GPS Trajectory Data Enrichment Based on a Latent Statistical Model

CPS-IIP project

Akira Kinoshita[1], Atsuhiro Takasu[2], Kenro Aihara[2], Jun Ishii[3], Hisashi Kurasawa[3], Hiroshi Sato[3], Motonori Nakamura[3], and Jun Adachi[2].

1) The University of Tokyo, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan.
2) National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan.
3) NTT Network Innovation Laboratories, 3-9-11 Midoricho, Musashino, Tokyo, Japan.
{kinoshita,takasu,kenro.aihara,adachi}@nii.ac.jp, {ishii.jun,kurasawa.hisashi,sato.hiroshi,nakamura.motonori}@lab.ntt.co.jp

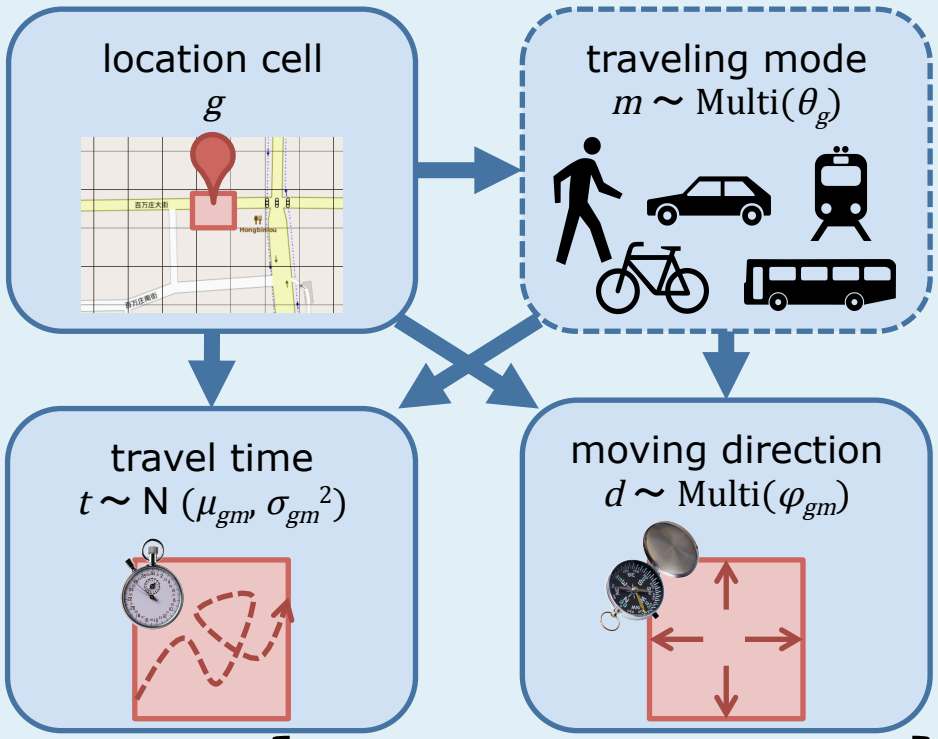
Goal

GPS trajectory data enrichment
- Interpolation
- Traveling mode estimation

Key Idea

Moving behavior can be changed according to the traveling mode even between the two same locations.

Proposed Latent Statistical Model

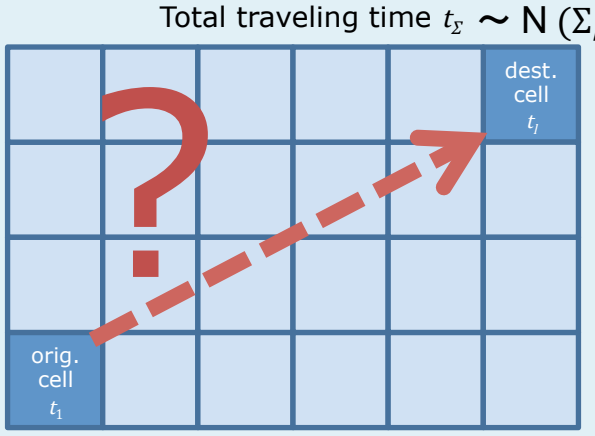


Trajectory $x = \begin{bmatrix} \text{location sequence} & <g_1, g_2, ..., g_l> \\ \text{travel time sequence} & <t_1, t_2, ..., t_l> \\ \text{direction sequence} & <d_1, d_2, ..., d_l> \end{bmatrix}$

Latent $y =$ traveling mode seq. $<m_1, m_2, ..., m_l>$

Interpolation Method

Total traveling time $t_\Sigma \sim \mathrm{N}(\Sigma\mu_{gm}, \Sigma\sigma_{gm}^2)$



Given:
- Possible set of complete-data trajectories $\{(\boldsymbol{x}, \boldsymbol{y})\}$
- Total traveling time t_Σ

Estimate:

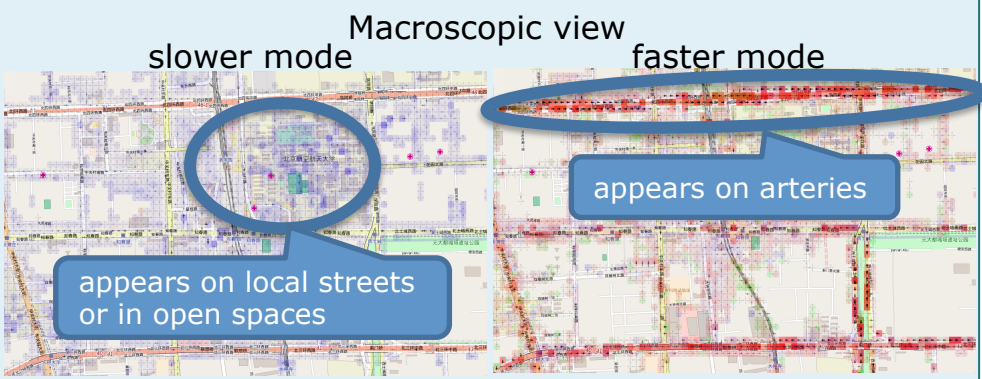
The route as the most probable trajectory

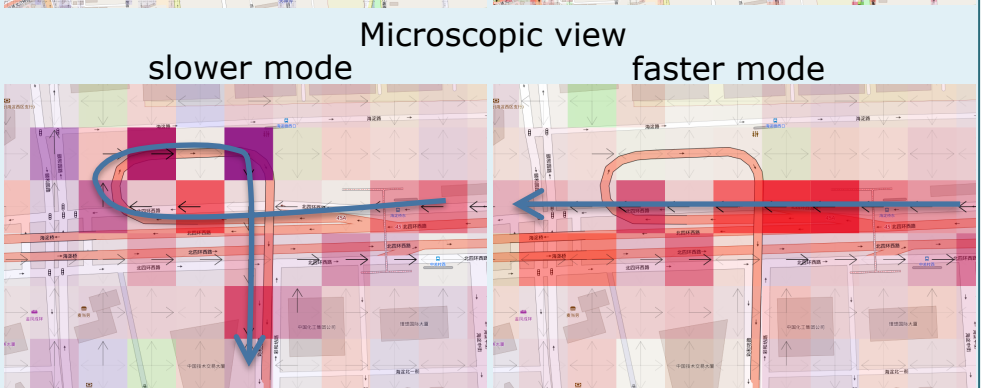
$$p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(t_\Sigma; \textstyle\sum_i \mu_{g_i m_i}, \sum_i \sigma_{g_i m_i}^2\right) \cdot \prod_i \theta_{g_i m_i} \cdot \phi_{g_i m_i d_i}$$

Experimental Results

Dataset: GeoLife GPS Trajectories [1–3]
- 90% for training, 10% for test.
- 9 traveling modes:
  walk, run, bike, bus, taxi, car, subway, train, airplane

MAP parameters estimated by semi-supervised learning





Interpolation

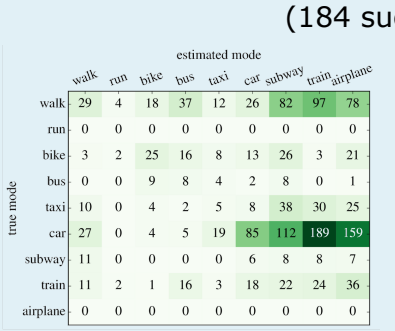
Test in 3 x 5 grid settings assuming shortest path.

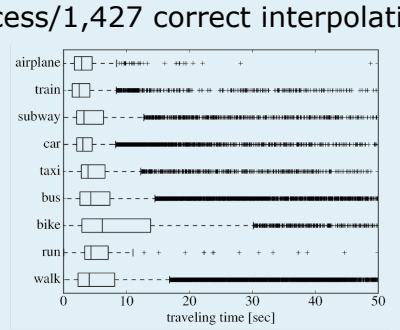
Interpolation accuracy **78.8%**

(38,695 success/49,092 cells)

Mode estimation accuracy 12.9%

(184 success/1,427 correct interpolations)





Future work

1. More computationally efficient interpolation algorithm.
2. Optimization of the set of traveling modes. 3. Feature selection.

References

[1]Y. Zheng et al. Understanding mobility based on GPS data. In *UbiComp '08*, 312–321, 2008.
[2]Y. Zheng et al. GeoLife: A collaborative social networking service among user, location and trajectory. *Bull. Tech. Comm. Data Eng.*, 33(2):32–39, 2010.
[3]Y. Zheng et al. Mining interesting locations and travel sequences from GPS trajectories. In *WWW '09*, 791–800, 2009.