Hybrid Analytics: *The Value of Machine and Human Supervision*

Written by
Noah Bethel, Vice President-Product Development
PdMA® Corporation

Big Data is a power to be harnessed. For nearly three decades since the concept of Big Data entered the public consciousness, researchers, programmers and data scientists have struggled with the best way to leverage the information and take full advantage of it. In the interim, technology has evolved to accommodate and differentiate even more copious data generated every second.

Some argue that the current level of technological sophistication is sufficient for machine supervision without the need for review or interpretation from a data analyst given Big Data's capacity to drive new applications and limit human interaction. The belief that it's the beginning of the end for the data analyst appears to be finding advocates in the industrial market, particularly in motor and facility management. They prophesize such expertise will be relegated to history thanks to technology's ever expanding capabilities. As analysts retire, there seems to be little effort to replenish the supply. Perhaps this is because more executives think Big Data is the proverbial handwriting on the wall—the death knell for analyst supervision of machines.

In some ways, this condition illustrates a utopian belief that unsupervised machine learning will eventually ensure reliability of the motor specifically and the facility, generally, without the need for human involvement. It is based on an assumption that a combination of predictive analytics to extract information from the data to configure outcomes followed by prescriptive analytics to generate the optimal solution will resolve all issues—a risky and short-sighted belief.

A hybrid analytical solution that includes numbers' crunching from Big Data and supervised machine learning provided by the skilled data analyst is a more practical and workable approach.

Big Data and the Industrial Market

In Oct. 1997, "Big Data" was the focal point of a highly-publicized paper written by researchers Michael Cox and David Ellsworth. The two referred to "the problem of big data" to describe data sets too large to be accommodated by core memory, local or remote disks.
Fast forward to 2011 when information specialists at the University of California, Davis, determined that nearly 95 percent of the world's storage information capacity is digital—a figure that probably seemed unfathomable at the dawn of the digital age a decade earlier. By this time, its importance had not been lost on business and industrial markets, which recognized the benefits and competitive advantage possibilities through harnessing Big Data.

Now, late in 21st Century's second decade, the development of the Internet of Things (IoT) offers a solution for leveraging the seemingly amorphous mass of Big Data. A paper authored by the Technology Strategy Board in 2013 included a comprehensive definition of IoT: "a state where Things (e.g. objects, environments, vehicles and clothing) will have more and more information associated with them and may have the ability to sense, communicate, network and produce new information becoming an integral part of the Internet." Research envisions an all-powerful Internet of Things with physical objects seamlessly integrated into an information network.

Advocates of IoT foresee an environment that is all encompassing and self-contained—an environment composed solely of machines that communicate under all circumstances and conditions. Converging technologies from different types of physical objects are expected to allow for greater understanding of processes and problem solving. For these advocates, machine supervision reaches its fulfillment with the concept of Industry 4.0, the industry protocol in which all of the data sources in IoT communicate with smart sensors and accurately conduct diagnostics and remedies when needed. Some envision Industry 4.0 development as going beyond self-sustaining. They see it as advanced to the point that machines can learn and apply new data, which explains why they consider the concept revolutionary.

What happens to the data analyst in this scenario? The position would be relegated to obsolescence assuming all of the machines do what advocates claim they will eventually do especially when it comes to diagnostics. If this prediction of a machine-supervised future actually comes to fruition, what could possibly go wrong?

The answer: quite a bit, which is why sounding the death knell for human supervision of machines by data analysts is premature at best and shortsighted at worst. Ask any facility manager or supervisor whose operations have been either hampered or halted because of seemingly inexplicable motor or other technology issues. To expect these issues to disappear in the utopia envisioned with Industry 4.0 is to base operational decision-making on what should be viewed as a false assumption.

Data Sources and Motor Reliability

To understand the ramifications of these views of a future of unsupervised machines, start with a comprehensive overview of the analytical evolution of IT and its most basic factor: descriptive analytics. Because this is the lowest level, descriptive analytics can only tell system operators that something has happened—information that, in most cases, should be evident. The next step, diagnostic analytics, answers the first question likely asked by the operators when a problem surfaces: why? What comes next is not as clear cut. It's where the analysis and remedies are not as well defined, especially if all these connected machines are relied upon solely for answers without any human interaction.

If the advocates for self-contained machine supervision are to be believed, IoT and Industry 4.0 will be the only source needed for the next step in the analytical revolution: predictive analytics. Big Data, in terms of the industrial market, drives predictive analytics—the forecast of what is likely to happen—enabling facility operators and managers to make better choices that impact plant reliability positively and cost-effectively. Predictive analytics are the resources for prioritizing assets for review as well as evaluating potential issues, but are they adequate enough if generated only by machine-provided data sources? Not really, especially in the motor industry where directors and operators work every day to achieve the goal of zero motor breakdowns.

Despite its capabilities, predictive analytics is insufficient for an environment of completely non-supervised machine learning and is likely to remain an incomplete response regardless of IoT and Industry 4.0 expansion. Process changes, environment changes, power distribution changes, sensor changes, and even operator changes throughout the life of the motor will require changes in decision-making. These changes are inevitable during the motor's useful life.

The next step is prescriptive analytics. A closer look at this level reveals a further justification for human analysis concurrent with machine supervision. With prescriptive analytics, the analyst takes the data gleaned from predictive analytics and designs actions and processes to achieve four desired outcomes: greater reliability, longer-lasting optimal performance, enhancing the motor or machine's useful life and improving plant reliability.

Hybrid Analytics

The optimal solution for ensuring the reliability of motors and the facilities where they operate requires the supervision of skilled data analysts trained to examine, recognize and recommend actions that the machines by themselves are incapable of generating. Thus the term hybrid analytics is used to describe a combination of supervised and unsupervised machine learning. Hybrid analytics maximizes both without relying solely on one or the other. Both require Big Data for their foundation and analysis. Yet, the argument can be made convincingly that it is the analyst's contribution that offers the better option for developing process changes to benefit the facility and its assets. The analyst's conclusions are, of course, based on the data sourced from the motors or other machines.

In this context, consider supervised learning, which is composed of the following elements:

- *Data that compares trends on a local by local basis.* The goal is to ensure the data is relevant in all cases.
- *Comparison of local operations to similar regional operations.* As is the case with trend evaluation, comparisons with local operations are designed to ensure the information applies closely to both without any irrelevant data skewering the findings and outcomes.
- *Time/application*. Applications vary as does the time a specific motor runs at its rated speed.
- *Characteristics/Plotting*. Motor characteristics tend to vary depending on type as well as the power supply for the motor under analysis.
- *Third-party assistance.* There are no hybrid analytics without the machine supervisor—the data analyst. This expert analyst may be an employee, contractor or vendor such as PdMA Technical Support.
- *The Cloud*. A cloud-based database can provide easy access to data for a data analyst anywhere in the world without needing to enter the corporate network or firewall. This easy access can yield reports that can be applied to predictive and prescriptive analytics.
- *The Dashboard*. A Site Condition Dashboard is vital for determining the status of motors placed in caution or severe condition in the determination of a site's electrical reliability condition. The dashboard is a macro view of plant reliability from management's perspective that should be considered a resource in the decision-making process. It identifies assets such as motors that have been tested but have not been assigned a condition—information that should be made available to the data analyst.

Role of the Data Analyst

In a hybrid analytics environment, the data analyst's role does not change from the traditional: to improve machine function and eventual learning through supervision. Despite some futurists' predictions of unlimited machine learning rendering human evaluation and occasional intervention unnecessary, the analyst remains the key to machine capability expansion and, especially, its reliability. Processes are subject to incremental change. As these changes occur, reliability and accuracy of the data the machine produces are impacted no matter how advanced Industry 4.0 technology may be. Process changes underscore the need for supervised learning. The analyst establishes guidelines so machines learn in a more controlled manner—a form of guidance that directs the machine away from data that is flawed or erroneous.

When issues or emergencies occur, supervision expedites machine removal of irrelevant data to clear a path for finding a cause and generating prescriptive analytical solutions. The machine output in this scenario is more efficient. So is the analyst who will not waste the facility's time or budget by having to delve into excessive and extraneous data.

None of this should be viewed as some last gasp to justify the usage of data analysts either in facilities or through the Cloud. The fact is that motors and other machines still require supervision. The most common occurrence is the nuisance alarm that is not taken seriously because an unsupervised machine cannot detect the cause despite the sophistication of IoT in tandem with Industry 4.0. One example occurred on the east coast in the utility industry in which an electrical alarm detected by an unsupervised machine learning analytics software was determined to be a nuisance alarm caused by a failed sensor. The alarm was ignored due to over reliance on the analytics software, which failed to cross compare the nuisance alarm with common motor application knowledge and other sensor-driven data points like vibration analysis. The analytics process totally missed an open phase in the power supply.

The motor had been lightly loaded so temperature was not alarming despite elevated current flow on the two phases still conducting current. Had the plant fully loaded the motor, the situation could have rapidly deteriorated into an emergency. Fortunately for the facility, a passing analyst with application knowledge was able to recognize a pattern between unrelated data points and point out that there were reasons to believe that this was not a nuisance alarm. At that time, the plant was planning a shutdown and the motor would not have restarted, extending the outage. The incident should be considered a classic example how and why the intervention of a data analyst can save a company substantial repair and replacement costs.

Prevent Problems, Avoid Risks

Executives and managers who believe that Big Data without data analysts is capable of preventing future issues and solving problems as they occur are placing their company and its assets at risk. Over-reliance on the machines in an unsupervised data interpretation environment can have a deleterious impact on plant reliability and the ever-present bottom line.

That is the whole point of hybrid analytics—allowing the machines to crunch the data, which is then leveraged by the supervision of a data analyst whose skills and technological expertise add significantly to machine learning. There is too much at stake to risk performance, efficiency and profit on machine-produced data alone.

Facility managers are coming to this realization in growing numbers. Supervised machine learning is essential for maintaining motor and plant reliability.

About the Author:
Noah Bethel is vice president of product development for PdMA Corporation, Tampa, FL, the leader in the field of predictive maintenance, condition monitoring applications, and development of electric motor test equipment for motor circuit analysis. Tel: (800) 476-6463 or www.pdma.com.