MINING SMARTPHONE MOBILITY DATA

Introduction

Spiros Papadimitriou, Tina Eliassi-Rad, Katharina Morik
Rutgers University   Northeastern   TU Dortmund

Tutorial plan

8:30 – 9:25 Spiros
- Mobile technology overview
- Mobile sensing: localization

9:25 – 10:05 Tina
- Local-based social networks
- Mobile advertising

10:05 – 10:30 Katharina
- Resource-constrained Graphical Models for App Usage Mining

Mobile devices

Smartphones

"IoT"

Network
(Cellular, WiFi, Bluetooth, ZigBee, …)

Sensors

Medical

Mobile devices: smartphones

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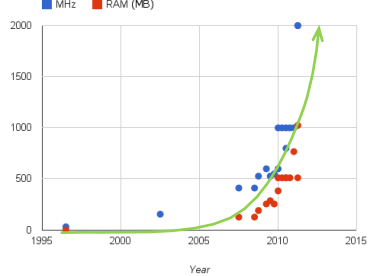- (Humidity, Temperature, Barometer/altimeter)
- …

(more later)

So what?

Smartphone characteristics
- MHz
- RAM (MB)

…you have a pretty powerful computer in your pocket!
…and it's connected!

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:

Identity / Authentication

Frontend ⟷ API ⟷ Backend

- Browser
  (e.g., HTML5 + Javascript)
- Smartphone
- …

- Typically JSON
  (increasincly, authenticated)

- Amazon EC2
- Microsoft Azure
- Google GCE
- …

Mobile "vs" web

Identity / authentication (e.g., OAuth):
- Users
- Applications & developers

- Mobile APIs for managing identity/accounts & content

Mobile "vs" web

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…)

"App"

- So… "app" vs "non-app" is maybe a better distinction...

- App has well-defined:
  - API (w/ semantics)
  - Entry points (controlled)
  - User identity (controlled)

- No longer entirely free (cf. web crawler vs Twitter firehose)
- Provide better UX and integration (cf... vs FB OpenGraph)
- Trade-off / balance: distributed and centralized (in organizational sense)

"Killer app for privacy/identity is 'social'" …

Example applications

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

Geo-location
Example applications

What most people think (mainstream applications):
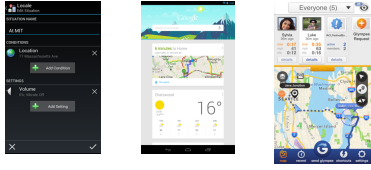
Google Maps   Waze   Yelp   Foursquare

- Maps
- Navigation
- Local search (+ social)

Geo-location
Example applications

Locale / Tasker · Google Now · Glympse

- 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.

Urban computing

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

Much of this data comes from traces of mobile activity!

[ICWSM 2016 Tutorial: "The Web of Cities and Mobility"]

Quantified self
Example applications

Withings devices · Sleep Cycle · Instant HR

- 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

Steve Mann's "wearable computer" and "reality mediator" inventions of the 1970s have evolved into what looks like ordinary eyeglasses.

(a) 1980 · (b) Mid 1980s · (c) Early 1990s · (d) Mid 1990s · (e) Late 1990s

Quantified self: log everything

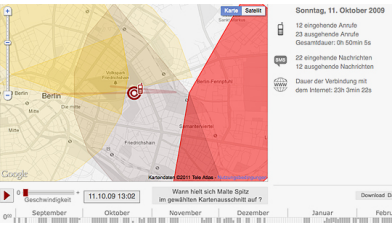Medical applications: glucose, asthma, ECG, …

- Related to quantified self
  - Many of these services can send data to your doctor
- Distinction: specific goal vs. "log everything" approach
- Micro-level (personal) and macro-level (population)

Privacy
Examples

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

"Tell-all telephone" – Die Zeit & Malte Spitz

Security & Malware
Examples

- Mobile malware: 6x [Juniper]
- E.g., BadNews: malware on Google Play (30+ apps, 2M downloads, fake app update prompts, mobile "pickpocketing")

Some challenges:
- Role mining: characterize groups of permissions more meaningfully
- Unusual activity detection

- Better: iOS-style permissions (now also on Android)
  - User asked when permission needed
  - Can grant/deny individual permissions

Mobile mining

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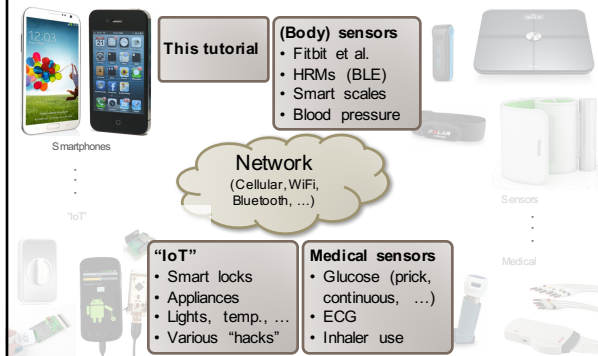- So, we had to draw some (occasionally arbitrary) divisions, and leave several things out

This tutorial focuses on:
- Work with a substantial analytics component
- Data collected via smartphones (although we'll touch on others sensors briefly, but 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

Smartphones

This tutorial

(Body) sensors
- Fitbit et al.
- HRMs (BLE)
- Smart scales
- Blood pressure

Network
(Cellular, WiFi, Bluetooth, …)

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

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

Sensors

Medical

Coming everywhere:

E.g., wearable mobile devices

…even in the shower!

Cheap…

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

Today:
e.g., RFduino ($21)
- Nordic ARM Cortex-M0 (32bit)
- Bluetooth 4.0 (BLE)

or, ESP8266 ($2-3 !!)
- 32bit processor (Xtensa IP core)
- 2-4GB flash !
- Full WiFi & IP stack !!

10-100x cheaper
More capable
Popular*

*RFduino raised $373K out of $5K goal on Kickstarter

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, IFTT, 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
  - New urban computing conferences
- Networking
  - PhoneSense
  - MobiCASE

- Many of these areas are starting "analytics" workshops

The rest of this tutorial

   – 9:25 Spiros
- Mobile technology overview
- Mobile sensing: localization

9:25 – 10:05 Tina
- Local-based social networks
- Mobile advertising

10:05 – 10:30 Katharina
- Resource-constrained Graphical Models for App Usage Mining

Very interdisciplinary area, we *had* to leave many things out

---

Tutorial resources

http://mobilemining.clusterhack.net/

- Link also on conference website

- These slides (handouts)
- Links
  - References,
  - Datasets,
  - Other useful material

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

MINING SMARTPHONE MOBILITY DATA

Introduction

Spiros Papadimitriou, Tina Eliassi-Rad, Katharina Morik
Rutgers University   Northeastern   TU Dortmund

License