TIME-SERIES CLASSIFICATION TO PREDICT UTILITY GRID FAULT EVENTS

Case Study

Highlights

Industries
- Utilities, Telecommunications

Outcomes
- Machine learning provided insights on transmission health
- Spark powers and automates the model for transmission engineers to examine
- Feedback loop allows engineers to update incorrect classifications
- Enabled improved asset renewal via improved fault detection and identification of at-risk locations

Statistics
- Trained on 5 fault causes
- Trained on 2 relay types
- Each file represents ~1 second of data (~3,000 rows)
- 20,000 files per day on average

Technical Areas
- Data Science
- Data Engineering
- Machine Learning Classifier
- Time-Series Analysis
- Discrete Fourier Transformations
- Extraction, Transform, Load

Technologies
- Hadoop
- Hive
- Spark
- PySpark
- Zeppelin

Data
- Sensors
- Spark-generated features

Using data science to gain insight from electrical sensor data placed on the grid of a utility company

Business Challenge

A utility company needed an automated method for classifying faults that were occurring regularly on their transmission grid

While a major utility company is aware when something goes wrong on their transmission grid (a fault), they often have limited ability to understand what caused the fault. For this utility company, their only means to understand the root cause of a fault was manual inspection of power system files by an engineer. Understanding these files takes very specialized skills. The company asked ILW to produce a machine learning algorithm capable of interpreting the data similarly to the manual interpretation by specialized engineers to determine if faults occurred.

Innovative Solution

ILW created a machine learning classifier capable of detecting faults with accuracy exceeding initial client expectations

ILW built an end-to-end solution, from data arrival through exposing insights, to drastically reduce the manual analysis completed by transmission engineers while increasing the accuracy of predictions. Time-series measurements of the oscillography voltage and current data for different phases running along a particular wire were collected. ILW developed an automated system that ingested the data into Hadoop to produce a much more readily consumable data schema than the software previously used by the transmission engineers. A PySpark application then generated features from each file, which were fed into a classifier powered by machine learning techniques to determine if a fault occurred. The company's goal was to accurately predict 80% of lightning and galloping conductor events on the network. ILW's model consistently identified faults with an accuracy between 92% to 95% and accuracy is continuing to rise with new feature generation work currently being performed. This process highlighted the capability of a machine to classify types of time series data using discrete Fourier transformations to identify and extract significant features.

ILLUMINATION WORKS

CIO Review 100 MOST PROMISING BIG DATA SOLUTION PROVIDERS-2017

www.illuminationworksllc.com