Geomorphology 295 (2017) 115–125

Contents lists available at ScienceDirect

Geomorphology

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

Landslide susceptibility mapping & prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India

Deepak Kumar[b], Manoj Thakur[b], Chandra S. Dubey[c], Dericks P. Shukla[a,*]

[a] School of Engineering, Indian Institute of Technology Mandi, Kamand, Mandi 175001 (HP), India
[b] School of Basic Sciences, Indian Institute of Technology Mandi, Kamand, Mandi 175001, HP, India
[c] Department of Geology, Centre of Advanced Studies, University of Delhi, Delhi 110007, India

ARTICLE INFO

Keywords:
Landslide susceptibility mapping
Support Vector Machine (SVM)
Remote sensing
GIS
Mandakini Basin
Uttarakhand

ABSTRACT

In recent years, various machine learning techniques have been applied for landslide susceptibility mapping. In this study, three different variants of support vector machine viz., SVM, Proximal Support Vector Machine (PSVM) and L_2-Support Vector Machine - Modified Finite Newton (L_2-SVM-MFN) have been applied on the Mandakini River Basin in Uttarakhand, India to carry out the landslide susceptibility mapping. Eight thematic layers such as elevation, slope, aspect, drainages, geology/lithology, buffer of thrusts/faults, buffer of streams and soil along with the past landslide data were mapped in GIS environment and used for landslide susceptibility mapping in MATLAB. The study area covering 1625 km^2 has merely 0.11% of area under landslides. There are 2009 pixels for past landslides out of which 50% (1000) landslides were considered as training set while remaining 50% as testing set. The performance of these techniques has been evaluated and the computational results show that L_2-SVM-MFN obtains higher prediction values (0.829) of receiver operating characteristic curve (AUC-area under the curve) as compared to 0.807 for PSVM model and 0.79 for SVM. The results obtained from L_2-SVM-MFN model are found to be superior than other SVM prediction models and suggest the usefulness of this technique to problem of landslide susceptibility mapping where training data is very less. However, these techniques can be used for satisfactory determination of susceptible zones with these inputs.

1. Introduction

Natural disasters like landslides, earthquakes, flash floods etc., happen very frequently in mountainous regions especially in Himalayas. They cause huge damage to the property and economy of the region. Over the few decades, several predictive models have been used for landslide susceptibility mapping such as statistical technique (Guzzetti et al., 1999; Lee and Pradhan, 2007; Akgun and Türk, 2010) and probabilistic models (Gokceoglu et al., 2000; Lee and Pradhan, 2006; Pourghasemi, Pradhan et al., 2013). However, it is very difficult to accurately predict the behavior of landslides. Recently, many machine learning techniques have been put forward to predict the occurrence of landslides. Among them, Artificial Neural Networks (ANNs) are one of the most common methods used in landslide prediction. Unlike other statistical models, ANNs are non-parametric models and are capable of approximating any non-linear function arbitrarily without any prior assumption about the given dataset, the dimension of the input space and the input space environment (Hornik et al., 1989; Haykin and Lippmann, 1994). Various empirical studies shows that ANNs are useful techniques for landslide susceptibility mapping (Lee et al., 2004; Pradhan and Lee, 2010a,b), but they suffer from a number of shortcomings like the demand of a large number of controlling parameters, the choice of the number of hidden layers, over-fitting problem and the non-convex optimization problem which may lead to local minima.

Recently, new machine learning approaches such as Support Vector Machines (SVMs) (Cortes and Vapnik, 1995), have been developed that were tried to predict the landslide occurrences. SVMs were developed by Cortes & Vapnik based on the Structural Risk Minimization (SRM) principle (Vapnik, 1995; Vapnik and Vapnik, 1998), which has been proved to be superior to traditional Empirical Risk Minimization (ERM) principle, used by conventional neural network techniques. SRM tries to minimize an upper limit on the expected risk, whereas ERM minimizes the error on the training data. This difference in approach equips SVM with better generalization ability. In SVMs, the classification problem is formulated as convex quadratic optimization problem, and the solution of this problem is always a global optimal solution rather than local optimal solution. Over fitting in SVM is less as compared to empirical based approaches and also has better empirical performance. Due to these characteristics, SVMs are becoming more prominent alternatives

* Corresponding author.
E-mail address: dericks@iitmandi.ac.in (D.P. Shukla).

http://dx.doi.org/10.1016/j.geomorph.2017.06.013
Received 1 October 2016; Received in revised form 15 April 2017; Accepted 16 June 2017
Available online 23 June 2017
0169-555X/

to a classification problem in many domains (Pang et al., 2002; Zhou et al., 2004; Moguerza and Muñoz, 2006; Brenning, 2009; Ruppert et al., 2009). Specifically, SVMs have been used in landslide susceptibility mapping (Yao et al., 2008; Yilmaz, 2010; Pourghasemi, Jirandeh et al., 2013; Pradhan, 2013) and is found to show higher performance against other learning paradigms. However, improving the correctness and efficiency of landslide susceptibility mapping still remains the premier subject of concern. In order to improve the efficiency, Fung & Mangasarian introduced simple and computationally fast proximity based classifier known as the Proximal Support Vector Machine (PSVM) (Fung and Mangasarian, 2001). In PSVMs, the solution of classification problem is obtained from solving a set of linear equations instead of quadratic problem, this decreases the computational time. The empirical findings also demonstrate that PSVM is much faster than SVM-Light (Joachims, 1999) and LS-SVM. Furthermore, Keerthi and DeCoste (2005) have given more efficient and scalable technique for training L_2-SVMs, a Modified Finite Newton (L_2-SVM-MFN) method for fast solution of large scale linear SVMs. Hence many SVM techniques have been developed and their applicability in the area of landslide susceptibility mapping needs to be carried out. Moreover in the Himalayan terrain, where landslide phenomena is very common, the model that have been applied include conventional methods (Sarkar et al., 1995; Rautela and Thakur, 1999), statistical methods (Saha et al., 2005), logistic regression (Mathew et al., 2009; Devkota et al., 2013), Frequency Ratio (Anbalagan et al., 2015), Artificial Neural Network (ANN) (Arora et al., 2004; Kanungo et al., 2006), Analytic Hierarchy Process (AHP), Fuzzy methods (Kanungo et al., 2006; Kumar and Anbalagan, 2016), etc. (Shukla et al., 2016). Many of these works for LSZ mapping employ rating system based on the expertise experience which tend to be dependent on the knowledge of expert or mapping geomorphologists and less on the learning capabilities from the data provided, especially the utilization of the machine learning SVM techniques. Hence in this study three different machine learning models namely SVM, PSVM and L_2-SVM-MFN have been applied and compared for mapping the landslide susceptibility in parts of Garhwal Himalaya, India.

The rest of the paper is structured in various sections. In Section 2, brief description is given about the models that have been used namely SVM, PSVM and L_2-SVM-MFN. The details about the study area and experimental data including training and testing is explained in Section 3. Section 4 describes the data preparation, model selection for preparation of LSZ. The results obtained are explained in Section 5, which is also evaluated based on the testing dataset. Finally this work is concluded in Section 6.

2. Methodology

2.1. Support Vector Machine

SVM is a binary classification problem, of classifying m data samples in the n dimensional real space R^n, denoted by data matrix A, data sample x_i is the ith row of A. The membership of each point x_i can be labeled as $y_i \in \{1, -1\}$ or a diagonal matrix Y with y_i along its diagonal. The data points can be expressed by

$$x_i.\ w + b \geq 1 \quad y_i = +1 \tag{1}$$

$$x_i.\ w + b \leq -1 \quad y_i = -1 \tag{2}$$

The distance between planes (1) and (2) is known as 'margin'. The optimal hyperplane $f(w,b)$, which maximizes the margin is defined as

$$w^T x + b = 0, \quad w \in R^n, \quad b \in R \tag{3}$$

where w is normal to the optimal hyperplane (3), termed as *weight vector* and b is known as bias. The optimal hyperplane can be obtained by maximizing the margin $\frac{1}{||w||}$, equivalently given by the following quadratic program:

$$\begin{aligned} &\textit{minimize} \quad \tfrac{1}{2}||w||^2 + C\textstyle\sum_{i=1}^{m} \xi_i \\ &\textit{subject to:} \\ &\qquad y_i(wx_i+b) \geq 1-\xi_i, \quad \forall_i, \quad \xi_i \geq 0, \quad \forall\ i \end{aligned} \tag{4}$$

where C is the regularization parameter and ξ_i is the non-negative slack variables for $i = 1,\ldots,m$. The above formulation of Eq. (4) is the Quadratic Programming (QP) optimization problem with linear constraints which can be solved by using standard QP solver. The computation cost of problem (4) is $O(m^3)$. Thus, the time required to train a classifier is high and it increases polynomially as the training pattern increases.

2.2. Proximal Support Vector Machine

Fung and Mangasarian (2001) proposed simple and fast version of standard SVM known as Proximal Support Vector Machine (PSVM). The standard SVMs formulation (Eq. (4)) is modified in two ways. First, inequality constraints are replaced by equality constraints. Second, the bias term b is also regularized with w, which is the margin between the bounding planes that gets maximized with respect to both w and b instead of w. In PSVM, classification problem is formulated as

$$\begin{aligned} &\textit{minimize} \quad \tfrac{1}{2}(w'w+b^2) + \tfrac{C}{2}\textstyle\sum_{i=1}^{m} \xi_i^2 \\ &\textit{subject to:} \\ &\qquad y_i(wx_i+b) = 1-\xi_i, \end{aligned} \tag{5}$$

The nature of optimization problem (Eq. (5)) is changed after these two simple modifications in Eq. (4). Note that Vapnik's SVMs (Eq. (4)) penalize slacks (ξ_i) linearly while PSVM penalize slacks quadratically (ξ_i^2). Problem (5) has strong convex quadratic objective function with equality linear constraints. The solution of problem (5) can be obtained analytically, whereas, it is not possible in standard SVM.

Formulation (5) is the Quadratic Programming (QP) problem with linear constraints, is solved using Karush Kuhn Tucker (KKT) necessary and sufficient optimal conditions. The solution of Eq. (5) for w and b involves the inversion of massive $m \times m$ matrix. Inversion of this massive matrix can be done by using Sherman-Morrison-Woodbury formulation (Parlett, 1980), which converts it into much smaller dimension matrix of the order $(n + 1) \times (n + 1)$. The expression for w and b is given as follows:

$$[\mathbf{w}\ \mathbf{b}] = \left(\frac{I}{C},+,M',M\right)^{-1} M'Ye \tag{6}$$

where $M = [Ae]$

After obtaining w and b, the new data point (test) is classified using the optimal decision surface as given by: $f(x) = sign(w^T x + b)$.

If the value of $f(x)$ is positive (>0), the new data point assigned to class 1, otherwise it is labelled as class -1.

2.3. Finite Newton method for Support Vector Machine (MFN-SVM)

Mangasarian (2002) developed a finite Newton method for L_2-SVMs which is computationally fast and also considers the sparsity property of SVMs. Furthermore, Keerthi and DeCoste (2005) modified this finite Newton method to transform it into a fast and scalable technique for solving large scale linear problems, called a modified finite Newton method for linear L_2-SVMs (L_2-SVM-MFN). L_2-SVM solves the following optimization problem:

$$\textit{minimize} \quad \frac{1}{2}||w'w + b^2|| + \frac{C}{2}\sum_{i=1}^{m} (1-y_i(x_i w+b))^2\ \forall\ i \tag{7}$$

The problem formulation (7) is converted into an equivalent formulation as follows:

$$\textit{minimize} \quad f(z) = \frac{\delta}{2}||z||^2 + \frac{1}{2}\sum_{i\in e(z)} (1-y_i x_i z)^2 \tag{8}$$