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

Spatio-Temporal Data Mining

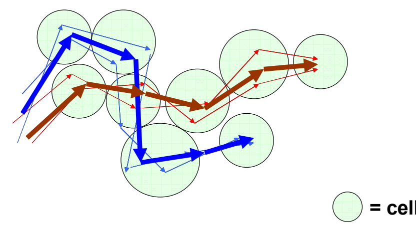
Vania Bogorny
Universidade Federal do Rio Grande do Sul
www.inf.ufrgs.br/~vbogorny
vbogorny@inf.ufrgs.br

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

Mining Trajectories: Clustering

Fosca Giannotti 2007 – www.geopkdd.eu

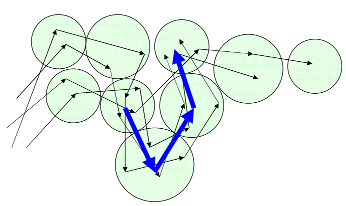
- Group together similar trajectories
- For each group produce a summary



Mining Trajectories : Frequent patterns

Fosca Giannotti 2007 – www.geopkdd.eu

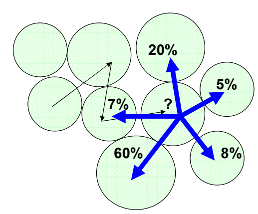
- Frequent followed paths



Mining Trajectories: classification models

Fosca Giannotti 2007 – www.geopkdd.eu

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



Spatio-Temporal Data Mining Methods

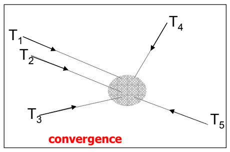
- Two approaches:
- *Geometry-based spatio-temporal data mining:*
  - Density-based clustering methods
  - Focus on similarity
  - Consider only geometrical properties of trajectories
- *Semantic-based spatio-temporal data mining*
  - Deal with sparse data and dense data
  - Independent of spatial locations
  - Patterns are computed based on the semantics of the data

Geometry-based Spatio-temporal Data Mining Methods

Laube (2004)

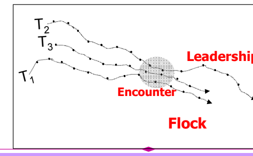
Proposed 5 trajectory patterns based on movement, direction, and location: convergence, encounter, flock, leadership, and recurrence

- *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)



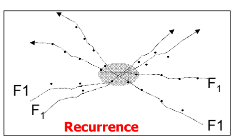
Laube [2004]

- *Flock pattern*: At least *m* entities are within a region of radius *r* and move in the same direction during a time interval >= *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 *t* 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)



Laube (2004)

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

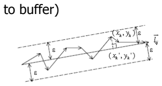


Extension of the work proposed by [Laube 2004, 2005]

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

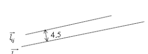
Frequent Sequential Patterns (Cao, 2005)

- Three main steps:
  1. **Transforms each trajectory in a line with several segments**
     - A distance tolerance measure is defined (similar to buffer)
     - All trajectory points inside this distance are summarized in one segment



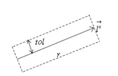
(a) Segment complies with l_{ij}^{e}

  2. **Similar segments are grouped**
     - **Similarity** is based on the **angle** and the **spatial lenght** of the segment
       - Segments with same angle and length have their distance checked based on a given distance *d* threshold



(b) Example for closeness

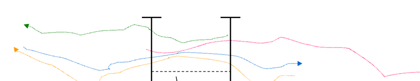
     - From the resultant groups, a medium segment is created
       - From this segment a region (buffer) is created
  3. **Frequent sequences of regions are computed** considering a *minSup* threshold



(c) Region r determined by $(\vec{l}, tol)$

Frequent Mobile Group Patterns (Hwang, 2005)

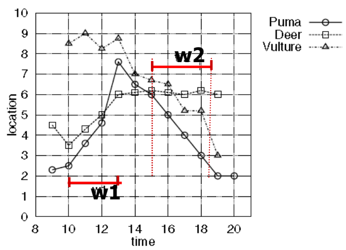
- A *group pattern* is a set of trajectories close to each other (with distance less than a given *minDist)* for a minimal amount of time (*minTime)*
- Direction is not considered
- Frequent groups are computed with the algorithm Apriori



Group pattern: *time, distance, and minsup*

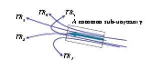
Co-Location Patterns (Cao 2006)

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



Traclus (Han, 2007)

- Clustering algorithm (TraClus-Trajectory Clustering)
- Group sub-trajectories



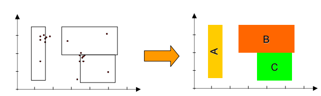
- Density-based
- Partition-and-group method
  - 1) each trajectory is partitioned into a set of line segments (sub-trajectories) with lenght *L* defined by the user
  - 2) similar segments (close segments) are grouped
    - Similarity is based on a distance function
- Clustering is based on spatial distance
  - time is not considerd
- Interesting approach for trajectories of hurricanes

T-Patterns (Giannotti, 2007)

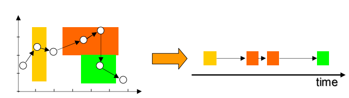
- Sequential Trajectory Pattern Mining
- Considers both space and time
- Objective is to describe frequent behaviour
  - considering visited regions of interest
  - During *movements* and the *duration* of movements
- Steps:
  1. compute or find regions of interest, based on dense spatial regions (no 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

T-Patterns (Giannotti, 2007)

- Fix a set of pre-defined regions



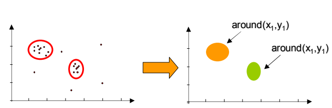
- Map each (x,y) of the trajectory to its region



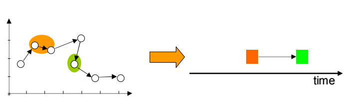
- Sample pattern: $A \xrightarrow{20\,\text{min.}} B$

T-Patterns (Giannotti, 2007)

- Detect significant regions thru spatial clustering



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



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

Summary

- These 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	...		...

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)

Changqing Zhou, Dan Frankowski, Pamela J. Ludford, Shashi Shekhar, and Loren G. Terveen. Discovering personally meaningful places: An interactive clustering approach. ACM Trans. Inf. Syst., 25(3), 2007.

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.

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.

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.

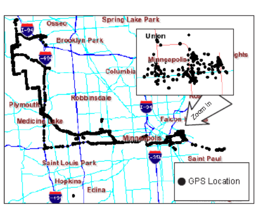
Hwang, S.-Y., Liu, Y.-H., Chiu, J.-K., and Lim, E.-P. (2005). Mining mobile group patterns: A trajectory-based approach. In Ho, T. B., Cheung, D. W.-L., and Liu, H., editors, PAKDD, volume 3518 of Lecture Notes in Computer Science, pages 713–718. Springer.

Cao, H., Mamoulis, N., and Cheung, D. W. (2006). Discovery of collocation episodes in spatiotemporal data. In ICDM, pages 823–827. IEEE Computer Society.

Semantic-based Spatio-temporal Data Mining Methods

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



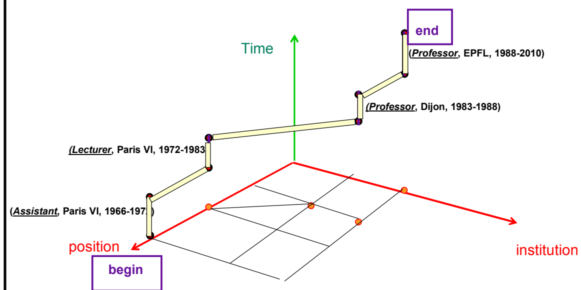
- Time is not considered

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

Different Kinds of Trajectories (Spaccapietra 2008)

- **Metaphorical Trajectory**:
  - travel in abstract space, e.g. a 2D career space <position, institution>



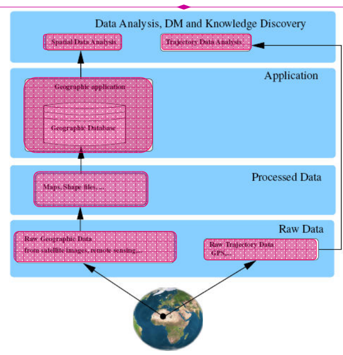
Semantic Trajectories - Motivation

Trajectory Samples (x,y,t)

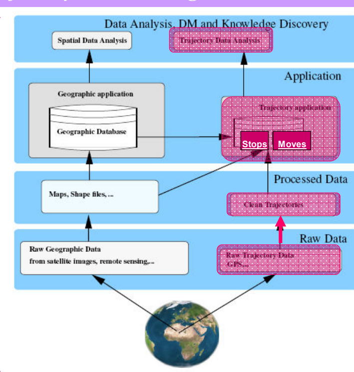
Geographic Data

Geographic Data +
Trajectory Data =
Semantic Trajectories

Geometry-based Trajectory DM Methods



Semantic Trajectory Data Mining Methods



Stops and Moves (Spaccapietra 2008)

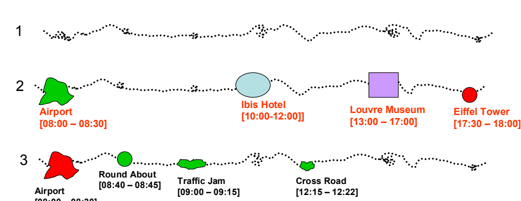
STOPS

- 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

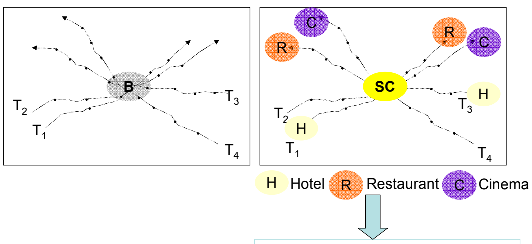
- Are the parts that are not stops

Stops and Moves are Application Independent



Geometric Patterns X Semantic Patterns (Bogorny 2008)

Geometric Pattern



Semantic trajectory Pattern

(a) Hotel to Restaurant, passing by SC
(b) go to Cinema, passing by SC

Geometric Patterns X Semantic Patterns (Bogorny 2008)

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

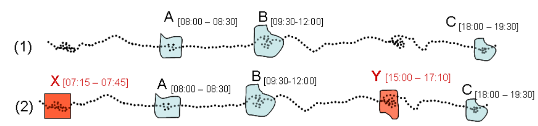
Consequence:

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

Methods for Adding Semantics to Trajectories (computing Stops and Moves)

Methods to Compute Stops and Moves

1) SMoT (intersection-based)



2) CB-SMoT (clustering-based)

SMoT: Candidate Stops and Application (Alvares 2007a)

A *candidate stop* C is a tuple (R_C, Δ_C), where

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

E.g. [Hotel - 3 hours]

An application A is a finite set

$A = \{C_1 = (R_{C1}, \Delta_{C1}), \dots, C_N = (R_{CN}, \Delta_{CN})\}$ of *candidate stops* with non-overlapping geometries $R_{C1}, \dots, R_{CN}$

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

SMoT: Stops and Moves (Alvares 2007a)

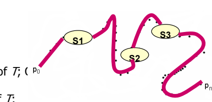
A *stop* of a trajectory T with respect to an *application* A is a tuple (R_{Ck}, t_j, t_{j+n}), such that a maximal subtrajectory of

$$T\{(x_i, y_i, t_i) \mid (x_i, y_i) \textit{ intersects } R_{Ck}\} = \{(x_j, y_j, t_j), (x_{j+1}, y_{j+1}, t_{j+1}), \dots, (x_{j+n}, y_{j+n}, t_{j+n})\}$$

where R_{Ck} is the geometry of C_k and $| t_{j+n} - t_j | \geq \Delta_{Ck}$

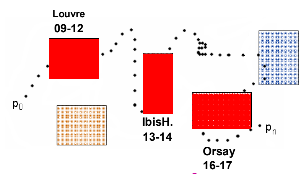
A *move* of T with respect to A is:

- a maximal contiguous subtrajectory of T:
  - between the starting point of T and the first stop of T; OR
  - between two consecutive stops of T; OR
  - between the last stop of T and the ending point of T;
- or the trajectory T itself, if T has no stops.



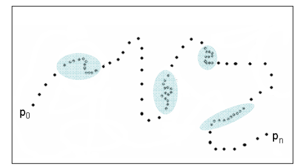
SMoT: Stops and Moves
(Alvares 2007[a])

- Input: $A = (\{C_1 = (R_{C_1}, \Delta_{C_1}), \ldots, C_N = (R_{C_N}, \Delta_{C_N})\})$ // Application
  $T = \langle (x_0, y_0, t_0), (x_1, y_1, t_1), \ldots, (x_N, y_N, t_N) \rangle$ // trajectory samples
- Output: S // Stops
  M // Moves
- Method:
  - For each trajectory in T

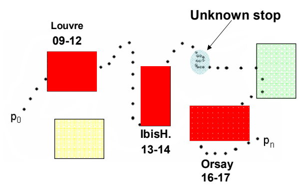


CB-SMoT: Stops and Moves
(Palma 2008)

- Based on stops and moves
- Cluster single trajectories based on speed:
  low speed → important place



Stops and Moves (CB-SMOT)
(Palma 2008)



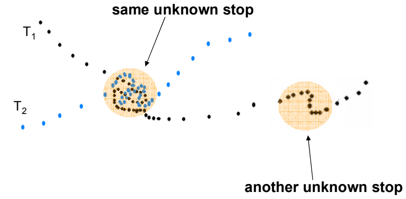
Step 1: find clusters

Step 2: Add semantics to each cluster

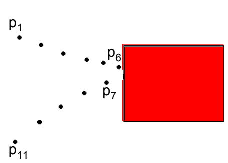
2.1: If intersects α during $\Delta t\alpha$ → stop α

2.2: If no intersection during Δt → *unknown* stop

Unknown Stops (CB-SMOT)
(Palma 2008)



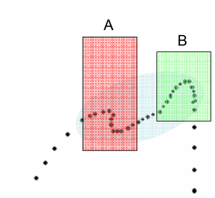
Can Find Clusters Inside Buildings
(Palma 2008)



t6= 10:10AM

t7= 10:32AM

Labeling clusters
(Palma 2008)

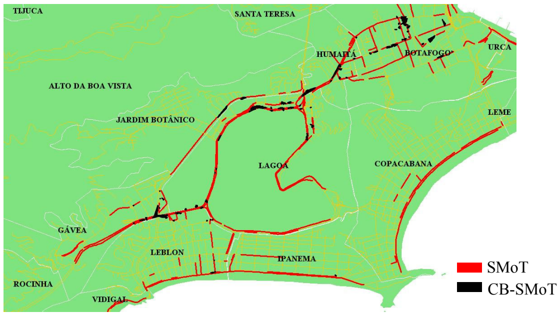


If the intersection time between the cluster and A $\geq \Delta t_A$ then A is a stop

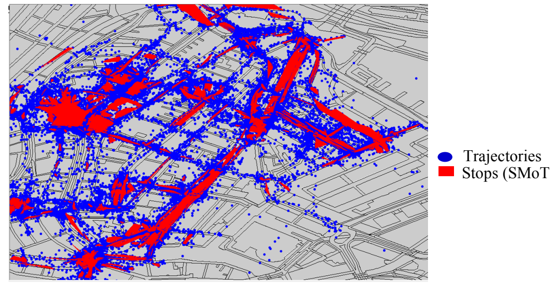
If the intersection time between the cluster and B $\geq \Delta t_B$ then B is a stop

the subtrajectory between A and B is a move

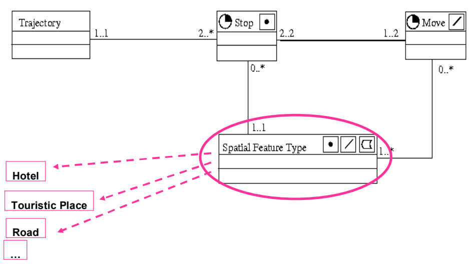
Stops in a Real Dataset (Transportation Application) (Bogorny 2008a)



Stops in a Real Dataset (Recreation Application)



Conceptual Schema of Stops and Moves



Schema of Stops and Moves (Alvares 2007a)

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
...	...		...
2	...		...

Stops

Tid	Sid	StopName	StopGid	Sbegint	Sendt
1	1	Hotel	1	08:25	08:40
1	2	TouristicPlace	2	09:05	09:30
1	3	TouristicPlace	3	10:01	14:20

Moves

Tid	Mid	S1id	S2id	geometry		timest
1	1	1	2	48.888880	2.246102	08:41
1	1	1	2	48.885732	2.255031	08:42
...	...	...	...	...	...	
1	1	1	2	48.860021	2.336105	09:04
1	2	2	3	48.860515	2.349018	09:41
...	...	...	...	...		...
1	2	2	3	48.861112	2.334167	10:00

Hotel

Id	Name	Stars	geometry
1	Ibis	2	48.890015 2.246100, ...
2	Meridien	5	48.880005 2.283889, …

Touristic Place

Id	Name	Type	geometry
1	Notre Dame	Church	48.853611 2.349167,…
2	Eiffel Tower	Monument	48.858330 2.294333,…
3	Louvre	Museum	48.862220 2.335556,…

Queries: Trajectory Samples X Stops and Moves (Alvares 2007a)

Q1: Which are the places that moving object A has passed during his trajectory?

```
SELECT ‘Hotel’ as place
   FROM trajectory t, hotel h
 WHERE t.tid='A' AND
            intersects (t.movingpoint.geometry,h.geometry)
UNION
SELECT ‘TouristicPlace’ as place
   FROM trajectory t, touristicPlace p
 WHERE t.tid='A' AND
            intersects (t.movingpoint.geomtetry,p.geometry)
UNION
…
```

```
SELECT stopName  as place
   FROM stop
WHERE tid='A‘
```


Sequential Patterns (Bogorny 2008b)

- Large Sequences of Length 2
- (41803_ruas_5,41803_ruas_5) Support: 7
- (41803_ruas_4,41803_ruas_4) Support: 9
- (41803_ruas_4,66655_ruas_4) Support: 5
- (41803_ruas_2,41803_ruas_2) Support: 6
- (41803_ruas_8,41803_ruas_8) Support: 5
- (41803_ruas_3,0_unknown_3) Support: 5

month

gid

Spatial feature type (stop name)

Sequential Patterns *(Bogorny 2008b)*

Large Sequences of Length 2

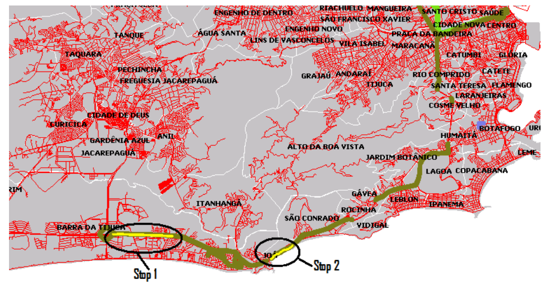
(41803_ruas_tuesday,41803_ruas_tuesday)	Support: 9
(41803_ruas_tuesday,66655_ruas_tuesday)	Support: 5
(41803_ruas_monday,66655_ruas_monday)	Support: 5
(41803_ruas_monday,41803_ruas_monday)	Support: 11
(41803_ruas_monday,0_unknown_monday)	Support: 5
(41803_ruas_thursday,41803_ruas_thursday)	Support: 13
(41803_ruas_thursday,0_unknown_thursday)	Support: 6
(41803_ruas_wednesday,41803_ruas_wednesday)	Support: 7

Day of the week

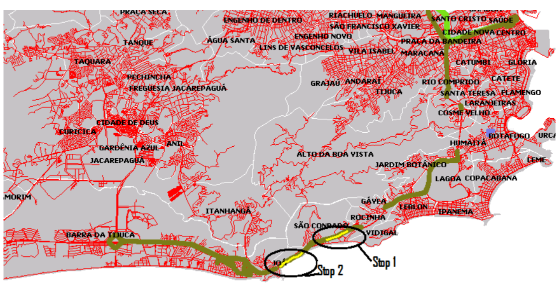
gid

Spatial feature type (stop name)

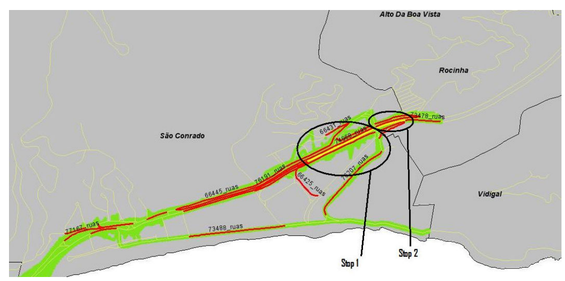
Sequential Patterns (Transportation Application)



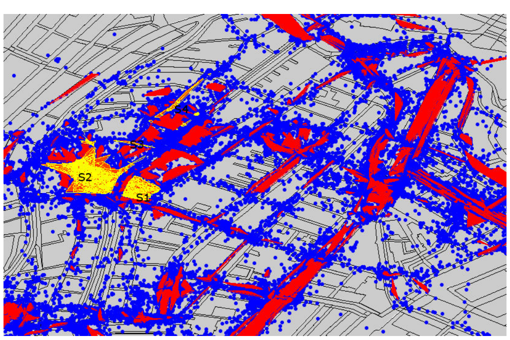
Sequential Patterns (Transportation Application)



Sequential Patterns (Transportation Application)



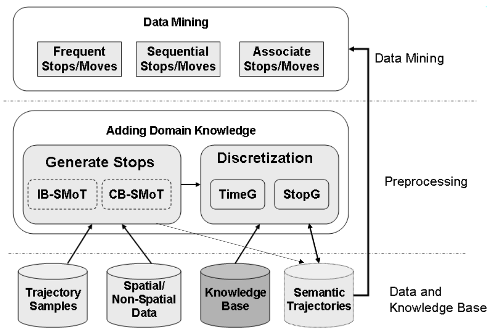
Sequential Patterns (Recreation Application)



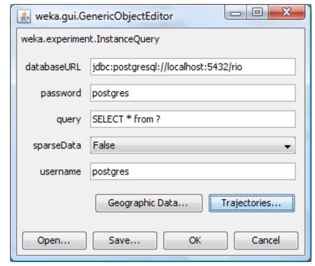
Tools for Semantic Trajectory Data Mining

Weka-STDM

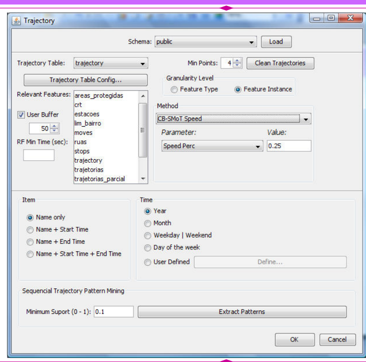
(Bogorny 2008b)



Weka-STDM



Weka-STDM



References

- Bogorny, V. ; Wachowicz, M. A Framework for Context-Aware Trajectory Data Mining. In: Longbing Cao, Philip S. Yu, Chengqi Ahang, Huaifeng Zhang. (Org.). Domain Driven Data Mining: Domain Problems and Applications. 1 ed. : Springer, 2008a.
- Bogorny, V., Kuijpers, B., and Alvares, L. O. (2008b). St-dmql: a semantic trajectory data mining query language. International Journal of Geographical Information Science. Taylor and Francis, 2008.
- 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., de Macedo, J. F., and Moelans, B. (2007a). Dynamic modeling of trajectory patterns using data mining and reverse engineering. In Twenty-Sixth International Conference on Conceptual Modeling - ER2007 - Tutorials, Posters, Panels and Industrial Contributions, volume 83, pages 149–154. CRPIT.
- 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.

Challenges and Open Issues in Spatial Data Mining

Challenges and Open Issues in Spatial Data Mining

- Focus on *clustering* methods
- Spatial association rules have received some attention
- Classification is still in its infancy
- Among quantitative approaches, co-location mining and outlier detection have been addressed

Challenges and Open Issues in Spatial Data Mining

- Several works focus on performance issues
- The QUALITY of the patterns has rarely been addressed
- Challenge: new intelligent methods are needed
- Semantics has to be considered to discover more interesting patterns

Challenges and Open Issues in Spatio-Temporal Data Mining

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

Challenges and Open Issues in Spatio-Temporal Data Mining

- Trajectory Similarity
  - Focus relies on different similarity measures
    - Shape, direction, closeness
  - Needs: semantic similarity

Challenges and Open Issues in Spatio-Temporal Data Mining

- Need for data mining methods using:
  - Metadata
  - Domain knowledge
  - Semantics
  - Ontologies
- For:
  - pattern pruning
  - improve the quality of the patterns
  - pattern interpretation

Challenges and Open Issues in Spatio-Temporal Data Mining

- Major NEEDS
- Image Mining (remote sensing images), from temporal and spatial perspective, has little been explored
- A lot of data exist, but only a few data mining methods
  - Aksoy, S., Koperski, K., Tusk, C., and Marchisio, G. (2004). Interactive training of advanced classifiers for mining remote sensing image archives. ACM International Conference on Data Mining.
  - Silva, M., Câmara, G., Souza, R., Valeriano, D., and Escada, M. (2005). Mining Patterns of Change in Remote Sensing Image Databases. Proceedings of the Fifth IEEE International Conference on Data Mining.

More needs

- There is a need for collaboration between data miners and domain experts (environmental experts, transportation managers, metheorologists, 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

Thank You!