Mining Mobility Data

Spiros Papadimitriou
Rutgers University
100 Rockafeller Road
Piscataway, NJ 08854, USA
spapadim@business.rutgers.edu

Tina Eliassi-Rad
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08854 USA
eliassi@cs.rutgers.edu

ABSTRACT

The fairly recent explosion in the availability of reasonably fast wireless and mobile data networks has spurred demand for more capable mobile computing devices. Conversely, the emergence of new devices increases demand for better networks, creating a virtuous cycle. The current concept of a smartphone as an always-connected computing device with multiple sensing modalities was brought into the mainstream by the Apple iPhone just a few years ago. Such devices are now seeing an explosive growth. Additionally, for many people in the world, such devices will be the first computers they use. Furthermore, small, cheap, always-connected devices (standalone or peripheral) with additional sensing capabilities are very recently emerging, further blurring the lines between the Web, mobile applications (a.k.a. apps), and the real world. All of this opens up countless possibilities for data collection and analysis, for a broad range of applications.

In this tutorial, we survey the state-of-the-art in terms of mining mobility data across different application areas such as ads, geo-social, privacy and security. Our tutorial consists of three parts. (1) We summarize the possibilities and challenges in the collection of data from various sensing modalities. (2) We cover cross-cutting challenges such as real-time analysis and security; and we outline cross-cutting algorithms for mobile data mining such as network inference and streaming algorithms. (3) We focus on how all of this can be usefully applied to broad classes of applications, notably mobile and location-based social, mobile advertising and search, mobile Web, and privacy and security. We conclude by showcasing the opportunities for new data collection techniques and new data mining methods to meet the challenges and applications that are unique to the mobile arena (e.g., leveraging emerging embedded computing and sensing technologies to collect a large variety and volume of new kinds of "big data").

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2741987.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining; E.1 [**Data Structures**]: Graphs and networks; B.4.1 [**Data Communication Devices**]

General Terms

Algorithms, Design, Performance, Experimentation

Keywords

Mobile devices, mobile sensing, mobility data, mobile applications, data mining

1. TUTORIAL DESCRIPTION

Mobile devices and applications have been as important and disruptive as the personal computer and the Web. This tutorial focuses on analytics, mining, and sensing; and presents an in-context survey of background and methods useful to both practitioners and researchers. The tutorial consists of the following three one-hour sessions:

- Session 1: Overview of mobile technologies
- Session 2: Algorithms for mining mobility data
- Session 3: Applications of mobility data

Session 1: Overview of Mobile Technologies.

In the first session, we overview mobile technologies, summarizing the possibilities and challenges in collection of data, with a particular emphasis on sensing. This part includes:

- iOS and Android basics
- App structure
- Identity management
- Sensing capabilities and interfaces, APIs, power management, *etc*

Mobile app development follows a different paradigm than traditional desktop applications, and combines device-native and Web-based aspects. The paradigm is heavily event-driven, and many apps use standard HTTP APIs to retrieve and transmit data. Furthermore, mobile devices offer richer identity and activity tracking APIs, that can be leveraged by apps. These include device and user information, as well as sensing modalities that can collect all kinds of data (e.g., location, acceleration, sound, *etc*).

Session 2: Algorithms for Mining Mobility Data.

In the second session, we cover cross-cutting challenges such as real-time analysis, security, and outline cross-cutting algorithms for mobile data mining such as network inference, streaming algorithms, *etc.* This part primarily focuses on the following areas:

- Location-aware analytics: Where are you? What resources are nearby?
- Social analytics: Who is accompanying you? With whom are your communicating?
- Context-aware (activity) analytics: What are you doing?

We will also cover algorithms that combine the offline and online worlds in order to capture the user's intent better (e.g., by tracking customers in a physical store for analytics or promotion targeting).

Session 3: Applications of Mobility Data.

In the third (and last session), we focus on how algorithms for mobility data can be usefully applied to broad classes of applications. This part covers these areas:

- Mobile and location-based social
- Mobile advertising and search
- Urban computing and healthcare
- Privacy and security

We also overview "smart" applications that allow for better understanding of users at both the individual and aggregate levels. For instance, urban computing (understanding and summarizing how a city works) and mobile-based healthcare (understanding users behavior as it relates to health) are two such examples. We conclude Session 3 by discussing trends in mobile and Web apps; and how apps may meld the mobile and Web worlds.

Wrap up.

We conclude the tutorial by discussing prospects for new data collection techniques and new data mining methods that meet the unique challenges and applications in the mobile arena, and by suggesting additional promising directions for future research.

To summarize the tutorial, we aim to introduce data mining and machine learning researchers and practitioners to the broadly emerging mobile data mining and sensing technologies (focusing primarily on smartphones, but providing a brief list of other possibilities); show how analytic techniques can be developed and/or adapted for the exciting new class of applications that is quickly arising; and finally, cover some of the hottest emerging application areas.

2. RESOURCES AND REFERENCES

A list of resources−including tutorial slides and mobility data sets−and references categorized by topic area are available at `http://mobilemining.clusterhack.net/info`.

3. PRESENTERS

Spiros Papadimitriou is mainly interested are data mining for graphs and streaming data, clustering, time series, large-scale data processing, and mobile applications. His interests span from the very small (embedded devices, and sensors; Arduino) to the very large (large-scale data processing and analysis; Hadoop). He has published more than forty papers on these topics in refereed conferences and journals. He received the best paper award in SDM 2008, has three invited journal publications in best paper issues, several book chapters and he has filed multiple patents. He has also been invited to give keynote talks on graph and social network analysis (WAAMD 2008, and ADN 2009) and tutorials on time series stream mining (University of Maine Summer School, 2008) and large-scale analytics (Carnegie Mellon University, 2012). In the past, he has also developed and released a number of Android applications (including live-view mobile OCR, and Web service clients) that have 50,000 downloads . He is currently an assistant professor at Rutgers University (MSIS-RBS). Prior to that, he was a research scientist at Google, and a research staff member at IBM Research. He was a Siebel scholarship recipient in 2005. He obtained his BSc in Computer Science from the University of Crete, Heraclion and his MSc and PhD degrees from Carnegie Mellon University.

Tina Eliassi-Rad is an Associate Professor of Computer Science at Rutgers University. Before joining academia, she was a Member of Technical Staff and Principal Investigator at Lawrence Livermore National Laboratory. Tina earned her Ph.D. in Computer Sciences (with a minor in Mathematical Statistics) at the University of Wisconsin-Madison. Her current research lays at the intersection of graph mining, network science, and computational social science. Within data mining and machine learning, Tina's research has been applied to the World-Wide Web, text corpora, large-scale scientific simulation data, complex networks, fraud detection, and cyber situational awareness. She has published over 60 peer-reviewed papers (including a best paper runner-up award at ICDM'09 and a best interdisciplinary paper award at CIKM'12); and has given over 100 invited presentations. Tina is an action editor for the Data Mining and Knowledge Discovery Journal and a member of the editorial board for the Springer Encyclopedia of Machine Learning and Data Mining. In 2010, she received an Outstanding Mentor Award from the US DOE Office of Science.

4. ACKNOWLEDGMENTS

This work was supported in part by NSF CNS-1314603, by DTRA HDTRA1-10-1-0120, and by DAPRA under SMISC Program Agreement No. W911NF-12-C-0028.