Improved support vector machine regression in multi-step-ahead prediction for rock displacement surrounding a tunnel

B. Yao\textsuperscript{a,*}, J. Yao\textsuperscript{b}, M. Zhang\textsuperscript{a} and L. Yu\textsuperscript{c}

\textsuperscript{a} School of Automotive Engineering, Dalian University of Technology, Dalian, 116024, P.R. China.
\textsuperscript{b} School of Civil Engineering, Beijing Jiaotong University, Beijing, 100044, P.R. China.
\textsuperscript{c} Yanqing Institute of Technology, Beijing, 065201, P.R. China.

Received 6 July 2013; received in revised form 6 September 2013; accepted 26 November 2013

KEYWORDS
Multi-step-ahead prediction; Tunnel; Surrounding rock displacement; SVM; Forgetting factor.

Abstract. A dependable long-term prediction of rock displacement surrounding a tunnel is an effective way to predict rock displacement values in the future. A multi-step-ahead prediction model, which is based on a Support Vector Machine (SVM), is proposed for predicting rock displacement surrounding a tunnel. To improve the performance of SVM, parameter identification is used for SVM. In addition, to treat the time-varying features of rock displacement surrounding a tunnel, a forgetting factor is introduced to adjust the weights between new and old data. Finally, data from the Chijiangchong tunnel are selected to examine the performance of the prediction model. Comparative results presented between SVMFF (SVM with a forgetting factor) and an Artificial Neural Network with a Forgetting Factor (ANNFF) show that SVMFF is generally better than ANNFF. This indicates that a forgetting factor can effectively improve the performance of SVM, especially for time-varying problems.

© 2014 Sharif University of Technology. All rights reserved.

1. Introduction

1.1. Background

Rock deformation surrounding a tunnel and tunnel lining cracks may lead to instability of the tunnel. When analyzing the stability of the surrounding rock mass of a tunnel, deformation is often used as the security index to denote the stability of the surrounding rock mass of the tunnel. The rate of deformation around the tunnel depends on geological and geotechnical conditions, and many techniques have been presented to estimate the conditions of the tunnel during tunnel construction. The need for statistical process control is important for quality assurance [1]. An efficient way is to use displacement statistics and analyses of the rock surrounding the tunnel to estimate the future displacement of the surrounding rock. However, the conditions of a tunnel during construction are varied, which will have an effect on the displacement of the tunnel. If the future displacement values of the surrounding rock of the tunnel could be predicted for reference, it could help to quantitatively evaluate the stability of the surrounding rock mass. Moreover, project managers could identify tunnel conditions and effectively operate their construction facilities. Furthermore, when the weak rock layer begins to become unstable and these failures could be acquired in advance, some corresponding measures could be taken to avoid these dangers.

Thus, displacement of tunnel surroundings is a common way to reflect the condition of the tunnel, and finding an effective method to predict the displacement of the surrounding rock of the tunnel is important.
for planning and management activities during tunnel construction. Displacement prediction during tunnel construction has intrigued experts for years. Lack of prediction will result in great difficulties in completing underground construction, with increased costs and delays. Some literature about the displacement prediction of the surrounding rock of a tunnel can be found in Li et al. [2] and Sellmer [3].

1.2. Literature review
From the standpoint of engineering applications, it is necessary to learn the values to be predicted many time-steps into the future. During the past few decades, a great deal of research has been devoted to multi-step-ahead (MS) techniques to deal with these problems [4-6]. In this study, the MS technique is also used to predict the displacement of surrounding tunnel rock for planning and management activities during tunnel construction. However, ground conditions change greatly during tunnel construction, which makes it more difficult to predict the displacement of rock surrounding tunnels accurately. The Support Vector Machine (SVM) is a relatively new kind of learning machine which has been applied successfully to some time series forecasting problems [7-11]. The numerical results indicate that SVM shows much resistance to the overfitting problem and can provide a high generalization performance. These successful applications suggest that SVM is an acceptable tool to provide accurate displacement prediction of rock surrounded tunnels. Unlike the empirical risk minimization principle implemented in most traditional ANN models, SVM implements the structural risk minimization principle. The most important feature in the structural risk minimization principle is minimizing an upper bound to the generalization error instead of minimizing the training error. The parameter, C, can be adjusted to control the tradeoff between errors of training data and margin maximization. For an SVM, the value of ε in the ε-insensitive loss function affects the complexity and the generalization capability of SVM. Fixing the parameter, ε, can be useful for specifying the desired accuracy of the approximation in advance. Another parameter, σ, is the range at which the generalization performance is stable. Considering the parameters will greatly affect the performance of SVM, and some literature has attempted to determine the proper parameter values for these problems. Hou and Li [12] proposed an evolution strategy with covariance matrix adaptation to determine the parameters in SVM. Hsu et al. [13] introduced a grid-search to determine the adaptive values for the parameters in SVM. Lin et al. [14] presented SVM for hydrological prediction, and a Shuffled Complex Evolution Algorithm (SCE-UA) was used to identify appropriate parameters in SVM. Lorena and Carvalho [15] optimized the parameter values for SVM based on genetic algorithms. Thus, this paper also applies some methods like grid-search. SCE-UA and GA to determine the appropriate parameters for SVM.

1.3. Contributions
The purpose of this paper is to build on the prediction model by Yao et al. [16], and extend it to predict rock displacement surrounding a tunnel, which is related to long-term prediction. Thus, the contributions of this paper are to apply a multi-step-ahead prediction for rock displacement surrounding a tunnel based on SVM, which has been successfully applied in the literature [16-17]. Then, we use a forgetting factor to improve the prediction accuracy of SVM. In addition, to examine the performance of this algorithm, a comparison between SVMFF (SVM with a Forgetting Factor) and an Artificial Neural Network with a Forgetting Factor (ANNFF) is used in this study.

This paper is organized as follows. In Section 2, we describe the MS prediction problem of rock displacement surrounding a tunnel, the SVM model for MS prediction, parameter identification for SVM and a brief introduction to SVM with a forgetting factor. In Section 3, some computational results are discussed and, lastly, the conclusions are provided in Section 4.

2. SVMFF for predicting rock displacement surrounding a tunnel

2.1. Formulation of the prediction
The properties of the prediction of rock displacement surrounding a tunnel with MS techniques not only depend on the observation values but also on the previous prediction. Thus, the recursive relation between inputs and outputs in MS prediction can be expressed using general nonlinear input-output models, as the following:

\[
\begin{align*}
\hat{x}_{t+p} &= F(x_{t+p-m}, \ldots, x_t, \hat{x}_{t+1}, \ldots, \hat{x}_{t+p-2}, \hat{x}_{t+p-1}) \\
&= p < m \\
\hat{x}_{t+p} &= F(\hat{x}_{t+p-m}, \ldots, \hat{x}_{t+p-2}, \hat{x}_{t+p-1}) \\
&= p \geq m
\end{align*}
\]

where \( p \) is the number of steps ahead of the \( p \)-step-ahead prediction model, \( F(\cdot) \) (the horizon of MS prediction): \( m \) is defined as the number of inputs; \( \hat{x}_{t+p} \) which has a “hat”, represents an estimate of the output at time-step \( t + p \); and \( x_{t+p-m} \), without a “hat”, represents an observation. Obviously, if \( p < m \), the model input consists of observation and prediction values, and if \( p \geq m \), it consists of all prediction values.

Referring to multi-step prediction [18-19], the one-step prediction and multi-step prediction for tunnel surrounding prediction can be described in Figure 1.
2.2. SVM for regression

SVM is a machine learning method proposed by Vapnik [20, 21]. Given a set of data points, \( \{x_k, y_k\} \), \( k = 1, 2, ..., s \), \( x_k \in \mathbb{R}^n \), \( y_k \in \mathbb{R} \), \( k \) is the number of training samples. SVM estimates the function by the following function:

\[
f(x) = \langle w, x \rangle + b, \quad w, x \in \mathbb{R}^n, \quad b \in \mathbb{R},
\]

(2)

Here, \( \langle w, x \rangle \) is the feature of the inputs. The coefficients, \( w \) and \( b \), are estimated by the regularized risk functional.

To get the estimation of \( w \) and \( b \), Eq. (2) can be transformed to a primal objective function:

\[
\text{Min } J = \frac{1}{2} ||w||^2 + C \sum_{i=1}^{s}(\xi_i^+ + \xi_i^-),
\]

s.t. \( y_k < \langle w, x \rangle + b \leq \varepsilon + \xi_i^+ \)

\( y_k - \langle w, x \rangle + b - \varepsilon \leq \xi_i^- \)

\( \xi_i^+, \xi_i^- \geq 0 \)

(3)

where \( C \) is a regularization constant to determine the trade-off between training error and generalization performance; \( \varepsilon \) is a tube to define the range of the observation and prediction values. Both \( C \) and \( \varepsilon \) are user-determined parameters. Two positive slack variables, \( \xi_i \), \( \xi_i^- \), are used to cope with infeasible constraints of the optimization problem. The formula is an optimization problem and its minimum can be evaluated by Lagrange multipliers, \( \alpha_i \) and \( \alpha_i^- \), in most cases:

\[
\text{Max } J = -\frac{1}{2} \sum_{i,j=1}^{s}(\alpha_i^- - \alpha_i)(\alpha_j^- - \alpha_j) < x_i, x_j >
\]

\[+ \sum_{i=1}^{s}\alpha_i^+(y_i - \varepsilon) - \sum_{i=1}^{s}\alpha_i(y_i + \varepsilon), \]

s.t. \( \sum_{i=1}^{s}\alpha_i = \sum_{i=1}^{s}\alpha_i^+ \)

\( 0 \geq \alpha_i \geq C \)

\( 0 \geq \alpha_i^- \geq C \)

(4)

Let \( w = \sum_{i=1}^{s}(\alpha_i - \alpha_i^-)x_i \).

(5)

Thus, \( f(x) = \sum_{i=1}^{s}(\alpha_i - \alpha_i^-) < x, x_j > + b \).

(6)

By introducing kernel function, \( K(x_i, x_j) \), Eq. (4) can be rewritten as follows:

\[
f(x) = \sum_{i=1}^{s}(\alpha_i - \alpha_i^-)K(x_i, x_j) + b,
\]

(7)

where \( K(x_i, x_j) \) is the kernel function which is proven to simplify the use of mapping. The value of \( K(x_i, x_j) \) is equal to the inner product of two vectors, \( x_i \) and \( x_j \) in the feature space \( \varphi(x_i) \) and \( \varphi(x_j) \), that is \( K(x_i, x_j) = \varphi(x_i)\varphi(x_j) \). By the use of kernels, all necessary computations can be performed directly in the input space, without having to compute the map, \( \varphi(x) \). More details on SVM can be seen in [7, 20].

2.3. Application of SVM for prediction of rock displacement surrounding a tunnel

Due to the rate of rock mass deformation varying with geological and geotechnical conditions, it is difficult to predict the displacement of rock surrounding a tunnel. The MS technique based on SVM is adopted to predict the future displacements with \( m \) observation data. In other words, the \( m + 1 \) value will be predicted based on the \( m \) observation data. The predicted output will be an input to the following predictions. Thus, future values will be recursively predicted in the same way. A formal definition of MS based on SVM can be summarized as follows:

1. A one-step-ahead predictor based on SVM is performed to provide a prediction output.

2. Future values can be predicted based on the recursive principles of \( p \)-step-ahead prediction by embedding a one-step-ahead estimator based on SVM.

Since tunnel conditions are complicated, rock
displacement surrounding a tunnel will be changed as
time goes on. However, a standard SVM does not
consider the time-varying features of the data, since
it refers to the data of memory (window) with the
same weights. A forgetting factor, $\lambda^i(0 < \lambda^i \leq 1, 1 \leq i \leq m)$, is applied to put exponentially less emphasis
on past data. In this study, $\lambda^i$ is used to reflect the
weights between new and old data, that is, the weight
of former displacements (e.g. $m, m - 1, \ldots$). When
$\lambda = 1$, the SVM with the forgetting factor is the
same as the SVM without the forgetting factor. When
$\lambda = 0$, only the former rock displacement
surrounding the tunnel is used to predict current displacement.
Other former displacements have no influence on the
current displacement. Thus, the structure of SVMFF
is illustrated in Figure 2.

3. Case study

The Chjiangchong tunnel of the Wuhan-Guangzhou railway, which is a high-speed rail line between Wuhan
and Guangzhou city in China, is chosen as the
study site. The length of the tunnel is about 385 m
and its location is from DK1650+720 to DK1660+105
of the Wuhan-Guangzhou railway. Three sections of
the tunnel were selected to acquire the data on the
rock displacement surrounding the tunnel. In general,
the measurement frequency should be once per day at
the beginning of the experiment and the frequency may
be once every other day later [22]. This is because the
deformation rate at the beginning of the experiments is
obviously more than that later. Thus, in this paper,
the measurement frequency is once every day in the
first thirteen days, and then, once every other day
after the thirteenth day. The experiment continues
until the rock displacement surrounding the tunnel is
almost stable. Here, we take the difference between
two consecutive measurements, $< 0.1$ mm, as the
termination condition. Thus, we acquired three sets
of data; each set with thirty-two samples from the
experiment, from September 8 to October 28, 2007.

![Figure 2. Structure of SVMFF model.](image)

### Table 1. The parameters in GA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$(p_e)$</th>
<th>$(P_m)$</th>
<th>$(P_{i.e})$</th>
<th>$(T_{max})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>0.6</td>
<td>0.05</td>
<td>80</td>
<td>1000</td>
</tr>
</tbody>
</table>

### Table 2. The parameters in SCE-UA.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$p$</th>
<th>$q$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>4</td>
<td>11</td>
<td>1</td>
<td>21</td>
</tr>
</tbody>
</table>

3.1. Result comparison with grid-search/SCE-UA/GA for parameter identification

The parameter selection is important for the performance of the algorithm. In this section, the objective
is to identify good parameters $(C, \varepsilon, \sigma)$ in order to predict the unknown data accurately. Referring to
previous literature on parameter selection for SVM, some methods, like grid search [13], SCE-UA [14] and
genetic algorithm [15], were used to optimize the values of the parameters for SVM.

The Genetic Algorithm (GA) is inspired by evolu-
tional biology, like inheritance, selection, crossover,
and mutation. A fitness function is used as a measure
for determining the relative superiority of one solution
compared to a second solution. Then, GA attempts to
retain relatively good genetic information from
generation to generation.

The SCE-UA algorithm attempts to look for the
optimum solution by combining the strengths of the
simplex procedure, deterministic and probabilistic
approaches, competitive evolution and complex shuffling.

Thus, these three methods are also analyzed for
parameter optimization for our problem, respectively.
The parameters of GA and SCE-UA can be seen in
Tables 1 and 2. To test the performance of the three
methods, it is necessary to divide all the data into
different sets. Here, the data are divided into three
sets: training samples, testing samples and inspection
samples. In this study, the data from the first section
and the third section are used for training and testing,
and the data from the second section are used for
inspection. To reduce the search space, according to
previous literature [13], the constraints of the three
parameters are suggested as $C \in [2^{-10}, 2^6], \varepsilon \in [2^{-13}, 2^{-1}]$
and $\sigma \in [0, 2]$. Then, to examine the prediction errors,
an objective function should be considered:

$$RMSE = \left( \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - p} \right)^{1/2},$$

where $n$ is the number of testing samples, $p$ is the
number of model parameters which refers to the liter-
ature [23]. Then, the three methods continue running
ten times under the same condition. The average
solution, the number of the best solutions and average
computation time of the three methods are shown in
Table 3.
Table 3. Results obtained by three methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average solution</th>
<th>The number of the best solutions</th>
<th>Average computation time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid search</td>
<td>1.712</td>
<td>10</td>
<td>55.27</td>
</tr>
<tr>
<td>SCE-UA</td>
<td>1.733</td>
<td>6</td>
<td>26.57</td>
</tr>
<tr>
<td>GA</td>
<td>1.726</td>
<td>8</td>
<td>31.15</td>
</tr>
</tbody>
</table>

By comparing the results from the three methods, it is obvious that all three methods can attain the best solution. However, the grid search attained the ten best solutions in the ten computing times. The GA acquired the best solutions 8 times, while SCE-UA attained the best solutions 6 times. This is because the grid search method simply determines the word search solver at each point on the grid of the parameter values. It may offer some protection against local minima, while the other two algorithms often tap into local minima. However, it is not very efficient. When more parameters are included in the model, the number of determinations can be excessive. When compared with SCE-UA, GA seems more suitable to determine the parameters for SVM, for our problem. It is because GA is a search heuristic inspired by natural evolution, which is often used to generate useful solutions to optimization. In addition, compared with the average solutions of GA and SCE-UA, the stabilization performance of SCE-UA (in which the difference between the best solution and the average solution is about 1.2%), is the worst of the three methods. However, the average time of the grid search is the highest. This could be due to the fact that the grid search computes the performance at all combinations of \( C, \varepsilon, \sigma \) to get the performance surface. We can also see that the time consumed by GA is more than that of SCE-UA. This may be because it is a fitness-based process, and using natural evolution, such as inheritance, mutation, selection, and crossover, will involve high computational time. Thus, for the practical prediction model of rock displacement surrounding the tunnel, GA is used to determine the parameters for SVM, and the optimal values are attained as \( C = 6.4371, \varepsilon = 0.0032, \sigma = 1.4129 \).

3.2. The determination of forgetting factors and the number of prediction steps

The forgetting factor is used to weigh the influence from the new and the old data. The choice of the forgetting factor is typically a compromise between the ability to track changes in the parameters and the need of suppressing stochastic behaviors of the estimates. A large forgetting factor (effectively a large memory of data) is used when the learning is in the steady state and there is no obvious model variation, while a small one (to fade away the very old data) is applied when the model error is large. A too large and too small forgetting factor will affect prediction accuracy and, even worse, the prediction ability of SVM. In addition, in the multi-step-ahead prediction model, the number of prediction steps is an important factor affecting prediction accuracy. It can be due to the fact that the MS prediction technique needs the recursive use of Single-Step (SS) predictors for reaching the endpoint. Even small errors from preceding predictions are accumulated and propagated, thus, resulting in poor prediction accuracy in following predictions. To determine the value of forgetting factors and the number of prediction steps, the prediction errors of SVMFF under various forgetting factor and prediction steps are shown in Figure 3.

From Figure 3, it can be observed that the changing trends of RMSEs on the two sections are almost the same, and prediction accuracy is the highest when prediction step is \( n = 6 \) and forgetting factor is \( \lambda = 0.85 \). When the forgetting factor is from 0.8 to 0.9, prediction accuracy is relatively less than others. Furthermore, as we predict ahead, the errors accumulate and propagate. There are both increasing trends of RMSEs of the displacement predictions on two sections, as the increase of the step ahead. However, at the beginning, the prediction errors increase slightly with the increase of the step ahead, while they start to increase greatly after the number of prediction steps is 8. It can be attained that too long a step will increase prediction errors. Therefore, the forgetting factor and the number of prediction steps ahead are determined as 0.85 and 6 in this study, respectively.
3.3. Results

To further analyze the characteristics of the MS prediction for the tunnel surrounding rock, we select the 6th to the 12th data from the second section, as test bed I (note: the first five data as the inputs), and the 27th to 33rd data from the second section, as test bed II. Here, the two test beds reflect, respectively, two typical cases: One is the phase where tunnel surrounding displacement increases obviously, and the other is the phase when tunnel surrounding displacement changes smoothly.

Then, to evaluate the performance of the proposed model, a three-layer ANN model with forgetting factor (ANNFF) is also introduced in this paper. To get a good comparison, the same input and output variables of ANNNF should be the same as the SVMFF. In the ANNNF, a scaled conjugate gradient algorithm [24] is employed for training. To prevent overtraining and improve the generalization ability, the hidden neurons are generally optimized by a trial and error procedure. In this study, the final ANNNF architecture consists of five hidden neurons that yield the best performance. Then, we compare the performance of the SVMFF with that of ANNNF using RMSE. Figure 4 depicts the prediction performance of the two models on the two test beds. It can be found that the two models obtain more accurate values at the former rather than at the latter data in each test bed. It can be attributed to the fact that the MS predictor is based on the recursive use of the SS predictor for reaching the end-point on the horizon. The prediction errors at the beginning of the horizon accumulate and propagate till the end prediction. It is true that prediction errors increase. However, in some real-world applications, especially for the prediction of rock displacement surrounding a tunnel, which has a relatively smaller time period, it requires enough time to take preventive measures to combat danger. Thus, it is acceptable to adapt to MS techniques to predict future displacements.

The relationship between observations and predictions for the two test beds is also illustrated in Figure 5. It is obvious that the errors from SVMFF models, generally, are smaller than those of ANNNF. This can be explained by the fact that SVMFF uses the structural risk minimization principle to minimize the generalization error, while ANNNF uses the empirical risk minimization principle to minimize the training error. Furthermore, SVMFF always seeks to find the global solution, while ANNNF may tend to fall into a local optimal solution. Therefore, it is feasible to use our model to solve the displacement prediction of rock surrounding a tunnel.

4. Conclusions

The deformation of rock mass surrounding a tunnel is an effective factor to reflect the stability of a tunnel. Proving accurate displacement prediction in advance is a method to identify potential danger and some effective measures are used to reduce losses. In this paper, an MS prediction based on SVM is used to predict surrounding rock deformation. To improve the training efficiency of SVM, grid-search, SCE-UA and GA are implemented for determining the parameters in SVM. To deal with the time-varying features of rock displacement surrounding a tunnel, a forgetting factor is used to weigh the effects from old and new data. In addition, considering the fact that long-time prediction by the MS technique will worsen the prediction errors, an experiment from the Chijianghong tunnel is applied to determine the prediction horizon. The result shows that 7 is the suitable prediction horizon for our problem.
in this study. Then, compared with ANNFF, the proposed SVMFF can provide a better performance in most situations than ANNFF. Thus, SVMFF has proved an effective method for prediction of rock displacement surrounding a tunnel.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51206079), the Major State Basic Research Development Program of China ( ‘973’ Program: 2013CB036203), Fundamental Research Funds for the Central Universities 2011JBM009 and the Fundamental Research Funds for the Central Universities (3013-S52019).

References


Biographies

**Baozhen Yao** received her PhD degree from Beijing Jiaotong University, Beijing, China, in 2011, and is currently postdoctoral research student in the School of Automotive Engineering at the Dalian University of Technology, China. Her current research interests include artificial intelligence, and logistics systems.

**Jinbao Yao** received his PhD degree, in 2010, from Beijing Jiaotong University, Beijing, China, where he is currently lecturer in the School of Civil Engineering & Architecture. His research interests include vibration effects on buildings, field measurement of environmental vibrations, environmental vibration monitoring and evaluation, and engineering project management.

**Mingheng Zhang** received his PhD degree from Jilin University, Changchun, China, in 2007. Currently he is lecturer in the School of Automotive Engineering at Dalian University of Technology, China. His research interests include artificial intelligence, intelligent Vehicle navigation, image processing.

**Lan Yu** received her MS degree from Helsinki School of Economics, Finland, and is currently lecturer at the Yanshan Institute of Technology, China. Her current research interests include artificial intelligence, and data processing and analysis.