Landslide Displacement Prediction of WA-SVM Coupling Model Based on Chaotic Sequence

Chao Zhou
China University of Geosciences (Wuhan), 388th, Lumo Road, Wuhan 430074, China
e-mail: zhouchao_rs@163.com

Kunlong Yin*
Professor, China University of Geosciences (Wuhan), 388th, Lumo Road, Wuhan 430074, China.
*Corresponding Author. e-mail: yinkl@126.com

ABSTRACT
Confronted with the chaotic characteristics of landslide displacement and the deficiencies of traditional time series prediction model, the wavelet analysis-support vector machine model (WA-SVM) based on chaotic time series for landslide displacement prediction is proposed. On the basis of the analysis of chaotic characteristics, landslide displacement is decomposed in to components with different frequency characteristics by wavelet analysis; Then each component is phase space reconstructed and predicted by the support vector machine (SVM) separately; Finally, the predicted value of the original sequence is obtained by superimposing all the components. Bazimen landslide in the Three Gorges Reservoir area is taken as an example, and a comparison between the proposed model and the wavelet analysis- BP Neural Network (WA-BP) and single SVM model is implemented, where the root mean square error (RMSE) of WA-SVM model is 10.45, the mean absolute percentage error (MAPE) is 1.06%, and the correlation coefficient is 0.989, which are better than the WA-BP and single SVM model. Results show that, the WA-SVM coupling model based on chaotic sequence is of high prediction accuracy, which is an effective prediction model for landslide displacement.

KEYWORDS: wavelet analysis, support vector machine, chaotic time series, phase space reconstruction, landslide displacement

INTRODUCTION
Landslide is a very significant geological hazard, which results in a large number of casualties and property losses every year. In final analysis, the research of landslide is to make a scientific assessment of its occurrence possibility and an accurate prediction of the time[1]. However, the landslide hazard prediction is still a worldwide problem in the exploration stage.
With the development of pattern recognition and artificial intelligence techniques, some of the non-linear models for landslide prediction have achieved great applications\cite{2}, such as gray model\cite{3}, Verhls model\cite{4}, neural network model\cite{5-6}, support vector machine model (SVM)\cite{7-9} and various coupling model (empirical mode decomposition-neural network model\cite{10}, etc.). Chaotic time series is a nonlinear prediction method developed in the late 1980s, which has been applied to the economic system, physiological system and power system effectively\cite{11}. In the study of landslide, the chaos theory is mainly used in the identification of chaotic characteristics and prediction of landslide displacement\cite{12}.

Landslide is a nonlinear dynamical system which is affected by geological conditions, groundwater, human engineering activities and many other factors\cite{2}. Seen as a complex multidimensional nonlinear system, the evolution of landslide displacement contains multi-level information, which makes it difficult to be predicted effectively with a single model\cite{10}. Wavelet analysis (WA) can deal with various non-linear signals effectively, and shows the ability in describing the local features of time series in both time and frequency domains\cite{13}. Wavelet transform can extract the trend, periodic and random characteristics of time series, it can smooth de-noising the chaotic sequence as well, which recently has been widely used for the prediction of chaotic time series.

The neural network technique provides a useful tool for the study of the prediction of landslide displacement-time series, but its theoretical basis is not perfect, the defects of which weakened the prediction ability, such as local minimum and over learned, etc.\cite{14} Support Vector Machine is a machine learning method, which is established based on the small sample statistical learning theory and the principle of structural risk minimization. SVM shows great advantages in solving the prediction of small sample and nonlinear regression\cite{15}. Considering the characteristics of landslide deformation and the advantages and disadvantages of forecasting model comprehensively, by the coupling application of Support Vector Machine model, the WA-SVM prediction model based on the principle of chaotic time series is proposed. This model is applied to the displacement prediction of Bazimen landslide in the Three Gorges Reservoir in this paper, and improved to be an effective method for the prediction of landslide displacement.

THE IDENTIFICATION OF CHAOTIC CHARACTERS

The phase space reconstruction of landslide displacement sequence

Phase space reconstruction is the base for the research of chaotic time series, which is aim to extract and recover the original laws of dynamic system from chaotic time series, and reconstruct an equivalent state space.
The landslide displacement sequence is defined as \( x_i, \ i = 1, 2, \cdots, n \); the appropriate delay time \( \tau \) and embedding dimension \( m \) are selected, and the one-dimensional displacement sequence is reconstructed as a multi-dimensional stage space, as shown in Eq. 1:

\[
y'_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, \cdots, x_{t+(m-1)\tau}\}
\]

where \( t = 1, 2, \cdots, n-(m+1)\tau \), \( y_t \) is the Phase point of \( m \) dimensional phase space.

By the Takens embedding theorem, when \( m \geq 2d + 1 \) (\( d \) is the dimension of dynamic system), the reconstructed dynamic system is topologically equivalent to original system. How to choose the appropriate embedding dimension and delay time is the key to the phase space reconstruction. Because of the presence of noise interference and estimated error in the actual displacement sequence, the delay time should not be too large\(^{[16]}\), and considering the limited length of the time series, 1 is generally selected as the delay time\(^{[17]}\). The commonly used methods for solving the delay time include autocorrelation and mutual information method. Similarly, the embedding dimension should not be too large as well, the commonly used methods for determining the embedding dimension in the practical application include test algorithm, saturated embedding dimension method and false neighboring points method. In this paper, the false neighboring point method is selected for calculating embedding dimension.

The identification of chaotic characters

The sensitivity of initial value of chaotic systems means that two initially proximal trajectories will diverge at an exponential rate in the phase space, while the Lyapunov index is to discriminate the chaotic characters of the system based on whether the diffusion motion features of the phase trajectory is existed or not. If the maximum Lyapunov exponent of the time series is greater than 0, the chaotic characters can be determined. In order to determine the chaotic characters, small data sets method\(^{[18]}\) is proposed to calculate the maximum Lyapunov exponent of the landslide displacement series.

THE THEORETICAL BASIS OF WA-SVM COUPLING MODEL

Wavelet Analysis

Wavelet analysis is a time-frequency localization analysis method with a window fixed in size but variable in shape, which makes a multi-scale refinement of the signals through the telescopic translation operation to focus on any signal details. Mallat algorithm is a fast wavelet transform algorithm based on multi-resolution analysis, including two parts of decomposition and reconstruction\(^{[19]}\). In this paper, Mallat algorithm is proposed to
decomposed landslide displacement sequence, then the sequence is decomposed into high-frequency $d_1$ and low-frequency $a_1$ through the wavelet transform, the low frequency part is decomposed again, any scale on the high frequency and low frequency can be obtained after the repeated processes. In general, decomposition for 3 or 4 layers can achieve the ideal result. By comparison and analysis, the landslide displacement series is decomposed into 4 layers in this paper.

**Support Vector Machine**

SVM (support vector machine) is a non-linear regression forecasting method proposed by Vapnik et al. in 1995. The input variables are mapped into a high-dimensional linear feature space through a non-linear transformation, then the optimal decision function is constructed. The dot product operation in the higher dimensional feature space is replaced by the kernel function in original space, and through the training of the finite sample, the global optimal solution will be obtained[15]. The regression function for SVR is:

$$f(x) =< W \cdot \Phi(x) > + b$$

Transforming the estimation function into function minimization problem by the $\varepsilon$ insensitive loss function:

$$R_{\min} = \frac{1}{2} ||W||^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i^*)$$

The constraint conditions are:

$$\begin{cases} 
W^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i; \\
y_i - W^T \phi(x_i) - b, \leq \varepsilon + \xi_i^*; \\
\xi_i, \xi_i^* \geq 0, i = 1, \ldots, l. 
\end{cases}$$

where $C$ is penalty factor; $\xi_i$ and $\xi_i^*$ is relaxation factor; $b$ is offset quantity. The Lagrange multiplier is introduced at last, and with the Wolf duality theory, it is converted into an equivalent dual problem as follows:

$$\min \frac{1}{2} (\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (\alpha_i - \alpha_i^*)$$

The constraint conditions are:

$$\begin{cases} 
\sum_{i=1}^{l} (\alpha_i + \alpha_i^*) = 0; \\
0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2 \ldots l. 
\end{cases}$$
where $Q_{ij} = k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$.

SVR regression prediction model can be obtained by quadratic programming:

$$f(x, \alpha^*, \alpha_i) = \sum_{i=1}^{l} (\alpha^*_i - \alpha_i) K(x_i, x) + b$$  \hspace{1cm} (7)$$

where $K(x_i, x)$ is the kernel function of SVM. There are four commonly used kernel functions currently: linear kernel, polynomial kernel function, radial basis kernel function (RBF) and sigmoid function.

**WA-SVM COUPLING PREDICTION MODEL**

WA-SVM coupling prediction model is based on the chaos theory. At first, the delay time $\tau$ and embedding dimension $m$ of the landslide displacement time series are determined, and then phase space reconstruction and the identification of chaotic character are made as well, which are able to provide theory basis for the establishment of the model; Secondly, the time series of landslide displacement is decomposed into low-frequency part $CA$ and high frequency part $CD$, ($t$ is decomposition scale) by wavelet decomposition, and through making phase space reconstructions of $CA$ and $CD$, respectively, $A$ and $D$, are obtained; Finally, $A$ and $D$, are predicted respectively by SVM, and the final predicted value can be obtained through the superposition of all components. The flowchart of the model is shown in Figure 1.
The commonly used indicators for prediction accuracy test include root mean square error $RMSE$ (Eq. 7) and the mean absolute percentage error $MAPE$ (Eq. 8). It’s not considered a credible result if the values of $RMSE$ and $MSE$ are very large, the result is effective only if the values are both small. However, the indicator values are closely related to the observation data\textsuperscript{11}. The correlation coefficient (Eq. 9) is one test indicator independent to observation data, which can reflect the degree of correlation between variables. Therefore, correlation coefficient is applied to the test of the prediction accuracy of the model as well.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{x}_i - x_i)^2}$$  \hspace{1cm} (7)
\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right|
\]

(8)

\[
R_{xx} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (\hat{x}_i - \bar{\hat{x}})^2}}
\]

(9)

where \( x_i \) is measured value; \( \hat{x}_i \) is predicted value; \( N \) is the number of predicted values; \( \bar{x} \) is the mean of measured values; \( \bar{\hat{x}} \) is the mean of predicted values.

CASE STUDY

Introduction of Bazimen landslide

Bazimen landslide is located in the west side of Xiangxi River in Zigui county, Three Gorges Reservoir in China, which is the chief tributary of the Yangtze River. The volume of the landslide is \( 2\times10^6 \) m\(^3\), with 380m in the maximum longitudinal length, and 100~300m in width, the elevation of the leading and trailing edges are 110m and 250m respectively. Seen from the borehole data (Figure 3), the sliding body is mainly composed of Quaternary deposits and fragmented rubble, and two sliding surface levels were found: The lower one was the initial sliding surface at the interface between deposit and bedrock at a depth of 10 to 35 m. The upper sliding surface was a secondary sliding surface at a depth of 6 to 18 m at the top and middle of the sliding mass and at a depth of 27 to 33 m at the bottom. The secondary sliding surface formed after the early deformation phase along the initial sliding surface.

There are three stacking platforms on Bazimen landslide. The first and second platforms are of small scales and distribution areas, the elevation are 222~225m and 202~205m respectively, and the slope is 5°~10°, which are the serious deformation parts of the landslide, the monitoring point ZG111 of larger surface displacement is just on the junction area of the two steps; The third platform is most widely distributed throughout the landslide, with the horizontal distance of 100~150m and 140~162m in elevation, ZG110 and ZG112 are both on the third steps, the surface displacement of which is small relatively[20].
The deformation data in different parts of the landslide can reflect the interactions between various parts of the sliding bodies, therefore, it contributes to analyzing the developing process and evolution trend of landslide to understand the relationship between them. Seen from the displacement monitoring curves, Bazimen landslide is an advancing landslide with serious deformations in the trailing part, so the monitoring point on the most
seriously deformed trailing part of the landslide (ZG111) is selected to establish the prediction model. The data from August 2004 to December 2007 is taken as the training sample, while the data from January 2008 to December 2008 as the prediction test sample.

**Figure 4:** Curves of cumulative displacement of GPS monitoring points

**The chaos identification of landslide displacement sequence**

*The phase space reconstruction of landslide displacement sequence*

In this paper, the delay time $\tau$ of Bazimen landslide displacement sequence is 1, while the Lyapunov exponent $m$ is calculated by the false neighboring point method, the results are shown in Figure 5. According to the chaos theory, when the percentage of the false nearest point is smaller than 5% or the false neighboring point is no longer reduced with the increase of $m$, $m$ is the best embedding dimension in this condition\(^{[11]}\). Seen from figure 5, the best embedding dimension of Bazimen landslide sequence is 3.
Figure 5: The result of FNN calculating embedding dimension

The chaos characters determination of landslide displacement sequence

The maximum lyapunov exponent of the displacement series of Bazimen landslide is calculated by small data sets method (figure 6). \( i \) is the dispersion stepping time, and \( y(i) \) is the index divergence rate of all reference points corresponding to \( i \). Selecting a section of linear region from \( y(i) \sim i \), and deriving the regression line with the use of least square method, the slope of the line is the largest Lyapunov exponent \( \lambda_1 \). The largest Lyapunov exponent of Bazimen landslide displacement series is 0.0967 (greater than zero), so the sequence shows chaotic characters. Therefore, the WA-SVM model based on chaos theory can be used for the displacement prediction of Bazimen landslide.

Figure 6: The calculation chart of Lyapunov value
The prediction of landslide displacement

With the method of Mallat algorithm, db4 wavelet function is proposed to decomposed landslide displacement sequence into low frequency sequence \( a_4 \) and high frequency sequence \( d_4, d_3, d_2, d_1 \), in order to eliminate the influence of dimension, the data are normalized, as shown in figure 7.

![Wavelet decomposition of each component of the normalized sequence](image)

**Figure 7:** Wavelet decomposition of each component of the normalized sequence

1 is taken as the delay time of decomposed component sequences, and the embedding dimensions are calculated separately with the saturated embedding dimension. When predicting with the SVM method, different embedding dimensions are attempted to achieve the best prediction effect\(^{[12]}\). The results show that, the best embedding dimensions of low frequency sequence \( a_4 \) and high frequency sequence \( d_4, d_3, d_2, d_1 \) are 3,3,4 and 4. Each component is predicted by the SVM model separately, and the final result can be obtained by anti-normalizing and superimposing the predicted values. Making a composition of the WA-SVM model, the WA-BP model and single SVM model, the final results are shown as figure 8, figure 9 and Table 1.
Figure 8: The comparison of WA-BP and WA-SVM prediction model

Figure 9: The comparison of WA-SVM and SVM prediction model

Table 1: The comparison of predicted results of each model

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAPE (%)</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA-SVM</td>
<td>10.4555</td>
<td>1.0619</td>
<td>0.989</td>
</tr>
<tr>
<td>WA-BP</td>
<td>11.5678</td>
<td>1.3677</td>
<td>0.987</td>
</tr>
<tr>
<td>SVM</td>
<td>24.6562</td>
<td>2.4131</td>
<td>0.941</td>
</tr>
</tbody>
</table>
As shown in Table 1, WA-SVM mode is superior to WA-BP model in the main accuracy evaluation indexes, which demonstrates that SVM has a very good generalization performance. The prediction accuracy of SVM model is much lower than WA-SVM and WA-BP model, which demonstrates that, it is effective to improve forecast accuracy by extracting different features of landslide displacement sequence with wavelet decomposition, and predicting sequences of different characteristics respectively.

Seen from Figure 8 and Figure 9, predicted values of WA-SVM model keep in a great agreement with measured values, which indicates that the WA-SVM model based on chaos theory is of high prediction accuracy. Therefore, the proposed model is an effective model for the prediction of landslide displacement.

CONCLUSION

Landslide is a complex nonlinear system, on the basis of chaotic time series analysis of landslide displacement, the WA-SVM prediction model based on chaotic time series is proposed in this paper. By the application of wavelet analysis theory, landslide displacement sequence is decomposed into components with different characteristics in this model, and each of the components is phase space reconstructed and predicted by the support vector machine (SVM) model respectively. Results show that, the proposed model can effectively improve the prediction accuracy, which is an effective prediction model for landslide displacement.

Comprehensive influenced by the geotechnical property, the scheduling of reservoir water level and the intensity of rainfall, landslide displacement sequence is a chaotic series with multiple features. Wavelet analysis can extract different characteristic components of landslide displacement sequence, and smooth de-noising the chaotic sequence while simplifying the prediction as well, which is able to improve the prediction accuracy effectively.

ACKNOWLEDGMENT

This research work is supported by the National Natural Science Foundation of China (No. 41240023), and the China Geological Survey Project (No. 1212011220173).

REFERENCES


© 2014 ejge