Solving the Jigsaw Puzzle: An Analytics Framework for Context Awareness in the Internet of Things

by Anurag Agarwal, Ramakrishna Govindu, Sunita Lodwig, and Fawn T. Ngo

The Internet of Things (IoT) is a paradigm wherein ubiquitous, context-aware devices equipped to identify, sense, and process data communicate over the Internet to accomplish some intended objective(s). The Internet of Things Global Standards Initiative (IoT-GSI) defines IoT as:

> A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.[1]

Several technologies that have evolved independently over the last two decades (e.g., sensor networks, RFIDs, microchips, intelligent agents, the Internet, mobile computing) are now converging to enable the IoT paradigm. For example, it used to be the case that devices equipped with RFID chips could only be employed locally due to their limited near-field communication abilities. With advances in communication and Internet technologies, the same devices have now expanded their geographic reach globally. The development and adoption of these technologies have been so rapid that the number of active IoT devices themselves are estimated to reach 50 billion by 2020, up from an estimated 25 billion in 2015.[2] According to one estimate,[3] 40% of all data generated by 2020 will come from interconnected sensors and devices. Thanks to IoT and other related technologies, massive quantities of both structured and unstructured data are being generated on a continuous basis at a phenomenal rate, leading to a big data revolution, which in turn is providing opportunities for big data analytics.

The increasing business and social impact of the IoT paradigm is prompting researchers and practitioners to bring together and integrate more and more "things," resulting in a vast network of autonomous, self-organizing, and intelligent devices.[4] Consequently, the IoT holds the promise of creating a global network, supporting ubiquitous computing and context awareness among devices.[5] Ubiquitous computing and context awareness have become critical requirements of ambient intelligence, one of the key promises of the IoT.[6] The IoT helps embed technologies into everyday products/devices, such as audio/video receivers, wristwatches, smoke detectors, and home appliances, which not only enables them to communicate online, but also to receive and process data and information from other devices in a dynamic fashion, in real time. Thus, the real revolution of IoT goes beyond embedding a sensor and sending signals over the Internet to developing a 360-degree context awareness by analyzing data from multiple sensors or sources using complex advanced algorithms, in real time, for improved decision making.

For example, RFID technologies previously enabled organizations to track the location of products through the supply chain at various destination points, such as the warehouse or the retail outlet. With IoT, though, products can now be tracked while in motion and in real time, resulting in dynamic tracking and improved inventory management. As another example, with traditional RFID technologies, a machine stationed at a location could be monitored through sensors and a potential maintenance issue communicated to manufacturers to preempt failures (e.g., elevator manufacturers getting notification alerts on impending failures so that preventive maintenance can be scheduled, thus improving customer safety). With the evolution to IoT, this advanced monitoring feature can now be extended to machines that are mobile, such as engines in trucks, ships, and planes. Such intelligent behavior requires IoT devices to be aware of their context or surroundings.

CONTEXT AWARENESS

The IoT literature provides many definitions of context awareness.[7] For the purpose of this article, we define context awareness as the information necessary and sufficient to perform the intended function of the device effectively and efficiently. Typically, but not always, the context can be ascertained comprehensively by answers to some or all of what we like to call "the four W's": Where, When, Who, and What. A simple IoT device with limited functionality may only need to answer

Parameter	Description	Examples of Data
Where	Location of events	The latitude, longitude, and altitude of the IoT device
When	Time of events	The timestamp of all the signals/data received by the IoT device
Who	Person(s) or object(s) of interest associated with the events	Identification data for the object or person (e.g., biometric readings for a person or the IP address of another IoT device in network)
What	Various measures of interest besides location, time, and objects	Temperature, pressure, speed, level, weight, duration, sentiment, demographics, distance

Table 1 — Parameters defining context.

one or two W's, while a more complex IoT device may need answers to all four, and perhaps even to additional questions such as How, Why, Which, How Much, and so on. See Table 1 for some examples of contextual data needed for the parameters defining the context awareness of an IoT device.

AN ANALYTICS FRAMEWORK FOR CONTEXT AWARENESS

Figure 1 illustrates our proposed analytics framework for a typical IoT device. As shown in Figure 1a, within one's environment of interest, there is typically a network of IoT and non-IoT devices interacting with each other over the Internet to accomplish certain objective(s) for an entity (person or organization). Each IoT device has three main components: a set of sensors, a context awareness engine, and a solution engine (see Figure 1b).

Through the set of sensors, the necessary data is collected. There are typically many different types of sensors that generate data for a specific application, ranging from a simple thermometer for measuring the room temperature to a radio telescope for sensing radio waves from faraway galaxies. Sensors can be designed to capture data for different environmental characteristics, such as latitude and longitude, time, temperature and other weather-related characteristics, the presence of objects, motion detection, speed, and so on. These sensors provide data to answer questions for the four W's (see Table 1). The context awareness engine then analyzes the data to model the context. Taking into account the objective of the application, the solution engine uses the context information and determines the best possible solution, which could be some action. The action might simply be a notification to an appropriate person or device or it could be a corrective action that alters the state of some object of interest in the environment.

The context awareness engine of the IoT device has two main components — the data representation engine and the context modeling and analytics engine (see Figure 1c). The data representation engine acquires the sensor data from various heterogeneous sources and represents and stores that data in the appropriate formats. The context analytics and modeling engine combines data in heterogeneous formats and applies suitable algorithms to model the context. The output of the context awareness engine is the context, which is used by the solution engine. The context is also stored within the context awareness engine using the data representation engine, to enable dynamic update of the context and further refinement of context if needed.

Modeling a 360-degree view of the context is like solving a jigsaw puzzle, where each puzzle piece comes from a different source in the form of data.

As shown in Figure 1c, there is bidirectional communication between the data representation engine and the context analytics and modeling engine, because the way the data is represented influences the modeling algorithms and vice versa. Depending on the nature of the input data, it can be stored in a variety of different formats within the data representation engine, such as the standard two-dimensional relational table, as a class of objects with properties and methods, or as one of the NoSQL database formats (key-value pairs, column family database, graph database, etc.). For instance, for high-velocity, single-column data, the key-value pair format would be the most appropriate. For low-volume, multi-attribute, structured data with no built-in methods, the tabular format would be the most appropriate.

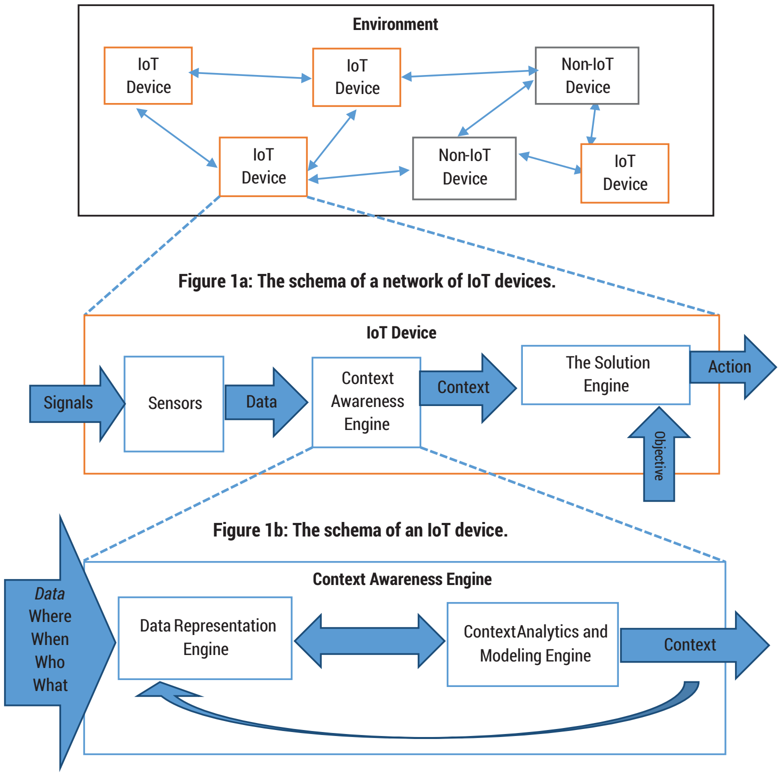

Figure 1a: The schema of a network of IoT devices.

Figure 1b: The schema of an IoT device.

Figure 1c: The schema of the context awareness engine.

Figure 1 — An analytics framework for context awareness of an IoT device.

For low-volume, multi-column, structured data with built-in methods, the object-oriented format would be the most appropriate.

Modeling a 360-degree view of the context is like solving a jigsaw puzzle, where each puzzle piece comes from a different source in the form of data. The completed puzzle may look completely different from any of the individual pieces. This puzzle solving happens within the context analytics and modeling engine by using the data collected to answer the four W's (in Table 1). Depending on the nature of the data and the application, it may employ a variety of analytics techniques, ranging from simple arithmetic operations to complex pattern recognition (facial, voice, image, or text). Other techniques might include data mining methods such as classification, clustering, association rule mining, and so forth. This context modeling is usually performed in real time, since the actions taken by an IoT device are typically time-sensitive.

To illustrate our proposed analytics framework for context awareness, let's look at a practical example of an IoT system that helps law enforcement agents keep the roads free of drunk drivers.

KEEPING IMPAIRED DRIVERS OFF THE ROAD: A PRACTICAL EXAMPLE

A breath-alcohol analyzer (BAA) is an IoT device used to monitor the breath-alcohol level of DUI offenders during their probationary period. DUI offenders are required to install a BAA device in their vehicle and provide their breath sample periodically while driving. The BAA analyzes the breath sample for its alcohol level, and if the level exceeds the legal limit, the device alerts the appropriate law enforcement agency. However, this approach is not foolproof. Drivers may potentially cheat the system by having a sober person provide the breath sample on their behalf, or they might adopt measures to manipulate their breath sample using some suppressants. Furthermore, DUI offenders may be risky drivers even while within the legal alcohol level.

To better assess the riskiness of an offender's driving, a more comprehensive context awareness is needed, one that builds from multiple sources of data. For example, in addition to breath-alcohol level, the smoothness of the drive itself can be assessed using appropriate sensors, such as a telematics device and the vehicle's curb camera. Thus, the functionality of the BAA as an IoT device can be enhanced by enabling it to gather data from multiple sources, thus developing a more comprehensive context by applying analytics on these multiple sources of data. Figure 2 shows the operational schema of an environment in which the BAA interacts with other IoT and non-IoT devices.

The BAA installed in the DUI offender's vehicle receives input from the driver (biometric readings, breath sample, voice sample), a telematics device (speed, timestamp), a curb camera (distance from curb, timestamp), satellite (longitude and latitude), and the law enforcement agency (voice sample and other biometric patterns of the offender). In addition to these external sources, the BAA itself may contain some internal sensors that provide data for the context awareness engine. Table 2 shows the details of the data collected for the BAA to answer the four W's.

Data Representation for the BAA

Since the data sources for the BAA are quite heterogeneous, the appropriate data representation formats also differ, depending on the data characteristics. In our example, the telematics data and the curb camera data are considered high-velocity, since we are collecting the data every second. It is best to represent this data as key-value pairs. For the telematics device data, the key will be the timestamp, and the value will be the vehicle speed. For the curb camera data, the key will again be the timestamp, and the value will be distance from the curb. The offender data from the law enforcement agency is very low-volume, multi-attribute with no in-built methods, and therefore a tabular data format is the most appropriate for this data. The data for each trip and for each breath analysis is also low-volume, multi-attribute, structured data with in-built methods, and thus the most appropriate format would be an object-oriented class.

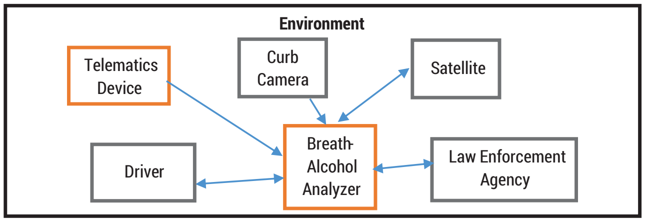

Figure 2 — The schema of the breath analyzer IoT device and its environment.

Figure 3 shows the data representation for each type of data discussed above. For the class diagram, we identify two classes: Trip and BreathTest. The Trip class captures data for each trip, such as the date, time, driver's biometric reading, and origin and destination coordinates. Its methods confirm the identity of the offender. The BreathTest class captures the results of the breath test multiple times for each trip. Its methods are designed to confirm the validity of tests and whether there was a violation. The data from the telematics device and the curb camera is shown as key-value pairs. This data will be used by the context modeler to determine if the offender is driving in a jerky manner or weaving within his or her lane.

Parameter	Description	Example of Data	Source
Where	Location	The longitude and latitude of the vehicle	Satellite
When	Time	The timestamp of the breath analysis Timestamp for speed Timestamp for curb distance Time of request for breath Time of breath input	BAA Telematics device Curb camera BAA BAA
Who	Person(s) or object(s) of interest	Offender's identity (through some biometric sensing)	Driver
What	Measures of interest	Alcohol level in the breath Speed of vehicle Curb distance Biometric data for the offender	BAA Telematics device Curb camera Law enforcement agency

Table 2 — Parameters defining context for a breath analyzer.

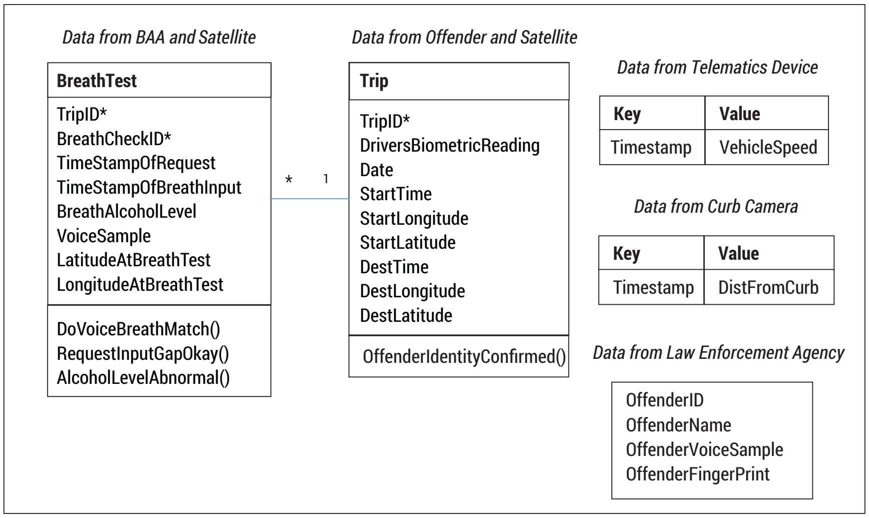

Figure 3 — Data representation for BAA data.

Context Analytics and Modeling for the BAA

The telematics data on vehicle speed, which is captured every second, is used to evaluate whether the drive is smooth or swervy. An algorithm calculates the acceleration and deceleration every second and determines if the variations are above normal in level and frequency. If they are, then the BAA device becomes aware that the drive is abnormal. Similarly, another algorithm calculates differences in the distance from the curb every second. If these differences are beyond a certain threshold and in both directions, and sustained over a period time, then the BAA device becomes aware that the offender is swerving.

> **Multidisciplinary skills will be needed in order to develop and deploy context-aware engines for IoT-driven applications.**

When the BAA becomes context aware of an abnormal drive, it may send a notification to the law enforcement agency and a new request for a breath test may be triggered to the driver. The driver must provide a breath sample within a predetermined duration of time; if the gap between the request and the receipt of the breath sample exceeds that duration, the RequestInputGapOkay() method (see Figure 3) sends an alert to law enforcement. If the breath-alcohol level is above a pre-set threshold, the AlcoholLevelAbnormal() method sends another alert. While providing the breath sample, the offender is also required to make a sound. The sound characteristics are then compared with the offender's voice sample that the law enforcement agency has on file. If the voice does not match, the method DoVoiceBreathMatch() sends an alert to the agency. This check is used to ensure that the offender is not using another person's breath to circumvent the test. If he or she is, the law enforcement agency will receive this notification and charge the offender with the violation. If the voice matches, then no alert is sent.

POTENTIAL CHALLENGES

There are a few challenges associated with our framework. As our example shows, the need for a comprehensive context requires heterogeneous data collection and storage. Further, the analytics is being performed in real time on disparate data representations. Multidisciplinary skills will be needed in order to develop and deploy context-aware engines for IoT-driven applications.

CONCLUSION

The analytics framework we have proposed in this article is designed to handle multiple and heterogeneous sources of data to develop accurate models for 360-degree context awareness for IoT-based applications. By capturing and analyzing the details of the context, the IoT device enhances context awareness, which in turn leads to improvements in the operational effectiveness of an IoT-based application. For instance, the BAA example illustrates how law enforcement agencies could benefit as enhanced context awareness provides more accurate and comprehensive monitoring of DUI offenders on probation. This, in turn, reduces the time, cost, and effort spent to achieve the objective of keeping unsafe drivers off the road. We believe application of the proposed framework will result in a much richer suite of data points for many IoT devices, thereby providing more value to businesses and other organizations in their decision making.

ENDNOTES

[1] "Overview of the Internet of Things." ITU-T (www.itu.int/ITU-T/recommendations/rec.aspx?rec=y.2060).

[2] Evans, Dave. "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything." Cisco Internet Business Solutions Group (IBSG), April 2011 (www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf).

[3] Taslim, Allison. "12 Facts You Need to Know About the Internet of Things." Servicemax (blog), 17 July 2015 (http://blog.servicemax.com/12-facts-you-need-to-know-about-the-internet-of-things).

[4] Whitmore, Andrew, Anurag Agarwal, and Li Da Xu. "The Internet of Things — A Survey of Topics and Trends." *Information Systems Frontiers*, Vol. 17, No. 2, April 2015 (http://link.springer.com/article/10.1007%2Fs10796-014-9489-2).

[5] Hong, Jong-Yi, Eui-Ho Suh, and Sung-Jin Kim. "Context-Aware Systems: A Literature Review and Classification." *Expert Systems with Applications*, Vol. 36, No. 4, May 2009 (www.sciencedirect.com/science/article/pii/S0957417408007574).

[6] Bellavista, Paolo, et al. "A Survey of Context Data Distribution for Mobile Ubiquitous Systems." *ACM Computing Surveys*, Vol. 44, No. 4, August 2012 (http://dl.acm.org/citation.cfm?id=2333119).

[7] Perera, Charith, et al. "Context Aware Computing for The Internet of Things: A Survey." *IEEE Communications Surveys & Tutorials*, Vol. 16, No. 1, 2014 (http://arxiv.org/pdf/1305.0982).

Anurag Agarwal is a faculty member in the College of Business at the University of South Florida Sarasota-Manatee. He has a PhD in information systems from The Ohio State University and an MBA from the University of Wisconsin. His teaching interests lie broadly in the field of analytics, data mining, and information systems. Dr. Agarwal has taught courses on big data analytics, business analytics, business statistics, operations management, information systems in organizations, systems analysis and design, database systems, programming languages, and decision support systems. His primary research interests are in predictive and prescriptive analytics. He has published in journals such as INFORMS Journal on Computing, Naval Research Logistics, European Journal of Operational Research, Omega: The International Journal of Management Science, Computers and Operations Research, Annals of Operations Research, Information Systems Horizons, *and* Information Technology and Management. *Dr. Agarwal serves on the editorial board of* Information Technology and Management *and* Enterprise Information Systems. *He can be reached at agarwala@sar.usf.edu.*

Ramakrishna Govindu is a faculty member in the College of Business at the University of South Florida Sarasota-Manatee. He has a PhD in industrial engineering from Wayne State University and an MTech (Honors) in quality, reliability, and operations research from the Indian Statistical Institute, Calcutta, India. Dr. Govindu has taught courses in applied statistics, operations and supply chain management, Lean Six Sigma, and operations research and decision sciences. His primary research areas include predictive and prescriptive analytics, supply chain modeling and management, and multi-agent systems. Dr. Govindu has published in Computers & Industrial Engineering, International Journal of Modelling and Simulation, Wiley's Handbook on Technology Management, *and* American Journal of Criminal Justice. *He has about two decades of industry experience in the US and India that includes more than a decade of consulting work. Dr. Govindru has executed, led, and managed projects for organizations such as Caterpillar, Ford, Visteon, GE, and others. He has implemented solutions involving supply chain analytics, predictive and prescriptive analytics solutions for decision support, business reengineering, and Lean Six Sigma that saved millions of dollars for clients in the manufacturing, service, and healthcare sectors. Dr. Govindu is a member of several professional associations. He can be reached at rgovindu@sar.usf.edu.*

Sunita Lodwig is a faculty member in the College of Business at the University of South Florida Sarasota-Manatee. She has a PhD in theoretical plasma physics from the Indian Institute of Technology, Delhi, India. Dr. Lodwig has served as a faculty member at Northern Illinois University and as a Research Associate at Washington State University. She currently teaches a range of courses in information technology, such as UML, program design, programming languages, security, data structures, and the capstone course. Dr. Lodwig also served in management positions at AT&T Bell Labs and held various roles in Motorola's Global Software Group, where she was involved in technical marketing, project/product management, and globalization issues. She can be reached at slodwig2@sar.usf.edu.

Fawn T. Ngo is a faculty member in the College of Arts and Sciences at the University of South Florida Sarasota-Manatee. She has a PhD in criminology and criminal justice from the University of Maryland and an MS in criminal justice from the California State University, Long Beach. Dr. Ngo's teaching interests include crime analysis for problem solvers, applied statistics, research methods, and program evaluation in criminal justice. Her primary research interests include predictive analytic applications in criminology and criminal justice, survey methods, quantitative methods, and evaluative research. Dr. Ngo's most recent publication involves a comparative study of four statistical techniques — logistic regression, classification and regression tree, chi-squared automatic interaction detection, and neural networks — for their utility in predicting inmate misconduct. She can be reached at fawnngo@sar.usf.edu.