Catena 133 (2015) 266–281

Contents lists available at ScienceDirect

Catena

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

Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines

Haoyuan Hong [a,b], Biswajeet Pradhan [c], Chong Xu [a], Dieu Tien Bui [d,*]

[a] Key Laboratory of Active Tectonics and Volcano, Institute of Geology, China Earthquake Administration, #1 Huayanli, Chaoyang District, PO Box 9803, Beijing 100029, PR China
[b] Jiangxi Provincial Meteorological Observatory, Jiangxi Meteorological Bureau, No. 109 ShengfuBeier Road, Nanchang 330046, PR China
[c] Department of Civil Engineering, Geospatial Information Science Research Center (GISRC), Faculty of Engineering, University Putra Malaysia, Serdang, Selangor Darul Ehsan 43400, Malaysia
[d] Geographic Information System Group, Department of Business Administration and Computer Science, Faculty of Arts and Sciences, Telemark University College, Hallvard Eika Plass 1, N-3800 Bø i Telemark, Norway

ARTICLE INFO

Article history:
Received 29 September 2014
Received in revised form 19 April 2015
Accepted 20 May 2015
Available online 6 June 2015

Keywords:
Two-class kernel logistic regression
Support vector machines
Alternative decision tree
Landslide
Yihuang area
China

ABSTRACT

Preparation of landslide susceptibility map is the first step for landslide hazard mitigation and risk assessment. The main aim of this study is to explore potential applications of two new models such as two-class Kernel Logistic Regression (KLR) and Alternating Decision Tree (ADT) for landslide susceptibility mapping at the Yihuang area (China). The ADT has not been used in landslide susceptibility modeling and this paper attempts a novel application of this technique. For the purpose of comparison, a conventional method of Support Vector Machines (SVM) which has been widely used in the literature was included and their results were assessed. At first, a landslide inventory map with 187 landslide locations for the study area was constructed from various sources. Landslide locations were then spatially randomly split in a ratio of 70/30 for building landslide models and for the model validation. Then a spatial database with a total of fourteen landslide conditioning factors was prepared, including slope, aspect, altitude, topographic wetness index (TWI), stream power index (SPI), sediment transport index (STI), plan curvature, landuse, normalized difference vegetation index (NDVI), lithology, distance to faults, distance to rivers, distance to roads, and rainfall. Using the KLR, the SVM, and the ADT, three landslide susceptibility models were constructed using the training dataset. The three resulting models were validated and compared using the receive operating characteristic (ROC), Kappa index, and five statistical evaluation measures. In addition, pairwise comparisons of the area under the ROC curve were carried out to assess if there are significant differences on the overall performance of the three models. The goodness-of-fits are 92.5% (the KLR model), 88.8% (the SVM model), and 95.7% (the ADT model). The prediction capabilities are 81.1%, 84.2%, and 93.3% for the KLR, the SVM, and the ADT models, respectively. The result shows that the ADT model yielded better overall performance and accurate results than the KLR and SVM models. The KLR model considered slightly better than SVM model in terms of the positive prediction values. The ADT and KLR are the two promising data mining techniques which might be considered to use in landslide susceptibility mapping. The results from this study may be useful for landuse planning and decision making in landslide prone areas.

1. Introduction

Landslides are considered to be one of the most widespread geologic hazards in many areas of the world and can be defined as a downslope movement of soil and rock under the influence of gravity (Malamud et al., 2004). Globally, around 17% of the fatalities occurred due to landslides with approximately 66 million people living within the high risk landslide areas (Sassa and Canuti, 2008). According to the report of the International Landslide Centre of the University of Durham recorded in 2007 (Petley, 2010), China was the most seriously affected country with 695 landslide-induced deaths, followed by Indonesia (465), India (352), Nepal (168), and Bangladesh (150). With the development of the economics, the urbanization of mountainous areas is continuing and this leads to the instability of slopes, thus increases the potential for landslides. Therefore, understanding landslide mechanisms and preventing them from future occurrence are considered as an important task that may help government, decision makers, and engineers in slope management and landuse planning (Jaafari et al., 2014; Pourghasemi et al., 2012b; Pourghasemi et al., 2014; Pradhan, 2013; Tien Bui et al., 2012e; 2013c; Youssef et al., 2015).

Landslide hazard is defined as the probability of a mass movement taking place in a certain area and in a specified period of time (Varnes, 1984). It means that the procedure of the landslide hazard mapping

* Corresponding author.

http://dx.doi.org/10.1016/j.catena.2015.05.019
0341-8162/

incorporates both the spatial and temporal predictions of landslides and spatial prediction of landslide hazards that is not taken into account triggering factors are known as susceptibility analyses. Although landslide susceptibility map is considered to be the first step in the hazard and risk assessment, the map is accepted as an end product that can be used in landuse zoning and environmental planning (Manzo et al., 2012).

Landslides occur due to complex and relate to various factors i.e. topography, geology, hydrogeological conditions, landuse, vegetation, and rainfall, producing reliable landslide susceptibility maps which are not easy (Lee and Pradhan, 2007; Pourghasemi et al., 2012c; Tien Bui et al., 2012e). Various methods and techniques have been proposed for mapping landslide susceptibility, such as logistic regression (Bai et al., 2010; Devkota et al., 2012; Pourghasemi et al., 2013a; Tien Bui et al., 2011), multivariate regression (Akgun and Turk, 2011; Chung et al., 1995; Suzen and Doyuran, 2004), discriminant analysis (Dong et al., 2009), analytical hierarchy process (Ju et al., 2012; Pourghasemi et al., 2012a), artificial neural network (Choi et al., 2012; Conforti et al., 2014; Pavel et al., 2011; Tien Bui et al., 2013d; Zare et al., 2013), neuro-fuzzy (Pradhan et al., 2010; Tien Bui et al., 2012d), evidential belief function (Nampak et al., 2014; Pradhan et al., 2014; Tien Bui et al., 2012e), decision tree (Nefeslioglu et al., 2010; Tien Bui et al., 2012a; Tsai et al., 2013; Yeon et al., 2010), and support vector machines (Pourghasemi et al., 2013b; Pradhan, 2013; Saito et al., 2009; San, 2014; Tien Bui et al., 2012a; 2012b; 2015; Xu et al., 2012; Hong et al. 2015). Although the accuracy of aforementioned methods is still debated, however, logistic regression, support vector machines, and ensemble methods are reported to have outperformed the conventional methods (Jebur et al., 2014; Pradhan, 2013; Tien Bui et al., 2013a; 2012a).

The recent development of machine learning and Geographic Information System has resulted in some new powerful techniques i.e. Two-class Kernel Logistic Regression (KLR), Alternative Decision Tree (ADT) with a capability to improve the prediction performance of models (Maalouf and Trafalis, 2011; Rokach, 2010; Were et al., 2015). Literature review shows that the KLR and the ADT have seldom been explored in landslide modeling; therefore an investigation of the two methods in comparison with conventional methods should be carried out. Only 2 or 3% of the increment of prediction capability may control the resulting landslide susceptibility areas and therefore it is highly sought after to find high performance based models which can accurately predict these areas (Jebur et al., 2014; Mohammady et al., 2012; Pradhan, 2013; Tien Bui et al., 2014; 2015).

The main objective of the present study is to explore potential application of the KLR and the ADT in landslide susceptibility mapping at Yihuang area, the Yihuang City of China. The KLR is a kernelized version of logistic regression that maps the original input space into a new high-dimensional feature space using a kernel function, i.e. Radial Basis Function. The main improvement of the KLR compared with the logistic regression is that the KLR has the ability to classify data with non-linear boundary (Zhu and Hastie, 2001). In addition, a KLR model can be constructed with very few training instances and can process very high dimensional data. In the case of the ADT, this is a machine learning method in which decision trees are combined with the LogitBoost to generate interpretable classification rules (Holmes et al., 2002). The main advantages of ADT are that ADT can generate simpler decision-tree structures and easier-to-interpret classification rules (Freund and Mason, 1999a). The computation process was carried out using Matlab 7.11, Weka 3.66, and ArcGIS 10.0.

2. Study area and data used

2.1. Study area

The Yihuang area is located in the central of the Jiangxi Province, in the west of the Wuyishan mountain and the north of the Ganfu plain. The study area lies between latitude 27°0′N and 27°43′N, and longitude 116°1′E and 116°28′E. It covers an area of about 1944 km^2. The altitude of the area ranges from 45.6 to 1728.4 m above sea level. Around 36% of the study area has a slope gradient less than 15° whereas areas with a slope gradient larger than 45° account for 0.3% of the total study area. Areas fall into the slope category 15°–25° account for 37.8% of the total study area. Areas with the slope category 25°–35° account for 21% of the study areas and the remaining area falls into the slope category 35°–45°. More than 36 geologic groups and units are recognized (Table 1). The main lithologies are monzonitic granite, tonalite diorite, porphyritic moyite, gray brown granulite (Fig. 4).

The study area belongs to a subtropical monsoon climate. According to the Jiangxi Province Meteorological Bureau (http://www.weather.org.cn), the average annual rainfall for the period 1960–2012 years is from 1060 mm to 2660 mm. The average annual temperature is 17.5 °C. The rainy season is from February to September that accounts for 83.6% of the yearly rainfall. In May and June, the average rainfall varies between 250 mm and 310 mm per month.

In the Yihuang area, no information about earthquake-induced landslides had been reported, the high amount of rainfall is considered as the main triggering factor for the occurrence of landslides (Huang and Li, 2011). According to the statistics of the Yihuang City, a total of 2421 people in the study area are affected by landslides. The damages to properties are estimated about 3 million USD. However, very few attempts have been made to forecast their location and prevent their damages (Lin et al., 2006; Lin et al., 2010; Wu et al., 2014).

Table 1
Lithological classification of the study area.

Category	Main lithology	Geologic group and unit
A	Gray, gray purple quartz conglomerate, pebbly sandstone, sandstone	Zishan group; Yunshan group; Zhongpeng groups
B	Granodiorite	Xinquan Super unit
	Monzonitic granite	The Ancient Yin Zhai unit; Huang Pichao unit
	K-feldspar granite	Silver Factory Super unit; Chen Fang unit; Diao Qiao unit; the Li family unit
	Monzonitic, K-feldspar granite	Ge Tanshan Super unit; Ge Xianyuan Unit; Ken former Unit; Xishan Row unit; Moon shaped super unit
	Monzonitic, K-feldspar granite	Jiuxian Decoction; Mufu mountain; Changshan; Match Yangguan Super unit, Huang Xie, Xihua Mountain Super unit
	Monzonitic granite	Huang Xie; the Xihua mountain; Changshan;
C	Brick red, purple red conglomerate, pebbly sandstone	Lianhe group; Tanbian group; Hekou group
D	Granulite, schist, marble variable conglomerate	Hongshan group
E	Gray brown granulite clip two mica schist, quartz schist	Wanyuan group
F	Tonalite diorite, porphyritic granodiorite, granite	Tang Huchao unit; Fu Fangchao unit; Car Brain unit; High Delta unit
	Monzonitic granite	Fu Fangchao unit; Soup Huchao unit; Gaoping unit; Cat Nasal Yin unit
G	Monzonitic granite	Triassic granite
	K-feldspar granite	The Super element; Qingxi Over unit, Fu Super unit; Tu Qiao Ao unit
	Sand, Shale, chert conglomerate with bottom seam, tuffaceous sandstone, tuff	Anyuan group
H	Dark gray, gray and black carbonaceous siliceous slate	The Outer Tube group
	Fine tonalitic granite diorite granite gneiss	Middle Cambrian tonalities and diorite