Carnegie Mellon University

Google

Baidu IDL

Inferring Movement Trajectories from GPS Snippets

Mu Li

Joint work with Amr Ahmed and Alex Smola

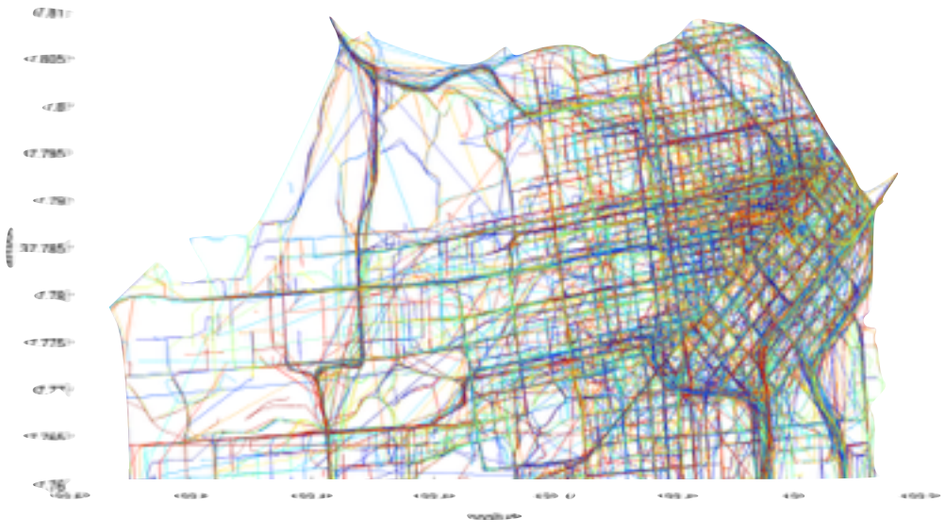
Motivation

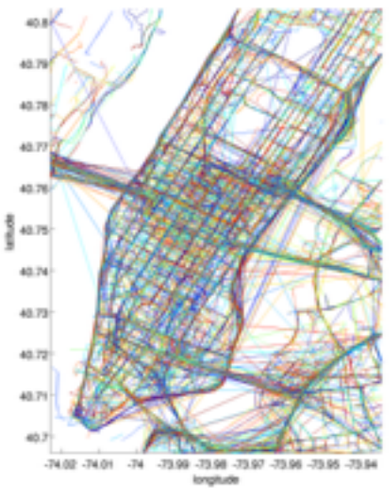
- Every 2/3 has a smartphone/tablet nowadays, typically has GPS

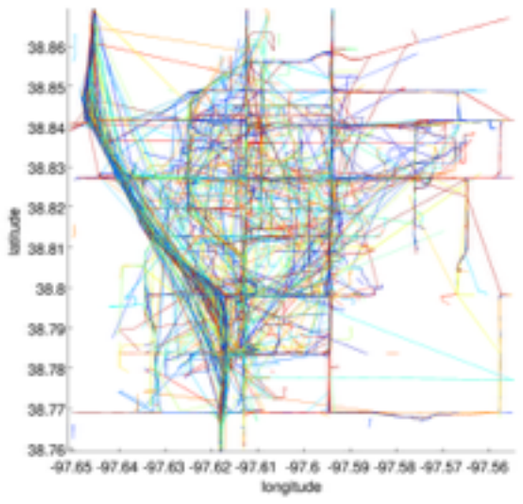
- Not only learn the current position, but also predict where people will go, and when arrive
- It benefits mobile apps
  - navigation
  - shop, restaurant recommendation
  - context-aware assistance
  - contextual metadata

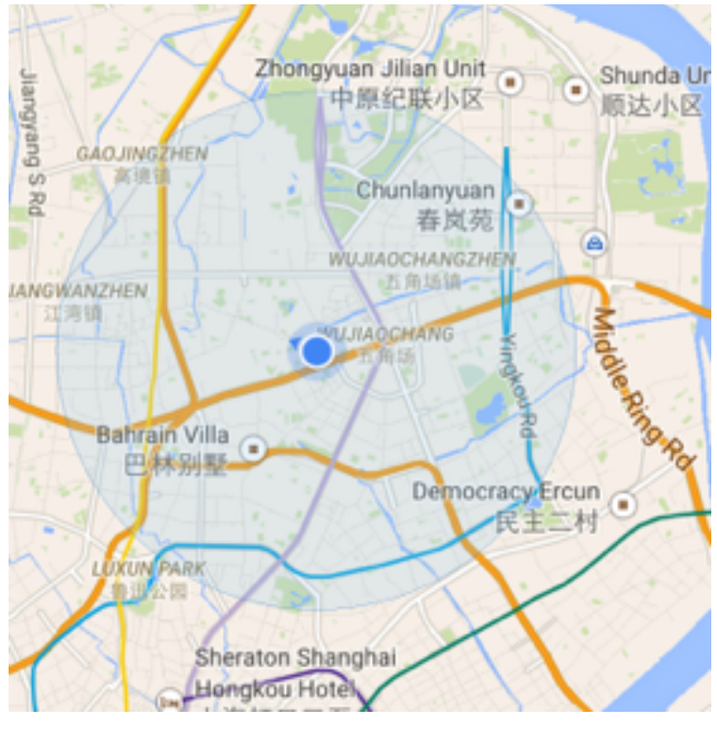
Challenges

- Data is big because of the huge amount of users
  - trillions sequence, worldwide coverage
- Data is sparse due to energy and privacy constraints
  - GPS sequence is short, only has several points
- Data is noisy
  - inexact positions in city
  - irrational path planing
  - travel speeds vary









Model and Inference

Model

$$p(O, S|\theta) = \prod_{k=1}^{n} p(o_k|s_k, \theta)p(s_{k+1}|s_k, \theta)$$

observation model

motion model

- Observation model:

$$p(o|s) \propto \exp\left(-\frac{1}{2\sigma_{\mathrm{d}}^2}\left\|o^{\mathrm{loc}} - s^{\mathrm{loc}}\right\|^2 - \frac{1}{2\sigma_{\mathrm{l}}^2}\left\|o^{\mathrm{dir}} - s^{\mathrm{dir}}\right\|^2\right)$$

- Motion model:

$$p(s'|s, \theta) = \sum_{\xi} p(\xi|s, \theta)p(s'|s, \xi, \theta)$$

all possible paths

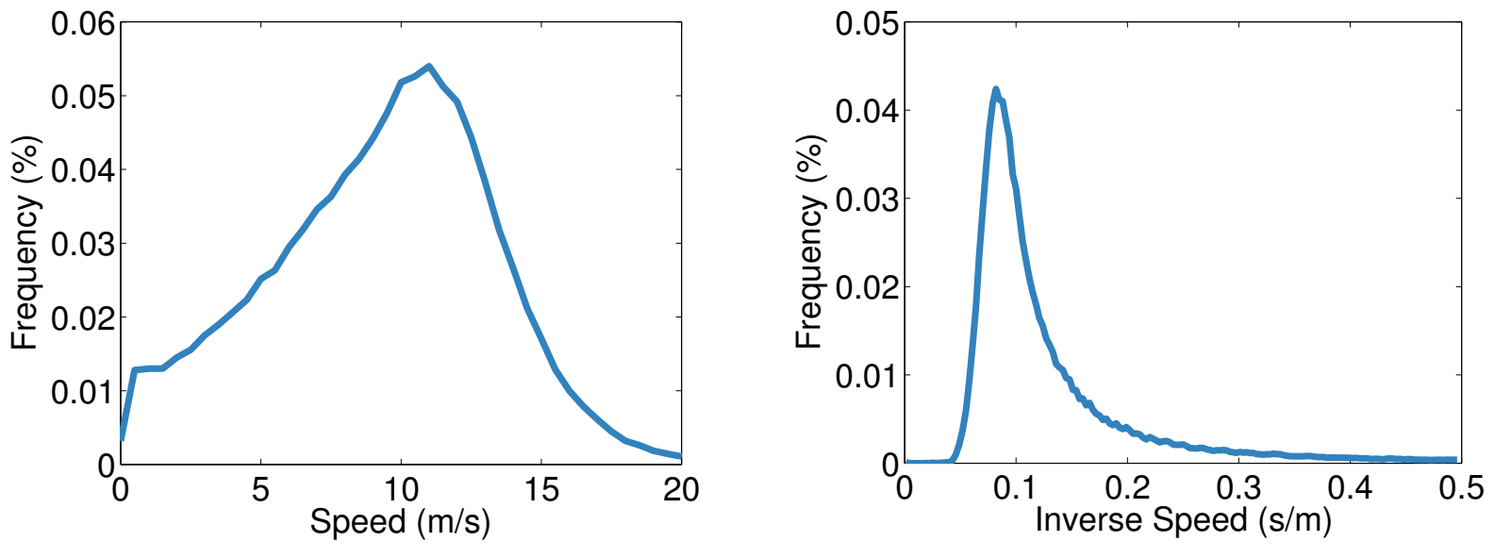
$$= \sum_{\xi}\left[\prod_{\iota=1}^{n} \pi(i_\iota, i_{\iota+1})\right] p(s'|s, \xi, \theta)$$

transition probability

from s to s' along path ξ

Travel time of a road segment

Histogram of speed and travel time



- Key observation: speed somewhat follows Gaussian, and travel time follows an inverse Gaussian (IG) distribution
  - Time from s to s' ~ IG(length/speed, $\delta^2 \cdot \text{length}^2$)

Travel time of a path

- We made two assumptions: the mean and variance of the speed
  - for a path is a weighted linear combination of all road segments
  - for a road segment is a weighted linear combination of all associated attributes
    - ★ road type, #lanes, speed limit, location, time, etc…

Inference method

- Solve the non-convex optimization problem

$$\text{maximize}_{\pi,\omega,\gamma} \log p(O|\pi,\omega,\gamma),$$

  - where O is the training data, π transition probability, ω and γ are coefficient about speed mean and variance

- Repeat until convergence
  - randomly sample several GPS sequence
  - run an optimized dynamic programming to update π
  - update ω,γ by subgradient descent

Experiment

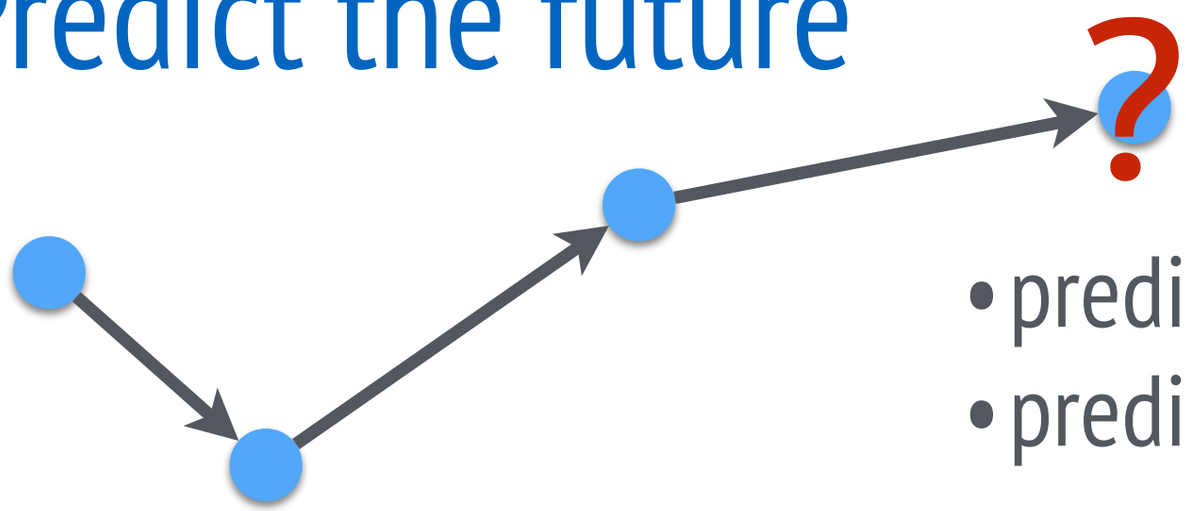
Setup

- Dataset

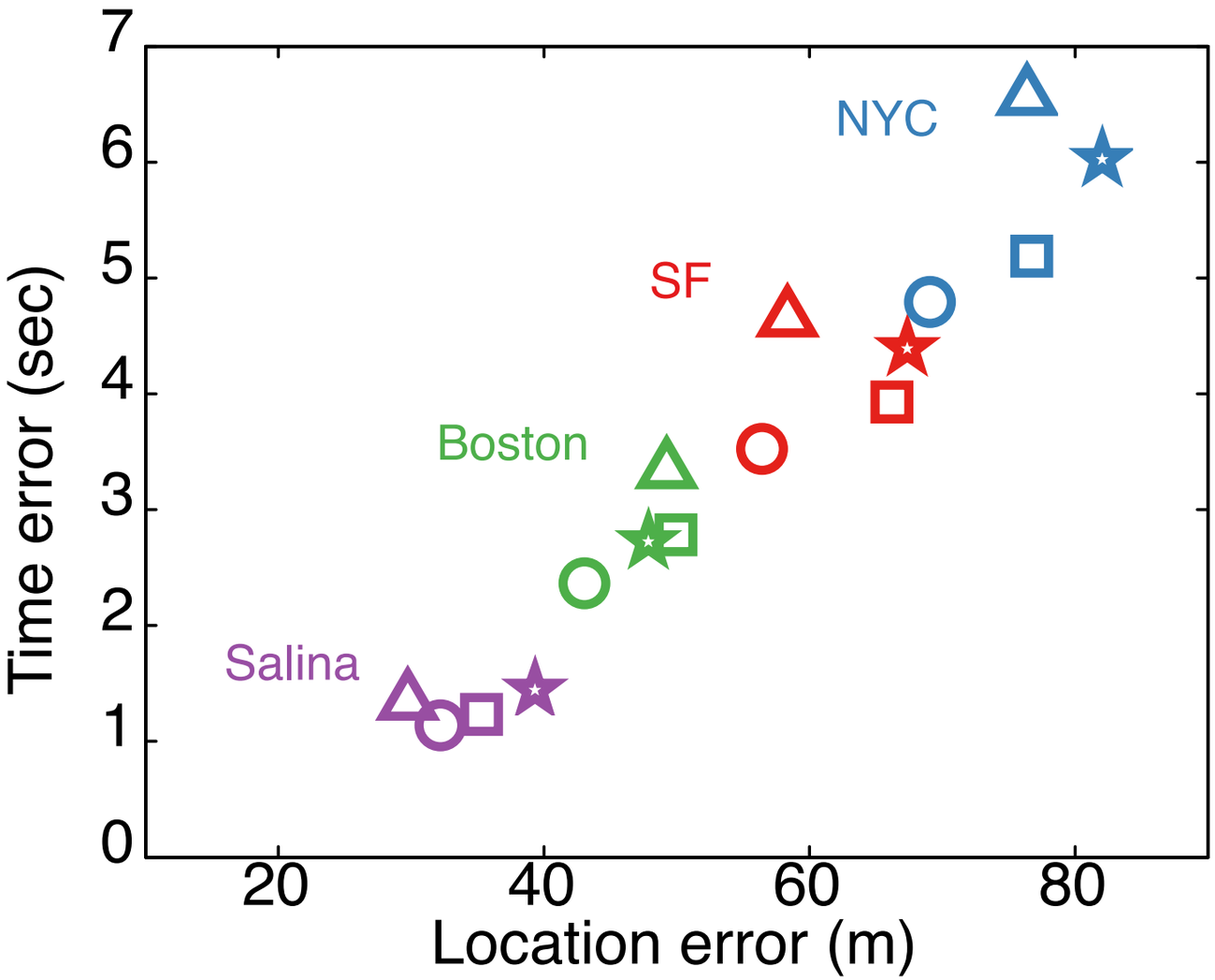
	SF	Boston	NYC	Salina
Road segment	18K	7K	17K	9K
Intersection	35K	10K	29K	23K
Trajectores	8M	7M	4M	3M

- Binary feature
  - road features: road attributes such as major road/high way, number of lanes, speed limit
  - temporal features: slice workday and weekend hours
  - personalized speed: use trajectory ID as a feature

Predict the future



- predict the position based on time
- predict the travel time based on position



△ use GPS recored speed

☆ only use the shortest path

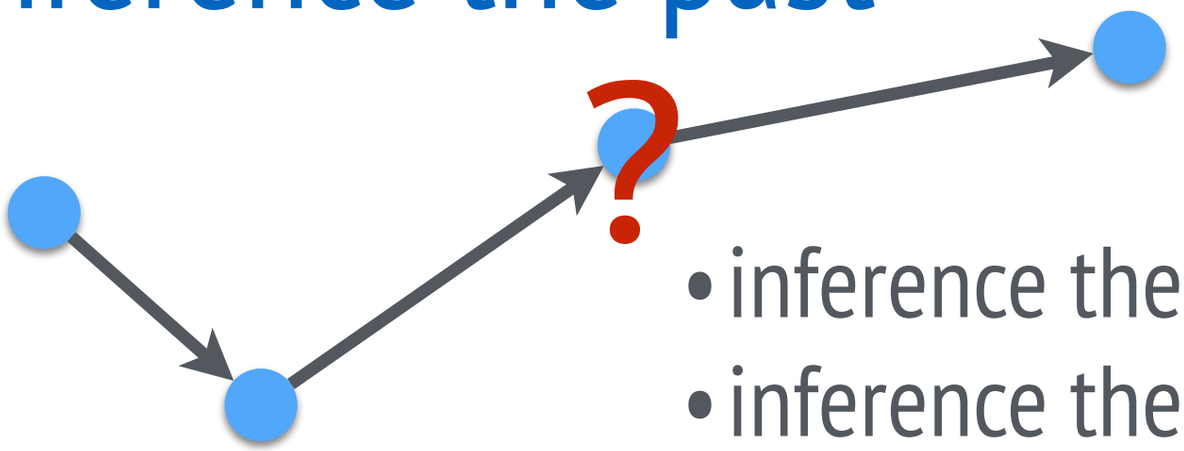
□ no personalized modeling

○ the full model

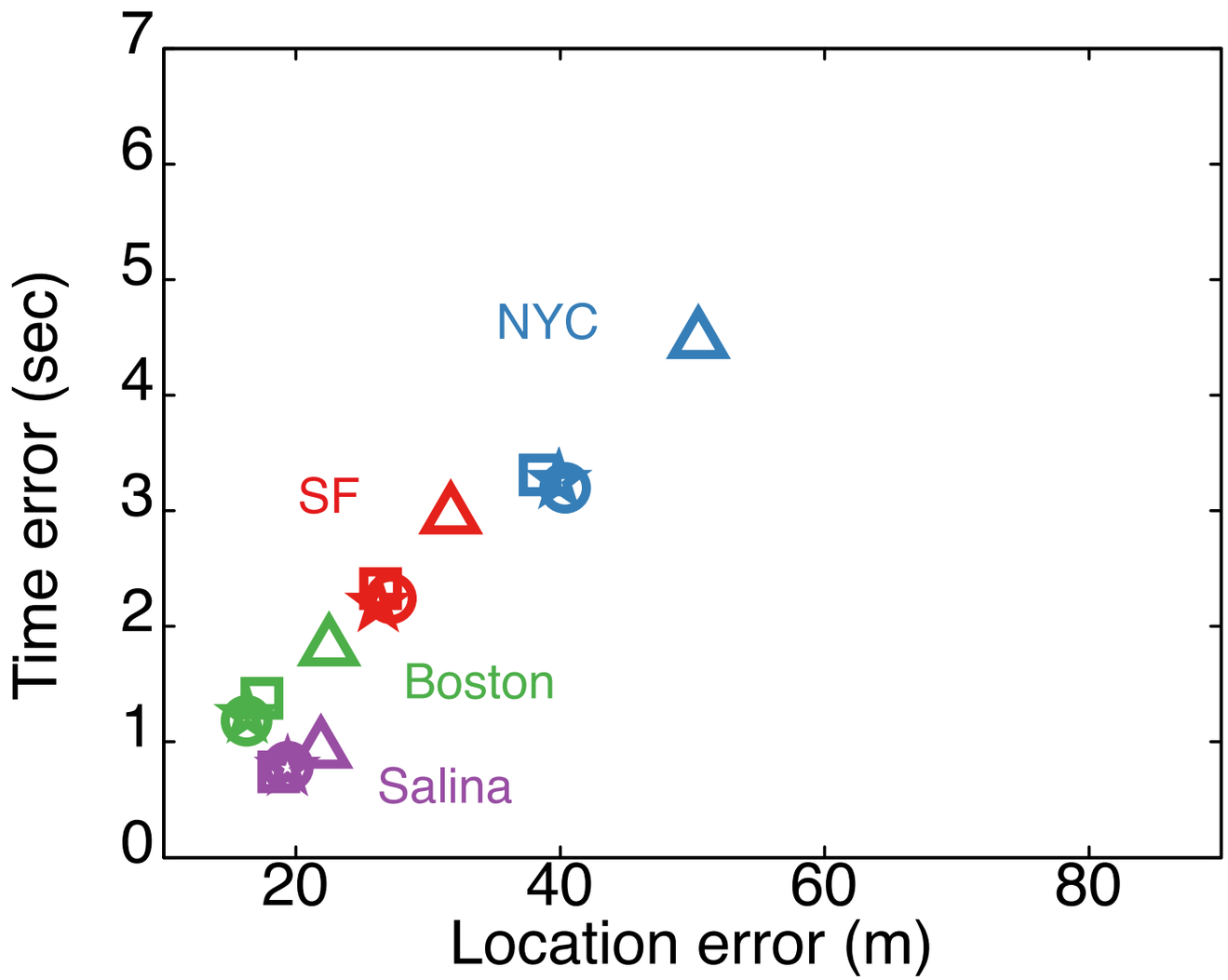
Conclusion

- A joint model: map trajectory + predict time/position + model road speed/variance
- Focus on sparse, noise, and anonymous GPS sequence
- A simple yet powerful model and efficient inference method

Inference the past



- inference the position based on time
- inference the travel time based on position



△ use GPS recored speed

☆ only use the shortest path

□ no personalized modeling

○ the full model