

携帯電話データによる災害時のリアルタイムな人の分布の推定のための ナッジング項を持つパーティクルフィルタの提案

Estimating Distribution of Disaster Victims in Real-time using Particle Filter with Nudging Term

○須藤 明人¹, 榎山 武浩², 矢部 貴大³, 関本 義秀⁴

○Akihito Sudo¹, Takehiro Kashiyama², Takahiro Yabe³ and Yoshihide Sekimoto⁴

災害発生の直後から被災者の分布をリアルタイムに時々刻々推定することは減災に直結する公益性の高い問題である。携帯電話データを観測データとするデータ同化手法によって交通状態の推定を行う研究が近年活発になっているが、大規模災害時の都市圏全体の人々の流動や分布を推定対象とし、かつ、現実的に入手可能な携帯電話のメッシュ集計データを観測データとして、リアルタイムなデータ同化を試みた研究はこれまでなかった。本研究では、人流推定問題の特徴をふまえたナッジング項を時間発展に持つパーティクルフィルタを提案し、メッシュ集計された携帯電話データを観測データとして災害時の都市圏での人の分布の推定を試みた。実験により、ナッジング項の効果により高精度な推定が行えること及びリアルタイムな推定が関東圏においても可能であることが示唆された。

Keywords: 減災、人流推定、携帯電話データ、パーティクルフィルタ

1. はじめに

災害が発生した直後から、被災者がどこにどれだけ存在するか(以下、単に人の分布と呼ぶ)をリアルタイムに時々刻々推定することは公益性の高い問題である。逃げ遅れている人や危険なエリアに留まっている人々を防災の担当部署が把握し、いち早く救助することで損害を軽減できると考えられるからである。例えば先の東日本大震災では津波により多くの貴重な人命が失われたが、当局が逃げ遅れ等を把握できていれば被害が軽減できていた可能性がある。また、30年以内に70%の確率で発生するといわれている首都直下地震では、死因の第一位が火災によるものと見積もられている¹⁾が、人の分布と火災の状況と突き合わせて危険な状況にある人々を発見できれば、死傷者を減らせる可能性がある。

人の分布を推定するためのアプローチとして、データ主導とシミュレーションの2つが考えられる。まず、個々人の移動履歴が記録されたデータとして携帯電話のデータが近年注目を集めており、Ratti ら²⁾や Gonzalez ら³⁾以降、携帯電話のデータを活用して、人々の移動特性の理解を試みる研究が増えている。プライバシーへの配慮から、許諾を得たユーザーから得たデータのみを用いるため利用者全員のデータを取れるわけではないが、それでも関東地方で数十万人規模のユーザーから取得した位置情報付きのデータの活用が始まりつつあり、人の分布を推定するためのデータとして有力である。また、人の

移動の数理モデルに基づくシミュレーションで人の分布を推定するというアプローチも考えられる。近年、人の行動をある程度予測できる可能性が示唆される研究が報告され始めている。携帯電話データによって人の行動の規則性を見出した Gonzalez ら³⁾の研究は人の行動に関する基礎的な法則の理解に向けたメルクマールの研究と考えられ、今後の研究の進展とともに、数理モデルに基づくシミュレーションの精度が向上することが期待される。実際、道路交通シミュレーションにおいては避難行動だけでも Pel ら⁴⁾がまとめたように様々な進展がある。

ただ、これらいずれかのアプローチだけで災害時の人の分布を推定することは現状では精度等に問題があるため、データとシミュレーションを組み合わせるデータ同化技術^{5) 6)}が有力である。ノイズの混入やスパース性が問題になりやすいデータ主導アプローチと、対象の完全なモデル化が困難なシミュレーションアプローチを組み合わせより良い推定を目指すデータ同化のメリットは交通工学分野でも近年注目され、データ同化で交通状態を推定する研究が増えてきている。データ同化技術と位置づけられる手法はいくつかあるが、交通のモデルは非線形であることが多いためパーティクルフィルタ(PF)やアンサンブルカルマンフィルタを用いることが多い。Cheng ら⁷⁾は車の流れのマクロモデルと携帯電話データで多数の交差点やランプが存在しない仮定のもとで PF を用いて高速道路における交通状態の推定を行った。

1 正会員, 博士 (工学), 東京大学生産技術研究所

〒153-8505 東京都目黒区駒場 4-6-1 Ce509 e-mail: sudoa@iis.u-tokyo.ac.jp Phone: 03-5452-6415

2 非会員, 博士 (情報学), 東京大学生産技術研究所

3 学生会員, 学士 (工学), 東京大学大学院工学系研究科

4 正会員, 博士 (工学), 東京大学生産技術研究所

観測データとして、位置情報が付与されている携帯電話データを集計したメッシュ単位の人の分布を用いる。携帯電話データには、GPS で推定したユーザーの位置が記録されている GPS データと、通信した基地局が記録されている CDR の2種類がある。プライバシー保護の観点から、災害の発生時であっても匿名化されたデータしか観測データとして用いることができないと考えるべきであるため、個々人の情報は全く付与されていないメッシュ単位の集計値を観測データとして用いることを前提にした。本研究の実験では GPS と CDR から作成した以下の2種類のデータを観測データとして用いた。

GPS 集計データ

複数個の GPS 衛星が発信する信号を携帯電話で受け取ることで、その携帯電話の位置を知ることができる。本州の開けているエリアでは高精度な推定に必要な数の衛星からの信号を受信できると言われている。本研究では、利用許諾を得たユーザーだけの GPS データをもとにゼンリンデータコム社が作成・販売している「混雑統計」を実験で用いた。混雑統計は1時間単位で250m メッシュ内に存在する人数が記録されており、国勢調査等を用いて実態に合うように拡大処理がされている。図2にデータの一部を、図3にある時刻の観測値を示した。

date	hour	longitude 1	latitude 1	longitude 2	latitude 2	count
2011-03-11	1	128400000	502830000	128407500	502841250	123

(a) GPS集計データ

mesh code	Datetime	count
5339450411	20140901010000	5

(b) CDR集計データ

図2 携帯電話データのレコードのサンプル

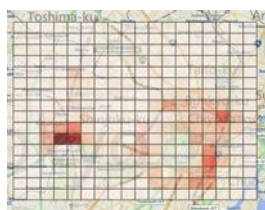


図3 ある時刻におけるGPSデータのメッシュ集計値。メッシュ内の人数が多いほど濃い色で表示。

CDR データ

携帯キャリアが保有する基地局ネットワークに自動記録される CDR データは、事実上、全ユーザー分の通信履歴が長期間に渡って存在しており、プライバシーに配慮して用いられれば非常にポテンシャルの大きいデータである。ただし、利用できる位置情報は、ユーザーが通信を行った位置ではなく、基地局の位置であるため数百メートル〜数キロメートル程度の誤差があり、GPS の数十メートルスケールの誤差と比べると大きな誤差である。

本研究では、利用許諾を得たユーザーだけを抽出した CDR データから作成されたメッシュ単位の集計値を実験に用いた。このデータは、5分ごとに250m 地域メッシュに存在する人数が記録されている。ユーザーのサンプル率は約1%である。混雑統計のような全数拡大はされていないが、通信が無い時間に存在したと思われる場所は推定して割り当ててあるので、全てのメッシュで人数を足しあわせた総数は、全ての時間でほぼ同じである。ただし、 $k=5$ のもとで k -匿名化処理が施されているので、集計した結果が5に満たない場合はゼロが割り当てられている。データは国内の携帯キャリアから提供を受けた。図2にデータの一部を示した。

(3) 道路ネットワークデータ

マルチエージェントシミュレーションがエージェントを移動させる際に道路及び鉄道ネットワークの情報が必要である。本研究では、日本デジタル道路地図協会が販売している「デジタル道路地図²⁾」を用いた。

2.3. 提案手法の詳細

(1) マルチエージェント・シミュレータと初期配置

本研究のマルチエージェント・シミュレーションには街路版 SOUND と呼ばれる岡村ら¹⁵⁾の手法に、エージェントのルートを決める機能を付加したシミュレータを用いた。エージェントのルートは、混雑を加味した通過時間をネットワークの重みとしたダイクストラ法で決める。通過時間は混雑による待ち時間 $\text{ceil}((q_l^l + 1)/V^l)$ と走行時間 L^l/v^l の和とする。 $\text{ceil}(\cdot)$ は天井関数、 q_l^l はルート探索実行時刻 t にリンク l の待ち行列にいるエージェント数、 V^l, L^l, v^l はそれぞれリンクの容量、道のり、通過速度である。シミュレータの実行には、これらのパラメータに加えて、道路ネットワーク、エージェントの初期配置、OD のセットを与える必要がある。道路と鉄道ネットワークは「デジタル道路地図」のデータを、OD は後述の行動モデルから生成したものを与える。シミュレーション開始時刻におけるエージェントの配置は人の流れデータから作成する。ただし、ゾーンごとのデータをメッシュごとに配分する必要がある。なお、人の流れデータは平常時の人の分布を与えるデータなので、発災後をシミュレーションの開始時刻とする場合は別の方法でエージェントの初期位置を決める必要がある。

(2) 震災時行動モデル

前節で述べた通り、マルチエージェントシミュレーションを実行する際にはエージェントごとの OD を与える必要がある。行動モデルは OD を生成できるものならば、任意のモデルで本研究のアプローチを実行できる。そこで本研究では、東日本大震災の発災直後の首都圏での行

¹ <http://www.zenrin-datacom.net/business/other/#statistic>

² <http://www.drm.jp/>

動をモデル化したシンプルな行動モデルを採用し、今後の首都直下地震等に対応できるより精緻なモデルの構築は今後の課題とする。本研究で用いる行動モデルは、確率 p_1 で移動を開始し、その際の移動先を最も近い駅、近くの商業施設、自宅のいずれかを確率 p_2, p_3, p_4 で選択するというものである。

(3) 状態空間モデル

一般に、状態空間モデルは状態変数 x_t 、観測データ y_t 、状態変数の時間発展を与えるシステムモデル $f_t(x)$ 、状態変数から観測データを与える観測モデル $h_t(x)$ を用いて

$$x_t = f_t(x_{t-1}, v_t), y_t = h_t(x_t, w_t)$$

と書ける。ここで、 v_t と w_t はそれぞれシステムノイズと観測ノイズである。状態変数 x_t はマルチエージェントシミュレーションのエージェント i の位置 $a_t = (a_t^1)$ と行動モデルのパラメータ $\theta_t = (\theta_t^1)$ からなり、

$$x_t = (a_t^1, a_t^2, \dots, a_t^n, \theta_t^1, \theta_t^2, \dots, \theta_t^m)$$

とする。ここで、 n と m はそれぞれエージェントの総数と行動モデルのパラメータの数である。エージェントの位置は道路ネットワークのノードの番号で与えるが、他にも様々な与え方があり得る。 a_t の時間発展はシミュレーションを実行した結果で与え、行動モデルのパラメータの時間発展は、前時刻のパラメータにガウシアンノイズを加える事で与える。すなわち、

$$f_t(x_{t-1}, v_t) = (a_t^1, a_t^2, \dots, a_t^n, \theta_t^1, \theta_t^2, \dots, \theta_t^m)$$

である。ただし、

$$\begin{aligned} a_t &= s_t(a_{t-1}) + v_t^a \\ \theta_t &= \theta_{t-1} + v_t^\theta \end{aligned} \quad (1)$$

であり、 $s_t(a_{t-1})$ はマルチエージェントシミュレーションを時刻 $t-1$ から t まで動かして得られるエージェントの配置、 v_t は平均0のガウシアンノイズである。

(4) 観測モデル

観測データとして用いる携帯電話データは種々の誤差を含んでいる。携帯電話データには様々な種類があるため、観測モデルはデータに合わせて構成する必要があるが、本研究では多くの携帯電話データで共通すると思われる観測過程として、センサによる滞在地の推定、通信間隔の内挿、利用を許諾したユーザーのサンプリング、サンプリングをキャンセルするための拡大処理についてモデル化した。2.2(2)のCDRデータに施されているk匿名化は無視している。

あるメッシュに存在する人数を x とし、その人たちが単位時間の間に最低1回でも通信する確率を β とする。データ提供元の携帯キャリアの普及率を κ 、データの利用を許諾したユーザーの比率 γ とすると、 β, κ, γ が時空間で偏りが全く無いと仮定のもとでそのメッシュで観測データに記録される人の数 d^1 の確率分布は二項分布に従うので、 $d^1 \sim B(x, \beta\kappa\gamma)$ となる。ここで、 $B(n, p)$ は試行数 n 、成功確率 p における成功回数を与えるような二項

分布、チルダ記号は左辺の確率変数が右辺の確率分布に従うことを示す。携帯キャリアが取得する位置情報はGPSによる精度や基地局とユーザーの距離の分だけ誤差が含まれているが、 d^1 はその誤差が無い時に携帯キャリアが生データを集計して得られるメッシュ集計値と解釈できる。実際には誤差は存在するのでメッシュ集計値に対してガウシアンノイズの和で誤差が影響すると仮定すれば、携帯キャリアが許諾ユーザーの通信ログから得るメッシュ集計値 d^3 は、 $d^2 \sim \mathcal{N}(0, \sigma_s)$ のもとで $d^3 = d^1 + d^2$ となる。ここで、 $\mathcal{N}(0, \sigma_s)$ は平均0分散 σ_s の正規分布である。この集計値に対し、通信が無い時間帯のユーザーの位置を推定することであたかもユーザーが全ての時間で通信を行っているような補間を携帯キャリアが行うとする。その処理が全てのメッシュで一律に定数 β で割る処理で良く近似されると仮定すれば、携帯キャリアから提供されるデータ d は $d = 1/\beta(d^1 + d^2)$ と書け、ここから尤度 $p(d|x)$ を計算すればよい。データによってはさらに全数が合うように拡大しているので、その場合は $d = 1/\beta\kappa\gamma(d^1 + d^2)$ のもとで尤度を計算する。

(5) パーティクルフィルタと提案分布

前節で導入したシステムモデルと観測モデルが非線形・非ガウスモデルであることから、本研究ではデータ同化の手法としてPFを用いる。ただし素朴なPFではなく、ナッジング項を持つことで観測値に近いパーティクルが得られやすい性質を持つ提案分布を用いる。ナッジング項を持つ提案分布を採用した理由は、素朴なPFを災害時の人流推定問題に適用する際の懸念として、いずれのパーティクルも観測値と大きく異なる推定値になってしまうという問題があるからである。人の流れを推定する問題の特徴として、高次元であることに加えて、高い精度での予測が可能なモデルが知られていないことや、急激にパラメータが変化し得ることがあげられる。重みを調整すればパーティクルの時間発展を自由に与えられることはAppendix 1を参照されたい。

以下、具体的にパーティクルの時間発展を与える。 $\varphi(a)$ 及び $\varphi^i(a)$ をエージェントの配置 a をメッシュ単位に集計した整数値ベクトル(以下、メッシュ集計量とも呼ぶ)とその i 番目の要素とし、 $\Delta = (\Delta_{i,j})$ を再配置によってメッシュ i から j に移動させるエージェント数を表す行列とする。メッシュ内のエージェントを区別しなくてよければ、 a_t と Δ が与えられれば再配置を施したエージェントの位置が定まるので、その配置と移動前のエージェントの配置の差を $\chi(a_t, \Delta)$ と書く。 $\delta_{i,j}$ をメッシュ i と j の間の距離とすると、 Δ による再配置でエージェントが移動する総距離 D_Δ は $\sum_{i,j} \Delta_{i,j} \delta_{i,j}$ であるので、 D_Δ を最小にするようにエージェント配置 a_1 から a_2 にエージェントを再配置させる Δ は、

$$\Delta = \underset{\Delta}{\operatorname{argmin}} D_{\Delta} = \underset{\Delta}{\operatorname{argmin}} \sum_{i,j} \Delta_{ij} \delta_{ij} \quad (2)$$

$$s.t. \sum_j \Delta_{ij} \leq \varphi^i(a), \sum_i \Delta_{ij} \leq \varphi^j(a'),$$

$$\sum_{ij} \Delta_{ij} = \min \left\{ \sum_i \varphi^i(a), \sum_j \varphi^j(a') \right\}, \Delta_{ij} \geq 0$$

で得られる。ここで、1 番目の制約は移動するエージェントが移動元のメッシュのエージェント数を超えないという制約で、2 番目の制約は移動してくるエージェント数が再配置後として望ましいエージェント数をこえないという制約である。それぞれの不等号は、エージェントの総数が一致しない場合に対応する。式(2)は a と a' のメッシュ集計から Δ を与える関数とみなせれる³ので、元のエージェント配置 a と再配置後のエージェント配置のメッシュ集計 $\varphi(a')$ から式(2)を満たす Δ を与える関数を $\psi(a, \varphi(a'))$ と書く。

また、シミュレーションが与えるエージェント配置のメッシュ分布 $\varphi(s_t(a_{t-1}))$ を観測値 y_t に近づけたメッシュ分布は、 K を 0 から 1 の任意の実数として $\varphi(s_t(a_{t-1})) + K(y_t - \varphi(s_t(a_{t-1})))$ で与えられる。 K はシミュレーションの結果をどれだけ観測値に近づけるかを定めるパラメータである。

以上から、パーティクル i の状態量 a_t^i の時間発展を、

$$a_t^i = s_t(a_{t-1}^i) + \chi(a_{t-1}^i, \Delta) + v_t^a \quad (3)$$

s.t.

$$\Delta = \psi(s_t(a_{t-1}^i, \tilde{a}^i),$$

$$\tilde{a}^i = \varphi(s_t(a_{t-1}^i)) + K(y_t - \varphi(s_t(a_{t-1}^i)))$$

とすれば、パーティクルの次ステップの状態量が、シミュレーション結果をエージェントの移動量を最小にする制約のもとで観測値に近づけるように再配置したものになることがわかった。この式(3)を、位置の状態量 a_t の時間発展として採用する。 $\chi(a_{t-1}, \Delta)$ がナッジング項である。この場合の a_t に関する提案分布は 式(3)の v_t^a の確率密度関数を平均が a_t になるよう平行移動したものである。

ただし、交通工学のマルチエージェントシミュレーションでは、ひとつのエージェントが複数人を表現するために拡大係数を持っていることがあり、その場合は、同じメッシュに存在しているエージェントでも拡大係数が違えば同一視できないので、どのエージェントをフロー

に従って再配置するかを決めねばならない。その問題は、ナッジング問題の変種とみなすことができ、ナッジング問題の貪欲法での解法の自然な拡張で解を得られる。我々が一般的なコンピュータで実験した限りでは実用的な時間で解を得ることができている。

また、エージェントの総数がシミュレーション結果と観測値で異なる場合、式(3)に従ってエージェントを再配置させた結果のメッシュ分布は $\tilde{a}^i$ に一致しない。 $\tilde{a}^i$ と一致するようなエージェントの配置を得たい場合は、総数を一致させるために、エージェントの拡大係数に $\sum_j \tilde{a}^j / \sum_i \varphi^i(a)$ をかけたものをその時刻の拡大係数とする方法や、対象エリアの内外にエージェントを移動させて総数を一致させるという処理を追加する。

(6) Earth Mover's Distance による時間発展の計算

前節で、式(3)をパーティクルの時間発展として用いることを説明したが、実際に PF を実行する際には式(2)で与えられる最適化問題の解法が必要である。Appendix2 より EMD を求める最適化問題と式(2)が同等とわかるので、式(2)の Δ を求めるためには、再配置前のメッシュ集計 $\varphi(a)$ を供給側の分布、再配置後のメッシュ集計 $\varphi(a')$ を需要側の分布、 a のエージェントを移動させる荷物、メッシュ間の距離を分布間の距離として EMD を求め、その EMD を与えるフローを Δ とすればよい。EMD の解法は 13) が与える解法以外にもいくつか提案されており、それらを用いることもできる。

2.4 提案手法のアルゴリズム

以上をまとめて、提案手法のアルゴリズムの手順は以下のフローとなる。ただし、 J はシミュレーション終了時刻、 T_j は、 $j = 1$ ならばシミュレーションを開始した時刻から最初の観測値が得られるまでの時間間隔、 $j > 1$ ならば前回観測値が得られてから次の観測値が得られるまでの時間間隔である。

提案手法のフロー

1. N 個のマルチエージェント・シミュレータ $\{S_i\}_{i=1}^N$ に、道路と鉄道のネットワークデータ、エージェントの初期配置、行動モデルのパラメータの初期値をセットする。
2. i 番目のパーティクルの状態量 $\{x^i\}_{i=1}^N$ に、シミュレータ S_i のエージェント初期値配置と行動モデルの初期パラメータを割り当てる。 j に 1 をセットする。
3. 全ての i について 3-1 から 3-3 を実行する
 - 3-1. 行動モデルからエージェントの OD セットを生成し時間 T_j だけシミュレーション S_i を進める
 - 3-2. 式(1)で θ_j を、式(3)で a_j を更新することで、パーティクルの状態量を更新する。 a_j の更新には観

³式(2)の最適解は一意に定まらないことがあるので、厳密にはそのままでは関数とはみなせないが、2.3(6)で与える Rubner の解法は解を 1 つだけ出力するため、ここではその解を Δ として ψ を $a, \varphi(a')$ から Δ を定める関数とする。

- 測値 y_{j+1} と3-1のシミュレーション結果を使う
- 3-3. 観測値 y_{j+1} と s_i のエージェントの配置を用いて
 i 番目のパーティクルの重みを計算する。
4. 重みに従ってアンサンブルをリサンプリングする
5. $j < J$ ならば終了. そうでなければ j に $j+1$ をセットしてステップ3に戻る。

3. 実験

提案分布の効果を確認するため、パーティクル数を1としてふたつの携帯電話データについて予備実験を行った。 $n > 1$ の実験結果は口頭発表時に報告予定である。

3.1 実験の設定

災害時の人の分布の推定を GPS データから作成された「混雑統計」と CDR から作成された2.2(2)のデータを観測値として実施した。すなわち観測データは250mメッシュの人の分布である。災害時の推定には3.5の行動モデルから得たODを用いたが、任意の行動モデルで本アプローチが有効であることを示すため、人の流れデータをそのままODとして用いるシミュレーションで平常時の人の分布推定も試みた。対象とするエリアは千代田区、中央区、新宿区とした。約41000のエージェントでおよそ210万人の行動をシミュレートした。EMDのフローを求める部分はC++で、それ以外はJava 1.8で構築し、AWS EC2のc3.8xlargeで実行した。

エージェントは、1日のうちに1回でも対象エリアに入るトリップを持つものを全てを用いており、外から対象エリア入ってきて推定時刻の終わりには対象エリアからいなくなるエージェントも含んでいる。行動モデルのパラメータ (p_1, p_2, p_3, p_4) は、発災から1時間毎に $(0,0,2,0,0,5)$, $(0,15,0,2,0,0,6)$, $(0,15,0,2,0,0,7)$, $(0,2,0,2,0,0,8)$ と変化させた。4時間後以降は $(0,2,1,2,0,0,8)$ で固定した。式(1)の v_t は無視した。駅や自宅に向かう人が発災後徐々に増えていくという設定である。ナッジング項の係数 K は1とし、前述の通りパーティクル数を1とした。また、シミュレータのエージェントが持つ拡大係数の総計 $X_{t,sum}^i$ と、観測値の総計 Y_{sum} が一致しないので、エージェントの拡大係数に一律に $Y_{sum}/X_{t,sum}^i$ の積を取る処理を観測値が得られる度に行った。シミュレータは5秒単位でエージェントを動かすように設定した。

3.2 GPS 集計データによる実験

東日本大震災当日2011年3月11日の15:30から22:30のGPSデータ(混雑統計)を用いて人の分布の推定を行った。また、その前日にあたる3月10日のデータを用い、ODを人の流れデータとした実験もあわせて行った。図3に式(3)の a_t と観測値 y_t の相関係数と、RMSEをメッシュ当たり人数で割った値をそれぞれ $K=1$ と $K=0$ の場合について示した。観測値に近づけることで精度が向上していることがわかる。ナッジング項により観測値を与

えた時点では観測値とほぼ同じ値をとることがわかる。図5に15:00時点の推定結果を示した。

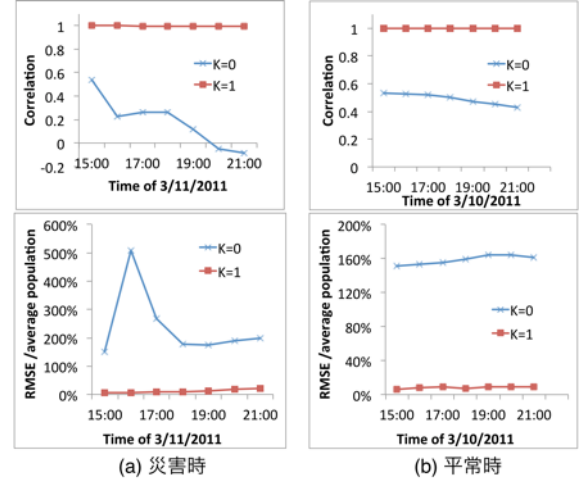


図4 1時間ごとに観測値が得られる前提でのGPSデータによる実験結果。

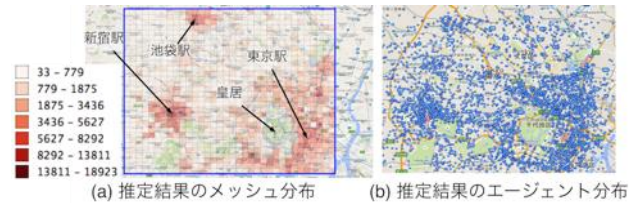


図5 2011年3月11日15:00の推定結果の(a)メッシュ分布(色の濃いメッシュに多くの人が存在)と(b)エージェント分布。(a)の青枠が推定対象のエリアであり、(b)は視認性向上のため30分の1程度にエージェントを間引いて表示している。

3.3 CDR 集計データによる実験

観測値として携帯キャリアがCDRデータから作成した3.2(2)のデータを用いて2014年9月1日の実験を行った。データには $k=5$ での k -匿名化処理が施されているため、 $s_t(a_t)$ が5以上ならば5に、5未満ならば $s_t(a_t)$ に観測データが0のメッシュの値を更新した。観測データが平常時のものなので、ODは人の流れデータのODをそのまま用いた。3.2と同様の実験を行い、図4に結果を示した。結果の傾向はGPSデータとほぼ同じである。

また、観測値が得られない間の誤差を評価するため、3時間に1回観測値を与え、観測値を得たあとの1時間毎の観測値との誤差を評価し、図7に示した。これは次の観測値が得られるまでの予測を行う際の誤差と解釈できる。パーティクル数が1であっても、ナッジング項の効果によって、シミュレーションだけで推定するよりも良い精度で予測できている。ただ実際の運用を見据えると予測誤差はまだ改善が必要であり今後の課題である。

3.4 処理時間

リアルタイム性の評価のため、3.2, 3.3の実験の処理時間を評価した結果を図8(a)(b)に示す。1時間分の処理を

1200 秒以下で行えており、リアルタイムな推定が可能であることを示唆している。また、エリアを関東に広げて

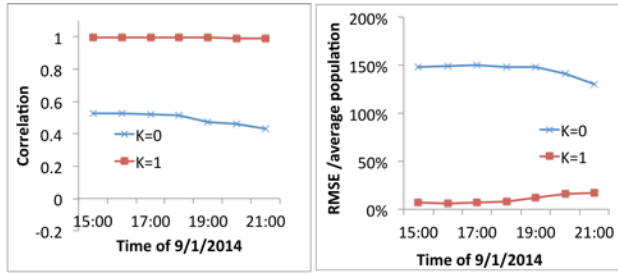


図 6 1 時間ごとに観測値が得られる前提での CDR データによる実験結果.

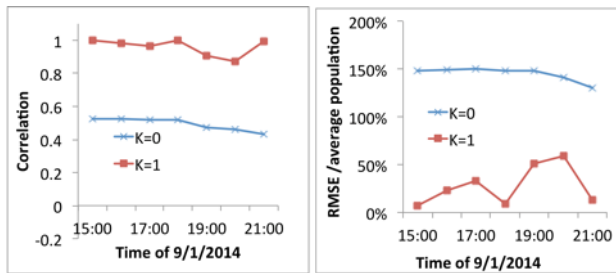


図 7 CDR データを用いた際の予測誤差の評価

実験した場合の処理時間を図 8(c)(d)に示す。関東エリアでの実験は観測データが 1km メッシュで得られる前提で実施し、DRM 道路ネットワークのうち幅員 5.5m 未満の”その他道路”を削除した。EMD の算出処理は高速化のため EMD- L_1 と呼ばれる手法を拡張したものをを用いた。3 次メッシュでリンクをある程度間引けば、関東エリアでもリアルタイムに実施できることを示唆している。メッシュサイズを 500m メッシュ等にしての実験は今後の課題である。

4. おわりに

災害時の人の分布をリアルタイムに推定するため、携帯電話データと人の流れデータを用いた震災時行動モデルを組み合わせてパーティクルフィルタ(PF)で推定を行う手法を提案した。問題の特徴をふまえ、素朴な PF ではなく、EMD を与えるフローで計算されるナッジング項を持つ提案分布を用いた。今後の課題として、災害時に想定されるインフラの被害を考慮、モデルの洗練と予測誤差の改善、パーティクルの縮退への対処があげられる。前述の通り、災害時のインフラ被害の考慮は必須だが、本論文ではインフラへの被害がない前提で議論した。本論文で提示した行動モデル、観測モデル、パラメータのダイナミクスはいずれもシンプルまたは大胆な仮定をおいたもので、洗練させる余地が十分残っている。予測誤差は行動モデルの洗練に伴って改善すると考えられる。また、パーティクルの縮退を防ぐため、例えば 12)で提案されている方法の適用が考えられる。このアイデアはナ

ッジング項の K の自由度に着目した手法であり、本研究にも適用できる可能性がある。また、ナッジング項を移

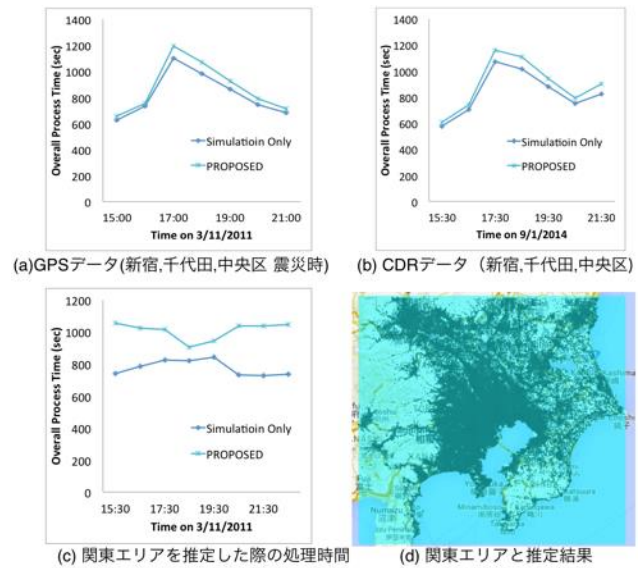


図 8 1 時間ごとの処理時間。(a)(b)新宿,千代田,中央区(c)関東エリア (d) 水色のエリアが関東エリアの対象地域. 黒の部分が 1 人以上存在するとされたエリア.

動量最小という制約条件で求めたが、この制約はどのような条件下で最適なのか検討が必要である。少なくとも行動モデルが完全なランダムな行動よりも良い予測性能を持つなら、移動量最小という制約下で得られる解が予測の意味で最適と考えられるが、厳密な理論的解析や実験による検証が必要である。

謝辞

本研究は、総務省「G 空間プラットフォームにおけるリアルタイム情報の利活用技術に関する研究開発」や JST CREST「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化」の支援を受けた。また、後者の CREST では、データ同化の部分において、統計数理研究所の樋口知之所長、中野慎也博士、齋藤正也博士から有益な助言を頂いた。

Appendix 1 パーティクルフィルタの提案分布の任意性

PF の時間発展は重みを変えることで任意に与えてよいことの確認をする。フィルタ分布 $p(x_t|y_t) = (1/A)p(y_t|x_t)p(x_t|y_{1:t-1})$ に一期先予測 $p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})$ を代入すると、 $p(x_t|y_t) =$

$$\frac{1}{A} \int p(y_t|x_t)p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \quad (A-1)$$

が得られる。ここで A は規格化定数である。 $p(x_t|x_{t-1})$ は事前遷移確率と呼ばれ、素朴な PF のアルゴリズムで

はパーティクルの状態変数の時間発展を与える。式(A-1)は任意の提案分布 $q(x_t|x_{t-1})$ を用いて、以下のように書きなおすことができる。

$$p(x_t|y_t) = \frac{1}{A} \int p(y_t|x_t) \frac{p(x_t|x_{t-1})}{q(x_t|x_{t-1})} q(x_t|x_{t-1}) p(x_{t-1}|y_{t-1}) dx_{t-1}$$

ここから、パーティクルの状態変数 x_t の時間発展を $p(x_t|x_{t-1})$ のかわりに提案分布 $q(x_t|x_{t-1})$ で与え、その辻褄を合わせるため、尤度 $p(y_t|x_t)$ と $\frac{p(x_t|x_{t-1})}{q(x_t|x_{t-1})}$ をかけた

値を正規化して新たな重みとみなしそれ以外は素朴なPFと同様に計算すればフィルタ分布 $p(x_t|y_t)$ が得られることがわかる。このときパーティクル i の重み w_i は新たな正規化係数を A' として

$$w_i = \frac{1}{A'} p(y_t|x_t^i) \frac{p(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i)} \quad (A-2)$$

で与えればよい。

Appendix 2 EMD と同等であることの説明

EMD はふたつの分布の違いを測る指標と解釈できる。EMD の文脈では分布をシグネチャと呼ばれる形式で表現し、次元数が N のシグネチャは $S = \{s_j = (m_j, w_j)\}_{j=1}^N$ と定義される。ここで、 m_j は分布の j 番目の要素の位置、 w_j はその位置の重みである。いま次元数が m, n のふたつの分布が与えられ、それらのシグネチャが $P = \{(p_j, u_j)\}_{j=1}^m, Q = \{(q_j, v_j)\}_{j=1}^n$ であるとし、また、位置 p_j と q_j の距離を d_{ij} とする。このとき、シグネチャ P と Q の間のEMD (以下、 $EMD(P, Q)$ と記載) は、 P を供給側、 Q を需要側と解釈した次のような輸送問題の解を用いて定義される。すなわち、 P の位置 $\{p_j\}$ にそれぞれ u_j 個の荷物が積まれており、 Q では位置 $\{q_j\}$ にそれぞれ v_j 個の荷物が入る倉庫があるもとで、荷物あたりの平均移動距離を最小にするように供給側 i から需要側 j に運ぶ荷物の量 f_{ij}^* を求め、規格化された荷物の総移動距離 $\sum_{i,j} f_{ij}^* d_{ij} / \sum_{i,j} f_{ij}^*$ を $EMD(P, Q)$ と定義する。ただし、 P の荷物の総数と Q で受け入れられる荷物の数が異なる場合は、 P の荷物を全て運びきるか Q の受入数がいっぱいになるかのいずれかまで荷物を運ぶとする。つまり、 $EMD(P, Q)$ は次の最適化問題の解である。

$$EMD(P, Q) = \min_{F=\{f_{ij}\}} \frac{\sum_{i,j} f_{ij} d_{ij}}{\sum_{i,j} f_{ij}} \quad (A-3)$$

ただし、 f_{ij} には

$$\sum_j f_{ij} \leq p_i, \sum_i f_{ij} \leq q_j, \sum_{i,j} f_{ij} = \min\{\sum_i p_i, \sum_j q_j\}, f_{ij} \geq 0. \quad (A-4)$$

という制約がつく。 f_{ij} はフローと呼ばれる量であり、位置 i から j へ運ぶ荷物の総数である。この最適化問題は線

形計画問題 $\min c^T x \quad s.t. \quad Ax = b, x \geq 0$ の行列 A がスパースである特殊な場合であり、13)に効率的な解法が示されている。

ここで、式(A-3)を、分母が定数であることに注意しつつ、EMDを与えるフローを求める式に書き直すと、

$$f_{ij} = \operatorname{argmin}_{f_{ij}} \sum_{i,j} f_{ij} d_{ij} \quad (A-4)$$

となる。式(2)の $\Delta_{ij}, \delta_{ij}, \varphi^i(a_1), \varphi^j(a_2)$ を式(A-3)(A-4)の f_{ij}, d_{ij}, p_i, p_j とおきかえれば、ふたつの最適化問題は全く同じ式で与えられるので、式(2)の Δ_{ij} はEMDの解法で求めることができる。

参考文献

- 1) 中央防災会議：首都直下地震の被害想定と対策について、2013年
- 2) Ratti, C. et al.: Mobile Landscapes: using location data from cell phones for urban analysis, *Environment and Planning B: Planning and Design*, 33(5), 727-748, 2006.
- 3) Gonzalez, M. et al.: Understanding individual human mobility patterns, *Nature*, 453, 779-782, 2008.
- 4) Pel, J. et al.: "A review on travel behaviour modeling in dynamic traffic simulation models for evacuations." *Transportation* 39.1 (2012): 97-123.
- 5) 樋口編：データ同化入門：次世代のシミュレーション技術。朝倉書店、2011.
- 6) 樋口知之。"粒子フィルタ。" 電子情報通信学会誌 88.12 (2005): 989-994.
- 7) P. Cheng, Z. Qiu, and B. Ran. Particle filter based traffic state estimation using cell phone network data. *ITSC*, 2006.
- 8) 佐々木ら。"ベイズ型状態空間モデルを用いた交通量変動の分析。" *交通工学* 47.2 (2012): 27-32.
- 9) Herring, Ryan, et al. "Estimating arterial traffic conditions using sparse probe data." *ITSC*, 2010.
- 10) 中村ら：パーティクルフィルタを用いた都市圏レベルの人の流れの推定手法の構築, 土木学会論文集 D3(土木計画学), Vol.69, No.3, pp.227-236, 2013
- 11) Doucet, Arnaud, and Adam M. Johansen. "A tutorial on particle filtering and smoothing: Fifteen years later." *Handbook of Nonlinear Filtering* 12 (2009): 656-704.
- 12) van Leeuwen, Peter Jan. "Nonlinear data assimilation in geosciences: an extremely efficient particle filter." *Quarterly Journal of the Royal Meteorological Society* 136.653, 2010.
- 13) Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. "The earth mover's distance as a metric for image retrieval." *International journal of computer vision* 40.2 (2000): 99-121.
- 14) Sekimoto, Yoshihide, et al. "Pflow: Reconstructing people flow recycling large-scale social survey data." *IEEE Pervasive*

Computing 10.4 (2011): 0027-35.

検証.” 交通工学研究発表会論文報告集 (1996): 93-96

15) 岡村ら. “一般街路網シミュレーションモデルの開発と