Knowledge-Based Systems 76 (2015) 256–263

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A Swarm-Optimized Fuzzy Instance-based Learning approach for predicting slope collapses in mountain roads

Min-Yuan Cheng [a,1], Nhat-Duc Hoang [b,*]

[a] Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, #43, Section 4, Keelung Road, Daan District, Taipei 8862, Taiwan
[b] Institute of Research and Development, Faculty of Civil Engineering, Duy Tan University, K7/25 Quang Trung, Danang 550000, Viet Nam

ARTICLE INFO

Article history:
Received 11 August 2014
Received in revised form 22 December 2014
Accepted 23 December 2014
Available online 30 December 2014

Keywords:
Instance-based learning
Swarm intelligence
Fuzzy k-Nearest Neighbor
Slope collapse prediction
Firefly Algorithm

ABSTRACT

Due to the disastrous consequences of slope failures, forecasting their occurrences is a practical need of government agencies to develop strategic disaster prevention programs. This research proposes a Swarm-Optimized Fuzzy Instance-based Learning (SOFIL) model for predicting slope collapses. The proposed model utilizes the Fuzzy k-Nearest Neighbor (FKNN) algorithm as an instance-based learning method to predict slope collapse events. Meanwhile, to determine the model's hyper-parameters appropriately, the Firefly Algorithm (FA) is employed as an optimization technique. Experimental results have pointed out that the newly established SOFIL can outperform other benchmarking algorithms. Therefore, the proposed model is very promising to help decision-makers in coping with the slope collapse prediction problem.

1. Background

Road is the major transportation construction that support modern society; many researchers have found significant positive correlations between roads and economic growth at both local and regional levels [1,2]. Accordingly, in various countries around the world, extensive networks of mountain roads has recently been built to catch up with the population expansion and the economic development [3,4].

Furthermore, natural hazards coupled with rugged terrains lead to the fact that slope collapses possibly occur in many sections of the road network. These catastrophic events are often triggered by earthquakes or heavy rainfalls during typhoons or monsoon storms [5,6]. Slope collapses are very undesirable since they inflict damages to man-made structures, disruption of traffic, and indispensible losses of human lives. Hence, slope stability assessment is an inevitable task which should be regularly conducted by roadway maintenance authorities [7,8]. The analysis results can be utilized for identifying collapse-prone areas as well as allocating scarce resources to establish an overall disaster prevention program [9]. In order to analyze slope stability, physical model, expert evaluation, and machine learning are the three common methods [10].

The physical model method is based on the slope displacement model which can analyze the slope stability by identifying of the most dangerous sliding surface and calculating the factor of safety [11]. Although this approach can deliver accurate analytical results, it requires input parameters for every calculation point of the investigated area. Therefore, the physical model method is only appropriate for evaluating stability in small areas, and its capacity for analysis over large areas is inapplicable [12].

The expert evaluation approach utilizes expert judgments and information of slope collapse events occurred in the past [8,13]. Using expert knowledge, the main influencing features and possible triggering factors can be identified [14]. Based on that information, the stability of a slope can be evaluated by expert knowledge. Obviously, requiring many subjective judgments and inconsistency of the prediction results are the main drawbacks of this method.

Recently, machine learning approaches have been utilized to automate the slope assessment process due to their better flexibilities and prediction capabilities compared to the traditional approaches. Generally, machine learning based models are established by combining artificial intelligence (AI) techniques and historical databases [15]. Using these models, the slope evaluation can be considered as a classification task in which prediction outputs are either "stable" or "unstable".

* Corresponding author. Tel.: +84 0511 3827111.
E-mail addresses: myc@mail.ntust.edu.tw (M.-Y. Cheng), hoangnhatduc@dtu.edu.vn (N.-D. Hoang).
[1] Tel.: +886 2 27336596/27301074.

http://dx.doi.org/10.1016/j.knosys.2014.12.022
0950-7051/

Lu and Rosenbaum [16], Zhou and Chen [17], Jiang [18], Das et al. [7], Cho [19], Lee et al. [20], and Wang et al. [21] applied the Artificial Neural Network (ANN) to predict the slope condition. Zhao et al. [22] employed the Relevance Vector Machine (RVM) to explore the nonlinear relationship between slope stability and its influence factors. Slope stability forecasting models based on the Support Vector Machine (SVM) were developed by Li and Wang [23], Cheng et al. [24], Zhao [25], Samui [26], and Li and Dong [27]; these studies found that SVM based models are very effective under the condition of limited data.

Although the ANN has been extensively applied for predicting slope collapse, the implementation of this approach has several drawbacks. The major disadvantage of the ANN is that its training process is achieved through a gradient descent algorithm on the error space, which can be very complex and may contain many local minima [28]. Moreover, the SVM training requires solving a quadratic programming problem subjected to inequality constraint; this means that the training process for large data sets requires expensive computational cost [29]. Most importantly, the black box nature of the ANN, SVM, and RVM algorithms makes them difficult for practical engineers or government agencies to comprehend how they predict slope collapses.

Different from the aforementioned AI methods, the Fuzzy k-Nearest Neighbor (FKNN) algorithm [30] belongs to the class of instance-based learning. This algorithm utilizes the whole collected data to establish its memory. A FKNN classifier utilizes the information obtained from the k nearest neighbors of a sample vector and assigns class memberships to it. The vector's membership values provide a level of assurance to accompany the resultant classification. Moreover, the algorithm also assigns fuzzy memberships as a function of the vector's distance from its k nearest neighbors and those neighbors' memberships in the possible classes [31]. Needless to say, this approach is simple to implement and its classification outcomes are also easily interpretable. In addition, the competitive prediction performance of the FKNN has been demonstrated in various studies [30,32–34]. Nevertheless, none of previous works has evaluated the capability of the FKNN method in slope collapse assessment.

Additionally, the implementation of the FKNN requires a proper setting of two tuning parameters: the neighboring size (k) and the fuzzy strength (m). Furthermore, this parameter selection process can be modeled as an optimization problem. Meta-heuristic approaches have been illustrated to be feasible to tackle the optimization problem at hand [35–39]. Recently developed by Yang [40], the Firefly Algorithm (FA) is a fast and effective meta-heuristic for solving global optimization in continuous space. Numerical experiments in previous researches have demonstrated the superior performance of the FA over other meta-heuristic methods [41–43]. Nonetheless, few research works have investigated the capability of this algorithm in optimizing the parameter selection process of the FKNN. Thus, this study proposes to hybridize the FKNN with FA [40] to automatically search for appropriate hyper-parameters of the prediction model.

Thus, this research employs the FKNN classifier as the machine learning technique to construct a prediction model for slope collapse assessment. We propose to hybridize the FKNN algorithm with the FA [40] to automatically search for an appropriate combination of tuning parameters for the prediction model. The newly established approach is named as Swarm-Optimized Fuzzy Instance-based Learning (SOFIL). The remaining part of this paper is organized as follows. The second section of this paper presents the research methodology. The framework of the proposed SOFIL is described in the third section. The fourth section demonstrates the experimental results. Conclusions of the study are stated in the final section.

2. Methodology

2.1. Fuzzy k-Nearest Neighbor algorithm

The FKNN algorithm is an instance-based classifier that incorporates the fuzzy set theory into the classification process [30]. In the FKNN, the fuzzy memberships of samples are assigned to different classes. The class which possesses the maximum membership degree can be chosen as the winner. The first step of the FKNN algorithm is to calculate the fuzzy partition matrix $U = [u_{ij}]$ from the memory which stores a set of n training sample vectors $[x_1,\ldots,x_n]$. Herein, we denote j as the vector index ($j = 1,2,\ldots,n$), where n is the number of training samples. And, the variable i represents the class index ($i = 1,2,\ldots,C$), where C is the number of classes. For each training case x, we identify its k nearest neighbors by calculating Euclidean distances. The membership degree of the sample vector x_j in the class i is given as follows:

$$u_{ij}(x) = u_i(x_j) = \begin{cases} 0.51 + (n_i/k) \times 0.49, & \text{if } c(x_j) = i \\ (n_i/k) \times 0.49, & \text{if } c(x_j) \neq i \end{cases} \tag{1}$$

where n_i is the number of neighbors found which belong to the class i and $c(x_j)$ represents the class label of the sample vector x_j. It is obvious that u_{ij} is an element of the C-by-n matrix U. Moreover, it is also worth noticing that the purpose of Eq. (1) is to assign higher fuzzy membership grades to the training samples that stay away from the decision boundary and lower fuzzy memberships grade to the patterns that lie in the vicinity of the decision boundary [30]. It is because the information supplied by the samples in the region close to the decision surface is more uncertain than that provided by other samples.

Since u_{ij} is a fuzzy membership grade of the sample x_j in the class i, u_{ij} must satisfy the following properties:

$$u_{ij} \in [0, 1] \tag{2}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \tag{3}$$

$$0 < \sum_{j=1}^{n} u_{ij} < n \tag{4}$$

The second step of the FKNN approach is to assign fuzzy memberships of the unknown sample x to different classes according to the following equation:

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij}\left(1/\|x - x_j\|^{2/(m-1)}\right)}{\sum_{j=1}^{k}\left(1/\|x - x_j\|^{2/(m-1)}\right)} \tag{5}$$

where $i = 1,2,\ldots,C$, and $j = 1,2,\ldots,k$. j represents the jth sample vector among the k nearest neighbors of x. C is the number of classes; k denotes the neighboring size. The fuzzy strength m is used to determine how heavily the distance is weighted when computing each neighbor's contribution to the membership value. $\|x - x_j\|$ represents the distance between x and its jth nearest neighbor x_j. In this study, Euclidean metric is used as the distance measurement. u_{ij}, denotes the membership degree of the sample vector x_j in the class i and is computed in the first step of the algorithm (refer to Eq. (1)).

2.2. Firefly Algorithm (FA)

In order to commence the training process of the FKNN, two tuning parameters (k, m) are required to be determined. A proper setting of these tuning parameters is necessary to achieve a desirable performance of the prediction model [29]. Thus, in this study, we utilize the FA as a means for tuning the FKNN parameters. The