RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

MINING DATA FROM MOBILE DEVICES

Applications: Urban, Healthcare

Spiros Papadimitriou, Tina Eliassi-Rad

Overview

- **Urban applications**
- Healthcare

Urban computing / “smart cities”

Emerging field of study on:

- Use of technology in public environments
- Interaction between humans and environments
- Leveraging heterogeneous sensing technologies

Multi-disciplinary: CS/EE, architects, urban planners, social scientists, artists, interaction designers, …

Some technical challenges:

- Integration of heterogeneous data
- Data management and analysis
- Visualization methods

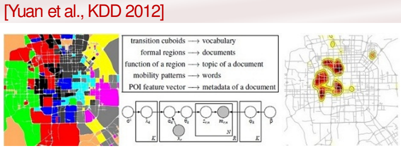
Urban computing

This part

- Again, we necessarily need to leave many things out
- Overview of some work that provides good examples of what is possible/interesting, with emphasis on smartphone/mobile data
- Far from exhaustive!

City zone characterization

[Yuan et al., KDD 2012]

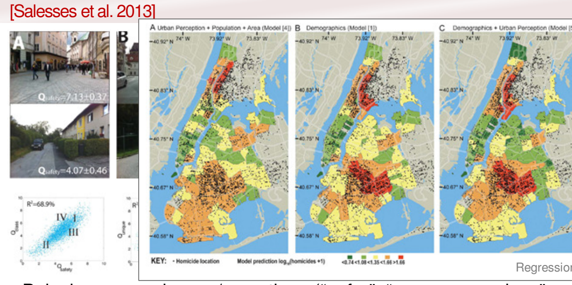


- Partition city in regions (based on major roads)
- Apply topic models:
  - Document = region
  - Mobility patterns = terms
  - POIs = metadata
  - Region function = topic
- Result:
  - Zone → Mixture of functions (topics)
  - Function → Distribution over mobility patterns (terms)

Jing Yuan, Yu Zheng, Xing Xie. *Discovering regions of different functions in a city using human mobility and POIs*, KDD 2012.

Urban zones and perception inequality

[Salesses et al. 2013]

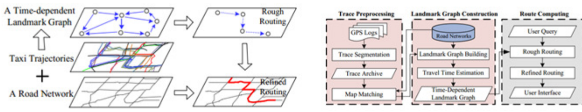


- Pairwise comparison w/ questions (“safer”, “more upper-class”, etc) on city images: Boston, New York, Linz, Saltzburg
- Correlated with violent crime rates (after controlling for income, population, area, and age)

Philip Salesses, Katja Schechtner, and César A. Hidalgo. *The Collaborative Image of the City: Mapping the Inequality of Urban Perception*, PLoS ONE 8(7): e68400, 2013

Traffic pattern inference

[Yuan et al., GIS 2010]



- Travel time depends on:
  - Route features (distance, traffic lights, etc)
  - Time-dependent traffic flow
  - Driver behavior
- Leverage GPS trajectory logs from taxis: experienced drivers, make good decisions
- Save 5min per 30min
- Landmark-based hierarchical routing (common approach)

Jing Yuan, Yu Zheng, et al., *T-Drive:Driving Directions Based on Taxi Trajectories*,GIS 2010 (best paper runner up).

Public transit – taxi analytics

[Ge et al., KDD 2011 (demo)]

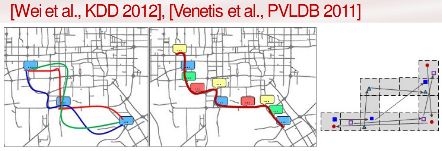
Interface Fraud detection

- Integrates various analytics (SF & NYC taxis):
  - Pick up and route recommendation (for fleet)
  - Driving fraud detection
  - Driver performance (occupancy rate)

Yong Ge, Chuanren Liu, Hui Xiong, Jian Chen: A taxi business intelligence system. KDD 2011: 735-738

Local search/social data analysis

[Wei et al., KDD 2012], [Venetis et al., PVLDB 2011]



- GPS position is fairly accurate
- However, much position data is uncertain (coarse); e.g.: check-ins, photo geo-tags, etc.
- Aggregate uncertain trajectories, to infer more accurate top-k trajectories
- Another example: POI ratings are strongly correlated with how frequently it appears as driving directions endpoint

Ling-Yin Wei, Yu Zheng, Wen-Chih Peng, *Constructing Popular Routes from Uncertain Trajectories*, KDD 2012

Petros Venetis, Hector Gonzalez, Christian S. Jensen, Alon Y. Halevy: *Hyper-local, directions-based ranking of places*. PVLDB 4(5): 290-301, 2011

Overview

- Urban applications
- **Healthcare**

mHealth

Mobile health (mHealth) seeks to improve individuals' health and well-being by:

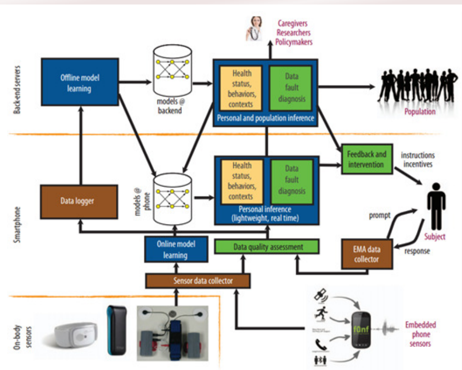
- Continuously monitoring their status
- Rapidly diagnosing medical conditions
- Recognizing behaviors
- Delivering just-in-time interventions [Kumar et al., 2013]

Multi-disciplinary, involving at least:

- Sensing
- On-device analytics
- Off-device (backend) analytics

Santosh Kumar, Wendy Nilsen, Misha Pavel, Mani Srivastava, *Revolutionizing Healthcare Through Transdisciplinary Research*, IEEE Computer, 2013

mHealth



mHealth

Some numbers [AlliedHealthWorld.com]:

- About 40,000 mobile health apps
- Over 500 projects with mobile emphasis
- Number of people who downloaded health apps doubled (124M in 2011, to 247M in 2012)
- Growing business; global revenue $1.2B in 2011 (projected to $12B in 2018)

Where is most of the action?

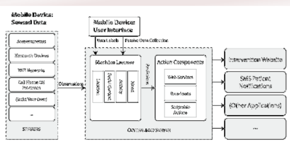
- Access to health data & tracking
- Use of peripheral sensors (e.g., BLE)
- Social / gamification

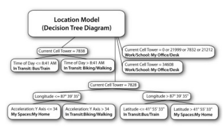
mHealth

This part

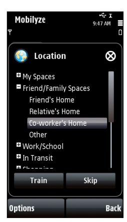
- This tutorial: smartphone and ML/DM bias
  - Much of the action is in peripheral sensors
- A far from exhaustive list of examples

Context-sensing for depression intervention





- Machine learning models (ie, learners) for:
  - Mood, emotions, cognitive/motivational states,
  - Activities, environmental context, and social context
  - >38 concurrent sensor values (eg, GPS, light, calls)
- Accuracy: 60-91% context, poor for mood
- Still, participants improved significantly, less likely to suffer depressive episodes



MN Burns, M Begale,J Duffecy, D Gergle, CJ Karr, E Giangrande, DC Mohr, *Harnessing Context Sensing to Develop a Mobile Intervention for Depression*, JMIR 2013

Activity recognition (again)

(off-board processing)

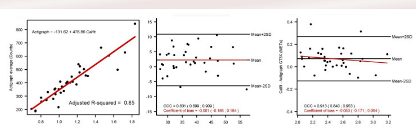
Activity	kNN[a]	J48[b]	MLP[c]	Logistic	NB[d]	Boosting	Bagging	Accel.	Accel. + Gyro	Diff.
C1. Slow walking	94.1%	86.3%	90.8%	88.3%	61.3%	94.1%	94.1%	89.6%	94.1%	+4.5%
C2. Normal walking	92%	80.9%	84.6%	74.2%	55.7%	92%	92.2%	85.8%	92%	+6.2%
C3. Brisk Walking	90.1%	82.2%	85%	68.7%	64.9%	89.9%	90.1%	78%	90.1%	+12.1%
C4. Jogging	91.7%	91.7%	91.5%	92.2%	79%	92.2%	91.7%	85.4%	91.7%	+6.3%
C5. Sitting	100%	99.6%	100%	100%	98.5%	100%	100%	100%	100%	0%
C6. Normal upstairs	69.8%	51%	42.7%	47.9%	30.2%	69.8%	69.8%	65.6%	69.8%	+4.2%
C7. Normal downstairs	79.4%	64.9%	54.6%	46.4%	32%	79.4%	77.3%	66%	79.4%	+13.4%
C8. Brisk upstairs	70.4%	69%	33.8%	19.7%	22.5%	70.4%	69%	64.8%	70.4%	+5.6%
C9. Brisk downstairs	52.3%	44.6%	24.6%	33.8%	35.4%	52.3%	43.1%	49.2%	52.3%	+3.1%
Weighted average	90.2%	83.0%	83.4%	77.2%	63.2%	90.2%	89.9%	83.7%	90.2%	+6.5%

$\underbrace{\text{Accel. / Accel. + Gyro / Diff.}}_{k\text{-NN}}$

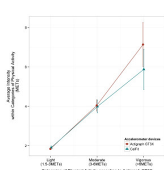
- This time from “medical” perpective
- Extensive experimental study of various out-of-the box techniques

W Wu, S Dasgupta, EE Ramirez, C Peterson, CJ Norman, *Classification Accuracies of Physical Activities Using Smartphone Motion Sensors*, JMIR 2013

Activity measures



Q: How good is a smartphone vs. a dedicated accelerometer?
A: Not bad!



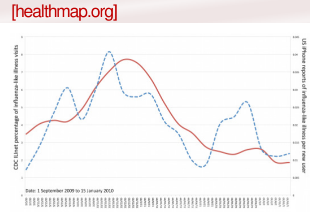
Differences:

- 10 vs 30 Hz
- 158 vs 27 g (but: always with you)
- 18 hours vs 31 days battery

D Donaire-Gonzalez, A de Nazelle, E Seto, M Mendez MJ Nieuwnhuijsen M Jerrett, *Comparison of Physical Activity Measures Using Mobile Phone-Based CalFit and ActiGraph*, JMIR 2013

Disease spread monitoring

[healthmap.org]



- Larger effort, started 2006, to integrate informal sources for
  - Disease outbreak monitoring
  - Real-time surveillance of emerging public health threats
- “Outbreaks Near Me” app (2009, 110K D/L, ~2.5K reports)
  - 95% H1N1-related, closely track CDC (but: real-time vs ~1 week lag)

C Freifeld, R Chunara, SR Mekaru, EH Chan, T Kass-Hout, AA Iacucci, JS Brownstein , *Participatory Epidemiology: Use of Mobile Phones for Community-Based Health Reporting*, PloS Medicine. 2010.

Disease spread monitoring

Other examples

- Search query and Twitter feed tracking
- Craigslist data to identify behavior patterns associated with increases in syphilis cases
- FourSquare + Twitter for food poisoning
- …

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

MINING DATA FROM MOBILE DEVICES

Applications: Urban, Healthcare

Spiros Papadimitriou, Tina Eliassi-Rad