

時空間ネットワーク解析を用いた主動線抽出

浅原 彰規^{1,a)} 寺本 やえみ¹ 丸山 貴志子¹ 柴崎 亮介²

受付日 2012年4月6日, 採録日 2012年10月10日

概要: 本研究は蓄積された位置情報の時系列から主動線を自動抽出する方式の研究である. ここで主動線とは位置情報の時系列を要約し, 代表的な動きを抽出したものを指す. 従来, 動線形状の類似度に基づく主動線抽出が知られているが, 交通需要や渋滞箇所を把握するには交通網に合わせた分析が必要となる. そこで本研究では, 動線を離散化してその遷移グラフから時空間ネットワークを構築し, 経路探索を適用して主動線を抽出する方式を提案する. 最後に, 東京都市圏の人の流れのデータを用いた主動線抽出実験により, 提案方式の有効性を確認した.

キーワード: 動線解析, GIS, 位置情報, 交通情報

Representative Path Extraction with Spatio-temporal Network Analysis

AKINORI ASAHARA^{1,a)} YAEMI TERAMOTO¹ KISHKO MARUYAMA¹
RYOSUKE SHIBASAKI²

Received: April 6, 2012, Accepted: October 10, 2012

Abstract: In this research, a method to extract representative paths from stored positioning data is proposed. The representative paths are defined as summaries about stored many time-series of positioning data. Existing methods to extract representative paths are based on similarity of shapes between trajectories. However, other approaches are needed to extract abstract information about the traffic situation. Thus the proposed method is a representative-paths-extraction algorithm which is based on the path finding over the spatio-temporal network. These extraction algorithms are evaluated experimentally, using the PFLOW data. As the result, the effect of the proposed method is confirmed.

Keywords: trajectory analysis, GIS, positioning data, traffic information

1. はじめに

近年, ICT (情報通信) 技術を用いて都市を最適化するスマートシティという概念が注目されている. スマートシティにおいて大きな位置を占める取り組みの1つに交通最適化がある. ここでいう交通最適化とは, 交通需要や設備投資, 環境負荷などを考慮した最適化な交通手段の提供をさす. 交通需要はある地点からある地点への人や物の移動に対する要求であり, 当該地点間の交通手段の提供によっ

て満たされる. 交通手段を提供するためのインフラ整備には設備投資費用や環境負荷をとまなうため, 過剰な交通手段の提供は効率的ではない. そこで, 定期的に交通需要を調査し, それにあわせて最適に交通手段を提供していくことが求められる.

従来, 交通需要を把握するためにアンケートや目視確認による交通需要調査が行われているが, これらの調査には多大な時間と費用を要するため, 数年から数十年に1度調査を行い都市計画を策定するのが一般的であった. スマートシティでは, 様々な都市のセンシング情報を集めて分析することが可能になるため, 時々刻々と変化する交通需要の把握が可能になる. たとえば, プローブカーなどの道路交通情報 [1], 鉄道やバスで用いられる IC カードによる乗車記録, 携帯電話に搭載されている GPS による人の位置

¹ 日立製作所中央研究所
Hitachi Ltd., Central Research Laboratory, Kokubunji,
Tokyo 185-8601, Japan

² 東京大学空間情報科学研究センター
Center of Spatial Information Science, the University of
Tokyo, Kashiwa, Chiba 277-8568, Japan

a) akinori.asahara.bq@hitachi.com

情報などを用いて、日々の都市の交通需要や実際の交通量を把握・予測し、次の計画立案に反映させるのである。

本研究では特に移動体の位置情報に着目する。ここで、移動体は人や車両などの移動する物体であり、移動体の位置情報は、緯度経度など移動体の位置を示す情報とその位置にあった時刻の組で表される。たとえば、スマートフォンなどのモバイル機器 [2], [3], プローブカーの GPS から得られる情報は車両や人の位置情報に相当する。これらの人や車両などの移動体の位置情報は、それぞれの移動の出発地や目的地、その間の交通手段の情報などを含んでいる。そのため、多数の移動体の位置情報を集積することにより、交通最適化に活用できると考えられる。たとえば、移動の出発地と目的地が分かれば交通需要を得られ、その交通手段が分かれば移動体の移動経路が集中し交通のよどみない流れを妨げる箇所（ボトルネック）が検出できるようになる。

位置情報は、単に大量に集積して表示するだけでは極端に数が多く、そのままでは情報を把握しきれないという問題がある。図 1 にその例を示す。(a) は関東近辺の行政区と鉄道路線網の地図 [4] 上に東京都市圏の AM7:00–10:00 の人の流れデータ [5] (PFLOW データ) を可視化したものである。この図からは交通網の形状に沿って人々が動くことが分かるが、どの地点からどの地点へ移動する人が多いか、人々が集中している区間がどこかなどの情報は把握しにくい。たとえば拡大表示やアニメーション表示などによって目視確認し交通需要や施策について検討することも考えられるが、日々全箇所のデータに関してこのような検討を行うには時間がかかりすぎる。そこで、(b) のように移動体の位置情報の集合に対して、いくつかの代表的な移動経路（以降、主動線と呼ぶ）を提示し、変化があった主動線を中心に詳細な検討を進める方法が考えられる。主動

線の重畳表示により、交通需要の大きい移動経路や、それらの間の衝突地点などのボトルネックが容易に把握できるようになり、交通最適化のための対策を講じられるようになる。このように、蓄積された位置情報を要約して主動線を抽出する技術は、日々の都市の交通状況を把握しやすくするために有用である。

本論文では、位置情報の時系列（以降、動線と呼ぶ）を要約して代表的な動きを抽出する主動線抽出処理について述べる。今回、特に位置情報を離散化させてから確率モデルにあてはめる方法に着目し、この確率モデルとして混合準マルコフモデルを用いる方法を提案する。さらに、準マルコフモデルを時空間的なネットワークと見なして経路探索を行うことにより主動線を抽出する方式を提案する。最後に、提案された主動線抽出方式について人の流れデータ [5] を用いて評価し、その特徴と有用性について示す。

2. 関連研究と本研究の位置づけ

2.1 位置情報履歴の分析技術

蓄積された位置情報の履歴を分析して意志決定に結びつける研究は数多くある。その 1 つが、位置情報の履歴からその場所に関する指標を算出する [6] 研究である。たとえば、観光客の位置情報を蓄積し観光動態を把握する [7], [8] 研究が知られている。この研究では、滞在時間などを用いて観光の魅力度を評価する評価指標が提案されている。また、屋内の位置情報履歴からそれがどのような行動であるかを判別する研究もある [9], [10]。これらは、多数の位置情報を、1 つの指標や判別結果という把握しやすい量に変換する方法に関する研究であるといえる。これらの方法による出力結果には、たとえばどの地点に問題が多いかなど空間的な特徴が失われているため、都市交通のインフラをどこに設置すべきかなどの意志決定には適さないことが多い。

そこで、多数の動線から主動線を抽出して可視化する方法が考えられる。これにより、多数の移動体と同じ区間を同じ方向に移動していることが把握できるようになる。ただし、図 1 の例に限らず、現実には 1 つの動きだけで状況が把握できる例は少ないと考えられるため、交通需要の把握という目的では複数の主動線抽出を想定する必要がある。なお、屋内作業など、同じ箇所を何度も往復するような動きを抽出する目的で、アソシエーションルール [11] などを用いて通行頻度の多い主動線を抽出する研究もある [12] が、これは多数の移動体の主要な移動経路を把握するという本研究の目的とは異なる。

2.2 主動線の種類

交通需要を把握する目的では、動線が集中していることが交通需要の高さを意味するので、類似する動線が多い経路が主動線と考えられる。ここで、主動線の種類について

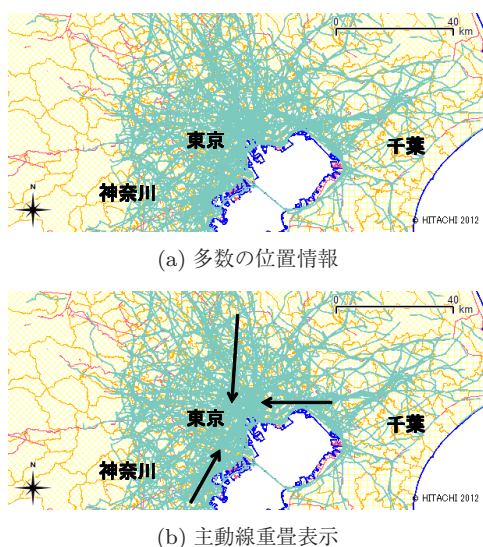


図 1 多数の位置情報

Fig. 1 Huge amount of positioning data.

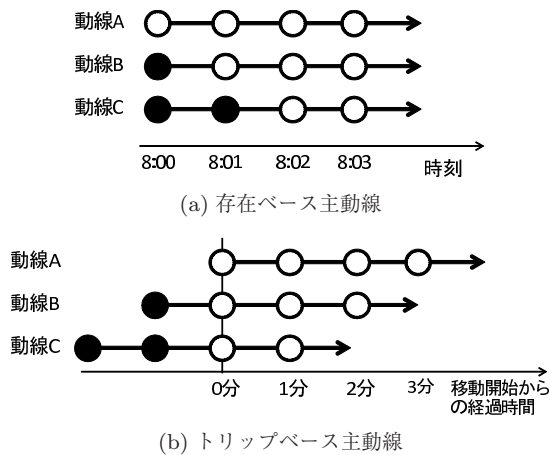


図 2 存在ベースとトリップベースの比較

Fig. 2 Comparison between existence-base path and trip-base path.

図 2 の模式図を用いて議論する．図中の動線 A, B, C は 3 つの移動体の動線であり，黒円は停留中，白円は移動中の位置情報をそれぞれ意味する．円の並びは時間経過を意味し，右方向に時間が進んでいる．主動線は A, B, C の座標値から算出されるが，この方法には以下の 2 種類が考えられる．

存在ベース主動線 図 2(a) に示すように移動体の動線を時刻を基準に並べ，同時刻の存在位置を重ねて得られる主動線である．

トリップベース主動線 図 2(b) のように移動体の移動開始からの経過時間を基準に並べ，経過時間が同じ存在位置を重ねることで得られる．なお，この移動開始から終了までの区間をトリップと呼ぶ．

スマートフォンなどを用いて得られる測位結果の履歴は，定期的に全移動体から得られるので存在ベース主動線抽出に適している．その反面，移動中か停留中かを問わず一律にデータが得られるので，移動停留の判定処理などによりトリップ区間を抽出しなければトリップベース主動線の主動線は抽出できない．また，アンケートや IC カードによる乗車履歴からはトリップの開始と終了，場合によってはその途中経路の情報のごく一部だけが得られるので，トリップベース主動線が抽出しやすい．しかし，補間処理などを経なければ，存在ベースの主動線は抽出できない．

交通需要の把握のためには，同時に多くの移動体が類似する移動をしていることを提示する必要がある．同じ経路を移動する移動体が多い場合でも，それが異なる時刻であれば小さい輸送力を持った交通機関のみで交通需要を満たせるためである．しかし，トリップベース主動線では同時に多数の移動体と同じ経路を同じ方向に移動しようとしているかは判断できない．そのため，交通需要の把握には存在ベースの主動線を複数抽出することが有効と考えられる．

複数の主動線を抽出する方法としては，多数の動線を類似するグループに分類する処理（クラスタリング）を実行

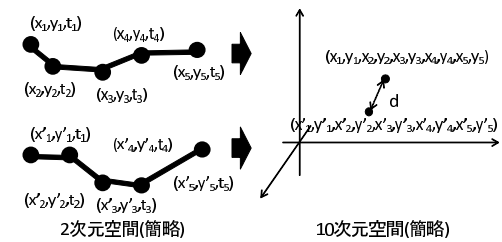


図 3 多変量クラスタリング

Fig. 3 Multivariate clustering.

し，各グループの代表を主動線とする方法が考えられる．その 1 つとして，位置の座標値の羅列を多変量と見なしてクラスタリングを実行する方法が提案されており，混合ガウス分布に基づくクラスタリング [13] や DBSCAN によるもの [14] が知られている．以降，本論文ではこれらを多変量クラスタリングと呼ぶ．一般に多変量クラスタリングを用いた存在ベースの主動線抽出は容易であるとされている．図 3 に存在ベースの主動線を抽出する場合の多変量クラスタリングの例を示す．この方法では，各軌跡を単純にならべたベクトルを多次元空間（図中では 10 次元）と見なし，一般的な多変量のクラスタリング手法（たとえば k-means）により構築されたクラスタの代表（セントロイド）を主動線としている．なお，多変量クラスタリングをトリップベース主動線抽出に適用するには，各軌跡が同一経路上で移動を開始した時刻も推定しなければならず，ベイズ推定を適用する方法が提案されている [13]．

2.3 本研究の位置づけ

本研究の目的は，交通需要を把握しやすく提示するために存在ベースの主動線を複数抽出することである．前節に述べたとおり，多変量クラスタリングによって存在ベースの主動線を複数抽出することができる．しかし，交通需要の把握という目的においては多変量クラスタリングが適さない場合も存在する．この点について図 4 を用いて示す．(a) は図 1 の背景地図の交通網である．図中破線で囲んだ都心域と，図中実線で囲んだ都心外の領域では交通網の細かさが大きく異なっており，人の移動はこのような詳細さの異なる交通網にまたがって起こる．したがって，詳細な交通網のある地域では詳細に，そうでない地域では大まかに特徴を抽出するのが望ましい．しかし，多変量クラスタリングの方法ではそのような結果が得られない場合が多い．(b) にその原因として交通網と多変量クラスタリングの関係を示す．鉄道路線 A と B に沿って動く人のうち B の方が多い場合，多変量クラスタリングによる主動線は図中点線で描かれている主動線のように，両路線の間で B 側に寄った位置となる．しかし，交通網に沿った主動線としては鉄道路線 B の線上にあることが望ましい．このように，多変量クラスタリングでは交通網の特徴を考慮できていない．

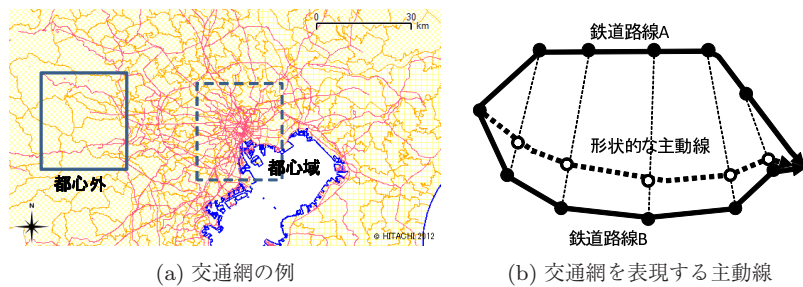


図 4 交通網に関する主動線

Fig. 4 Representative paths over the traffic network.

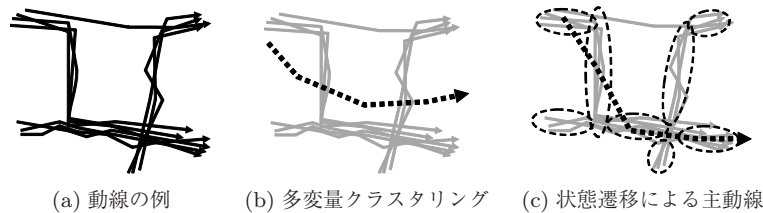


図 5 主動線の比較

Fig. 5 Contrast among representative paths.

そこで、動線を構成する各位置座標値を離散値に変換し、離散値の系列に変換した後、その集合に対して主動線抽出を行う方法が考えられる。通常、位置がわずかに異なっても人の意志決定には影響しない場合が多いが、たとえば駅の改札内外のように、意味的にまったく異なる状況を指すこともある。それに対応するため、位置を移動体の状況に対応付くように離散化する方法が用いられる [15]。以降、この位置情報を意志決定に意味のある離散的な値に変換したものを状態と呼び、状態の時系列変化を状態遷移と呼ぶ。動線を状態遷移に変換する方法としては、k-means や DBSCAN などの一般的なクラスタリング手法や、駅やインターチェンジなどの交通結節点付近のみを抽出し道路にマッチングさせて離散化するなどのヒューリスティックな手法が考えられる。

本論文では、状態遷移を表現する確率モデルを求め、そこから複数の主動線を抽出する方式を提案する。従来、状態遷移から求められた確率モデルは、たとえば尤度の低い動線は異常行動であると判定する [16] 目的などに用いられていたが、本研究は確率モデルからの主動線抽出が目標である。図 5 に従来の主動線と状態遷移から得られる主動線の比較を示す。(a) の動線の集合に対し多変量クラスタリングを適用すると、(b) の点線のように各動線との距離が最小になる主動線が求まる。本研究は、(c) の破線円で囲んだ交通の集中している箇所を通る主動線(点線)を抽出することが目的である。状態遷移の確率モデルとしては、人々の移動がマルコフ連鎖であるとしたもの [17], [18] や、人々の行動選択が潜在的なマルコフ連鎖に基づくとした HMM (隠れマルコフモデル) を用いた研究 [19] が知られているが、これらのモデルは遷移確率が時刻によらず一定であるので、同じ時間帯に交通需要が集中する経路を可

視化するという目的には適さない。そこで今回、時間帯によって遷移確率が変化する混合準マルコフモデルを用いて存在ベース主動線を抽出する方式について提案する。

3. 時空間ネットワークを用いた主動線抽出

3.1 空間的状态遷移モデル

位置情報とは移動体の各時刻における位置を示す情報であり、測位によって得られるのはそれに対する観測結果である座標値と時刻の組の集合である。本研究の対象とする状態遷移は位置情報を離散的な値に変換したものであるが、この変換の方法に応じて状態が定義される。最も単純な定義としては、座標値を離散化しそのまま状態とするものが考えられる。以降では、離散化された座標値を状態とする状態遷移モデルを空間的状态遷移モデルと呼ぶ。

空間的状态遷移モデルでは、人の移動を単純なマルコフ連鎖でモデル化することが多い。マルコフ連鎖では現在位置のみを状態とすると現在の滞留箇所のみに基づき次の場所へ移動する確率が定まる。ところが、実際には場所が定まっても次の行き先選択の確率は一定ではないと考えられる。そこで、MMM (混合マルコフモデル) を用いた先行研究 [17] がある。MMM では、人物の特徴を示す潜在変数を 1 つ導入し、この変数の値によって、行き先選択の確率が異なるとしている。この MMM の潜在変数は、観測不可能で時間変化しない人物行動の特徴を表している。潜在変数の値によって人物を分類することは、人物をそれまでの行動によりグループ化することに相当する。すなわち、当該人物がどのグループに属するかによって遷移確率を変化させることになる。

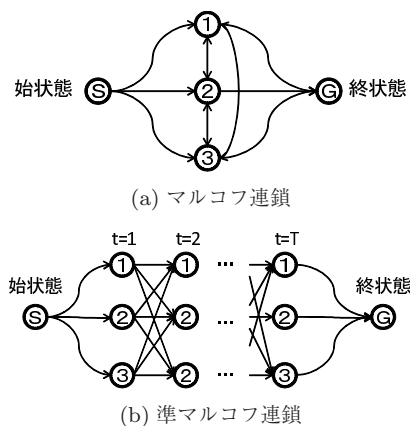


図 6 マルコフ連鎖と準マルコフ連鎖

Fig. 6 Markov chain and semi-Markov chain.

3.2 時空間的状态に基づく遷移モデル

空間的状态遷移モデルは空間的な特性しか加味していないため、状態遷移すなわち移動確率は時間が経過しても変化しない。そのため、適切に主動線が抽出できない場合がある。図 6(a) にマルコフモデルの状態遷移を図示する。この図は、状態遷移が状態遷移開始前を意味する始状態、状態遷移の終了後を意味する終状態からなると仮定し、状態 1, 2, 3 間の遷移を模式的に示したグラフである。グラフ中のノードが状態を意味し、ノード間のエッジが状態遷移を意味する。この表現では、始状態から終状態までグラフのエッジをたどっていき、通過したエッジの遷移確率の積が最も高くなるような経路を選ぶと最尤な状態遷移列が得られる。すなわち状態 i から j への遷移確率を $\omega_{i,j}$ とすると、経路上の遷移の対数尤度の和

$$L = \log \omega_{S,i} + \log \omega_{i,j} + \dots + \log \omega_{k,G} \quad (1)$$

が最大になるように経路探索すればよい。マルコフモデルでは状態間の遷移確率は時間によらず一定であるため、そのため S, 1, G などのように遷移の回数が少ない方が尤度が大きくなり、主動線は非常に短くなる。存在ベースの主動線を抽出する目的で一定回数の遷移をする経路のうち最尤なものを求める場合でも、 $\omega_{i,i}$ が大きい場合や $\omega_{i,j}$ と $\omega_{j,i}$ が大きい場合など、確率が最大の遷移のみをたどると同じ動きの繰返ししの尤度が高くなってしまう。

そこで、状態遷移を時空間条件に基づくものにするのが考えられる。その 1 つが、観測できる状態遷移とは別に観測できない潜在的な状態遷移を用意し、その状態にともない観測可能な状態の遷移確率も変化させていく方法である。たとえば、MARHMM (混合自己回帰隠れマルコフモデル) [19] は MMM の潜在変数を状態遷移するように修正したモデルであり、潜在変数の値の変遷によって、時間変化を表現する。しかし、潜在状態遷移を用いたとしても、潜在状態遷移自体はマルコフ連鎖であるので、上述のマルコフモデルと同じ問題は発生する。

そこで、状態遷移確率を明示的に時間依存するように

する方法が考えられる。時間依存するマルコフ連鎖は準マルコフ連鎖 (Semi Markov-chain) [20] と呼ばれている。図 6(b) に準マルコフモデルの状態遷移を示す。準マルコフモデルでは、毎ステップ遷移確率が変わるため、終状態への遷移するのはつねに元の動線集合を構成する動線の長さになる。そのため、存在ベースの主動線を抽出するのに適している。そこで本研究では、MMM を準マルコフ連鎖に拡張し、混合準マルコフモデル (MSMM; Mixed Semi Markov-chain Model) を用いたモデル化を提案する。

3.3 モデルパラメータの推定

以降では N 個の空間状態列を $\{d_n\}$ とし、各系列の t 番目の値を $d_{n,t}$ とする。この $d_{n,t}$ は空間状態 (離散化された位置) の種類数を M としたとき M 次元のベクトルとなる。 n 番目の空間状態遷移系列の t 番目の状態が μ のとき、 $d_{n,t,\mu} = 1$ それ以外は $d_{n,t,\mu} = 0$ となるものとする。

このモデルでは、状態間の遷移確率が時間に依存しているとしている。 k 番目の混合要素における状態 μ から状態 ν への遷移確率を $\omega_{k,\mu,\nu,t}$ とすると、MSMM の確率分布は

$$P_k(\{d_{nt}\}) = \prod_t \prod_{\mu} \prod_{\nu} \omega_{k,\mu,\nu,t}^{d_{n,t,\mu} d_{n,t-1,\nu}} \quad (2)$$

で定義される P_k を用いて

$$P(d) = \sum_k \pi_k P_k(\{d_{nt}\}) \quad (3)$$

と記述できる。ここで、 $\{\pi_k\}$ は各モデルの混合比率を表すパラメータである。なお、確率の規格化のため、 $\sum_k \pi_k = 1$, $\sum_{\mu} \omega_{k,\mu,\nu,t} = 1$ を満たす。

MSMM によるモデル化のためには、このモデルに登場するパラメータ π_k , $\omega_{k,\mu,\nu,t}$ と同時に、各状態遷移列の所属グループを意味する潜在変数 z_{nk} を定める必要がある。MMM では、EM アルゴリズムを用いて同様の計算を行っていたが、これを MSMM にも適用することを考える。ここで状態の定義を置き直し、時刻 t と位置 μ の組で 1 つの状態を示すとして $\xi = (\mu, t)$ および $\eta = (\nu, t-1)$ とすると、遷移確率は $\omega_{k,\xi,\eta}$ となり

$$P_k(\{d_{nt}\}) = \prod_{\xi} \prod_{\eta} \omega_{k,\xi,\eta}^{d_{n,\xi} d_{n,\eta}} \quad (4)$$

と MMM に帰着させることができる。したがって、MMM で用いられるのと同じ EM (Expectation-Maximization) アルゴリズムがそのまま適用できる。ただし、 $t \neq t' + 1$ のときの $\xi = (\mu, t)$ と $\eta = (\nu, t')$ については $\omega_{k,\xi,\eta} = 0$ に固定される。この制約は EM アルゴリズムに自然と含まれており、実際には d_n に $\xi \rightarrow \eta$ という遷移が含まれることはなく、必ず $\omega_{k,\xi,\eta} = 0$ となるため $\omega_{k,\xi,\eta}$ に対する M ステップは省略することができる。図 7 にこの MSMM に対する

初期化 $\{\pi_k\}, \{\omega_{k,\mu,\nu,t}\}$ を制約条件を満たすようにランダムに初期化する。

E ステップ 現在のパラメータを用い、以下の式で負担率 $\gamma_k(d_n)$ を求める。

$$\gamma_k(d_n) = \frac{\pi_k P_k(d_n | \{\omega_{k,\mu,\nu,t}\})}{\sum_{k'} \pi_{k'} P_{k'}(d_n | \{\omega_{k',\mu,\nu,t}\})}$$

M ステップ E ステップで求められた負担率によりパラメータを更新する。

$$\pi_k = \frac{1}{N} \sum_n \gamma_k(d_n),$$

$$\omega_{k,\mu,\nu,t} = \frac{\sum_n d_{n,\mu,t-1} d_{n,\nu,t} \gamma_k(d_n)}{\sum_{n,\nu'} d_{n,\mu,t-1} d_{n,\nu',t} \gamma_k(d_n)}$$

収束計算 上記 E ステップと M ステップを尤度が収束するまで繰り返す。

図 7 MSMM に対する EM アルゴリズム

Fig. 7 EM algorithm for MSMM.

EM アルゴリズムを示す。これにより、遷移確率の時間変化を加味した確率モデルを構築できる。

3.4 遷移モデルからの主動線抽出

マルコフモデルは遷移に関するモデルなので、遷移確率をもとに尤度計算し主動線を算定する方法が考えられる。その場合、混合している各マルコフモデルに対して、式 (1) で示したような尤度の高い動線を求めればよい。つまり、

$$\ln L_k(\mathbf{d}) = \sum_{t,\mu,\nu} d_{t,\mu} d_{t-1,\nu} \ln \omega_{k,\nu,\mu,t} \quad (5)$$

が最大になる遷移 $\mathbf{d}$ を求めることになる。前述のとおり、MSMM の状態遷移は各状態を時空間的なものと置き換えることによりマルコフ過程に帰着できる。すると、各状態をノード、状態間の遷移をエッジとした双方向ネットワークに帰着でき、各エッジのコストを $-\ln \omega_{k,\nu,\mu,t}$ とした最短経路探索問題と見なせる。これに類似する問題としては、HMM の最尤状態列を求める Viterbi アルゴリズム [21] が知られている。Viterbi アルゴリズムは観測されていないマルコフ連鎖の最尤系列を推定する方法であるので、MSMM の最尤状態遷移を求める目的でも類似の計算方法を用いることができる。図 8 にそのアルゴリズムを示す。

この遷移確率に関する尤度計算は、移動に関する確率を定義するため、トリップベースの主動線を抽出する目的に適すると考えられる。しかし、存在ベースの主動線としては、人の存在確率によって尤度を定めることも考えられる。 k 番目の混合要素に関し、時刻 t における各状態 μ の人の存在確率 $P_{k,\mu,t}$ は、

$$P_{k,\mu,t} = \sum_{\nu} \omega_{k,\mu,\nu,t-1} \quad (6)$$

と漸化式によって求まる。図 8 におけるエッジのコストを $-\ln P_{k,t,\mu}$ で置き換えることにより、人の存在確率に関す

初期化 状態 μ の最小コスト経路の配列 D_μ とコスト C_μ の組を状態数 M 個ずつ準備、 $D_\mu = (\text{始状態 } 0, \text{状態 } \mu)$, $C_\mu = -\ln \omega_{k,0,\mu,0}$ を代入。

検索 各状態 ν に対し、 $-\ln \omega_{k,\mu,\nu,t} + C_\mu$ が最小となる μ を求める。

更新 検索ステップを反映、 $D_\nu = (D_\mu, \nu)$, $C_\nu = -\ln \omega_{k,\mu,\nu,t} + C_\mu$ とする。

反復計算 検索と更新のステップを終状態に到達するまで繰り返す。

図 8 主動線算定アルゴリズム

Fig. 8 Representative paths extraction algorithm.

る主動線が得られる。ところが、人の存在確率のみによって主動線を抽出しようとする、単に人の多い場所をつなぐ形になってしまう。結果として、極端に遠方への移動など実際には存在しない移動を抽出してしまうことが考えられる。そこで、本研究では以下に示す μ から ν へ移動中の人の存在確率 $P_{k,\mu,\nu,t}^{\text{mov}}$ を用いて主動線を抽出することを提案する。

$$P_{k,\mu,\nu,t}^{\text{mov}} = P_{k,\mu,t} \omega_{k,\mu,\nu,t-1} \quad (7)$$

$\omega_{k,\mu,\nu,t-1}$ は実際には発生しない移動では 0 になるため、前述の問題は解決される。この $P_{k,\mu,\nu,t}^{\text{mov}}$ は、ランダムに人を選んだときにどの状態からどの状態へ遷移中であるかという確率である。つまり、この確率分布は移動中の人の存在確率分布に相当する。

4. 実験

4.1 実験条件

これまでに述べた主動線抽出方式の有効性を検証するため、PFLOW データ [5] に対して主動線抽出処理を行い、その結果を比較した。表 1 に今回用いたデータの概要を示す。今回、平成 20 年の東京都市圏のデータを用いた。本データには全体でおよそ 60 万人の 1 分おきの位置情報が 24 時間分含まれている。このデータをすべて用いると極端にデータ量が多くなるため、今回は対象人物の職種がサービス業従事者となっているデータのみを用い、時間帯も 7:00–10:00 に限定した。図 1 は、これを可視化したものである。

このデータに対し、MMM, MSMM による主動線抽出と多変量クラスタリングによってそれぞれ 10 本ずつ主動線を抽出した。MMM と MSMM の主動線抽出には、前処理として離散化が必要なので、最初にすべての点データに対して k-means および混合ガウス分布 (GMM; Gaussian Mixture Model) による離散化を行った。図 9 (a) に $k = 400$ とした k-means の結果を、図 9 (b) には混合数 $k = 200$ の GMM の結果をそれぞれ示す。図中の白い円が各クラスターの中心 (セントロイド) を示しており、円の間を結ぶ青い線は離散化された移動軌跡である。東京都心部は交通網が非常に混

み合っており、この付近に多数のクラスタが配置されていることが分かる。また、多変量クラスタリングに関しても同様に、k-means によるものと GMM によるものの 2 種類でクラスタリングを実行した。

4.2 結果と考察

図 10 に MSMM による処理結果を示す。MSMM の主動線は東京方面へ移動する動線が多数得られている。k-means を用いた MSMM では千葉、神奈川方面から東京方面へ交

表 1 本実験にて用いたデータ

Table 1 Summary of experimental dataset.

属性	内容
使用データ	人の流れプロジェクト「平成 20 年 東京都市圏」
時間帯	AM7:00–10:00 の間、5 分おき 37 点
対象人物	職種が「サービス従事者」で常時存在する 4686 人

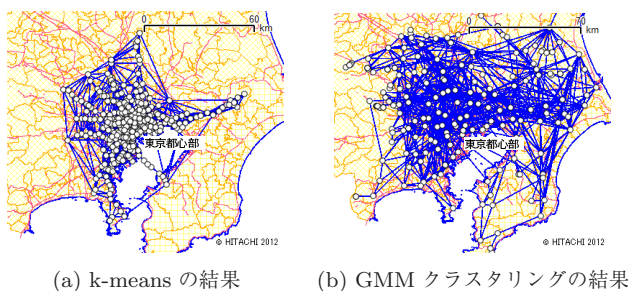


図 9 位置クラスタリング結果

Fig. 9 Clustering results.

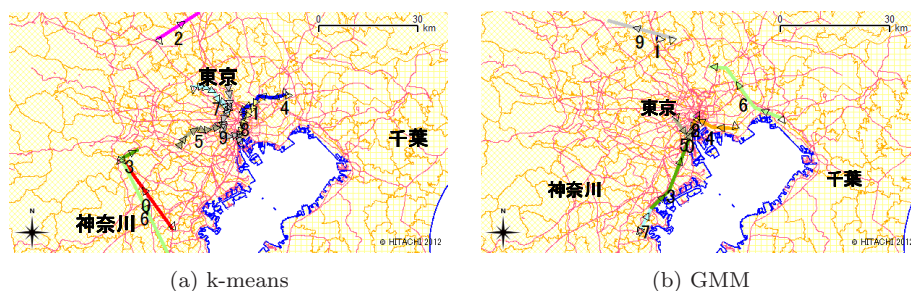


図 10 MSMM (提案方式) を用いた抽出結果

Fig. 10 Extraction results with MSMM.

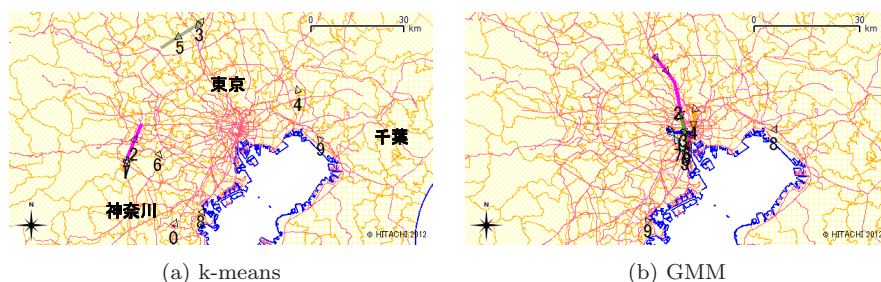


図 11 MMM を用いた抽出結果

Fig. 11 Extraction results with MMM.

通網に沿った動きが主動線として抽出されている。また、図中 2 や 6 と示される主動線など、郊外区域間での移動が抽出されている。k-means におけるクラスタ生成では、クラスタは距離と件数によって定まるが、交通量の多い鉄道路線沿線は件数が多く、長い距離を移動するために路線に沿って多数のクラスタが生成される。結果として、鉄道路線沿いの動きは細かく分類される。一方、郊外ではクラスタが広がるため、多数の動きが 1 つのクラスタに集約される。結果として、郊外付近の移動件数が多くなり、主動線も郊外付近が抽出されたということになる。郊外での振舞いは図 11 に示された MMM による主動線抽出結果でも同様である。異なるのは滞留の動きが多い点である。MMM では、どの時間帯であっても同一に扱うため、自宅から出発前の動きや目的地到着後の滞留の動きが大きく影響し、最尤な移動は常時滞留する動きになってしまうのである。結果として、MMM では郊外に常時滞留するか都心部に滞留するかという動きのない主動線が多くなってしまっている。次に、図 12 に多変量クラスタリングによる主動線抽出の結果を示す。多変量クラスタリングでは、鉄道路線や道路との対応がつかない線の表示になっているが、全体の流れは把握可能である。東京の中心域、路線の密集域での動きは平均化されており、全体としての流れが把握できる反面、交通網に沿った流れは把握しにくい。表 2 にこれらの比較についてまとめた表を示す。提案方式は都心域の交通網に沿った主動線が抽出できる点が長所であった。また MMM は郊外での滞留が多く抽出される点が、多変量クラスタリングでは交通網と独立した主動線が得られた点が短所であった。

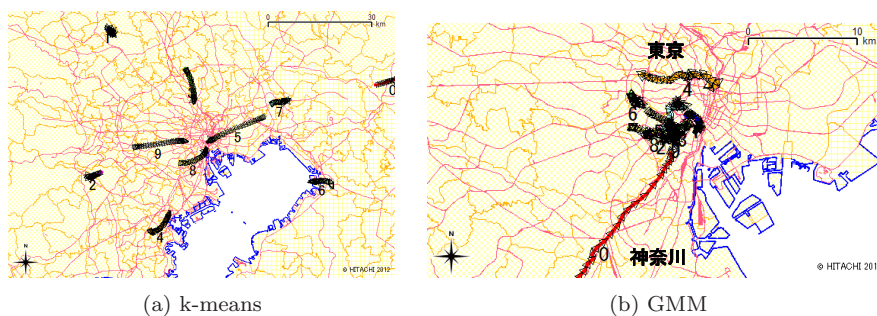


図 12 多変量クラスタリングによる抽出結果

Fig. 12 Extraction results with multidimensional clustering.

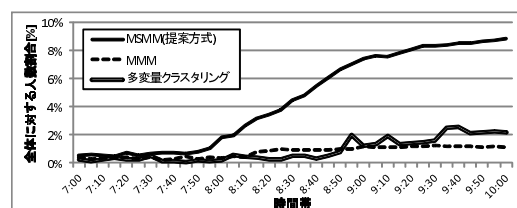
表 2 方式間の比較

Table 2 Contrast among methods.

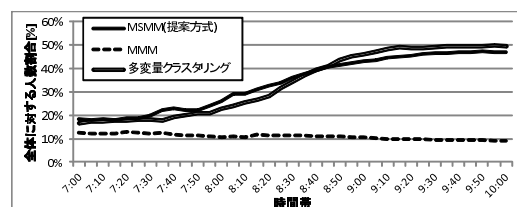
方式	MSMM (提案方式)	MMM	多変量クラスタリング
代表画像			
特徴	都心域で交通網に沿った主動線	郊外の滞留が多い	交通網との対応がとられていない

これらの結果を比較するため、図 13 に k-means を用いた各方式で抽出された 10 本の主動線付近の人数比較を示す。(a) は全体の人数に対する主動線位置を中心とする半径 500 m 内にいる人数の割合を示す。なお、複数の主動線の近傍に同一人物がいる場合は重複しないように算出している。どの方式であっても、7:00 から 8:00 くらいの時間帯ではあまり主動線付近には人がいない。しかし、それ以降、出勤のため都心域へ向かう人数が多くなると、MMM や多変量クラスタリングに比べて提案方式による主動線付近の人数は特に増大している。MMM は時間帯によらず状態遷移確率が一定であるので、都心域へ向かう人数が多い時間帯でも都心域への移動確率が多くはならない。実際には、人々は移動よりも滞留している時間の方が長いので、滞留の確率が高くなる。結果として郊外に滞留しつづける主動線が得られたと考えられる。また、(b) には半径 5,000 m 内の人数の割合を示す。(a) の場合と異なり、多変量クラスタリングの結果も提案方式と同様に、後半に人数割合が増大している。確認のため、10:00 時点での主動線周囲の人数割合を横軸に半径をとって比較したグラフを図 14 に示す。このグラフから、半径を大きくするに従って、提案方式と多変量クラスタリングの差が小さくなっていることが確認できる。この結果は、多変量クラスタリングは図 15 (a) のように近傍に複数ある交通網の間に主動線を抽出してしまい、提案方式は図 15 (b) に人の多い領域上に主動線を抽出していることを示している。

図 16 に GMM による多変量クラスタリングと MSMM の GMM に関する主動線の拡大図を示す。(a) の多変量ク



(a) 500 m 内の人数割合



(b) 5,000 m 内の人数割合

図 13 主動線付近の人数割合の時系列変化

Fig. 13 Time series of people number ratio around representative paths.

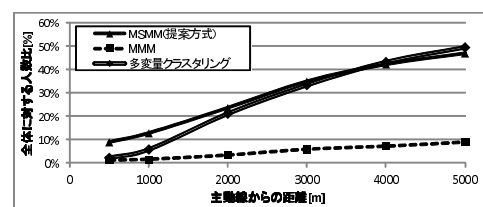


図 14 10:00 の主動線付近の人数割合

Fig. 14 People number ratio around representative paths at 10:00.

ラストリングの結果からは、都心部域の流れの傾向が把握できるが、(b)のMSMMの結果からは、鉄道路線に沿った動きが抽出されている。図中、点線にて表示した線はこの地域の主な鉄道路線である山手線と京浜東北線の軌道であり、抽出されたMSMMの主動線がこの路線に沿っていることが分かる。このように、多変量クラスタリングでは人の多い交通網から少し離れた場所に主動線を引いてしまっているが、提案方式は人の多い交通網のすぐ近傍に主動線を引いているものと考えられる。これは、提案方式が都心域のように複数の交通網が近傍に存在する場合に交通網に沿った主動線を抽出するのに適することを示唆している。

図17に図10(b)で示したMSMMと同じモデルを用いて、主動線抽出の方法のみを変えた結果を示す。図17(a)は遷移確率のみに関する尤度、つまり $-\ln \omega_{k,\nu,\mu,t}$ をコストとして主動線を算出した結果である。この結果では滞留している動きや、同じ動きが多くなりがちな都心外地域での短い動きばかりが抽出されている。図17(b)は滞留人数、すなわち $P_{k,\mu,t}$ によって主動線を抽出した結果である。この結果は、単純に人数の多いクラスタ間をつないだ結果に

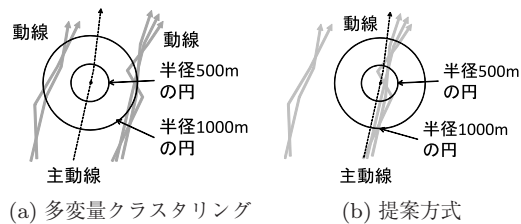


図15 主動線付近の人数の概念図

Fig. 15 People number around representative-paths.

相当しており、住宅地区の多い都心外地域のうち人口の多い地域から開始され、出勤などで人口が減っていき人数が逆転したら次の地域へ移動するという動きが表示されているが、実際にはここに図示されているような都心外地域間の移動は発生していないと考えられる。このように、主動線抽出では移動中の人の存在確率という観点から経路探索することが効果的である。

上記の結果から、提案手法である状態遷移確率モデルに基づく主動線抽出手法が人の行動単位での特徴把握に有効であると考えられるが、ここで重要なのは状態の抽出方法である。今回の実験では、k-meansとGMMによる抽出を実行したが、k-meansではクラスタが円状であることを仮定しているため、1つの路線上で移動が分割されてしまう問題があった。他方、GMMでは細長いクラスタなどが抽出されるため、この問題は起きにくくなっている。しかしながら、GMMによる状態抽出でも主動線抽出の目的によっては必ずしも適切ではない。たとえば、鉄道の乗り換えに関する傾向把握のためには、別途準備した鉄道路線図との対応付けなどによって、把握したい特徴が明確化される状態抽出が必要である。目的に応じた状態抽出の方法については、今後の課題として研究が必要と考えられる。

5. おわりに

本研究では、多くの位置情報履歴から移動体の移動傾向を把握するために有用な主動線抽出処理について検討し、混合準マルコフモデル (MSMM) を用いた主動線抽出の方法を提案した。次にPFLOWデータを用いて主動線抽出

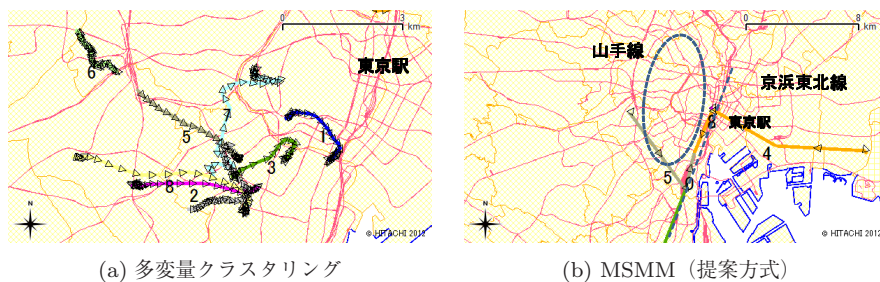


図16 主動線の拡大図

Fig. 16 Closeups to representative paths.

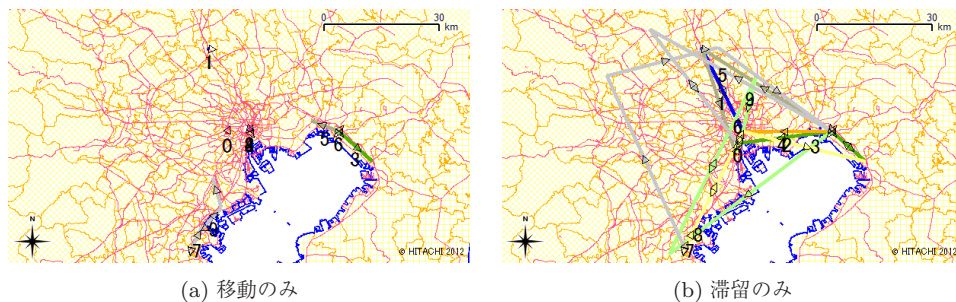


図17 主動線抽出方法の比較

Fig. 17 Comparison among representative paths extraction methods.

方式を評価し、提案手法に関して有用なケースがあることを確認した。

今後、位置情報履歴が蓄積できる状況はより整備されていき、現在以上に多くのデータが蓄積されていくと考えられる。そのため、巨大なデータに対して概要を把握できるような情報抽出を行う技術は、必要性が増大していく。さらに情報が蓄積されてきた場合、本研究で述べた方法のみでは処理時間がかかりすぎることが考えられる。このような、いわゆるビッグデータに対する主動線抽出処理の高速化方式については、今後の課題である。

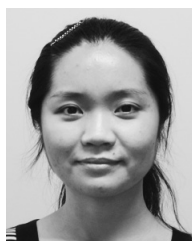
参考文献

- [1] 三輪富生, 森川高行: プローブカーデータを利用した経路選択行動に関するモデル分析, 土木計画学研究・論文集, Vol.21, No.2, pp.553-560 (2004).
- [2] 伊藤誠悟, 吉田廣志, 河口信夫: locky.jp: 無線 LAN を用いた位置情報・測位ポータル, 情報処理学会研究報告 2005-MBL-34 (4), pp.25-31 (2005).
- [3] 辻川雅夫, 雅楽隆基, 小檜山智久: ハンズフリー型入退室管理システム—日立 AirLocation-MJ の技術概要と応用分野 (特集 人をまもる電波センシング技術), 計測技術, Vol.38, No.12, pp.26-30 (2010).
- [4] 国土地理院: 国土基盤地図情報 1/25000, 入手先 (<http://www.gsi.go.jp/kiban/index.html>).
- [5] 東京大学空間情報科学センター (共同研究「ダイナミックデータ統合可能な都市空間情報基盤の研究」): 人の流れプロジェクト, 入手先 (<http://pflow.csis.u-tokyo.ac.jp/index-j.html>).
- [6] 吉田琢史, 小磯貴史, 服部可奈子, 豊島伊知郎, 今崎直樹: 顧客動線に基づく店舗内環境評価手法の提案, 情報処理学会研究報告 2005-ICS-139 (21) (2005).
- [7] 長尾光悦, 川村秀憲, 山本雅人, 大内 東: 観光動態情報の獲得を意図した GPS ログデータマイニング (社会システムにおける知能: 社会システムのモデル化), 情報処理学会研究報告 2004-ICS-135 (2), pp.7-12 (2004).
- [8] 鈴木恵二, 松原 仁, 鈴木克也: はこだてまちナビにおける観光客動態の推定, 情報処理学会研究報告, UBI, Vol.2008, No.40, pp.41-46 (2008).
- [9] 青木茂樹, 大西正輝, 小島篤博, 福永邦夫: HMM による行動パターンの認識, 電子情報通信学会論文誌 D-II, Vol.J85-D-II, No.7, pp.1265-1270 (2002).
- [10] 豊島伊知郎, 服部可奈子, 吉田琢史, 板倉豊和, 小磯貴史: 多段階のパターン認識を用いた歩行軌跡データからの顧客行動判別, 情報処理学会研究報告 2006-CVIM-153 (25) (2006).
- [11] Agrawal, R. and Srikant, R.: Fast algorithms for mining association rules, *Proc. 20th Int. Conf. Very Large Data Bases*, Vol.1215, pp.487-499 (1994).
- [12] 小磯貴史, 服部可奈子, 吉田琢史, 今崎直樹: 歩行者動線分析システムを用いた大型家電量販店での行動分析, 情報処理学会研究報告 2003-UBI-2 (11) (2003).
- [13] Chudova, D., Gaffney, S., Mjolsness, E. and Smyth, P.: Translation-invariant mixture models for curve clustering, *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, ACM, pp.79-88 (2003).
- [14] Lee, J., Han, J. and Whang, K.: Trajectory clustering: a partition-and-group framework, *Proc. 2007 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, ACM, pp.593-604 (2007).
- [15] 神田崇行, 塩見昌裕, 野村竜也, 石黒 浩, 萩田紀博: RFID タグを用いた科学館来館者の移動軌跡の分析, 情報処理学会論文誌, Vol.49, No.5, pp.1727-1742 (2008).
- [16] 鈴木直彦, 平澤宏祐, 田中健一, 小林貴訓, 佐藤洋一, 藤野陽三: 人物動線データ分析による逸脱行動人物の検出, 情報処理学会研究報告 2007-CVIM-158 (15) (2007).
- [17] 浅原彰規, 丸山貴志子, 佐藤暁子: 混合マルコフモデルに基づく歩行者動線解析方式, 情報処理学会論文誌, Vol.52, No.1, pp.187-196 (2011).
- [18] Asahara, A., Maruyama, K., Sato, A. and Seto, K.: Pedestrian-movement prediction based on mixed Markov-chain model, *Proc. 19th ACM SIGSPATIAL GIS 2011*, New York, NY, USA, ACM, pp.25-33 (2011).
- [19] 浅原彰規, 佐藤暁子, 丸山貴志子: 混合自己回帰隠れマルコフモデルによる歩行者行き先予測, 情報処理学会論文誌, Vol.53, No.1, pp.342-351 (2012).
- [20] Smith, W.L.: Regenerative stochastic processes, *Proc. Royal Society of London. Series A. Mathematical and Physical Sciences*, Vol.232, No.1188, pp.6-31 (1955).
- [21] Viterbi, A.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Information Theory*, Vol.13, No.2, pp.260-269 (1967).



浅原 彰規

2002 年北海道大学理学部物理学学科卒業。2004 年北海道大学大学院理学研究科物理学専攻修士課程修了。同年 (株) 日立製作所入社, 以来, 中央研究所にて空間情報システムの研究に従事。電子情報通信学会員。



寺本 やえみ (正会員)

2003 年早稲田大学理工学部電子情報通信学科卒業。2005 年早稲田大学大学院理工学研究科情報・ネットワーク専攻修士課程修了。同年 (株) 日立製作所入社, 以来, 中央研究所にてテキスト解析システムおよび空間情報シ

ステムの研究に従事。



丸山 貴志子 (正会員)

1988 年お茶の水女子大学大学院理学部物理学科修士課程修了。1992 年総合研究大学院大学数物科学研究科統計科学博士課程修了。博士（学術）。同年（株）日立製作所入所。以来、中央研究所にて、空間情報システムの研究に従事。現在、同研究所主任研究員。電子情報通信学会会員。



柴崎 亮介 (正会員)

1980 年東京大学工学部土木工学科卒業，1982 年東京大学大学院修了。建設省土木研究所（1982～1988 年），東京大学工学部助教授（1988～1991 年），同大学生産技術研究所助教授（1991～1998 年）を経て，1998 年より空間情報科学研究センター教授。2005 年から 2010 年にはセンター長を務めた。空間情報，位置情報の生成，利用等に関する研究に従事。