geoconvention
Virtual Event
September 21-23
2020
CSPG • CSEG • CWLS • GAC • MAC • IAH

Microseismic Data Quality Control Using Supervised Machine Learning

Brandon Reid, Ellie Ardakani, Ted Urbancic*
Meta Innovation Technologies

Summary

Microseismic data is one of the major geophysical data streams that is used in evaluation and characterization of hydraulic fracturing programs. To perform conclusive and unbiased interpretation, it is important to identify reliable versus unreliable microseismic data since using unreliable data leads to incorrect conclusions and faulty interpretations. Traditionally microseismic events have been quality controlled (QC'd) manually to identify a portion of the data that is free of biases and processing artifacts. In this work, we propose a supervised machine learning algorithm referred to as DQI (Data Quality Intelligence) algorithm which enables automated classification of high- versus low-quality data.

Methodology

Gradient boosting is a form of ensemble supervised learning where multiple weak learners are sequentially added together to construct a stronger meta-learner. This ML algorithm starts off with a single weak learner applied to the input data to estimate the value of the specified loss function. It then minimizes this loss function to calculate a residual from this learner. A new weak learner is fit to this residual and added to the model, lowering the total loss. This process is repeated until some threshold condition is met. In this study, we use gradient boosting to develop a model which can classify microseismic events for their quality based on measured/calculated attributes in an event catalogue. These events were initially classified as either low- or high-quality by an expert, allowing the use of these classifications to train our model with the 'true' quality values to assess the accuracy of this approach for data QC purposes.

Results

To perform the training for the DQI algorithm (our machine learning model), we first split the microseismic data (approximately 40,000 events) into three categories of training (30%), validation (35%), and testing data (35%). The reason behind this data splitting is to maximize performance while safeguarding against overfitting. Through the training process, we identified several key attributes that play a major role in determining the microseismic data quality; moment magnitude, moment, corner frequency, signal to noise ratio, number of P and S arrival picks, azimuth error, and distance from detection array and hydraulically stimulated stage. These attributes are ranked in the gradient boosting model in Figure 1, based on their importance, i.e. the number of times the model uses this attribute to make classification decision.

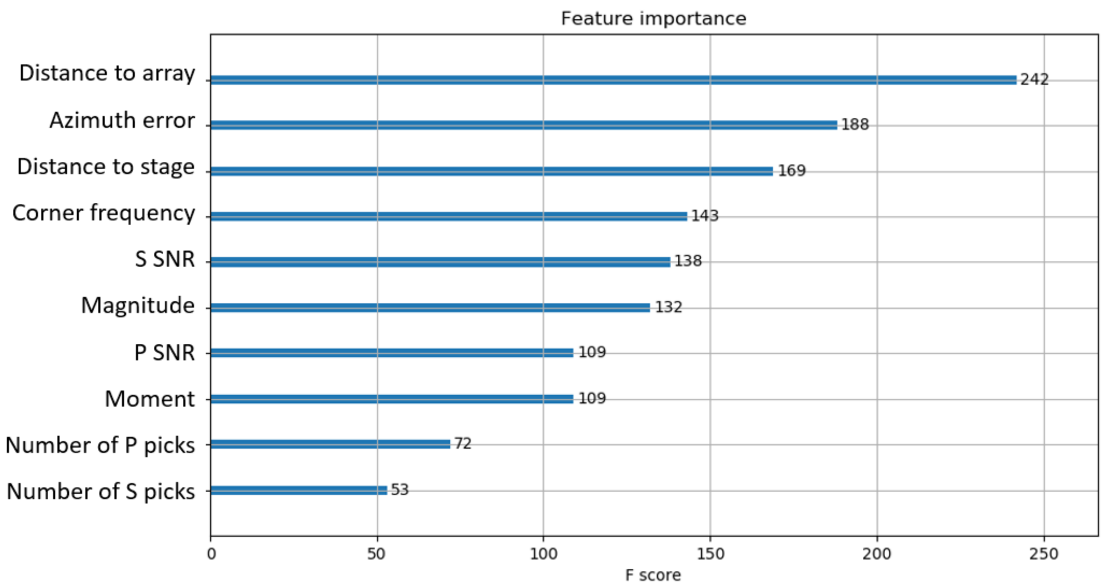

Figure 1: Features ranked by their importance in training the gradient boosting model.

After the model was trained, we evaluate it's accuracy by making predictions using the test set, the results of which are shown in the confusion matrix in Figure 2. The developed model has an accuracy of 94.4%, with a precision and recall of 93.2% and 98.2%, respectively.

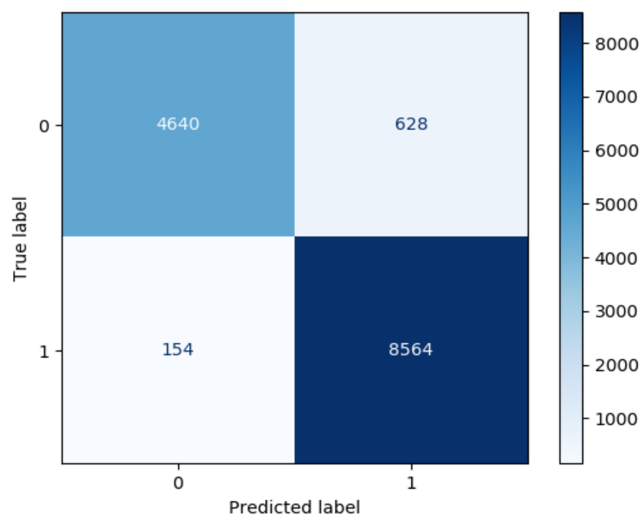

Figure 2: Confusion matrix showing the quantity of correct vs incorrect classification results of the gradient boosting model on the test set.

Conclusions

The Data Quality Intelligence (DQI) algorithm based on gradient boosting decision trees, shows promising results in automating the grunting task of manual microseismic data quality control by an expert with an accuracy of 94.4%.

References

T. Chen and C., Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

Y. Xie, C. Zhu, W. Zhou, Z. Li, X. Liu, M. Tu, 2018. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. Journal of Petroleum Science and Engineering, 160, 182-193.