ETH zürich

Learning individual behaviors from vehicle trajectory data

Dr. Kenan Zhang

29.03.2022

AMLD AI & Mobility

Background

- Vehicle trajectory as a probe of urban dynamics
  - Mobility profiling[1]

    1 Liu et al. Understanding intra-urban trip patterns from taxi trajectory data. J. Geogr. Syst., 2012

  - Community & hotspot detection[2,3]

    2 Liu et al. Revealing travel patterns and city structure with taxi trip data. J. Transp. Geogr., 2015
    3 Chang, Tai and Hsu. Context-aware taxi demand hotspots prediction. Int. J. Bus. Intell. Data Min., 2010

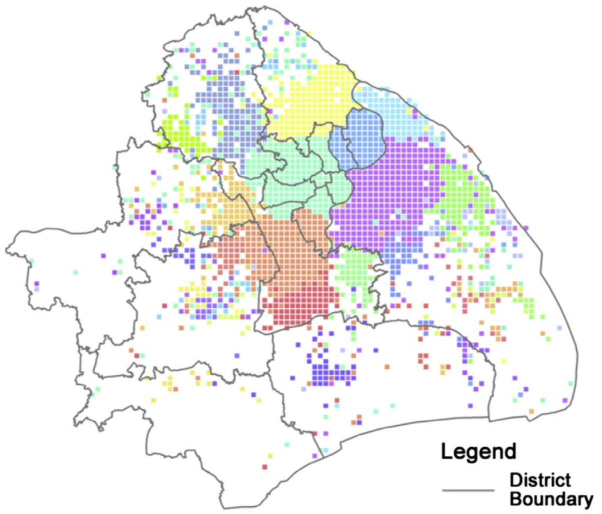
  - Traffic prediction[4]

    4 Zhang et al. Deep spatiotemporal residual networks for city wide crowd flows prediction. AAAI, 2017

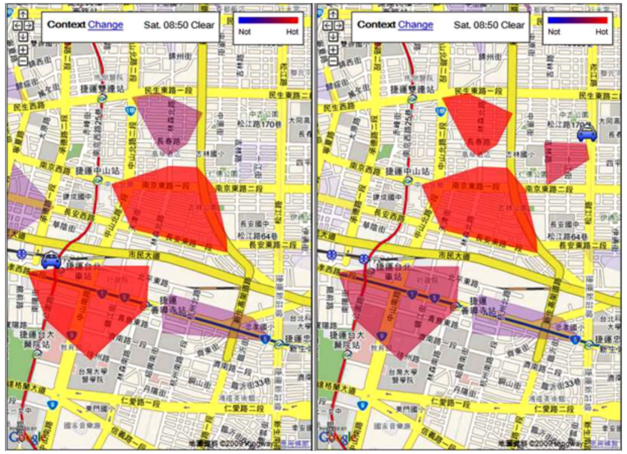
  - Location/Route recommendation[5,6]

    5 Lee, Shin and Park. Analysis of the passenger pick-up pattern for taxi location recommendation. IEEE NCM, 2008
    6 Yu et al. A Markov decision process approach to vacant taxi routing with e-hailing. TR-B, 2019

⇒ Analysis at the aggregate level rather individual level



Community detection results (Liu et al., 2012)



Personalized hotspot score (Chang, Tai and Hsu., 2010)

Outline

- Data & preprocessing
- Learn behavioral patterns from trajectories
  - Taxi search strategy recognition
- Learn latent state from trajectories
  - Taxi occupancy status correction
- Discussions

Data

- Taxi GPS trajectories
  - All registered taxis in Shenzhen, China (~20K by 2020)
  - Average record intervel is 20 sec (~20B rows per week)

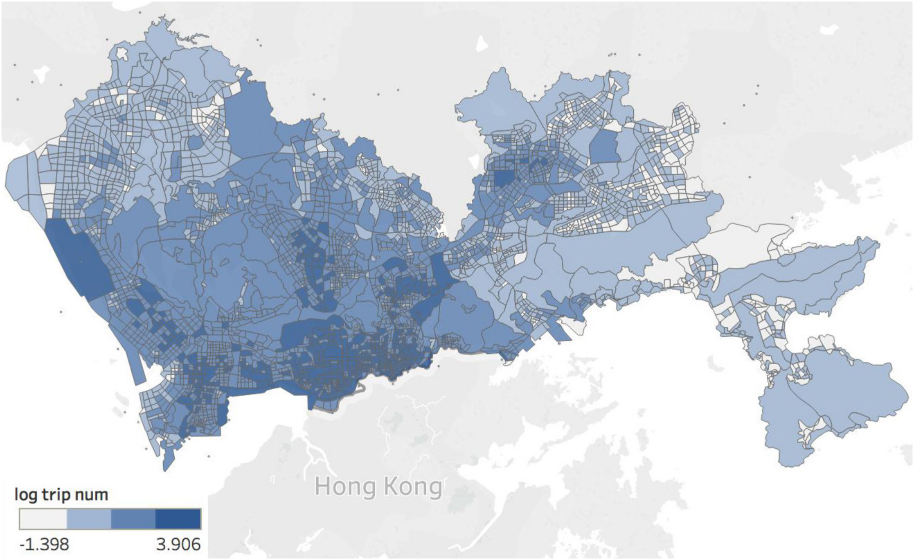
Taxi ID | Timestamp | Coordinates | Instantaneous speed and orientation | Occupancy

taxiid character varying (12)	time timestamp without time zone	lon double precision	lat double precision	velocity integer	angle integer	passenger smallint	
UUUB0C0M7	2020-01-01 00:19:43	114.125206	22.567154	0	171	0	
UUUB0C0M7	2020-01-01 00:19:43	114.125206	22.567154	0	171	0	
UUUB0C0M7	2020-01-01 00:19:58	114.125206	22.567154	0	171	0	
UUUB0C0M7	2020-01-01 00:20:13	114.125206	22.567154	0	171	0	*vacant*
UUUB0C0M7	2020-01-01 00:20:28	114.125206	22.567154	0	171	0	
UUUB0C0M7	2020-01-01 00:20:28	114.125206	22.567154	0	171	0	
UUUB0C0M7	2020-01-01 00:20:30	114.125206	22.567154	0	171	0	
UUUB0C0M7	2020-01-01 00:22:55	114.124718	22.562443	19	89	1	*pickup*
UUUB0C0M7	2020-01-01 00:24:23	114.125198	22.559134	22	60	1	*occupied*
UUUB0C0M7	2020-01-01 00:24:26	114.125412	22.559196	35	31	1	

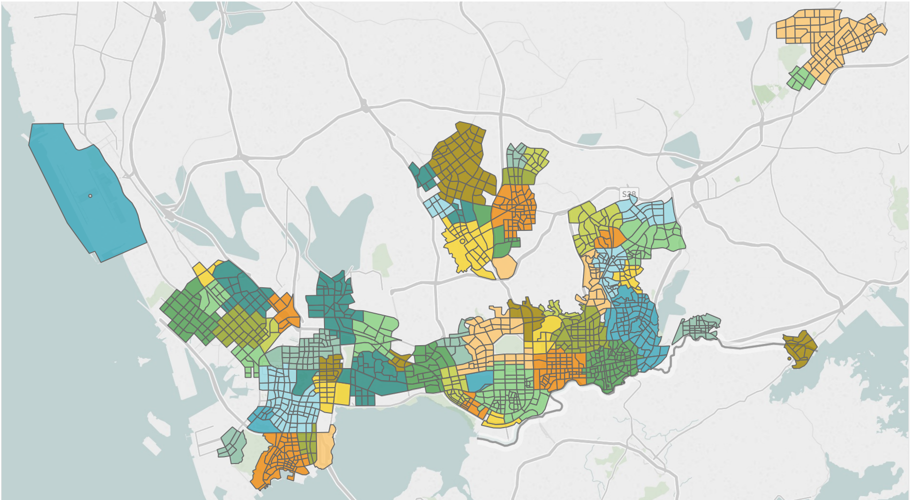
Preprocessing

- Geographic information augmentation
  - Mapping to Transportation Analysis Zone (TAZ)
  - Mapping to road segment[1]
- Trip segmentation
  - Based on the change in occupancy status[2]
- Hotspot extraction
  - TAZ clustering by iDBSCAN[3]
    - Distance: centroid distance
    - Weight: pickup and dropoff number

Distribution of pickups and dropoffs



Hotspot areas



1 Wu et al. Map matching based on multi-layer road index. TR-C, 2020
2 Nie. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. TR-C, 2017
3 Pan et al. Land-use classification using taxi GPS traces. IEEE ITS, 2013

ETH zürich

Outline

- Data & preprocessing
- Learning patterns from trajectories
  - Taxi search strategy recognition
- Learning latent state from trajectories
  - Taxi occupancy status correction
- Discussions

Motivation

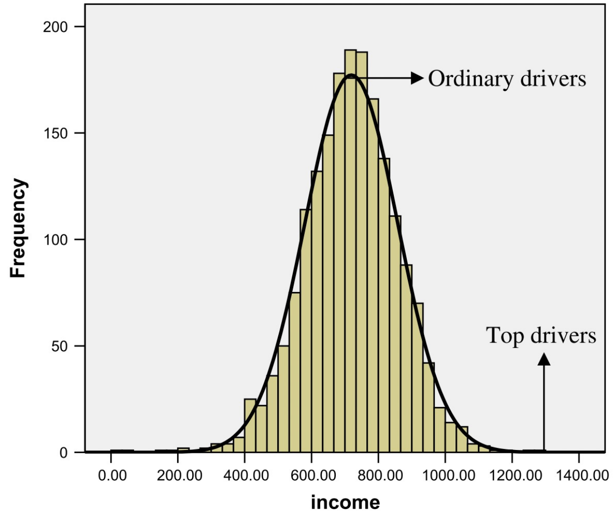
- Identify different search strategies among idividual taxi drivers
  - Previous work
    - Preclassified driver groups[1,2]
    - Predefined search strategies[2]

1 Liu, Andris and Ratti. Uncovering cabdrivers' behavior patterns from their digital traces. Comput. Environ. Urban Syst., 2010
2 Zhang et al. Understanding taxi service strategies from taxi GPS traces. IEEE ITS, 2015

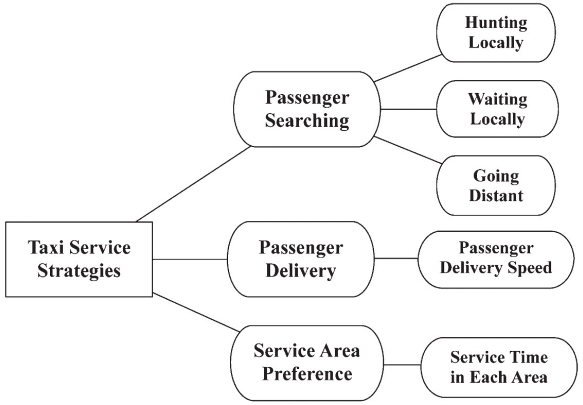
  - Related work
    - Model drivers as rational agents with full information[3]
    - Design optimal search path based on historical data[4]

3 Wong et al. Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing Markov chain approach. J. Adv. Transp., 2005
4 Yu et al. A Markov decision process approach to vacant taxi routing with e-hailing. TR-B, 2019

⇒ Directly learn taxi search strategies from data



Distribution of driver income (Li et al., 2011)



Predefined service strategies (Zhang et al., 2015)

Hunting image

- An image-based representation of individual behaviors
  - Pixel ⇒ Spatiotemporal index
  - Channel ⇒ Features

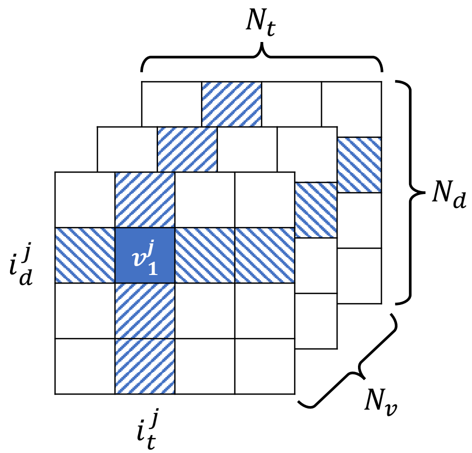
Search trips of each taxi

$$\mathcal{H}^j = \left\{\tau_k^j\right\}_{k=1}^{K^j}$$

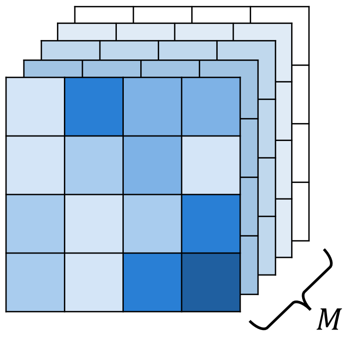
Features for each spatiotemporal index

$$x_{i_t,i_d}^j = \left\{v_l^j\right\}_{l=1}^{N_v}$$

Hunting image of each taxi



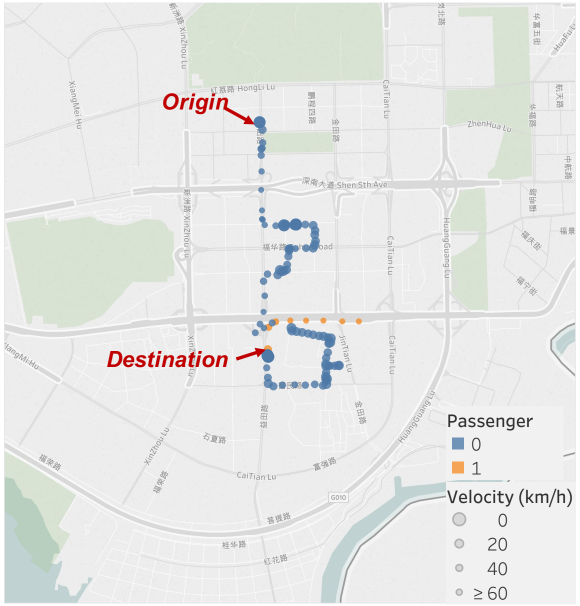
Collection of hunting images



Feautre engineering

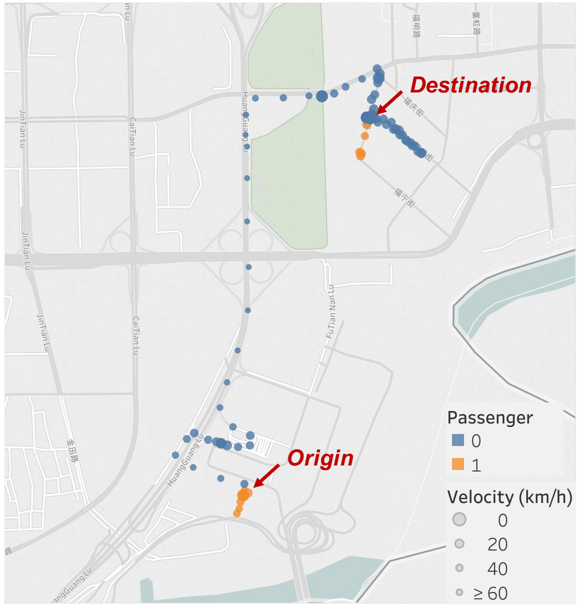
- Cruising characteristics
  - Cruising speed $v_c = d_s/t_s$, where d_s, t_s are search distance and time, respectively
  - Cruising ratio $r_c = \begin{cases} 1, & d_s = 0 \\ d_s/l_s\,, & d_s > 0 \end{cases}$, where l_s is line distance between search origin and destination

$v_c = 12.65\ mph,\ r_c = 2.64$



$v_c = 22.49\ mph,\ r_c = 1.39$

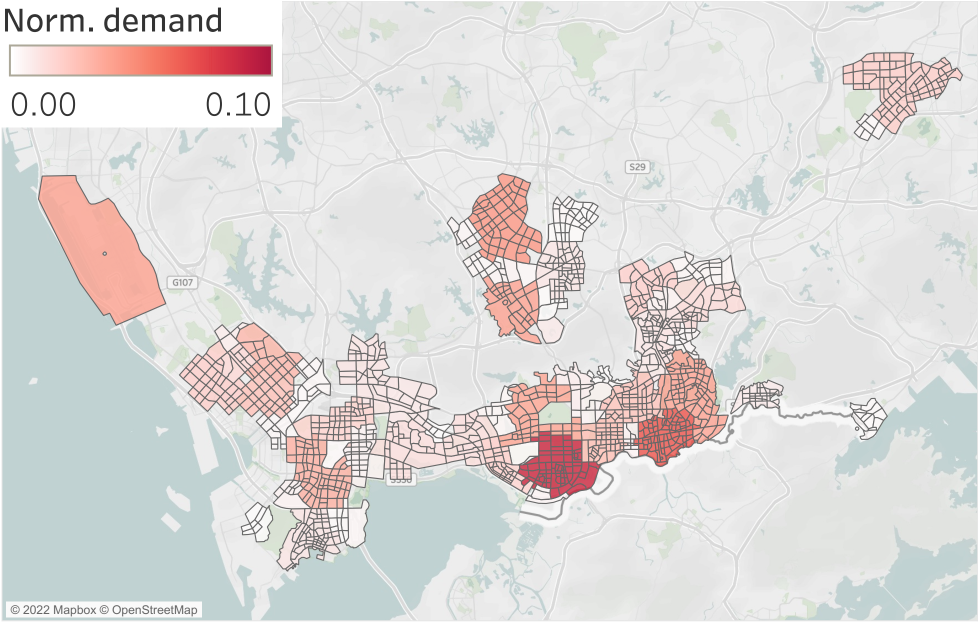
$v_c = 12.71\ mph,\ r_c = 2.38$

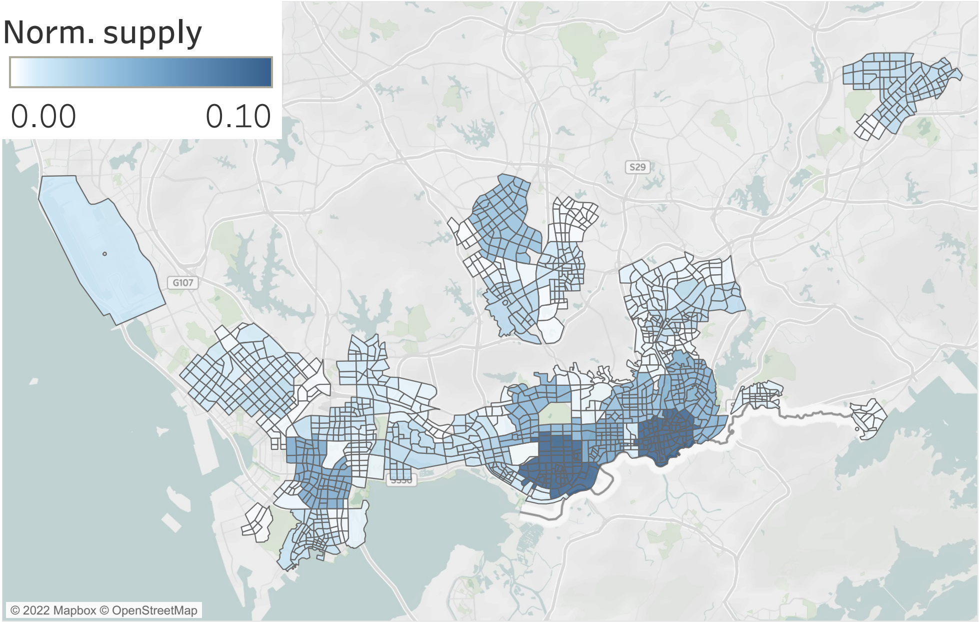


Feautre engineering

- Search driven by pickup probability
  - Normalized demand and supply by location and time $D^{a,t} = \frac{Q^{a,t}}{\sum_j Q^{j,t}}$, $S^{a,t} = \frac{T^{a,t}}{\sum_j T^{j,t}}$
  - Demand-supply ratio $R^{a,t} = D^{a,t}/S^{a,t}$
  - Difference in demand-supply ratio: $\Delta R = \log\left(R^{d,t}\right) - \log(R^{o,t})$

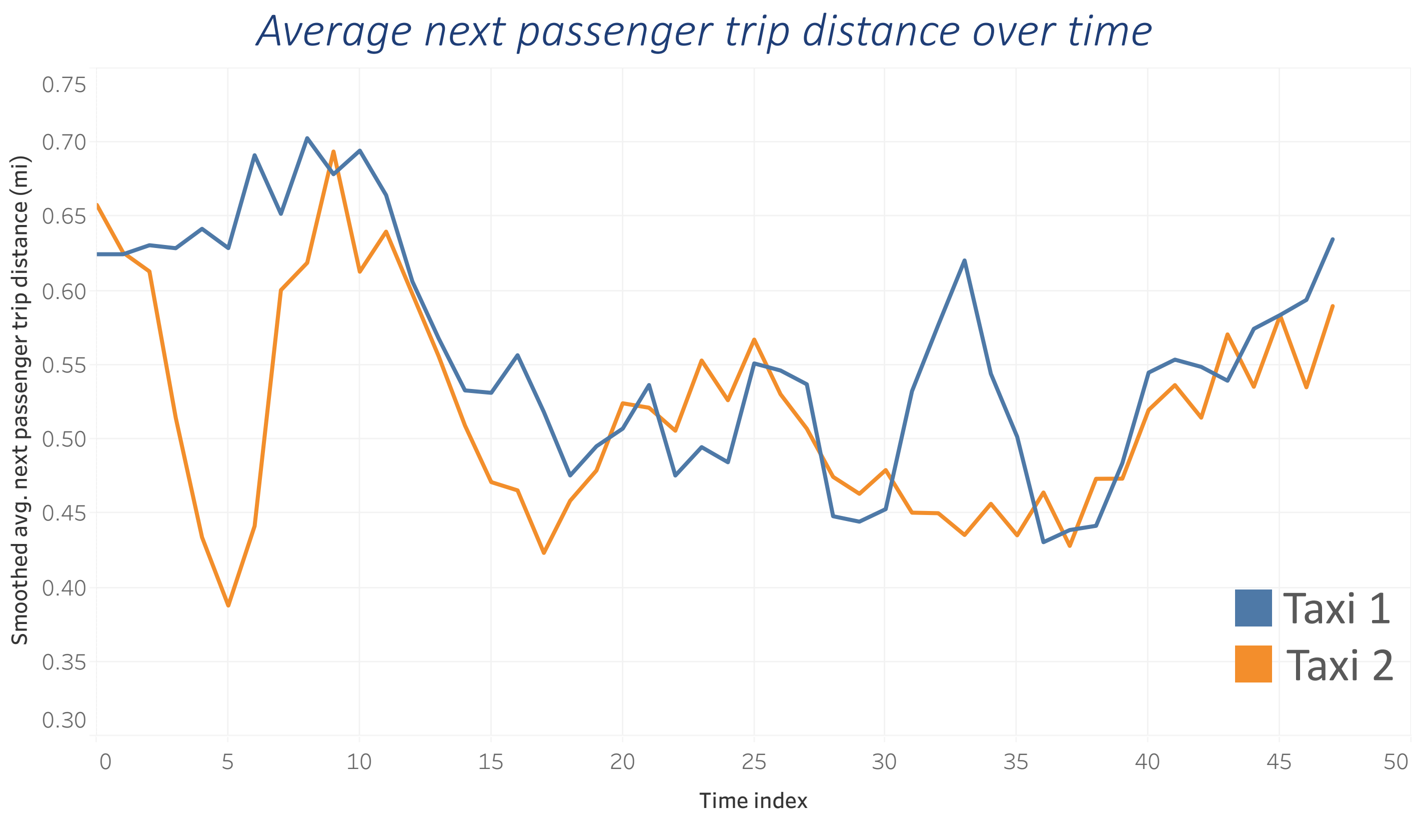
Normalized demand and supply on daily average





Feautre engineering

- Search driven by expected revenue
  - Distance of next passenger trip d_p



Strategy recognition

- Sparse space clustering (SSC)[1]
  - Problem: Identify low-dimensional structure embedded in a high-dimensional feature space
  - Idea: Represent each taxi driver as linear sparse combination of others

    ⇒ *similar drivers have non-zero weights in the representation matrix*

  - Sparse representation problem
    - Feature matrix $Y \in \mathbb{R}^{N \times M}$
    - Representation matrix $C \in \mathbb{R}^{M \times M}$

Original formulation

$$\min_{C} ||C||_0 \quad s.t.\ Y = YC,\ diag(C) = 0$$

➡

Relaxed to l_1-norm

$$\min_{C} ||C||_1 \quad s.t.\ Y = YC,\ diag(C) = 0$$

➡

Consider outlying entries

$$\min_{C,Z} ||C||_1 + \gamma ||Z||_1 \quad s.t.\ Y = YC + Z,\ diag(C) = 0$$

1 Elhamifar and Vidal. Sparse subspace clustering. CVPR, 2009

Strategy recognition

- Sparse space clustering (SSC)
  - Solve the representation matrix with ADMM[1]
    - A more compact form

$$\min_{X} ||X||_1 \quad s.t.\ Y = PX,\ \ X_{ii} = 0, i = 1, \dots, M$$

where $X = [C, \gamma Z]^T$, $P = [Y, I/\gamma]$

$$\Rightarrow \quad \min_{X,B} ||B||_1 \quad s.t.\ \ Y = PX,\ \ X = B,\ \ X_{ii} = 0, i = 1, \dots, M$$

    - Augmented Lagrangian

$$\mathcal{L}_{\rho_1,\rho_2}(X, B, \lambda_1, \lambda_2) = ||B||_1 + \frac{\rho_1}{2}||Y - PX||_2^2 + \frac{\rho_2}{2}||X - B||_2^2 + \mathrm{tr}\left(\lambda_1^T(Y - PX)\right) + \mathrm{tr}\left(\lambda_2^T(X - B)\right)$$

    - Iterative rules

$$X^{k+1} = (\rho_1 P^T P + \rho_2 I)^{-1}(\rho_1 P^T Y + \rho_2 B^k + P^T \lambda_1^k - \lambda_2^k)$$

$$B^{k+1} = S_{1/\rho_2}\left(X^{k+1} + \lambda_2^k/\rho_2\right)$$ *soft-thresholding with parameter $1/\rho_2$*

$$\lambda_1^{k+1} = \lambda_1^k + \rho_1(Y - PX^{k+1})$$

$$\lambda_2^{k+1} = \lambda_2^k + \rho_2(X^{k+1} - B^{k+1})$$

1 Boyd et al. Distributed optimization and statistical learning via the alternative direction methods of multipliers, 2011

Strategy recognition

- Sparse space clustering (SSC)
  - Spectral clustering on the representation matrix
    - Construct similarity matrix
      - Normalize C^* by column $c_i^* = c_i^*/|c_i^*|_\infty$
      - Ensure symmetry $\bar{C} = |C^*| + |C^*|^T$
    - K-means clustering on the normalized graph Laplacian

$$L = I - D^{-1/2}\bar{C}D^{-1/2},$$

where $D = diag\{\sum_j \bar{C}_{ij}\}$

      - Cluster number determined based on the number of zero-eigenvalaues of L

1 Boyd et al. Distributed optimization and statistical learning via the alternative direction methods of multipliers, 2011

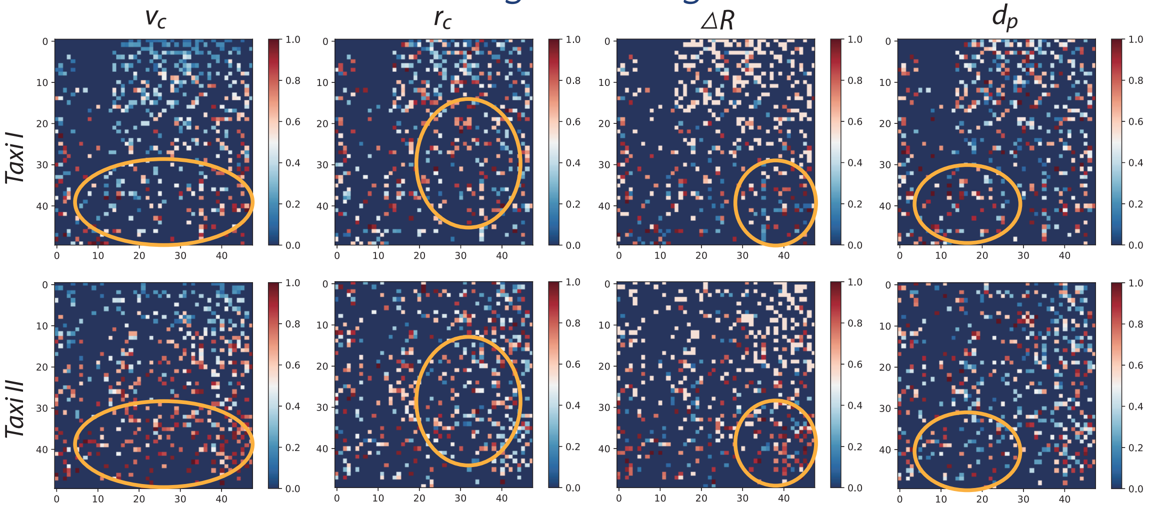
Experiment

- Setting
  - Weekday trips during five weeks in 2016, each from a different month
  - 885 taxis are selected, which
    - continuously operate in the analysis period
    - have at least 450 valid search trips (~18 trips per day)
  - Image construction
    - Time horizon: discretized by half-hour intervals
    - Space horizon: discretized by search distance percentile with 2% increment
    - Pixel value: average feature value of trips in corresponding space-time slot

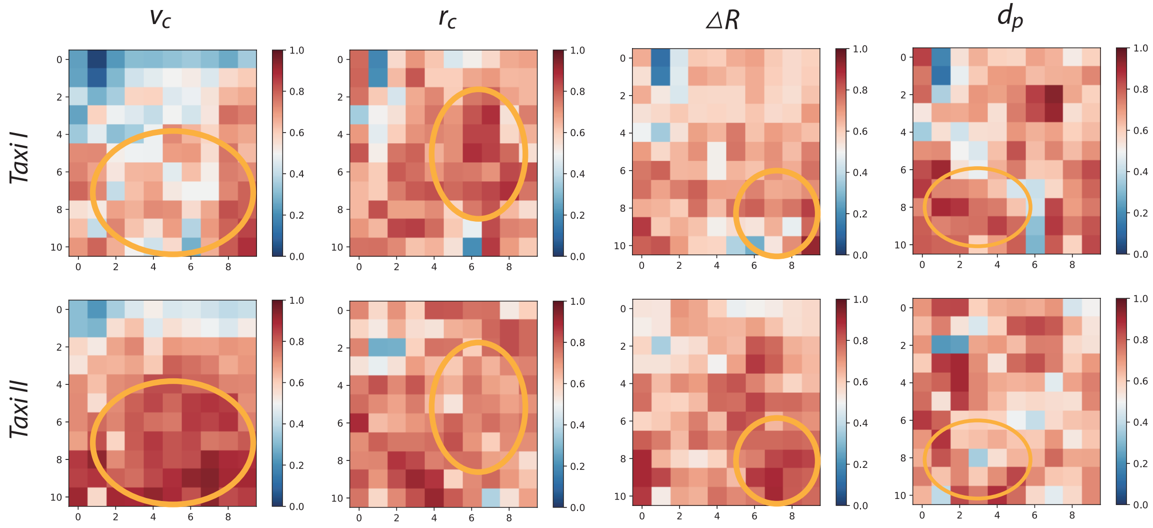
Experiment

- Image smoothing
    - Gaussian filtering + max-pooling

Original images

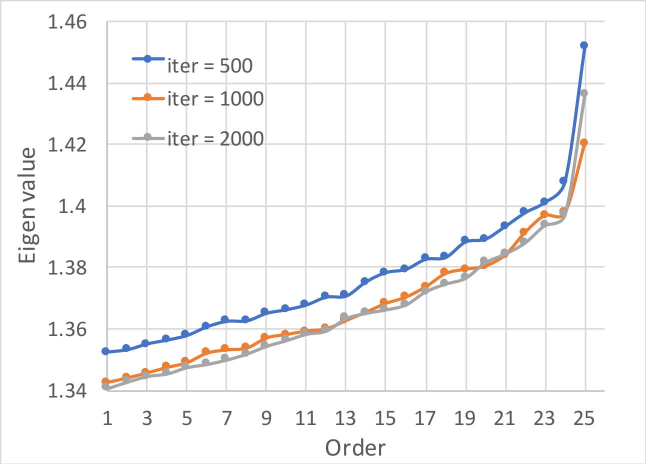


Smoothed images



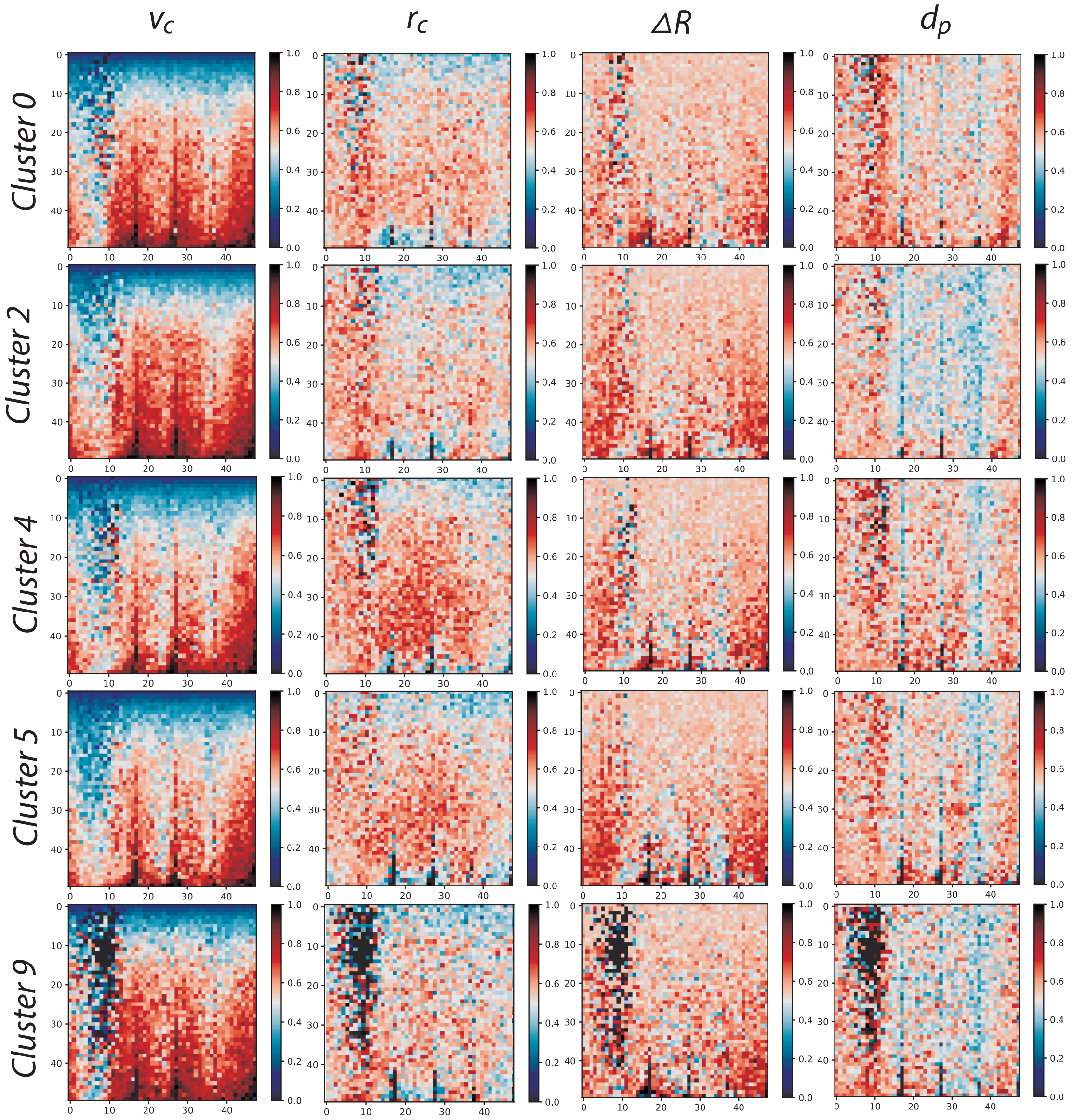
- Clustering results

Eigenvalues of L



Experiment

- Results of selected taxi clusters
  - Common patterns
    - Feature distribution
    - Special time windows
  - Difference in strategies
    - Direct search vs local cruising
      - e.g., Cluster 2 and Cluster 4
    - Long-distance vs short-distance
      - e.g., Cluster 0 and Cluster 2
    - Search at night
      - e.g., Cluster 0 and Cluster 5



Takeaways

- Image-based represention of individual behaviors
  - Enable analysis of behavioral patterns using image processing and learning methods

- First expand the feature space then do subspace clustering
  - Make full use of the information
  - Avoid the curse of dimensionality

Outline

- Data & preprocessing
- Learning patterns from trajectories
  - Taxi search strategy recognition
- Learning latent state from trajectories
  - Taxi occupancy status correction
- Discussions

Motivation

- Mislabeled occupancy status
  - Lead to incorrect trip segmentation
  - Unable to fix through simple filtering

- Learning with label noise
  - Data cleansing: detect anomaly/outliers (e.g., SVM[1])
  - Noise-robust model: use model robust to noise (e.g., risk minimization[2])
  - Noise-tolerant model: incorporate label noise in learning (e.g., Bayesian[3])

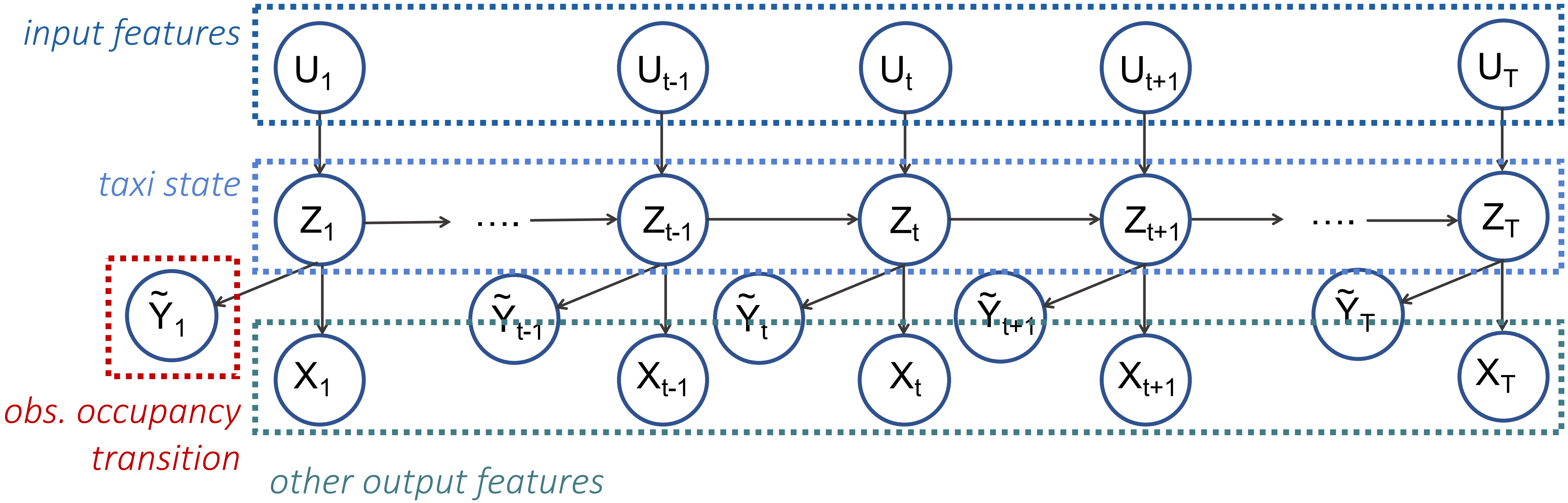
1 Thongkam et al. Support vector machine for outlier detection in breast cancer survivability prediction. Asia-Pacific Web Conf., 2008
2 Manwani and Sastry. Noise tolerance under risk minimization. IEEE Cyber, 2013
3 Swartz et al. Bayesian identifiability and misclassification in multinomial data. Can. J. Statist., 2004

Problem statement

- Definitions
  - Trajectory sequence $s \coloneqq \{e_t\}_{t=1}^T$
    - a series of GPS records such that the time between any two consecutive points is less than a threshold
  - Occupancy transition $\tilde{y}_t = (o_{t-1}, o_t)$
    - transition in occupancy status between two consecutive points
  - State $z_t \in \{\text{vacant}, \text{pickup}, \text{occupied}, \text{dropoff}\}$
    - each produce different occupancy transitions
  - Input features u_t
    - independent on but affect state z_t (e.g., time of day, location)
  - Output features x_t
    - dependent on state z_t (e.g., vehicle movement)
- Main objective
  - Predict state sequence $\{z_t\}_{t=1}^T$ based on the input and output features, along with observed occupancy transitions, i.e., $\{u_t, x_t, \tilde{y}_t\}_{t=1}^T$

IO-HMM formulation

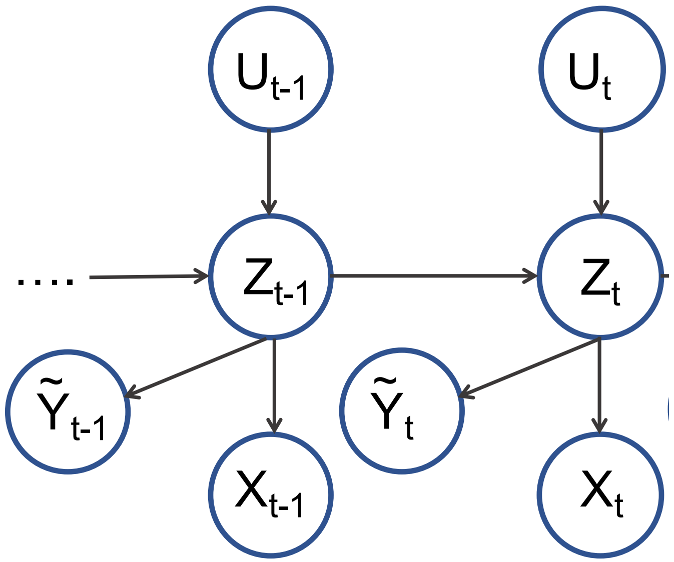
- Input-output hidden Markov model
  - Causal graph



IO-HMM formulation

- Input-output hidden Markov model
  - Parameterization $\Theta(q, A, b, e)$
    - Initial state distribution $q_i = P(Z_1 = i)$
    - State transition $a_{i,j,u} = P(Z_t = j | Z_{t-1} = i, U_t = u)$
    - Emission probability $b_{i,l,x} = P\left(X_{t,l} = x | Z_t = i\right)$

      $e_{i,y} = P\left(\tilde{Y}_t = y | Z_t = i\right)$



  - Solution algorithm: expectation-maximization (EM)
    - E step: compute posterior via forward-backward algorithm
    - M step: update parameters via MLE

Feature engineering

- Input features
  - “night” (binary)
  - “hotspot” (binary)
  - “highway” (binary)
  - “no-trans” (binary): transition is not likely to happen due to infeasible acceleration
- Output features
  - “unf-acc-dec” (categorical): moving at a uniform/increasing/decreasing speed
  - “straight-return-largeturn” (categorical): moving straight/regular turn/sharp turn
  - “move-shortstop-longstop” (categorical): moving/short-time stop/long-time stop
  - “obstrans” (categorical): observed occupancy transition

Experiment

- Trajectory preprocessing
  - Speed and orientation filtering using both instantaneous values and consecutive coordinates
  - Segment data into trajectory sequences with time gap threshold 300 sec
- Training
  - 2,000 sequences (~800K GPS points) divided into 10 batches, 20 taxis per batch and 10 sequences per taxi
  - One model is trained for each batch with 100 times random initialization
- Testing
  - 115 sequences (~40K GPS points) with manually corrected labels
  - Small sample of trajectories with a large amount of errors

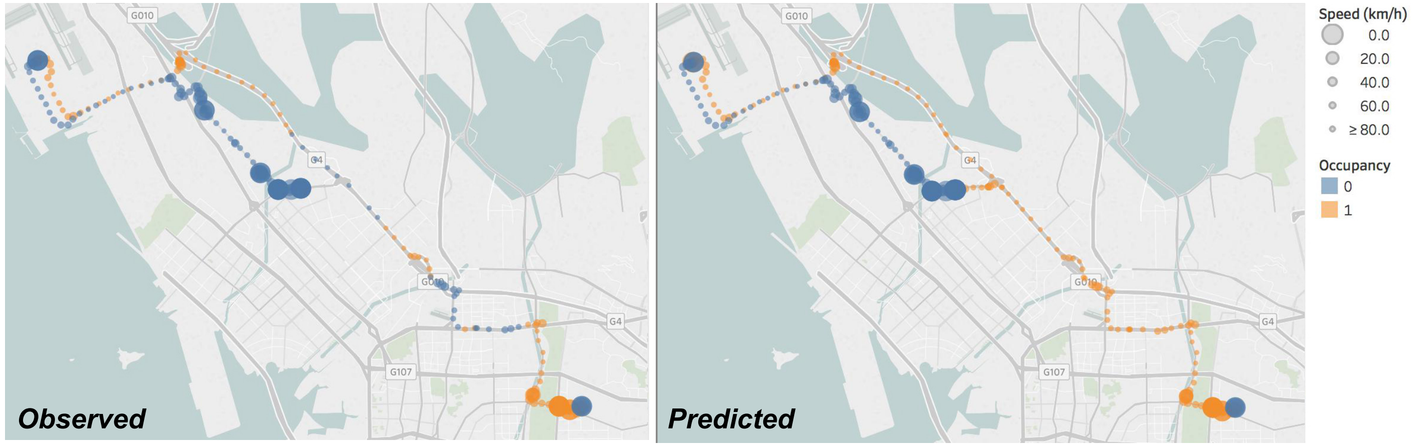
Experiment

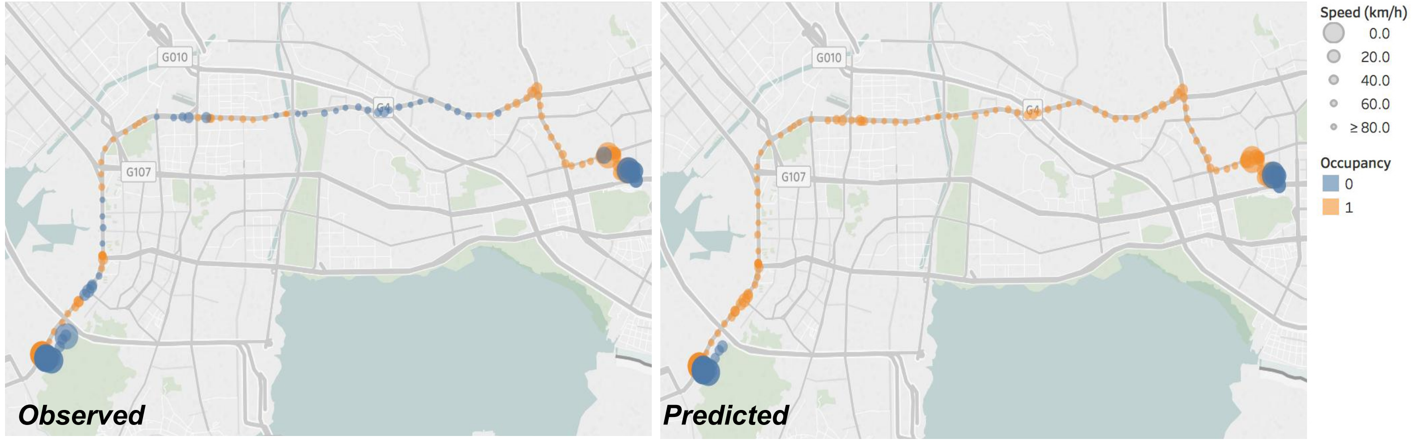
- Prediction accuracy and robustness
  - Transform predicted states back to occupancy status
    - assume flip in occupancy happens at the end of pickup/dropoff interval
  - Evaluate predictions on occupancy flips rather occupancy status
    - true-positive (TP): an occupancy flip is observed in a predicted pickup/dropoff interval
    - false-positive (FP): within a predicted pickup/dropoff interval, no occupancy flip is observed
  - Robustness analysis
    - manually added noise with Poisson occurrence (with prmt. λ) and Exponential duration (with prmt. β)

	λ	β	Recall		Precision		F1 score	
Baseline*	-	-	0.9699		0.9699		0.9699	
Rand.	0.05	5	0.8955	-0.0744	0.5679	-0.4020	0.6950	-0.2749
Rand.	0.05	10	0.9130	-0.0569	0.7022	-0.2677	0.7938	-0.1761
"highway" = T	0.1	5	0.9686	-0.0013	0.9507	-0.0192	0.9596	-0.0103

Experiment

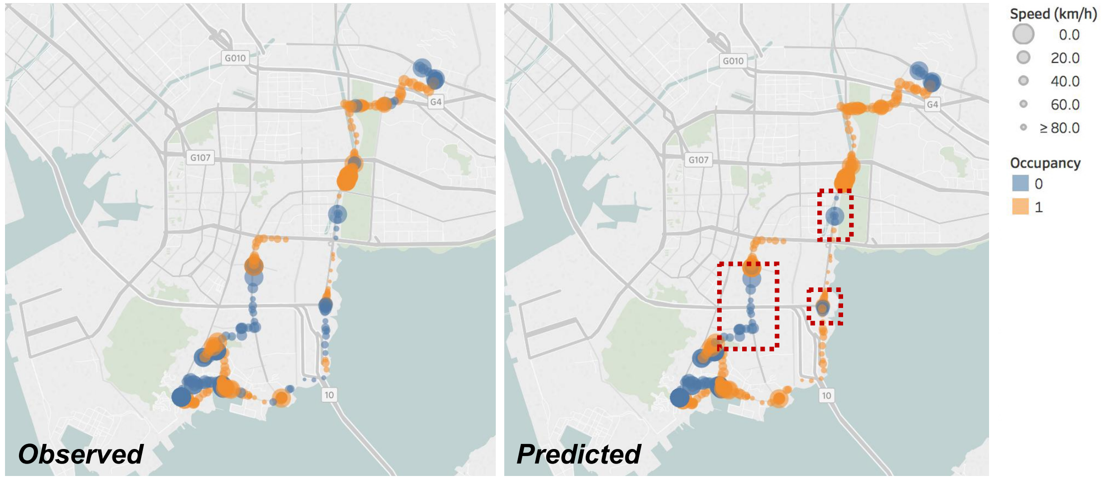
- Performance of label correction
  - Works well with simple and long trips with several mislabeled segmemts



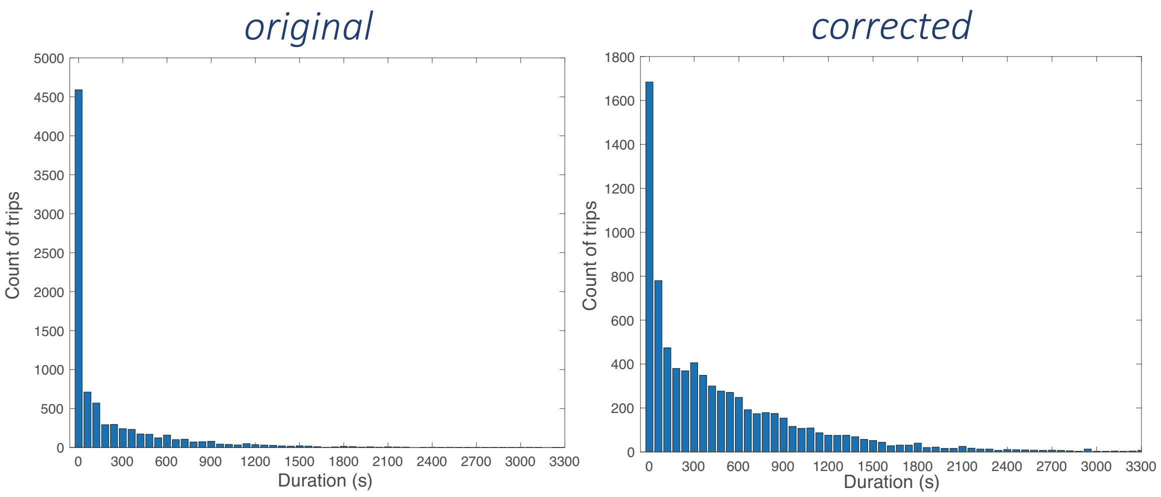


Experiment

- Performance of label correction
  - Fail to handle trips with complex behaviors and fix errors in highly corrupted data



Distribution of trip duration



Takeaways

- Sequential representation of individual behaviors
    - With latent states and observable contexts (“input”) and behaviors (“output”)

- Deal with noisy and missing label
    - Model it as a feature and learn the true label based on it

Outline

- Data & preprocessing
- Learning patterns from trajectories
  - Taxi search strategy recognition
- Learning latent state from trajectories
  - Taxi occupancy status correction
- Discussions

Discussions

- Issue of data sparsity
  - 10M GPS points ⇒ 450K trips ⇒ 500 trips per taxi ⇒ 0.2 trip per space-time slot
- Representation of behaviors
  - Spatial and temporal interdependence
- Model interpretability
  - Statistic model vs neural network
- Problem-driven
  - Study impact of certain factors ⇒ discriminative
  - Recognize behavioral pattern ⇒ clustering
  - Predict behaviors ⇒ deep learning, imitation learning

Thank you!

Please reach out if you have any questions

kenzhang@control.ee.ethz.ch

Papers presented in this talk

- **Zhang**, Chen and Nie. Hunting image: Taxi search strategy recognition using Sparse Subspace Clustering. TR-C, 2019
- **Zhang**, Zhong and Nie. Correcting mislabeled taxi trajectory occupancy status using Input-Output Hidden Markov model. TRB Annual Meeting, 2018