Proceedings of the 1st *Annual International Geotechnical Innovation Conference (IGIC 2024),* Jeddah Hilton, Kingdom of Saudi Arabia, May 06–07, 2024. Published and Indexed by Gulf Scholar.

Advancing Data-Driven Site Characterization Techniques in the Age of Artificial Intelligence

Prof. Robert Y. Liang
rliang1@udayton.edu
Department of Civil and Environmental Engineering and Engineering Mechanics, University of Dayton, Ohio USA

ARTICLE INFO

Published on 24th of October 2024.
Doi:10.54878/nwd1wv13

KEYWORDS:

Data-Driven, Site Characterization, Artificial Intelligence, Stratification Modeling, Uncertainty Quantification, Reliability Assessment

HOW TO CITE:

Advancing Data - Driven Site Characterization Techniques in the Age of Artificial Intelligence. (2024). *1st International Geotechnical Innovation Conference*, *1*(1)

ABSTRACT

Data-driven site characterization is one of the cornerstones of geotechnical engineering design and construction. Current practice of interpreting site investigation data to establish stratification models is inadequate in handling the complexities and uncertainties. As a result, the subsequent geotechnical reliability analysis cannot explicitly consider the impact of these uncertainties. This presentation introduces a novel approach for generating a subsurface stratification model with the ability to quantify model uncertainty using structured site data, stochastic modeling, and Bayesian machine learning. Examples of applications are presented.

1. INTRODUCTION

Data-driven site characterization (DDSC) is one of the cornerstones of geotechnical engineering design and construction. As elucidated in Phoon and Zhang (2023), there are at least three fundamental challenges in DDSC: (1) ugly data, (2) site recognition, and (3) stratification. With the advancement in artificial intelligence, geotechnical engineers have developed AI based algorithms for a variety of applications. However, for DDSC, existing site-specific data sets are too small for any machine learning methods to produce useful insights for decision support. Geotechnical site-specificity limits the applicability of combining several different sites to increase sample size. Therefore, the author and his colleagues applied Bayesian machine learning and stochastic simulations to address the limited and complex nature of real-world data and at the same time quantify uncertainty in the inference process. On the data reporting side, there is a concurrent development and adoption of DIGGS (Data Interchange for Geotechnical and Geo-environmental Specialists) for standardizing the grammar and base vocabulary in an Extensible Markup Language (XML) file for encoding geo-referenced geotechnical and geo-environmental data in a format that is both human-readable and machine -readable (Wang et al. 2021). The framework of using AI to advance DDSC techniques developed by the authors' group are described.

2. METHODOLOGIES AND APPLICATION EXAMPLES

Structured data (e.g., DIGGS compliant borehole log, CPT), stochastic modeling, and Bayesian machine learning are the main ingredients for pattern recognition in geotechnical stratigraphic modeling with the ability to quantify the inference-related uncertainty. The concept is that we work on two spaces: the physical space and features space. In physical space, for example for 1D CPT interpretation, the model contains two layers: a hidden layer including soil types and an observed layer containing the measured soil properties. The spatial constraint is defined by a local neighborhood system. Elements in the observed layer are projected into the feature space. The statistical correlation is represented by clusters of similar points which can be modeled by Gaussian mixture. The center of each cluster is controlled by the feature average and shape is controlled by the covariance matrix. In this way, we can extract spatial and statistical patterns from the data. The underlying philosophy is that the soil stratification can be considered as 1D vertical spatial pattern of categorical data. The spatial pattern is encoded into a neighborhood system using potential function named Potts model. If two neighborhood elements have different soil types, the energy will increase by a parameter called beta. Since the entire system prefers lower energy, nearby pixels tend to have the same soil type, which is consistent with the nature's soil forming process. Then we can compute the local energy and calculate the local conditional probability. Based on that, we can perform Bayesian learning and stochastic simulation. Since the learning algorithm is fully unsupervised, as a user, we just need to input the measured soil mechanical properties, for example, tip resistance and friction ratio for CPT sounding results, then the learning algorithm will automatically report the spatial pattern, stratification pattern of different soil layers, as well as the quantified uncertainty indicated by an index called information entropy.

For 2D space, we introduce the anisotropic Potts model; the concept of which is that the soil profiles can be considered as 2D spatial pattern of categorical data. To reflect the anisotropic effect, the energy increment is different in corresponding directions. By integrating the anisotropic Potts model and Gaussian mixture model, we can extract the spatial pattern and the associated statistical patterns from the observed data. For simulation purpose, the 2D space is discretized into a lattice and parallel stochastic simulation on the entire field is performed using the in-house developed chromatic sampler techniques. Open-source computational packages based on the developed algorithms are made available in GitHub:

PyCPT: https://github.com/hwang051785/pyCPT

BaySeg: https://github.com/cgre-aachen/bayseg

Applications of the methodologies for various real-world examples are presented (Wang et al. 2017, Wang et al. 2018, Wang, et al. 2019, Wang et al. 2020). The reliability analysis for two example cases involving shield tunneling soil-structure interaction problem and slope stability evaluation using the uncertainty quantified stratigraphic models are presented (Wang, et. al. 2016, 2017).

To conclude, structured data, stochastic models, and Bayesian machine learning are adopted as the cornerstones of uncertainty aware pattern recognition techniques in subsurface stratigraphic modeling.

REFERENCES

1. Phoon, K. K. and Zhang, W. (2023) Future of machine learning in geotechnics, Georisk, 17:1, 7-22, DOI: 10.1080/17499518.2022.2087884

2. Wang, H., Wellmann, J. F., Li, Z., Wang, X., & Liang, R. Y. (2017). A segmentation approach for stochastic geological modeling using hidden Markov random fields. Mathematical Geosciences, 49(2), 145-177.

3. Wang, H., Wang, X., Wellmann, F. and Liang, R.

Y. (2019). A Bayesian unsupervised learning approach for identifying soil stratification using cone penetration data. Canadian Geotechnical Journal, 56(8), 1184-1205

4. Wang, H., Wang, X., Liang, R. Y. (2020) Study of AI based methods for characterization of geotechnical site investigation data" Report No. FHWA/OH-2020-3, Ohio Department of Transportation

5. Wang, H., Wang, X., Liang, R. Y., Merklin, C, and Taliaferro, S. (2021), When AI meets DIGGS, The birth of a new site characterization paradigm? Geostrata, January/February, 39-45

6. Wang, X., Li, Z., Wang, H., Rong, Q., and Liang, R. Y. (2016) Probability analysis of shield-driven Tunnel in Multiple Strata Considering Stratigraphic Uncertainty, Journal of Structural Safety, Vol. 62, 88-100, http://dc.doi.org/10.1016/j.strusafe.2016.06.007

7. Wang, X., Wang, H., and Liang, R. Y. (2017) A method for slope stability analysis considering subsurface stratigraphic uncertainty, Journal of Landslides, DOI: 10.1007/s10346-017-0925-5

8. Wang, X., Wang, H., Liang, R. Y., Zhu, H, Di, H. (2018). "A hidden Markov random field model-based approach for probabilistic site characterization using multiple cone penetration test data" Structural Safety, 70, 128-138