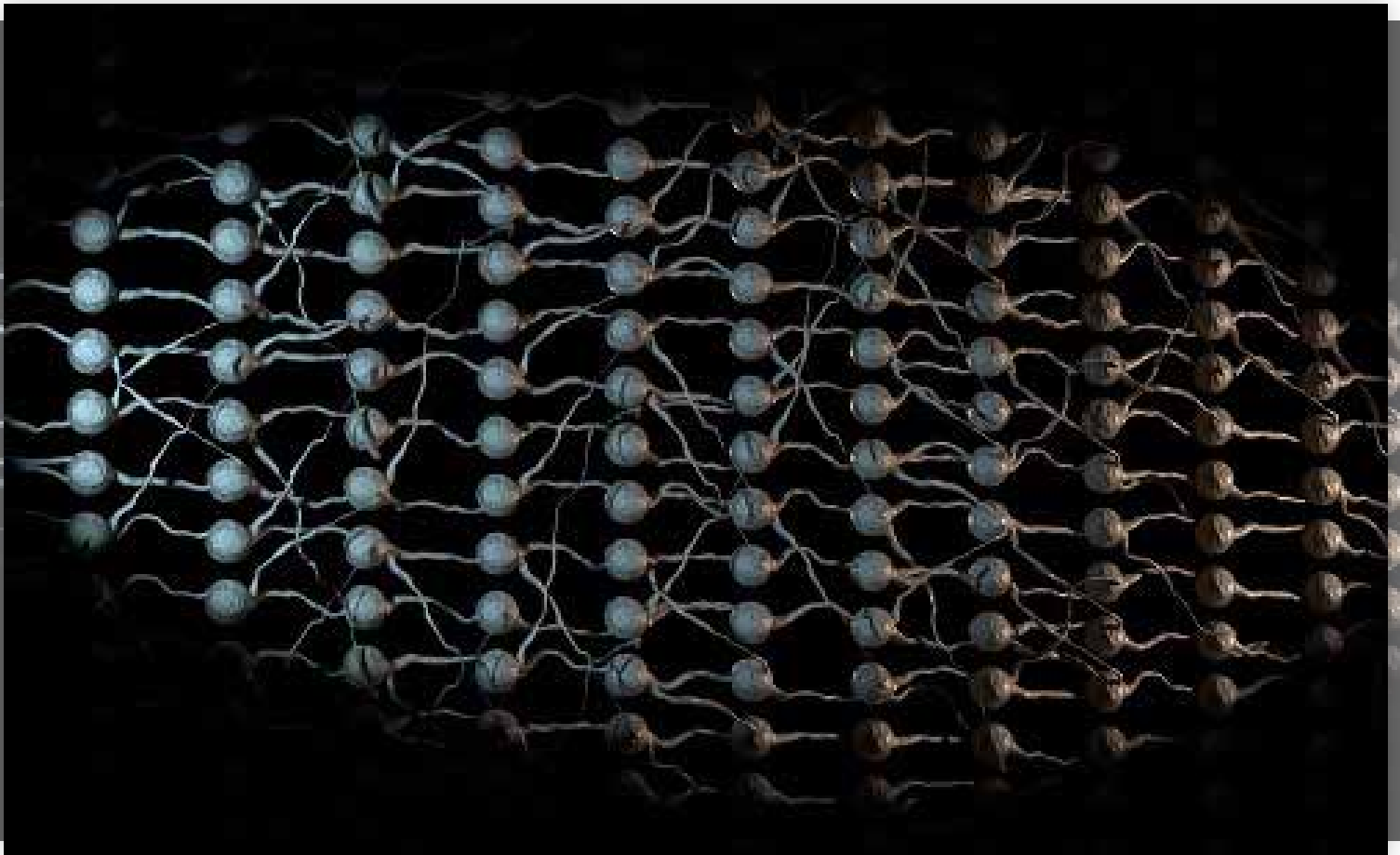


Deep Learning Neural Networks For IoT:

Leveraging Big Data, Advanced Machine Learning, & Complex Event Processing Technologies
Written by Bruce Ho, Chief Big Data Scientist, BigR.io, LLC



Deep Learning Neural Networks For IoT:

Leveraging Big Data, Advanced Machine Learning, & Complex Event Processing Technologies
Written by Bruce Ho, Chief Big Data Scientist, BigR.io, LLC

About the Author

Bruce is a top rated problem solver, communicator, and data scientist. Bruce has a unique strength in bridging software engineering with mathematical algorithms. He has over 15 years of IT experience, engaging in high visibility, large scale projects in marketing automation, cloud solutions, Big Data engineering and predictive analytics for top technology companies like Amazon, TeraData, and Life Technologies. He is a certified AWS solutions Architect, with a specialty in Big Data practices. Bruce published over 100 scientific papers during his academic career, and continues to draw the best ideas from start-of-the-art research to fuel his Data Science practice.

Bruce has a deep rooted passion in using statistics modeling to improve the way business is done in the fashion of "Moneyball". His fascination with the use of mathematics to solve real world problems dates back to the early days of



[View Bruce's LinkedIn](#)

Vector Calculus and Newtonian physics. After witnessing the growth of computing power in the mode of Moore's law, Bruce fully embraces the vision that data driven decision making will soon become a basic tenet of business practice. He loves playing the advocate-in-chief for predictive analysis when resolving client companies' challenges.

Bruce holds a BS in Physics from MIT, and an MS in Electrical Engineering & PhD in Applied Physics from Caltech.

Abstract

The global Internet of Things (IoT) market will grow to \$1.7 trillion in 2020 from \$656 billion in 2014, according to IDC Insights Research. IoT is forecast to generate a staggering 500 zettabytes of data per year by 2019, coming from 50 billion connected devices (up from 134.5 ZB per year in 2014), according to a report from Cisco. Massive challenges arise in managing this data, making it "useful". This is due not just to the sheer volume of data being generated, but also the inherent complexity in the data. Fortunately, there are great open source applications and frameworks such as Spark & Hadoop that have emerged to address these challenges. Similarly, advances in Neural Networks, Deep Learning, and Complex Event Processing help drive ever-more sophisticated analyses. BigR.io can help you take on these new challenges at any stage of the adoption lifecycle: strategy/planning phase, infrastructure design and implementation, or production operations and data science.



About BigR.io

BigR.io is a technology consulting firm empowering data to drive analytics for revenue growth and operational efficiencies. Our teams deliver software solutions, data science strategies, enterprise infrastructure, and management consulting to the world's largest companies. We are an elite group with MIT roots, shining when tasked with complex missions: assembling mounds of data from a variety of sources, building high-volume, highly-available systems, and orchestrating analytics to transform technology into perceivable business value.

With extensive domain knowledge, BigR.io has teams of architects and engineers that deliver best-in-class solutions across a variety of verticals. This diverse industry exposure and our constant run-in with the cutting edge, empowers us with invaluable tools, tricks, and techniques. We bring knowledge and horsepower that consistently delivers innovative, cost-conscious, and extensible results to complex software and data challenges. Learn more at www.bigr.io

Overview

Potential applications of IoT range from health maintenance and remote diagnoses at an individual level to grandiose world-changing scenarios like smart semi-automated factories, buildings, homes, and cities. IoT systems generate serious amounts of data. For example:



a Boeing 787 aircraft generates 40TB per hour of flight



a Rio Tinto mining operation can generate up to 2.4TB of data per minute

Data is ingested with or without schema, in textual, audio, video, imagery and binary forms, sometimes multi-lingual and often encrypted, but almost always with real-time velocity. While the initial technology challenge in harnessing IoT is an infrastructural upgrade to address the data storage, integration, and analytic requirements, the end goal is to generate meaningful business insights from the ocean of data that can translate to strategic business advantages.

The first step is making sure your infrastructure is ready for the influx of this increased data volume. It is imperative to productionize Hadoop and reap the benefits of technologies such as Spark, Hive, and Mahout. BigR.io has specialists who can evaluate your current systems and provide any architectural direction necessary to update your infrastructure to embrace "Big Data", while leveraging your existing investments. Once the environment is fully implemented, BigR.io will then help you capitalize on your investment with Machine Learning experts who can help you to start mining, surfacing insights, and automating the process of notifications and autonomous actions based on data insights.





The branch of machine learning most central to IoT is automated rule generation; BigR.io uses the term Complex Event Processing (CEP). These rules represent causal relations between the observed events (e.g. noise, vibration) and the phenomena to be detected (worn washer). Human experts can be employed to create user-defined rules within reasonable limits of complexity. In sensor data terms, that limit is the first millimeter in the journey to Mars. The raw events themselves rarely tell a clear story. Reliable and identifiable signs of trouble generally consist of a combination of low-level events masqueraded in irregular temporal patterns. Individual events that make up the valid signal can exhibit temporal behaviors over impossibly wide ranges from sub seconds to months or longer, each further confounded by anomalies such as sporadicity or outliers. Only machine learning techniques can overcome both the challenge of collecting, preparing and fusing the massive data into useful feature sets, and extract the event patterns that can be inducted as readable rules for predicting a future recurrence of a suspect phenomenon.

As in any nascent field of endeavor, there are multiple candidate approaches inspired by techniques proven in related past experiences, each with their promises and handicaps. While abundant rule-based classifiers are reported in literature and have gone through extended efforts of improvement, they were generally applied to classes of problems that are narrower in scope, of an offline nature, and lack explicit temporal attributes. At BigR.io, we reach beyond these more established classification approaches in favor of innovations that deal more effectively with the greater levels of volume and complexity typically found in the IoT context. As usually is the case in machine learning, we find that better final results are obtained by using an ensemble of models that are optimally combined using proven techniques like Super Learner.

Deep Learning Neural Networks

For problems of this complexity, Neural Networks are a natural fit. In statistical terms, a neural network implements regression or classification by applying nonlinear transformation to linear combinations of raw input features. Because of the typically 3 or 4 layers and potentially high number of nodes per layer, it is generally untenable to interpret the intermediate model representations, even when good prediction results are achieved, and the computational load requires a dedicated engineering effort.

Neural Networks have many key characteristics which make it an attractive and typically the default option for very complex modeling such as those found in IoT applications. Sensor data is voluminous with complex patterns (especially temporal patterns); both fall under the strengths of neural networks. The variety of data representations makes feature engineering difficult for IoT, but neural networks automate feature engineering. Neural Networks also excel in cross-modality learning, matching the multiple modalities found in IoT.



Adding Deep Learning to Neural Network architectures takes the sophistication and accuracy of machine-generated insights to the next level and is BigR.io's preferred method. Deep Learning Neural Networks differ from "normal" neural networks by adding in more hidden layers and can be trained in both an unsupervised and supervised manner (although we suggest employing unsupervised learning tasks as often as feasible).

There are numerous additional strengths in the Deep Learning approach:

- Full expressiveness from non-linear transformations
- Robustness to unintended feature correlations
- Allows extraction of learned features
- Can stop training anytime and reap the rewards
- Results improve with more data
- World class pattern recognition capabilities

Because of its richness in expressiveness, Deep Learning can be counted on to tackle IoT rule extraction from a modeling perspective. However, the complexity and risks associated with the implementation should be weighted carefully. Consider some well known challenges:

- Slow to train – high iterations and many hyper parameters translate to significant computing time
- Black box paradigm – subject matter experts cannot make sense of the connections to improve results
- Over fitting is a common problem that requires attention
- Still requires preprocessing steps to handle dirty data problems such as missing values
- Practitioners generally resort to special hardware to achieve desired performance

BigR.io's team of highly-trained specialists is well-equipped to take on these implementation challenges. We select from a host of available platforms including Apache Spark, Nvidia CUDA, or HP Distributed Mesh Computing. Often, having the necessary intuitions derived from experience can expedite the completion of training by 10 times.

In certain cases, the performance cost associated with Neural Networks, especially with Deep Learning motivates other approaches. One alternative BigR.io often champions is the use of a specialty CEP engine which is optimized for flowing sensor data.

Specialty CEP Engine

In this approach, we look at the IoT rule extraction challenge not as a generalized machine learning problem, but rather one which is characterized by some unique aspects:



- Voluminous and flowing data
- The input is one or more event traces
- Temporal pattern plays a prominent role besides event types and attributes
- A decomposable problem into time window, sequence and conjunctive relationships
- The event sequence and their time relationship forms large grains of composite events
- The conjunction of the composite events formulates describable rules for predicting suspect phenomenon

This CEP engine represents a practical tradeoff between expressiveness and performance. Where a comparable IoT study may require days to process, this specialized engine may complete its task in under an hour. Parallelization based on in-memory technologies such as Apache Spark may soon lead to real-time or near real-time IoT analysis. Unlike the case of a Neural Network, a subject matter expert can make sense of the results from this engine, and may be able to manually optimize the rule through iterations.

These two approaches are complementary in a number of ways. For example, a prominent derived feature involving an obscure non-linear combination of raw events may be extracted from the Neural Network study and fed into the CEP engine and vastly improve the quality of prediction. The CEP engine might drive an initial effort of any study, extracting most of the low hanging fruit rules. This leaves Neural Networks to detect the remaining rules after pruning either the sample data or event types from the first phase. In some cases, the two techniques can simply be used for cross-validation when inconsistent results are obtained.

Ensemble of Models

Running more than one modeling approach is more the norm than the exception in today's machine learning best practices. Recent work has demonstrated that an ensemble of a collection of algorithms can outperform a single algorithm. The stacking algorithm and combined weak classifiers are two examples of formal research where the ensemble approach produces better results.

In this context, the two model approach can lead to a final result in several ways:

- Voluminous and flowing data
- The input is one or more event traces
- Temporal pattern plays a prominent role besides event types and attributes
- A decomposable problem into time window, sequence and conjunctive relationships
- The event sequence and their time relationship forms large grains of composite events
- The conjunction of the composite events formulates describable rules for predicting suspect phenomenon

A Super Learner is a loss-based supervised learning method that finds the optimal combination of a collection of prediction algorithms. It is generally applicable to any project with either diverse models or a single model which leverages different feature sets and modeling parameters. Such provisions can mean significant improvements in terms of reduced false alarms or increased accuracy of detection. Depending on the context of the application, one or both of such improvements can have a strong impact on the perceived success of the project.