Privacy and anonymity in mobility data analysis

Fosca Giannotti, Dino Pedreschi, Franco Turini
KDD Lab Pisa, Italy
Knowledge Discovery & Data Mining Lab
ISTI/CNR – Dipartimento di Informatica/ University of Pisa

http://www-kdd.isti.cnr.it
Programma dell’incontro

11:00-12:30

- L’analisi dei dati di mobilità e lo scenario del progetto GeoPKDD, Fosca Giannotti
- Privacy ed anonimato nell’analisi dei dati e nel data mining, Dino Pedreschi
- Analisi previsionale e discriminatorietà, Franco Turini
- L’Osservatorio sulla Privacy di GeoPKDD

12:30-13:30

- Discussione
Plan of the talk (Mobility Data Analysis)

- The wireless explosion:
  - Location Based Services vs Mobility Data Analysis
  - Analytic opportunities

- GeoPKDD vision and goals
  - The movement patterns
  - The impact

- The scenario of ubiquitous computing

- The privacy challenge
The Wireless Explosion

Do you use any of these devices?

Do you ever feel that you are tracked?
The Wireless Network

- The pervasiveness of mobile and ubiquitous technologies is increasing day after day
  - GSM wireless phone networks
    - 1.5 billions in 2005, still increasing at a high speed
    - Italy: # mobile phones ≈ # inhabitants
  - GPS and Galileo positioning systems
  - Wi-Fi and Wi-Max wireless networks
  - RFID’s and sensor networks

- miniaturization
- positioning accuracy
  - location technologies capable of providing increasingly better estimate of user location
Which new opportunities?

- Location based services:
  - A certain service that is offered to the users based on their locations

- Mobility data analysis:
  - Discovering knowledge from the digital traces of our mobile activity to support decision making in mobility related issues.
Location-based Services: Then

- Limited to fixed traffic signs
Location-based Services: Now

- Location-based traffic reports:
  - **Range query:** How many cars in the free way
  - **Shortest path query:** What is the estimated time travel to reach my destination

- Location-based store finder:
  - **Range query:** What are the restaurants within five miles of my location
  - **Nearest-neighbor query:** Where is my nearest fast (junk) food restaurant

- Location-based advertisement:
  - **Range query:** Send E-coupons to all customers within five miles of my store
Mobility data analysis

- How people move around in the town
  - During the day, during the week, etc.
- Are there typical movement behaviours?
- Are there typical movement behaviours in a certain area at a certain time?
- How frequently people access the network?
- How are people movement habits changing in this area in last decade-year-month-day?
- Are there relations between movements of two areas?
- Are there periodic movements?
From *time-geography*

The representation of Napoleon’s Russian campaign of 1812 produced by Charles Joseph Minard in 1861.
to interactive (recent/real) time-geography
Real-time density estimation in urban areas

The senseable project: http://senseable.mit.edu/grazrealtime/
Plan of the talk (Mobility Data Analysis)

The wireless explosion:
- Location Based Services vs Mobility Data Analysis
- Analytic opportunities

GeoPKDD vision and goals
- The movement patterns
- The impact

The scenario of ubiquitous computing

The privacy challenge
A paradigmatic example:
GeoPKDD
A European FP7 project
www.geopkdd.eu
Geographic Privacy-aware
Knowledge Discovery and Delivery
GeoPKDD general goal

is to develop

- theory, techniques and systems for geographic knowledge discovery and delivery,

- based on new automated privacy-preserving methods for extracting user-consumable forms of knowledge from large amounts of raw data referenced in space and in time.
Geographic privacy-aware Knowledge Discovery process

Geographic privacy-aware Knowledge Discovery process

Aggregative Location-based services

Telecommunication company (WIND)

Privacy-aware Data mining

Trajectory reconstruction

Trajectories warehouse

Privacy enforcement

Bandwidth/Power optimization

Mobile cells planning

... interpretation visualization ST patterns

Traffic Management

Accessibility of services

Mobility evolution

Urban planning

Public administration or business companies

GeoKnowledge
From movement data to movement patterns
From movement data to movement patterns
Mining Trajectories: Clustering

- Group together similar trajectories
- For each group produce a summary
Mining Trajectories: Frequent patterns

- Discover frequently followed (sub)paths
Mining Trajectories: classification models

- Extract behaviour rules from history
- Use them to predict behaviour of future users

![Diagram showing distribution of user behavior with percentages: 20%, 7%, 60%, 5%, 8%, and ?]
The GeoPKDD impact

Improving decision-making in mobility-related issues:

- Planning traffic and public mobility systems in metropolitan areas;
- Planning physical communication networks
- Localizing new services in our towns
- Forecasting traffic-related phenomena
- Organizing logistics systems
- Avoid repeating mistakes
- Timely detecting changes.
Plan of the talk (Mobility Data Analysis)

The wireless explosion:
- Location Based Services vs Mobility Data Analysis
- Analytic opportunities

GeoPKDD vision and goals
- The movement patterns
- The impact

The scenario of ubiquitous computing

The privacy challenge
From wireless networks to Ubiquitous Computing environments

- Log data from mobile phones, i.e. sampling of localization points in the GSM/UMTS network.
- Log data from GPS-equipped devices
- Log data from
  - peer-to-peer mobile networks
  - intelligent transportation environments
  - ad hoc sensor networks, RFIDs
- Increasing precision and pervasiveness
Plan of the talk (Mobility Data Analysis)

The wireless explosion:
- Location Based Services vs Mobility Data Analysis
- Analytic opportunities

GeoPKDD vision and goals
- The movement patterns
- The impact

The scenario of ubiquitous computing

The privacy challenge
Why emphasis on privacy?

- More, better data are gathered, more vulnerability from linkage
- On the other hand, more and new data bring new opportunities
  - Public utility, new markets/paradigms, new services
- Need to maintain privacy without giving up opportunities
- Need to obtain social acceptance through demonstrably trustworthy solutions
Privacy in Mobility Data and Services

- Trusted/secure storage/Management of Mobility Data

- Privacy in Location Based Services:
  - the right of a user to receive a service without revealing his/her identity
  - Trade-off between quality of service and privacy protection

- Privacy and Anonymity in Mobility Data Analysis
  - Trade-off between privacy protection and analysis opportunities
Privacy in GeoPKDD

- How to design Data Analysis methods that, by construction, meet the the privacy constraints?

- How to develop trustable data mining technology capable of producing
  - provably/measurably privacy-preserving patterns
  - which may be safely distributed
Scientific Privacy Issues in GeoPKDD

Is there any specific challenge/risk/opportunity in the context of ST data?

- New threats from traces analysis: learning who you are from where you have been (Malin et al. 2003)
- Space and Time in a trajectory can act as quasi-identifiers (Bettini and Jajodia 2005)

How to formalize privacy measures over Spatio-Temporal data and Spatio-Temporal patterns?

- E.g., anonymity threshold on clusters of individual trajectories
Ethical, Legal and Societal Privacy Issues in GeoPKDD

- Harmonization with national privacy regulations and authorities – privacy observatory
- Privacy Observatory (see our last talk)