Paper Number: 2630

Evaluation of automated lithology classification architectures using highly-sampled wireline logs for coal exploration

Horrocks, T., Holden, E.-J. and Wedge, D.

Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia; e-mail: tom.horrocks@research.uwa.edu.au

_____

Automatic lithology classification from wireline logs gives geologists objective downhole lithology interpretations which can be used for validating their own interpretations, or alternatively as an in-lieu interpretation where core inspection is not possible. Advances in machine learning have resulted in classification algorithms that utilise datasets with both (i) large volume and (ii) high dimensionality, which corresponds to well logs that are (i) highly sampled and (ii) have numerous wireline logs. The work presented here evaluates three machine learning algorithms (classifiers) for coal classification from such well logs, while considering the hyperparameter optimisation metric, committee architecture, and post-processing of the predicted intervals.

Seven well logs were used from the Juandah East project area located 60 km north-west of Wandoan (Queensland, Australia), which is well known for coal mineralisation and reported in the Queensland Digital Exploration database [1]. Each well log contains a set of nineteen common wireline logs that are highly sampled (1 cm^{-1}), and lithology logs which describe ten lithologies including coal. Depth intervals containing coal account for 4.7% of the total length covered by all holes, with the majority of the depth intervals comprising sandstone (69.6%) and siltstone (18.6%).

A Naïve Bayes classifier (NB), Support Vector Machine (SVM), and Artificial Neural Network (ANN) were selected from a pool of six machine learning algorithms based on the accuracy of pilot lithology classification methods. Each of the three machine learning algorithms were configured in two architectures: *singular* wherein one instance of the algorithm was trained on data from all well logs, and *committee* wherein a different instance of the algorithm was trained for each well log (with voting at classification-time). The configured classifiers were evaluated on each well log by training on the other six well logs (a form of cross-validation), with input taking the form of a vector of depth-aligned wireline log values. Hyperparameters for the ANN and SVM were selected by considering performance on coal identification (measured by the g-mean) to ensure accurate coal identification.

Figure 1 shows one set of lithology intervals predicted using an ANN committee. It was found that the committee architecture generally increases overall accuracy over the singular architecture by increasing both the accuracy and classification rate of the dominant lithology
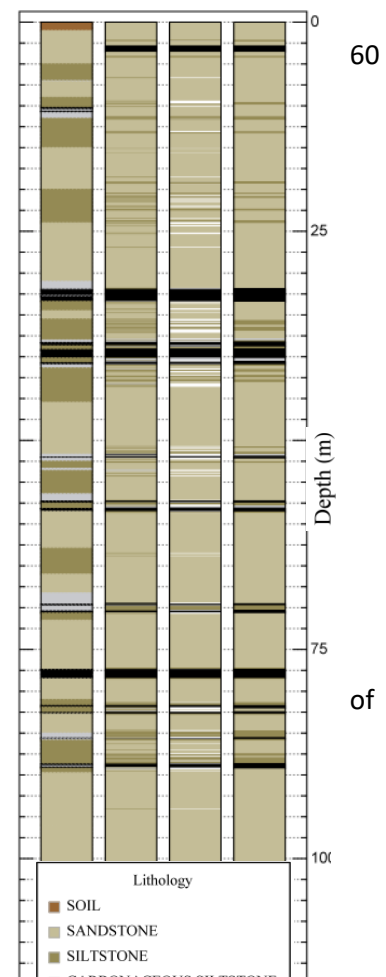

Fig. 1: Lithology intervals from (a) geologist, (b) wireline logs classified by a committee ANN, with thin intervals (c) removed and (d) merged.

(sandstone), with the ANN achieving 73.2% overall accuracy. Post-processing the predictions to merge thin intervals (<10 cm thick) decreased overall prediction error by 6.9%. Further results and explanations of the machine learning algorithms used are given in [2].

References:
[1] Green, D, Chestnutt, C, Mackie, S, (2006). EPC791 Juandah East, Queensland, QDEX 42776.
[2] Horrocks T, Holden E-J, Wedge D (2015) Comput Geosci 83:209-218