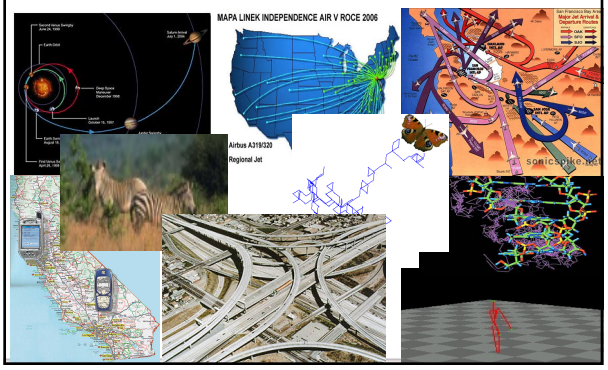
Mining Data From Mobile Devices

Dimitrios Gunopulos
University of Athens
dg@di.uoa.gr

Trajectories are everywhere



Smart-phones: Novel applications

Smart-phones as a platform for sense and share information applications.

- Personalized traffic, weather information
- Social networking
- Smart city applications
- Similar trajectories (e.g. for popular tourist routes)

Data come from GPS, also other sources (e.g. GPS tagged images, crowdsourcing mechanisms)
Users are becoming producers and consumers of data

Interesting Applications

- **Intelligent Transportation Systems:** "*Find new possible bus routes which would help at least K users.*"
- **Social Networks:** *"Find new cycling routes from MOMA to the Julliard"*
  - GeoLife, GPS-Waypoints, Sharemyroutes, etc. offer centralized counterparts.
- **Habitant Monitoring:** Find zebras that moved more similarly to zebra X before it got injured.

Analyzing GPS trajectory data

- App deployment allows location data collection
- How to use trajectory data
  - Integrate with other (e.g. twitter/sensor) data
  - Supplement flow based traffic data
    - Where is the improvement?
  - Building better/more general traffic models?
  - Better and more accurate maps?
  - Demand modeling?
    - Finding patterns, traffic modes
  - Building modeling tools for communities?
  - Real-time situation understanding and reaction

A Word on GPS Trace Collection

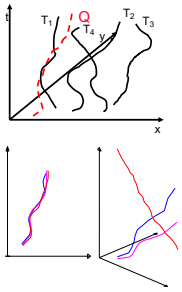
- Popular Smartphones are already collect positional information. Same applies to **Social Networking** Applications (e.g., Latitude, Gowalla, Twitter, etc.)
- **iPhone User Position Logging:**
  - iPhone collects coarse-grain positional information (i.e., triangulated Cell tower) locally on your smartphone (and iTunes backup).
  - The **unencrypted log** file is even **migrated between devices**!
  - Displaying your iPhone trace history on a map: http://petewarden.github.com/iPhoneTracker/

Problems

- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

Trajectory Similarity

- Given a query trajectory *Q* and a trajectory database $D=\{T_i\}$ search among them for the most similar trajectory
  - Determine the most similar by using not only the spatial shape of the trajectories but also their evolution in time
- Similarity applications:
  - Clustering
  - Classification
  - Frequent subtrajectories
  - Indexing

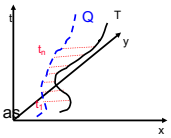


[FGP+07] Frentzos, E., Gratsias, K., Pelekis, N., Theodoridis, Y.: Algorithms for Nearest Neighbor Search on Moving Object Trajectories. GeoInformatica 11(2): 159-193 (2007).

Metrics for Measuring Similarity

- Definition of Dissimilarity *DISSIM(Q,T)* between trajectories *Q* and *T* being valid during the period $[t_1, t_n]$: the definite integral of the function of time of the Euclidean distance between the two trajectories during the same period

$$DISSIM(Q,T) = \int_{t_1}^{t_n} D_{Q,T}(t)dt$$



- In real world, each trajectory is represented as a collection of discrete points
  1. Apply linear interpolation in between
  2. Use a discrete measure

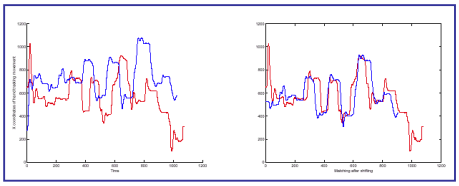
$$DISSIM(Q,T) = \sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} D_{Q,T}(t)dt$$

[FGP+07]

Requirements for Similarity Model

We need to address the following issues:

- Similar Motions in different space Regions



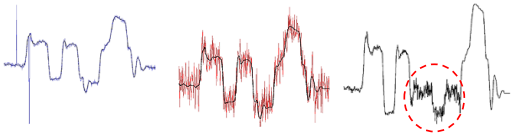
[VGK02] Vlachos, M., Gunopulos, D., Kollios, G.: Discovering Similar Multidimensional Trajectories. ICDE 2002

Requirements for Similarity Models

We need to address the following issues:

- Outliers

Random Peaks | Noise Everywhere | Non Recoverable Part

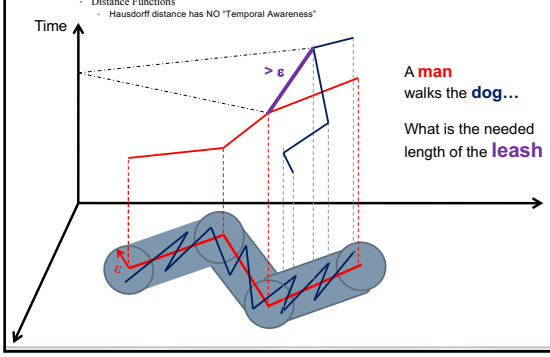


- Different Lengths

[VGK02]

Similarity Functions

- Distance Functions
  - Hausdorff distance has NO "Temporal Awareness"



A **man** walks the **dog…**

What is the needed length of the **leash**

Frechet distance

- Fréchet distance (length of the leash)

Let C1: $[a_1,b_1] \rightarrow \mathbf{R}^2$ and C2: $[a_2,b_2] \rightarrow \mathbf{R}^2$ denote two curves
their Fréchet distance is defined as

$$\delta_f(C_1,C_2) = \inf_{\alpha,\beta} \max_{t\in[0,1]} \| C_1(\alpha(t)) - C_2(\beta(t)) \|$$

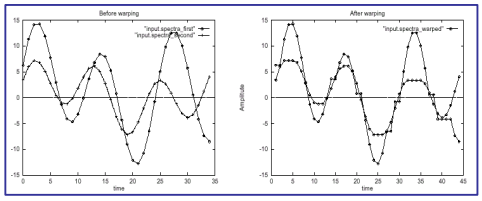
α (respectively β) range over all the continuous and monotonically increasing functions $[0,1] \rightarrow [a_1,b_1]$ (respectively $[0,1] \rightarrow [a_2,b_2]$)

Clearly, a lot more general – but also harder to compute than the Hausdorff distance

Computation: O(mn log(mn)) [AG95]

Dynamic time warping (DTW)

- **Time Warping**
  - **Allows stretching in time axis**
  - **Matches all elements**
  - **Extensively used in speech recognition domain**



[VGK02]

Dynamic time warping (DTW)

- Euclidean (L2) distance is

$$D_2(\mathbf{x},\mathbf{y}) := \sum_{t=1}^{N} (x_t - y_t)^2$$

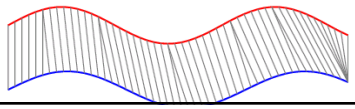
or, recursively,

$$D_2(\mathbf{x}_{1:t},\mathbf{y}_{1:t}) = (x_t - y_t)^2 + D_2\left(\mathbf{x}_{1:(t-1)},\mathbf{y}_{1:(t-1)}\right).$$

- Dynamic time warping distance is

$$D_w(\mathbf{x}_{1:i},\mathbf{y}_{1:j}) := (x_i - x_j)^2 + \min \begin{cases} D_w\left(\mathbf{x}_{1:(i-1)},\mathbf{y}_{1:(j-1)}\right) \\ D_w\left(\mathbf{x}_{1:(i-1)},\mathbf{y}_{1:j}\right) & \leftarrow \text{shrink } \mathbf{x} \text{ / stretch } \mathbf{y} \\ D_w\left(\mathbf{x}_{1:i},\mathbf{y}_{1:(j-1)}\right) & \leftarrow \text{stretch } \mathbf{x} \text{ / shrink } \mathbf{y} \end{cases}$$

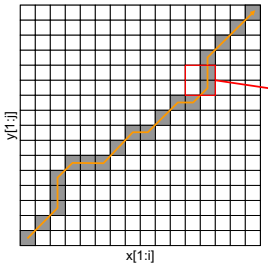
where $\mathbf{x}_{1:i}$ is the subsequence $(x_1, \ldots, x_i)$

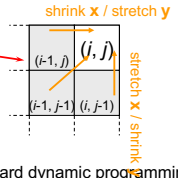


[VP06]

Dynamic time warping algorithm

- DTW allows any path
- Examine all paths:





- Standard dynamic programming to fill in table—top-right cell contains final result

[VP06] Vlachos, M., Papadimitriou, S.:Hands-on Time-Series Analysis with Matlab. Tutorial, ICDM 2006

Longest Common Subsequence (LCSS)

Arithmetic Example:

- **t1=[0, 4, 6, 8, 7, 4, 6, 5, 6, 4, 6]**
- **t2=[0, 3, 4, 6, 7, 6, 3, 6, 4, 6]**

- **Definition**

$$LCSS(A,B) = \begin{cases} 1 + LCSS(Head(A), Head(B)), & \text{if } a_n = b_n \\ \max(LCSS(Head(A), B), LCSS(A, Head(B))), & \text{otherwise} \end{cases}$$

- **Dynamic Programming Solution**

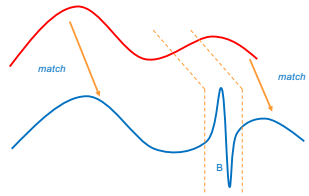
$$LCSS[i,j] = \begin{cases} 0, & \text{if } i = 0 \\ 0, & \text{if } j = 0 \\ 1 + LCSS[i-1, j-1], & \text{if } a_i = b_i \\ \max(LCSS[i-1, j], LCSS[i, j-1]), & \text{otherwise} \end{cases}$$

[VGK02]

Differences between Frechet, DTW & LCSS

- **DTW has to much every element, so is more sensitive to outliers**
  LCSS matches only the similar parts

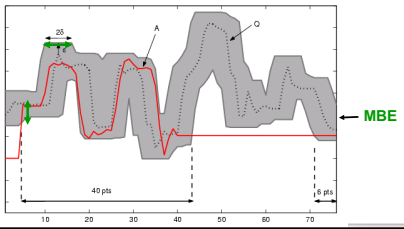
[VGK02]



[DTSWK08] provides experimental comparison of 8 representation methods; 9 similarity measures (+variants); 38 datasets. [WMDTSK12] provides further extensions.

LCSS for real values

$$LCSS_{\delta,\varepsilon}(A,B) = \begin{cases} 0 & \text{if A or B is empty} \\ 1 + LCSS_{\delta,\varepsilon}(Head(A), Head(B)) & \text{if } |a_n - b_m| < \varepsilon \text{ and } |n-m| \le \delta \\ \max(LCSS_{\delta,\varepsilon}(Head(A),B), LCSS_{\delta,\varepsilon}(A, Head(B))) & \text{otherwise} \end{cases}$$

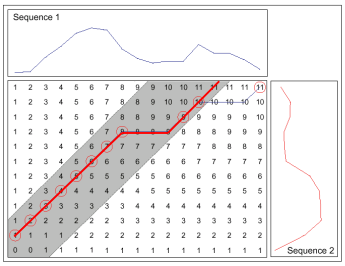


[VGK02]

LCSS Dynamic Programming algorithm

LCSS can be computed in O($\delta(l_1+l_2)$) by dynamic programming algorithm.

$$LCSS[i,j] = \begin{cases} 0 & \text{if } i = 0 \\ 0 & \text{if } j = 0 \\ 1 + LCSS[i-1, j-1] & \text{if } a_i = b_i \\ \max(LCSS[i-1,j], LCSS[i,j-1]) & \text{otherwise} \end{cases}$$



$$S1(\delta,\varepsilon,A,B) = \frac{LCSS_{\delta,\varepsilon}(A,B)}{min(n,m)}$$

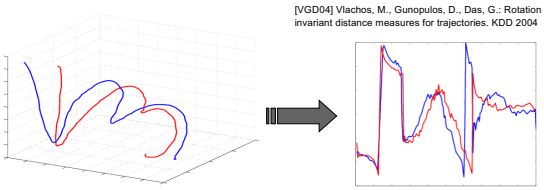
[VGK
Simil

Space Transformations: Rotation

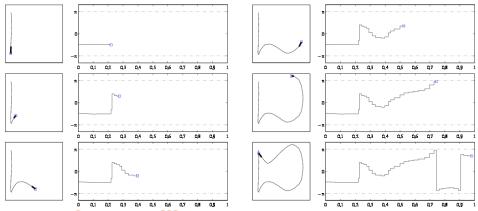
Method:

- Transform trajectory to rotation invariant space
  - *Positional information -> Angle/ Arc length information*
- Proper normalization ensures also:
  - *Translation invariance*
  - *Scale Invariance*
- Perform matching in new space
  - *Use Warped Matching to compensate for shape variations*

[VGD04] Vlachos, M., Gunopulos, D., Das, G.: Rotation invariant distance measures for trajectories. KDD 2004



Mapping Example



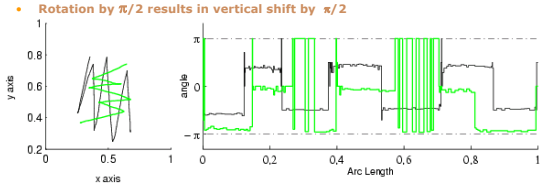
Angles to utilize:

- Exact angles (*as in figure*)
- Relative Angles (to previous segment)
- Angles from center of mass

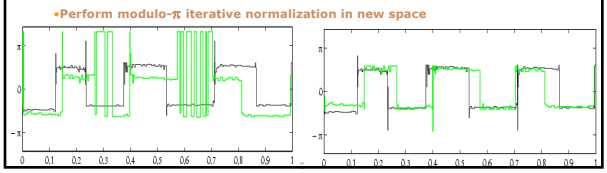
[VGD04]

Mapping Refinement (Vlachos et al 04)

- **Rotation by π/2 results in vertical shift by π/2**

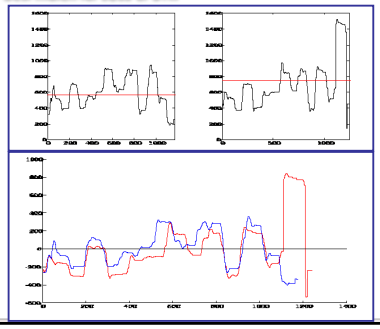


- **Perform modulo-π iterative normalization in new space**



Parallel Movements: Normalization

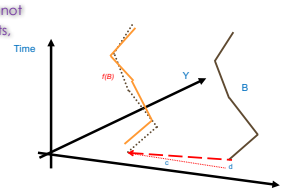
- **Normalization (average value subtraction) does not guarantee best match for LCSS or DTW**



[VGK02]

Extending LCSS: Translation Invariant Similarity

- Standard techniques cannot detect parallel movements,



- So, we define S2:

$$S2 = \max_{f_{c,d} \in F} S1(\delta, \varepsilon, f_{c,d}(B))$$

- S2 can detect parallel movements
- Better accuracy than simple normalization
- Distance *D1= 1-S1* & distance *D2 = 1-S2*

[VGK02]

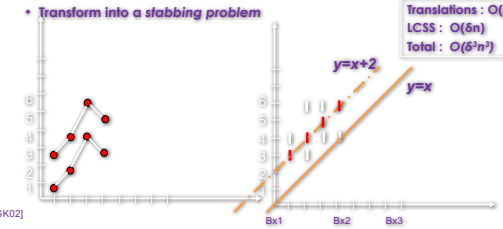
Exact Algorithm for Translation Invariant Similarity

For trajectories A, B with length n we want to find:

- translation $f_{c,d}$ that maximizes LCSS between A and $f_{c,d}$ *(B)*
- Not infinite translations.
  - Each dimension separately
  - A translation in 1D: *fc(bi) = bi + c* (line with slope 1)
  - *fc(bi)* will allow bi to be matched to all aj: $|i-j| < \delta$ & $ai-\varepsilon \le fc(bi) \le (bi, aj+\varepsilon)$

- **Transform into a *stabbing problem***

Translations : $O(\delta^2 n^2)$
LCSS : $O(\delta n)$
Total : $O(\delta^2 n^3)$

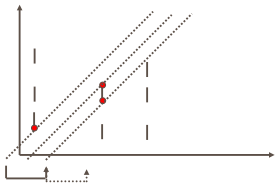


[VGK02]

Approximate Algorithm

A translation corresponds to a line *fc(x)* = x+c.

- **Sort translations by c → THEY DIFFER IN HOW MANY SEGMENTS?**



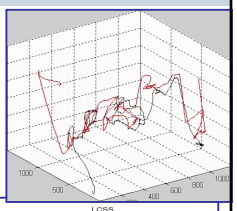
- **If we can afford to be within β of *max(Sim)* → we can afford to lose *βn* elements**
- Don't take all translations → we can examine every βn translations each time
- So, if we examine every βn, we lose *at most* βn elements (1D)
- So, for 2D, we can skip every βn/2 translations

[VGK02]

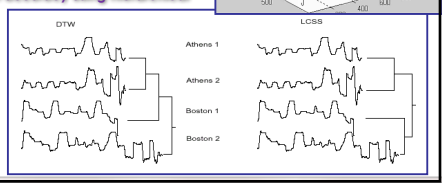
Classification Accuracy

Example using Video Tracking data

- **The 2D time series of a human tracking feature (e.g. the end of a finger). In conjunction with a "spelling program" the user can "write" various words.**
- **The data we used correspond to: *'athens', 'berlin', 'london', 'boston', 'paris'*.**



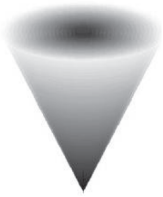
Test classification accuracy using Hierarchical Clustering



[VGK02]

Similarity and Uncertainty

1. **Uncertainty in AI and Databases**
   - Model-Theoretic issues (Possible Models)
   - NULLs (open vs. closed world assumption)
   - Probabilistic databases (and possible worlds)
2. **Time Geography and Inexact Geometries**
   - Uncertain location over time
   - Re-evaluation of Euclidian postulates



[Roberts73,T11,WF10,WY10]

Vector representation of approximate/uncertain trajectories (Pelekis et al, 09)

- Assume a regular grid $G(m \times n)$ consisting of cells $c_{k,l}$, a trajectory $T_i = <(x_{i,0}, y_{i,0}, t_{i,0}), \ldots, (x_{i,n_i}, y_{i,n_i}, t_{i,n_i})>$ and a target dimension $p << n_i$,
- The "**approximate trajectory**"

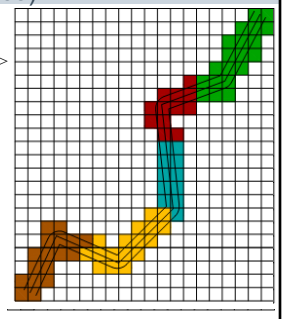
$$\overline{T}_i = <r_{i,1}, \ldots, r_{i,p}>$$

consists of *p* regions (i.e. sets of cells) crossed by T_i during period p_j

$$p_j = \left[\frac{ls \cdot (j-1)}{p}, \frac{ls \cdot j}{p}\right]$$

- The "**Uncertain Trajectory**" is the *ε*-buffer of

$$UnTra(\overline{T}_i) = <ur_{i,1}, \ldots, ur_{i,p}>$$

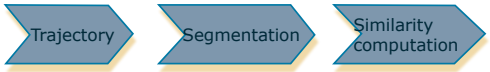


[PKK+09] Pelekis, N., Kopanakis, I., Kotsifakos, E.E, Frentzos, E., Theodoridis, Y.: Clustering Trajectories of Moving Objects in an Uncertain World. ICDM 2009

Problems

- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

Similarity in Parts of Trajectories

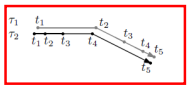


- (a) Trajectory Segmentation
  - segments of homogeneous movement characteristics
  - to convert a trajectory into a simplified and yet revealing structure
  - trajectory is transformed to a movement parameter class sequence
- (b) Similarity Computation
  - Normalized Weighted Edit Distance (NWED)
  - Based on edit distance
  - applies weighted conversion costs (insertion/deletion, substitution)

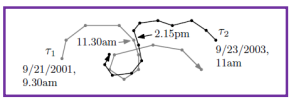
[DLW12]

Similarity of Parts of Trajectories

- Finding longest similar part (time-shifts) [BBvK+09]



Similar shapes (routes), but different speeds at different times

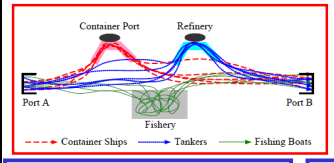


End-part of τ1 similar to the middle-part of τ2

$$\frac{\int_{t_s}^{t_s+T} d(\tau_1(t), \tau_2(t+t_{shift}))\, dt}{T}$$

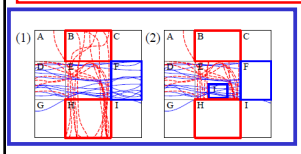
Average distance in a time-interval T, with time-shift…

Trajectories Clustering

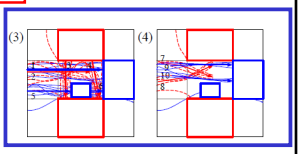


TraClass [LHL+08]

Partition trajectories, and then perform hierarchical
-Region-based;
-Trajectory-based clustering.

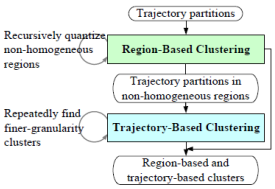


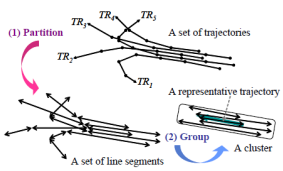
Region-based



Trajectory-based

Trajectories Clustering

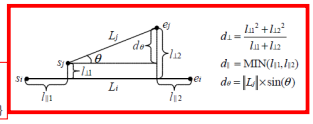




Each of di's is weighted and added…

$$sep\_gain(C_i) = \frac{1}{|\Upsilon(C_i)|} \sum_{C_j \in \Upsilon(C_i)} dist(C_i, C_j),$$

where $\Upsilon(C_i) = \{C_j \mid C_j \in \mathcal{T} \wedge CL(C_i) \neq CL(C_j)\}$



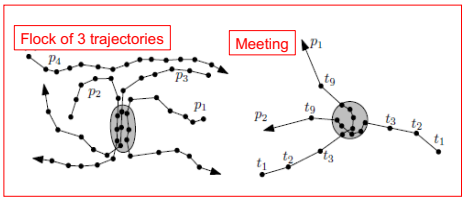
$$d_\perp = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\perp 1} + l_{\perp 2}}$$

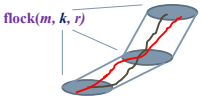
$$d_\parallel = \mathrm{MIN}(l_{\parallel 1}, l_{\parallel 2})$$

$$d_\theta = |L_l| \times \sin(\theta)$$

Flocks

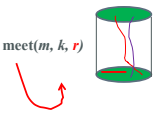
- Trajectories that "stay together for a while"





Maximize ***k*** for
- Fixed
- Varying

"versions" of ***m***



[GvK06]

Flocks

- Trajectories that "stay together for a while"

Approximate variants of both flock/meet and fixed/varying: for the same duration, make sure that there exists a disk of radius ≤ ***cr*** where ***c = 1 + ε***

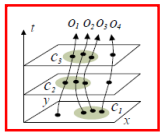
Complexity Results

Subset \ Pattern	Meeting		Flock	
Varying subset	$O(n^4\tau^2 \log n + n^2\tau^3)$	(Exact)	$O(n^3\tau \log n)$	(Exact)
	$O(\frac{n^2}{m\varepsilon^2}\tau \log n \log \tau)$	$((1+\varepsilon)$-apx. $(*))$	$O(\frac{n^2}{\varepsilon^2}\tau \log n)$	$((2+\varepsilon)$-apx.)
	$\Omega(n^2\tau)$	(Any apx.)	$\Omega(n^2\tau)$	(Any apx.)
Fixed subset	$O(n^4\tau^2 \log n + n^2\tau^3)$	(Exact)	NP-hard	(Any $(2-\varepsilon)$-apx.)
	$O(\frac{n^2}{m\varepsilon^2}\tau \log n \log \tau)$	$((1+\varepsilon)$-apx. $(*))$	$O(n^2\tau \log n)$	(2-apx.)
	$\Omega(n^2\tau)$	(Any apx.)	$\Omega(n^2\tau)$	(Any apx.)

Other predicates (patterns): ***flock, leadership, convergence, encounter***
[LvKI04]

Convoys

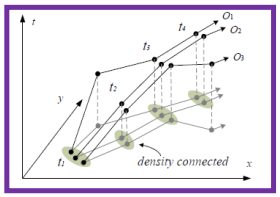
- Moving Clusters and "lossy flocks"…



$(C_t \cap C_{t+1}) / (C_t \cup C_{t+1}) \geq \theta$

[KMB05,SNT+06]]

Given N trajectories, a distance threshold ε, integer m, and integer-lifetime k
convoy query = {all groups of trajectories, consisting of a (maximal) set of *density-connected objects* with respect to ε and m, during at least k consecutive time instants}



[JLX+08]

Problems

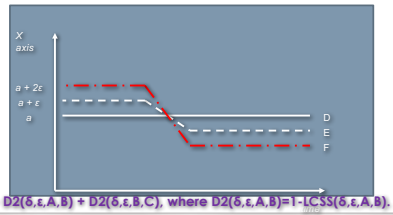
- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

Metric Properties of Similarity Measures

Triangle inequality does not hold for DTW or LCSS
(but holds for Edit Distance or DTW with gaps).

Example 1: A = 1,1,5
B = 2,2,5
C = 2,2,2,2,2,5
Then: DTW(A,B) = 2, DTW(B,C) = 0, DTW(A,C) = 5

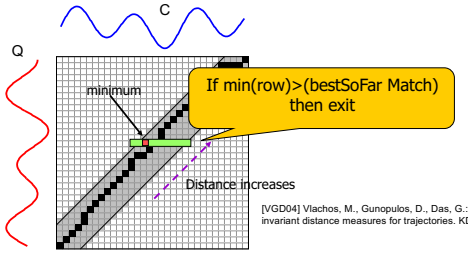
Example 2:



D1(δ,ε,A,C) > D2(δ,ε,A,B) + D2(δ,ε,B,C), where D2(δ,ε,A,B)=1-LCSS(δ,ε,A,B).

Sequential Scan

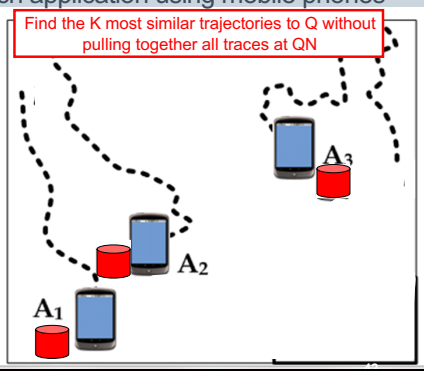
- Compare the Query to all the trajectories in the database
- Use Lower/Upper Bounds or Early Termination to speed up the search:



[VGD04] Vlachos, M., Gunopulos, D., Das, G.: Rotation invariant distance measures for trajectories. KDD 2004

Fully distributed, privacy preserving trajectory search application using mobile phones

Find the K most similar trajectories to Q without pulling together all traces at QN



Constraints and Objectives

Don't Disclose the User's Trajectories (privacy constraints)

- Social sites are already undergoing significant privacy restructuring
- Trajectories are large (270MB/year with 2s samples)

Minimize Net Traffic and Optimize Local Processing (algorithmic constraints)

- 3G/4G and WiFi traffic: i) depletes smartphone battery and ii) degrades network health

Basic Operation on Smartphone	Power (mW = mJ/s)
CPU Idle (OS running)	175 mW
CPU Busy (Processing)	**369 mW**
WiFi Idle (Connected)	38 mW
WiFi Busy (Uplink 123Kbps, -58dBm)	**600 mW**
LCD Brightness (economy mode)	300 mW

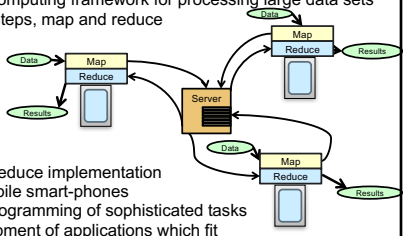
43

Constraints and Objectives

Quick and efficient deployment (Programmability constraints)

MapReduce programming model

- Flexible distributed computing framework for processing large data sets
- Consists of 2 major steps, map and reduce



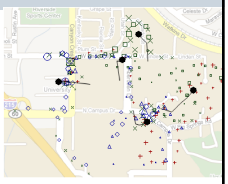
Misco (DEBS'10)

- Distributed MapReduce implementation for a cloud of mobile smart-phones
- Allows efficient programming of sophisticated tasks
- Simplifies development of applications which fit into this framework
- Hides difficult low-level implementation details
  - Code distribution, failures, heterogeneity

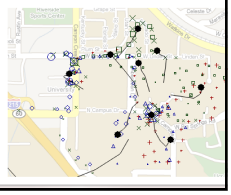
Clustering in a distributed MapReduce setting

Dowser - a Connectivity Discovery system

- Identify areas of high WiFi connectivity
- Processing computed in a online fashion
- Data from multiple cooperating users
- Preserves data privacy between users



Clustering is performed on individual devices without the need to distribute data between devices



Each Worker:

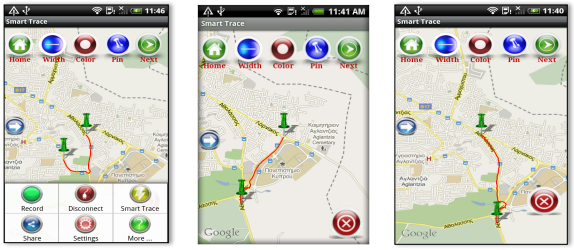
- selects K points or uses centroids from server
- performs local clustering
- sends centroids to server

Server:

- performs weighted clustering on centroids
- new centroids are sent back to workers
- Repeats until centroids no longer move.

Trajectory Similarity Search in a Distributed Environment

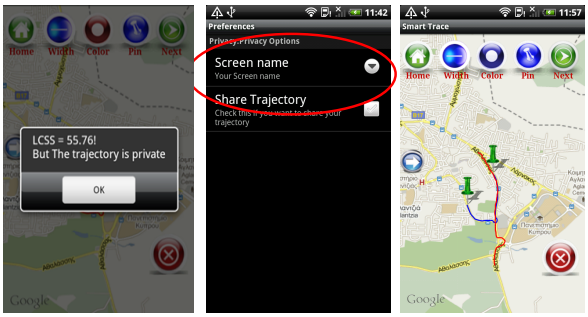
- SMARTTRACE: Server implemented in JAVA (4,500 LOC)
- Client implemented in JAVA on Android (2,500 LOC + XML files)



Query Device B Device C

[CLZ+11]

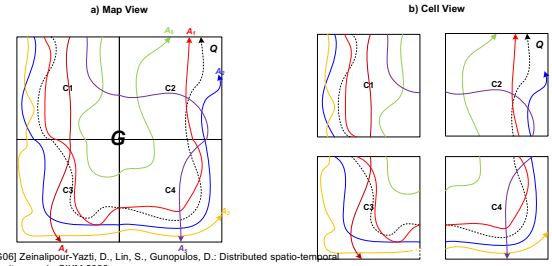
SmartTrace



Answer Privacy Setting Answer With Trace

Sensor Networks Setting

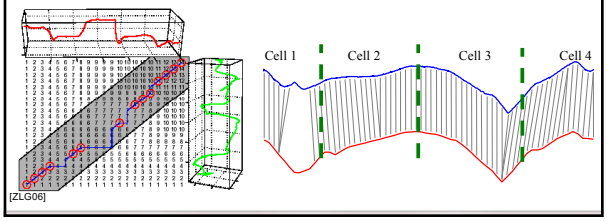
- Monitoring area G, m objects moving inside
- n cells each has a camera sensor:C1,C2,C3,C4
- Each record of the trajectory is stored locally at the closest sensor



[ZLG06] Zeinalipour-Yazti, D., Lin, S., Gunopulos, D.: Distributed spatio-temporal similarity search. CIKM 2006

Problems with distributed computation of DTW/LCSS

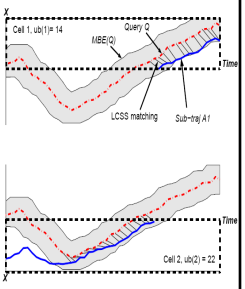
- Computing DTW/LCSS without sharing all data is difficult, because
  - Sequential matching problem
  - Matching may occur across cells



[ZLG06]

Distributed LCSS Upper Bound

- Each cell uses $LCSS_{\delta,\varepsilon}(MBE(Q), A_{ij})$ to calculate the similarity of each local sub_trajectory A_{ij} to $MBE(Q)$
- Upper bound $DUB\_LCSS(Q,A_i)$ is computed by adding the n local results



Theorem 1

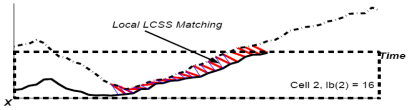
$$\sum_{j=1}^{n} LCSS_{\delta,\varepsilon}(MBE(Q), A_{ij}) \geq LCSS_{\delta,\varepsilon}(Q, A_i)$$

[ZLG06]

Distributed LCSS Lower Bound

- For each trajectory A_i, cell c_j finds the time region T_{ij} = {ts(p)|p in A_{ij}} when A_i stays in cell c_j. Filter Q into Q_{ij}' such that Q_{ij}' is in the same time intervals as A_{ij} , Q_{ij} = {p|p in Q and ts(p) in T_{ij}}.
- Each cell performs a local computation of $LCSS_{\delta,\varepsilon}(Q'_{ij}, A_{ij})$
- The lower bound $DLB\_LCSS(Q,A_i)$ is computed by adding the n local results

Theorem 2 $\sum_{j=1}^{n} LCSS_{\delta,\varepsilon}(Q'_{ij}, A_{ij}) \leq LCSS_{\delta,\varepsilon}(Q, A_i)$



[ZLG06]

Distributed top-K computation with bounds

- Now we have the **Lower** and **Upper Bounds** rather than **Exact scores**.
- e.g. instead of sim(A0,Q)=20 it gives us [A0,15,25]

v1 id,lb,ub	v2 id,lb,ub	v3 id,lb,ub	METADATA id,lb,ub
A2,3,6	A4,4,5	A4,1,3	A4,10,18
A0,4,8	A2,5,6	A0,6,10	A2,13,19
A4,5,10	A0,5,7	A2,5,7	A0,15,25
A7,7,9	A3,5,6	A9,6,7	A3,20,27
A3,8,11	A9,8,10	A3,7,10	A9,22,26
A9,8,9	A7,12,13	A7,11,13	A7,30,35
....	....	....	....

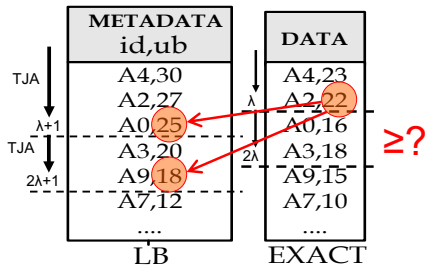
m

n

- Extend current work on Top-K algorithms: TA (Fagin et al, 01), TPUT (Cao et al, 04), TJA (Zeinalipour et al, 05)

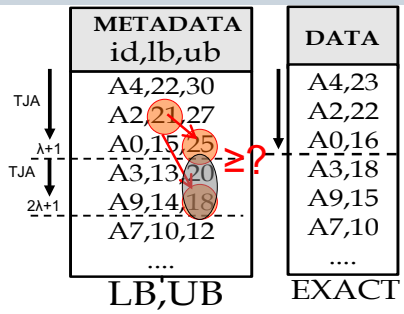
UB-K Algorithm

Query: Find the K=2 highest ranked answers



Why not stop at 25?
Because we might have another object X [UB:24, Real:23]

UBLB-K Algorithm



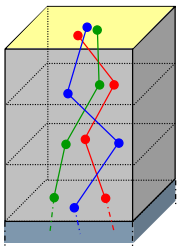
Note: Kth highest LB is: 21
Therefore A3 (UB:20) and below are not necessary

Problems

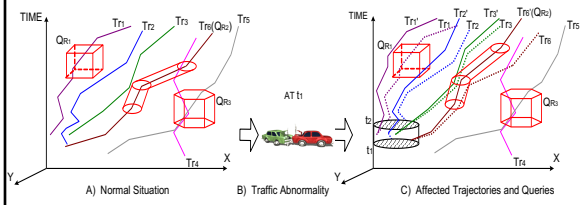
- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

Querying Full-Trajectories (Past)

- Examples:
  - Querying Along the Temporal Dimension: *What was the location of a certain object from 7:00 AM to 10:00 AM yesterday?*
  - Querying Along the Spatial Dimension: *Find all objects that were in a certain area at 7:00 AM yesterday*
  - Querying Along the Spatio-temporal Dimension: *Find all objects that were close to each other from 7:00 AM to 8:00 AM yesterday*
- Features:
  - Large number of historical trajectories
  - Persistent read-only data
  - The ability to query the spatial and/or temporal dimensions



Querying Full-Trajectories (Future)



Disturbance zone affects some moving objects' trajectories in the future

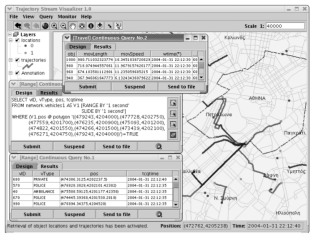
Hence, queries need re-evaluation

Triggers; Ordering (space-filling curves); Indexing of Queries

[DTS08,JLO04]

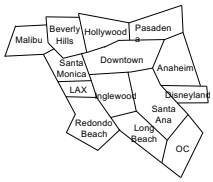
Continuous queries over trajectories

- A framework was proposed for managing trajectory streams
- Continuous spatiotemporal queries could be posed over trajectories
- Real time responses of the given queries [PS04]
- A sql like interface was proposed to declare the spatiotemporal queries
- A graphical interface was proposed, were users could specify queries [PKS08]



Motion Pattern Queries

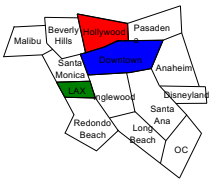
Example: "find trajectories that were in **downtown LA**, then sometime later went as *close as possible* to the **Hollywood sign**, then later ended up at **LAX**"



[VBT10]

Motion Pattern Queries

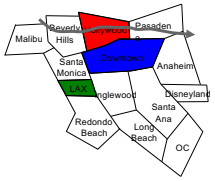
Example: "find trajectories that were in **downtown LA**, then sometime later went as *close as possible* to the **Hollywood sign**, then later ended up at **LAX**"



[VBT10]

Motion Pattern Queries

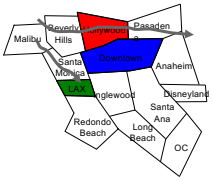
Example: "find trajectories that were in **downtown LA**, then sometime later went as *close as possible* to the **Hollywood sign**, then later ended up at **LAX**"



[VBT10]

Motion Pattern Queries

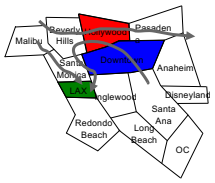
Example: "find trajectories that were in **downtown LA**, then sometime later went as *close as possible* to the **Hollywood sign**, then later ended up at **LAX**"



[VBT10]

Motion Pattern Queries

Example: "find trajectories that were in **downtown LA**, then sometime later went as *close as possible* to the **Hollywood sign**, then later ended up at **LAX**"



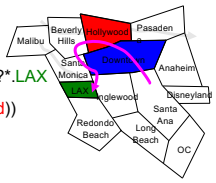
[VBT10]

Motion Pattern Queries

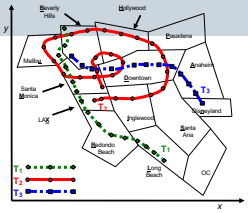
Example: "find trajectories that were in **downtown LA**, then sometime later went as *close as possible* to the **Hollywood sign**, then later ended up at **LAX**"

P=downtown LA.?*.@x.?*.LAX

min(*d*(@x,Hollywood))



[VBT10]



(a) Region-based trajectory representation

region-list

Downtown: T_1(7,9); T_2(21,23); T_3(5,10) ...
Santa Monica: T_1(10,13); T_2(18,21); T_3(25,27); T_4(10,11) ...
Beverly Hills: T_1(14,16); T_2(13,15); T_3(24,25); T_4(11,13) ...
LAX: T_1(9,10); T_4(1,3) ...
...

(b) region-*list* index

trajectory-list

T_1: *L*(1,5) *R*(5,9) *X*(9,10) *M*(10,13) *U*(13,14) *B*(14,16)
T_2: *X*(1,3) *R*(3,5) *S*(5,7) *D*(7,9) *P*(9,10) *H*(10,13) *B*(13,15) *U*(15,18) *M*(18,21) *D*(21,23) *H*(23,24) *B*(24,25) *M*(25,27)
T_3: *M*(1,2) *A*(2,5) *D*(5,10) *M*(10,11) *B*(11,13)
...

(c) *trajectory-list* index

raw trajectories

T_1: (7.5,2.2,1) (6.8,2.4,2) (6.4,2.8,3) (6.0,3.0,4) (5.2,3.1,5) ...
T_2: (3.9,5.9,1) (4.1,6.1,2) (5.0,6.2,3) (5.8,6.4,4) (6.5,6.4,5) ...
T_3: (9.3,5.2,1) (8.9,6.1,2) (8.7,6.7,3) (8.0,7.2,4) (7.7,7.8,5) ...
...

(d) raw trajectory archive

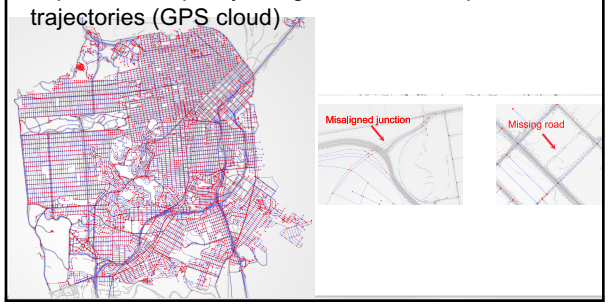
[VBT10]

Problems

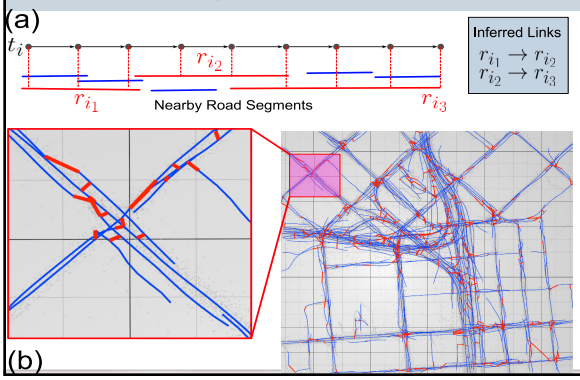
- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

The Map creation problem

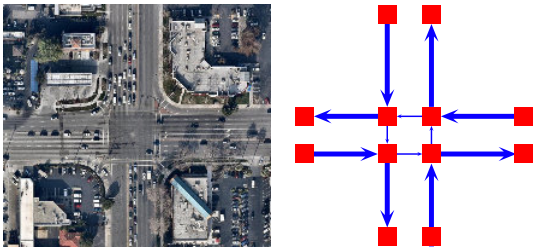
Maps are incomplete and often out of date
Improve their quality using collections of sparse trajectories (GPS cloud)



Combine segments to form roads

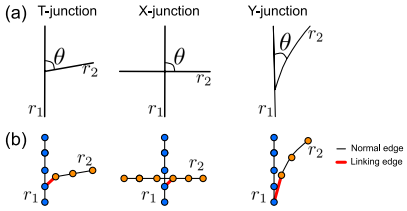


How to model street networks

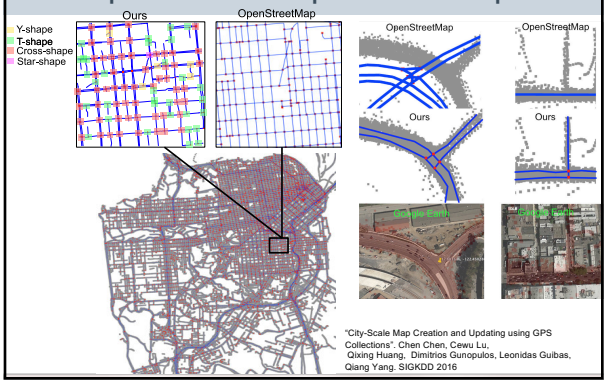


Combine road segments into a network

The basic problem: identify junctions (where segments combine)
The main idea: use prior knowledge of how roads intersect



Comparison with OpenStreetMap



"City-Scale Map Creation and Updating using GPS Collections". Chen Chen, Cewu Lu, Qixing Huang, Dimitrios Gunopulos, Leonidas Guibas, Qiang Yang. SIGKDD 2016

Problems

- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

Accelerometer-based transportation mode detection on smartphones

Segment-based features

- The segment-based features we consider are the frequency of acceleration and breaking periods, the frequency and duration of the intermittent stationary periods, and the variance of individual peak-based features.

Domain	Features
Segment	Variance of peak features (10 features), Peak frequency (2 features), Stationary duration, Stationery frequency

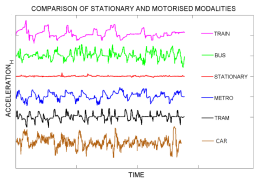


Figure 4: Comparison of the horizontal accelera-

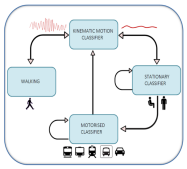


Figure 1: Overview of the classifiers used in our system and their dependencies.

[HNT13]

Using mobile phones to determine transportation mode

Classifier able to identify the transportation mode of an individual using:
- the phones' GPS receiver sensor
- accelerometer sensor

The **target classes**:
- {stationary, walking, running, biking, motorized transport}.

Features Used:
- GPS speed
- Accelerometer Variance
- Accelerometer DFT(3Hz)
- Accelerometer DFT(2Hz)
- Accelerometer DFT(1Hz)

Classifies used
- C4.5 Decision Trees
- **two stage system with a Decision Tree combined with an HMM**

Checked the accuracy of the model when the user had the phone at a different position (Arm, Bag, Chest, Hand, Pocket, Waist) [RMEDS10]

Understanding transportation modes based on GPS data for web applications

Three part approach:

1. A change point segmentation method

Distinguish segments as walking or not walking (based on the idea that walking should be the transmission between different transportation modes).
Distinguish segments using a threshold of velocity and acceleration

2. An inference model

Received as input extracted features:
- Heading Change Rate
- Stop Rate
- Velocity Change Rate

Used supervised classification methods
- **Decision Tree**
- Support Vector Machine
- Bayesian Net
- Conditional Random Field

3. A graph based post processing algorithm
- Collect start/end points and change points and applied a hierarchical clustering algorithm on them
- Build a Graph connecting the different clusters
- Use spatial indexing to identify quickly the clusters
- Calculate the probabilities changing from one transportation mode to another using the labeled data

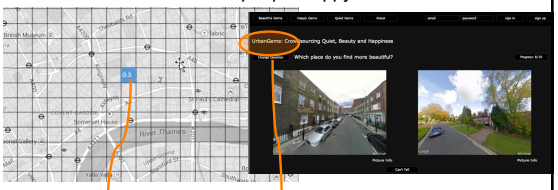
[ZCLXM10]

Problems

- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

Capturing the Aesthetic Capital

- Urban spaces are related to the ***emotional perceptions***
- Automatically generate routes that are not only ***short*** but also ***emotionally pleasant***
- Which urban elements make people happy?

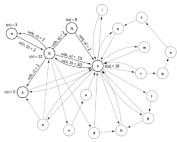


- Walkable cell of 200x200 meters
- Urbangems value for each emotional dimension

Daniele Quercia, Rossano Schifanella, Luca Maria Aiello:
The shortest path to happiness: recommending beautiful, quiet, and happy routes in the city. HT 2014: 116-125

What is the purpose of people motion

- Create an Individual Mobility Network (IMN) for each user
- Extract a set of features:
  - Network Features
    - Centrality (clustering coefficient, avg path length)
    - Predictability (entropy)
    - Hubbiness (degree betweenness)
    - Volume (edge weight, flow per location)
  - Trip Features
    - Length, duration, time interval
- Use the Activity-Based Cascade (ABC) classifier that at each step creates a binary classifier aiming to distinguish a particular class (i.e. Going home) among the others

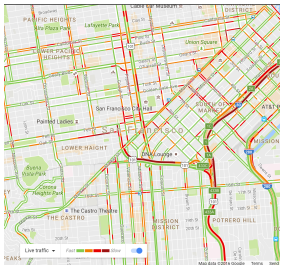


[RGNG14]

Problems

- Learning from trajectories
  - Distance measures
  - Clustering
  - Indexing, Distributed Indexing
  - Pattern Queries
  - Applications
    - Map Creation
    - Mode of Transportation
    - Social networks applications
- Crowdsourcing Applications
  - Privacy issues

Mobile Crowdsourcing Systems

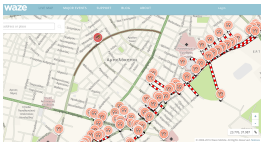


<carID, timestamp, latitude, longitude, speed >

- Recent mobile systems "sense" and collect sensor data provided by humans
  - *Google Maps, Moovit, Uber, etc.*
- Current solutions focus on centralized databases to:
  - Maintain all user sensed data
  - Preserve user privacy
  - Provide events of interest
- Bringing all data to a central location
  - Involves high cost
  - May not be necessarily beneficial
- In the distributed setting each user
  - Maintains an individual local database that contains frequently-sampled sensed data
  - Issues queries on data stored across the system to retrieve sensed data and identify events of interest

Athens Authentic Marathon

- Waze Crowdsourcing App
  - Allows users to report traffic incidents
- Athens Authentic Marathon
  - Identified closed and congested roads
  - Users can use alternative routes
  - Traffic Events update in real-time



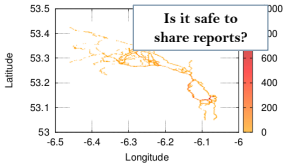
13-11-2016 12:00

13-11-2016 13:00

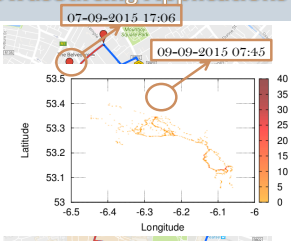
Challenges in Mobile Crowdsourcing

- **The selection of the most appropriate users to perform a task is fundamental to the quality of the result**
  - Human users have different abilities and characteristics
    - Objectiveness, Perspective, Response Delays , Location
    - Spammers perform tasks randomly
    - Users might be unavailable
- **User privacy should be protected when users contribute to crowdsourcing systems**
  - Mobile Crowdsourcing assumes that the shared data are coupled with the user mobility traces
  - Sophisticated attacks can expose the user identity, her sensitive locations and the places that the user has visited

Privacy Exposure in Crowdsourcing Applications



Complete set of mobility traces



User Shared Data

Exploiting public information can lead to leak of private information
Tracking: Identify the user's mobility patterns
Identification: Isolate her frequently visited locations
Profiling: Identify places that could profile the user

Privacy Preservation Problem

How to design a system that will allow users to use crowdsourcing apps and share geo-located content without revealing sensitive information?

- **Design Objectives**
  - Users post data coupled with their identifiers, location and timestamp
  - Users tune their privacy levels
  - User mobility should be kept locally
- **Goals**
  - Evaluate the user privacy exposure on resource constrained devices
  - Analyze large amounts of spatiotemporal data in real-time
  - Identify critical points that may expose user privacy and alert the user
- **Privacy Definition:** The set of shared data R_i preserves user privacy when
  - An attacker cannot generate an approximation of any user trajectory from T_i
  - The user sensitive locations cannot be determined from the shared data R_i based on their amount and frequency

Deal with attacks on Android OS

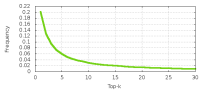
- Each application in Android can be composed of the following **Components**
  - **Activities, Services, Content Providers, Broadcast Receivers**
- Components are invoked and communicate through messages called **Intents**

Action	Data	Type	Category	Component	Extra
ACTION_DIAL	tel:*+302108203911*	-	-	SERVICE	-

- Intent messages can be:
  - **Explicit**, when the naming of the target component is provided
  - **Implicit,** when the Android runtime resolves the target component by matching the message against the components that can handle it
- **Service Selection attack:** Trigger a component that does not belong to the scope of the application
  - Solution:
    - Use Explicit Intents and Interception to investigate the target component
    - Examine Intents at the application level to reduce the overhead
- **Query selection attack**
  - A malformed query can cause exceptions and application or device misbehavior
  - The attacker execute well-formed queries in order to corrupt the database
  - Solution:
    - Investigate the SQL command for each query before instantiating the Intent
    - Develop regex for legimate commands, to ensure that they have a valid syntax

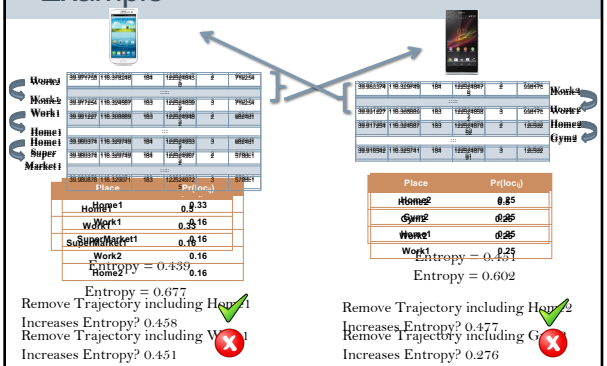
Deal with attacks on user trajectories

- Problem
  - Local data contain a lot of similar trajectories
  - An attacker can identify the user reveal frequent locations or track the user
- Each smartphone periodically
  - Defines the k most frequent locations loc_{ij}



  - We choose m trajectories that contain m from the k (m<k) most frequent locations
    - Trajectory is selected when removing the trajectory increases the Entropy
    - Measures the quality of trajectories with respect to the anonymity it provides
  - We select n trajectories that do not contain frequent locations
  - These trajectories can be forwarded to a specific neighbor or to different neighbors
  - **Both types of trajectories may contain sensitive data** (e.g. political meetings)

Example



Remove Trajectory including Home1
Increases Entropy? 0.458
Remove Trajectory including Work1
Increases Entropy? 0.451

Remove Trajectory including Home2
Increases Entropy? 0.477
Remove Trajectory including Gym2
Increases Entropy? 0.276

Trajectory Anonymizing

- Given a database of trajectories representing user mobility for a number of users, generate an anonymized version of this database such that :
  - The identities of all individuals are protected in the anonymized database from adversaries who have certain knowledge at their disposal
  - The effect of sanitization to the original database is minimal, i.e. the original database is minimally distorted to produce its anonymized counterpart so that utility is maximized

Algorithms for trajectory anonymization

- Several algorithms have been proposed for anonymizing a database of user trajectories
- Existing approaches can be categorized as follows:
  - Approaches that simplify user trajectories into sequences of POIs → spatiotemporal data is simplified to sequential data (*suppression approaches*)
  - Approaches that take into consideration the spatiotemporal nature of user trajectories when anonymizing the database (*clustering & perturbation*)
- The assumed knowledge of the attackers also varies from approach to approach

PLS: Protecting privacy when publishing location sequences (Terrovitis et al 08)

- User trajectories are simplified into sequences of POIs (a1, b1, …)
- Different adversaries own different parts of the POIs (e.g., all "a" can denote branches of the same bank, where the adversary is the bank owner)
- The portions of trajectories can be used for a linkage attack to reveal user identity (e.g., only user 5 goes through a1 and a3 so attacker "a" easily learns the identity of this user as well as that he also went to b1)

id	trajectory
t_1	$a_1 \to b_1 \to a_2$
t_2	$a_1 \to b_1 \to a_2 \to b_3$
t_3	$a_1 \to b_2 \to a_2$
t_4	$a_1 \to a_2 \to b_2$
t_5	$a_1 \to a_3 \to b_1$
t_6	$a_3 \to b_1$
t_7	$a_3 \to b_2$
t_8	$a_3 \to b_2 \to b_3$

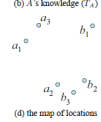
(a) exact data (T)

id	trajectory
t_1^A	$a_1 \to a_2$
t_2^A	$a_1 \to a_2$
t_3^A	$a_1 \to a_2$
t_4^A	$a_1 \to a_2$
t_5^A	$a_1 \to a_3$
t_6^A	a_3
t_7^A	a_3
t_8^A	a_3

(b) A's knowledge (T_A)

id	trajectory
t'_1	$a_1 \to b_1 \to a_2$
t'_2	$a_1 \to b_1 \to a_2$
t'_3	$a_1 \to b_2 \to a_2$
t'_4	$a_1 \to a_2 \to b_2$
t'_5	$a_3 \to b_1$
t'_6	$a_3 \to b_1$
t'_7	$a_3 \to b_2$
t'_8	$a_3 \to b_2$

(c) transformed database (T')

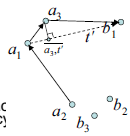


(d) the map of locations

[TM08]

PLS: Anonymization strategy

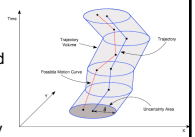
- Identify possible linkage attacks (threats) for each attacker
  - Consider the projection of the original database with respect to a symbol: "a", "b", … This corresponds to an attacker's knowledge
  - Identify the probability of the attacker to successfully reconstruct a POI (other than the ones he knows) in eac sequence. When this is above P_{br} it constitutes a privacy threat
- Suppress location information (POIs) whenever a possible privacy leak is found to block the threat
- Maximize the utility of the published data by measuring the similarity of the original trajectories to their corresponding transformations



t = a2 → a1 → a3 → b1
t' = a1 → b1

NWA: Protecting privacy when publishing complete user trajectories (Abul et al 08)

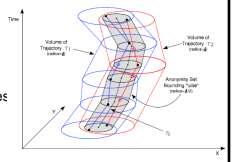
- User trajectories are represented as cylindrical volumes (of radius δ) that capture the location imprecision due to sampling and positioning systems
- All trajectories are assumed to have the same sampling time, to move along a straight line and at a constant speed between two consecutive observations
- Two objects that move within the same cylinder are followed almost the same route are considered to be indistinguishable
- The goal is to anonymize trajectories as a whole by assuming attackers that can identify each user in any location and at any time



[ABN08]

NWA: Anonymization strategy

- Co-localization: 2 trajectories defined in the same time interval are co-localized iff for each (x, y) point [same t] it holds that Euclidian distance dist((x1,y1), (x2,y2)) < δ
- (k, δ)–anonymity: based on co-localization
  - A set of at least k trajectories that are co-localized with respect to a radius δ
- Anonymization is performed in 3 steps
  - Preprocessing (enforce large equivalence classes of trajectories w.r.t. same time span)
  - Clustering of trajectories into equivalence classes of same time span
  - Space translation to transform the identified clusters into (k,δ)–anonymity sets

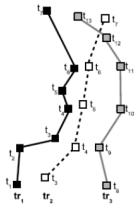


An anonymity set can be bounded by a cylindrical volume of radius δ/2

[ABN08]

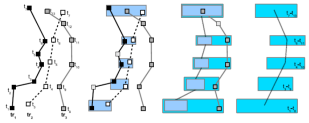
TGA: Trajectory anonymization based on grouping and reconstruction (Nergiz et al 09)

- TGA assumes that space is discretized into εxε grids and time is discretized in buckets of size τ
- User trajectories are represented as temporarily ordered sets of spatiotemporal points
- Adversaries with the following types of knowledge (and malicious intends) are assumed:
  - Already know some portion of the trajectory of an individual in the database and want to learn the rest
  - Already know the whole trajectory of an individual and want to learn some sensitive information (e.g., requests done to LBSs)



[NAS+09] Nergiz, M.E., Atzori, M., Saygin, Y., Güç, B.: Towards Trajectory Anonymization: a Generalization-Based Approach. Transactions on Data Privacy 2(1): 47-75 (2009).

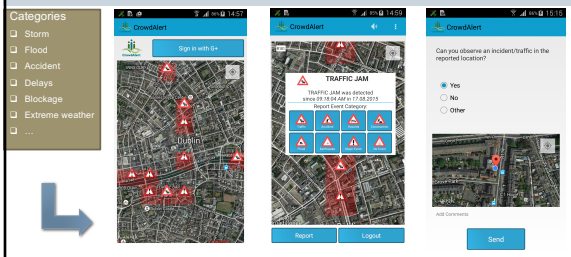
TGA: Anonymization strategy



- TGA offers trajectory anonymization by
  - Generalizing sets of k trajectories to group them together by enclosing their respective locations into MBRs
  - Representing each group of k trajectories as an anonymized trajectory
  - Releasing atomic trajectories sampled randomly from the area covered by the anonymized trajectories

[NAS+09]

CrowdAlert app deployed in Dublin City



- CrowdAlert app deployed in Dublin Smart City
- Using the mobile crowdsourcing app to reach humans users during an event:
  - Queries are generated by the Insight system using ontology
  - Implement messaging system for communicating the queries to human users and collecting responses
  - Receive responses asynchronously and extract user replies

Thanks

- VaVeL H2020 project: www.vavel-project.eu
- Collaborators: G. Trajevski, M. Vlachos, L. Guibas, V. Kalogeraki, G. Das, D. Zeinalipour-Yazti, T. Palpanas, P. Papapetrou, I. Katakis, N. Zygouras, A. Saravanou, N. Panagiotou, C. Chen, Y. Li, G. Papadopoulos, T. Lappas, G. Valkanas, D. Kotsakos

References – Trajectories; Moving Objects

- [AAE00] P. K. Agarwal, Lars Arge, and Je Erickson. Indexing Moving Points. In Proceeding *ACM PODS, 2000.*
- [ABN08] O. Abul, F. Bonchi, M. Nanni, "Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases" (ICDE 2008)
- [BHK+05] P. Bakalov, M. Hadjieleftheriou, E.J. Keogh, and V.J. Tsotras. Efficient trajectory joins using symbolic representations. In *Proceedings of Mobile Data Management (MDM'05),* 2005.
- [BGL+03] Veronica Biazzo, Rosalba Giugno, Thomas Lukasiewicz, and V. S. Subrahmanian. Temporal probabilistic object bases. *IEEE Trans. Knowl. Data Eng.*, 15(4):921–939, 2003.
- [BHK+05] P. Bakalov, M. Hadjieleftheriou, and V.J. Tsotras. "Time-Relaxed Spatiotemporal Trajectory Joins". In *Proceedings ACM GIS, 2005.*
- [C+11] Simone Campora, Jose Antonio Fernandes de Macedo, Laura Spinsanti. St-Toolkit: A Framework for Trajectory Data Warehousing. In AGILE, 2011.
- [CWT06] Hu Cao, Ouri Wolfson, and Goce Trajcevski. Spatio-temporal data reduction with deterministic error bounds. *VLDB Journal*, 15(3), 2006.
- [CKS+09] Y.-Y. Chiang, C. A. Knoblock, C. Shahabi and C.-C. Chen. "Automatic and Accurate Extraction of Road Intersections from Raster Maps". *GeoInformatica, 13(2), 2009.*
- [CKP04] R. Cheng, D.V. Kalashnikov, S. Prabhakar. "Querying Imprecise Data in Moving Object Environments". IEEE Trans. Knowl. Data Eng. 16(9), 2004.
- [DTS08] H. Ding, G. Trajcevski and P. Scheuermann, "Efficient Maintenance of Continuous Queries for Trajectories". *GeoInformatica, 12(3), 2008.*
- [DLW12] S. Dodge, P. Laube, R. Weibel. Movement Similarity Assessment Using Symbolic Representation of Trajectories. International Journal on Geographic Information Systems, 2012 (to appear).

References – Trajectories; Moving Objects

- [FPN+08] E. Frentzos, N. Pelekis, I. Ntoutsi, and Y. Theodoridis, "Trajectory Database Systems". *In Mobility, Data Mining and Privacy (F. Giannotti and D. Pedreschi – eds.), Springer, 2008.*
- [GL06] B. Gedik and L. Liu. "MobiEyes: A Distributed Location Monitoring Services using Moving Location Queries". In *Proceedings EDBT, 2006.*
- [GvK06] J. Gudmundsson and M. van Kreveld. "Computing Longest Duration Flocks in Trajectory Data". In *Proceedings ACM GIS, 2006.*
- [GBE+01] R.H. Guting, M.H. Bohlen, M. Erwig, C.S. Jensen, N.A. Lorentzos, M. Schneider, and M. Vazirgiannis. "A foundation for representing and querying moving objects". *ACM Transactions Database Systems, 25(1)m 2000.*
- [GBA+04] R. Guting, T. Behr, V. Almeida, Z. Ding, F. Hoffmann, and M. Spiekermann. "SECONDO: An extensible dbms architecture and prototype". *Fernuniversitat Hagen, Informatik-report 313, 2004.*
- [GAD06] Ralf Hartmut Güting, V.T. deAlmeida and Z. Ding: "Modeling and querying moving objects in networks". *VLDB J., 15(2), (2006)*
- [GS05] R.H. Guting and M. Schneider. "Moving Objects Databases". *Morgan-Kauffman, 2005.*
- [ISS06] G.S. Iwerks, H. Samet and K.P. Smith. "Maintenance of K-nn and Spatial Join Queries on Moving Points". *ACM TODS, 31(2), 2006.*
- [HKB+05] M. Hadjieleftheriou, G. Kollios, P. Bakalov and V. Tsotras. "Complex Spatio Temporal Pattern Queries". In *Proceedings VLDB, 2005.*
- [HE02] K. Hornsby, M.J. Egenhofer "Modeling Moving Objects over Multiple Granularities". *Ann. Math. Artif. Intell. 36(1-2), 2002*

References – Trajectories; Moving Objects

- [Jensen04] C. S. Jensen. Database Aspects of Location-based Services. In Location-based Services, pages 115-148. Morgan Kaufmann, 2004
- [JLO04] C. S. Jensen, Dan Lin, and Beng Chin Ooi. "Query and Update Efficient B+-Tree Based Indexing of Moving Objects". In *Proceeding of the International Conference on Very Large Data Bases, VLDB, 2004.*
- [JLX+08] H.Jeung , M.L. Yiu, X. Zhou, C.S. Jensen and H.T. Shen. "Discovery of Convoys in Trajectory Databases". In *Proceedings VLDB, 2008.*
- [KMB05] P. Kalnis, N. Mamoulis, and S. Bakiras." On discovering moving clusters in spatio-temporal data". In *Proceedings SSTD, 2005.*
- [KPG+05] G. Kollios, D. Papadopoulos, D. Gunopulos, and Vassilis J. Tsotras. Indexing Mobile Objects Using Dual Transformations. *VLDB Journal, 14(2), April 2005*
- [KO09] B. Kuijpers and W. Othman. "Trajectory databases: data models, uncertainty and complete query languages". *Journal of Computer and System Sciences, 2009. doi:10.1016/j.jcss.2009.10.002*
- [KPG00] B. Kuijpers, J. Paredaens, and D.V. Gucht. Towards a theory of moving database queries. In *Proceedings of the 7th International Workshop on Temporal Representation and Reasoning (TIME), 2000.*
- [KMN+10] Bart Kuijpers, Harvey J. Miller, Tijs Neutens, and Walied Othman. Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science, 24(8):1223–1248, 2010.*
- [KMO+09] Kuijpers, B., Moelans, B., Othman, W., Vaisman, A.A.: Analyzing Trajectories Using Uncertainty and Background Information. *In Proceedings SSTD 2009.*
- [LJZ11] Z. Lei, L. Jing and W. Zhichao. Uncertain Trajectory Clustering based on Uneven Two-Step Sampling. *International Journal on Convergence Information Technology, 6(5), 2011.*

References – Trajectories; Moving Objects

- [LWG+09] Ralph Lange, HaraldWeinschrott, Lars Geiger, Andr´e Blessing, Frank D¨urr, Kurt Rothermel, and Hinrich Sch¨utze. On a generic uncertainty model for position information. In *QuaCon*, pages 76–87, 2009.
- [LvKI04] P. Laube, M. van Kreveld and S. Imfeld. Finding REMO – detecting relative motion patterns in geospatial lifelines. In Proc. *11th International Symp. on Spatial Data Handling, 2004.*
- [LHL+08] J.-G. Lee, J. Han, X. Li and H. Gonzales. "TraClass: Trajectory classification using hierarchical region-based and trajectory-based clustering". In *Proceedings VLDB, 2008.*
- [MVO+08] J. Macedo, C. Vangenot,W. Othman, N. Pelekis, E. Frentzos, B. Kuijpers, I.Ntoutsi, S. Spaccapietra, and Y. Theodoridis. "Trajectory Data Models". *In Mobility, Data Mining and Privacy (F. Giannotti and D. Pedreschi – eds.), Springer, 2008.*
- [MA05] Mohamed F. Mokbel and Walid G. Aref. PLACE: A Scalable Location-aware Database Server for Spatiotemporal Data Streams. *IEEE Data Engineering Bulletin, 28(3),2005.*
- [MA08] Mohamed F. Mokbel, Walid G. Aref. "SOLE: scalable on-line execution of continuous queries on spatio-temporal data streams. VLDB J. 17(5), 2008.
- [MXH+05] Mohamed F. Mokbel, Xiaopeng Xiong, Moustafa A. Hammad, and Walid G. Aref. Continuous Query Processing of Spatio-temporal Data Streams in PLACE. *GeoInformatica, 9(4),2005.*
- [NAS+09]] Nergiz, M.E., Atzori, M., Saygin, Y., Güç, B.: Towards Trajectory Anonymization: a Generalization-Based Approach. Transactions on Data Privacy 2(1): 47-75 (2009).
- [PHT+08] Jian Pei, Ming Hua, Yufei Tao, and Xuemin Lin. Query answering techniques on uncertain and probabilistic data: tutorial summary. In *ACM SIGMOD*, 2008.
- [PJ00] Dieter Pfoser and Christian S. Jensen. Trajectory Indexing Using Movement Constraints. *GeoInformatica, 9(2), 2005.*
- [PJ99] Dieter Pfoser, Christian S. Jensen. "Capturing the Uncertainty of Moving-Object Representations". In *SSTD, 1999.*

References – Trajectories; Moving Objects

- *[RCK+10]* Matthias Renz, Reynold Cheng, Hans-Peter Kriegel, Andreas Z¨ufle, and Thomas Bernecker. Similarity search and mining in uncertain databases (tutorial). *PVLDB*, 3(2):1653–1654, 2010.
- [Roberts73] F.S. Roberts. Tolerance geometry. Notre Dame Journal of Formal Logic, 14(1):68–76, 1973
- [SJL+00] Simonas Saltenis, Christian S. Jensen, Scott T. Leutenegger, and Mario A. Lopez. Indexing the Positions of Continuously Moving Objects. In *Proceeding of the ACM International Conference on Management of Data, SIGMOD, 2000*.
- [SWC+97] A.P. Sistla, O.Wolfson, S. Chamberlain, and S. Dao.Modeling and Querying Moving Objects. In *Proceedings of the 13th International Conference on Data Engineering (ICDE)*,1997
- [SC08] Mehdi Sharifzadeh, Cyrus Shahabi. "Processing Optimal Sequenced Route Queries Using Voronoi Diagrams. GeoInformatica 124), 2008.
- [SNT+06] M. Spiliopoulou, I. Ntoutsi, Y. Theodoridis, and R. Schult. Monic: modeling and monitoring cluster transitions. In Proceedings *SIGKDD* 2006
- [SXI01] J. Su, H. Xu, and O.H. Ibarra. Moving Objects: Logical Relationships and Queries., *Proceedings of 7th International Symposium on Advances in Spatial and Temporal Databases (SSTD), 2001*
- [TPS03] Y. Tao, D. Papadias, and J. Sun. The TPR*-Tree: An Optimized Spatio-temporal Access Method for Predictive Queries. In *Proceeding of the International Conference on Very Large Data Bases, VLDB, 2003*.
- [TXC07] Yufei Tao, Xiaokui Xiao, and Reynold Cheng. Range search on multidimensional uncertain data. *ACM Trans. Database Syst.*, 32(3):15, 2007.
- [TM08] Terrovitis, M., Mamoulis, N.: Privacy Preservation in the Publication of Trajectories. MDM 2008
- [TCW+10] G. Trajcevski, A. Choudhary, O. Wolfson, Y. Li and G. Li. Uncertain Range Queries for Necklaces. *In Mobile Data Management, 2010.*

References – Trajectories; Moving Objects

- [TWH+04] G. Trajcevski, O. Wolfson, K. Hinrichs and S. Chamberlain. "Managing Moving Objects Databases with Uncertainty". *ACM TODS, 29(3), 2004*.
- [TDST07] G. Trajcevski, H. Ding, P. Scheuermann and R. Tamassia. "Dynamics Aware Similarity of Moving Objects Trajectories". In *Proceedings ACM GIS, 2007*.
- [TCS+06] Goce Trajcevski, Hu Cao, Peter Scheuermann, Ouri Wolfson, and Dennis Vaccaro. On-line data reduction and the quality of history in moving objects databases. In *MobiDE*, 2006.
- [TTC+11] Goce Trajcevski, Roberto Tamassia, Isabel Cruz, Peter Scheuermann, David Hartglass, and Christopher Zamierowski. Ranking continuous nearest neighbors for uncertain trajectories. *VLDB Journal*, 2011.
- [TPP03] N. Tryfona, R. Price, and C. Price. *Spatiotemporal Conceptual Modeling., chapter 3*, Berlin, 2003.
- [V10] A. Vaisman. Trajectory Warehousing and Mining: a state of the art and Research Directions. In MDM 2010.
- [VLR+08]] Michail Vlachos, Claudio Lucchese, Deepak Rajan, Philip S. Yu: Ownership protection of shape datasets with geodesic distance preservation. EDBT 2008
- [WF10] Gwen Wilke and Andrew F. Frank. Tolerance geometry – euclid's first postulate for points and lines with extension. In *Proceedings of the 18th ACM SIGSPATIAL GIS Conference*, pages 162–171, 2010.
- [WY10] Stephan Winter and Zhang-Cai Yin. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science*, 24(9), 2010.
- [WSC+99] O. Wolfson, A.P. Sistla, S. Chamberlain, and Y. Yesha. Updating and querying databases that track mobile units. *Distributed and Parallel Databases, 7(3), 1999.*

References – Trajectories; Moving Objects

- Zheng, K., Zheng, Y., Yuan, N. J., Shang, S., & Zhou, X. (2014). Online discovery of gathering patterns over trajectories. IEEE Transactions on Knowledge and Data Engineering, 26(8), 1974-1988.
- Ge, Y., Xiong, H., Zhou, Z. H., Ozdemir, H., Yu, J., & Lee, K. C. (2010, October). Top-eye: Top-k evolving trajectory outlier detection. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 1733-1736). ACM.
- Pelekis, N., Kopanakis, I., Panagiotakis, C., & Theodoridis, Y. (2010, September). Unsupervised trajectory sampling. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 17-33). Springer Berlin Heidelberg.
- Banerjee, P., Ranu, S., & Raghavan, S. (2014, December). Inferring uncertain trajectories from partial observations. In 2014 IEEE International Conference on Data Mining (pp. 30-39). IEEE.
- Su, H., Zheng, K., Huang, J., Wang, H., & Zhou, X. (2015). Calibrating trajectory data for spatio-temporal similarity analysis. The VLDB Journal, 24(1), 93-116.
- Li, Y., Chow, C. Y., Deng, K., Yuan, M., Zeng, J., Zhang, J. D., ... & Zhang, Z. L. (2015, October). Sampling Big Trajectory Data. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 941-950). ACM.
- Chen, Z., Shen, H. T., Zhou, X., Zheng, Y., & Xie, X. (2010, June). Searching trajectories by locations: an efficiency study. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 255-266). ACM.
- Zheng, K., Trajcevski, G., Zhou, X., & Scheuermann, P. (2011, March). Probabilistic range queries for uncertain trajectories on road networks. In Proceedings of the 14th International Conference on Extending Database Technology (pp. 283-294). ACM.
- Niedermayer, J., Züfle, A., Emrich, T., Renz, M., Mamoulis, N., Chen, L., & Kriegel, H. P. (2013). Probabilistic nearest neighbor queries on uncertain moving object trajectories. Proceedings of the VLDB Endowment, 7(3), 205-216.
- Vieira, M. R., Bakalov, P., & Tsotras, V. J. (2010, March). Querying trajectories using flexible patterns. In Proceedings of the 13th International Conference on Extending Database Technology (pp. 406-417). ACM.

References – Trajectories; Moving Objects

- Lee, J. G., Han, J., & Whang, K. Y. (2007, June). Trajectory clustering: a partition-and-group framework. In Proceedings of the 2007 ACM SIGMOD international conference on Management of data (pp. 593-604). ACM.
- Zhang, C., Han, J., Shou, L., Lu, J., & La Porta, T. (2014). Splitter: Mining fine-grained sequential patterns in semantic trajectories. Proceedings of the VLDB Endowment, 7(9), 769-780.
- Li, Z., Han, J., Ding, B., & Kays, R. (2012). Mining periodic behaviors of object movements for animal and biological sustainability studies. Data Mining and Knowledge Discovery, 24(2), 355-386.
- Liu, S., Wang, S., Jayarajah, K., Misra, A., & Krishnan, R. (2013, October). TODMIS: Mining communities from trajectories. In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management (pp. 2109-2118). ACM.
- Patel, D., Sheng, C., Hsu, W., & Lee, M. L. (2012, April). Incorporating duration information for trajectory classification. In 2012 IEEE 28th International Conference on Data Engineering (pp. 1132-1143). IEEE.
- Au, T. S., Duan, R., Kim, H., & Ma, G. Q. (2010, December). Spatiotemporal event detection in mobility network. In 2010 IEEE International Conference on Data Mining (pp. 28-37). IEEE.
- Dai, J., Yang, B., Guo, C., & Ding, Z. (2015, April). Personalized route recommendation using big trajectory data. In 2015 IEEE 31st International Conference on Data Engineering (pp. 543-554). IEEE.
- Xue, A. Y., Zhang, R., Zheng, Y., Xie, X., Huang, J., & Xu, Z. (2013, April). Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In Data Engineering (ICDE), 2013 IEEE 29th International Conference on (pp. 254-265). IEEE.
- Renso, C., Baglioni, M., de Macedo, J. A. F., Trasarti, R., & Wachowicz, M. (2013). How you move reveals who you are: understanding human behavior by analyzing trajectory data. Knowledge and information systems, 37(2), 331-362.
- Zheng, K., Zheng, Y., Yuan, N. J., & Shang, S. (2013, April). On discovery of gathering patterns from trajectories. In Data Engineering (ICDE), 2013 IEEE 29th International Conference on (pp. 242-253). IEEE.

References – Trajectories; Moving Objects

- Chen Chen, Cewu Lu, Qixing Huang, Qiang Yang, Dimitrios Gunopulos, Leonidas J. Guibas: City-Scale Map Creation and Updating using GPS Collections. KDD 2016: 1465-1474
- Knowledge-Based Trajectory Completion from Sparse GPS Samples Yang Li, Yangyan Li, Dimitrios Gunopulos and Leonidas Guibas, Knowledge-Based Trajectory Completion from Sparse GPS Samples", ACM GIS 2016
- Liu, X., Biagioni, J., Eriksson, J., Wang, Y., Forman, G., & Zhu, Y. (2012, August). Mining large-scale, sparse GPS traces for map inference: comparison of approaches. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 669-677). ACM.
- Biagioni, J., & Eriksson, J. (2012, November). Map inference in the face of noise and disparity. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems (pp. 79-88). ACM.
- Davies, J. J., Beresford, A. R., & Hopper, A. (2006). Scalable, distributed, real-time map generation. IEEE Pervasive Computing, 5(4), 47-54.
- Cao, L., & Krumm, J. (2009, November). From GPS traces to a routable road map. In Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems (pp. 3-12). ACM.
- Chen, D., Guibas, L. J., Hershberger, J., & Sun, J. (2010, January). Road network reconstruction for organizing paths. In Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms (pp. 1309-1320). Society for Industrial and Applied Mathematics.
- Karagiorgou, S., & Pfoser, D. (2012, November). On vehicle tracking data-based road network generation. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems (pp. 89-98). ACM.
- Su, H., Zheng, K., Zeng, K., Huang, J., Sadiq, S, Yuan, N. J., & Zhou, X. (2015, April). Making sense of trajectory data: A partition-and-summarization approach. In 2015 IEEE 31st International Conference on Data Engineering (pp. 963-974). IEEE.

References – Mode of Transportation

- [HNT13] Hemminki, S., Nurmi, P., & Tarkoma, S. (2013, November). Accelerometer-based transportation mode detection on smartphones. In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (p. 13). ACM.
- [ZCLXM10] Zheng, Y., Chen, Y., Li, Q., Xie, X., & Ma, W. Y. (2010). Understanding transportation modes based on GPS data for web applications. ACM Transactions on the Web (TWEB), 4(1), 1.
- [RMEDS10] Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2010). Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks (TOSN), 6(2), 13.

References

- [PS04] Patroumpas, K., & Sellis, T. K. (2004, August). Managing Trajectories of Moving Objects as Data Streams. In STDBM (pp. 41-48).
- [PS06] Patroumpas, K., & Sellis, T. (2006, March). Window specification over data streams. In International Conference on Extending Database Technology (pp. 445-464). Springer Berlin Heidelberg.
- [PS07] Patroumpas, K., & Sellis, T. (2007, May). Semantics of spatially-aware windows over streaming moving objects. In 2007 International Conference on Mobile Data Management (pp. 52-59). IEEE.
- [PPS06] Potamias, M., Patroumpas, K., & Sellis, T. (2006, July). Sampling trajectory streams with spatiotemporal criteria. In 18th International Conference on Scientific and Statistical Database Management (SSDBM'06) (pp. 275-284). IEEE.
- [PKS08] Patroumpas, K., Kefallinou, E., & Sellis, T. (2008, November). Monitoring continuous queries over streaming locations. In Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems (p. 81). ACM.
- [GNPP07] Giannotti, F., Nanni, M., Pinelli, F., & Pedreschi, D. (2007, August). Trajectory pattern mining. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 330-339). ACM.
- [MPTG09] Monreale, A., Pinelli, F., Trasarti, R., & Giannotti, F. (2009, June). Wherenext: a location predictor on trajectory pattern mining. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 637-646). ACM.
- [AARN[G09] Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., & Giannotti, F. (2009, October). Interactive visual clustering of large collections of trajectories. In Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on (pp. 3-10). IEEE.
- [RGN[[G14] Rinzivillo, S., Gabrielli, L., Nanni, M., Pappalardo, L., Pedreschi, D., & Giannotti, F. (2014, October). The purpose of motion: Learning activities from individual mobility networks. In Data Science and Advanced Analytics (DSAA), 2014 International Conference on (pp. 312-318). IEEE.
- [PROPG13] Pappalardo, L., Rinzivillo, S., Qu, Z., Pedreschi, D., & Giannotti, F. (2013). Understanding the patterns of car travel. The European Physical Journal Special Topics, 215(1), 61-73.

References – Application Domains

- [AntoniniT06] G. Antonini and J.P. Thiran. "Counting of Pedestrians in Video Sequences Using Trajectory Clustering". IEEE Transactions on Circuits and Systems for Video Technology, 16(8), 2006.
- [BBM+10] R.A. Boller, S.A. Braun, J. Miles and D.H. Laidlaw. Application of uncertainty visualization methods to meteorological trajectories. Earth Sci. Inform. 3, 2010.
- [DodgeWF09] S. Dodge, R. Weibel and E. Foroothan. "Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects".
- [FanXW09] Jialue Fan, Jiang Xu and Ying Wu. "Context-aware Tracking of Small Targets in Video". SPIE Symp. On Optical Engineering and Applications, 2009.
- [Hamacher09] K. Hamacher. "Information Theoretical Measures to Analyze Trajectories in Rational Molecular Design". J Comput Chem 28, 2007
- [JiangWK07] F. Jiang, Y.Wu, and A. K. Katsaggelos, "Abnormal event detection from surveillance video by dynamic hierarchical clustering," in Proc. IEEE Int'l Conf. on Image Processing, 2007.

References – Application Domains

- [KeunEB+04] H.C. Keun, T. M. D. Ebbels, M. E. Bollard, O. Beckonert, H. Antti, E. Holmes, J. C. Lindon and J. K. Nicholson. "Geometric Trajectory Analysis of Metabolic Responses To Toxicity Can Define Treatment Specific Profiles". Chemical Research in Toxicology, 17(5), 2004.
- [LiZP07] C.Li, S. Q. Zheng and B. Prabhakaran. "Segmentation and Recognition of Motion Streams by Similarity Search". ACM Trans. on Multimedia Computing, Communications and Applications, 3(3), 2007.
- [MosigWN+09] A. Mosig, S. Jäger, C. Wang, S. Nath3, I. Ersoy, K.-.p Palaniappan and S.-S. Chen1 "Tracking cells in Life Cell Imaging videos using topological Alignments". Algorithms for Molecular Biology 2009, 4:10
- [TangLK+08] J. K. T. Tang, H. Leung, T. Komura and H. P. H. Shum. "Emulating Human Perception of Motion Similarity". Computer Animation and Virtual Worlds, 19, 2008.
- [WuL08] S. Wu and Y.F. Li "On Signature Invariants for Effective Motion Trajectory Recognition". The International Journal of Robotics Research, 27(8), 2008.

References – Similarity of Spatial Data/Similarity and Distance Functions

- [AAR94] H. Alt, O. Aicholzer and G. Rote. "Matching Shapes with a Reference Point". Symp. Computational Geometry, 1994.
- [AG95] H. Alt and D. Godau. "Computing Frechet Distance between Two Polygonal Curves". International Journal on Computational Geometry and Applications". 5, 1995.
- [AltG96] H. Alt and L. Guibas. "Discrete Geometric Shapes: Matching, Interpolation and Approximation – a Survey", 1996.
- [ABB91] H. Alt, B. Behrends and J. Blomer. "Approximate Matching of Polygonal Shapes". Symp. On Discrete Computational Geometry, 1991.
- [AltMW+88] H. Alt, K. Mehlhorn, H. Waegner and E. Welzl. "Congruence, Similarity and Symmetries of Geometric Objects", Discrete Comput. Geom, 3, 1988.
- [AK98] F. Aurenhamer and R. Klein. "Voronoi Diagrams (Survey)" Tech. Report 1-98, Univ. Vienna, 1998
- [ArkinKM92] E.M. Arkin, K. Kedem, J. B. Mitchell, J. Sprinzak and M. Werman. "Matching Points into Pairwise-Disjoint Noise Regions". ORSA Journal Comput. 4(4), 1992.
- [AKW01] H. Alt, C. Knauer and C. Wenk. "Matching Polygonal Curves with Respect to Frechet Distance". STACS, 2001.
- [AST99] P. K. Agarwal, M. Sharir and S. Toledo, "Applications of parametric searching in geometric optimization". Journal of Algorithms 17, 1994.
- [Atk87] M.D. Atkinson. "An Optimal Algorithm for Geometrical Congruence". J. Algorithms, 8, 1987.

References – Similarity of Spatial Data/Similarity and Distance Functions

- [BBW08] K. Buchin, M. Buchin and C. Wenk. "Computing the Frechet Distance between Simple Polygons". Computational Geometry, 41(1-2, 2008.
- [CohenG97] S.C. Cohen and L. Guibas. "Partial Matching Under Similarity Transformations". SODA, 1997.
- [CG97] S. Cohen and L. Guibas. "Partial Matching of Planar Polylines under Similarity Transformations". SODA, 1997.
- [CGH+93] P. Chew, M. Goodrich, D. Huttenlocher, K. Kedem, J. Kleinberg and D. Kravets. "Geometric Pattern Matching under Euclidian Motions", Canadian Conf. on Computational Geometry, 1993.
- [EI96] A. Efrat and E. Itati. "Improvements on Bottleneck Matching and Related Problems Using Geometry". Symp. Computational Geometry, 1996.
- [EIV04] A. Efrat, P. Indyk and S. Venkatasubramanian. "Pattern Matching for Sets of Segments". Algorithmica, 40(3), 2004
- [GD04] K. Grauman and T. Darrel. "Fast Contour Matching Using Approximate Earth Mover's Disatnce", Proceedings ICVRP, 2004.
- [GMO99] M.T. Goodrich, J.S.B. Mitchell and W. Orletsky. "Approximate Geometric Pattern Matching Under Rigid Motions". IEEE Trans. On Pattern Recognition and Machine Intelligence, 21(4), 1999.
- [Hausdorff18] F. Hausdorff. "Dimension und äußeres Maß ". Matematiche Annalen, 1918

References – Similarity of Spatial Data/Similarity and Distance Functions

- [HKK92] D.P. Huttenlocher, K. Kedem and J.M. Kleinberg. "On Dynamic Voronoi Diagrams and the Minimum Hausdorff Distance for Point Sets Under Euclidean Motion in the Plane", Symp. Comput. Geometry, 1992.
- [HKS93] D.P. Huttenlocher, K. Kedem and M. Sharir. "The Upper Envelope of Voronoi Surfaces and its Applications" . Discrete Comput. Geometry, 9, 1993.
- [ImaiSI98] K. Imai, S. Sumino and H. Imai. "Minimax Geometric Fitting of Two Corresponding Sets of Point and Dynamic Furthest Voronoi Diagrams". Proc. IEICE Trans. In. & Syst. E81(11), 1988.
- [IV03] P. Indyk and S. Venkatasubramanian. "Approximate congruence in nearly linear time". Computational Geometry, 24(2), 2003.
- [Kocurek07] M. Kocurek. Car Shape Recognition, MS Thesis, 2007.
- [RTG98] Yossi Rubner; Carlo Tomasi, Lonidas J. Guibas ."A Metric for Distributions with Applications to Image Databases". Proceedings ICCV 1998.

References

- NWA08] O. Abul, F. Bonchi, M. Nanni, "Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases" (ICDE 2008)
- [AA10] Andrienko, N., Andrienko, G.: Spatial Generalization and Aggregation of Massive Movement Data. IEEE Transactions on Visualization and Computer Graphics (TVCG), 2010.
- [AAR+09] Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., Giannotti, F.: Interactive Visual Clustering of Large Collections of Trajectories. IEEE Visual Analytics Science and Technology (VAST 2009), Proceedings, IEEE Computer Society Press, 2009, pp.3-10.
- [AAW07] Andrienko, G., Andrienko, N., Wrobel, S.: Visual Analytics Tools for Analysis of Movement Data. ACM SIGKDD Explorations, 2007, v.9 (2), pp.38-46.
- [ABK+07] Alvares, L.O., Bogorny, V., Kuijpers, B., Macêdo, J.A.F., Moelans, B., Vaisman, A.A.: A model for enriching trajectories with semantic geographical information. GIS 2007: 22.
- [ABM+07] Alvares, L.O., Bogorny, V. Macêdo, J.A.F., Moelans, B., Spaccapietra, S.: Dynamic Modeling of Trajectory Patterns using Data Mining and Reverse Engineering. ER (Tutorials, Posters, Panels & Industrial Contributions) 2007: 149-154.
- [AFS93] Agrawal, R., Faloutsos, C., Swami, A.: Efficient similarity search in sequence databases. In Proc. of FODO, pp. 69–84(1993).
- [AG05] Almeida, V.T., Güting, R.H.: Indexing the Trajectories of Moving Objects in Networks. GeoInformatica 9(1): 33-60 (2005).

References

- [APP+08] Athitsos, V. Papapetrou, P., Potamias, M., Kollios, G., Gunopulos, D.: Approximate embedding-based subsequence matching of time series. SIGMOD Conference 2008: 365-378.
- [AVM+06] Anagnostopoulos, A., Vlachos, M., Hadjieleftheriou, M., Keogh, E.J., Yu, P.S.: Global distance-based segmentation of trajectories. KDD 2006: 34-43.
- [BC94] Berndt, D., Clifford, J.: Using dynamic time warping to find patterns in time series. In AAAI Workshop on Knowledge Discovery in Databases, pp. 229–248(1994).
- [BCF+09] Bu, Y., Chen, L., Fu, A.W.-C., Liu, D.: Efficient anomaly monitoring over moving object trajectory streams. KDD 2009: 159-168.
- [BDG97] Bollobás, B., Das, G., Gunopulos, D., Mannila, H.: Time-series similarity problems and well-separated geometric sets. In Proc. of SCG (1997).
- [BHK+05] Bakalov, P., Hadjieleftheriou, M., Keogh, E.J., Tsotras, V.J.: Efficient trajectory joins using symbolic representations. Mobile Data Management 2005: 86-93.
- [BKT07] Bakalov, P., Keogh, E.J., Tsotras, V.J.: TS2-tree - an efficient similarity based organization for trajectory data. GIS 2007: 58.
- [BMR+09] Baglioni, M., Macêdo, J.A.F., Renso, C., Trasarti, R., Wachowicz, M.: Towards Semantic Interpretation of Movement Behavior. AGILE Conf. 2009: 271-288.
- [BYÖ97] Bozkaya, T., Yazdani, N., Özsoyoglu, Z. M.: Matching and Indexing Sequences of Different Lengths. CIKM 1997: 128-135.
- [CGS00] Cadez, I.V., Gaffney, S., Smyth, P.: A general probabilistic framework for clustering individuals and objects. KDD 2000: 140-149.

References

- [CN04] Chen, L., Ng, R.T.: On The Marriage of Lp-norms and Edit Distance. VLDB 2004: 792-803.
- [CÖO04] Chen, L., Özsu, M.T., Oria, V.: Symbolic representation and retrieval of moving object trajectories. Multimedia Information Retrieval 2004: 227-234.
- [CÖO05] Chen, L., Özsu, M.T., Oria, V.: Robust and Fast Similarity Search for Moving Object Trajectories. SIGMOD Conference 2005: 491-502.
- [DTSWK08] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang and E. Keogh. "Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures". In Proceedings VLDB, 2008.
- [FGP+07] Frentzos, E., Gratsias, K., Pelekis, N., Theodoridis, Y.: Algorithms for Nearest Neighbor Search on Moving Object Trajectories. GeoInformatica 11(2): 159-193 (2007).
- [FJM+97] Faloutsos, C., Jagadish, H.V., Mendelzon, A., Milo, T.: Signature technique for similarity-basedq ueries. In Proc. of SEQUENCES (1997).
- [FRM94] Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast subsequence matching in time series databases. In Proc. of ACM SIGMOD, (1994).
- [GK95] Goldin, D., Kanellakis, P.: On similarity queries for time-series data. In Proc. of CP'95, September(1995).

References

- [GV08] Goulalas-Divanis, A., Verykios, V.S.: A Free Terrain Model for Trajectory K-Anonymity. DEXA 2008: 49-56.
- [GVM09] Goulalas-Divanis, A., Verykios, V.S., Mokbel, M.F.: Identifying Unsafe Routes for Network-Based Trajectory Privacy. SDM 2009: 942-953.
- [GKM+09] Ghys, K., Kuijpers, B., Moelans, B., Othman, W., Vangoidsenhoven, D., Vaisman, A.A.: Map matching and uncertainty: an algorithm and real-world experiments. GIS 2009: 468-471.
- [GNP+09] Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F.: Trajectory pattern analysis for urban traffic. GIS-IWCTS 2009: 43-47.
- [JSZ07] Jeung, H., Shen, H.T., Zhou, X.,: Mining Trajectory Patterns Using Hidden Markov Models. DaWaK 2007: 470-480.
- [JSZ08] Jeung, H., Shen, H.T., Zhou, X.: Convoy Queries in Spatio-Temporal Databases. ICDE 2008: 1457-1459.
- [Keo02] Keogh, E.: Exact indexing of dynamic time warping. In Proc. of VLDB (2002).
- [KWX+06] Eamonn J. Keogh, Li Wei, Xiaopeng Xi, Sang-Hee Lee, Michail Vlachos: LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures. VLDB 2006.
- [KAP+11] Alexios Kotsifakos, Vassilis Athitsos, Panagiotis Papapetrou, Jaakko Hollmén, Dimitrios Gunopulos: Model-based search in large time series databases. PETRA 2011: 36
- [KPH+11] Alexios Kotsifakos, Panagiotis Papapetrou, Jaakko Hollmén, Dimitrios Gunopulos: A Subsequence Matching with Gaps-Range-Tolerances Framework: A Query-By-Humming Application. PVLDB 4(11): 761-771 (2011)

References

- [KPT09] Kellaris, G., Pelekis, N., Theodoridis, Y.: Trajectory Compression under Network Constraints. SSTD 2009: 392-398.
- [KPZ+04] Keogh, E.J., Palpanas, T., Zordan, V.B., Gunopulos, D., Cardle, M.: Indexing Large Human-Motion Databases. VLDB 2004: 780-791.
- [KVG08] Kollios, G., Vlachos, M., Gunopulos, D.: Trajectories, Discovering Similar. Encyclopedia of GIS 2008: 1168-1173.
- [KMO+09] Kuijpers, B., Moelans, B., Othman, W., Vaisman, A.A.: Analyzing Trajectories Using Uncertainty and Background Information. SSTD 2009: 135-152.
- [KMW06] Kuijpers, B., Moelans, B., Weghe, N.V.: Qualitative polyline similarity testing with applications to query-by-sketch, indexing and classification. GIS 2006: 11-18
- [LVR+10] Claudio Lucchese, Michail Vlachos, Deepak Rajan, Philip S. Yu: Rights protection of trajectory datasets with nearest-neighbor preservation. VLDB J. 19(4): 531-556 (2010)
- [MPT+09] Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F.: WhereNext: a location predictor on trajectory pattern mining. KDD 2009: 637-646.
- [Lev66] Levenshtein, V.: Binarycodes capable of correcting deletions, insertions, and reversals. In Soviet Physics –Doklady 10, 10, pp. 707–710 (1966).
- [LHG08] Lee, J.G., Han, J., Li, X., Gonzalez, H.: TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. PVLDB 1(1): 1081-1094 (2008).
- [LHW07] Lee, J.G., Han, J., Whang, K.-Y.: Trajectory clustering: a partition-and-group framework. SIGMOD Conference 2007: 593-604.

References

- [LKS07] Lee, W., Kara, L.B., Stahovich, T.F.: An efficient graph-based recognizer for hand-drawn symbols. Computers & Graphics 31(4): 554-567 (2007).
- [LLL+10] Li, Z., Lee, J.G., Li, X., Han, J.: Incremental Clustering for Trajectories. DASFAA (2) 2010: 32-46.
- [LVR+08] Lucchese, C., Vlachos, M., Rajan, D., Yu, P.S.: Rights Protection of Trajectory Datasets. ICDE 2008: 1349-1351..
- [MPT+10] Anna Monreale, Fabio Pinelli, Roberto Trasarti, Fosca Giannotti: Location Prediction through Trajectory Pattern Mining (Extended Abstract). SEBD 2010: 134-141
- [NAS08] Nergiz, M.E., Atzori, M., Saygin, Y.: Towards trajectory anonymization: a generalization-based approach. SPRINGL 2008: 52-61.
- [NAS+09] Nergiz, M.E., Atzori, M., Saygin, Y., Güç, B.: Towards Trajectory Anonymization: a Generalization-Based Approach. Transactions on Data Privacy 2(1): 47-75 (2009).
- [SMC05] Shou, Y., Mamoulis, N., Cheung, D.W.: Fast and Exact Warping of Time Series Using Adaptive Segmental Approximations. Machine Learning 58(2-3): 231-267 (2005).
- [SPD+08] Spaccapietra, S., Parent, C., Damiani, M.L., Macêdo, J.A.F, Porto, F., Vangenot, C.: A conceptual view on trajectories. Data Knowl. Eng. 65(1): 126-146.

References

- [STH+08] Sun, J., Tsourakakis, C.E., Hoke, E., Faloutsos, C., Rad, T.E.: Two Heads Better Than One: Pattern Discovery in Time-Evolving Multi-aspect Data. ECML/PKDD (1) 2008: 22.
- [TDS+07] Trajcevski, G., Ding, H., Scheuermann, P., Tamassia, R., Vaccaro, D.: Dynamics-aware similarity of moving objects trajectories. GIS 2007: 11.
- [NP06] Nanni, M., Pedreschi, D.: Time-focused clustering of trajectories of moving objects. J. Intell. Inf. Syst. 27(3): 267-289 (2006).
- [NTR+10] Nanni, M., Trasarti, R., Renso, C., Giannotti, F., Pedreschi, D.: Advanced knowledge discovery on movement data with the GeoPKDD system. EDBT 2010: 693-696.
- [PSF05] Spiros Papadimitriou, Jimeng Sun, Christos Faloutsos: Streaming Pattern Discovery in Multiple Time-Series. VLDB 2005: 697-708
- [PDA09] Papapetrou, P., Doliotis, P., Athitsos, V.: Towards faster activity search using embedding-based subsequence matching. PETRA 2009.
- [PAP+11] Panagiotis Papapetrou, Vassilis Athitsos, Michalis Potamias, George Kollios, Dimitrios Gunopulos: Embedding-based subsequence matching in time-series databases. ACM Trans. Database Syst. 36(3): 17 (2011)
- [PJT00] Pfoser, D., Jensen, C.S., Theodoridis, Y.: Novel Approaches in Query Processing for Moving Object Trajectories. VLDB 2000: 395-406.

References

- [PSF05] Papadimitriou, S., Sun, J., Faloutsos. C.: Streaming Pattern Discovery in Multiple Time-Series. VLDB 2005: 697-708.
- [PKK+09] Pelekis, N., Kopanakis, I., Kotsifakos, E.E, Frentzos, E., Theodoridis, Y.: Clustering Trajectories of Moving Objects in an Uncertain World. ICDM 2009: 417-427.
- [PGV+11] Nikos Pelekis, Aris Gkoulalas-Divanis, Marios Vodas, Despina Kopanaki, Yannis Theodoridis: Privacy-aware querying over sensitive trajectory data. CIKM 2011: 895-904
- [PMP09] Pensa, R.G., Monreale, A., Pinelli, F., Pedreschi, D.: Anonymous Sequences from Trajectory Data. SEBD 2009: 361-372.
- [PRT08] Pedreschi, D., Ruggieri, S., Turini, F.: Discrimination-aware data mining. KDD 2008: 560-568.
- [PWZ+00] Perng, S., Wang, H., Zhang, S., Parker, D.S.: Landmarks: a new model for similarity-based pattern querying in time series databases. In Proc. of IEEE ICDE, pp.33–42 (2000).
- [QWW98] Qu, Y., Wang, C., Wang, X.S.: Supporting fast search in time series for movement patterns in multiple scales. In Proc. of ACM CIKM, pp. 251-258 (1998).
- [RM00] Rafiei, D., Mendelzon, A.: Querying time series data based on similarity. IEEE Transactions on Knowledge and Data Engineering, Vol. 12, No 5., pp. 675–693 (2000).

References

- [RPF+08] Renso, C., Puntoni, S., Frentzos, E., Mazzoni, A., Moelans, B., Pelekis, N., Pini, F.: Wireless Network Data Sources: Tracking and Synthesizing Trajectories. Mobility, Data Mining and Privacy 2008: 73-100.
- [RPN+08] Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., Andrienko, G.: Visually–driven analysis of movement data by progressive clustering. Information Visualization, 2008, v.7 (3/4), pp. 225-239.
- [SEA+03] Stefanidis, A., Eickhorst, K., Agouris, P., Partsinevelos, P.: Modeling and comparing change using spatiotemporal helixes. GIS 2003: 86-93.
- [SKP06] Subramaniam, S., Kalogeraki, V., Palpanas, T.: Distributed Real-Time Detection and Tracking of Homogeneous Regions in Sensor Networks. RTSS 2006: 401-411.
- [TM08] Terrovitis, M., Mamoulis, N.: Privacy Preservation in the Publication of Trajectories. MDM 2008: 65-72.
- [TPN+09] Tiakas, E., Papadopoulos, A., Nanopoulos, A., Manolopoulos, Y., Stojanović, D., Kajan, S.D., "Searching for similar trajectories in spatial networks", Journal of Systems and Software, Elsevier, Volume 82, Issue 5, pp. 772-788, 2009, ISSN: 0164-1212

References

- [TTD+09] Trajcevski, G., Tamassia, R., Ding, H., Scheuermann, P., Cruz, I.F.: Continuous probabilistic nearest-neighbor queries for uncertain trajectories. EDBT 2009: 874-885.
- [TWZ+02] Trajcevski, G., Wolfson, O., Zhang, F., Chamberlain, S.: The Geometry of Uncertainty in Moving Objects Databases. EDBT 2002: 233-250.
- [T11] Goce Trajcevski: Uncertainty in Spatial Trajectories. Computing with Spatial Trajectories 2011: 63-107
- [URT11] Muhammad Reaz Uddin, Chinya V. Ravishankar, Vassilis J. Tsotras: Finding Regions of Interest from Trajectory Data. Mobile Data Management (1) 2011: 39-48
- [V10] Olivier Verscheure: Smart Cities: How Data Mining and Optimization Can Shape Future Cities. ECML/PKDD (1) 2011: 11
- [VBT10] Vieira, M.R., Bakalov, P., Tsotras, V.J.: Querying trajectories using flexible patterns. EDBT 2010: 406-417.
- [VGD04] Vlachos, M., Gunopulos, D., Das, G.: Rotation invariant distance measures for trajectories. KDD 2004: 707-712.
- [VGK02] Vlachos, M., Gunopulos, D., Kollios, G.: Discovering Similar Multidimensional Trajectories. ICDE 2002: 673-684.
- [VHG+03] Vlachos, M., Hadjieleftheriou, M., Gunopulos, D., Keogh, E.J.: Indexing multi-dimensional time-series with support for multiple distance measures. KDD 2003: 216-225.

References

- [VKG05] Vlachos, M., Kollios, G., Gunopulos, D.: Elastic Translation Invariant Matching of Trajectories. Machine Learning 58(2-3): 301-334 (2005).
- [VKY09] Vlachos, M., Kozat, S.S., Yu, P.S.: Optimal Distance Bounds on Time-Series Data. SDM 2009: 109-120.
- [VLR+08] Michail Vlachos, Claudio Lucchese, Deepak Rajan, Philip S. Yu: Ownership protection of shape datasets with geodesic distance preservation. EDBT 2008: 276-286.
- [VKY10] Michail Vlachos, Suleyman Serdar Kozat, Philip S. Yu: Optimal distance bounds for fast search on compressed time-series query logs. TWEB 4(2): (2010)
- [VKY09] Michail Vlachos, Suleyman Serdar Kozat, Philip S. Yu: Optimal Distance Bounds on Time-Series Data. SDM 2009: 109-120
- [WMD+12] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann and E. Keogh. Experimental Comparison of Representation Methods and Distance Measures for Time-Series Data. DMKD (to appear), 2012.
- [YGK+11] Zhixian Yan, Nikos Giatrakos, Vangelis Katsikaros, Nikos Pelekis, Yannis Theodoridis: SeTraStream: Semantic-Aware Trajectory Construction over Streaming Movement Data. SSTD 2011: 367-385

References

- [VKG05] Vlachos, M., Kollios, G., Gunopulos, D.: Elastic Translation Invariant Matching of Trajectories. Machine Learning 58(2-3): 301-334 (2005).
- [[YLG06] Zeinalipour-Yazti, D., Lin, S., Gunopulos, D.: Distributed spatio-temporal similarity search. CIKM 2006: 14-23.
- [KTZ+11] Kai Zheng, Goce Trajcevski, Xiaofang Zhou, Peter Scheuermann: Probabilistic range queries for uncertain trajectories on road networks. EDBT 2011: 283-294
- [ZLA+11] Demetrios Zeinalipour-Yazti, Christos Laoudias, Maria I. Andreou, Dimitrios Gunopulos: Disclosure-Free GPS Trace Search in Smartphone Networks. Mobile Data Management (1) 2011: 78-87

Mining Data From Mobile Devices

Dimitrios Gunopulos
University of Athens
dg@di.uoa.gr