THE APPLICATION OF ENSEMBLE CLASSIFICATION TECHNIQUES IN SOIL CLASSIFICATION SYSTEM ON THE BASIS OF DMT AND CPT DATA

Jarosław Kurek[1], Michał Kruk[2], Piotr Bilski[3], Oguz Akpolat[4], Simon Rabarijoely[5] and Grzegorz Wieczorek[6]

[1]Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences – SGGW, ul. Nowoursynowska 159, 02-767, Warsaw, Poland
jaroslaw_kurek@sggw.pl

[2]Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences – SGGW, ul. Nowoursynowska 159, 02-767, Warsaw, Poland
michal_kruk@sggw.pl

[3]Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences – SGGW, ul. Nowoursynowska 159, 02-767, Warsaw, Poland
piotr_bilski@sggw.pl

[4]Chemistry Department, Faculty of Science, Mugla University - Kotekli Campus, 48000 Mugla, Turkey
oakpolat@gmail.com

[5]Faculty of Engineering and Environmental Sciences, Warsaw University of Life Sciences – SGGW, ul. Nowoursynowska 159, 02-767, Warsaw, Poland
simon_rabarijoely@sggw.pl

[6]Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences – SGGW, ul. Nowoursynowska 159, 02-767, Warsaw, Poland
grzegorz_wieczorek@sggw.pl

ABSTRACT

The paper presents the application of some of novel ensemble classification techniques to classify data derived from soil probes. The ensemble learning methods can be applied as an effective classification technique for any common issue. In an ensemble classification system many base classifiers are merged to obtain a classifier with higher performance. The authors take up issue to apply common used and effective ensemble techniques to classification of soil data to one of the soil profile layers. So the goal is to propose some stable method classification based on which we can create soil profile in chosen place. Then we compare the soil profile created automatically from these in-situ tests. It will help geotechnical experts to create such soil profile automatically. Proposed ensemble classification methods will be compared to other applied methods such us SVM, KNN. The results of research will be discussed at the end of article.

***Keywords**: soil profiles, soil classification, ensemble techniques classification, SVM, KNN*

1. Introduction

One of the important issues of the construction such us: buildings, houses, tunnels, bridges, etc. is to validate of soil where the construction will be planning. To find out what is the soil profile in place when the construction will take place, usually geotechnical probes are well-know used. The analysis of soil layers based on these probes such us: cone penetration tests (CPT) and dilatometer tests (DMT's) are cheaper and faster than traditional drilling boreholes. But to recognize correctly soil layers the huge knowledge base is required from geotechnical engineering. But the limitation of mentioned approach is that can be used only in nearby area. For example where you learn system to classify data probes derived from Poland to one of the soil layer, that system probably cannot be applied in any other countries caused different geology structure. Hence the usage of these probes can be only locally e.g. in area of one town where we know that geological structure is comparable. Of course traditional approach of soil profile layers creating is still acceptable and even mandatory when we have obtain very precision result and we do not put attention on time and money. It means that regular, traditional approach usually takes too much time to collect data and then analysis of data in special laboratory by geotechnical engineers. The further aspects are: quite expensive procedure and the last one: too invasive (too many drilled borehole) – to many boreholes can have impact of stability of the planned foundation and the ground.

To avoid of mentioned disadvantages, computer algorithms can be applied to automatically create soil layers profile. But application of this learned system are limited to local area only e.g. one town or bigger area where we know that geological structure can be comparable. We can assess accuracy of learned system comparing the results to the most reliable method such us traditional drilling of boreholes.

Computer algorithm is able to create soil profile layers and classifies the layers to appropriate soil type. If we create that algorithm the output can be analyzed by geotechnical specialists to verify accuracy of generated profile. Created in this way database can be used as the soil identification module for further geotechnical system. Similar works were done before (Hashash et al. (2004), Shahin et al. (2005)), but system should be based on reliable classifiers such as ensemble classification techniques.

The measurement data have been acquired from Warsaw University of Life Sciences campus (WUoLS) during expansion of the university. Before the university obtain building permit to start build the new campus building objects, the multiples tests have been performed. Geotechnical specialists collect measurement data from CPT and DMT probes and also using traditional approaches such as drilling boreholes. The latter let us to treat as the reference method to compare with new the novel method presented in this paper and calculate the accuracy ratio.

The research presented here is a continuation of the experiments published before (Rabarijoely et al. 2007), Rabarijoely and Bilski 2009), Kruk et al. 2014) and Kurek et. al. 2014).

2. Database

The database consists of trials derived from CPT and DMT probes. The cone penetration test (CPT) is a standard and well established method widely used to recognize and analyze geotechnical conditions (Lunne et al. 1997, Młynarek 2007, Huang A & Mayne 2008). The probe is presented in Fig. 1. It is inserted into the ground with the constant speed of 2cm/s. During that process the measurement data of four parameters are obtained: depth (d), the resistance of the cone (qc), sleeve friction resistance (fs) and friction coefficient (Rf). We used all values in the presented experiment.

Figure 1 The cone penetration test probe

The latter probe is flat dilatometer test (DMT) which was developed in Italy in 1980, is currently used in dozens countries both for research and practical applications. The sample of DMT probe is depicted on figure 2.

The dilatometer is build of a steel blade having a thin, expandable, circular steel membrane mounted on one face. When at rest, the membrane is flush with the surrounding flat surface of the blade. The latter is connected, by an electro-pneumatic tube running through the insertion rods, to a control unit on the surface (Fig. 2 -left). The control unit is equipped with pressure gauges, an audio-visual signal, a valve for regulating gas flow (provided by a tank) and vent valves. The blade (Fig. 2b - right) is advanced into the ground using common field equipment, i.e. push rigs normally used for the cone penetration test (CPT) or drill rigs. Pushing the blade with a 20 ton penetrometer truck is the most effective (up to 100 m of profile per day). The test starts by inserting the dilatometer into the ground. Soon after the penetration, the operator inflates the membrane and takes, in about 1 min, two readings: the A pressure, required to just begin to, move the membrane ("lift-off"), and the B pressure, required to move the centre of the membrane 1.1 mm against the soil. A third reading C ("closing pressure") can also optionally be taken by slowly deflating the membrane soon after B is reached. The blade is then advanced into the ground of one depth increment (typically 20 cm) (Totani et al. 2001).

Figure 2 The DMT control unit (left) and the probe (right)

The results of soil profile manually obtained from traditional drilled boreholes approach which performed at Warsaw University of Life Sciences campus, are reference method to our automatically methods. The profiles and probing places are presented in Fig. 3. In this example on the basis of borehole drilling (OW-8, OW-11, OW-6), CPT probing (CPT-4) and DMT probing (DMT-2) the soil profiles were generated by human expert manually. Our main task in this paper is to describe the automatic method which could do this automatically on the basis CPT and DMT probing only (non invasive).

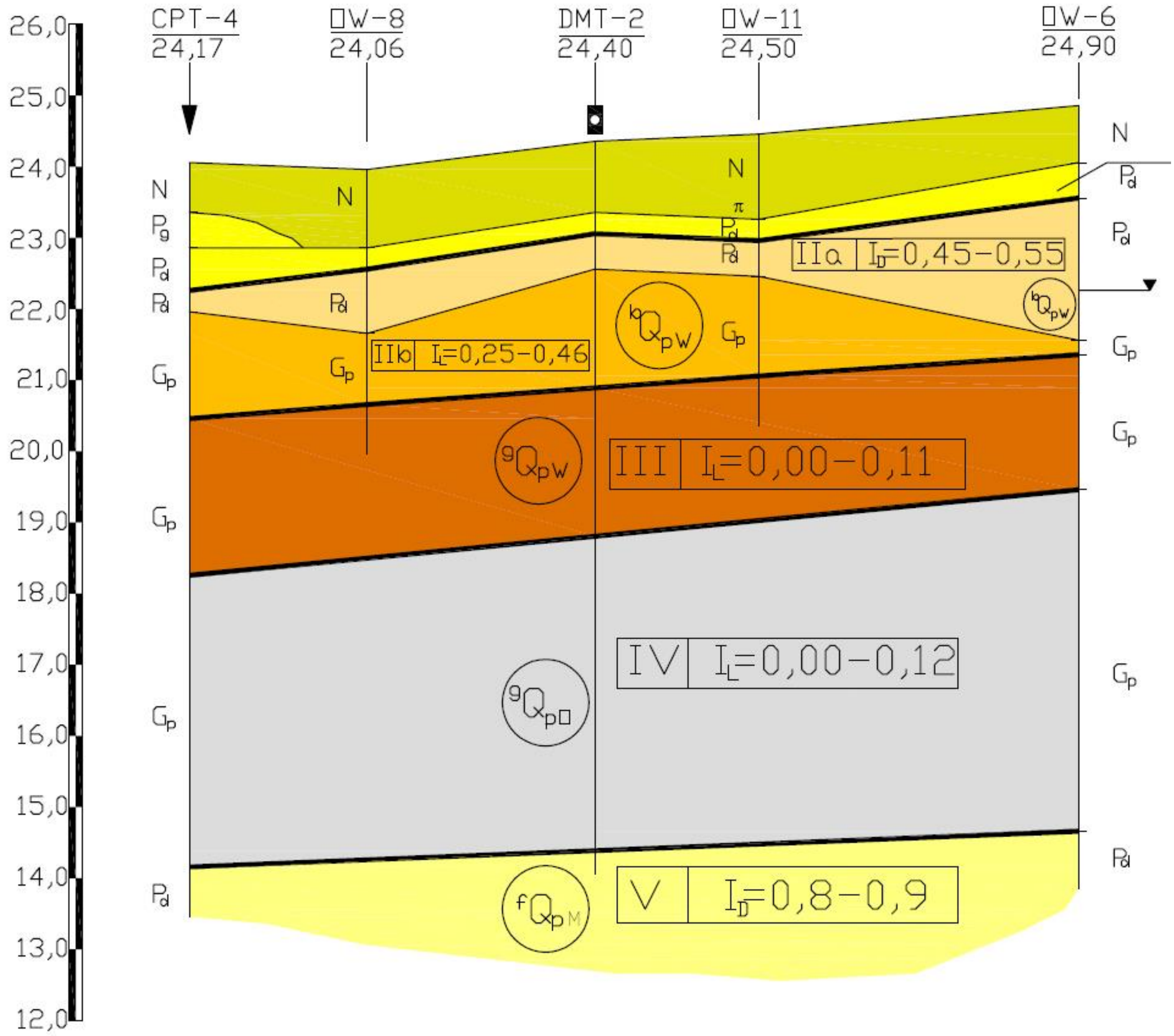

Figure 3 A standard geotechnical cross-section: OW – borehole test (reference method), CPT – cone penetration test, DMT – dilatometer test; (N – fill, Gp – sandy clay, Pd –fine sand, wn – moisture content, ID – relative density, IL –liquidity index)

Database has been generated based on trials obtained from 19 measurements by means of CPT and DMT probes. Detailed description how to the trials have been collected can be found in previous article (Rabarijoely et al. 2007 , Rabarijoely and Bilski 2009, Kurek et al. 2014, Kruk et. al. 2014).

The database consists of 625 trials and 7 features. Hence we can have matrix in size of 625x7. Then we have added additional column described soil layer class treated as output of learning system.

Detailed description of features (input) and output learning system is listed in table 1.

Table 1 Detailed description of database structure.

Column name	Type	Description
Depth	Input	the depth of measurement
□□[kN/m^3]	Input	volumetric weight of soil
□$_{vo}$ [MPa]	Input	total vertical stress
u [Mpa]	Input	pore water pressure
A pressure	Input	the pressure, required to just begin to, move the membrane ("lift-off")
B pressure	Input	the pressure, required to move the centre of the membrane 1.1 mm against the soil
C pressure	Input	("closing pressure") can be taken by slowly deflating the membrane soon after B is reached
soil layer class	Output	soil class was obtained by the drilling boreholes which is our referenced method

Due to small number of features we did not take up the stage of features selection for learning system.

3. Ensemble classification techniques

3.1. Bootstrap aggregating (bagging) approach

Bootstrap aggregating well-known as bagging is the approach when the whole classifier consists of many individual classifiers which can take up decision individually. All together create one ensemble vote with equal weight. Every individual classifier trains in the ensemble by means of randomly drawn subset of the whole training set. As an example, the random forest algorithm combines random decision trees with bagging to achieve very high classification accuracy.

3.2. Boosting approach

Boosting approach is a little different to Bootstrap aggregating. Common feature is that create ensemble classification but in other way. Boosting approach is based on incrementally build an ensemble classifier by training each new model instance to emphasize the training instances that previous models misclassified. In some cases, boosting has been shown to yield better accuracy than bagging, but it also tends to be more likely to over-fit the training data. One of the most common implementation of boosting approach having very high classification accuracy is Adaboost.

3.2. Simulation of ensemble classification technique

All simulation we performed by means of Matlab 2015a which has native functionality to run ensemble algorithms. The structure of Matlab approach to ensemble classification is depicted in Figure 4. The whole set of data set has to be partitioning into a held-in data set and held-out data set. Then train the models on the held-in data and then choosing whichever of those trained models performs best on the held- out data. This is the cross-validation technique.

For all ensemble approach in Matlab we must provide the following parts:

- put predictor data in a matrix
- prepare the response data
- choose an applicable ensemble method
- set the number of ensemble members
- prepare the weak learners
- call fitensemble function

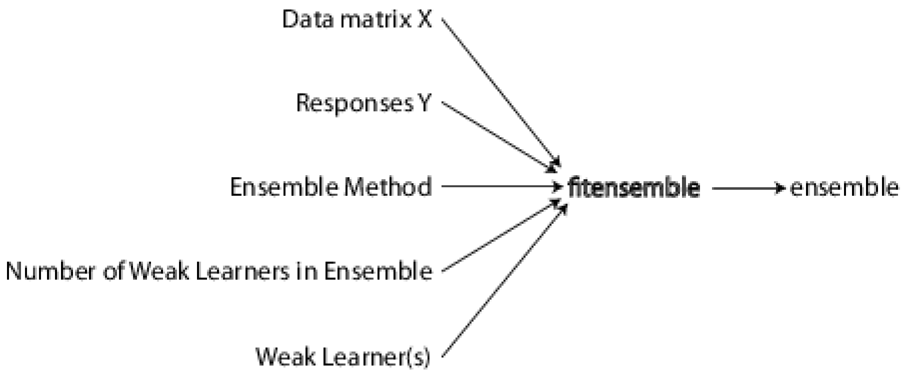

Figure 4 Ensemble classification approach in Matlab

In our approach we have tested the following algorithms:

- AdaBoostM2
- Bag

5. The results

During the numerical experiments we have assumed 100 decision trees as weak learner. The numerical experiments have been repeated 10 times and the result is depicted in figure 5.

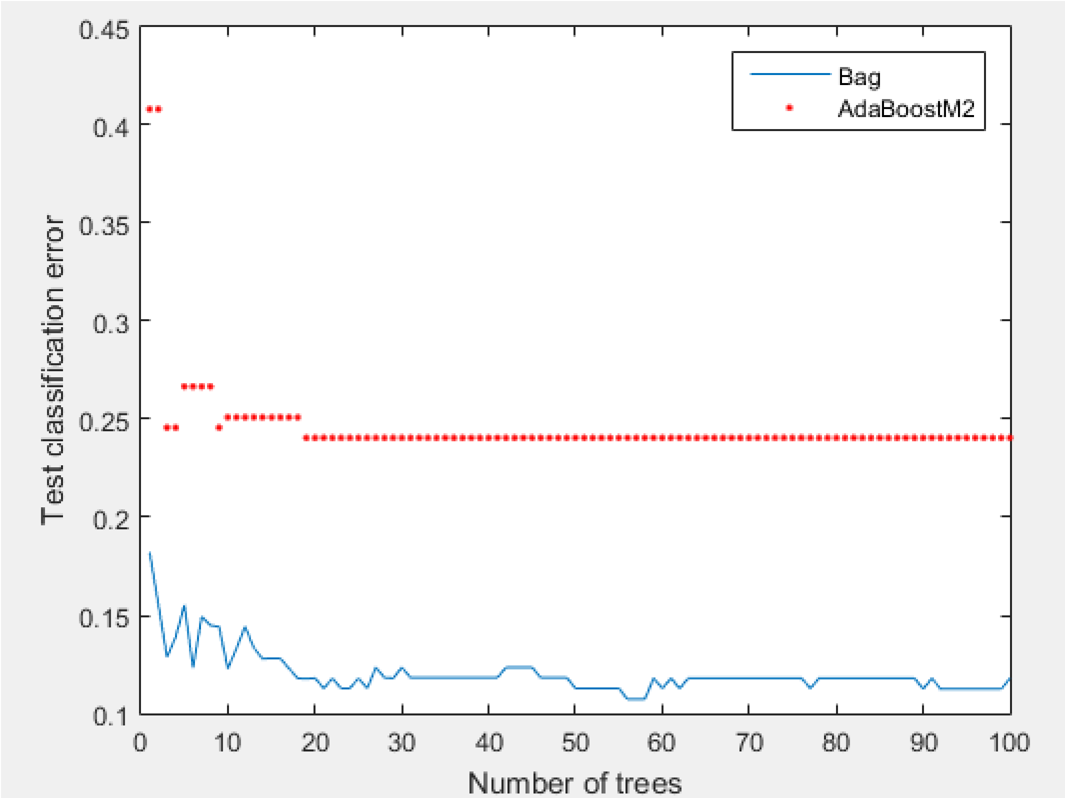

Figure 5 Comparison of results two ensemble classification techniques (Bagging and Bossting).

The best results were obtained by bagging algorithm – 89.20% accuracy in comparison to boosting approach -75.95% accuracy. We have compared to SVM technique where result was better (93.6% accuracy) than above ensemble classification techniques but is less stable than ensemble approaches. Hence for production environment is suggested to use ensemble classification techniques.

6. Conclusion

The paper has presented the research directed to the automatic recognition of soils on the basis of CPT and DMT data gethered from university campus. The proposed approach presents comparison of two different ensemble classification approaches boosting and bootstrap aggregating. Both are stable more that typically SVM approaches when based on parameters can be over-fitted to data set. These results confirm that an automatic learning system based on ensemble classification techniques can reach the efficiency comparable to the geotechnical human expert results.

Acknowledgements

This work is supported by the Polish National Centre of Science, Grant No. 2011/03/D/ST8/04309.

References

Darwish Dina,(2013) Assessment of Offline Digital Signature Recognition Classification Techniques, International Journal of Computer Networks and Communications Security, VOL. 1, NO. 4, SEPTEMBER 2013, 143–151

Forgy E. W. (1965). Cluster analysis of multivariate data: Effciency vs. interpretability of classification. Biometrics, 21:768-769

Guyon, I. & Elisseeff A. (2003), An introduction to variable and feature selection, Journal of Machine Learning Research, 2003, vol. 3, pp. 1158-1182.

Hall M. (2000), Correlation-based feature selection for discrete and numeric class machine learning. Proceedings of the Seventeenth International Conference on Machine Learning Morgan Kaufmann Publishers, San Francisco.

Hashash, Y.M.A., et al. (2004). Numerical implementation of a neural network based material model in finite element analysis, International Journal for Numerical Methods in Engineering, 59, 989-1005.

Haykin S. 1999, Neural Networks, Comprehensive Foundation, Prentice-Hall, Englewood Cliffs, NJ.

Huang A., Mayne P. W. (2008). Geotechnical and Geophysical Site Characterization. Proc. of the 3 rd inter. Conf. on Site characterization, Taipei, Taiwan. Published by: Taylor & Francis Group, London, UK

Kruk, M. et al. (2014), Automated soil profile generation methods on the basis of DMT and CPT data, Proceeding of the International Conference on Artificial Intelligence and Computer Science (AICS 2014), 15 - 16 September 2014, Bandung, INDONESIA

Kurek J. et al. (2014), Automatic estimation of the number of soil profile layers using bayesian information criterion, Proceeding of the International Conference on Artificial Intelligence and Computer Science (AICS 2014), 15 - 16 September 2014, Bandung, INDONESIA

Liu H & Yu L (2003), Feature selection for high-dimensional data: A fast correlation-based based filter solution. Proceedings of The Twentieth International Conference on Machine Leaning (ICML-03), Washington, D.C.

Lunne, T., Robertson, P.K., Powell, J.M. (1997). Cone penetration testing in geotechnical practice. Blackie Academic and Professional, London, England

Marchetti S. (1980). In Situ Tests by Flat Dilatometer. J. Geotech. Eng. Div., ASCE, 106, GT3, 299-321.

Matlab (2014) user manual , MathWorks, Natick.

Młynarek Z. (2007). Site investigation and mapping in urban area. Proc. of the 14th European Conference on Soil Mechanics and Geotechnical Engineering. Madrid, Vol. 1, 175-202.

Rabarijoely S. et al. (2007). The usage of the graph clustering algorithm to the recognition of geotechnical layers. Annals of Warsaw University of Life Sciences – SGGW. Ann. Warsaw Univ. of Life Sciences – SGGW, Land Reclam., No 38, 2007, 57 - 68.

Rabarijoely S. & Bilski P. (2009). Automated soil categorization using CPT and DMT investigations, 2nd International Conference on New Developments In Soil Mechanics and Geotechnical Engineering, 28-30 May 2009, Near East University, Nicosia, North Cyprus

Scholkopf B.& Smola A. (2002). LearningwithKernels, MITPress, Cambridge, MA.

Shahin, et al. (2005). Neural network based stochastic design charts for settlement prediction, Can. Geotech. Jour. (42), 110-120.

Tan P.N et al. (2006) Introduction to data mining, Pearson Education Inc.,Boston.

Vapnik V. (1998), Statistical Learning Theory, Wiley, New York.