All-Scale Trajectory Clustering for Moving Behavior Detection with Spatiotemporal Recurrent Convolutional Neural Networks[1]

Yao-Yi Chiang[*]
Associate Professor
Computer Science and Engineering
University of Minnesota
yaoyi@umn.edu

[1] An NGA Boosting Innovative GEOINT (BIG) project

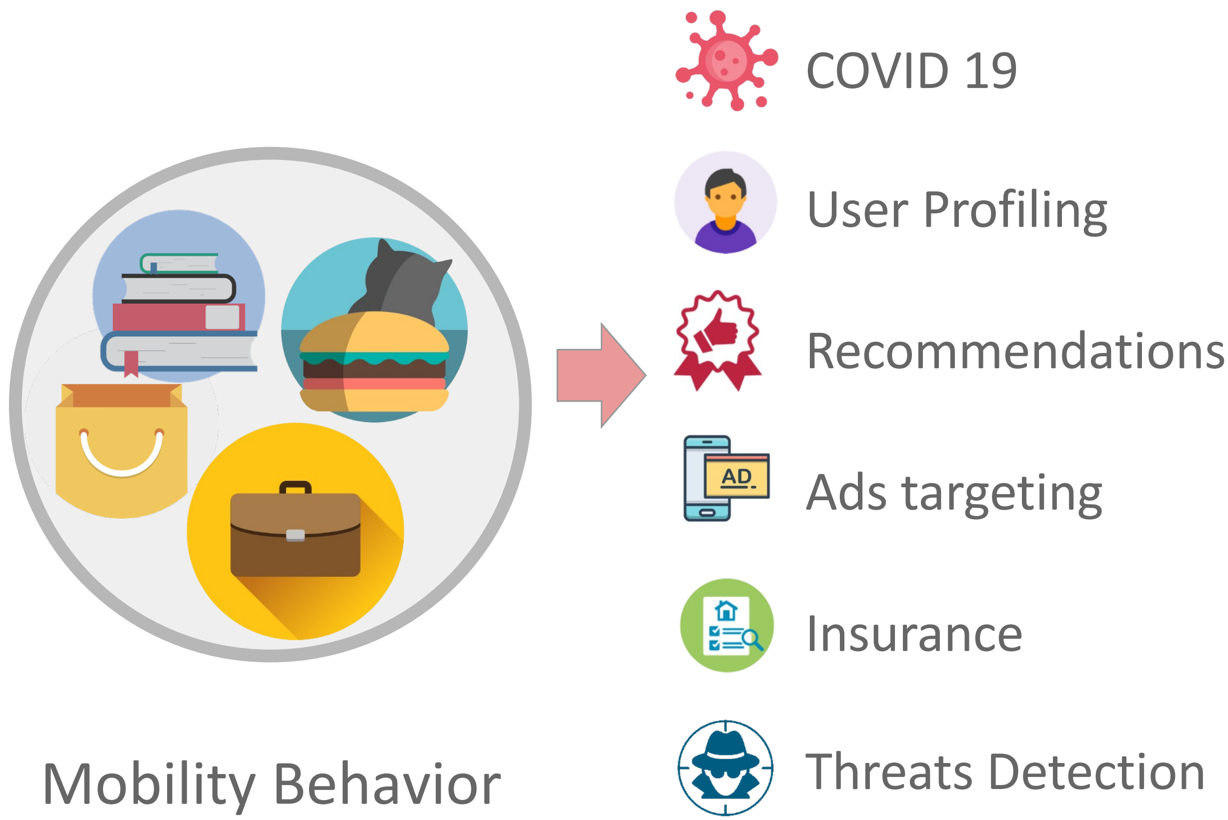
[*] With Cyrus Shahabi and Mingxuan Yue, University of Southern California; slides aopted from Mingxuan Yue

OUTLINE

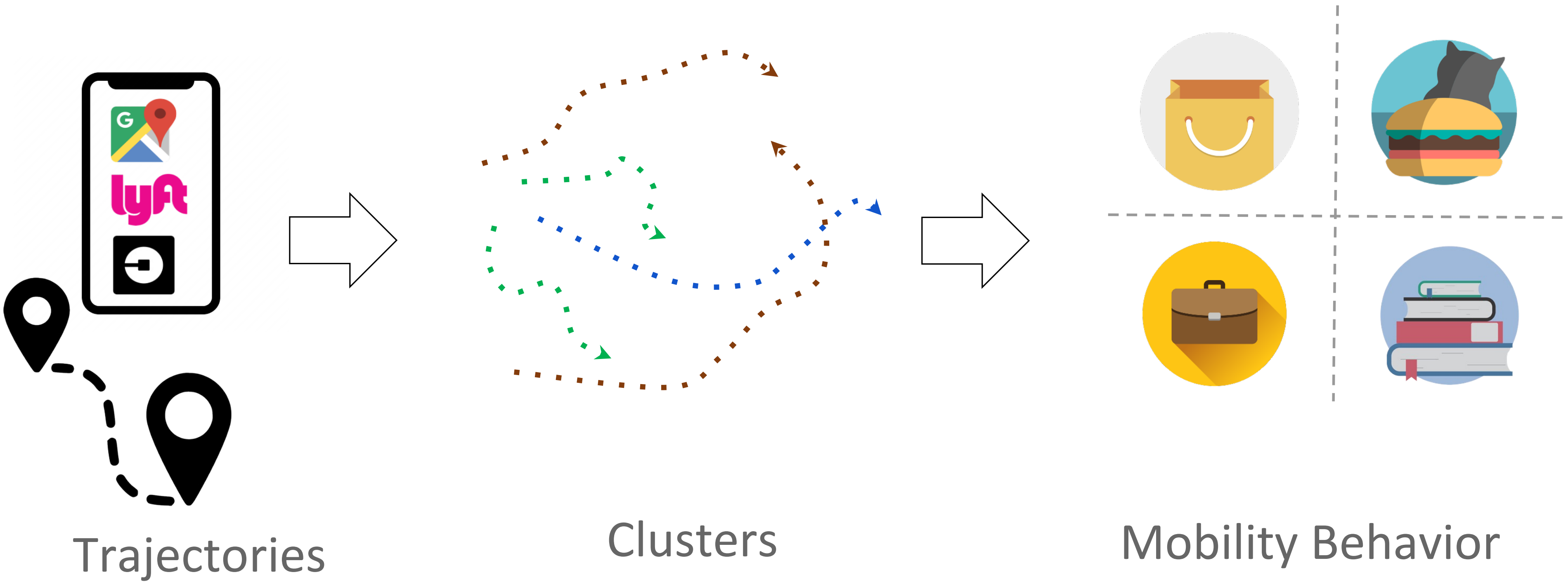
- Overview
- Approach: DETECT
  - Convert trajectories to sequences of contexts
  - Fixed-size representation (embedding) with RNN
  - Embedding Clustering and Optimization
- Experiments
- Future Work

Motivation

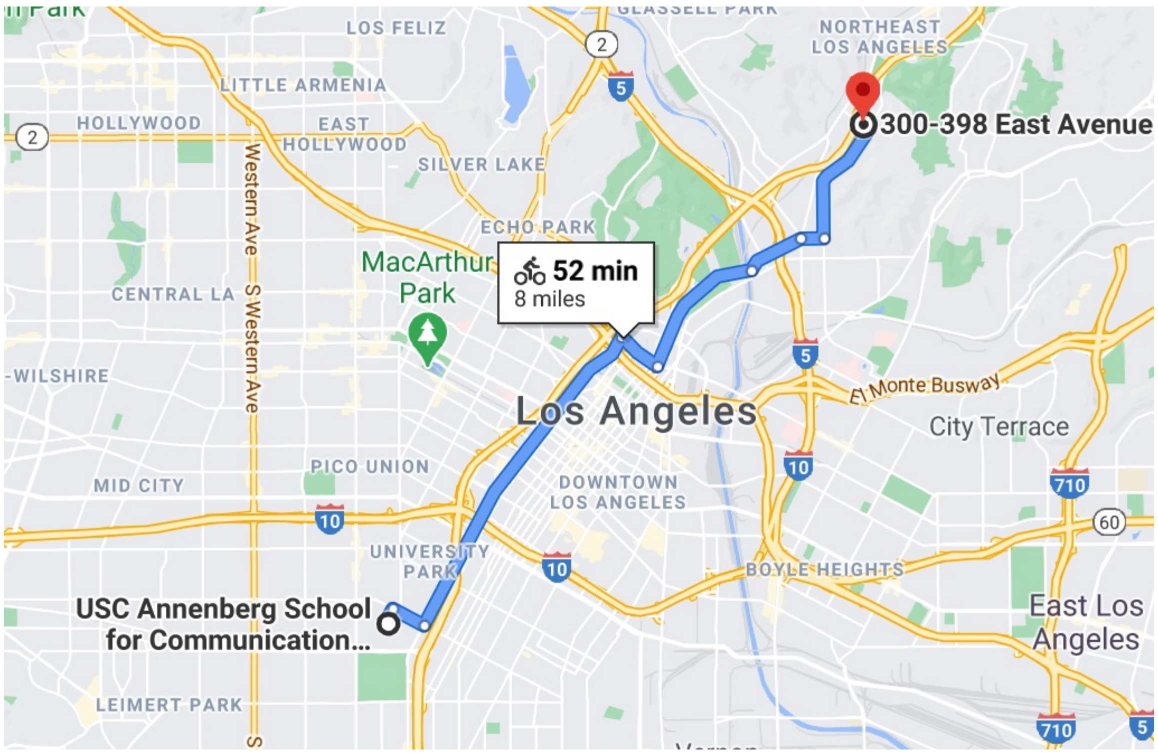
Mobility behavior:
travel activity describing a user's movements, e.g., work commute, shopping, school commute, dining



Idea



Challenge: Multi-scale Trajectories

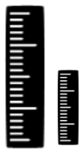


- Various temporal and spatial scales may represent the same mobility behavior
- 50 minutes work commute:
  - 14 miles, 44 miles, 8 miles
- 14 miles work commute
  - 20 min, 50 min, 1.5 hour

Trajectory Clustering Techniques

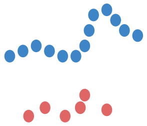
- Based on similarity of raw spatiotemporal features **[AIR'17]**
- Sequence distance measurement
  - Dynamic Time Warping (**DTW**), Longest Common SubSequence (**LCSS**)
- Clustering based on the distances
  - kMeans-DBA **[ICDM'14]**, DBSCAN **[CVPR'09]**, Hierarchical Clustering

Limitations of Traditional Trajectory Clustering Techniques



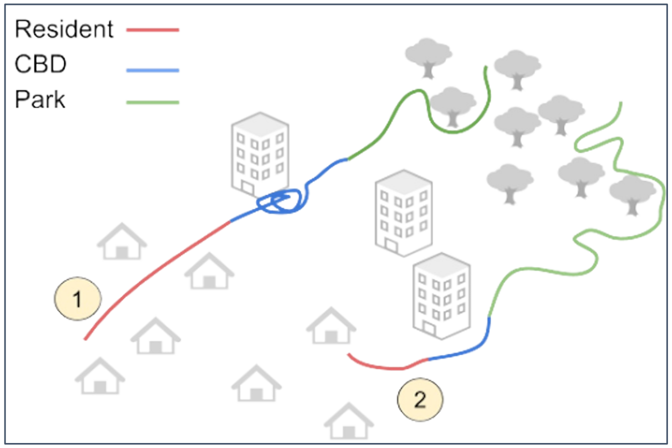
Prone to scales & noises

No activity context information

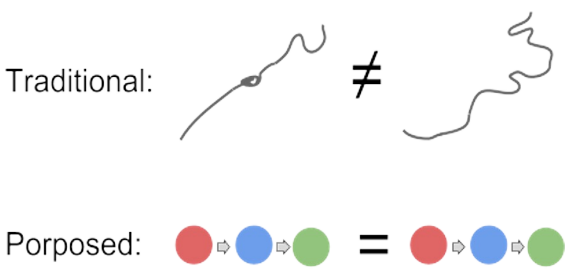


Pre-defined similarity vs. data-driven

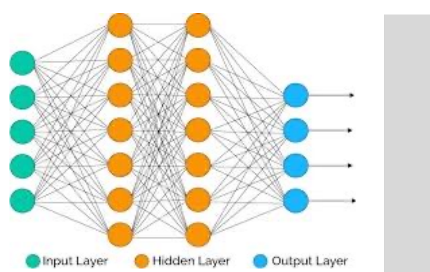
DETECT: Deep Trajectory Clustering for Mobility-Behavior Analysis



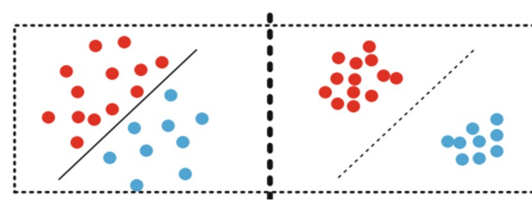
- Address varying spatial & temporal **scales**
- Aware of **geographical context**
- Work for variable lengths of sequences
- Learn useful properties driven by data
- Avoid expensive manual labeling



Convert trajectories to sequences of contexts



Fixed-size representation (embedding) with RNN



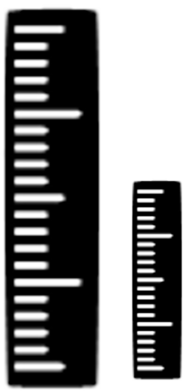
Embedding Clustering and Optimization

OUTLINE

- Overview
- Approach: DETECT
  - Convert trajectories to sequences of contexts
  - Fixed-size representation (embedding) with RNN
  - Embedding Clustering and Optimization
- Experiments
- Future Work

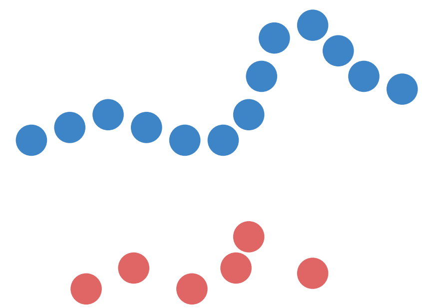
Approach

DETECT **[BigData 19]**



All-scale

Context-aware



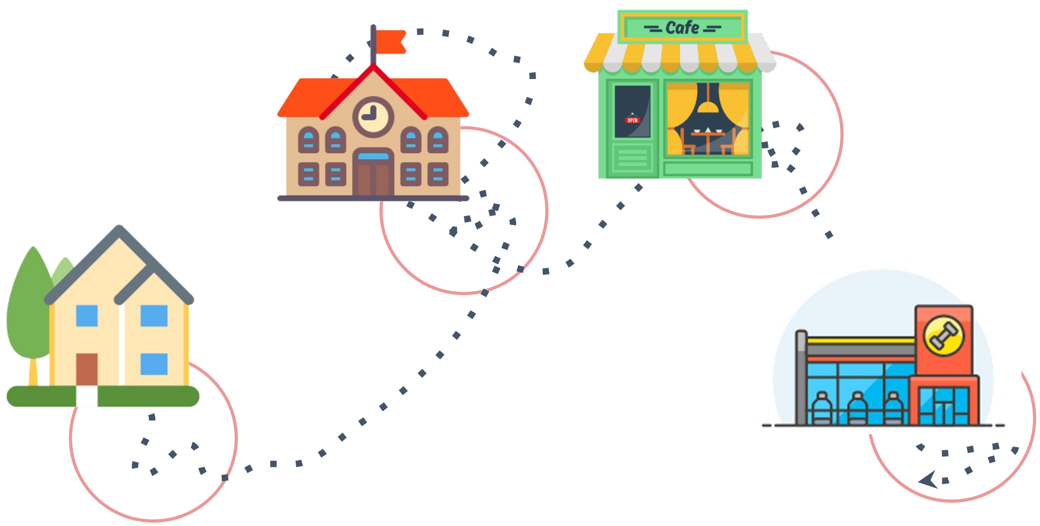
Sequence Dynamics

Yue, M., Li, Y., Yang, H., Ahuja, R., **Chiang, Y.-Y.**, and Shahabi, C. (December 2019). DETECT: Deep Trajectory Clustering for Mobility-Behavior Analysis. In *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*, pp. 988–997, Los Angeles, CA, USA

All-scale: Stay Points

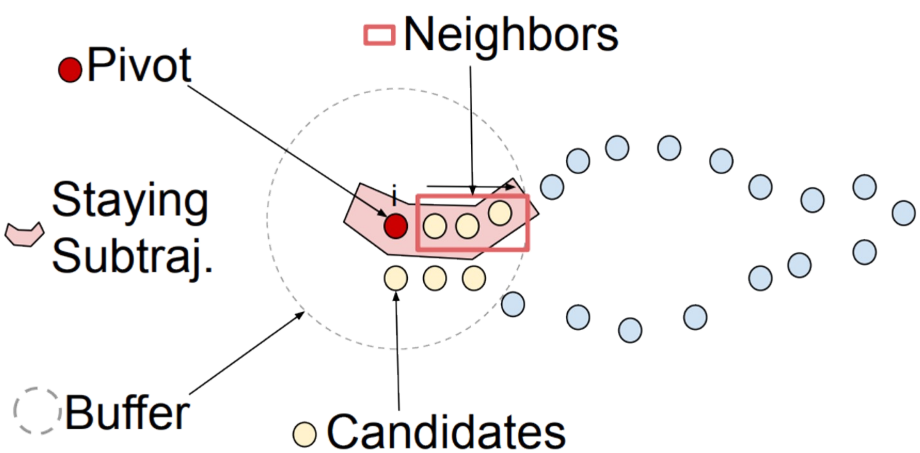
Stay points **[SIGSPATIAL'08]** are meaningful locations where:

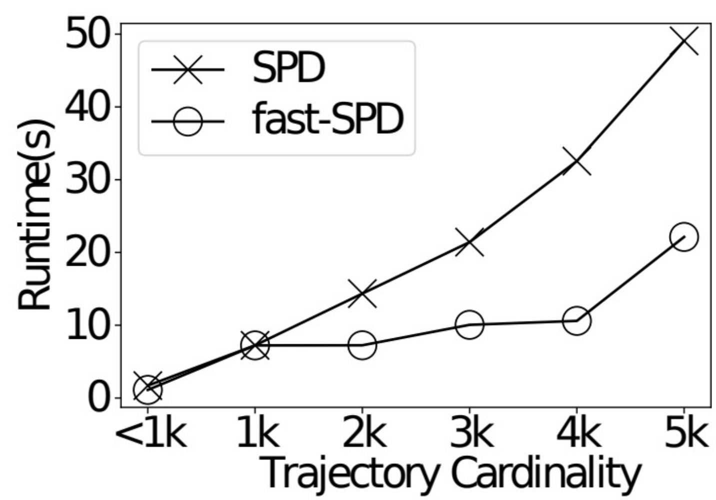
1. the user travels within a small range of space
2. the user stays in this range for some time



All-scale: Fast Stay Point Extraction

- Fast-SPD:
  - Scan each trajectory to find consecutive sub-trajectories that the user travel within a limited range but stay for a long time
  - Use the centers of such sub-trajectories as stay points





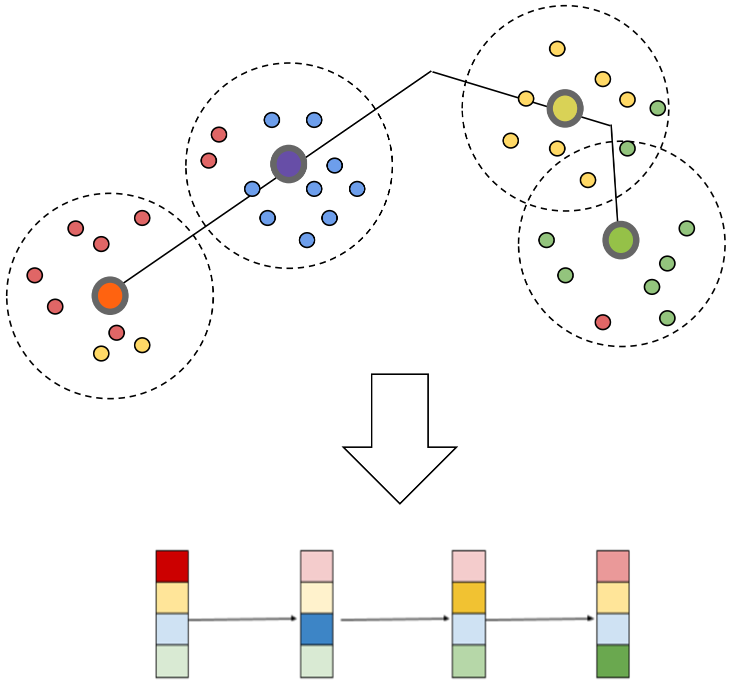
Context-aware: Geographical Augmentation

For each extracted stay point $\dot{s}^{(t)}$:

1. create a spatial buffer $b(r_{poi}, \dot{s}^{(t)})$
2. search a gazetteer for POI's in the buffer
3. count POIs in the buffer
4. generate a normalized vector

$$x^{(t)} = \{0.3, 0.09, ...0.55\}$$

Normalized number of POI categories, e.g., business area

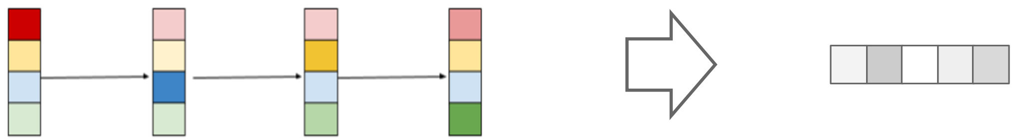


OUTLINE

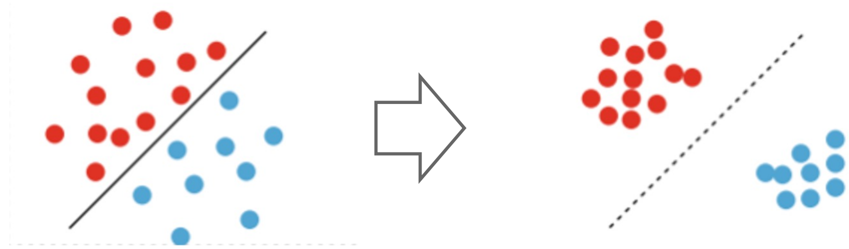
- Overview
- Approach: DETECT
  - Convert trajectories to sequences of contexts
  - Fixed-size representation (embedding) with RNN
  - Embedding Clustering and Optimization
- Experiments
- Future Work

Sequence Dynamics: RNN-AE + Clustering

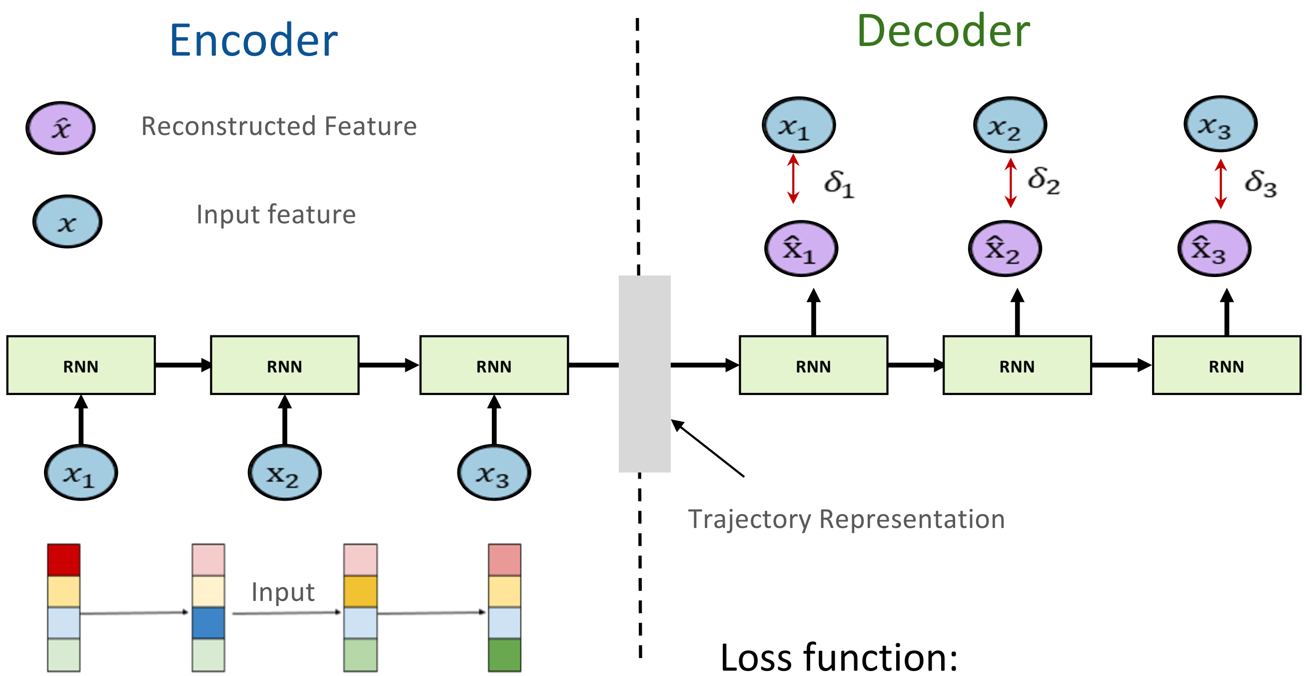
Phase I:



Phase II:



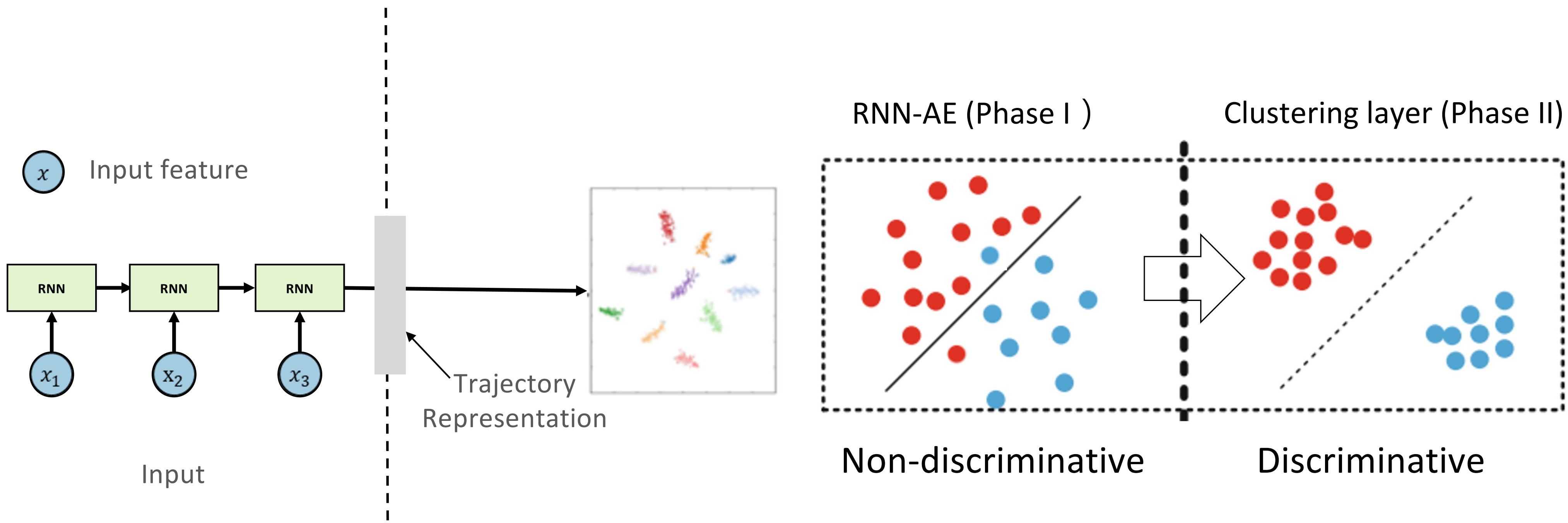
Phase I: RNN Autoencoder



Loss function:

$$\ell = \frac{1}{T}\sum_{t=1}^{T}(x_i^{(t)} - \hat{x}_i^{(t)})^2$$

Phase II: Refine for clean clusters



Reconstruction

Cluster-aware

Phase II: Unsupervised clustering

p_{ij} and q_{ij} could be interpreted as the probability of trajectory *i* is assigned to cluster *j*

Current t-distribution:

$$q_{ij} = \frac{(1+||z_i-\mu_j||^2)^{-1}}{\sum_{j'}(1+||z_i-\mu_{j'}||^2)^{-1}}$$

Auxiliary distribution:

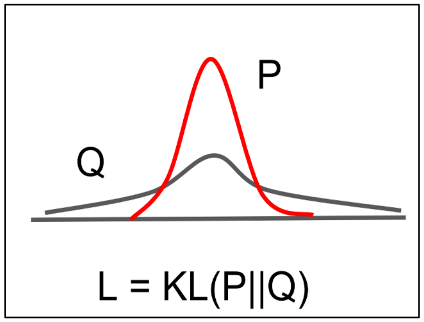
$$p_{ij} = \frac{\boxed{q_{ij}^2} / \sum_{i'} q_{i'j}}{\sum_{j'}(q_{ij'}^2 / \sum_{i'} q_{i'j'})}$$

Loss function:

$$\ell = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

If q_{ij} is small, q_{ij}^2 will be even smaller

- punish uncertain cluster assignments
- high certain cluster assignments remain high



Minimizing the KL distance to compact the clusters

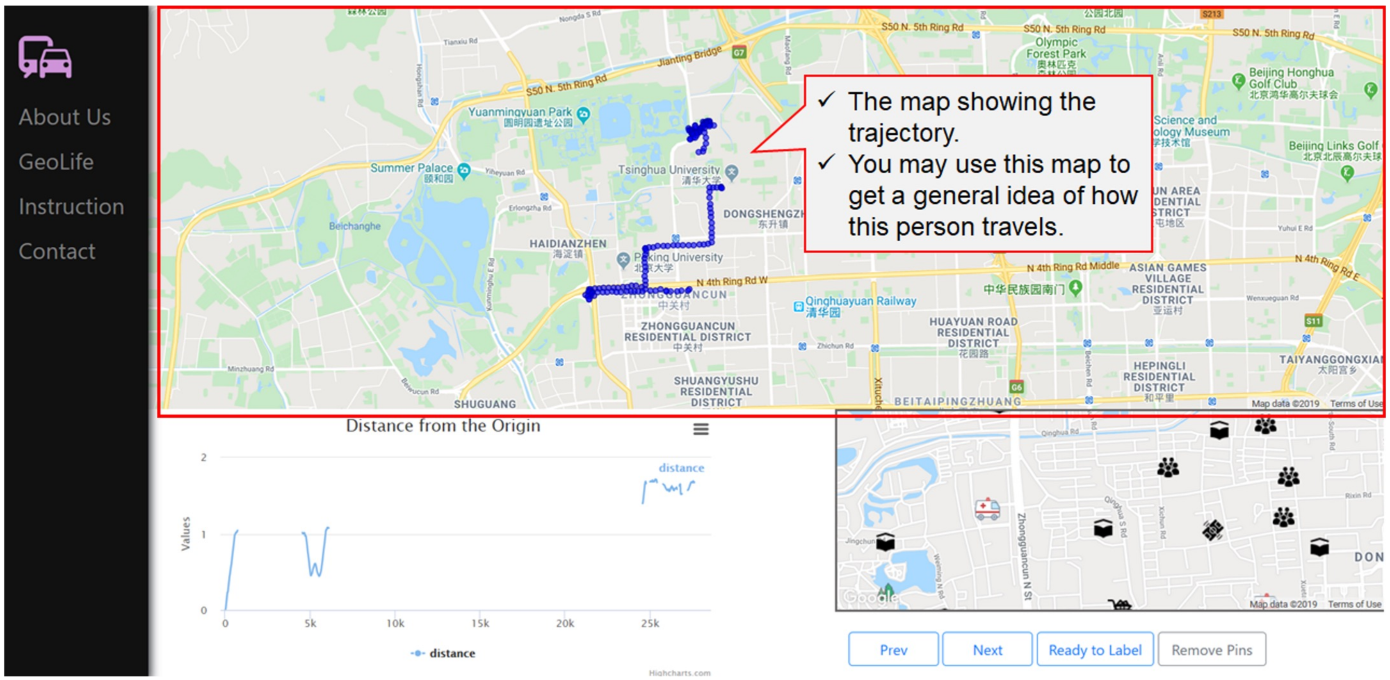
OUTLINE

- Overview
- Approach: DETECT
  - Convert trajectories to sequences of contexts
  - Fixed-size representation (embedding) with RNN
  - Embedding Clustering and Optimization
- Experiments
- Future Work

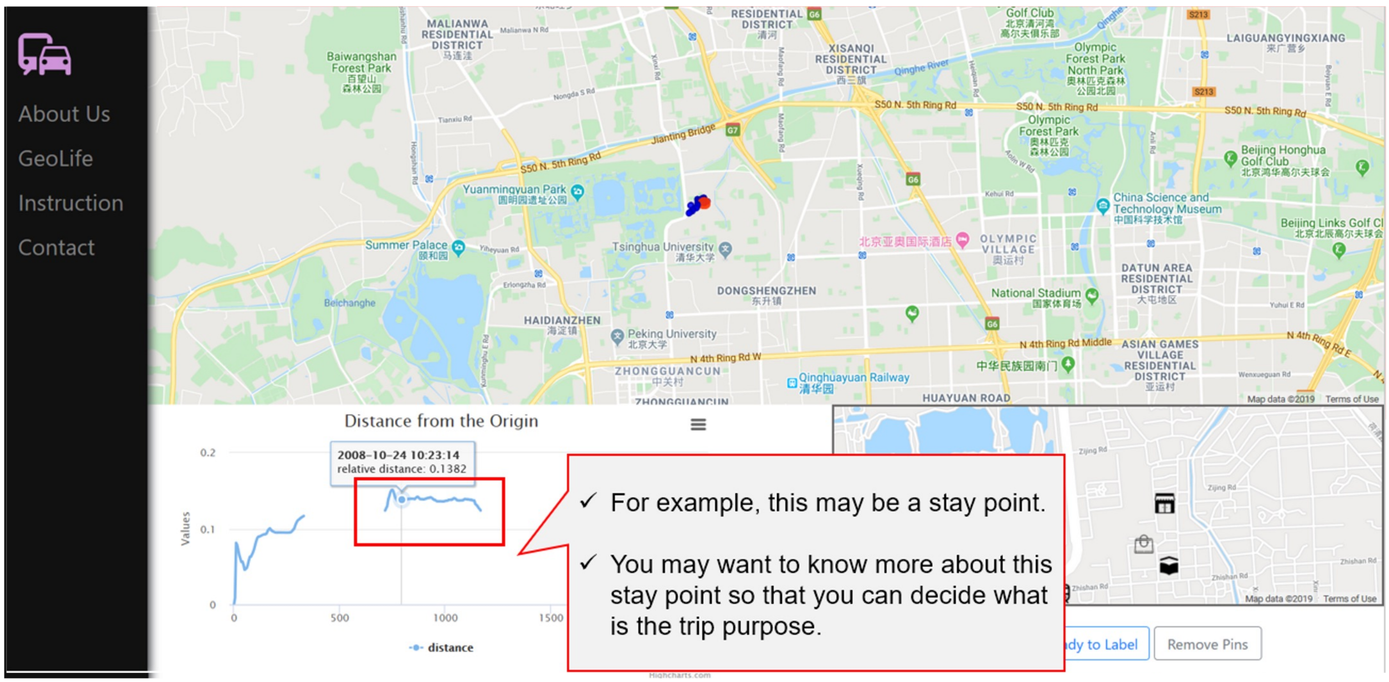
Experimental settings

- Dataset: GeoLife
  - 17,621 trajectories (601 labeled).
  - 6 labels: “dining activities”, “working commutes”, etc.
  - 14,000 POIs in Beijing
- Evaluation Metrics
  - With label: Rand Index (RI), Mutual Information (MI), Purity Fowlkes-Mallows Index (FMI)
  - Without label: Silhouette Score, Dunn index, Within-like Criterion, Between-like Criterion

Labeling Platform



Labeling Platform



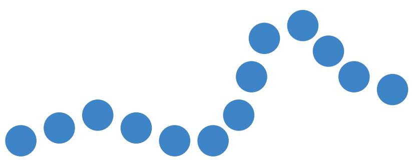
With-label: quantitative results

Distance		Clustering
DTW		K-Means
LCSS	+	DBSCAN
SSPD		Hierarchical clustering

Method	RI	MI	Purity	FMI
KM-DBA	0.33	0.64	0.58	0.58
DB-LCSS	0.22	0.55	0.51	0.56
RNN-AE	0.39	0.46	0.56	0.53
SSPD-HCA	0.52	0.93	0.66	0.67
KM-DBA*	0.51	0.91	0.74	0.63
DB-LCSS*	0.5	0.95	0.64	0.66
DETECT Phase I	0.65	1.06	0.84	0.73
DETECT	**0.76**	**1.26**	**0.89**	**0.81**

RI (Rand Index) (0,1)
MI (Mutual Information) (0, inf)
Purity (0,1)
FMI (Fowlkes-Mallows Index) (0,1)
Higher values for all metrics mean better results.

With-label: quantitative results



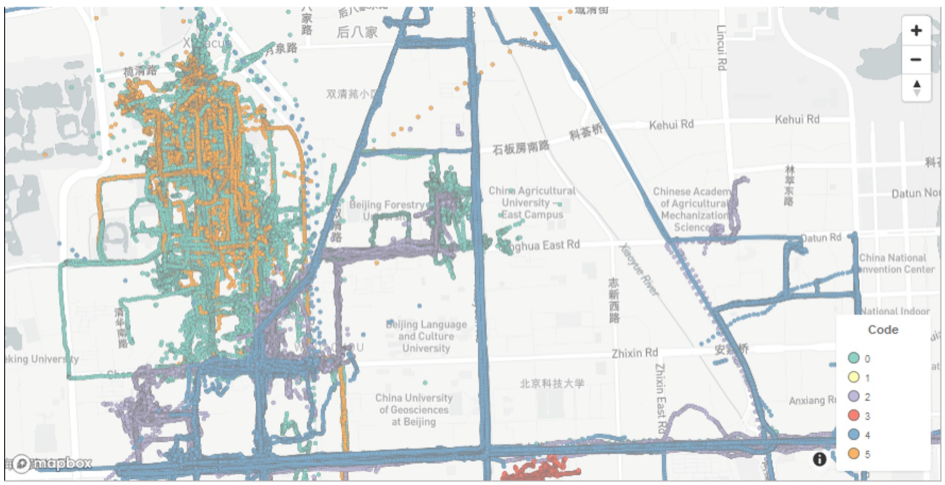
Raw trajectories

Augmented trajectories

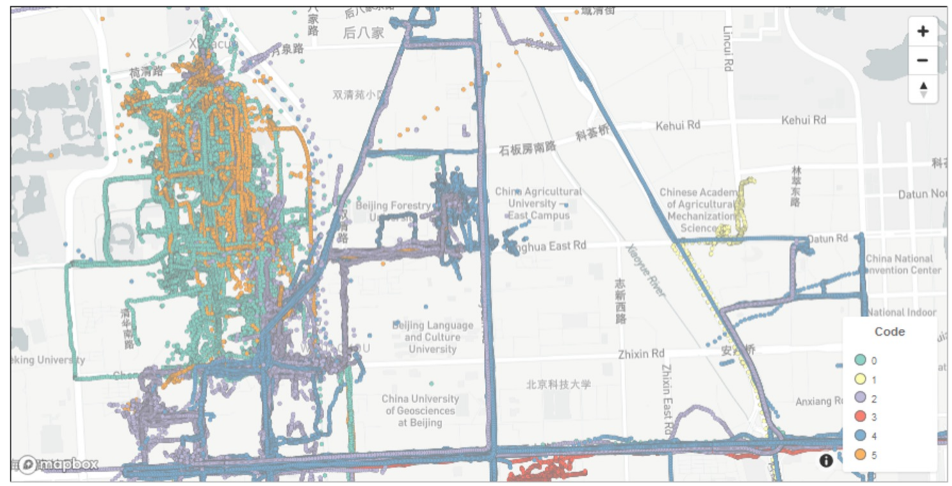
RI (Rand Index) (0,1)
MI (Mutual Information) (0, inf)
Purity (0,1)
FMI (Fowlkes-Mallows Index) (0,1)
Higher values for all metrics mean better results.

Method	RI	MI	Purity	FMI
KM-DBA	0.33	0.64	0.58	0.58
DB-LCSS	0.22	0.55	0.51	0.56
RNN-AE	0.39	0.46	0.56	0.53
SSPD-HCA	0.52	0.93	0.66	0.67
KM-DBA*	0.51	0.91	0.74	0.63
DB-LCSS*	0.5	0.95	0.64	0.66
DETECT Phase I	0.65	1.06	0.84	0.73
DETECT	**0.76**	**1.26**	**0.89**	**0.81**

With-label: qualitative results



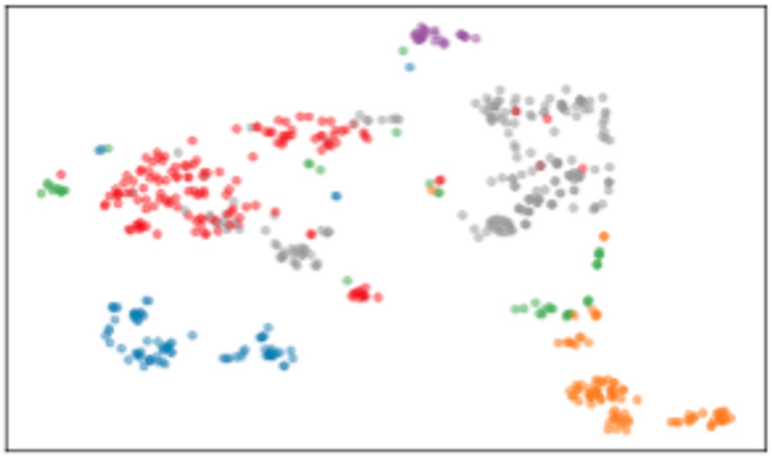
Ground Truth



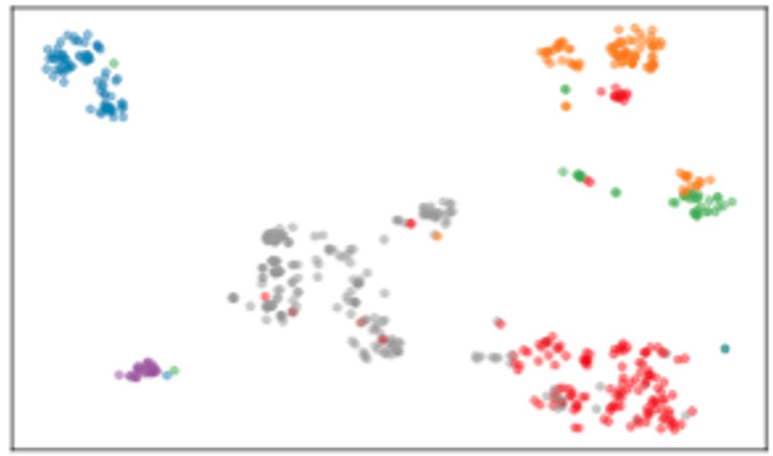
Our Results

Colors indicate different clusters.

With-label: qualitative results

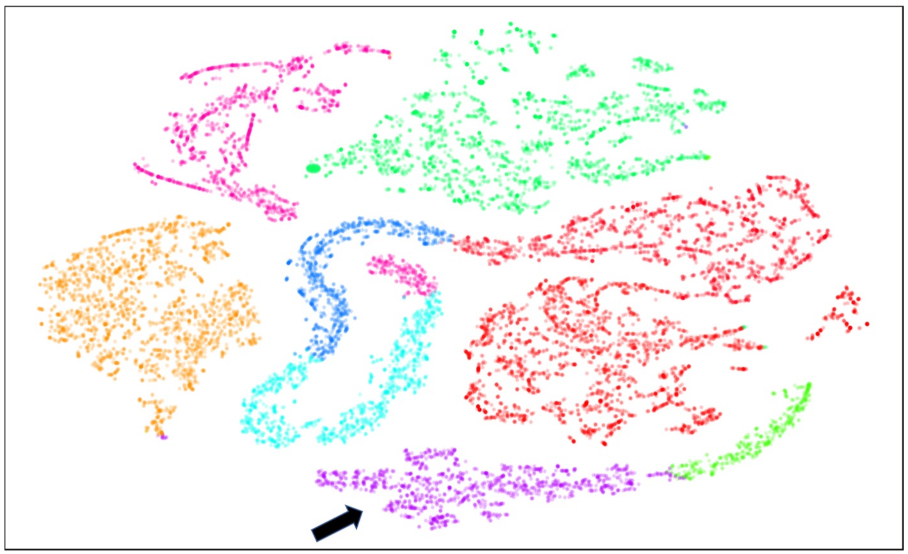


Embedding after Phase I

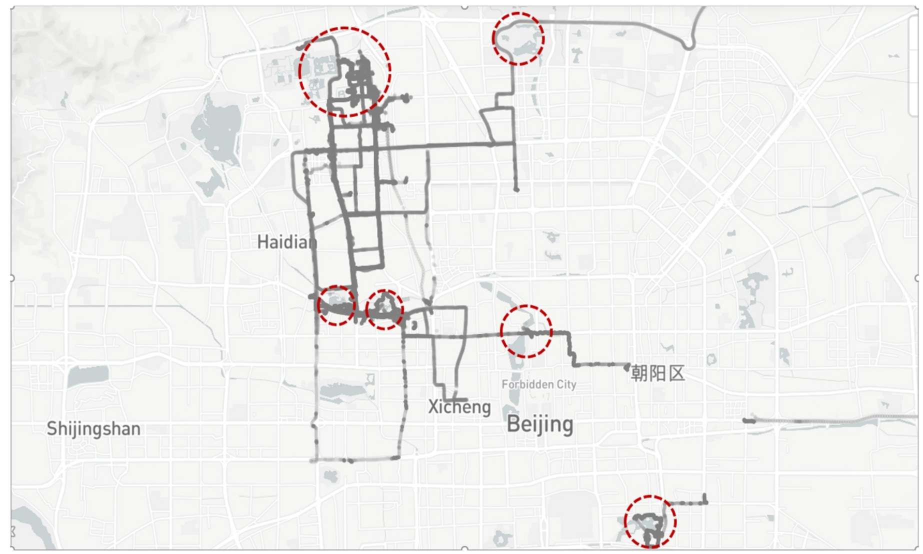


Embedding after Phase II

Without-label: qualitative results



Embedding of the full dataset

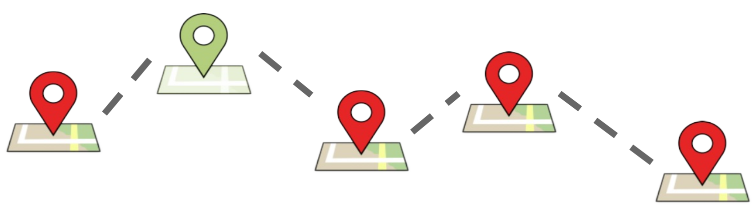


Recreation Activities

DETECT Extension: context learning

- What if we don't have a gazetteer for the area, e.g., boat trajectories?
- Idea: Learn the context from trajectories. **[ECML 20]**

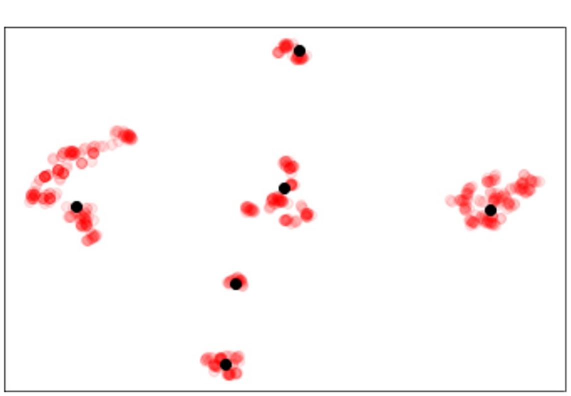
Sequence of words



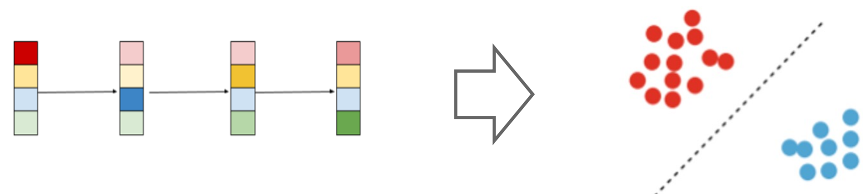
Sequence of locations

Also, one-phase generative model

- Using a generative model to directly learn a cluster-aware hidden space rather than a 2-phase procedure
- Can also be used for synthetic trajectory creation and anomaly detection



Embedding by generative model

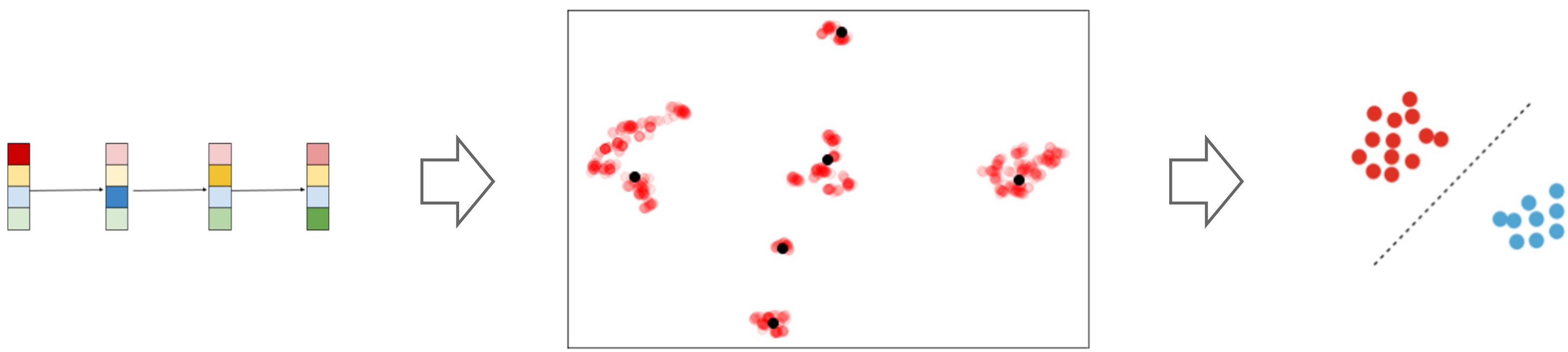


Directly learns a cluster-aware hidden space

Interpolation between two augmented trajectories

Future Work: Explainability

- How to generate meaningful (explainable) embeddings to explain the clustering results
  - provide semantic meanings of individual clusters
  - understand outlier trajectories



References

[BigData 19] Yue M, Li Y, Yang H, Ahuja R, **Chiang YY**, **Shahabi C**. DETECT: Deep Trajectory Clustering for Mobility-Behavior Analysis. In Big Data 2019.
[ECML 20] Yue M, Sun T, Wu F, Wu L, Xu Y, **Shahabi C**, Learning a Contextual and Topological Representation of Areas-of-Interest for On-Demand Delivery Application, ECML-PKDD 2020
[ITS 16] Besse, Philippe C., et al. "Review and perspective for distance-based clustering of vehicle trajectories." IEEE Transactions on Intelligent Transportation Systems 17.11 (2016): 3306-3317.
[AIR 17] Yuan, Guan, et al. "A review of moving object trajectory clustering algorithms." Artificial Intelligence Review 47.1 (2017): 123-144.
[ICDM 14] Petitjean, François, et al. "Dynamic time warping averaging of time series allows faster and more accurate classification." 2014 IEEE international conference on data mining. IEEE, 2014.
[CVPR 09] Morris, Brendan, and Mohan Trivedi. "Learning trajectory patterns by clustering: Experimental studies and comparative evaluation." 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009.
[SIGSPATIAL 08] Li, Quannan, et al. "Mining user similarity based on location history." Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. ACM, 2008.

Acknowledgements

- Gil, Yolanda (Ed.) Introduction to Computational Thinking and Data Science. Available from http://www.datascience4all.org