



K-means algorithm applied to synthetic well log data for reconstruction of lithologic logs

Ibsen P. Gomes¹, Victor R. Carreira², Rodrigo Bijani¹

¹ - Departamento de Geologia e Geofísica, Universidade Federal Fluminense (UFF)

² - Coordenação em Geofísica, Observatório Nacional (ON/MCTIC)

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

Artificial intelligence (AI) is an extremely important technique that is growing rapidly in the recent decades. Several different areas of research are dealing with AI, such as imaging recognition, voice identification, gaming, healthcare, robotics, automobiles, E-commerce and many others. In Geophysics, (AI) has been quite used in Seismic, well logging and Geoprocessing problems. In well logging, AI is increasingly relevant for predicting well log data, completion of missing logs and also for classification problems. However, the well log data suffer from severe acquisition difficulties, as caliper destruction and drilling failures, which make the well-log data unfeasible. Recently, several alternatives have been applied to overcome these drawbacks. In this work, we implement a user-friendly python code for modeling well-log data from an interpreted sedimentary basin. In our case, an interpreted section of Namorado Oilfield, in Campos Basin is considered. Basically, the user draws a 2D geologic section with specific lithologic information and values of log properties to be modeled (i.e., gamma-ray, sonic, density, neutrons porosity and others). After that, using clicks of the mouse, the top and the bottom of the well are defined and, consequently, the desired logs are computed, using a Gaussian probability density function. Once the well-log data is accurately modeled, we decided to go further in our application and use the K-means algorithm of scikit-learn to reconstruct lithologic logs. K-means is a widespread type of non-supervised algorithm commonly used for data clustering. The mathematical simplicity and the rapid convergence are two outstanding aspects of K-means clustering method. Basically, in our problem, lithologies in the synthetic log are referred to as clusters and each modeled well-log data is referred to as properties or attributes. To better analyze the k-means performance for reconstruction of lithologic logs, we performed three tests using different initializing configurations for the K-means code. In the first test, we run K-means in a “default” mode, in which all the centroids are randomly defined in a set of internal runs of the code. For this configuration, eight lithologies were accurately recovered and one is missing, from a total of nine lithologies in the synthetic well. In the second test, we set the (k-means++) mode, which is basically a different internal run based on statistical performance for defining the clusters. For this configuration, eight lithologies are still recovered and one missing. In the last test, we previously set four clusters as prior information. With this assumption, the K-means algorithm could reconstruct all nine lithologies of our synthetic well. The importance of this work shows how to take maximum advantage of k-means algorithms in a geophysical controlled synthetic scenario. As a future perspective, we intend to apply k-means technique in more complex geological synthetical scenarios and apply new machine learning techniques for classification problems.