电子科技大学
University of Electronic Science and Technology of China

Similarity Measurement of Trajectories

Wang Wei

Data Mining Lab,
Big Data Research Center, UESTC
Email :wweichn@163.com

Outline

1 Introduction

- Derivation
- Application

2 Trajectory compression

- Douglas-Peucker
- Discrete Fourier Transform

3 Similarity measurement

- Euclidean space

----Dynamic Time Warpping(DTW)

----Edit Distance on Segment(EDS)

----Flatly structured grids (FSG)

- Road network space

----Trajectory point filter

----Trajectory point matching

1.1 Derivation

1.2 Application

1.1 Derivation

- mobility of people

----active recording
joggers record trails for sports analysis
----passive recording
transaction records of a credit card

- mobility of transportation vehicles

Taxis with a GPS sensor

- mobility of animals

Moving trajectories of tigers and birds

- mobility of natural phenomena

Trajectories of hurricanes,tornados, and ocean currents

1.2 Application

数据挖掘实验室
Data Mining Lab

- behavior pattern mining

----personalized recommendation
(potential friends)

- route planning

----travel routes
----heavy-traffic roads

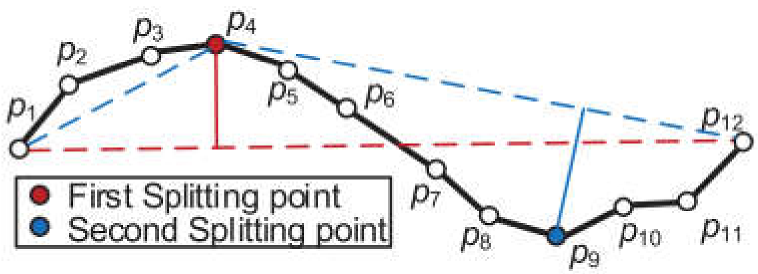
- location prediction

2.1 Douglas-Peucker algorithm

2.2 Discrete Fourier Transform

2 Trajectory compression

- Douglas-Peucker algorithm



Input: The trajectory T;
Start point i;
End point j;
Distance threshold ε

Output: Trajectory after compression

function $DOUGLASPEUCKER(T, i, j, \varepsilon)$

Find point farthest from the line $\overline{p_i p_j}$ between $\{p_{i+1} : p_{j-1}\}$, set the distance $dist$

If $dist > \varepsilon$, set p_f as the separation point

$DOUGLASPEUCKER(T, i, f, \varepsilon)$

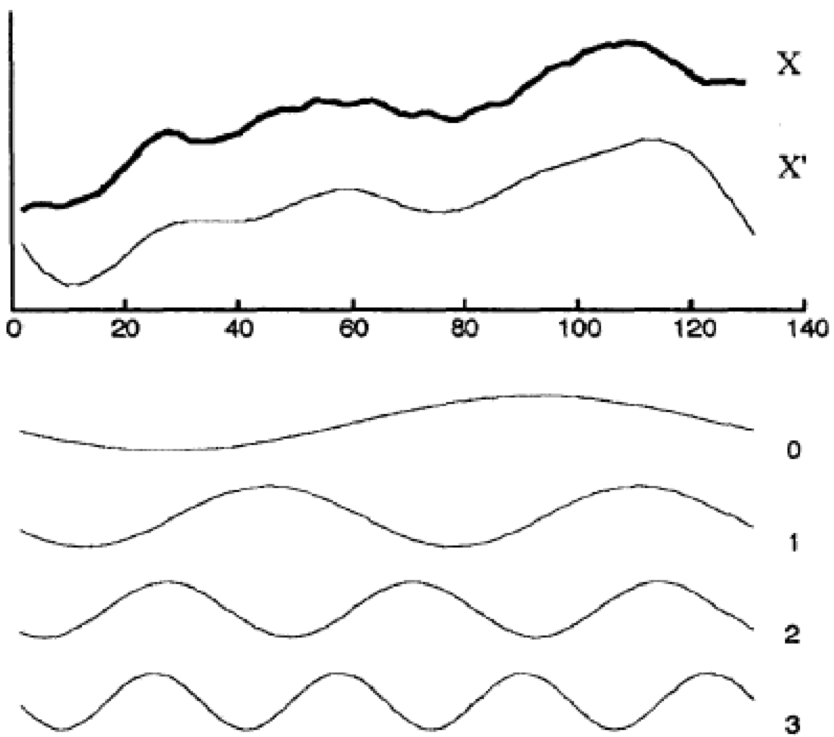
$DOUGLASPEUCKER(T, f, j, \varepsilon)$

Else output $\overline{p_i p_j}$

end

- Discrete Fourier transform (DFT)

DFT convert signals from the time domain to the frequency domain.
Any signal can be represented by a sum of sine waves(or cosine waves).



3 Similarity measurement

数据挖掘实验室
Data Mining Lab

3.1 Euclidean space

----Dynamic Time Warping(DTW)

----Edit Distance on Segment(EDS)

----Flatly Structured Grids(FSG)

3.2 Road network space

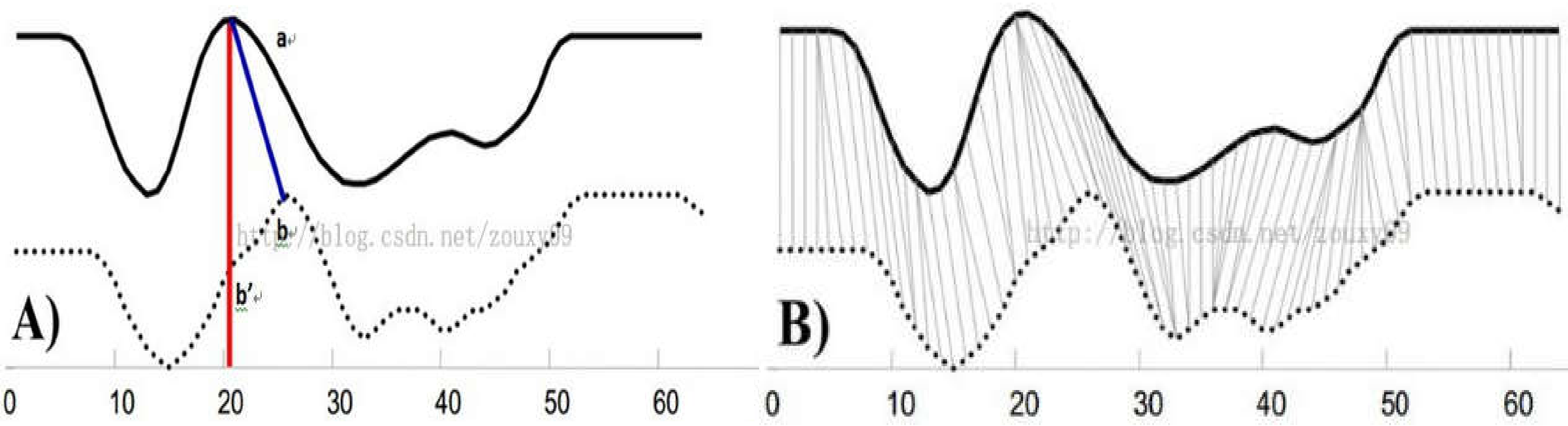
----Trajectory point filter

----Trajectory point matching

3.1 Euclidean space

- Dynamic Time Wrapping(DTW)

  to measure similarity between two temporal sequences which may vary in time or speed.



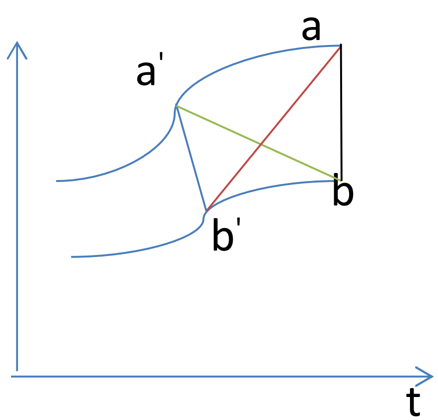
1)boundary conditions: $w_1=(1,1)$ and $w_k=(m,n)$

2)continuity: if $w_{k-1}=(a',b')$,then the next pair $w_k=(a,b)$ need to satisfy $(a-a')<=1$ and $(b-b')<=1$

3)monotonicity: if $w_{k-1}=(a',b')$, then the next pair $w_k=(a,b)$ need to satisfy $0<=(a-a')$ and $0<=(b-b')$

For two time series A、B, |A|=m,|B|=n

$A=(a_1,a_2,\ldots,a_m),B=(b_1,b_2,\ldots,b_n)$



$$DTW(A,B) = \begin{cases} \infty, |m| = 0 \text{ or } |n| = 0 \\ dist(a_1, b_1) + \min\{ \\ DTW(Rest(A), Rest(B)), \\ DTW(Rest(A), B), \\ DTW(A, Rest(B))\} \end{cases}$$

$$Rest(A) = (a_2, a_3, \ldots, a_m), Rest(B) = (b_2, b_3, \ldots, b_n)$$

- Edit Distance on Segment(EDS)

Transformation cost

$$cost = w_{dis} * c_{dis}(s,s') + w_{str} * c_{str}(s,s') + w_{rot} * c_{rot}(s,s')$$

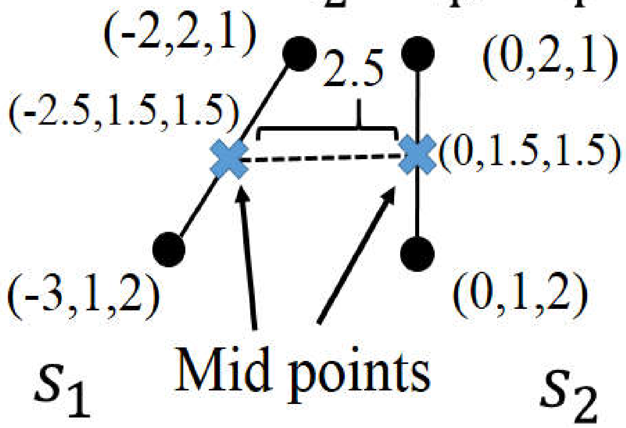
displacing cost $c_{dis}(s,s') = d(mid(s), mid(s')) / d_{\max}$

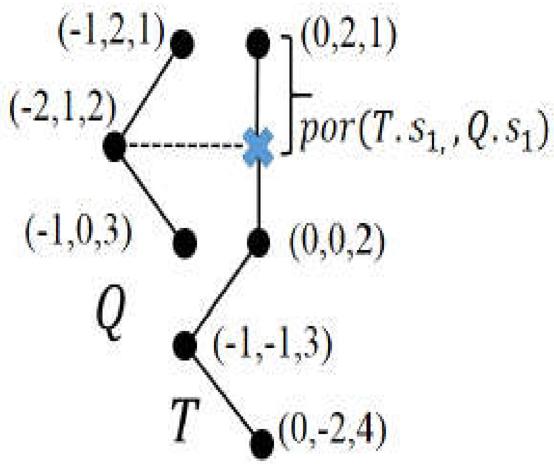
$d_{\max}$ is the maximum distance between s and s'

stretching cost $c_{str} = 1 - \min\{|s|,|s'|\} / \max\{|s|,|s'|\}$

rotating cost $c_{rot} = \theta(s,s') / \pi$

$$\theta(s_1, s_2) = \left(\frac{\pi}{2} - \frac{\pi}{4}\right) = \frac{\pi}{4}$$





1)insert, Q.s1 is transformed to the portion of T.s1 This portion of T.s1 is denoted by por(T.s1, Q.s1).

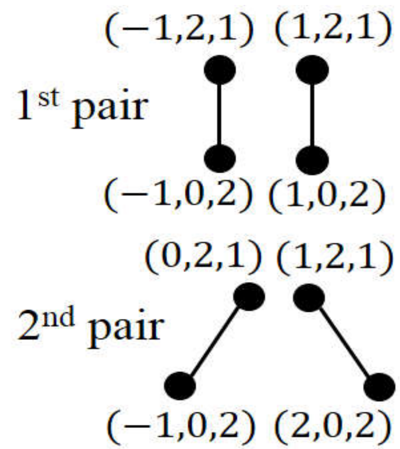
EDS(Q,T) = cost(Q.s1,por(T.s1,Q.s1)) + EDS(Q\Q.s1,T\por(T.s1,Q.s1))

2)replace, Q.s1 is transformed to T.s1

EDS(Q,T) = cost(Q.s1,T.s1) + EDS(Q\Q.s1,T\T.s1)

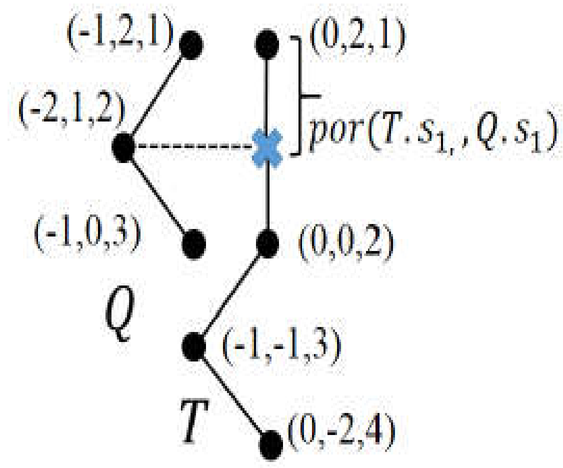
$$
\mathrm{EDS(T,Q)} = \begin{cases} 0, |Q| = 0 \\ \infty, |T| = 0 \, and \, |Q| \neq 0 \\ \min\{ \\ \mathrm{cost}(T.s1, por(Q.s1, T.s1)) + EDS(T \setminus T.s1, Q \setminus por(Q.s1, t.s1)) \\ \mathrm{cost}(por(T.s1, Q.s1), Q.s1) + EDS(T \setminus por(T.s1, Q.s1), Q \setminus Q.s1), \\ \mathrm{cost}(T.s1, Q.s1) + EDS(T \setminus T.s1, Q \setminus Q.s1)\} \end{cases}
$$

- **Features**



- EDS is a segment-based distance measure
(better than a point-based distance measure)

- EDS only enumerates all possible suffixes of the trajectory
(Computation cost: O(n) better than O(n^2))
(for a trajectory consisting of n sampled points)



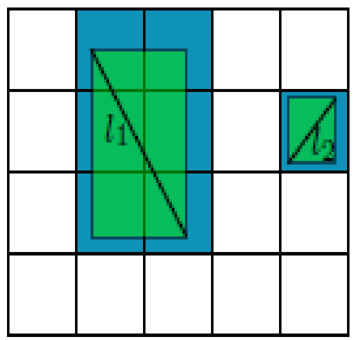
- Flatly Structured Grids (FSG)

to index trajectory data with grid cells

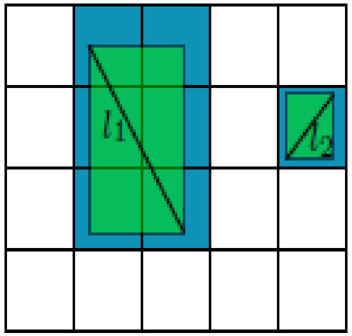
- R-Tree

to group nearby objects and represent them with their minimum bounding rectangle

Each line segment is assigned to the FSG by rasterizing its MBB to grid cells.



Spatial domain



minimum bounding boxes

$$MBB_l^{min} = (\min(x_i^{start}, x_i^{end}), \min(y_i^{start}, y_i^{end}), \min(z_i^{start}, z_i^{end}))$$

$$MBB_l^{max} = (\max(x_i^{start}, x_i^{end}), \max(y_i^{start}, y_i^{end}), \max(z_i^{start}, z_i^{end}))$$

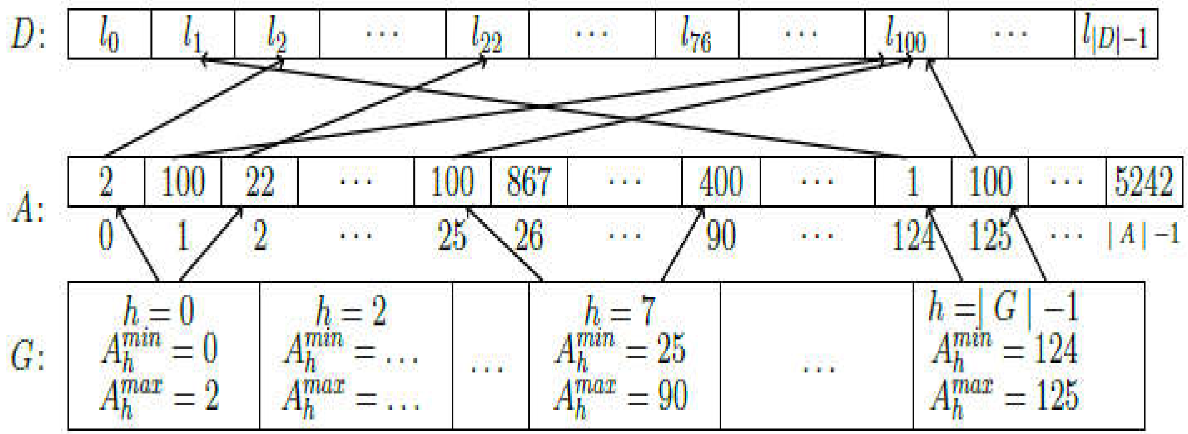
The database
of entry line segments

Integer array

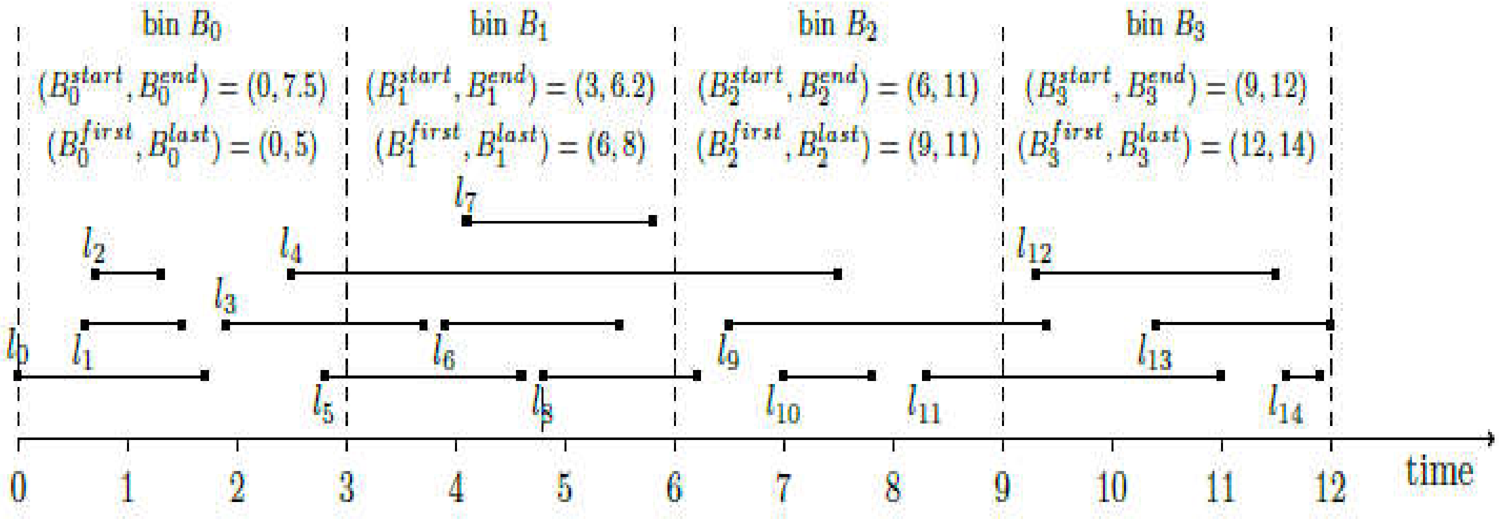
indices of the line segments
whose MBBs overlap cell

Grid cells array

Index range in A



Temporal domain



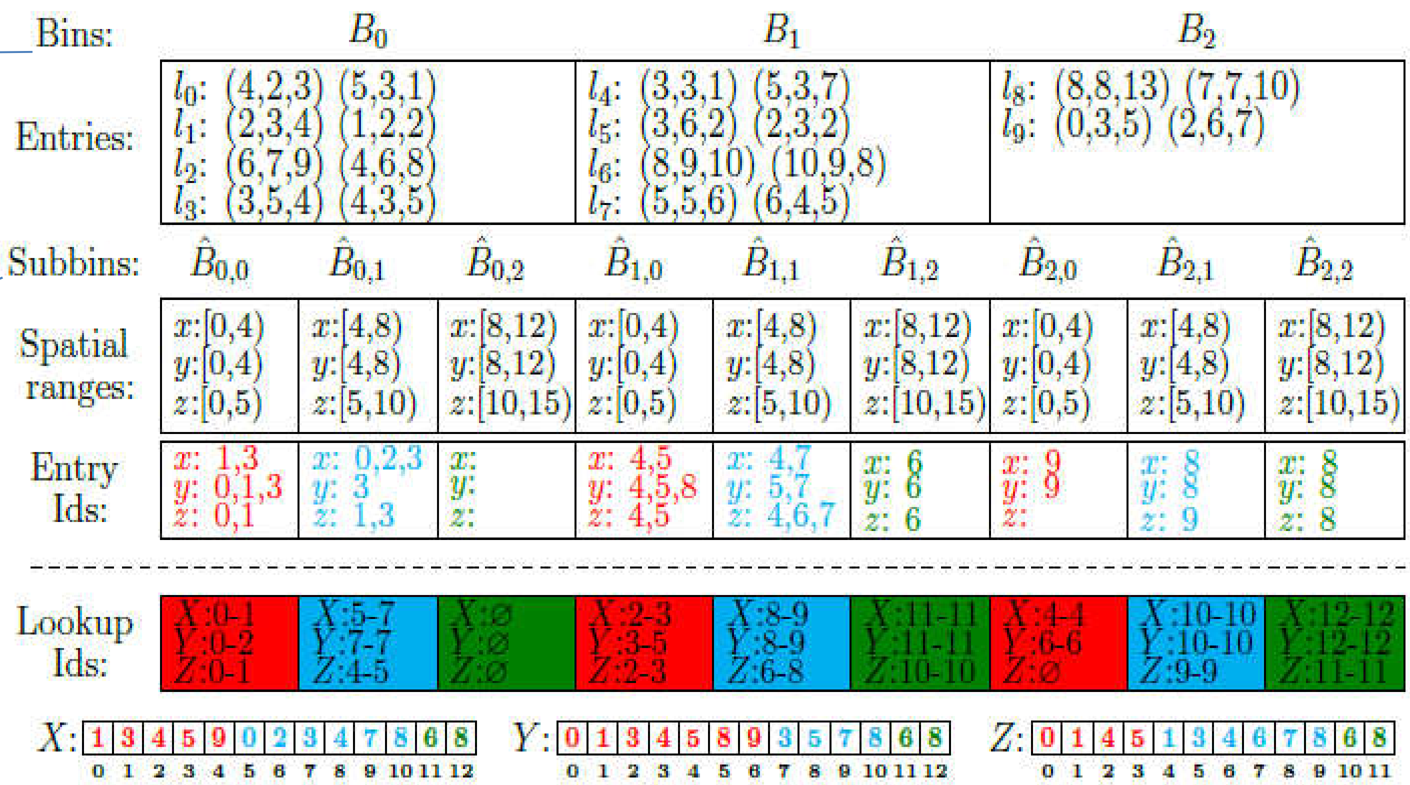
divide the temporal range into m segments

$$b = (t_{\max} - t_{\min}) / m \qquad B_j^{start} = j * b, B_j^{end} = \max((j+1) * b, \max_{l_i \in B_j} t_i^{end})$$

$[B_j^{first}, B_j^{last}]$ index range of entry segments in B_j

3 temporal bins

9 temporal spatial subbins



Each array stores the ids of the line segments that overlap the subbins in one spatial dimension

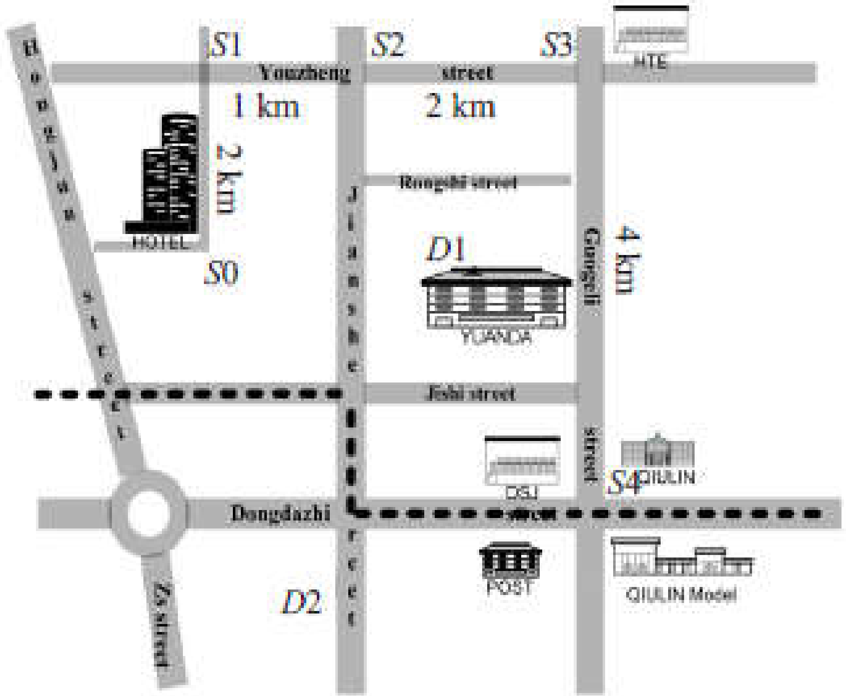
Spatial network

- Trajectory point filter

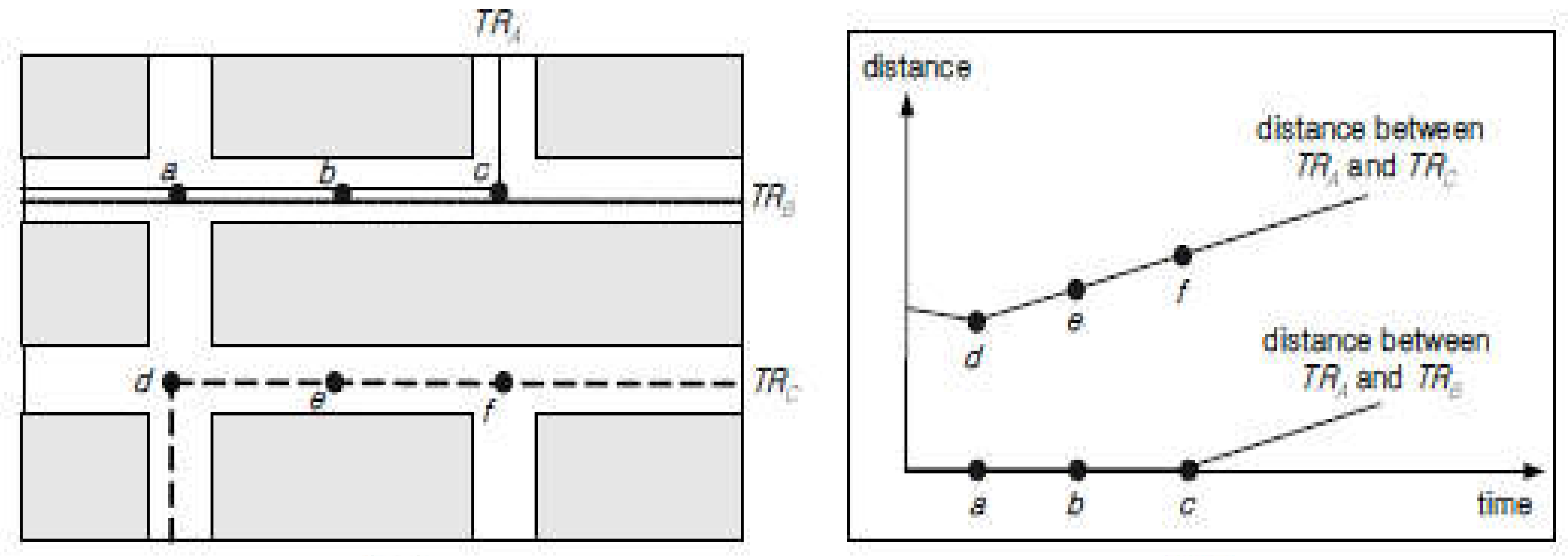
1)Coordinate system (Sid,d,t)

Sid: road sector identifier
d: the offset from the starting point of the road sector
t: time

2)Point of Interest(POI)
Important intersections of roads or places



It may be meaningless to compute distance between two moving objects if they are on different road sectors.



(a) (b)

1)The filtering step based on spatial similarity step

Spatial similarity between two trajectories TR_A and TR_B

$$Sim_{POI}(TR_A, TR_B, P) = \begin{cases} 1, & if\ \forall p \in P,\ p\ is\ on\ TR_A\ and\ TR_B \\ 0, & otherwise \end{cases}$$

(P is a set of POIs on a given road networks)

2) The refinement step based on temporal distance

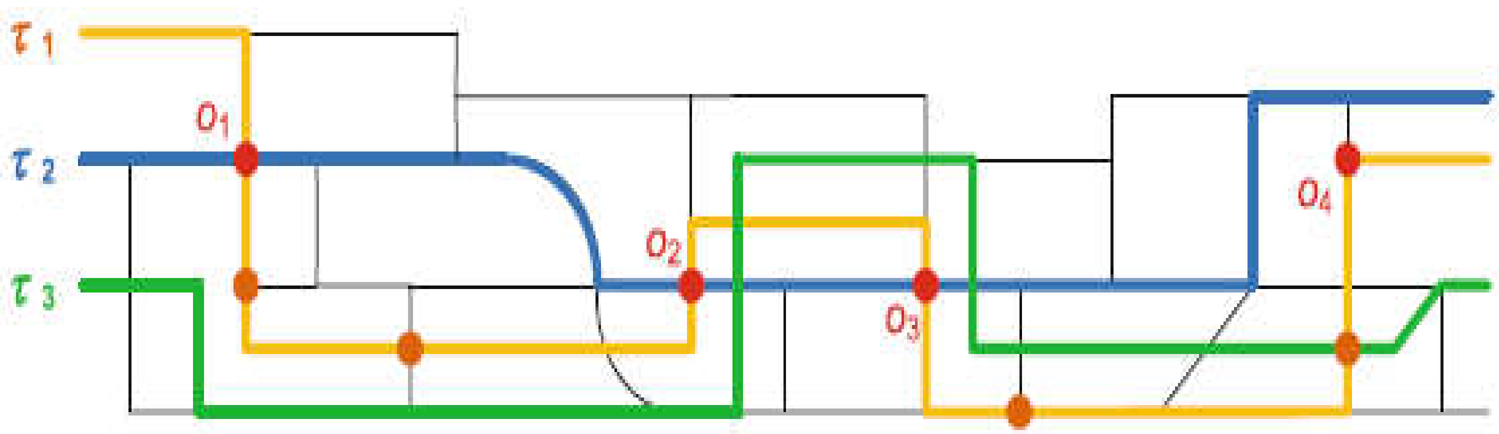
Temporal Distance

$$dist_T(TR_A, TR_B, P) = L_p(TR_A, TR_B, p) = \left(\sum_{i=1}^{k} |(p_i(TR_A) - p_i(TR_B)|^p\right)^{\frac{1}{p}}$$

- Trajectory point match

If only spatial similarity is considered, trajectory τ_3 will be returned

This approach takes into account the significance of each sample point.

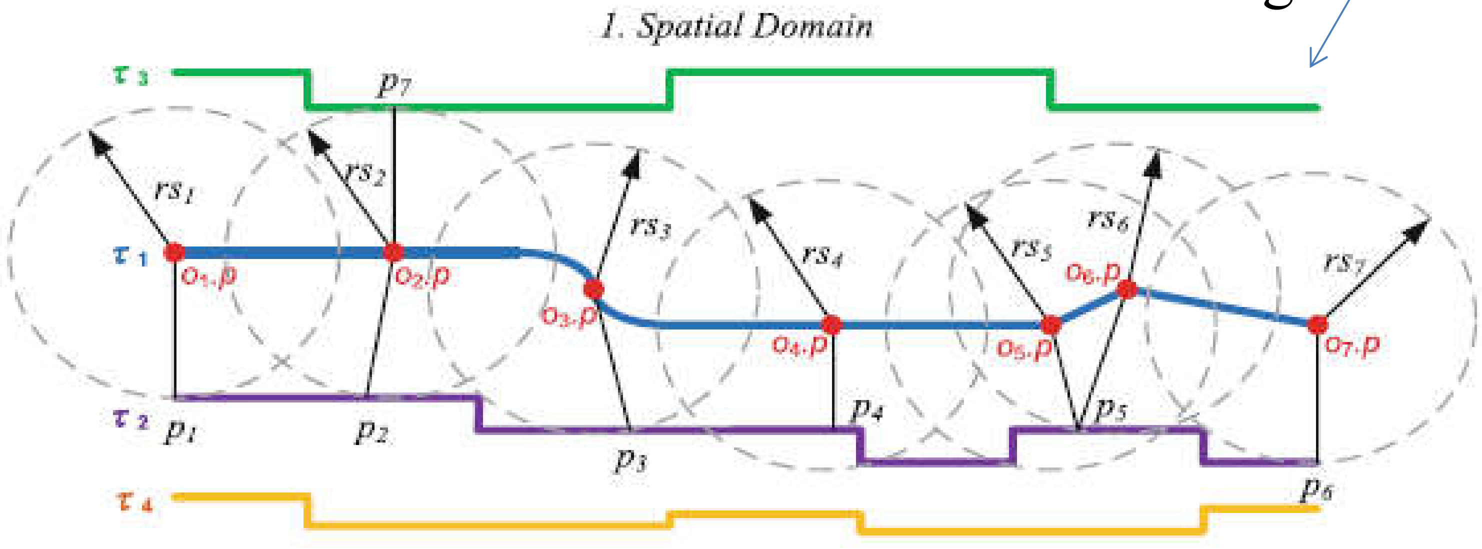


Spatial similarity

$$I_S(v1, v2) = \begin{cases} 0, & \text{if } sd(v_1.p, v_2.p) > \varepsilon_s \\ e^{-sd(v1.p, v2.p)}, & \text{otherwise} \end{cases}$$

$$S_{sim}(q, \tau) = \max \begin{cases} q.head.w \cdot I_s(q.head, \tau.head) + S_{sim}(q.tail, \tau) \\ S_{sim}(q, \tau.tail) \end{cases}$$

Dijkstra algorithm



The upper bound on the spatial similarity

$$S_{sim}(q,\tau).ub = \max\{a, S_{sim}(q,\tau.tail)\}$$

$$a = \begin{cases} q.head.w.e^{-sd(q.head.p,\tau.head.p)} + S_{sim}(q.tail,\tau), & \text{if } C_1 \\ q.head.w.e^{-rs_i} + S_{sim}(q.tail,\tau), & \text{if } C_2 \end{cases}$$

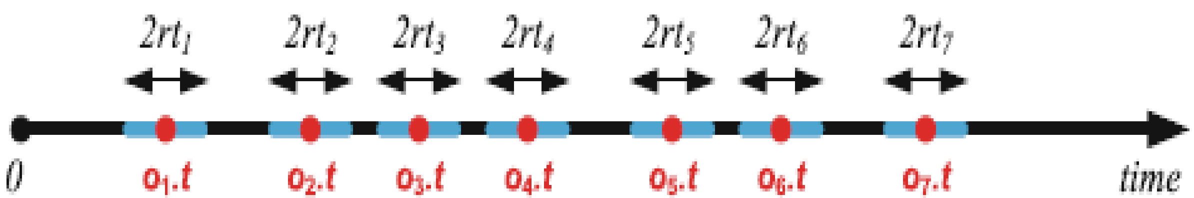
C1 The value of sd(q.head.p, τ.head.p) is available (can be obtained from the current network expansion).
C2 The value of sd(q.head.p, τ.head.p) is not available and is replaced by the value of rsi .

Temporal distance

$$I_t(v1, v2) = \begin{cases} 0, & \text{if } |\text{v1.t} - \text{v2.t}| > \varepsilon_t \\ e^{-|v1.t - v2.t|}, & \text{otherwise} \end{cases}$$

$$T_{sim}(q, \tau) = \begin{cases} q.head.w \cdot I_t(q.head, \tau.head) + T_{sim}(q.tail, \tau) \\ T_{sim}(q, \tau.tail) \end{cases}$$

2. Temporal Domain



The upper bound on the temporal similarity

$$T_{sim}(q,\tau).ub = \max\{\beta, T_{sim}(q,\tau.tail)\}$$

$$\beta = \begin{cases} q.head.w.e^{-|q.head.t-\tau.head.t|} + T_{sim}(q.tail,\tau), & \text{if C1} \\ q.head.w.e^{-rt_i} + T_{sim}(q.tail,\tau), & \text{if } C2 \end{cases}$$

C1 The value of |q.head.t − τ.head.t| is available.

C2 The value of |q.head.t − τ.head.t| is not available and is replaced by the value of rti global lower-bound LB

$$LB = \max_{\forall \tau \in T_f} \{ST_{sim}(q, \tau)\}$$

τ availiable in spatial domain and time domain

global upper-bound UB

$$UB = \max_{\tau \in T_n} \{ST_{sim}(q, \tau).ub\}$$

τ partly touched in the spatial or the temporal domains

UB<LB the corresponding trajectory is the most similar trajectory

Thanks

Wang Wei
Yingcai Experimental School
wweichn@163.com