 4].

Currently, the track inspection vehicles (TIVs) are used to detect the vertical track irregularity. Nevertheless, the TIVs are expensive and unable to achieve online monitoring. Therefore, many researches have been conducted on in-service track irregularity identification approaches based on the vibration signals of car body, bogie and axle box. However, due to the noise in observation environment, the change of driving condition and the delay effect of vibration signals in comparison to displacement, the corresponding relationship usually

✉ Xiaobin Xu
xuxiaobin1980@163.com

1 China-Austria Belt and Road Joint Laboratory on Artificial Intelligence and Advanced Manufacturing, Hangzhou Dianzi University, Hangzhou, Zhejiang, China

2 School of Automation (School of Artificial Intelligence), Hangzhou Dianzi University, Hangzhou, Zhejiang, China

3 Sino-Austria Research Institute for Intelligent Industries (Nanjing) Co., Ltd, Gupinggang Rd 4, Gulou District, Nanjing, China

4 Distributed Systems Group, TU Wien, Vienna, Austria

presents nonlinearity and uncertainty. Moreover, because of the routine inspection and maintenance, the abnormal points (samples) on the entire test railway are far less than the normal points (samples). How to precisely establish the complex relationship between vibration signals and track irregularity in the uncertain environment is a challenging problem.

Fortunately, evidential reasoning (ER) rule method provides an effective mechanism for nonlinear relationship modeling and inference to process uncertain information [5]. It has been successfully applied in fault diagnosis [6], medical diagnosis [7], environmental protection [8] and so on. Actually, in our previous work [9], the ER rule model was firstly introduced in the identification of vertical track irregularity. However, it should be noted that samples on severe track irregularity are difficult to be acquired. How to use these limit samples sufficiently and how to avoid the effect of small samples on final identification result are two key issues directly influencing the irregularity identification result.

Therefore, a multi-level ER (M-ER) rule model is proposed in this paper. In the modeling process of M-ER, the referential evidence matrix (REM) and fusion parameters (i.e., reliability factors and importance weights) are determined and optimized. In the model, the reliability factor of evidence is determined through trend analysis, while the importance weight of evidence and REM are optimized by sequential quadratic programming (SQP). In the inference process of M-ER, sample expansion strategy and two-level evidence fusion mechanism are designed. Specifically, in the first-level, samples on each vibration signal are fused with their nearest neighboring historical samples obtained by *K*-NN method. In the second-level, the results generated in the first-level are integrated by ER rule. The main contributions of this paper are as follows:

1) The M-ER rule model is a data-driven model which do not need to make any hypothesis between irregularity amplitude (level) and vibration data.
2) Based on the likelihood function normalization method and two-level evidence fusion mechanism, the small samples problem in vertical track irregularity identification can be solved. Thus higher identification rate can be achieved with fewer samples.
3) By involving more relevant historical samples with *K*-NN method in the fusion process, more useful information can be added in irregularity identification. With this method, the problem on insufficient data samples can be solved and the identification accuracy can increase as well.

The rest of the paper is organized as follows. Section 2 discusses the related works. Section 3 briefly introduces the ER rule theory. The multi-level ER rule model for track irregularity identification is developed in Section 4. In Section 5, the performance of multi-level ER rule model is verified and compared with other identification models. Section 6 gives the conclusions.

2 Related works

The TIVs are widely used to monitor the condition of railway infrastructure, such as GJ-4 and GJ-5 TIVs in China [9–11]. Vibration signals and displacement signals reflecting the interaction between train and track are acquired by accelerator and displacement sensors, and then the geometric deformation of track can be identified by using inertial reference (IR) method. In the IR method, the track irregularity is the sum of the inertial displacement of car body, the relative displacement between car body and axle box as well as the inertial displacement on bogie [12]. It should be adjusted according to the car body angle measured by clinometer and gyroscope to increase the identification accuracy. Although the TIVs have high measurement accuracy, it needs many additional sensors, especially the expensive clinometer and gyroscope. Meanwhile, TIVs should be modified and strengthened to meet the requirement of precise instrument installation. Due to the high cost, the number of TIVs applied in health condition monitoring of railway network is significantly insufficient. Moreover, the detection cycle for the same track is long and the real-time monitoring is hard to realize.

Compared with TIVs, the online condition monitoring methods by using low-cost sensors mounted on the in-service trains can effectively increase inspection frequency and reduce measurement cost.

Measuring irregularity based on double integration of axle box acceleration (ABA) was most popular. For example, it was used in the measurement of the vertical track irregularity in the RAIDARSS system, which was a condition monitoring device installed in the N700 train sets in Japan [13, 14]. Real et al. proposed an identification method of track irregularities using the inverse Fourier transform technique based on the measured accelerations of axle boxes. Later, they developed inspection systems to detect vertical and lateral track geometry defects based on axle-box accelerations registered from in-service trains [15, 16]. To detect the light squats, Li et al. made some improvements in the traditional ABA method, including enhancement of ABA instrumentation and signal processing [17]. Sun et al. proposed an on-board detection system for longitudinal irregularity via axle box acceleration signal [18]. However, the numerical error induced by the double integration would be inevitable [12].

To reduce the numerical error, signal pre-processing approaches such as Fourier transform [14, 19, 20], wavelet analysis [3, 14] and Kalman filter [4, 12] were always used. For example, Tsunashima et al. [14] combined short-time Fourier transform (STFT) with wavelet based multi-resolution analysis to extract the frequency-domain features from the vibration signals [14]. Bhardwaj et al. designed method that ensemble averages the individual FFTs (EA-FFT) from the approximately equal length and position aligned inertial signals to enhance the clarity of the underlying pattern [19]. Xiao et al. presented a Kalman filter algorithm to identify the track irregularities of railway bridges using vehicle dynamic responses considering the VB interactions in real-time [12]. Meanwhile, mixed filtering methods were also applied to improve the identification accuracy. Lee et al. presented a mixed filtering approach which consists of a Kalman filter for displacement estimation, bandpass filter for waveband classification and compensation filter for amplitude and phase compensation [4]. Wei et al. applied the DC filter and a low pass filter to process the vibrate signals [21]. However, most of these signal processing methods were limited by such assumptions as linear model and Markovian process. Furthermore, the signals were usually non-stationary and signal processing results were not normally satisfied with the ordinary way [22].

In short, the above studies have the following limitations. Firstly, the trains are always operating under non-stationary conditions [23], in which the sensors are easily disturbed by external environment or human factors, increasing the uncertainty of the acquired signals. However, most of the existing research ignore the uncertainty in track irregularity identification. Secondly, these methods can only detect the track irregularity degrees, but cannot estimate the specific amplitude of the irregularity displacement. Thirdly, the relationship between the vibration signals and the vertical track irregularity are assumed to be linear, while in fact it is non-linear. Although some machine learning methods such as support vector machine and neural network have been used to process the nonlinearity [24, 25], the identification accuracy cannot be guaranteed.

3 ER rule theory

In the ER rule theory, suppose $\Theta=\{h_1,h_2,...,h_N\}$ is the frame of discernment (FoD), consisting N mutually exclusive and collectively exhaustive hypotheses. The power set of Θ includes all subsets of Θ, represented by $P(\Theta)$ or 2^Θ. There are three main elements in ER rule theory which are the belief distribution of evidence, evidence reliability and evidence important weight. In the FoD, the belief distribution of one piece of evidence is shown as Eq. (1).

$$e_j = \left\{ (\theta, p_{\theta,j}) | \forall\theta \subseteq \Theta | \sum\nolimits_{\theta\in\Theta} p_{\theta,j} = 1 \right\} \tag{1}$$

where $(\theta, p_{\theta,j})$ represents that the belief degree of evidence e_j supporting to the proposition θ is $p_{\theta,j}$, θ can be any element of 2^Θ.

Evidence reliability r_j indicates that how the evidence e_j provides correct assessment or solution for a given problem. Evidence important weight w_j reflects the relative importance of evidence e_j compared to other evidence. r_j is the inherent attribute of evidence influenced by information source or evidence acquisition method, while w_j is determined by other evidence to be fused and the subjective experience of decision maker. The belief distribution of evidence e_j can be modified by r_j and w_j, which is expressed as Eq. (2).

$$m_j = \left\{ \left(\theta, \tilde{m}_{\theta,j}\right) | \forall\theta \subseteq \Theta | \left(P(\Theta), \tilde{m}_{P(\Theta),j}\right) \right\} \tag{2}$$

where $\tilde{m}_{\theta,j}$ is the supporting degree of e_j to hypothesis θ considering r_j and w_j. $\tilde{m}_{\theta,j}$ is defined as Eq. (3).

$$\tilde{m}_{\theta,j} = \begin{cases} 0 & \theta = \emptyset \\ c_{rw,j} m_{\theta,j} & \theta \subseteq \Theta, \theta \neq \emptyset \\ c_{rw,j}\left(1 - r_j\right) & \theta = P(\Theta) \end{cases} \tag{3}$$

In Eq. (3), $m_{\theta,j} = w_j p_{\theta,j}$ is the basic probability mass, $c_{rw,j} = 1/(1 + w_j - r_j)$ is normalization factor to ensure $\sum_{\theta\in\Theta} \tilde{m}_{\theta,j} + \tilde{m}_{P(\Theta),j} = 1$ when $\sum_{\theta\in\Theta} p_{\theta,j} = 1$, and (1-$r_j$) denotes the unreliability of evidence e_j.

Joint supporting degree of two independent pieces of evidence e_1 and e_2 to hypothesis θ is $p_{\theta,e(2)}$ which can be calculated by Eq. (4).

$$p_{\theta,e(2)} = \begin{cases} 0 & \theta = \emptyset \\ \frac{\hat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta} \hat{m}_{D,e(2)}} & \theta \subseteq \Theta, \theta \neq \emptyset \end{cases}$$
$$\hat{m}_{\theta,e(2)} = \left[\left(1 - r_2\right)m_{\theta,1} + \left(1 - r_1\right)m_{\theta,2}\right] + \sum\nolimits_{B\cap C=\theta} m_{B,1} m_{C,2} \forall\theta \subseteq \Theta \tag{4}$$

When L independent pieces of evidence are fused, recurrent ER rule can be used to generate the supporting degree of L pieces of evidence to hypothesis θ, as shown in Eq. (5).

$$p_{\theta,e(n)} = \frac{\hat{m}_{\theta,e(n)}}{\sum_{A\subseteq\Theta} \hat{m}_{A,e(n)}} \tag{5a}$$

$$\hat{m}_{\theta,e(j)} = \left[\left(1 - r_j\right)m_{\theta,e(j-1)} + m_{P(\Theta),e(j-1)} m_{\theta,j}\right] + \sum_{B\cap C=\theta} m_{B,e(j-1)} m_{C,j} \tag{5b}$$

$$m_{\theta,e(j-1)} = \left[m_1 \oplus \cdots \oplus m_{j-1}\right](\theta) = \frac{\hat{m}_{\theta,e(j-1)}}{\sum_{D\subseteq\Theta} \hat{m}_{D,e(j-1)} + \hat{m}_{P(\Theta),e(j-1)}} \tag{5c}$$

$$\hat{m}_{P(\Theta),e(j-1)} = \left(1 - r_{j-1}\right) m_{P(\Theta),e(j-2)} \tag{5d}$$

4 Multi-level ER rule model for vertical track irregularity identification

In this section, a multi-level ER (M-ER) rule model is proposed to estimate the track irregularity displacement amplitude as well as the irregular level, as shown in Fig. 1.

The M-ER rule model mainly contains two parts, namely modeling process and inference process. In the modeling process, the referential evidence matrix (REM) and fusion parameters (i.e., reliability factors and importance weights) are determined and optimized. In the inference process, sample expansion strategy and two-level evidence fusion mechanism are designed. Specifically, for a new sample $(a_1(t), a_2(t), a_3(t))$ in time domain, three steps are conducted to determine the track irregularity in the M-ER rule model as shown in Fig. 3. Firstly, the sample $(a_1(t), a_2(t), a_3(t))$ is converted into $(f_1(t), f_2(t), f_3(t))$ in frequency domain via short-time Fourier transform (STFT). Then, the K-nearest

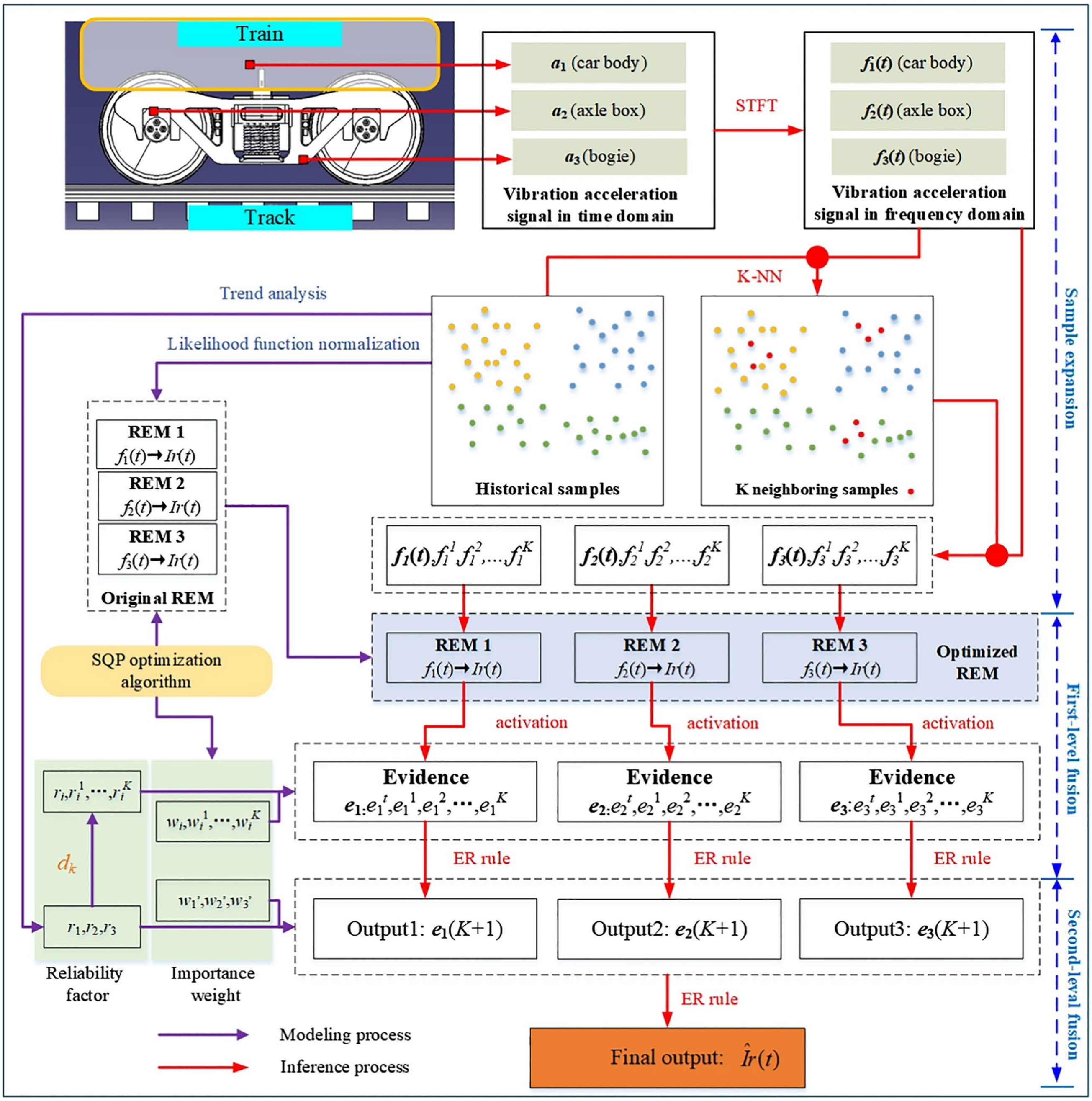


Fig. 1 The vertical track irregularity identification model based on multi-level ER rule method

neighbor (K-NN) method is used to select the K nearest neighbors ($f_1^1, f_1^2, \ldots, f_1^K$), ($f_2^1, f_2^2, \ldots, f_2^K$) and ($f_3^1, f_3^2, \ldots, f_3^K$) for ($f_1(t), f_2(t), f_3(t)$) in the historical datasets. Secondly, the new combined samples ($f_1(t), f_1^1, f_1^2, \ldots, f_1^K$), ($f_2(t), f_2^1, f_2^2, \ldots, f_2^K$) and ($f_3(t), f_3^1, f_3^2, \ldots, f_3^K$) are input to REM1 for f_1, REM2 for f_2, and REM3 for f_3 respectively to activate the relevant evidence ($e_1^t, e_1^1, e_1^2, \ldots, e_1^K$), ($e_2^t, e_2^1, e_2^2, \ldots, e_2^K$) and ($e_3^t, e_3^1, e_3^2, \ldots, e_3^K$). All pieces of evidence on f_1, f_2 and f_3 are fused by ER rule to generate three pieces of evidence $e_1(K+1)$, $e_2(K+1)$ and $e_3(K+1)$. Thirdly, the three pieces of evidence from three information sources are integrated in the second-level fusion to obtain the final result.

4.1 Modeling process of M-ER rule model

4.1.1 REM construction

Likelihood function normalization is applied to construct the REM by using historical samples. In the training dataset, each sample is expressed as $S(t) = \{[f_1(t), f_2(t), f_3(t), Ir(t)] | f_i(t) \in S_{fi}, Ir(t) \in S_{Ir}, t = 1, 2, \ldots, T_s, i = 1, 2, 3\}$, where T_s is the number of samples, S_{fi} and S_{Ir} represent the range of f_i and Ir respectively. To develop the mapping relationship between f_i and Ir, the relationships between $A_i = \{A_j^i | j = 1, \ldots, J_i\}$ and $D = \{D_n | n = 1, \ldots, N\}$ are built, where A_i is the referential point set of f_i, and D is the referential point set of Ir. All referential points of f_i and Ir are generally determined by experts initially, and then optimized by the training dataset.

For each sample $S(t)$, f_i can be transformed into a similarity distribution $S_I(f_i(t)) = \left\{ \left(A_j^i, \alpha_{i,j} \right) | j = 1 | \ldots | J_i | i = 1 | 2 \right\}$, where $\alpha_{i,j}$ is the similarity between $f_i(t)$ and its jth referential point A_j^i, and J_i is the number of referential points for the ith input. The similarity distribution can be generated by piecewise function as Eq. (6).

$$\begin{cases} \alpha_{i,j} = \frac{A_{j+1}^i - f_i(t)}{A_{j+1}^i - A_j^i}, \alpha_{i,j+1} = 1 - \alpha_{i,j} A_j^i \le f_i(t) \le A_{j+1}^i \\ \quad \alpha_{i,j\prime} = 0 j' = 1, \ldots, J_i, j' \ne j, j+1 \end{cases} \tag{6}$$

Similarly, Ir can be transformed into similarity distribution $S_O(Ir(t)) = \{(D_n, \gamma_n) | n = 1, 2, ..., N\}$ as well, and the similarity distribution can be acquired by piecewise function as Eq. (7).

$$\begin{cases} \gamma_n = \frac{D_{n+1} - Ir(t)}{D_{n+1} - D_n}, \gamma_{n+1} = 1 - \gamma_n D_n \le Ir(t) \le D_{n+1} \\ \quad \gamma_{n\prime} = 0 n' = 1, \ldots, N, n' \ne n, n+1 \end{cases} \tag{7}$$

where γ_n is the similarity between $Ir(t)$ and its nth referential point D_n, N is the number of referential points for the output.

According to Eqs. (6) and (7), each sample pair ($f_i(t)$,$Ir(t)$) in the training dataset can be represented by integrated similarity distribution($\alpha_{i,j}\gamma_n, \alpha_{i,j+1}\gamma_n, \alpha_{i,j}\gamma_{n+1}, \alpha_{i,j+1}\gamma_{n+1}$). Obviously, $\alpha_{i,j}\gamma_n + \alpha_{i,j+1}\gamma_n + \alpha_{i,j}\gamma_{n+1} + \alpha_{i,j+1}\gamma_{n+1} = 1$, where$\alpha_{i,j}\gamma_n$indicates the integrated similarity degree of $f_i(t)$ matching input referential point A_j^i while $Ir(t)$ matching output referential point D_n simultaneously. By calculating the integrated similarity degree of all sample pairs in training dataset S, the casting results can be obtained as shown in Table 1. It reflects the relationship between each input referential point and output referential point.

In Table 1, $a_{n,j}$ is the sum of the integrated similarity degree of all sample pairs ($f_i(t)$,$Ir(t)$) matching the input referential point A_j^i and meanwhile matching the output referential point $Ir(t)$. $\delta_n = \sum_{j=1}^{J_i} a_{n,j}$represents the sum of the integrated similarity degree of all sample pairs that their output $Ir(t)$ matches D_n, while$\eta_j = \sum_{n=1}^{N} a_{n,j}$represents the sum of the integrated similarity degree of all sample pairs that their input $f_i(t)$ matches A_j^i. It can be clearly found that $\sum_{n=1}^{N} \delta_n = \sum_{j=1}^{J_i} \eta_j = T_s$.

Based on Table 1, the likelihood function, denoted as $c_{n,j}$, can be calculated as follows.

$$c_{n,j} = p\left(A_j^i | D_n \right) = \frac{a_{n,j}}{\delta_n} \tag{8}$$

Then a piece of evidence $e_j^i = \left\{ \left(D_n, \beta_{n,j}^i \right) | n = 1 | \ldots | N \right\}$ corresponding to A_j^i can be defined as shown Table 2.

In Table 2, the evidence e_j^i can be simply represented as $e_j^i = \left[\beta_{1,j}^i, \beta_{2,j}^i, \ldots, \beta_{N,j}^i \right]$, where $\beta_{n,j}^i$is the belief degree of evi-

Table 1 Casting results of sample pair (fi(t), Ir(t)) in the training dataset

Ir \ f_i	A_1^i	…	A_j^i	…	$A_{J_i}^i$	Total
D_1	$a_{1,1}$	…	$a_{1,j}$	…	a_{1,J_i}	δ_1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
D_n	$a_{n,1}$	…	$a_{n,j}$	…	a_{n,J_i}	δ_n
⋮	⋮	⋮	⋮	⋮	⋮	⋮
D_N	$a_{N,1}$	…	$a_{N,j}$	…	a_{N,J_i}	δ_N
Total	η_1	…	η_j	…	η_{J_i}	T_s

dence $e_j^{\,i}$. It represents the probability that $Ir = D_n$ given that $f_i = A_j^{\,i}$ and can be calculated by normalization of likelihood function $c_{n,j}$, as shown in Eq. (9).

$$\beta_{n,j}^{i} = \frac{c_{n,j}}{\sum_{k=1}^{N} c_{k,j}} \tag{9}$$

Apparently, $\sum_{n=1}^{N} \beta_{n,j}^{i} = 1$.

4.1.2 Determination of reliability factor using trend analysis

In the first-level ER rule model, the reliability factor of the evidence $e_i^{\,t}$ corresponding to the newly acquired sample is related to the reliability factors of evidence $\{e_i^{\,1}, e_i^{\,2}, e_i^{\,3}\}$ corresponding to historical samples, but they have some differences. Specifically, the reliability factor r_i of $e_i^{\,t}$ describes the ability of the ith information source correctly evaluating the vertical track irregularity. The higher reliability of the information source, the more sensitive of the irregularity variation. It means that the higher irregularity variation corresponds to a higher variation of frequency characteristic $f_i (i = 1, 2, 3)$, and vice versa. The relative changes of $f_i(t)$ and $Ir(t)$ are defined as Eq. (10).

$$Mf_i(t) = \frac{f_i(t) - \min\left(f_i(t)\right)}{\max\left(f_i(t)\right) - \min\left(f_i(t)\right)}, t \in \left\{1, 2, \ldots, T_S\right\} \tag{10a}$$

$$MIr(t) = \frac{Ir(t) - \min(Ir(t))}{\max(Ir(t)) - \min(Ir(t))}, t \in \left\{1, 2, \ldots, T_S\right\} \tag{10b}$$

And then the capability of f_i reflecting Ir variation is defined as Eq. (11). Apparently, the smaller cf_i indicates that f_i can reflect the Ir variation more correctly.

$$cf_i = \sum_{t=1}^{T_s} \left| MIr(t) - Mf_i(t) \right| \mathrm{af_k} = \sum_{\mathrm{t=1}}^{\mathrm{T}} \left| \mathrm{CIr(t) - Cf_k(t)} \right| \tag{11}$$

Based on the above analysis, the reliability factor of information source f_i can be calculated as Eq. (12), and it can be seen from the equation that f_i is the most reliable (i.e., $r_i = 1$) when $\min\limits_{k,k\in\{1,2,3\}} \left(cf_k\right) = cf_i$.

$$r_i = \frac{\min\limits_{k,k\in\{1,2,3\}} \left(cf_k\right)}{cf_i} \tag{12}$$

Additionally, evidence $e_i^{\,1}$, $e_i^{\,2}$ and $e_i^{\,3}$ generated by historical samples are also from information source f_i and they are selected by comparing their similarity with $e_i^{\,t}$. As a result, the reliability factors of $\{e_i^{\,1}, e_i^{\,2}, e_i^{\,3}\}$ should be determined by considering both the characteristic of information source f_i and their relative reliability with $e_i^{\,t}$. The Euclidean distance d_k between the historical sample and the newly acquired sample is used to determine the reliability factors of evidence $\{e_i^{\,1}, e_i^{\,2}, e_i^{\,3}\}$, as shown in Eq. (13).

$$r_i^k = \left(1 - d_k\right) r_i, i = 1, 2; k = 1, 2, 3 \tag{13}$$

It can be found that the evidence with smaller Euclidean distance has a higher reliability factor.

4.1.3 Model optimization

Generally, the parameters of the multi-level ER rule model are determined by domain knowledge and may be inaccurate. To improve the accuracy of the non-linear relationship between the input f_i $(i = 1,2,3)$ and output Ir, the parameters should be optimized by using historical samples S. These parameters include input referential points $A_i = \{A_j^{\,i} | j = 1,\ldots,J_i\}$, output referential points $D = \{D_n | n = 1,\ldots,N\}$, importance weights of all pieces of evidence in the multi-level models $W = \{w_i, w_i', w_i^{\,k} | i = 1,2,3;\ k = 1,2,3\}$. Here, mean square error (MSE) is used as the objective function of the optimization model as shown in Eq. (14a).

Table 2 REM of input f_i

Ir \ f_i	e_1^i / A_1^i	…	e_j^i / A_j^i	…	$e_{J_i}^i$ / $A_{J_i}^i$
D_1	$\beta_{1,1}^i$	…	$\beta_{1,j}^i$	…	β_{1,J_i}^i
⋮	⋮	⋮	⋮	⋮	⋮
D_n	$\beta_{n,1}^i$	…	$\beta_{n,j}^i$	…	β_{n,J_i}^i
⋮	⋮	⋮	⋮	⋮	⋮
D_N	$\beta_{N,1}^i$	…	$\beta_{N,j}^i$	…	β_{N,J_i}^i

Table 3 Vertical track irregularity levels

(160 km/h ~ 200 km/h)	Acceptable	Uncomfortable	Occasional repair	Speed limiting
Levels	I	II	III	IV
Range of Ir (mm)	$0 \le Ir \le 5$	$5 < Ir \le 8$	$8 < Ir \le 12$	$12 < Ir$

$$\min_{\boldsymbol{P}} \xi(P) = \frac{1}{T_s}\sum_{t=1}^{T_s}\left(\hat{I}r(t) - Ir(t)\right)^2 \tag{14a}$$

$$\begin{gathered} 0 \le w_i, w_i', w_i^{\,k} \le 1, i = 1, 2; k = 1, 2, 3 \\ A_{j-1}^i < A_j^i < A_{j+1}^i, j = 2, \dots, J_i - 1 \\ D_2 < D_3 < \cdots < D_{N-1} \end{gathered} \tag{14b}$$

where $P = \{A_i, D, W \mid i = 1,2,3; j = 2,\ldots,J_i\text{-}1; n = 2,\ldots,N\text{-}1\}$ is the set of optimized parameters, and Eq. (14b) is the constrains that the parameters should meet. w_i' and $w_i^{\,k}$ are the importance weight of the evidence activated by input sample $f_i(t)$ and its kth nearest neighboring samples which are used in first-level model, and w_i is the importance weight of the evidence for f_i in the second-level model. Except for the parameters in P, the other parameters are fixed boundary values. The parameters $\mathrm{D}_1, D_N, A_1^1, A_{J_1}^1, A_1^2, A_{J_2}^2, A_1^3, A_{J_3}^3$ are set to be $\min_{t,t\in S_{Ir}}(Ir(t))$, $\max_{t,t\in S_{Ir}}(Ir(t))$, $\min_{t,t\in S_{f1}}\left(f_1(t)\right)$, $\max_{t,t\in S_{f1}}\left(f_1(t)\right)$, $\min_{t,t\in S_{f2}}\left(f_2(t)\right)$, $\max_{t,t\in S_{f2}}\left(f_2(t)\right)$, $\min_{t,t\in S_{f3}}\left(f_3(t)\right)$, $\max_{t,t\in S_{f3}}\left(f_3(t)\right)$ respectively. The model described by Eq. (14) is optimized by SQP algorithm. With the variation of the parameters in P, REM shown in Table 3 is optimized as well.

4.2 Inference process of M-ER rule model

4.2.1 Sample expansion based on K-NN method

The aim of the multi-level ER rule model as shown in Fig. 1 is to make the full use of the historical samples in the fusion process. K-NN method is applied to find the samples from the training dataset S which are similar to the newly acquired sample, and then the evidence activated by all these similar samples are fused together. In the process of seeking for the similar samples, the K nearest neighboring samples to the newly acquired sample are selected by calculating their Euclidean distance.

To avoid the influence of different measurement units and magnitudes on input features and ensure the information offered by each input feature can be fully used, all of the input data should be normalized before selecting the nearest neighboring samples by using K-NN method. In vertical track irregularity identification, the value ranges of input feature f_2 and f_3 are relatively smaller than that of feature f_1. Therefore, f_i is normalized via *Min-Max* method to avoid the roles of f_2 and f_3 being covered by f_1. The normalization of $f_i(t)$ can be conducted as Eq. (15).

$$Mf_i(t) = \frac{f_i(t) - \operatorname{Min}_t\left(f_i(t)\right)}{\operatorname{Max}_t\left(f_i(t)\right) - \operatorname{Min}_t\left(f_i(t)\right)}, t \in \{1, 2, \dots, T\} \tag{15}$$

After normalization, K-NN method is used to find out the K nearest neighboring samples from the historical data samples. For a newly acquired sample $F'(t) = (f_1(t), f_2(t), f_3(t))$ and each historical sample $F(\tau) = (f_1(\tau), f_2(\tau), f_3(\tau))$, their Euclidean distance can be obtained by Eq. (16).

$$d_\tau\left(F', F\right) = \sqrt{\sum_{i=1}^{3}\left(Mf_i(t) - Mf_i(\tau)\right)^2} \tag{16}$$

where $\tau = 1,2,\ldots,T_S$, and T_S is the number of samples in historical dataset. Sorting d_τ form small to large and choosing the first K samples, denoted as $F_k = \left\{\left(f_1^k, f_2^k, f_3^k\right) \mid k = 1|2| \dots |K\right\}$. Apparently, the calculation amount will increase with the rising value of K, and therefore $K = 3$ in this paper.

4.2.2 Evidence fusion based on multi-level ER rule

For the ith input feature, the value of $f_i(t)$ is located into $\left[A_j^i, A_{j+1}^i\right]$. Consequently, it will active two adjacent

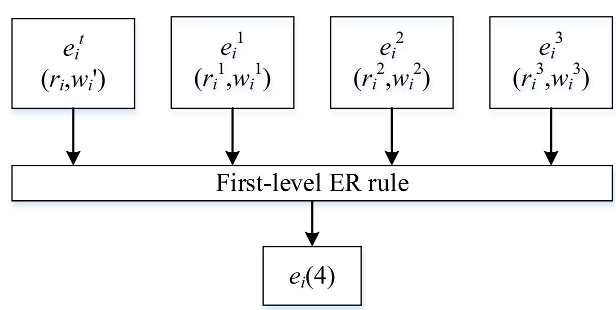


Fig. 2 Evidence fusion in the first-level ER rule model for the ith information source

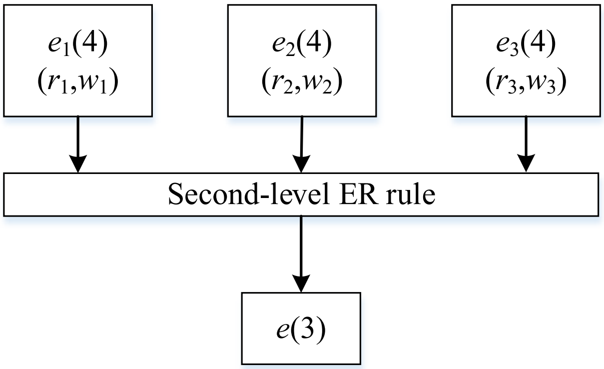


Fig. 3 Evidence fusion in the second-level ER rule model for three information sources

pieces of evidence e_j^i and e_{j+1}^i described in Table 2. The final evidence activated by $f_i(t)$ is the weighted sum of e_j^i and e_{j+1}^i. The activated evidence can be represented as $e_i = \{(D_n, p_{n,i}), n = 1, \ldots, N\}$, where $p_{n,i}$ is the belief degree of that the value of $Ir(t)$ is D_n when e_j^i and e_{j+1}^i are activated by $f_i(t)$, and it can be calculated by Eq. (17).

$$p_{n,i} = \alpha_{i,j}\beta_{n,j}^i + \alpha_{i,j+1}\beta_{n,j+1}^i \tag{17}$$

According to Eq. (17), three pieces of evidence $e_1{}^t$, $e_2{}^t$ and $e_3{}^t$ are activated by $F'(t) = (f_1(t), f_2(t), f_3(t))$ and the evidence activated by the three nearest neighboring samples $\{e_1{}^1, e_1{}^2, e_1{}^3\}$, $\{e_2{}^1, e_2{}^2, e_2{}^3\}$ and $\{e_3{}^1, e_3{}^2, e_3{}^3\}$ can be acquired. After that, for each information source, $e_i{}^t$ is fused with $\{e_i{}^1, e_i{}^2, e_i{}^3\}$ $(i = 1,2,3)$ by using Eq. (5), generating the fused result of the first-level ER rule model, denote as $e_i(4) = \left\{\left(D_n, p_{n,e_i(4)}\right) | n = 1| \ldots |N| i = 1|2|3\right\}$, as shown in Fig. 2.

The fused result of the first-level ER rule model will be used as an integrated piece of evidence in the second-level ER rule model. The obtained three pieces of evidence $e_1(4)$, $e_2(4)$ and $e_3(4)$ from three different information sources are fused to generate the final result, denoted as $e(3) = \{(D_n, p_{n,e(3)}) | n = 1| \ldots | N\}$, as illustrated in Fig. 3.

The estimated vertical track irregularity displacement $\hat{I}r(t)$ can be obtained with Eq. (18).

$$\hat{I}r(t) = \sum\nolimits_{n=1}^{N} D_n p_{n,e(3)} \tag{18}$$

Then the vertical track irregularity level can also be obtained according to Table 3.

As shown in Table 3, Ir is the absolute value of vertical track irregularity. According to the standard of China railway infrastructure maintenance [26], it can be graded into four levels. Specifically, level I represents the health condition of track is acceptable, and only routing maintenance should be conducted. When Ir is between 5 mm and 8 mm (i.e. level II), abnormal vibration of train can be detected reducing railway comfort, but it has slight influence on the normal operation of trains. If Ir is between 8 mm and 12 mm (i.e. level III), train safety will decrease and occasional maintenance must be carried out to avoid severe geometry deformation. When Ir is over 12 mm (i.e., level IV), track has been deformed seriously, the train should speed down and tracks need to be repaired urgently.

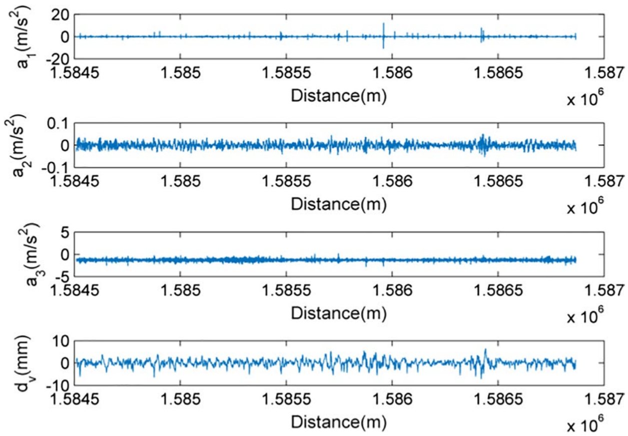


Fig. 4 Vibration signals in time domain a_1, a_2, a_3 and vertical track irregularity d_v

5 Experimental results and discussions

To verify the effectiveness and superiority of the irregularity identification approach based on the M-ER rule model, the experiment is conducted in this section.

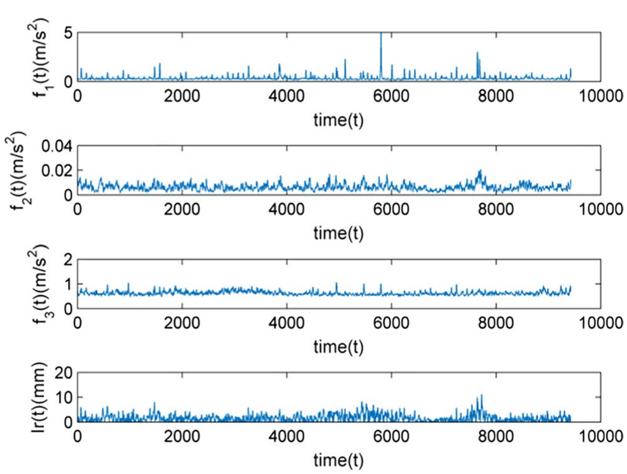


Fig. 5 Mean values of $f_1(t), f_2(t), f_3(t)$ and absolute value of $Ir(t)$

5.1 Description of dataset

In this paper, the train vibration data and the corresponding irregularity data collected by TIVs which work from 1584.5103 km to 1586.8673 km in Beijing-Guangzhou Line. In detail, sensors installed in the car body, axle box and bogie of GJ-4 collect the vibration acceleration signals in time domain which are represented by a_1, a_2 and a_3 respectively, and vertical track irregularity d_v is acquired by IR method, as shown in Fig. 4. Then the short-time Fourier transform (STFT) is conducted for a_1, a_2 and a_3, respectively denoted as $f_1(t)$, $f_2(t)$ and $f_3(t)$, and the absolute value of d_v denoted as $Ir(t)$, as shown in Fig. 5.

Compared with the time domain vibration signals in Fig. 4, the variation trends of $f_1(t)$, $f_2(t)$, $f_3(t)$ and $Ir(t)$ are more consistent as shown in Fig. 5. When $Ir(t)$ rises, $f_1(t)$, $f_2(t)$ and $f_3(t)$

Table 4 Details of dataset

Data source	1584.5103 km to 1586.8673 km in Beijing-Guangzhou Line		
Sampling interval	0.25m		
Sliding window length of STFT	5.25m		
Total number of samples (*T*)	9429	Number of samples on **level I**	9221
		Number of samples on **level II**	193
		Number of samples on **level III**	15
		Number of samples on **level IV**	0

Table 5 Initial casting result of sample pairs ($f_1(t)$, $Ir(t)$) in training dataset

Ir \ f_1		A_1^1 0	A_2^1 0.3	A_3^1 0.6	A_4^1 0.9	A_5^1 1.2	A_6^1 5	Total
D_1	0	62.1668	290.4692	41.6610	8.6979	5.1613	0.1863	408.3425
D_2	2	74.4521	324.4076	50.1912	14.0472	7.8935	2.4458	473.4375
D_3	4	23.6380	86.7704	11.4314	2.3383	4.0978	1.4416	129.7175
D_4	6	11.4995	44.7335	9.3129	5.0296	2.5662	0.7358	73.8775
D_5	8	1.4671	10.6221	3.0214	1.3066	0.4103	0	16.8275
D_6	10	0.0064	3.4685	1.1268	0.0434	0	0	4.6450
D_7	12	0	0.9850	0.1675	0	0	0	1.1525
Total		173.2299	761.4562	116.9122	31.4630	20.1292	4.8096	1108

Table 6 Initial casting result of sample pairs ($f_2(t)$, $Ir(t)$) in training dataset

Ir \ f_2		A_1^2 0	A_2^2 0.002	A_3^2 0.004	A_4^2 0.006	A_5^2 0.007	A_6^2 0.008	A_7^2 0.01	A_8^2 0.015	A_9^2 0.021	Total
D_1	0	208.4924	176.2621	21.9402	1.6478	0	0	0	0	0	408.3425
D_2	2	48.2550	150.7187	93.1313	86.2153	74.0069	19.6400	1.4703	0	0	473.4375
D_3	4	0	0	0	5.9117	32.9101	38.4227	36.8750	15.4589	0.1391	129.7175
D_4	6	0	0	0	0	0	0.1203	10.6754	57.4488	5.6330	73.8775
D_5	8	0	0	0	0	0	0	0	8.9656	7.8619	16.8275
D_6	10	0	0	0	0	0	0	0	2.2935	2.3515	4.6450
D_7	12	0	0	0	0	0	0	0	0.1658	0.9867	1.1525
Total		256.7474	326.9808	115.0715	93.7748	106.9170	58.1830	49.0207	84.3326	16.9723	1108

Table 7 Initial casting result of sample pairs ($f_3(t)$, $Ir(t)$) in training dataset

Ir \ f_3		A_1^3 0.4	A_2^3 0.55	A_3^3 0.60	A_4^3 0.70	A_5^3 0.75	A_6^3 1.1	Total
D_1	0	6.5042	109.6234	169.9034	78.6708	39.4515	4.1892	408.3425
D_2	2	9.0064	134.5631	185.2954	97.9683	39.6610	6.9432	473.4375
D_3	4	4.2297	37.6570	49.4082	23.0157	12.4565	2.9505	129.7175
D_4	6	1.5667	25.4212	24.8076	14.1358	6.5756	1.3706	73.8775
D_5	8	0.2001	9.3301	2.5579	3.8650	0.8569	0.0175	16.8275
D_6	10	0.1890	1.2183	0.0052	2.4207	0.8118	0	4.6450
D_7	12	0	0	0	0.9041	0.2484	0	1.1525
Total		21.6961	317.8130	431.9778	220.9805	100.0617	15.4710	1108

Table 8 Initial REM of input f_1

Ir \ f_1		e_1^1 A_1^1 0	e_2^1 A_2^1 0.3	e_3^1 A_3^1 0.6	e_4^1 A_4^1 0.9	e_5^1 A_5^1 1.2	e_6^1 A_6^1 5
D_1	0	0.2069	0.1451	0.1031	0.0951	0.1053	0.0171
D_2	2	0.2137	0.1397	0.1071	0.1324	0.1389	0.1935
D_3	4	0.2476	0.1364	0.0890	0.0804	0.2632	0.4163
D_4	6	0.2115	0.1235	0.1274	0.3038	0.2894	0.3731
D_5	8	0.1185	0.1287	0.1814	0.3465	0.2032	0
D_6	10	0.0019	0.1523	0.2451	0.0417	0	0
D_7	12	0	0.1743	0.1468	0	0	0

Table 9 Initial REM of input f_2

Ir \ f_2		e_1^2 A_1^2 0	e_2^2 A_2^2 0.002	e_3^2 A_3^2 0.004	e_4^2 A_4^2 0.006	e_5^2 A_5^2 0.007	e_6^2 A_6^2 0.008	e_7^2 A_7^2 0.01	e_8^2 A_8^2 0.015	e_9^2 A_9^2 0.021
D_1	0	0.8336	0.5755	0.2145	0.0174	0	0	0	0	0
D_2	2	0.1664	0.4245	0.7855	0.7859	0.3812	0.1223	0.0072	0	0
D_3	4	0	0	0	0.1967	0.6188	0.8729	0.6582	0.0576	0.0006
D_4	6	0	0	0	0	0	0.0048	0.3346	0.3762	0.0400
D_5	8	0	0	0	0	0	0	0	0.2577	0.2450
D_6	10	0	0	0	0	0	0	0	0.2388	0.2655
D_7	12	0	0	0	0	0	0	0	0.0696	0.4490

are increase accordingly, and vice versa. Therefore, the frequency domain data are used as the training and testing samples for the proposed multi-level model ($f_1(t)$, $f_2(t)$ and $f_3(t)$ as inputs and $Ir(t)$ as output). The details are shown in Table 4.

Table 10 Initial REM of input f_3

Ir \ f_3		e_1^3	e_2^3	e_3^3	e_4^3	e_5^3	e_6^3
		A_1^3	A_2^3	A_3^3	A_4^3	A_5^3	A_6^3
		0.4	0.55	0.60	0.70	0.75	1.1
D_1	0	0.1127	0.1340	0.2481	0.0836	0.1198	0.1525
D_2	2	0.1346	0.1418	0.2333	0.0898	0.1039	0.2180
D_3	4	0.2307	0.1449	0.2271	0.0770	0.1191	0.3382
D_4	6	0.1500	0.1717	0.2002	0.0831	0.1103	0.2758
D_5	8	0.0841	0.2767	0.0906	0.0997	0.0631	0.0155
D_6	10	0.2879	0.1309	0.0007	0.2262	0.2167	0
D_7	12	0	0	0	0.3405	0.2672	0

Table 11 The optimized importance weight of each piece of evidence

i	w_i	$w_i^{'}$	w_i^1	w_i^2	w_i^3
1	0.8617	0.6782	0.5686	0.3078	0.2772
2	0.4943	0.9656	0.8096	0.4382	0.3947
3	0.9148	0.7314	0.6132	0.3319	0.2990

5.2 The identification process for vertical track irregularity based on M-ER

5.2.1 Modeling of M-ER

In our experiment, 1108 samples are randomly selected to constitute the training dataset, denoted as $S=\{[f_1(t),f_2(t),f_3(t),Ir(t)]| f_i(t)\Pi\epsilon S_i, Ir(t)\epsilon S_{Ir}, t=1,2,\ldots,1108, i=1,2,3\}$, where $S_1=[0,5]$, $S_2=[0,0.021]$, $S_3=[0.4,1.1]$, $S_{Ir}=[0,12]$. In the training dataset S, there are 1000 samples on irregular level I(h_1), 100 samples on irregular level II(h_2), and 8 samples on irregular level III(h_3). Since irregular level IV is quite dangerous and significantly influences the safe operation of trains, it should be avoided to happen. Therefore, no sample on irregular level IV is considered in this paper. The rest 8321 samples are used to test the irregularity identification model, where there are 8221 samples on h_1, 93 samples on h_2, and 7 samples on h_3.

By analyzing the variation of input and output values in S, 6 referential points for f_1 are set as $A_1=\{0,0.3,0.6,0.9,1.2,5\}$, 9 referential points for f_2 are set as $A_2=\{0,0.002,0.004,0.006,0.007,0.008, 0.01,0.015,0.021\}$, 6 referential points are set as $A_3=\{0.4,0.55,0.60,0.70,0.75,1.1\}$, and 7 referential points for output Ir are set as $D=\{0,2,4,6,8,10,12\}$. Based on Eqs. (6) and (7) in section 4.1, the input and output are transformed into integrated similarity distribution, and the initial casting results of sample pairs in training dataset are generated as shown in Tables 5, 6 and 7.

Based on Eqs. (8) and (9), the REMs for the three input features f_1, f_2 and f_3 are formed by likelihood function normalization as shown in Tables 8, 9 and 10.

However, REMs for f_1, f_2 and f_3 in Tables 8, 9 and 10 may be inaccurate, and the importance weights of different evidence should be optimized. Therefore, the parameters of the identification model are optimized according to Eq. (14) with the training dataset. In the optimization process, input features

Table 12 Optimized REM_1 of input f_1

Ir \ f_1		e_1^1	e_2^1	e_3^1	e_4^1	e_5^1	e_6^1
		A_1^1	A_2^1	A_3^1	A_4^1	A_5^1	A_6^1
		0	0.4294	0.6890	1.0286	1.3916	5
D_1	0	0.1641	0.1494	0.0702	0.0740	0.1107	0.0057
D_2	2.3650	0.1643	0.1441	0.0869	0.0963	0.1016	0.2060
D_3	2.8562	0.1698	0.1372	0.0705	0.0617	0.4184	0.4209
D_4	5.9750	0.1649	0.1333	0.0821	0.1819	0.3693	0.3674
D_5	6.9850	0.0943	0.1028	0.5455	0.1085	0	0
D_6	7.2422	0.1401	0.1587	0.0363	0.3247	0	0
D_7	12	0.1025	0.1745	0.1085	0.1529	0	0

Table 13 Optimized REM_2 of input f_2

Ir \ f_2		e_1^2 A_1^2 0	e_2^2 A_2^2 0.0039	e_3^2 A_3^2 0.0057	e_4^2 A_4^2 0.0061	e_5^2 A_5^2 0.0088	e_6^2 A_6^2 0.0098	e_7^2 A_7^2 0.0116	e_8^2 A_8^2 0.0176	e_9^2 A_9^2 0.021
D_1	0	0.7899	0.4552	0.0562	0	0	0	0	0	0
D_2	2.3650	0.2101	0.5415	0.5376	0.0054	0	0	0	0	0
D_3	2.8562	0	0.0034	0.4019	0.7266	0.2878	0.0275	0.0000	0	0
D_4	5.9750	0	0	0.0044	0.2680	0.7116	0.6502	0.0222	0.0005	0
D_5	6.9850	0	0	0	0	0.0006	0.2729	0.5137	0.1375	0
D_6	7.2422	0	0	0	0	0	0.0246	0.2222	0.5053	0.1602
D_7	12	0	0	0	0	0	0.0248	0.2419	0.3567	0.8398

Table 14 Optimized REM_3 of input f_3

Ir \ f_3		e_1^3 A_1^3 0.4	e_2^3 A_2^3 0.5371	e_3^3 A_3^3 0.5488	e_4^3 A_4^3 0.5755	e_5^3 A_5^3 0.7480	e_6^3 A_6^3 1.1
D_1	0	0.1147	0.1097	0.0926	0.1838	0.1321	0.1422
D_2	2.3650	0.1218	0.1330	0.0914	0.1780	0.1299	0.1850
D_3	2.8562	0.2558	0.1905	0.0702	0.1655	0.1265	0.3593
D_4	5.9750	0.2056	0.1774	0.0928	0.1649	0.1240	0.2736
D_5	6.9850	0.0201	0.0446	0.1991	0.1825	0.1036	0.0400
D_6	7.2422	0.0968	0.1446	0.3438	0.0658	0.1333	0
D_7	12	0.1852	0.2002	0.1101	0.0595	0.2507	0

Table 15 Evidence activated by $f_1(t), f_2(t), f_3(t)$ in testing dataset and their nearest neighboring samples

f	Ir	D_1	D_2	D_3	D_4	D_5	D_6	D_7
$f_1(t)$	e_1^t	0.0702	0.0870	0.0705	0.0823	0.5446	0.0369	0.1086
f_1^1	e_1^1	0.0775	0.0922	0.0766	0.0868	0.5047	0.0476	0.1146
f_1^2	e_1^2	0.0866	0.0988	0.0844	0.0928	0.4534	0.0618	0.1223
f_1^3	e_1^3	0.1204	0.1232	0.1128	0.1146	0.2648	0.1139	0.1503
$f_2(t)$	e_2^t	0	0	0	0.0169	0.4216	0.2916	0.2700
f_2^1	e_2^1	0	0	0	0.0177	0.4357	0.2810	0.2657
f_2^2	e_2^2	0	0	0	0.0214	0.5007	0.2320	0.2458
f_2^3	e_2^3	0	0	0	0.0158	0.4024	0.3060	0.2759
$f_3(t)$	e_3^t	0.1804	0.1749	0.1630	0.1623	0.1775	0.0701	0.0717
f_3^1	e_3^1	0.1819	0.1763	0.1641	0.1634	0.1797	0.0682	0.0662
f_3^2	e_3^2	0.1704	0.1653	0.1515	0.1543	0.1850	0.1066	0.0669
f_3^3	e_3^3	0.0962	0.0948	0.0740	0.0956	0.1985	0.3328	0.1081

Table 16 The reliability factors of the evidence activated by nearest neighboring samples

i	r_i^1	r_i^2	r_i^3
1	0.2571	0.1392	0.1253
2	0.8384	0.4538	0.4088
3	0.1942	0.1051	0.0947

in training dataset S are expressed in belief distribution by using Eq. (6), and the nearest neighboring samples of each input sample are selected by K-NN algorithm. These samples activate the corresponding evidence in Tables 8, 9 and 10.

The reliability factors for $f_1(e_1^t)$, $f_2(e_2^t)$ and $f_3(e_3^t)$ are calculated by using Eqs. (10) - (12), which are $r_1 = 0.3066$, $r_2 = 1$ and $r_3 = 0.2317$. The reliability factors of evidence activated by the nearest neighboring samples are calculated according to Eq. (13). All the importance weight factors are pre-defined to be equal to their corresponding reliability factors. The importance weights of each piece of evidence after optimization are

Table 17 The confusion matrix

		Estimated level			Total number	Accuracy
		h_1	h_2	h_3		
True level	h_1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	M_1	u_1
	h_2	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	M_2	u_2
	h_3	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$	M_3	u_3

Table 18 ACM obtained by M-ER model

		Estimated levels		
		h_1	h_2	h_3
True levels	h_1	8035.6	185.4	0
	h_2	0	78.3	14.7
	h_3	0	1.1	5.9

Table 19 ACM obtained by S-ER model

		Estimated levels		
		h_1	h_2	h_3
True levels	h_1	8015.2	205.8	0
	h_2	0	76.2	16.8
	h_3	0	1.1	5.9

Table 20 ACM obtained by BP-NN model

		Estimated levels		
		h_1	h_2	h_3
True levels	h_1	8101.5	119.5	0
	h_2	32.2	54.8	6
	h_3	1.3	4.6	1.1

Table 21 ACM obtained by RBF-NN model

		Estimated levels		
		h_1	h_2	h_3
True levels	h_1	8139.7	81.3	0
	h_2	19.3	63.4	10.3
	h_3	0	4.3	2.7

Table 22 ACM obtained by GPR model

		Estimated levels		
		h_1	h_2	h_3
True levels	h_1	8192.3	28.7	0
	h_2	45.2	43.5	4.3
	h_3	2.5	2	2.5

Table 23 ACM obtained by SVM model

		Estimated levels		
		h_1	h_2	h_3
True levels	h_1	8167.4	53.6	0
	h_2	17.2	64.9	10.9
	h_3	0	3.8	3.2

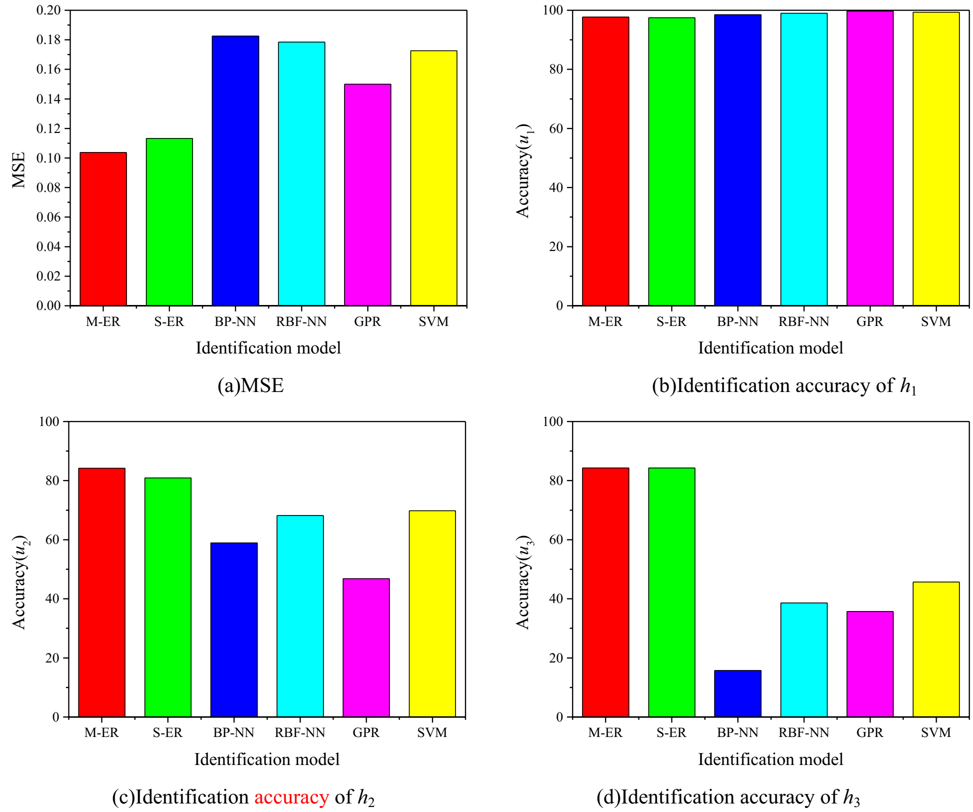


Fig. 6 Experimental results obtained by different identification models

illustrated in Table 11. The optimized REMs for f_1, f_2 and f_3 as well as the referential points are as shown in Tables 12, 13 and 14.

5.2.2 Inference process based on M-ER

A sample in testing dataset ($f_1(t)=0.6897$, $f_2(t)=0.0131$, $f_3(t)=0.5866$) is used to describe the inference process of the multi-level ER rule model. The sample is transformed into belief distributions and activate the corresponding evidence in Tables 12, 13 and 14. Specifically, $f_1(t)$ activates evidence e_3^1 and e_4^1 with similarity $\alpha_{1,1}=0.9979$ and $\alpha_{1,2}=0.0021$ respectively, $f_2(t)$ activates evidence e_7^2 and e_8^2 with similarity $\alpha_{2,1}=0.7552$ and $\alpha_{2,2}=0.2448$, and $f_3(t)$ activates evidence e_4^3 and e_5^3 with similarity $\alpha_{3,1}=0.9358$ and $\alpha_{3,2}=0.0642$. According to Eq. (17), the final evidence activated by $f_1(t)$, $f_2(t)$ and $f_3(t)$ is obtained, as shown in Table 15. Meanwhile, three nearest neighboring samples of ($f_1(t)=0.6897$, $f_2(t)=0.0131$, $f_3(t)=0.5866$) are selected according to Eqs. (15) and (16), which are ($f_1^1=0.6651$, $f_1^2=0.0128$, $f_1^3=0.5816$), ($f_2^1=0.6350$, $f_2^2=0.0118$, $f_2^3=0.5716$), and ($f_3^1=0.5244$, $f_3^2=0.0134$, $f_3^3=0.5499$). Evidence activated by the three samples is also shown in Table 15.

Then, the reliability factors of the three evidence r_i^1, r_i^2 and r_i^3 activated by the nearest neighboring samples are calculated by Eq. (13), as shown in Table 16.

All of the four pieces of evidence for f_1, f_2 and f_3 are fused by the first-level ER rule model respectively, and the outputs of the first-level model are $e_1(4)=\{(D_1,0.0698), (D_2,0.0824), (D_3,0.0684), (D_4,0.0771), (D_5,0.5529), (D_6,0.0452), (D_7,0.1042)\}$, $e_2(4)=\{(D_1,0), (D_2,0), (D_3,0), (D_4,0.0044), (D_5,0.5549), (D_6,0.2332), (D_7,0.2075)\}$, and $e_3(4)=\{(D_1,0.1696), (D_2,0.1640), (D_3,0.1486), (D_4,0.1521), (D_5,0.1878), (D_6,0.1079), (D_7,0.0699)\}$. The outputs of the first-level model $e_1(4)$, $e_2(4)$ and $e_3(4)$ are the three pieces of evidence to be fused in the second-level model, of which the reliability factors are $r_1=0.3066$, $r_2=1$, $r_3=0.2317$ and the importance weights are $w_1=0.8617$, $w_2=0.4943$, $w_3=0.9148$. The final output of the second-level model is $e(3)=\{(D_1,0), (D_2,0),(D_3,0), (D_4,0.0034), (D_5,0.6806), (D_6,0.1652), (D_7,0.1508)\}$. Finally, the estimated vertical track irregularity $\hat{I}r(t)$ is 7.7805 which is calculated by Eq. (18). Consequently, the final irregularity level is determined to be level II which is the same with the true irregularity level.

5.3 Comparison with different identification models

To further illustrate the superiority of the M-ER rule model, five typical machine leaning models are selected to make comparison. They are single-level ER rule model (S-ER) [10, 11], BP neural network model (BP-NN), RBF neural network model (RBF-NN), support vector machine (SVM) and Gaussian process regression (GPR). The ten-fold cross-validation is used to evaluate the experimental results. Specifically, the experiment is repeated ten times. In each time, the historical dataset S is randomly divided into a training dataset with 1108 samples and a testing dataset with 8321 samples. We compare the results from two aspects: model accuracy and engineering practicality.

The model accuracy includes the track irregularity displacement amplitude and the track irregularity level which are evaluated by the MSE (as shown in Eq. (14a)) and confusion matrix (as shown in Table 17) respectively.

In Table 17, $h_l(l=1,2,3)$ represents irregularity level I, II, III, $n_{l,s}(l, s=1,2,3)$ represents the number of samples that are estimated as h_s by the model while the true irregularity level is h_l, M_i represents the total number of samples, here $M_1=8221$, $M_2=93$ and $M_3=7$. The identification accuracy for the lth irregularity level is defined as Eq. (19).

$$u_l = \frac{n_{l,s}}{M_i} \times 100\% l = s = 1, 2, 3 \tag{19}$$

Based on the vertical track irregularity levels in Table 1, the average confusion matrix (ACM) obtained by different identification models are shown from Tables 18, 19, 20, 21, 22 and 23.

The experimental results obtained by different identification models are shown in Fig. 6.As shown in Fig. 6a, the values of MSE of the ER-based models (i.e., M-ER and S-ER) are much smaller than the other models (i.e., BP-NN, RBF-NN, SVM and GPR). It means that the track irregularity displacement amplitude evaluated by the ER-based models is more accurate. Since the introduction of samples expansion strategy and the two-level fusion mechanism, the identification accuracy of M-ER is improved compared with the traditional S-ER. In Fig. 6b, we can see that all of the identification models achieve high identification accuracy of h_1. However, the identification accuracy of high-level irregularity (i.e., h_2 and h_3) of M-ER model and S-ER model are much higher than other models, as shown in Fig. 6c and d. Further more, M-ER model performs best. The reason is that the information for irregularity identification is enriched by involving the similar historical samples in the fusion process. It indicates that the likelihood function normalization can naturally highlight the roles of these small samples in belief distribution when the referential evidence is generated. Therefore, the M-ER is more effective and superior to estimate both of the track irregularity displacement amplitude and track irregularity level.

Moreover, considering the requirements in practical application, the following criteria are made to evaluate the engineering practicality of the identification models. The priorities of the criteria are: R1, R2 and R3.

R1: The levels with high risk h_3 cannot be wrongly identified as low risk h_1 since it is very dangerous (the red cell as shown Table 17). Therefore, $n_{3,1}$ must be equal to 0.
R2: $n_{3,2}$ and $n_{2,1}$ should be as smaller as possible since it is also dangerous (the yellow cells as shown Table 17).
R3: The total accuracy $(u_1+u_2+u_3)$ should be as higher as possible.

Based on the Tables 18, 19, 20, 21, 22 and 23, the values of $n_{3,1}$ of GPR model and BP-NN model are 2.5 and 1.2, respectively. According to **R1**, these two methods cannot be used in practical engineering. Then by comparing the values of $n_{3,2}$ the other four models according to **R2**, we can see that the RBF-NN model and SVM model are much poorer than M-ER model and S-ER model. Finally, according to **R3**, M-ER model is better than S-ER model. Therefore, the M-ER model is most suitable in practical engineering.

In conclusion, the M-ER model performs best in both model accuracy and engineering practicality. Therefore, it can be applied in identifying the tack irregularity.

6 Conclusions

To solve the problems in vertical track irregularity identification, a multi-level ER rule model is proposed in this paper. In the modeling process of M-ER, the referential evidence matrix and fusion parameters (i.e., reliability factors and importance weights) are determined and optimized. In the inference process of M-ER, sample expansion strategy and two-level evidence fusion mechanism are designed. The experimental results show that the M-ER model is more superior than other classical identification models and can be used in the practical engineering.

Further experimental research can be conducted to optimize the M-ER model. For example, in the K-NN method, the selection criteria of the model parameter K is worthy of study. In our experiments, we just give one feasible solution (i.e., $K=3$). Different values of K may affect the identification accuracy of the model. How to balance the identification accuracy and efficiency is a complicated problem which needs to be solved in the future work.

Acknowledgements We acknowledge financial support from NSFC (62103121), Zhejiang Province Outstanding Youth Fund (LR21F030001), Zhejiang Province Public Welfare Technology Application Research Project (LGG22F020023), NSFC (61903108), Zhejiang Province Key R&D projects (2021C03015), Open Fund of National Engineering Research Centre for Water Transport Safety, China (A2020002), Zhejiang Province Public Welfare Technology Application Research Project (LGF20H270004, LGF19H180018), Key project of Zhejiang Provincial Medical and health Science and Technology Plan (WKJ-ZJ-2038).

References

1. Choi II-Y, Um J-H, Lee JS, Choi H-H (2012) The influence of track irregularities on the running behavior of high-speed trains. Proc Inst Mech Eng Part F-J Rail Rapid Transit 227(1):94–102
2. Weston PF, Ling CS, Roberts C, Goodman CJ, Li P, Goodall RM (2007) Monitoring vertical track irregularity from in-service railway vehicles. Proc Inst Mech Eng Part F-J Rail Rapid Transit 221(1):75–88
3. Cantero D, Basu B (2015) Railway infrastructure damage detection using wavelet transformed acceleration response of traversing vehicle. Struct Control Health Monit 22(1):62–70
4. Lee JS, Choi S, Kim S-S, Park C, Kim YG (2012) A mixed filtering approach for track condition monitoring using accelerometers on the axle box and bogie. IEEE Trans Instrum Meas 6(3):749–758
5. Yang JB, Xu DL (2013) Evidential reasoning rule for evidence combination. Artif Intell 205:1–29
6. Wang G, Zhang F, Cheng BY, Fang F (2021) DAMER: a novel diagnosis aggregation method with evidential reasoning rule for bearing fault diagnosis. J Intell Manuf 32(1):1–20
7. Almaghrabi F, Xu DL, Yang JB (2021) An evidential reasoning rule based feature selection for improving trauma outcome prediction. Appl Soft Comput 103:107112
8. Fu C, Chang W, Xu D, Yang S (2019) An evidential reasoning approach based on criterion reliability and solution reliability. Comput Ind Eng 128:401–417
9. Xu XB, Zheng J, Yang JB, Xu DL, Sun XY (2016) Track irregularity fault identification based on evidence reasoning rule. IEEE International Conference on Intelligent Rail Transportation (ICIRT), 16378174
10. Xu XB, Zhang DQ, Bai Y, Chang LL, Li JN (2020b) Evidence reasoning rule-based classifier with uncertainty quantification. Inf Sci 516:192–204
11. Xu W, Peng LL, Zhong QW, Zheng SB, Huang RY (2020a) Prediction model and method of train body vibration based on bagged regression tree. 13th Asia Pacific transportation development conference. https://doi.org/10.1061/9780784482902.061
12. Xiao X, Sun Z, Shen WN (2020) A Kalman filter algorithm for identifying track irregularities of railway bridges using vehicle dynamic responses. Mech Syst Signal Process 138:106582
13. Naganuma Y, Kobayashi M, Okumura T (2010) Inertial measurement processing techniques for track condition monitoring on Shinkansen commercial trains. J Mech Syst Transp Logist 3(1):315–325
14. Tsunashima H, Naganuma HY, Matsumoto A, Mizuma T, Mori H (2011) Japanese railway condition monitoring of tracks using in-service vehicle. 5th IET Conference on Railway Condition Monitoring and Non-Destructive Testing. https://doi.org/10.1049/cp.2011.0587
15. Real JI, Montalbán L, Real T, Puig V (2012) Development of a system to obtain vertical track geometry measuring axle-box accelerations from inservice trains. J Vibroeng 14(2):813–826
16. Real T, Montrós J, Montalbán L, Zamorano C, Real JI (2014) Design and validation of a railway inspection system to detect lateral track geometry defects based on axle-box accelerations registered from in-service trains. J Vibroeng 16(1):234–248
17. Li ZL, Molodova M, Núñez A, Dollevoet R (2015) Improvements in axle box acceleration measurements for the detection of light squats in railway infrastructure. IEEE Trans Ind Electron 62(7):4385–4397
18. Sun XF, Yang F, Shi J, Ke ZT, Zhou YL (2021) On-board detection of longitudinal track irregularity via axle box acceleration in HSR. IEEE Access 9:14025–14037
19. Bhardwaj B, Bridgelal R, Chia L, Lu P, Dhingra N (2020) Signal filter cut-off frequency determination to enhance the accuracy of rail track irregularity detection and localization. IEEE Sensors J 20(3):1393–1399
20. Real J, Salvador P, Montalbán L, Bueno M (2010) Determination of rail vertical profile through inertial methods. Proc Inst Mech Eng Part F-J Rail Rapid Transit 225(1):14–23
21. Wei XK, Liu F, Jia LM (2016) Urban rail track condition monitoring based on in-service vehicle acceleration measurements. Measurement 80:217–228
22. Ning J, Lin JH, Zhang B (2016) Time-frequency processing of track irregularities in high-speed train. Mech Syst Signal Process 66-67:339–348
23. Xie T, Li ZC, Wang TZ, Shi M, Wang YD (2021) An integration fault detection method using stator voltage for marine current turbines. Ocean Eng 226:108808
24. Jiang J, Wang FG, Zhang Y, Qin Y, Gao XD (2015) Estimation of vertical track irregularity based on NARX neural network. Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation. https://doi.org/10.1007/978-3-662-49370-0_17
25. Rosa AD, Kulkarni R, Qazizadeh A, Berg M, Gialleonardo ED, Facchinetti A, Bruni S (2020) Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms. Proc Inst Mech Eng Part F-J Rail Rapid Transit 235(1):107–120
26. Ministry of Railways of the Peoples Republic of China (2006) Railway line repair rules. Chinese Railway Press, Beijing

Zhenjie Zhang was born in Haining, Zhejiang, China in 1989. He received the B.S. and M.S. degrees in mechanical engineering from the Hangzhou Dianzi University in 2012 and 2015 respectively, and the Ph.D. degree in mechanical engineering from Zhejiang University of Technology in 2019.

Since 2019, he has been a lecturer with School of Automation (School of Artificial Intelligence) and the Belt and Road information technology research institute, Hangzhou Dianzi University, Hangzhou, China. His research interests include uncertain information fusion, fault diagnosis and complex networks.

Xiaobin Xu was born in Zhengzhou, Henan, China in 1980. He received the B.S. and M.S. degrees from the Henan University in 2003 and 2006 respectively, and the Ph.D. degree in power electronics and power drives from Shanghai Maritime University in 2009.

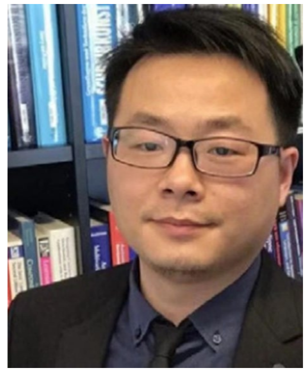
He is currently a professor with School of Automation (School of Artificial Intelligence) and the Belt and Road information technology research institute, Hangzhou Dianzi University, Hangzhou, China. From 2014 to 2015, he was a a visiting professor in University of Manchester, Manchester, UK. His research interests include fuzzy set theory, evidence theory and its Applications in the processing of uncertain information, the reliability analysis, safety evaluation and condition monitoring of complex industrial systems.

Xuelin Zhang was born in Xinyang, Henan, China in 1992. He is currently studying for a PhD in the School of Automation (Artificial Intelligence College) of Hangzhou Dianzi University. His research directions include uncertain information fusion, fault diagnosis and intelligent control.

Xiaojian Xu was born in Zibo, Shandong, China in 1988. She received the B.S., M.S. and ph.D degree in school of energy and power engineering from Wuhan University of Technology in 2011, 2014, and 2018.

From 2018 to 2020, she was a lecturer with school of Automation (School of Artificial Intelligence) and the Belt and Road information technology research institute, Hangzhou Dianzi University, Hangzhou, China. Since 2021, She has been a post doctor in China Waterborne Transport Research Institute, Beijing, China. Her research interests include intelligent fault diagnosis, decision support system, multiple information fusion.

Zifa Ye was born in Shangrao, Jiangxi, China in 1995. He received the M.S. degrees in Control Science and Engineering from the Hangzhou Dianzi University in 2021. His research interests include intelligent information processing, evidence theory and its applications in the processing of uncertain information.



Guodong Wang received his Ph.D. degree in computer science from Vienna University of Technology at January 2019.

He is currently the deputy chief engineer of Sino-Austria Research Institute for Intelligent Industries. Before that, he was a machine learning research fellow at Cyber Physical Systems Group, Vienna University of Technology, Austria. His research interests include automated machine learning, learning representation, hyper-parameter searching. He is also fascinated by designing machine learning and deep learning solutions for solving real-world problems e.g., ML/DL for Cyber-Physical Production Systems, predictive analytics, object detection with multi-source data fusion, data analysis for medical usage.

Schahram Dustdar (Fellow, IEEE) received his M.Sc. (1990) and PhD. degrees (1992) in Business Informatics from the University of Linz, Austria. In 2003 he received his Habilitation degree for his work on Process-aware Collaboration Systems-Architectures and Coordination Models for Virtual Teams.

He was an Honorary Professor of Information Systems at the Department of Computing Science, University of Groningen, Groningen, The Netherlands, from 2004 to 2010. From 2016 to 2017, he was a visiting professor at the University of Sevilla, Sevilla, Spain. In 2017, he was a visiting professor at the University of California at Berkeley, Berkeley, CA, USA. He is currently a professor of computer science with the Distributed Systems Group, TU Wien, Vienna, Austria.

He was an elected member of the Academy of Europe, where he is the Chairman of the Informatics Section. He was a recipient of the ACM Distinguished Scientist Award in 2009, the IBM Faculty Award in 2012, and the IEEE TCSVC Outstanding Leadership Award for outstanding leadership in services computing in 2018. He is the CoEditor-in-Chief of the ACM Transactions on Internet of Things and the Editor-in-Chief of Computing (Springer). He is also an Associate Editor of the IEEE Transactions on Services Computing, the IEEE Transactions on Cloud Computing, the ACM Transactions on the Web, and the ACM Transactions on Internet Technology. He serves on the Editorial Board of IEEE Internet Computing and the IEEE Computer Magazine.