**MINING SMARTPHONE AND MOBILITY DATA**

**Introduction**

Spiros Papadimitriou  
Rutgers University

Tina Eliassi-Rad  
Northeastern University

Katharina Morik  
TU Dortmund

Dimitrios Gunopulos  
University of Athens

**Tutorial plan**

11:00 – 11:55 Spiros  
• Introduction  
• Mobile technology overview

11:55 – 13:00 Katharina  
• Resource-constrained graphical models  
• App usage mining and traffic prediction  
• Lunch break (13:30-14:30)

14:30 – 15:15 Tina  
• Local-based social networks  
• Mobile advertising and search

15:15 – 16:10 Dimitrios  
• Learning from trajectory data  
• Crowdsourcing and applications

**Mobile devices**

Embedded sensors:  
• GPS & compass  
• Accelerometer & gyro  
• Proximity  
• Camera  
• Speech recognition  
• (Humidity, Temperature, Barometer/altimeter)  
• …

(more later)

**So what?**

It’s what I and many others have worked towards our entire careers. It’s just happening now:  
– Eric Schmidt (on cloud computing)

• The same could be said about mobile sensing and mining  
• Sensing & sensor networks  
• Ubiquitous computing  
• Mobility tracking  
• …

• But all are becoming mainstream now!
Mobile “vs” web

Modern applications:

- Browser (e.g., HTML5 + Javascript)
- Smartphone
- …
- Typically JSON (increasingly, authenticated)
- Amazon EC2
- Microsoft Azure
- Google GCE
- …

Identity / Authentication

E.g.: what is the difference between Facebook in your web-browser, vs Facebook on your smartphone

Not much:
- It’s the same backend & API, just running a different frontend

A lot:
- Access to content and data only on the device (e.g., photos, location, accelerometer, etc…)

Identity/authentication (e.g., OAuth): both users & apps

Example applications

- Geo-location
- Urban computing
- Recommendation
- Quantified self
- Healthcare
- Security
- …many more!

Geo-location

Example applications

What most people think (mainstream applications):

- Maps
- Navigation
- Local search (+ social)

- Context-based:
  - Locale: e.g., “if I’m within 0.5mi of work address and I have a meeting on my calendar, then set my phone to silent”
  - Google Now: “if I have a dentist appointment on my calendar, notify me when I need to leave, based on current traffic conditions, to be on time” or “if my email contains records of a booked flight, show flight status”
- Location reporting and sharing: Glympse, Google Latitude, etc.
Crowdsourcing
Example applications

Urban computing
Use broadly collected data for urban planning and analytics:
- Traffic monitoring and management
- Public transportation planning
- Zoning and planning
- Crisis detection and management
- Energy consumption sensing
- Air quality monitoring
- …

Much of this data comes from traces of mobile activity!

App usage / recommendation
- Smartphone activity is much more than a geolocation trace!
  - What apps you use and when
  - What actions you take within those apps

  Recommendations
  - Much richer data to leverage for better recommendations, e.g.,
  - Other things to do on or with your smartphone
  - …and beyond

Quantified self
Example applications
- Measure “self”, visualize, and correlate
- Idea dates back to 70s; term coined ~2007 by Kevin Kelly
- Both peripheral sensors as well as just apps; e.g.
  - Heart rate, Sleep quality
  - Weight, Activity
  - …

http://quantifiedself.com/

Healthcare

Privacy
Examples
- Vast data that allows quite accurate activity tracking or inferences
- Clearly raises privacy concerns
- Policy ( & technology ?)
Mobile mining

- The mobile "revolution" (like the "PC revolution") brings together many disciplines and touches many areas
- So, we had to draw some (occasionally arbitrary) divisions, and leave several things out

This tutorial focuses on:
- Work with a substantial analytics component
- Data collected via smartphones (although we’ll touch upon others sensors, we won’t go into sensing or ubiquitous computing territories—much)

Looking forward…

- Mobile phone penetration rapidly increasing
  - For many people, a smartphone will be their first computer
- All of these technologies are becoming mainstream
- Sensors are becoming cheaper and easier to hook up

So, what’s beyond (just) the mobile (smart)phone?

Mobile devices

- "IoT" (Body) sensors
  - Fidji et al.
  - HRMs (BLE)
  - Smart scales
  - Blood pressure

Network (Cellular, WiFi, Bluetooth, …)

Medical sensors
- Glucose (prick, continuous, …)
- ECG
- Inhaler use

Mobile devices

- IoT
- Smart locks
- Appliances
- Lights, temp., …
- Various "hacks"

"IoT" sensors

Medical sensors

Cheap...

Today:
- e.g., RFduino ($21)
  - Nordic ARM Cortex-M0 (32bit)
  - Bluetooth 4.0 (BLE)
  - or, ESP8266 ($2-3 !!)
- 32bit processor (Xlensia IP core)
- 2-4GB flash !
- Full WiFi & IP stack !!

Ten years ago:
- Mica Mote (Crossbow)
  - Atmel ATmega 103L @4MHz
  - 128KB flash / 4KB SRAM
  - 916MHz radio transceiver (38.4Kbps)
  - ~$300 per mote w/sensors

Cheap… and ubiquitous

- It’s easier than you think!
- Proliferation of open-source, open-hardware tools:
  - Arduino ecosystem (AVR and ARM), mbed, BeagleBone, RasPi, …
  - ESP8266, NodeMCU, …
  - Sensors in forms for easy prototyping (breakout boards, etc)
  - Wireless modules (BLE, Xbee, …)
- Very active hacker (maker) communities
- Cloud platforms (e.g., Imp, Xively, IFTTT, Spark Core, …)
- Some mainstream interest (e.g., Android Accesory APIs)
  - Co-design of sensing and analytics
    - Already a trend in mHealth

Larger picture: venues

In addition to data mining / web + social media venues:
- Medical health informatics
  - Many…
    - Good collection: http://mhealth.jmir.org/collection/view/51
- Ubiquitous computing
  - Mobile sensing workshop
  - Urban computing workshops & tutorials
  - New urban computing conferences
- Networking
  - PhoneSense
  - MobiCASE
- More: please check tutorial website!
- Many of these areas are starting "analytics" workshops
Tutorial plan

– 11:55 Spiros
  • Mobile technology overview
11:55 – 13:00 Katharina
  • Resource-constrained graphical models
  • App usage mining and traffic prediction

Lunch break
(13:30–14:30)

14:30 – 15:15 Tina
  • Local-based social networks
  • Mobile advertising and search
15:15 – 16:10 Dimitrios
  • Learning from trajectory data
  • Crowdsourcing and applications

Very interdisciplinary area, we had to leave many things out

Tutorial resources

http://mobilemining.clusterhack.net/

• Link to tutorial website also on conference website

• Please check for links to:
  • Slides,
  • References,
  • Datasets,
  • Other useful material

License

These slides are made available under a Creative Commons Attribution-ShareAlike license (CC BY-SA 3.0):
http://creativecommons.org/licenses/by-sa/3.0/

You can share and remix this work, provided that you keep the attribution to the original authors intact, and that, if you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

© 2013–6 Spiros Papadimitriou, Tina Eliassi-Rad