GIGAOM RESEARCH

Spreading the nexus of intelligence: blending massive connectivity with cognitive computing to enable the insightful fog

David Loshin

August 25, 2014

This report is underwritten by Saffron Technology.

TABLE OF CONTENTS

Executive summary

As our world becomes more attuned to the current generation, and more importantly, the use of massive amounts of data, information technology (IT) professionals are increasingly looking to new technologies to help focus on deriving value from the velocity of data streaming from a wide variety of data sources.

The breadth of the internet and its connective capabilities has enabled the evolution of the internet of things (IoT), a dynamic ecosystem that facilitates the exchange of information among a cohort of devices organized to meet specific business needs. It does this through a growing, yet intricate interconnection of uniquely identifiable computing resources, using the internet's infrastructure and employing internet protocols. Extending beyond the traditional system-to-system networks, these connected devices span the architectural palette, from traditional computing systems, to specialty embedded computer modules, down to tiny micro-sensors with mobile-networking capabilities.

In this paper, geared to the needs of the C-suite, we'll explore the future of predictive analytics by looking at some potential use cases in which multiple data sets from different types of devices contribute to evolving models that provide value and benefits to hierarchies of vested stakeholders. We'll also introduce the concept of the "insightful fog," in which storage models and computing demands are distributed among interconnected devices, facilitating business discoveries that influence improved operations and decisions. We'll then summarize the key aspects of the intelligent systems that would be able to deliver on the promise of this vision.

Key takeaways

- The data created and communicated by a myriad of *different* devices and operational contexts can lead to synergy for data collection, exchange, analysis, insight, and opportunity on a new scale.
- Combining knowledge about each and every "thing" in the data from a multitude of heterogeneous sources can adapt to meet individual needs in a variety of ways, particularly when links can be established at the entity level.
- Identifying and learning about the connection between and among entities exposes new opportunities for intelligent personalization.
- Aside from the typical analytical methods used for marketing and sales, some emerging analytic techniques use the connections among "things" at the entity level that are inherent in data so that those connections can be used to naturally identify patterns and infer insights that can improve our lives.

Insightful knowledge

Predictive-analytics methods are extremely useful when a series of events can be recognized as sentinels for anticipating behavior or for scenarios requiring attention. These situations might include fraudulent activity, imminent hardware failures, or measuring credit risk for financial loans. But these analytics algorithms are still somewhat limited because they map activity to known behaviors, but don't anticipate unknown behaviors.

Discovering contexts and behaviors that lend insight to action beyond the typical prediction scenario can lead to understanding potential connections across a broad spectrum of events, and even to differentiating between *coincidences* versus *correlated incidences*.

Here are some examples:

- **Automobile safety**. In a predictive environment, air pressure sensors in a car's tires continuously monitor whether the pressure levels meet a safety standard. A rapid decrease in air pressure could indicate an imminent tire deflation, which would alert the driver to have the tire pressure checked. Alternatively, sensor readings from many cars' tires — collected with driver demographics and geographic road conditions — might be analyzed to reveal that tires from the same manufacturer, but installed on different automobile makes and models, are failing under similar driving circumstances when driven by a particular class of driver, within specific geographic regions, times of day, route types, and under similar traffic conditions. With this insightful knowledge, the automobile manufacturer could investigate the quality criteria for the tire manufacturer and possibly issue a preemptive safety recall.
- **Complex civic management**. Managing services within a municipality involves overseeing diverse resources such as waste management, road maintenance, traffic management, transportation, police, etc. Predictive methods might provide guidance for general planning, such as how much road salt to purchase. Data combined from motion sensors, power lines, mobile telephony, social media, traffic cameras, weather predictions, and transcripts of 911 calls might lead to the stationing of emergency vehicles in locations that are discovered (or anticipated) to be affected by violent weather such as hurricanes and tornados.
- **Quality health care**. Prediction is common in the health care arena. Comparing vital-sign measurements to baseline expectations can identify health risks such as high blood pressure,

which can signal heart disease and strokes. However, adding visibility into data that might traditionally be obscured from the health provider (such as detailed demographics and psychographics, individual lifestyle data, and geographic location) and combining that data with collections of clinical health information from millions of interconnected scanners, sensors, and other diagnostic devices, can expose associations and patterns that act as sentinels of potential risks or improvements. Analyzing the results of combining these data streams can help providers diagnose health problems; continuously monitored data can identify emerging health issues such as concentrations of incidence of viral infections or certain unexpected concentrations of incidence of cancer. That data might trigger investigations to identify and address root causes.

There are many potential applications. The growing number of devices generating data has created a need for processing and filtering massive data sets. More critically, the complex interdependence among these devices requires algorithms that can filter the noise from the signal and automatically select the relevant data. This drives the potential for identifying when patterns are converging, or inferring beneficial insights from new emergent signals. This approach to knowledge discovery can enable senior managers at massive corporations to make profitable decisions, as well as allow individuals to increase their personal comfort as they relax at the end of a tough day.

Balancing generalized and localized intelligence

The above examples all share key characteristics that contribute to discovering insight.

- A rampant increase in the operational deployment of computing devices that generate different kinds of data (i.e., structured data from machine sensors, text data from social media channels, streaming video from cameras, or radiology scans from medical devices)

- A combination of captured history, human expertise, and discovered interdependence among devices beyond their original purposes

- Volumes of data generated far beyond what can be naturally comprehended by a human

- Patterns of relevant knowledge may be localized within a specific context (such as a particular type of machine sensor, focused on a specific operational activity, or within a limited geographic region), so absorbing all of the data into a single computing resource for discovery is always necessary.

Algorithms deployed using the computing power of intelligent components across this mesh network can be combined in different ways at different times to yield both localized intelligence and generalized knowledge to benefit a variety of parties.

For example, "smart dust" sensors can be mixed with fertilizer that farmers spread across their fields. Each device is smart enough to measure and immediately communicate soil temperature, chemical balances, humidity, density, salinity, etc. The mesh of sensors can tell the farmer whether he should water more frequently, add additional fertilizer or other chemicals, treat pests, etc. Each of these operational tasks is relevant within each farmer's domain. At the same time, the breadth of information the numerous sensor networks collect can allow fertilizer manufacturers to assess the success of the chemical mix within the product — the degree to which it increased the key characteristics of the soil to enhance crop growth, inhibited pests, or maintained resistance to certain types of plant diseases.

Simultaneously, a mesh of sensors can reveal what was communicated to a group of farmers and how each (or all) reacted to the provided guidance, when, and at what level of success. The fertilizer manufacturer can then improve or customize the product as well as drive education and training for farmers so that they improve their use of the product.

Drivers for reconsidering analytical methods

Despite the logical assumption that benefitting from broad interconnectivity must be tightly coupled to centralized big-data analytics, potential sweet spots for discovered knowledge can be manifested by taking advantage of the synergy of local mesh-connected computational units.

Following are some key drivers for reconsidering the methodology that facilitates the discovery of insightful knowledge:

- **Breadth of the mesh**: The IoT implies a broad lattice of devices with varying degrees of connectivity, computational power, and "persistence of memory." The explosion of devices generating data leads to the next driver.
- **Data acceleration**: The rate at which data is created will continue to accelerate, as will the volumes of generated data. Big-data platforms can handle the volume, but the bottleneck remains the bandwidth for communicating massive amounts of data.
- **Bandwidth eclipse**: The need to share *information* will exceed the bandwidth capacity for *data*. Sharing relevant knowledge discovered locally can reduce the volume of data while increasing the signal-to-noise ratio. s
- **Embedded intelligence**: Devices in this mesh combine sensors, storage, communication, and programmatic analysis that can contribute to knowledge discovery with both local and global value.
- **Context**: Local context will influence local decisions, while summarized and shared knowledge can influence global decisions.
- **Intuition horizon**: Blending these factors enables more effective inferring. Data generated as a byproduct of a specific purpose can be employed for inferring knowledge about numerous aligned contexts that benefit hierarchies of interested stakeholder communities.

The insightful fog and its implications

Massive interconnected environments such as the IoT are not content pumps attached to a centralized repository — even if that repository is cloud-based. Instead, since we can perform analyses at various orders of computational organization across this lattice, we might see opportunities for generating insight that can take advantage of intersecting and overlapping collections of intelligent components. Discoveries and conclusions derived within one domain of this insightful fog can be diffused across other domains to inspire the determination of choices, decisions, and guidance that, in turn, can lead to increasingly desirable outcomes.

Four key implications of the insightful fog will significantly alter the way people think about benefiting from analytics:

- **Bridging the localized/centralized domains**: The breadth of interconnectivity suggests a need for balancing aspects of knowledge discovery that can be implemented within a constrained domain using limited amounts of shared data with those analytical methods that must accumulate and centralize data from many points. Knowing when distinguishing partial analyses have local benefit and when their results can contribute to a more global business context is valuable.
- **Diffusion**: Using analytics for discovering and deriving influential insight and guidance must be easily deployed, and the corresponding results must be easily and efficiently transmitted across the lattice of interconnected devices.
- **Hierarchical influence**: Relevant information accumulated at key points of centralization will influence a broader realm of business contexts.
- **Intersection**: The intersection of different mesh networks can leverage operational measures and metrics to derive insights beyond the operational environment.

Our previous example of smart dust demonstrates each of these implications. Locally analyzed conclusions about soil properties help the farmer, while the combined results of many farms help the manufacturer find opportunities for improving the product. The partial results calculated within each farmer's domain could be summarized when transmitted back to the manufacturer, thereby reducing the bandwidth demand while preserving the power required for communication. The local results influence the individual farmer, but the collected results can contribute to the collective benefit (for both the manufacturer and the community of farmers). Data from different sources can be used in the centralized analyses, including farmer demographics, purchasing behaviors, and other relevant sources.

The model for cognition

All of the examples provided thus far share some common characteristics, and we can abstract a model that can be leveraged to evolve an ecosystem that integrates data sharing, analysis, and, importantly, a closed feedback loop that enables learning. This model encompasses what is envisioned by the IoT and goes beyond it by adding inferring and intelligence at all levels of the hierarchy. The model consists of:

- **Smart devices** that generate data used to derive knowledge about the device's operation, as well as information about when and how the device is used.
- **Collaborative connections** among both homogeneous and heterogeneous devices blend data generation, data exchange, and analysis associated with deploying and monitoring operational business objectives.
- **Actors** within the ecosystem include all automated systems and humans that will interact with one or more of the devices over the life cycle of one or more business processes.
- **Decision process**. The benefits of the insights are related to different business contexts, ranging from immediate interactions to broader systemic decisions spanning the breadth of the business-process life cycle.
- **Radii of insight and influence** both define and derive knowledge from context, including:
    - The immediate user experience associated with direct interactions can be enhanced by adaptive techniques for collecting and sharing relevant information associated with different contexts.
    - A network of "close-by" influencers can be probed to identify alternatives not limited to the users' own behaviors.
    - Global opportunities can be communicated to the process owner in relation to the ways the various actors interact with different contexts of the business process.

Blending the physical lattice of connected devices with local analytical insight that includes associative memory and cognitive analytics can create scenarios in which "sense" can be made of data in relation to the necessary radius of insight and influence.

Associative memory and cognitive analytics

Typically, three modes of analytics are used in industry.

- Traditional reporting and ad hoc analysis organizes structured data pulled from one or more sources in a way that is more easily reported and queried. The ability to "slice and dice" gives analysts visibility into the past so that they can understand root causes.

- Descriptive analysis establishes algorithms for assessing what has happened, describing why it has happened, and using that description or classification for rudimentary prediction. For example, clustering, classification, and segmentation techniques use algorithms to infer similarities among entities in a data set, and then use those inferences to predict behavior of other similar entities:

  - Behavior prediction: "A series of transactions like these are potentially indicative of money laundering."

  - Recommendation engines: "People like you have bought these types of products."

  These techniques seek to draw conclusions about what has happened in order to predict future behaviors, but are largely limited to supervised learning, i.e., looking for the combination of variables that most closely matches some desired model.

  While these traditional algorithmic approaches add value, they can neither capture all aspects of the decision processes nor always provide reasonable transparency in discerning the analytic results. These analytics methods depend on models that, by definition, reduce the information available for analysis, rather than expand the knowledge horizon and enable incremental learning.

- The connected world — telephone networks, highways, utility grids, and the internet — defines how its components interact with each other. Individuals connect in different ways, and established connections reflect relationships among different entities. The scope of these relationships expands beyond individuals to include relationships between individuals, organizations, locations, items — a virtual plethora of things that can be related together. Often-used contexts and descriptions embody the semantic meaning.

  Each of the following pieces of information represents a semantic connection of a particular type between two entities that can be represented many ways, both as structured and unstructured data. Incorporating text analytics with an infrastructure for documenting connectivity — and correspondingly, the nature and other characteristics of the connections — yields the entities and their links.

- Matthew works for ABC Corporation.
- Joe married Elaine.
- Jason lives in Seattle.
- Mary and Jennifer both went to Michigan State.

In isolation, these content tidbits are minimally interesting. However, in aggregate over time, patterns of association emerge. Mary and Jennifer may both be alumnae of the same school, and may participate in activities that other Michigan State alumni attend. Statistical dependencies such as these can be used for understanding known behavior patterns and for their anticipatory value. Continuing the example, after some time we can anticipate that other individuals graduating from Michigan State may show up where Mary and Jennifer hang out.

The third type of analysis, derived from the connected world, organizes information so that it leads to inferring concepts that were not defined from the start.

One analytics algorithm that can capture these connections uses "associative memory." This method employs a graph approach for capturing the breadth of connections and their frequencies, as well as calculating the degree of correlation between different semantic concepts within the graph. Associative memory, which is intended to imitate a human's memory, captures and reasons from links and associations among different types of entities. The links and associations reveal common patterns that can be evaluated and analyzed for emergent patterns. The emergent patterns enable unsupervised learning and cognitive analytics, in which the strength, closeness, and frequency of identified connections pave the way for drawing new conclusions and insights.

Sense-making drives beneficial decisions

The future of analytics demands the ability to “make sense” out of all the data being generated. As the amount explodes, the ability to pinpoint relevance becomes increasingly difficult, driving the need for ever finer precision. Model-based methods are good at matching against known patterns, but often need directed input to grow their pattern inventory. On the other hand, within a cognitive analytics environment, sense-making allows predictive models to grow by absorbing and learning from the actual experience in collected data. They can then anticipate how individuals and relationships are changing and the corresponding implications for the way we live and work.

For example, an airline knows which passengers are aboard each flight and can track the result of each interaction during a flight between the flight attendants and the passengers: food and drink preferences, when a passenger used the television console, and what channels or movies that passenger watched, etc. Over time, an associative model can evolve that links individual passengers with specific preferences. Additional links such as the flight times, directions, durations, and destinations can provide information about *types* of passengers and other variables that might affect their behaviors and preferences. The airline can then determine such things as which items are stocked and which entertainment packages to offer so that it can improve the overall passenger experience in ways that appear to be personalized for each passenger. This appearance of individualized care can improve passenger experience and increases passenger loyalty so that both the passenger and the airline benefit.

Insight and cognitive analytics and true personalization

Existing methods for personalization are good examples of applications that rely on logic-based or pattern-based techniques. While logic engines are good for recognizing patterns they are programmed to find, they can't understand changing priorities based on real-time context. So, they have limited use for personalization.

The collective knowledge generated from the many devices and systems people use or interact with on a daily basis offers opportunities for effecting personalization that improves individual lives. Mobile devices, published artifacts (e.g., calendars), automobile sensors, fitness devices, GPS devices, and RFID card readers all shed light on what people do as well as when and where they do it.

Personalization is a fertile area for improvement. Fusing intelligent assistants composed of engines using cognitive computing can augment the traditional logic-based approach. By unifying the accumulated data at the individual level, adaptive learning can identify emerging patterns and personalize recommendations based on anticipatory thinking at the individual level in relation to context: at home, in the office, in the car, at a place of entertainment, or any combination thereof.

Personalization can also influence and improve health and wellness. Collecting knowledge about the things in data from the many devices and machines that generate it — portable fitness devices, exercise machines, pedometers, heartbeat and pulse monitors, etc. — and blending it with data from health reports, medical information, medical histories, sleep schedules, TV watching times, and work schedules, can help understand the relationship between stress events, increased blood pressure and lack of sleep, or other imminent potential risks. Collecting information and using tools to make sense of it can provide a person with recommendations for improving his or her day and maintaining optimal health.

Enabling the insightful fog

Making the insightful fog a reality means coupling cognitive analytics techniques tightly with the naturally evolving organization of the types of device meshes implied by concepts like the IoT. This suggests innovative tools and technologies that overcome some key technical hurdles. The tools require finesse and elegance rather than brute force:

- **Big-data scalability.** The rates at which data volumes grow will continue to accelerate, implying the need for intelligently absorbing huge amounts of structured and unstructured data as a prelude to attaching context, relevance, and discernment of connectivity about every "thing" from a wide variety of source streams in real time.
- **Flexible information/knowledge organization.** The ability to assess statistical relevance and correlation embedded within networks of connected entities requires graph-based data organization in which the links are attributed with semantic meanings inferred from explicit derivations and context. Associative memories are examples of data organization that capture what is known and enable inferring what is currently unknown.
- **Sense-making.** Algorithmic savviness, not data mining, is necessary for making sense of the data and the contexts — who the entities are, how they are related, where the similarities are, when those similarities are relevant, consideration of similar scenarios and their outcomes, connecting performance measures to inferring models, and as a result, deriving insight.
- **Incremental learning.** The means by which the cognitive models become more refined over time as more information is collected and feedback and outcomes from user interactions are captured in the knowledge layer
- **Personal control.** Enabling individuals to exercise control over what data sets are used and who has the right to use the data and the conclusions

While traditional reporting, ad hoc analysis, and even many data-mining techniques provide some of these required characteristics, few, if any, are immediately adaptable to meet all of these needs. More likely to fit the bill are approaches that support cognitive analytics and embody memory-based reasoning, and associative memories using high-performance computing techniques for managing both the space and computation needs. As cognitive-computing tools are blended within the massively connected IoT, we will begin to experience the benefits of the insightful fog.

Key takeaways

- The data created and communicated by a myriad of different devices and operational contexts can lead to synergy for data collection, exchange, analysis, insight, and opportunity on a larger scale than ever before.
- Combining knowledge about each and every "thing" in the data from a multitude of heterogeneous sources can feed different types of analytics algorithms for discovery, exploration and prediction, and even what might be perceived to be a rudimentary awareness of context can adapt to meet individual needs in a variety of ways.
- Identifying and learning about the connection between and among entities exposes new opportunities for intelligent personalization.
- Aside from the typical analytical methods used to drive corporate marketing and sales, there are emerging analytic techniques that blend massive connectivity with ubiquitous data generation to naturally absorb and consolidate content in ways that can make sense of each individual's personal context, anticipate that person's needs and desires, and communicate that insight with the right devices at the right time to make that person's life better.

About David Loshin

David Loshin is a Gigaom Research analyst, who is recognized worldwide as an information-management industry thought leader. He has dedicated many years to popularizing the best practices for business intelligence, data governance, performance computing, master-data management, predictive analytics, and data quality.

About Gigaom Research

Gigaom Research gives you insider access to expert industry insights on emerging markets. Focused on delivering highly relevant and timely research to the people who need it most, our analysis, reports, and original research come from the most respected voices in the industry. Whether you're beginning to learn about a new market or are an industry insider, Gigaom Research addresses the need for relevant, illuminating insights into the industry's most dynamic markets.

Visit us at: research.gigaom.com.