METHOD, SYSTEM AND STORAGE MEDIUM FOR CALCULATING TUNNEL PASSING DURATION OF TRAIN

Field of the Invention

The present invention relates to the field of artificial intelligence learning, and specifically to a method, system and a storage medium for calculating the tunnel passing duration of a train.

Background of the Invention

Improving the passenger experience of passenger trains throughout the journey is conducive to enhancing the competitiveness of high-speed rail passenger transport. Comprehensive information prompts are conducive to allowing passengers to grasp the current train operation status and improving the passenger experience. At present, the prompt information in a train includes temperature, train speed and so on, but lacks relevant information prompts for special conditions such as the train passing through a tunnel.

After the train enters the tunnel, communication information may be lost, so that passengers in the Internet age are at loose ends, and the comfortable sensation of passengers declines rapidly. In addition, the dark airtight environment in the tunnel, the eardrum discomfort caused by the deterioration of airtightness of the train, and the unknown length of the tunnel cause great anxiety. For this reason, effective and accurate positioning methods for tunnel passing durations are designed to timely inform passengers of the duration they need to wait in the tunnel, which is very important for improving the passenger experience. The core of calculating the tunnel passing duration of the train is accurate positioning of the train in the tunnel. The common train positioning methods are as follows:

1. Beidou/GPS positioning type, the Beidou satellite positioning can provide all-round, all-meteorological, and all-day positioning information, but is not suitable for train positioning in the tunnel.

2. Wireless base station type, wireless base stations at both ends of a tunnel provide train information, which effectively reduces the use of trackside equipment along the line, but is not suitable for train positioning in the tunnel.

3. Responding type, a plurality of trackside equipment is arranged along the railway, and corresponding on-board equipment is mounted on the train, which has better

accuracy, but the engineering cost and maintenance cost are huge, so the practicability is poor.

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, system and a storage medium for calculating the tunnel passing duration of a train, so as to realize intelligent perception of the distance between the train in a tunnel and the exit of the tunnel, and to accurately estimate the running time of the train in the tunnel.

In order to solve the above technical problem, the technical solution adopted by the present invention is: a method for calculating the tunnel passing duration of a train, the remaining duration that the train passes through the current tunnel is estimated by using the following equation: $t=O(i)/f_{Hz}(O(i)-O(i-1)))$; wherein the mileage value from the current position of the train to the tunnel exit is $O(i)= O^{M}(i)+\varepsilon(i)$; $O(i-1)$ is the mileage value from the last position of the current position of the train to the tunnel exit; $O^{M}(i)$ is the best matching position of the current position in a typical sample joint template library, which is obtained after the correlations between current position feature images and images in the typical sample joint template library are calculated, the current position feature images are obtained by phase space reconstruction and RGB color space combination of acquired sequences of meteorological parameters of sample points at the head and tail of the train at the current position and meteorological parameters with a period of time ahead of the current position, and the template images in the typical sample joint template library are color images about the evolution of meteorological parameters at the head and tail of the train, and the meteorological parameters at the head and tail of the train are obtained by phase space reconstruction of typical samples of each category of tunnel groups; the typical samples of each category of tunnel groups refer to samples obtained by classifying the meteorological parameters according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel and further classifying the classified homogeneous samples in a region; $\varepsilon(i)$ is a compensation error obtained by inputting the meteorological parameters acquired during the operation of the train into a tunnel mileage prediction error compensation model, and the tunnel mileage prediction error compensation model refers to a model obtained by training a least squares support vector machine using data of homogeneous tunnel groups, with a temperature sequence,

a humidity sequence, and an air pressure sequence as input, and a prediction error of an RGB color joint matching model as output; and f_{Hz} is a position update frequency.

The present invention realizes intelligent perception of the distance between the train in the tunnel and the tunnel exit by using artificial intelligence big data analysis technology through temperature sequence, humidity sequence, and air pressure sequence data at both ends of the train, and the present invention does not require any trackside equipment, so the cost is low and the practicability is strong.

The specific determination process of the final output value $O(i)$ of the distance between the current position of the train and the tunnel exit includes:

1) acquiring tunnel meteorological parameters to construct a tunnel meteorological parameter database;

2) classifying, based on the tunnel meteorological parameter database, the meteorological parameters in the tunnel meteorological parameter database according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel to obtain homogeneous samples in a region, so as to realize the classification of tunnel groups with similar attributes;

3) further classifying the homogeneous samples in the region to obtain typical samples of each category of tunnel groups;

4) performing phase space reconstruction on the typical samples of each category of tunnel groups to obtain color images about the evolution of barometric and meteorological parameters at the head and tail of the train, that is, template images; wherein the template images corresponding to all the typical samples constitute the typical sample joint template library;

5) collecting sample points at the head and tail of the train at the current position and sample sequences with a period of time ahead the current position respectively, and performing phase space reconstruction and RGB color space combination to construct current position feature images $\{c_h, c_f\}$; calculating the correlations between the current position feature images and the template images in the typical sample joint template library to determine the best matching position of the current position in the typical sample joint template library;

6) training a least squares support vector machine using data of the homogeneous tunnel groups, with a temperature sequence, a humidity sequence, and an air pressure sequence as input, and a prediction error of the best matching position as output, to establish the tunnel mileage prediction error compensation model; and

7) acquiring meteorological parameters during the operation of the train in real time, and obtaining a final mileage value to the tunnel exit by using the tunnel mileage prediction error compensation model, that is, obtaining $O(i)$.

The method of the present invention can make full use of the changes of meteorological parameters in the tunnel to realize intelligent perception on the position of the train, does not require trackside equipment, and therefore, has the advantages of simplicity and practicality.

The specific implementation process of step 2) includes:

1) classifying the tunnel groups into N categories according to the geographical position of the tunnel samples; and

2) acquiring average temperature distributions within T min before the train enters the tunnel for one year in the current region, fitting the average temperature distributions in the tunnel of the current region by using a Gaussian distribution function, dividing the fitted average temperature distributions into K equal parts according to the probability, defining the temperature samples that belong to the same part as homogeneous temperature samples in the current region, and defining the temperature sequence, humidity sequence and air pressure sequence corresponding to the homogeneous temperature samples as homogeneous samples.

The present invention uses the geographical position to roughly classify tunnels, and then uses meteorological parameters to finely classify the tunnels in the same region, thereby ensuring the reliability of the template samples.

The specific implementation process of step 3) includes:

1) creating an autoregressive integrated moving average model for each time sequence of the homogeneous samples in the current region, and extracting characteristic quantities of each time sequence, that is, parameters of autoregressive terms, integrated terms and moving regression terms of each time sequence; wherein all the characteristic quantities of the homogeneous samples in the region constitute a characteristic matrix A; wherein the time sequence include a temperature sequence, a humidity sequence and an air pressure sequence;

2) reducing dimensions of the characteristic matrix A, and selecting M principal components with the greatest contribution to characterize the information of the original characteristic matrix , thus obtaining a transformed matrix A';

3) defining a kernel function $k=\alpha k_{rbf}+\beta k_{linear}+(1-\alpha-\beta)k_{laplace}$; wherein k_{rbf} is a radial basis kernel function, k_{linear} is a linear kernel, and $k_{laplace}$ is a Laplace kernel function;

randomly determining initial values of the coefficients α, β and the number n_0 of categories of the kernel function, wherein $\alpha, \beta \in [0, 1]$, and n_0 is a positive integer less than 30;

4) mapping the characteristic values in the matrix A' to a characteristic space corresponding to the kernel function k to obtain a new characteristic matrix A'', according to the initial number n_0 of categories, using the characteristic quantities in A'' as input to cluster characteristic quantities in A'' by means of a k-means clustering algorithm to classify the homogeneous samples in the region into n categories, wherein each category of samples constitute a sample cluster; calculating an optimization objective function $fitness=\frac{1}{n}\sum_{i=1}^{n}\max_{i\neq j}\left(\frac{avg(C_i)+avg(C_j)}{d_{cen}(C_i,C_j)}\right)$, wherein $avg(C_i)$ is an average distance of samples in the sample cluster C_i, $avg(C_j)$ is an average distance of samples in the sample cluster C_j, and $d_{cen}(C_i, C_j)$ is a distance between center points of the sample cluster C_i and the sample cluster C_j;

5) updating, according to the current optimization objective function value *fitness*, the values of α, β and n_0 by gray wolf optimization;

6) repeating step 4) and step 5) until the set number m=100 of optimizations is reached, wherein the values of α, β and n_0 at this time are final optimized parameter values, and the clustering results under the final optimized parameter values are final clustering results; and

7) classifying the homogeneous samples in the current region into E sample clusters according to the final clustering results, and acquiring a cluster center of each sample cluster and an original time sequence $\{X_i^n\}$ (i=0, 1, 2…P, n=0, 1, 2…E) corresponding to P samples closest to the cluster center, wherein $X_i^n=(x_i^1, x_i^2, x_i^3, x_i^4, x_i^5, x_i^6)_n$ corresponds to a temperature sequence acquired at the head of the train, a temperature sequence acquired at the head of the train, an air pressure sequence acquired at the head of the train, a temperature sequence acquired at the tail of the train, a temperature sequence acquired at the tail of the train, and an air pressure sequence acquired at the tail of the train, respectively; and defining an original time sequence corresponding to (P+1) samples as a typical sample of the current category of tunnel groups.

The present invention determines typical samples through an adaptive method, and optimizes the feature mapping space and feature clustering parameters, which is beneficial to obtaining more representative typical samples and guarantees the

effectiveness of templates.

The specific implementation process of step 5) includes:

1) determining current sample points and previous G sample points in the temperature sequence, humidity sequence and air pressure sequence of the head and tail of the train at the current position;

2) performing phase space reconstruction by means of the delay coordinate method to obtain six two-dimensional reconstruction matrices representing the evolution characteristics of temperature, humidity and air pressure at the head and tail of the train, dividing the six matrices into matrices corresponding to the head of the train and matrices corresponding to the tail of the train, and combining the six matrices according to the RGB color space to form current position feature images $\{c_h, c_f\}$;

3) performing convolution operation on the current position feature images and the images in the typical sample joint template library to obtain a plurality of one-dimensional sequences;

4) sorting elements in all the one-dimensional sequences from large to small, to determine L largest elements as candidate elements, wherein the positions corresponding to the candidate elements before sorting are candidate positions, and the mileage values to the tunnel exit corresponding to the candidate positions are s_j (j=1, 2...L); and

5) averaging the mileage values to the tunnel exit corresponding to the candidate positions, wherein the average value is the best matching position O^M: $O^M = \frac{1}{L}\sum_{j=1}^{L} s_j$ of the current position in the typical sample joint template library.

The present invention transforms one-dimensional time sequences into two-dimensional images by phase space reconstruction, which is beneficial to mining detailed information in the time sequences, and is beneficial to accurate template matching of a plurality of time sequences.

The specific implementation process of step 6) includes:

1) defining input samples $I=(T_h, H_h, P_h, T_f, H_f, P_f)$, wherein $T_h=(t_1, t_2 \ldots t_G)$ is a temperature sequence of current sample points and G previous sample points at the head of the train in the tunnel, H_h is a train head humidity sequence with a length of G, P_h is a train head air pressure sequence with a length of G, T_f is a train tail temperature sequence with a length of G, H_f is a train tail humidity sequence with a length of G, and P_f is a train tail air pressure sequence with a length of G; defining output samples as

mileage error values ε corresponding to the current position; wherein the combinations $Y=\{I, \varepsilon\}$ of the input samples and the output samples constitute modeling samples;

2) classifying the modeling samples into a training set and a validation set;

3) performing binary coding on the feature of each dimension in the input samples I, wherein when the coded value corresponding to the feature of a certain dimension is 1, the feature is selected as an input variable of the least squares support vector machine; when the coded value corresponding to the feature of a certain dimension is 0, the feature of this dimension is discarded;

4) updating the input samples based on the coded values of the current features to obtain a updated training set and validation set, and training the least squares support vector machine with data of the updated training set; inputting data of the updated validation set into the trained least squares support vector machine to obtain an output sequence ε_k^v of the least squares support vector machine, wherein k=1, 2..., N_1, and N_1 is the number of sample pairs in the validation set; and

5) repeating step 3) and step 4) to determine input features and a least squares support vector machine that minimize the optimization objective function $f=\sum_{k=1}^{N_1}(\varepsilon_k^t-\varepsilon_k^v)^2$, that is, obtaining optimal input features and an optimal least squares support vector machine, wherein the optimal least squares support vector machine is the mileage prediction error compensation model; wherein ε_k^t is a real output value in the validation set.

The present invention establishes an error correction model by means of data driving, which is beneficial to reducing the error caused by the deviation of templates and actual values, and further improves the mileage prediction accuracy.

The specific implementation process of step 7) includes:

1) acquiring average temperature distribution data within T min before the train enters the tunnel, obtaining temperature sequences and humidity sequences of the head and tail of the train at the current position, obtaining a template matching output value $O^M(i)$, and obtaining a compensation error output result $\varepsilon(i)$ by using the tunnel mileage prediction error compensation model; and

2) obtaining a mileage value from the current position of the train to the tunnel exit by means of the following equation: $O(i)=O^M(i)+\varepsilon(i)$.

The present invention calculates the tunnel mileage by a combination of template matching rough prediction and error correction compensation, which is beneficial to

improving the mileage calculation accuracy.

The present invention further provides a system for calculating the tunnel passing duration of a train, including:

a best matching position obtaining module, configured to calculate correlations between current position feature images and templates in a typical sample joint template library to obtain a best matching position of the current position in the typical sample joint template library; wherein the current position feature images are obtained by phase space reconstruction and RGB color space combination of acquired sequences of meteorological parameters of sample points at the head and tail of the train at the current position and meteorological parameters with a period of time ahead of the current position, the template images in the typical sample joint template library are color images about the evolution of meteorological parameters at the head and tail of the train, and the meteorological parameters at the head and tail of the train are obtained by phase space reconstruction of typical samples of each category of tunnel groups, and the typical samples of each category of tunnel groups refer to samples obtained by classifying the meteorological parameters according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel and further classifying the classified homogeneous samples in a region;

a tunnel mileage prediction error compensation model, obtained by training a least squares support vector machine using data of homogeneous tunnel groups, with a temperature sequence, a humidity sequence, and an air pressure sequence as input, and a prediction error of the best matching position as output, wherein the output of this model is a compensation error $\varepsilon(i)$; and

a remaining duration calculation module, configured to estimate the remaining duration t that the train passes through the current tunnel according to the following equation: $t=O(i)/(f_{Hz}(O(i)-O(i-1)))$; wherein $O(i)=O^M(i)+\varepsilon(i)$, and $O(i-1)$ is a mileage value from a previous position relative to the current position of the train to the tunnel exit; $O^M(i)$ is the best matching position; and f_{Hz} is the position update frequency of the train.

The best matching position obtaining module includes:

a data acquisition unit, configured to determine current sample points and previous G sample points in the temperature sequence, humidity sequence and air pressure sequence of the head and tail of the train at the current position;

a feature image calculation unit, configured to perform phase space reconstruction by means of a delay coordinate method to obtain six two-dimensional reconstruction

matrices representing the evolution characteristics of temperature, humidity and air pressure at the head and tail of the train, divide the six matrices into matrices corresponding to the head of the train and matrices corresponding to the tail of the train, and combine the six matrices according to an RGB color space to form current position feature images $\{c_h, c_f\}$;

a convolution unit, configured to perform convolution operation on the current position feature images and the images in the typical sample joint template library to obtain a plurality of one-dimensional sequences;

a sorting unit, configured to sort elements in all the one-dimensional sequences from large to small, to determine L largest elements as candidate elements, wherein the positions corresponding to the candidate elements before sorting are candidate positions, and the mileage values to the tunnel exit corresponding to the candidate positions are s_j (j=1, 2...L); and

an output unit, configured to average the mileage values to the tunnel exit corresponding to the candidate positions, wherein the average value is the best matching position O^M: $O^M = \frac{1}{L}\sum_{j=1}^{L} s_j$ of the current position in the typical sample joint template library.

The training process of the tunnel mileage prediction error compensation model of the present invention includes:

1) defining input samples $I=(T_h, H_h, P_h, T_f, H_f, P_f)$, wherein $T_h=(t_1, t_2\ldots, t_G)$ is a temperature sequence of current sample points and G previous sample points at the head of the train in the tunnel, H_h is a train head humidity sequence with a length of G, P_h is a train head air pressure sequence with a length of G, T_f is a train tail temperature sequence with a length of G, H_f is a train tail humidity sequence with a length of G, and P_f is a train tail air pressure sequence with a length of G; defining output samples as mileage error values ε corresponding to the current position; wherein combinations $Y=\{I, \varepsilon\}$ of the input samples and the output samples constitute modeling samples;

2) classifying the modeling samples into a training set and a validation set;

3) performing binary coding on the feature of each dimension in the input samples I, wherein when the coded value corresponding to the feature of a certain dimension is 1, the feature is selected as an input variable of the least squares support vector machine; when the coded value corresponding to the feature of a certain dimension is 0, the feature of this dimension is discarded;

4) updating the input samples based on the coded values of the current features to obtain a updated training set and a validation set, and training the least squares support vector machine with data of the updated training set; inputting data of the updated validation set into the trained least squares support vector machine to obtain an output sequence ε_k^v of the least squares support vector machine, wherein k=1, 2..., N_1, and N_1 is the number of sample pairs in the validation set; and

5) repeating step 3) and step 4) to determine input features and a least squares support vector machine that minimize the optimization objective function $f=\sum_{k=1}^{N_1}(\varepsilon_k^t-\varepsilon_k^v)^2$, that is, obtaining optimal input features and an optimal least squares support vector machine, wherein the optimal least squares support vector machine is the mileage prediction error compensation model; wherein ε_k^t is a real output value in the validation set.

As an inventive concept, the present invention further provides a computer-readable storage medium storing a program; and the program is 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 uses artificial intelligence big data analysis technology to fully mine the potential law of changes of barometric and meteorological parameters in the tunnel with the depth of the tunnel. Through the temperature, humidity and air pressure sequence data acquired at both ends of the train, intelligent perception of the distance between the train in the tunnel and the tunnel exit is realized, and on this basis, effective estimation of stay time of the train in the tunnel is realized.

2. The present invention can provide passengers with accurate information prompts (such as temperature, humidity, and remaining time for staying in the tunnel), which is beneficial to improving the passenger experience.

3. After the modeling is completed, only train-mounted temperature, humidity, and air pressure sensors are required to acquire input data without any trackside equipment, so the preset invention has great promotion value.

Brief Description of the Drawings

Fig. 1 is a main flowchart of the present invention;

Fig. 2 is a schematic diagram of time sequence phase space reconstruction and RGB joint template matching according to the present invention.

Detailed Description of Embodiments

As shown in Fig. 1, the specific steps of Embodiment 1 of the present invention are as follows:

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

On-board sensors distributed at both ends of a train acquire, in real time, temperature, humidity and air pressure sequences when the train passes a tunnel, and response devices pre-mounted in the tunnel and on the train acquire mileage time sequences of the train in the tunnel, with a sampling frequency of 10 Hz. Average values of temperature and humidity acquired at T min before the train enters the tunnel are used as estimated values of local average temperature and humidity. The temperature sequence, humidity sequence and air pressure sequence acquired by the sensors at both ends of the train when the train passes through a tunnel at one time, and the estimated values, acquired before entering the tunnel, of average temperature and average humidity of a region where the tunnel is located constitute a set of tunnel meteorological parameter samples. The tunnel meteorological parameter samples acquired during the operation of all trains in the region within one year constitute a tunnel meteorological parameter database. The value range of T is [10, 20], and T=10 in the present invention.

Step 2: classification of barometric and meteorological parameters of tunnel groups

The meteorological parameters in the database are classified based on the tunnel meteorological parameter database according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel, so as to realize the classification of tunnel groups with similar attributes. The specific implementation process is as follows:

Step A1: the tunnel groups are classified into 7 categories according to the classification of China's climate characteristics by China's architectural zoning and according to the geographical position of the tunnel samples in the tunnel meteorological parameter database.

Step A2: average temperature distributions within T min before the train enters the tunnel for one year in the current region are acquired, and the average temperature distributions are fitted with a Gaussian distribution function. The average temperature distributions are divided into 10 equal parts according to the probability, the temperature samples that belong to the same part are defined as homogeneous

temperature samples in the current region, and the temperature sequence, humidity sequence and air pressure sequence corresponding to the homogeneous temperature samples are defined as homogeneous samples.

Step 3: representation of homogeneous samples and typical samples in the region

Each homogeneous sample in the region, which is obtained based on rough tunnel group category classification and by inputting a set of attributes, is further classified. Specifically, this process includes the following sub-steps:

Step B1: evolution characteristics of the temperature, humidity and air pressure sequences when the train passes through the tunnel in the homogeneous samples are extracted. The specific step includes modeling an autoregressive integrated moving average (ARIMA) model for each time sequence, and extraction of parameters of autoregressive terms, integrated terms and moving regression terms of each sequence. Specifically, 6 time sequences can be acquired when the train passes through the tunnel at one time, so a characteristic quantity extracted when the train passes through the tunnel at one time can be expressed as $\alpha=(p_1, d_1, q_1, p_2, d_2, q_2, \ldots, p_6, d_6, q_6)$, where p_1 is the autoregressive terms in time sequence 1, d_1 is the integrated terms in time sequence 1, and q_1 is the moving regression terms in time sequence 1. All the characteristic quantities of the homogeneous sample in the region constitute a characteristic matrix A.

Step B2: dimension reduction is performed by means of principal component analysis (PCA) on the characteristic matrix A constituted by all the characteristic quantities of the homogeneous sample in the region. 5 principal components with the greatest contribution are selected to characterize the information of the original characteristic matrix A, and a transformed matrix A' is obtained.

Step B3: a kernel function is defined as:

$$k = \alpha k_{rbf} + \beta k_{linear} + (1-\alpha-\beta) k_{laplace} \tag{1}$$

In the equation, k_{rbf} is a radial basis kernel function, k_{linear} is a linear kernel, and $k_{laplace}$ is a Laplace kernel function. Characteristic values in the matrix A' are mapped to a characteristic space corresponding to the kernel function k.

Step B4: optimized objects are determined, that is, the coefficients α, β and the number of categories of the kernel function are optimized by means of gray wolf optimization (GWO). $\alpha, \beta \in [0,1]$, and the number of categories is a positive integer less than 20.

Step B5: an optimization objective function is determined as:

$$fitness = \frac{1}{n}\sum_{i=1}^{n}\max_{i\neq j}\left(\frac{avg\left(C_i\right)+avg\left(C_i\right)}{d_{cen}\left(C_i,C_j\right)}\right) \quad (2)$$

In the equation, $avg(C_i)$ is an average distance of samples in cluster C_i, and $d_{cen}(C_i, C_j)$ is a distance between center points of cluster C_i and cluster C_j.

Step B6: characteristics after dimension reduction are clustered according to the above set parameters by means of a k-means clustering algorithm optimized by the gray wolf optimization. Cluster centers of each sample cluster and an original time sequence { X_i^1 } (i=0, 1, 2, …, 5) corresponding to 5 samples closest to the cluster center are acquired. $X_{i=}^1(x_i^1, x_i^2, x_i^3, x_i^4, x_i^5, x_i^6)$ corresponds to a temperature sequence acquired at the head of the train, a humidity sequence acquired at the head of the train, an air pressure sequence acquired at the head of the train, a temperature sequence acquired at the tail of the train, a humidity sequence acquired at the tail of the train, and an air pressure sequence acquired at the tail of the train. The time sequence set { X_i^1} (i=0, 1, 2…5) is defined as a typical sample of the current category.

Step 4: construction of a typical sample joint template library

Phase space reconstruction is performed on the typical sample, a delay time is set to 1, a window length is set to 5, phase space reconstruction is performed on the temperature, humidity and air pressure sequences by means of a delay coordinate method to obtain three two-dimensional reconstruction matrices representing the evolution characteristics of temperature, humidity and air pressure at the head of the train and three two-dimensional reconstruction matrices representing the evolution characteristics of temperature, humidity and air pressure at the tail of the train, the respective three matrices are combined according to an RGB color space to form color images representing the evolution characteristics of barometric and meteorological parameters at the head or tail of the train, and the color images are template images {h_i, f_i} (i=0, 1, 2…5).

Step 5: training of a template RGB color joint matching model (as shown in Fig. 2)

Meteorological data of sample points at the head and tail of the train at the current position and meteorological sample sequences ahead a period of time are acquired respectively, and phase space reconstruction and RGB color space combination are performed to construct current position feature images {c_h, c_f}. The correlations between a current position feature module and images in the template library are calculated to determine a best matching position of the current position in the template

library. This process includes the following steps:

Step C1: relevant meteorological data of current sample points in the temperature, humidity and air pressure sequences of the head and tail of the train at the current position and previous 19 sample points is acquired.

Step C2: a delay time is set to 1, a window length is set to 5, phase space reconstruction is performed by means of the delay coordinate method to obtain six two-dimensional reconstruction matrices representing the evolution characteristics of temperature, humidity and air pressure at the head and tail of the train, the six matrices are divided into matrices corresponding to the head of the train and matrices corresponding to the tail of the train, and the six matrices are combined according to the RGB color space to form current position feature images $\{c_h, c_f\}$.

Step C3: convolution operation $g_i=h_i\otimes c_h+f_i\otimes c_f$ (i=0, 1, 2...5) is performed on the current position feature images and the images in the template library, wherein each g_i is a one-dimensional sequence.

Step C4: elements in all the g_i sequences are sorted from large to small, 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 to the tunnel exit corresponding to the candidate positions are s_j (j=1, 2...5).

Step C5: the mileage values to the tunnel exit corresponding to the candidate positions are averaged, wherein the average value is a current template matching output value, that is, the calculation formula of the template matching output value is:

$$O^M = \frac{1}{5}\sum_{j=1}^{5} s_j \tag{3}$$

Step 6: establishment of a joint template matching error compensation model

A least squares support vector machine (LSSVM) is trained using data of the homogeneous tunnel groups, with the temperature sequence, humidity sequence, and air pressure sequence as input, and a prediction error of the template matching model as output, to establish a tunnel mileage prediction error compensation model. This process includes the following steps:

Step D1: training samples are defined, and input samples $I=(T_h, H_h, P_h, T_f, H_f, P_f)$ are defined, wherein $T_h=(t_1, t_2\ldots, t_{19}, t_{20})$ is a temperature sequence of current sample points and 19 previous sample points at the head of the train in the tunnel. Similarly, H_h is a train head humidity sequence with a length of 20, P_h is a train head air pressure sequence with a length of 20, T_f is a train tail temperature sequence with a length of 20, H_f is a

train tail humidity sequence with a length of 20, and P_f is a train tail air pressure sequence with a length of 20. Output samples are mileage error values ε corresponding to the current position. Input and output combinations $Y = \{I, \varepsilon\}$ constitute modeling samples. For each homogeneous tunnel group, 3000 samples are selected to establish the mileage prediction error compensation model.

Step D2: the modeling samples are classified into a training set and a validation set. By means of random sampling without replacement, 70% of 3000 sample pairs among the modeling samples are selected as the training set and 30% as the validation set.

Step D3: optimization objects are determined and optimization values are initialized. Input features of the model are optimized by means of a binary ant lion optimizer 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 LSSVM model. When the coded value corresponding to the feature of a certain dimension is 0, the feature of this dimension is discarded. 60 dimensional features are randomly initialized and coded as 0 or 1.

Step D4: an optimization objective function is determined. Input features are determined based on the coded values of the current features, and the LSSVM model is trained with data of the training set. Data of the validation set is input to the trained LSSVM model to obtain a model output sequence ε_k^v (k=1, 2…900). The optimization objective function is defined as:

$$f = \sum_{k=1}^{900} (\varepsilon_k^t - \varepsilon_k^v)^2 \tag{4}$$

In the equation, ε_k^t (k=1, 2…900) is a real output value of the validation set.

Step D5: an optimized prediction model is outputted. Iterative optimization operations are performed by means of the binary ant lion algorithm to determine optimal input features and LSSVM model, the LSSVM model is the LSSVM mileage prediction error compensation model.

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

During the operation of the train, real-time barometric and meteorological parameters are acquired to construct current status feature images, and three-dimensional image templates are searched and matched; and error compensation input variables and an error compensation model are determined. The template matching output is fused with

the error compensation output to obtain a final mileage value to the tunnel exit. This process includes the following steps:

Step E1: an average temperature distribution within T min before the train enters the tunnel is acquired. A sample category of the current status (that is, meteorological parameters of the current position of the train) is determined. A current temperature sequence and a current humidity sequence are acquired by means of temperature and humidity sensors mounted at the head and tail of the train.

Step E2: a corresponding template library is extracted according to the sample category, determined according to the current average temperature, of the current status.

Step E3: the best matching position is determined in the template library with reference to the process in step 3, and template matching output values $O^M(i)$ of current sample points are outputted.

Step E4: model input vectors I of the current status are obtained with reference to the process C1 in step 6, and substituted into the trained LSSVM model to obtain compensation error output results $\varepsilon(i)$.

Step E5: the three-dimensional template matching output values and the output values of the LSSVM model are fused to obtain a final output value of the distance between the current position and the tunnel exit:

$$O(i) = O^M(i) + \varepsilon(i) \tag{5}$$

Step 8: calculation of predicted remaining tunnel exit duration

The remaining duration that the train passes through the current tunnel can be estimated according to the following equation:

$$t = O(i) / \left(f_{Hz} \left(O(i) - O(i-1) \right) \right) \tag{6}$$

In the equation, f_{Hz} is a position update frequency, and f_{Hz}=50Hz in the present invention.

Corresponding to the method in Embodiment 1, Embodiment 2 of the present invention provides a system for calculating the tunnel passing duration of a train, including:

a best matching position obtaining module, configured to calculate correlations between current position feature images and templates in a typical sample joint template library to obtain a best matching position of the current position in the typical sample joint template library; wherein the current position feature images are obtained by phase space reconstruction and RGB color space combination of acquired sequences of meteorological parameters of sample points at the head and tail of the train at the current

position and meteorological parameters with a period of time ahead of the current position, the template images in the typical sample joint template library are color images about the evolution of meteorological parameters at the head and tail of the train, and the meteorological parameters at the head and tail of the train are obtained by phase space reconstruction of typical samples of each category of tunnel groups, and the typical samples of each category of tunnel groups refer to samples obtained by classifying the meteorological parameters according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel and further classifying the classified homogeneous samples in a region;

a tunnel mileage prediction error compensation model, obtained by training a least squares support vector machine using data of homogeneous tunnel groups, with a temperature sequence, a humidity sequence, and an air pressure sequence as input, and a prediction error of the best matching position as output, wherein the output of this model is a compensation error $\varepsilon(i)$; and

a remaining duration calculation module, configured to estimate the remaining duration t that the train passes through the current tunnel according to the following equation: $t=O(i)/(f_{Hz}(O(i)-O(i-1)))$; wherein $O(i)=O^M(i)+\varepsilon(i)$, and $O(i-1)$ is a mileage value from a previous position relative to the current position of the train to the tunnel exit; $O^M(i)$ is the best matching position; and f_{Hz} is the position update frequency of the train.

The best matching position obtaining module includes:

a data acquisition unit, configured to determine current sample points and previous G sample points in the temperature sequence, humidity sequence and air pressure sequence of the head and tail of the train at the current position;

a feature image calculation unit, configured to perform phase space reconstruction by means of a delay coordinate method to obtain six two-dimensional reconstruction matrices representing the evolution characteristics of temperature, humidity and air pressure at the head and tail of the train, divide the six matrices into matrices corresponding to the head of the train and matrices corresponding to the tail of the train, and combine the six matrices combine the six matrices at the head and tail according to an RGB color space to form current position feature images $\{c_h, c_f\}$;

a convolution unit, configured to perform convolution operation on the current position feature images and the images in the typical sample joint template library to obtain a plurality of one-dimensional sequences;

a sorting unit, configured to sort elements in all the one-dimensional sequences from

large to small, to determine L largest elements as candidate elements, wherein the positions corresponding to the candidate elements before sorting are candidate positions, and the mileage values to the tunnel exit corresponding to the candidate positions are s_j (j=1, 2...L); and

an output unit, configured to average the mileage values to the tunnel exit corresponding to the candidate positions, wherein the average value is the best matching position O^M: $O^M = \frac{1}{L}\sum_{j=1}^{L} s_j$ of the current position in the typical sample joint template library.

The data acquisition unit in this embodiment includes various sensors mounted at the head and tail of the train, such as temperature sensors, humidity sensors, and air pressure sensors.

The training process of the tunnel mileage prediction error compensation model includes:

1) Input samples I=(T_h, H_h, P_h, T_f, H_f, P_f) are defined, wherein T_h=(t_1, t_2…, t_G) is a temperature sequence of current sample points and G previous sample points at the head of the train in the tunnel; H_h is a train head humidity sequence with a length of G, P_h is a train head air pressure sequence with a length of G, T_f is a train tail temperature sequence with a length of G, H_f is a train tail humidity sequence with a length of G, and P_f is a train tail air pressure sequence with a length of G; output samples are defined as mileage error values ε corresponding to the current position; and combinations Y ={I, ε} of the input samples and the output samples constitute modeling samples;

2) The modeling samples are classified into a training set and a validation set;

3) Binary coding is performed on the feature of each dimension in the input samples I, wherein when the coded value corresponding to the feature of a certain dimension is 1, the feature is selected as an input variable of the least squares support vector machine; when the coded value corresponding to the feature of a certain dimension is 0, the feature of this dimension is discarded;

4) The input samples are updated based on the coded values of the current features to obtain a updated training set and validation set, and the least squares support vector machine is trained with data of the updated training set; data of the updated validation set is input into the trained least squares support vector machine to obtain an output sequence ε_k^v of the least squares support vector machine, wherein k=1, 2..., N_1, and N_1 is the number of sample pairs in the validation set;

5) Step 3) and step 4) are repeated to determine input features and a least squares support vector machine that minimize the optimization objective function $f=\sum_{k=1}^{N_1}(\varepsilon_k^t-\varepsilon_k^v)^2$, that is, optimal input features and an optimal least squares support vector machine are obtained, wherein the optimal least squares support vector machine is the mileage prediction error compensation model; wherein ε_k^t is a real output value in the validation set.

Embodiment 3 of the present invention provides a computer-readable storage medium storing a program; and the program is configured to execute the steps of the method according to Embodiment 1 of the present invention.

Claims

1. A method for calculating the tunnel passing duration of a train, wherein the remaining duration that the train passes through the current tunnel is estimated by using the following equation: $t=O(i)/f_{Hz}(O(i)-O(i-1)))$; wherein the mileage value from the current position of the train to the tunnel exit is $O(i)= O^{M}(i)+\varepsilon(i)$; $O(i-1)$ is the mileage value from the last position of the current position of the train to the tunnel exit; $O^{M}(i)$ is the best matching position of the current position in a typical sample joint template library, which is obtained after the correlations between current position feature images and images in the typical sample joint template library are calculated; the current position feature images are obtained by phase space reconstruction and RGB color space combination of acquired sequences of meteorological parameters of sample points at the head and tail of the train at the current position and meteorological parameters with a period of time ahead of the current position; the template images in the typical sample joint template library are color images about the evolution of meteorological parameters at the head and tail of the train; the meteorological parameters at the head and tail of the train are obtained by phase space reconstruction of typical samples of each category of tunnel groups; the typical samples of each category of tunnel groups refer to samples obtained by classifying the meteorological parameters according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel and further classifying the classified homogeneous samples in a region; $\varepsilon(i)$ is a compensation error obtained by inputting the meteorological parameters acquired during the operation of the train into a tunnel mileage prediction error compensation model, and the tunnel mileage prediction error compensation model refers to a model obtained by training a least squares support vector machine using data of homogeneous tunnel groups, with a temperature sequence, a humidity sequence, and an air pressure sequence as input, and a prediction error of an RGB color joint matching model as output; and f_{Hz} is a position update frequency; and the meteorological parameters refer to a temperature sequence corresponding to the head of the train, a humidity sequence corresponding to the head of the train, an air pressure sequence corresponding to the head of the train, a temperature sequence corresponding to the tail of the train, a humidity sequence corresponding to the tail of the train, an air pressure sequence corresponding to the tail of the train;A specific determination process of the mileage value $O(i)$ from the current position of the train to the tunnel exit comprises:

1) acquiring tunnel meteorological parameters to construct a tunnel meteorological parameter database;

2) classifying, based on the tunnel meteorological parameter database, the meteorological parameters in the tunnel meteorological parameter database according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel to obtain homogeneous samples in a region, so as to realize the classification of tunnel groups with similar attributes;

3) further classifying the homogeneous samples in the region to obtain typical samples of each category of tunnel groups;

4) performing phase space reconstruction on the typical samples of each category of tunnel groups to obtain color images about the evolution of barometric and meteorological parameters at the head and tail of the train, that is, template images; wherein the template images corresponding to all the typical samples constitute the typical sample joint template library;

5) collecting sample points at the head and tail of the train at the current position and sample sequences with a period of time ahead the current position respectively, and performing phase space reconstruction and RGB color space combination to construct current position feature images $\{c_h, c_f\}$; calculating the correlations between the current position feature images and the template images in the typical sample joint template library to determine the best matching position of the current position in the typical sample joint template library;

6) training a least squares support vector machine using data of the homogeneous tunnel groups, with a temperature sequence, a humidity sequence, and an air pressure sequence as input, and a prediction error of the best matching position as output, to establish the tunnel mileage prediction error compensation model; and

7) acquiring meteorological parameters during the operation of the train in real time, and obtaining a final mileage value to the tunnel exit by using the tunnel mileage prediction error compensation model, that is, obtaining $O(i)$.

2. The method for calculating the tunnel passing duration of a train according to claim 1, wherein the specific implementation process of step 2) comprises:

1) classifying the tunnel groups into N categories according to the geographical position of the tunnel samples; and

2) acquiring average temperature distributions within T min before the train enters the

tunnel for one year in the current region, fitting the average temperature distributions in the tunnel of the current region by using a Gaussian distribution function, dividing the fitted average temperature distributions into K equal parts according to the probability, defining the temperature samples that belong to the same part as homogeneous temperature samples in the current region, and defining the temperature sequence, humidity sequence and air pressure sequence corresponding to the homogeneous temperature samples as homogeneous samples; wherein T is a set time.

3. The method for calculating the tunnel passing duration of a train according to any one of claims 1, wherein the specific implementation process of step 3) comprises:

1) creating an autoregressive integrated moving average model for each time sequence of the homogeneous samples in the current region, and extracting characteristic quantities of each time sequence, that is, parameters of autoregressive terms, integrated terms and moving regression terms of each time sequence; wherein all the characteristic quantities of the homogeneous samples in the region constitute a characteristic matrix A; wherein the time sequence comprises a temperature sequence, a humidity sequence and an air pressure sequence;

2) reducing dimensions of the characteristic matrix A, and selecting M principal components with the greatest contribution to characterize the information of the original characteristic matrix , thus obtaining a transformed matrix A';

3) defining a kernel function $k=\alpha k_{rbf}+\beta k_{linear}+(1-\alpha-\beta)k_{laplace}$; wherein k_{rbf} is a radial basis kernel function, k_{linear} is a linear kernel, and $k_{laplace}$ is a Laplace kernel function; randomly determining initial values of the coefficients α, β and the number n_0 of categories of the kernel function, wherein $\alpha, \beta \in [0, 1]$, and n_0 is a positive integer less than 30;

4) mapping the characteristic values in the matrix A' to a characteristic space corresponding to the kernel function k to obtain a new characteristic matrix A''; according to the initial number n_0 of categories, using the characteristic quantities in A'' as input to cluster characteristic quantities in A'' by means of a k-means clustering algorithm , that is, to classify the homogeneous samples in the region into n categories, wherein each category of samples constitute a sample cluster; calculating an optimization objective function $fitness=\frac{1}{n}\sum_{i=1}^{n}\max_{i\neq j}\left(\frac{avg(C_i)+avg(C_j)}{d_{cen}(C_i,C_j)}\right)$, wherein $avg(C_i)$

is an average distance of samples in the sample cluster C_i, $avg(C_j)$ is an average distance of samples in the sample cluster C_j, and $d_{cen}(C_i, C_j)$ is a distance between center points of the sample cluster C_i and the sample cluster C_j;

5) updating, according to the current optimization objective function value *fitness*, the values of α, β and n_0 by gray wolf optimization;

6) repeating step 4) and step 5) until the set number m=100 of optimizations is reached; wherein the values of α, β and n_0 at this time are final optimized parameter values, and the clustering results under the final optimized parameter values are final clustering results; and

7) classifying the homogeneous samples in the current region into E sample clusters according to the final clustering results; and acquiring a cluster center of each sample cluster and an original time sequence $\{X_i^n\}$ (i=0, 1, 2…P, n=0, 1, 2…E) corresponding to P samples closest to the cluster center, wherein $X_i^n = (x_i^1, x_i^2, x_i^3, x_i^4, x_i^5, x_i^6)_n$ corresponds to a temperature sequence acquired at the head of the train, a humidity sequence acquired at the head of the train, an air pressure sequence acquired at the head of the train, a temperature sequence acquired at the tail of the train, a humidity sequence acquired at the tail of the train, and an air pressure sequence acquired at the tail of the train, respectively; and defining an original time sequence corresponding to (P+1) samples as a typical sample of the current category of tunnel groups.

4. The method for calculating the tunnel passing duration of a train according to any one of claims 1, wherein the specific implementation process of step 5) comprises:

1) determining current sample points and previous G sample points in the temperature sequence, humidity sequence and air pressure sequence of the head and tail of the train at the current position;

2) performing phase space reconstruction by means of the delay coordinate method to obtain six two-dimensional reconstruction matrices representing the evolution characteristics of temperature, humidity and air pressure at the head and tail of the train; dividing the six matrices into matrices corresponding to the head of the train and matrices corresponding to the tail of the train, and combining the six matrices according to the RGB color space to form current position feature images $\{c_h, c_f\}$;

3) performing convolution operation on the current position feature images and the images in the typical sample joint template library to obtain a plurality of one-

dimensional sequences;

4) sorting elements in all the one-dimensional sequences from large to small, to determine L largest elements as candidate elements; wherein the positions corresponding to the candidate elements before sorting are candidate positions, and the mileage values to the tunnel exit corresponding to the candidate positions are s_j (j=1, 2...L); and

5) averaging the mileage values to the tunnel exit corresponding to the candidate positions , wherein the average value is the best matching position O^M : $O^M = \frac{1}{L}\sum_{j=1}^{L} s_j$ of the current position in the typical sample joint template library.

5. The method for calculating the tunnel passing duration of a train according to any one of claim 1, wherein the specific implementation process of step 6) comprises:

1) defining input samples I=(T_h, H_h, P_h, T_f, H_f, P_f), wherein T_h=(t_1, t_2…t_G) is a temperature sequence of current sample points and G previous sample points at the head of the train in the tunnel, H_h is a train head humidity sequence with a length of G, P_h is a train head air pressure sequence with a length of G, T_f is a train tail temperature sequence with a length of G, H_f is a train tail humidity sequence with a length of G, and P_f is a train tail air pressure sequence with a length of G; defining output samples as mileage error values ε corresponding to the current position; wherein the combinations Y ={I, ε} of the input samples and the output samples constitute modeling samples;

2) classifying the modeling samples into a training set and a validation set;

3) performing binary coding on the feature of each dimension in the input samples I; wherein when the coded value corresponding to the feature of a certain dimension is 1, the feature is selected as an input variable of the least squares support vector machine; when the coded value corresponding to the feature of a certain dimension is 0, the feature of this dimension is discarded;

4) updating the input samples based on the coded values of the current features to obtain a updated training set and validation set; and training the least squares support vector machine with data of the updated training set; inputting data of the updated validation set into the trained least squares support vector machine to obtain an output sequence ε_k^v of the least squares support vector machine, wherein k=1, 2..., N_1, and N_1 is the number of sample pairs in the validation set; and

5) repeating step 3) and step 4) to determine input features and a least squares support

vector machine that minimize the optimization objective function $f=\sum_{k=1}^{N_1}(\varepsilon_k^t-\varepsilon_k^v)^2$, that is, obtaining optimal input features and an optimal least squares support vector machine, wherein the optimal least squares support vector machine is the mileage prediction error compensation model; wherein ε_k^t is a real output value in the validation set.

6. The method for calculating the tunnel passing duration of a train according to any one of claims 1, wherein the specific implementation process of step 7) comprises:

1) acquiring average temperature distribution data within T min before the train enters the tunnel; obtaining temperature sequences and humidity sequences of the head and tail of the train at the current position; obtaining a template matching output value $O^M(i)$; and obtaining a compensation error output result $\varepsilon(i)$ by using the tunnel mileage prediction error compensation model; and

2) obtaining a mileage value from the current position of the train to the tunnel exit by means of the following equation: $O(i)=O^M(i)+\varepsilon(i)$.

7. A system for calculating the tunnel passing duration of a train, comprising:

a best matching position obtaining module, configured to calculate correlations between current position feature images and templates in a typical sample joint template library to obtain a best matching position of the current position in the typical sample joint template library; wherein the current position feature images are obtained by phase space reconstruction and RGB color space combination of acquired sequences of meteorological parameters of sample points at the head and tail of the train at the current position and meteorological parameters with a period of time ahead of the current position; the template images in the typical sample joint template library are color images about the evolution of meteorological parameters at the head and tail of the train, and the meteorological parameters at the head and tail of the train are obtained by phase space reconstruction of typical samples of each category of tunnel groups; and the typical samples of each category of tunnel groups refer to samples obtained by classifying the meteorological parameters according to the geographical position of the tunnel and the meteorological conditions inside and outside the tunnel and further classifying the classified homogeneous samples in a region;

a tunnel mileage prediction error compensation model, obtained by training a least squares support vector machine using data of homogeneous tunnel groups, with a

temperature sequence, a humidity sequence, and an air pressure sequence as input, and a prediction error of the best matching position as output, wherein the output of this model is a compensation error $\varepsilon(i)$; and

a remaining duration calculation module, configured to estimate the remaining duration t that the train passes through the current tunnel according to the following equation: $t=O(i)/(f_{Hz}(O(i)-O(i-1)))$, wherein $O(i)=O^M(i)+\varepsilon(i)$, and $O(i-1)$ is a mileage value from a previous position relative to the current position of the train to the tunnel exit; $O^M(i)$ is the best matching position; and f_{Hz} is the position update frequency of the train.

8. The system according to claim 7, wherein the best matching position obtaining module comprises:

a data acquisition unit, configured to determine current sample points and previous G sample points in the temperature sequence, humidity sequence and air pressure sequence of the head and tail of the train at the current position;

a feature image calculation unit, configured to perform phase space reconstruction by means of a delay coordinate method to obtain six two-dimensional reconstruction matrices representing the evolution characteristics of temperature, humidity and air pressure at the head and tail of the train; divide the six matrices into matrices corresponding to the head of the train and matrices corresponding to the tail of the train, and combine the six matrices according to an RGB color space to form current position feature images $\{c_h, c_f\}$;

a convolution unit, configured to perform convolution operation on the current position feature images and the images in the typical sample joint template library to obtain a plurality of one-dimensional sequences;

a sorting unit, configured to sort elements in all the one-dimensional sequences from large to small, to determine L largest elements as candidate elements; wherein the positions corresponding to the candidate elements before sorting are candidate positions, and the mileage values to the tunnel exit corresponding to the candidate positions are s_j (j=1, 2...L); and

an output unit, configured to average the mileage values to the tunnel exit corresponding to the candidate positions, wherein the average value is the best matching position O^M: $O^M=\frac{1}{L}\sum_{j=1}^{L}s_j$ of the current position in the typical sample joint template library;

preferably, a training process of the tunnel mileage prediction error compensation model of the present invention comprises:

1) defining input samples $I=(T_h, H_h, P_h, T_f, H_f, P_f)$, wherein $T_h=(t_1, t_2\ldots, t_G)$ is a temperature sequence of current sample points and G previous sample points at the head of the train in the tunnel, H_h is a train head humidity sequence with a length of G, P_h is a train head air pressure sequence with a length of G, T_f is a train tail temperature sequence with a length of G, H_f is a train tail humidity sequence with a length of G, and P_f is a train tail air pressure sequence with a length of G; defining output samples as mileage error values ε corresponding to the current position; wherein combinations $Y=\{I, \varepsilon\}$ of the input samples and the output samples constitute modeling samples;

2) classifying the modeling samples into a training set and a validation set;

3) performing binary coding on the feature of each dimension in the input samples I; wherein when the coded value corresponding to the feature of a certain dimension is 1, the feature is selected as an input variable of the least squares support vector machine; when the coded value corresponding to the feature of a certain dimension is 0, the feature of this dimension is discarded;

4) updating the input samples based on the coded values of the current features to obtain a updated training set and validation set; and training the least squares support vector machine with data of the updated training set; inputting data of the updated validation set into the trained least squares support vector machine to obtain an output sequence ε_k^v of the least squares support vector machine, wherein k=1, 2..., N_1, and N_1 is the number of sample pairs in the validation set; and

5) repeating step 3) and step 4) to determine input features and a least squares support vector machine that minimize the optimization objective function $f=\sum_{k=1}^{N_1}(\varepsilon_k^t-\varepsilon_k^v)^2$, that is, obtaining optimal input features and an optimal least squares support vector machine, wherein the optimal least squares support vector machine is the mileage prediction error compensation model; wherein ε_k^t is a real output value in the validation set.

9. A computer-readable storage medium storing a program; wherein the program is configured to execute the steps of the method according to any one of claims 1-6.