Supporting mobility analysis with crowd-sourced data - Opportunities and challenges

Jean Damascene Mazimpaka
University of Augsburg, Germany
Problem and objective (1)

Mobility in real world

Individual trajectory
Problem and objective (2)

- Mobility data are used in many applications
  - Transportation: traffic management, routing
  - Urban planning: land use and infrastructure utilisation monitoring
  - Environment: air pollution and noise pollution control
  - Business: advertisement, choice of new business location
  - ...

- However, such applications require high level information:
  - characteristics of places visited
  - events that took place at the time of visit
  - characteristics of the geographic space where the movement took place
  - ...

- Surveys are expensive to deliver such information
- Manual annotation by the moving individuals is not feasible
- What about crowd-sourcing?
Opportunities (1)

- Increasing people’s participation to producing geo-referenced crowdsourced data
Opportunities (2)

• Produced crowdsourced data are rich in semantics
  ▪ social network data
    — main data + social structure
    — Social structure given a high importance
    — Example:
      check-ins (https://foursquare.com/)
      mainly text (http://www.twitter.com/)
      mainly text (http://www.facebook.com/)
      photos (http://www.flickr.com/)
  ▪ contribution-focused data
    — main data (+ social structure)
    — If available, social structure given low importance
    — Example:
      noise level data (http://www.noisetube.net/#&panel1-1)
      vehicle emission data (https://www.envirocar.org/)
      pleasantness of urban locations (http://urbangems.org/)
      geo-referenced pages (http://en.wikipedia.org/wiki/Main_Page)
      geographic features (http://www.openstreetmap.org)
Opportunities (3)

- Availability of APIs for accessing crowd-sourced data

---

### API Methods

**Read these first:**
- Developer Guide
- Overview
- Encoding
- User Authentication
- Dates
- Tags
- URLs
- Buddyicons
- Flickr APIs Terms of Use

**API Keys**

**Developers’ mailing list**

---

The Flickr API is available for non-commercial use by outside developers. Commercial use is possible by prior arrangement.

**activity**
- flickr:activity.userComments
- flickr:activity.userPhotos

**auth**
- flickr:auth.checkToken
- flickr:auth.getRob
- flickr:auth.getFullToken
- flickr:auth.getToken

**auth.oauth**
- flickr:auth.oauth.checkToken
- flickr:auth.oauth.getAccessToken

---

- Root
- Friends
  - /users/username/friends
  - /users/username/friends/:friend
- Tracks
  - /tracks
  - /tracks/:trackid
  - /users/username/tracks
  - /users/username/tracks/:trackid
- Phenomenons
  - /phenomenons
  - /phenomenons/:phenomenon
- Sensors
  - /sensors
  - /sensors/:sensor
Challenges (1)

- Data sparsity
- Data representativeness
Challenges (2)

• Limitations associated with the available APIs, e.g.:
  – In Foursquare, view friends of an individual but cannot view the check-ins of a specific user
  – In Flickr, limit at 3600 queries per hour and per key

• Motivation
  – contribution-focused data may require:
    o acquisition of specific devices (e.g. sensors)
    o more efforts
  – for social network data
    • Establishing and maintaining a network is a one motivation

  Need for (further) motivation
Challenges (3)

• Incompleteness due to limited quality control

• Location uncertainty due to:
  – accuracy of the device used for location recording
  – processing done at the crowdsourcing platform that can filter or modify geographic information
  – the credibility of the user generating the data who can change geographic coordinates intentionally
  – the difference between user and content locations e.g. a Flickr photo showing Mount Everest with coordinates on another mountain
Addressing the challenges (1)

• **Collaborative filtering (CF)** for addressing data sparsity
  – commonly used in recommendation systems
  – idea behind: similar users make ratings in a similar manner for similar items.
  
  → similarity is determined between users and between items, a prediction can be made to the rating of a user about future items (Nakamura and Abe, 1998)

• **Data pre-processing** for addressing the uncertainty
  – *Filtering* outlying data items and those of obviously wrong locations or values
  – *density of contributions* as one indicator of the level of certainty
Addressing the challenges (2)

- Integrating data from multiple sources for addressing data representativeness, incompleteness, and sparsity
  - But addressing also the problems due to the integration
    e.g. Conflicting information solved by a voting strategy (Li et al. 2013)
  - Evaluate different integration approaches to choose most effective and best performing
Addressing the challenges (3)

• Implementing incentive mechanisms for addressing the motivation challenge:
  
  ▪ Choose suitable incentive mechanisms from available proposals (examples in Quinn and Bederson, 2011)

  ▪ Further examples:
    - Crowd-sourcing platforms imparting reputation to contributors indirectly
    - Providing interesting applications based on contributed data with additional features for contributors, and advertising these features
Conclusion

• Crowd-sourced data have a high potential to support mobility analysis.

• However, several challenges need to be addressed.

• Our future work will concentrate on challenges related to integrating data from multiple sources:
  
  – Applying ontological modelling approaches for simplifying the integration
  
  – Integrating social network activities of a user from different social platforms to fill the gap caused by using only one source
Thank you for your attention
References (1)


References (2)

