Introduction to Trajectory Data Mining

Zhe Zhang
10.05.2016
Maa-123. 3580 Spatial Data Mining
Introduction

- Various types of positioning and identification methods eg. GPS enable us to track human beings, animals, vehicles´ (cars, ships) positions

- AIS (automatic identification system): is an automatic tracking system used on ships and by vessel traffic services (VTS) for identifying and locating vessels by electronically exchanging data with other nearby ships, AIS base stations, and satellites.

- Mobile data is multivariate (spatial and temporal dimensions + attribute data)

- Mobile data can be analysed by spatial data mining methods like clustering and identifying regular patterns as well as abnormalities from the regular bahaviour
Trajectory data, definition

• trajectory(en) = liikerata(su)

• Trajectory of a moving object is a continuous function $\tau(t)$ of time $t$ such that given a time instant $t$ it returns the position of the moving object

• In reality, the moving objects trajectory is recorded by a finite set of observations at discrete times $t_1, t_2, \ldots t_n$
Examples of TD and applications

- Examples of trajectory data:
  - Tracked animals (reindeers, moose, birds, ...), people (elderly people), GSMs (e.g. for traffic purposes)
  - Tracked vessels; real time tracking in “MarineTraffic.com”
  - Trajectories of tornadoes
  - Sports scene analysis (players on a field)

- Trajectory data mining has many important, real-world applications
  - Homeland security (e.g., border monitoring), situation pictures
  - Law enforcement (e.g., video surveillance)
  - Weather forecast
  - Traffic control, marine traffic (VTS., Vessel Traffic Services)
  - Location-based services
1) Example pattern in trajectories

• What is the location visited by most entities? Useimmin käyty paikka?

location = circular region of specified radius
Example pattern in trajectories

• What is the location visited by most entities?

location = circular region of specified radius

4 entities
Example pattern in trajectories

• What is the location visited by most entities?

location = circular region of specified radius
2) Trajectory pattern types

(Gudmundsson et al., 2006)

• Flock (lauma)
  – Entities moving in the same direction while being close to each other. "being close" refers to being inside a circle of some specified radius $r$.

• Leadership (johto)
  – One entity is heading a flock in a specified direction for some time before the flock pattern occurs.

• Convergence (lähentyminen)
  – Moving to the same location, given that direction of motion does not change.

• Encounter (kohtaaminen)
  – Moving to and meeting at the same location so it is a convergence pattern where the entities arrive at the same time.
Fig. 1  *Left,* a flock pattern for $p_1$, $p_2$, $p_3$ at the eighth time step. It is also a leadership pattern with $p_2$ as the leader. *Right,* a convergence pattern if $m = 4$ for $p_2$, $p_3$, $p_4$, $p_5$. 
Patterns in trajectories

- **Longest flock:** given a radius \( r \) and subset size \( m \), determine the longest time interval for which \( m \) entities were within each other’s proximity (circle radius \( r \))

\[
\text{Time} = 0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8
\]

\( m = 3 \)

longest flock in \([1.8, 6.4]\)
How to find the patterns

• Clustering
• Identifying the regular behaviour
• Detecting the anomalies/abnormal behaviour
Trajectory similarity
(Kreveld and Luo)

• Similarity analysis:
  – subtrajectories with smallest dissimilarity

• Variants
  – 1. Duration is fixed, starting time the same. Find the most similar subtrajectory of exactly 3 hours in two animal trajectories recorded during a week.
  – 2. Duration is not fixed, but it must be not less than some minimum length, starting times are the same
3. The duration $T$ is fixed and the starting times may be different. For example, we want to find the most similar subtrajectories of two hurricanes for the duration of exactly four days, but the hurricanes occurred at different times (and may have different lengths).

4. The duration $T$ is not fixed but $T \geq T_{\text{min}}$, and the starting times may be different.
Similarity measure

• For the similarity measure the distance between two trajectories is calculated
  – See Fig. 1 in Kreveld & Luo

• With time shift
  – See Fig 2. in Kreveld & Luo
(Kreveld and Luo)
For the similarity measure the distance between two trajectories is calculated.

Figure 2 shows that a time shift can help to find a subtrajectory with smaller distance. If the duration is specified to be two time units, then without a time shift, interval \([t_2, t_4]\) will be most similar, but with a time shift, \([t_7, t_9]\) on \(\tau_1\) is more similar to \([t_8, t_{10}]\) on \(\tau_2\).

Figure 2: Subtrajectory similarity with time shift.
Potential applications

• For all kinds of tracking applications the collected trajectory data offers great opportunities to use spatial data mining methods for knowledge extraction

• The starting point should be:
  – What kind of knowledge we are interested in?

• Not:
  – What kind of calculations and computations are possible?

• Examples: various kinds of situation picture applications; questions "what is normal" "what is abnormal"
Reference

• Gudmundsson and Kreveld. Computing longest duration flocks in trajectory.
• Dodge et al., 2008. Towards a taxonomy of movement patterns. Information visualization. 7, 240-252
• Kreveld and Luo. Trajectory similarity of moving objects.

Articles will be given in Mycourse.
Space-time density of trajectories: exploring spatio-temporal patterns in movement data

Dr Urška Demšar,
National Centre for Geocomputation, NUI Maynooth, Ireland

Prof Kirsi Virrantaus,
Department of Surveying, Aalto University, Finland
AIS data

• AIS = automatic identification system used by ships, for identification and positioning, for safety
• AIS data can be viewed for example in http://www.marinetraffic.com/ais/
• same for planes http://www.flightradar24.com/

• International Maritime Organization (I.M.O.). Regulations for carriage of AIS, 1998
• URL: http://www.imo.org/Legal/mainframe.asp?topic_id=754#regulations
Example

A.I.S. data at 2009/04/03 12h00 (local time, Z+3) (Legouge, R.)
Data used as accident history source (Legouge, R.)
Time in 3D – Space-Time Cube and its variations

**Space-time cube** – developed in time geography to show people’s movements through both geographic time and space. Applications: GPS traces of people/animals/objects – ships, vehicles, airplanes, etc.

Movement of an object shown as **trajectory** in space and time:
- 3D points \((x_1,y_1,t_1), (x_2,y_2,t_2), \ldots, (x_n,y_n,t_n)\),
- points linked with straight linear lines (linear segments)
What is the problem – clutter and overprinting in space-time cube

Space-time cube becomes messy when there are too many trajectories (clutter & overprinting)

Vessel trajectories for 58 passenger ships

Space-time density of trajectories

Volume visualisation of importance of each trajectory in space and time
The algorithm

- the density around each trajectory is calculated as a volume.
- the value is assigned to each voxel according to the distance of the central point of the respective voxel to the trajectory.
- this distance is normalized using user specified Kernel size and the kernel function is linear.
- the densities for each separate trajectory are added up into a total combined density.
- calculation implemented by 3d-map algebra, Matlab
- visualizations by various means by Voxler software
Algorithm for calculation of space-time density of a set of trajectories

\[ \text{TotalDensity} = 0; \]

for each trajectory \( T \)

\[ \text{TrajectoryDensity} = 0; \]

calculate KernelArea around the trajectory;

for each voxel \( P \) in the KernelArea

calculate DistanceToTrajectory \( T \);

\[ \text{TrajectoryDensity} = \text{normalised DistanceToTrajectory with kernel size}; \]

end

\[ \text{TotalDensity} = \text{TotalDensity} + \text{TrajectoryDensity}; \]

end

Normalise \text{TotalDensity} with number of trajectories;
Densities around each of eight simulated trajectories

- Density = 1 on trajectory
- Density = 0 on kernel surface

Add up into total density and normalise with no. of trajectories
Total density of eight simulated trajectories

Shown with isosurfaces at 0.8, 0.6, 0.4, 0.2 and 0.

Same colour for all isosurfaces, different transparency.
Visualisation possibilities – volume visualisation

3 methods:
Direct volume rendering
Isosurfaces
Clipping planes

Direct volume rendering

- Transparency according to the scalar field values
Isosurfaces

- 2d surfaces that share the same scalar value; distribution of one particular value

Clipping planes

- cross section of a volume
Space-time density of real data – vessel trajectories
In the Gulf of Finland

Two data sets:
- one day AIS data (the 26th of January, 2008)
- one month AIS data (January 2008)

AIS = Automatic Identification System based on GPS positioning

Densities for 2 vessel types were produced by using trajectory data of:
- passenger ships (ferries)
- tankers
Visualisation possibilities and spatio-temporal patterns

Same three possibilities as for simulated data

Isosurfaces

Direct volume rendering
Clipping planes
Results – passenger ships – monthly density
Results – tankers – monthly density – DVR
Results – tankers – monthly density - isosurfaces
Spatio-temporal patterns to look for

**Spatio-temporal hotspots**
- convergence of trajectories in both space and time

**Temporal towers**
- many objects in near proximity at all times

**Temporal bridges**
- discrete temporal pattern at certain times
Conclusions

New spatio-temporal visualisation method and its implications
- a method was developed and implemented with some tests on AIS data
- a couple of new pattern types of the vessel behaviour was identified
- the developed tool seems promising linked with the maritime situation picture (Gulf of Finland)

Three more types of issues to consider:
- algorithm efficiency
- visualisation design
- usability testing
- more data (weather conditions)