

Mobile Big Data Analytics: Research, Practice and Opportunities

Demetrios Zeinalipour-Yazti

Department of Computer Science, University of Cyprus, 1678 Nicosia, Cyprus

Email: dzeina@cs.ucy.ac.cy

Shonali Krishnaswamy

Faculty of IT, Monash University, VIC3145, Australia and Institute for Infocomm Research, 138632 Singapore

Email: shonali.krishnaswamy@infotech.monash.edu.au

Abstract—The rapid expansion of broadband mobile networks by Telecom Operators, has introduced a versatile global infrastructure that internally generates vast amounts of spatio-temporal *network-level* data (e.g., user id, location, device type, etc.) At the same time, mobile app vendors have nowadays at their fingertips massive amounts of *app-level* data collected through implicit or explicit crowdsourcing schemes with multi-sensing smartphones that have become a commodity. Mobile big data analytics refers to the discovery of previously unknown meaningful patterns and knowledge from a few dozen terabytes to many petabytes of data collected from mobile users at the network-level or the app-level. Example analytics range from high-level metrics and summaries (e.g., through clustering, classification and association rule mining) useful to executive managers to alert-based analytics (e.g., anomaly detection) useful to front-line engineers and users. This panel will explore how the academia and industry are tackling mobile big data analytic challenges. It will also identify and debate the key challenges and opportunities, in terms of applications, queries, architectures, to which the mobile data management and mobile data mining communities should contribute to.

Keywords—*Big Data, Analytics, Smartphones, Query Processing, Telecom Infrastructures.*

I. INTRODUCTION

The rapid expansion of broadband mobile networks by Telecom Operators, has introduced a versatile global infrastructure that is bringing unprecedented communication possibilities to the human mankind. Mobile telecoms have traditionally generated vast amounts of spatio-temporal mobile broadband data about their customers (e.g., user id, location, device type, etc.), but this data has been kept internal to the telecoms for a variety of reasons, including economic advantage, privacy and scale. The advent of mobile apps on the other hand, executing on powerful computational devices equipped with multitude of sensors that are capable of generating vast amounts of data (geo-location, audio, video, etc.), has brought data collection and crowdsourcing to the fingertips of virtually any of millions of Mobile App Vendors. In addition, other sources of mobile user location data is emerging through infrastructure level sensors (e.g., iBeacons) and external transactional level information obtained through interactions that involve location (e.g., location-based social networks).

Big data refers to data sets or streams that strain the ability of commonly used relational DBMSs to capture, manage, and

process the data within a tolerable elapsed time. Big data sizes commonly range from a few dozen terabytes to many petabytes in a single database and their underlying data model might be anything from structured (relational or tabular) to semi-structured (XML or JSON) or even unstructured (Web text and log files). Mobile big data analytics refers to the discovery of previously unknown meaningful patterns and knowledge from data collected from mobile users. In many applications and use cases, the aim is typically to understand deeply the following aspects:

- *who is the user (e.g., demographics and aggregate / summarization queries)?*
- *what are they likely to do at that location and time (e.g., activity)?*
- *where are they likely to go next (e.g., prediction of next place / place-of-interest)?*

Example mobile big data analytics range from high-level metrics and summaries (e.g., through clustering, classification and association rule mining) useful to executive managers to alert-based analytics (e.g., anomaly detection) useful to front-line engineers and users. A typical constraint associated with such mobile user analytics is that it needs to be near real-time since the focus is typically on delivering location-based and highly personalized information and services to a mobile user. Mobile app vendors (social, location, games, marketing enterprises like flurry, etc.) might nowadays know much more about the customers of a telecom than the telecom itself, as they have access to finer-grain data collected right on the *app-level* rather than at the *network-level*.

This panel will explore how academia and industry are tackling mobile big data analytic challenges. It will also identify and debate the key challenges and opportunities, in terms of applications, queries, architectures, to which the mobile data management community should contribute. Our panelists are expected to bring wealth of experience from the telecom sector, the mobile app sector but most importantly the mobile data analytic sector. The intent is not to overview individual products and or solutions, nor to provide a background on mobile data analytic solutions. The panelists will attempt to clarify where mobile data analytics are expected to take place in the future, the kind of architectures necessary, potential applications and killer queries, privacy and ethical aspects as well as commercial uptake possibilities.

II. POTENTIAL QUESTIONS

The goal of this panel is to convey an advanced understanding of the unique characteristics, challenges and opportunities of mobile data analytics to its audience. The panel is targeted to scientists with a basic understanding of mobile data management, but no knowledge of mobile data analytics in particular should be expected. Through the short positioning statements of panelists and the active engagement of the audience, the goal will be to derive the future directions in the emerging domain of mobile data analytics.

During the panel, the panelists will be challenged with one or more of the following questions:

- **What are the right big-data architectures for mobile big data analytics?** Big data architectures are usually highly parallel and distributed, in order to cope with the inherent I/O and CPU limitations (e.g., Hadoop/Hive data warehouses for purely analytic scenarios but also Apache Storm/Spark for streaming scenarios). Such systems typically perform on mid-scale internal/private clouds, offering higher privacy, to large-scale public clouds, both exposing operational and analytic functionality stand-alone or as-a-Service.
- **What are the killer (top) queries to answer in mobile big data analytics?** A few examples include: Spatio-Temporal Analytics, Anomaly detection, Association Rule Learning, Clustering, Classification, Privacy Issues. The panelists are expected to expand on specific projects they have worked on in the past, which are related to this theme and that can trigger discussions on what is important to be tackled in the future.
- **What is the role for distribution of analytics between the mobile device and the cloud itself to achieve scalability and privacy?** The more data analytics is pushed to the device, the more privacy could potentially be provided with only aggregate information being sent to the Cloud. Thus, how are users provided with privacy controls beyond the proverbial question of whether they are acceptable to sharing location data? What are typical characteristics of mobile data management and data mining techniques that are operating at the device level?
- **Who is winning the mobile big-data analytics war? telecoms or mobile app vendors?** Traditional telecom applications like SMS are severely hampered by the proliferation of Over-the-Top (OTT) messaging apps (e.g., WhatsApp, WeChat, Viber, Line and Skype), which are pushing core telecom functionality to the app-level. This has triggered a backfire act from telecoms that are trying to catch up with those developments by opening up internal telecom functionality to application development processes and developers (e.g., Huawei's Service Delivery Platform).
- **Do you believe that mobile big-data analytics will have a commercial/real impact?** What are specific industries where mobile big-data analytics will play a crucial role in the future (e.g., marketing, urban planning, social networks)? The panelists are expected

to expand on the business domains they expect this paradigm to have the earliest commercial/practical impact.

- **What are the most prominent Privacy and Ethical aspects you see with mobile data analytics?** Enabling mobile data analytics might have a legal dimension in many areas. For example, Europe has a strict Data Protection Policy (e.g., Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.) The panelists are expected to expand on whether those aspects present a legal barrier in the uptake of mobile data analytics.

SHORT BIOGRAPHIES OF PANEL CO-ORGANIZERS



Demetrios Zeinalipour-Yazti is an Assistant Professor of Computer Science at the University of Cyprus, directing the Data Management Systems Laboratory (DMSL). He got his Ph.D. and M.Sc. in Computer Science and Engineering from the University of California - Riverside in 2005 and 2003, respectively. He got

his B.Sc. in Computer Science from the University of Cyprus in 2000. He has also been a visiting researcher at the Network Intelligence Lab of Akamai Technologies Cambridge, USA in 2004. His research interests include Data Management in Systems and Networks, including: Distributed Query Processing, Storage and Retrieval Methods for Sensor, Smartphone and Peer-to-Peer Systems, Mobile Data Management and Network Data Management. He has been a PC Co-Chair for IEEE MDM'10, DMSN'10 (with VLDB'10) and ACM MobiDE'09 (with SIGMOD'09). He has also been the Contest Chair for IEEE ICDM'10, the Demo Co-chair for IEEE MDM'13 and the Panel Co-chair for IEEE MDM'14. For more information please visit: <http://www.cs.ucy.ac.cy/~dzeina/>.



Shonali Krishnaswamy is an Associate Professor at the Faculty of Information Technology of Monash University, Australia and is also the Head of the Data Analytics Department at the Institute for Infocomm Research (I2R), Singapore. She holds a PhD (Computer Science, 2003) and Master of Computing (1998)

from Monash University, a Bachelor Science (Computer Science, 1996) from the University of Madras, India, and a Graduate Certificate in Higher Education (2006) from Monash University. Her research is broadly defined in the area of Distributed, Mobile and Pervasive Computing Systems, where she focuses on the development of intelligent applications that aim to service real-time information needs while having to function in highly dynamic and resource-constrained environments. Her specific expertise is in Mobile Data Stream Mining, Large Scale Distributed Data Mining, Mobile Crowdsensing and Mobile Activity Recognition. She was part of teams that won ALL Four Challenge Tasks in the 2011 EU OPPORTUNITY Activity Recognition Challenge. For more information please visit: <http://www.csse.monash.edu.au/~shonali/>.