Tutorial on Spatial and Spatio-Temporal Data Mining (SBBD-2008)

Introduction to Moving Data and Moving Object Databases

Vania Bogorny
Universidade Federal do Rio Grande do Sul
www.inf.ufrgs.br/~vbogorny
vbogorny@inf.ufrgs.br

Shashi Shekhar
University of Minnesota
www.cs.umn.edu/~shekhar
shekhar@cs.umn.edu

Outline

- The wireless explosion
- Moving Object Data and Mobility Data Analysis
- Moving Object Databases
- Spatio-Temporal Data Mining for Moving Object Trajectories

The Wireless Explosion

- A lot of effort has been devoted to the development of efficient software for mobile devices as
  - Palm-tops
  - GPS
  - Cell Phone
  - Sensor networks, etc.
- ....while....
- very little has been done to analyze the data generated by mobile devices: *Trajectories of Moving Objects* **(new kind of data)**

The Wireless Explosion (Fosca Giannotti 2007 – www.geopkdd.eu)

Do you use any of these devices ?

Do you ever feel that you are tracked?

The Wireless Explosion

Trajectories of Moving Objects

Location-based Services: 10 years ago (Fosca Giannotti 2007 – www.geopkdd.eu)

- **Limited to fixed traffic signs**

Location-based Services: Now (Fosca Giannotti 2007 – www.geopkdd.eu)

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

- Store finder:
  - ***Range query:*** What are the restaurants within five KM of my location?

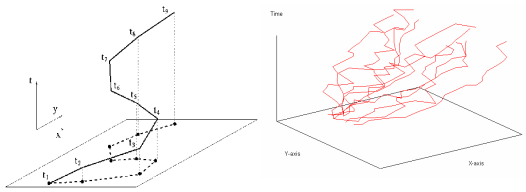
- Advertisement:
  - ***Range query:*** Send E-coupons to all customers within five KM of my store

Mobility Data Analysis

- How people move around in the town
  - During the day, during the week, etc.
- Are there typical movement behaviours? In a certain area at a certain time?
- 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?

Trajectory Data (Fosca Giannotti 2007 – www.geopkdd.eu)

- Spatio-temporal Data
- Represented by a set of points located in space and time (time-stamped coordinates)
- $T=(t_1,x_1,y_1), \ldots, (t_n, x_n, y_n)$ => position at time t_i was (x_i,y_i)



Trajectories: Basic Concepts (Adrienko 2008)

- Trajectories have to be represented by finite sequences of time-referenced locations. There are several ways to observe movements:
  - **time-based recording:** positions of entities are recorded at regularly spaced time moments
    - e.g. every 5 minutes
  - **change-based recording:** a record is made when the position of an entity differs from the previous one;
  - **location-based recording:** records are made when an entity comes close to specific locations
    - e.g. where sensors are installed
  - **event-based recording:** positions and times are recorded when certain events occur
    - e.g. calling by a mobile phone
- Typically, positions are measured with uncertainty.

Trajectories: Overall Characteristics (Adrienko 2008)

1. Geometric shape
2. Length (traveled distance)
3. Duration (in time)
4. Speed
   - Mean, median, and maximal Speed
   - Periods of constant speed, acceleration, deceleration
5. Direction:
   - Periods of straight, curvilinear, circular movement;
   - Major turns ('turning points') in: time, position, angle, initial and final directions, and speed in the moment of the turn;

Relationships

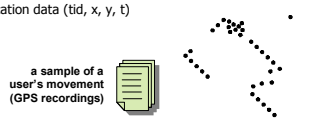
Many types of relationships may be of interest, depending on the problem in hand:

- similarity or difference of the overall characteristics of the trajectories
  e.g. shapes, travelled distances, durations, dynamics of speed and directions)
- spatial and temporal relationships:
  - co-location in space (i.e. the trajectories consist of the same positions or have some positions in common):
  - co-existence in time (i.e. the trajectories are made during the same time period or the periods overlap);
  - co-incidence in space and time (i.e. same positions are attained at the same time);
  - distances in space and in time.

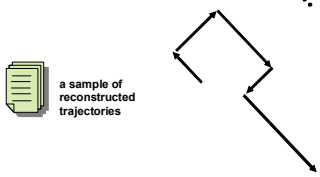
Raw Trajectory Data: Problems and Solutions

The trajectory reconstruction problem (Theodoridis and Peleikis 2007)

- From raw location data (tid, x, y, t)

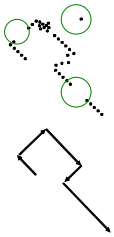


- To trajectory data (obj-id, traj-id, (x, y, t))



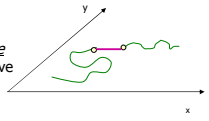
Trajectory stream manager (Theodoridis and Peleikis 2007)

- Trajectory stream manager operations
  - reconstructs trajectories (excluding noise, etc.) and posts trajectory data to a MOD (Moving Object Database)
- Results so far – 2 alternatives
  - Assumptions about trajectory 'birth' (for spatial/temporal gaps between traces)
  - Studying the notion of 'stop' (suspension of an entity's movement)

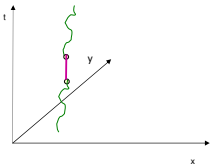


Trajectory stream manager (now…) (1) (Theodoridis and Peleikis 2007)

- When will an object have assigned a new trajectory-id?
  - When there is sufficiently *large gap in the spatial dimension* between two consecutive recorded positions



  - When there is sufficiently large gap in the temporal dimension between two consecutive recorded positions



Moving Object Databases

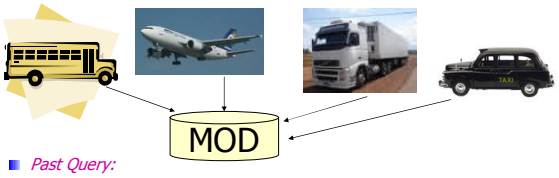
Moving Object Databases: History

- Is a quite new research field
  - Main researchers: Ralf Harmut Guting and Ouri Wolfson
  - More recently, Yannis Theodoridis/ Nikos Pelekis

Moving Object Databases: Applications

- Traffic Analysis and Management
- Analysis of Movements of People (Customers)
- Environmental Studies
- Biology (E.G. Animal Behaviour, Tracking)
- Meteorology
- History
- ....

Moving Object Database Technology (Wolfson 1999)



- *Past Query:*
  - During the past year, how many times was bus *5* late in more than 10 minutes at station *X*?

  Trigger:
  Send me a message when an airplane arrives in a given geographic area
- *Future Query*:
  Which trucks will reach destination within 20 minutes?
- *Present Query*: Where are the taxis within 1 KM of my location?

Moving Object Database: Operations

- Research on *Moving Object Databases* (MODs) has addressed the need for representing movements of objects to:
  - **perform ad-hoc querying**
  - **analysis**
  - **data mining**
- During the last decade, research has focused on:
  - **data models**
  - **query languages**
  - **implementation aspects (e.g. efficient indexing, query processing and optimization techniques)**

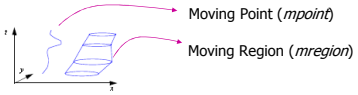
Moving Object Database Prototypes

- SECONDO – Ralph Guting (Germany)
- HERMES – Yannis Theodoridis and Nikos Pelekis (Greece)

Secondo

University of Hagen

Data Types (Guting 1999)

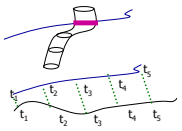


- Data Types *mpoint* and *mregion* are mappings from time into space:
  - *mpoint = time → point*
  - *mregion = time → region*
- Examples:
  - **flight (id: string, origin: string, dest: string, route: mpoint)**
  - **weather (id: string, kind: string, area: mregion)**

Spatio-Temporal Operations (Guting 1999)

The data types include suitable operations such as:

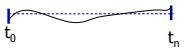
- ***Intersection*** (*mpoint, mregion*) → ***mpoint***



- ***distance*** (*mpoint, mpoint*) → ***mreal***

- ***Trajectory*** (*mpoint*) → ***line***

- ***Deftime***(*mpoint*) → ***period***



- ***length*** (*line*) → ***real***

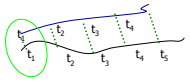
Spatio-Temporal Queries

- **flight (id: string, origin: string, dest: string, route: mpoint)**
- Query 1: "Find all flights from Sao Paulo that are longer than 4000 km."

```
SELECT *
FROM flights
WHERE origin = 'SP' AND lenght(trajectory(route)) > 4000
```

- Query 2: "Retrieve any pairs of air planes that during their flight came closer to each other than 500 meters!"

```
SELECT f.id, g.id
FROM flights  f, flights  g
WHERE f.id <> g.id AND min(distance(f.route, g.route)) < 0.5
```

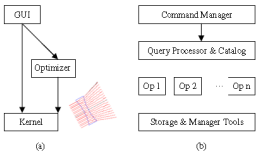


General Characteristics

- An environment for research prototyping and teaching
- Developed in the last ten years at University of Hagen, Germany
  - system frame can be filled with implementations of different data models,
    - e.g. relational, object-relational, etc
- Implemented in algebra
- Basically a relational system with several advanced data type packages
- Open source software, available at
- **http://dna.fernuni-hagen.de/Secondo.html/index.html**

Architecture

- Consists of:
  - (a) a *kernel*, which offers query processing over a set of implemented algebras,
  - (b) an *optimizer*, which implements the essential part of an SQL-like language, and
  - (c) an *extensible* GUI where new data types and models can provide specialized views for moving objects



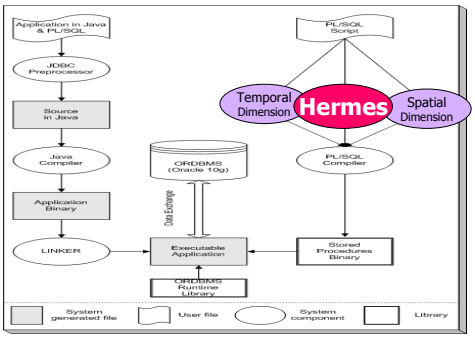
Hermes

University of Pireaus

Hermes - Engine (Theodoridis and Peleikis 2007)

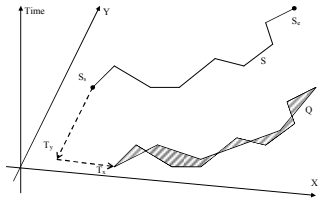
- Principles
  - Spatial and temporal dimensions
  - HERMES Moving Data Cartridge (MDC) On top of Oracle extensible DBMS
- Trajectory management
  - Insert/Update/Delete a moving object or a segment of its trajectory
  - Works over trajectories or sets of trajectories
- Data management
  - Supported indices: R-tree
  - Ongoing task: Development of a specialized index (TB-tree)
- Flexible system

HERMES prototype architecture (Theodoridis and Peleikis 2007)



Hermes - Similarity Queries (Theodoridis and Peleikis 2007)

- *Spatial similarity* (time is not taken into consideration):
  - Find objects whose *route* is quite similar to that of object id=132 (irrespective of time).
- *Spatio-temporal similarity* (time is taken into consideration):
  - Find objects that follow a route similar to that of object id=132 during the same time interval, e.g. from 3 to 6 pm.
- *Speed-pattern similarity*:
  Find objects moving with a similar speed pattern
  - e.g. mid speed for about half an hour and high speed for the next two hours
- *Direction-pattern similarity*:
  Find objects moving with a similar direction pattern
  - e.g. NE during the first half of the route and then W



Distance-Based & NN Queries (Theodoridis and Peleikis 2007)

- Distance-based query
  - stationary reference object
    - Find shops close to me (e.g. less than 100m) offering sportswear
  - moving reference object
    - Find humans who have passed close (e.g., less than 100 m) to me and have already requested sportswear"
- Nearest-Neighbor query
  - stationary reference object
    - Find the two nearest shops to my current location
  - moving reference object
    - Find the two nearest trajectories to my route

Distance and buffer

- Generates a buffer polygon around a point at a specific timepoint
  - **f_buffer** (distance NUMBER, tolerance NUMBER, tp tau_tll.d_timepoint_sec) RETURN MDSYS.SDO_GEOMETRY,
- Computes the distance between two moving points at a specific timepoint. The distance between two geometry objects is the distance between the closest pair of points or segments of the two objects
  - **f_distance** (moving_point moving_point, tolerance NUMBER, tp tau_tll.d_timepoint_sec) RETURN NUMBER,
- Determines if a moving point is within some specified Euclidean distance from other geometry objects at a specific timepoint
  - **f_within_distance** (distance NUMBER, geom MDSYS.SDO_GEOMETRY, tolerance NUMBER, tp tau_tll.d_timepoint_sec) RETURN VARCHAR2,

Direction Relationships

- Returns TRUE for objects being east of moving point at the given timepoint
  - ***f_east*** (geom MDSYS.SDO_GEOMETRY, tp tau_tll.d_timepoint_sec, angle_min NUMBER, angle_max NUMBER) RETURN NUMBER,
- Returns TRUE for objects being west of moving point at the given timepoint
  - ***f_west*** (geom MDSYS.SDO_GEOMETRY, tp tau_tll.d_timepoint_sec, angle_min NUMBER, angle_max NUMBER) RETURN NUMBER,
- Returns TRUE when the moving point is between the multi-geometry at the given timepoint
  - ***f_between*** (geom MDSYS.SDO_GEOMETRY, tp tau_tll.d_timepoint_sec) RETURN NUMBER,
- Returns TRUE for objects being in front of moving point at the given timepoint
  - ***f_front*** (geom MDSYS.SDO_GEOMETRY, tp tau_tll.d_timepoint_sec, angle_min NUMBER, angle_max NUMBER) RETURN NUMBER,
- Returns TRUE for objects being behind of moving point at the given timepoint
  - ***f_behind*** (geom MDSYS.SDO_GEOMETRY, tp tau_tll.d_timepoint_sec, angle_min NUMBER, angle_max NUMBER) RETURN NUMBER,

Topological Queries

Theodoridis and Peleikis 2007

- Find humans *crossing* a street
- Find areas whose boundaries *touch* humans' routes
- When did a human enter the area of his departing gate?
- Which was the entering point?

Topological Relationships

- Returns a geometry object that is the topological intersection (AND operation) of an instanced point at a specific timepoint with another geometry object
  - **f_intersection** (geom MDSYS.SDO_GEOMETRY, tolerance NUMBER, tp tau_tll.d_timepoint_sec) RETURN MDSYS.SDO_GEOMETRY,
- Returns a moving point that is the restriction (intersection) of the calling moving point inside the polygon argument
  - **f_intersection** (geom MDSYS.SDO_GEOMETRY, tolerance NUMBER) RETURN moving_point,
- Returns a geometry object that is the topological union (OR operation) of an instanced point with this moving point at a specific timepoint
  - **f_union** (moving_point moving_point, tolerance NUMBER, tp tau_tll.d_timepoint_sec) RETURN MDSYS.SDO_GEOMETRY,
- Examines current Moving_Point to determine its spatial relationship with another moving point
  - **f_relate** (MASK VARCHAR2, moving_point moving_point, tolerance NUMBER, tp tau_tll.d_timepoint_sec) RETURN VARCHAR2,

General Spatio-temporal Relationships

- Return the enter and leave points of the moving point for a given geometry
  - get_enter_leave_points (geom MDSYS.SDO_GEOMETRY) **RETURN MDSYS.SDO_GEOMETRY,**
- Returns the Point (spatial coordinates) of the Moving_Point at a specific timepoint
  - at_instant (tp tau_tll.d_timepoint_sec) **RETURN MDSYS.SDO_GEOMETRY**,
- Returns a moving point restricted at a specific period
  - at_period (per tau_tll.d_period_sec) **RETURN moving_point**,
- Restricts the moving point at the space specified by the linestring parameter which is supposed to be part of his route
  - at_linestring (line MDSYS.SDO_GEOMETRY) **RETURN moving_point**,

References

Martin Erwig, Ralf Hartmut Güting, Markus Schneider, Michalis Vazirgiannis: Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases. GeoInformatica 3(3): 269-296 (1999)

P. Sistla, O. Wolfson, S. Chamberlain, S. Dao, "Modeling and Querying Moving Objects", *Proceedings of the Thirteenth International Conference on Data Engineering (ICDE13),* Birmingham, UK, Apr.97, pp. 422-432.

O. Wolfson, S. Chamberlain, P. Sistla, B. Xu, J. Zhou, "DOMINO: Databases for Moving Objects tracking", *Proceedings of the ACM-SIGMOD 1999, International Conference on Management of Data,* Philadelphia, PA, June 1999, pp. 547-5 49

Victor Teixeira de Almeida, Ralf Hartmut Güting, Thomas Behr: Querying Moving Objects in SECONDO. MDM 2006: 47, 2006

Ralf Hartmut Güting, Markus Schneider: Moving Objects Databases Morgan Kaufmann 2005

G. Trajcevski, O. Wolfson, K. Hinrichs, and S. Chamberlain. Managing uncertainty in moving objects databases. ACM Transactions on Database System, 29(3):463–507, 2004.

N. Andrienko, G. Andrienko, N. Pelekis, and S. Spaccapietra. Basic concepts on movement data. In: Mobility, Data Mining and Privacy: Geographic Knowledge Discovery. Giannotti, Fosca; Pedreschi, Dino (Eds.) 2008.