

SEG 2023 Conference: Resourcing the Green Transition

Data Capture from Geoscience Documents Using Transformers

Amit Juneja, [Henri Blondelle](#)

Agile Data Decisions, Inc., Houston, TX, USA

When it comes to prospect or field evaluation in mature geological areas, accessing the key information from thousands or millions of reports and logs may be time challenging and the risk to miss a key observation may be very high. Artificial intelligence (AI) and particularly machine learning may be beneficial in capturing data and sentiment on the contents of reports or logs at scale and detecting the sentences supporting the key information. This paper will discuss use of computer vision and NLP (natural language processing) for the following:

- Detection of lithological descriptions and assigning them to appropriate depths in logs;
- Classifying the geological description on composite logs according to several criteria;
- Detecting some key positive or negative information in geological reports.

Text modelling has been done using the BERT Algorithm (Bidirectional Encoder Representations from Transformers), and visual object detection has been done using convolutional neural networks (CNNs) . For our case study, the inferences obtained from a set of exploration reports in using a BERT model have shown very interesting results with a minimum of training. We emphasize the importance of “human-in-the-loop” in building successful AI models for capturing structured data from different modalities of documents—tables, images, sentences, and semi-structured layouts. We show how AI models continue to improve with feedback from geologists in these different modalities and how this feedback and subject matter expertise is critical in building useful AI models.