A Multisource Data Fusion Modeling Prediction Method for Operation Environment of High-Speed Train

Decang Li,1,2,3 Juhui Zhang,1 Ruxun Xu,1,2,3 Jianjun Meng,1,2,3 Jianming Wang,1 Xiaoqiang Chen,1,2,3 Xin Jia,1 and Junhui Ma1

1Mechatronics & R Institute, Lanzhou Jiaotong University, Lanzhou 730070, China
2School of Mechatronic Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
3Gansu Provincial Engineering Technology Center for Informatization of Logistics & Transport Equipment, Lanzhou 730070, China

Correspondence should be addressed to Ruxun Xu; 365261637@qq.com

Received 20 December 2021; Revised 5 March 2022; Accepted 17 March 2022; Published 30 March 2022

Academic Editor: Luca Pancioni

Copyright © 2022 Decang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Providing accurate and reliable railway regional environmental data is a key consideration in operation control and dynamic dispatching of high-speed train. However, there are problems of low reliability and high uncertainty in the single data processing of high-speed train operating area environment. Therefore, this paper proposes a novel multisource sensor data fusion method based on a three-level information fusion framework. Firstly, the feature of the same kind of sensor data is extracted by the Kalman Filter (KF) algorithm as the input of back propagation neural network (BPNN). Then input the sample site into the BPNN for training and recognition, the feature fusion of heterogeneous sensor data is carried out, the decision output of BPNN is obtained, the output results are normalized, and its output is used as the basic probability assignment of Dempster–Shafer (D–S) evidence theory and synthesis rules. Finally, the decision fusion of multisource data is realized by D-S evidence theory. The simulation results show that compared with the traditional single fusion algorithm, the algorithm improves the accuracy of the prediction of high-speed train operation environment and reduces the MAPE from 13.82% to 7.455%, and the RMSE from 0.77 to 0.69, and meanwhile, increases the $R^2$ from 0.87 to 0.97.

1. Introduction

Nowadays world railway has entered the “high-speed railway age.” With a large number of new high-speed trains driving, a large number of new technology and new equipment coming into use, the safety of high-speed train operation is becoming particularly important. Take China’s high-speed railway as an example, a high-speed railway network with the fastest operation speed and the longest mileage has been built in the world. An “eight vertical and eight horizontal” railway net of China with reasonable layout, wide coverage, clear hierarchy, safety and efficiency has been basically formed [1]. With the increase of the train speed, which makes the train more sensitive to the complex and changing environment, the complex and harsh environment (such as strong wind and sand, earthquake, mountain torrents, avalanches, landslides) seriously threatens the safety of train operation. There are many derailments and overturning of train accidents caused by the harsh environment in the world rail transportation system [2–4]. Therefore, people have done a lot of research on the safety performance of trains under harsh conditions [5–8].

With the rapid development of high-speed railway and the large-scale construction of railway network, the operation environment of train has become more and more complex. To ensure the safe operation of high-speed train under complex environment, higher requirements are put forward for the integration technology of high-speed train operation control and dynamic dispatching [9]. Reliable multi-sensor data fusion technology plays an important role in predicting the train operation environment, the data fusion algorithm of multi-sensor system has a direct and
deep impact on the decision-making of information system, which is the key technology to improve the integrated mode of high-speed train operation control and dynamic dispatching.

Data fusion technology has always been one of the research fields that have received many attentions in the WSN field. Now, there are many data fusion methods used in life. These methods are currently widely studied in the WSN network layer and application layer [10].

At present, for the research of multisource data fusion, scholars at home and abroad have proposed a large number of data fusion algorithms and reliable fusion structures. In this work, Jafari proposed the fuzzy c-means classifier algorithm is used to calculate the partial matching of each feature, and the features with strong diagnostic ability are fused by fuzzy integral. The experimental results are given [11]. Wang aiming at the problem of multisource heterogeneous data fusion in landslide monitoring, a multisource heterogeneous monitoring data fusion algorithm based on BPNN is proposed. The algorithm takes the temperature, humidity and precipitation affecting landslide deformation as the input and the change of landslide displacement data as the expected output data, which can effectively improve the prediction accuracy of the algorithm [12]. Huda tested and compared the ability of three classification algorithms (maximum likelihood (ML), decision tree (DT) and support vector machine (SVM)) in image inference and extraction of urban land use, so as to obtain the classification map of Newcastle, UK. The above three classifiers are used and applied to the combined data of 33 bands to evaluate their effectiveness in distinguishing urban land cover/land use types [13]. Xu used Bayesian theory is used to analyze the reliability of data and the credibility of data sources, and a trust evaluation model based on Dempster shaper theory is established to study the dynamics, distribution and uncertainty of emergencies [14]. Rajendra proposed a virtual sensing algorithm based on KF is proposed to estimate the response of the position of interest. The algorithm also carries out multi-sensor data fusion to improve the estimation accuracy under non-stationary tidal load. Through numerical analysis and laboratory experiments, the results show that the unmeasured response can be reasonably recovered from the measured response [15]. Martino developed a general framework, the connections between them are drawn through dual formulas, and their applications in the main tasks (regression, smoothing, interpolation, and filtering) are discussed [16].

To sum up, these data fusion algorithms have their own advantages and disadvantages, and they are more or less complementary. Therefore, in the actual use process, researchers will use one or more algorithms according to their respective advantages and disadvantages, or establish new algorithm models based on these algorithms to solve practical problems. At present, in the multisource data fusion of railway operation environment monitoring, the traditional algorithm or single algorithm makes the data fusion results inaccurate, inconsistent, incomplete, scattered, and unreliable. Therefore, in order to improve the accuracy and efficiency of data fusion results of train operation environment along the railway, based on the analysis and research of existing data fusion algorithms, this paper proposes a data analysis and processing model based on multi-sensor three-level information fusion. Firstly, multiple sensors are used to collect the data signals to be monitored. Then, the collected data are fused at three levels. In the first stage, Kalman filter fusion algorithm is used to preprocess similar data, and BP neural network algorithm is used to fuse heterogeneous sensor data in the region. Finally, the decision fusion of multi-sensor fusion results is carried out by using D-S evidence theory, and the safety of environmental monitoring is judged according to the fusion results. The effectiveness and accuracy of this method are verified by experiments.

2. Multisource Data Fusion Model for High-Speed Train Operation Environment

2.1. Multi Source Data Fusion Architecture. To monitor the environment along the railway, it is necessary to divide and arrange regions to arrange sensor networks in each region, and then form a clustering structure in each region according to the routing protocol. The network structure is shown in Figure 1.

The sensor node is responsible for collecting environmental data, and the cluster head node is responsible for collecting the data transmitted by the sensor nodes in the region. The cluster head node performs first fusion after collecting data, and then sends the results of first fusion to the sink node. The sink node performs feature fusion after collecting the data of each region. After the secondary feature fusion, taking the data characteristics of the fusion results as the input of three-level decision fusion, comprehensively evaluate the decision fusion results, obtaining the prediction results of high-speed train operation environment along the railway. Then, put the prediction and early warning information into the high-speed train intelligent control and dynamic dispatching system to realize the management and control integration of high-speed train control and dispatching in complex environment.

2.2. Multisource Data Fusion Model. In this paper, taking the prediction of wind speed along the regional railway as an example, when the cluster head node receives the data collected by the sensor nodes in the region, first preprocessing the similar sensor data in the region by using the Kalman filter fusion algorithm, and then fusing the heterogeneous sensor data in the region by using the BPNN algorithm. Send the fusion results to the sink node for fusion, and use D-S evidence theory to fuse the results fused by BPNN at the decision level so as to predict and judge the environmental conditions along the railway. The multi-sensor fusion model of complex environment along the regional railway is shown in Figure 2.

2.2.1. First Fusion Based on KF Algorithm. KF can be divided into the continuous system and discrete system according to the time characteristics of the system. The environmental
monitoring of high-speed train operation region is to observe the monitoring points at a certain time interval and determine the periodic variation of the monitoring points. Its processing process is a typical discrete linear system filtering problem. The mathematical model of discrete linear system is described by a linear differential equation and discrete observation equation with random initial state and dynamic noise [18]. The state equation and observation equation of discrete linear system are expressed by (1) [19], as follows:

\[
\begin{align*}
X_k &= \Phi_{k,k-1}X_{k-1} + \Gamma_{k,k-1}\Omega_{k-1}, \\
L_k &= B_kX_k + \Delta_k,
\end{align*}
\]

where \(X_k\) and \(X_{k-1}\) is the state vector, \(\Phi_{k,k-1}\) is the state transition matrix, \(\Gamma_{k,k-1}\) is the dynamic noise figure, \(\Omega_{k-1}\) is the dynamic noise, \(L_k\) is the observation vector, \(B_k\) is the coefficient of observation equation, and \(\Delta_k\) is the observing noise.

Firstly, the similar sensor data collected by cluster head nodes in each region are fused by the KF fusion method. Then, the KF is used to analyze the relationship between the estimated values of the previous time and the later time of the environment along the railway, and the estimated value at the next time is predicted by the known a priori value. Finally, depending on the established system model, if the previous state is known, the next state can be predicted by (2), as follows:

\[
\begin{align*}
\tilde{x}_k &= A\tilde{x}_{k-1} + Bu_{k-1}, \\
P_k &= AP_{k-1}A^T + Q,
\end{align*}
\]

where \(\tilde{x}_k\) is the priori state estimation value of time \(k\), that is, the result of time \(k\) predicted according to the optimal estimation of \(k-1\) time, \(A\tilde{x}_{k-1}\) is the best posteriori state estimate at \(k-1\) time, \(Bu_{k-1}\) is the control quantity of state equation, \(P_k\) is the priori estimated covariance at \(k\) time, \(P_{k-1}\) is the posteriori estimated covariance at \(k-1\) time, \(A\) is the state transition matrix, \(A^T\) is the transpose matrix, and \(Q\) is the covariance of system process, that is, the error between the state transition matrix and the actual process.

The KF state update equation by equation (3) is as follows:

\[
\begin{align*}
\tilde{x}_k &= \tilde{x}_{k-1} + K_k(z_k - H\tilde{x}_k), \\
P_k &= (I - K_kH)P_{k-1},
\end{align*}
\]

where \(K_k\) is the Kalman gain, \(I-H\tilde{x}_k\) is the transformation matrix from state variable to observation, \(H^T\) is the transpose matrix, \(R\) is the covariance of observation noise, and \(I\) is the identity matrix.

The weighting factor of Kalman gain by (4) is as follows:

\[K_k = \frac{P_kH^T}{HP_kH^T + R}.
\]

It is assuming that \(n\) wind speed sensors are placed in the region, the state at \(k\) time can be predicted by the state at \(k-1\) time. Then, the \(P_k\) is estimated at \(k\) time by the system prediction error at \(k-1\) time, therefore, the \(K_k\), the \(\tilde{x}_k\) and the \(P_k\) are obtained by calculation. Further, the data fusion results are obtained. The flow of KF data fusion algorithm is shown in Figure 3.
2.2.2. Secondary Fusion Based on BPNN Algorithm. BPNN is modeled according to the internal relationship of the data itself, with good self-organization, self-adaptability, strong learning ability and anti-interference ability. The BPNN fusion method is used to fuse the heterogeneous sensor data in each region, according to the fusion results obtained by KF algorithm, the wind speed prediction of the first 10 hours’ data was made by time series analysis, so as to obtain the secondary prediction of the regional environmental conditions along the railway. The specific fusion steps are as follows:

Step 1: the environmental parameters were processed by feature processing

Step 2: standardize the processing of each characteristic signal and provide a unified form of BPNN input

Step 3: input the sample data into the BPNN for model training until the requirements are met, then, the trained network is taken as the known network, the operation environment monitoring, the designed BPNN is shown in Figure 4. where \( x \) is the input vector, that is, \( x = \{x_1, x_2, x_3, \ldots, x_{10}\} \), \( w_{ij} \) is the connection weight between the input layer and the hidden layer, \( v_{ij} \) is the link weight between the hidden layer and the output layer, and \( Y_j \) is the output of BPNN.

The parameter setting of each layer of BPNN includes: the number of neurons in the input layer is determined according to the dimension of the input signal, and the wind speed parameter is selected as the input in this paper. The number of neurons in the hidden layer is generally calculated by equation (8), and the network output error is finally determined through experiments until the minimum value.

As long as the number of hidden layer neurons is enough, the BPNN with a hidden layer can approach the nonlinear function of any complexity with any accuracy. The nonlinear transfer function of BPNN by Eq. (6), as follows:

\[
f(x) = \frac{1}{1 + e^{-\beta x}}
\]  

(6)

To standardize the data, the mean and standard deviation of the original data are given. The processed data conforms to the standard normal distribution, that is, the mean value is 0 and the standard deviation is 1, the transformation function by equation (7), as follows:

\[
x = \frac{\chi - \mu}{\sigma},
\]  

(7)

where \( \chi \) is the raw data, \( \mu \) is the mean of all data, and \( \sigma \) is the standard variance.

According to the actual situation of high-speed train operation environment monitoring, the designed BPNN is shown in Figure 4. where \( x \) is the input vector, that is, \( x = \{x_1, x_2, x_3, \ldots, x_{10}\} \), \( w_{ij} \) is the connection weight between the input layer and the hidden layer, \( v_{ij} \) is the link weight between the hidden layer and the output layer, and \( Y_j \) is the output of BPNN.

The parameter setting of each layer of BPNN includes: the number of neurons in the input layer is determined according to the dimension of the input signal, and the wind speed parameter is selected as the input in this paper. The number of neurons in the hidden layer is generally calculated by equation (8), and the network output error is finally determined through experiments until the minimum value.

\[
h = \sqrt{m + n + a},
\]  

(8)

where \( h \) is the number of hidden layer nodes, \( m \) is the number of input layer nodes, \( n \) is the number of output layer nodes, and \( a \) is the adjustment constant between 1 and 10. Through the experimental results, it is found that when the number of hidden layer nodes is 6, the network error is the
smallest. Therefore, this paper sets the number of hidden layer nodes to 6 and the number of output layer nodes to 1. The output is the first-class judgment of the environmental conditions of each region.

After the design of the above number of nodes is completed, the BPNN trained, 744 groups of data were collected as sample data, and BPNN was established by MATLAB. After normalizing the data, take the error difference ε = 0.01 for training. After training, the normalized monitoring data are input, and the output of BPNN is the regional environmental status along the railway.

2.2.3. Three Level Fusion Based on D-S Evidence Theory.

In order to enhance the accuracy of environmental monitoring along the railway, the D-S evidence theory is used for regional data fusion. The D-S evidence theory can not only solve the uncertainty of BPNN, but also set the confidence interval with the help of mass function to ensure the effectiveness of each subset of data. At the same time, the BPNN can solve the problem of serious and complete conflict of evidence in the D-S evidence theory. After the first level fusion, the preliminary judgment of each region can be obtained, and then the decision-making level fusion of the whole region can be carried out by using the D-S evidence theory.

We assume that the wireless sensor network along the railway is divided into n regions, and the result of region i fused by the BPNN is recorded as $R_i$. The fusion results of all regions constitute the recognition framework, and the focal elements of each trust function correspond to the fusion results of each region. Then, the BPNN of each region is standardized to obtain the basic probability distribution value $m$ of each focal element. Finally, the whole region fusion is carried out by using the synthesis rules of the D-S evidence theory, so as to obtain the environmental prediction results along the railway.

The D-S evidence theory is expressed in the form of sets, assuming that $N$ mutually exclusive independent events form a set, which is called the recognition framework, as shown in equation.

$$\Theta = \{x_1, x_2, \ldots, x_n, \ldots, x_N\},$$  \hspace{1cm} (9)

where $x_n$ is the nth element identification of the framework $\Theta$.

The form of identification framework $\Theta$ subsets a set and becomes a power set $2^\Theta$. Then the basic probability distribution function $m$ is defined by Eq. (10), as follows:

$$m: 2^\Theta \rightarrow [0, 1].$$  \hspace{1cm} (10)

And the Eq. (10) should meet the Eq. (11), as follows:

$$\begin{align*}
m(\emptyset) &= 0, \\
\sum_{A \in \Theta} m(A) &= 1,
\end{align*}$$  \hspace{1cm} (11)

where $\emptyset$ is the impossible event, $m(A)$ is the basic probability of $A$.

The trust function $Bel$ and likelihood function $Pl$ in the D-S evidence theory are defined by Eq. (12), as follows:

$$\begin{align*}
Bel(A) &= \sum_{B \in A} m(B), \\
Pl(A) &= \sum_{B \cap A \neq \emptyset} m(B).
\end{align*}$$  \hspace{1cm} (12)

Multiple evidence sources can be synthesized through the above the Eq of the D-S evidence theory. This paper assumes that the wireless sensor network along the railway is divided into n regions, $n$ regions correspond to $m$ pieces of evidence, and the corresponding basic probability distribution functions are $m_1, m_2, \ldots, m_n$. The evidence synthesis formula by Eq. (13), as follows:

$$m(A) = \frac{1}{K} \sum_{A_1 \cap A_2 \cap \ldots \cap A_n = A} m_1(A_1) \cdot m_2(A_2) \ldots m_n(A_n),$$  \hspace{1cm} (13)

where $K$ is the normalization factor, it is defined by Eq. (14), as follows:

$$K = \sum_{A_1 \cap A_2 \cap \ldots \cap A_n \neq \emptyset} m_1(A_1) \cdot m_2(A_2) \ldots m_n(A_n).$$  \hspace{1cm} (14)

2.3. Evaluation Method of Model Accuracy.

In order to verify the prediction effect of the three-level fusion model on the wind speed in the target area, three statistical indexes of mean absolute percentage error (MAPE), root mean square error (RMSE) and determined coefficient ($R^2$) are selected as the evaluation criteria of the accuracy of the three-level fusion model.
The MAPE and the RMSE are comprehensive indicators of model error analysis. It is expressed as the average of the absolute value of the error between the predicted value and the observed value. The MAPE is expressed by Eq. (15), as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{f(x_i) - y_i}{y_i} \right|,$$

where $f(x_i)$ is the predicted value of the $i$ sample, $y_i$ is the true measured value of the $i$ sample, and $n$ is the number of samples.

The RMSE represents the degree of dispersion of the predicted value, which is 0 in the case of best fit. The RMSE is expressed by Eq. (16), as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (f(x_i) - y_i)^2}{n}}.$$

The determined coefficient is an important index to investigate the fitting degree between the sample regression line and the observed value. In multivariate regression fitting, it is usually called multiple determination coefficient, which is expressed by $R^2$, that is, the ratio of the sum of regression squares to the sum of squares of total deviations. The larger $R^2$, the higher the fitting degree of the model. The $R^2$ is expressed by Eq. (17), as follows:

$$R^2 = \frac{\sum_{i=1}^{n} (f(x_i) - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$

where $\bar{y}$ is the mean value of the measured value of the sample. The closer $R^2$ is to 1, the smaller the error, the better the prediction effect of the model.

### 3. Simulation Experiments and Results Analysis

#### 3.1. Data Level Fusion

This paper takes the data detected in an area with wind speed monitoring station on Lanxin high-speed railway as an example, collects the wind speed data in the area, and divides the data information by region. The hourly wind speed data of 31 days (1–31 days) in the area are selected as the research object, and specific monitoring values are shown in Table 1.

At 01:00 on the 1st, take region 1 as an example, select the wind speed value at 01:00 on the 1st at the adjacent time, substitute the data into Eq (1)–Eq (4), that is, $[v_{(1)}^{[01:00]}] = 7.521$ m/s, and so on, the fusion value $[v_{(1)}]_{2}$ of other meshes at each time can be obtained. According to the data level fusion value of the known grid, spatial interpolation is carried out for the area to be predicted, and the first fusion data results are shown in Table 2.

#### 3.2. Feature Level Fusion

Due to the complex factors affecting the change of wind speed, BPNN is used as the secondary fusion model and wind speed data as the input variable. The specific implementation steps are as follows:

Step 1: sort out the collected data from 01:00 on the 1st to 24:00 on the 31th, and process the labels of the input samples into a unified format.

Step 2: the experimental results show that when the number of hidden layer nodes is 6, the network error is small, so this paper sets the number of hidden layer nodes to 6. Until the network error is satisfied $\varepsilon = 0.01$, end the network training. At this time, the sample set can be input into the neural network for identification, the decision output of the BPNN can be obtained, and the output results can be normalized to obtain the basic probability distribution of focus elements.

Step 3: the wind speed prediction value of each area is stored in table form. $[v_{(0)}]_{2}$ is calculated according to Eq (5)–(7), as shown in Table 3.

#### 3.3. Decision Level Fusion

According to the data obtained above, the data level fusion value $[v_{(0)}]_{1}$ and characteristic level fusion value $[v_{(0)}]_{2}$ are taken as the two influencing factors affecting wind speed, the $[v_{(0)}]_{1}$ and $[v_{(0)}]_{2}$ are taken as the input variables of the decision level fusion model, and the wind speed is input into the decision level fusion model as the prediction output variable. The decision level fusion value that is the wind speed prediction value, is calculated according to (9)–(14), as shown in Table 4. Using the three-level fusion model to predict the wind speed in the unknown area can obtain the real-time change information of wind speed. Through this model, we can more accurately grasp the wind speed change information at different locations.

#### 3.4. Model Comparison and Result Analysis

In order to test the fusion accuracy of the three-level fusion model, the predicted values and real values of different models are compared and analyzed, and the prediction results of the first, second and third level fusion are compared at the same time. The results are shown in Figures 5–8.

Figures 5–8 respectively reflect the comparison between the fusion value of the first level, second level third level and the real value of wind speed. Through the comparison, it can be seen that the gap between the fusion value of the third level and the real value of wind speed is small. After training, the fitting curves of the first level, second level and third level fusion models reflect that the third fusion model is better than the first level and second level fusion models in the test data set, the change trend of the tertiary fusion value is consistent with the real value of wind speed, and the fitting effect is better. It shows that the third level data fusion model has good fusion performance.

Further evaluate the prediction effect of the fusion model, and evaluate the model by calculating the statistical measurement index by Eq. (15)–(17). The calculation results are shown in Table 5.

The results show that for MAPE, the third level fusion model is 6.36% and 4.45% lower than the first level and second level fusion models respectively. Overall, the three indexes of the third level fusion are better than the first and secondary fusion indexes, indicating that the prediction
Table 1: Wind speed value at monitoring point (m/s).

| Time              | Region |
|-------------------|--------|
|                  | 1     | 2       | ... | 10 | ... | 20 |
| 01:00 on the 1st  | 8.0   | 5.1     | ... | 3.8 | ... | 6.6 |
| 02:00 on the 1st  | 8.1   | 5.5     | ... | 2.2 | ... | 6.3 |
| 03:00 on the 1st  | 7.5   | 6.1     | ... | 3.2 | ... | 5.6 |
| ...               | ...   | ...     | ... | ... | ... | ... |
| 24:00 on the 31th | 6.7   | 4.4     | ... | 2.0 | ... | 7.5 |

Table 2: Wind speed value of first fusion (m/s).

| Time              | Region |
|-------------------|--------|
|                  | 1     | 2       | ... | 10 | ... | 20 |
| 02:00 on the 1st  | 7.521 | 5.658   | ... | 2.822 | ... | 6.936 |
| 03:00 on the 1st  | 8.224 | 6.973   | ... | 3.774 | ... | 5.879 |
| 04:00 on the 1st  | 6.817 | 6.076   | ... | 4.188 | ... | 6.273 |
| ...               | ...   | ...     | ... | ... | ... | ... |
| 24:00 on the 31th | 6.367 | 4.46    | ... | 2.14 | ... | 7.647 |

Table 3: Wind speed value of secondary fusion (m/s).

| Time              | Region |
|-------------------|--------|
|                  | 1     | 2       | ... | 10 | ... | 20 |
| 02:00 on the 1st  | 8.208 | 4.482   | ... | 3.072 | ... | 6.306 |
| 03:00 on the 1st  | 8.097 | 5.521   | ... | 2.988 | ... | 5.455 |
| 04:00 on the 1st  | 6.954 | 6.075   | ... | 4.388 | ... | 5.351 |
| ...               | ...   | ...     | ... | ... | ... | ... |
| 24:00 on the 31th | 6.808 | 4.322   | ... | 2.761 | ... | 6.519 |

Table 4: Wind speed value of third level fusion (m/s).

| Time              | Region |
|-------------------|--------|
|                  | 1     | 2       | ... | 10 | ... | 20 |
| 02:00 on the 1st  | 8.204 | 5.973   | ... | 2.612 | ... | 6.615 |
| 03:00 on the 1st  | 7.737 | 5.803   | ... | 3.385 | ... | 5.599 |
| 04:00 on the 1st  | 6.972 | 6.164   | ... | 4.477 | ... | 5.815 |
| ...               | ...   | ...     | ... | ... | ... | ... |
| 24:00 on the 31th | 6.164 | 4.321   | ... | 2.412 | ... | 7.424 |

Figure 5: First-level fusion values.
performance of the model is better. Train the model and obtain multiple sets of MAPE, RMSE and $R^2$ index values of the third level fusion model through 20 tests, as shown in Figure 9.

Where shows the experimental results of the third level fusion model after 20 times of training. The experiments show that the MAPE and RMSE of the third level fusion model are less than the optimal values of the first and second level fusion models, while $R^2$ is greater than the optimal values of the first and second level fusion models. Therefore, the first mock exam of the wind speed prediction using the third level fusion method is better than the single model, and it proves that the data fusion can improve the quality of the wind speed prediction.

4. Conclusion

According to the actual situation of the environment along the railway, a three-level fusion model is established to analyze the monitoring values to predict the wind speed values, and the following conclusions are obtained:

Considering the abnormal data collected by multiple sensors, firstly, Kalman filter fusion is used for data-level fusion of sensor data in each region, and then BPNN is used for feature-level fusion, the output at this point is the real-time status of each area. Due to the uncertainty of the outputs, we still use the D-S evidence theory for decision-making level fusion of the results of the second-level fusion. The model is designed to eliminate the uncertainties in multi-sensor data acquisition along the railway. And it turns out to be feasible by simulation experiments.

This paper only takes wind speed as the research object. It is worth noting that we do not consider the influence of ultra-short-term wind speed on data fusion results, which in fact has an important influence on optimization results.
Therefore, our future work will input the ultra-short-term wind speed into the model, and at the same time, considering earthquakes, landslides and other complex environments, put forward a safer prediction method for the regional operation environment of high-speed trains.

Data Availability

The data used to support the findings of this study were supplied under license and so cannot be made freely available. Requests for access to these data should be made to the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors acknowledge the support provided by the National Natural Sciences Foundation of China (nos. 72061021 and 62063013), Science and Technology Plan Project of Gansu (nos. 21JR7RA284 and 20JR10RA251), and Youth Fund Project of Lanzhou Jiaotong University (no. 20210018).

References

[1] Q. Peng, J. Li, Y. Yang, and C. Wen, “Influences of high-speed railway construction on railway transportation of China,” Journal of Southwest Jiaotong University, vol. 51, no. 3, pp. 525–533, 2016.
[2] D. Li, J. Meng, H. Bai, and R. Xu, “Active control strategy for the running attitude of high-speed train under strong crosswind condition,” Vehicle System Dynamics, vol. 56, no. 7, pp. 1028–1050, 2017.
[3] H. Zhang, L. Meng, and X. Dai, “Analysis and countermeasures of railway high impact weather,” China Railway, vol. 62, no. 5, pp. 12–16, 2019.
[4] Z. Tan, “Reports on emergencies test the actual effect of media integration and development—thoughts on Chenzhou Daily’s all media report on the rollover accident of train T179 on Beijing-Guangzhou line,” China City Newspaper, vol. 5, pp. 10–13, 2020.
[5] M. Yu, J. Liu, T. Li, and Q. Zhang, “Operational safety characteristics of high-speed train under strong wind and rainfall environment,” Journal of Mechanical Engineering, vol. 57, no. 20, pp. 172–180, 2021.
[6] F. Cheli, S. Giappino, and L. Rosa, “Experimental study on the aero dynamic forces or railway vehicles in presence of turbulence,” Journal of Wind Engineering and Industrial Aerodynamics, vol. 12, no. 3, pp. 311–316, 2013.
[7] Y. Misu and T. Ishihara, “Prediction of frequency distribution of strong crosswind in a control section for train operations by using onsite measurement and numerical simulation,” Journal of Wind Engineering and Industrial Aerodynamics, vol. 17, no. 4, pp. 69–79, 2018.
[8] H. Steele, S. Noda, S. Yamamoto, and M. Korenaga, “Improved epicentral distance estimations for railway earthquake early warning,” Quarterly Report of RTRI, vol. 61, no. 3, pp. 222–227, 2020.
[9] B. Ning, H. Dong, and W. Zheng, “Integration of train control and online rescheduling for high-speed railways: challenges and Future,” Acta Automatica Sinica, vol. 45, no. 12, pp. 2208–2217, 2019.
[10] X Zhao and B Li, “The countermeasures of urban energy risk control oriented to machine learning and data fusion,” Energy Reports, vol. 8, pp. 2547–2557, 2022.
[11] H. Jafari, J. Poshtan, and H. Sadeghi, “Application of fuzzy data fusion theory in fault diagnosis of rotating machinery,” Journal of Systems and Control Engineering, vol. 232, no. 8, pp. 1015–1024, 2018.
[12] Z. Wang, L. Wang, G. Huang, Q. Han, and F. Xu, “Research on multi-source heterogeneous data fusion algorithm of landslide monitoring based on BPNN,” Journal of Geomechanics, vol. 26, no. 4, pp. 575–582, 2020.
[13] M. Huda, “The implementation of image classification and analysis of Mrsd using three different classifiers: a Case Study of Newcastle-UK,” Divyal Journal of Engineering Sciences, vol. 8, no. 4, pp. 521–537, 2015.
[14] X. Xu, Y. Lou, and C. Yu, “Research on the requirements of multi-source data fusion emergency decision-making based on Dempster-Shafer theory,” Information Studies: Theory & Application, vol. 42, no. 8, pp. 67–72, 2019.
[15] P. R. Palanisamy, B. J. Jung, S. H. Sim, and J. H. Yi, “Quasi real-time and continuous non-stationary strain estimation in bottom-fixed offshore structures by multimetric data fusion,” Smart Structures and Systems, vol. 23, no. 1, pp. 61–69, 2019.
[16] L. Martino and J. Read, “A joint introduction to Gaussian processes and relevance vector machines with connections to Kalman filtering and other kernel smoothers,” Information Fusion, vol. 74, pp. 17–38, 2021.
[17] Z. Dong, C. Zheng, and G. Zhang, “Self-adaption of weighted average research for data fusion with different precision,” Ship Electronic Engineering, vol. 34, no. 10, pp. 31–33, 2014.
[18] B. Wang, “Application of Kalman filter algorithm in noises reduction of tunnel structure settlement analysis,” Journal of Beijing University of Civil Engineering and Architecture, vol. 37, no. 3, pp. 64–69, 2021.
[19] L. Feng, “Data processing model of adaptive Kalman filter for high-speed rail settlement observation,” Geomatics & Spatial Information Technology, vol. 43, no. 2, pp. 208–212, 2020.