Catena 149 (2017) 52–63

Contents lists available at ScienceDirect

Catena

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

Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS

Binh Thai Pham [a,b,*], Dieu Tien Bui [c], Indra Prakash [d], M.B. Dholakia [e]

[a] Department of Civil Engineering, Gujarat Technological University, Nr. Visat Three Roads, Visat - Gandhinagar Highway, Chandkheda, Ahmedabad 382424, Gujarat, India
[b] Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Ha Noi, Viet Nam
[c] Geographic Information System Group, Department of Business Administration and Computer Science, University College of Southeast Norway, Hallvard Eikas Plass 1, N-3800 Bø i Telemark, Norway
[d] Department of Science & Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar, India
[e] Department of Civil Engineering, LDCE, Gujarat Technological University, Ahmedabad 380015, Gujarat, India

ARTICLE INFO

Article history:
Received 17 October 2015
Received in revised form 9 July 2016
Accepted 7 September 2016
Available online xxxx

Keywords:
Landslides
Ensemble techniques
Multilayer Perceptron Neural Network
Himalaya
India

ABSTRACT

The main objective of this study is to evaluate and compare the performance of landslide models using machine learning ensemble technique for landslide susceptibility assessment. This technique is a combination of ensemble methods (AdaBoost, Bagging, Dagging, MultiBoost, Rotation Forest, and Random SubSpace) and the base classifier of Multiple Perceptron Neural Networks (MLP Neural Nets). Ensemble techniques have been widely applied in other fields; however, their application is still rare in the assessment of landslide problems. Meanwhile, MLP Neural Nets, which is known as an artificial neural network, has been applied widely and efficiently in landslide problems. In the present study, landslide models of part Himalayan area (India) have been constructed and validated. For the evaluation and comparison of these models, receiver operating characteristic curve and Chi Square test methods have been applied. Overall, all landslide models performed well in landslide susuceptibility assessment but the performance of the MultiBoost model is the highest (AUC = 0.886), followed by Dagging model (AUC = 0.885), the Rotation Forest model (AUC = 0.882), the Bagging and Random SubSpace models (AUC = 0.881), and the AdaBoost model (AUC = 0.876), respectively. Moreover, machine learning ensemble models have improved significantly the performance of the base classifier of MLP Neural Nets (AUC = 0.874). Analysis of results indicates that landslide models using machine learning ensemble frameworks are promising methods which can be used as alternatives of individual base classifiers for landslide susceptibility assessment of other prone areas.

1. Introduction

Himalaya is known as a landslide affected region in India. Approximately 80% landslide occurrences in India belong to this region (Onagh et al. 2012). Recently, occurrences of landslides affecting the life and properties have increased in this region due to anthropogenic and natural causes (Mukane 2014; Sarkar et al. 1995). There are various methods to assess and plan mitigation of landslide damages. Landslide susceptibility map is one of the standard useful tool for proper land use planning and mitigation decision making (Pourghasemi et al. 2013c; Tien Bui et al. 2016b). This map helps in visualization and spatial prediction of landslides in certain areas (Dou et al. 2015). Spatial prediction of landslides is being carried out on the basis of an assumption that future landslides might occur under same conditions which has caused past landslides (Dou et al. 2014; Tsangaratos et al. 2013). Therefore, evaluation of the spatial relationship between a set of affecting factors and previous landslide occurrences is desirable (Ilia and Tsangaratos 2015; Tsangaratos et al. 2015).

Researchers have developed many methods, in recent decades, using soft computing approaches to produce landslide susceptibility maps of different regions of the world. These approaches include logistic regression (Akgun 2012; Bai et al. 2010; Das et al. 2012; Devkota et al. 2013; Tsangaratos and Ilia 2016), neuro-fuzzy systems (Oh and Pradhan 2011; Sezer et al. 2011; Sezer et al. 2013; Vahidnia et al. 2010), decision trees (Alkhasawneh et al. 2014; Lee and Park 2013; Pradhan 2013; Saito et al. 2009; Tsangaratos and Ilia 2015), fuzzy logic (Aksoy and Ercanoglu 2012; Pourghasemi et al. 2012b; Pradhan 2010; Pradhan 2011; Saboya et al., 2006), support vector machines (Peng et al. 2014; Pourghasemi et al. 2013b; Pradhan 2013; Xu et al. 2012;

* Corresponding author at: Department of Civil Engineering, Gujarat Technological University, Nr. Visat Three Roads, Visat - Gandhinagar Highway, Chandkheda, Ahmedabad 382424, Gujarat, India.
E-mail addresses: phambinhgtvt@gmail.com, binhpt@utt.edu.vn (B.T. Pham).

http://dx.doi.org/10.1016/j.catena.2016.09.007
0341-8162/

Yilmaz 2010), evidential belief function (Jebur et al. 2015; Lee et al. 2013; Nampak et al. 2014; Pradhan et al. 2014), and artificial neural networks (Chauhan et al. 2010; Choi et al. 2010; Conforti et al. 2014; Poudyal et al. 2010; Tsangaratos and Benardos 2014). Though the performance of these techniques in landslide susceptibility assessment is relatively good, it can be further improved using ensembles techniques to generate machine learning ensemble frameworks for classification (Althuwaynee et al. 2014; Jebur et al. 2015; Tien Bui et al. 2014).

Ensemble techniques are data mining methods. These techniques use machine learning algorithms to combine multiple base classifiers to improve their performance, and are considered promising techniques for classification in recent years (Pham et al. 2016c). Some of the ensemble techniques such as AdaBoost, Bagging, Dagging have been widely used to solve a lot of problems of classification in real world during recent decades. Other methods namely Random SubSpace, MultiBoost, and Rotation Forest are relatively new ensemble approaches. Use of the ensemble techniques in the field of landslide is rare though these techniques have been widely used in other fields such as medical (Kelarev et al. 2012), banking (Chen et al. 2015), computer science (Abawajy et al. 2014).

The main objective of present study is to evaluate and compare the performance of different machine learning ensemble frameworks using Multilayer Perceptron Neural Networks (MLP Neural Nets) and ensemble techniques (AdaBoost, Bagging, Dagging, MultiBoost, Rotation Forest, and Random SubSpace) for spatial prediction of landslides. Out of these methods, MLP Neural Nets is one of the artificial neural networks which have been applied successfully and efficiently in landslide problems (Gomez and Kavzoglu 2005; Pradhan and Buchroithner 2010; Zare et al. 2013). Receiver operating characteristic curve and Chi Square test methods have been selected to evaluate and compare these landslide models. Data have been collected in a part located in Himalaya (India). Analysis of landslide data and model study of part of Himalaya has been carried out using ArcMap 10.2 and Weka 3.7.12 software.

2. Background theory of methods

2.1. Base classifier of Multilayer Perceptron Neural Network (MLP Neural Nets)

MLP Neural Nets is known as the artificial neural network technique which has been widely used in classification (Haykin et al. 2009). It has several advantages such as the distribution of training dataset is not dependent on pre-assumptions, no decision requires to be set in relation with the relative importance of the various input measurements, and the most input measurements are selected based on the adjustment of the weight during training process (Gardner and Dorling 1998).

Input layers, hidden layers, and output layers are three main compositions that construct the MLP Neural Nets (Fig. 1). Input layers are understood as landslide affecting factors, the output layers are viewed as the classified results which infer landslide or non-landslide classes, and the hidden layers are the classifying layers in order to transform inputs to outputs.

MLP Neural Nets is trained in two main steps (Tien Bui et al. 2015): (i) the inputs are propagated forward through the hidden layers to produce the output values, and then the output values are compared to pre-values in order to estimate the difference, (ii) the connection weights are adjusted to optimize the best results with the least difference.

Let $x = x_i$, $i = 1, 2, \ldots, 15$ is the vector of the fifteen landslide affecting factors, $y = 1$ (landslide class) or 0 (non-landslide class). MLP Neural Nets function for classification is as below:

$$y = f(x) \tag{1}$$

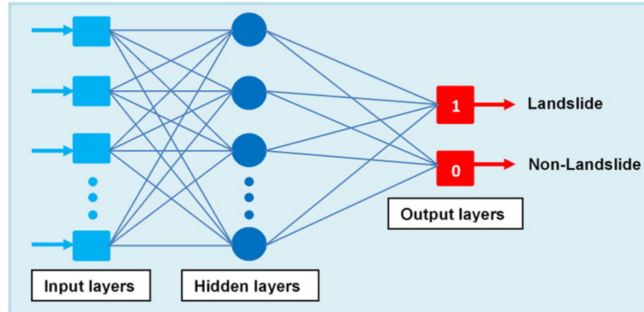

Fig. 1. A Multilayer Perceptron Neural Networks structure.

where f(x) is an hidden function that is optimized by the adjustable weights during training process for given network architecture.

In the present study, MLP Neural Nets has been trained with the number of hidden layers of 1 and 500 epochs and the validation threshold of 20 which have been obtained from trial-and-error process in order to avoid over-fitting problem.

2.2. Machine learning ensemble techniques

2.2.1. AdaBoost

AdaBoost is one of the boosting algorithms which have been utilized to enhance the predictive capability of the learning classifier algorithms. It was introduced by Freund and Schapire (1997). AdaBoost algorithm has been widely utilized in classification that usually focuses on difficult data points. The weight values are first assigned to instances in training dataset. These weights are then replaced during iterations of training process based on the performance of the previous base classifier (Freund and Schapire 1997). The training process would be stopped if the optimal weights have been assigned to instances to obtain the best performance of the base classifier.

2.2.2. Bagging

Bagging was first introduced by Breiman (1996) that is one of the earliest ensemble techniques. It uses bootstrap samples to drive individual classifiers. Firstly, the new sub-training sets are obtained by simple random sampling from learning sets with replacements. These sub-training sets are used to train base classifiers. Subsequently, majority voting (weighted majority voting) is utilized to combine the results of base classifiers (Breiman 1996).

2.2.3. Dagging

Dagging is well known re-sampling ensemble technique using majority vote to combine a diversity of classifiers in order to improve predictive accuracy of base classifiers (Kotsianti and Kanellopoulos 2007). It was first introduced by Ting and Witten (1997). Dagging uses a number of disjoint samples instead of bootstrap samples to derive the base classifiers (Ting and Witten 1997).

2.2.4. MultiBoost

MultiBoost was first proposed by Webb (2000) that is a combination of AdaBoost and Wagging techniques to reduce both variance and bias and avoid the over-fitting problem. The AdaBoost algorithm has been used firstly to assign the weights for instances, and then the Wagging algorithm has been applied to replace the weights for training samples based on the performance of the previous base classifier (Webb 2000).

2.2.5. Rotation Forest

Rotation Forest was first introduced by Rodriguez et al. (2006) that is known as a newest ensemble techniques so far. It uses Principal Component Analysis (Wold et al. 1987) to extract the features from learning