



Classificação litológica por *Machine Learning* utilizando dados de poços da Bacia de Campos

Ana Naujokat Tavares, Marcus Vinicius Aparecido Gomes de Lima, Ítalo Gomes Gonçalves, Felipe Guadagnin e Everton Frigo, Universidade Federal do Pampa

Copyright 2021, SBGf - Sociedade Brasileira de Geofísica.

This paper was prepared for presentation during the 17th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 August 2021.

Contents of this paper were reviewed by the Technical Committee of the 17th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

Abstract

A sondagem é um dos processos mais fundamentais e de maior custo nas atividades de exploração de óleo e gás natural. Geralmente os poços exploratórios de petróleo são perfilados com ferramentas geofísicas, mas uma pequena parcela dos poços é testemunhada, obtendo amostras de rocha de subsuperfície. Sendo assim, a determinação das litologias em subsuperfície é parcialmente realizada de forma indireta, pela análise de perfis geofísicos de poços (*well-logs*). A análise de *well-logs* evoluiu da interpretação em meio analógico para meio digital por intérprete humano. Com o advento e a popularização das técnicas de Inteligência Artificial, a interpretação passa a ser realizada por Aprendizagem de Máquina (*Machine Learning* - ML). A classificação de litologias a partir de *well-logs* utilizando ML é realizada pelo reconhecimento de padrões nos dados geofísicos com constante melhoria dos resultados. Neste trabalho, foi utilizado o método supervisionado para a classificação litológica a partir de dados geofísicos e geológicos de poços da Bacia de Campos, disponibilizados pelo Banco de Dados de Exploração e Produção da Agência Nacional do Petróleo (ANP). Antes de aplicar uma rotina computacional para treinar os algoritmos com base em dados de entrada e saída, foi realizada a compilação dos mesmos utilizando a linguagem de programação Python no *software* Anaconda. Nesta fase, a ferramenta *striplog* foi necessária para a leitura dos arquivos PDF para PNG, onde foram criadas individualmente as legendas dos poços devido a diferença entre elas. Dessa forma foi possível analisar estatisticamente a consistência dos dados, onde a ocorrência de dados incompletos tornou-se evidente nos poços, sendo crítico em um poço e resultando na exclusão do mesmo em etapas posteriores. Assim os dados de entrada, inseridos e rotulados pelo usuário, consistiram em oito poços e totalizaram 124.091 pontos oriundos da profundidade e das variáveis geofísicas de resistividade, radioatividade natural, velocidade sísmica, densidade e porosidade, com 26 categorias litológicas. Antes de treinar o modelo, os dados foram analisados do ponto de vista estatístico, estabelecendo a correlação entre as propriedades físicas e o tipo de rocha. Seis poços foram escolhidos para a implementação do método estatístico denominado de floresta aleatória (*Random Forest*). Na aplicação do modelo aos outros dois poços resultou em maior precisão para as litologias descritas como rochas carbonáticas a despeito das demais litologias, devido a fatores como dados incompletos, a falta de representatividade de algumas litologias e a distância entre os poços. A perspectiva é de que a exatidão da técnica melhore à medida em que se obtenha dados mais completos.