ISFOG 2020 Data science prediction event

1. Introduction

Data science techniques are rapidly transforming businesses in a broad range of sectors. While marketing and social applications have received most attention to date, geotechnical engineering can also benefit from data science tools that are now readily available.

In the context of the ISFOG2020 conference in Austin, TX, a prediction event is launched which invites geotechnical engineers to share knowledge and gain hands-on experience with machine learning models.

This document outlines the scope of the ISFOG2020 data science prediction event which formulates pile driving predictions as a machine learning problem.

2. Problem Statement

2.1 Prediction problem

The problem proposed for the prediction event is the prediction of blowcount vs depth for jacket piles in North Sea soil conditions. In current engineering practice, blowcounts are derived using a three-step approach:

- Predict the soil resistance vs depth from geotechnical data (Alm and Hamre, 2001; Stevens et al., 1982);
- Calculate a bearing graph from 1D wave equation analysis for selected depths (Smith, 1960);
- Interpolate the bearing graphs to derive blowcounts.

This is a fairly complicated approach to relate soil properties, pile dimensions and hammer settings to the target variable, being the blowcount.

This prediction event will explore whether machine learning models can be trained on a set of pile driving records to provide reliable estimates of blowcount for unseen locations.

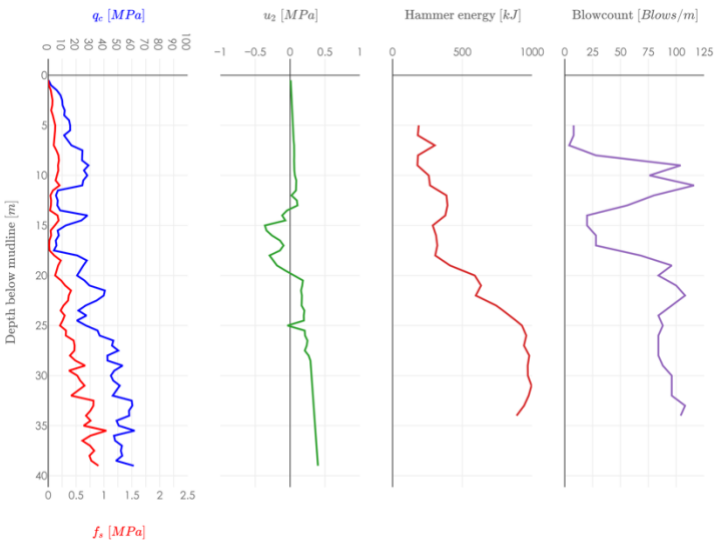
2.2 Data set

The available dataset consists of installation records for 114 piles in the North Sea and contains the following data sources:

- CPT profiles with cone tip resistance, sleeve friction and pore pressure measurements;
- Pile outer diameter and wall thickness;
- Blowcount and hammer energy.

To facilitate participation and keep a focus on the data science techniques, he data has already been processed (Figure 1) and mapped to a regular grid with nodes every 0.5m.

Figure 1: Results of data processing and grid mapping routine for one location



For the prediction event, the data set is partitioned into a training dataset of 94 piles and a validation data set of 20 piles. Only the training data will be provided, the validation set will be used to assess the quality of participant's models.

The data science prediction event is hosted on Kaggle, (https://www.kaggle.com/c/isfog2020-pile-driving-predictions/) where the data files can be downloaded.

3. GUIDELINES FOR PARTICIPATION

3.1 Target audience

This prediction event targets individuals or groups with an interest in data science techniques. Although some knowledge of the techniques can be beneficial, we also invite people who have no prior experience with machine learning.

To facilitate participation, a tutorial which demonstrates the basics of getting started with data science and building a machine learning model for the pile driving problem is made available to participants through the Kaggle page (https://www.kaggle.com/brunostuyts/isfog-2020-linear-model-demo). The tutorial uses Python and the scikit-learn machine learning library.

Participants are encouraged to use a combination of data science techniques and engineering knowledge to come up with the best possible model.

3.2 Submitting contributions

Contributions can be submitted via the Kaggle competition page.

Participants are also requested to submit a summary of their reasoning for developing the model. A short explanation on which type of model was used and whether any engineering knowledge was used is required.

Although participants are not required to submit kernels containing their models, they are invited to do so to further share knowledge and insights.

Submissions will be ranked based on a RMSE metric. Note that only a public leaderboard which will be visible during the competition, which is only based on a fraction of the test dataset. The final scores based on the full test dataset will be revealed once the competition ends.

3.3 Summary and presentation of results

All predictions will be evaluated using the validation dataset of 20 piles excluded from the data provided. The lessons learnt from the prediction event and the best models will be discussed in an ISFOG2020 paper.

3.4 Timeframe

The prediction event runs for a period of 8 months, from April to December 2019. This time frame is long enough to allow teams to build up their models iteratively and learn from other submissions.

3.5 Communication and support platform

During the prediction event, there will be frequent communication to encourage contributors, provide links to machine learning tutorials, provide hints for improving models and discuss problems people might be facing.

A discussion can be started via the Kaggle platform.

4. REFERENCES

Alm, T., Hamre, L., 2001. Soil model for pile driveability predictions based on CPT interpretations. Presented at the International Conference On Soil Mechanics and Foundation Engineering.

Smith, E.A.L., 1960. Pile driving analysis by the wave equation. Soil mechnanics and foundations division - ASCE 24/01/1900.

Stevens, R.S., Wiltsie, E.A., Turton, T.H., 1982. Evaluating Pile Drivability for Hard Clay, Very Dense Sand, and Rock. Presented at the OTC 4205, 14th Annual Offshore Technology Conference, Houston, Texas, pp. 465–469.

5. ACKNOWLEDGEMENT

The dataset is kindly provided by Cathie Group (http://www.cathiegroup.com/).