Expert Systems with Applications 36 (2009) 12491–12499

Contents lists available at ScienceDirect

Expert Systems with Applications

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

# Generating prediction rules for liquefaction through data mining

Adil Baykasoğlu [a,*], Abdülkadir Çevik [b], Lale Özbakır [c], Sinem Kulluk [c]

[a] University of Gaziantep, Department of Industrial Engineering, Gaziantep, Turkey
[b] University of Gaziantep, Department of Civil Engineering, Gaziantep, Turkey
[c] Erciyes University, Department of Industrial Engineering, Kayseri, Turkey

ARTICLE INFO

*Keywords:*
Liquefaction
Neural networks
Ant colony optimization
Data mining

ABSTRACT

Prediction of liquefaction is an important subject in geotechnical engineering. Prediction of liquefaction is also a complex problem as it depends on many different physical factors, and the relations between these factors are highly non-linear and complex. Several approaches have been proposed in the literature for modeling and prediction of liquefaction. Most of these approaches are based on classical statistical approaches and neural networks. In this paper a new approach which is based on classification data mining is proposed first time in the literature for liquefaction prediction. The proposed approach is based on extracting accurate classification rules from neural networks via ant colony optimization. The extracted classification rules are in the form of IF–THEN rules which can be easily understood by human. The proposed algorithm is also compared with several other data mining algorithms. It is shown that the proposed algorithm is very effective and accurate in prediction of liquefaction.



## 1. Introduction

In soil deposits in undrained condition, earthquakes cause cyclic shear stresses, which may lead to liquefaction. Liquefaction indicates a condition where a soil will subject to a deformation, because of the build-up and maintenance of high pore water pressure. In a soil deposit subjected to dynamic load the pore water generation increases, leading to a reduction in its strength and then liquefaction of the soil. Field evidence of liquefaction usually consisted of observed sand boils, ground fissures or lateral spreads. Liquefaction in loose and saturated sands and silty sands is one of the causes of tilting or subsiding in civil engineering structures. In the literature, there are several methods developed to determine the liquefaction potential of soil deposits using the in situ tests results, such as; standard penetration test (SPT), and cone penetration test (CPT). Simplified methods have been used by scientists to evaluate nonlinear liquefaction potential of soil. Derived from several laboratory and field tests, various simplified procedures (e.g., strain-based, stress-based, Chinese criteria) have been developed by using case studies and undisturbed soil specimens.

Over the past 35 years, a stress-based procedure, termed as 'simplified procedure', has evolved for evaluating the liquefaction resistance of soils. Following the earthquakes in Alaska and Niigata, Japan in 1964, Seed and Idriss (1971) developed this method. The simplified procedure was developed from evaluations of experimental data and field observation. The number of cycles and the shear stress level are the major requirements in this method. Equivalent stress intensity and the number of cycles need to be determined to correlate actual motion of an earthquake to laboratory harmonic loading conditions. This equivalent shear stress was selected as 65% of the maximum shear stress induced in an earth structure in the study by Seed, Idriss, Makdisi, and Banerjee (1975). However, Ishihara and Yasuda (1975) said that the equivalent shear stress is 57% for 20 cycles of loading. Although the procedure has been corrected and augmented periodically since then with the studies by Seed (1979), Seed and Idriss (1982) and Seed et al. (1985), the uncertainty concerning random loading still exists (Baziar & Jafarian, 2007).

The results of research by Dobry, Ladd, Yokel, Chung, and Powell (1982) were employed to assess the liquefaction potential of soils which is known as strain-based approach. According to the strain-based approach, pore water pressure initiates to develop when shear strain surpass a threshold shear strain level. Dobry et al. (1982) concluded that the threshold shear strain is approximately 0.001%, and independent of relative density, sand type, initial effective confining pressure and sample preparation technique. The same uncertainty is available for this approach as well, since the Dobry et al. (1982) thought the same equivalent number of earthquake loading as the stress-based approach.

Over last three decades, various researchers have studied on an energy-based liquefaction assessment approaches. The amount of frictional energy loss needed to liquefy a soil is dependent on contact density, confining stress, and frictional characteristics of the soil. The cumulative energy loss up to liquefaction has been identified as a useful index for liquefaction potential of soils (Berrill &

* Corresponding author. Tel./fax: +90 342 3604383.
*E-mail address:* baykasoglu@gantep.edu.tr (A. Baykasoğlu).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.04.033

Davis, 1985; Nemat-Nasser & Shokooh, 1979; Okada & Nemat-Nasser, 1994; Thevanayagam, Liang, & Shenthan, 2000; Trifunac, 1995). This idea has also come up with the development of energy based liquefaction potential evaluation procedures (Green, 2001; Kayen & Mitchell, 1997) based on penetration resistances and arias intensity. Researches have proposed energy-based models relating pore water pressure increment ratio, to dissipated strain energy density, loading parameters (cyclic stress ratio, strain level), initial void ratio, or relative density, initial effective confining pressure, and some calibration parameters from curve fitting of experimental data. Parameters used in this approach can be directly related to seismological parameters (Green, 2001). The main advantage of the energy-based approach is that the shear energy required to liquefy a soil deposit is not dependent on the stress history.

Geotechnical engineers often solve complex problems having a serious of interacting parameters. However, in some problems, these parameters are not well defined and it is not possible to identify a relationship between the parameters, or the problem could be too complex to be described in a mathematical function. One of the most common approaches to solve these problems uses experimental data to develop empirical models that relate the variables in the system. Modern techniques such as fuzzy systems and neural networks have been considered to develop models based on data owing to their ability to learn and recognize trends in data patterns (Goh & Goh, 2007). For example, artificially intelligence (AI) applications have been used in several applications in civil engineering. Artificial neural network (ANN), one of the AI approaches, has the capability to mimic the learning abilities of human brain by processing the data. ANN models were recently adopted by researchers in the area of geotechnical engineering (e.g., Abu Keifa, 1998; Baziar & Jafarian, 2007; Chen, Hsieh, Chen, & Shih, 2005; Goh, 1994, 1995; Hanna, Morcous, & Helmy, 2004, 2007a, Hanna, Ural, & Saygili, 2007b; Teh, Wong, Goh, & Jaritngam, 1997; Wang & Rahman, 2002). Such as, Baziar and Jafarian (2007) developed an ANN model to correlate some of the soil parameters with the strain energy required liquefaction triggering. They carried out parametric study and confirmed the results using previously published cyclic triaxial, simple shear and torsional shear test results. Chen et al. (2005) assessed the liquefaction probability using an energy-based method with back-propagation neural network. The method proposed by them has a capability in evaluating the probability of soil liquefaction. Hanna, Ural, and Saygili (2007a, 2007b) proposed an alternative general regression neural network model for the two major earthquakes in Turkey and Taiwan. The proposed model was found to be a viable tool to assess the seismic conditions susceptible to liquefaction. Khozaghi and Choobbasti (2007) made a prediction for the liquefaction potential in southeast of Tehran city in Iran using neural network approach. Comparing their results with the results of Seed (1979) method, they found that the method they proposed using neural network made a prediction with 92% accuracy for the studied area. Apart from the recent studies, Wang and Rahman (2002) made one of the considerations that can be particularly connected with the neural network model for horizontal ground displacement caused by liquefaction. They had indicated that the neural network model serves as a reliable and simple predictive tool for the amount of ground displacement.

As it can be seen from the literature review, ANNs have found application potential in prediction of liquefaction. Although, ANN can achieve high prediction accuracies, an important drawback of them is their very limited explanation capability. ANNs are generally considered as black boxes as it is very difficult to understand how they learn and solve a given problem. In order to overcome this drawback several approaches have been proposed in data mining/pattern recognition research area. In these approaches, trained ANN is processed by another algorithm (*which is generally a meta-heuristic algorithm*) in order to extract accurate classification rules in the form of **IF–THEN** rules which can be easily understood by human decision makers. In the present paper, a data mining algorithm which is recently proposed by the authors is used to predict liquefaction first time in the literature. The proposed data mining algorithm works on trained ANN to extract accurate prediction (*classification*) rules by making use of an ant colony optimization algorithm. The algorithm is named as TACO-miner (Özbakır, Baykasoğlu, & Kulluk, in press). The details of the proposed ant colony algorithm are given in the following sections of this paper. Several other classical data mining algorithms along with a powerful genetic programming based technique (*MEPAR-miner, which is also recently proposed by authors*; Baykasoğlu & Özbakır, 2007) are also applied to liquefaction prediction problem in order to compare the performance of the proposed algorithm and to show how classification data mining techniques can be used.

## 2. Classification rule extraction from trained ANN via ant colony optimization

ANN is a powerful data modeling tool that is able to capture and represent complex input/output relationships. ANNs resemble the human brain in two aspects; firstly an ANN acquires knowledge through learning and secondly an ANN's knowledge is stored within inter-neuron connection strengths known as synaptic weights.

The knowledge acquired by an ANN is codified on its connection weights, which in turn are associated to both its architecture and activation functions (Andrews, Diederich, & Tickle, 1995). In this context, the process of knowledge acquisition from ANNs usually implies the use of algorithms based on the values of either connection weights or hidden unit activations. The algorithms designed to perform such task are generally called algorithms for rule extraction from ANNs (Hruschka & Ebecken, 2006).

In the present study, multi-layer perceptron (MLP) which is one of the most widely used ANN is considered. MLP is especially useful for classification problems. MLP uses *sigmoid* or *tanh* functions in general. ANN is trained on the encoded vectors of the input attributes and the corresponding vectors of the output classes until the convergence rate between actual and the desired output is achieved. In this study L-36 Taguchi Design is carried out for determining effective structural parameters of MLP and improving convergence rates for liquefaction dataset. The dataset used in this study is based on the cone penetration tests case records compiled by Juang, Yuan, Lee, and Lin (2003). The dataset contains of 226 records, comprising 133 liquefied cases and 93 non-liquefied cases. There are six input variables and one output variable which are the cone tip resistance $q_c$, the sleeve friction ratio $R_f$, the effective stress at the depth of interest $S_1$, the total stress at the same depth $S_2$, the maximum horizontal ground surface acceleration $a_{max}$, the earthquake moment magnitude $M_w$, and binary variable representing the existence of liquefaction $L$. The range of values associated with each input and output variable are shown in Table 1. Similar

**Table 1**
Attribute names and descriptions.

| Attribute | Description | Code | Range of attributes |
|---|---|---|---|
| $q_c$ | Cone tip resistance | $x_1$ | 0.9–25 |
| $R_f$ | Sleeve friction ratio | $x_2$ | 0.1–5.2 |
| $S_1$ | Effective stress | $x_3$ | 22.5–215.2 |
| $S_2$ | Total stress | $x_4$ | 26.6–274 |
| $a_{max}$ | Maximum horizontal ground surface acceleration | $x_5$ | 0.08–0.8 |
| $M_w$ | Earthquake moment magnitude | $x_6$ | 6–7.6 |
| $L$ (class) | Liquefaction exists or not (1 = exists, 0 = not) | $\lambda$ | 0 and 1 |