

Evaluation of liquefaction potential based on CPT results using C4.5 decision tree

A. Ardakani* and V. R. Kohestani

Faculty of Engineering & Technology, Imam Khomeini International University, Qazvin, Iran.

Received 5 September 2014; Accepted 18 February 2015

*Corresponding author: a.ardakani@eng.ikiu.ac.ir (A.Ardakani)

Abstract

The prediction of liquefaction potential of soil due to an earthquake is an essential task in civil engineering. The decision tree has a structure consisting of internal and terminal nodes, which process the data to ultimately yield a classification. C4.5 is a known algorithm widely used to design decision trees. In this algorithm, a pruning process is carried out to solve the problem of the over-fitting. This article examines the capability of C4.5 decision tree for the prediction of seismic liquefaction potential of soil based on the Cone Penetration Test (CPT) data. The database contains the information about cone resistance (q_c), total vertical stress (σ_0), effective vertical stress (σ'_0), mean grain size (D_{50}), normalized peak horizontal acceleration at ground surface (a_{max}), cyclic stress ratio (τ/σ'_0) and earthquake magnitude (M_w). The overall classification success rate for all the data set is 98%. The results of C4.5 decision tree have been compared with the available artificial neural network (ANN) and relevance vector machine (RVM) models. The developed C4.5 decision tree provides a viable tool for civil engineers to determine the liquefaction potential of soil.

Keywords: Soil Liquefaction, Cone Penetration Test, Artificial Intelligence, C4.5 Decision Tree.

1. Introduction

Seismically induced liquefaction in saturated soils is a phenomenon in which soil loses much of its strength or stiffness due to rising pore water pressure for a generally short period of time but nevertheless long enough for it to cause ground failure. Liquefaction of saturated sandy soils during the earthquakes causes building settlement or tipping, sand blows, lateral spreading, ground cracks, landslides, dam and high embankment failures and many other hazards. Liquefaction occurrence depends on the mechanical characteristics of the soil layers on the site, the depth of the water table, the intensity and duration of the ground shaking, the distance from the source of the earthquake and the seismic attenuation properties of in situ soil [1]. Determination of liquefaction potential of soil due to an earthquake is an important step for earthquake hazard mitigation. Because of the participation of a large number of factors that affect the occurrence of liquefaction during earthquake, the determination of liquefaction potential is a complex geotechnical engineering

problem and has attracted considerable attention of geotechnical researchers in the past three decades. Several methods have been proposed to predict the occurrence of liquefaction. Many of these methods, are based on correlations between the in situ test measurements and observed field performance data, and are extensions of the "simplified procedure" pioneered by Seed and Idriss [2].

Amongst in situ tests, many researchers have adapted Cone Penetration Test (CPT) results as the basis for evaluation of liquefaction potential of the test method [3,4]. A primary advantage of the CPT is the nearly continuous information provided along the depth of the explored soil strata. The CPT is also considered more consistent and repeatable than other in situ test methods.

Artificial intelligence (AI) techniques such as artificial neural network (ANN) [4-6], support vector machine (SVM) [6-8] and relevance vector machine (RVM) [9,10] have been used to develop liquefaction prediction models based on in situ test database.

These models have the ability to operate on large quantities of data and learn complex model functions from examples, i.e., by training on sets of input and output data.

The greatest advantage of AI techniques over traditional modeling techniques is their ability to capture non-linear and complex interaction between variables of the system without having to assume the form of the relationship between input and output variables.

In the context of determination of liquefaction occurrence, this method can be trained to measure the relationship between the soil and earthquake characteristics with the liquefaction potential, requiring no prior knowledge of the form of the relationship.

Even though most of the introduced AI techniques have been successfully applied to CPT data, they do have shortcomings. For example, in the ANN approach, the optimum structure (e.g., number of inputs, hidden layers, and transfer functions) must be identified as a priori. This is usually done through a trial and error procedure.

The other major shortcoming is the black box nature of ANN model and the fact that the relationship between input and output parameters of the system is described in terms of a weight matrix and biases that are not accessible to the user [11].

Decision trees algorithms are quite transparent and also do not need optimization of model and internal parameters. Either a decision tree partitions or the input space of data set into mutually exclusive regions is assigned a label (classification tree) or a value to characterize its data points (regression tree). The decision tree has a structure consisting of internal and external nodes connected by branches. Each internal node is associated with a decision function to determine which node to visit next. Meanwhile, each external node, known as a terminal node or leaf node, indicates the output of a given input vector.

Figure 1 shows partitions of the input space into four non-overlapping rectangular regions, and each of which is assigned a labeled class 'Ci'.

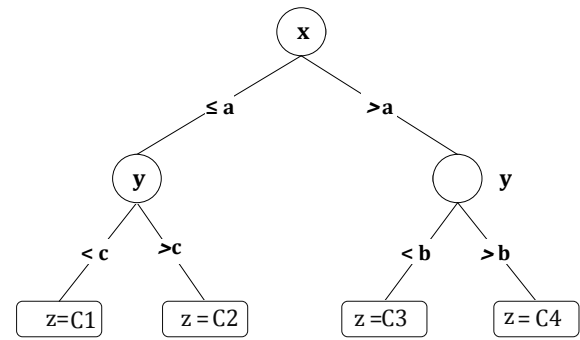
C4.5 introduced by Quinlan [12] is a known algorithm widely used to design decision trees. This paper investigates the capability of C4.5 decision tree for the prediction of liquefaction potential of soil based on CPT data.

2. Materials and methods

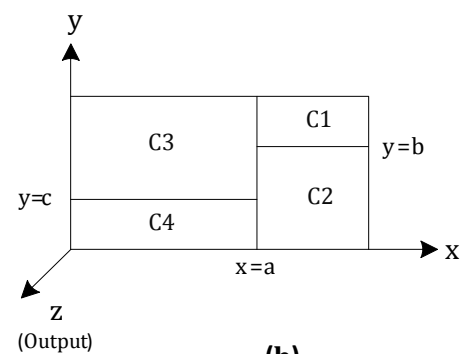
2.1. Decision trees

Decision trees are fast and easy to use. The rules generated by decision trees are simple and accurate for most problems. Therefore, decision

trees are very popular and powerful tools in data mining [14].



(a)



(b)

Figure 1. An example of a decision tree for classification (a) binary decision tree (b) feature space partitioning [13].

In general, a decision tree is a tree in which each branch node represents a choice between a number of alternatives and each leaf node represents a classification or decision [15]. An unknown (or test) instance is routed down the tree according to the values of the attributes in the successive nodes. When the instance reaches a leaf, it is classified according to the label assigned to the corresponded leaf.

In the first stage of model construction, a decision-tree induction algorithm is used to build the tree. Many algorithms for decision tree induction exist. Interactive Dichotomizer version 3 (ID3) and Commercial version 4.5 (C4.5) [13,16] are the most widely used with the classification and regression tree (CART) algorithm [17]. C4.5 algorithm is an extension of ID3 algorithm and the divide-and-conquer approach [12] whose main improvements included the pruning methodology and the processing of numeric attributes, missing values and noisy data.

The construction phase is begun at the root node where each attribute is evaluated using a statistical test to determine how well it can classify the training samples. The best attribute is chosen as

the test at the root node of the tree. A descendant of the root node is then created for each either possible value of this attribute if it is a discrete-valued attribute or possible discretized interval of this attribute if it is a continuous-valued attribute. Next, the training samples are sorted to the appropriate descendant node.

The process is repeated using the training samples associated with each descendant node to select the best attribute for testing at that point in the tree. This forms a greedy search for a decision tree, in which the algorithm never backtracks to reconsider earlier node choices. Although it is possible to add a new node to the tree until all samples assigned to one node belong to the same class, the tree is not allowed to grow to its maximum depth. A node is only introduced to the tree only when there are a sufficient number of samples left from sorting. After the complete tree is constructed, a tree pruning is usually carried out to avoid data over-fitting.

A statistical test used in C4.5 for assigning an attribute to each node in the tree also employs an entropy-based measure. The assigned attribute is the one with the highest information gain ratio among attributes available at that tree construction point. The information gain ratio Gain Ratio(A, S) of an attribute 'A' relative to the sample set S is defined as

$$Gain\ Ratio(A, S) = \frac{Gain(A, S)}{Split\ Information(A, S)} \quad (1)$$

Where

$$Gain(A, S) = Ent(S) - \sum_{a \in A} \frac{|S_a|}{|S|} Ent(S_a) \quad (2)$$

and

$$Split\ Information(A, S) = - \sum_{a \in A} \frac{|S_a|}{|S|} \log_2 \frac{|S_a|}{|S|} \quad (3)$$

S_a is the subset of S for which the attribute A has the value a. Obviously, the information gain ratio can be calculated straightaway for discrete-valued attributes. In contrast, continuous-valued attributes are needed to be discretised prior to the information gain ratio calculation.

2.2. Database

The database [5] used in this study consists of total 109 cases, 74 of them are liquefied cases and 35 of them are non-liquefied cases. The database contains: cone resistance (q_c), total vertical stress (σ_0), effective vertical stress (σ'_0), mean grain size (D_{50}), normalized peak horizontal acceleration at ground surface (a_{max}), cyclic stress ratio (τ/σ'_0) and earthquake magnitude (M_w). The range of

values associated with each input variable is shown in table 1.

Generally in pattern recognition procedures (e.g., ANN, SVM or GP) it is common that the model construction is based on adaptive learning over a number of cases and the performance of the constructed model is then evaluated using an independent validation data set. Therefore, in the present study, a total of 74 datasets are considered for the training dataset, and other datasets are considered for the testing dataset. The training and testing datasets are the same as the ones used by Goh [5] and Samui [9].

Table 1. Range of values associated with the input variables used in liquefaction analysis.

Input variable	Range
Cone tip resistance, q_c (MPa)	1.0-31.4
Total vertical stress, σ_0 (kPa)	17-122
Effective vertical stress, σ'_0 (kPa)	17-249
Mean grain size, D_{50} (mm)	0.06-0.67
Maximum horizontal ground surface acceleration, a_{max} (g)	0.1-0.8
Cyclic stress ratio, τ/σ'_0	0.06-0.72
Earthquake moment magnitude, M_w	6.6-7.8

3. Result and discussion

For building the model based on training data set, C4.5 algorithm implemented in WEKA software [18] was used. WEKA is written in Java and is freely available from Waikato University website [19].

The Decision tree generated by C4.5 algorithm is shown in figure 2, table 2 and table 3 illustrate the performance of C4.5 decision tree for training and testing dataset respectively. For the training patterns, three cases of liquefaction were wrongly classified. For the testing patterns, one case of liquefaction was wrongly classified.

Training and testing performance have been determined by using (4).

$$Training = Testing\ performance\ (\%) = \left(\frac{\text{No of data predicted accurately by C4.5}}{\text{Total data}} \right) \times 100 \quad (4)$$

As the results presented in table 2, the performances of C4.5 decision tree for training and testing data are comparable.

The successful prediction values are 95.9% for training and 97.1% for testing data whereas the overall success rate in predicting liquefaction in all cases is 96.3%. The overall classification success rate for the entire data set is slightly lower than the overall

rate of ANN and RVM models reported by Goh [5] and Samui [9], respectively (see Table 4).

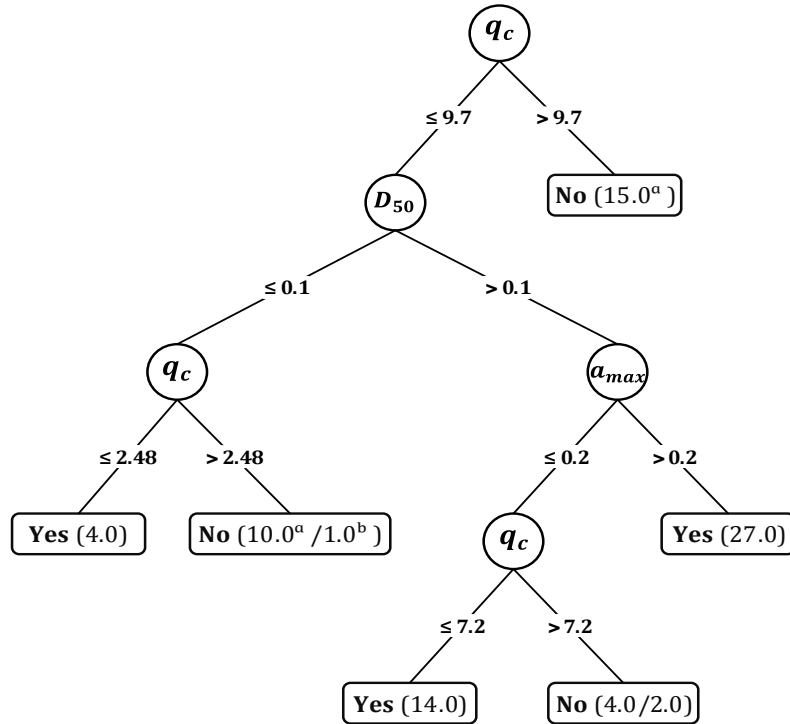


Figure 2. Decision tree generated by C4.5 algorithm.

^a Number of cases in this partition.

^b Number of cases misclassified.

Table 2. Performance of C4.5 for training dataset.

M_w	σ'_0 (kPa)	σ_0 (kPa)	q_c (MPa)	a_{max} (g)	τ/σ'_0	D_{50} (mm)	Actual liquefied	Predicted liquefied?
7.5	53	36	3.20	0.16	0.15	0.331	Yes	Yes
7.5	87	52	1.60	0.16	0.16	0.331	Yes	Yes
7.5	99	58	7.20	0.16	0.17	0.331	Yes	Yes
7.5	152	83	5.60	0.16	0.17	0.331	Yes	Yes
7.5	91	63	5.45	0.16	0.14	0.331	Yes	Yes
7.5	127	80	8.84	0.16	0.15	0.331	Yes	No
7.5	211	120	9.70	0.16	0.15	0.331	Yes	No
7.5	86	46	8.00	0.16	0.19	0.30	No	No
7.5	95	50	14.55	0.16	0.18	0.30	No	No
7.7	58	48	10.00	0.23	0.18	0.32	No	No
7.7	73	54	16.00	0.23	0.20	0.32	No	No
7.7	96	65	15.38	0.23	0.21	0.32	No	No
7.7	54	46	1.79	0.23	0.17	0.32	Yes	Yes
7.7	64	52	4.10	0.23	0.19	0.32	Yes	Yes
7.7	96	67	7.96	0.23	0.21	0.32	Yes	Yes
7.7	114	75	8.97	0.23	0.22	0.32	Yes	Yes
7.8	57	42	1.70	0.40	0.35	0.06	Yes	Yes
7.8	114	69	9.40	0.40	0.41	0.25	Yes	Yes
7.8	148	85	5.70	0.40	0.42	0.25	Yes	Yes
7.8	162	92	7.60	0.40	0.42	0.30	Yes	Yes
7.8	17	17	1.50	0.40	0.27	0.17	Yes	Yes
7.8	25	21	1.00	0.40	0.32	0.17	Yes	Yes
7.8	34	25	5.00	0.40	0.36	0.17	Yes	Yes
7.8	38	34	2.50	0.40	0.29	0.14	Yes	Yes

M_w	σ'_0 (kPa)	σ_0 (kPa)	q_c (MPa)	a_{max} (g)	τ/σ'_0	D_{50} (mm)	Actual liquefied	Predicted liquefied?
7.8	57	43	2.60	0.40	0.34	0.14	Yes	Yes
7.8	76	52	3.20	0.40	0.37	0.16	Yes	Yes
7.8	89	58	5.80	0.40	0.39	0.16	Yes	Yes
7.8	122	74	3.50	0.40	0.40	0.16	Yes	Yes
7.8	181	102	8.40	0.40	0.41	0.16	Yes	Yes
7.8	38	29	1.70	0.40	0.35	0.12	Yes	Yes
7.8	40	29	3.50	0.40	0.36	0.12	Yes	Yes
7.8	51	35	4.10	0.40	0.38	0.12	Yes	Yes
7.8	29	27	5.50	0.40	0.29	0.17	Yes	Yes
7.8	57	40	9.00	0.40	0.37	0.32	Yes	Yes
7.8	23	21	7.00	0.40	0.29	0.48	Yes	Yes
7.8	34	26	1.18	0.40	0.35	0.48	Yes	Yes
7.8	48	33	4.24	0.40	0.38	0.48	Yes	Yes
7.8	76	71	11.47	0.40	0.27	0.16	No	No
7.8	160	111	15.76	0.40	0.34	0.20	No	No
7.8	59	56	11.39	0.20	0.14	0.21	No	No
7.8	78	65	12.12	0.20	0.15	0.21	No	No
7.8	99	75	17.76	0.20	0.17	0.14	No	No
7.8	74	49	2.65	0.20	0.19	0.14	Yes	Yes
7.8	53	35	4.40	0.20	0.20	0.16	Yes	Yes
7.8	61	39	3.00	0.20	0.20	0.16	Yes	Yes
7.8	156	81	9.00	0.20	0.23	0.08	Yes	No
7.8	99	55	2.00	0.10	0.11	0.14	Yes	Yes
7.8	95	52	1.10	0.20	0.23	0.07	Yes	Yes
7.8	209	106	15.50	0.10	0.11	0.08	No	No
7.8	217	110	6.50	0.10	0.11	0.08	No	No
7.8	91	53	9.00	0.10	0.11	0.10	No	No
7.8	101	58	2.50	0.10	0.11	0.10	No	No
7.8	112	63	16.50	0.10	0.11	0.10	No	No
7.8	91	68	13.65	0.10	0.06	0.25	No	No
7.8	114	58	8.47	0.20	0.24	0.062	No	No
7.8	228	112	4.55	0.20	0.23	0.067	No	No
7.8	249	122	5.79	0.20	0.22	0.067	No	No
7.8	121	55	2.48	0.20	0.25	0.062	Yes	Yes
7.8	114	56	1.57	0.20	0.25	0.062	Yes	Yes
7.8	213	103	1.45	0.20	0.23	0.67	Yes	Yes
7.8	220	106	2.15	0.20	0.23	0.67	Yes	Yes
7.8	230	111	2.60	0.20	0.23	0.67	Yes	Yes
7.8	213	103	2.73	0.20	0.23	0.67	Yes	Yes
7.8	219	106	1.78	0.20	0.23	0.67	Yes	Yes
7.8	211	108	7.64	0.20	0.22	0.67	No	No
6.6	29	29	25.60	0.80	0.44	0.11	No	No
6.6	48	36	24.70	0.80	0.57	0.11	No	No
6.6	64	42	31.40	0.80	0.64	0.11	No	No
6.6	29	29	1.43	0.80	0.44	0.11	Yes	Yes
6.6	64	42	2.48	0.80	0.64	0.11	Yes	Yes
6.6	96	54	4.03	0.80	0.72	0.11	Yes	Yes
6.6	29	29	3.30	0.80	0.44	0.06	No	No
6.6	48	36	8.80	0.80	0.57	0.06	No	No
6.6	64	42	6.70	0.80	0.64	0.06	No	No

Table 3. Performance of C4.5 for testing dataset.

M_w	σ'_0 (kPa)	σ_0 (kPa)	q_c (MPa)	a_{max} (g)	τ/σ'_0	D_{50} (mm)	Actual liquefied	Predicted liquefied?
7.8	89	78	1.65	0.20	0.41	0.17	Yes	Yes
7.8	99	83	3.65	0.20	0.15	0.17	Yes	Yes
7.8	29	25	1.03	0.20	0.15	0.19	Yes	Yes
7.8	55	37	5.00	0.20	0.19	0.31	Yes	Yes
7.8	76	47	2.91	0.20	0.21	0.18	Yes	Yes
7.8	105	61	6.06	0.20	0.21	0.18	Yes	Yes
7.8	23	22	13.24	0.20	0.14	0.17	No	No
7.8	32	26	13.06	0.20	0.16	0.17	No	No
7.8	40	30	16.59	0.20	0.18	0.17	No	No
7.8	59	59	10.59	0.20	0.13	0.26	No	No
7.8	63	61	9.12	0.20	0.13	0.26	No	No
7.8	76	67	11.29	0.20	0.15	0.26	No	No
7.8	70	41	1.94	0.20	0.22	0.16	Yes	Yes
7.8	76	44	5.00	0.20	0.22	0.16	Yes	Yes
7.8	70	47	2.24	0.20	0.19	0.14	Yes	Yes
7.8	114	79	14.12	0.20	0.09	0.25	No	No
7.8	162	102	18.94	0.20	0.09	0.28	No	No
7.8	44	44	3.52	0.20	0.13	0.16	Yes	Yes
7.8	59	51	2.73	0.20	0.15	0.16	Yes	Yes
7.8	57	50	3.29	0.20	0.15	0.21	Yes	Yes
7.8	61	52	4.12	0.20	0.15	0.21	Yes	Yes
7.8	72	57	2.94	0.20	0.16	0.21	Yes	Yes
7.8	61	52	3.00	0.20	0.15	0.15	Yes	Yes
7.8	95	68	5.85	0.20	0.18	0.32	Yes	Yes
7.8	106	73	9.00	0.20	0.18	0.32	Yes	No
7.8	49	48	1.88	0.20	0.13	0.13	Yes	Yes
7.8	74	64	2.55	0.20	0.15	0.17	Yes	Yes
7.8	76	65	4.50	0.20	0.15	0.17	Yes	Yes
7.8	106	79	4.24	0.20	0.17	0.17	Yes	Yes
7.8	114	77	8.00	0.20	0.18	0.22	No	No
7.2	80	48	5.22	0.22	0.21	0.20	Yes	Yes
7.2	95	55	3.73	0.22	0.22	0.20	Yes	Yes
7.2	114	64	3.11	0.22	0.22	0.20	Yes	Yes
7.2	133	73	1.32	0.22	0.22	0.20	Yes	Yes
7.2	152	82	5.22	0.22	0.22	0.20	Yes	Yes

Table 4. Comparison of results of developed C4.5 decision tree with available ANN [5] and RVM models [9].

Method	Performance in terms of successful prediction (%)		
	Training	Testing	Overall
ANN [5]	98.6	94.3	97.2
RVM [9]	100	100	100
C4.5 decision tree	95.9	97.1	96.3

The ANN uses many parameters, such as the number of hidden layers, number of hidden nodes,

learning rate, momentum term, number of training epochs, transfer functions, and weight initialization methods. Though the RVM has lower parameters compared with ANN, but RVM requires a selection of a suitable kernel function first and then setting of the specific parameters and these processes are time consuming. Moreover, these techniques will not produce an explicit relationship in the variables and thus, the developed model provides very little insight into the basic mechanism of the problem. Decision

trees algorithms are quite transparent and also do not need optimization of model and internal parameters. The developed C4.5 decision tree, figure 2, can be used by geotechnical engineering professionals with the help of a spreadsheet to evaluate the liquefaction potential of soil for a future seismic event without going into complexities of model development whereas the available ANN and RVM models do not provide any explicit equations for professionals. Also in the C4.5 approach normalization or scaling of the data is not required, but is an advantage over ANN and RVM approach.

The limitations of the C4.5 decision trees need to be mentioned as well. Similar to other artificial intelligence techniques, decision trees have a limited domain of applicability and are mostly case dependent. Therefore, their generalization is limited and they are only applicable in the range of training data. However, the C4.5 model can always be updated to yield better results, as new data becomes available.

4. Conclusions

Liquefaction in soil is one of the major causes of concern in geotechnical engineering. The cone penetration test has proven to be an effective tool in characterization of subsurface conditions and analysis of different aspects of soil behavior, comprising estimating the potential for liquefaction on a specific site. In this paper, the C4.5 decision tree is used to predict the liquefaction potential of soil based on CPT data. The C4.5 model was trained and validated using a database of 109 liquefaction and non-liquefaction field case histories for sandy soils based on CPT results. The overall classification success rate for the entire data set is 96.3% and is comparable with those calculated using ANN and RVM models which were taken in the literature. Unlike available ANN and RVM models, the proposed model provide easily interpretable tree structure that can be used by geotechnical engineering professionals with the help of a spreadsheet to predict the liquefaction potential of soil for future seismic event without going into the complexities of model development using C4.5 decision tree. This model can be adopted for modeling different problems in geosciences.

References

[1] Rezania, M. Javadi, A. A. & Giustolisi, O. (2010). Evaluation of liquefaction potential based on CPT results using evolutionary polynomial regression. *Computers and Geotechnics*, vol. 37, no. 1, pp. 82-92.

[2] Seed, H. B. & Idriss, I. M. (1971). Simplified procedure for evaluating soil liquefaction potential. *Journal of the Soil Mechanics and Foundations*, vol. 97, no. 9, pp. 1249-1273.

[3] Youd, T. Idriss, I. Andrus, R. D. Arango, I. Castro, G. Christian, J. T. Dobry, R. Finn, W. L. Harder, L. F. & Hynes, M. E. (2001). Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 127, no. 10, pp. 817-833.

[4] Juang, C. H. Yuan, H. Lee, D. H. & Lin, P. S. (2003). Simplified cone penetration test-based method for evaluating liquefaction resistance of soils. *Journal of geotechnical and geoenvironmental engineering*, vol. 129, no. 1, pp. 66-80.

[5] Goh, A. T. (1996). Neural-network modeling of CPT seismic liquefaction data. *Journal of Geotechnical engineering*, vol. 122, no. 1, pp. 70-73.

[6] Samui, P. & Sitharam, T. (2011). Machine learning modelling for predicting soil liquefaction susceptibility. *Natural Hazards and Earth System Science*, vol. 11, pp. 1-9.

[7] Goh, A. T. & Goh, S. (2007). Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data. *Computers and Geotechnics*, vol. 34, no. 5, pp. 410-421.

[8] Pal, M. (2006). Support vector machines based modelling of seismic liquefaction potential. *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 30, no. 10, pp. 983-996.

[9] Samui, P. (2007). Seismic liquefaction potential assessment by using relevance vector machine. *Earthquake Engineering and Engineering Vibration*, vol. 6, no. 4, pp. 331-336.

[10] Samui, P. & Karthikeyan, J. (2014). The Use of a Relevance Vector Machine in Predicting Liquefaction Potential. *Indian Geotechnical Journal*, vol. 44, no. 4, pp. 458-467.

[11] Javadi, A. A. Ahangar-Asr, A. Johari, A. Faramarzi, A. & Toll, D. (2012). Modelling stress-strain and volume change behaviour of unsaturated soils using an evolutionary based data mining technique, an incremental approach. *Engineering Applications of Artificial Intelligence*, vol. 25, no.5, pp. 926-933.

[12] Quinlan, J. R. (1993). C4.5: programs for machine learning. San Francisco: Morgan Kaufmann.

[13] Cho, J. H. & Kurup, P. U. (2011). Decision tree approach for classification and dimensionality reduction of electronic nose data. *Sensors and Actuators B: Chemical*, vol. 160, no. 1, pp. 542-548.

- [13] Quinlan, J. R. (1993). C4.5: programs for machine learning. San Francisco: Morgan Kaufmann.
- [14] Duch, W. Setiono, R. & Zurada, J. M. (2004). Computational intelligence methods for rule-based data understanding. Proceedings of the IEEE, vol. 92, no. 5, pp. 771-805.
- [15] Hand, D. J. Mannila, H. & Smyth, P. (2001). Principles of data mining. London: MIT press.
- [16] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, vol. 1, no. 1, pp. 81-106.
- [17] Breiman, L. Friedman, J. H. Olshen, R. A. & Stone, C. J. (1984). Classification and regression trees. Belmont: Wadsworth.
- [18] Witten, I. H. & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. San Francisco: Morgan Kaufmann.
- [19] Waikato University website (2015), Available: <http://www.cs.waikato.ac.nz/~ml/>.

ارزیابی پتانسیل روانگرایی بر مبنای نتایج CPT با استفاده از درخت تصمیم C4.5

علیرضا اردکانی* و وحیدرضا کوهستانی

دانشکده فنی و مهندسی، دانشگاه بین المللی امام خمینی (ره)، قزوین، ایران.

ارسال ۲۰۱۴/۰۹/۰۵؛ پذیرش ۲۰۱۵/۰۲/۱۸

چکیده:

پیش‌بینی پتانسیل روانگرایی خاک‌ها هنگام وقوع زلزله یکی از وظایف اصلی در مهندسی عمران است. درختان تصمیم یک ساختار درختی شامل گره‌های داخلی و ترمینال‌ها هستند که با پردازش روی داده‌ها، در نهایت منجر به طبقه‌بندی آن‌ها می‌شود. C4.5 یکی از الگوریتم‌های پرکاربرد در طراحی درختان تصمیم است. در این الگوریتم، به منظور حل مشکل بیش‌پردازش، یک مرحله هرس انجام می‌شود. این مقاله توانایی درخت تصمیم C4.5 را برای پیش‌بینی پتانسیل روانگرایی خاک‌ها بر مبنای داده‌های آزمایش نفوذ مخروط (CPT) بررسی می‌کند. پایگاه داده شامل اطلاعاتی درباره مقاومت مخروط (q_c)، تنش قائم کل (σ_0)، تنش قائم مؤثر (σ'_0)، میانگین اندازه دانه‌ها (D_{50})، شتاب افقی ماکزیمم در سطح زمین (a_{max})، نسبت تنش سیکیلی (τ/σ'_0) و بزرگی زلزله (M_w) است. درصد موفقیت کلی طبقه‌بندی داده‌ها، ۹۸ درصد به دست آمده است. نتایج درخت تصمیم C4.5 با مدل‌های موجود شبکه عصبی مصنوعی (ANN) و ماشین بردار ارتباط (RVM) مقایسه شده است. درخت تصمیم C4.5 شکل گرفته، یک ابزار مفید برای مهندسان عمران فراهم می‌کند تا با استفاده از آن پتانسیل روانگرایی خاک‌ها را تعیین کنند.

کلمات کلیدی: روانگرایی، آزمایش نفوذ مخروط، هوش مصنوعی، درخت تصمیم C4.5.