Improving Machine Learning Workflow and Business Value for Oil and Gas Applications

Fahad Ahmad (Senior Data Scientist, Halliburton HDS)
Bo Liang (Data Scientist, Halliburton HDS)
Patrick Brody (Data Scientist, Avanade)
Chao Yang (Data Scientist, Halliburton HDS)
Meng Zhang (Senior Data Scientist, Halliburton HDS)
DISCLAIMER 2019

The contents of this presentation are for informational purposes only. Halliburton** makes no representation or warranty about the accuracy or suitability of the information provided in this presentation and any related materials. Nothing in this presentation constitutes professional advice or consulting services. No contractual relationship with Halliburton is established by attending or viewing this presentation. No rights to intellectual property of Halliburton are granted through this presentation. The opinions expressed in this presentation are those of the author and do not necessarily represent the views of Halliburton. **Halliburton means Halliburton Energy Services, Inc., Landmark Graphics Corporation, and their affiliates.
Presentation Outline

- Problem statement
- Data preprocessing pipeline and tree-based models
- Deep learning model with minimal feature engineering
- Auto encoder for anomaly detection
- Business value optimization
- Auto ML
Problem Statement

- Application: Predictive maintenance on Oil and Gas service equipment

- Business impact:
  - Reduce NPT time for customer projects
  - Reduce overall customer project cost
  - Improve field operation efficiency
  - Increase customer satisfaction
Data Engineering, Machine Learning at Scale with Azure Databricks and Apache Spark

- **What is it:** on-demand, Spark clusters with interactive code notebooks
- **Advantage:** scalable computation, easy collaboration with notebooks, job scheduling, and machine learning pipelines
- **Example:** High Resolution Data Processing
  - Data Source: Azure Blob Storage
  - Format: 1,000’s of parquet files
  - Size: 650gb raw data per year
  - Time: 2 – 12hr run time to clean depending on volume
- **Success:** cleaned data (missing value, null, outlier, aggregation) ready for sub-second level machine and deep learning
Tree-based machine learning models for initial prediction implementation

▪ **What is it:** tree-based regression & classification models implementing equipment failure prediction

▪ **Advantage:** distributed training, deployable pipelines with Spark ML

▪ **Example:** Equipment Failure Prediction
  ▪ Data Source: Azure Blob Storage
  ▪ Format: cleaned parquet files
  ▪ Size: 4m data points train, 1m test
  ▪ Time: 2 – 6hr train time depending on grid depth

▪ **Success:** initial predictions deployed & updated daily to end users through simple Power BI dashboard; model tuning on-going
Deep Learning Hybrid Approach

- Combine Fully Convolutional Network (FCN) and Long Short Term Memory Network (LSTM) with attention mechanism
- Temporal convolutions capture the local variations in the data whereas LSTMs capture the long-term variations in the data
Autoencoder-Decoder for Anomaly Detection

- Usage:
  - Anomaly detection in equipment
  - Neural network
  - Unsupervised

- Types
  - Classical autoencoder
  - Sequence to Sequence LSTM autoencoder
  - Autoencoder-classification
Autoencoder-Decoder for Anomaly Detection
Business Value Optimization

- **Challenges**: competing metrics to maximize the business value for any machine or deep learning model when balancing model precision and recall
- **Advantage**: customized value analysis to maximize the value (>50%) of machine learning
- **Example**: Equipment Failure Prediction
  - Improved machine learning algorithm with a custom cost function, fully optimizing model performance considering gain and penalty
  - Required methodology due to make tradeoffs between accuracy and recall

**PROCESS**

1. Classifier prediction
2. Transformation to unit level
3. Cross Validation
4. Apply value savings for business

![Graph showing recall vs. average hour loss with model, hour-based, run-to-fail, and saving contour lines.](image-url)
Automated Machine Learning Application

- **What is it:** automated tool for model selection, hyperparameter tuning, and feature engineering
- **Advantage:** less coding effort, save development time & resources, and leverage data science best practice
- **Example:** Drilling Equipment Repair Prediction
  - Data Source: Oracle, SQL Server
  - Format: csv files
  - Size: 20gb raw data, 1gb train
  - Time: 2 week build that can be reused in future projects
- **Success:** reduced project cycle time by 20% to achieve >90% accuracy
Conclusion

- Improved workflow and collaboration through Azure Databricks platform
- Implemented both tree-based ML algorithms as well as deep learning algorithms for different resolution of the data sets and ETL processes
- Solved business problem using optimized ML pipeline
- Maximized business impact through customized cost functions for ML cross validation
- Increased project efficiency through Auto ML tools
Thank You
Your feedback is very important to us. Please open the LIFE2019 app to answer a few short questions on this presentation.