

## Part III – Trajectory Data Mining

Vania Bogorny  
 Universidade Federal de Santa Catarina, Brazil  
 Department of Informatics and Statistics  
[www.inf.ufsc.br/~vania](http://www.inf.ufsc.br/~vania)  
[vania@inf.ufsc.br](mailto:vania@inf.ufsc.br)

12/15/2010

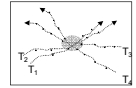
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

1 of 90

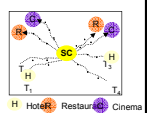
## Outline

- The wireless explosion
- Moving Object Data and Mobility Data Analysis
- Trajectory Patterns

## ■ Geometric Trajectory Pattern Mining Methods:



## ■ Semantic Trajectory Pattern Mining Methods:



- Trajectory Data mining Tools

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

2 from 90

The Wireless Explosion (Fosca Giannotti 2007 – [www.geopkdd.eu](http://www.geopkdd.eu))

*Have you ever feel to be tracked?*

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

3 of 90

## The Wireless Explosion

- The world becomes more and more mobile with the easy access to smart phones, GPS, etc
- Satellite services, sensors and wireless technologies are rapidly improving
- ◆ *lots of spatio-temporal data is being generated*

12/15/2010

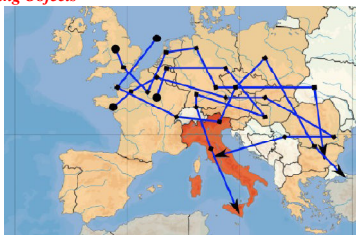
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

4 of 90

The Wireless Explosion (Fosca Giannotti 2007 – [www.geopkdd.eu](http://www.geopkdd.eu))

Mobile devices leave behind digital traces that are collected as **trajectories**, describing the movement of its users

Mobile devices generate a new type of data, called “**Trajectories of Moving Objects**”



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

5 of 90

## Mobility Data Analysis

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

6 of 90

## Mobility Data Analysis

Several analysis may be done over trajectories:

- How people move around 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?
- .....



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

7 of 90

## Mobility Data Analysis: Applications

■ **Trajectory data analysis may be useful in several application domains**

### ■ **Veicule Monitoring**

- ◆ Transportation Companies monitor their trucks
- ◆ Insurance companies use GPS devices to monitor insured vehicles to reduce insurance price

### ■ **Traffic Analysis**

- ◆ To alert people about traffic jams, accidents, etc...
- ◆ Identify/predict low traffic regions in a city



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

8 of 90

## Mobility Data Analysis

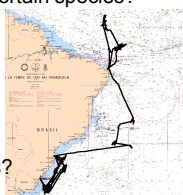
### ■ **Animal Migration / Behaviour Analysis**

- ◆ Which are the trajectories of a given migration bird? Where do birds stop? For how long?
- ◆ Which is the migration pattern of certain species?



### ■ **Fishing Analysis and Control**

- ◆ Are boats really fishing in allowed areas?
- ◆ Can we classify vessel trajectories?



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

9 of 90

## Mobility Data Analysis

### ■ **Weather prediction and movement analysis**

### ◆ **Hurricane tracking**



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

10 of 90

## Trajectory Data

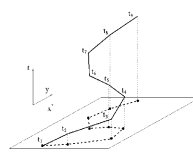
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

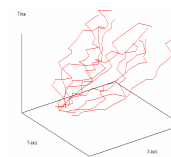
11 of 90

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

- ◆ Spatio-temporal Data
- ◆ Represented as a set of points, located in space and time
- ◆  $T = (x_1, y_1, t_1), \dots, (x_n, y_n, t_n) \Rightarrow$  position in space at time  $t_i$  was  $(x_i, y_i)$



Tid	position (x,y)	time (t)
1	48.890018 2.246100	08:25
1	48.890018 2.246100	08:26
...	...	...
1	48.890020 2.246102	08:40
1	48.889880 2.248208	08:41
1	48.885732 2.255031	08:42
...	...	...
1	48.888434 2.236105	09:04
1	48.853611 2.349190	09:05
...	...	...
2	...	...



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

12 of 90

### Trajectories: Overall Characteristics (Adrienko 2008)

1. Geometric shape
2. Length (traveled distance)
3. Duration (in time)
4. Speed
  - ♦ Mean and maximal Speed
  - ♦ Acceleration, deceleration
5. Direction:
  - ♦ Periods of straight, curvilinear, circular movement
- ♦ More.....

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 13 of 90

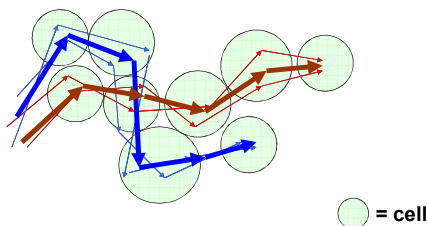
### Trajectory Patterns

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 14 of 90

### Mining Trajectories: Clustering

Fosca Giannotti 2007 – www.geopkdd.eu

- Group together similar trajectories
- For each group produce a summary

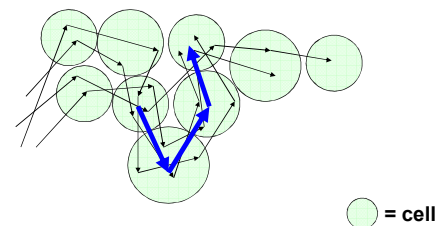


12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 15 of 90

### Mining Trajectories : Frequent patterns

Fosca Giannotti 2007 – www.geopkdd.eu

- Frequent followed paths

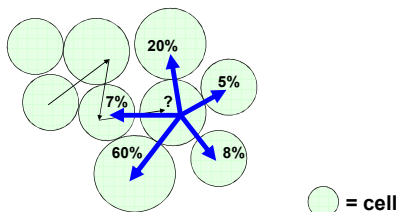


12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 16 of 90

### Mining Trajectories: classification models

Fosca Giannotti 2007 – www.geopkdd.eu

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



12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 17 of 90

### Trajectory Data Mining Methods

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 18 of 90

## Spatio-Temporal Data Mining Methods

### Two approaches:

#### Geometry-based spatio-temporal data mining:

- ◆ Density-based clustering methods
- ◆ Focus on physical similarity
- ◆ Consider only geometrical properties of trajectories (space and time)

#### Semantic-based spatio-temporal data mining

- ◆ Deal with sparse data also
- ◆ Patterns are computed based on the semantics of the data
- ◆ Trajectories are pre-processed to enrich the data

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

19 of 90

## Geometry-based Trajectory Data Mining Methods

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

20 of 90

## General Geometric Trajectory Patterns

12/15/2010

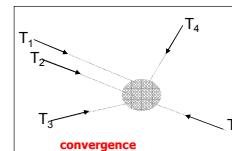
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

21 of 90

## Relative Motion Patterns (Laube 2004)

Proposed 5 kinds of trajectory patterns based on movement, direction, and location: convergence, encounter, flock, leadership, and recurrence

- ◆ **Convergence:** At least  $m$  entities pass through the same circular region of radius  $r$ , not necessarily at the same time (e.g. people moving to train station)



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

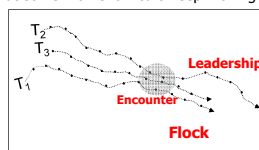
22 of 90

## Relative Motion Patterns (Laube 2004)

- ◆ **Flock pattern:** At least  $m$  entities are within a region of radius  $r$  and move in the same direction during a time interval  $\geq s$  (e.g. traffic jam)

- ◆ **Leadership:** At least  $m$  entities are within a circular region of radius  $r$ , they move in the same direction, and at least one of the entities is heading in that direction for at least  $\ell$  time steps. (e.g. bird migration, traffic accident)

- ◆ **Encounter:** At least  $m$  entities will be concurrently inside the same circular region of radius  $r$ , assuming they move with the same speed and direction. (e.g. traffic jam at some moment if cars keep moving in the same direction)



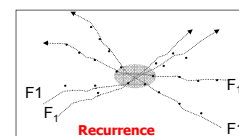
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

23 of 90

## Relative Motion Patterns (Laube 2004)

- ◆ **Recurrence:** at least  $m$  entities visit a circular region at least  $k$  times



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

24 of 90

### Extension of the work proposed by [Laube 2004, 2005]

- Gudmundsson(2006)
  - ◆ Computes the longest duration flock patterns
  - ◆ The longest pattern has the longest duration
  - ◆ And has at least a minimal number of trajectories
- Gudmundsson (2007)
  - ◆ proposes approximate algorithms for computing the patterns leadership, encounter, convergence, and flock
  - ◆ Focus relies on performance issues

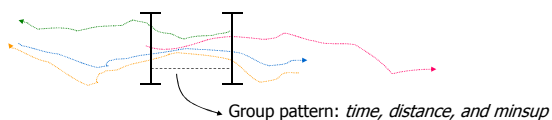
12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 25 of 90

### Frequent Trajectory Patterns

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 26 of 90

### Frequent Mobile Group Patterns (Hwang, 2005)

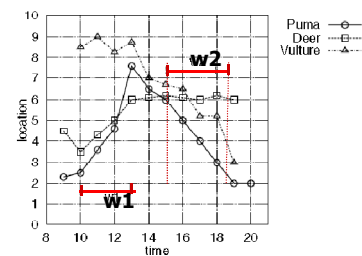
- A *group pattern* is a set of trajectories close to each other (with distance less than a given *minDist*) for a minimal amount of time (*minTime*)
- Direction is not considered
- Frequent groups are computed with the algorithm Apriori



12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 27 of 90

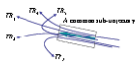
### Co-Location Patterns (Cao 2006)

- Co-location episoids in spatio-temporal data
- Trajectories are spatially close in a time window and move together



12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 28 of 90

### Traclus (Han, 2007)

- Clustering algorithm (TraClus-Trajectory Clustering)
- Group sub-trajectories
  - ◆ 
- Density-based
- Partition-and-group method
  - ◆ 1) each trajectory is partitioned into a set of line segments (sub-trajectories) with length  $L$  defined by the user
  - ◆ 2) similar segments (close segments) are grouped
    - Similarity is based on a distance function
- Interesting approach for trajectories of hurricanes
- **Main drawback:** Clustering is based on spatial distance
  - ◆ time is not considered

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 29 of 90

### Trajectory Sequential Patterns

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 30 of 90

### T-Patterns (Giannotti, 2007)

- Sequential Trajectory Pattern Mining
- Considers both space and time
- Objective is to describe frequent movement
  - Considering visited regions of interest
  - During *movements* and the *duration* of movements

#### Steps:

1. Compute or find regions of interest, based on dense spatial regions (no time is considered)
2. Select trajectories that intersect two or more regions in a sequence, annotating travel time from one region to another
3. Compute sequences of regions visited in same time intervals

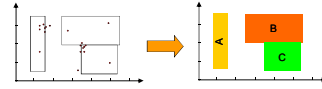
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

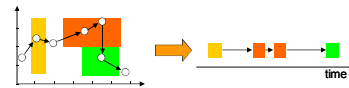
31 of 90

### T-Patterns (Giannotti, 2007)

- Fix a set of pre-defined regions



- Map each (x,y) of the trajectory to its region



- Sample pattern:

A  $\xrightarrow{20 \text{ min.}}$  B

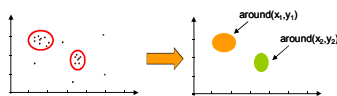
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

32 of 90

### T-Patterns (Giannotti, 2007)

- Detect significant regions thru spatial clustering



- Map each (x,y) of the trajectory to its region



- Sample pattern:  $\text{around}(x_1, y_1) \xrightarrow{20 \text{ min.}} \text{around}(x_2, y_2)$

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

33 of 90

### Trajectory Classification

The idea is to classify types of trajectories

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

34 of 90

### TraClass Algorithm (Lee 2008)

- Two main steps algorithm:
  - First: region – based clustering
  - Second: trajectory-clustering

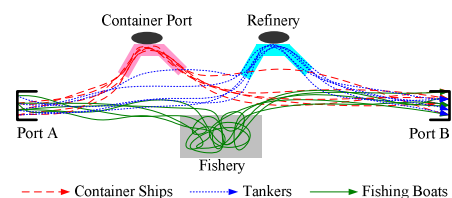
- Time is not considered

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

35 of 90

### TraClass Algorithm (Lee 2008)



Classify subtrajectories instead of whole trajectories

Examples:

- Red trajectories move from Port A to Container Port and then to Port B
  - Blue trajectories move from Port A to Refinery and then to Port B
- Classifying whole trajectory would classify all trajectories as moving from Port A to Port B

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

36 of 90

## Trajectory Outlier Detection

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

37 of 90

## Trajectory Outlier Detection

- The objective is to find trajectories that have different behavior in relation to other trajectories
- For instance:
  - A fishing vessel that has a behaviour different from other fishing vessels in the same area
  - A hurricane that may change behaviour in certain parts of its trajectory
  - Cars or pedestrians with suspicious behaviour

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

38 of 90

## TraOD - Trajectory Outlier Detection (Lee 2008)

- Partition trajectories into subtrajectories
- Compare subtrajectories based on:
  - distance** and **length**
- If a subtrajectory is not close to other trajectories for a minimal length
  - It is an outlier

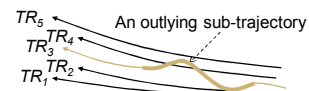
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

39 of 90

## TraOD - Trajectory Outlier Detection (Lee 2008)

- Example:
  - Looking to the **whole trajectory**,  $TR_3$  is not detected as an outlier since its overall behavior is similar to neighbour trajectories
  - Looking at the **subtrajectories**,  $T_3$  can be an outlier



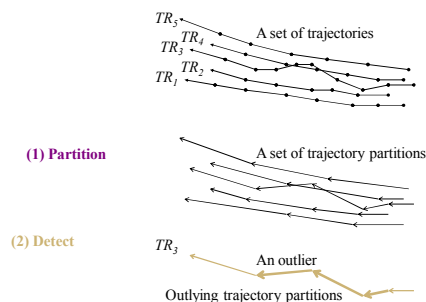
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

40 of 90

## TraOD - Trajectory Outlier Detection (Lee 2008)

- Two phases: **partitioning** and **detection**

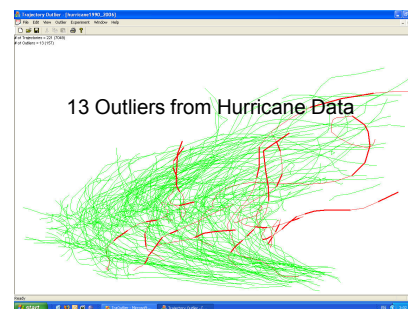


12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

41 of 90

## TraOD - Trajectory Outlier Detection (Lee 2008)



12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

42 of 90

## Summary

- These data mining approaches deal with *Trajectory Samples*

Tid	geometry	timest
1	48.890018 2.246100	08:25
1	48.890018 2.246100	08:26
...	...	...
1	48.890020 2.246102	08:40
1	48.888880 2.248208	08:41
1	48.885732 2.255031	08:42
...	...	...
1	48.858434 2.336105	09:04
1	48.853611 2.349190	09:05
...	...	...
1	48.853610 2.349205	09:40
1	48.860515 2.349018	09:41
...	...	...
1	48.861112 2.334167	10:00
1	48.861531 2.336018	10:01
1	48.861530 2.336020	10:02
...	...	...
2	...	...

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

43 of 90

## References

- Laube, P. and Imfeld, S. (2002). Analyzing relative motion within groups of trackable moving point objects. In Egenhofer, M. J. and Mark, D. M., editors, GIScience, volume 2478 of Lecture Notes in Computer Science, pages 132–144. Springer.
- Laube, P., Imfeld, S., and Weibel, R. (2005a). Discovering relative motion patterns in groups of moving point objects. International Journal of Geographical Information Science, 19(6):639–668.
- Laube, P., van Kreveld, M., and Imfeld, S. (2005b). Finding REMO: Detecting Relative Motion Patterns in Geospatial Lifelines. Springer.
- Lee, J.-G., Han, J., and Whang, K.-Y. (2007). Trajectory clustering: a partition-and-group framework. In Chan, C. Y., Ooi, B. C., and Zhou, A., editors, SIGMOD Conference, pages 593–604. ACM.
- Li, Y., Han, J., and Yang, J. (2004). Clustering moving objects. In KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 617–622, New York, NY, USA. ACM Press.
- Nanni, M. and Pedreschi, D. (2006). Time-focused clustering of trajectories of moving objects. Journal of Intelligent Information Systems, 27(3):267–289.

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

44 of 90

## References

- Verhein, F. and Chawla, S. (2006). Mining spatio-temporal association rules, sources, sinks, stationary regions and thoroughfares in object mobility databases. In Lee, M.-L., Tan, K.-L., and Wuwongse, V., editors, DASFAA, volume 3882 of Lecture Notes in Computer Science, pages 187–201. Springer.
- Gudmundsson, J. and van Kreveld, M. J. (2006). Computing longest duration flocks in trajectory data. In [de By and Nittel 2006], pages 35–42.
- Gudmundsson, J., van Kreveld, M. J., and Speckmann, B. (2007). Efficient detection of patterns in 2d trajectories of moving points. GeoInformatica, 11(2):195–215.
- Hwang, S.-Y., Liu, Y.-H., Chiu, J.-K., and Lim, E.-P. (2005). Mining mobile group patterns: A trajectory-based approach. In Ho, T. B., Cheung, D. W.-L., and Liu, H., editors, PAKDD, volume 3518 of Lecture Notes in Computer Science, pages 713–718. Springer.
- Cao, H., Mamoulis, N., and Cheung, D. W. (2006). Discovery of collocation episodes in spatiotemporal data. In ICDM, pages 823–827. IEEE Computer Society.
- Zhenhui Li, Jae-Gil Lee, Xiaolei Li, Jiawei Han: Incremental Clustering for Trajectories. DASFAA (2) 2010: 32–46

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

45 of 90

## More...

- Huiping Cao, Nikos Mamoulis, David W. Cheung: Discovery of Periodic Patterns in Spatiotemporal Sequences. IEEE Trans. Knowl. Data Eng. 19(4): 453–467 (2007)
- Panos Kalnis, Nikos Mamoulis, Spiridon Bakiras: On Discovering Moving Clusters in Spatio-temporal Data. SSTD, 364–381 (2005)
- Florian Verhein, Sanjay Chawla: Mining spatio-temporal patterns in object mobility databases. Data Min. Knowl. Discov. 16(1): 5–38 (2008)
- Florian Verhein, Sanjay Chawla: Mining Spatio-temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases. DASFAA, 187–201 (2006)
- Cao, H., Mamoulis, N., and Cheung, D. W. (2005). Mining frequent spatio-temporal sequential patterns. In ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining, pages 82–89, Washington, DC, USA. IEEE Computer Society.
- Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez, "TraClass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering", Proc. 2008 Int. Conf. on Very Large Data Base (VLDB'08), Auckland, New Zealand, Aug. 2008.
- Jae-Gil Lee, Jiawei Han, and Xiaolei Li, "Trajectory Outlier Detection: A Partition-and-Detect Framework", Proc. 2008 Int. Conf. on Data Engineering (ICDE'08), Cancun, Mexico, April 2008.

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

46 of 90

## Semantic-based Trajectory Data Mining Methods

## Semantic Trajectory Data Mining

- The main idea is to enrich trajectories with domain semantic information in preprocessing steps
  - This task can be done using data mining
- Apply data mining as a second step
  - Mining is on semantic rich trajectories

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

47 of 90

12/15/2010

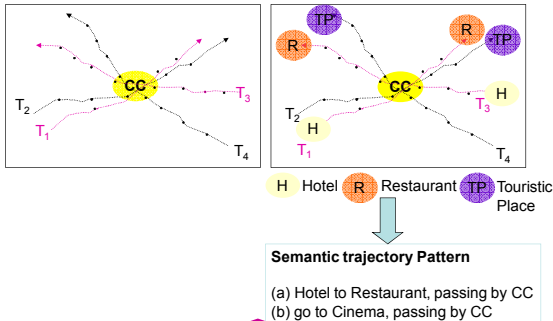
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

48 of 90



### Geometric Patterns X Semantic Patterns (Bogorny 2008)

Geometric Pattern



12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 49 of 90

### Geometric Patterns X Semantic Patterns (Bogorny 2008)

- There is very little or no semantics in most DM approaches for trajectories

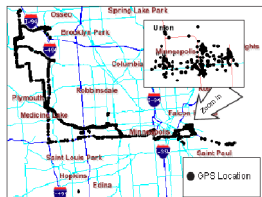
Consequence:

- Patterns are purely geometrical
- Difficult to interpret from the user's point of view
- Do not discover *semantic patterns*, which can be independent of *spatial* location

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 50 of 90

### DJ-Cluster (Zhou 2007)

- DJ-Cluster is a variation of DBSCAN
  - Focus relies on performance issues
- Objective: find interesting places of individual trajectories
  - Clusters are computed from a SET of trajectories of the same object



- Time is not considered

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 51 of 90

### A Conceptual View on Trajectories (Spaccapietra 2008)

- A trajectory is a spatio-temporal thing (an object) that
  - has *generic* features
    - generic: application independent
  - has *semantic* features
    - semantic: application dependent
- A trajectory is more than a moving object

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 52 of 90

### The Model of Stops and Moves (Spaccapietra 2008)

#### STOPS

- Important parts of trajectories
- Where the moving object has stayed for a minimal amount of time
- Stops are application dependent
  - Tourism application
    - Hotels, touristic places, airport, ...
  - Traffic Management Application
    - Traffic lights, roundabouts, big events...

#### MOVES

- Are the parts that are not stops

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 53 of 90

### Semantic Trajectories

- A *semantic trajectory* is a set of **stops** and **moves**
  - Stops** have a *place*, a *start time* and an *end time*
  - Moves** are characterized by two consecutive stops



12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 54 of 90

## Methods for Adding Semantics to Trajectories

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

55 of 90

## Methods to Compute Stops and Moves

- 1) **IB-SMoT** (INTERSECTION-based)  
Interesting for applications like tourism and urban planning
- 2) **CB-SMoT** (SPEED-based clustering)  
Interesting for applications where the **speed** is important, like traffic management
- 3) **DB-SMoT** (DIRECTION-based clustering)  
Interesting in application where the **direction** variation is important like fishing activities

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

56 of 90

## IB-SMoT (Alvares 2007a)

A **candidate stop**  $C$  is a tuple  $(R_C, \Delta_C)$ , where

- $R_C$  is the geometry of the candidate stop (spatial feature type)
- $\Delta_C$  is the *minimal time duration*

*E.g. [Hotel - 3 hours]*

An **application**  $A$  is a finite set  
 $A = \{C_1 = (R_{C_1}, \Delta_{C_1}), \dots, C_N = (R_{C_N}, \Delta_{C_N})\}$  of *candidate stops* with non-overlapping geometries  $R_{C_1}, \dots, R_{C_N}$

*E.g. [Hotel - 3 hours, Museum - 1 hour]*

12/15/2010

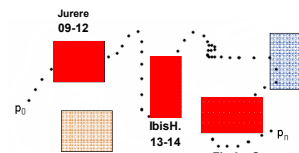
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

57 of 90

## IB-SMoT

(Alvares 2007a)

- Input: candidate stops // Application  
trajectories // trajectory samples
- Output: Semantic rich trajectories
- Method:
  - For each trajectory
    - Check if it intersects a candidate stop for a minimal amount of time



12/15/2010

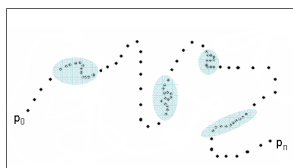
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

58 of 90

## CB-SMoT: Speed-based clustering

(Palma 2008)

- Clusters single trajectories based on the speed variation:  
low speed → important place



12/15/2010

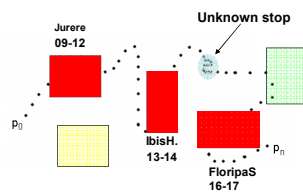
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

59 of 90

## CB-SMoT: Speed-based clustering

(Palma 2008)

- Input: Trajectory samples  
Speed variation  
minTime
- Output: stops and moves
- Step 1: find clusters
- Step 2: Add semantics to each cluster
  - 2.1: If intersects  $\alpha$  during  $\Delta\alpha$  → stop  $\alpha$
  - 2.2: If no intersection during  $\Delta\alpha$  → unknown stop



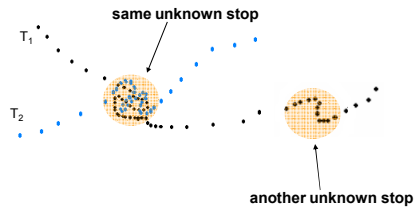
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

60 of 90

**CB-SMoT: Speed-based clustering**

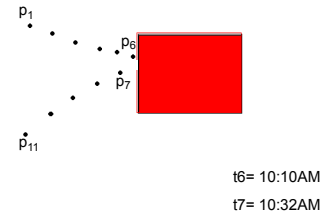
(Palma 2008)

**Unknown Stops (CB-SMOT)**

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 61 of 90

**CB-SMoT: Speed-based clustering**

(Palma 2008)

**Can Find Clusters Inside Buildings**

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 62 of 90

**DB-SMOT : Direction-based Clustering (Manso 2010)**

- Input: trajectories // trajectory samples
- minDirVariation // minimal direction variation
- minTime // minimum time
- maxTolerance

- Output: semantic rich trajectories

**Method:**

- For each trajectory
  - Find clusters with direction variation higher than minDirVariation
  - For a minimal amount of time
- 

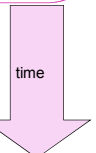
12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 63 of 90

**Examples of semantic trajectory patterns**

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 64 of 90

**Multiple-granularity semantic trajectory pattern mining**

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 65 of 90

**STOPS at Multiple-Granularities (Bogorny 2009)**Stop at **Ibis Hotel** from **6:04PM to 7:42PM, september 16, 2010****IbisHotel or Hotel or Accommodation****Afternoon or Thursday or 6:00PM – 8:00PM or RUSH-HOUR**

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 66 of 90

### ITEMS - the building blocks for semantic pattern discovery

- An *item* is generated either from a stop or a move
- An *item* is a set of complex information (space + time), that can be defined in many formats/types and at different granularities

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 67 of 90

### Building an ITEM for Data Mining (Bogorny 2009)

- **Formats/types** for an *item*:
- **NameOnly**: is the name of the stop/move
  - ◆ **STOPS**: name of the spatial feature instance
    - IbisHotel
  - ◆ **MOVES**: name of the two stops which define the move
    - SydneyAirport – IbisHotel
- **NameStart**: is the name of the stop/move + start time
  - ◆ IbisHotel [morning] --stop
  - ◆ LouvreMuseum [weekend] --stop
  - ◆ IbisHotel-SydneyAirport [10:00AM-11:00AM] --move

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 68 of 90

### Building an ITEM for Data Mining (Bogorny 2009)

- **NameEnd**: name of a stop/move + end time
  - ◆ IbisHotel[morning] → stop
  - ◆ IbisHotel-SydneyAirport[10:00AM-11:00AM] → move
- **NameStartEnd**: name of a stop/move + start time + end time
  - ◆ IbisHotel[08:00AM-11:00AM][1:00pm-6:00pm] → stop
  - ◆ LouvreMuseum[morning][afternoon] → stop
  - ◆ SydneyAirport – IbisHotel [10:00AM-11:00PM] [10:00AM-6:00PM]

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 69 of 90

### Multiple-Granularity Semantic Trajectory DMQL (Bogorny 2009)

- ST-DMQL is an approach to semantically enrich trajectories with domain information
- Automatically transforms these semantic information into different space and time granularities
- Extracts frequent patterns, association rules and sequential patterns from semantic trajectories

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 70 of 90

### Multiple Level Semantic Sequential Patterns

Large Sequences of Length 2 (ITEM=SPACE+Start\_Time)

- (41803\_street\_5, 41803\_street\_5) Support: 7
- (41803\_street\_4, 41803\_street\_4) Support: 9
- (41803\_street\_4, 66655\_street\_4) Support: 5
- (41803\_street\_2, 41803\_street\_2) Support: 6
- (41803\_street\_8, 41803\_street\_8) Support: 5
- (41803\_street\_3, 0\_unknown\_3) Support: 5

time unit = month  
gid  
Spatial feature type (stop name)

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 71 of 90

### Multiple Level Semantic Sequential Patterns

Large Sequences of Length 2 (ITEM=SPACE+Start\_Time)

- (41803\_street\_tuesday, 41803\_street\_tuesday) Support: 9
- (41803\_street\_tuesday, 66655\_street\_tuesday) Support: 5
- (41803\_street\_monday, 66655\_street\_monday) Support: 5
- (41803\_street\_monday, 41803\_street\_monday) Support: 11
- (41803\_street\_monday, 0\_unknown\_monday) Support: 5
- (41803\_street\_thursday, 41803\_street\_thursday) Support: 13
- (41803\_street\_thursday, 0\_unknown\_thursday) Support: 6
- (41803\_street\_wednesday, 41803\_street\_wednesday) Support: 7

Time unit = Day of the week  
gid  
Spatial feature type (stop name)

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010) 72 of 90

Tools:  
Weka- STPM  
Athena

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

73 of 90

## Trajectory Behaviour Patterns

Recent works have emerged on mining behaviour patterns from trajectories

12/15/2010

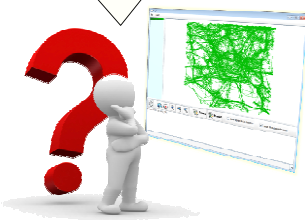
Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

74 of 90

### Athena (Baglioni 2009)

#### Semantic-rich movement analysis

Which are the home-work trajectories? And the common behaviors of them?



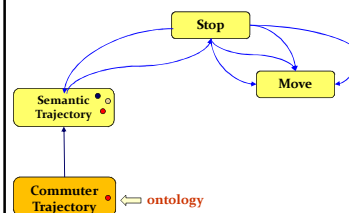
To answer these questions the idea is to get the **home-work trajectory** (or pattern) from a knowledge base, for then to discover the trajectories that frequently follow this pattern

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

75 of 90

### Athena (Baglioni 2009)



**Commuter trajectory**  $\equiv$  a trajectory frequently starting outside the city, stopping inside the city for a long time and going back outside the city

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

76 of 90

### Athena (Baglioni 2009)

- Based on ontologies to represent domain knowledge and to infer the semantic types of the patterns/trajectories.
- A trajectory pattern is given as input and the method checks the trajectories that support the pattern to classify the trajectory

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

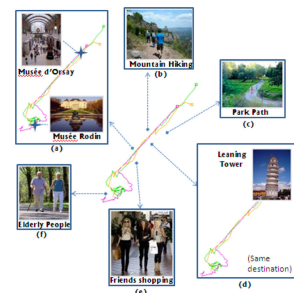
77 of 90

### Pattern Interpretation (Ong 2010)

This work focuses on post-processing, trying to interpret the patterns

Considering that the movement context is essential to correctly interpret and understand the patterns

CONTEXT = geography + thematic attributes



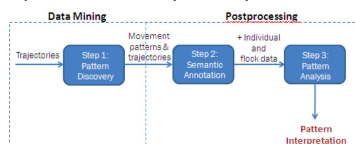
12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

78 of 90

## Pattern Interpretation (Ong 2010)

1. Mining flock movement **patterns**
2. **Semantic enrichment**: **annotates** patterns with information about the moving object (e.g. age, gender)
3. Mine the enriched patterns with hierarchical clustering to help the user to interpret the patterns



12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

79 of 90

## Works Summarized in this part of the Tutorial

<b>Geometric Pattern Mining Methods</b> (mining is on sample points)	<b>Semantic Pattern Mining Methods</b> (Generate Semantic Trajectories using DM - mining is on Semantic Trajectories)	<b>Behaviour Pattern Mining and Interpretation Methods</b>
■ Laube 2004, 2005 ■ Hwang 2005 ■ Gudmundson 2006, 2007 ■ Giannotti 2007 ■ Lee 2007 ■ Cao 2006, 2007 ■ Lee 2007, 2008a, 2008b ■ Li 2010	■ Alvares 2007 ■ Zhou 2007 ■ Palma 2008 ■ Bogorny 2009 ■ Bogorny 2010 ■ Manso 2010 ■ Alvares 2010	■ Giannotti 2009 ■ Baglioni 2009 ■ Ong 2010

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

80 of 90

## References

- Bogorny, V.; Bart Kuijpers, Luis Otávio Alvares: ST-DMQL: A Semantic Trajectory Data Mining Query Language. International Journal of Geographical Information Science 23(10): 1245-1276 (2009)
- Palma, A. T; Bogorny, V.; Kuijpers, B.; Alvares, L.O. *A Clustering-based Approach for Discovering Interesting Places in Trajectories*. In: 23rd Annual Symposium on Applied Computing, (ACM-SAC'08), Fortaleza, Ceara, 16-20 March (2008) Brazil. pp. 863-868.
- Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., and Vangenot, C. (2008). A conceptual view on trajectories. Data and Knowledge Engineering, 65(1):126-146.
- Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B., and Vaisman, A. (2007b). A model for enriching trajectories with semantic geographical information. In ACM-GIS, pages 162-169, New York, NY, USA. ACM Press.
- Rebecca Ong, Monica Wachowicz, Mirco Nanni, Chiara Renso, *From Pattern Discovery to Pattern Interpretation in Movement Data - IEEE SDM 2010*

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

81 of 90

## References

- Manso, J. A. ; TIMES, V. C. ; Oliveira, G. ; ALVARES, L. O. ; BOGORNÝ, V. . DB-SMoT: A Direction-Based Spatio-Temporal Clustering Method. In: IEEE International Conference on Intelligent Systems (IS), 2010, Londres. Proceedings of the IEEE International Conference on Intelligent Systems, 2010. p. 114-119. 3.
- Alvares, Luis O. ; PALMA, Andrey ; Oliveira, G. ; BOGORNÝ, V. . Weka-STPM: from trajectory samples to semantic trajectories. In: Workshop de Software Livre, 2010, Porto Alegre. WSL, 2010.
- Zhou, C.; Nupur Bhatnagar, Shashi Shekhar, Loren G. Terveen: Mining Personally Important Places from GPS Tracks. ICDE Workshops 2007: 517-526
- Bogorny, V. ; Carlos Alberto Heuser, Luis Otávio Alvares: A Conceptual Data Model for Trajectory Data Mining. GIScience 2010: 1-15
- Miriam Baglioni, José Antônio Fernandes de Macêdo, Chiara Renso, Roberto Trasarti, Monica Wachowicz: Towards Semantic Interpretation of Movement Behavior. AGILE Conf. 2009: 271-288
- Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Chiara Renso, Roberto Trasarti: Mining Mobility Behavior from Trajectory Data. CSE (4) 2009: 948-951

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

82 of 90

## Summary, Challenges and Open Issues in Spatio-Temporal Data Mining

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

83 of 90

## Challenges and Open Issues in Spatio-Temporal Data Mining

- Trajectory Clustering
  - ◆ Most works are density-based clustering methods
  - ◆ Most are adapted spatial or non-spatial clustering algorithms
  - ◆ Consider either time or space, only a few consider both dimensions

12/15/2010 Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

84 of 90

### Challenges and Open Issues in Spatio-Temporal Data Mining

#### ■ Trajectory Similarity

- ◆ Focus relies on **objective** similarity measures
  - ▢ Shape, direction, closeness
- ◆ **Needs: semantic** similarity
  - ▢ Higher abstraction level similarity
- ◆ Example:
  - groups of trajectories going together for shopping
  - Groups of trajectories going together to the University two times a week

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

85 of 90

### Challenges and Open Issues in Spatio-Temporal Data Mining

#### ■ Need for data mining methods using:

- ◆ Metadata
- ◆ Domain knowledge
- ◆ Semantics
- ◆ Ontologies

#### ■ For:

- ◆ Trajectory data pre-processing
- ◆ Pattern pruning
- ◆ Improve the quality of the patterns
- ◆ Pattern interpretation

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

86 of 90

### More needs

- There is a need for collaboration between data miners and domain experts (environmental experts, transportation managers, meteorologists, etc)
  - ◆ to evaluate data mining methods and the discovered patterns
- Post-Processing: almost no spatial or spatio-temporal data mining methods evaluate the patterns and their interestingness

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

87 of 90

[Thank You !](#)

12/15/2010

Tutorial on Spatial and Spatio-Temporal Data Mining (ICDM 2010)

88 of 90