Regressão de dados Geotérmicos de FORGE Utah por Deep Learning (DL)

Marcus L. A. do Amaral (UFPA) e José J. S. de Figueiredo (INCT-GP e UFPA)

Copyright 2024, SBGf - Sociedade Brasileira de Geofísica.
Este texto foi preparado para a apresentação no X Simpósio Brasileiro de Geofísica, Salvador, 8 a 10 de outubro de 2024. Seu conteúdo foi revisado pelo Comitê Técnico do X SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Resumo

O trabalho aborda a aplicação de técnicas de predição de dados geotérmicos do projeto FORGE em Utah, nos EUA, por modelos de Deep Learning (DL) comparados com convencionais. Com intuito de atender a demanda da indústria, no que tange a transição energética, é necessário aprimorar o mapeamento do potencial geotérmico dessas áreas. Dito isso, dados de poços da região como profundidade, velocidade compressional (V_P), velocidade cisalhante (V_S), densidade (ρ_B), foram aplicadas no modelo para prever dados de temperatura (℃) em subsuperfície. Os modelos de DL utilizados foram, Multilayer Perceptron (MLP) e Rede Neural Convolucional (CNN - 1D), comparando-os com método convencional como Regressão Múltipla Linear (RML).

A priori, fizemos a seleção dos dados de poços, selecionando 3 poços verticais (58-32, 56-32, 78B-32) e 2 horizontais (16A(78)-32 e 16B(78)-32). Utilizamos bibliotecas de acesso livre, como TensorFlow e Scikit-Learn para facilitar a reprodução da metodologia. Para o pré-processamento, reunimos dados de arquivos distintos, retiramos dados faltantes, redimensionamos para terem a mesma quantidade de amostragem, transformamos suas grandezas para padrões nacionais. Acerca da análise de estatística, analisamos de forma quantitativa e qualitativa os dados, medindo sua média, desvios padrões, mínimos, máximos, além de calcular o coeficiente de correlação de Pearson (r) e fazer a análise de dispersão dos dados. Antes de iniciar o treino nos modelos, concatenamos os dados, dividimos em 70% treino e 30% teste, selecionamos um poço para validação, que não passe pelo treinamento do modelo, o ”poço cego” e normalizamos os dados para terem a mesma ordem de grandeza. As métricas utilizadas para avaliar a eficácia dos modelos foram, coeficiente de determinação (R^2), correlação de Pearson (r) e RMSE. Foram utilizados dados de treino, de teste e do ”poço cego” para avaliação do modelo. Para uma boa arquitetura das redes neurais, foram testados diversos hiperparâmetros e conjuntos de camadas tanto densas quanto convolucionais.

Os resultados foram bastante expressivos para as Redes Neurais, apesar de RML ter se mostrado eficaz, mesmo com padrões de dados não lineares. Correlação de Pearson, RMSE tiveram valores acima de 90% e RMSE valores próximos a zero, indicando boas previsões dos modelos. A análise qualitativa se deu por gráficos de correlação entre dados reais e preditos, a maioria dos dados indicam predições perfeitas. Com os resultados é possível notar que tanto os valores das métricas quanto as análises qualitativas dos modelos foram melhores nos dados do ”poço cego” o que indica generalização na predição dos modelos. Apesar dos bons valores das métricas, é possível notar ruídos nas predições dos dados de temperatura, provavelmente referentes as características dos dados de poço, o que nos faz considerar apenas a tendências dos valores preditos, como uma média móvel.

Em suma, nossa metodologia proporciona boas predições de dados de temperatura a partir de dados petrofísicos. O trabalho destaca a importância da generalização dos modelos e da validação dos resultados para garantir a precisão das previsões, representando um avanço significativo no estudo e exploração de recursos geotérmicos.