

# 携帯電話のGNSSとWi-Fiデータの統合による 滞在人口分布と移動人口の同時推定手法の構築

米澤 実保<sup>1</sup>・瀬尾 亨<sup>2</sup>・布施 孝志<sup>3</sup>

<sup>1</sup>非会員 学士（工学） 東京大学大学院 工学系研究科社会基盤学専攻（〒113-8656 東京都文京区本郷 7-3-1）

E-mail: yonezawa@trip.t.u-tokyo.ac.jp

<sup>2</sup>正会員 博士（工学） 東京大学助教 工学系研究科社会基盤学専攻（〒113-8656 東京都文京区本郷 7-3-1）

E-mail: seo@civil.t.u-tokyo.ac.jp

<sup>3</sup>正会員 博士（工学） 東京大学教授 工学系研究科社会基盤学専攻（〒113-8656 東京都文京区本郷 7-3-1）

E-mail: fuse@civil.t.u-tokyo.ac.jp

近年、街中の様々なセンサから人物位置データを取得・分析し、流動人口分布を推定することへの関心が高まっている。本研究では高い時空間分解能を持ち、取得が容易である携帯電話の Global Navigation Satellite System (GNSS) データと Wi-Fi データに着目し、滞在人口分布と移動人口の同時推定手法を構築する。具体的には、人口の保存則に基づき、データ同化により GNSS・Wi-Fi データを統合する手法を定式化した。そのため、人口移動の物理的制約を考慮しつつ、詳細な行動モデルが不要な簡便な手法とした。そして、提案手法をシミュレーションデータに適用して精度を検証したうえで、実データに対して適用しその応用可能性を示した。

**Key Words:** GNSS, Wi-Fi, spatio-temporal data, data assimilation, conservation law

## 1. はじめに

流動人口分布の把握は交通・空間計画策定や交通管理上重要である。情報技術の発展により、人々の所持する携帯電話端末や街中の様々なセンサから、時空間的に広範囲で高分解能なデータを低コストで入手可能となった。そのようなデータを用いて混雑や人の移動が推定できれば、例えば都市計画・交通計画分野において、混雑集中箇所や時間帯をより高分解能に把握でき、混雑対策、災害、マーケティング、観光のための具体的な施策への貢献が期待される。

ここで、流動人口分布推定に対して、人物位置情報を取得するセンサの側から着目する。センサの種類によって取得可能データの性質が異なるため、それぞれのデータ特性に沿った活用がなされているが、本研究ではこれらのデータの中で、携帯電話の Global Navigation Satellite System (GNSS) および Wi-Fi データの2種類に着目する。両データとも対象地域全人口のうち一部の人物位置情報を観測したサンプリングデータであり、得られたデータを対象地域全人口の値に拡大することで、流動人口分布が推定できる。携帯電話によりデータを取得するため、高空間分解能で広範囲の時空間情報が容易に取得できるという長所がある。典型的には、GNSS データは端末位置情報がある時間ごとに取得して人物の移動軌跡を把握する。Wi-Fi データはアクセスポイント (AP) とよばれる固定局にアクセスした端末

情報を取得することで、当該 AP 周辺の人口密度が把握できる。

本研究では、以上で述べた性質を持つ GNSS (軌跡) データ・Wi-Fi (密度) データを統合し、エリア内の滞在人口分布と隣接エリア間の移動人口を同時に推定する手法を構築・検証する。本手法の独自性として、エリア間のマクロな人口保存則をデータ同化を用いて考慮した点がある。これにより、従来からよく用いられる、単純に拡大係数を乗じる手法などと比較して精度の向上が見込まれる。また、アクティビティモデルなどの行動モデルを用いる手法も存在するが、計算コストや精緻なキャリブレーションが必要になるため、本手法はより簡便な手法として位置づけることができる。さらには、滞在人数自体を機械学習で予測する手法も提案されているが、保存則はより高次の原則に還元しているため、より精度が高いことが期待される。以上に加え、GNSS データと Wi-Fi データを統計的に合理的な形で統合できる点も長所である。

本稿の構成は以下の通り。第2章にて既存の流動人口推定手法を整理する。第3章にて本研究における滞在人口分布と移動人口の同時推定手法を示す。第4章にて作成データによる提案手法の精度検証と、実データへの手法の適用を行う。第5章にて結論をまとめる。

## 2. 流動人口推定手法の整理

本章では、人物位置情報取得センサを用いた流動人口推定手法の既往研究を整理する。なお、センサには、カメラデータなど、全数データも存在するが、ここでは、本研究で用いるようなモバイルセンシングに基づくサンプリングデータについて整理する。センサの種類全般の説明は関本ら<sup>1)</sup>を参照されたい。既往研究においては携帯電話の基地局データを用いた流動人口分布推定がある<sup>2)</sup>。これは、複数の携帯電話の基地局で携帯端末の位置情報を取得し、集計データとすることで、流動人口を推定しているものである。基地局データを用いた流動人口推定は様々な研究がなされているが<sup>3)</sup>、基地局を単位としているために、空間分解能が数百mとなる。そこで、より高空間分解能を持つデータとして、以下ではGNSSデータとWi-Fiデータについての既往研究を示す。

典型的なGNSSデータでは、流動人口推定の研究が盛んであり、観測データをもとに相対的に混雑状況を判定するものなどすでに実用化されているものもある。研究例として、GNSSデータと行動モデルと組み合わせで流動人口を推定したものが存在する<sup>4)</sup>。GNSSデータの長所は非集計データとして軌跡データが取得できる点であるが、軌跡データは機敏な個人情報であるため端末所持者の明示的な同意が必要な場合が多く、データ数が少ないという欠点がある。

典型的なWi-Fiデータの特徴は取得可能な滞在人数データ数の多さである。これは、APへのアクセス人数は集計状態量であり、個人情報としては特に機敏ではないためである。一方、個人の移動軌跡を取得するためには各個人が複数のAPに検知される必要があり、必ずしも容易ではない。観光地や建物内など複数地点に網羅的にAPを置くことで人物位置情報を取得し、移動・滞在人口分布を推定した研究はなされているもの<sup>5)6)7)</sup>、一般の都市部でパッシブにデータ収集する場合を考えると移動人口の把握は難しい。

推定の方法論についてレビューする。行動モデルと観測データを組み合わせて流動人口を推定する研究は多い<sup>4)8)</sup>。しかし、このようなアプローチは事前に高精度な行動モデルを用意する必要があり、特に観光地といった人物によって異なる動きをする場所では適用が難しい。一方、機械学習により滞在人口を推定する研究もみられる<sup>9)</sup>。しかし、機械学習を単純に滞在人口に適用すると、人口のダイナミクスをうまく表現できず精度が良くない場合があるという問題がある。

本研究では、それぞれ長所と短所があるGNSS・Wi-Fiデータを組み合わせて流動人口分布を推定する。また、エリア間のマクロな人口保存則を考えることにより、人

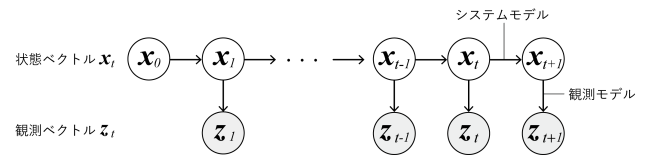


図-1 状態空間モデル

口のダイナミクスの物理的制約を考慮しつつ、行動モデルを必要とせずに観測データのみで流動人口分布を推定する手法とする。

## 3. 滞在人口分布と移動人口の同時推定手法の構築

ある地域を複数のエリアに分割（例：メッシュ状）し、流動人口推定として、それぞれのエリアの滞在人口とエリア間の移動人口を推定する手法を構築する。ここで、滞在人口分布、及び隣接セル間の移動人口とはそれぞれ、ある時刻の滞在人数、及び1タイムステップにおける隣接エリア間の移動人数の空間分布を指す。GNSSデータ・Wi-Fiデータのそれぞれをタイムステップごとにメッシュ集計データの形に変換し、変換した両データを統合することによって、滞在人口分布、及び隣接エリア間の移動人口を同時に推定するものとする。

### (1) 状態空間モデル

線形・ガウス型状態空間モデルは

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{v}_t, \quad (1a)$$

$$\mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t, \quad (1b)$$

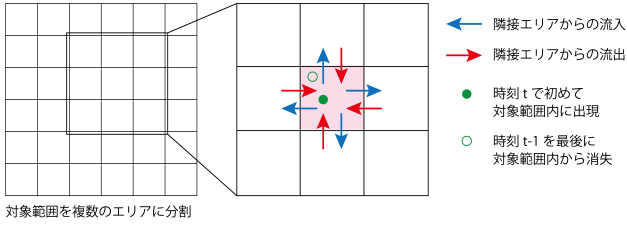
と表せる。式(1a)、(1b)はそれぞれシステムモデル、観測モデルである（図-1）。ここに、 $\mathbf{x}_t$ は状態ベクトル、 $\mathbf{F}$ は遷移行列、 $\mathbf{v}_t$ は $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$ に従うシステムノイズ（ $\mathcal{N}(\mu, \sigma^2)$ は平均 $\mu$ 、分散 $\sigma^2$ の正規分布）、 $\mathbf{z}_t$ は観測ベクトル、 $\mathbf{H}_t$ は観測行列、 $\mathbf{w}_t$ は $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{R}_t)$ に従う観測ノイズである。

状態空間モデルを用いることによって、GNSSデータとWi-Fiデータをひとつの枠組みの中で統合することができ、逐次的に状態を推定することができる。特に、状態空間モデルを構成する式が線形・ガウス分布のときには、カルマンフィルタを用いて計算が可能である。次節でそれぞれの状態・観測ベクトルの定義と、システムモデル・観測モデルの定式化を行う。

### (2) システムモデル

対象範囲を複数のエリアに分割し、そのうちのひとつのエリアをエリア*i*とする。状態ベクトルを

$$\mathbf{x}_t = \left( \dots, n_t^i, a_t^i, e_t^i, q_t^{k,i}, \dots \right)^T \quad (2)$$



対象範囲を複数のエリアに分割

図-2 保存則の説明

と置く。ここで、 $n$  は滞在人口、 $a$  は発生人口、 $e$  は消失人口、 $q$  は移動人口を表す。添字の  $i$  はエリア番号を表し、 $k$  は、エリア  $i$  に隣接するエリアの集合  $C_i$  のうちの一つのエリアを表す。 $t$  はタイムステップである。

システムモデルとして、エリア間の人口保存則を考える (図-2)。これは、状態ベクトルを構成する状態変数を用いた形で表すと、

$$n_t^i = n_{t-1}^i + \sum_{k \in C_i} q_{t-1}^{k,i} - \sum_{k \in C_i} q_{t-1}^{i,k} + a_t^i - e_t^i \quad (3)$$

$$a_t^i = \gamma_t a_{t-1}^i + \sigma_{a_t^i}^2 \quad (4)$$

$$e_t^i = \gamma_t e_{t-1}^i + \sigma_{e_t^i}^2 \quad (5)$$

$$q_t^{k,i} = \gamma_t q_{t-1}^{k,i} + \sigma_{q_t^{k,i}}^2 \quad (6)$$

と定式化できる。これをまとめて状態ベクトルの関係で記したものがシステムモデルとなる。ここで、 $\gamma_t$  は時刻  $t$  におけるトレンドであり、 $\sigma_{a_t^i}, \sigma_{e_t^i}, \sigma_{q_t^{k,i}}$  は、発生・消失・移動人口のシステムノイズである。トレンドとシステムノイズの設定は、時刻による推定人口の時間増加を複数日のデータから集計し、前の時刻の人数と比較したときの人口の増加割合の平均をトレンドに、分散に対してひとつ前の時刻の推定人口の2乗を掛けたものをシステムノイズとして用いている。

### (3) 観測モデル

携帯電話の GNSS データから個人の軌跡が、Wi-Fi データから AP 周辺の滞在人口が把握され、それぞれのサンプル率が既知の場合を考える。得られたデータを集計すると、観測ベクトルを

$$z_t = \left( \dots, \hat{n}_{w,t}^i, \hat{n}_{g,t}^i, \hat{a}_{g,t}^i, \hat{e}_{g,t}^i, \hat{q}_{g,t}^{k,i}, \dots \right)^\top \quad (7)$$

と定義できる。 $\wedge$  は各センサからの観測値に拡大係数をかけた値であることを意味している。各変数の添え字の  $w$  は Wi-Fi による観測データを表し、 $g$  は GNSS による観測データを表している。 $\hat{n}$  は滞在人口、 $\hat{a}$  は発生人口、 $\hat{e}$  は消失人口、 $\hat{q}$  は移動人口を表す。これを全エリアに対して並べたものを観測ベクトルと定義する。

観測モデルの式は、GNSS データと Wi-Fi データを統合する形で状態・観測ベクトルを用いて以下のように表せる：

$$z_t = H_t x_t + \delta_t \quad (8)$$

ここに、 $H_t$  は時刻  $t$  におけるエリア毎の GNSS・Wi-Fi データ有無をまとめた指示行列である。観測ノイズ  $\delta_t$  は以下のように設定する：観測データによる全人口は、観測された人数が成功回数  $r$ 、全人口のうちセンサを持っている人、つまり抽出率が成功確率  $p$  となる負の二項分布  $NB(r, p)$  に従う。この負の二項分布に従う確率変数を  $X$  としたときの分散  $V[X]$  を観測ノイズとして考えるが、今回は線形・ガウス状態空間モデルを扱うため、これと同じ平均と分散を持つ正規分布で近似したものを観測ノイズとして用いる (式 (9))。ただし、各センサによる観測値がそれぞれ 0 人のときは、GNSS・Wi-Fi のそれぞれの最小観測可能人数の値をとった際と同値の観測ノイズを設定した。

$$\hat{n}_{w,t}^i \sim NB(r, p) \simeq \mathcal{N} \left( \frac{r}{p}, \frac{r(1-p)}{p^2} \right) \quad (9)$$

この設定により、GNSS と Wi-Fi データを統計的に合理的な形で統合できる。

## 4. シミュレーションデータによる検証と実データへの適用

提案手法の性質を検証する。一般に現実世界の人口流動の真値は未知であるので、提案手法の精度検証のためにまずはシミュレーションにより作成したデータに対して手法を適用する。次に、提案手法の実データに対する挙動を確認するため、鎌倉市中心部で得られた GNSS・Wi-Fi の実データに対して適用した結果を示す。

### (1) 各種データと事前処理

ここでは、シミュレーションデータ・実データの概要とデータの前処理について述べる。

実データとして、Agoop 社の提供している GNSS ポイントデータと Softbank 社が提供している Wi-Fi メッシュ集計データの 2018 年 6 月分を用いる。GNSS データはデイリー ID・時刻・座標の情報を用い、Wi-Fi データは、125m メッシュで時間幅 10 分で集計されている滞在人数のデータを用いた。ここで、この Wi-Fi データはプライバシー保護の観点から、1つのメッシュに対し集計人数が 10 人以下であった場合には欠損として扱われることに留意する。データの適用範囲は鎌倉駅や鶴岡八幡宮といった人が多く集まる場所を含む約 1km 四方の範囲とする。ただし、Wi-Fi データは AP の設置がある場所のみデータが取得可能となるので、適用範囲は図-3 のようになる。

GNSS データは、メッシュ集計データにするために、事前処理を行う。ID ごとに取得された座標のポイントを時系列順に直線で結んで軌跡の形にし、1分ごとになるように座標を補完する。そしてこの軌跡データをメッ

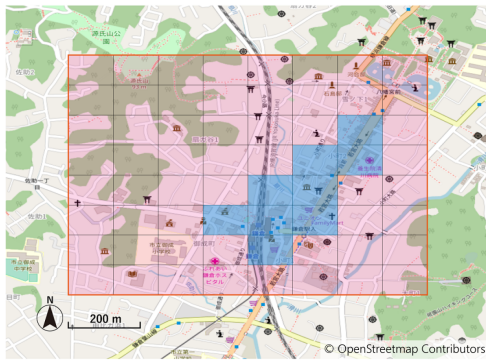


図-3 対象範囲（集計対象範囲が赤枠内、Wi-Fiデータの取得可能範囲が青枠内）

シユサイズ 125m・時間幅 10 分で集計し、滞在・移動・発生・消失人口の集計データとする。

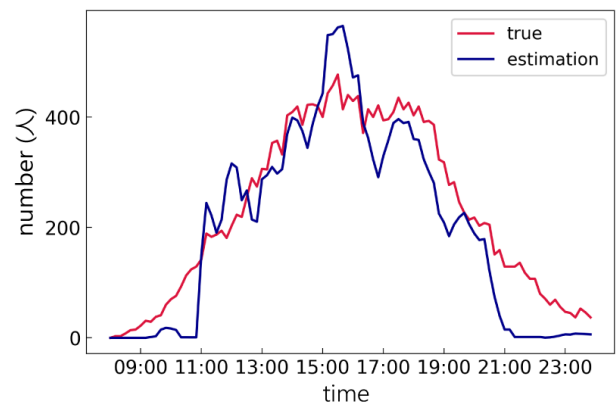
精度検証のために用いるシミュレーションデータは、実データと同スケールの対象範囲である約 1km 四方の範囲において、約 5 万人の 1 日の移動軌跡をランダムウォークで作成したものをを用いる。このデータは時刻ごとに発生可能性人数を変化させ、発生した人ごとにランダムウォークで移動あるいは消失するように時刻ごとに座標を生成することによって約 5 万人の 1 日分軌跡データとなっている。この作成したポイントデータから、それぞれ GNSS データとして 0.6%、Wi-Fi データとして 6% の ID のデータを抽出し、メッシュサイズ 125m、時間幅 10 分のメッシュデータに集計することで、観測データとした。また、プライバシー保護のための Wi-Fi データの欠損を再現するため、Wi-Fi データとして抽出したデータのうち、集計人数が 10 人以下であった時間・エリアに対しては欠損とした。

## (2) 精度検証

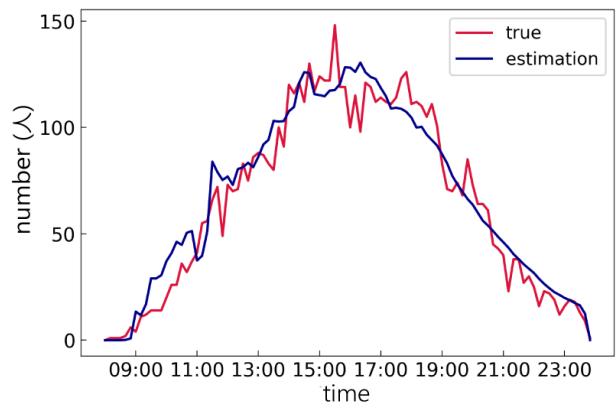
シミュレーションデータに対して提案手法を適用した結果は滞人口・移動人口について図-4 のようになる。なお、システムモデルに用いたトレンドとシステムノイズの決め方は、4.(1) で示したものと同条件でデータを 10 日分作成し、そのデータからひとつ前の時刻と比較した時間変化の平均と分散をとることで決定した。平均をトレンド、分散をシステムノイズとして推定を行った。

図-4 からわかるように、滞人口・移動人口の推定結果はともに、おおむね真値と似たような傾向を示すが、滞人口については朝や夜は推定結果が真値と比べて大きく低くなっていることがわかる。これは Wi-Fi データの観測データの欠損の影響である。

次に、推定結果の精度を数値的に検証する。提案手法の誤差 (MAE: Mean Absolute Error, MAPE: Mean Absoute Percent Error) を求め、さらに保存則を考慮し



(a) 滞人口



(b) 移動人口

図-4 シミュレーションデータにおける滞人口・移動人口の推定結果

ない従来手法（表中では GNSS 拡大手法および Wi-Fi 拡大手法と呼称）の誤差と比較した結果を表-1、表-2 に示す。滞人口・移動人口ともに、提案手法による結果は従来手法と比較して精度の向上が確認された。また、表-1 の数値は Wi-Fi データの欠損の部分も含まれているため、全体的に数値が大きいものとなっているが、Wi-Fi データの欠損がないセル・時間帯に対する同様の精度評価の数値は同表の ( ) で示されているように小さくなっていることがわかり、これらのことから本手法は従来手法と比較した際の手法の有効性が示される。以上の結果から、朝や夜など Wi-Fi データの観測数が欠損となるセルが多い場合は、推定値が大きすぎるが、昼の観測が可能な時間については、従来手法と比較して少なくとも同程度以上の推定精度を持つことが確認された。

## (3) 鎌倉実データへの適用結果

鎌倉への 2018 年 6 月 17 日 (日) の 1 日分のデータへの適用結果のうち、一部時刻のものを抜粋したものが図-5 のようになる。対象範囲において、メッシュサイズ 125m・時間幅 10 分での滞人口分布・移動人口

表-1 滞在人数の精度評価結果

	MAE (人)	MAPE (%)
提案手法	54.98 (55.93)	42.74 (19.87)
GNSS 拡大手法	160.83 (160.83)	90.54 (90.54)
Wi-Fi 拡大手法	59.59 (55.67)	49.35 (21.08)

表-2 移動人数の精度評価結果

	MAE (人)	MAPE (%)
提案手法	36.73	60.26
GNSS 拡大手法	96.59	170.60
Wi-Fi 拡大手法	推定不可	推定不可

について安定して推定できたことが確認された。滞在人口分布・移動人口の推定が同時に行えるようになったため、流動人口の移動の方向をつかむことができる。今回の例でいえば、若宮大路から鎌倉駅に流入してくる人口は、鎌倉駅から若宮大路に流出する人口よりも1日を通じて多いことが読み取れた(図-6)。このように、滞在人口分布・移動人口の同時推定により、メッシュデータで細かく流動人口を把握することが可能となり、実データへの応用可能性が示された。

## 5. おわりに

本研究では、流動人口分布の推定として、携帯電話のGNSS・Wi-Fiデータを用い、システムモデルに保存則を用いた滞在人口分布・移動人口の推定手法を構築した。そして、シミュレーションデータに基づき精度検証を行い、提案手法の有効性を示した。さらに、鎌倉のGNSS・Wi-Fiの実データに対して適用を行うことにより、手法の応用可能性を示した。

今後の展開として以下が考えられる。まず、システムモデルの高度化が望まれる。本手法は移動人数の傾向を過去データの平均としてモデル化したが、機械学習に基づく予測手法<sup>9)</sup>も有望と考えられる。この場合、森田ら<sup>9)</sup>の手法と比較し、滞在人数をデータから直接説明するのではなく、より高次の原則である保存則を通して推定するため、精度の向上が見込まれる。次に、Wi-Fiの欠損が精度に悪影響を及ぼしていることが確認

されたため、その適切な補正が必要である。最後に、サンプル率をカメラなどの全数観測データ<sup>10)</sup>を用いて内生的に決定する手法も重要である。

謝辞：本研究は国土交通省新道路技術会議の研究課題「学習型モニタリング・交通流動予測に基づく観光渋滞マネジメントについての研究開発」の助成を受けた。ここに謝意を表します。

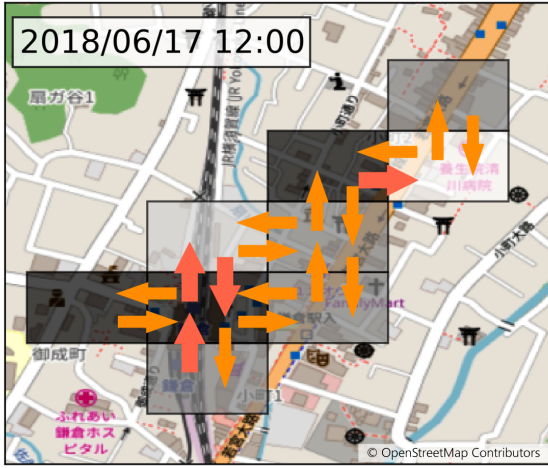
## 参考文献

- 1) 関本義秀, Horanont.T, 柴崎亮介: 携帯電話を活用した人々の流動解析技術の潮流, 情報処理, Vol. 52, No. 12, pp. 1522-1530, 2011.
- 2) 今井龍一, 藤岡啓太郎, 新階寛恭, 池田大造, 永田智大, 矢部努, 重高浩一, 橋本浩良, 柴崎亮介, 関本義秀: 携帯電話網の運用データを用いた人口流動統計の都市交通分野への適用に関する研究, 土木計画学研究・講演集, Vol. 52, pp. 1010-1021, 2015.
- 3) 大藪勇輝, 寺田雅之, 山口高康: モバイル空間統計の信頼性評価(社会・産業の発展を支える「モバイル空間統計」: モバイルネットワークの統計情報に基づく人口推計技術とその活用), NTT DoCoMo テクニカル・ジャーナル, Vol. 20, No. 3, pp. 17-23, 2012.
- 4) 矢部貴大, 関本義秀, 榎山武浩, 金杉洋, 須藤明人: パーティクルフィルタを用いた災害時におけるリアルタイムな人流推定手法, 交通工学論文集, Vol. 2, No. 2, pp. A19-A27, 2016.
- 5) Schauer, L., Werner, M., and Marcus, P.: Estimating crowd densities and pedestrian flows using wi-fi and bluetooth, in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pp. 171-177, ICST (Institute for Computer Sciences, Social-Informatics and telecommunications engineering), 2014.
- 6) 森本哲郎, 辻本悠佑, 白浜勝太, 上善恒雄: Wi-fi パケットセンサを用いた人流解析と可視化, データ工学と情報マネジメントに関するフォーラム (DEIM2015) 論文集, 2015.
- 7) 中西航, 小林巴奈, 都留崇弘, 松本拓朗, 田中謙大, 菅芳樹, 神谷大介, 福田大輔: Wi-Fi パケットセンサによる観光周遊パターンの把握可能性: 沖縄・本部半島における検討, 土木学会論文集 D3 (土木計画学), Vol. 74, No. 5, pp. I\_787-I\_797, 2018.
- 8) 布施孝志, 原田遼: 詳細な交通行動推定のためのアクティビティシミュレーションと観測データの統合に関する研究, 土木計画学研究・講演集, Vol. 57, 2018.
- 9) 森田仁美, 神谷啓太, 布施孝志: メッシュ単位のデータ駆動型シミュレーションモデル作成手法に関する研究, 土木計画学研究・講演集, Vol. 55, 2017.
- 10) 石井健太, 瀬尾亨, 布施孝志: 深層学習を導入したデータ同化による複数物体追跡手法の構築, 土木計画学研究・講演集, Vol. 59, 2019, forthcoming.

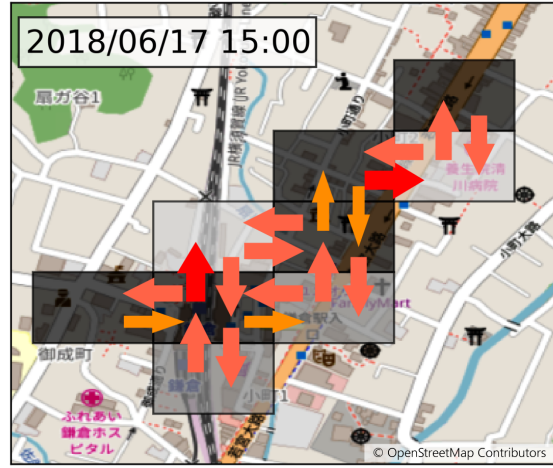
(2019. 3. 10 受付)

## ESTIMATION OF POPULATION AND PEOPLE FLOW BY USING GNSS AND WI-FI DATA FROM MOBILE PHONES

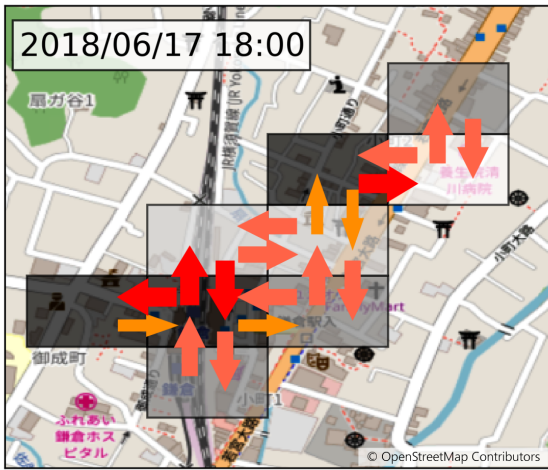
Miho YONEZAWA, Toru SEO and Takashi FUSE



(a) 12 時



(b) 15 時



(c) 18 時

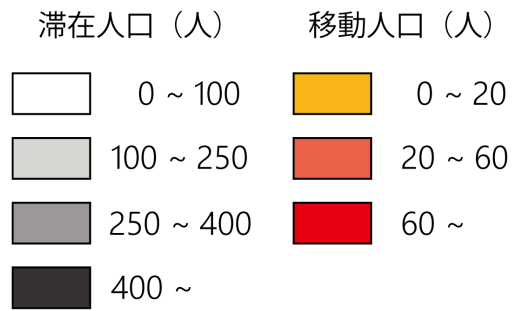
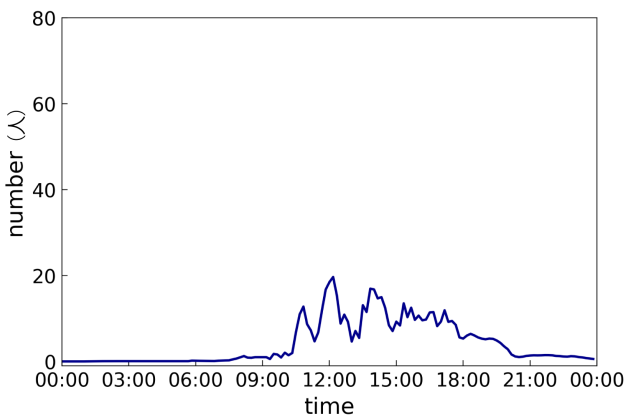
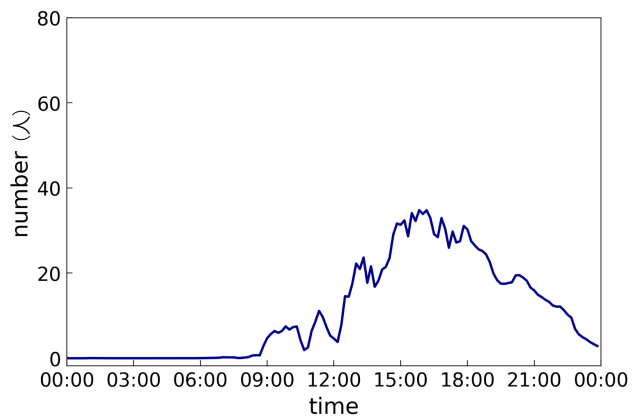


図-5 鎌倉実データへの推定滞在人口分布



(a) 鎌倉駅から若宮大路への移動人口



(b) 若宮大路から鎌倉駅への移動人口

図-6 鎌倉実データへの推定結果