Using Deep Learning and Distributed Machine Learning Algorithms to Forecast Missing Well Log Data*

Chijioke Ejimuda[1] and Emenike Ejimuda[1]

Search and Discovery Article #42234 (2018)**
Posted July 2, 2018

*Adapted from oral presentation given at 2018 Pacific Section AAPG Convention, Bakersfield, California, April 22-25, 2018

[1]Hybridata.us, Los Angeles, California (cj@hybridata.us)

Abstract

With limited amounts of prospects available within producing fields, the holistic and realistic evaluation of well logs represent an important parameter to guide development, production and investment decisions within the field. However, some wells with log data have missing curves or erroneous values, while some wells have no log data at all. We propose a deep learning and distributed machine learning model which effectively models the non-linearity within well log files to forecast a realistic estimate of missing well log data.

We used python programming language, Apache Spark and Tensorflow API's to perform this work. We scraped and downloaded about 985 well log data (log ascii files) of Cook Inlet Basin wells to the cloud. We extracted their actual data and metadata for all the wells. We subsequently saved the file into a feather format file which allows for quick file reads and writes. Furthermore, we preprocessed the well log data, normalized and scaled it to ensure all features were within the same scale and thus, reduce computational expense.

Initially, we applied Apache Spark's distributed machine learning stochastic linear regression model. The model accuracy was very low (about 10%). However, currently we are using auto encoder and convolutional neural network ResNet deep learning architecture. Preliminary results are promising and expect further model architecture and parameter tuning to yield better results. These and more will forecast well logs with missing values and wells without log data with high accuracy.

Authors of previous articles use commercial applications to generate such missing log data. Such applications are limited in their analysis and cannot handle big data. However, our methodology used python programming language, cloud distributed computing resources and open source tools to develop the deep learning and machine learning algorithm. Additionally, we are applying it to about 15,000,000 rows of data where such commercial applications are unable to scale or model complexities that exist amongst their variables. Additionally, our work will further assist petrophysicists, reservoir engineers and geologists build a more robust geologic and reservoir simulation model and guide management's current or future investment decisions in the field.

Selected References

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, 2015, Deep residual learning for image recognition: Cornell University Library. Website accessed June 25, 2018.
https://arxiv.org/abs/1512.03385

Huang, Jonathan, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy, 2017, Speed/accuracy trade-offs for modern convolutional object detectors: Cornell University Library. Website accessed June 25, 2018.
https://arxiv.org/abs/1611.10012

Stanley, R., B. Pierce, and D.W. Houseknecht, 2011, U.S. Geological Survey 2011 Assessment of Undiscovered Oil and Gas Resources of the Cook Inlet Region, South-Central Alaska: U.S. Geological Survey Open-File Report 2011-1237. Website accessed June 25, 2018.
https://pubs.usgs.gov/of/2011/1237/

Zaharia, Matei, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica, 2010, Spark: Cluster computing with working sets: University of California, Berkeley. Website accessed June 25, 2018.
https://people.csail.mit.edu/matei/papers/2010/hotcloud_spark.pdf

Using Deep Learning and Distributed Machine Learning Algorithms to Forecast Missing Well Log Data

CJ Ejimuda-MS; Emenike Ejimuda-PhD

A little about me...

CJ Ejimuda

Full Stack Data Scientist - hybriData

Msc. Petroleum Engineering - USC

BEngr. Electrical Electronics Engineering - FUTO

ExxonMobil and Aera Energy

Reservoir and Production Engineering

Outline

- Challenges
- Benefits
- Technical Basis
- Conclusion

Challenges

- Re-evaluate old well logs for opportunities
- Unable to effectively assess well / field reserves
- Challenge with inferring geological features

Benefits

- Obtain pertinent missing well log data
- Estimate reserves more accurately
- Better understand geology of play

Technical Basis

- Data Acquisition
- Data Preprocessing
- Data Augmentation
- Data Exploration
- Model Architecture
- Model Evaluation

Technical Basis

- Data Acquisition

- 995 raw log data from Cook Inlet Basin(5 fields)

- Python script to extract and download data to the cloud

Technical Basis

- Data Preprocessing

- Apache Spark to extract and join actual and metadata

- Saved to Apache Arrow format (feather)

- Dropped from 152 to 15 features

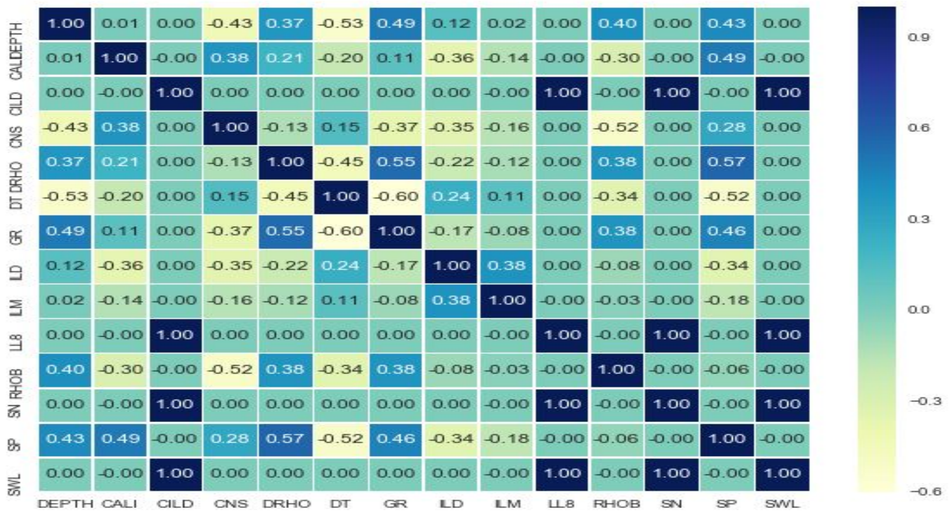
Technical Basis

- Data Augmentation
- Handled missing data

- Scaled and normalized data

Technical Basis

- Data Exploration - 50133203580000



Technical Basis

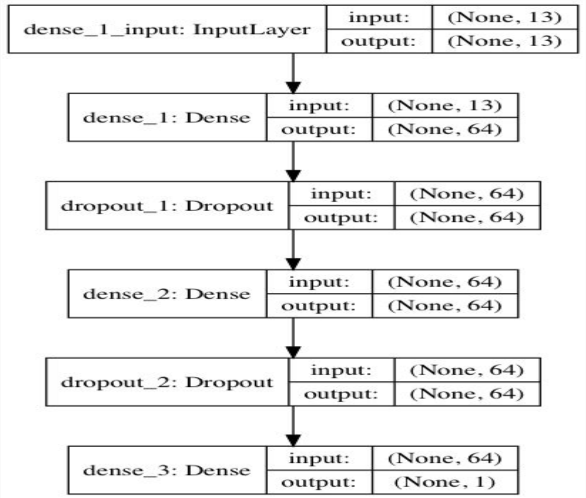
- Model Architecture

- Apache Spark Mllib :

- Very low accuracy

Technical Basis

- Model Architecture - ANN (MLP)

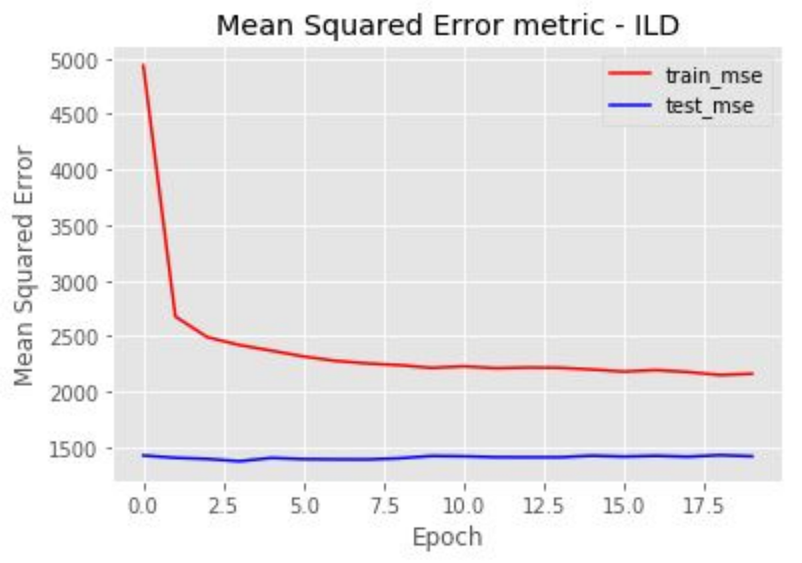


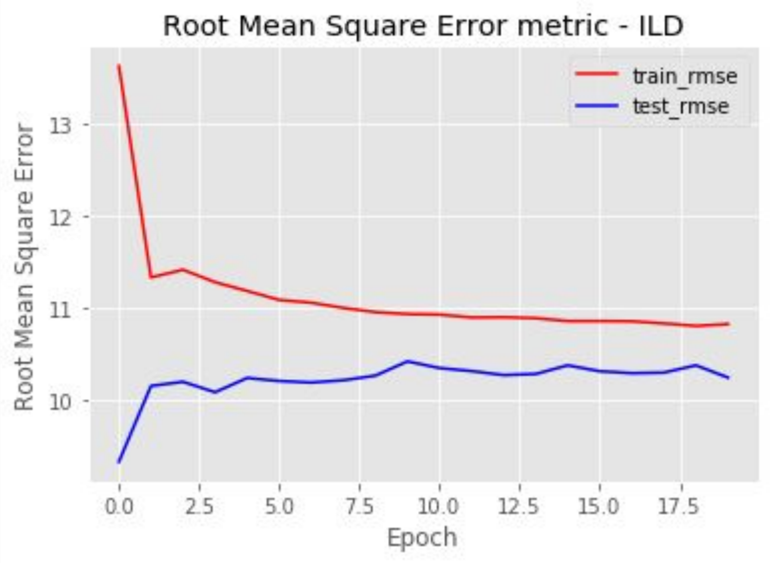
- 2 hidden layers
- Dropout and Relu applied
- Output layer-No activation function
- 2 regression error metrics
  - MSE, RMSE

- Resnet with Auto encoder didn't improve our model

Technical Basis

- Model Evaluation





Conclusion

- Deep Learning better approach
- Reduce cost of expensive logging suites
- Estimate lithology, fractures and more
- Room for model improvements