Fault Diagnosis Method of Section Track Circuit Based on Support Vector Machine

Pengrui Liu, Wenchang Li
China Academy of Railway Sciences Corporation Limited standards & Metrology Research Institute, Beijing, 100081, China

Abstract: Support Vector Machines (SVM) is a Machine Learning Algorithm Which, Based on statistical Learning Theory. This algorithm accomplishes structural risk-minimization principle. Most predominance. SVM. proper. limited samples decision.. Nature. algorithm. acquiring connotative class information. great extent form limited samples. This article will support vector machine

Keywords: support; diagnosis; cross-subject

1. An overview

The most in recent years China of Railway career development quickly. Thanks to Sensing Technology, Big Data Analysis Technology, Cloud computing and cross-subject of development, The intelligence level of Railway Signal Monitoring System is not high at present. Low troubleshooting and set maintenance efficiency, Rely heavily on artificial experience, Support Vector Machine (SVM) Artificial intelligence algorithm ZPW-2000 Intelligent Diagnosis of track circuit, and the reliability and feasibility of the scheme are verified by experiments..

2. What is support vector machine?

The problem of machine learning is mainly solved by the traditional statistics based on the asymptotic theory, that is, when there are infinite training samples, the optimization of learning machine parameters is made use of the principle of empirical risk minimization. However, due to the differences between the empirical risk and the actual risk, and in the case of small samples, this difference is particularly obvious. In practical engineering, due to the fuzzy signal and fault relationship and other factors, the number of typical fault samples in fault diagnosis is limited. Like the artificial neural network based on the principle of empirical minimization, the problem to be solved is how to get the decision function with strong generalization ability from less fault samples, for small sample problem, the result of fault diagnosis is not ideal, and it is difficult to get valuable application. Therefore, in practical engineering, there is an urgent need to solve the small sample problem.

Century 90 In the middle of the decade, the theory of machine learning for small sample processing gradually matured, forming a more perfect theoretical system. 1992 Year to year 1995 A new learning machine Support Vector Machine (SVM) based on Statistical Learning Theory (Support vector machine, SVM) Came into being. Compared with the artificial neural network, SVM has several advantages. Can solve the small sample problem; second, Solving High Dimension and local extremum problems.
In addition, SVM has a very simple structure and is an effective tool to apply statistical learning theory. In low-dimensional space, for hard-to-classify data, SVM can be mapped into a high-dimensional space by some kind of mapping, making it classification. This machine learning algorithm is based on statistical learning theory. VCDimensional theory (Vapnik-Chervonenkis dimension) and structural risk minimization.

Structural Risk Mining (SRM) Based on the training principle, it has stronger generalization ability than neural network, can do the best tradeoff between model complexity and learning ability, and requires small sample training data. Therefore, this study uses SVM Basic Algorithms for Diagnosis. The basic principles are as follows:

Function \( \phi \) As a nonlinear mapping function, the vector \( X \) from the original data collection space to the high-dimensional feature space; \( \langle A, B \rangle \) Operator representation variable \( A \) With \( B \) The Inner Product. After the above parameter mapping, the original data which can not be classified in the low dimensional original space can be divided into high dimensional feature space. Classification hyperplane in upper-type high-dimension space by weight parameter \( \Omega \) And offset Parameter \( B \) Functional Theory shows that the optimal classification hyperplane must satisfy the formula (5.) Qualifying condition.

Can be understood as between two classification plane "Functional distance". Therefore, an appropriate transformation of the above problems can be expressed as follows (6.), Type (7.) Optimization Problem.

\[ \text{For Lagrange Operator. Function L Respectively on } \Omega \text{ And } B \text{ Please partial-make it zero and solving the can get its extreme value. Such (9),(10).} \]

Corresponding \( B^* \) Also can be accordingly derived by the classification super flat of decision-making function can be said such (12):

\[ F(X) = \text{Sign} \left[ \sum_{i=1}^{L} \alpha_i^* Y_i \langle X_i \phi(X) \rangle B^* \right] \]

Will monitoring Vector \( X \) Of value into the-according \( F(X) \) Of output results can be the input sample to classification. However actual data collection, Transmission and process will not avoid to make monitoring Vector \( X \) Mixed with certain of noise makes data difficult to completely the linear methods the classification then can through and relaxation Parameters \( \zeta > 0 \) The methods of the original methods the correction the original problem override such (13):

\[ \langle \Omega \phi(X_i) \rangle B \geq 1 - \zeta I = 12...L \]

At this time of classification problem can describe for such (14) Or (15) Shown in the Optimization Problem Which, C For punishment Parameters, \( \xi \) For relaxation variable.

According to the similar of analysis solving process at this time of decision-making function can be said for such (16) The form:

\[ F(X) = \text{Sign} \left[ \sum_{i=1}^{C} \alpha_i^* \langle X_i \phi(X) \rangle B^* \right] \]

I SVS

Which, C For punishment Parameters, \( \xi \) For relaxation variable, \( K(X_i) = \langle \phi(X)\phi(X_i) \rangle \) For nuclear Function.

The above mathematical model have many kinds of solving methods this paper the classic of order sequential minimum optimization (SMO) Algorithm to solve.

3. Model Establishment

This research based on support vector machine the algorithm Modeling. The model is divided into data read, Training Model and diagnosis three part. First import training diagnosis model used of training data set in import training data set after on its the training get to Fault Diagnosis Model. Training Before the model required of punishment Parameters C And Nuclear Parameters Gamma The estimation.

3.1 Data set

In order to make research object simplified this project only study track circuit adjustment state under the monitoring data. Data divided into training set
Traindata) And Test Set (Testdata) Two part training set used to training diagnosis model test set that is for diagnosis of actual

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Monitoring of name</th>
<th>Serial number</th>
<th>Monitoring of name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Work the voltage</td>
<td>14</td>
<td>By end cable side small system Carrier Frequency</td>
</tr>
<tr>
<td>2</td>
<td>Work the current</td>
<td>15</td>
<td>By end cable side small system Low Frequency</td>
</tr>
<tr>
<td>3</td>
<td>Work the carrier frequency</td>
<td>16</td>
<td>By end equipment side main system voltage</td>
</tr>
<tr>
<td>4</td>
<td>Work the low frequency</td>
<td>17</td>
<td>By end equipment side main system Carrier Frequency</td>
</tr>
<tr>
<td>5</td>
<td>Send end cable Side Voltage</td>
<td>18</td>
<td>By end equipment side main system Low Frequency</td>
</tr>
<tr>
<td>6</td>
<td>Send end cable side current</td>
<td>19</td>
<td>By end equipment side small system voltage</td>
</tr>
<tr>
<td>7/</td>
<td>Send end cable side Carrier Frequency</td>
<td>Natural 20</td>
<td>Receive entrance main system voltage</td>
</tr>
<tr>
<td>8</td>
<td>Send end cable side Low-Frequency</td>
<td>21</td>
<td>Receive entrance main system Carrier Frequency</td>
</tr>
<tr>
<td>9</td>
<td>Ballast Resistance</td>
<td>22</td>
<td>Receive entrance main system Low Frequency</td>
</tr>
<tr>
<td>10</td>
<td>By end cable side main system voltage</td>
<td>23</td>
<td>Receive entrance small system voltage</td>
</tr>
<tr>
<td>11</td>
<td>By end cable side main system Carrier Frequency</td>
<td>24</td>
<td>Receive entrance small system Carrier Frequency</td>
</tr>
<tr>
<td>12</td>
<td>By end cable side main system Low Frequency</td>
<td>25</td>
<td>Receive entrance small system Low Frequency</td>
</tr>
<tr>
<td>13</td>
<td>By end cable side small system voltage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data. Training set data and test set data points two kind of situation:
1) Existing signal Centralized Monitoring System of indoor monitoring data (25Monitoring)
2) In addition to the above 25A data the increase outdoor monitoring device after the data 11 The total 36 The.

2 Outdoor monitoring data list

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Monitoring of name</th>
<th>Serial number</th>
<th>Monitoring of name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Send end cable side current</td>
<td>7/</td>
<td>By end lead wire long die of wiring current</td>
</tr>
<tr>
<td>2</td>
<td>Send end lead wire long die of wiring current</td>
<td>8</td>
<td>By end lead wire long introduced wiring current</td>
</tr>
<tr>
<td>3</td>
<td>Send end lead wire long introduced wiring current</td>
<td>9</td>
<td>By end lead wire short die of wiring current</td>
</tr>
<tr>
<td>4</td>
<td>Send end lead wire short die of wiring current</td>
<td>10</td>
<td>By end lead wire short introduced wiring current</td>
</tr>
<tr>
<td>5</td>
<td>Send end lead wire short introduced wiring current</td>
<td>11</td>
<td>Protection box temperature</td>
</tr>
<tr>
<td>6</td>
<td>By end cable side current</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Taking into account the track circuit actual fault data difficult to get so first the mathematical modeling of methods established track circuit of Mathematical Model, and then in Model of based on the need for a diagnosis of all kinds of fault simulation to get corresponding data. Track circuit equivalent circuit model as shown in Figure1 Shown in the province went to the little track part.
Use uniform transmission line theory can be the track circuit the accurate modeling in modeling process in each unit module of four end network coefficient can be (17),(18)Calculation get.

According to each module of actual circuit structure can get track circuit the module of four end network parameters.Track Circuit of fault simulation can be by in Fault Location access resistance of style implementation change access resistance of resistance can be simulation fault of serious degree resistance infinite when equivalent to break resistance for Zero Equivalent to short-circuit.Figure 2,3 Shown in for use track circuit mathematical model simulation get of fault influence Curve.Limited to space only given part curve results by the curve can see different of fault on different of monitoring of has a different of influence this also show that for actual application process in May multi-fault situation.

1) RCSaid connection resistance by changeRCSimulation fault serious degree.
2) RGSaid ballast resistance simulation in Set1<RG<Natural 20OMEGAIn km(According to the field situation adjustment) when Ballast resistance belongs to normal state;RG>Natural 20OMEGAIn kmBelongs to ballast resistance increased;RG<1OMEGAIn km Belongs to ballast resistance reduce.
3) By parameters is set assurance training set data and test set data of different.
Actual when data format in will add sample collection of time a bar in to training add new fault sample when algorithm will according to the fault mode and sample collection of time will new sample. Appropriate location.

Due to system equipment physical topology on the connectivity and equipment of each other influence indoor Device 25A monitoring of and outdoor Device
A monitoring measuring tools have is high correlation and coupling.

3.3 Training Model

Based on the above equipment system and fault characteristics of describe this section will based on support vector machine methods to establish ZPW-2000 Track Circuit of intelligent fault diagnosis model; In model in two stay Optimization of parameters the Weight Parameters C And Radial-based nuclear function of parameters Sigma Which punishment parameters Main Influence"An interval maximum of hyperplane" And"Assurance data point deviation minimum" Between the weight; Radial-based nuclear Parameters Sigma Main Influence mapping after the data sample in high-dimensional space in the complex degree this two parameters of select will directly influence diagnosis model of generalization ability.

The parameters C And Sigma The correction and optimization can be divided into artificial style and automatic adjustment of style. If the artificial correction style the first training set data of diagnosis model the test if Model Diagnosis performance does not meet the requirements again on the above calculation get of parameters C And Gamma The appropriate artificial correction after again re-training model; General for parameters C To in Parameters Gamma To minor. But in actual test in artificial Correction Adjustment of errors of up 30% There error big, Trouble effort and difficult to meet the actual demand and Problems. Therefore, this paper based on genetic algorithm of parameters Regulation. Genetic algorithm is American J. Holland Professor put forward of a kind of by simulation natural evolution theory of optimization methods according to survival of the fittest and the principle of survival of the fittest-by-s evolution the more and more accurate of approximate solution final implementation Optimal Solution. Main including 3 A basic of operation: Select, Cross and Variation. First, "Select" Of Process main is individual string in accordance with their respective of adaptation value to select copy adaptation value equivalent to various biological In order to survival the necessary of ability decided to the individual string of select or not; "Cross" This a operation main can be divided into two steps: The first step is will select operation of individual random of two combined with the second step is the cross-and reproductive its results is constantly to recombinant to generate new of individual; finally, "Variation" Said random of to small of probability change a string of value and generate new of individual and take population of scale size Natural 20 Cross-probability 0.9 Evolution of algebra 100. Its algorithm process as shown in Figure 4 Shown in:

The Genetic Algorithm of Support Vector Machine Parameters Optimization get optimal C Value 2018 Optimal
GammaValue1.

Particle Group algorithm also known as birds foraging algorithm is a kind of new of evolution algorithm from random solutions to the by iterative evolution search optimal solution. Figure 5Is based on particle group algorithm SVM Parameters Optimization process.

Which PiFor ParticleI After the history best location, PGGroup in all particle the experience of best location the Genetic Algorithm of Support Vector Machine Parameters Optimization get optimal C Value 1987 Optimal Gamma Value 0.84.

3.4 Diagnosis test and Results Analysis

This system prototype diagnosis divided into two kind of mode: Off-line Diagnosis and real-time diagnosis. Off-line diagnosis is the record of history data the diagnosis. Real-time diagnosis is the Monitoring System Real-time Monitoring data the diagnosis model of signal centralized monitoring and transmission of monitoring data the diagnosis and have to visits Broken results.

In the diagnosis when can according to need to select previous training save of diagnosis model (whether contains probability output) such as select probability output of Form, diagnosis results also given each group sample belongs to the fault category of probability value in "Off-line Diagnosis" Mode under, in view of historical data sample quantity huge of reason don't will each sample of probability output value the graphics display user can according to own of interest select a sample the probability display.

In "Real-time diagnosis" Mode under diagnosis model will real-time display each monitoring sample diagnosis results of probability output, taking into account the diagnosis system without also no necessary to Monitoring System of sampling rate of Signal Monitoring System of each sampling data sample the diagnosis so can is set time window such 1 Minutes on the time window in the sample values average after again the diagnosis.

This research in the field to complete the no insulation Track Circuit ZPW-2000A Of the parts connection loose fault mode under monitoring points signal of test to analysis verify Track Circuit Theory Model of accuracy and the fault mode of the monitoring signal of Influence Law. By in different fault location access variable rheostat of Methods simulation fault its fault of degree by change resistance of size to Simulation. Experimental measured of monitoring data in 244 Group (2928A sample points) Which including different of fault mode and different fault of serious degree. The fault mode of the sample under the data are shown in the following table which ballast resistance including 3 OMEGA In km And 4 OMEGA In km Two.

<table>
<thead>
<tr>
<th>Fault mode</th>
<th>F1</th>
<th>F2</th>
<th>F4</th>
<th>F6</th>
<th>F7</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious degree</td>
<td>550</td>
<td>550</td>
<td>550</td>
<td>550</td>
<td>550</td>
<td>550</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>Sample quantity</td>
<td>Each fault mode Natural 20 Group in 220 Group</td>
<td>$3F5F8$</td>
<td>$3F12F14F15$</td>
<td>$F16$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to field-proven most situation under the measurement results and theory calculation results very close to which based on genetic algorithm and particle swarm optimization SVM Diagnosis accuracy respectively as shown in Figure 6 And figure 7 Shown in.

Average error of difference not very individual situation under Genetic Algorithm of maximum error is big but its in run time has the absolute advantage; in addition this experimental data total contains 2928A sample points however taking into account system equipment performance with time will happen migration in actual test in reduce training sample 40% When particle group algorithm of diagnosis accuracy quickly decreased 47% And genetic algorithm still keep 81% The diagnosis accuracy show the genetic algorithm in processing small sample when the high precision characteristics. Therefore priority select Based on Genetic Algorithm SVM Parameters Optimization. In addition in actual fault diagnosis process in on Data "To mean" And processing operation can get data of evolution trend further eliminate measured data and calculation data between the error.
4. Conclusion

This paper use Support Vector Machine on ZPW-2000 Track Circuit Equipment The the Fault Diagnosis Analysis and use genetic algorithm to the Optimization get optimal CValue 2018 Optimal Gamma Value 1. Use particle group algorithm get of get optimal CValue 1987 Optimal Gamma Value 0.84 For actual application system of development and development lay the foundation. ZPW-2000 Fault Diagnosis of Track Circuit Equipment. The main content of the follow-up research is to extend the research on Fault Diagnosis of other equipment of station signal, and then develop the practical application system.

References