Data Processing and Semantics for Advanced Internet of Things (IoT) Applications: modeling, annotation, integration, and perception

Pramod Anantharam
Kno.e.sis - Ohio Center of Excellence in Knowledge-enabled Computing
Wright State University
Dayton, Ohio, USA
pramod@knoesis.org

Payam Barnaghi
Centre for Communication Systems Research
University of Surrey
Guildford, UK
p.barnaghi@surrey.ac.uk

Amit Sheth
Kno.e.sis - Ohio Center of Excellence in Knowledge-enabled Computing
Wright State University
Dayton, Ohio, USA
amit@knoesis.org

1. PRESENTERS

Pramod Anantharam is a researcher at Kno.e.sis - the Ohio Center of Excellence in Knowledge-enabled Computing (http://knoesis.org) where he is a primary member of the Semantic Sensor Web (http://knoesis.org/research/semsci /application_domain/sem_sensor) project. Currently he is pursuing research in Physical-Cyber-Social (PCS) computing (http://wiki.knoesis.org/index.php/PCS) with Prof. Sheth. His collaborations include researchers from IBM Research, University of Surrey, and electrical engineering at Wright State University. His broader research interests include knowledge representation, knowledge extraction and discovery from heterogeneous sensory observations, exploiting complementary nature of citizen and sensor observations for reasoning under uncertainty in PCS systems, and surprise modeling. He has interned twice at IBM Research and his internship work was awarded the best research showcase during summer of 2012. He has also served as external reviewer for WWW2012, EKAW2012, ER2012, HTS2013, and WWW2013.
Home page:
http://knoesis.org/researchers/pramod/

Payam Barnaghi is a Lecturer in the Centre for Communication Systems Research (CCSR) at the University of Surrey. He has been involved in several European projects in the Internet of Things area, including Internet of Things Environment for Service Creation and Testing (IoT.est), Internet of Things Architecture (IoT-A) (http://www.iot-a.eu/), and EXALTED (EXpAnding LTE for Devices) (http://www.ict-exalted.eu/). He was an invited expert in the W3C Incubator Group on Semantic Sensor Networks and was one of the contributors to the W3C Semantic Sensor Networks (SSN) ontology. His research interests include Knowledge Engineering, machine learning, large-scale semantic enabled systems, service platforms, linked-data, semantic sensor networks, Internet of Things, intelligent knowledge-based systems, and information search and retrieval. He co-edited a special issue of International Journal on Semantic Web and Information Systems on semantic technologies for Sensor Networks, Internet of Things and Smart Devices, and he is currently co-Guest editor of a special issue of IEEE Intelligent Systems on Web of Things. He is a senior member of IEEE.
Home page:
http://personal.ee.surrey.ac.uk/Personal/P.Barnaghi/

Amit Sheth is an educator, researcher and entrepreneur. He is the LexisNexis Ohio Eminent Scholar at the Wright State University, Dayton OH. He directs Kno.e.sis - the Ohio Center of Excellence in Knowledge-enabled Computing (http://knoesis.org) which works on topics in Semantic, Social, Sensor, and Services computing over the Web, with the goal of advancing from the information age to meaning age. Prof. Sheth is an IEEE fellow and is one of the most highly cited authors in Computer Science (h-index = 75) and World Wide Web (http://j.mp/www-0113). He is the EIC of the International Journal of Semantic Web & Information Systems, joint-EIC of Distributed & Parallel Databases, series co-editor of two Springer book series and serves on several editorial boards. By licensing his funded university research, he has also founded and managed two successful companies. Several commercial products and many operationally deployed applications have resulted from his R&D.
Home page:
http://knoesis.org/amit

WIMS'13, June 12-14, 2013 Madrid, Spain

ABSTRACT

This tutorial presents tools and techniques for effectively utilizing the Internet of Things (IoT) for building advanced applications, including the Physical-Cyber-Social (PCS) systems. The issues and challenges related to IoT, semantic data modelling, annotation, knowledge representation (e.g. modelling for constrained environments, complexity issues and time/location dependency of data), integration, analysis, and reasoning will be discussed. The tutorial will describe recent developments on creating annotation models and semantic description frameworks for IoT data (e.g. such as W3C Semantic Sensor Network ontology). A review of enabling technologies and common scenarios for IoT applications from the data and knowledge engineering point of view will be discussed. Information processing, reasoning, and knowledge extraction, along with existing solutions related to these topics will be presented. The tutorial summarizes state-of-the-art research and developments on PCS systems, IoT related ontology development, linked data, domain knowledge integration and management, querying large-scale IoT data, and AI applications for automated knowledge extraction from real world data.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Theory, Algorithm

Keywords

Knowledge Engineering, Internet of Things (IoT), Modelling, Cyber-Physical-Social Systems, Semantic Sensor Web, Ontology, annotation, integration, inference, reasoning, healthcare, traffic analytics

2. MOTIVATIONS

The World Wide Web has advanced beyond HTML documents and the links between them evolving toward billions of devices and sensors connected to the internet called "Internet of Things". The National Intelligence Council has named IoT as one of the six disruptive technologies. The report states that "By 2025 Internet nodes may reside in everyday things – food packages, furniture, paper documents, and more. Today's developments point to future opportunities and risks that will arise when people can remotely control, locate, and monitor even the most mundane devices and articles" [2, 1]. This tutorial is thus timely and important for the web community in general and semantic web community in particular. Internet of Things is an extension to the current Internet in which physical objects (e.g. sensors) or "Things" are seamlessly integrated into information networks which have the ability to communicate and interact collectively. IoT is enabled by various devices and resources (such as sensors) that report status of real world objects, provide observations and measurements from the physical environment, and communicate via Internet with other devices and/or parties that are interested in the data. With the diversity and heterogeneity of devices and resources that participate in capturing, communicating, and consuming the IoT data, new challenges such as interoperability, privacy, scalability, and reliability emerge in addition to storage, management, query, and discovery issues in dealing with potentially large amount of data captured from the physical world.

IoT data describes features of interest of entities in the physical world and is often dynamic, time and space dependent. There may be correlations between observation and measurement data collected from different devices and resources in an environment. Understanding correlations between observations from different modalities is the key in associating meaning to the variations in observations. Semantics in the form of modeling, annotation, integration, and perception will lead to deeper insights to synthesize actionable information from data.

Users and IoT data consumers frequently need to query status of physical objects and observe situations in different environments. Accessing, storing, querying, and managing large volumes of highly dynamic data already exhibits significant challenges. Further, to make low-level IoT data useful and meaningful to high-level applications and services, methods for semantic modelling, knowledge engineering, and abstraction over the raw observation and measurement data are required. These techniques are used to identify patterns, to create meaningful abstractions, and to extract machine-interpretable, and human-understandable knowledge. The need for these methods and the tools supporting them has become increasingly important due to extraordinarily large number of devices and smart Things being connected to the internet forming the IoT.

3. RELATION TO THE CONFERENCE TOPICS

This tutorial will cover a timely topic of semantics for IoT highly relevant to the theme of the conference as it will discuss the requirements, recent research and industry developments, best practices, and existing solutions to acquire, model, and manage data and knowledge in the IoT domain at a web scale. It will emphasise the importance of semantic technologies, data mining techniques, social data and crowd sourcing, and knowledge-based systems and services for IoT. Such innovative technologies provide essential support for intelligent information processing and integration, autonomous management of resources, perception creation, and automated decision making processes.

4. OVERVIEW OF CONTENT

The first part of the tutorial will cover the background information on Internet of Things, PCS systems, and some of their application scenarios for general audiences and for those interested in expanding their knowledge in these areas. The second part presents the state-of-the-art semantic techniques for data and knowledge engineering methods for the PCS domains. In particular, we will discuss the semantic frameworks and ontology models, interoperability issues, linked data solutions, annotation of the real world data, and reasoning with background knowledge. The third part of the tutorial will describe the available tools, pilot projects, and ongoing works and research projects. In addition, it will define the open issues and further highlight research and engineering challenges for data and knowledge engineering, knowledge extraction, and management in PCS systems.

The main topics of the tutorial include:

- Introduction to the Internet of Things based systems

- Data and knowledge modelling requirements, semantic annotation
- Semantics and data modelling for IoT
- Storage and access of IoT data and metadata
- User interaction and application scenarios
- Knowledge representation, querying, and interfaces
- Semantic abstractions/perception, reasoning
- Knowledge engineering and acquisition, Linked-data and semantic enabled systems
- Social and cognitive aspects of knowledge representation, reasoning and perception creation using physical-cyber-social data

Technology discussions will be complemented by demonstration of a number of systems and applications, several involving data from real-world deployments.

4.1 Introduction to IoT Based Systems

This part of the tutorial reviews the landscape of IoT and PCS systems, their proliferation into various application domains such as health care, sustainability applications (e.g. traffic, water, power management), logistics, and emergency response. Different types of sensors used in building IoT applications, their characteristics, and limitations will be presented along with a broader view of research issues encountered in real world IoT based system. These issues include heterogeneity, interoperability, energy efficiency, scalability, and security. Addressing these issues and the role of semantics based technologies in addressing these issues will be presented.

4.2 Data and knowledge modelling requirements, semantic annotation

The challenges portrayed in the introduction will drive the need for data and knowledge modeling. Data and knowledge modeling for discovery, access, search, integration, and interpretation will be presented with the description of Semantic Sensor Network (SSN) ontology [3, 4]. The data modeling involves representation of observations from various sensors such as temperature, humidity, video cameras, etc. The annotation techniques rely on the model of observations for annotating raw sensor observations. The annotated observations will enable integration of heterogeneous observations and interpretation of these observations in a contextually meaningful way. A comprehensive list of applications and research projects that have adopted the SSN ontology will be presented.

4.3 Semantics and data modelling for IoT

Use of semantics as an enabler for achieving the goals of IoT based applications such as scalability, interoperability, efficiency, analytics, timeliness, mobility, and discovery will be presented. Desirability of a model in terms of expressiveness, consistency, and modularity in the context of IoT will provide valuable insights into the tradeoffs that may be needed in the design of IoT based systems. A detailed view of representation techniques for sensors and sensor networks, observations and measurements, and platforms and systems using the SSN ontology will be presented. IoT representation vocabulary such as IoT-A projectÕs entity-resource-service ontology and IoT.est projectÕs service description framework will be presented. A concrete example of representing sensors as services and annotation standards for services will be presented.

4.4 Linked-data and semantic enabled systems

Use of URIs to represent unique "things" such as resources (e.g. sensors) and their observations (e.g. temperature) on the IoT leads to well linked and interoperable data. Further, it facilitates querying across data sets leading to innovative applications. Experiences in modeling sensors and their observations using the SSW ontology, publishing linked sensor data with over 1.7 billion triples will be presented. This provides an overall view of the process leading to better adoption of web standards and associated techniques.

4.5 Knowledge engineering and acquisition

Analytics over IoT observations plays an important role in decision making. This section of the tutorial focuses on a wide range of knowledge acquisition techniques including building ontology by domain experts, statistical and machine learning techniques of knowledge acquisition. The domain knowledge acquired in this process may be expressed in the form of an ontology. We provide concrete examples of such an acquisition from heterogeneous traffic sensor observations spanning machine and citizen sensor observations. Some of the challenges in representation of uncertainty in domain knowledge will be presented which is one of the focus areas of the conference.

4.6 Physical-Cyber-Social Computing

The data being generated from IoT encompasses observations from physical, cyber, and social worlds. In this part of the tutorial, we present a novel approach toward processing observations from physical-cyber-social worlds. We introduce horizontal operators for integration of heterogeneous observations and vertical operators for creating abstractions transcending the DIKW triangle. We present next generation application scenarios in healthcare and traffic that takes a holistic view toward computing.

4.7 Cognitive aspects of knowledge representation, reasoning and perception

As perceptual beings, humans have evolved to process observation streams and make sense of it in an efficient manner. The cognitive theories of perception are based on the background knowledge that humans have evolved to process sensory observations. In this part of the tutorial we present computational theories of machine perception for making sense of machine sensor observations.

4.8 IoT/PCS Systems: Applications

4.8.1 k-Health

kHealth is a knowledge-enabled semantic platform to enhance decision making and improve health, fitness, and well-being. It supports contextual (e.g., condition specific) annotation, integration, and interpretation of sensor and mobile data from individuals using deep domain (e.g., disease) specific knowledge base.

k-Health is an example of a cyber-physical-social system.

The physiological parameters of a human is the physical component. The environmental factors such as temperature, pollen count, and pollution constitutes the cyber component. The collective intelligence or the community knowledge forms the social component. We will discuss our experience and the challenges in designing, implementing, and deploying a real world system for reducing readmissions of cardiac patients. The focus will be on use of semantics for representation of observations, cardiology domain knowledge, and perceptual reasoning in deriving abstractions from raw sensor data.
Demo: http://knoesis.org/healthApp/

4.8.2 k-Traffic

k-Traffic is a knowledge-enabled hybrid platform that combines semantics for conceptual representation of domain knowledge and the graphical models to deal with incomplete and uncertain observations. The system analyzes heterogeneous traffic observations such as machine sensor observations (e.g. speed, volume) and textual observations (e.g. incident reports) to glean correlations.

The focus will be on the principles of knowledge extraction from PCS systems with a novel view of complementing the extraction with existing domain knowledge. These principles will be presented in such a way that it can be beneficial in gleaning insights from other physical-cyber-social systems such as health care. The challenges and experiences in implementing a system that processes streaming heterogeneous traffic observations will be portrayed.

4.8.3 SECURE

This project demonstrates a Semantic Web enabled system for collecting and processing sensor data within a rescue environment. Heterogeneous raw sensor data from sensors on a rescue robot are collected in real time through a wireless sensor network. The raw sensor data is converted to RDF using the Semantic Sensor Network (SSN) ontology and further processed to generate abstractions used for event detection in emergency scenarios. These abstractions when presented to first responders will lead to intuitive understanding of the situation and better response. Demo: `http://www.youtube.com/watch?v=in2KMkD_uqg`

4.8.4 Real Time Feature Streams

The emergence of dynamic information sources - including sensor networks - has led to large streams of real-time data on the Web. Research studies suggest, these dynamic networks have created more data in the last three years than in the entire history of civilization, and this trend will only increase in the coming years. With this coming data explosion, real-time analytics software must either adapt or die. This paper focuses on the task of integrating and analyzing multiple heterogeneous streams of sensor data with the goal of creating meaningful abstractions, or features. These features are then temporally aggregated into feature streams. We will demonstrate an implemented framework, based on Semantic Web technologies, that creates feature streams from sensor streams in real-time, and publishes these streams as Linked Data. The generation of feature streams can be accomplished in reasonable time and results in massive data reduction. Demo: `http://www.youtube.com/watch?v=_ews4w_eCpg`

4.8.5 SemMOB

SemMOB enables dynamic registration of sensors via mobile devices, search, and near real-time inference over sensor observations in ad-hoc mobile environments (e.g., fire fighting). We demonstrate SemMOB in the context of an emergency response use case that requires automatic and dynamic registrations of sensor devices and annotation of sensor observations, decoding of latitude-longitude information in terms of human sensible names, fusion and abstraction of sensor values using background knowledge, and their visualization using mash-up. Demo: `http://knoesis-twit.cs.wright.edu/SidfotWebapp/Demo/`

4.8.6 LOKILAS

LOst Key Identification, Localization and Alert System (LOKILAS) is an IoT based solution to track physical resources that are often misplaced. This application involves real time monitoring of keys to shared facilities such as conference rooms, study areas, and kitchen at Kno.e.sis research lab. The system demonstrates monitoring of physical objects (keys) and spaces (rooms) by instrumenting them for logistics and optimization of resource usage.

5. PREVIOUS PRESENTATION AND MODIFICATION TO WIMS'13

A version of this tutorial was presented at The 18th International Conference on Knowledge Engineering and Knowledge Management, Galway City, Ireland, held from October 8 - 12, 2012 (EKAW tutorial can be seen as more infrastructure centric, WIMS tutorial will have more advance data processing and application focus). The tutorial will be framed in the context of web data and semantics and covers many relevant topics of the conference such as linked data, web mining, uncertain reasoning, and ubiquitous intelligence. The tutorial includes recent developments in the state-of-the-art work on IoT and PCS system analytics. The insights into processing observations from PCS systems will be rendered in the form of application case studies spanning healthcare, sustainability, logistics, and emergency response.

The challenges in modeling, integrating, and reasoning with observations from PCS systems will be discussed by presenting real world problem scenarios and data. Role of semantics in dealing with IoT/PCS data will be exemplified in the tutorial.

6. DURATION

The tutorial is proposed for 3 hour session, however it can be offered in 2 hours if needed.

7. LEARNING OUTCOMES AND JUSTIFICATION OF HIGH QUALITY PRESENTATION

The tutorial will present the state-of-the-art in modeling, annotation, integration, and perception for processing IoT data to derive insights. The tutorial balances between theory and practical applications that have been developed along with some on-going projects. These practical applications will help attendees gain a unique perspective on the use of semantics in the context of IoT data modeling and analytics in practice. The applications intended for demonstration spans healthcare, sustainability, logistics, and emergency response. With the interdisciplinary nature of the tutorial topic and applications, the tutorial will foster discussions and exchange of ideas among the attendees.

The tutorial goes beyond state-of-the-art in providing a unique perspective on contextual observations spanning physical-cyber-social domains for generating abstractions and insights for holistic decision making. Application demonstrations may involve active participation of the attendees especially for the healthcare application and will be captivating for the attendees.

The presenters of the tutorial and their colleagues have presented many well attended tutorials in the past:

- Krishnaprasad Thirunarayan, Pramod Anantharam. Trust Networks: Interpersonal, Sensor, and Social. Tutorial Presentation at International Conference on Collaborative Technologies and Systems (CTS 2011), Philadelphia, Pennsylvania, USA. (http://knoesis.org/library/resource.php?id=1103)

- Wei Wang, Cory Henson, Payam Barnaghi. Data Modeling and Knowledge Engineering for the Internet of Things. Tutorial presented at the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012), Galway City, Ireland, October 8-12, 2012. (http://knoesis.org/library/resource.php?id=1800)

- M. Nagarajan,A. Sheth,S. Velmurugan Citizen Sensor Data Mining, Social Media Analytics and Development Centric Web Applications. Proc. of the WWW 2011, March 28 - April 1, 2011, Hyderabad, India, ACM.

- M. Nagarajan, C. Ramakrishnan, and A. Sheth. Text Analytics for Semantic Computing - the good, the bad and the ugly, Second IEEE International Conference on Semantic Computing, 2008.

- A. Sheth and S. Stephens. Semantic Web: Technologies and Applications for the Real-World, WWW2007. Tutorial notes at: http://www2007.org/tutorial-T11.php [Highest attendance at this conference.]

The presenters are also actively involved in research and development of IoT projects which involves interaction with domain experts (e.g. medical staff and researchers at Ohio State Medical School), and real-world deployment of IoT and physical-cyber-social applications (e.g. k-Health), and EU research projects on IoT (e.g. FP7 FI-WARE, FP7 IoT.est, FP7 IOT-A).

8. AUDIO-VISUAL OR TECHNICAL REQUIREMENTS

No specific requirements- a video projector and Internet connection during the tutorial will be enough.

9. INTENDED AUDIENCE AND EXPECTED NUMBER OF PARTICIPANTS

We expect audience from the Internet of Things and Web of Things communities as well as participants from the related EU projects. The main targeted audience will be PhD students and researchers from industry and academia who would like to have more insight on the state-of-the-art solutions and developments on data modelling and knowledge engineering in IoT and other related domains. We expect around 30 to 40 participants for this tutorial. The presenters are involved in a number of EU and other international projects in this area and will use the project communities and other links to disseminate the call for participation.

10. ACKNOWLEDGMENTS

We would like to thank Cory Henson (Kno.e.sis) and Wei Wang (CCSR) for their contribution to the tutorial.

11. REFERENCES

[1] National intelligence council, disruptive civil technologies - six technologies with potential impacts on us interests out to 2025 - conference report cr 2008-07. April 2008.

[2] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer Networks*, 54(15):2787–2805, 2010.

[3] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, et al. The ssn ontology of the w3c semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2012.

[4] L. Lefort, C. Henson, K. Taylor, P. Barnaghi, M. Compton, O. Corcho, R. G. Castro, J. Graybeal, A. Herzog, K. Janowicz, H. Neuhaus, A. Nikolov, and K. Page. Semantic Sensor Network XG Final Report. Technical report, W3C Semantic Sensor Network Incubator Group (SSN-XG), June 2011.