

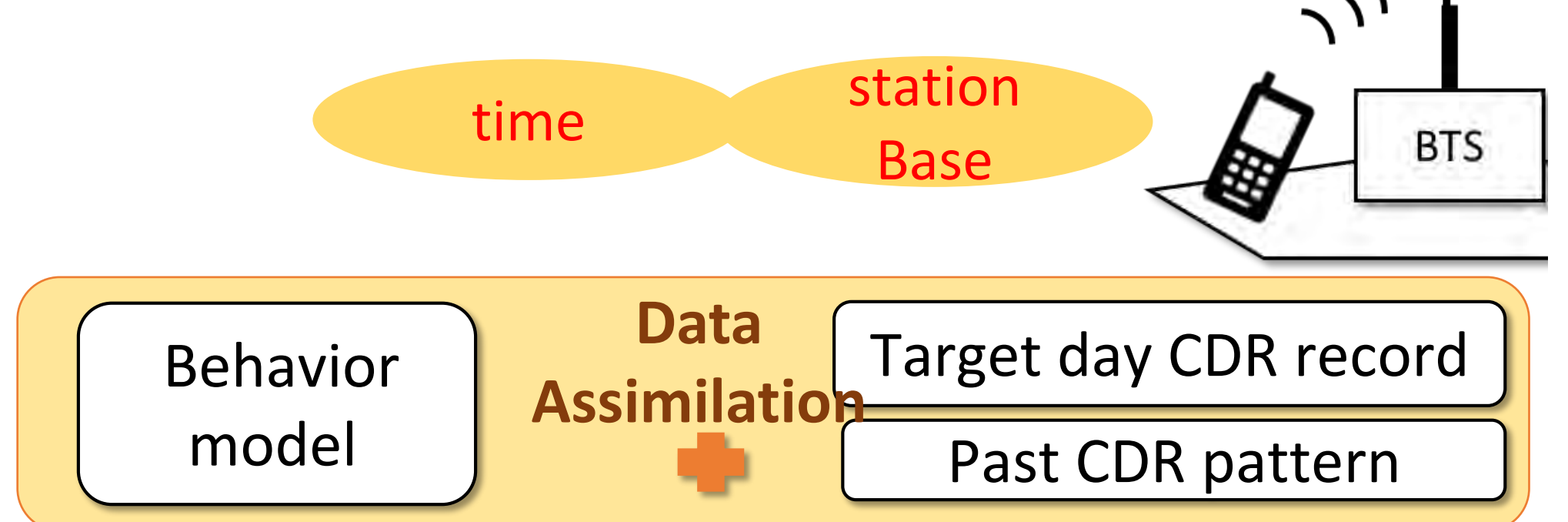
People Movement Estimation Using Sparse CDR Data

Yoko Hasegawa, Yoshihide Sekimoto, Takehiro Kashiya, Hiroshi Kanasugi

Background

Detailed and up-to-date information of the current traffic conditions is needed for efficient traffic management, especially in expanding cities with traffic networks still under rapid development. Analysis of user-anonymized mobile phone billing records, including the Call Detail Records (CDR), have an especially high potential for effective traffic conditions estimation, due to their wide population and area coverage. Another benefit of using this data is that there is no need for additional infrastructure, because mobile phones have already become one of the most important lifelines in many countries.

CDR: Call Detail Records



Methods

Estimation Method for People's Movement

Particle Filtering

posterior distribution likelihood prior distribution

$$p(X_t|Z_t) \propto p(Z_t|X_t) p(X_t|Z_{t-1})$$

$$p(X_t|Z_{t-1}) = \int p(X_t|X_{t-1}) p(X_{t-1}|Z_{t-1}) dX_{t-1}$$

transition probability

State Vector

$$x_{(i)}(t) = (n_{(i)}(t), m_{(i)}(t))^T$$

Node Transportation mode

$$x_{(i)}(0) = (n_{home(i)}, 'stay')^T$$

Transition Model

$$x_{(i)}(t) = F(x_{(i)}(t-1))$$

$$F(x_{(i)}(t)) = \left(\begin{array}{c} n_{(i)}(t) + u(m_{(i)}(t)) \\ S(m_{(i)}(t)) \end{array} \right)$$

Speed
Transportation mode transfer function

Likelihood Function Consistency with past pattern

$$W_{(i)}(t) = \delta_t(n_{(i)}(t)) \cdot w_{(i)}(t)$$

Target day observation

$$\delta_t(N) = \begin{cases} 1 & (N \text{ in observed base area}) \\ 0 & (\text{Out of area}) \end{cases}$$

(a) Initialize

(b) Predict

(c) Observation (Set importance weight)

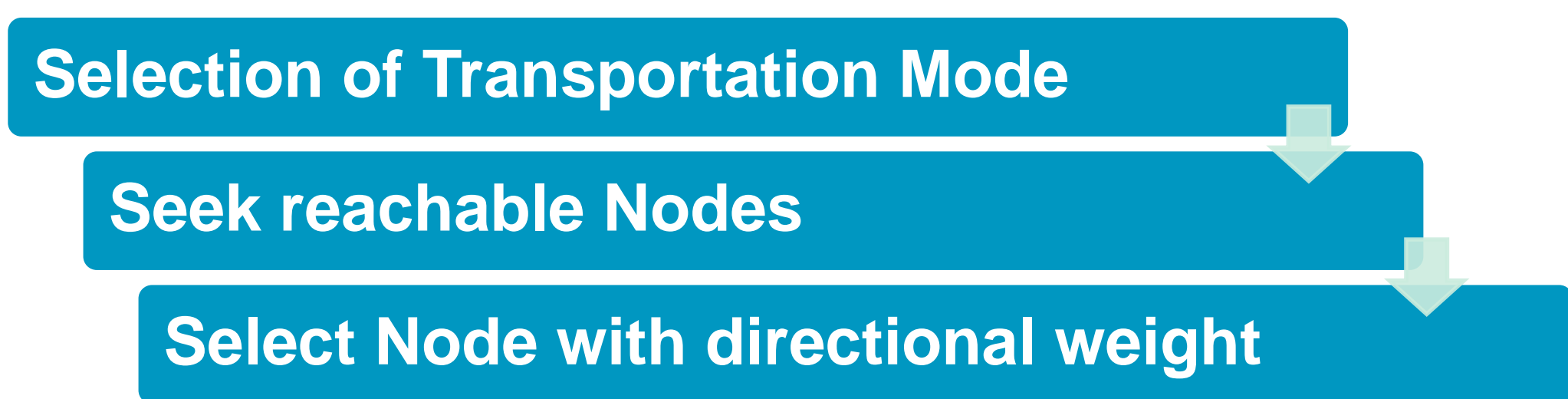
(d) Filtering

状態

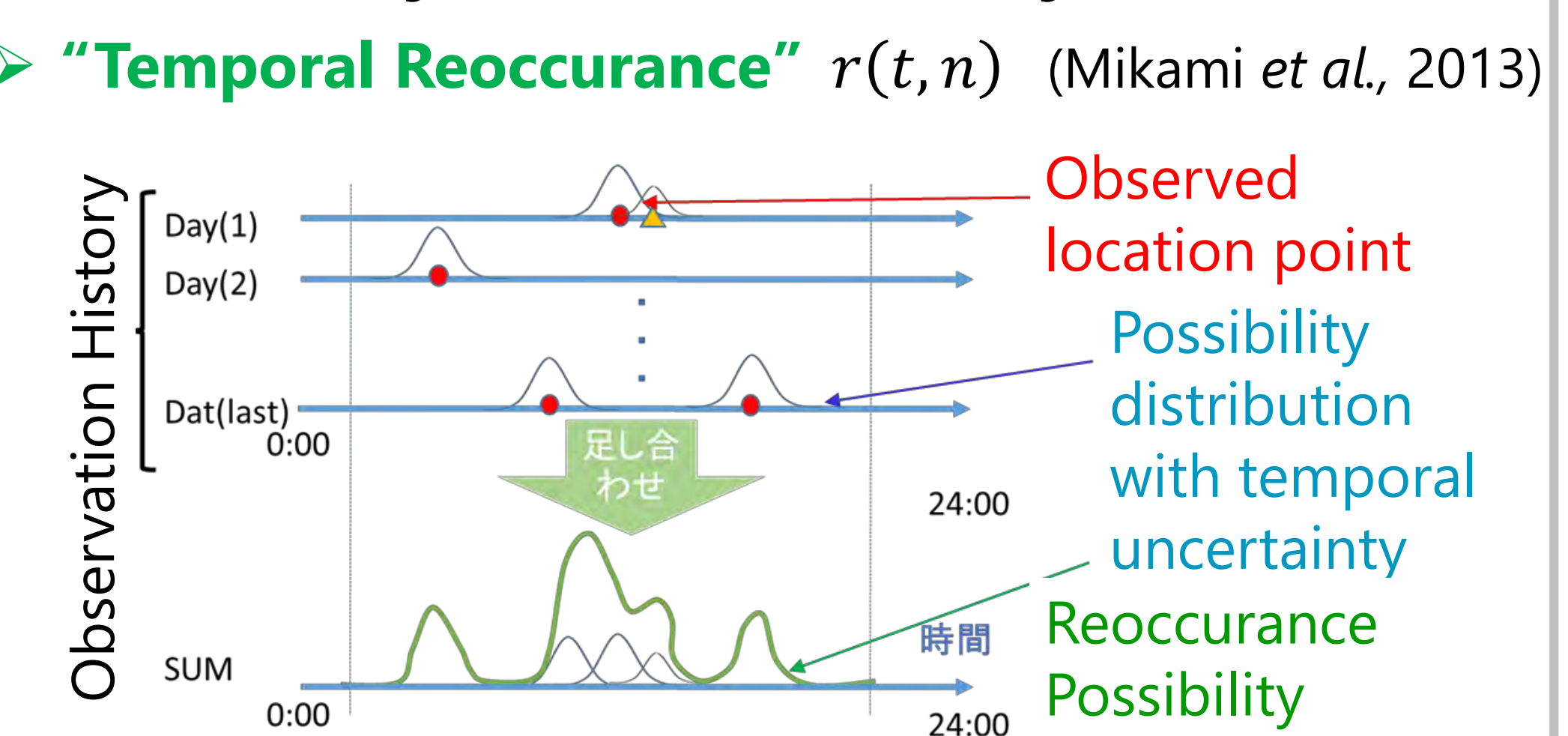
時間

- Full day movement estimation applicable to sparse CDR
- Estimation result matches
 - Observational data of the target day
 - Past locational pattern for each time stamp

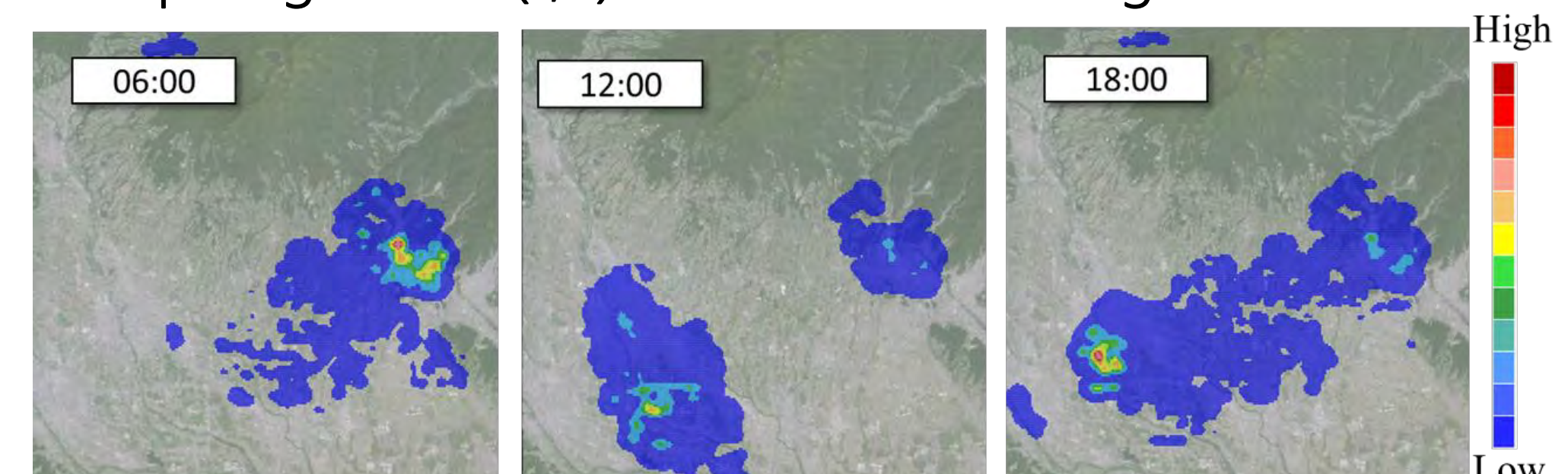
Transition Model Flow



Consistency with Past Mobility Pattern



Example figure of $r(t, n)$ distribution of a single user



Test with Actual Dataset

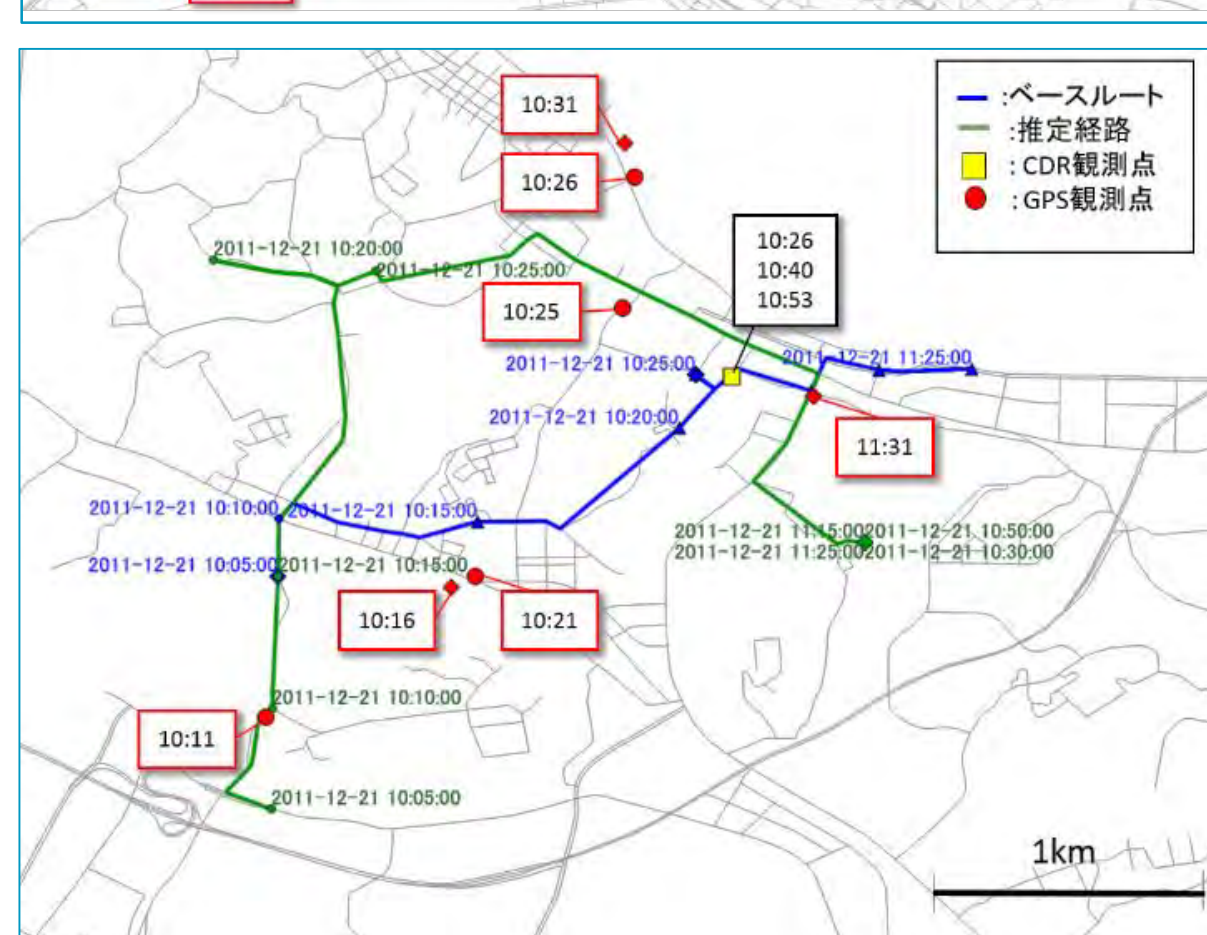
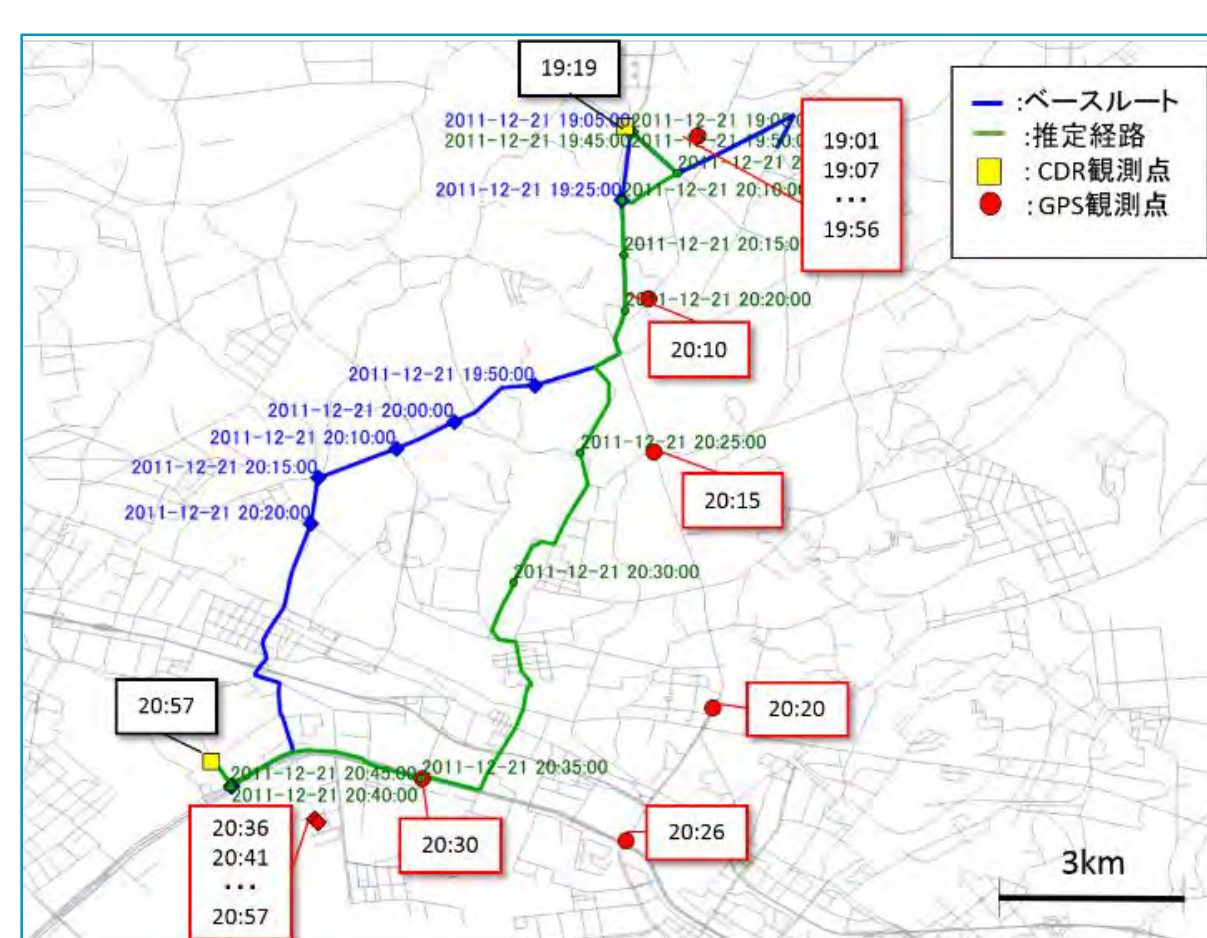
Used Data

- Sample User N: 17 (non-driver, Kanto region)
- Past observation data: CDR(2011/11/23-12/20)
- Target day observation data: CDR(2011/12/21)
- Road network data: Japan DRM

Evaluation Method

- Estimation using CDR datasets with different ave. communication interval (Default, 15min, 30min, 60min)
- Evaluation by spatiotemporal distance (d) from GPS record and 'stay' concordance
- Comparison with 'shortest path' estimation

Main Parameters	Value
Time step Δt	5 min
Start of activity t_{start}	06:00:00
End of activity t_{end}	22:00:00
Mobility mode transfer matrix	
$P = \begin{pmatrix} p_{s \rightarrow s} & p_{s \rightarrow w} & p_{s \rightarrow v} \\ p_{w \rightarrow s} & p_{w \rightarrow w} & p_{w \rightarrow v} \\ p_{v \rightarrow s} & p_{v \rightarrow w} & p_{v \rightarrow v} \end{pmatrix}$	$\begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.05 & 0.90 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{pmatrix}$
N of particles	1000



Example of estimation with closer (upper) & further (below) route selection from GPS records, compared with shortest path

Result for each CDR dataset

		Default Ver.	15min Ver.	30min Ver.	60min Ver.
d	Base Route	1044m	1004m	1183m	1462m
	Estimated Route	1012m	1069m	1230m	1443m
Stay concordance		76.2%	75.5%	75.1%	72.1%

※ ※ Red, Blue: Comparison with shortest path

- The estimation result showed a steady precision with the **stay concordance higher than 72%**, regardless of the resolution of dataset. The distance between the estimation result and the GPS data were comparatively small; **smaller than 1.5km in average.**

- The precision of our estimation varied between the users. The main reason for this is assumed to be **the difference in the number of CDR logs during trips and consistency of the target day movement and daily mobility patterns.**