

METHOD AND SYSTEM FOR LOCATING HIGH-SPEED TRAIN IN NAVIGATION BLIND ZONE BASED ON METEOROLOGICAL PARAMETERS

FIELD OF THE INVENTION

The present invention relates to a locating technology for trains in long and large tunnels, and specifically to a method and system for locating a high-speed train in a navigation blind zone based on meteorological parameters.

BACKGROUND OF THE INVENTION

In recent years, more and more railway lines have been laid to the western region of China. With the changes in topography, the lines inevitably pass through many high mountain regions, and navigation satellite signals cannot be searched in tunnels, which will cause temporary information loss and form navigation blind zones to endanger driving safety. The accurate locating of trains in the tunnels is of great significance to ensure the safety of the trains.

At present, domestic studies on train locating directions in the long and large tunnels (tunnels with a single hole length of more than 10 kilometers), which are typical navigation blind zones, are still in the preliminary stage. In order to avoid blind zones in locating and train operation accidents caused by signal loss, locating devices and methods that are accurate in locating, cost-effective and easy to implement need to be developed. The existing train tunnel locating methods are as follows:

A tunnel navigation information simulation system acquires information through a satellite signal simulator, then generates simulated navigation information and sends same to a target train by optical cables, which can realize continuous navigation and locating simulation in a tunnel, and achieve the purpose of reducing the cost and solving the problems of information loss and time discontinuity in the tunnel. However, a lot of optical cable groups are required, hardware conditions are highly required, and special design is required for different terrain positions in actual applications, so the universality is poor.

A trackside train locating device acquires trackside equipment images during train operation, and realizes the locating of a train through a trackside intelligent identification device combined with an electronic map. The identification accuracy increases with the increase of the frame rate of a camera, but trackside equipment

needs to be arranged along the lines, so the maintenance cost is high.

In addition, there are technologies such as speed measurement calculation type locating and response type locating, which also have problems of low locating accuracy or high maintenance costs. Based on the above, it can be seen that the existing technologies for locating trains in tunnels are difficult to popularize in large region while ensuring high locating accuracy.

SUMMARY OF THE INVENTION

The technical problem to be solved by the present invention is to provide, in response to the shortcomings in the prior art, a method and system for locating a high-speed train in a navigation blind zone based on meteorological parameters, so as to solve the problem that the train is difficult to be located in a long and large tunnel which is a typical navigation blind zone and reduce the locating cost.

In order to solve the above technical problem, the technical solution adopted by the present invention is: a method for locating a high-speed train in a navigation blind zone based on meteorological parameters, including the following steps:

S1, acquiring tunnel meteorological parameters in a tunnel when the train passes to construct a tunnel meteorological parameter database;

S2, classifying, based on the tunnel meteorological parameter database, tunnel groups with similar attributes, to obtain typical tunnel samples of each category of tunnel groups;

S3, constructing a HSV color space template library of a typical sequence by using the typical tunnel samples;

S4, training an HSV template matching model with the HSV color space template library; training a relevance vector machine with data of the same category of tunnel groups to establish a tunnel mileage prediction model;

S5, constructing a fusion model of the HSV template matching model and the tunnel mileage prediction model to obtain a mileage prediction fusion model; and

S6, acquiring tunnel meteorological parameter data during the operation of the train, and calling the mileage prediction fusion model to predict the position of the train.

The present invention establishes a mileage prediction model by means of artificial intelligence big data analysis technology. After the modeling is completed, only on-board sensors are required to acquire input data without any trackside equipment, so the system construction cost and maintenance cost are reduced.

In step S1, the specific process of constructing a tunnel meteorological parameter database includes: acquiring a temperature sequence and a humidity sequence that are acquired when the train passes through a tunnel at one time, longitude and latitude of the region where the tunnel is located, and predicted values of hourly average temperature, hourly average humidity, and hourly average solar radiation to constitute a group of tunnel meteorological parameter samples; wherein the tunnel meteorological parameter samples acquired during the operation of all trains in the region within one year constitute the tunnel meteorological parameter database. The database construction process of the present invention can ensure sufficient sample data to improve the prediction accuracy of the prediction fusion model.

The specific implementation process of step S2 includes:

- a) transforming the latitude and longitude coordinates of the region where the tunnel is located into plane coordinates; normalizing the tunnel meteorological parameters, wherein the processed latitude and longitude coordinates and the tunnel meteorological parameters form a set of input attributes of rough tunnel category classification;
- b) performing output sequencing of tunnel group samples with the input attributes of rough tunnel category classification as objects by means of an OPTICS algorithm, and comparing reachability-distances of samples in the sequence after sequencing with a set neighborhood distance parameter ε , wherein the consecutive samples whose reachability distances are smaller than the set neighborhood distance parameter ε in the sequence constitute a sample cluster; acquiring a cluster center X_o^1 of each sample cluster and T1 samples $\{X_{T1}^1\}, i=1,2,\dots,T1$ closest to the cluster center; wherein $X_i^1 = (x_i^1, x_i^2, x_i^3, \dots, x_i^T)$, x_i^1 and x_i^2 correspond to the processed longitude and latitude, and $x_i^3, \dots, x_i^T$ correspond to the processed tunnel meteorological parameters; defining the cluster center X_o^1 and the T1 samples as representative samples of a current category of tunnel groups, wherein the T1 samples in the representative samples are typical tunnel samples of the current category of tunnel groups.

Through the above process, rough classification of tunnel groups with similar attributes is realized.

The specific implementation process of step S2 includes:

- 1) performing mirror extension on the temperature time sequence and humidity

sequence when the train passes through the tunnel in the sample cluster to transform the temperature sequence and humidity sequence in the sample cluster into sequences whose lengths are equal to respective longest sample lengths;

2) setting a delay time and a window length, and performing phase space reconstruction on the temperature and humidity sequences with respective longest sample lengths by means of a Delay-coordinate method to obtain a two-dimensional reconstruction matrix representing the evolution characteristics of temperature and humidity;

3) inputting the two-dimensional reconstruction matrix into a training autoencoder to obtain a set of depth feature attributes for further classification; and

4) acquiring a cluster center X_0^2 of each sample cluster of secondary clustering and T1 samples closest to the cluster center by means of the set of depth feature attributes as an input according to the process in step b); defining the cluster center X_0^2 and the T1 samples as representative samples of secondary clustering of the current category of tunnel groups, wherein the T1 samples in the representative samples of secondary clustering are typical tunnel samples of the current category of tunnel groups.

The implementation process of step S2 above realizes accurate classification of the tunnel groups, which can ensure that the data of the subsequent HSV color space template library is more accurate, thereby improving the prediction accuracy of the prediction fusion model.

In step S3, the specific implementation process of constructing a HSV color space template library of a typical sequence by using the typical tunnel samples includes: using the temperature and humidity parameter sequences which are in the typical samples and vary with the tunnel mileage as template sequences of the corresponding category of tunnel groups, setting a delay time and a window length, performing phase space reconstruction on the temperature, humidity and temperature difference time sequences by means of the Delay-coordinate method to obtain three two-dimensional reconstruction matrices representing the evolution characteristics of temperature and humidity, and combining the three two-dimensional reconstruction matrices according to an HSV color space to form a color image, that is, a template image in the HSV color space template library of the typical sequence is obtained. This implementation constructs two-dimensional reconstruction matrices through temperature, humidity and temperature difference, so the calculation process is simple.

In step S4, the specific implementation process of training an HSV template matching model includes:

- 1) collecting current sample points in the temperature, humidity and temperature difference time sequences and N sample points in front of the current sample points;
- 2) setting a delay time and a window length, performing phase space reconstruction by means of the Delay-coordinate method to obtain three two-dimensional reconstruction matrices representing the evolution characteristics of temperature and humidity, and combining the three matrices according to the HSV color space to form a feature image of a current position ;
- 3) performing convolution operation on the feature image of the current position and the template image in the HSV color space template library of the typical sequence to obtain a plurality of one-dimensional sequences;
- 4) sorting elements in all the one-dimensional sequences in descending order to determine T2 largest elements as candidate elements, wherein the positions corresponding to the candidate elements before sorting are candidate positions; acquiring mileage values corresponding to the candidate positions; and
- 5) averaging all the mileage values corresponding to all the candidate positions to obtain a current output sequence of the HSV template matching model.

Through the above process, the best matching position of the current position of the train in the template library can be determined, and the prediction accuracy can be further improved.

In step S4, the specific implementation process of training a relevance vector machine to establish a tunnel mileage prediction model includes:

- 1) defining input samples $I = (T, H, t_0, h_0, r_0)$, wherein $T = (t_1, t_2 \cdots, t_M)$ is a temperature time sequence of current sample point and previous M sample points in the tunnel, $H = (h_1, h_2 \cdots, h_{19}, h_M)$ is a humidity sequence of current sample point and previous N sample points in the tunnel, and t_0 , h_0 and r_0 are respectively predicted values of hourly average temperature, hourly average humidity and hourly average solar radiation acquired from a meteorological station; defining an output sample to be a mileage value O corresponding to the current position, wherein input and output combinations $Y = \{I, O\}$ constitute modeling samples; selecting M modeling samples for each tunnel group of the same category;
- 2) randomly dividing the M modeling samples into a training set, a verification set

and a test set;

3) performing binary coding on the feature of each dimension in the input samples I ; when the coded value corresponding to the feature of a certain dimension is 1, selecting the feature as an input variable of the RVM model; when the coded value of the feature of a certain dimension is 0, discarding the feature of this dimension;

4) determining new input features based on the coded values of the current features, updating the training set, the verification set and the test set, training the RVM model with data of the updated training set, and inputting data of the updated verification set into the trained RVM model to obtain a model output sequence $O_k^v, k=1, 2 \cdots M_1$;

wherein $M_1=0.3M$; and

5) repeating step 3) and step 4) to determine optimal input features and RVM model that minimizes an optimization objective function $f = \sum_{k=1}^{M_1} (O_k^t - O_k^v)^2$, wherein the RVM model is the tunnel mileage prediction model; wherein O_k^t is a real output value in the verification set.

A relevance vector machine (RVM) is used to obtain the tunnel mileage prediction model, which can improve the prediction accuracy of the tunnel mileage prediction model.

The fusion model of the HSV template matching model and the tunnel mileage prediction model is $O_{k1}^F = a_R O_{k1}^R + a_M O_{k1}^M$; wherein $k1=1, 2 \cdots M_2$, $M_2=0.1M$;

$a_R = \frac{E^M}{E^R + E^M}$; $a_M = \frac{E^R}{E^R + E^M}$; $E^R = \sum_{k1=1}^{M_2} |O_{k1}^M - O_{k1}^E|$; $E^M = \sum_{k1=1}^{M_2} |O_{k1}^M - O_{k1}^E|$; O_{k1}^R is an output sequence obtained after the data of the test set is input into the tunnel mileage prediction model; O_{k1}^E is a real output result in the test set; and O_{k1}^M is an output sequence obtained after the test set is input into the HSV template matching model.

In step S6, the specific implementation process of predicting the position of the train includes: calculating input vectors of the RVM model of the current sample points, and substituting the input vectors of the RVM model into a target model to obtain an output value of the tunnel mileage prediction model; acquiring input values of the HSV template matching model of the current sample points, and substituting same into the HSV template matching model to obtain an output value of the HSV template matching model; substituting the output value of the mileage prediction model and the output value of the HSV template matching model into the mileage prediction fusion

model to obtain a final train position prediction result; wherein the target template refers to a model trained under a secondary clustering target sample cluster; the secondary clustering target sample cluster refers to a sample cluster corresponding to minimum value between the current sample point and the secondary clustering representative samples subordinate to all primary clustering target sample clusters; the primary clustering target sample clusters refers to sample clusters corresponding to minimum value between the current sample point and all clustering representative samples; the clustering representative samples refer to samples in the sample cluster, and the sample cluster refers to a sample cluster composed of consecutive samples in a sequence obtained by performing output sequencing of tunnel group samples with the input attributes of rough tunnel category classification as objects and by means of an OPTICS algorithm, value of the sample in the consecutive samples which is between the reachability-distance and the set neighborhood distance parameter ε is smaller than the neighborhood distance parameter ε .

Correspondingly, the present invention further provides a system for locating a high-speed train in a navigation blind zone based on meteorological parameters, including sensors mounted on the train for acquiring meteorological parameters in a tunnel; the sensors communicate with computer equipment; and the computer equipment is programmed or configured to execute the steps of the method according to the present invention.

Compared with the prior art, the present invention has the following beneficial effects:

1. The present invention makes full use of artificial intelligence big data analysis technology, fully mines the potential law of changes of tunnel environment parameters with tunnel depth, and solves the problem of difficult train locating in a long and large tunnel which is a typical navigation blind zone from the perspective of data-driven modeling;
2. After the modeling is completed, the present invention only needs the on-board temperature and humidity sensors to acquire input data without any trackside equipment, thereby reducing the system construction cost and maintenance cost.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a flowchart of data acquisition and rough category classification;

Fig. 2 is a flowchart of secondary category classification and modeling;

Fig. 3 shows a time sequence, phase space reconstruction and HSV color space

combination;

Fig. 4 is a flowchart of model calling during test.

DETAILED DESCRIPTION OF EMBODIMENTS

Embodiment 1 of the present invention provides a method for locating a high-speed train in a navigation blind zone, which specifically includes the following steps:

Step 1: acquisition of tunnel meteorological parameters, and construction of a tunnel meteorological parameter database

Temperature, humidity and mileage time sequences in a tunnel when the train passes are acquired in real time by on-board sensors, with a sampling interval of 0.1 s. Latitude and longitude of a current position, and predicted values of hourly average temperature, hourly average humidity, and hourly average solar radiation are acquired from a meteorological station in the region where the tunnel is located. The temperature sequence and humidity sequence acquired when the train passes through the tunnel at one time, the latitude and longitude of the region, and the predicted values of hourly average temperature, hourly average humidity, and hourly average solar radiation constitute a group of tunnel meteorological parameter samples. The tunnel meteorological parameter samples acquired during the operation of all trains in the region within one year constitute the tunnel meteorological parameter database.

Step 2: multi-scale hierarchical classification of the tunnel meteorological parameters

Based on the tunnel meteorological parameter database, category classification of multiple types of attributes, multiple feature scales, and multiple levels is performed using the temperature sequence and humidity sequence acquired when passing through a tunnel, the latitude and longitude of the region, and the predicted values of hourly average temperature, hourly average humidity, and hourly average solar radiation (respectively corresponding to hour-over-hour average temperature, hour-over-hour average humidity, and hour-over-hour average solar radiation, that is, hourly average values) as information input, to realize the classification of tunnel groups with similar attributes. As shown in Fig. 2 and Fig. 3, the specific implementation process is as follows:

Step A1: latitude and longitude coordinates are transformed into plane coordinates by means of a spherical coordinate conversion relationship. The transformed longitude and latitude, hourly average temperature, hourly average humidity, and hourly average solar radiation are normalized to form a set of input attributes of rough tunnel

category classification.

Step A2: an initial neighborhood distance parameter ε is set to 0.1, an initial neighborhood sample number parameter $MinPts$ is set to 5, and output sequencing of tunnel group samples is performed with the input attributes of rough tunnel category classification as objects by means of an OPTICS (Ordering Points To Identify the Clustering Structure) algorithm. Reachability-distances (Reachability-distance, a concept defined by the OPTIC algorithm, is a minimum value of a core distance and an Euclidean distance, see <https://blog.csdn.net/han1202012/article/details/105936710>) of samples in the sequence are compared with the set neighborhood distance parameter ε , and consecutive samples whose reachability distances are smaller than a set value in the sequence constitute a sample cluster. A cluster center X_0^1 of each sample cluster and 5 samples $\{X_i^1\}, i=1,2,\dots,5$ closest to the cluster center are acquired. $X_i^1 = (x_i^1, x_i^2, x_i^3, x_i^4, x_i^5)$ corresponds to the five attributes including the transformed longitude, latitude, hourly average temperature, hourly average humidity, and hourly average solar radiation. The cluster center X_0^1 and the 5 samples $\{X_i^1\}, i=1,2,\dots,5$ are defined as representative samples of a current category.

Step A3: each sample cluster obtained based on the set of input attributes of rough tunnel category classification is further classified. Specifically, step A3 includes the following sub-steps: ① Mirror extension is performed on the temperature time sequence and humidity sequence when the train passes through the tunnel in the sample cluster to transform the temperature sequence and humidity sequence in the sample cluster into sequences whose lengths are equal to respective longest sample lengths. ② A delay time is set to 1, a window length is set to 5, and phase space reconstruction is performed on the temperature sequence and humidity sequence by means of a Delay-coordinate method to obtain a two-dimensional reconstruction matrix representing the evolution characteristics of temperature and humidity. ③ The reconstruction matrix of the samples is input into a training autocoder to obtain a set of depth feature attributes for further classification. ④ A cluster center X_0^2 of each sample cluster of secondary clustering and 5 samples $\{X_i^2\}, i=1,2,\dots,5$ closest to the cluster center are acquired with the set of depth feature attributes as an input according to the process in step A2. $X_i^2 = (y_i^1, y_i^2, y_i^3, y_i^4, y_i^5)$ corresponds to 5 dimensional variables in the set of depth feature attributes. The cluster center X_0^2 and

the 5 samples $\{X_i^2\}, i=1,2,\dots,5$ are defined as representative samples of secondary clustering of the current category.

Step 3: construction of a HSV (Hue, Saturation, Value) color space template library of a typical sequence

The 5 tunnel samples $\{X_i^2\}, i=1,2,\dots,5$ closest to the cluster center corresponding to the group in each category of tunnel groups are used as typical tunnel samples under this category. As shown in Fig. 3, the temperature and humidity parameter sequences varying with the tunnel mileage in the typical samples are used as template sequences of this category, the delay time is set to 1, the window length is 5, phase space reconstruction is performed on the temperature, humidity and temperature difference time sequences by means of the Delay-coordinate method to obtain three two-dimensional reconstruction matrices representing the evolution characteristics of temperature and humidity, and the three matrices are combined according to an HSV color space to form a color image, that is, a template image $f_i, i=1,2,\dots,5$.

Step 4: training of an HSV template matching model

Based on current position sample point and previous sample sequences, phase space reconstruction and HSV color space combination are performed to construct a feature image of the current position. Correlations between a feature module of the current position and images in the template library are calculated to determine the best matching position of the current position in the template library. This process includes the following steps:

Step B1: current sample point and previous 19 sample points in the temperature, humidity and temperature difference time sequences are collected.

Step B2: the delay time is set to 1, the window length is set to 5, phase space reconstruction is performed by means of the Delay-coordinate method to obtain three two-dimensional reconstruction matrices representing the evolution characteristics of temperature and humidity, and the three matrices are combined according to the HSV color space to form a feature image of a current position h .

Step B3: convolution operation $g_i = f_i \otimes h, i=1,2,\dots,5$ is performed on the feature image of the current position and the images in the template library, wherein each g_i is a one-dimensional sequence.

Step B4: elements in all the g_i sequences are sorted in descending order, to determine 5 largest elements as candidate elements, wherein the positions

corresponding to the candidate elements before sorting are candidate positions, and the mileage values corresponding to the candidate positions are $l_j, j=1,2,\dots,5$.

Step B5: an average value of the mileage values corresponding to the candidate positions is determined as a current output value of template matching model, that is, a output value of the template matching model is $O^M = \text{mean}(l_j), j=1,2,\dots,5$.

Step 5: training of an RVM recognition model

A relevance vector machine (RVM) is trained with the data of the same category of tunnel groups, the temperature sequence, humidity sequence, hourly average temperature, hourly average humidity, and hourly average solar radiation as input, and the mileage data in the current tunnel as output, to establish a tunnel mileage prediction model. This process includes the following steps:

Step C1: training samples are defined, and input samples $I = (T, H, t_0, h_0, r_0)$ are defined, wherein $T = (t_1, t_2, \dots, t_{19}, t_{20})$ is a temperature time sequence of a current sample point and previous 19 sample points in the tunnel, $H = (h_1, h_2, \dots, h_{19}, h_{20})$ is a humidity sequence of a current sample point and previous 19 sample points in the tunnel, and t_0, h_0 and r_0 are respectively predicted values of hourly average temperature, hourly average humidity and hourly average solar radiation acquired from the meteorological station. The output sample is a mileage value O corresponding to the current position. Input and output combinations $Y = \{I, O\}$ constitute modeling samples. For each tunnel group of the same category formed by secondary clustering, M (M is 5000 in the present invention) samples are selected to establish the mileage prediction model.

Step C2: training samples, verification samples, and test samples are classified. By means of random sampling without replacement, 60% of the M samples are selected to form a training set, 30% of the M samples are selected to form a verification set, and 10% of the M samples are selected to form a test set.

Step C3: optimized objects are determined, and optimized values are initialized. The input features of the model are optimized by means of a binary whale algorithm, that is, binary coding is performed on the feature of each dimension in the input samples I . When the coded value corresponding to the feature of a certain dimension is 1, the feature is selected as an input variable of the RVM model. When the coded value corresponding to the feature of a certain dimension is 0, the feature of this dimension is discarded. 43 dimensional features are randomly initialized and coded as 0 or 1.

Step C4: an optimization objective function is determined. Input features are determined based on the coded values of the current features, and the RVM model is trained with data of the training set. Data of the verification set is input to the trained RVM model to obtain a model output sequence $O_k^v, k=1, 2 \cdots M_1$, wherein $M_1=0.3M$.

The optimization objective function is defined as:

$$f = \sum_{k=1}^{M_1} (O_k^t - O_k^v)^2 \quad (1)$$

In the formula, $O_k^t, k=1, 2 \cdots M_1$ is a real output value of the verification set.

Step C5: the optimized prediction model is output. Iterative optimization operations are performed by means of the binary whale algorithm to determine optimal input features and RVM model, the RVM model is the RVM mileage prediction model.

Step 6: construction of an HSV template matching and RVM recognition fusion model

Data of the test set is substituted into the RVM mileage prediction model to obtain a model output sequence as $O_k^R, k=1, 2 \cdots M_2$, wherein $M_2=0.1M$. Taking the data of the test set as input, according to the operation process in step 4, a output result of template matching model obtained is $O_k^M, k=1, 2 \cdots M_2$, while the real output result in the data of the test set is $O_k^E, k=1, 2 \cdots M_2$. The error between the output value of the RVM model and the real value is calculated as follows:

$$E^R = \sum_{i=1}^{M_2} |O_k^M - O_k^E| \quad (2)$$

The error between the output result of template matching model and the real value is calculated as follows:

$$E^M = \sum_{i=1}^{M_2} |O_k^M - O_k^E| \quad (3)$$

Distances between one or more current template samples and the current samples are simultaneously calculated. Then, a model fusion coefficient of the RVM model is defined as:

$$a_R = \frac{E^M}{E^R + E^M} \quad (4)$$

A model fusion coefficient of the template matching model is defined as:

$$a_M = \frac{E^R}{E^R + E^M} \quad (5)$$

The final output result of the model is:

$$O_k^F = a_R O_k^R + a_M O_k^M \quad (6)$$

Step 7: acquisition of input data and calling of the mileage prediction fusion model

During the operation of the train, current temperature, air pressure and solar radiation data outside the tunnel are acquired from the meteorological station in the region. Current temperature and humidity sequences are acquired by means of temperature and humidity sensors mounted at the head and tail of the train. This process includes the following steps:

Step D1: input attributes X_t^1 for primary clustering are acquired with reference to the process of step 2, wherein $X_t^1 = (x_t^1, x_t^2, x_t^3, x_t^4, x_t^5)$ corresponds to 5 attributes including the transformed longitude, latitude, hourly average temperature, hourly average humidity, and hourly average solar radiation. Input attributes X_t^2 for secondary clustering are acquired with reference to the process of step 2, wherein $X_t^2 = (y_t^1, y_t^2, y_t^3, y_t^4, y_t^5)$ corresponds to 5 attributes in the set of depth feature attributes.

Step D2: distances between the feature values X_t^1 of the current sample points and the primary clustering representative samples are calculated as follows:

$$d_1 = \sum_{i=0}^5 (X_t^1 - X_0^1)^2 \quad (7)$$

The sample cluster corresponding to the minimum value between the current sample point and all the clustering representative samples is selected as a primary clustering target sample cluster.

Step D3: distances between the feature values X_t^2 of the current sample points and the secondary clustering representative samples are calculated as follows:

$$d_2 = \sum_{i=0}^5 (X_t^2 - X_0^2)^2 \quad (8)$$

The sample cluster corresponding to the minimum value between the current sample point and the secondary clustering representative samples subordinate to all the primary clustering target sample clusters is selected as a secondary clustering target sample cluster. The model and template library trained under this sample cluster are a target model and a target template library.

Step 8: Prediction of the position of the train

Input vectors of the RVM model of the current sample points are calculated with reference to the process of step 5, the input vectors of the RVM model are substituted

into the target model to obtain output values of target RVM model. Input values of the template matching model of the current sample points are acquired with reference to the process of step 4, and substituted into the target template matching model to obtain target template matching model output values. A final train position prediction result is obtained with reference to equation 6.

Embodiment 2 of the present invention provides a system for locating a high-speed train in a navigation blind zone based on meteorological parameters, including sensors mounted on the train for acquiring meteorological parameters in a tunnel; the sensors communicate with computer equipment; and the computer equipment is programmed or configured to execute the steps of the method according to Embodiment 1 of the present invention.

Claims

1. A method for locating a high-speed train in a navigation blind zone based on meteorological parameters, comprising the following steps:

S1, acquiring tunnel meteorological parameters in tunnels when the train passes to construct a tunnel meteorological parameter database;

S2, classifying, based on the tunnel meteorological parameter database, tunnel groups with similar attributes, to obtain typical tunnel samples of each category of tunnel groups;

S3, constructing a HSV color space template library of a typical sequence by using the typical tunnel samples; training a relevance vector machine with data of the same category of tunnel groups to establish a tunnel mileage prediction model;

S4, training an HSV template matching model with the HSV color space template library;

S5, constructing a fusion model of the HSV template matching model and the tunnel mileage prediction model to obtain a mileage prediction fusion model; and

S6, acquiring tunnel meteorological parameter data during the operation of the train, and calling the mileage prediction fusion model to predict the position of the train.

2. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to claim 1, wherein in step S1, the specific process of constructing a tunnel meteorological parameter database comprises: acquiring a temperature sequence and a humidity sequence that are acquired when the train passes through a tunnel at one time, longitude and latitude of the region where the tunnel is located, and predicted values of hourly average temperature, hourly average humidity, and hourly average solar radiation to constitute a group of tunnel meteorological parameter samples; wherein the tunnel meteorological parameter samples acquired during the operation of all trains in the region within one year constitute the tunnel meteorological parameter database.

3. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to claim 1, wherein the specific implementation process of step S2 comprises:

a) transforming the latitude and longitude coordinates of the region where the tunnel is located into plane coordinates; normalizing the tunnel meteorological parameters,

wherein the processed latitude and longitude coordinates and the tunnel meteorological parameters form a set of input attributes of rough tunnel category classification;

b) performing output sequencing of tunnel group samples with the input attributes of rough tunnel category classification as objects by means of an OPTICS algorithm, and comparing reachability-distances of samples in the sequence after sequencing with a set neighborhood distance parameter ε , wherein the consecutive samples whose reachability distances are smaller than the set neighborhood distance parameter ε in the sequence constitute a sample cluster; acquiring a cluster center X_0^1 of each sample cluster and T1 samples $\{X_{T1}^1\}, i=1,2,\dots,T1$ closest to the cluster center; wherein $X_i^1 = (x_i^1, x_i^2, x_i^3, \dots, x_i^T)$, x_i^1 and x_i^2 correspond to the processed longitude and latitude, and $x_i^3, \dots, x_i^T$ correspond to the processed tunnel meteorological parameters; defining the cluster center X_0^1 and the T1 samples as representative samples of a current category of tunnel groups, wherein the T1 samples in the representative samples are typical tunnel samples of the current category of tunnel groups.

4. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to claim 3, wherein the specific implementation process of step S2 comprises:

- 1) performing mirror extension on the temperature time sequence and humidity sequence when the train passes through the tunnel in the sample cluster to transform the temperature sequence and humidity sequence in the sample cluster into sequences whose lengths are equal to respective longest sample lengths;
- 2) setting a delay time and a window length, and performing phase space reconstruction on the temperature and humidity sequences with respective longest sample lengths by means of a Delay-coordinate method to obtain a two-dimensional reconstruction matrix representing the evolution characteristics of temperature and humidity;
- 3) inputting the two-dimensional reconstruction matrix into a training autocoder to obtain a set of depth feature attributes for further classification; and
- 4) acquiring a cluster center X_0^2 of each sample cluster of secondary clustering and T1 samples closest to the cluster center by means of the set of depth feature attributes

as an input according to the process in step b); defining the cluster center X_0^2 and the T1 samples as representative samples of secondary clustering of the current category of tunnel groups, wherein the T1 samples in the representative samples of secondary clustering are typical tunnel samples of the current category of tunnel groups.

5. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to any one of claims 1-4, wherein in step S3, the specific implementation process of constructing a HSV color space template library of a typical sequence by using the typical tunnel samples comprises: using the temperature and humidity parameter sequences which are in the typical samples and vary with the tunnel mileage as template sequences of the corresponding category of tunnel groups, setting a delay time and a window length, performing phase space reconstruction on the temperature, humidity and temperature difference time sequences by means of the Delay-coordinate method to obtain three two-dimensional reconstruction matrices representing the evolution characteristics of temperature and humidity, and combining the three two-dimensional reconstruction matrices according to an HSV color space to form a color image, that is, a template image in the HSV color space template library of the typical sequence is obtained.

6. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to any one of claims 1-5, wherein in step S4, the specific implementation process of training an HSV template matching model comprises:

- 1) collecting current sample points in the temperature, humidity and temperature difference time sequences and N sample points in front of the current sample points;
- 2) setting a delay time and a window length, performing phase space reconstruction by means of the Delay-coordinate method to obtain three two-dimensional reconstruction matrices representing the evolution characteristics of temperature and humidity, and combining the three matrices according to the HSV color space to form a feature image of a current position ;
- 3) performing convolution operation on the feature image of the current position and the template image in the HSV color space template library of the typical sequence to obtain a plurality of one-dimensional sequences;

- 4) sorting elements in all the one-dimensional sequences in descending order to determine T2 largest elements as candidate elements, wherein the positions corresponding to the candidate elements before sorting are candidate positions; acquiring mileage values corresponding to the candidate positions; and
- 5) averaging all the mileage values corresponding to all the candidate positions to obtain a current output sequence of the HSV template matching model.

7. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to any one of claims 1-6, wherein in step S4, the specific implementation process of training a relevance vector machine to establish a tunnel mileage prediction model comprises:

- 1) defining input samples $I = (T, H, t_0, h_0, r_0)$, wherein $T = (t_1, t_2 \cdots, t_M)$ is a temperature time sequence of current sample point and previous M sample points in the tunnel, $H = (h_1, h_2 \cdots, h_{19}, h_M)$ is a humidity sequence of current sample point and previous N sample points in the tunnel, and t_0 , h_0 and r_0 are respectively predicted values of hourly average temperature, hourly average humidity and hourly average solar radiation acquired from a meteorological station; defining an output sample to be a mileage value O corresponding to the current position, wherein input and output combinations $Y = \{I, O\}$ constitute modeling samples; selecting M modeling samples for each tunnel group of the same category;
- 2) randomly dividing the M modeling samples into a training set, a verification set and a test set;
- 3) performing binary coding on the feature of each dimension in the input samples I ; when the coded value corresponding to the feature of a certain dimension is 1, selecting the feature as an input variable of the RVM model; when the coded value of the feature of a certain dimension is 0, discarding the feature of this dimension;
- 4) determining new input features based on the coded values of the current features, updating the training set, the verification set and the test set, training the RVM model with data of the updated training set, and inputting data of the updated verification set into the trained RVM model to obtain a model output sequence $O_k^v, k = 1, 2 \cdots M_1$; wherein $M_1 = 0.3M$; and
- 5) repeating step 3) and step 4) to determine optimal input features and RVM model that minimizes an optimization objective function. $f = \sum_{k=1}^{M_1} (O_k^t - O_k^v)^2$, wherein the

RVM model is the tunnel mileage prediction model; wherein O_k^t is a real output value in the verification set.

8. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to claim 7, wherein the fusion model of the HSV template matching model and the tunnel mileage prediction model is

$$O_{k1}^F = a_R O_{k1}^R + a_M O_{k1}^M; \text{ wherein } k1=1, 2 \dots M_2, \quad M_2 = 0.1M; \quad a_R = \frac{E^M}{E^R + E^M}; \quad a_M = \frac{E^R}{E^R + E^M};$$

$$E^R = \sum_{k1=1}^{M_2} |O_{k1}^M - O_{k1}^E|; \quad E^M = \sum_{k1=1}^{M_2} |O_{k1}^M - O_{k1}^E|; \quad O_{k1}^R \text{ is an output sequence obtained after the data}$$

of the test set is input into the tunnel mileage prediction model; O_{k1}^E is a real output result in the test set; and O_{k1}^M is an output sequence obtained after the test set is input into the HSV template matching model.

9. The method for locating the high-speed train in a navigation blind zone based on meteorological parameters according to any one of claims 1-8, wherein in step S6, the specific implementation process of predicting the position of the train comprises: calculating input vectors of the RVM model of the current sample points, and substituting the input vectors of the RVM model into a target model to obtain an output value of the tunnel mileage prediction model; acquiring input values of HSV template matching model of the current sample points, and substituting same into the HSV template matching model to obtain an output value of the HSV template matching model; substituting the output value of the mileage prediction model and the output value of the HSV template matching model into the mileage prediction fusion model to obtain a final train position prediction result; wherein the target template refers to a model trained under a secondary clustering target sample cluster; the secondary clustering target sample cluster refers to a sample cluster corresponding to minimum value between the current sample point and the secondary clustering representative samples subordinate to all primary clustering target sample clusters; the primary clustering target sample clusters refers to sample clusters corresponding to minimum value between the current sample point and all clustering representative samples; the clustering representative samples refers to samples in the sample cluster, and the sample cluster refers to a sample cluster composed of consecutive samples in a sequence obtained by performing output sequencing of tunnel group samples with

the input attributes of rough tunnel category classification as objects and by means of an OPTICS algorithm, value of the sample in the consecutive samples which is between the reachability-distance and the set neighborhood distance parameter ε is smaller than the neighborhood distance parameter ε .

10. A system for locating a high-speed train in a navigation blind zone based on meteorological parameters, comprising sensors mounted on the train for acquiring meteorological parameters in a tunnel; the sensors communicate with computer equipment; and the computer equipment is programmed or configured to execute the steps of the method according to any one of claims 1-9.