Spatio-temporal Mobility Summary using Location Traces

Manasa J M

Master of Science (by Research)

Under the supervision of

Dr. Soumya K Ghosh and Dr. Vinay Kolar (IBM Research)

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Section Overview

Introduction
- Problem Definition
- Motivation
- Objectives

Literature Survey

Trajectory Preprocessing

Trajectory Similarity

Trajectory Clustering

Conclusion and Future Work

Introduction

PROBLEM DEFINITION, MOTIVATION, OBJECTIVES

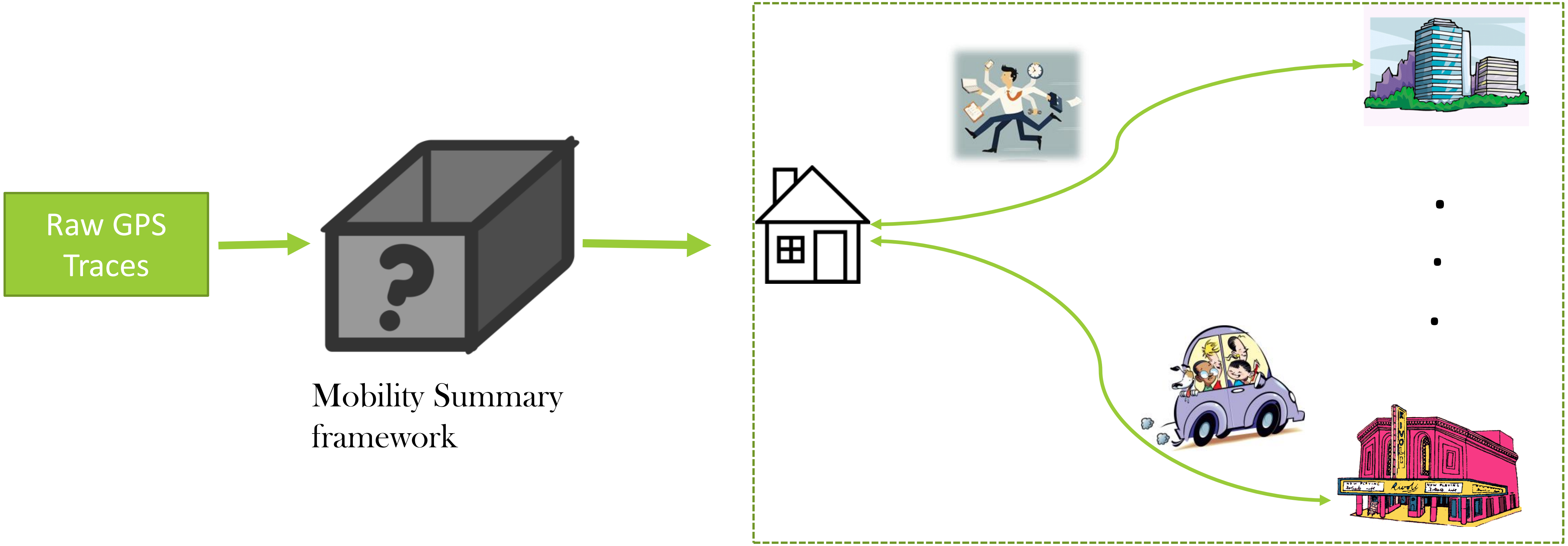
- Mobility Data of people collected at a large scale
- Various sources of data – GPS traces, Call Detail Records, Location Based Social Networks
- Analysis at both individual and community level is possible
- Other applications of trajectory mining includes animal cluster movement prediction, hurricane movement, etc

Animal Movement Pattern

Satellite image showing the trajectories of 380 taxi trajectories in London

Problem Definition

Mobility Summary of an individual define the frequent paths taken by the user.



Mobility Summary of a user

Motivation

- Mobility Summary represents the movement pattern of a user
- Queries on the mobility of a person need not deal with all the trajectories
- Provides as an abstraction to the user's movement.
- Queries based on mobility summary include -

Users who frequently pass through a given place – ex- Coffee shop

Next path prediction

Anomaly Detection

Objectives

- Trajectory Preprocessing
- Trajectory Similarity Analysis
- Trajectory Clustering
- Trajectory Summarization
- Applications of Mobility Summary

Literature Survey

Trajectory Preprocessing

Authors	Work	Contribution
Li, Zheng et al.	Mining user similarity based on location history	Algorithm for Stay point detection
Zheng and Xie	Learning travel recommendations from user-generated GPS traces	Uses of stay point to define a trip
Yu Zheng	Trajectory Data Mining: An Overview	Trajectory segmentation

Trajectory Similarity Analysis

Authors	Work	Contribution
Agrawal et al.	Efficient similarity search in sequence databases	Euclidean distance similarity with Discrete Fourier Transform
Berndt et al.	Finding patterns in time series	Introduced Dynamic Time Warping similarity
Chen et al.	Robust and fast similarity search for moving object trajectories	Introduced Edit Distances on Real Sequences(EDR) similarity
Chen et al.	On the marriage of LP-norms and edit distance	Introduced Edit Distance with Real Penalty(ERP) similarity
Wang et al	An effectiveness study on trajectory similarity measures	Comparison of six widely used similarity measures

Trajectory Clustering

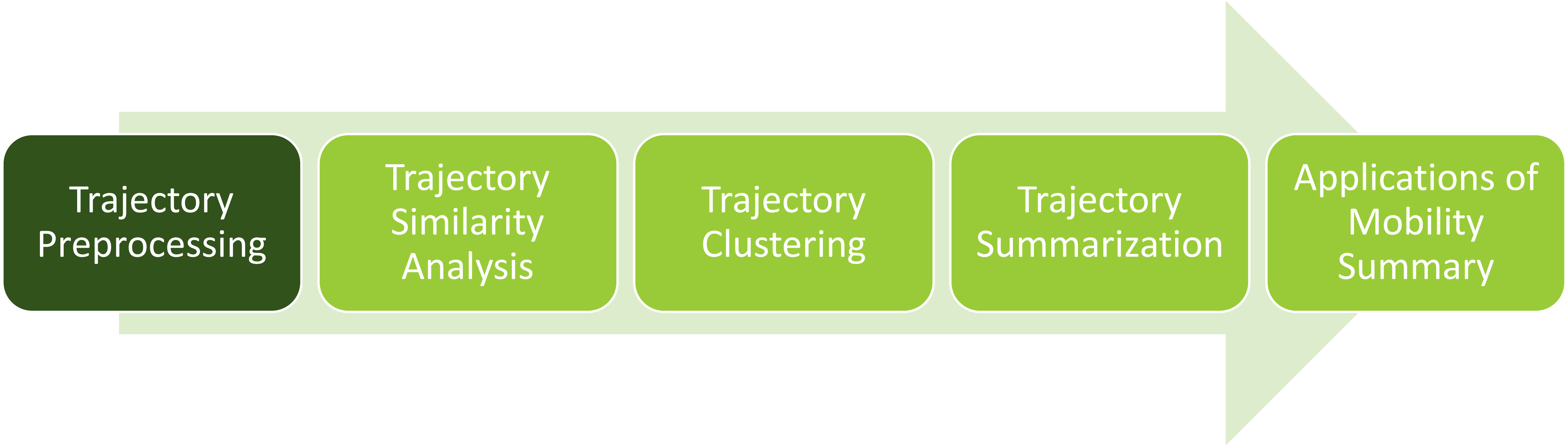
Authors	Work	Contribution
Gaffney et al.	Trajectory clustering with mixtures of regression models	Trajectory clustering using Expectation Maximization algorithm
Nanni et al.	Time-focused clustering of trajectories of moving objects	Proposed a density based clustering algorithm
Zhang et al.	Clustering spatio-temporal trajectories based on kernel density estimation	Kernel density estimation based approach for clustering
Han et al.	Neat: Road network aware trajectory clustering	Clustering after map matching

Sub-Trajectory Clustering

Authors	Work	Contribution
Li et al.	Swarm: Mining relaxed temporal moving object clusters.	Moving Clusters and Detecting closed Swarms
Lee et al.	Trajectory clustering: A partition-and-group framework	Partitioning the sub-trajectories and grouping them as a whole

Representative Trajectory

Authors	Work	Contribution
Buchin et al.	Median Trajectories	Different ways to find the representative or median trajectory for a cluster



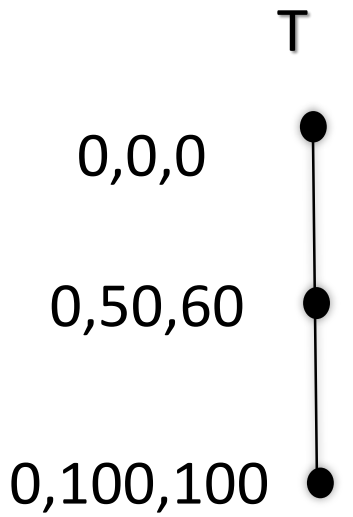
Trajectory Preprocessing

TRAJECTORY CUT, NORMALIZATION

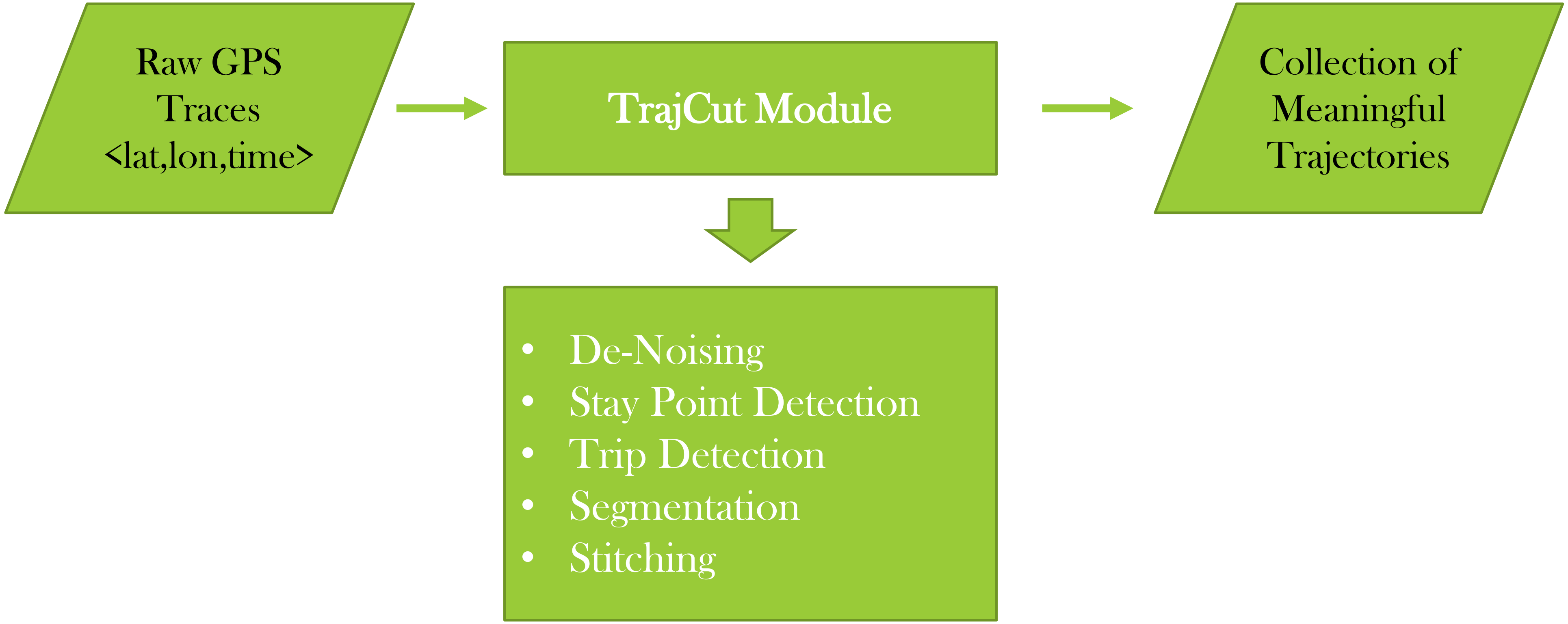
Trajectory

- A trajectory is an *ordered sequence* of *spatio-temporal points*.
  - $T=[s_1,\ldots,s_n]$

- A spatio-temporal point has three dimensions
  - $s_i = (x,y,t)$
    - $s_i.x$ and $s_i.y$ describes the space
    - $s_i.t$ describes the time, $s_i.t \leq s_{i+d}.t \; \forall d \geq 0$

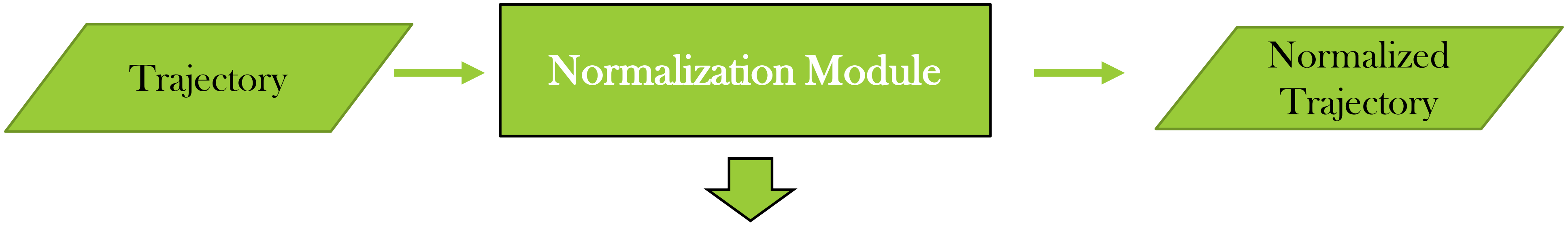


Trajectory Cut



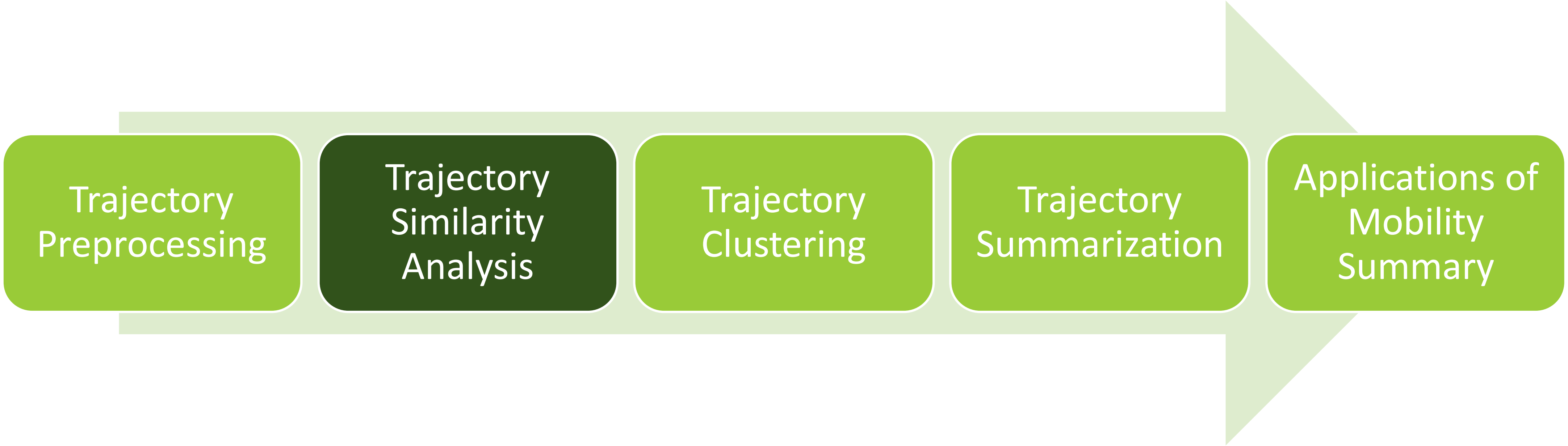
Normalization

- Each of the <lat, lon> pairs in the trajectory has to be normalized
- Small differences in latitude and longitude can cause great differences on distances.



Normalized latitude	Normalized longitude
$\mathrm{lat}_i = \dfrac{\mathrm{rlat}_i - \mathrm{min}_{\mathrm{lat}}}{\mathrm{max}_{\mathrm{lat}} - \mathrm{min}_{\mathrm{lat}}}$	$\mathrm{lon}_i = \dfrac{\mathrm{rlon}_i - \mathrm{min}_{\mathrm{lon}}}{\mathrm{max}_{\mathrm{lon}} - \mathrm{min}_{\mathrm{lon}}}$

Symbol	Explanation
rlat_i , rlon_i	Raw lat, raw lon
lat_i , lon_i	Normalized lat, lon
$\mathrm{min}_{\mathrm{lat}}$, $\mathrm{min}_{\mathrm{lon}}$, $\mathrm{max}_{\mathrm{lat}}$, $\mathrm{max}_{\mathrm{lon}}$	Bounding box of the user's trajectories



Trajectory Similarity Analysis

WEIGHTED CURVE DISTANCE, DTW

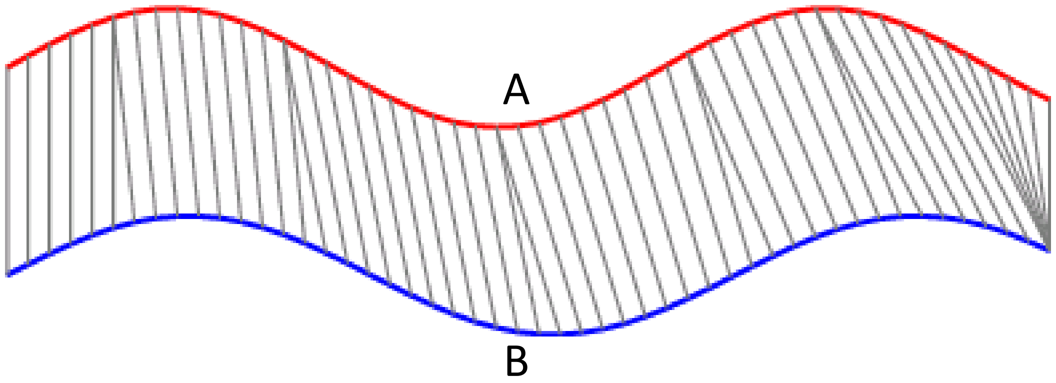
Trajectory Similarity: Dynamic Time Warping (DTW)

ϕ_d (p1,p2)=L_2(p1,p2)

$$D_{dtw}(A,B) = \begin{cases} 0 & \text{if both } A \text{ and } B \text{ are empty} \\ \infty & \text{if one of } A \text{ or } B \text{ is empty} \\ \phi_d(head(A), head(B)) + \\ \quad min \begin{cases} D_{dtw}(A, rest(B)), \\ D_{dtw}(rest(A), B), \\ D_{dtw}(rest(A), rest(B)) \end{cases} \\ \quad \text{otherwise.} \end{cases}$$

Stretch A

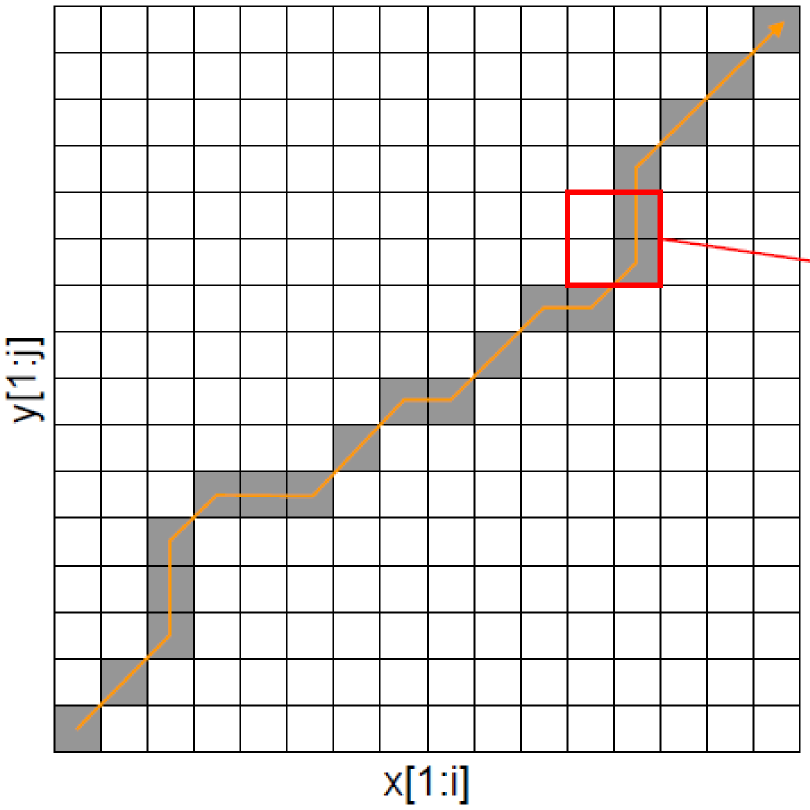
Stretch B



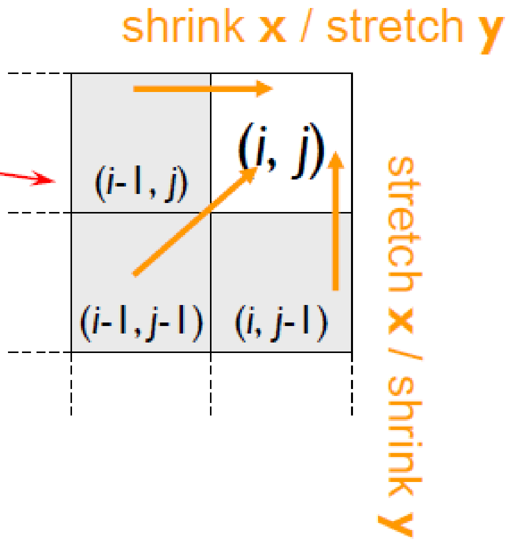
Every point in A has to be matched to some point in B

B. kee Yi, H. V. Jagadish, and C. Faloutsos. Efficient Retrieval of Similar Time Sequences Under Time Warping. In ICDE, 1998.

DTW: Computation using Dynamic Programming

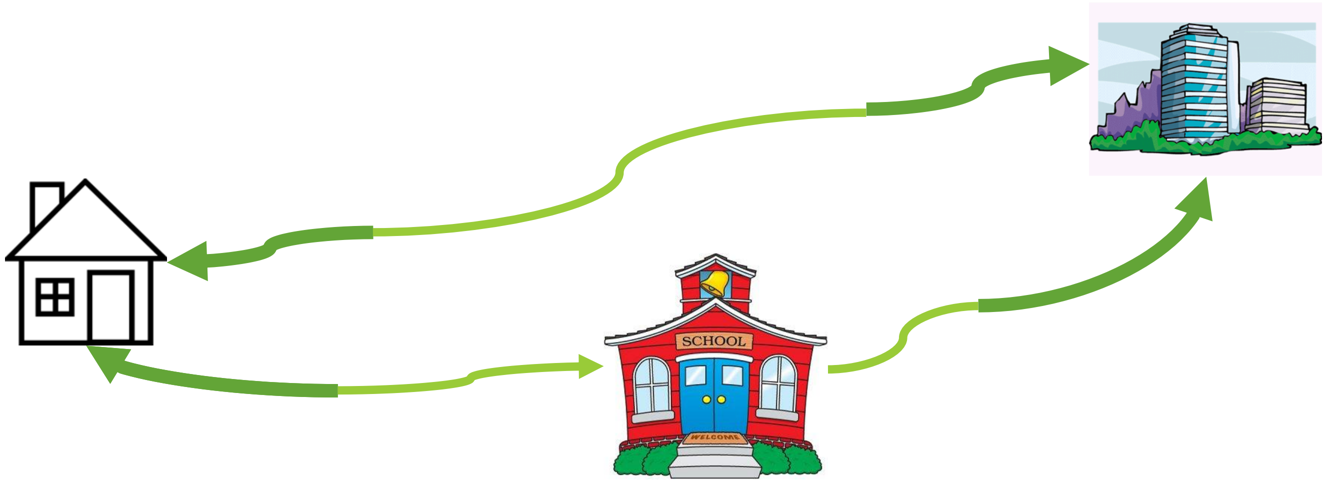


- Examine all paths
- Optimal matching is path connecting the two corners



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

Proposed Method: Motivation



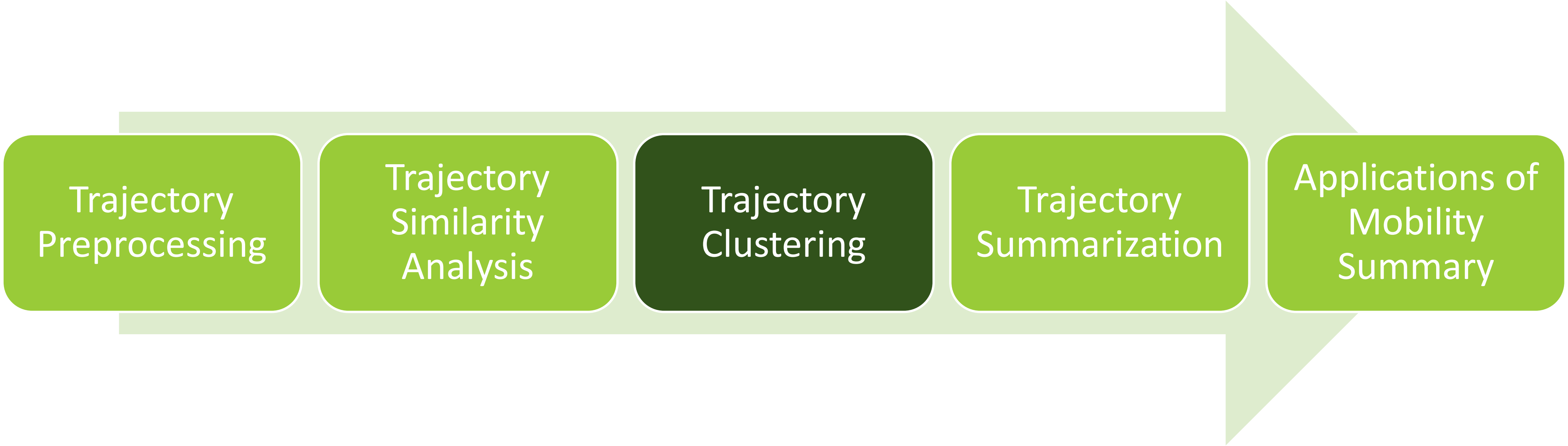
- We want to capture human mobility
- Humans have an intention behind the trips the make unlike hurricanes and animal movement
- This can be captured by giving more weightage to origin and destination

Proposed Method

User's Trajectories → **Weighted Curve Distance Similarity** → *nXn* Similarity Matrix

- Resample the Trajectories into N points
- Normalize the sample points
- Take pointwise LP Norm with giving more weight to the points closer to the origin and destination

$$\mathrm{WCD}(t_i, t_j) = \left[\int_0^1 w(x)\,(f_i(x) - f_j(x))^2 \,\mathrm{d}x\right]^{\frac{1}{2}}.$$



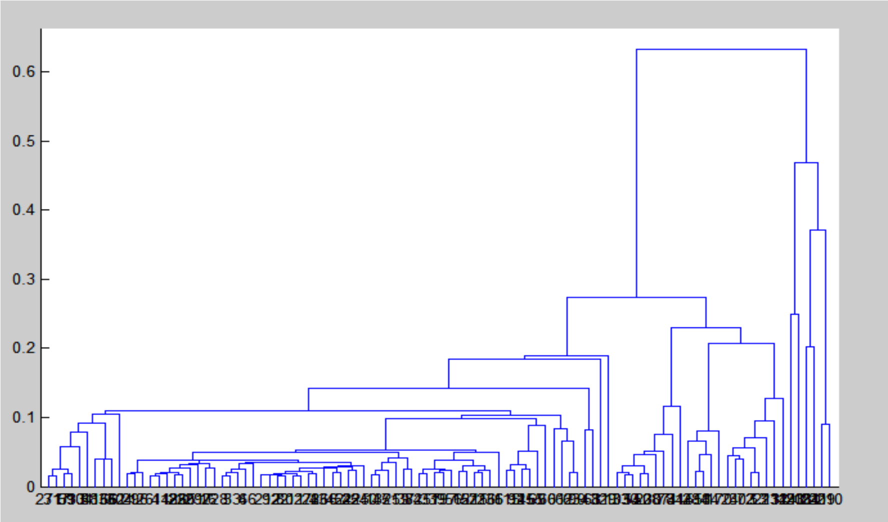
Trajectory Clustering

HIERARCHICAL CLUSTERING, OPTIMAL POINT

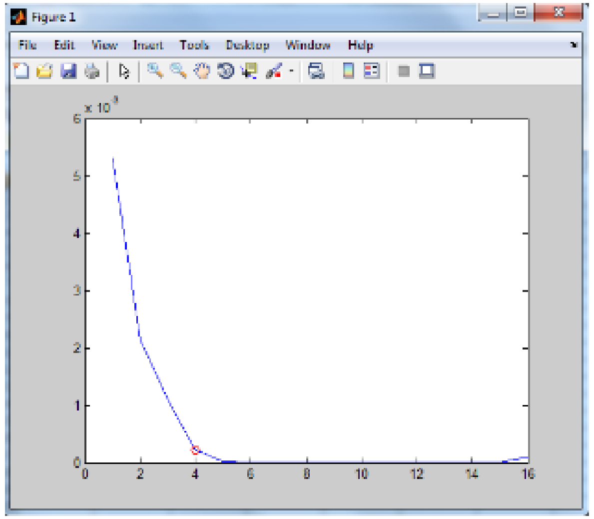
Hierarchical Clustering

- Perform hierarchical clustering on the distance matrix obtained from trajectory similarity.
- Evaluate a good measure for defining goodness of a cluster
- Devise a heuristic to stop at the optimal point

Similarity Matrix



Dendrogram



Elbow Curve

Heuristic to find optimal point

Final Clusters

Cluster goodness Measures

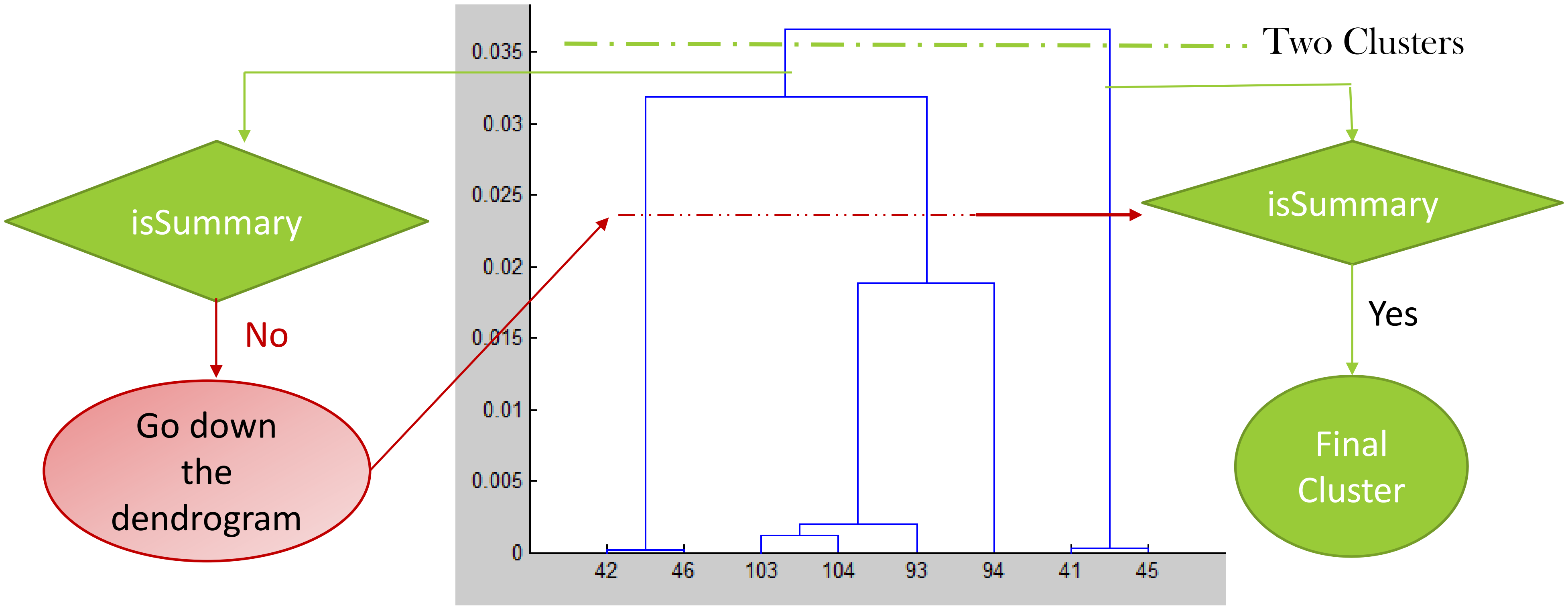
Table B.1: Formulas for internal indexes

Name	Formula
SSW	$SSW = \frac{1}{N}\sum_{i=1}^{N} \left\| x_i - C_{p_i} \right\|^2$
SSB	$SSB = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=1, j \neq i}^{M} \left\| C_i - C_j \right\|^2$
Calinski-Harabasz index	$CH = \frac{SSB/(M-1)}{SSW/(N-M)}$
Hartigan	$H_M = (\frac{SSW_M}{SSW_{M+1}} - 1)(N - M - 1)$ $or: H_M = \log(SSB_M / SSW_M)$
Krzanowski-Lai index	$diff_M = (M-1)^{2/D} SSW_{M-1} - M^{2/D} SSW_M$ $KL_M = \|diff_M\| / \|diff_{M+1}\|$
Ball&Hall	$BH_M = SSW_M / M$
Xu-index	$Xu = D \log(\sqrt{SSW_M / (DN^2)}) + \log M$
Dunn's index	$Dunn = \sum_{i=1}^{M} \frac{\max\{\left\| x_i - C_i \right\|^2\}_{i \in C_i}}{}$
Davies&Bouldin index	$R_{ij} = \frac{S_i + S_j}{d_{ij}}, i \neq j$ $where: d_{ij} = \left\| C_i - C_j \right\|^2, S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left\| x_j - C_i \right\|^2$ $and, R_i = \max_{j=1,\ldots,M} R_{ij}, i = 1, \ldots, M$ $DBI = \frac{1}{M} \sum_{i=1}^{M} R_i$

Finding the optimal number of Clusters

```
Find the elbow point from the SSW Plot over all levels
Set all trajectories as unmarked
for k=elbowPoint+1 to N do
    for Each non anomalous cluster do
        if Trajs(Cluster) are unmarked && isSummary(Cluster_i) then
            Report Cluster as a final Cluster
            Mark all Trajs in Cluster
        end
    end
end
isSummary(Cluster_i)
for All pairs of trajectories in Cluster do
    if Maximum Pointwise Distance≤ δ (kms) then
        return True
    end
end
return False
```

Optimal Cluster No. – Heuristic Explained…



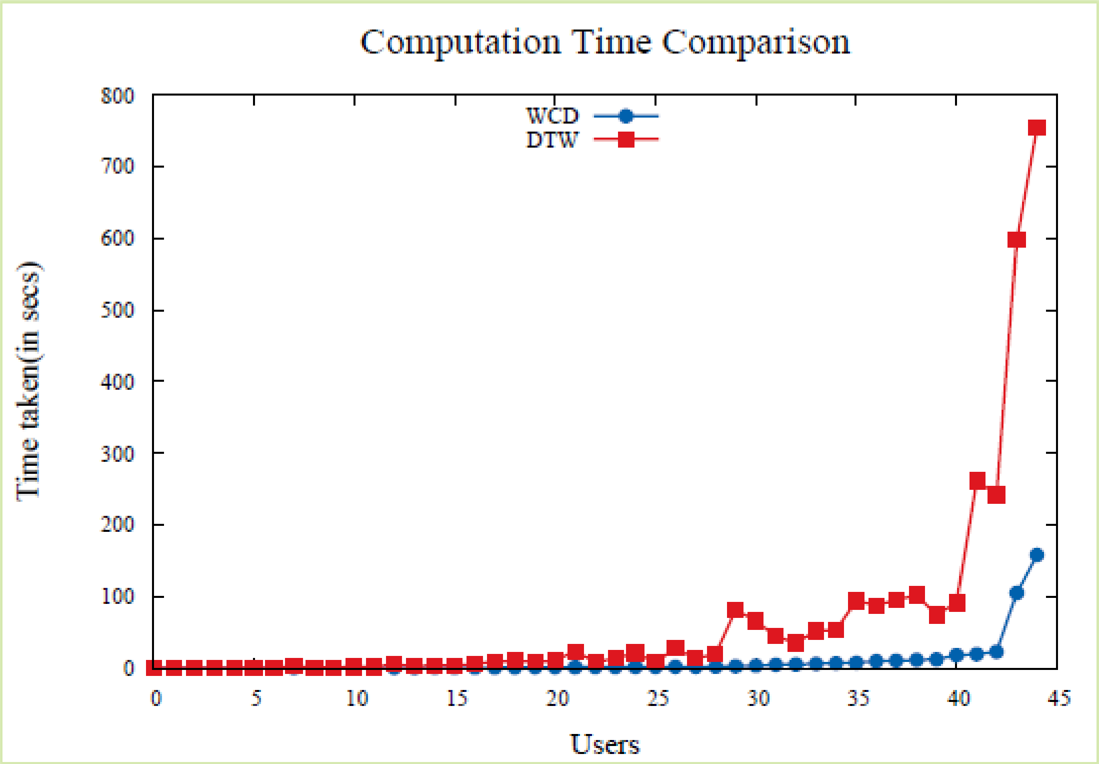
Experimentation

COMPARISONS WITH DTW

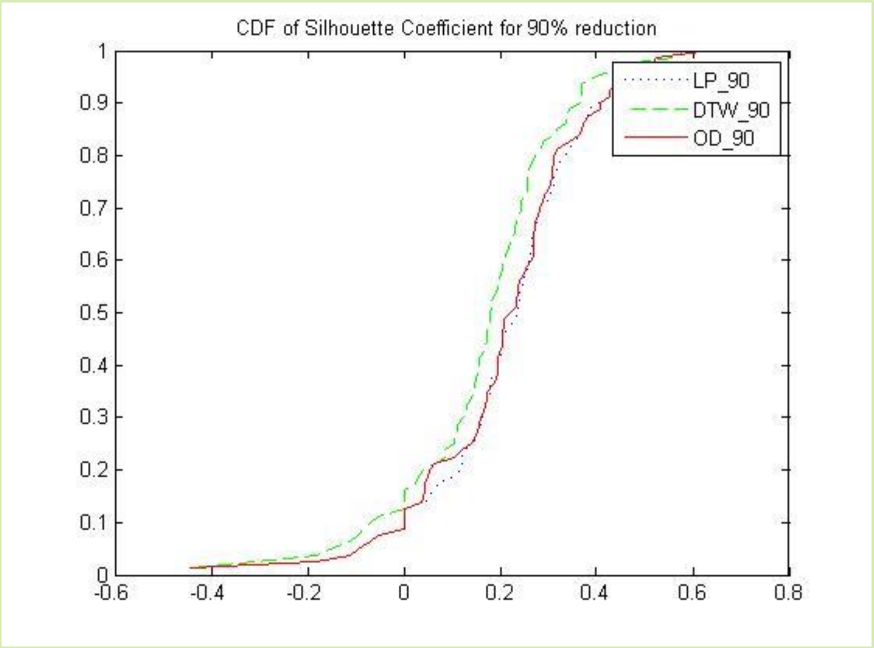
Comparisons with DTW Similarity

- DTW is not a metric as it violates triangle inequality –
  - Might affect the clusters formed in hierarchical clustering
- DTW is computationally much more expensive
- As the noise increases, effectiveness of DTW goes down much faster when compared to proposed method.

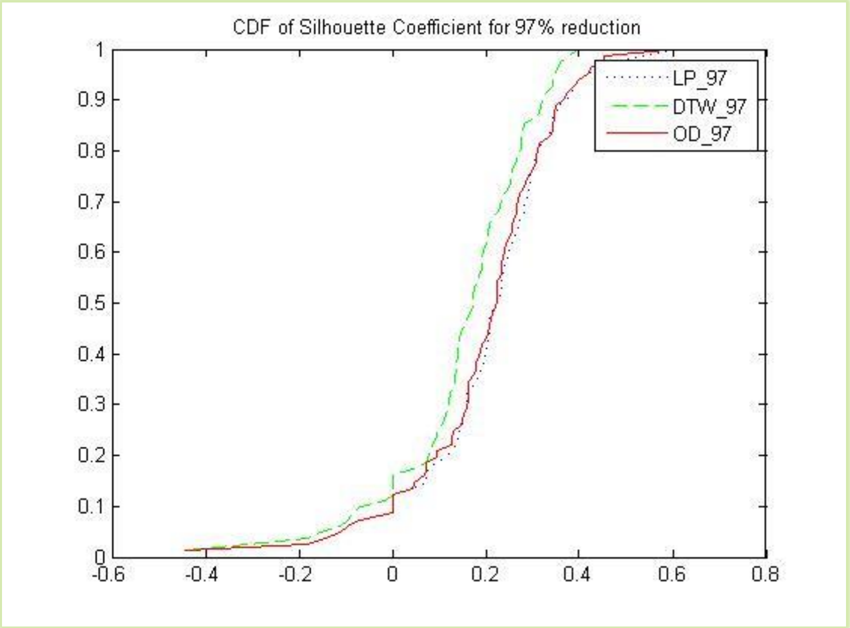
Comparisons with DTW Similarity



Comparisons with DTW Similarity



CDF for SC with 90% reduction in sample points



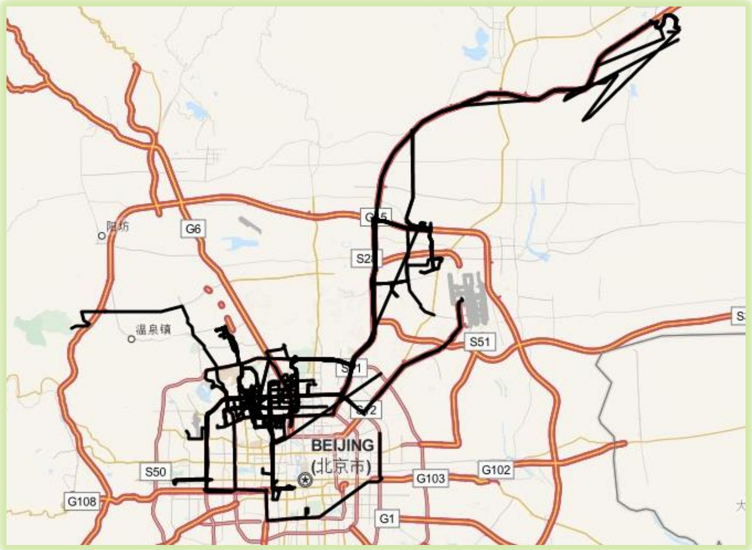
CDF for SC with 97% reduction in sample points

Experimentation

PROPOSED METHOD- A VISUAL JOURNEY…

Dataset

- Geolife Dataset –
  - 178 users
  - Over a period of over four years (from April 2007 to October 2011)
  - 17,621 trajectories
  - Around Beijing



All Trajectories

Cluster 1

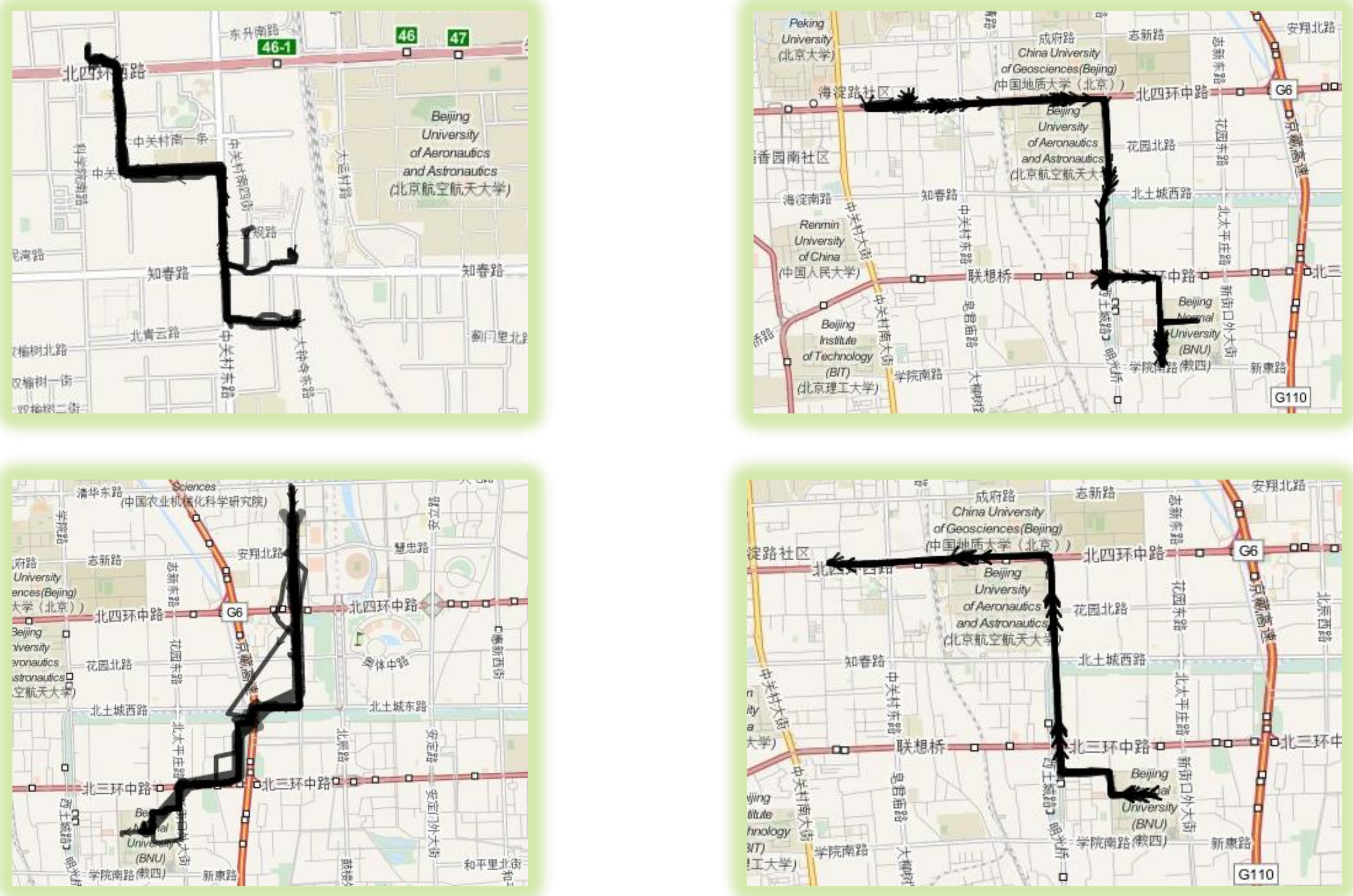
Cluster 2

Cluster 3

Cluster 4

Visualizations at Elbow Point

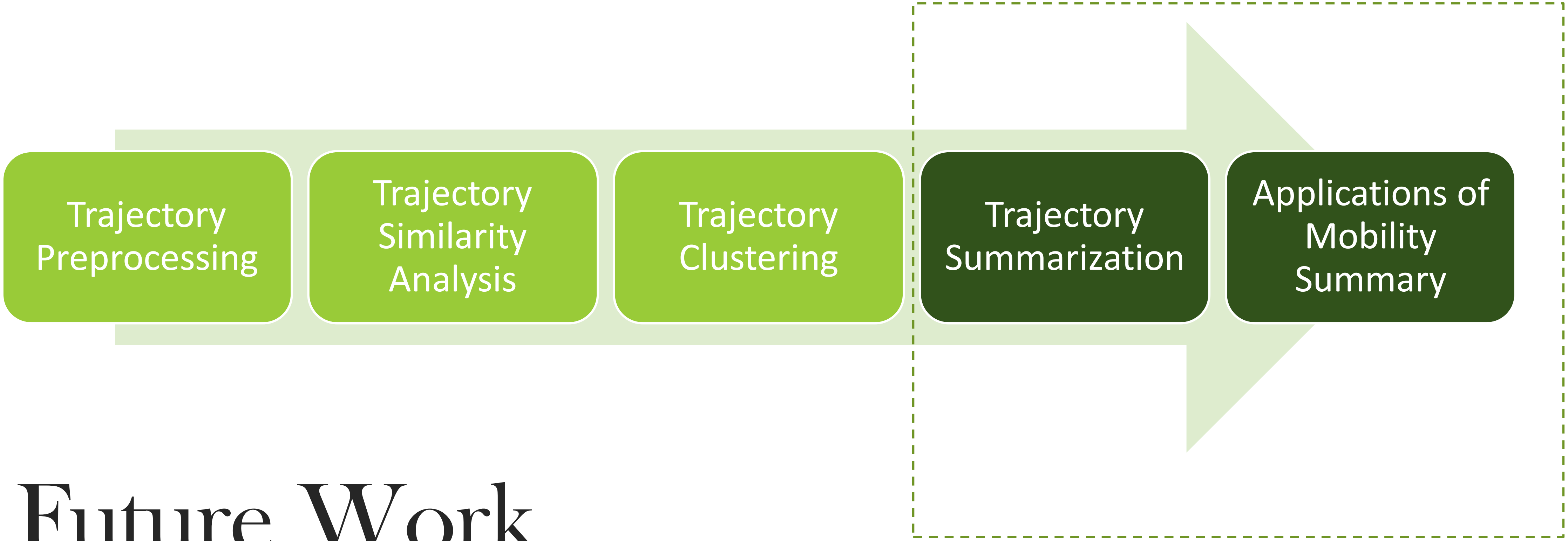
Final Clusters



Conclusion

Conclusion

- From a raw source of GPS traces, aim is to find the mobility summary of the individual
- The steps involved were Preprocessing, defining a Similarity, and Clustering.
- The mobility summary of a person can be useful for various application such as
  - Next Path Prediction
  - Anomaly Detection
- Future Work involves
  - Coming up with a better heuristic to find optimal number of clusters
  - Trajectory Summarization
  - Storing the Trajectory Summaries
  - Applications of mobility summary


Trajectory Preprocessing
Trajectory Similarity Analysis
Trajectory Clustering
Trajectory Summarization
Applications of Mobility Summary

Future Work

Thank You!

References

1. R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. 1994.
2. D. J. Berndt and J. Clifford. Finding patterns in time series, advances in knowledge discovery and data mining. In The Thirteenth National Conference on Artificial Intelligence,AAAI, Portland, Oregon, August 4- 8, 1996. MIT Press.
3. K. Buchin, M. Buchin, J. Gudmundsson, M. L¨offler, and J. Luo. Detecting commuting patterns by clustering subtrajectories. International Journal of Computational Geometry & Applications, 21(03):253–282, 2011.
4. K. Buchin, M. Buchin, M. Van Kreveld, M. L¨offler, R. I. Silveira, C. Wenk, and L. Wiratma. Median trajectories. Algorithmica, 66(3):595–614, 2013.
5. L. Chen and R. Ng. On the marriage of lp-norms and edit distance. In Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30, VLDB ’04, pages 792–803, Toronto, Canada, 29 August - 3 September 2004. VLDB Endowment.
6. L. Chen, M. T. O¨ zsu, and V. Oria. Robust and fast similarity search for moving object trajectories. In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, SIGMOD ’05, pages 491–502, Baltimore, Maryland, June 13-17, 2005. ACM.
7. S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 63–72, San Diego, CA, USA, August 15 - 18,1999. ACM.
8. J. Gudmundsson, M. van Kreveld, and B. Speckmann. Efficient detection of patterns in 2d trajectories of moving points. Geoinformatica, 11(2):195–215, 2007.
9. B. Han, L. Liu, and E. Omiecinski. Neat: Road network aware trajectory clustering. In Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on, pages 142–151, Macau,China, 18-21 June,2012. IEEE.

10. S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding. Statistical modeling of curves using shapes and related features. Journal of the American Statistical Association, 107(499):1152–1165, 2012.
11. J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: A partition-and-group framework. In Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, SIGMOD '07, pages 593–604, Beijing, China, June 11 - 14, 2007. ACM.
12. Q. Li, Y. Zheng, X. Xie, Y. Chen,W. Liu, andW.-Y. Ma. Mining user similarity based on location history. In proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, Irvine, CA, USA, November 5-7, 2008. ACM.
13. Z. Li, B. Ding, J. Han, and R. Kays. Swarm: Mining relaxed temporal moving object clusters. Proc. VLDB Endow., 3(1-2):723–734, Sept. 2010.
14. M. K. E. Mahrsi and F. Rossi. Modularity-based clustering for network-constrained trajectories. arXiv preprint arXiv:1205.2172, 2012.
15. M. Nanni and D. Pedreschi. Time-focused clustering of trajectories of moving objects. Journal of Intelligent Information Systems, 27(3):267–289, 2006.
16. H. Su, K. Zheng, K. Zeng, J. Huang, S. Sadiq, N. J. Yuan, and X. Zhou. Making sense of trajectory data: A partition-and-summarization approach. In IEEE 31st International Conference on Data Engineering (ICDE), pages 963–974, Seoul, South Korea, 13 Apr - 17 Apr, 2015. IEEE.
17. M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. In Proceedings of the 18th International Conference on Data Engineering, pages 673–684, San Jose, CA, 26 February - 1 March, 2002.

18. H.Wang, H. Su, K. Zheng, S. Sadiq, and X. Zhou. An effectiveness study on trajectory similarity measures. In Proceedings of the Twenty-Fourth Australasian Database Conference-Volume 137, pages 13–22, Adelaide, Australia, 2013. Australian Computer Society, Inc.
19. B.-K. Yi and C. Faloutsos. Fast time sequence indexing for arbitrary lp norms. Cairo, Egypt, September 10-14, 2000. VLDB.
20. B.-K. Yi, H. Jagadish, and C. Faloutsos. Efficient retrieval of similar time sequences under time warping. In proceedings of the Fourteenth International Conference on Data Engineering, ICDE, pages 201–208, Orlando, Florida, USA, February 23-27, 1998.
21. J. Yuan, Y. Zheng, X. Xie, and G. Sun. T-drive: Enhancing driving directions with taxi drivers' intelligence. Knowledge and Data Engineering, IEEE Transactions on, 25(1):220–232, Jan 2013.
22. P. Zhang, M. Deng, and N. Van deWeghe. Clustering spatio-temporal trajectories based on kernel density estimation. In Computational Science and Its Applications–ICCSA 2014, pages 298–311. Springer, Guimaraes, Portugal, June 30 - July 3, 2014.
23. Y. Zheng. Trajectory data mining: an overview. ACM Transactions on Intelligent Systems and Technology (TIST), 6(3):29, 2015.
24. Y. Zheng, Y. Chen, X. Xie, and W.-Y. Ma. Geolife2.0: A location-based social networking service. In Mobile Data Management: Systems, Services and Middleware, 2009. MDM '09. Tenth International Conference on, pages 357–358, Taipei, Taiwan, May May 18- May 21, 2009.
25. Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08, pages 312–321, New York, NY, USA, 2008. ACM.

18. H.Wang, H. Su, K. Zheng, S. Sadiq, and X. Zhou. An effectiveness study on trajectory similarity measures. In Proceedings of the Twenty-Fourth Australasian Database Conference-Volume 137, pages 13–22, Adelaide, Australia, 2013. Australian Computer Society, Inc.
19. B.-K. Yi and C. Faloutsos. Fast time sequence indexing for arbitrary lp norms. Cairo, Egypt, September 10-14, 2000. VLDB.
20. B.-K. Yi, H. Jagadish, and C. Faloutsos. Efficient retrieval of similar time sequences under time warping. In proceedings of the Fourteenth International Conference on Data Engineering, ICDE, pages 201–208, Orlando, Florida, USA, February 23-27, 1998.
21. J. Yuan, Y. Zheng, X. Xie, and G. Sun. T-drive: Enhancing driving directions with taxi drivers' intelligence. Knowledge and Data Engineering, IEEE Transactions on, 25(1):220–232, Jan 2013.
22. P. Zhang, M. Deng, and N. Van deWeghe. Clustering spatio-temporal trajectories based on kernel density estimation. In Computational Science and Its Applications–ICCSA 2014, pages 298–311. Springer, Guimaraes, Portugal, June 30 - July 3, 2014.
23. Y. Zheng. Trajectory data mining: an overview. ACM Transactions on Intelligent Systems and Technology (TIST), 6(3):29, 2015.
24. Y. Zheng, Y. Chen, X. Xie, and W.-Y. Ma. Geolife2.0: A location-based social networking service. In Mobile Data Management: Systems, Services and Middleware, 2009. MDM '09. Tenth International Conference on, pages 357–358, Taipei, Taiwan, May May 18- May 21, 2009.
25. Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08, pages 312–321, New York, NY, USA, 2008. ACM.

26. Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on GPS data. In Proceedings of the 10th international conference on Ubiquitous computing, pages 312–321, Seoul, South Korea, September 21 - 24, 2008. ACM.
27. Y. Zheng and X. Xie. Learning travel recommendations from user-generated GPS traces. ACM Transactions on Intelligent Systems and Technology (TIST), 2(1):2, 2011.
28. Y. Zheng, X. Xie, and W.-Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. IEEE Data Eng. Bull., 33(2):32–39, 2010.
29. Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of the 18th international conference on World wide web, pages 791–800, Madrid, Spain, April 20 - 24, 2009. ACM.