Chapter 41
Towards Automation of IoT Analytics:
An Ontology-Driven Approach

Sounak Dey
TCS Research and Innovation, India

Arijit Mukherjee
TCS Research and Innovation, India

ABSTRACT

The rapid growth in the number of sensors deployed at various scenarios and domains has resulted in the demand of smart applications and services which can take advantage of the sensor ubiquity and the Internet of Things paradigm. However, IoT analytic applications are grossly different from typical IT applications in the sense that in case of IoT, the physical world model is absolutely essential to understand the meaning of sensor data and context. It is also unreasonable to assume that application developers will possess all the necessary skills such as signal processing, algorithms, domain and deployment infrastructures. The scenario is more complicated because of overlapping domains and variety of knowledge. Researchers have attempted to automate parts of the development process, but, the area of feature engineering for sensor signals remain relatively untouched. In this chapter, the authors discuss about the use of semantic modeling for IoT application development with respect to a framework that is capable of largely automating parts of IoT application development.

INTRODUCTION

Nearly two decades ago, in 1999, Prof. Neil Gross predicted: "*In the next century, planet Earth will don an electronic skin. It will use the internet as a scaffold and transmit its sensations*" (Gross, 1999). Seventeen years later, in 2016, we are moving towards a direction of "Internet of Everything", a step forward from the era of "Internet of Things", as number of connected and deployed objects around us are mushrooming at a very high rate. They are predicted to reach a count of 4.9 billion from today to about 25 billion by 2020 (Gartner, 2014). Each of these objects is uniquely addressable and communicates

DOI: 10.4018/978-1-5225-3422-8.ch041

with each other and outer world based on standard communication protocols (INFSO, 2008). This novel paradigm of IoT, or IoE, or more generic *Cyber Physical Systems*, is able to discover context at personal and physical level and is able to answer questions like: who is doing what, when, where; what is happening now; how one is feeling now; and even what is probably going to happen next. These powerful sets of information have high impact on individual users (for assisted living, enhanced learning etc.) as well as industry and business users (like smart transportation, manufacturing, energy and utilities etc.).

However, every good thing comes at a price. This growth in number of smart objects translates into generation of a tremendous volume of data which in turn translates into a huge requirement of data storage, and an efficient and fast communication network. Moreover, all connected objects do not send data in similar format; they are controlled by different stake holders and cater to different use cases involving variety of knowledge domains. There is heterogeneity in terms of hardware and software versions which are again controlled by multiple independent communities. To sum up, the heterogeneity present in the IoT world due to obvious reasons is a major challenge facing researchers in this domain.

The data, produced by these smart objects, are however precious. Data are often referred as "oil" for the not too distant future. But data without any follow-on analysis and some meaningful insight do not have much utility. All useful applications in IoT domain perform this analytics and help end users to visualize the situation/context represented by data in a more meaningful way, thus enabling the user to take some decisions or actions based on the situation.

In order to develop such intelligent and useful IoT applications, one should (i) identify the most appropriate data sources together with the detailed specification of the sources, i.e. the sensors, (ii) properly collect the data, (iii) identify the data format thereof and (iv) understand which part of the data is useful for the application. At the same time, most applications involve usage of multiple algorithms in a sequence to extract a useful insight from the data. To take an informed decision about the algorithms to use and the usage sequence, the application developer requires an in-depth knowledge about algorithms, their usage, complexities, features, side-effects etc. Real-time or near-real-time execution is another common feature for an IoT application, which requires detailed knowledge about the resources involved e.g. edge devices, cloud platforms, gateways, network etc. On the other hand, to build a successful application, the knowledge about the particular domain is as important as the technical aspects. For example, an application in the healthcare domain, targeted towards the smart health paradigm, may require the developer to know certain aspects of the healthcare domain apart from the technical knowledge about processing sensor signals. Needless to say that, it is a tremendous ask for a programmer or developer.

Considering the enormity in the number of connected objects, the volume, velocity and variety of the resultant data, number of algorithms, number of use cases to solve, number of experts from different knowledge domain to be involved, and above all enormous heterogeneity involved at each level, the whole domain of IoT application development cannot be handled manually and thus requirement for machine automation at many levels becomes obvious. In other words, there is a very obvious requirement of modeling the physical world and the knowledge associated with it using semantic techniques.

To enable machines (which can compute) to help us, the knowledge involved at different stages must be captured in a machine-readable and machine-understandable format, and it has already been observed by researchers that Ontology is a useful instrument in this regard. Ontology can capture 'facts' specific to a domain and aims to replace a human domain expert by executing reasoning algorithms on top of the ontology elements. In other words, ontology is a knowledge and information store, represented as a connected graph, to which both machines and human can ask questions, like students normally do to a teacher or patients do to a doctor; more generally, it is similar to a situation where a non-expert asks

Related Content

Clustering Model for Microblogging Sites using Dimension Reduction Techniques
Soumi Dutta, Nilan Saha, Asit Kumar Dasand Saptarshi Ghosh (2019). *International Journal of Information System Modeling and Design (pp. 26-45).*
www.irma-international.org/article/clustering-model-for-microblogging-sites-using-dimension-reduction-techniques/231579

A Comparative Analysis of Software Engineering Approaches for Sequence Analysis
Muneer Ahmad, Low Tang Jungand Noor Zaman (2013). *Software Development Techniques for Constructive Information Systems Design (pp. 285-295).*
www.irma-international.org/chapter/comparative-analysis-software-engineering-approaches/75751

An Efficient Trajectory Representative Generation Moving Web-Based Data Prediction Using Different Clustering Algorithms
Vishnu Kumar Mishra, Megha Mishra, Bhupesh Kumar Dewanganand Tanupriya Choudhury (2022). *International Journal of Information System Modeling and Design (pp. 1-16).*
www.irma-international.org/article/an-efficient-trajectory-representative-generation-moving-web-based-data-prediction-using-different-clustering-algorithms/316132

E-Monitoring System: Analyzing the Benefits and Effects of an E-Monitoring System in the Banks of Kerala
Bharathiveena V.and Janardhanan Pillai (2022). *International Journal of Software Innovation (pp. 1-19).*
www.irma-international.org/article/e-monitoring-system/311507

Current Challenges in Intrusion Detection Systems
H. Gunes Kayacikand A. Nur Zincir-Heywood (2009). *Software Applications: Concepts, Methodologies, Tools, and Applications (pp. 458-466).*
www.irma-international.org/chapter/current-challenges-intrusion-detection-systems/29403