Engineering Geology 218 (2017) 213–222

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

Comparison of two optimized machine learning models for predicting displacement of rainfall-induced landslide: A case study in Sichuan Province, China

Xing Zhu [a,b], Qiang Xu [a,*], Minggao Tang [a], Wen Nie [b], Shuqi Ma [b], Zhipeng Xu [c]

[a] State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, PR China
[b] School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639789, Singapore
[c] China Railway Academy Co., Ltd., Chengdu 611731, PR China

ARTICLE INFO

Article history:
Received 20 August 2016
Received in revised form 15 January 2017
Accepted 18 January 2017
Available online 19 January 2017

Keywords:
Genetic Algorithm
Least Squares Support Vector Machines
Double Exponential Smoothing
Landslide
High-accuracy prediction

ABSTRACT

Evaluation and prediction of displacement by specific models help in forecasting geo-hazards. Among the various available predictive tools, Least Square Support Vector Machines (LSSVM) model optimized with Genetic Algorithm, namely GA-LSSVM, is commonly used to empirically forecast landslide displacement due to its capability of processing non-linear complex systems. Another improved hybrid model composed of Double Exponential Smoothing (DES) and LSSVM considers measured displacement and precipitation time series to estimate the one-step ahead displacement evolution of rain-induced landslide. Here, the modelling process and accuracy of these two models are presented, and their predictive performances are evaluated by the root mean squared error (RMSE), mean absolute percentage error (MAPE), accuracy factor (AF), and correlation coefficient (R). A slowly-moving landslide on gently dipping rocky slope located in Sichuan Province of China was chosen as the case study for its deformation triggered by intense seasonal rainfall. The application results indicated that both GA-LSSVM and DES-LSSVM models were suitable for accurately predicting the landslide displacement on the basis of precipitation and displacement observations. Furthermore, comparison results show that DES-LSSVM model can provide the better predictive accuracy, with RMSE and MAPE values of 0.059 mm and 0.004%, respectively.

1. Introduction

Landslide is one major worldwide geo-hazard, causing massive casualties and property damage every year. Prediction of the evolution process of landslide is an important issue of the safety assessment and dynamical behaviour investigation for landslide under the influence of external factors, for example, precipitation, water, earthquake, and human activities. It is well known that the evolution process of landslide is a complex non-linear process that is caused by the complex interaction of different factors (Huang et al., 2005; Zhang et al., 1994), e.g. the complicated geological settings, varying hydrological conditions. Displacement time series are generally appreciated as the direct representation of the complex and non-linear dynamical behaviour of landslide. Monitoring, prediction and early warning of landslide displacement are the effective and reliable methods in use to reduce the risk of landslide failure on human's lives and infrastructure.

Recently, numerous models have been proposed and widely used for landslide displacement, such as functional regression (Samui and Kurup, 2012; Yin et al., 2007), Artificial Neural Network (ANN) (Chen et al., 2015b; Jiang and Chen, 2016; Lian et al., 2015; Lv and Liu, 2012), and Support Vector Machines (SVMs) (Cai et al., 2015; Feng et al., 2004; Li and Kong, 2014; Samui and Kurup, 2012; Zhou et al., 2016). All those models tried to find the complex non-linear relationship between a training set of input vectors and corresponding output. The ANN-based methods have provided powerful tools to predict the displacement of landslide for their capability of processing non-linear problems. However, ANN has its own drawbacks such as arriving at the local minimum, over fitting, slow convergence speed that limit its predictive performance (Lian et al., 2012; Samui and Kurup, 2012). The SVM is a machine learning model based on the knowledge of statistical learning for small samples and structural risk minimization. Therefore, SVM becomes a more advanced method for dealing with the non-linear problems in predicting landslide displacement. With the rapid development of theory and technique, Least Squares Support Vector Machines (LSSVM) have been proposed for overcoming the defects of the SVM with high computational complexity due to quadratic programming (Suykens et al., 2002; Vapnik et al., 1997).

The predictive performance of those models is very crucial for early warning of landslide (Sassa et al., 2009). To improve the predictive performance, Genetic Algorithm (GA) was introduced to optimize the parameters of model for obtaining better predictive performance in

* Corresponding author.
E-mail address: xuqiang_68@126.com (Q. Xu).

http://dx.doi.org/10.1016/j.enggeo.2017.01.022
0013-7952/

recent achievements. Chen and Zeng (2013) improved the predictive capability of ANN by combining it with GA. Li and Kong (2014) presented an application of Genetic Algorithm and Support Vector Machines (GA-SVM) method with parameter optimization in landslide displacement rate prediction. Cai et al. (2015) presented a new model of LSSVM and GA for predicting the displacement of a landslide based on the multiple triggering factors. Besides the achievements above, some authors proposed a new work scheme for predicting the displacement of landslide under external influencing factors. Du et al. (2012) divided the accumulated displacement into a trend and a periodic component by the moving average method, and the back-propagation neural network (BPNN) was adopted to forecast the periodic component, while non-linear regression was used to predict the trend component. Zhou et al. (2016) used the Particle Swarm Optimization and Support Vector Machines (PSO-SVM) to predict the periodic component to improve the predictive accuracy.

In this paper, we have proposed two improved LSSVM models for high-precision prediction of the displacement of rainfall-induced landslide. The first one is a hybrid model composed of LSSVM and Double Exponential Smoothing (DES), namely DES-LSSVM in this study. The second one is an LSSVM model optimized by Genetic Algorithm for displacement rate prediction, namely GA-LSSVM. Kualiangzi landslide, a typical rainfall-induced deep-seated rocky landslide with gentle slid surface angle in the Sichuan Province, China, was taken as the case study to construct and validate those two models.

2. Methodology

2.1. A hybrid model composed of Least Square Support Vector Machines and Double Exponential Smoothing (DES-LSSVM)

Fig. 1 shows the flowchart of the hybrid model (DES-LSSVM), which includes three main parts of Hodrick-Prescott filter/decomposition, Least Square Support Vector Machines, and Double Exponential Smoothing (DES) method. Firstly, the original observed displacement time series are easily de-noised by wavelet de-noising method in order to reduce the random noises in GPS observations. And then, the Hodrick-Prescott filter is used to divide cumulative displacement time series into periodic component related to an external influencing factor (i.e. seasonal intense rainfall) and trend component representing the long-term dynamic evolution behaviour of landslide. On the one hand, the values of periodic component in the prior two days and the average rainfall intensity in the prior certain days are chosen as the input vector to construct the LSSVM model, and the one-step ahead periodic displacement is regarded as the output of the LSSVM model. The optimized LSSVM model is trained and obtained using grid search algorithm with cross-validation method on the basis of the previously observed periodic displacement and precipitation time series. On the other hand, the one-step ahead trend displacement can be estimated by the Double Exponential Smoothing method on the basis of the trend component value in the prior two days. Finally, the summation of the two components is considered as the cumulative displacement predicted.

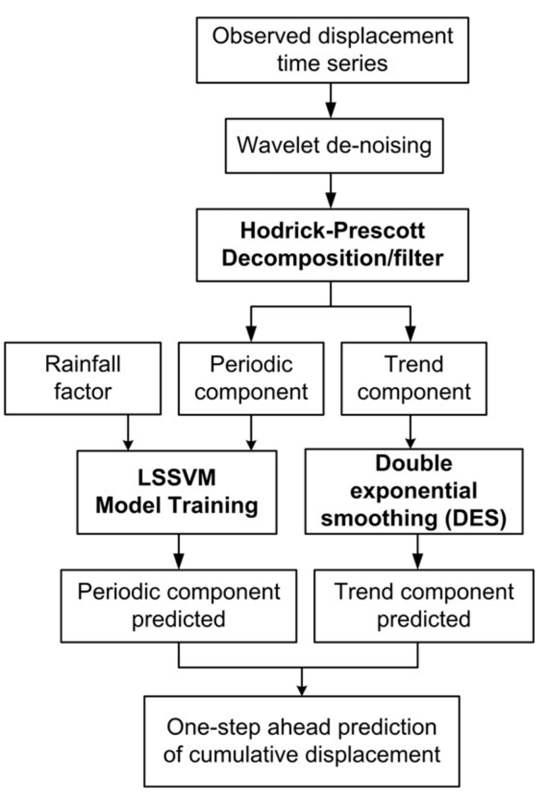

Fig. 1. Flowchart of DES-LSSVM model.

In this paper, the automatic 1-D de-noising MATLAB function *wden()* was chosen as the pre-processing approach to remove the random noises. It is applied to the originally observed displacement time series using soft heuristic SURE thresholding and scaled noise option on detail coefficients obtained from the decomposition of original data set at level 5 by 'sym8' wavelet.

The following sections will introduce the three main methods: Hodrick-Prescott decomposition/filter, LSSVM and DES.

2.1.1. Decomposition of cumulative displacement utilizing the Hodrick-Prescott filter

The Hodrick-Prescott filter is a mathematical tool used in macroeconomics, especially in real business cycle theory, to remove the cyclical component of a time series from raw data. The filter was popularized in the field of economics in the 1990s by economists Robert J. Hodrick and Nobel Memorial Prize winner Edward C. Prescott, and the details of the method can be found in the paper by Hodrick and Prescott (1997). It was used to obtain a smoothed-curve representation of a time series, one that is more sensitive to long-term than to short-term fluctuations. Therefore, it can be used to divide the cumulative displacement time series into fluctuation term owing to intense rainfall in every year and trend component in the long term as shown as follows:

$$S_i = \tau_i + \alpha_i \tag{1}$$

where S_i is the total cumulative displacement value at time *i*; α_i represents the fluctuation term and is also called periodic component because it is related to the seasonal rainfall in this study; τ_i is the trend term at time *i*. One smoothing parameter λ before applying the Hodrick-Prescott filter should be determined according to the period of the external trigger. In this study, 100 was determined as the value of λ according to the user guide of this filter function in MATLAB and the sharp increase characteristics of cumulative displacement in the rainfall seasons every year.

2.1.2. Construction of LSSVM model for predicting the periodic displacement component

Least Squares Support Vector Machines (LSSVM) is the improved formulation of the original SVM algorithm (Vapnik et al., 1997) proposed by Suykens and Vandewalle (1999). In LSSVM, given a training data set of *N* samples $\{x_i, y_i\}_{i=1}^{N}$ with input data $x_i \in R^n$ and output $y_i \in R$, where R^n is the *n*-dimensional vector space and *R* is the one-dimensional vector space. In this model, the three input variables of the LSSVM are α_i, α_{i-1} obtained by Eq. (1), and r_i representing the average intensity of rainfall in the prior certain *K* days. In this study, the value of *K* is set to 20 by considering the lagging-effect of rainfall influence on the physical characteristics and mechanical behaviour of geo-materials within landslide. The output of the LSSVM model is the one-step ahead periodic displacement Y_{i+1}. So, $x = [\alpha_i, \alpha_{i-1}, r_i]$ and $y = [Y_{i+1}]$.