The Influence of Geochemical Data Reconstruction Strategy on the Results of Automatic Stratigraphic Interpretation

Author Block: M. V. Sideltseva, 1Aramco Innovations LLC; N. A. Michael, 2Saudi Aramco, Dhahran.

Abstract

Geochemical data is a powerful instrument for stratigraphic correlation due to the strong relation between rock elemental composition and sedimentological environments and provenance. The cover of geochemical data is usually poor in exploration wells, adding to the difficulty of utilizing it in automatic interpretation workflows. This work aims to compare reliability of techniques for geochemical data imputation, with applications in stratigraphic interpretation.

In this work simple polynomial approximation, multi-directional recurrent neural network, and k-Nearest Neighbors imputation technique were compared as missing values reconstruction strategies. All the experiments were conducted on several wells from the Middle East region with available well logs and elemental composition data. The target for the reconstruction was to combine geochemical data with well logs measurements and to utilize them as features in stratigraphy prediction model. The stratigraphy interpretation was transferred into multiclass classification problem and solved by XGBoost classifier and simple Feedforward Neural Network. The total number of classes interpreted by the models was four. The result of the classification was evaluated based on classification quality metrics, which allowed us to conclude on the reliability of reconstructed data.

The baseline stratigraphy prediction models were the models trained on well logs data only. The F1-score of baseline models was up to 0.86. After the addition of geochemical data to the feature space the prediction quality was increased even when the geochemical data were reconstructed using simple polynomial interpolation techniques. Thus, confirming the importance of such type of data for stratigraphy interpretation task. Reconstruction of geochemical data with more complicated strategies (multi-directional recurrent neural network and k-Nearest Neighbors imputation technique) led to the increase of F1-score up to 0.9 for both XGBoost and Neural Network models. The complicated algorithms of missing values imputation are more sensitive to relations between features even with weak correlation, which typically characteristic of the relations between well logs and geochemical data. Therefore, multi-directional recurrent neural network, and k-Nearest Neighbors imputation technique resulted in more reliable prediction of geochemical data. It worth to out that reliability of geochemical data reconstruction can not be evaluated only based on the stratigraphy interpretation quality. Preservation of natural distributions and variance is crucial. The reconstruction observed from the multi-directional recurrent NN preserved natural trends of the original data, thus it can be considered as the best technique form the investigated ones.

Geochemical data is highly valuable for solving petrophysical tasks with machine learning algorithms. With the results presented in this work we have proven the reliability of existing missing data imputation techniques, which opens the
opportunity to enhance models quality with the value from thousands of new features generated from elements ratios. Results demonstrated that even for the limited number of wells reconstruction of the geochemical data is vital step for the stratigraphy interpretation workflows.