The Zonation of landslide occurrence Using of Support Vector Machines algorithm (Case Study: Darakeh Basin)

Mojtaba Yamani[1 *] Ahmadabadi. A[2] Zare. GH[3]
[1] Department of Geography, Tehran University, Tehran, IRAN
[2] Department of Geography, Alkharazmi University, Tehran, IRAN
[3] Ph.D Student Candidate of Geography, Tarbiat Modares University, Tehran, IRAN
Received 4 December 2012
Accepted 13 January 2013

Introduction

Landslide is a geologic process that occurs over a wide variety of spatial and temporal scales in many mountainous landscapes. Landslides have a correspondingly wide range of effects that depends strongly on their spatial pattern of occurrence and frequency and magnitude of movement. Mass movements can be the dominant source of erosion responsible for the long-term geomorphic evolution of hillslope morphology. A number of different models have been developed for landslide susceptibility mapping, such as heuristic, conditional probabilistic, logistic regression (LR), artificial neural network (ANN), support vector machine (SVM), and deterministic models. These approaches have been reviewed in detail in recent publications (Carrara et al., 1995, 1999; Aleotti and Chowdhury, 1999; Guzzetti et al., 1999; Dai et al., 2002; Guzzetti, 2003; van Westen, 2004; Brenning, 2005; Wang et al., 2005; Chacon et al., 2006; Alexander, 2008; Corominas and Moya, 2008; van Westen et al., 2008). Among these approaches, SVM modeling is becoming increasingly popular. The procedure is based on statistical learning theory, and involves a training phase with associated input and target output values. The trained model is then used to evaluate a separate set of test data (Yao et al., 2008). SVM modeling has been undertaken less frequently than other approaches to landslide susceptibility mapping. Yao et al. (2008) showed that two-class SVM modeling produced more accurate susceptibility maps than one-class SVM and LR modeling on the natural slopes of Hong Kong, China. Brenning (2005) showed the predictive power of LR, SVMs and bootstrap-aggregated classification trees in a case study of the Ecuadorian Andes. In that study, LR with stepwise backward-selection of variables yields the lowest error rates and demonstrates the best generalization capabilities. Landslide is among slope process dominant on South Alborz and especially in Darakeh basin. Identify area that are prone to landslide is very important, because near of basin to Tehran city, tourism aspect and human settlement in the basin. In this research we tried using SVM algorithm is determined and zonation landslide hazard and Areas susceptible to landslides in the basin. Seems that output of research has a role in sustainable environmental management and a document used in future planning for development of infrastructure.

Study Area

Darakeh Basin is located in north of Tehran city, of view point relative position and in southern slope of Alborz range and among 35^0 48' to 35^0 53' and 51^0 21' to 51^0 24' of viewpoint the absolute position. This basin is 24.22 km^2. In the east of Darake basin is located Darband valley, in the north Tochal peak and west Farahzad valley. Of viewpoint structural Geology is part of Alborz. Of viewpoint structural Geomorphology has occurred intense tectonic movements that has the last major move in the Alborz.

* Email: myamani@ut.ac.ir Corresponding Author 00989123197682

Material and Methods

SVM modeling is a recently developed method based on nonlinear transformations of covariates into a higher dimensional feature space. Two principal ideas underlie SVM modeling for discriminant-type statistical problems. The first is an optimum linear separating hyper plane that separates data patterns. The second is the use of kernel functions to convert the original nonlinear data patterns into a format that is linearly separable in a high-dimensional feature space. A detailed description of two-class SVM modeling is available in Yao et al. (2008), which can be summarized as follows: Consider a set of linear separable training vectors xi (i=1, 2, …,n). The training vectors consist of two classes, which are denoted as yi=±1. The goal of SVM is to search an n-dimensional hyper plane differentiating the two classes by their maximum gap. Mathematically, it can be expressed as: $\frac{1}{2}\|w\|^2$

Subject to the following constraints: $Y_i((w.x_i)+b) \geq 1$

Where ||w|| is the norm of the normal of the hyper plane, b is a scalar base, and (·) denotes the scalar product operation. Using the Lagrangian multiplier, the cost function can be defined as:

$$L = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \lambda_i (y_i((w \cdot x_i) + b) - 1)$$

Where λi is the Lagrangian multiplier. The solution can be achieved by dual minimization of Eq. (3) with respect to w and b through standard procedures. More details and discussion on the development of the method, have been given by Vapnik (1995) and Tax and Duin (1999). For non-separable case, one can modify the constraints by introducing slack variables ξi (Vapnik, 1995):

$$y_i((w \cdot x_i) + b) \geq 1 - \xi_i .$$

$$L = \frac{1}{2}\|w\|^2 - \frac{1}{vn}\sum_{i=1}^{n} \xi_i$$

Eq. (1) becomes:

Where v (0, 1] is introduced to account for misclassification (Scholkoph et al., 2000; Hastie et al., 2001). In addition, Vapnik (1995) introduces a kernel function K (xi, xj) to account for the nonlinear decision boundary. Selection of the kernel function is very important in SVM modeling. Although many kernel functions have been previously proposed and used, only some have been found to work well in a wide variety of applications. Those that have shown this skill are:

$$\text{Linear: } K(x_i, x_j) = x_i^T \cdot x_j$$

$$\text{Polynomial: } K(x_i, x_j) = (\gamma \cdot x_i^T \cdot x_j + r)^d, \gamma > 0$$

$$\text{Radial basis function: } K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2}, \gamma > 0$$

$$\text{Sigmoid: } K(x_i, x_j) = \tanh(\gamma \cdot x_i^T \cdot x_j + r)$$

Where γ, r, and d are parameters of the kernel functions and are entered manually; sometimes the kernel functions are parameterized using: $\gamma = 1/2\sigma^2$, where σ is an adjustable parameter that governs the performance of the kernel. If overestimated, the exponential behaves almost linearly and there is a loss of non-linear power in higher-dimensional projection; whereas, if underestimated, the decision boundary becomes sensitive to noise in the training data. In the present study, +1 and −1 represent landslide and stable slope locations, respectively. Note that the stable slope cases are not available and they have to be generated (Yao et al., 2008). Because Yao et al. (2008) showed that the two-class SVM possesses better prediction efficiency than LR and a one-class SVM; we use a two-class SVM in this study.

Results and Discussion

The result showed that Basin has moderate susceptibility to landslide occurrence based on linear function. Based on output of sigmoid function, the basin shows moderate to high sensitivity to the

landslide. Maximum and minimum distance of possible area susceptible to landslide is order sigmoid and radial functions with 30 and 20 kilometres. Among the performed functions, sigmoid functions showed the best accuracy for more conforming to reality and linear function showed the worst accuracy of performance viewpoint. Output maps result of performance function has indicated the accuracy of calculations with regard to the training points. It is worthy to note that the implementations of these functions are associated with the error and rate of errors in performing these functions in the present study Venial.

Conclusion

Based on linear function, regions that are near to the drainage and fault have more chance of occurrence than the landslides. On the other hand, the fault and drainage has most influence on the occurrence of landslide hazard. These two factors have an influence geology agent (shear andesitic lavas and lava flows). According to polynomial function, the distance from fault and drainage has a significant role in susceptibility of the basin to landslide. Also geological factors (crystal tuff and cutting and site lavas and lava flows, and hyalite tuffs with some shale strata) are effective on landslide. Based on radial function, distance from fault and drainage has significant role in susceptibility of the basin to the landslide hazard. Also, elevation and geological factors (shear andesitic lavas and lava flows, tuffs and hyalite strata in areas with shale, siltstone, and shale alternation of tuff, crystal tuff and ash green tuff, tuff cut and local limestone strata) has considerable influence on the landslides. According to sigmoid function, distance from fault and drainage has significant role in susceptibility of the basin to landslide hazard. Also geology and elevation agents has significant role in basin susceptibility to landslides. In the southern basin, there has seen a small area that is indicating the high sensitivity of this area to the occurrence of landslide hazard.

Key Word: Landslide, Artificial Intelligence, SVM Algorithm, Environmental Management, Darakeh Catchment.