CSCE Annual Conference

Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)

June 12 - 15, 2019

ENSEMBLE LEARNING FOR CLASSIFICATION OF SOIL LIQUEFACTION

Alobaidi, M.H.[1, 3], Meguid, M.A.[1], Chebana, F.[2]
[1] McGill University, Canada
[2] ETE-INRS, Canada
[3] mohammad.alobaidi@mail.mcgill.ca

Abstract: Seismic-induced liquefaction prediction is an important application of classification problems. Machine learning offers evolutionary prediction models. Moreover, Ensemble learning is a recent advancement in this field, where a number of learners are trained and their inferences are integrated to produce stable as well as improved generalization ability. There has been very limited work on the utilization of ensemble learning in the field of interest. In this work, a review of some of the useful ensemble learning approaches available in the literature is presented. A comprehensive analysis of eight ensemble models is performed using five machine learners and three ensemble approaches. The results demonstrate improved prediction performance and diminishing uncertainty of ensembles, compared with single models.

1 INTRODUCTION

Traditionally, techniques used to determine the liquefaction susceptibility of soil mainly rely on informed-determination of liquefaction threshold from observed soil behavior records. The state-of-the-art contributing to this field mainly rely on the development of better usage of such information. The major variables used in developing liquefaction prediction models are the Standard Penetration Test (SPT), Cone Penetration Test (CPT) and Shear Wave Velocity (Baez, Martin, & Youd, 2000; Bolton Seed, Tokimatsu, Harder, & Chung, 1985; Glaser, 1996; Juang, Chen, Tang, & Rosowsky, 2000; Kayen et al., 2013; Mola-Abasi, Kordtabar, & Kordnaeij, 2018; Robertson, 2015). At least one of these variables is included as a descriptor to infer the liquefaction event of the site of interest. Moreover, the literature utilizes these variables in different merits when developing their modeling technique.

Recently, machine learning techniques have been proposed in the broad literature and provided superior performance in learning complex relationships while maintaining a reliable generalization ability. The most common machine learning models used in the specific problem of seismic-induced liquefaction prediction are Support Vector Machines (SVMs) (Hoang & Bui, 2018), Decision Trees (DTs) (Ardakani & Kohestani, 2015), Artificial Neural Networks (ANNs) (Juang et al., 2000), and Extreme Learning Machines (ELMs) (Samui, Jagan, & Hariharan, 2016). These techniques have the common trait of being categorized under Supervised Learning techniques, where the target of the learning process is identified and a functional mapping between some explanatories and the target states is then produced by the specific model. Obviously, this process is of empirical nature and requires available intelligence, or information, that prompts the learning of functional mapping.

Ensemble learning is generally defined as the process of generating multiple prediction models, which are trained using subsets of the available training data and then fused to make a prediction, either for classification or regression applications. Figure (2) depicts the major identity components of Ensemble modeling and the general classification of Ensembles in the literature. Ensemble learning has received lately a significant interest. Continuous work has been conveyed in the broad literature to discuss the

effectiveness of ensemble modeling (Benkeser, Ju, Lendle, & van der Laan, 2018; Krawczyk, Minku, Gama, Stefanowski, & Woźniak, 2017; Luo, Wang, Yue, Liu, & Guo, 2018). Generally, for a learning framework to be considered an ensemble model, it should have three fundamental elements (or stages);

(1) Resampling: The process of generating multiple training samples, i.e. resamples, from the available sample. Common resampling methods are Bootstrap and Jackknife. Detailed information related to important resampling plans are reported by Bradly (Efron, 1982). The most common resampling technique in the literature is bootstrapping, where simple random sampling, with replacement, is carried out to create the resamples.

(2) Sub-ensemble model generation and training: Training and validating the individual models separately, in-parallel or in-series, depending on the adopted ensemble architecture. In this stage, each individual model will use the data from a corresponding resample in order to be trained. Because each resample is expected to comprise different sets of observations from the original dataset, the sub-ensemble models will each have a different optimal solution to their parameters. In this stage, if a new observation is inputted in the individual models, the corresponding estimate of the output observation, i.e. the individual estimates, will vary among these models.

(3) Ensemble integration: The process of fusing the sub-ensemble estimates into one result, called the ensemble estimate. In classification problems, the simplest, and most common, ensemble integration procedure is majority voting.

In addition, ensemble learning frameworks are divided into two broad clusters; Homogeneous ensembles and Non-Homogeneous ensembles (Brown, 2011; C. Zhang & Ma, 2012; Z.-L. Zhang, Luo, Yu, Yuan, & Tang, 2018). In Homogeneous ensemble frameworks, we adopt the same resampling technique, the same version of a certain model, and only one integration technique. For example, a learning framework that uses Jackknife resampling to generate the resamples, logistic regression with the same input-output configuration (generated multiple times and trained, each, using a different resample), and majority voting as a combiner, is considered an ensemble model. Non-Homogeneous ensemble frameworks are ensemble models which violate the definition of homogenous ensembles, while maintaining the three fundamental stages of ensemble learning. In the previously mentioned example, if individual models have different input-output configurations where some use all the available descriptor variables while others do not, the resulting ensemble model is of Non-homogeneous nature. While this group of ensembles is an open-ended one, the creation of non-homogeneous ensembles allows for innovative learning architectures. A common use of Non-Homogeneous ensemble frameworks is usually used by hyperspectral image classification studies called sub-space ensemble learning (Polikar, 2012; Z.-L. Zhang et al., 2018). In this work, Homogeneous learning frameworks are adopted to show the improved prediction performance of ensemble models. In addition, Homogeneous ensemble learning is recently picking up pace in the literature (Mitchell, 1997; Polikar, 2012). In following section, the diversity-in-learning concept is presented. Also, a description of the adopted ensemble models is provided.

2 METHODOLOGY

2.1 Bagging

Bagging, formally identified as Bootstrap Aggregation, is one of the most common Ensemble machine learning techniques (Breiman, 1996a). In this learning process, k sub-samples are created from the original sample set, D, available for training the individual models. Using the bootstrap resampling technique, each sub-sample, also called a resample d_i (i = 1, 2, …, k), is generated using random sampling with replacement. In addition, each sub-sample will have the same size as the original sample set, available for training the individual models. This resampling plan prompts a uniform distribution for the selection process. Each observation in the original dataset will have a probability 1/n of being chosen, where n is the size of the original sample set, D.

The type of the individual models is predetermined and their selection can be based on the type of problem tackled and experimental results on single learning trials. Each model will rely on one of the created

resamples in order to solve for its optimal configuration, during their training and validation. In this work, homogeneous ensembles are considered in general, and as a result all the individual models will be of the same type and initial input/output configuration. After all the ensemble members are generated and trained, a unique output for the ensemble can be derived by averaging the outputs from these individual models. Hence, an ensemble combiner is adopted in the final ensemble learning stage. Suppose that the descriptor variables' observations of a test instance *i* has been inputted into the *k* individual models, each of these models will have a unique output, or estimate, prediction, etc., and the ensemble output can be calculated using majority voting, as follows:

$$[1] \qquad \hat{s}_{i,ensemble} = round\left(\frac{1}{k}\sum_{j=1}^{k}\hat{s}_{i,j}\right), \hat{s}_{i,j} \in [0,1]$$

where $\hat{s}_{ensemble}$ is the resulting ensemble output, and $\hat{s}_{i,1}$ is the output from the first individual model, describing the estimate of i^{th} test instance.

It is important to note that equation (1) relies on the 1/0 configuration of the binary classification problem, and other modification can be used if the classes are assigned two different numbers. Because Bagging is essentially a parallel ensemble learning framework, the described algorithm can be parallelized in the computational environment. To this extent, it is obvious that the main diversity-in-learning effect is expected to manifest from the resampling plan adopted in Bagging. Distinct training data is used to enforce a spectrum of solutions to the individual models and, in return, provide various estimates most which are expected to provide improved prediction. The majority voting elects the ensemble output, eliminating the outlier estimates. Depending on the sub-ensemble models used, a diversity effect can also be created from the individual models if they are known to produce locally-optimum solutions to their configuration even if trained using the same dataset. The improved generalization ability of Bagging has been discussed and shown over many case studies in the broad literature (Alobaidi, Chebana, & Meguid, 2018; M. Cao et al., 2017; Erdal & Karakurt, 2013).

2.2 Stacking

Stacking is an ensemble technique that deals with the ensemble integration particularly (Wolpert, 1992). To create the stacking combiner, an additional model is used to learn how to combine the individual members, by tuning its weights over the feature space. To demonstrate the Stacking strategy, we start with the simplest linear combination. Suppose we derive *k* sub-samples using a particular resampling technique, such as Bootstrap Resampling, and then *k* individual models are created and trained using the generated resamples. The i^{th} pattern (i.e. an instance) has an observed value s_i and an output, or predicted value, $\hat{s}_{i,j}$ obtained from the j^{th} sub-ensemble model (*j = 1, 2, …, k*). Combining the individual predictions and deriving the ensemble output using simple averaging is defined in equation (2). Under stacking, we label the individual models as level 0 generalizers. At this point, the set of level 0 outputs, for a given pattern, is fed to a level 1 generalizer, which is a separate model that is trained to produce the appropriate output. Several stacking strategies are available in the literature, and all are based on level 0 and level 1 generalization approach.

For example, the stacking algorithm developed by Breiman (Breiman, 1996b) suggests minimizing the following error function:

$$[2] \qquad E(c_1, c_2, \dots, c_k) = \sum_{i=1}^{n}\sum_{j=1}^{k}\left[s_i - c_j \times \hat{s}_{i,j}\right]^2, \qquad c_j \geq 0$$

s_i is the i^{th} observation from the original training dataset. $\hat{c}_j$ is the estimate for the combiner coefficients, which are used to construct the ensemble prediction as follows:

$$[3] \qquad \hat{s}_{i,ensemble} = round\left(\sum_{j=1}^{k} c_j \times \hat{s}_{i,j}\right), \hat{s}_{i,j} \in [0,1]$$

3 EXPERIMENTAL SETUP

3.1 Data

The database used in this study has been originally compiled by Tokimatsu and Yoshimi (1983). Moreover, the specific dataset used to demonstrate the ensemble learning frameworks is selected by Goh (Goh, 1994). Eight variables are considered as explanatories to liquefaction potential. The total vertical stress (σ_o), effective vertical stress (σ_o'), Standardized SPT $(N_1)_{60}$, normalized peak horizontal acceleration (a/g), equivalent dynamic shear stress (τ_{av}/σ_o'), Fines content (F), and the average grain size (D_{50}), are used in this study. In addition, 85 observations are available from the final database, where 42 liquefaction observation and 43 non-liquefaction observations are recorded. More details on the significance of the database and the descriptor variables can be found in (Goh, 1994, 1996). The training and testing datasets comprised 59 observations and 26 observations, respectively, following the experimental setup by Goh (1996). The significance of the utilized explanatory variables in this study has been thoroughly investigated in the relevant literature and used, also in various forms and combination, for the liquefaction potential assessment.

It is important to note that the same input-output configuration is used for all the ensemble members in order to achieve an equivalent performance comparison. While this particular database is used in the current study, many other databases exist in the literature and can be used. However, a preliminary analysis should be carried out in order to determine the optimum explanatory variables from the available database as well as determine a class-balancing procedure in case the database has a relatively large difference in the number of observations for each class (Galar, Fernandez, Barrenechea, Bustince, & Herrera, 2012). The present database has a balanced number of observations in the liquefaction/no-liquefaction classes and previous work has already determined the optimality of the eight variables considered (Goh, 1994).

3.2 Individual Models

Three different ensemble members are considered in this study. Hence, for each ensemble architecture, an ensemble model is created using one of these single models. More precisely, Ensemble-based Logistic Regression (ELR), Support Vector Machines (ESVM), Decision Trees and Extreme Learning Machine (EELM) models are considered. The ensemble form of Decision Trees is commonly known in the broad literature as Random Forests (RF). The literature is abundant with research studies utilizing these five single models.

Because the description of each of the utilized ensemble members is widely available, we do not reiterate it here for optimal delivery of information. In addition, the focus of the current study is to demonstrate the application of ensemble learning approaches for liquefaction prediction. However, it is important to note that an optimal configuration of the single Artificial Neural Networks and Extreme Learning Machines should be decided before the ensemble model is created (Bolton Seed et al., 1985; J. Cao, Zhang, Luo, Yin, & Lai, 2016; Gelbart, Snoek, & Adams, 2014; Goh, 1994).

3.3 Evaluation Criteria

For each ensemble model configuration, the training and testing studies are carried out simultaneously. Figure (1) shows the flow diagram explaining the ensemble model creation, training and testing phases. In order to assess the classification performance of the presented ensembles, Cohen's Kappa coefficient ($\boldsymbol{\kappa}$) measure is used. In a binary classification setting, the $\boldsymbol{\kappa}$ statistic is computed as follows:

$$\kappa = \frac{\frac{TP+TN}{TP+TN+FP+FN} - \frac{(TP+FN)\times(TP+FP)+(TN+FN)\times(TN+FP)}{(TP+TN+FP+FN)^2}}{1-\frac{(TP+FN)\times(TP+FP)+(TN+FN)\times(TN+FP)}{(TP+TN+FP+FN)^2}} \quad (4)$$

which can be reduced to:

$$\kappa = \frac{Accuracy - RA}{1 - RA} \quad (5)$$

where *TP* and *TN* are the true positive and true negative values, respectively. In other words, the value of *TP* + *TN* is the number of correctly classified observations. *FP* and *FN* are the false positive and false negative values, respectively, making *FP* + *FN* equivalent to the number of incorrectly classified observations, and *TP* + *TN* + *FP* + *FN* as the total number of observations.

The *Accuracy* of a model is the ratio of correct predictions to the total number of predictions made by the model. *RA* is the probability of having a correct prediction due to chance, given mutually exclusive and independent events. It can be shown from Equation (12) that when the Accuracy is maximum, RA does not affect the *Kappa* statistic. This is usually apparent in the training performance of any ensemble model, as the training error can achieve a zero value. In all the other cases, RA value will reduce the accuracy and, hence, the *Kappa* statistic will be always less than or equal to *Accuracy*.

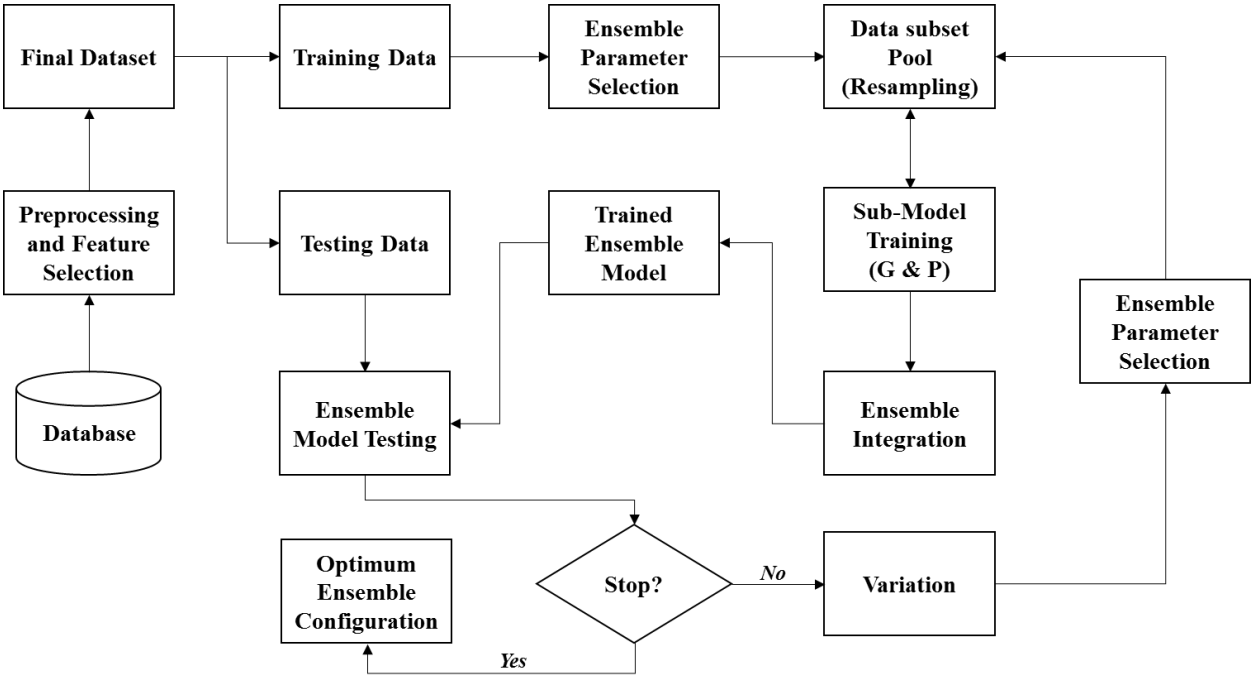

Figure 1: Proposed simulation approach for the case study.

4 CONCLUSION

In this work, a comprehensive study on the application of ensemble learning for seismic induced liquefaction prediction is presented. Three ensemble learning frameworks are utilized with five unique machine learning

models, summing to eight investigated ensemble models. The application of ensemble learning is discussed for the field of interest, and a step-by-step analysis is shown. The added benefit of ensemble learning is demonstrated throughout the various targeted testing schemes. A Monte Carlo simulation is carried out for each ensemble model in order to further investigate the improved generalization ability of ensemble learning. This work is aimed at motivating development of state-of-the-art ensemble approaches for regional liquefaction prediction. More research is required to adequately address ensemble learning in this field.

Liquefaction records are of regional and diverse. Hence, two future studies are recommended; the explicit utilization of diversity concept in developing ensemble learning models is expected to provide further improvement to the prediction. Such approach is expected to capture the patterns usually hidden or weaker than the ones on which hyperplanes are simulated by the previous models. Lastly, the binary classification nature of this problem allows for investigating models that can modify the magnitude of the explanatory variables using a novel learning strategy which links the diverse liquefaction records to the classification magnitude. This approach is able to transform the classification problem to a regression problem, which is favoured for its more flexible objectives used for deriving the best fit on the available training data. In other words, instead of finding the optimal separation plane for the binary classification problem, a magnitude of liquefaction index can be constructed with information on the modification required for the liquefaction records before including them into the dataset.

REFERENCES

Alobaidi, M. H., Chebana, F., & Meguid, M. A. (2018). Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Applied Energy, 212*, 997-1012.

Ardakani, A., & Kohestani, V. (2015). Evaluation of liquefaction potential based on CPT results using C4. 5 decision tree. *Journal of AI and Data Mining, 3*(1), 85-92.

Baez, J. I., Martin, G. R., & Youd, T. L. (2000). Comparison of SPT-CPT liquefaction evaluations and CPT interpretations *Innovations and Applications in Geotechnical Site Characterization* (pp. 17-32).

Benkeser, D., Ju, C., Lendle, S., & van der Laan, M. (2018). Online cross-validation-based ensemble learning. *Statistics in medicine, 37*(2), 249-260.

Bolton Seed, H., Tokimatsu, K., Harder, L., & Chung, R. M. (1985). Influence of SPT procedures in soil liquefaction resistance evaluations. *Journal of Geotechnical Engineering, 111*(12), 1425-1445.

Breiman, L. (1996a). Bagging predictors. *Machine learning, 24*(2), 123-140.

Breiman, L. (1996b). Stacked regressions. *Machine learning, 24*(1), 49-64.

Brown, G. (2011). Ensemble learning *Encyclopedia of Machine Learning* (pp. 312-320): Springer.

Cao, J., Zhang, K., Luo, M., Yin, C., & Lai, X. (2016). Extreme learning machine and adaptive sparse representation for image classification. *Neural networks, 81*, 91-102.

Cao, M., Pan, L., Gao, Y., Novák, D., Ding, Z., Lehký, D., & Li, X. (2017). Neural network ensemble-based parameter sensitivity analysis in civil engineering systems. *Neural Computing and Applications, 28*(7), 1583-1590.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans* (Vol. 38): Siam.

Erdal, H. I., & Karakurt, O. (2013). Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. *Journal of Hydrology, 477*, 119-128.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(4), 463-484.

Gelbart, M. A., Snoek, J., & Adams, R. P. (2014). Bayesian optimization with unknown constraints. *arXiv preprint arXiv:1403.5607*.

Glaser, S. (1996). Insight into liquefaction by system identification. *Geotechnique, 46*(4), 641-656.

Goh, A. T. (1994). Seismic liquefaction potential assessed by neural networks. *Journal of Geotechnical Engineering, 120*(9), 1467-1480.

Goh, A. T. (1996). Neural-network modeling of CPT seismic liquefaction data. *Journal of Geotechnical Engineering, 122*(1), 70-73.

Hoang, N.-D., & Bui, D. T. (2018). Predicting earthquake-induced soil liquefaction based on a hybridization of kernel Fisher discriminant analysis and a least squares support vector machine: a multi-dataset study. *Bulletin of Engineering Geology and the Environment, 77*(1), 191-204.

Juang, C., Chen, C., Tang, W., & Rosowsky, D. (2000). CPT-based liquefaction analysis, Part 1: Determination of limit state function. *Geotechnique, 50*(5), 583-592.

Kayen, R., Moss, R., Thompson, E., Seed, R., Cetin, K., Kiureghian, A. D., . . . Tokimatsu, K. (2013). Shear-wave velocity–based probabilistic and deterministic assessment of seismic soil liquefaction potential. *Journal of Geotechnical and Geoenvironmental Engineering, 139*(3), 407-419.

Krawczyk, B., Minku, L. L., Gama, J., Stefanowski, J., & Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. *Information Fusion, 37*, 132-156.

Luo, H., Wang, D., Yue, C., Liu, Y., & Guo, H. (2018). Research and application of a novel hybrid decomposition-ensemble learning paradigm with error correction for daily PM10 forecasting. *Atmospheric Research, 201*, 34-45.

Mitchell, T. M. (1997). Machine Learning (McGraw-Hill International Editions Computer Science Series).

Mola-Abasi, H., Kordtabar, B., & Kordnaeij, A. (2018). Liquefaction prediction using CPT data by triangular chart identification. *International Journal of Geotechnical Engineering, 12*(4), 377-382.

Polikar, R. (2012). Ensemble learning *Ensemble machine learning* (pp. 1-34): Springer.

Robertson, P. (2015). Comparing CPT and V s liquefaction triggering methods. *Journal of Geotechnical and Geoenvironmental Engineering, 141*(9), 04015037.

Samui, P., Jagan, J., & Hariharan, R. (2016). An alternative method for determination of liquefaction susceptibility of soil. *Geotechnical and Geological Engineering, 34*(2), 735-738.

Tokimatsu, K., & Yoshimi, Y. (1983). Empirical correlation of soil liquefaction based on SPT N-value and fines content. *Soils and Foundations, 23*(4), 56-74.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks, 5*(2), 241-259.

Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: methods and applications*: Springer.

Zhang, Z.-L., Luo, X.-G., Yu, Y., Yuan, B.-W., & Tang, J.-F. (2018). Integration of an improved dynamic ensemble selection approach to enhance one-vs-one scheme. *Engineering Applications of Artificial Intelligence, 74*, 43-53.