

Report on the First International Workshop on Personal Data Analytics in the Internet of Things (PDA@IOT 2014)

Vassilis Christophides
R&I Center, Technicolor, France
vassilis.christophides@technicolor.com

Themis Palpanas
Paris Descartes University, France
themis@mi.parisdescartes.fr

1. INTRODUCTION

In the last few years, we are witnessing a wave of connected things (i.e., devices) that are flooding our everyday living spaces. These networks of physical objects with *embedded sensors* and *actuators* that communicate with other objects, databases, and in some cases with people, are often described under the umbrella term *Internet of Things* (IoT). Recent studies¹ estimate that by 2020 the IoT will scale up to 50 billion daily-use objects, while its economic value will reach \$14.4 trillion dollars, spread across all sectors (both consumer² and industrial³ markets). According to the IoT vision, “smart things” will “disappear” in our living environments (aka *Pervasive Computing*), or be embodied in humans (aka *Wearable Computing*, with the goal of building disruptive services for human-beings, offering a *seamless* and *implicit interaction* between the *real* and the *virtual* worlds.

In this changing world, every human action and living context will create “information shadows” on the IoT via active, or passive monitoring. For example, *Quantified-self*⁴) devices enable self-tracking of any kind of biological, physical, or behavioral information, *smart home sensors*⁵ capture accurate indoor environmental information and energy consumption habits, *residential gateways* record in real-time usage information of communication and entertainment devices, while *smart cars and phones* trace the places and trajectories of our everyday life. Fine-grained personal data are already being massively created, permeating almost every facet of human interaction to objects (*user awareness*), to their physical environment (*ambient awareness*) and to other humans (*social awareness*). Such personal data are mostly analyzed today by application silos tightly coupled with the employed “smart things” (in health and well-being, home automation and entertainment, etc.). Individuals

are striving today for tools that can help them to gather, process and make sense of all the data they produce in private, as well as public spaces.

The 1st International Workshop on Personal Data Analytics in the Internet of Things (PDA@IOT), held in conjunction with VLDB 2014, aims at *sparkling research on data analytics, shifting the focus from business to consumers services*. While much of the public and academic discourse about personal data has been dominated by a focus on the privacy concerns and the risks they raise to the individual, especially when they are seen as the new oil of the global economy⁶, PDA@IOT focus on how persons could effectively exploit the data they massively create in Cyber-Physical worlds. It relies on the following assumptions⁷:

- My personal data is most *valuable to me!*
- People want to *benefit from their data*— *not hide it away* - but they need control, trust & transparency
- Personal data economy is primarily a *human experience economy*
- Personal data intend to *empower people*

We believe that the full potential of the IoT goes far beyond connecting “things” to the Internet: it is about using data to create new value for people. In a *People-centric computing paradigm*⁸, both small scale personal data and large scale aggregated data should be exploited to identify unmet needs and proactively offer them to users. PDA@IOT seeks to address current technology barriers that impede existing personal data processing and analytics solutions to empower people in *personal decision making*⁹.

The PDA@IOT ambition is to provide a unique forum for researchers and practitioners that approach personal data from different angles, ranging from data management and processing, to data mining and human-data interaction, as well as to nourish the interdisciplinary synergies required to tackle the

¹ D. Evans “The Internet of Things: How the Next Evolution of the Internet Is Changing Everything”, Cisco IBSG 2011.

² M. Vallve “The Internet Of Things' Will Change Everything About The Global Consumer Economy”, BII May 2014

³ P. C. Evans, M. Annunziata “Industrial Internet: Pushing the Boundaries of Minds and Machines”, GE, Nov. 26, 2012

⁴ Z. Yumak, P. Pu “Survey of Sensor-Based Personal Wellness Management Systems” *BioNanoScience* 3(3) Sept. 2013

⁵ X. Zhang, T.Kato, T.Matsuyama Learning a context-aware personal model of appliance usage patterns in smart home. *Innovative Smart Grid Technologies (ISGT) - Asia*, 2014

⁶ A. Pentland et al. “Personal Data: The Emergence of a New Asset Class”, *World Economic Forum*. January 2011.

⁷ A. Margolis “A people-centred approach to Data and the Internet of Things”, *Claro Partners*, 2014.

⁸ Similar paradigm shifts have been recently proposed in other domains such as *Vendor Relationship Management (VRM)* in contrast to *Customer Relationship Management (CRM)* or the *Intention* in contrast to the *Attention Economy*.

⁹ J. Duggan. “The Case for Personal Data-Driven Decision Making”. In *PVLDB*, vol. 7, 2014.

challenges and problems emerging in People-centric Computing. A list of core research issues considered by the first edition of the workshop includes:

- *Trusted architectures* for personal data collection balancing data privacy & utility
- *Interpretation of raw personal data streams* originating from heterogeneous sensors
- Descriptive, predictive & cognizant *data personal analytics*
- *Offline/online personal data processing & curation.*

The workshop program included 2 keynote talks, 7 research papers, and a panel discussion. In this report, we summarize the ideas that were presented and discussed in the workshop. The detailed program and the slides for the talks are available at the PDA@IOT web page: <http://www2.thlab.net/pda-iot2014/>.

2. Keynote Talks

The first keynote talk, “Objective Self”, was given by Prof. Ramesh Jain from the University of California at Irvine (USA). Prof. Jain presented how it is now possible to analyze and understand a person’s life style in order to build a model for an individual, called the “Objective Self”. Most people use phones with a multitude of sensors that continuously generate data streams related to several different aspects of their lives. Other powerful sources of information for a person are social networks, e-mail, calendar systems, and lately wearable and home sensors. By integrating these multi-sensory data streams, it is possible to create an accurate chronicle of a person’s life. In this talk, Prof. Jain focused on the context of a person’s health: by correlating life events to health related events obtained using wearable sensors and other sources of information, one can build the health persona of a person, which is a long-term objective characterization of a person’s health. Several data streams can be used in this respect, such as motion tracking, location tracking, activity level, wearable health sensors, and personal calendar data. Prof. Jain illustrated how recognition algorithms can be applied to the Life Event detection problem and then build an objective personal chronicle, called personicle. By building a personicle over a long period and applying data mining, it is possible to create a model of the person that could result in actionable insights and alerts in everyday life.

The second invited speaker was Prof. Dimitrios Georgakopoulos from the RMIT University (Australia) and his talk was entitled “Distilling High Value Personalized Information from the Internet of Things and Social Networks”. He highlighted that Cyber-Physical-Social (CPS) Computing encompasses real-time extraction of high value and personalized information from social networks and millions of

cyber-physical systems in the IoT, as well as the development of specialized cloud-based services. Despite the expanding array of applications that require distilling knowledge from big CPS data, there is currently no easy way to manage and exploit such big data, personalize the results, or develop cloud services that make this possible anywhere via mobile devices. Therefore, CPS requires the development of novel solutions for discovering on-line cyber-physical and social media sources, dynamically integrating such sources and their data, and analyzing or personalizing billions of data streams and tens of years of historical data from on-line sources anywhere and in real time. Prof. Georgakopoulos was particularly interested in context-aware techniques proposed by research and commercial systems that can help individuals understand IoT data. Finally, he outlined a unified approach for CPS data management and analysis involving real-time summarization and personalization as the main way of extracting high value information from big data in various domains: digital agriculture, smart energy grids, and disaster management.

3. Research Papers

3.1 Healthcare and Wellbeing

The first session of the workshop was devoted to research papers that addressed *personal decision making in healthcare and wellbeing*.

In the paper entitled “*kHealth: Proactive Personalized Actionable Information for Better Healthcare*”, Amit Sheth, Pramod Anantharam, and Krishnaprasad Thirunarayan focused on health aficionados and patients with chronic conditions that increasingly rely on sensors and mobile devices to track sleep, food, physical activity, and other physiological observations (e.g., weight, heart rate, blood pressure). They highlighted a paradigm shift in healthcare from reactive medicine to predictive, preventative, personalized, and participatory medicine, in which individuals are empowered to fully participate in health related decision making. To facilitate this transformation, there is a need to understand the richness and nuances of fine-grained healthcare data produced by a variety of mobile devices and wearable sensors, related to a person and the population in general. The majority of existing data analytics techniques focuses on finding discrepancies in a single stream of observations without much insight into the problem and actionable knowledge to the individuals. The authors presented the kHealth system for analyzing observations from passive (no human involvement in data collection) and active (human input involved in data collection) sensors in order to provide explanations that are intelligible to individuals and their clinicians for well-informed decision making.

In the same direction, Emmanouil G. Spanakis, Feng Dong, Manolis Tsiknakis, Vangelis Sakkalis, Kostas Marias and Dimitris Kafetzopoulos, in their work entitled “*MyHealthAvatar: The digital patient perspective using Internet of Things technologies*”, focused on how health-related information in the IoT can be locally aggregated and transmitted for remote monitoring and response. They stressed that the interrelation of such heterogeneous data sources opens new directions for providing a comprehensive picture related to health, thus triggering interventions by medical staff when necessary, and realizing preventive care. They described MyHealthAvatar (MHA), a personal digital health related collection bag, carried by individual citizens throughout their lifetime to sustain in a meaningful manner all information related to their health, empowering users through a number of patient-centred healthcare services. They outlined the MHA challenges for accessing, collecting and sharing long term multilevel personal health data through an integrated environment including: clinical data, genetic data, medical sensor data and devices, human behavior data and activity data for clinical data analysis, prediction and prevention for the individual citizen.

Finally, Xinpeng Zhang, Hiroki Kitabayashi, Yasuhito Asano and Masatoshi Yoshikawa, presented “*A Health-aware Pedestrian Navigation System by Analysis of Spatiotemporal Vital Signs Data*”, discussing how health-related data can empower individuals in their outdoor social activities. The authors focused on a health-aware pedestrian navigation system to assist the outdoor activities for vulnerable pedestrians, such as senior people, or people with disabilities. Examples of health related profiling of jogging itineraries include the recognition of when pedestrians get tired while climbing a slope, or if they feel stressed when walking some road. The system offers essentially two services. First, it monitors the personal health condition from individual vital signs data. Since a pedestrian’s health condition varies by location and time due to fatigue and stress, it is spatio-temporally contextualized. Second, the system recommends jogging itineraries according to the results of the previous analysis. It can therefore guide multiple pedestrians to walk together, promoting conversation and strengthening mental health.

3.2 Life Logging and Citizen Services

The second session of the workshop focused on extracting *individual and collective behaviors for life logging and citizen services*.

Laleh Jalali, Da Huo, Hyungik Oh, Mingfan Tang, Siripen Pongpaichet and Ramesh Jain, in their work entitled “*Personicle: Personal Chronicle of Life Events*”, presented an overview and early results of

their project aiming at integrating, aggregating, and analyzing heterogeneous personal data streams to build a persona, and use it to recognize evolving personal situations. In particular, the authors are interested in recognizing movements and personal situations in order to empower individuals with actionable information and insights. They demonstrated how high-level life events can be extracted through simultaneous use of asynchronous observations consisting of continuous GPS and accelerometer measurements. To create a chronicle of life events, called Personicle, they detect low level physical activities using data produced by various unobtrusive sensors embedded in a mobile phone, and use hierarchical classification techniques for identifying high level life events using location context and physical activity. The authors plan to collect, store, and analyze data from a large number of heterogeneous sensors, and make publicly available the test data sets.

Yuzuru Tanaka, Hajime Imura and Jonas Sjabergh, in their work entitled “*Exploratory Visual Analytics for Winter Road Management using Statistically Preprocessed ProbeCar Data*”, show how aggregated probe car data can be statistically preprocessed over road links for an urban-scale area, in order to visualize the dynamic change of traffic flow in terms of the divergence and the flow vector field. This analysis provides insights regarding the dynamic change of traffic hotspots, main traffic streams, and route selection preference. This work is motivated by new citizen services, in particular improving winter road management in Sapporo. They extend well-known exploratory visual analytics techniques to multiple coordinated views by integrating different analysis tools with their result visualization views into the same operational framework. These newly added views may coordinate with others, and allow users to directly select patterns calculated at runtime to further quantify the underlying database view. Exploratory visual analytics with such an environment enables us to detect road links for effective, pinpoint snow-removal.

3.3 Wearable Devices

The third session of the workshop program comprised papers *analyzing data from wearable devices* to recognize individuals’ emotional arousal, or train their brain’s fast and slow thinking abilities.

Julien Fleureau, Philippe Guillotel and Izabela Orlic in their work entitled “*Affective Profiles of Movies and Opera Based on the Physiological Responses of the Audience*” proposed an objective study of the emotional impact on real audience of various audiovisual shows and live events. An affective benchmarking solution was presented, making use of a low-intrusive measurement of the ElectroDermal

Activity (EDA), which is known to be linked to the unconscious reactions of the nervous system, and thus can be used as a measure of emotional and sympathetic arousal responses. A dedicated processing of this biosignal produces a time-variant and normalized affective profile related to the significant excitation variations of the audience. Besides the methodology, the originality of this work stems from the evaluation of the proposed framework on 62 real audience members during special screenings of a film festival and one opera. The authors show that the resulting "Affective Profile" is strongly correlated to the events occurring during the shows, that some shows have a higher intensity than others, and that the extracted affective profile is consistent with the artistic rules from the creative intent.

Finally, Melanie Swan, Takashi Kido and Minna Ruckenstein in their work entitled "BRAINY: Multi-modal Brain Training App for Google Glass" presented a new Google Glass application for brain training. This work is motivated by the observation that there are over 50 brain fitness training companies, but there is not yet a brain fitness training app for Glass. BRAINY initially targets the improvement of memory function, and additional cognitive training modules can be added later for attention, processing speed, flexibility, problem solving, and other areas. BRAINY takes advantage of the audio-visual and voice command functionality of the Glass platform to create multi-modal memory games that can be played easily by users during waiting time, or other down time. The more advanced memory modules of BRAINY target neuroplasticity and memory updating, and train the brain's fast and slow thinking systems.

4. Panel Discussion

Given that personal data processing and analytics is still in its infancy, the panel attempted to sketch a first roadmap of the emerging core research issues. The moderator, Vassilis Christophides (Technicolor R&I, France), challenged three panelists on the building blocks required to understand and make sense of Personal Data harvested in the IoT for enabling *smarter individual choices* and *behavioral changes*, as well as *resource-conscious automation and optimization of everyday life*: (a) Prof. Ramesh Jain (Univ. of California at Irvine, USA) (b) Prof. Yuzuru Tanaka (Hokkaido Univ., Japan), and (c) Prof. Themis Palpanas (Paris Descartes Univ., France).

Prof. Ramesh Jain emphasized the need for collecting data from various wearable devices, mobile phones, and social networks, in a way that users do not feel threatened for their privacy. Current data sources are very limited, because people do not see a good reason to take the risk and share their data. He

emphasized the need for developing simple applications that clearly show that data sharing results in tangible benefits, in order to encourage people to share their data for the long-term benefits to individuals, as well as to the society. Another issue that Prof. Jain raised was a sustainable business model to be used for personal data. Some people have talked about trading personal data for value depending on who wants access to it. This is a fundamental issue that assumes that personal data is a new currency of some kind that could be traded for the appropriate value.

Prof. Yuzuru Tanaka pointed out that urban-scale monitoring of collective human activities is nowadays conducted through private sector sensing (smart-phone, probe-car and card usage data), public sector sensing (traffic jam/accident, public transportation and surveillance-camera image data), and crowd-source sensing (smart-phone application usage and Social Network data). Unfortunately, these data reside in different silos, mainly due to data privacy concerns. Statistical data processing may be a solution. However, what kind of statistical processing may allow what kind of analyses to discover what kind of knowledge about collective human activities is still an open issue. More generally, he argued that there exists a big gap between the state-of-the-art analytical methodologies and the complex big data application requirements, such as optimization of urban-scale infrastructures. In these settings, we need to analyze systems of heterogeneous systems, where each component can be mathematically modeled, but the whole system cannot. To find an appropriate analysis scenario is itself a research task. In this respect, he advocated the need for new machine learning algorithms capable to deal with these challenges, and that, for the moment, a pragmatic alternative seems to be exploratory visual analytics.

Prof. Palpanas stressed the need for non-technical users to be able to easily manage their personal data, as well as perform a range of analysis tasks on them. This would give them insights with respect to their habits and personal choices, and enable them to make informed decisions about their lifestyles. He argued that this goal could be achieved by the development of a new generation of systems and applications having usability at the core of their design. He offered as an example the domain of data series: even though most of the popular personal monitoring devices produce data in the form of time series, there are no suitable systems that can empower non-technical users (or developers) to take full advantage of the wealth of information hidden in these data.

ACKNOWLEDGMENTS

Special thanks go to Technicolor R&I Center (France) for their support in the organization of this workshop.