Monetizing Utility Data With Cognitive Anomaly Detection & Prediction

BY Taj Darra
ABOUT THE AUTHOR

Taj Darra

Taj Darra is an experienced Data Scientist with a demonstrated history of working in both large-scale enterprises and smaller startups. Skilled in Machine Learning, Statistical Data Analysis, Statistical Modelling, Data Science, and Python. Strong engineering professional with a Bachelor of Science (B.Sc.) focused in Mathematical Sciences from The University of British Columbia / UBC. Work experience include grad-level research as well as working on IBM’s Watson in Canada.
The New Forces at Work - data

Explosion of Smart-enabled, connected Sensors

Big Data Platforms to Collect, Process & Store Data at Mass-Scale

A Tsunami of Mega Big Data

Groundbreaking Advancements in Data Science Technologies
Where Is The **Money** In Industrial IoT?

![Diagram showing the value at stake in industrial IoT.](source: Cisco Consulting Services, 2014)
**IIoT Examples: Stuxnet**

**What:** Malicious computer worm infects Iranian nuclear plant that intercepts communications between Step 7 software and PLC layer. Causes change in rotational speed of connected motors responsible for centrifugation.

**Cost:** 1,000 centrifuges destroyed as a result. 11% of daily output. Estimated losses per day = $1m. Total downtime = 8 months. $240m
IloT Examples: Ericsson

**What**: Automated scheduling of transport services. A user-specified, high-level objective such as “deliver cargo A to point B in time C” is automatically broken down by the system to be fully automated.

**Savings**: At least 200% efficiency increase. Reduction in human capital costs = $100m per year
Ilot Examples: Town of Olds, Alberta, Canada

What: In 2010, the town deployed a system that placed acoustic sensors in service pipes that analyzed sound patterns every day, detecting new, evolving and pre-existing leaks automatically. Over time, an expanding database of historical sensor information a cognitive approach to handling and preventing leaks.

Cost: The American Water Works Association indicated that the 237,600 water line breaks each year in the U.S. cost public water utilities approximately $2.8B annually.
Traditional approaches of Predictive Analytics don’t work for Predictive Maintenance
Generalized Models created using Samples simply Don’t Work

Only 20% of Asset Failures are Common & Predictable

80% of Asset Failures are seemingly Random
Why do they fail?

- Limitations of statistical approaches as assumptions about data generating distribution generally do not hold.
- Choosing a single statistical test proves difficult (Grubb’s Test, Kurtosis, AIC, Mahalanobis)
- ARIMA, Regression, Mixture Models perform poorly for noisy, dynamic series
Lack of labeled training data

- Unsupervised learning problem
- Most data collected lacks any symbolic representation or code
- Dynamic in nature as environmental states change over time
- Noise ~ anomalies

LOOK AT ALL THE LABELED DATA

I DON'T HAVE
**Hard to define** failures using data

What you **think** failures look like  
What failures **actually** look like
Hard to define failures using data

Fig. 1. A simple example of anomalies in a 2-dimensional data set.
Building **high complexity** ML systems

Figure 2: A high-level, conceptual view of the framework architecture

Machine Learning: The High-Interest Credit Card of Technical Debt

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young
{dsculley,g Holt, dgolovin, edavydov}@google.com
{toddphillips, ebner, vchaudhary, mwyoun g}@google.com
Google, Inc
The Cognitive Conclusion?

This is not a **HUMAN SCALE SOLUTION**
Can’t throw enough Data Scientists at it

This is a **MACHINE SCALE SOLUTION**
Requires Automation | Scalability | Repeatability
How we do it - CADP

1. Cluster like sensor signals into Groups

2. Define State of signal & use throughout as smallest unit of data

3. Merge like States in order to characterize Behavior.

4. Run Anomaly Detection ML on behavioral states of signal, where one state does not necessarily follow same pattern as another.

5. After scores are generated, label points as anomalous accordingly. Meta-learning is used to infer parameters for models

6. Use labeled anomalies to predict Future anomalies via supervised learning
Cognitive IoT Framework

Abstracting away the hard parts, extracted from Cognitive Internet of Things: A New Paradigm Beyond Connection
Think about the impact of scaling the creation & sustaining of Predictive Data Models across a business heavy in IoT or sensor data in an Automated fashion instead of manually; thinking of every piece of equip with sensors each getting its own customized Pred Data Models, and processing data from all the sensors instead of cherry picking while hoping you don’t miss any critical sensors, and trying all modeling approaches instead of hoping you didn’t eliminate any modeling approaches that could have been most effective.

That is the impact of Automated vs Manual.