

博士論文

人の移動履歴のデータマイニングのための
確率モデルの提案とその応用

浅 原 彰 規

要旨

題 目： 人の移動履歴のデータマイニングのための確率モデルの提案とその応用

氏 名： 浅 原 彰 規

近年のモバイルコンピュータ（容易に持ち運び可能な小型コンピュータ）分野の発展は著しいが、このモバイルコンピュータの多くは、測位の機能を有している。さらにこれらに類似するものとして、建物や道路などに設置された機器を用いてそこにいる人や車等の存在を検知する技術がある。これらの測位を利用したサービスは、近年になってさまざまな環境において用いられるようになってきており、人の位置に関する情報が社会に溢れている。

これらの人の位置に関する情報を活用すると、人の移動行動の調査の補完に用いることができると想定される。しかし、現時点で実用になっている事例は多くはない。その原因の一つが、データの入手の困難さの問題である。そこで、調査目的で集められたデータではなく、人が普段の行動をする範囲で収集された位置のデータを副次的に活用することが考えられる。本論文では、このような他の目的で収集された位置の履歴のデータ群から有益な情報を抽出することを移動履歴のデータマイニングと呼ぶ。

移動履歴のデータマイニングは、その特性の差異から、従来の調査とは異なった用途が考えられる。人の移動行動の調査が求められる用途の例として、交通インフラ整備のための交通需要の地点からある地点へ移動したいという需要）の調査を例とする。交通需要調査では、まず四段階推定法などの人の交通行動のモデルを設定し、そのモデルのパラメータを決めるために、パーソントリップ調査のアンケートなどの入手による調査が行われる。これによって人の交通行動のモデルが定まり、新設の道路の需要を求めることができる。このような調査で得られるデータは網羅的、精緻であるが、大きなコストや調査時間がかかる。よって、従来の調査は、道路建設の決定などの十年、二十年という長期的な意思決定には適するが、毎月、あるいは日々の変動に対応せねばならないケース、たとえばバスの臨時便追加などに関するケースには適さない。他方、移動履歴のデータマイニングでは、日々副産物として集まる情報を用いるので、コストはかなり抑えられ、データの鮮度も高い。その反面、データの量そのものは多くとも、網羅性については従来の調査よりも低いと考えられる。

移動履歴のデータマイニングは、従来の調査とは異なった性質をもつので、それに適した分析手法が新たに必要となる。網羅性が低いデータから、短期間で、できるだけ有益な情報を多数抽出しなければならないのである。交通需要調査では各個人に対して交通行動の出発地、目的地、交通手段などが得られることを前提にしているが、移動履歴データではその前提は満たされず、位置情報のみしか得られないことを想定せねばならない。また、目視確認により情報を補完していくようなことも、日々の運用に供することを想定すれば、現実的ではない。つまり、得られた網羅性の低い位置情報だけから、人の位置や移動の全体的な傾向を知るための様々な情報を短時間で自動的に抽出することが必要になるのである。

移動履歴のデータマイニングは、様々な用途が想定されるため、その処理技術も複数想定される。本論文では、この処理技術を要約、推定・予測の2つに分類する。要約とは、人が概況を把握するた

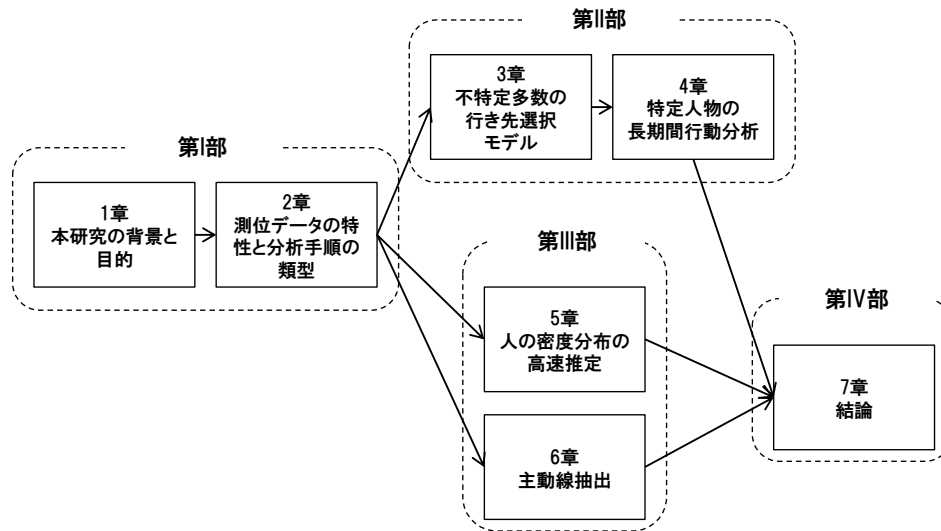


図 1 本論文の構成

めの技術であり、特徴の抽出や異常の検知、可視化などが含まれる。推定・予測とは現状知られていない事実を導き出す技術であり、人の行き先の予測や、移動のパターンにもとづく類似の行動をするグループの検出などが含まれる。意思決定には様々な事情を勘案しなければならないため、抽出される情報は多様なほうがよいが、多すぎると逆に把握が難しくなる。よって抽出情報の観点を絞り、必要な情報を取捨選択できるよう、分析のメニューを揃えていくことが求められる。

ただし、人の行動のように複雑で非決定論的な事象について、それを完全に予測できるような数理モデルをたてることは困難である。そのような場合、経験的な確からしさを定量的に評価する確率モデルが用いられる。本研究でも、確率モデルを用いて人の移動行動をより精緻にモデル化することを検討する。確率モデルは予測に用いることができるが、それだけでなく、確率モデルの中に内部的な状態を含むことにより、推定や要約にも適用できるという利点がある。この観点で先行研究で不足するものの一つが、不特定多数への対応性である。従来研究のうち、指標算出以外の多くの手法は、ある特定個人が十分長期間動線データを計測しつづけることを前提にしており、事前準備されていない不特定多数のデータに対する処理に適した手法はまだ不十分である。

そこで本研究は、蓄積された位置情報だけから情報を抽出するために、確率モデルにもとづく種々の推定・予測と要約の手法を提案して拡充することを目指した。人の移動を扱う確率モデルとして最も基本的なものとしては、マルコフモデルが想定される。これは、人の行動は非決定論的である状態でどう行動するかは一意には定まらないものの、その選択確率だけは定められると仮定したことに相当する。実際には不特定多数の人の行動を一つのマルコフモデルに当てはめられないと想定されるがマルコフモデルは単純ながら拡張性が高いモデルであり、人の行動選択の条件は拡張できる。

以上より、本論文では特にマルコフモデル的な確率モデルに着目し、その拡張によって種々の推定・予測、要約に適用する方法を提案することで、抽出される情報の拡充を目指す。図 1 に本論文の構成を示す。

第 I 部では、この研究の前提条件や目的、定義、問題設定、本研究のアプローチについて述べる。

まず、「第 1 章 本研究の背景と目的」では本研究の背景と位置づけについて概略を示す。「第 2 章 測位データの特性と分析手順の類型」では、測位に関して整理するとともに、分析方式の概略と要素について整頓する。第 I 部では、人の移動にまつわる意思決定に必要な基本的な情報として、対象地域内の移動する人たちに関し、どんな人たちが、いつ、どこにいて、どこへ、どう移動しようとしているかが重要であることを述べている。

次に、第 II 部では、人の行き先を他の人の移動履歴を用いて予測する確率モデルについて第 2 章で規定した手順にもとづき、事例を交えながら議論した。「第 3 章 不特定多数の人の行き先選択モデル」と「第 4 章 特定人物の長期間行動分析」は、予測推定の手法について、主にマルコフ連鎖の確率モデルを用いた方法について述べる。第 II 部では、人の個性をうまく確率モデルとして扱うことが精度向上の鍵になっていたと思われ、時間が経つにつれて個性が顕在化するため、時間変化を扱うモデルが高い性能を発揮していた。つまり、どんな人たちにがいるかに着目した分析であり、移動のパターンや個人ごとの行き先予測という形の情報が得られた。

そこで第 III 部の「第 5 章 人の密度分布の高速推定」と「第 6 章 主動線抽出」では、きの全体像を可視化するため、人の密度分布計算について直前の分布を用い高速化する方法と、マルコフ連鎖的なモデルを適用し主動線を抽出する方法を提案する。ここでは人の個性の情報は集約し、他の情報に焦点を当て可視化している。どこに人がいるかを時間ごとに可視化したのがヒートマップであり、どこへどうやって移動するかを可視化したのが主動線であった。

最後に、「第 7 章 結論」では、全体について俯瞰し、今後の展望について示す。第 II 部および第 III 部の結果を合わせると、マルコフ連鎖的なモデルによって第 I 部で示された人の移動に関する基本的な情報の抽出技法がおおむね網羅されたことになる。したがって、人の移動の平均にマルコフ連鎖的なモデルを適用することによる情報抽出は妥当であり、意思決定のための状況把握がしやすくなったと考えられる。

目次

要旨	i
第Ⅰ部 序論	4
第1章 本研究の背景と目的	5
1.1 背景	5
1.2 研究の動機	6
1.3 先行研究と課題	12
1.4 本研究の目的	14
第2章 測位データの特性と分析手順の類型	18
2.1 動線のデータと処理	18
2.2 動線解析とその構成	21
2.3 まとめ	22
第Ⅱ部 人の位置の遷移モデル	23
第3章 不特定多数の人の行き先選択モデル	24
3.1 はじめに	24
3.2 マルコフ連鎖モデル	25
3.3 混合マルコフ連鎖モデル	28
3.4 混合自己回帰隠れマルコフ連鎖モデル	30
3.5 状態の定義	35
3.6 実験による評価	35
3.7 関連研究	45
3.8 おわりに	46
第4章 特定人物の長期間行動分析	47
4.1 はじめに	47
4.2 人の行動モデル	48

4.3	状態遷移生成方式	50
4.4	実験	51
4.5	関連研究	53
4.6	おわりに	54
第 III 部 集団統計量の遷移モデル		58
第 5 章	人の密度分布の高速推定	59
5.1	本研究の位置づけと目的	59
5.2	従来方式の問題と提案方式のアプローチ	60
5.3	部分データによる分布生成方式	63
5.4	実験	71
5.5	おわりに	75
第 6 章	主動線抽出	77
6.1	はじめに	77
6.2	関連研究	78
6.3	時空間ネットワークを用いた主動線抽出	81
6.4	実験による評価	85
6.5	おわりに	90
第 IV 部 結言		93
第 7 章	結論	94
7.1	本論文のまとめ	94
7.2	結論と今後の課題	96
7.3	本研究成果の応用と今後の展開	97
付録 A	MMM に対する EM アルゴリズム	99
A.1	EM アルゴリズムの導出	99
A.2	局所解への対策	100
付録 B	MAR-HMM に対する BW アルゴリズム	103
B.1	EM アルゴリズムの導出	103
B.2	局所解への対策	105
付録 C	MSMM に対する EM アルゴリズム	106
C.1	局所解への対策	106

第 I 部

序論

第 1 章

本研究の背景と目的

1.1 背景

1.1.1 モバイルコンピューティングと位置情報

近年のモバイルコンピュータ（容易に持ち運び可能な小型コンピュータ）分野の発展は著しく、日本国内でも、スマートフォンやタブレット PC と呼ばれる高機能なモバイルコンピュータが急速に普及している。さらにモバイルコンピュータのための高速な通信網が拡充されてきており、外出先や移動中などにコンピュータを用いたサービスを楽しむこと（モバイルコンピューティングとよばれる）が容易になってきた。さらに、人が装着可能なコンピュータであるウェアラブルコンピュータにも普及の兆しがある。^{*1}このように、ICT は人々の生活により溶け込み、常時その機能を活用することが現実のものとなった。

モバイルコンピュータの多くは、ナビゲーションや情報配信などモバイルコンピュータを使用している人の現在位置にあわせた機能を提供するために、測位の機能を有している。測位とは機器の現在位置を計測し、緯度経度などの実際の位置を特定可能な記述形式にて得る機能であり、GPS(Global Positioning System) や準天頂衛星を用いた衛星による測位が最も広く用いられている。衛星を用いた方式は、衛星から発せられる電波が機器に到達するまでの時間を用いて位置が推定される方式であり、衛星からの電波を受信できる機器以外に新たな設備なしに測位できるという利点があるため、カーナビゲーションシステムをはじめとする様々な用途で用いられている。ただし、衛星からの電波が受信しにくい状況、例えば樹木や山岳、建築物など電波を反射、遮蔽する物があると十分な精度で測位ができないという欠点がある。特に需要の大きい都心部や屋内のための測位方式が多数開発されている。例えば、屋内に設置された無線 LAN からの電波を用いた方式 [1][2][3][4][5] や、屋内設置型の GPS 発信器 [6][7] などがあり、いずれも建物等に設置された機器を用いて位置が推定される。また、基地局に残っている携帯電話の通信記録 (Call Detail Records) をもとに位置情報を推定する方法も研究されている [8]。

これらに類似するものとして、建物や道路などに設置された機器を用いてそこにいる人や車の存在

^{*1} 特に Google Glass などのヘッドマウントディスプレイが注目を集めており、今後さらに類似のウェアラブルコンピュータが輩出してくることが予測される。

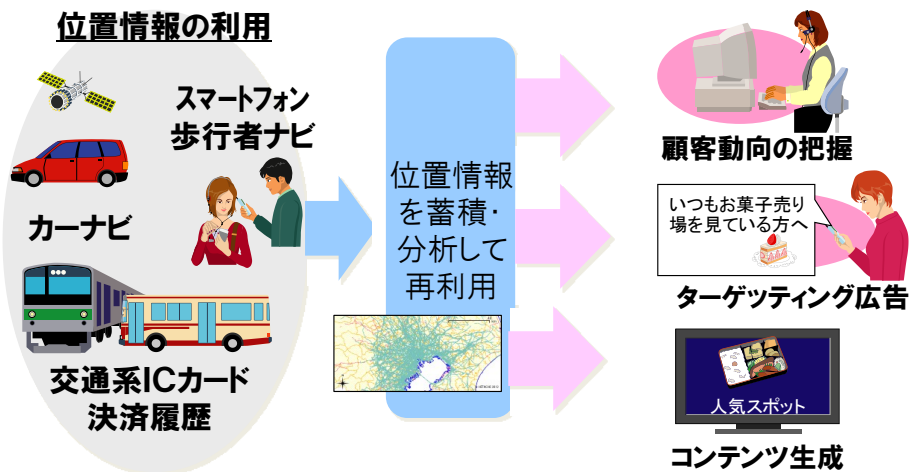


図 1.1 位置情報の活用

を検知する技術がある。上記の測位との大きな違いは計測対象となる人がモバイルコンピュータ等を持たなくてもよい、つまり、不特定多数を計測できるひいては網羅性が非常に高いという点にある。日本国内での普及しているものとしては、道路に設置された車両感知器を用いた渋滞情報の配信サービスが存在する [9]。^{*2}カメラやレーザなどを用いて多数の人や物の位置を計測・追跡する技術など屋内でも利用できる方式の研究開発が盛んに進められている [10][11]。

また、公共交通機関の運賃支払い用の IC カード（交通系 IC カード）を用いた決済履歴のデータのように、人が外出先で ICT を活用したときの記録も、広い意味では測位に類するデータといえる。それらはそれ自身には位置の情報そのものは含まれていないが、例えば降車駅と乗車駅の位置を用いるなどすれば、履歴が残ったときの位置を推定することが可能である。

これらの測位を利用したサービスは、近年になってさまざまな環境において用いられるようになってきた。たとえば、LBS(Location Based Services) と呼ばれる位置に基づき情報を配信するサービス、SNS やゲームなどのエンタテインメント、経路案内を行うなどのサービスが知られている [12]。このような測位を用いたサービスの普及にともない、より多くの情報が得られることになる。

1.2 研究の動機

1.2.1 位置情報の副次的利用による現状把握の可能性

前節にて述べたように、人の位置に関する情報は様々あり、社会に溢れている。これらの人の位置に関する情報を活用し、図 1.1 に示すように、人の移動に関する現状の把握とに用いることができると想定される。このような、蓄積されたデータ群から新たな事実や特徴などの知見を取り出すことを一般的にデータマイニングと呼ぶ。本論文では、特に人の位置情報の履歴を対象にしたデータマイニングによっていかに有益な情報を抽出するかについて議論する。

^{*2} 渋滞情報の配信サービスは渋滞を加味した経路選択の支援が目的であるため、交通量ではなく旅行時間（道路に入ってから出るまでにかかる時間）が計測される。

人の位置情報のデータマイニングの用途は多岐に渡ると考えられる。そのうち代表的なものの一つとして、交通行動を対象にした調査が考えられる。特に普及しているのが、渋滞情報の抽出である。前節にて述べたとおり、車両感知器を用いた車速の計測は広く活用されているが、それ以外にも、タクシーなどの特殊車両やカーナビゲーションシステムの利用者、スマートフォンの所有者など、実際に道路を走行している車両の位置情報を収集することにより渋滞を検出するプローブ交通情報 [13][14] とよばれる技術がある。プローブ交通情報により、詳細な道路の旅行時間推定による交通渋滞のボトルネック検出や、新規道路の建設による混雑の緩和効果定量化などの都市計画策定支援などに役立つ情報が得られる [15][16]。プローブ交通情報では個々の自動車を用いた移動の情報しか取得できないという問題があるが、前述の交通系 IC カードの情報と複合すれば、例えば新設された公共交通路線による渋滞緩和効果の検証 [17][18] など、公共交通機関の改善策の立案に有用な情報が得られると期待される。

近年では、携帯電話に定期的に位置情報を送付するソフトウェアを導入し、一般の人の交通行動を把握するプローブパーソンの試みが知られている [19][20]。従来、携帯電話に新たにソフトウェアを導入する点に困難があったが、新規ソフトウェアを導入しやすいスマートフォンが普及したことで、現実味を増してきている。プローブパーソンでは、アンケートであった記載漏れや時刻等の偏りが無い。また、情報を取得できる頻度が高く、うまくすればアンケート等を用いた調査よりも詳細な情報を得ることができると考えられる [21]。ただし、プローブパーソンではアンケート等では取得可能な情報が得られないという点に留意する必要がある。例えば、移動に用いた交通手段に冠する情報や移動の目的地などは何らかの手段で補って明確化する必要がある。これらは位置情報に前処理を加えることによって得られることも多い（例えば長く一箇所に滞留している場合、そこを目的地と推定するなど）が、GPS を用いた場合の計測誤差の影響を加味しつつ適切に推定処理を行うことが求められる [22]。また、アンケート等にくらべると調査対象となる人の傾向が偏りがちで、網羅性も低いため、アンケートの代替にはなっていないのが現状である。

また、位置情報はマーケティングなどの企画の用途での活用も検討されている。測位データには、例えば作業員の測位データにはその作業の特性が、店舗などの顧客の測位データではその店舗における顧客の嗜好が、一人の人物の測位データを蓄積し続ければ当該人物の特徴が、それぞれ含まれる。例えばマーケティング分野では、顧客に対する詳細な調査を行う手法としてエスノグラフィーと呼ばれる方法がよく知られているが、その中で、店舗等の来客がどのような移動をしたかまで仔細に調べられており、位置情報が活用できる可能性が高い。あるいは、駅などのレイアウト設計のために人の流れを計測することがある。それには調査員を配置したり、録画映像を目視で確認することが主流である。防災の観点からは、人の移動に関する概況把握が有用である。災害発生時に人がどこにどのくらいいるかの情報が得られると、自治体等が例えば物資の配分を決定する場合など、災害発生時の意思決定を行う上で有用である。あるいは、観光客の位置情報を収集できれば観光行動が把握できるため、プロモーション企画等の計画策定において有用な情報源となる [23][24][25][26]。これらは、一般の人の位置情報を収集、分析することによって価値を得るものであるが、不特定多数の位置情報を活用することに対する社会的なコンセンサス（特にプライバシーに関する合意）がまだないために十分普及したといえる状況にはない。また、例えば携帯電話からのデータ収集では通信や測位による消費電力が大きいなど、いまだ技術的な問題点もある。しかし、実際にポイントカードを用いて購買履歴

が収集されているなど、今後、データ収集の枠組み作りは進むと考えられる。

他にはプライバシーの問題のない用途として、作業員等の動きのモニタリングでも徐々に実用化が進められている。たとえば工場等の作業員や営業員の位置を計測して、作業中の事故等の早期発見につなげるとともに、その記録から作業のどこに時間を要しているかをみつけて効率化を図ることができる。この場合は、作業員の位置は業務上の情報であり、比較的、プライバシーの問題とは切り離される。また、個人の日常生活における位置を記録しつつ、その個人に対して日常の行動パターンに沿った情報配信を行うなどの用途が知られている。この用途では、データの用途を当該個人に限定することにより、プライバシーの問題を回避している。

1.2.2 現在実施されている調査の例： 交通調査

前節の通り、位置情報のデータマイニングの用途には様々な可能性がある。これに類するものとして、アンケート等を用いて行う調査があるが、アンケート等の調査のデータを分析するための方法が位置情報にそのまま適用できるとは限らない。その差異を明確化するため、本節では端的な例として交通調査との比較を示す。

現在、日本で交通需要の把握のために行われている調査としては、各地域の交通協議会などが10年に一度程度の頻度で行うパーソントリップ調査（以降、PT 調査）^[27] が最もよく知られている。PT 調査では、調査対象の地域の居住者 5～10 %程度を対象にアンケートを行い、ある地域から他の地域への移動（以降、トリップとよぶ）について、一日分の履歴が調査される。アンケートの調査項目は、回答者の属性（年齢、性別、職業など）とその日の回答者の全てのトリップの情報（出発地の住所、出発時刻、目的地、到着時刻、交通手段、移動の目的など）である。このアンケート結果は、国勢調査などで調査された居住地や年代ごとの人口全数と比較され、各回答者が何人分の集団を代表しているかを示す拡大係数を求めるなどして、補正された上で用いられるのが一般的である。PT 調査で得られるトリップの情報には、回答者の誤記や記入漏れといった回答不備、出発到着の時刻がきりのよい数字（13 時 00 分や 15 時 30 分など）への偏り、出発地や目的地の記載が粗い（丁目まで書かれていないなど）などの誤差もある^{*3}。PT 調査で得られたトリップの統計情報は、交通受容のモデルに当てはめて活用されており^[28]、交通流のシミュレーションシステム^[29]等のパラメータ推定にも活用されている^[30]。例えば、一般的に人の交通需要の予測に用いられる四段階推定法^[31]では、人の交通行動は表 1.1 に示す発生・集中モデル、分布モデル、分担モデル、配分モデルの4つの段階に分けてモデル化されている。それぞれのモデルには、関連する交通量のパラメータがあり、PT 調査の結果はこれらの交通量を定めるために活用されることが多い。

他に類似した大規模な調査として、主要道路の断面交通量を調べる道路交通センサス調査が知られている。断面交通量とは、ある道路を単位時間あたりに通行する車両や人の数を意味しており、各道路の利用率の評価指標としてよく用いられている。断面交通量の調査は、道路わきに調査員を配置して通行する車両や人の数を数えるという方法が主流であるが、近年では道路に設置された画像トラフィックカウンタを用いて計測された断面交通量も用いられはじめた^[32]。画像トラフィックカウン

^{*3} 回答数が十分多いので、統計的には十分な精度が得られる。

表 1.1 四段階推定法のモデルと関連する交通量

モデル	説明	関連交通量	意味
発生集中モデル	各地域から出発あるいは各地域に到着する人の数を決定するモデル	発生交通量	各地域を出発する人の数
		集中交通量	各地域に到着する人の数
		生成交通量	ある地域にいる人の全数(国勢調査などによって得られる)
分布モデル	交通行動をとる人それぞれについて出発地域(Origination)と到着地域(Destination)の組を決定するモデル	OD 交通量	出発地域と到着地域の間を移動する人数
分担モデル	移動時にどのような交通手段を用いるかを決めるモデル	分担交通量	各交通手段毎の移動人数
配分モデル	同一の交通手段の中で存在する複数の経路に対し何人がどの経路を選ぶかを配分するモデル	交通量配分	経路毎の交通量

タは、カメラ等を用いて道路を通過した車両を撮影して数を計るセンサであり、道路を走行する車両のみではあるが、高精度に低コストで常時計測ができるという利点がある。

このように、人の交通行動の調査はアンケート等の人手による調査が主流であった。このような調査で得られるデータは網羅的であり、非常に有用である。しかし人手による調査では大きなコストがかかるという問題がある。また、収集頻度が少ないという問題がある。特に急速に状況が変化するような場合は、アンケートに要する時間も加味すると、それだけで全てをカバーするのは難しい。よって、従来の調査は、十年、二十年という長期的な意思決定には適するが、毎月、あるいは日々の変動に対応せねばならないケース、たとえば交通機関の運用に関するケースには適さない。

そこで、ICT(情報通信技術)を用いた効率化が求められる。つまり、携帯電話等の機器を活用することで、上記の調査を補完するデータを収集することができれば、非常に有用であるといえ、交通調査の高度化や効率化が期待されている [33][34]。

実際、人の位置の情報をを用いて種々の調査を補完する試みは多数存在するが、継続的に使用される実用的な事例は必ずしも多くない。その原因の一つにデータの入手の困難さの問題がある。例えば、測定用の携帯電話用ソフトウェアを配布して交通行動の履歴を収集しようとしても、位置情報だけでは交通手段等を全て把握することはできない。PT 調査の結果を分析する際に用いられる四段階推定法は、各個人に対して十分な情報が得られることを前提にしているが、それを適用するには PT 調査などと同じ項目を持つアンケートが必要となってしまう。これはアンケートなどにかかる負担を低減したいという目的からみれば本末転倒である。また、各個人に日常の行動にない行動を長期にわたって要求するのは現実的ではなく、日々アンケートをとるようなことはできない。結果的に、たまたま PT 調査と同様のアンケートをオンラインで行うという、PT 調査を紙媒体からオンラインに変えただけの形となってしまう。

そこで人が普段の行動をする範囲で集められるデータ、言い換えれば、調査目的ではなく収集され

た位置のデータを P T 調査を補う形で活用することができれば、現実的に有益な情報を得られると期待できる。他の用途で集められた位置情報を副次的に利用することは、被調査対象にも、調査者にも負荷や費用をかけずに情報を得られる点で有用である。ただし、上述の通り、被調査対象にとってのプライバシーの問題があるため、他の個人に関する情報と位置情報を関連付けるのは実用上は難しい場合が多い。結果的に、個人を特定することなく数多くのデータを収集し、統計値としてのみ利用するのが想定可能な利用形態である。

1.2.3 移動履歴のデータマイニングの用途

位置情報の副次的利用により、日々変化する人の移動に関する情報が得られ、施設管理や交通機関についてのより運用に近い分野の意思決定に役立つと想定される。ここでいう、より運用に近い分野の意思決定としては、例えば、バスの臨時増便、タクシーの配車、来客対応スタッフの増員、警備や交通誘導の人員配置の改善等が挙げられる。調査目的と異なる理由で収集された位置情報を二次的に利用する場合、網羅性の低さなどの問題点はある。しかし、従来の調査のもとでは困難な短期的な状況把握には短時間、低コストで情報収集可能な位置情報の利用が効果的である。また、少ないコストで現状把握ができるため、まだコストをかけて調査を行っていない暗中模索の状態においては、現状を把握するための貴重な手がかりとなる。

ただし、上述のとおり、位置情報には従来の調査と同じ手法での分析は適さない。図 1.2 にその例として、展示会場における人の流れの例を示す。(a) は来訪者の動線をそのまま重畳表示したもので、雑然とした表示になっている。これを目視確認によって一つ一つ精査し情報を補完していくようなことも不可能ではないが、日々の運用に供することを想定すれば、時間的にも労力的にも現実的ではない。そこで (b) のような代表的な動線（主動線）の表示方式が考えられる。この表示では図中下側を巡回するパターンと、図中上側を巡回するパターンに分離できており、すべてを満遍なく回る人が少ないことが見て取れる。これらの複数のパターンに人の移動を分類すれば、どのような行動がどのくらいの比率であるのかが把握でき、たとえば図中下側の展示に図中上側の展示への案内表示を増やすなど、運用時の施策に関する意思決定につながる。このように事前情報なしに日々の副産物としての位置情報の履歴を利用するには、自動処理を加えて意思決定に役立つ概略的な情報を抽出すること、つまり移動履歴のデータマイニングが必要である。

このような用途の移動履歴のデータマイニングにおいては、対象地域における人々の移動にまつわる状況把握が求められる。図 1.3 に例を示す。スマートフォンのソフトウェアによって観光案内等を行い、それに付随して位置情報を収集する場合、どのくらいの観光客がどの順番で観光地を訪れ、どのくらい時間を費やしたかなどの情報を得ることができる。もしバス路線の整備されていないある観光地への人気は急上昇したような場合、タクシーでの移動が増えて台数不足が起こると予期される。そのような場合に、臨時的にシャトルバス等を運行して渋滞等を低減し、観光客の満足度を高める工夫ができる。展示会場の例では、入り口付近から奥側へ人があまり移動していないのであれば、奥側へ人が移動するように案内表示を追加したり、現場で誘導を行ったりするなどの用途が考えられる。

上述のように、本来は移動に関する問題の特定とその解決策を得ることが根本的な目的であるが、現実的に得られる情報は限られており、なんら前提をおかずに自動で解決策まで導き出すことは困難

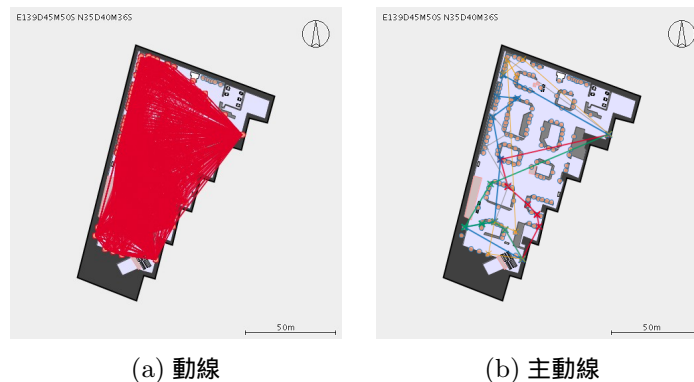


図 1.2 移動履歴をもちいた全体傾向把握のための情報抽出

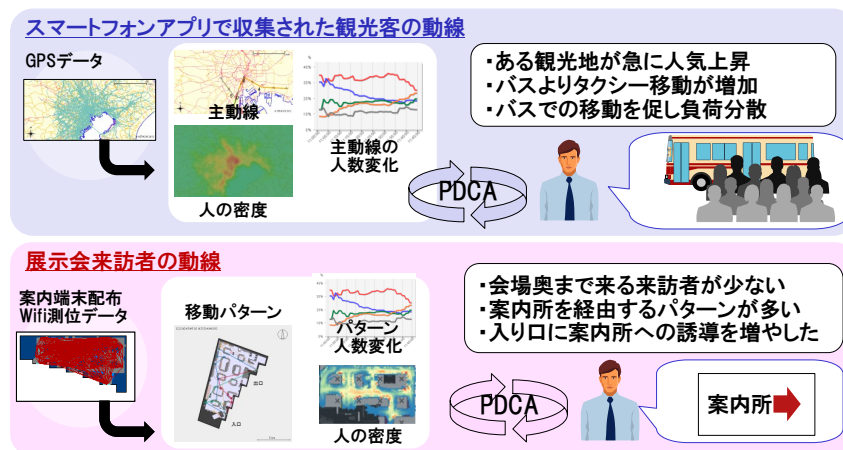


図 1.3 移動履歴のデータマイニングの応用

である。例えば、商業施設内の人の移動をスムーズにするために仕切り等を設置した結果、ある店舗の前を通る人が激減して問題になることがある。あるいは、通路の構造が変わり、車椅子の通行が困難になってしまうなどもありえる。このように、現場ごとの事情で考慮しなければならないことは無数にある。現実には、管理者によって現場ごとの様々な事情を勘案した意思決定がなされているが、それを完全に自動化するにはそれらの事情すべてを入力として与えればならず、技術的に非常に難しい問題になる(この種の問題は人工知能分野ではフレーム問題ともよばれる)。よって、自動で問題の検出や解決策の導出ができる範囲は限定的であり、人の判断を介することは避けられない。つまり、対象の地域や施設における人の移動の情報を多面的に管理者に提示することで問題解決に寄与することになる。例えば、どのくらいの人があるところにいて、どこへ行こうとするのかなど、位置情報から抽出できる種々の情報を時々刻々と表示することで、現状を一望できるようにする、などが考えられる。

なお、移動履歴のデータマイニングによって、追加のコストをほとんど発生させずに統計情報の更新頻度を高めていくことができる。これは従来の調査と相補的であるが、また相互に有益な情報が含まれるという利点も特筆に価する。例えば、新興国などの公共インフラ整備が十分進んでおらず、仔細な調査が行われていない地域では、位置情報だけから得られる知見であっても有用であり、たとえ調査そのものは行わざるを得ないとしても、その前段階の情報収集には役立つと考えられる。逆に、

移動履歴のデータマイニングにおいても、従来の調査結果との差異をみることで有益な情報が得られるのである。

上述の議論のとおり、社会活動において収集される位置情報を副次的に利用し、日々の概要を把握できるように表示する移動履歴のデータマイニングは意思決定に有益である。ただし、従来の調査とは異なり、事前情報なしに位置情報のデータのみから、移動に関する概略を自動抽出する必要がある。意思決定には様々な事情を勘案しなければならないため、抽出される情報は多様なほうがよいが、多すぎると逆に把握が難しくなる。よって抽出情報の観点を絞り、必要な情報を取捨選択できるよう、分析のメニューを揃えていくことが求められる。

そこで本研究では、人の位置データを大量に収集した場合に、事前情報なしにそこから知識を抽出するデータマイニングの種々の方法について議論し、分析の拡充をめざす。

1.3 先行研究と課題

1.3.1 現状把握の類型

位置情報に対する処理方式は、用途に応じて様々なものが知られているが、人の移動にまつわる意思決定に必要な情報を抽出することが必要である。

もっとも基本的な情報としては、対象地域内の移動する人たちに関する現状が必要である。人の移動に関する状況を 5W1H で整理すると、以下ようになる。すなわち、どんな人たちが (Who)、いつ (When)、どこに (Where) いて、どこへ (What)、何の目的で (Why)、どう移動しようとしているか (How) である。

将来的には、特定の範囲や条件下で、対策が必要な場所や、どんな対策が効果的かという施策まで提示することが望ましい。例えば、人の一箇所への集中過密が起きそうかを調べたり、人の行きそうな場所に関する情報を配信するなどして人を誘導したりすることがありえる。しかし、情報抽出の信頼性が十分でないと、人の誘導等を直接行うには事故を誘発しかねないなどの多大なリスクが伴う、よって、まずは基本的な情報を抽出して可視化し、意思決定に人を介在することで、リスクを低減することが考えられる。そこで、本研究では意思決定に役立つ情報抽出に焦点を当て、その方法について議論する。

本研究では、情報抽出の方法を 2 つに分類する。表 1.4 にその分類を示す。一つは、移動する人に関して現状知られていない事実 (移動そのもの以外の付随する情報) を導き出すことであり、人の行き先の予測や、移動のパターンにもとづく類似の行動をするグループの検出などが含まれる。すなわち、どんな人たちが (Who)、何の目的で (Why)、どう移動しようとしているか (How) に焦点を当てている。もう一つは、人が概況を把握するための技術であり、人の移動そのものに関する情報を抽出する技術である。つまり、どこに人がいて (Where)、いつ (When)、どういう交通手段で (How)、どこへ (What) 行こうとしているか、全体の平均的な傾向を可視化したり通知したりするものである。以降では、それぞれについて詳しく述べ、本研究の目的をより明確化する。

表 1.2 現状把握の要素

分類	抽出対象	把握したい情報
人の移動に 付随する情報 (Who, How, Why)	移動のパターン	人の移動に影響する環境要因
		人の属性と移動の関連
	人の移動の確率	人の行き先の選好
人の数や移動そのもの の傾向 (Where, When, What, How)	人の密度分布	混雑の度合い
	主動線	多くの人が進む動線

1.3.2 人の移動に付随する情報の推定

人の移動に付随する情報は個人のデータを用いる方法と不特定多数のデータを用いる方法がある。個人のデータを用いる方法は、ある個人の過去の動線データを用いて、その個人の次の行き先を予測したり、個人の状態を推定したりする方法である。この方法では、ある個人を特定して長期間データを取得しなければならない。不特定多数のデータを用いる場合は、移動のパターンが類似する他の人の情報を用いて推定や予測を行う。この方法では、同様の状況にいる人を多数集める（例えば、同じ地域の居住者のデータなど）必要がある。代わりに対象となる個人に対しては長期間のデータが必要ないという利点がある。

行き先の予測は、機器と連動することにより、例えば歩行者が近づく直前まで機器を待機状態にしてエネルギー消費を抑えるといった機器制御 [35] にも利用されている。個人を特定して長期間データを取得する事例としては、カーナビゲーションシステムなどを対象にした運転の目的地を予測する方式 [36][37] がある。

動線データを用いた推定の用途としては、事故等の通報すべき事象が起きた場合に管理者等に通報するという使い方がある。このような使い方をするためには、事前に通常の動線データを蓄え、統計処理等を行い数理モデルとして扱えるようにしておく。これにより、動線データをその数理モデルと比較してずれが大きい場合に、異常と判定できる。青木ら [38] は単一の人物の移動履歴を HMM(Hidden Markov Model; 隠れマルコフモデル) にて表現し、生成された HMM が当該人物のこれまでの行動と異なるかを調べ例外行動を検出している。また、移動のパターンから人の属性（特徴）を推定する技術もある。松尾ら [39] は赤外線センサにて多数の人物の地点訪問履歴を取得し、多変量解析により歩行者属性を推定する方式を提案した。この方式により、例えば喫煙所に何度も行った人が喫煙者であるなどの判定が可能である。同様に、動線データを分析することで、それがどのような行動であるかを判別する研究もある [40]。

1.3.3 人の数や移動そのものの傾向の要約

人の数や移動そのものの傾向として、最も基本的なものは人の密度である [41]。密度の可視化手法としてはヒートマップがよく用いられる。ヒートマップは、人の密度に応じて色分けした画像を生成

する方法である。人の密度は現在どのあたりに人がどのくらいいるかを把握するために有効であるが、交通分野などでは現在の人の数よりも人の移動を把握するほうが求められる場合がある。その場合、多数の動線から類型化された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、現実には一つの動きだけで状況が類型化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。

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）を実行し、各グループの代表を主動線とする方法が考えられる [42][43][44]。その一つとして、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いる。一般的な多変量のクラスタリング手法（例えば 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。Jae-Gil ら [45] は座標値列の構成する線分間に距離を定義し、当該距離にもとづく動線を DBSCAN によってクラスタリングするとともに、クラスタ内動線の平均速度方向の主動線を抽出する方式を提案した。また、同様に混合ガウス分布にもとづきベイズ推定を適用したクラスタリング [46] が知られている。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 [47] などを用いて通行頻度の多い動線を抽出する研究もある [48]。

他に類するものとしては、位置情報の履歴からその場所に関する指標を算出する研究がある。例えば、観光客の位置情報を蓄積し観光動態を把握する [25][26] 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。店舗等でも同様の研究 [49][50][51] がなされるなど、動線データを指標化する方法については、各応用分野にて検討がなされている。また、神田ら [35] は RFID タグを用いて博物館来館者の位置を計測し、各地点での多数の歩行者の滞留時間などから典型的な歩行者行動を抽出することにより、その地点が通路であるか立止まる場所であるかなど、地点の特徴を得る技術を報告した。これらは、多数の位置情報を、一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。

1.4 本研究の目的

1.4.1 先行研究の課題

これまでに議論した通り、移動履歴のデータマイニングでは、データ収集の前の綿密な準備等ができることを前提にすることはできず、詳細な状況が不明な状態で、位置情報だけから分析が可能でなければならない。したがって、本研究では原則として動線データのみからあらゆる分析を行うことを前提とする。この観点から先行研究を俯瞰し、前述の人の移動に関する 5W1H の観点から整理したのが表 1.3 である。

人の移動に付随する情報の推定にあたり重要なのが、不特定多数の人の移動からパターンを見出すことが必要であった。しかし、先行研究では移動のパターン抽出はある特定個人の移動を表現するためのものである。これはある特定個人が十分長期間動線データを計測しつづけることを前提にしており、単一のモデルで人の行動を表現している。しかし、不特定多数の人を対象とすると、その行動に

表 1.3 把握したい情報と先行研究の課題

把握したい情報	主抽出すべき情報	先行研究	先行研究の課題
移動に影響する要素や移動する人の特徴	人の移動のパターン (Who, How, Why)	個人ごとにマルコフ連鎖などでモデル化	不特定多数の人の動きを画一的なモデルで扱っており、人の個性を加味していない
人の数(混雑)や移動そのものの傾向	時間帯ごとの人の多いもしくは少ない場所 (Where, When)	カーネル密度推定法(スムージング)によるヒートマップ	全件密度計算はデータ件数に依存して処理時間が増大する
	人の代表的な移動経路・手段 (What, How)	軌跡の座標の発生分布のモデルによるクラスタリング	軌跡形状によるクラスタリングでは移動の意味的特徴が加味されない

は個性があり一様ではないことは自明である。よって、不特定多数の人の移動履歴のみから、人の個性を加味した複数のモデルを抽出することが必要となる。

人の数や移動そのものの傾向を要約するにあたり必要なのは、場所ごとの人の数と、地点間の移動の発生頻度の情報抽出である。場所ごとの人の数を示す方法としてはすでにカーネル密度推定法によるヒートマップがよく知られているが、人の数が極端に多くなったときにそれに比例して処理時間がかかるという課題がある。一方、地点間の移動に関しては、軌跡形状に着目したクラスタリングによる主動線の抽出法が知られている。しかし、人の移動においては地点の情報が重要なことも多く、地点間移動に着目した主動線の抽出が必要となる。

1.4.2 本研究の目的とアプローチ

これまでの議論では、蓄積された不特定多数の人の位置情報から意思決定に有益な概略的な情報を抽出するために、移動に影響する要素の推定と移動そのものの要約について、拡充が必要であることを示した。先行研究の多くは統計処理の立脚点として人の移動行動を数理モデル化しているが、上述の通り、不特定多数の人の移動についての概略を把握しやすく表現するには不足する部分があった。そこで本研究では、これらの先行研究における課題を解決できるような人の移動行動を数理モデルをたてることを目的とする。

人の行動のように複雑で非決定論的な事象について、それを完全に予測できるような数理モデルをたてることは困難である。そのような場合、経験的に確からしさを定量的に評価する確率モデルが用いられる。本研究でも、確率モデルを用いて人の移動行動をより精緻にモデル化する。確率モデルを予測に用いることができるのは明白であるが、それだけでなく、確率モデルの中に内部的な状態を含むことにより、推定や要約にも適用できるという利点がある。

人の移動を扱う確率モデルとして最も基本的なものとしては、マルコフモデルが想定される。これは、人の行動は非決定論的である状態でどう行動するかは一意には定まらないものの、その選択確率だけは定められると仮定したことに相当する。実際には不特定多数の人の行動を一つのマルコフモデ

表 1.4 課題へのアプローチ

抽出すべき情報	課題	アプローチ
移動に影響する要素や移動する人の特徴	不特定多数の人の動きを画一的なモデルで扱っており、人の個性を加味していない	複数のマルコフ連鎖モデル（複数の母集団）を組み合わせたモデルの提案
人の数（混雑）や移動そのものの傾向	全件密度計算はデータ件数に依存して処理時間が増大する	直前時刻の分布からの変化を扱う確率モデルを用いて高速化
	軌跡形状によるクラスタリングでは移動の意味的特徴が加味されない	地点間移動の発生確率モデルを複数組み合わせたモデルの提案

ルに当てはめられないと想定されるが（人によって選択確率が異なることは容易に想像できる）、マルコフモデルは単純ながら拡張性が高いモデルであり、人の行動選択の条件は拡張できる。

そこで本論文では特にマルコフモデル的な確率モデルに着目し、その拡張によって種々の推定・予測、要約に適用する方法を提案し、実験評価した結果について述べる。表 1.4 に本研究の課題に対するマルコフモデルを用いた解決方式について示す。

移動に影響する要素や移動する人の特徴を抽出するにあたっては、先行研究では不特定多数の人の動きを画一的なモデルで扱っていた点が課題であった。これは、多種多様な人を単一の母集団からのサンプルとみなして分析したことにあたる。実際には個人は各々固有の性質を持っているのだが、不特定多数の位置情報のみを扱っているため、個人をそれぞれモデル化することは出来ないためである。そこで本研究では、人の移動について複数の母集団のグループから得られたサンプルとみなし、移動のパターンからグループを推定しモデル化する方式として、複数のマルコフ連鎖の混合形を提案する。

人の数の傾向を表現するヒートマップ生成にあたっては、その処理時間に課題があった。本研究では、直前時刻の分布との差分が小さいことを利用し、密度の分布の変化をマルコフモデルとみなし、直前の分布からの変化のみを扱うようにすることで高速化を図ることとした。

人の移動そのものの傾向を示す主動線の抽出にあたっては、移動の意味的特徴を加味した主動線が抽出できないという課題があった。そこで本研究では、時間帯ごとの地点間移動の発生頻度を扱うモデルとして、ステップごとに地点間の移動確率の変わる準マルコフモデルを複数組み合わせたモデルを提案する。

以上のように、本論文では不特定多数の人の移動履歴から移動に関する概略的な情報を抽出する手段として、マルコフモデル的な確率モデルの拡張を提案し、それぞれを実験的に検証する。これにより、不特定多数の人の位置情報から意思決定に有益な概略的な情報を抽出可能であることを示す。なお、上述のとおり移動履歴の活用方法としては、移動に関する現状把握以外にも、異常検知等様々な用途があるが、本論文では今後の課題とした。

表 1.5 本論文の構成

部	章	題名	内容
第 I 部 序論	第 1 章	本研究の背景と目的	位置情報のデータマイニングの必要性
	第 2 章	測位データの特性と分析手順の類型	本論文で扱う位置情報の定義および分析の手順
第 II 部 人の位置の遷移モデル	第 3 章	不特定多数の人の行き先選択モデル	個人ごとの移動パターン (Who, How, Why) を扱う確率モデル
	第 4 章	特定人物の長期間行動分析	
第 III 部 集団統計量の遷移モデル	第 5 章	人の密度分布の高速推定	人がいつどこにどのくらいいるか (When, Where) を扱う確率モデル
	第 6 章	主動線抽出	人がどこへどう移動するか (What, How) を扱う確率モデル
第 IV 部 結言	第 7 章	結論	本論文のまとめと今後の課題

1.4.3 本論文の構成

本論文では、測位を用いた人の流動分析を対象とした技術について述べる。第 2 章では、測位に関して整理するとともに、分析方式の概略と要素について整頓する。第 3 章と第 4 章は、人の位置の遷移を確率モデルであらわす手法について、主にマルコフ連鎖の確率モデルを用いた方法について述べる。第 5 章と第 6 章では、集団統計量の変化を確率モデルであらわす手法について、人の数の分布のモデルを変化させて時間変化を表現する方法と、人の動きのうち主要な動線を得るのに適したモデルとおよび主動線の算定方法の 2 つについて述べる。最後に、第 7 章では、全体について俯瞰し、今後の展望について示す。

第 2 章

測位データの特性と分析手順の類型

2.1 動線のデータと処理

2.1.1 測位データと動線データ

測位データとは移動体の動線をサンプリングして得られる誤差を含んだデータであり、測位対象を一意に示す ID、測定を実行した時刻、および測定結果である座標値の組で表現される。測位データのデータ形式を表 2.1 に示す。また、以降では、動線データとは、単一の移動体に対する測位データを時間順に並べて得られるものを指す。本研究では、動線データに対して位置の情報以外の意味づけを行う処理を動線解析とよぶ。

収集したデータの流通促進のために動線データのデータ保存に用いられる形式は、できるかぎりにおいて標準的なものとしておくことが望ましい。ISO では、ISO19141:2008 として、移動する物体の軌跡の表現である Moving Features のデータモデルについての標準を規定している。図 2.1 にこの ISO19141:2008 で規定されている Moving Features のデータモデルの模式図を示す。ISO19141:2008 では、移動する物体の位置は特定時刻の位置を示す Leaf(葉) と、その時刻間の移動を表現し任意時刻の位置を再現できるデータを持つ Trajectory(軌跡) を組み合わせた、Prism(角柱) と呼ばれるモデルで表現される。この系全体を Foliation(葉層) とよぶ。ISO19141:2008 の Moving Features ではこの Foliation のモデルで移動する物体の集団全体を表現するように規定されており、任意の時刻の各物体の位置を再現できるようになっている。OGC(Open Geospatial Consortium) ではこの ISO19141:2008 Moving Features のデータモデルに従ったデータの記述形式の仕様 [52] が策定、標準化されており、データへのアクセス機能についても標準化されている。今後、このような

表 2.1 測位データの構成

項番	名称	意味
1	移動体 ID	測位対象を一意に示す ID。具体的には、ユーザの ID となる。
2	時刻	動線上の 1 点を表す時刻。
3	座標値	空間上の 1 点を表す値。2 次元の場合は 2 つの値の組、3 次元の場合は 3 つの値の組となる。

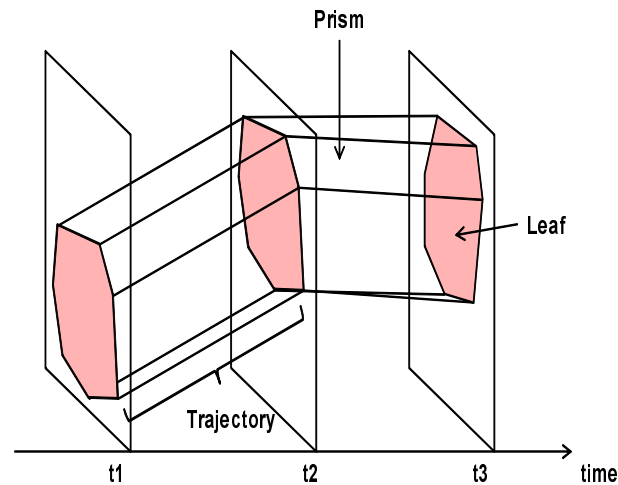


図 2.1 ISO19141:2008 の Prism モデル

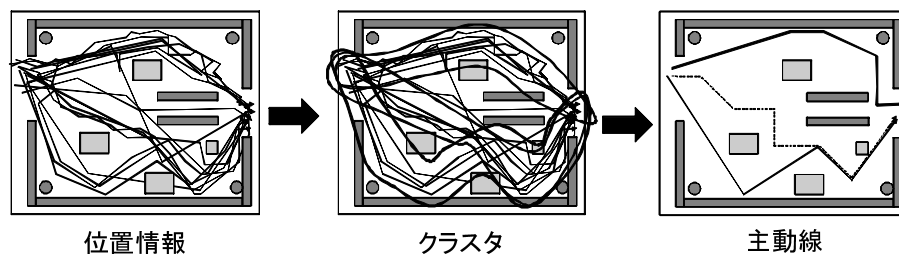


図 2.2 歩行者行動の例 [53]

標準形式を用いて記録され、多くの利用者の間で共通利用されていくと考えられる。

2.1.2 動線解析

代表的な動線解析の例として、主動線抽出処理を図 2.2 に示す。主動線抽出処理とは蓄積された測位データから代表的な動線を抽出する処理である。蓄積した測位データをもとに意志決定する場合、多数の動線を読み取りその意味を理解する必要があるが、図 2.2 左の例のように複雑なデータの目視確認は容易ではない。そこで、動線の集合を類型化し、代表的な動線（主動線）を抽出することが求められる。この主動線抽出は、以下の 2 つのステップにて構成される。最初のステップでは、図 2.2 中央、太線に囲まれた部分のように、類似する動線の集合に分類がなされる。このようなデータを類似性に基づき分類して得られる集合をクラスタ、クラスタを生成する処理をクラスタリングと呼ぶ。次のステップでは、図 2.2 右に示すように、各クラスタを代表する動線が生成される。この結果、各移動特徴を人が把握しやすく提示でき、それらがどの程度発生しているかが定量化できる。

クラスタリングの手法は多数提案されているが、これらの手法の多くはクラスタリングの目的に適うように定義された対象データ間の類似度を用い、類似しているものが同一のクラスタに含まれるようにする手法である。よって、動線間の類似度 [42] を定義すれば、動線にも同様のクラスタリング手法が適用でき、クラスタ内で最も平均的なデータを当該クラスタの代表とすれば、主動線の抽出が可

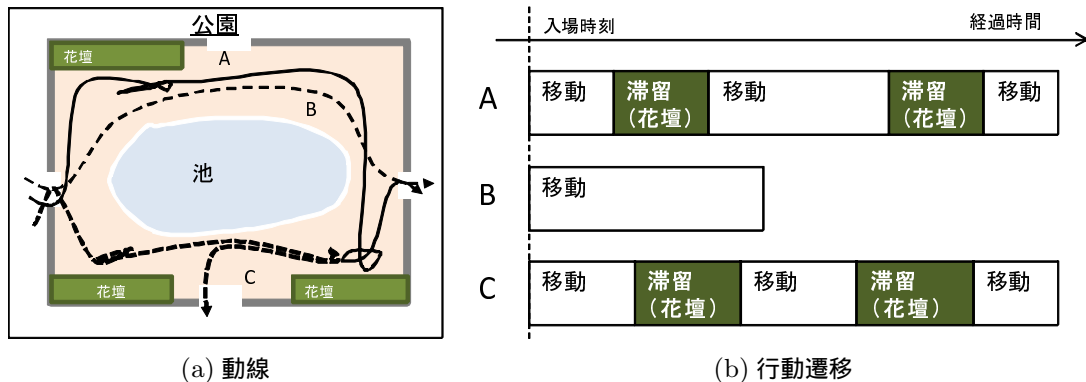


図 2.3 歩行者行動の例

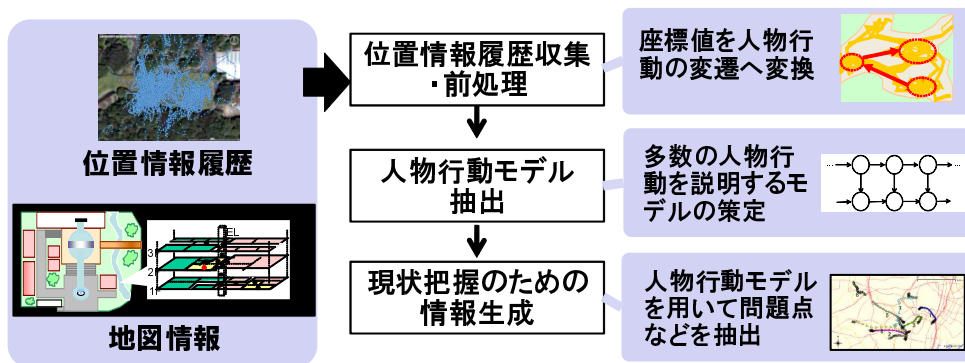


図 2.4 動線解析処理の全体構成

能であるとも考えられる。例えばカメラに写る人物動線から主動線を得る場合、動線形状間の距離を定義することによってこれらのクラスタリングが適用できる。

ところが、長時間広範囲に及ぶ歩行者の行動が類似していても動線形状は大きく異なる。例として図 2.3 に公園を散策する歩行者行動の例を示す。この場合、図 2.3-(a) のようにそれぞれの歩行者が散策する順路は異なり、各動線の形状はあまり類似しない。図 2.3-(b) に歩行者行動の時間遷移を示す。図中 A と C の動線は公園内の花壇 2 つに立ち寄る行動であり、B の動線は単に公園を通り抜ける行動である。A の行動は C の行動に類似しているにも拘らず、動線の形状はむしろ B に近く、動線形状をそのまま比較しても適切な結果は得られない。

そこで、本研究では図 2.4 の手順によってこの問題を解決する。最初に、位置情報履歴収集・前処理にて動線データを集め、その誤差などを低減するとともに、移動や滞留など状態遷移を意味する記号列に変換するなど行動のモデルに適した形に変換する。次に、人物行動モデルの抽出によって、人の行動を表現するモデルに収集データを当てはめる。主動線抽出の場合、記号列に対するクラスタリングに相当する。例えば公園で滞留を繰り返す歩行者と、公園を通り抜ける歩行者を区別するなど歩行者行動が分類される。最後に、現状把握のための情報生成を行う。この処理では、人の行動のモデルを利用者にわかりやすく提示することによって、現状把握ひいては意思決定を支援する。主動線抽出では、クラスターの代表を抽出して可視化するとともに、各クラスターの発生割合等を数値化することに相当する。

2.2 動線解析とその構成

以下では、図 2.4 の各処理について述べる。

2.2.1 前処理

図 2.4 の左図の前処理とは、意味づけされる前の時系列測位データ（元データ）に対して加えられる処理である。この処理としては、測定誤差の低減のための平滑化処理や、計測の失敗により極端に外れた位置となった測位データを除去する外れ値除去処理 [54]、測位データを地図情報と整合させるマップマッチング処理などがある。

また動線の区間に対し移動体の状態を付与して離散的な状態の遷移に変換する必要がある場合がある。その場合、動線のある区間を意味する動線データに対して、移動中や滞留中など状態を判別する状態判別処理が実行される。状態判別の方法としては、滞留地点間の遷移をもとにトリップを抽出する方法 [55]、地図情報などを用い、動線データの入っている領域ごとに状態を決める方法 [35] や、全動線データの集合の密集の度合いから自動的に抽出する方法 [56] などがある。このようにして得られた離散的な状態の遷移を表現する遷移データは、例えば、状態 a から状態 b、次に状態 c に遷移した場合には“abc”と記述するなど、記号列として記述できる。本論文ではこの記号列を状態遷移とよぶ。この記号列に離散系の統計処理を加えることにより、離散的なマルコフ連鎖とみなすことができ、「地点 A から地点 B への移動確率」などの形で遷移確率を得ることができる。これにより離散マルコフ連鎖の拡張（例えば隠れマルコフモデルなど）による潜在的な属性推定などができるようになる。

2.2.2 人の行動モデル抽出

図 2.4 の人の行動モデル抽出は、動線あるいは状態遷移に対する統計処理である。例えば、状態遷移のモデルを立てる、動線データ間の類似度を算定するなどである。本研究では、人の行動モデルは確率モデルとした。ここでいう確率モデルとは、未定な値（例えば、次の行き先）について、その確率を算定することができるモデルである。この確率モデルは歩行者行動の特徴を表現しており、人の類型や次の行き先など動線データには直接含まれていない情報の推定や予測、動線データの特徴を現す量の算出などの応用が可能である。

2.2.3 情報生成処理

情報生成処理では、人の行動モデルから有用な情報を抽出して提示する処理である。例えば店舗の顧客回転率向上を狙ってレイアウトを変更した場合、確率やクラスタの変化から、当該施策にどの程度効果があったかを確率分布のパラメータの変化をもって数値化できる。あるいは、倉庫や工場における作業者の動きの頻度を定量化して業務効率を数値化するなど、施設管理者の意志決定に重要な情報も得られる。他にも、当該確率分布にあてはまりにくい状態遷移列を得ることができれば、不審行

動やトラブルを検知することもできる。また、確率モデルに基づき歩行者が次に立ち寄るであろう地点を推測し、それに関する情報を配信することもできる。

2.3 まとめ

本章では、位置情報の計測手段である測位とそれを取り巻く状況についてまとめるとともに、位置情報を分析する手法の類型について述べ、本研究の位置づけを明確化した。以降の章では、このそれぞれについて詳細を述べる。

第Ⅱ部

人の位置の遷移モデル

第 3 章

不特定多数の人の行き先選択モデル

3.1 はじめに

本章では、ある領域に訪れた過去の不特定多数の人の時系列位置情報を用いて、新たにそこを訪れた人物が次に行くであろう箇所を予測するなど、移動する人に付随する情報を抽出する方式を提案する。

本方式の利用シーンの一例として歩行者が測位機能付きの携帯端末を持ち歩くことを想定する。この携帯端末の GPS 受信機能や測位基地局によって歩行者の位置情報が測定し、それを測位データとしてデータセンタへ送付する。データセンタではこれを動線データとして蓄積しており、地図情報を組み合わせた動線解析処理が実行できる。これにより、解析の結果を利用したサービスが提供できる。例えば、商業施設の管理者を想定すると、施設内歩行者の平均的移動パターン [57] を提示することで、設備の稼働率の見積もりや店員配置の改善を支援することができる。また、異常行動を検出 [44] すれば、施設の管理者は施設内で発生する問題に迅速に対処できるようになる。管理者だけでなく、例えば歩行者の行き先を予測 [38] することで行き先に関する情報を前もって配信したり、歩行者が来る前に機器類の待機状態を解除するといった機器制御 [35] も考えられる。これらのサービスの多くは蓄積された動線データを表現する確率モデルを用いて実装することができる。

そこで本章では、多数の動線から抽出された状態遷移を確率モデル化する方式について議論する。ここでいう状態遷移とは、「ある場所に滞留してから次のある場所へ移動する」などの位置に関連づいた離散的な歩行者の状態の時系列変化を指す。状態遷移の確率モデルを用いた動線解析として最も典型的な行き先予測の手順を図 3.1 に示す。この手順は大きく二つのステップに分割できる。(1) では、蓄積された大量の人々の状態遷移のデータから、その傾向をもっともよく表現する確率モデルとパラメータを算出する。これにより、人々の状態遷移を与えるとその発生確率が算出できるようになる。この処理は夜間等十分な時間の確保できるタイミングで実行すればよい。(2) は (1) の統計処理の結果を受けサービスを提供する処理ある。ここでは確率モデルを用いて実際に確率を計算し、次に人が移動する確率が最も高い地点を行き先の予測結果とする。予測の結果を活用するには、移動の前にこの予測処理が終わっている必要があり、一度の移動よりも短時間（分単位程度）で処理が終わることが求められる。

この予測の手順においては、確率モデルとしてどのようなモデルを選ぶかが重要であり、人々の行

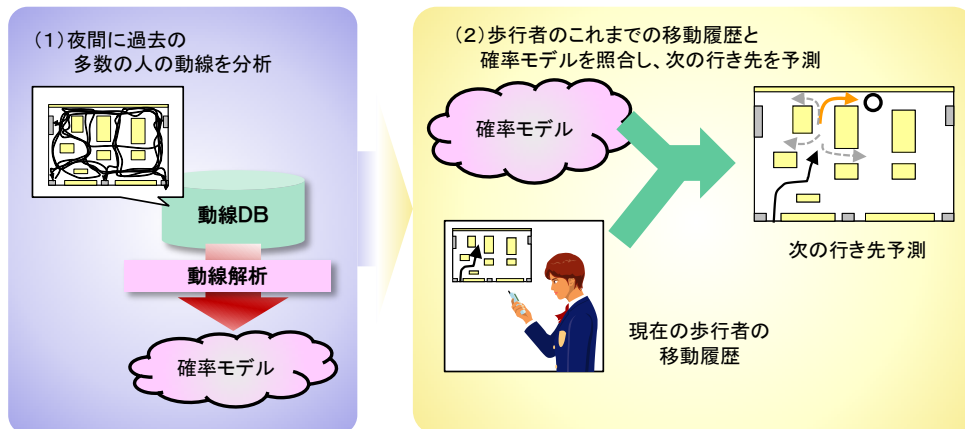


図 3.1 動線解析を用いたサービスシステム

動を表現するのに適切なモデルを選べば、より適切に人物の移動を表現できると考えられる。今回、歩行者の行動を表現する確率モデルとして、マルコフ連鎖にもとづくモデルを用いる。なかでも、混合マルコフモデルおよびそれに時間経過による状態の変化を導入した混合自己回帰隠れマルコフモデルを提案する。マルコフ連鎖は現在の状態をもとに次の行き先の確率分布が定まるという単純なモデルであり、事前にパラメータの計算を行っておけば予測等の処理が短時間で完了できるため、上述の用途に適している。混合マルコフモデルは歩行者の移動がいくつかのパターン（例えば、施設内を通り抜けるだけの人、明確な目的地がある人、目的なく彷徨している人など）に分類できると仮定して、このマルコフ連鎖を複数組み合わせさせたモデルであり、交通需要の推定でも同様の要因を加味したものが提案されている [58] など、予測等の精度は単純なマルコフ連鎖のモデルよりも高精度になると期待される。混合自己回帰隠れマルコフモデルでは、そのパターンが経過時間に伴って変化することもモデルに含めており、より高い精度での予測が期待できる。

さらに、本章ではいくつかのデータを用いて実験的にその精度を評価した結果についても述べる。ここで、これらの確率モデルの妥当性の基準として、人がこれまでに通った経路に関する動線データを元に、次の行き先を予測し、その正解率を用いた。この評価により、人の行動を記述する確率モデルとして、マルコフ連鎖にもとづくモデルの妥当性を確認できる。また、行動パターンの分類や時間経過がどの程度、人の行き先選択に寄与しているかが比較検証できると考えられる。

3.2 マルコフ連鎖モデル

3.2.1 従来のマルコフ連鎖モデル

本節では状態遷移を近似できる種々の確率モデルについて議論する。ここでいう確率モデルとは、分析対象となる歩行者の移動確率を表現したモデルであり、歩行者の移動確率の算定ができるモデルである。このモデルにより、遷移確率が最大である状態を計算すれば歩行者の行き先予測が可能となる。他にも、例えば新たな状態遷移を確率モデルにあてはめて尤度を求めれば状態遷移が例外的であるかの判別が可能となる。また、最尤な状態遷移系列を求めれば典型的な移動経路（主動線）生成が

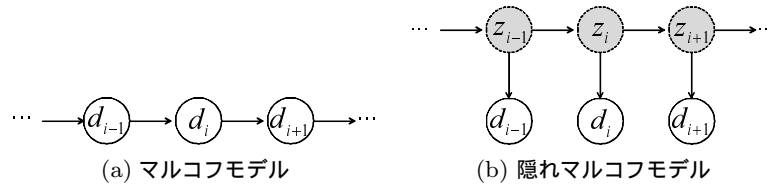


図 3.2 一般的なマルコフモデルの例

可能となる。

状態遷移を適切に表す確率モデルを生成するためには、歩行者が自身の行動を決定し、実際に行動に移すまでのプロセスを表すのに適したモデルを選択せねばならない。状態の遷移を表現できる簡単なモデルとして、マルコフ連鎖モデルがある。マルコフ連鎖とは、ある状態が次の状態に遷移するとき、現在の状態のみに依存して遷移確率が定まるモデルである。従来知られているマルコフ連鎖モデルのグラフィカルモデルを図 3.2 に示す。最も単純でよく用いられているのが (a) の単純なマルコフ連鎖モデル (SMC; Simple Markov Chain) である。グラフ中の $\{d_i\}$ は測位データに状態判別処理を実行することによって得られる観測可能な状態である。このモデルでは、人は今の状態（場所と移動、滞留）だけから次の状態を決定すると仮定されている。しかし、人の行動がそれほど単純ではないことは明らかである。

そこで、人が外部から観測できない内部的な状態（例えば、興味対象、感情、空腹度合いなど）を保有しており、この状態変数の影響を受けて行動を決定するというモデルが考えられる。以降では、測位データにより観測される状態 d_i を空間状態、観測不可能な内部的な状態を示す変数 z_i を内部状態と呼ぶ。このような隠れ変数を取り入れたモデルとしては、図 3.2(b) の隠れマルコフモデル (HMM; Hidden Markov Model) がよく知られている。HMM では内部状態のみによって空間状態の確率が決まる。内部状態は観測不可能であるので内部状態とその遷移確率は、観測可能な空間状態によって推定されなければならないが、各状態遷移の原因とな内部状態と内部状態間の遷移確率を同時に決定する方法 [59] が知られており、多くの状態遷移を収集できれば推定可能である。HMM を用いた動線解析の例としては例外行動の検出などが知られている [44, 38, 60, 61]。ただし、HMM における空間状態が実現する確率はその直前の空間状態に依存していない。実際には、人は短時間では空間的に近い地点にしか行けないという制約があるため、直前の空間状態によって次の歩行者がいる空間状態は大きく制約を受けるのは自明である。したがって、HMM よりも妥当性の高いモデルが必要である。

3.2.2 内部状態を持つマルコフ連鎖モデル

本節では、HMM とは異なるモデルとして、空間状態の実現確率が、直前の空間状態と内部状態の両方に依存するマルコフ連鎖のモデルについて議論する。図 3.3 にそのような確率モデルを示す。(a) は混合マルコフモデル (MMM; Mixed Markov Model) [53] である。MMM では内部状態が遷移することはなく、最初に定まった一つの状態のまま、空間状態のみが遷移する。空間状態は、固定の内部状態と直前の空間状態のみを用いて遷移確率が決まる。つまり内部状態を決定して以降は、固定の空間状態間の遷移確率にもとづき状態遷移するということであり、換言すると内部状態によってマ

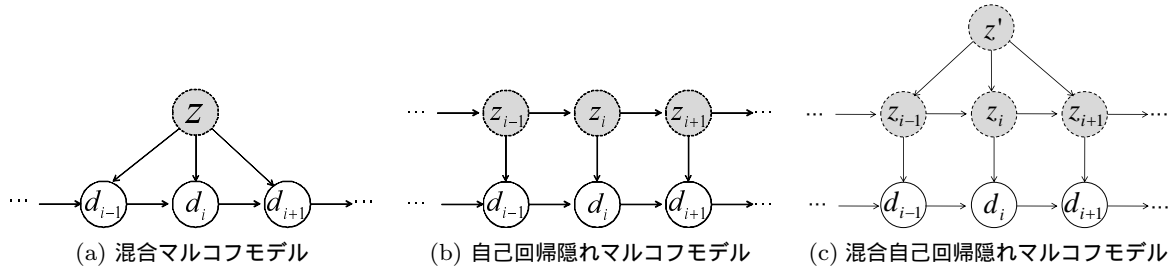


図 3.3 人の移動行動を表現する確率モデル

ルコフ連鎖の遷移確率が定まるモデルといえる^{*1}。つまり、MMM は内部状態に対応する複数のマルコフモデルを重ね合わせたモデルである。

従来、MMM は Web サイト等で次に見るページを選択するなどの行動を表現するのに用いられていた [62] が、空間的な制約（つまり、「遠隔地へ瞬間移動はできない」という制約）を表現するのにも適している。また、MMM では内部状態はただ一つに決まり、動線データの全ての区間において変化しない。したがって、内部状態は時間によらない特性を意味しており、人の動線をパターンとして分類したものとみなせる。これは HMM が時間変化を表現するのに適したモデルであるのと同様に、内部状態がほとんど変化しない場合や、人の挙動が個人の嗜好など時間によらない特性に強く依存する場合、MMM による近似が適する。しかし、分析の対象とする動線データが長時間であるような場合、内部状態が変化することも考慮に入れることが望ましい。

そこで図 3.3(b) に内部状態の変化を加味したモデルを示す。このモデルでは、空間状態 d_{i+1} は直前の空間状態 d_i と内部状態 z_i に依存する確率 $P(d_{i+1}|d_i, z_i)$ によって決定され、それと同時に内部状態 z_{i+1} も直前の内部状態 z_i に依存する確率 $P(z_{i+1}|z_i)$ にもとづき遷移する。このモデルは、自己回帰隠れマルコフモデル (AR-HMM; AutoRegressive Hidden Markov Model) と呼ばれるモデルの一例になっている [59]。行き先の予測精度を高めるには、この AR-HMM のように内部状態の変化を取り扱うことが望ましい。ところが、このモデルは複雑であり、パラメータ推定に多数のデータが必要という問題がある。そこで、本章ではパラメータ数の少ない AR-HMM を複数組み合わせた図 3.3(c) の混合自己回帰隠れマルコフモデル (MAR-HMM; Mixed AutoRegressive Hidden Markov Model) を提案する [63]。MAR-HMM は AR-HMM を複数混合させた形になっており、内部状態 z' によってどの AR-HMM にしたがって遷移するかが定まる。このモデルの構造は複雑に見えるが、遷移する内部状態 z_i と、どの AR-HMM にしたがっているかを示す内部状態 z' の組をもって一つの内部状態とみなせば、実質的には AR-HMM の内部状態の遷移に、AR-HMM 間の遷移を禁止するという制限をかけたものと同等である。つまり、結果的に同規模の AR-HMM との比較では、決定すべきパラメータの数を削減できていると考えられる。以降では、これらのモデルについて詳しく述べる。

^{*1} これには仮定すべき内部状態が少なく、比較的少ないデータでもモデルが計算できるという利点もある。

3.3 混合マルコフ連鎖モデル

3.3.1 モデルの定式化

MMM[64] は、モデルの内部状態を 1 つに固定し、それが時間変化しないとしたモデルである。MMM では、最初に決定される 1 つの内部状態のみにより空間状態間の遷移確率が決まり、以降、空間状態の遷移は直前の空間状態によってのみ定まるマルコフ過程となる。MMM では内部状態間の遷移がないため、推定すべき内部状態は 1 つしかない。仮定すべき内部状態が少ない分、HMM に比べて少ないデータでもモデルパラメータが計算できる。特に内部状態がほとんど変化しないと仮定できる場合や、人の挙動が時間によらない特性 (個人の嗜好など) に強く依存している場合、MMM による近似は妥当である。

MMM による人の移動のモデル化では、状態遷移のデータセットによる最尤法によって MMM のパラメータを決定する。最尤法は尤度が最大となるようパラメータを決める方法である。尤度とは確率モデルが観測データを説明する度合いを意味し、確率分布関数に観測データを代入して得られる。MMM の空間状態遷移の集合 $d_n \in D$ 、ただし $d_n (n = 0, \dots, N)$ に対する確率分布は、

$$P\left(d|\{\pi^{(k)}\}, \{\omega_{\mu\nu}^{(k)}\}\right) = \sum_k^K \pi^{(k)} \psi(d|\{\omega_{\mu\nu}^{(k)}\}) \quad (3.1)$$

である。ここで、 $\{\pi^{(k)}\}$ は k 番目のモデルに属するデータの発生確率 (以降、 k -混合比率とよぶ)、 $\{\omega_{\mu\nu}^{(k)}\}$ は k 番目のモデルの空間状態 μ から ν への遷移確率を表す。また、 ψ はマルコフモデルの確率分布、例えば、 d が状態 $\alpha \rightarrow \beta \rightarrow \gamma$ であるなら、 $\psi(d|\{\omega_{\mu\nu}^{(k)}\}) = \omega_{\alpha\beta}^{(k)} \omega_{\beta\gamma}^{(k)}$ となるものを指す。このモデルにおけるパラメータの概念図を図 3.4 に示す。図では、3 つのモデルが混合しており、左のモデルに属する人の割合が $\{\pi^{(1)}\}$ であり、 $\{\omega_{AD}^{(1)}\}$ が左のモデルの地点 A から地点 D への移動の確率を意味する。なお、確率の規格化のため、 $\sum_k \pi^{(k)} = 1$ 、 $\sum_{\mu} \omega_{\mu\nu}^{(k)} = 1$ を満たす。これらのパラメータを、状態遷移のデータセットを最もよく説明する、すなわち尤度が最大になるように定めれば、この確率分布関数を用いて次ステップの予測や、隠れ変数、つまり人の行動の類型を求めることができる。

3.3.2 EM 法によるパラメータ推定

隠れ変数を含む確率モデルを用いた場合の最尤法としては、EM (Expectation-Maximization) 法がよく用いられている。隠れ変数を含む確率モデルでは、パラメータの推定には隠れ変数が、隠れ変数の推定にはパラメータが、それぞれ互いに必要であり、それらを同時に最適化しなければならない。そこで EM 法では仮のパラメータのもとで隠れ変数の期待値を求める E ステップと、隠れ変数を期待値で置き換えてパラメータを求める M ステップを交互に繰り返し、両者の最適値へ収束計算させる。

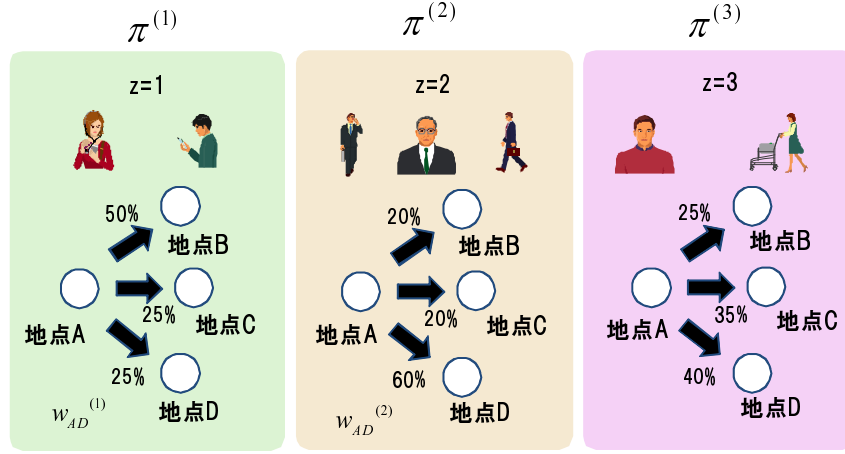


図 3.4 混合マルコフモデルの概念図

混合マルコフモデルに最尤法を適用するということは、データ D の対数尤度 $L(D)$

$$L(D) = \sum_n \log P(d_n | \{\pi^{(k)}\}, \{\omega_{\mu\nu}^{(k)}\}) = \sum_n \log \left(\sum_k \pi^{(k)} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) \right) \quad (3.2)$$

を制約条件 $\sum_k \pi^{(k)} = 1$ 、 $\sum_{\mu} \omega_{\mu\nu}^{(k)} = 1$ のもとで最大化することである。そこで Lagrange の未定定数法を適用すると（詳細は付録 A に記載）、

$$\pi^{(k)} = \frac{1}{N} \sum_n \frac{\pi^{(k)} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})} \equiv \frac{1}{N} \sum_n \gamma^{(k)}(d_n) \quad (3.3)$$

$$\omega_{\mu\nu}^{(k)} = \frac{\sum_n c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)}{\sum_{n,\nu} c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)} \quad (3.4)$$

が得られる。ただし、状態遷移 d_n において状態 μ から状態 ν への遷移が起こった回数を $c(d_n, \mu \rightarrow \nu)$ とおいた。この $\gamma^{(k)}(d_n)$ は d_n の k 番目のモデルへの適合の度合いを表し、また、定義から $\sum_k \gamma^{(k)}(d_n) = 1$ である。つまり、 d_n が $\mu \rightarrow \nu$ の遷移を 1 度含んでいるとき、その回数を各モデルに分配して $\gamma^{(k)}(d_n)$ 回ずつ数えたのに相当し、同遷移が $c(d_n, \mu \rightarrow \nu)$ 回あった場合には各モデルに $c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)$ 回と分配したことになる。 $\omega_{\mu\nu}^{(k)}$ は遷移の回数をすべての行き先 ν の回数の和で割った値であり、 $\gamma^{(k)}(d_n)$ にもとづき分配された頻度分布が $\omega_{\mu\nu}^{(k)}$ である。この特徴から $\gamma^{(k)}(d_n)$ を、他の混合分布における同様の値にならい、負担率と呼ぶ。

EM 法ではこの負担率を求める E ステップと最尤パラメータを求める M ステップを繰り返し収束させる。したがって、混合マルコフモデルの EM 法は図 3.5 のようになる。移動行動の予測では、この結果を用いて d_n の m 番目の空間状態を $d_{n,m}$ とすると、

$$P(z_{nk}, d_{n,m+1} | \{d_{n,m}\}) \propto \sum_n \gamma^{(k)}(\{d_{n,m}\}) \psi(\{d_{n,m+1}\} | \{\omega_{\mu\nu}^{(k)}\}) \quad (3.5)$$

が最大となる $d_{n,m+1}$ を計算すればよい。

なお、大域解への収束が保証されている混合ガウス分布の EM アルゴリズムとは異なり、マルコフ連鎖のパラメータ決定は離散的であるため、EM 法の計算結果が局所解に収束することが多い。最

初期化 $\{\pi^{(k)}\}$ 、 $\{\omega_{\mu\nu}^{(k)}\}$ 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。

$$\gamma^{(k)}(d_n) = \frac{\pi^{(k)} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})}$$

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。

$$\pi^{(k)} = \frac{1}{N} \sum_n \gamma^{(k)}(d_n)$$

$$\omega_{\mu\nu}^{(k)} = \frac{\sum_n c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)}{\sum_{n,\nu} c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)}$$

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 3.5 EM 法による混合マルコフモデルのパラメータ推定

適解を得るには、計算が収束した時に更新式に微小な乱数を加えて局所解から脱出できるようにする方法や多数の初期値で計算を行って尤度の最も高いものを選ぶなどすることが必要になる。また、通常、最尤な解が複数あることはほぼない（たまたま尤度が一致する 2 つの最尤解があるようなケースのみ）が、相互に干渉しない区画がある場合にはその入れ替えに相当する自由度がある点は注意を要する。たとえば、地点 A,B,C の間の移動と、地点 D,E,F の間の移動が独立しており、A から B や B から C に行くことはあっても A から D や C から E に行くことは絶対でない、というようなケースである。この場合、地点 A,B,C の遷移のモデルと地点 D,E,F の間のモデルはそれぞれ別個に計算してもよく、これらをまとめて計算することは別個に計算した結果を単に並べて連結させるのと同様である。すなわち、連結させる組み合わせの数だけ、解が存在することになる。ただし、それらは単に連結のさせ方の自由度によるものであり、本質的に別の解というわけではなく、位置の遷移として行き来がある空間のみを対象とすべきという点に注意すれば十分と考えられる。

3.4 混合自己回帰隠れマルコフ連鎖モデル

3.4.1 自己回帰隠れマルコフモデル

MMM では、各状態遷移が一つの変化しない内部状態によって定まる確率分布に従って発生するとしていたが、長時間の動きでは内部状態の変化を加味することが必要である。この内部状態の変化を取り入れたモデルが、AR-HMM(自己回帰隠れマルコフモデル) である。図 3.6 に AR-HMM のモデルの概念図を示す。AR-HMM は、HMM の一種であり、大きく分けて 3 種類のパラメータがある、図中に示すとおり、内部状態の初期値が m となる確率を π_m 、内部状態 l から内部状態 m への遷移確率 $A_{l,m}$ 、直前の空間状態 μ 、内部状態 m のときに空間状態 ν となる確率 $\phi_{m,\mu,\nu}$ の 3 種類である。ここで、内部状態 M 種類のとき m は $0 < m < M$ の整数、空間状態の K 種類のとき μ, ν は

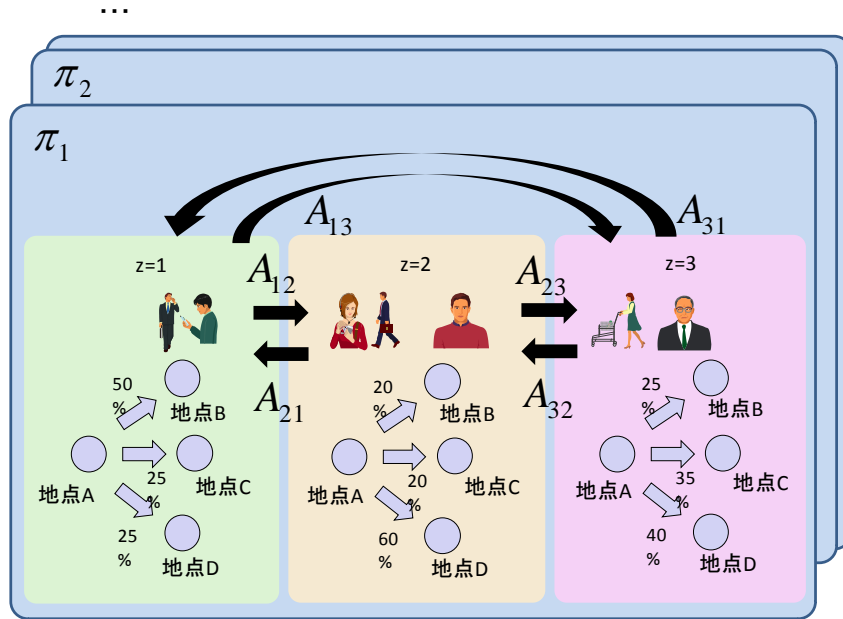


図 3.6 自己回帰隠れマルコフモデルの概念図

それぞれ $0 < \mu < K, 0 < \nu < K$ の整数である。

人の移動行動のモデル化では、これらのパラメータを状態遷移のデータセットから最尤法によって学習する。HMM の場合、パラメータを高速に推定する手法として、FB(Forward-Backward) 法が知られているが、この手法は AR-HMM にも適用できる。FB 法は HMM に EM 法を適用する場合のパラメータ推定を高速に行う手法である。EM 法は、仮のパラメータのもとで隠れ変数の期待値を求める E ステップと、隠れ変数がこの期待値であると仮定して最尤パラメータを求める M ステップを交互に繰り返し収束計算させる方法であるが、HMM では時間経過にしたがって変化する隠れ変数を扱っているため、E ステップでは各時刻の隠れ変数の期待値を全て求めなければならないが、この処理は非常に計算量が多い。その理由は、隠れ変数の確率はその直前の時刻における隠れ変数の値によって決まるためである。つまり、隠れ変数の期待値を求めるために必要な確率を計算するために、直前の隠れ変数の確率も計算が必要なのである。すると、その直前の隠れ変数の確率を計算するために二つ前の確率も必要であり、それを計算するのに三つ前の確率も必要であり、と繰り返し計算が必要であり、計算量が膨大になってしまう。FB 法はこの期待値計算の重複に着目し、一度計算した結果を保存しておくことにより、重複計算を避けている。AR-HMM は一般的な HMM とは少々異なっているが、この FB 法の原則は変わらないため、ほぼそのままの形で適用できる。

AR-HMM に対する FB 法のアルゴリズムを図 3.7 に示す。なお、学習に用いる N 個の空間状態列を $\{d_n\}$ 、各系列の t 番目の値を $d_{n,t}$ と記述する。この $d_{n,t}$ は K 次元のベクトルであり、 n 番目の状態遷移の t 番目の状態が k のとき、 $d_{n,t,k} = 1$ それ以外は $d_{n,t,k} = 0$ を満たす。また、各空間状態の長さ (t の上限) を T_n と書く。E ステップでは、まず各空間状態遷移列に対し、漸化式を用いて α_n と β_n を再帰的に計算する。式中の c_t は変数を計算機の計算可能な精度内に抑えるためのスケールリング係数と呼ばれる値で、 $\alpha_{t,j}(d_n)$ の再帰計算の中で同時に計算できる量である。次に α_n と β_n を用いて $\gamma_{t,i}(d_n)$ と $\xi_{t,t+1,i,j}(d_n)$ を求める。これらが E ステップで求める期待値に相当する。M ス

初期化

$\{\pi_i\}, \{\phi_{i,\mu,\nu}\}, \{A_{i,j}\}$ 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 $\alpha_{t,j}(d_n)$ と $\beta_{t,i}(d_n)$ を再帰的に求める。

$$\begin{aligned}\alpha_{t,j}(d_n) &= \frac{1}{c_t(d_n)} \sum_{i,\mu,\nu} \alpha_{t-1,i}(d_n) A_{i,j} \phi_{j,\mu,\nu} d_{n,t,\mu} d_{n,t-1,\nu}, \\ \beta_{t,i}(d_n) &= \frac{1}{c_{t+1}(d_n)} \sum_{j,\mu,\nu} \beta_{t+1,j}(d_n) A_{i,j} \phi_{j,\mu,\nu} d_{n,t,\mu} d_{n,t-1,\nu} \\ \text{ただし } c_t(d_n) &= \sum_i \alpha_{t,i}(d_n), \alpha_{0,i}(d_n) = \pi_i, \beta_{T,i}(d_n) = 1\end{aligned}$$

次に、以下の式で負担率 $\gamma_{t,i}(d_n)$ と $\xi_{t,t+1,i,j}(d_n)$ を求める。

$$\begin{aligned}\gamma_{t,i}(d_n) &= \alpha_{t,i}(d_n) \beta_{t,i}(d_n) \\ \xi_{t,t+1,i,j}(d_n) &= \frac{\alpha_{t,i}(d_n) \beta_{t+1,j}(d_n)}{c_{t+1}(d_n)} \sum_{\mu,\nu} A_{i,j} \phi_{j,\mu,\nu} d_{n,t,\mu} d_{n,t-1,\nu}\end{aligned}$$

M ステップ

E ステップで求められた負担率を用い、以下の式でパラメータを更新する。

$$\begin{aligned}\pi_i &= \frac{\sum_n \gamma_{0,i}(d_n)}{\sum_{n,i} \gamma_{0,i}(d_n)} & A_{i,j} &= \frac{\sum_{n,t} \xi_{t,t+1,i,j}(d_n)}{\sum_{n,t,j} \xi_{t,t+1,i,j}(d_n)} \\ \phi_{i,\mu,\nu} &= \frac{\sum_{n,t} \gamma_{t,i}(d_n) d_{n,t,\mu} d_{n,t-1,\nu}}{\sum_{n,t,\nu} \gamma_{t,i}(d_n) d_{n,t,\mu} d_{n,t-1,\nu}}\end{aligned}$$

収束計算

上記 E ステップと M ステップを収束するまで繰り返す。

図 3.7 自己回帰混合マルコフモデルのパラメータ推定アルゴリズム

テップではこの計算結果を用いてパラメータが更新され、これら E ステップと M ステップを繰り返られることによりモデルパラメータが定まる。

この計算において、 α_{ni} と β_{ni} の数は $M \sum T_n$ に比例しており、非常に多い。とくに長い空間状態列を扱うには、内部状態の数も多くしなければモデルが適切にならないと考えられる。一方で、モデルのパラメータ数も M に大きく依存し、特に内部状態間の遷移確率 $A_{i,j}$ は M^2 に比例する。よって、内部状態の種類を多くするとパラメータが非常に多くなり、極端に多くの状態遷移のデータがないとパラメータが決定できない。したがって、限られた動線データから確率モデルを構築するには、内部状態の数を増大させずに移動行動を表現できるようにする方法が必要になる。

3.4.2 混合された自己回帰隠れマルコフモデル

AR-HMM では非常に多くの内部状態を仮定せねばならず、計算すべき期待値や決定すべきパラメータが極端に増大してしまうという問題があった。そこで、本研究では、内部状態の遷移範囲に制限を設けることにより、パラメータの数を削減することを考える。一般的には、パラメータ数は多いほうが表現力が高いが、歩行者行動を表現できるような削減のしかたであれば、パラメータ数削減に

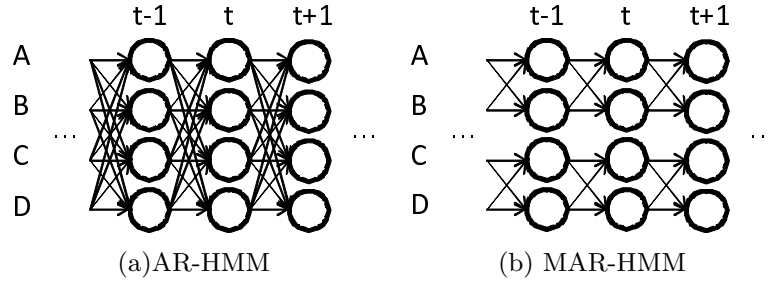


図 3.8 隠れ変数の状態遷移

	1	2	3	4	...
1	A_{11}	A_{12}	A_{13}	A_{14}	...
2	A_{21}	A_{22}	A_{23}	A_{24}	...
3	A_{31}	A_{32}	A_{33}	A_{34}	...
4	A_{41}	A_{42}	A_{43}	A_{44}	...
\vdots	\vdots	\vdots	\vdots	\vdots	

(a) AR-HMM

	1	2	3	4	...
1	A_{11}	A_{12}	0	0	...
2	A_{21}	A_{22}	0	0	...
3	0	0	A_{33}	A_{34}	...
4	0	0	A_{43}	A_{44}	...
\vdots	\vdots	\vdots	\vdots	\vdots	

(b) MAR-HMM

図 3.9 状態遷移行列 A_{ij} の比較

伴う性能劣化は大きくないはずである。それどころか、サンプル数が大きくない場合は、パラメータの推定精度が上がることによる性能向上も期待できる。例えば、AR-HMM を用いて推定すると 0 に近くなることが事前に期待できるパラメータを、計算によって求めるのではなく予め 0 と固定することにより、その分残りのパラメータの推定に割り振ることができ、計算量の削減と精度の向上が両立すると考えられる。

図 3.8(a) に AR-HMM の内部状態の遷移を模式的に示す。図中の横方向は時間遷移、縦方向は内部状態を示し、矢印は遷移確率が 0 でないことを意味する。例えば、時刻 t において状態 1 ならば時刻 $t+1$ までに状態 2 に遷移する確率が 0 でないことを、時刻 t 状態 1 の円から時刻 $t+1$ 状態 2 の円への矢印で表している。つまり、この矢印を減らすことが、内部状態間の遷移範囲に制限を加える、言い換えれば、いくつかのパラメータを 0 に固定することを意味する。本研究では、図 3.8(b) のように内部状態をいくつかのグループに分け、時間が経っても同じグループ内でしか移動しないようにする。すなわち、最初にどのグループに属するかが決まり、あとはそのグループ内で遷移を繰り返す。この仮定は、人物の行動パターンにはある程度のまとまったグループがあり、当該グループごとに特定の内部状態の時間変化がある、という仮定である。この仮定は、MMM のモデルと同質の仮定である。簡単のため、グループの大きさを均一と仮定すると、ちょうど複数の AR-HMM を混合させた形になる。これを混合自己回帰隠れマルコフモデル (MAR-HMM) とよぶ。

図 3.9 は AR-HMM と MAR-HMM のパラメータの比較である。図の 2 つの表は隠れ状態の遷移確率パラメータ A_{ij} を示しており、行は遷移前の状態、列は遷移先の状態を意味する。すなわち、1 の行 2 の列の値 A_{12} は状態 1 から 2 への遷移確率を表す。(a) の AR-HMM では M^2 個のパラメータが必要であるのに対し、(b) の MAR-HMM ではグループをまたがる遷移確率は 0 で固定されている。このグループ数を L とすると (b) のパラメータ数は $L(M/L)^2$ 個になるので、MAR-HMM

ではパラメータ数が $1/L$ に削減できたことになる。このように、MAR-HMM は AR-HMM の一部のパラメータを 0 とした特別な形であるとみなせる。なお、内部状態の種類を 1 にすると図 3.9(b) は対角成分のみの時間変化しない混合モデルとなり、このとき、MAR-HMM は MMM と等価になる。言い換えれば、MAR-HMM は MMM と AR-HMM の中間のモデルといえる。したがって、MAR-HMM のパラメータ推定のアルゴリズムは、図 3.7 に示した AR-HMM のパラメータ推定アルゴリズムにおいて一部のパラメータを 0 に固定したものとなる。ここで、図 3.7 のアルゴリズムにおける $\xi_{t,t+1,i,j}(d_n)$ の計算式と $A_{i,j}$ の更新式をくらべると、 $A_{i,j} = 0$ であるならば $\xi_{t,t+1,i,j}(d_n) = 0$ となり、 $\xi_{t,t+1,i,j}(d_n) = 0$ ならば次の $A_{i,j} = 0$ となることがわかる。すなわち、0 に固定されるべき $A_{i,j}$ の初期値を 0 とするだけで、 $A_{i,j}$ は 0 のままに固定される。また、同様に常に $\xi_{t,t+1,i,j}(d_n) = 0$ となる $\xi_{t,t+1,i,j}(d_n)$ も存在するが、このような常に 0 となる変数の記憶領域や計算を省略することにより、計算速度の改善も期待される。

なお、AR-HMM による行動予測は、これまで得られた状態遷移列 $\{d_{n,t}\}$ をもとに、次の状態 $\{d_{n,t+1}\}$ の尤度を求め、尤度が最大になる状態を予測値とすればよい。この計算には前述の $\alpha_{t,i}(d_n)$ を用いることができ、

$$p(d_{t+1,\mu}|\{d_t\}) = \sum_{i,j,\nu} \phi_{j,\mu,\nu} d_{n,t,\nu} A_{i,j} \alpha_{t,i}(d_n) \quad (3.6)$$

のように求められる。

MAR-HMM は AR-HMM の一種とみなせるので、十分なサンプル数のもとで適切にパラメータを決定できるならば、AR-HMM は MAR-HMM や MMM と同等、あるいはそれ以上の性能になるはずである。ところが、実用上はサンプル数が限られるため、適切なパラメータを決定するのが困難である。例えば図 3.9(b) において 0 に固定されているパラメータが実際に 0 に付近になるべきであったとする。このとき、十分な数のサンプルを用いて図 3.7 のパラメータ推定を行えば、それらの値は 0 に付近になるはずである。ところが、決定すべきパラメータが多いほど、多数のサンプルとそれを処理できる計算環境が必要になる。つまり、0 付近になるべきパラメータが 0 付近であると決定するために、それだけ余分なサンプルが必要となってしまう。したがって、もし利用できるサンプル数が多くないならば、パラメータをサンプルに過剰に適合させてしまい極度に予測性能が低下するという、一般に過学習とよばれる現象を引き起こしてしまうと考えられる。しかし、単に内部状態数を減らしてパラメータを少なくすると、今度はモデルの表現力が低下して予測精度も低下してしまう。MAR-HMM はこのパラメータ数に関するトレードオフの問題に対する解決策の一つであると考えられる。もし、MAR-HMM によるパラメータの削減がモデルの表現力を失わせない、つまり、もともと 0 であるべきパラメータが 0 に固定されたのならば、結果として MAR-HMM の方が AR-HMM よりも高い性能を得られる場合がありえるのである。MAR-HMM では混合数と内部状態数のバランスによってこの 0 に固定するパラメータが決定されるので、モデルの表現力を失わずにパラメータ数を削減するように調整できる。なお、MMM は MAR-HMM と同じ方法で極端にパラメータ数を削減した場合に相当する。

3.5 状態の定義

MMM にて歩行者の状態遷移を表現するためには、動線から得られた状態遷移をモデルに当てはめる必要がある。本研究の対象は、数 100m ~ 数 km 四方面程度の広さの領域中を 10 数分 ~ 1 時間程度の時間歩行する歩行者であるため、対象領域内で計測された動線に対し状態判別を実行して得られる状態遷移記号列はモデル化の対象となる。この際、対象領域から出るまで遷移は続くため、遷移回数は制限されない。ここで問題となるのが、対象領域からの出入りの表現である。一般的には、動線の計測は領域内でしかできないので、動線への状態判別の結果には領域へ入る前後は含まれない。もし、対象領域外でも計測が可能であれば、対象領域外の動線に対しても状態判別を加えることができるが、今度はその前の状態が必要となる。これを繰り返すと長期間の状態遷移列となってしまうが、そもそも混合マルコフモデルは隠れ状態の時間変化がないとしたモデルであるため、隠れ状態に変化が生じるほど長期間の状態遷移には適さない。

そこで本研究では対象領域外の動線は扱わず、対象領域に入る前後をそれぞれ一つの状態とする。すなわち、動線の対象領域内の部分だけを状態遷移に変換し、その前後に、対象領域に入る前の状態と出た後の状態を追加する。これは対象領域外を定数に繰込んだことを意味する。その結果、各状態遷移の最初の状態と最後の状態は全て同一状態となる。この変更に伴い、一般的な混合マルコフモデルにおける最初の状態の選択確率というパラメータは不要となる。ただし、対象領域に入る前の状態から次の状態への遷移確率は本パラメータと等価である。また、対象領域から出た後の状態を導入したことにより、状態遷移の終点が表現され、対象領域からの退場が予測可能となった。なお、HMM のような空間的な制約がないモデルにこれらの状態を導入すると、対象領域内の状態遷移中に対象領域に入る前が現れることがある。また、領域へ入る前後の状態は全てのデータに登場するため、領域への出入りの登場確率が高すぎる結果が得られてしまう。つまり、対象領域からの退場は HMM では予測困難である。

3.6 実験による評価

3.6.1 実験概要

本実験の目的は、MMM と MAR-HMM が人物行動をどの程度表現できるか評価することである。もし、サンプルから推定された確率モデルが適切に移動行動を表現していれば、当該確率モデルにより歩行者の行動が予測できるはずである。

そこで今回、3 つの実験を行い、状態遷移に対する予測精度を評価した。これにより、確率モデルの移動行動に対する妥当性を確認できる。一つ目の実験 (実験 1) は、群集シミュレーションデータに関する予測実験である。群集シミュレーションが出力する人の動きは、実際の人の動きに比べると非常に単純であり、予測がしやすいと考えられる。そのため、この実験により、各確率モデルの基本的な性能を測ることができる。次の実験 (実験 2) は、屋内 (商業施設) の人の動きに関する予測実験である。実際の人の動きは非常に複雑であるが、屋内では部屋からは廊下へしか移動できないなどの制

表 3.1 実験データの概要

項番	項目名	内容
1	データ内容	群集シミュレータ [65][66] による来訪展示の履歴
2	サンプル数	1337 件 × 10 セット
3	想定会場	技術展示会 (展示数 150 地点)
4	想定時間	11 時 ~ 12 時の 1 時間

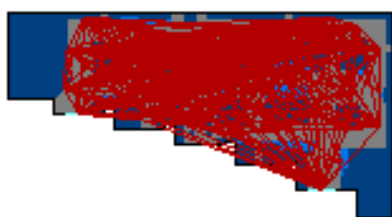


図 3.10 1 セット分の評価データ

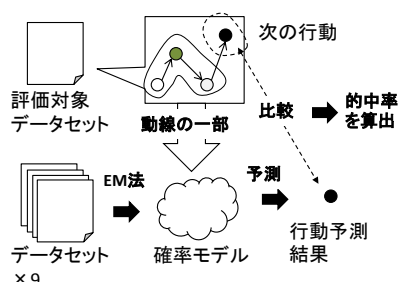


図 3.11 実験手順

約が強く、移動可能な範囲が限定的である。したがって、外乱が少なく、確率モデルごとの人の行き先選択の予測における特徴を比較しやすい。最後の実験 (実験 3) は、屋外、広域での人の動きに関する予測実験である。広域の人の動きでは、人が目的地とする地点の選択肢が多い。また、例えば同じ通勤という行動であっても、それぞれ自宅も勤務先も異なるなど、同じ意味の行動でも人によって異なる動きとなって現れるため、動きのパターンは複雑である。実験 3 はこの意味において複雑な行動のモデル化への対応力を評価できると考えられる。

以降では、これら 3 つの実験についてそれぞれ詳細を述べる。

3.6.2 実験 1: 群集シミュレーションデータ

実験 1 の条件

表 3.1 に実験 1 に用いたデータを示す。ここで用いたデータは群衆シミュレータ [65][66] により得られた、展示会の歩行者行動シミュレーションの結果である。本シミュレーションはセルオートマトンによって歩行者の移動をモデル化しており、歩行者一人一人の行動が特定のルールと乱数に基づき決定される。歩行者は展示物の近くを通ったとき、展示物ごとに決められた確率でその展示物に立ち寄るようになっており、このシミュレータには歩行者が滞留した展示物の記録を出力する機能があるため、歩行者の滞留した展示の履歴を歩行者の状態遷移とみなすこととした。なお、本シミュレーションは実際の技術展示会を想定した設定となっており、展示物の数は全体で 150 展示あり、入場者数や一部地点における歩行者数は現実と近くなるように調整されている。今回、1 時間分のシミュレーションを 10 回繰り返し、得られた 10 セットの展示来訪履歴データをサンプルとした。

本評価実験の評価手順を図 3.11 に示す。まず、10 セットのデータから 1 セットを抜き出し、評価対象データセットとし、残り 9 セットにより確率モデルを構築する。次に、評価対象セットから一つ

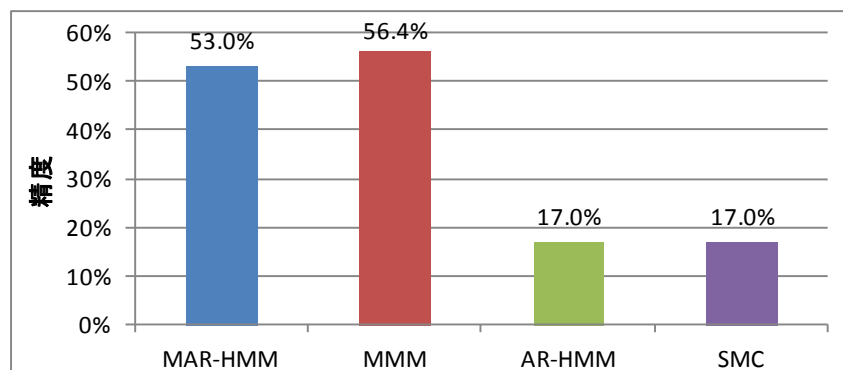


図 3.12 行動予測精度の比較

ずつ動線を取り出し、途中までの履歴から次の滞留箇所を予測する。例えば滞留展示番号が 1,12,17 の順であるサンプルに対しては、まず 1 の次の滞留箇所を予測し、次に 1,12 の履歴から次の予測、1,12,17 から次の予測、と次々と予測し、的中か否かを判定する。評価対象セットの全てに対し予測が終了したら、他のセットをサンプルとして評価を繰り返し、10 セット全てを 1 度ずつ評価し、総合的な的中確率を予測精度として評価した。

結果

次の行き先の予測精度について、SMC、MMM、AR-HMM、MAR-HMM の四つを評価した。各モデルのパラメータについては、内部状態のパターン数（図 3.8 の縦方向のパターン数）が等しくになるように、それぞれ MMM の混合数は 100、AR-HMM の隠れ状態は 100、MAR-HMM の混合数は 25 で隠れ状態を 4 つとした。

各モデルの精度比較を図 3.12 に示す。予測精度は SMC では 17.0%、MMM では 56.4%、AR-HMM では 17.0%、MAR-HMM では 53.0% であった。次候補をランダムに提示した場合を想定した予測精度の理論値は、候補数 150 なので $\frac{1}{150} \approx 0.7\%$ であり、いずれもランダムに提示するのに比べると十分高精度であるが、混合モデルを用いていない SMC と AR-HMM に比べると、MMM と MAR-HMM の精度が大きく優位な結果になった。他方、AR-HMM や MAR-HMM は時間変化を取り入れたモデルであるが、AR-HMM と SMC、MAR-HMM と MMM の差は小さく、その効果を確認できなかった。

考察

図 3.13 に、歩行者がこれまで立ち寄った展示数ごとの的中率、すなわち $n-1$ ステップ目までの滞留展示履歴を用いて n ステップ目の滞留展示を予測した場合の的中率をステップ数 n ごとに示す。横軸が予測に用いたステップ数 n 、縦軸が当該ステップ数における的中率を表す。図からもわかるように MMM、MAR-HMM は n が大きいほど予測精度が改善しており、特に $n=4$ までの精度向上が顕著である。ステップが進むことで遷移確率が変化する MAR-HMM だけでなく、遷移確率がステップ数に依存しない MMM も精度が向上しているが、この結果は、混合モデルにおけるグループを意味する潜在変数 z の推定精度が向上したためと解釈される。言い換えると、移動のステップ数が

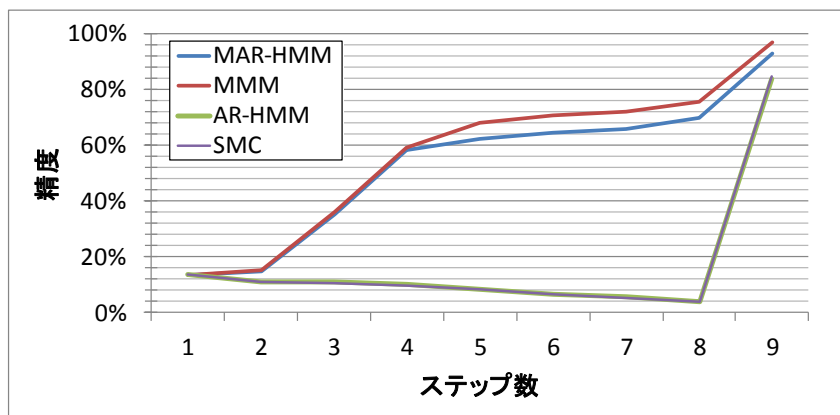


図 3.13 移動ステップ毎の行き先予測精度

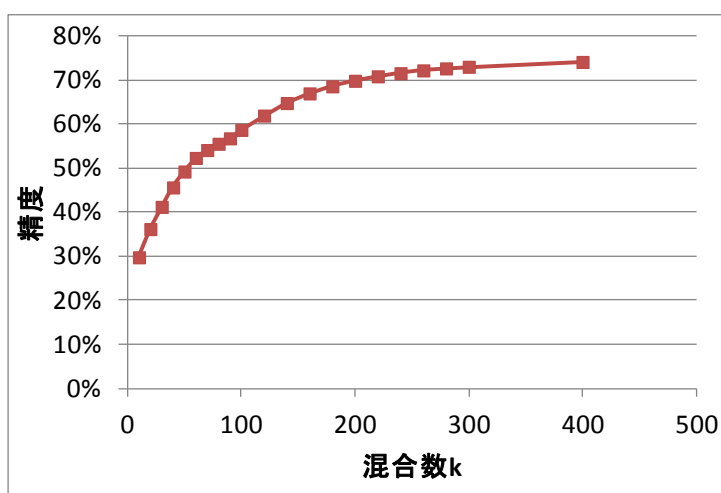


図 3.14 クラスタ数ごとの予測精度

少ない段階では当該人物がどのモデルに該当するか判断できないが、ステップが進むにつれて適切なモデルへと振り分けられたと理解できる。

もう一つ特筆すべきは、最後の行動、すなわち退場の予測精度である。今回用いたシミュレーションの履歴のほとんどのケースでは 9 ステップ目で退場しており、10 ステップ以降も会場に残っていたケースは存在しなかった。MMM にはステップ数の概念は存在しないため、歩行者が退場するステップ数そのものは予測対象にはならないが、実際には 90% 程度が的中している。ただし、精度の低いまま推移した SMC や AR-HMM についてもこの最後のステップだけは極端に精度が向上している。ここからこのシミュレーションは退場する直前の行動が非常に特徴的であったと想定できる。つまり、次に立ち寄る展示がない（他の展示に立ち寄る前に出口に達する）展示群が存在しており、そのような展示群の次は退場すると予測すれば 90% 程度の確率で予測が的中するということである。実際、この群集シミュレータでは来訪展示数に上限が設けられており、その部分だけランダム性が低くなっていたため、上記のような急激な精度向上が起きたと考えられる。

図 3.14 に、MMM の混合数を変えながら評価した結果のグラフを示す。横軸は混合数 k 、縦軸は

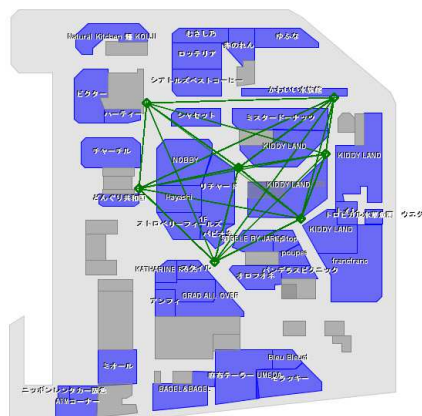


図 3.15 1 セット分の評価データ

図 3.16 実験の概要

項番	項目名	内容
1	データ源	e 空間 Kansai 実験 [67]
2	サンプル数	697 件
3	実験会場	大阪 阪急三番街 北館 1F、B1、B2
4	実験日時	2010 年 2 月 14 日 ~ 28 日
5	データ内容	イベント参加者の屋内 GPS 信号の受信記録

的中率すなわち予測精度を意味する。このグラフによると、クラスタ数が多ければ多いほど行動予測の精度が高まっているが、 $k = 400$ の場合で 74.1% の精度であり、おおむねその近辺に収束する傾向になっている。したがって、シミュレーションにおける歩行者行動を再現するためには数百程度のモデルを混合させなければならないことになる。

また、比較のため一般的な時系列データのモデルとして用いられる HMM による予測に関して同様の評価を行った。図 3.3(b) の確率モデルを用い、隠れ状態の数として 1 から 200 の範囲で最も精度が高くなるようにした。また、前述の通り HMM は退場を予測するには適さないため、HMM では退場予測は評価対象から除外した。その結果、予測精度は 4.7% とマルコフ連鎖に比べて極端に低い精度になった。人の移動は直前にいた場所に大きく影響を受ける、すなわち直前にいた場所の近くに行く、ということを反映していると考えられる。

3.6.3 実験 2: 屋内の人の移動データ

実験 2 の条件

実験 2 は屋内の人の行動の実測データを用いた実験である。表 3.1 に実測データを収集した実験の概要を示す。本実験は大阪にある阪急三番街という商業施設で行われた実証実験である。この実証実験では本商業施設への来訪者に屋内 GPS[7] の受信機能をもった端末を配布した。屋内 GPS の発信機はあらかじめ本商業施設内に設置されており、この端末は受信した位置情報などに応じたコンテンツを提示する。参加者はこのコンテンツを楽しみながら、商業施設内を自由に歩き回る。最後にこの端末を回収され、位置情報を含むログを取り出すことにより、参加者の行動を追跡し、提示コンテンツによる参加者の行動変容を確認することが、当該実証実験の目的であった。

本研究では、この実証実験にて収集された測位データを分析の対象とした。図 3.10 に収集された測位データを示す。この図は、本商業施設 1F のフロア図を背景とし、折れ線で全参加者の測位データを描画したものである。この実験で用いられた屋内 GPS では、受信した電波強度が最も強い屋内 GPS 発信機の位置を測位結果としている。したがって、得られる測位結果は屋内 GPS 電波の到達範囲に在る、という結果のみであり、測位データは離散的なものになる。そのため、図 3.10 に示され

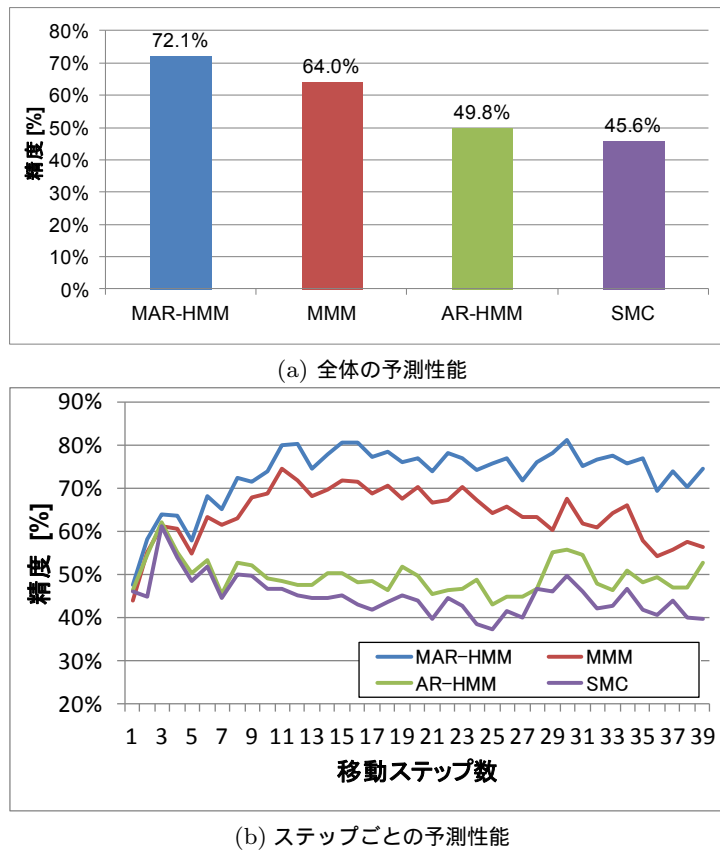


図 3.17 予測性能

た動線も離散的であり、結果として多くの動線が重なってしまい判読が困難になっている。また、計測は 5 秒に 1 回しか行われないため、どの発信機からの信号も受信できなかった時間を除去する前処理をかけた。提案方式は離散的な状態遷移列に対するものであるが、本測位データはもともと離散的であるため、そのまま状態遷移とみなして処理を実行した。

実験 2 では上記の実測移動履歴データ群の SMC、MMM、AR-HMM、MAR-HMM による行動予測精度を比較した。評価方法は 10 分割 Cross-Validation に類する方法を用いた。まず測位データの時系列群をランダムに 10 のグループに分割し、うち 9 グループをもとに確率モデルを構築、残る一つに対して行動予測を行うことを、全グループが一度ずつ予測対象になるようにして全パターン評価した。この行動予測では位置の変化 1 ステップごとにその次の行き先を予測して実際の行き先と的中したか否かを判定し、この的中率を行動予測の精度とした。

結果

図 3.17 に結果を示す。今回の評価では、状態遷移のステップごとに行動予測を実行しているが、最初のステップはほぼ受付地点にることになり、極端に予測的中率が高くなるので、評価対象からは除外した。なお、MMM では混合数を 190、AR-HMM では内部状態の数を 60 とした。また、提案方式の混合数は 95、内部状態は 4 とした。図 3.17(a) は全ステップを総合した予測的中精度の比較である。SMC は 45.6% の精度であったが、それを混合して得られた MMM は 64.0% であっ

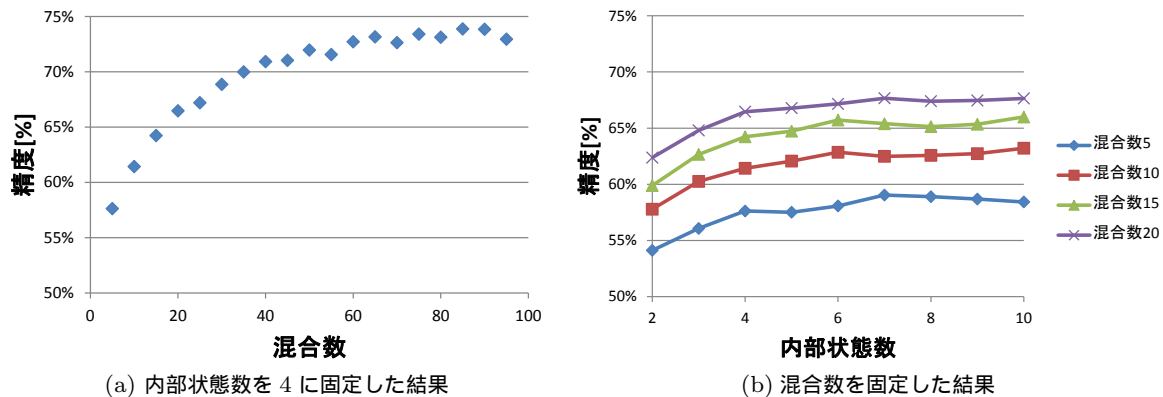


図 3.18 内部状態数と混合数の影響比較

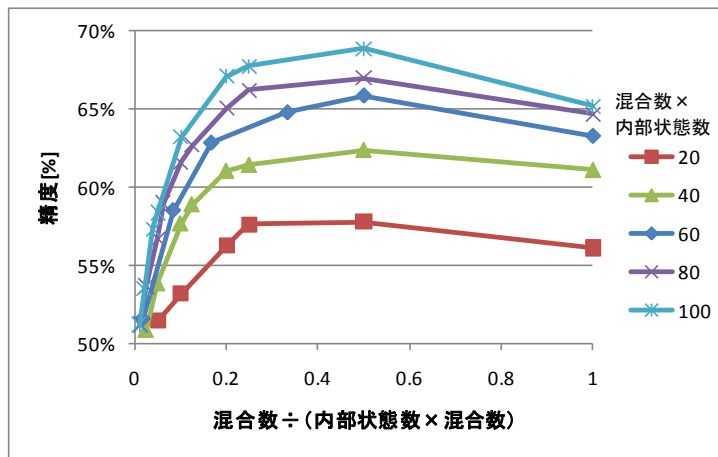


図 3.19 内部状態 × 混合数を固定した場合の精度比較

た。また、AR-HMM は 50.0% と SMC より高いものの、MMM より低い結果となった。提案方式では、これらすべてよりも精度が高く、72.0% の精度で行動予測ができることが確認できた。図 3.17(b) は、状態遷移ステップごとの予測精度の比較である。SMC では状態遷移の初期は精度が高めであるが、ステップが進むごとに精度が低下している。AR-HMM も SMC と類似の振る舞いをしているが、ステップが進んでも精度はあまり低下していない。これは、モデルに時間変化を取り入れたためであると考えられる。また、MMM では 10 ステップ目くらいまではステップが進むと精度が向上しており、SMC や AR-HMM よりも高い精度となっている。しかし、10 ステップ目以降は徐々に精度が低下している。一方、提案方式の結果は MMM と類似しているが、10 ステップ目以降の精度低下があまりないという特徴がある。これは、AR-HMM と SMC の関係と同様、時間変化をモデルに取り込んだために、ステップ数が進むことによる歩行者行動の変化をモデルに取り入れることができているためであると考えられる。

考察

図 3.18 に提案方式のパラメータと精度との関係を示す。図 3.18(a) は内部状態の数を 4 に固定し、混合数を変化させた場合の精度の変化である。グラフによると、混合数が多いほど精度は高いが、混

混合数が 60 程度以上の領域では精度はあまり変化していない。逆に、混合数を 20 に固定して内部状態を変化させた場合の精度を示したのが図 3.18(b) である。内部状態も混合数と同様に、多いほど精度が向上しているが、その向上度合いは混合数を変化したときほど小さくなく、4 程度以上の領域ではあまり改善がみられなくなっている。これらの比較から、予測精度向上においては混合数が支配的であることがわかる。また、内部状態の数も 4~6 くらいは必要と考えられる。

この関係を明示的に比較するため、図 3.19 に内部状態の数と混合数の積を一定にした場合の精度比較グラフを示す。グラフ中の各系列は内部状態数と混合数の積を一定にしたものであり、横軸に混合数と積の比を、縦軸にそのときの予測精度を示した。例えば、内部状態数と混合数の積が 60 の系列で横軸が 0.5 の点は、混合数 30、内部状態数 2 の MAR-HMM の精度を示している。図 3.9 にて示したとおり、混合数が 1、つまりグラフ左端の MAR-HMM は単なる AR-HMM であり、混合数が大きな MAR-HMM では多くのパラメータが 0 に固定される。また、内部状態数が 1、つまりグラフ右端の MAR-HMM は MMM に相当する。すなわち、図 3.19 のグラフは、グラフの左端を AR-HMM、右端を MMM として、0 に固定するパラメータ数を変化させながら MAR-HMM の精度を評価したものである。

一般的には自由に調節できるパラメータが多い AR-HMM の精度が高くなると考えられたが、実際にはむしろパラメータを減らしたほうが精度が向上するという結果になった。これは、前述のとおり、過学習が発生しているためであると考えられる。AR-HMM のパラメータは内部状態間の遷移確率 A_{ij} だけでも内部状態数 60 のときで 3600、内部状態数 100 では 10000 にもなる。本実験で用いた学習サンプルは 600 人分程度で 1 人あたりの移動の回数も高々数 10 回程度であり、決定するパラメータ数に比べて学習サンプル数が少ない。ところが MAR-HMM の場合、内部状態数 100 の AR-HMM に相当する内部状態数 2、混合数 50 の MAR-HMM でもパラメータ数は 500 程度である。これは同規模の AR-HMM に比べて非常に少ない。つまり MAR-HMM では、決めるべきパラメータ数に対する学習サンプル数が相対的に多く、高い精度が得られたと考えられる。

過学習が問題であるならばパラメータ数が少ない AR-HMM でも高い精度が得られるようにも考えられるが、実際には、パラメータ数 400 程度相当の内部状態数 20 の AR-HMM はパラメータ数 500 程度の MAR-HMM よりも精度が低い。これは、AR-HMM の内部状態数、つまり図 3.9 の表の大きさを減らすのに比べて、MAR-HMM のように一部を 0 にするほうが、より歩行者行動を表現するのに適していたためと推測される。これは 2 節で述べた歩行者行動のモデル化が妥当であったことを意味しており、歩行者行動の表現においては混合数が支配的であるということとも合致する。実際、AR-HMM で学習サンプルを用いて決定された A_{ij} の多くは 0 に近い値をとっていた。このような場合、それらをはじめから 0 に固定した方がサンプル数の面で有利と考えられる。

一方で、内部状態数を 1 にする、つまり時間変化がない MMM は、MAR-HMM に比べて精度が低い。これは、時間変化をモデルに取り入れることも意味があるためであると考えられる。つまり、MAR-HMM が歩行者行動の表現に適していたとしても、モデルを単純化しすぎると表現力が不足すると考えられる。

なお、上記計算に用いた混合数 × 内部状態数が 60 とした場合において、処理にかかった時間の例で MMM(混合数 60) では 30.8 秒、MAR-HMM(混合数 30) では 166.2 秒、AR-HMM(混合数 1) では 537.0 秒であった。この時間の計測においては、実行環境として Xeon 3.2GHz を 2 つ搭載しメモ

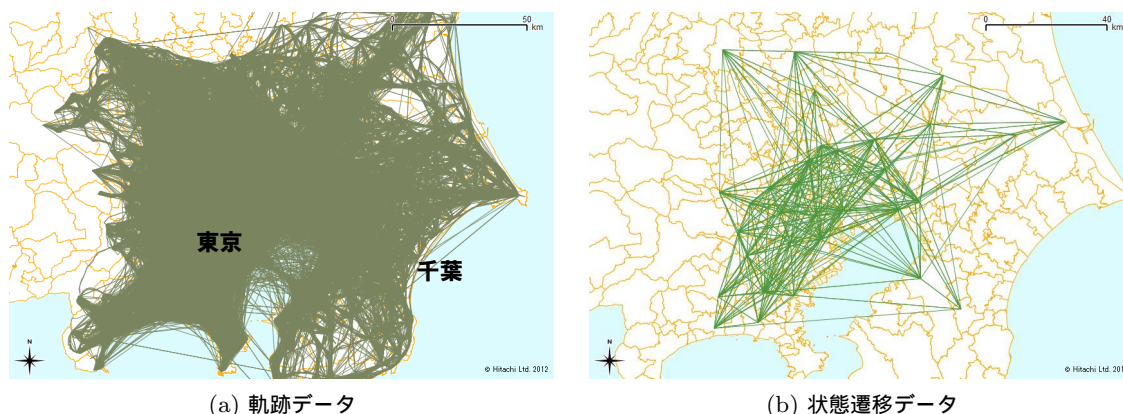


図 3.20 人の流れPJデータ

表 3.2 使用データ

項目名	詳細
使用データ	人の流れPJ[30] 東京都市圏 (H20) の位置情報 (緯度経度)
件数	毎分約 60 万人分
時間	午前 7 時～午前 9 時

リ 3GB、OS が WindowsXP 64bit のコンピュータを用い、Java(JDK6-Update 22) 環境に実装したものをを用いた。このように、パラメータ数削減に伴う計算量の削減効果も確認された。

以上のように、今回のサンプル数規模では MAR-HMM が有効であることがわかった。実際には、多数のサンプルを得ることは容易ではない場合が多い。特に来訪者人数そのものを超える数のサンプルを収集することは原理的に不可能である。したがって、混合モデルの特性と時間変化をともに取り入れた MAR-HMM は、歩行者行動をモデル化するにあたって、非常に効果的であると考えられる。

3.6.4 屋外の人移動データ

実験 3 の条件

実験 3 で評価に使われたデータセットは、屋外の推定動線である。今回用いたデータは「人の流れPJ」[27] のデータである。その概要を表 3.2 に示す。このデータは平成 20 年のパーソントリップ調査の OD(出発地と目的地) を基に作成されたもので、経路探索による補完がなされている。トリップとしては 60 万人分の位置情報が毎分 1 点記録される形になっている。

今回の評価に当たって、データが多すぎて処理時間がかかる点に配慮し、移動目的が「通勤」となっている 94,794 軌跡を対象とし、うち 2,000 軌跡を無作為に抽出して評価データとした。また、この評価では、サンプルを 10 分に 1 回に間引いた。これにより評価にかかる計算時間が大幅に短縮された。図 3.20(a) に青い線で軌跡を示す。また、この軌跡の構成点は k-means ($k=40$) によってクラスタリングされ、同一クラスにいることを同じ状態として状態遷移をつくり、マルコフ過程に当てはめた。図 3.20(b) の緑の線がその軌跡である。

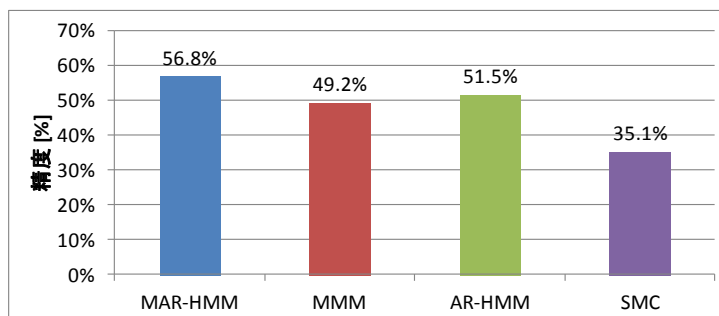


図 3.21 平均予測精度の比較

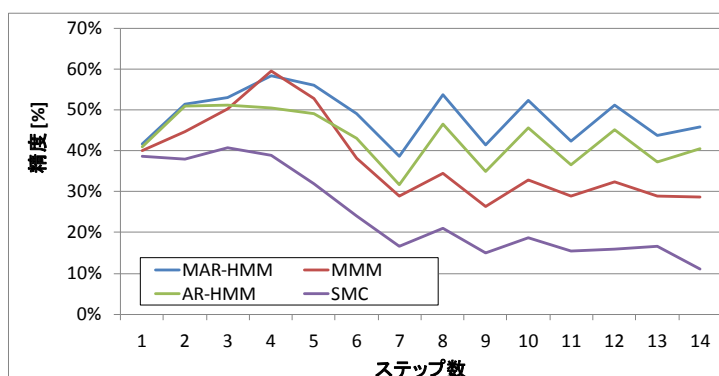


図 3.22 ステップごと予測精度

実験 3 の結果と考察

全体の精度比較を図 3.21 に示す。パラメータについては、それぞれ、MMM の混合数は 100、AR-HMM の隠れ状態は 60、MAR-HMM の混合数は 20 で隠れ状態は 2 つとした。結果としては、SMC は 35.1%、MMM は 49.2%、AR-HMM は 51.5%、MAR-HMM は 56.8% となっており、実験 1、2 と同様に、MAR-HMM がもっとも高精度であったことが確認された。

図 3.22 はステップが進むごとの精度比較を示している。最初のうちは予測精度はほぼ同等である。しかし、数ステップの後にはモデルによって傾向が異なっている。SMC は実験 2 の屋内のケースと同じように精度が落ちていく。AR-HMM も同様の挙動だが最初の 1 ステップで大きく精度向上しており、全体的な精度は SMC より高い。また、7 ステップ目以降、SMC が精度低下し続けるのに比べると精度を維持し続けている。MMM は 4 ステップ目で他のモデルよりも高い精度を達成するが、その後すぐに精度低下しており、6 ステップ目から AR-HMM よりも精度が低くなってしまい、総合的には AR-HMM より低い精度になった。MAR-HMM は 2 ステップ目まで AR-HMM と類似して精度が上がり、その後 5 ステップ目まで MMM と同様に精度が向上し続け、それ以降は AR-HMM と同様に精度を維持する傾向にある。このように、MAR-HMM は MMM と AR-HMM の良い点が両立できており、総合的にも高い精度が得られていると考えられる。

3.6.5 実験全体の考察

実験 1 は、群集シミュレーションデータに関する予測実験であり、シミュレーションにて生成されたデータに対して歩行者行動予測を試みたところ、次の行動を予測するという処理の精度は、MMM が最も高く、全体で 56.4%、4 つ以上の行動履歴があれば 98.6% と高い確率で予測できていた。このデータは実際の人の動きに比べると非常に単純であり、主に施設の物理的構造の効果が反映されていた。パターン分類の効果が大きいのは経路選択の形がモデル化できていたためと考えられる。

次に、実際の商業施設における歩行者行動履歴を用いて、行動予測の性能を評価した。実験 1 に実際の人の要素が加味され複雑であるが、屋内では部屋からは廊下へしか移動できないなどの制約が強く、移動可能な範囲が限定的である。MMM では 64% の精度であったところが、MAR-HMM では 72% の精度で予測できることがわかった。つまり、パターン分類の効果は大きいですが、時間変化を組み入れるとさらに予測精度が高まっていたとわかる。

最後に、より広域の場合の評価として、人の流れデータを用いた屋外の動線の予測を行った。これは実験 2 を広域、長時間化した結果に相当する。その結果では MMM は 49.2%、AR-HMM は 51.5%、MAR-HMM は 56.8% となり、MAR-HMM が最も高精度であるものの MMM よりも AR-HMM のほうが精度が高くなっていた。屋外は広域であり空間的な制限が減り、MMM のパターン分類の効果よりも時間変化を扱う AR-HMM の精度が高くなったことが想定される。ただし、両方加味する MAR-HMM が最も高精度であり、パターン分類そのものの効果は存在するが、時間が経つにつれてパターン分類が困難になったことを意味する。

広域の人の動きでは、人が目的地とする地点の選択肢が多くパターンは複雑である。本実験の結果では、広域であると時間変化を加味したモデルの精度が高くなっていた。広域さと時間変化の加味の両者の間には、直接的な関連はないように考えられるが、広域であるほどそこにいる人の個性が多様化していくため、移動を時間とともに変化するとしなければ表現しきれないということが想定される。

3.7 関連研究

本研究に類似するものとしては、カーナビゲーションシステムを対象にした運転の目的地を予測する方式 [36][37] が知られている。また、携帯電話で収集した歩行履歴に基づく行動予測の研究も知られている [68]。これらの研究は、対象を特定個人に限定しており、同一人物の動線を数カ月にとまって収集して、その行動からパターンを見つけ出すというものである。それに対し、本研究では特定個人の動線を収集する期間はごく短くてもよい代わりに、多人数の動線を収集するようにした点が異なる。

3.8 おわりに

本章では、移動体の位置を計測して得られる測位データに処理を加えて新たな情報を得る動線解析の一環として、状態遷移の履歴を混合マルコフモデルに当てはめる方式を提案した。特に歩行者の行動をモデル化する方法として、人の行動パターンの分類の機構を組み入れた混合マルコフ連鎖モデル (MMM) と、さらにそれに時間経過に伴う行動変容を組み入れた混合自己回帰隠れマルコフモデル (MAR-HMM) を用いた方式を提案した。

また、3 種類のデータに対して、移動の予測というタスクの性能をもって有効性検証を行った。この実験から、それぞれの効果の度合いに差こそあれ、基本的に MMM および MAR-HMM が高い性能を発揮することが確認できた。これらは、不特定多数の人の行動をモデル化するにあたっての主要候補の一つとして考えることができる。

また、広域の人の動きを表現するには時間変化を考慮に入れる必要がありそうだという結果が得られた。これが人の個性のばらつきに起因するのであれば、例えばある個人の長時間の行動を対象とした実験が有益な洞察を与えると期待される。そこで次章では、特定個人の長時間の行動分析を対象にした分析を行う。

第4章

特定人物の長期間行動分析

4.1 はじめに

本章では、ある特定の人の時系列位置情報が過去にわたって蓄積されているときの、その人物の交通行動を分析する方法について述べる。

不特定多数の人の交通行動は、従来の交通調査のように、短期間に集中的に多くの人の動線を計測することを前提においている。しかし、個人はそれぞれ行動傾向が異なるため、一つのモデルで表現しようとする、不整合が生じる。したがって、不特定多数の人の交通行動を一つのモデルで表現することは、異なる行動傾向を一つにまとめるという計算に相当する。それに対し、特定の人の交通行動を長期間にわたって計測できる場合では、その人の個性を加味したモデル化ができ、より個々人に固有の特性を反映したモデルが構築できるはずである。したがって、個人の行き先予測の精度は不特定多数の計測に比べて高精度になると想定され、このような個人行動を長期間計測する試みが行われている。その種のデータを取りあつかうことにより、よりよい交通行動分析が可能となることが期待される。

そこで本章では、特定人物の位置情報が長期間に渡って計測可能である場合に、交通行動のモデル化を行う方法について述べる。特定の個人行動の分析では、不特定多数の人の分析と異なり、計測対象の個人に関する情報を得ることが可能な場合もある。そのため、先行研究での特定人物の位置情報では、当該個人の行動範囲などが得られることを前提にしていることが少なくない[38]。しかし、匿名化された携帯電話のデータなど、個人が誰かが特定できない長期間のデータを複数人分得られることがある。このデータを分析する場合、モデル化は個人ごとに行うとしても、個々人の行動に関する事前知識を反映することは必ずしもできず、事前知識なく分析をすることも必要である。

特定人物の長期間行動分析では、得られたデータをまず適切な単位に分解することを行わなければならない。また、長期間のうちに何度も同じような交通行動が見られる場合（典型的には、通勤通学が当てはまる）には、同一の交通行動を繰り返しているとしてモデル化しなければならない。パーソントリップ調査のような不特定多数を対象とした調査では、たいてい得られるデータが1日分の交通行動などに限定されており、複数日にまたがるデータを分割する処理は必要とされない。長期間の個人の調査としては、個人ごとの交通需要の時間変動[69]を捉える目的で、交通日誌を用いた調査[70]やパネル調査[71]が行われており、マルコフモデル等での分析がなされているが、標本数や調査の頻

度は不特定多数のものと比べると小さく、データは疎である。

一方、人の持つ移動体端末から得た測位データを用いた調査手法の一つであるプローブパーソン [19] では、交通需要や観光行動 [23][24][25][26] の把握のために、交通行動の単位であるトリップ、すなわち出発地から目的地への移動について、対象者に端末操作による入力を求めるなどしている。また、入力忘れ等に対応するため GPS の記録から移動滞留を判定し、できるだけ正確なトリップを抽出し、従来のパーソントリップ調査に近い精度の情報を収集しようとしている。特定人物の長期間行動分析では、このような入力を長期にわたって要求するのは困難である。よって、従来は精度向上のための要素であった GPS の記録から移動滞留を抽出する方法は、軌跡の分割を行う主な手段になる。ただし、特定人物の長期間行動分析では位置情報の収集期間が長いため、自宅と通勤先の往復などの習慣的な移動が判断できるなど、より精密な判定ができると期待される。このように、特定人物の長期間行動分析のための軌跡の分割方式については仔細な検討が必要である。

そこで、本章では、ある特定人物の位置情報が長期間にわたって得られたとして、それらを事前知識なく自動的に分析し、概要を把握可能な統計量を得ることを目的として、位置情報の履歴を移動の単位に分割し、その遷移を表現する確率モデルを構築する方式について述べる。

4.2 人の行動モデル

4.2.1 交通行動の分析単位

ある特定の個人の位置情報を計測する手段として最も簡便なのが、GPS を用いた計測システムである。近年普及しているスマートフォンをはじめとする携帯電話は、GPS や基地局の情報を用いて、常時、その位置を計測できる。ただし、一般に計測には多大な電力を要するため、サンプリングレートは 5 分に一度程度を上限としていることが多い [72][73]。

GPS によって計測された人の位置情報は、緯度、経度などの連続的な座標値によって示される。そのため、前章までの議論で示したとおり、座標値から人の状態を判別して離散的な状態の遷移への変換した後、マルコフモデル (MMM, MAR-HMM など) を適用することを考える。ただし、特定個人の位置情報を全てつなげると一つの動線になる。前章までで扱っていた混合モデルにもとづくパターンの分類は、多数の動線の中から類似するものをまとめるアプローチであるため、そのままでは適用できず、動線をパターンに分類可能な単位で分割する方法が必要になる。

パーソントリップ調査をはじめとする既存の交通調査は、一日分の交通行動を調査している場合が多く、交通行動の分析の単位として一日を扱うのが一般的である。そこで、本章では一日の交通行動を一つの状態遷移列とみなし、日々がパターンに分類できるとして日々の状態遷移系列を入力としてマルコフモデルの学習を行う。これにより、その人の日々の交通行動のパターンを再現する確率モデルを構築できると考えられる。

4.2.2 移動滞留判別

不特定多数に対する分析の場合、多数の人が滞留する地点を基準として状態を判定した。これと同じ方法を特定個人に適用しようとする、個人が何度も滞留する地点を参照することになる。しか

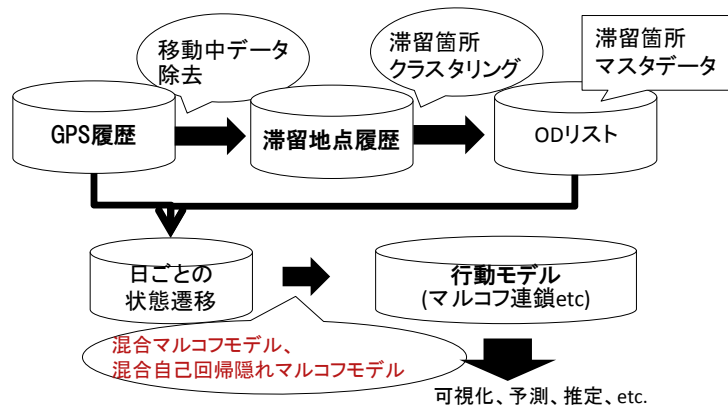


図 4.1 提案方式の模式図

し例えば自宅と通勤先の往復など、一般的に人の行動は日々繰り返しており、偏りが大きい。するとめったに訪れない地点は状態の判定基準には適用されないことになり、自宅と勤務先だけの状態遷移を分析することになってしまい、情報の大半を失ってしまう。そこで、何度も滞留が発生したという基準ではないアプローチが必要である。

データの件数が十分でない場合に、人の位置情報を人の離散的な状態の単位に分類する方法としては、地図情報（自宅の位置、店舗の位置など）を用いて判別する方法（ルールベース）と位置情報の履歴から長期間滞留した実績のある地点を抽出し、その滞留地点に滞留していたことをもって状態と判定する方法（挙動ベース）の二種類が考えられる。

ルールベースの方法では、過去の位置情報が少ない状況でも十分な判別精度で判別ができる。しかし、ルールを仔細に設定していくことは非常に難しく、対象人物の行動についての事前知識が求められる。例えば、大学生の行動进行分类するには、通学先の大学、自宅、アルバイト先などの位置を入力して用いることになるが、それには当該学生の行き先がおおむね把握できていることが必要である。本来、本研究の目的はデータだけがある状況でも概略が把握できるようにすることであるので、この方法は適さない。

挙動ベースの方法では事前知識を導入することなく移動の有無のみをもって判定するため、その滞留にどのような意味があるかは別に考慮する必要がある反面、位置情報を投入するだけである程度状態が抽出できる。本研究では、この挙動ベースの方法を元に、できるだけ滞留地点の意味に配慮した状態遷移列を生成することを目指す。図 4.1 に本研究にて提案する状態遷移列の抽出手順を模式的に示す。まず、GPS の座標値の履歴から滞留地点を抽出する。この滞留地点の抽出により、GPS の履歴は移動の開始点と終了点、そしてそれらの間の移動、の三種類に分類されることになる。次に滞留地点を集めてクラスタリングのアルゴリズムを適用することで、類似する滞留地点を一つの代表値に集約する。以降では、この代表値の一覧を OD リストとよぶ。すなわち、GPS の履歴は、OD リストの地点集合とその間を結ぶ移動という形に変換できることになる。最後に GPS の履歴を OD リストを参照しながら日ごとの状態遷移列に変換し、その状態遷移列を用いてマルコフモデルによる分析を行う。

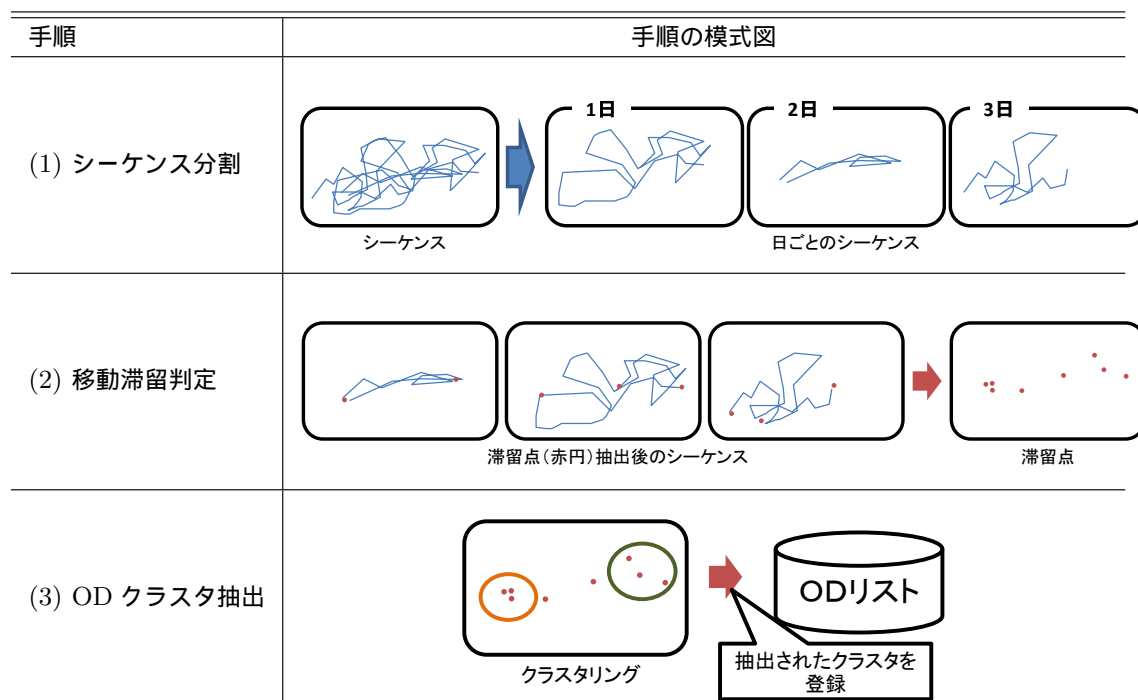


図 4.2 状態遷移生成手順

4.3 状態遷移生成方式

ここまで述べてきたアプローチを元にした状態判定方式を図 4.2 に示す。

(1) シーケンス分割は、GPS の軌跡データを連続する移動の単位に分割する処理である。軌跡における連続する移動のかたまりを以降ではシーケンスと呼ぶ。1 時間以上データ間隔があった場合、そこでシーケンスを分割する。また、シーケンスの端点は交通行動の OD になっている可能性があるとして OD リストに登録する。深夜に日付をまたがって行動するようなこともあるが、今回は「深夜に帰宅しなかった日」のようにそれを前提においたパターンでのモデル化を想定する。

(2) 移動滞留判定は、各シーケンスに対してそれが移動中かどうかを判定してラベル付けする処理である。各トリップ内の点について滞留点か否かを判定し、連続する滞留点の塊にたいして平均位置を算出して、OD リストに登録する。本研究で用いた挙動ベースの移動滞留判定を図 4.3 に各トリップ内の点について、前後の平均速度および移動距離から滞留点か否かを判定する。具体的には、一つの点について、その一定時間幅（滞留の基準となる時間幅、例えば 30 分）内の他の点を集め、平均速度が 5m/s、移動距離 20m 以下に収まっている場合、それを滞留とみなした。仮に 30 分程度滞留が続くものと仮定すると、少なくとも一点、そのような条件を満たす点があるはずである。

最後に、(3)OD クラスタ抽出では、移動体流判定で滞留と判定された地点のリストをクラスタリングして、同一の地点（地域）への移動をグループ化する。このグループ化は、以下の手順で行う。まず、DBSCAN（距離が閾値以下の点を併合していくクラスタリングアルゴリズム）によりクラスタを生成する。次に、クラスタ構成点数が少ない（2 点以下）の場合、例外的措置として除去する。これに

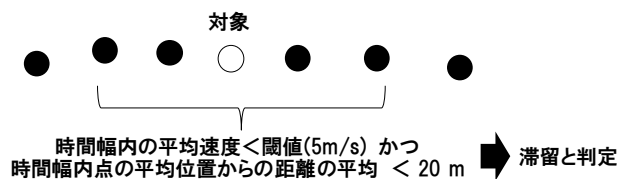
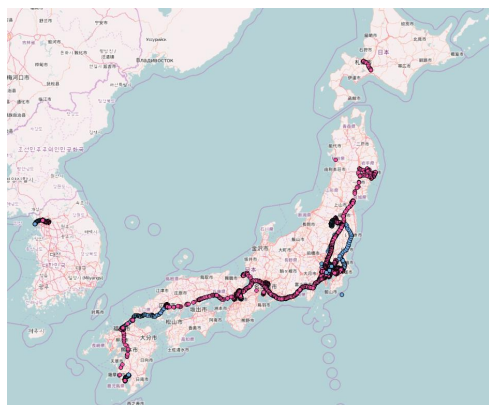
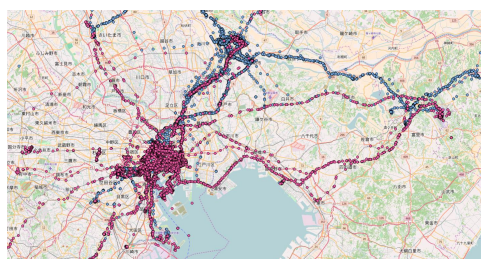


図 4.3 移動滞留判定



(a) 広域図



(b) 東京付近の拡大図

背景地図: (c) Open street map

図 4.4 評価に用いたデータ

より、はずれ値を排除しつつ一つの塊となっている OD を抽出できる。ここでは閾値として 50m を用いた。

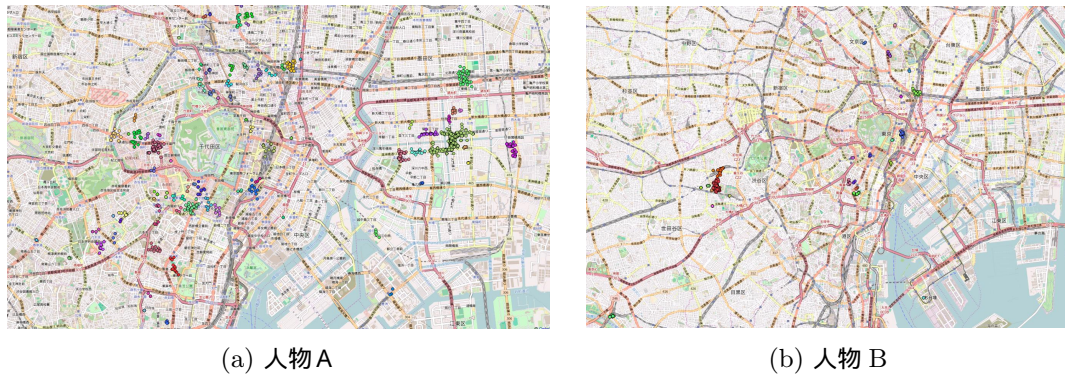
以上の方法によって、GPS のような緯度経度の座標値のデータを地点間の遷移に変換するためのテンプレートを構築することができる。あとは、構築されたテンプレートに位置情報を当てはめ、OD の地点間の遷移に帰着させることで、従来のマルコフ連鎖等のモデル化を適用できるようになる。

4.4 実験

4.4.1 実験条件

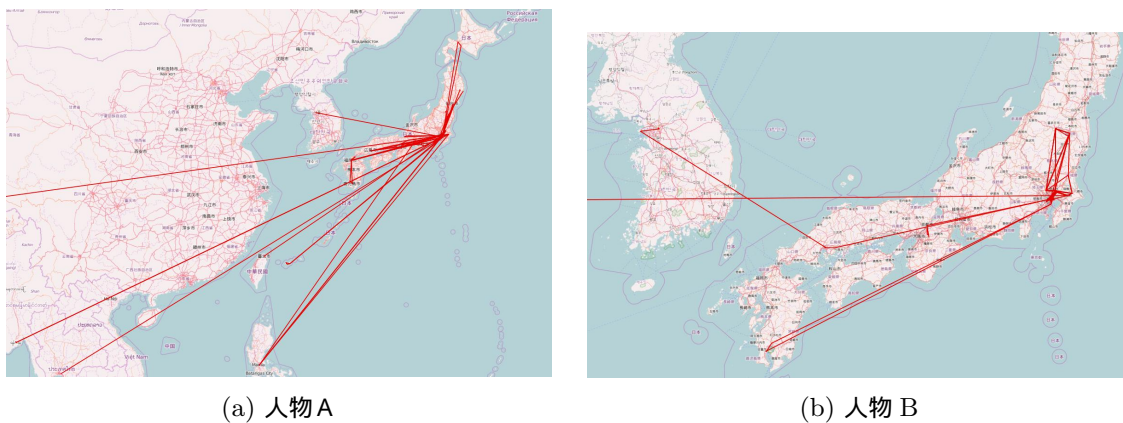
前節にて検討された方式を検証するため、実際に人の位置情報データを用いて評価を行った。図 4.4 に使用したデータを示す。このデータはある 2 人の人物の一年分の位置情報の履歴であり、人物ごとに色分けされている（以降赤を人物 A、青を人物 B とする）。この位置情報は携帯電話の GPS 機能により収集されたものである。常時計測を行うと記録容量や消費電力の面で負担が大きいため、この計測データは 5 分に一度程度の粗いサンプリングで、さらに大きな移動等があった際に限り位置情報を取得するよう、データの収集頻度を少なくする方式で収集されている。

本研究では、このデータに対し、提案方式の状態遷移抽出を行い、それにもとづく分析の妥当性について前章同様の行き先予測等を用いて評価した。



背景地図: (c) Open street map

図 4.5 滞留エリアデータ



背景地図: (c) Open street map

図 4.6 状態遷移列

4.4.2 実験結果

図 4.5 に提案方式において抽出された滞留地点を示す。図中の色の塗られた円が滞留地点であり、同じクラスタに分類されたものは同じ色を用いて描画されている。図 4.6 にこのクラスタによる OD リストを用いて得られた状態遷移列を図示する。両者とも日本の関東地方を中心に移動しているがときおり海外へ移動している。このような一度しかない海外への移動も適切に OD の一つとして抽出できていることがわかる。なお、人物 A については 4569 サンプルの状態遷移データが得られた。

図 4.7 にモデルごとの行き先予測精度の比較を示す。この比較では、MMM の混合数は 100、AR-HMM の内部状態数も 100、MAR-HMM は混合数 4、内部状態数 25 とし、内部状態と混合数の積が 100 になるようにした。なお、MAR-HMM については最も精度が高くなるように選んだ。結果としては、MAR-HMM が最も高い精度で予測できており、77.2% となった。その他については、SMC で 45.0%、MMM で 68.3%、AR-HMM で 56.2% となっている。この結果は不特定多数の場合の性能とほぼ同じ程度であり、不特定多数を対象にしたモデルが個人の長期間分析でも同様に適用

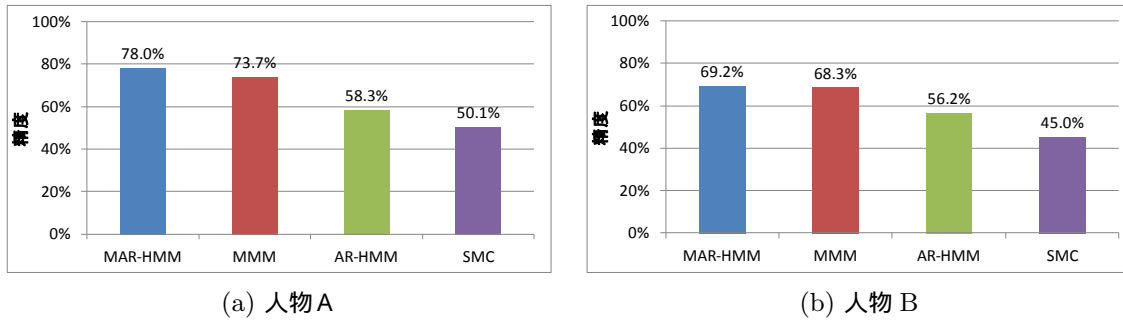


図 4.7 行き先の予測精度

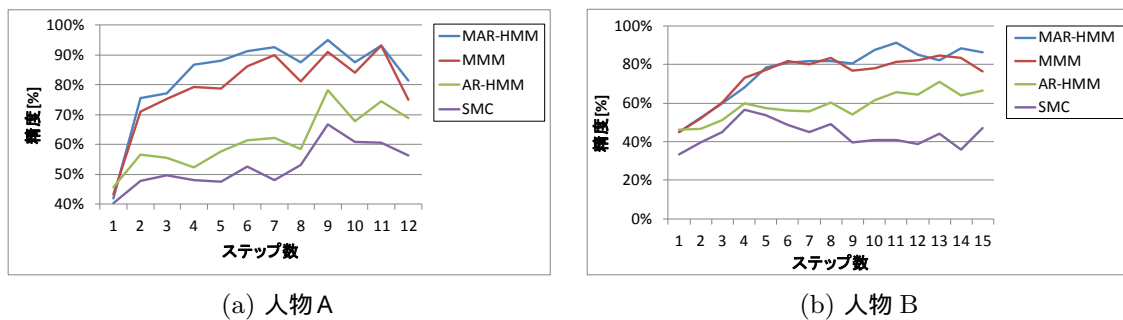


図 4.8 移動ステップごとの行き先の予測精度

できることがわかる。

図 4.7 にステップごとの予測精度を示す。不特定多数の人の位置情報の場合と同様に、後半ほど精度が上がり、 k が 50 を越えたあたりでは 80% もの精度で予測できることがわかる。この傾向は不特定多数の場合と類似している。また、精度が不特定多数の場合よりも高くなっている。特定個人の行動をその個人の過去の行動から予測することは不特定多数の場合の、他人の行動履歴の集合から予測する場合に比べると、同一のものに対するモデル化がより妥当である点が反映していると考えられる。

以上の考察から、提案方式による状態遷移抽出によって、不特定多数の場合と同様の分析が可能であり、それは時として不特定多数の場合よりも正確である、と考えることができる。

4.5 関連研究

本研究に類似する研究としては、山田らによる携帯電話で収集した歩行履歴に基づく行動予測の研究が知られている [68]。これらの研究は、対象を特定個人に限定しており、同一人物の動線を数カ月 にわたって収集して、その行動からパターンを見つけ出すというものである。

また、田中らはカーナビゲーションシステムを対象に、運転の目的地を予測する方式について発表した [36][37]。カーナビゲーションシステムの場合、一般的に自動車が道路を走行することから、走行中の道路の ID と方面をもって状態とみなすことができるため、本研究で行ったような滞留点の抽出は行わなくともマルコフモデル等は適用できる。ただし、道路の ID を持って状態とした場合、走行道路の選択をもとに次の走行道路を予測することになるため、出発地、目的地等の分析による行動

パターンを把握するには必ずしも適さず、むしろ道路の選択傾向に関する分析（たとえば、主要道を好み細街路を嫌う、混雑を嫌う、など）に適する。

4.6 おわりに

本章では、ある特定の人の位置情報を長きに渡って計測して蓄積しておき、それを分析する方法について、不特定多数に対して用いられるマルコフモデルを援用するために、位置情報に移動滞留判定を加えて出発地目的地のクラスタリングを行うことで、GPS で得られた動線を状態遷移列に変換する方法を提案した。

また、実験的に 2 人の人物の一年間の位置情報に提案方式を適用して分析したところ、不特定多数の場合と同様の分析が可能であることが確認できた（なお、6 章ではそのような分析の一例として主動線抽出について説明する）。したがって、特定人物であっても長期間のデータを収集して本章で提案した方式を用いて分析すれば、不特定多数の場合と同様のさまざまな分析ができると考えられる。

第 II 部のまとめ

第 II 部では、位置情報のみから得られる一人一人の交通行動を確率モデル化するため、人の行動パターンの分類の機構を組み入れた混合マルコフ連鎖モデル (MMM) と、さらにそれに時間経過に伴う行動変容を組み入れた混合自己回帰隠れマルコフモデル (MAR-HMM) を用いた方式を提案した。

また、不特定多数 3 種類、特定個人 1 種類のデータに対して、SCM、AR-HMM、MMM、MAR-HMM の 4 種のモデルを用いて移動の予測というタスクの性能を評価した。その結果、いずれの場合においても、MAR-HMM を用いたときに最も精度の高い結果が得られた。これらの実験を、データ（人の個性、空間の広さ）と精度（パターン分類と時間変化の効果）に関してプロットした関係図を図 4.9 に示す。実験 1 のデータは、屋内の群集シミュレーションデータであり、そこに移動する人の個性はないに等しい。このとき精度の高かった MMM は個性を除く要因、つまり経路選択の効果をうまく加味できていると理解できる。実験 2 は屋内の人の行動の実測データを用いており、扱っているのは種々の個性をもった人の移動データであるが、施設内の移動を対象としており移動可能な範囲が限定的である。この場合に時間変化を加味した MAR-HMM の予測精度が高いということには、そこに人の個性が反映されているためと考えられる。実験 3 は屋外の推定動線で、実験 2 を広域、長時間化した結果に相当する。この実験では時間変化の効果だけを取り込んだ AR-HMM がパターン分類の効果を取り込んだ MMM の精度を上回っており、広域であるほどそこにいる人の個性が多様化していくため、移動を時間とともに変化するとしなければ表現しきれなかったものと想定される。個性の極端に少ない例として扱った個人の長時間行動では、やはり MMM の精度がかなり高くなった。このことから、より個性に富んだ人の動きを扱うには、時間変化を取り入れる必要があると考えられる。

行動のパターン分類を取り入れた MMM であるが、個性が豊富すぎると時間変化を取り入れることのほうが精度に寄与した。この要因の一つとしては、人の行動範囲が広すぎるとパターンが多くなりすぎて、MMM では不十分なことが考えられる。一方、時間変化を扱うモデルの精度が高いということは、時間が経つにつれて個性が顕在化するものと解釈できる。つまり、瞬間瞬間での人の移動の選択のパターンはさほど多くないが、人の移動はその積み重ねであり、膨大な選択の積み重ねによって個性が現れていくという理解である。この理解は、一人ずつの移動を予測し続けることは、その当人の移動に関する情報なしには難しいことを示唆する。ある人の移動が他の人の平均的な移動と似ているならば予測可能であるが、実際にはどこかに平均との差異があり、その差異の現れ方が人によって異なるため、すべての予測は困難である。

上述の通り、行き先の予測についての精度が評価され、移動のパターン分類がおおむね人の移動に

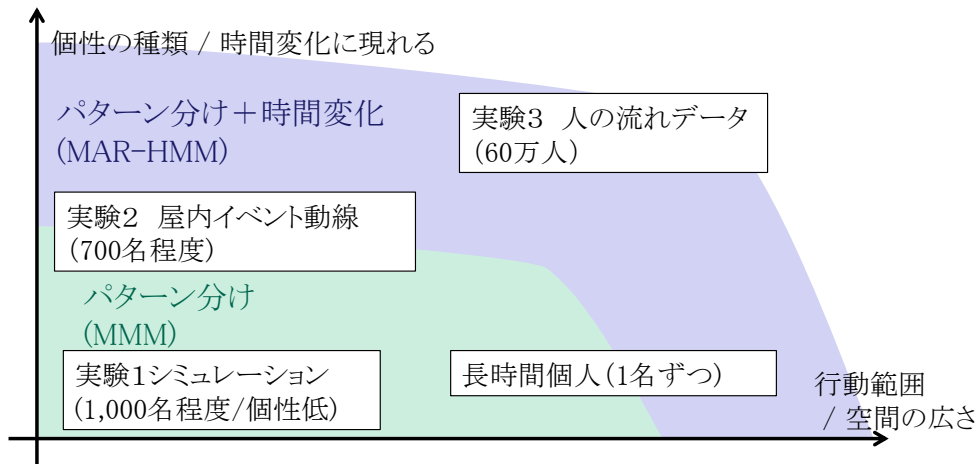


図 4.9 実験の位置づけ

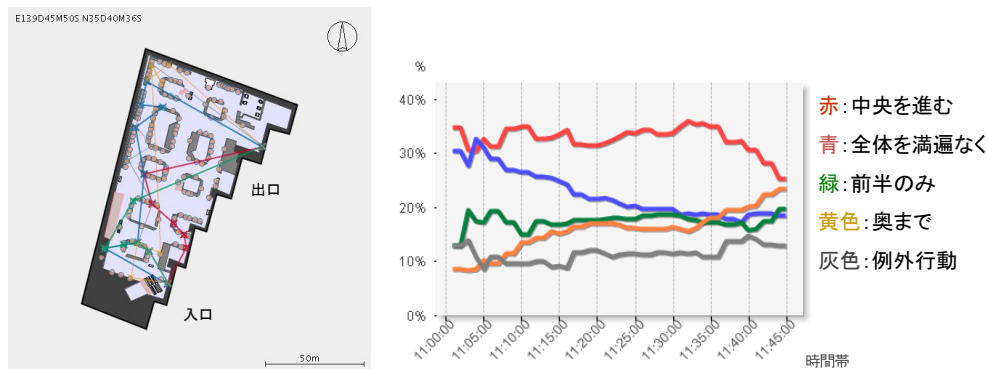


図 4.10 3章の実験1のパターン抽出

みられる個性を表現しているようであると考えられる。それを可視化することで、本研究の目的である現状把握に繋がることが期待できる。図 4.10 および図 4.11 は、3章の実験における不特定多数の人の移動について、パターン分類の可視化を行った例である。図 4.10 の左の図は実験1のデータに対して混合数を5としてMMMの計算を行い、各モデルについて入り口から出口まで地点間の遷移確率が最も高いものをつないで得られた動線を表示したものである。右のグラフは、各時刻に出現した人が、それぞれどのパターンに分類される動きをしたかを求め、その人数比をグラフとして表現したものである。図 4.11 は、実験2のデータについて混合数2としてMMMでモデル化し、その遷移確率を線として表現したものである。左の図（パターンA）は全体の85%ほどの人が所属するパターンであり、単調に周囲を回る動きをしている。右の図（パターンB）は全体の15%ほどの人が所属するパターンで、動きが複雑で一定しない。このように、人の移動を混合モデルによってパターン分類し、その特徴を可視化したり、所属する人数比を見ることで、単に動線を見ることよりも多くの情報を把握できるようになる。ここから、Who, Why, Howに関連する事実を推察できる可能性が見出せたと考えられる。

第II部の結果によると、MAR-HMMの予測精度が示すように、人の移動は平均的には高い精度で予測ができることがわかった。すると、人の集団の傾向、つまり平均的な動きは、比較的単純なマル

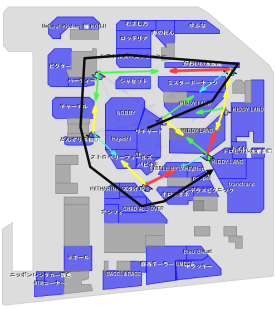
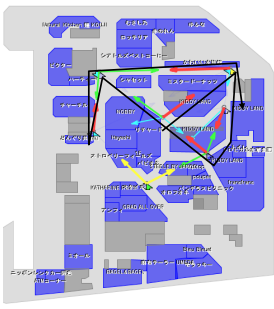
	パターン A	パターン B
		
人数割合	83.9%	16.1%
動き	外周部を一周	方向性の一定しない動き

図 4.11 3章の実験 2 のパターン抽出

コフ連鎖のモデルで予測できる可能性がある。第 III 部では、人の集団としての動きをモデル化し、可視化することについて述べる。

第 III 部

集団統計量の遷移モデル

第 5 章

人の密度分布の高速推定

前述の通り、測位データは、人や物の動きの現状把握に有用である。ただし、単に位置情報を可視化しただけでは情報が多すぎて、例えばある地点間の移動が特に多くなっている、などを短時間で理解するのは困難である。そのため、多くの人の測位データに対し、その概要（移動そのものの情報）をわかりやすく表示することが求められる。移動そのものの概略として基本的なのは、どのくらいの人が、どう動いているかという情報と考えられる。そこで本章では、まず人の密度の空間分布を算出する方法について述べる。

5.1 本研究の位置づけと目的

近年、測位データを用いて人の交通行動を把握する取り組みの一つとして人の持つ移動体端末から測位データを収集し有益な情報を得るプローブパーソン [19] などの試みがある。その一形態として、多大な費用をかけたアンケート調査の代わりに一般市民の協力者が所有するスマートフォンに位置情報を集めるソフトウェアを配信する形態をとることが考えられる。これにより低コスト高精度な交通需要や観光行動 [23][24][25][26] などの人流把握ができると考えられる。スマートフォンの通信負荷や電力消費の大きさや、プライベートな情報を提供させるインセンティブが必要であったり、個人の位置情報を収集するが故のセキュリティなどの課題はあるが、それらを解決できれば有用な情報が集まるはずである。

収集された測位データは、ただそれだけでは多すぎて、その特徴や傾向を把握して意思決定に活用するのは困難である。そこで、人の位置について、概要を把握できるように可視化する方法が必要である。図 5.1 に例を示す。測位データは (a) のように点によって地図上に表示することができる。この表示では移動体の個別の動きが把握できる一方、高密度な地域では点が重なり合い、その数までは把握できない。プローブパーソンにおいては個人の動きは重要ではなく、その統計量、たとえば人口密度の把握が重要であるが、点をそのまま表示しても統計的な把握はできないのである。そこで、点の密度にもとづく分布を表示することが考えられる。(b) のその例を示す。この表示は人の密度に応じて色の異なる矩形を描画したもので、緑 黄 赤の順に密度が多いことを示す。このような表示は一般的にヒートマップ表示などと呼ばれている。本表示の目的は人流の概況を可視化し、把握可能にすることである。これにより、例えば過密地域を把握して保安・防災上の計画立案に役立てたり、人

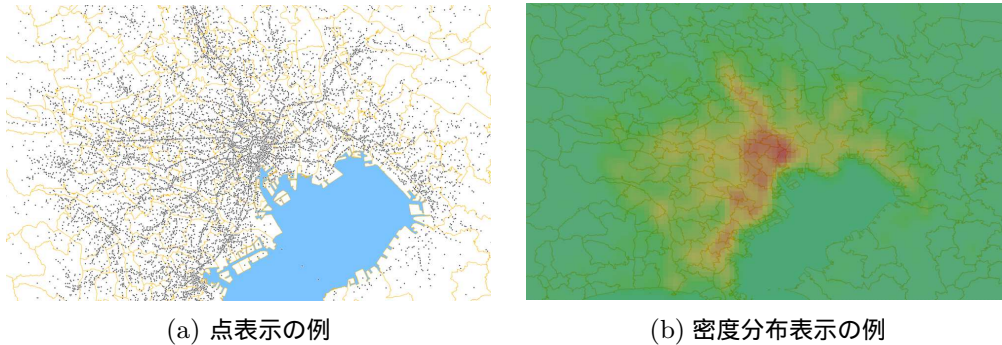


図 5.1 測位データの表示方法

の流動を把握して交通需要の確認やマーケティングに使用することが考えられる。また、点表示では個人の位置情報が露わになってしまうのに対し、分布表示では個人の位置は伏せたまま傾向だけを提示できるという点で、プライバシーの問題も少なくなるという利点もある。

スマートフォンなどからリアルタイムに測位データを収集し、それを逐次的にヒートマップ表示すれば、リアルタイムな概況把握ができると考えられる。しかしそれには、ヒートマップ表示の処理にかかる時間が十分高速でなければならない。すなわち、あるときに取得した測位データからヒートマップ表示をする処理は、その次の測位データを取得するまでに終了させなければならない。例えば、数百万人規模の人口をもつ都市と周辺地域の一千万人近くの人々について、その 5 % の位置情報を 1 分に一度収集する想定であれば、五十万件の測位データからヒートマップ表示をする処理を 1 分で終えねばならない。つまり、1 件あたり $120\mu\text{s}$ という一般的なコンピュータでは達成困難な処理速度を求められる。

そこで、本研究ではすべての測位データを使用せずに、密度分布を推定することを目指す。今回、移動体の群の分布をガウス分布にてモデル化し、当該モデルを少数の測位データで補正することにより、少ない測位データから全体を推定する方式を提案する。今回、着目したのは移動体の位置の分布は急激には変化しないという点である。すなわち、時系列測位データの密度を逐次的に推定していく場合、前時刻の推定結果は次の時刻の推定結果と類似すると想定できる。そこで、現在の時刻の分布は、直前の推定結果である密度分布をわずかに変更させたものであると考え、直前の密度分布推定結果を補正して現在の時刻の密度分布を推定するアプローチをとる。これにより、現在のすべての測位データを使用しなくとも精度よく分布が推定できるのである。最後に、アンケートから推定された人の流れデータを用いて提案方式を評価し、その有効性を示す。

5.2 従来方式の問題と提案方式のアプローチ

5.2.1 従来の密度分布推定法

空間的に分布する大規模データに対する統計解析手法は様々な目的のものが盛んに研究されている。例えば、空間上に分布する点で気温などの値が得られたとき、その値の空間分布を推定する方法が研究されている [74]。一方、人の動きに着目した研究としては、人の移動パターンを見つける分析

[46][75] や、移動パターンの確率モデルを用いた行き先予測 [68][36][63][76]、移動パターンからの人物の属性推定 [39][48] などが知られている。

本研究は点の密度分布の分析手法 [77] の研究である。密度分布の推定手法としては、大きく分けてパラメトリックな推定法とノンパラメトリックな推定法の 2 種類が知られている。パラメトリックな推定法とは、予め定められた確率密度分布関数などのパラメータを決定することによる確率密度分布推定方式である。一方、ノンパラメトリックな推定法では、測定されたデータ集合の数から確率密度分布を推定するなどデータに基づいて分布を推定する。

ノンパラメトリックな推定法のうち最も単純な方法の一つは、空間領域を矩形に分割し各矩形内の点の数を密度とするヒストグラムによる推定法である。ただし、矩形内に十分多数のサンプルデータがなければ推定結果の精度が低くなる、領域の境界で密度が不連続になるなどの問題があり、一般的にはカーネル密度推定法 [41, 78, 79] がよく用いられる。カーネル密度分布推定ではサンプルデータの周囲に分布を発生させ、それを重ね合わせた結果を密度分布とする。 N 個の各点 $\{\mathbf{x}_n\}$ を中心とするガウス分布を仮定したガウシアンカーネルを用いた確率分布 $P(\mathbf{x})$ は

$$P(\mathbf{x}) = \frac{1}{N} \sum_n \frac{1}{2\pi h^2} \exp\left(-\frac{|\mathbf{x} - \mathbf{x}_n|^2}{2h^2}\right), \quad (5.1)$$

となる。ここで、 h は分布の半径に相当する量である。2 次元空間上の分布をより高速に生成できる方法として以下の Quadratic カーネルも用いられることがある。

$$P(\mathbf{x}) = \frac{1}{Nh^2} \sum_n Q\left(\frac{|\mathbf{x} - \mathbf{x}_n|}{h}\right) \quad (5.2)$$

$$Q(r) \equiv \begin{cases} \frac{6}{\pi} (1-r)^2 & r < 1 \\ 0 & r \geq 1 \end{cases} \quad (5.3)$$

カーネル密度分布推定法には、密度分布の値を得るためにサンプル数に比例する処理時間がかかるという問題がある。仮に 640×480 画素の画像を 10 万件のデータから生成する場合、 $640 \times 480 \times 10$ 万回つまり 30 億 7200 万回もの計算が必要となり、処理速度が低下する要因となる。

パラメトリックな推定法としては、パラメータをもつ確率分布を用い、尤度が最大になるパラメータを決定する最尤法が知られている。最尤法でよく用いられるのは、観測不可能な内部確率変数を仮定する確率モデルを適用して精度を高める方法である。その場合、パラメータと同時に内部確率変数も推定が必要なため、内部確率変数の期待値推定とパラメータの最適化を交互に実行し収束させる EM アルゴリズム [59] が用いられる。内部確率変数をもつ確率モデルとして最も代表的な分布は混合ガウス分布である。混合ガウス分布は複数のガウス分布を組合せた分布であり、各確率変数が生成元のガウス分布を特定する内部変数を持つ。パラメトリックな推定法では、使用した確率分布が推定結果の制約となるため、適切かつ表現力のあるモデルを選択しなければ精度が低下してしまう。表現力の高いモデルは一般に複雑であるが、EM アルゴリズムを適用すると反復演算に処理時間がかかるという問題がある。

パラメトリック推定とノンパラメトリック推定の特徴を示すため、分布画像の生成手順の模式図を図 5.2 に示す。最初に移動体端末が自身の位置を測位データの形にして、分布推定を行う分析サーバ

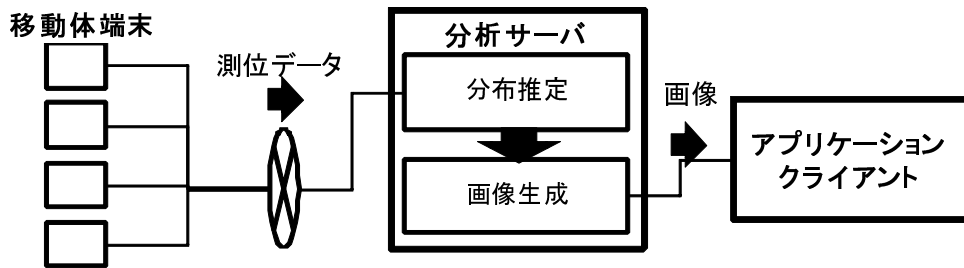


図 5.2 分布画像生成の手順

へ送付する。分析サーバでは、密度分布を推定した後、密度分布を画像を生成するという2つの手順を実行する。アプリケーションクライアントは生成された画像を受け取り表示などをする。アプリケーションクライアントで画面解像度や地図の縮尺変更が起こると、画像生成だけを再度実行すればよい。

ノンパラメトリック推定においては、分布生成は全データを読み出すだけで実質的には何も処理を行わないが、画像生成は全ての測位データを用いて画素値を決定する処理になる。この処理は測位データの量が多いほど多くの時間を要する。パラメトリック推定では、分布推定処理に処理時間を要するかわりに、画像生成に要する時間は非常に小さくなる。図 5.2 のシステム構成を想定するとアプリケーションクライアント側で地図画面のズームなどを行うたびに画像生成だけを何度も呼び出すことになるが、その場合はパラメトリック推定のほうが高速に処理が実行できる。また、分析サーバの負荷を分散するために画像生成をアプリケーションクライアント側で実行することもできるが、ノンパラメトリック推定では全ての測位データをアプリケーションクライアントへ送付する必要があり、膨大な通信量を要する。これは密度分布画像を生成する目的での比較であるが、別の情報、例えば竜巻等の経路情報と重ねて被害者数の期待値を計算したい場合のように人口分布を分布としてそのまま用いる際にも、画像生成処理に相当する演算を繰り返し行うことになる。このように、パラメトリック推定はノンパラメトリック推定にはない有用な特徴を多数兼ね備えている。しかし、パラメトリック推定では分布推定に反復演算が含まれており、全体としての計算時間はノンパラメトリック推定よりもかかる場合が多い。ただし、入力となる測位データが多くなるとノンパラメトリック推定でも処理時間は増大する。

5.2.2 本研究のアプローチ

本研究の提案手法は、人の移動速度の上限を前提にして分布更新頻度を十分短くすることで、前時刻の分布をわずかに補正して新たなデータの分布を推定する方法である。これによりパラメトリック推定に要する計算量を削減し高速な分布推定ができる。図 5.3 にその模式図を示す。パラメトリック推定によりある初期の時刻 $t-1$ の分布を推定した結果を用意しそのパラメータを補正して次の時刻 t の測位データと整合する分布に更新する。この更新の処理では、前時刻の分布情報を事前分布とみなした、新たなデータを用いた事後確率分布推定を行うベイズ推定を用いる。これにより、一部の情報だけを用いて全体の分布をリアルタイムに得られる。

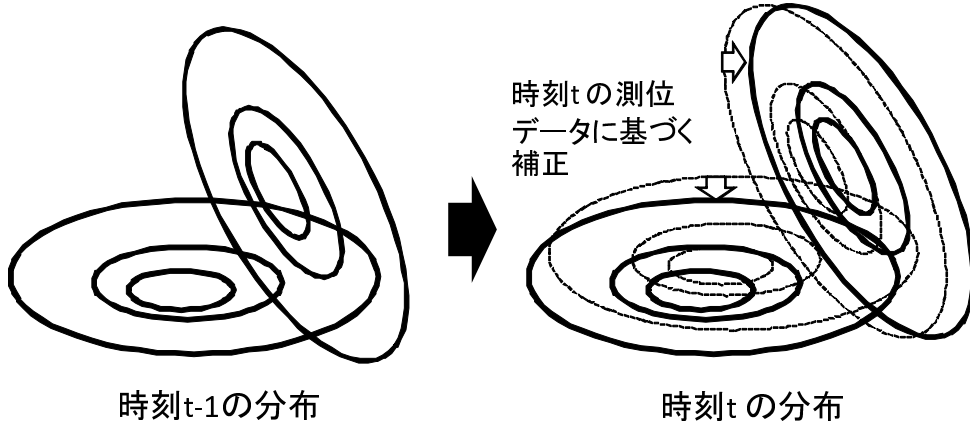


図 5.3 分布補正の概念図

5.3 部分データによる分布生成方式

5.3.1 変分ベイズ法

変分ベイズ法 [80] は確率分布のパラメータの確率分布を事前分布として導入し、パラメータの確率分布に独立性に関する近似を取り入れることにより、真の分布に十分近い分布を求める方法である。本研究では測位データの発生分布を一般的によく用いられる混合ガウス分布を用いて近似する。混合ガウス分布の分布関数を以下に示す。

$$P(\mathbf{x}|\{\pi_k\}, \{\mu_k\}, \{\Lambda_k\}) = \sum_k \pi_k \mathcal{N}(\mathbf{x}|\Lambda_k, \mu_k) \quad (5.4)$$

ここで、 \mathbf{x} が観測データ (つまり、位置座標) であり、 $\mathcal{N}(\mathbf{x}|\Lambda, \mu)$ は分散共分散行列 Λ 、期待値 μ の 2 次元ガウス分布、 π_k は k 番目の要素の混合比である。この分布のパラメータ $\{\pi_k\}, \{\mu_k\}, \{\Lambda_k\}$ について以下の分布を事前分布として与える。

$$\begin{aligned} &P(\{\pi_k\}, \{\mu_k\}, \{\Lambda_k\} | W_0, \nu_0, \beta_0) \\ &= \text{Dir}(\{\pi_k\}|\alpha_0) \prod_k \mathcal{N}(\mu_k | (\beta_0 \Lambda_k)^{-1}, \mathbf{m}_{0,k}) \mathcal{W}(\Lambda_k | W_0, \nu_0) \end{aligned} \quad (5.5)$$

ここで、 $\alpha_0, W_0, \mathbf{m}_0, \nu_0, \beta_0$ は事前分布のパラメータであり、ハイパーパラメータと呼ばれるアプリアリに与えられるパラメータである。また、 $\mathcal{W}(W, \nu)$ は分散 W 、自由度 ν のウィシャート分布、 $\text{Dir}(\{\pi\}|\alpha_0)$ は α_0 デリクレ分布である。ガウス-ウィシャート分布はガウス分布の、デリクレ分布は多項分布の共役事前分布である。共役事前分布とは観測値にもとづきベイズ推定を行った結果が、再び事前分布と同じ形になる分布であり、結果的に推定結果はハイパーパラメータの更新として表現される。なお、ガウス分布の中心 (期待値) の事前分布であるガウス分布は、 \mathbf{m}_0 が期待値、 $\beta_0 \Lambda_k$ (元のガウス分布の分散の β_0 倍) が分散である。なお、これらの表記は [81] に記載のものと同

一のものを用いた。ハイパーパラメータ $\alpha_0, \beta_0, \mathbf{m}_0, W_0$ に対し、 $\{\mathbf{x}_n\}$ を推定に用いることで、

$$\alpha_k = \alpha_0 + N_k \quad (5.6)$$

$$\beta_k = \beta_0 + N_k \quad (5.7)$$

$$\mathbf{m}_k = \frac{1}{\beta_k} (\beta_0 \mathbf{m}_0 + N_k \bar{\mathbf{x}}_k) \quad (5.8)$$

$$W_k^{-1} = W_0^{-1} + S_k N_k + \frac{N_k \beta_0}{N_k + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_k)(\mathbf{m}_0 - \bar{\mathbf{x}}_k)^T \quad (5.9)$$

$$\nu_k = \nu_0 + N_k \quad (5.10)$$

のように更新される。ここで N_k 、 \mathbf{x}_k 、 S_k はそれぞれ

$$N_k = \sum_n r_{n,k}, \quad (5.11)$$

$$\bar{\mathbf{x}}_k = \frac{1}{N_k} \sum_n r_{n,k} \mathbf{x}_n, \quad (5.12)$$

$$S_k = \sum_n r_{n,k} (\mathbf{x}_n - \bar{\mathbf{x}}_k)(\mathbf{x}_n - \bar{\mathbf{x}}_k)^T \quad (5.13)$$

である。この式中の $r_{n,k}$ は負担率と呼ばれる量であり、

$$r_{n,k} = \frac{\rho_{n,k}}{\sum_n \rho_{n,k}}, \quad (5.14)$$

$$\rho_{n,k} \equiv \tilde{\pi}_k |\tilde{\Lambda}_k|^{\frac{1}{2}} \exp \left[-\frac{1}{\beta_k} - \frac{1}{2} (\mathbf{x}_n - \mathbf{m}_k)^T W_k (\mathbf{x}_n - \mathbf{m}_k) \right] \quad (5.15)$$

$$\ln \tilde{\pi}_k \equiv \ln \psi(\alpha_k) - \ln \psi\left(\sum_k \alpha_k\right) \quad (5.16)$$

$$\ln |\tilde{\Lambda}_k| \equiv \ln \psi\left(\frac{\nu_k}{2}\right) + \ln \psi\left(\frac{\nu_k - 1}{2}\right) + 2 \ln 2 + \ln |W_k|. \quad (5.17)$$

と計算される。ここで $\psi(\cdot)$ はダイガンマ関数である。負担率は各サンプルからの各混合ガウス分布に対する所属の度合に相当し、あるサンプルが特定のガウス分布に完全に所属するとしたとき負担率は 1 となる。よって、負担率の和 N_k は、 k 番目のガウス分布に所属するサンプルの数を一般化したものに相当する。この負担率の計算にハイパーパラメータが含まれており、ハイパーパラメータの更新式に負担率が必要であるが、この計算は繰返し処理によって収束させることができる。したがって、ハイパーパラメータに適当な初期値を与えて負担率を計算し、次に負担率にもとづきパラメータを更新、再び新たなハイパーパラメータから負担率を計算し、と繰返すことにより、最適なハイパーパラメータが得られる。このハイパーパラメータを用いた予測分布は以下ようになる。

$$P(\mathbf{x}|W_k) = \frac{1}{\sum_k \alpha_k} \sum_k \alpha_k \text{st}(\mathbf{x}|\mathbf{m}_k, L_k, \nu_k - 1), \quad (5.18)$$

$$L_k = \frac{\beta_k(\nu_k - 1)}{1 + \beta_k} W_k \quad (5.19)$$

ここで、 $\text{st}(\cdot)$ は t 分布である。以上の計算が一般的な変分ベイズ法による混合ガウス分布推定の手続きである。

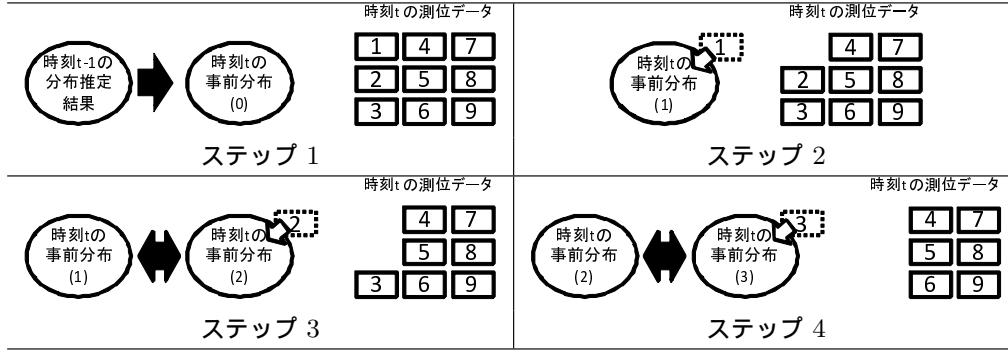


図 5.4 逐次投入の概念図

5.3.2 カルバックライブラー情報量による計算打ち切り

上記手続きにおいてアプリオリに与えられたハイパーパラメータ $\alpha_0, \beta_0, \mathbf{m}_0, W_0$ を前時刻の分布推定結果にもとづき合成する。結果的に、現時刻の測位データを一部だけ用いたとしても、全測位データを用いて推定したのに近い分布が得られる。

ここで問題となるのは、現時刻のデータのうち、使用するデータ数をどのようにして決定するかである。今回、人流分布の変化が激しくないという前提のもと、人流分布の変化が激しい時ほど多くの測位データを用い、変化が緩やかな場合には少数の測位データを用いるようにする。図 5.4 にこの処理の概念図を示す。まず、直前の時刻の分布をもとに事前分布を構成する (図中、ステップ 1)。同時に、入力サンプルをいくつかのブロックに分割する。次に、ステップ 2 のようにそれらのブロックの内一つを事前分布に投入し事後推定する。さらに、ステップ 3、4 に示すように直前のステップの結果に対して新たな測位データを投入し事後推定する。この際、直前のステップの結果と比較して、分布の変化があまり大きくなかった場合には、測位データが十分な量投入されたと判断して、計算を打ち切る。これにより、測位データ数を削減できる。本研究では、事前分布に新たな測位データを追加した場合の確率分布の変化を算出し、変化が十分小さくなったら、それ以上の測位データ追加を打ち切る方式をとる。この際、確率分布の変化を評価する量として、以下の $KL(p(x)||q(x))$ によって定義されるカルバックライブラー情報量 (以降、KL 情報量) を用いる。

$$KL(p(x)||q(x)) = \int p(x) \ln \frac{q(x)}{p(x)} dx \quad (5.20)$$

KL 情報量は複数の確率分布間の距離に相当する量である。今回のケースに KL 情報量を適用すると

$$\begin{aligned} K &\equiv KL(p(\{\theta_i\}|x_{m+1})||p(\{\theta_i\}|x_m)) \\ &= \int p(\{\theta_i\}|\{\mathbf{x}_n\}_m) \ln \frac{p(\{\theta_i\}|\{\mathbf{x}_n\}_{m+1})}{p(\{\theta_i\}|\{\mathbf{x}_n\}_m)} \prod_i d\theta_i \end{aligned} \quad (5.21)$$

ここで $\{\theta_i\} = \{\{\pi_k\}, \{\Lambda_k\}, \{\mu_k\}\}$ とし、 $\{\mathbf{x}_n\}_m$ は m 番目に投入される測位データのブロックを意味する。この K が十分小さいとき、確率分布が収束したと見なして計算を打ち切ればよい。式中の

$p(\{\theta_i\}|\{\mathbf{x}_n\}_m)$ は変分ベイズ法の推定結果 $q(\{\theta_i\}|\{\mathbf{x}_n\}_m) = q(\{\pi_k, \Lambda_k, \mu_k\}|\{\alpha_k, \beta_k, \mathbf{m}_k, \nu_k, W_k\})$ に置き換えられるので、

$$K = \int q(\{\theta_i\}|\{\alpha_k, \beta_k, \mathbf{m}_k, \nu_k, W_k\}) \ln \frac{q(\{\theta_i\}|\{\alpha'_k, \beta'_k, \mathbf{m}'_k, \nu_k, W'_k\})}{q(\{\theta_i\}|\{\alpha_k, \beta_k, \mathbf{m}_k, \nu_k, W_k\})} \prod_i d\theta_i \quad (5.22)$$

と書ける。ここで $\{\alpha_k, \beta_k, \mathbf{m}_k, \nu_k, W_k\}$ は $\{\mathbf{x}_n\}_m$ を用いた場合の推定結果、 $\{\alpha'_k, \beta'_k, \mathbf{m}'_k, \nu_k, W'_k\}$ は $\{\mathbf{x}'_n\}_{m+1}$ を用いた場合の推定結果とした。この K は式 (5.17) の定義を用いて簡単に書き表せ、

$$\begin{aligned} & \int q(\{\theta_i\}|\{\alpha_k, \beta_k, \mathbf{m}_k, \nu_k, W_k\}) \ln q(\{\theta_i\}|\{\alpha'_k, \beta'_k, \mathbf{m}'_k, \nu_k, W'_k\}) \prod_i d\theta_i \\ &= \ln \alpha_0 + \sum (\alpha_0 - 1) \ln \tilde{\pi}_k - \ln \beta'_k - \frac{\beta'_k}{2} \left(\frac{2}{\beta_k} + \nu_k (\mathbf{m}_k - \mathbf{m}'_k)^T W_k (\mathbf{m}_k - \mathbf{m}'_k) \right) \\ &+ \ln B(W'_k, \nu'_k) + \frac{1}{2} (\nu'_k - 1) \ln |\tilde{\Lambda}_k| - \sum_{i,j} \frac{W'^{-1}_{i,j}}{W^{-1}_{i,j}} \end{aligned} \quad (5.23)$$

となる。ただし、 $B(\cdot)$ は以下の通りである。

$$B(W, \nu) \equiv |W|^{-\frac{1}{2}} \left(2^\nu \pi^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{\nu-1}{2}\right) \right)^{-1} \quad (5.24)$$

5.3.3 事前分布の生成

逐次的に測位データを投入する方法として最も単純に考えられるのは、前ステップにて算定された推定結果の事後分布をそのまま次の事前分布とする、逐次ベイズ推定的な方法である。例えば β に関する更新式 (5.7) について、時刻 t の寄与率の和 $N_{t,k}$ を用いた推定結果 $\beta_{t,k} = \beta_0 + N_{t,k}$ を、そのまま寄与率和 $N_{t+1,k}$ の時刻 $t+1$ における推定処理の事前分布とした場合、 $\beta_{t+1,k} = \beta_0 + N_{t,k} + N_{t+1,k}$ となり、これを繰返していくと $\beta_{t+n,k} = \beta_0 + N_{t,k} + N_{t+1,k} + \dots + N_{t+n,k}$ と際限なく大きな値になる。すると、新しいデータの寄与 $N_{t+n,k}$ が相対的に小さくなり推定結果にほとんど影響しなくなってしまう。本来、時刻 t の分布という事象と時刻 $t+1$ の分布という事象は別の事象なので、一つの事象の観測値を追加するような事後推定は適さないのである。そこで変分ベイズ法のパラメータ更新式に登場する負担率の和 $N_{t,k}$ に着目する。負担率とは各測位データの混合分布の各成分 (k と表記されている) に対する寄与の大きさを表しており、分布に寄与するサンプル数 (測位データ数) を連続化したものに相当する。つまり、前述の β の更新式において、新しい測位データの寄与が相対的に小さくなるのは、過去のあらゆる時刻のすべての測位データ数の累積値がハイパーパラメータに含まれているためと解釈される。

そこで本研究では、新しい測位データを追加するときに同時に過去の測位データの寄与を減衰させ、ハイパーパラメータに含まれている測位データ数を一定に保つようにする。まず、時刻 t の測位データセットを十分大きな M 個のブロックに分解する。ここで各ブロックに含まれる測位データ数を $n_t^{(1)}, n_t^{(2)}, \dots, n_t^{(i)}, \dots, n_t^{(M)}$ とする。これらのブロックを逐次投入し、事後確率推定を行う。このとき、 j 番目までのブロックを投入した時点での投入済み測位データの全測位データに対する比

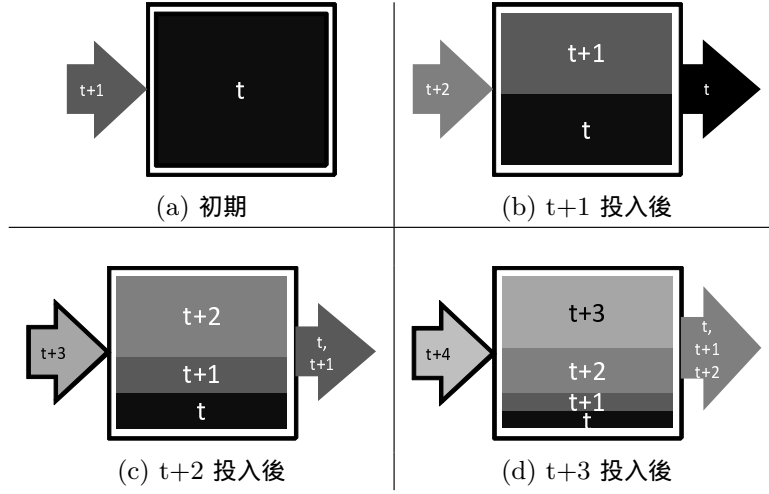


図 5.5 逐次投入における負担率調整の模式図

率は $\frac{\sum_i^j n_t^{(i)}}{\sum_i^M n_t^{(i)}}$ となる。そこで、以下のように減衰率 $\gamma^{(j)}(t)$ を定義する。

$$\gamma_t^{(j)} = 1 - \frac{\sum_i^j n_t^{(i)}}{\sum_i^M n_t^{(i)}} \quad (5.25)$$

この減衰率を用いて、過去の測位データの数を決減させてから新たに現時刻の測位データを投入する。これによって、推定されるハイパーパラメータに含まれる過去のデータの寄与はおおむね一定に保たれると考えられる。

この手順の効果を図 5.5 を用いて説明する。図中 (a) の四角形の箱には時刻 t の寄与が詰め込まれている。これは、 t における測位データの負担率のみから事前分布が作られている状態を意味している。測位データの寄与は (a)、(b)、(c) にて箱に左から入る矢印で示されているように随時投入され、それに伴い、箱内の寄与は投入されたのと同程度程度の寄与が押し出される。図中、箱から右へ出る矢印は前時刻の寄与投入により押し出された分の寄与を示しており、その量は投入データの全データに対する比率 $\gamma_t^{(j)}$ にしたがって、箱内に既存の寄与を同比率で減衰させたのと同様である。すると、(d) にあるように古い時刻の寄与は指数関数的に減衰していくことになる。つまり、現在時刻の箱の中にある寄与のみを保存し、新たなデータ投入時にその投入比率に応じて減衰させるだけで、古いデータの寄与を急激に減らしていくことができるのである。

この減衰の手続きを適用すると、時刻 t の推定結果を利用して生成される時刻 $t+1$ の事前分布は

以下ようになる。

$$\alpha_{0,t+1}^{(0)} = \alpha_0 + \gamma_t^{(0)} N_{t,k} \quad (5.26)$$

$$\beta_{0,t+1}^{(0)} = \beta_0 + \gamma_t^{(0)} N_{t,k} \quad (5.27)$$

$$\mathbf{m}_{0,t+1}^{(0)} = \frac{1}{\beta_{0,t+1}^{(0)}} \left(\beta_0 \mathbf{m}_0 + \gamma_t^{(0)} N_{t,k} \bar{\mathbf{x}}_{t,k} \right) \quad (5.28)$$

$$W_{0,t+1}^{-1(0)} = W_0^{-1} + \gamma_t^{(0)} S_{t,k} N_{t,k} + \frac{\gamma_t^{(0)} N_{t,k} \beta_0}{\gamma_t^{(0)} N_{t,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})^T \quad (5.29)$$

$$\nu_{0,t+1}^{(0)} = \nu_0 + \gamma_t^{(0)} N_{t,k} \quad (5.30)$$

ここで時刻 t において全データを用いた変分ベイズ法の推定結果があるとしており、 $N_{t,k}$, $\bar{\mathbf{x}}_{t,k}$, $S_{t,k}$ はそれぞれ時刻 t における式 (5.13) の値を示している。この事前分布のもとでの推定結果は以下のようになる。

$$\alpha_{t+1,k}^{(0)} = \alpha_0 + \gamma_t^{(0)} N_{t,k} + n_{t,k}^{(0)} \quad (5.31)$$

$$\beta_{t+1,k}^{(0)} = \beta_0 + \gamma_t^{(0)} N_{t,k} + n_{t,k}^{(0)} \quad (5.32)$$

$$\mathbf{m}_{t+1,k}^{(0)} = \frac{1}{\beta_{t+1,k}^{(0)}} \left(\beta_0 \mathbf{m}_0 + \gamma_t^{(0)} N_{t,k} \bar{\mathbf{x}}_{t,k} + n_{t,k}^{(0)} \bar{\mathbf{x}}_{t,k}^{(0)} \right) \quad (5.33)$$

$$\begin{aligned} W_{t+1,k}^{-1(0)} &= W_0^{-1} + \gamma_t^{(0)} S_{t,k} N_{t,k} + s_{t,k}^{(0)} n_{t,k}^{(0)} \\ &\quad + \frac{\gamma_t^{(0)} N_{t,k} \beta_0}{\gamma_t^{(0)} N_{t,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})^T \\ &\quad + \frac{n_{t,k}^{(0)} \beta_0}{n_{t,k}^{(0)} + \beta_0} (\mathbf{m}_{0,t}^{(0)} - \bar{\mathbf{x}}_{t,k}^{(0)})(\mathbf{m}_{0,t}^{(0)} - \bar{\mathbf{x}}_{t,k}^{(0)})^T \end{aligned} \quad (5.34)$$

$$\nu_{t+1,k}^{(0)} = \nu_0 + \gamma_t^{(0)} N_{t,k} + n_{t,k}^{(0)} \quad (5.35)$$

ここで、

$$n_{t,k}^{(0)} = \sum_n r_{t,n,k}^{(0)}, \quad (5.36)$$

$$\bar{\mathbf{x}}_{t,k}^{(0)} = \frac{1}{n_{t,k}^{(0)}} \sum_n r_{t,n,k}^{(0)} \mathbf{x}_{t,n}^{(0)}, \quad (5.37)$$

$$s_{t,k}^{(0)} = \sum_n r_{t,n,k}^{(0)} (\mathbf{x}_{t,n}^{(0)} - \bar{\mathbf{x}}_{t,k}^{(0)})(\mathbf{x}_{t,n}^{(0)} - \bar{\mathbf{x}}_{t,k}^{(0)})^T \quad (5.38)$$

であり、 $\mathbf{x}_{t,n}^{(0)}$, $r_{t,n,k}^{(0)}$ はそれぞれ n 番目のブロックに含まれる測位データとその負担率を意味する。

次にこの推定結果と初期の分布との間の KL 情報量が大きかった場合、新たにデータを追加する。

このとき、事前分布のハイパーパラメータを

$$\alpha_{0,t+1}^{(1)} = \alpha_0 + \gamma_t^{(1)} N_{t,k} + n_{t,k}^{(0)} \quad (5.39)$$

$$\beta_{0,t+1}^{(1)} = \beta_0 + \gamma_t^{(1)} N_{t,k} + n_{t,k}^{(0)} \quad (5.40)$$

$$\mathbf{m}_{0,t+1}^{(1)} = \frac{1}{\beta_{0,t+1}^{(1)}} \left(\beta_0 \mathbf{m}_0 + \gamma_t^{(1)} N_{t,k} \bar{\mathbf{x}}_{t,k} + n_{t,k}^{(0)} \bar{\mathbf{x}}_{t,k}^{(0)} \right) \quad (5.41)$$

$$\begin{aligned} W_{0,t+1}^{-1(1)} &= W_0^{-1} + \gamma_t^{(1)} S_{t,k} N_{t,k} + s_{t,k}^{(0)} n_{t,k}^{(0)} \\ &\quad + \frac{\gamma_t^{(1)} N_{t,k} \beta_0}{\gamma_t^{(1)} N_{t,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})^T \\ &\quad + \frac{n_{t,k}^{(0)} \beta_0}{n_{t,k}^{(0)} + \beta_0} (\mathbf{m}_{0,t}^{(0)} - \bar{\mathbf{x}}_{t,k}^{(0)})(\mathbf{m}_{0,t}^{(0)} - \bar{\mathbf{x}}_{t,k}^{(0)})^T \end{aligned} \quad (5.42)$$

$$\nu_{0,t+1}^{(1)} = \nu_0 + \gamma_t^{(1)} N_{t,k} + n_{t,k}^{(0)} \quad (5.43)$$

とすればよい。これを繰り返すと i 回目の更新における事前分布は

$$\alpha_{0,t+1}^{(i)} = \alpha_0 + \gamma_t^{(i)} N_{t,k} + \sum_j^{i-1} n_{t,k}^{(j)} \quad (5.44)$$

$$\beta_{0,t+1}^{(i)} = \beta_0 + \gamma_t^{(i)} N_{t,k} + \sum_j^{i-1} n_{t,k}^{(j)} \quad (5.45)$$

$$\mathbf{m}_{0,t+1}^{(i)} = \frac{1}{\beta_{0,t+1}^{(i)}} \left(\beta_0 \mathbf{m}_0 + \gamma_t^{(i)} N_{t,k} \bar{\mathbf{x}}_{t,k} + \sum_j^{i-1} n_{t,k}^{(j)} \bar{\mathbf{x}}_{t+1,k}^{(j)} \right) \quad (5.46)$$

$$\begin{aligned} W_{0,t+1}^{-1(i)} &= W_0^{-1} + \gamma_t^{(i)} S_k N_k + \sum_j^{i-1} s_{t+1,k}^{(j)} n_{t+1,k}^{(j)} \\ &\quad + \frac{\gamma_t^{(i)} N_{t,k} \beta_0}{\gamma_t^{(i)} N_{t,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})^T \\ &\quad + \sum_j^{i-1} \frac{n_{t+1,k} \beta_{0,t+1}^{(j)}}{n_{t+1,k} + \beta_{0,t+1}^{(j)}} (\mathbf{m}_{0,t+1}^{(j)} - \bar{\mathbf{x}}_{t+1,k})(\mathbf{m}_{0,t+1}^{(j)} - \bar{\mathbf{x}}_{t+1,k})^T \end{aligned} \quad (5.47)$$

$$\nu_{0,t+1}^{(i)} = \nu_0 + \gamma_t^{(i)} N_{t,k} + \sum_j^{i-1} n_{t,k}^{(j)} \quad (5.48)$$

となる。ここで、更新式を得るために i 番目と $i+1$ 番目の式の差をとる。 β_0 に関する式は

$$\beta_{0,t+1}^{(i+1)} = (\gamma_t^{(i+1)} - \gamma_t^{(i)}) N_{t,k} + \beta_{0,t+1}^{(i)} + n_{t,k}^{(i)} \quad (5.49)$$

となる。ここで $\beta_k^{(i)} = \beta_0^{(i)} + n_k^{(i)}$ なので

$$\beta_0^{(i+1)} = \beta_{t,k}^{(i)} + (\gamma_t^{(i+1)} - \gamma_t^{(i)}) N_{t,k} \quad (5.50)$$

と記述できる。同様にして

$$\alpha_{0,t+1}^{(i)} = \alpha_{t,k}^{(i)} + (\gamma_t^{(i+1)} - \gamma_t^{(i+1)})N_{t,k} \quad (5.51)$$

$$\mathbf{m}_{0,t+1}^{(i)} = \frac{1}{\beta_{0,t+1}^{(i)}} \left(\beta_{t+1,k}^{(i)} \mathbf{m}_{t+1,k}^{(i)} + (\gamma_t^{(i+1)} - \gamma_t^{(i+1)})N_{t,k} \bar{\mathbf{x}}_{t,k} \right) \quad (5.52)$$

$$\begin{aligned} W_{0,t+1}^{-1(i)} &= W_{t,k}^{-1(i)} + (\gamma_t^{(i+1)} - \gamma_t^{(i+1)})N_{t,k} S_{t,k} \\ &+ \left(\frac{\gamma_t^{(i+1)} N_{t,k} \beta_0}{\gamma_t^{(i+1)} N_{t,k} + \beta_0} - \frac{\gamma_t^{(i)} N_{t,k} \beta_0}{\gamma_t^{(i)} N_{t,k} + \beta_0} \right) \\ &\times (\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})^T \end{aligned} \quad (5.53)$$

$$\nu_{0,t+1}^{(i)} = \nu_{t,k}^{(i)} + (\gamma_t^{(i+1)} - \gamma_t^{(i+1)})N_{t,k} \quad (5.54)$$

となる。このように、 $i + 1$ 番目の事前分布は直前の推定結果 $\alpha_{t,k}^{(i)}, \beta_{0,t+1}^{(i)}, \mathbf{m}_{t+1,k}^{(i)}, W_{t,k}^{-1(i)}, \nu_{t,k}^{(i)}$ と減衰率 $\gamma_t^{(i+1)}, \gamma_t^{(i+1)}$ から算出することができる。この更新式を KL 情報量が十分小さくなるまで繰り返した後、最後に得られる結果が時刻 $t + 1$ の推定結果となる。推定結果を得たのが $i = l$ のときであったとすると、

$$N_{t+1,k} = \alpha_{t,k}^{(l)} - \alpha_0 = \beta_{t,k}^{(l)} - \beta_0 \quad (5.55)$$

$$\bar{\mathbf{x}}_{t+1,k} = \frac{1}{N_{t+1,k}} \left(\beta_{t+1,k}^{(l)} \mathbf{m}_{t+1,k}^{(l)} - \beta_0 \mathbf{m}_0 \right) \quad (5.56)$$

$$\begin{aligned} S_{t+1,k} &= \frac{1}{N_{t+1,k}} \left(W_{t,k}^{-1(l)} - W_0^{-1} \right. \\ &\left. - \frac{N_{t+1,k} \beta_0}{N_{t+1,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t+1,k})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t+1,k})^T \right) \end{aligned} \quad (5.57)$$

のように定義した $N_{t+1,k}, \bar{\mathbf{x}}_{t+1,k}, S_{t+1,k}$ を用いて

$$\alpha_{0,t+1,k} = \alpha_0 + N_{t+1,k} \quad (5.58)$$

$$\beta_{0,t+1,k} = \beta_0 + N_{t+1,k} \quad (5.59)$$

$$\mathbf{m}_{0,t+1,k} = \frac{1}{\beta_{0,t+1,k}} (\beta_0 \mathbf{m}_0 + N_{t+1,k} \bar{\mathbf{x}}_{t+1,k}) \quad (5.60)$$

$$\begin{aligned} W_{0,t+1,k}^{-1} &= W_0^{-1} + S_{t+1,k} N_{t+1,k} \\ &+ \frac{N_{t+1,k} \beta_0}{N_{t+1,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t+1,k})(\mathbf{m}_0 - \bar{\mathbf{x}}_{t+1,k})^T \end{aligned} \quad (5.61)$$

$$\nu_{0,t+1,k} = \nu_0 + N_{t+1,k} \quad (5.62)$$

と書き換えることができる。この形は変分ベイズ法における更新式と同等の形になっている。したがって、次の時刻 $t + 2$ に対しても同様の更新手続きが実行できる。なお、ここで定義した値は以下

表 5.1 使用データ

項目名	詳細
使用データ	人の流れ PJ[30] 東京都市圏 (H20) の位置情報 (緯度経度)
件数	毎分約 60 万人分
時間	午前 7 時 ~ 午前 9 時

のようにも書ける。

$$N_{t+1,k} = \gamma_t^{(l)} N_{t,k} + \sum_j^l n_{t+1,k}^{(j)}, \quad (5.63)$$

$$\bar{\mathbf{x}}_{t+1,k} = \gamma_t^{(l)} N_{t,k} \bar{\mathbf{x}}_{t,k} + \sum_j^l n_{t+1,k}^{(j)} \bar{\mathbf{x}}_{t+1,k}^{(j)}, \quad (5.64)$$

$$S_{t+1,k} = \gamma_t^{(l)} S_{t,k} N_{t,k} + \sum_j^l s_{t+1,k}^{(j)} n_{t+1,k}^{(j)} + \delta W \quad (5.65)$$

ここで用いた δW は

$$\begin{aligned} \delta W = & \frac{\gamma_t^{(i)} N_{t,k} \beta_0}{\gamma_t^{(i)} N_{t,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k}) (\mathbf{m}_0 - \bar{\mathbf{x}}_{t,k})^T \\ & - \frac{N_{t+1,k} \beta_0}{N_{t+1,k} + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_{t+1,k}) (\mathbf{m}_0 - \bar{\mathbf{x}}_{t+1,k})^T \\ & + \sum_j^l \frac{n_{t+1,k} \beta_{0,t+1}^{(j)}}{n_{t+1,k} + \beta_{0,t+1}^{(j)}} (\mathbf{m}_{0,t+1}^{(j)} - \bar{\mathbf{x}}_{t+1,k}) (\mathbf{m}_{0,t+1}^{(j)} - \bar{\mathbf{x}}_{t+1,k})^T \end{aligned} \quad (5.66)$$

である。式 (5.13) の定義を参照すると、 N および $\bar{\mathbf{x}}$ については過去のデータの寄与を減衰させて加算をとった自然な形になっている。 S についても同様であるが、 δW の項だけが残っている。ただし、 δW は逐次計算に伴って発生する誤差であり、 N や n が十分大きい場合には、ほとんど計算には影響しない。したがって、上記計算式による更新で分布が推定できる。

5.4 実験

5.4.1 実験条件

提案方式の評価のための実験を行った。表 5.1 に実験条件を示す。今回用いたのは、アンケートにより収集された人がいつどこへ移動したかという情報に対して交通網の経路探索などの内挿処理を施した位置情報のデータ [30] であり、1 分おきに全人物がどの地点にいたかが格納されている。このデータのうち午前 7 時から午前 10 時までのデータについて分布を生成する。なお、各データは緯度経度で表記されているため、初期時刻のデータの平均座標値を原点にし、外接矩形の大きさが 10.0 になるように正規化したものを入力データとした。

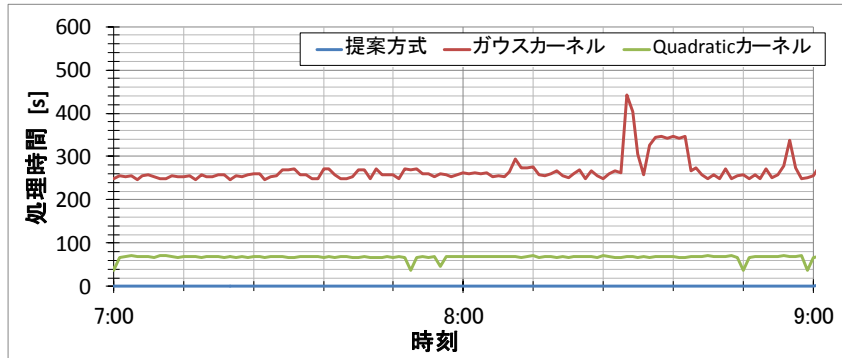


図 5.6 処理速度の比較

以下に評価の手順を述べる。まず、各時刻の人の流れデータをランダムに並び変え、最初の 50% のデータから分布を推定し、後半 10% のデータを対象に精度評価を行った。なお、後半 10% を評価対象としたのは、データ件数が多すぎて処理時間が過剰にかかるためで、10% つまり 6 万点のデータを対象とすれば評価には十分な数が得られていると考えられる。精度の評価基準としては、各点に関する予測分布の値の対数の平均値、つまり N 点の測位データ \mathbf{x}_n について

$$Ev = \frac{1}{N} \sum_{n=1}^N \ln P(\mathbf{x}_n) \quad (5.67)$$

によって定義される Ev を精度とした。なお、この値は予測分布から \mathbf{x}_n が得られる同時観測確率を規格化したものを意味している。比較のため今回の対象領域 10.0×10.0 の範囲の一様分布をこの評価基準で評価すると、 $Ev = \frac{1}{N} \sum_{n=1}^N \ln P(\frac{1}{10^2}) \simeq -4.61$ となる。

提案方式では、ガウス分布の混合数を 50 とし、測位データの逐次投入ブロックの大きさは 1,500 とした。ハイパーパラメータの初期値は $\alpha_0 = 2.1, \beta_0 = 1.0$ とし、 W_0 は対角成分が 0.4、対角成分以外は 1.0×10^{-10} 、 $\nu_0 = 2.2$ とした。また、 \mathbf{m}_0 は初期時刻の点からランダムに選択した座標値を用いた。なお、計算を始める初期時刻を対象時刻の 3 時間前の午前 4 時とし、逐次投入に伴う密度分布更新が十分に行われた後の状態を評価対象としている。比較対象として、ガウスクーネル法、Quadratic カーネル法に関しても同様の評価を行った。ガウスクーネル法の平滑化パラメータは $h=0.15$ とした。また、Quadratic カーネル法の平滑化パラメータは $P(\mathbf{x}_n) = 0$ とならないように $h=3$ とした。これは $P(\mathbf{x}_n) = 0$ ならば $Ev = -\infty$ であり比較に適さないためである。なお、実行環境としては Java (JDK7 Update 13) で実装されており、Intel Core i7-3980K 3.2GHz、RAM 32.0GB のコンピュータを使用した。

5.4.2 実験結果

図 5.6 に処理速度の比較を示す。この処理速度としては、測位データから 150×100 画素の画像を生成するのにかかる時間を計測して用いた。この計算では横軸の時刻における測位データから画像生成するのにかった時間を縦軸とした。提案方式は 0.5 ~ 1.0 秒程度、平均 0.67 秒で処理を終えているのに対し、従来方式はガウスクーネルで 250 ~ 270 秒で平均 260 秒、Quadratic カーネルでも

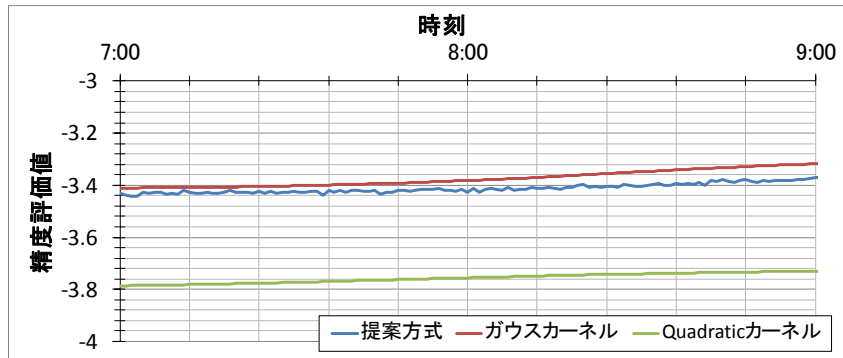


図 5.7 分布推定精度の比較

30～70 秒、平均 67 秒程度の処理時間を要している。提案方式の処理速度は従来のいずれの方式に比べても十分高速である。1 分のデータを処理するのに 1 分以上の時間をかけるとリアルタイムではなくなるが、提案方式だけがその要求を満たしており、提案方式は従来比 100 倍以上高速であったといえる。

図 5.7 に精度に関する実験結果を示す。精度に関しては、ガウスカーネルが最もよい結果を示した。Quadratic カーネルは処理時間が高速化された反面、精度はガウスカーネルよりも低い値となった。一方、提案方式はガウスカーネルより精度が低いものの、Quadratic カーネルよりは大きく改善している。図 5.8 に実際に生成された分布の比較を示す。本評価で生成した画像は 150×100 画素の画像であるが、これを 1500×1000 画素の背景地図 [82] に重畳した結果を示している。(a) が提案方式、(b) はガウスカーネル法、(c) は Quadratic カーネル法で生成された画像である。ガウスカーネル法で想定するガウス分布は常に円形であるのに対し、提案方式におけるガウス分布では楕円状の分布を許している。そのため、提案方式の生成画像は鉄道路線などに沿って線状の分布が得られている。また、(c) の左側は精度評価に用いたパラメータ $h = 3$ の Quadratic カーネル法の結果であるが、薄く全体に広がってしまっている。これは、予測分布が 0 にならないように h を決めたためであり、例えば $h=1$ にすると右側の画像が得られ、他の方式と近い分布となる。ただし、このときの精度指標は $-\infty$ である。このように Quadratic カーネル法で生成された分布は予測分布として扱うには適さない。ただし、可視化のためにパラメータ調整すれば有用である。

5.4.3 考察

上記の実験結果から、提案方式は従来よりも非常に高速に処理ができると判明した。これは、提案方式のデータ数削減の影響が大きいと考えられる。実際には、通信などにかかる時間が発生することを考慮すれば、提案方式の効果はさらに大きくなると推定される。図 5.9 に提案方式において使用されたデータ数を示す。平均 4,500 件程度データを使用しており、最大でも 30,000 件程度、すなわち全体の 1/10 まで使用データを削減したことになる。今回、逐次投入されるブロックは 1500 件ずつであるので、平均的には 2～3 回逐次投入が行われたといえる。なお、精度について、全方式とも 9:00 に近づくにつれて値が大きくなっているが、これは当該時間帯が通勤時間帯であり、人口が都心部に密集してきているために予測がしやすくなったということが原因と考えられる。また提案方式は

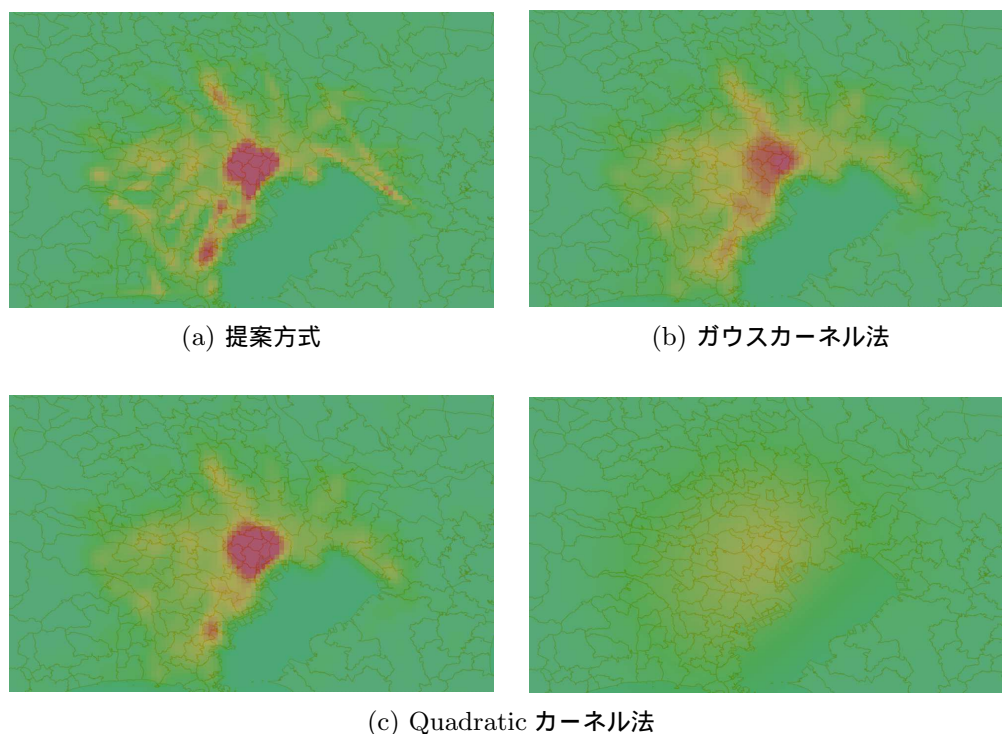


図 5.8 生成された分布画像の比較

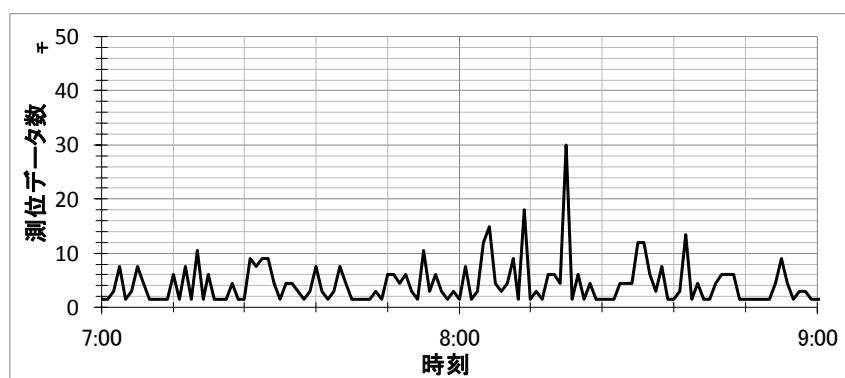
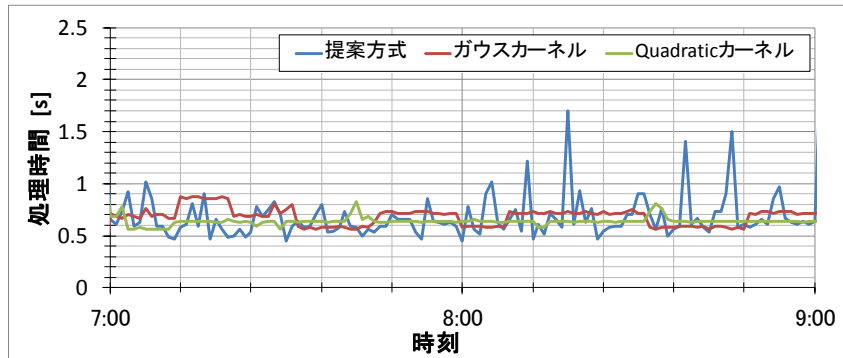


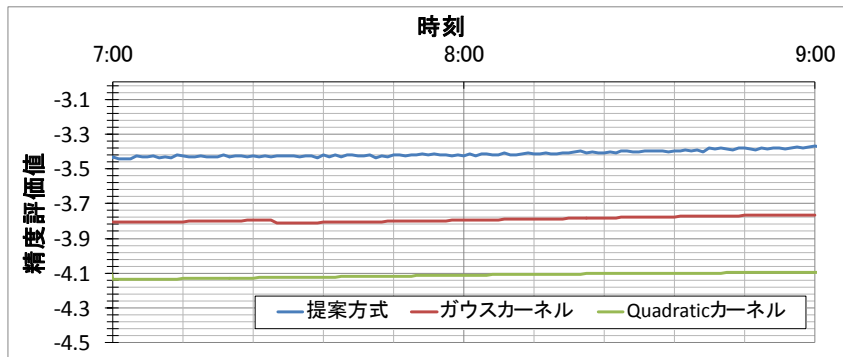
図 5.9 提案方式の使用データ数

従来のガウシアンカーネル法にくらべてこの精度向上が低く差が大きくなっているという挙動を示している。これは、提案方式の処理中で過去のデータを使用しているため、人口分布が急激に変化した場合でも変化前のデータの寄与が残り、分布の変化に追従するのが遅くなっているためであると考えられる。

従来方式で使用するデータ数を削減する方式としてランダムにデータを間引く方法が考えられる。そこで、提案方式と同程度の速度になるまで測位データを間引きしてから従来方式を適用した場合の結果を図 5.10 に示す。このグラフに表示した提案方式は、比較のため図 5.6 にて示した結果を再掲したものである。ガウシアンカーネル法で用いたのは 1100 件、Quadratic カーネル法で用いたのは 3000 件であり、(a) に示すとおり、それぞれ平均 0.68 秒、0.64 秒で処理を終えており、提案方式



(a) 使用データ削減時の従来方式速度



(b) 使用データ削減時の従来方式精度

図 5.10 データ削減による精度低下

(0.67 秒)と同程度の処理速度になった。このとき、(b) のとおり提案方式が最も高精度な結果が得られた。なお、ガウシアンカーネル法、Quadratic カーネル法では、データを削減した結果、前節の評価時に用いたパラメータでは精度が $-\infty$ になってしまった。これは、点間引きによって極端に尤度の低い点が現れてしまい、計算機精度を下回ったためである。そこで、ガウシアンカーネル法では $h = 3.0$ 、Quadratic カーネル法では $h = 5.0$ とした時の結果を示している。以上のように、提案方式と同程度の速度で処理ができるように間引きを行うと、精度が著しく低下していることが確認でき、提案方式は間引きを行っても精度を保つ方式であるといえる。

図 5.11 に提案方式の処理時間の内訳を示す。提案方式はパラメトリックな推定方式であり、密度分布生成と画像生成の 2 つの手順それぞれで処理時間がかかる。画像生成の処理時間はおよそ 0.5 秒で一定しているのに対し、分布推定はほとんど時間がかからないこともあるが 1 秒以上の時間がかかることもあった。これは、データの逐次投入による影響であり、データを多数投入する必要がある場合には処理時間がやや大きい、計算が打ち切れた場合は極端に小さくなっているといえる。

5.5 おわりに

本研究では、移動体端末から収集された測位データの密度分布を高速に精度よく推定する方法として、変分ベイズ法にもとづき一部の測位データだけを利用して全体分布を推定する分布推定方式を

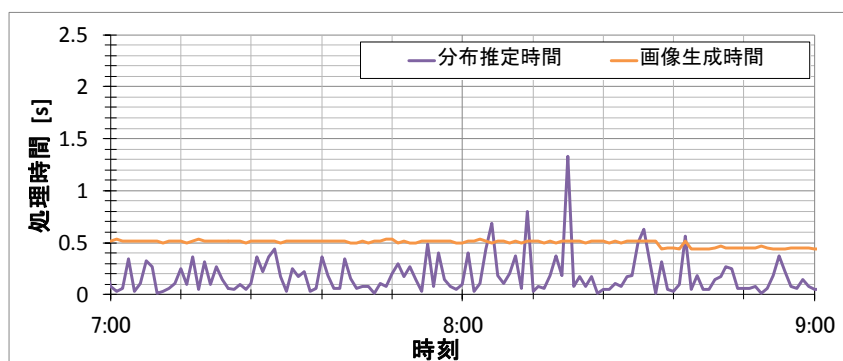


図 5.11 提案方式における処理時間の内訳

提案した。また、提案方式について東京都市圏の人の流れデータを用いて評価したところ、従来に比べて約 100 倍の速度で分布が推定できた。さらに全データの半数を用いて生成された分布と残りのデータの比較による精度評価を行い、提案方式は従来方式と遜色ない精度が得られ、実際に生成される分布画像も従来と大きくは異ならないことが確認できた。提案方式において分布生成に用いる測位データは全体のおよそ 1/10 程度でよいことが判明した。

本研究で対象としたのはもっとも基本的な密度分布推定であるが、今後、空間上に分布する量的データに関する推定処理 [74] についても同様のアプローチをとれる可能性があり、それは今後の課題である。

第 6 章

主 動 線 抽 出

6.1 はじめに

本章では、人の位置情報の履歴をから移動そのものの概要をタイムリーにわかりやすく表示することを目的とし、移動そのものの概略として基本的な、人がどう動いているかという情報を抽出する方法を提案する。なかでも特に主な移動を抽出して可視化し、交通最適化のための意思決定を支援する技術について述べる。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伴うため、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。

一般的には、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。しかし、これらの調査には多大な時間と費用を要するため、数年から十年に一度調査して都市計画を策定するのが一般的である。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 [13][14]、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するのである。

本研究では特に人や車両などの移動体の位置情報に着目する。例えば、スマートフォンなどのモバ

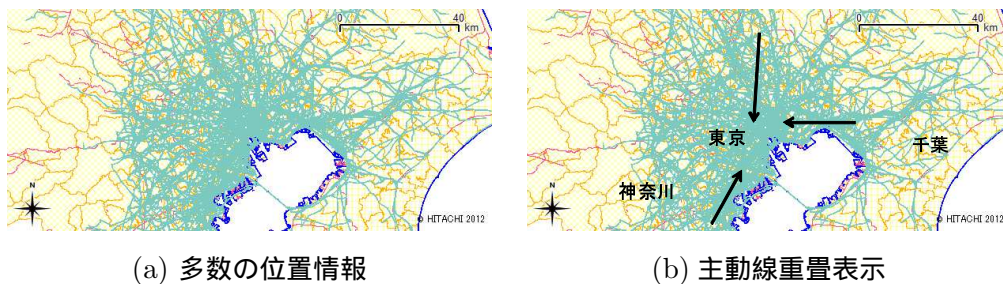


図 6.1 多数の位置情報

イル機器 [2][1]、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わかれば交通需要を得られ、その交通手段がわかれば移動体の移動経路が集中して交通がよどんでいる箇所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 6.1 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 [82] 上に東京都市圏の AM7:00-10:00 の人の流れデータ [30](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わかるが、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、流れの特徴は把握しにくい。例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と時間がかかりすぎて速やかな意思決定に結びつかない。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、主動線とよぶ)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。

本章では、位置情報の時系列(以降、動線とよぶ)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当てはめる方法に着目し、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み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。

6.2 関連研究

6.2.1 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が、その定義は明確ではない。そこで、想定しうる主動線の種類について図 6.2 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であり、右方向に時間が進むものとする。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 6.2(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 6.2(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よぶ。

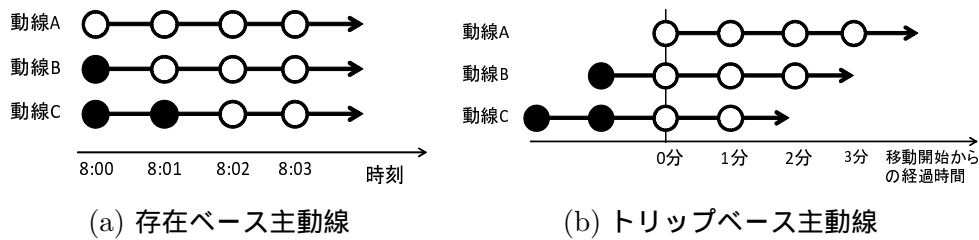


図 6.2 存在ベースとトリップベースの比較

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、移動中か滞留中かを問わず一律にデータが得られるので、移動滞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。その一つとして、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 [46] や DBSCAN によるもの [45] 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 6.3 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）とみなし、一般的な多変量のクラスタリング手法（例えば k-means）により構築されたクラスターの代表（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せねばならず、ベイズ推定を適用する方法が提案されている [46]。

6.2.2 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 6.4 を用いて示す。(a) は図 6.1 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

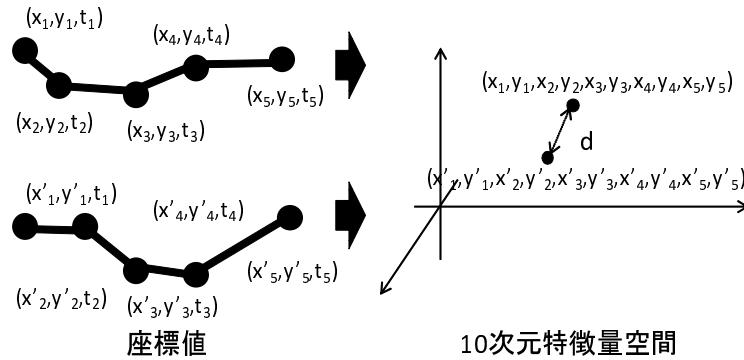


図 6.3 多変量クラスタリング

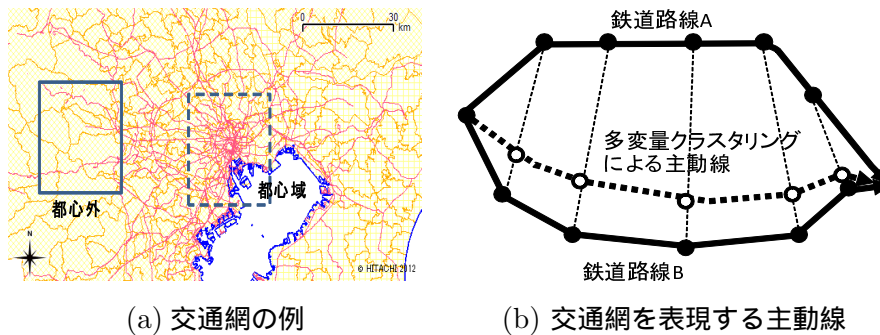


図 6.4 交通網に関する主動線

移動はこのような詳細さの異なる交通網に跨って起こる。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。

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、例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 [35]。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。

記号列にクラスタリングを適用しようとした場合、記号列の距離は動線形状の距離の定義に比べて自然な定義が難しいという点が問題になる。例えば、地点 A と地点 B の距離が何 m であるかを算出することは容易にできる。しかし、地点 A で停留中であることと地点 B を通行中であることの間の

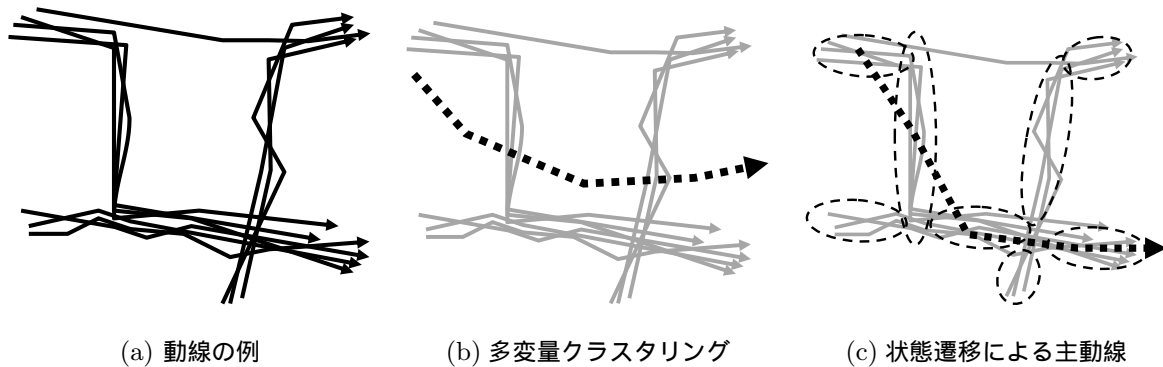


図 6.5 主動線の比較

距離が何 m に相当するかは不明確である。また、地点は異なるが意味的には類似しているケースもある。例えば、駅構内のベンチで休憩している、というような場合はそれがベンチの間の距離にかかわらず、非常に類似した行動といえる。他にも、アプリケーションによって同一視すべき状態は異なり、状態遷移記号列間の距離を定義するのは容易ではない。

そこで、類似度を直接的に用いないクラスタリング手法の一つである混合確率分布によるクラスタリングを適用する。混合確率分布は、複数種類の確率的事象によってデータが生成される時に各データがどの確率的事象によって生成されたか区別がつかない場合のデータの生成確率を表現する確率分布である。混合確率分布にてデータを表現することができれば、データがどの確率的事象から生成されたかを推定することにより、クラスタリングができることになる。また、この推定の確からしさに基づき各クラスターの代表的な状態遷移列、ひいてはその動線を得ることができる。

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、例えば尤度の低い動線は異常行動であると判定する [44] 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 6.5 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 [53][76] や、人々の行動選択が潜在的なマルコフ連鎖にもとづくとした HMM (隠れマルコフモデル) を用いた研究 [63] 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。

6.3 時空間ネットワークを用いた主動線抽出

6.3.1 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にてモデル化することが多い。例えば、3 章の MMM(混合マルコフモデル)[53] がある。MMM では、人物の特徴を示す潜在的な変数を 1 つ導入し、この変数の値によって行き先選択の確率が定まる。つまり、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。

6.3.2 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 6.6(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辿っていき、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べば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

$$L = \log \omega_{S,i} + \log \omega_{i,j} + \cdots + \log \omega_{k,G} \quad (6.1)$$

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辿ると同じ動きの繰り返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ことが考えられる。その一つが、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。例えば、MAR-HMM(混合自己回帰隠れマルコフモデル)[63] 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に依存するように変更することが考えられる。時間依存するマルコフ連鎖は準マルコフ連鎖 (Semi Markov-chain)[83] と呼ばれている。図 6.6(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化するため、終状態への遷移するのは常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

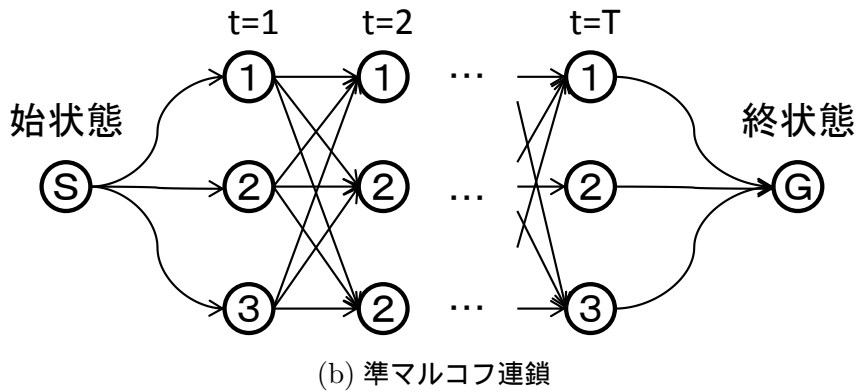
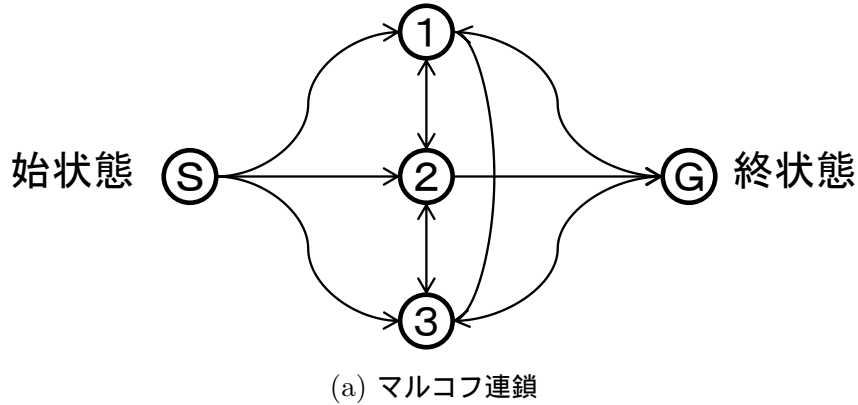


図 6.6 マルコフ連鎖と準マルコフ連鎖

(MSMM; Mixed Semi Markov-chain Model) を用いたモデル化を提案する。

6.3.3 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するとしている。 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

$$P_k(\{d_{nt}\}) = \prod_t \prod_{\mu} \prod_{\nu} \omega_{k,\mu,\nu,t}^{d_{n,t,\mu} d_{n,t-1,\nu}} \quad (6.2)$$

にて定義される P_k を用いて

$$P(d) = \sum_k^K \pi_k P_k(\{d_{nt}\}) \quad (6.3)$$

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。

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 6.7 MSMM に対する EM アルゴリズム

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 $\pi_k, \omega_{k,\mu,\nu,t}$ 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 の組で一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

$$P_k(\{d_{nt}\}) = \prod_{\xi} \prod_{\eta} \omega_{k,\xi,\eta}^{d_{n,\xi} d_{n,\eta}} \quad (6.4)$$

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、 $t \neq t' + 1$ の時の $\xi = (\mu, t)$ 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 6.7 にこの MSMM に対する 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。この計算は MMM の EM 法と同等であるので、最尤解の唯一性などについても MMM の場合に準ずる。すなわち、局所解への対応は MMM 同様に必要である。

6.3.4 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、数式 (6.1) で示したような尤度の高い動線を求めればよい。つまり、

$$\ln L_k(\mathbf{d}) = \sum_{t,\mu,\nu} d_{t,\mu} d_{t-1,\nu} \ln \omega_{k,\nu,\mu,t} \quad (6.5)$$

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 6.8 主動線算定アルゴリズム

が最大になる遷移 d を求めることになる。前述の通り、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 [84] 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 6.8 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、

$$P_{k,\mu,t} = \sum_{\nu} \omega_{k,\mu,\nu,t-1} \quad (6.6)$$

と漸化式によって求まる。図 6.8 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と、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。

$$P_{k,\mu,\nu,t}^{\text{mov}} = P_{k,\mu,t} \omega_{k,\mu,\nu,t-1} \quad (6.7)$$

$\omega_{k,\mu,\nu,t-1}$ 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。

6.4 実験による評価

6.4.1 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 [30] に対して主動線抽出処理を行い、その結果を比較した。表 6.1 に今回用いたデータの概要を示す。今回、平成 20 年

表 6.1 本実験にて用いたデータ

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「サービス業従事者」で常時存在する 4686 人

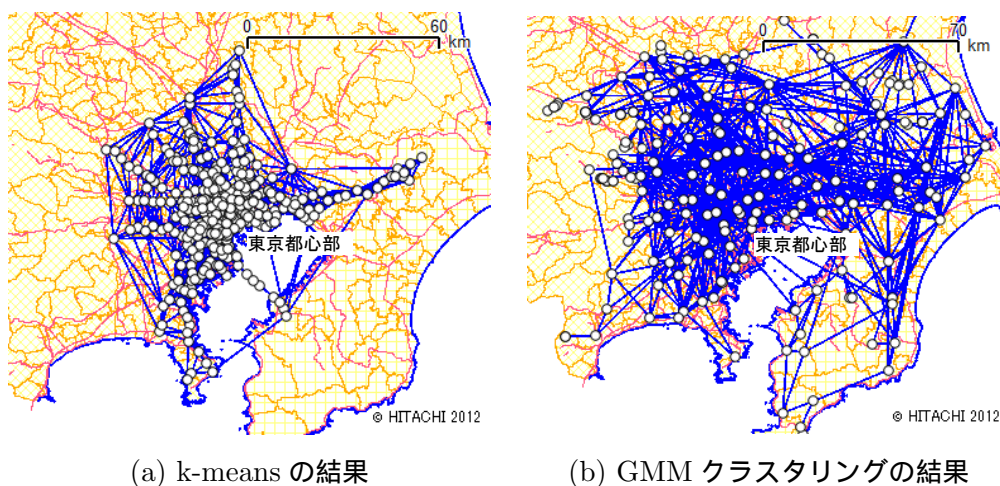


図 6.9 位置クラスタリング結果

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 6.1 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 6.9(a) に $k=400$ とした k-means の結果を、図 6.9(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ーが配置されていることがわかる。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。

6.4.2 結果と考察

図 6.10 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 におけるクラスター生成では、クラスターは距離と件数によって定まるが、交通

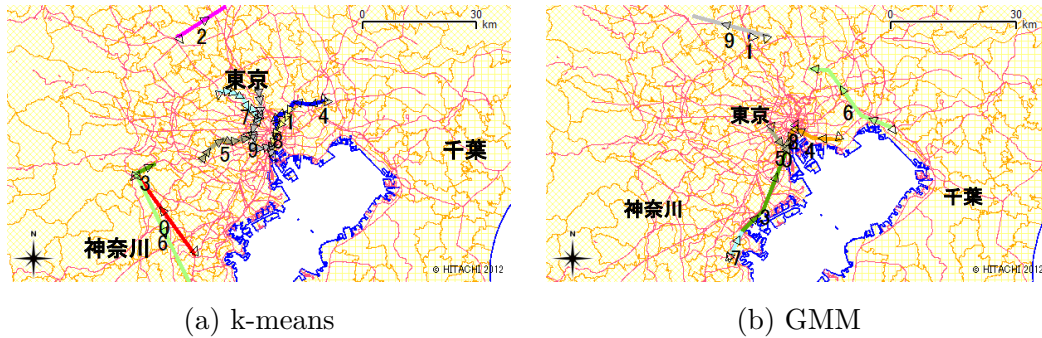


図 6.10 MSMM(提案方式)を用いた抽出結果

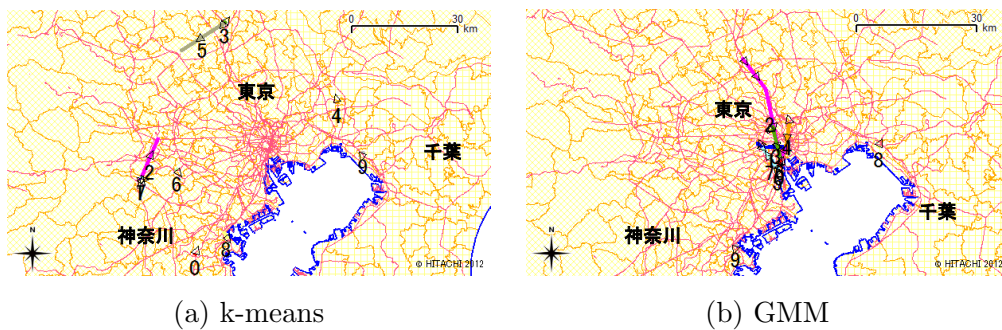


図 6.11 MMMを用いた抽出結果

量の多い鉄道沿線は件数が多く、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、多数の動きが一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。郊外での振る舞いは図 6.11 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 6.12 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 6.2 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。

これらの結果を比較するため、図 6.13 に k-means を用いた各方式にて抽出された 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 500m 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

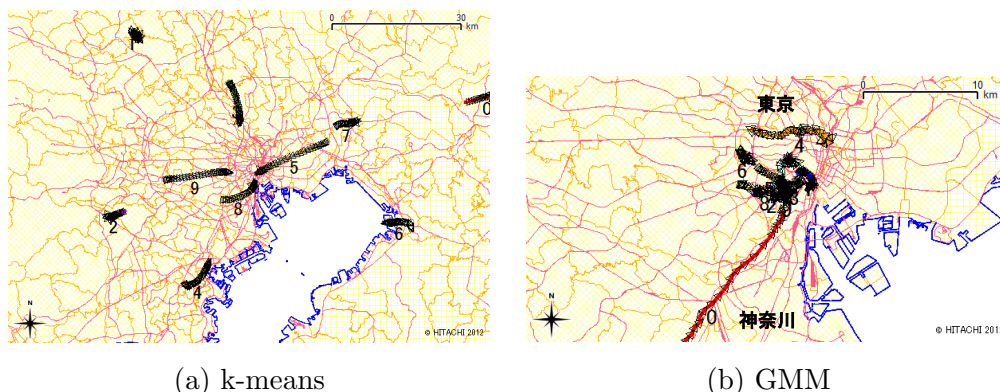


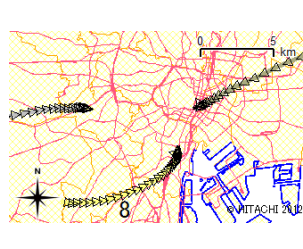


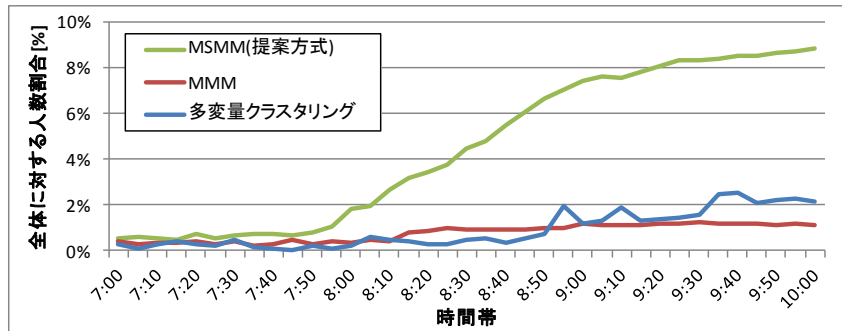
図 6.12 多変量クラスタリングによる抽出結果

表 6.2 方式間の比較

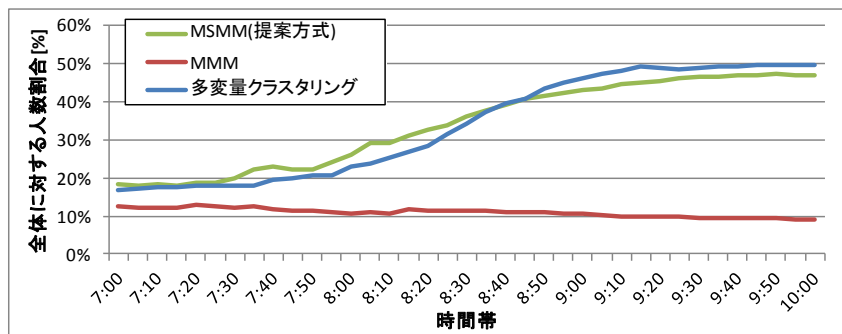
方式	MSMM(提案方式)	MMM	多変量クラスタリング
代表画像			
特徴	都心域で交通網に沿った主動線	郊外の滞留が多い	交通網との対応がとられていない

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、都心域に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 5000m 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 6.14 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 6.15(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 6.15(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 6.16 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) の多変量クラスタ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わかる。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



(a) 500m 内の人数割合



(b) 5000m 内の人数割合

図 6.13 主動線付近の人数割合の時系列変化

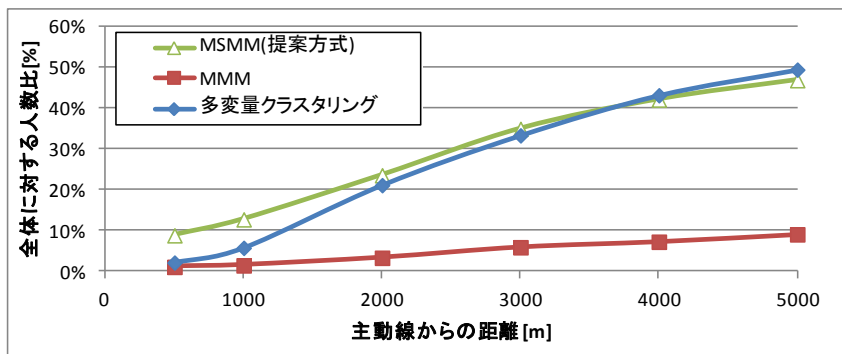


図 6.14 10:00 の主動線付近の人数割合

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 6.17 に図 6.10(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 6.17(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の短い動きばかりが抽出されている。図 6.17(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

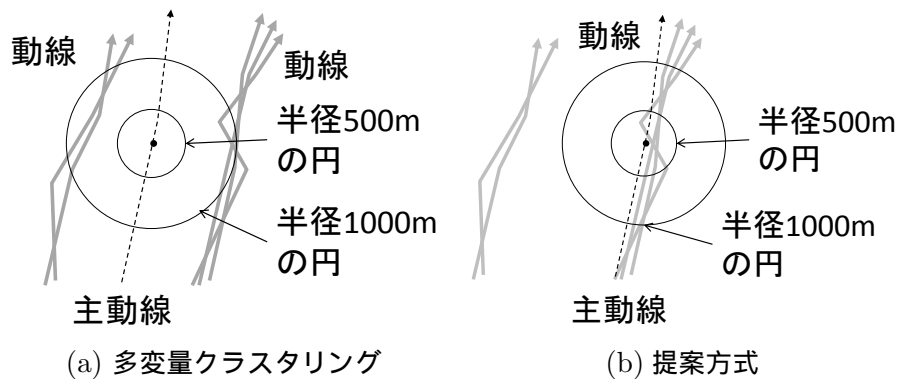


図 6.15 主動線付近の人数の概念図

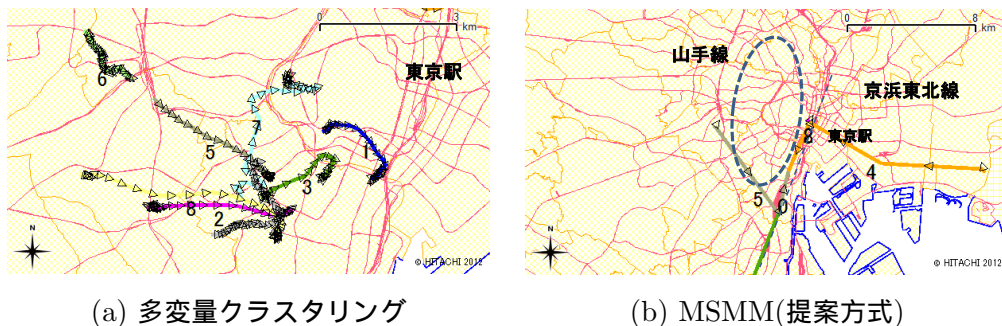


図 6.16 主動線の拡大図

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、一つの路線上での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。例えば、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。

6.5 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

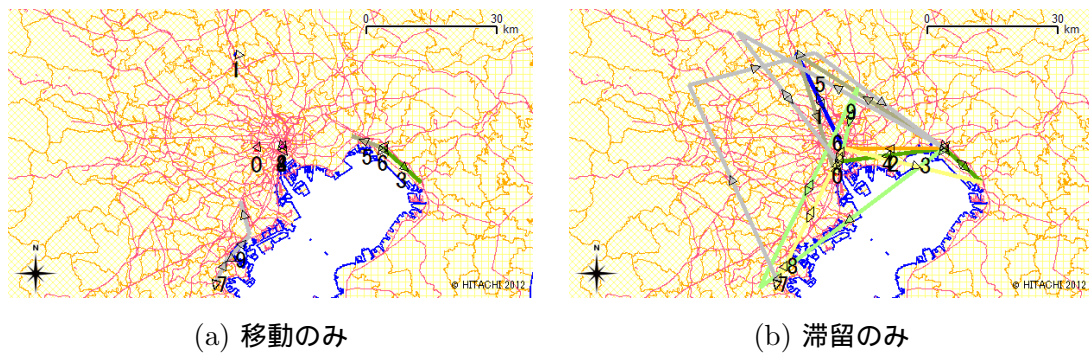


図 6.17 主動線抽出方法の比較

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。

第 III 部のまとめ

第 III 部では、多数の人の位置情報により移動傾向を把握できる表示のための確率モデルとして人の集団としての傾向をモデル化することを対象として、位置と移動の 2 つについてその全体の傾向を確率モデルで表現する方法を提案した。第 II 部で扱った MAR-HMM や MMM は、人の個性を折りこみ、その推定も可能な確率モデルで表現していた。しかし、全体の傾向を可視化するために、第 III 部では人の平均的な移動を表現するモデルを用いた。

人の位置については、人の密度という指標を用いたヒートマップ表示により、各個人の特徴を平均化した。この場合、各個人の位置は時間断面で集約されるが、その元が位置情報であることを加味してマルコフ性を想定すると、直前の人の密度分布が次の人の密度分布と関連をもつことが推定できる。変分ベイズ法にもとづく密度推定をもちい、ベイズ推定における事前確率として直前の密度分布の一部を使うことでその精度を落とさずに処理時間を短縮できた。

人の移動については、主動線抽出という複数の人の移動パターンを可視化することを目的とした。第 II 部までの議論が示すように、人の移動では時間変化を扱うことが肝要であるが、AR-HMM のような個人ごとの状態遷移を前提とした方式は、平均的なパターンを扱うには適さない。そこで、混合準マルコフ連鎖モデルという、あるパターン内での時間変化は前人物同一としたモデルを提案した。これにより、全体の遷移確率の時間変化の代表的な動きを探すと主動線が抽出できた。

これらにより、ある特定の時間の人の移動の平均はマルコフ的なモデルである程度は表現可能であることを示すと考えられる。第 II 部までの議論では、個人の移動のパターンに現れる個性がの豊富さが、時間経過に伴う差の現れ方に寄与するという仮説が提示された。しかし、同時にある瞬間の移動には平均的なパターンが存在していることも示唆された。すなわち、平均的な位置の分布を可視化したのがヒートマップであり、平均的な移動をつなぎ合わせたのが主動線である。これらの平均的な可視化は、無数にありすぎて理解が困難な人の個性を平均として理解可能にするという点で効果があると考えられる。

第Ⅳ部

結言

第7章

結論

7.1 本論文のまとめ

本論文では、人の位置情報履歴に対しマルコフ連鎖の確率モデルを用いて分析する手法について議論した。これは、都市域や公共空間での交通の最適化を目的とし、事前情報の利用できない位置情報だけで、人の流れを把握できる有益な情報をえることを想定したものである。

第I部では、この研究の前提条件や目的、定義、問題設定、本研究のアプローチについて述べた。

「第1章 本研究の背景と目的」では本研究の背景と位置づけについて、位置情報のデータに対して、個別の事前情報なく概要や特徴を把握するための分析手法が必要であり、本研究の目的をそこに設定した。

「第2章 測位データの特性と分析手順の類型」では、まず前提となる測位に関して定義、整理するとともに、動線データの分析方式の概略と各要素について整頓した。本章で定義した分析の手順では、最初に、位置情報の履歴を収集し、それに前処理を加えて測位の誤差などを低減しつつ、移動や滞留など状態遷移を意味する記号列に変換するなど行動のモデルに適した形に変換する。次に、人物行動モデルの抽出によって、人の行動を表現するモデル、例えばマルコフ連鎖のモデルに当てはめる。最後に、人の行動のモデルを利用者にわかりやすく提示して現状把握を促し、ひいては意思決定を支援する。

次に、第II部では、人の行き先の選択の確率モデルについて第2章で規定した手順にもとづき、事例を交えながら議論した。

「第3章 不特定多数の人の行き先選択モデル」では、予測推定などに適用できる方法として、状態遷移の履歴をいくつかのマルコフモデルに当てはめる方式を提案した。特に、従来の方法に比べて、人の行動パターンの分類を加味しつつ人の時間経過に伴う行動変容を適切に扱うことができるMAR-HMMを用いた方式提案した。また、本方式の効果について、シミュレーションにて生成されたデータ、実際の商業施設における歩行者行動履歴のデータ、アンケートにもとづく人の流れのトリップデータの3種類のデータを用いた実験を行い、有効性を確認した。これにより、歩行者の位置情報を収集できる環境が整えば、そこから価値を抽出して各種機能を実現することができると考えられる。

「第4章 特定人物の長期間行動分析」では、ある特定の人の位置情報を長きに渡って計測して蓄積

しておき、それを分析する方法について議論した。特に、不特定多数に対して用いられるマルコフモデルを援用するために、位置情報に移動滞留判定を加えて出発地目的地のクラスタリングを行うことで、GPS で得られた動線を状態遷移列に変換する方法を提案した。また、実験的に 2 人の人物の一年間の位置情報に提案方式を適用して分析したところ、不特定多数の場合と同様の分析が可能であることが確認できた。したがって、特定人物であっても長期間のデータを収集して本章で提案した方式を用いて分析すれば、不特定多数の場合と同様のさまざまな分析ができると考えられる。

第 II 部では、人の個性をうまく確率モデルとして扱うことが精度向上の鍵になっていたと思われる、時間が経つにつれて個性が顕在化するため、時間変化を扱うモデルが高い性能を発揮していた。人の移動にはパターンがあるものの、どこかにある平均との差異が積み重なり、個人としての予測が難しくなるようであった。

そこで第 III 部では、人の集団としての動きをモデル化し可視化することを対象とし、人の動きの全体像を可視化するため、人の密度分布計算について直前の分布を用い高速化する方法と、主動線を抽出する方法を提案した。全体像の可視化では人の個性は扱わなくともよいが、個々の人がマルコフ連鎖的に移動していることから、これらもマルコフ連鎖的なモデルを適用した方式である。

「第 5 章 人の密度分布の高速推定」では、移動体端末から収集された測位データの密度分布を高速に精度よく推定する方法として、変分ベイズ法にもとづき一部の測位データだけを利用して全体分布を推定する分布推定方式を提案した。これについて東京都市圏の人の流れデータを用いて評価したところ、従来に比べて約 100 倍の速度で分布が推定できた。さらに全データの半数を用いて生成された分布と残りのデータの比較による精度評価を行い、提案方式は従来方式と遜色ない精度が得られ、実際に生成される分布画像も従来と大きくは異ならないことが確認できた。

「第 6 章 主動線抽出」では、人の動きの概略を把握する方法として、主動線抽出処理を検討した。特に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、PFLOW データを用いて有用性を評価、その有効性が確認できた。

第 II 部でも、個人の移動のパターンには個性な豊富さがあるものの平均的な予測には成功していたが、第 III 部でも同様に、ある瞬間の移動の平均的なパターンとして扱うことは効果的であった。第 II 部のモデルも第 III 部の可視化も、無数にある人の個性を平均として理解可能にするが、「平均的」は必ずしもそれが正しいことを意味しない点に注意を要する。例えば、平均的な動きをつなぎ合わせてできる主動線であっても、実際にその通りに動いた人がいないこともありえる。つまり、あくまで全体を捉えるための指標であって、個人ごとの交通行動への理解は別途必要である。

とはいえ、事前の調査設計なく収集された位置情報に対し、全体的な傾向を理解するための情報抽出を行うという観点では、様々なマルコフ連鎖的確率モデルを適用することで、有益な情報が得られるケースはあると考えられる。マルコフ連鎖は、MAR-HMM や MMM のように拡張しやすく、事前情報なしでも個性等を折り込むような形にすることができる。このことから、事前の調査設計なく収集された位置情報だけでも、確率モデルの工夫により何らかの情報が引き出せる場合があることが確認できたといえる。また、さらなる確率モデルの工夫により、より多くの情報が得られるようになることも期待される。

7.2 結論と今後の課題

本論文の第 I 部では、施設や交通機関等の意思決定を支援するための情報抽出を想定し、不特定多数の人の位置情報から意思決定に有益な概略的な情報を抽出するための移動履歴のデータマイニングの手法の拡充が必要であることを示した。その上で、移動に影響する要素の推定と移動そのものの要約について、先行研究における課題を解決する数理モデルをたてることを目的と設定した。この観点から本論文で提案した分析手法をまとめたものを表 7.1 に示す。第 1 章では、人の移動にまつわる意思決定に必要な基本的な情報として、対象地域内の移動する人たちにに関する 5W1H、すなわち、どんな人たちが (Who)、いつ (When)、どこに (Where) いて、どこへ (What)、どう移動しようとしているか (How) を挙げた。

第 II 部では主に人の個性、つまり「どんな人たちが (Who)」に着目したモデルを提案した。その目的で重要なのが、不特定多数の人の移動からパターンを見出すことであるが、先行研究では単一のモデルですべての人の行動を表現している点が課題であった。本論文では、混合マルコフ連鎖、および、混合自己回帰隠れマルコフ連鎖のモデルを提案し、不特定多数の人の移動履歴のみから、人の個性を加味した複数のモデルを抽出することができることを、展示会来訪者のシミュレーション、ゲームイベント参加者の動線、人の流れデータ、個人の長期間行動の履歴データの 4 種類を用いて検証した。その結果、移動のパターンや個人ごとの行き先予測という形の情報が得られることが確認できた。

第 III 部では、人の個性の情報は集約し、移動そのものに関する統計量の変動に焦点を当てた確率モデルとその応用を示した。どこに (Where) 人がいるかを時間ごとに (When) 可視化したのが人の密度分布であり、どこへ (What) どうやって (How) 移動するかを可視化したのが主動線であった。人の密度分布を得る方法として知られているカーネル密度推定法では、人の数が多くなったときにそれに比例して処理時間がかかるという課題があったが、人の密度分布を混合ガウス分布として表現し、直前の分布を事前分布としてベイズ推定にかけることにより、人の位置情報の全数を使わなくとも密度分布が推定でき、精度と速度が両立できることを示した。地点間の移動に関しては、地点間移動に着目した主動線の抽出が必要であった。本論文では、マルコフ連鎖によって得られる確率を時間帯で変化させたモデルである準マルコフ連鎖を複数混合した混合準マルコフ連鎖を提案し、複数の主動線を生成できることを確認した。以上から、人の移動の平均にマルコフ連鎖的なモデルを拡張して適用することによって、人の移動に関する概略的な情報を抽出でき、すくなくとも基本的な情報である 5W1H について、意思決定のための状況把握がしやすくなったと考えられる。

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処理の高速化方式については、今後の課題である。ただし、現在、多数のコンピュータを並列利用するさまざまなビッグデータ分析プラットフォームが用いられ始めており、それらは地理空間情報にも適用され始めた。本研究で提案した方式の大半は並列性が高く、これらを用いて高速化していくことが容易である。

表 7.1 本論文にて提案された分析と今後の課題

部	章	抽出する情報	確率モデル	用途
第 II 部 推定・予測	第 3 章, 第 4 章	個人ごとの移動パターン (Who,How,Why)	混合マルコフ連鎖, 自己回帰混合マルコフ連鎖	パターン可視化、行き先予測 (情報配信など)
第 III 部 要約	第 5 章	人がいつどこにどのくらいいるか (When, Where)	混合ガウス分布	ヒートマップ表示
	第 6 章	どこへどう移動するか (What,How)	混合準マルコフ連鎖	主動線表示

表 7.2 確率モデルの応用

抽出する情報	応用	想定される適用分野
移動に影響する要素や移動する 人の特徴の推定、次の行き先予測 (Who, How, Why)	人の行き先予測	予測配信、LBS
	問題検出、問題箇所の特定、異常検知	エレベータ、空調等の機器制御 防犯、事故検知
人の数や移動そのものの傾向の要約 (Where, When, What, How)	指標算出	空間評価
	人の数、密度のヒートマップ生成	概況把握、混雑具合がわかり、過剰なら対策
	人の動きの傾向	主動線抽出、概況把握、案内表示の仕方を変えるなどの対応ができる
今後の課題	問題箇所の特定	警報 (アラート)
	問題の解決策	自動制御

7.3 本研究成果の応用と今後の展開

本研究の成果である各種モデルには、様々な応用が考えられる。表 7.2 にその例を示す。たとえば行き先の予測は、情報を配信する際に、これからその人が行きそうな地点に関する情報を配信するなどの使い方が出来る。また、多くの人が同じ地点に行くと予測できる場合には、そこが混雑する箇所であるという予測もできる。また、エネルギー消費の低減 [35] や効率化などにも利用できる。あるいは、行き先の推薦など [36][37] も考えられる。

さらに応用を広げるためには、抽出する情報の多様化が重要である。本研究で抽出可能になった情報はごく一部にすぎない。たとえば、人の属性の推定補完、場所や気候など環境の情報補完、人の分布の予測や推定など、本研究の成果をもとにさらなる発展が想定できる。また、実用的な観点からは、いかにして意思決定を支援するかの方法論が求められる。現時点では、位置情報から意味のある

特徴（統計量）を抽出しているにすぎない。実用上は、これをもとに問題箇所（対策すべき場所）の自動抽出や要調査項目の選定を支援、位置以外の情報を追加した分析などが必要となる。

データは多ければ多いほど読み解くのが困難になるため、今後、ビッグデータ分析プラットフォームを用いねばならないほどの規模のデータが集まった場合、本研究のような事前情報なく概要、特徴を把握するための技術は重要性を増してくる。これらが、今後のより良い社会基盤づくりの一助となれば幸いである。

付録 A

MMM に対する EM アルゴリズム

A.1 EM アルゴリズムの導出

以下では、EM 法による混合マルコフモデルのパラメータ推定の導出を示す。混合マルコフモデルの D に対する対数尤度 $L(D)$ は、以下の通りである。

$$L(D) = \sum_n \log P(d_n | \{\pi^{(k)}\}, \{\omega_{\mu\nu}^{(k)}\}) = \sum_n \log \left(\sum_k^K \pi^{(k)} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) \right) \quad (\text{A.1})$$

これを制約条件 $\sum_k \pi^{(k)} = 1$, $\sum_{\mu} \omega_{\mu\nu}^{(k)} = 1$ のもとで最大化する。Lagrange の未定定数として $\pi^{(k)}$ の制約を表す α , および $\omega_{\mu\nu}^{(k)}$ の制約を表す $\beta_{\mu}^{(k)}$ を導入すると対数尤度 $L'(D)$ は以下ようになる。

$$L'(D) = L(D) + \alpha \left(\sum_k \pi^{(k)} - 1 \right) + \sum_{k,\mu} \beta_{\mu}^{(k)} \left(\sum_{\nu} \omega_{\mu\nu}^{(k)} - 1 \right) \quad (\text{A.2})$$

L' が最大となるのは微分係数が 0 のときなので、 π の偏微分に関する方程式は

$$\frac{\partial L(D)}{\partial \pi^{(k)}} = \sum_n \frac{\psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})} + \alpha = 0 \quad (\text{A.3})$$

となる。この両辺に $\pi^{(k)}$ をかけ k について和をとると $\alpha = -N$ となる。この結果を (A.3) に代入して両辺に $\pi^{(k)}$ をかけると

$$\pi^{(k)} = \frac{1}{N} \sum_n \frac{\pi^{(k)} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})} \equiv \frac{1}{N} \sum_n \gamma^{(k)}(d_n) \quad (\text{A.4})$$

が得られる。次に、(A.2) に関し $\omega_{\mu\nu}^{(k)}$ の微分係数が 0 となる条件の方程式は、

$$\frac{\partial L(D)}{\partial \omega_{\mu\nu}^{(k)}} = \sum_n \frac{\pi^{(k)} \frac{\partial}{\partial \omega_{\mu\nu}^{(k)}} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})} + \beta_{\mu}^{(k)} = 0 \quad (\text{A.5})$$

となる．この方程式を解くため， $\frac{\partial}{\partial \omega_{\mu\nu}^{(k)}} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\})$ を求める．軌跡 d_n において状態 μ から状態 ν への遷移が起こった回数を $c(d_n, \mu \rightarrow \nu)$ とすると，

$$\psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) = \prod_{\mu, \nu} \omega_{\mu\nu}^{(k) c(d_n, \mu \rightarrow \nu)} \quad (\text{A.6})$$

$$\frac{\partial}{\partial \omega_{\mu\nu}^{(k)}} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) = \frac{c(d_n, \mu \rightarrow \nu)}{\omega_{\mu\nu}^{(k)}} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) \quad (\text{A.7})$$

これを用いて (A.5) を書き換え、両辺に $\omega_{\mu\nu}^{(k)}$ をかけて ν について和をとると，

$$\beta_{\mu}^{(k)} = - \sum_{n, \nu} c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n) \quad (\text{A.8})$$

が得られる．(A.5) にこれを代入し $\omega_{\mu\nu}^{(k)}$ について解くと

$$\omega_{\mu\nu}^{(k)} = \frac{\sum_n c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)}{\sum_{n, \nu} c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)} \quad (\text{A.9})$$

となる．

A.2 局所解への対策

上述の EM アルゴリズムは、最急降下法的なアルゴリズムであり、単純に実行しても局所解で収束してしまい、最尤解に到達しないことが多い。これを回避するには、収束した後にランダムなノイズを加えて再計算することを繰り返すことが効果的である。

この種の局所解を避ける技法としては、シミュレーテッドアニーリング (SA) が知られている。SA では、ある局所解が得られたら、それに近いランダムな他の解を生成し、そちらのほうがより評価がよければ遷移し、評価が悪い場合でもある確率で遷移するようにし、局所解に陥らないようにしている。繰り返すうちに評価が悪い解への遷移確率を減らすことで、最適解に収束させる。これを上記の EM アルゴリズムに適用すると以下ようになる。まず適当な初期値で EM アルゴリズムを実行し、それが収束したらそのモデルパラメータ θ と尤度 l を評価しておく。次にモデルパラメータに乱数で遷動を加え、再度 EM アルゴリズム計算して収束したら、そのモデルパラメータ θ' の尤度 l' を評価する。もしその尤度 l' が前回の尤度 l より低ければ θ' は棄却して、モデルパラメータ θ に乱数で遷動を加えるところからやり直す。もし尤度 l' が前回の尤度 l より高いならそれを採択とし、新たなモデルパラメータ θ' を起点として、上記の乱数で遷動を加える手順を実行する。この手順を繰り返し行うことで徐々にモデルパラメータが最適な値へと収束する。この棄却の判定の際に、繰り返し初期の場合には尤度が低くても一定確率で採択するようにすることで、より局所解から脱出しやすくなる。

図 A.1 に 4 章のデータに対し混合数 $K=10$ として、10 通りの初期値 (10 種類の乱数シードでの計算) でモデル化した結果を示す。グラフの縦軸は前節の EM アルゴリズムによる収束計算の結果での学習データの対数尤度平均値を示し、横軸は EM アルゴリズムの計算の回数を示す。このグラフから読み取れるように、ランダムな初期値の対数尤度がいずれも -10.0 程度であるのに対し、繰り返す

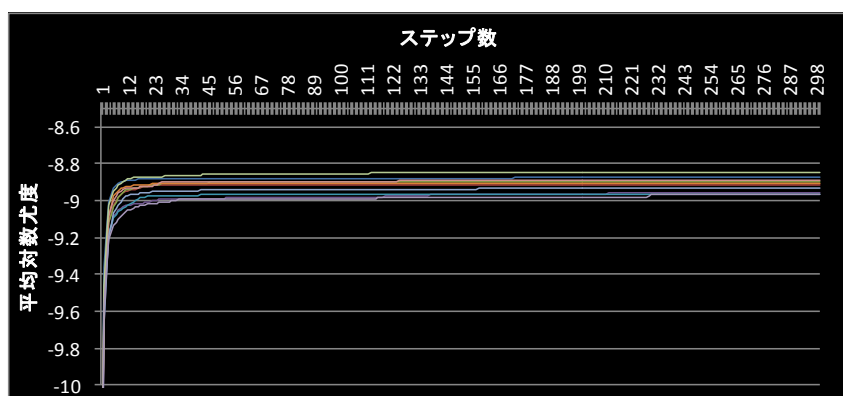


図 A.1 MMM の EM アルゴリズムの SA 適用による収束改善

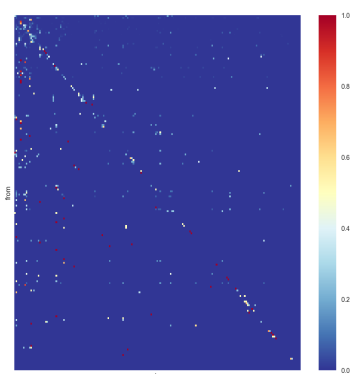


図 A.2 SMC の遷移行列

ことで-8.9 付近になっており、SA は最尤な解を得るという観点で効果があることがわかる。より安定した解を得るには、この例が示すように同様の計算を何度か行い、尤度の高いものを選ぶことが効果的である。

一つ注意すべき点としては、この安定性の評価は学習データの尤度に関するものであり、学習データの中にほとんどない地点の遷移確率は尤度への寄与が小さいという点がある。よって、サンプル数 f の少ない遷移の確率は安定しないが、発生頻度が小さいため全体としての結果には影響が少ない。図 A.2 にこの例での SMC の遷移行列を示す。この図は地点間の遷移確率をの行列を遷移確率に応じて色分けしたものであり、もともと全体としての遷移確率は疎、つまりある地点からの遷移先はかなり限定的であることがわかる。これを MMM でモデル化したときの混合要素ごとの遷移行列のうち2つを図 A.3 に示す。MMM によるモデル化はこの図のように遷移行列をより疎なものに分割することに相当しており、尤度の収束値にあるばらつきはどの遷移をどのグループに振り分けるかのばらつきを意味している。SA 等を用いて尤度がある程度収束した場合に残る尤度のばらつきは、めったに起きない遷移をどのグループに入れるかという選択の自由度に起因していることになり、全体としての結果（予測精度など）には大きな影響は与えないと考えられる。

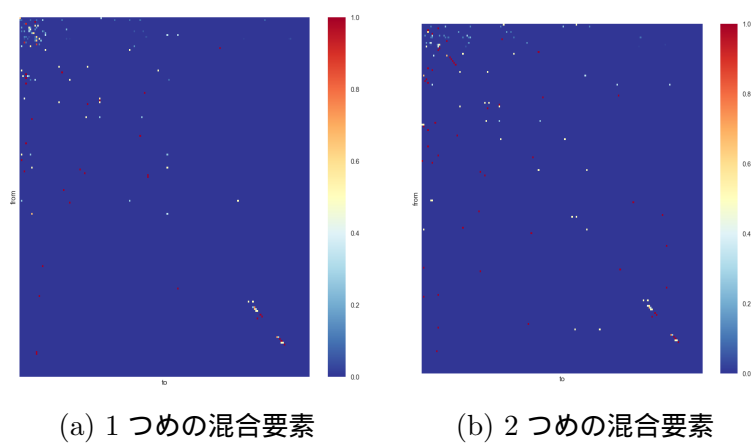


図 A.3 MMM の遷移行列

付録 B

MAR-HMM に対する BW アルゴリズム

B.1 EM アルゴリズムの導出

MAR-HMM は, AR-HMM の一種とみなせる. AR-HMM のパラメータは大きく分けて 3 種類ある. 一つは, 初期の内部状態の発生確率であり, 初期内部状態が m となる確率を π_m とおく. 次に, 内部状態間の遷移確率であり, 内部状態 l から内部状態 m への遷移確率を $A_{l,m}$ とおく. 最後に, 空間状態の発生確率であり, 直前の空間状態 μ , 内部状態 m のときに空間状態 ν となる確率を $\phi_{m,\mu,\nu}$ とおく. なお, 仮定する内部状態の種類を M 種類とする. また, 学習に用いる 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 $d_{n,t}$ は空間状態の種類数を K としたとき K 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 k のとき, $d_{n,t,k} = 1$ それ以外は $d_{n,t,k} = 0$ となるものとする. また, 各空間状態の長さを T_n とする.

HMM のパラメータを高速に推定する手法として, FB(Forward-Backward) 法が知られているが, この手法は ARHMM にも適用が可能である. FB 法は HMM に対し, EM(Expectation-Maximization) 法によるパラメータ推定を高速に行う手法である. EM 法は, 隠れ変数を含む確率モデルに対し, 最尤となるように隠れ変数とパラメータを決定する計算手法の一つである. 隠れ変数を含む確率モデルのパラメータ推定では, 隠れ変数とパラメータの両方を同時に推定しなければならない. EM 法は, 仮のパラメータのもとで隠れ変数の期待値を求める E ステップと, 隠れ変数がこの期待値であると仮定して最尤パラメータを求める M ステップを交互に繰り返し収束計算させることにより, これを実現する方法である. HMM では, 時間経過にしたがって変化する隠れ変数を扱っているため, E ステップでは各時刻における期待値をそれぞれ求めなければならない. それらの隠れ変数の発生確率はその直前の時刻における隠れ変数の値に依存するので, 隠れ変数の期待値を求めるためには, 直前の隠れ変数の発生確率も必要となる. すると, さらにその直前の隠れ変数の値も必要となり, 繰り返しの計算が必要になるが, この計算量は膨大である. FB 法はこの期待値計算が各時刻で重複する点に着目し, 一度計算した結果を保存しておくことにより, 重複計算を避ける方法である. ARHMM は一般的な HMM とは少々異なっているが, この FB 法がほぼそのままの形で適用できる.

ARHMM に対する FB 法のアルゴリズムを図 3.7 に示す. E ステップでは, まず各空間状態遷移列に対し, 漸化式を用いて α_n と β_n を再帰的に計算する. 式中の c_t は変数を計算機の計算可能な精

度内に抑えるためのスケーリング係数と呼ばれる値で、 $\alpha_{t,j}(d_n)$ の再帰計算の中で同時に計算できる量である。次に α_n と β_n を用いて $\gamma_{t,i}(d_n)$ と $\xi_{t,t+1,i,j}(d_n)$ を求める。これらが E ステップで求める期待値に相当する。M ステップではこの計算結果を用いてパラメータが更新され、これら E ステップと M ステップを繰り返られることによりモデルパラメータが定まる。

この計算において、 α_{ni} と β_{ni} の数は $M \sum T_n$ に比例しており、非常に多い。とくに長い空間状態列を扱うには、内部状態の数も多くしなければモデルが適切にならないと考えられる。一方で、モデルのパラメータ数も M に大きく依存し、特に内部状態間の遷移確率 $A_{i,j}$ は M^2 に比例する。よって、モデル化の精度を向上しようとして内部状態の種類を多くすると、さらに多数のパラメータを決定するために極端に多くの空間状態列を収集せねばならない。したがって、限られた時系列測位データから確率モデルを構築するには、内部状態の数を増大させずに歩行者行動を表現できるようにする方法が必要になる。

初期化

パラメータ $\{\pi_i\}$, $\{\phi_{i,\mu,\nu}\}$, $\{A_{i,j}\}$ 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 $\alpha_{t,j}(d_n)$ と $\beta_{t,i}(d_n)$ を再帰的に求める。

$$\begin{aligned}\alpha_{t,j}(d_n) &= \frac{1}{c_t(d_n)} \sum_{i,\mu,\nu} \alpha_{t-1,i}(d_n) A_{i,j} \phi_{j,\mu,\nu} d_{n,t,\mu} d_{n,t-1,\nu}, \\ \beta_{t,i}(d_n) &= \frac{1}{c_{t+1}(d_n)} \sum_{j,\mu,\nu} \beta_{t+1,j}(d_n) A_{i,j} \phi_{j,\mu,\nu} d_{n,t,\mu} d_{n,t-1,\nu} \\ \text{ただし } c_t(d_n) &= \sum_i \alpha_{t,i}(d_n), \alpha_{0,i}(d_n) = \pi_i, \beta_{T,i}(d_n) = 1\end{aligned}$$

次に、以下の式で負担率 $\gamma_{t,i}(d_n)$ と $\xi_{t,t+1,i,j}(d_n)$ を求める。

$$\begin{aligned}\gamma_{t,i}(d_n) &= \alpha_{t,i}(d_n) \beta_{t,i}(d_n) \\ \xi_{t,t+1,i,j}(d_n) &= \frac{\alpha_{t,i}(d_n) \beta_{t+1,j}(d_n)}{c_{t+1}(d_n)} \sum_{\mu,\nu} A_{i,j} \phi_{j,\mu,\nu} d_{n,t,\mu} d_{n,t-1,\nu}\end{aligned}$$

M ステップ

E ステップで求められた負担率を用い、以下の式でパラメータを更新する。

$$\pi_i = \frac{\sum_n \gamma_{0,i}(d_n)}{\sum_{n,i} \gamma_{0,i}(d_n)} \quad A_{i,j} = \frac{\sum_{n,t} \xi_{t,t+1,i,j}(d_n)}{\sum_{n,t,j} \xi_{t,t+1,i,j}(d_n)} \quad \phi_{i,\mu,\nu} = \frac{\sum_{n,t} \gamma_{t,i}(d_n) d_{n,t,\mu} d_{n,t-1,\nu}}{\sum_{n,t,\nu} \gamma_{t,i}(d_n) d_{n,t,\mu} d_{n,t-1,\nu}}$$

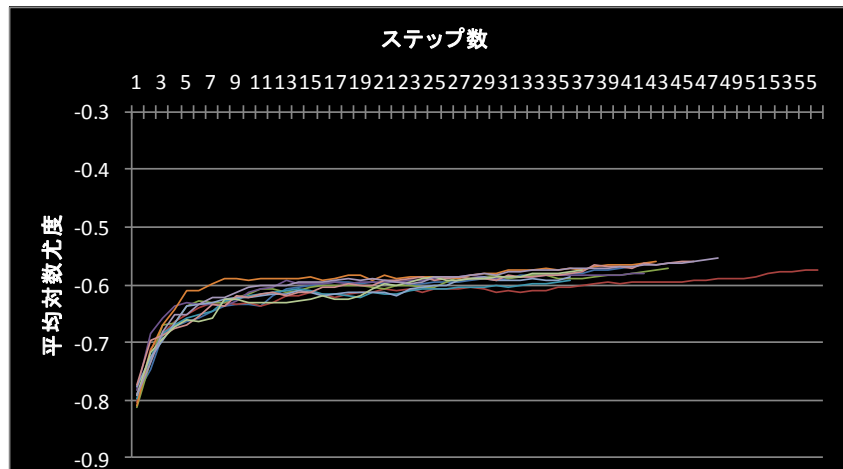


図 B.1 MMM の EM アルゴリズムの SA 適用による収束改善

収束計算

上記 E ステップと M ステップを収束するまで繰り返す。

B.2 局所解への対策

MAR-HMM の EM アルゴリズムも MMM のものと同様、局所解に陥ることが多い。その対策も同様であり、SA を用いた最適化が適する。図 B.1 は 4 章のデータに対し内部状態 20、混合数 10 とした場合の、10 通りの初期値 (10 種類の乱数シードでの計算) でモデル化した結果を示す。縦軸は前節の EM アルゴリズムによる収束計算の結果での学習データの対数尤度平均値を示し、横軸は EM アルゴリズムの計算の回数を示す。MAR-HMM ではパラメータ数が多いため、収束しにくい傾向にあるものの、SA による収束の傾向は初期値によらず同様であることがわかる。

付録 C

MSMM に対する EM アルゴリズム

C.1 局所解への対策

MSMM は MMM の遷移確率を時間帯ごとに割り当てることで主動線抽出に適するようにしたモデルであり、その収束性や結果の安定性については MMM に準ずると考えられる。したがって、SA 等による局所解への対策が必要である。

図 C.1 に 6 章のデータに対し、10 通りの初期値 (10 種類の乱数シードでの計算) を行った結果を示す。このグラフは、適当な初期値で EM アルゴリズムを実行し、それが収束したらモデルパラメータに乱数で揺動を加え、再度 EM アルゴリズム計算、という手順を繰り返し行うことでより最適な値へと収束させた過程を示すものである。縦軸は EM アルゴリズムによる収束計算の結果得られたモデルによって学習データの対数尤度平均値を示し、横軸は EM アルゴリズムの計算の回数を示す。

このグラフからは初期の乱数ではいずれも -18 程度だったのが、繰り返しにより 16.3 ~ 16.4 程度に収束していく様子が読み取れる。なお、揺動の加え方 (強さ) が小さいと不十分な解に収束することがあるため、十分強い揺動を加えつつ、十分な回数繰り返し計算を行うことが必要である。

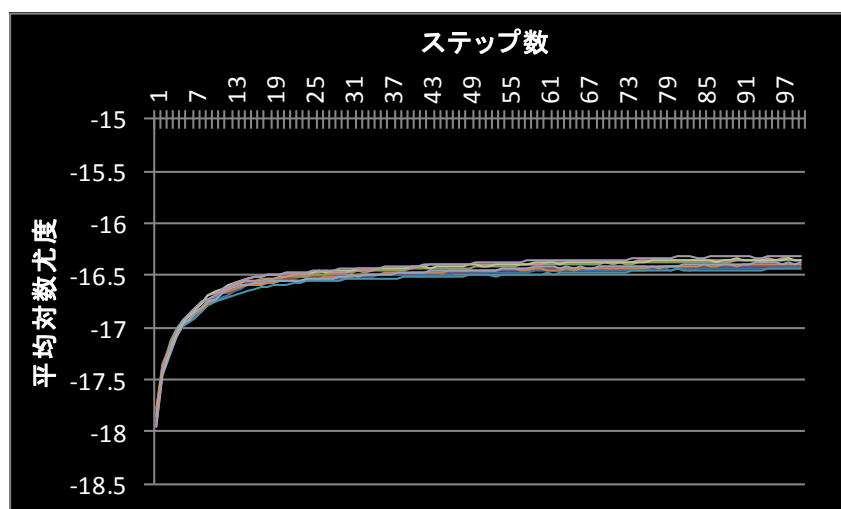


図 C.1 MMM の EM アルゴリズムの SA 適用による収束改善

参考文献

- [1] 辻川雅夫, 雅樂隆基, 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.
- [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] 寺本やえみ, 浅原彰規. 鏡像法による電界強度分布モデリングに基づく位置推定手法. 情報処理学会論文誌, Vol. 54, No. 1, pp. 116–124, jan 2013.
- [4] 日立ワイヤレスインフォベンチャーカンパニー. Hitachi Air Location™. <http://www.hitachi.co.jp/wirelessinfo/airlocation/index.html>, 2007.
- [5] クウジット株式会社. PlaceEngine®. <http://www.placeengine.com/>.
- [6] D. Manandhar, Seiya Kawaguchi, Masayuki Uchida, Makoto Ishii, and Hideyuki Tomohiro. IMES for mobile users. social implementation and experiments based on existing cellular phones for seamless positioning. *Proc. of the Int. Symposium on GPS/GNSS 2008*, Nov 2008.
- [7] 日立製作所. ニュースリリース:2月18日. <http://www.hitachi.co.jp/New/cnews/month/2009/02/0218.html>, 2009.
- [8] Yoko Hasegawa, Yoshihide Sekimoto, Takehiro Kashiya, and Hiroshi Kanasugi. Transportation melting pot dhaka: Road-link based traffic volume estimation from sparse cdr data. In *Proceedings of the First International Conference on IoT in Urban Space, URB-IOT '14*, pp. 105–107, ICST, Brussels, Belgium, Belgium, 2014. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [9] 一般財団法人道路交通情報通信システムセンター. VICS. <http://www.vics.or.jp/index1.html>.
- [10] 興梠正克, 大隈隆史, 蔵田武志. 歩行者ナビのための自蔵センサモジュールを用いた屋内測位システムとその評価. モバイル学会シンポジウム モバイル 08 予稿集, pp. 151–156, 2008.
- [11] 中村克行, 邵肖偉, 趙卉菁, 柴崎亮介. レーザスキャナを用いた歩容解析に基づく非定常歩行の認識. 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 127, No. 4, pp. 537–545, 2007.
- [12] 株式会社 ROAGroup. 日本の位置情報サービス (LBS) の動向と展望, Dec. 2007.
- [13] 日立製作所. ニュースリリース:2007年7月23日. <http://www.hitachi.co.jp/New/cnews/>

- month/2007/07/0723a.html, 2009.
- [14] 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.
 - [15] 橋本浩良, 河野友彦, 門間俊幸, 上坂克巳. 交通円滑化対策のためのプローブデータの分析方法に関する研究. 平成 22 年度国土交通省国土技術研究会, 2010.
 - [16] 丹下智之, 廣森聡仁, 梅津高朗, 山口弘純, 東野輝夫ほか. 車両プローブ情報に基づく信号待ち車列長変動の分析モデル. マルチメディア, 分散協調とモバイルシンポジウム 2014 論文集, Vol. 2014, pp. 1751–1759, 2014.
 - [17] 今井龍一, 井星雄貴, 中村俊之, 牧村和彦, 濱田俊一. 複数の動線データの組合せ分析によるバス停留所付近の走行改善の検討支援に関する研究. 土木学会論文集 D3 (土木計画学), Vol. 68, No. 5, pp. I.1287–I.1296, 2012.
 - [18] 井星雄貴, 今井龍一, 濱田俊一, 千葉尚, 牧村和彦. 複数の動線データを用いた道路整備の効果検証に関する基礎的研究. 土木計画学研究・講演集, Vol. 43, , 2011.
 - [19] 羽藤英二. モデルとデータ: プローブパーソン調査の課題と今後. 土木計画学研究・講演集, Vol. 27, p. 259, 2003.
 - [20] 熊本大学交通計画系研究グループ. 「スマくま」. <http://www.cps.kumamoto-u.ac.jp/smart-pt/>.
 - [21] 眞浦靖久, 塚田幸広, 井坪慎二. Pp データによる tx 開通に伴う交通行動への影響に関する基礎的分析. 土木計画学研究・講演集, Vol. 33, , 2005.
 - [22] 貞廣雅史, 堀口良太, 松本修一. 携帯 gps 測位ログデータを用いた交通行動調査に関する基礎的研究. 土木計画学研究講演集, Vol. 35, , 2007.
 - [23] 杉野勝敏, 矢野晋哉, 羽藤英二, 朝倉康夫. プローブパーソン調査を用いた観光行動分析. 土木計画学研究・講演集, Vol. 32, p. 294, 2005.
 - [24] Yasuo Asakura and Eiji Hato. Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C: Emerging Technologies*, Vol. 12, No. 3-4, pp. 273 – 291, 2004.
 - [25] 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.
 - [26] 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, 2008-05-08.
 - [27] 東京都市圏交通協議会事務局. パーソントリップ調査とは. <http://www.tokyo-pt.jp/person/>.
 - [28] 北村隆一. 交通需要予測の課題. 土木学会論文集, Vol. 1996, No. 530, pp. 17–30, 1996.
 - [29] Japan International Cooperation Agency. 交通需要予測パッケージ STRADA. <http://www.intel-tech.co.jp/strada/strada.html>.
 - [30] 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. <http://pflow.csis.u-tokyo.ac.jp/index-j.html>.

- [31] 交通計画システム研究会. 都市の交通計画 総合交通体系調査と交通需要の分析・予測. 共立出版, 2006.
- [32] 門間俊幸, 橋本浩良, 河野友彦, 上坂克巳. 常時観測データを用いた新たな道路サービスレベル指標に関する研究. 土木計画学研究・講演集, Vol. 43, , 2011.
- [33] 井坪慎二, 羽藤英二, 中嶋康博. 情報技術の活用による交通行動調査の効率化・高度化に関する研究. 土木計画学研究・講演集, Vol. 31, , 2005.
- [34] 上坂克巳, 門間俊幸, 橋本浩良, 松本俊輔, 大脇鉄也. 道路交通調査の新たな展開~ 5 年に 1 度から 365 日 24 時間へ. 土木計画学研究・講演集, Vol. 43, , 2011.
- [35] 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.
- [36] 田中宏平, 寺田努, 岸野泰恵, 西尾章治郎. カーナビゲーションシステムのための走行履歴と運転状況を用いた車両目的地予測手法. 情報処理学会論文誌, Vol. 51, No. 1, pp. 141–151, 2010-01-15.
- [37] 宮下浩一, 寺田努, 田中宏平, 西尾章治郎. 目的予測型カーナビゲーションシステムのためのマップマッチング手法. 情報処理学会論文誌, Vol. 50, No. 1, pp. 75–86, 2009-01-15.
- [38] 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.
- [39] 松尾豊, 岡崎直観, 中村嘉志, 西村拓一, 橋田浩一, 中島秀之. 位置履歴からのユーザ属性の推定 (モバイルコンピューティング). 情報処理学会論文誌, Vol. 48, No. 6, pp. 2106–2117, 20070615.
- [40] 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.
- [41] Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC, 1986.
- [42] Goce Trajcevski, Hui Ding, Peter Scheuermann, Roberto Tamassia, and Dennis Vaccaro. Dynamics-aware similarity of moving objects trajectories. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, pp. 1–8, New York, NY, USA, 2007. ACM.
- [43] J. Nascimento, M. Figueiredo, and J. S. Marques. Segmentation and classification of human activities. In *Workshop on Human Activity Recognition and Modelling*, 2005.
- [44] 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.
- [45] Jae-Gil Lee, Jiawei Han, and Kyu-Young Whang. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pp. 593–604, New York, NY, USA, 2007. ACM.
- [46] Darya Chudova, Scott Gaffney, Eric Mjolsness, and Padhraic Smyth. Translation-invariant mixture models for curve clustering. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 79–88, New York, NY, USA, 2003. ACM.

- [47] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. *Proc. 20th Int. Conf. Very Large Data Bases*, Vol. 1215, pp. 487–499, 1994.
- [48] 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.
- [49] 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.
- [50] 若宮翔子, 李龍, 角谷和俊. 位置ベース SNS を通した群衆の移動経験に基づく都市空間の近接性分析. 情報処理学会論文誌. データベース, Vol. 6, No. 3, pp. 159–176, jun 2013.
- [51] 角田史記, 加藤学, 大塚理恵子, 助田浩子, 大関一博. 交通系 IC カードを利用した鉄道輸送障害時の影響を定量化する方法の研究. 情報処理学会論文誌. データベース, Vol. 6, No. 3, pp. 187–196, jun 2013.
- [52] Open Geospatial Consortium. OGC(R) Moving Features. <http://www.opengeospatial.org/standards/movingfeatures>, 2015.
- [53] 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, 2011-01-15.
- [54] 雨宮有, 齋藤知範, 雨宮護, 島田貴仁, 原田豊. 時刻と位置の連続記録から滞留場所と移動経路を抽出する時空間的アルゴリズムの開発. 地理情報システム学会 講演論文集, 第 17 巻, pp. 75–79, 2008.
- [55] 堀口良太, 長岡亨, 畑成年. Gps 携帯電話による大規模パーソンプローブ調査のためのトリップ情報抽出手法に関する研究. 土木計画学研究 講演集, Vol. 33, , 2006.
- [56] Akinori Asahara, Akiko Sato, and Kishiko Maruyama. Evaluation of trajectory clustering based on information criteria for human activity analysis. *Mobile Data Management, IEEE International Conference on*, Vol. 0, pp. 329–337, 2009.
- [57] 日本アイ・ビー・エム株式会社. 2007 年 10 月 12 日プレスリリース. <http://www-06.ibm.com/jp/press/20071012001.html>, 2007.
- [58] 森川高行, 佐々木邦明. 主観的要因を考慮した非集計離散型選択モデル. 土木学会論文集, Vol. 1993, No. 470, pp. 115–124, 1993.
- [59] Christpher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.
- [60] Nam T. Nguyen, Dinh Q. Phung, Svetha Venkatesh, and Hung Bui. Learning and detecting activities from movement trajectories using the hierarchical hidden markov models. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, Vol. 2, pp. 955–960, 2005.
- [61] Jie Yang, Yangsheng Xu, and C.S. Chen. Human action learning via hidden markov model. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, Vol. 27, No. 1, pp. 34–44, January 1997.
- [62] Carsten Stig Poulsen. Mixed Markov and latent Markov modelling applied to brand choice behaviour. *Intern. Journal of Research in Marketing*, Vol. 7, pp. 5–19, 1990.

- [63] 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, 2012-01-15.
- [64] José G. Dias and Joren K. Vermunt. Latent class modeling of website user's search patterns: implications for online market segmentation. *Journal of Retailing and Consumer Services*, Vol. 14, pp. 359–348, 2007.
- [65] T.Hamada, T.Hagiwara, T.Teramoto, S.Morishita, M.Umetsu, and M.Ohgama. A tree-dimensional pedestrian-flow simulation for high-rising buildings. In *ACRI2008*, Vol. LNCS 5191, pp. 417–424. Springer, 2008.
- [66] 浜田朋之, 岡田祐子, 森下信, 梅津充幸, 大釜みち代. 公共空間における滞留行動評価のための大規模群集シミュレーション. 情報処理学会第72回全国大会公演予稿集 (第2分冊), pp. 83–84, 2010.
- [67] 株式会社電通. 2010年2月10日ニュースリリース. <http://www.dentsu.co.jp/news/release/2010/pdf/2010012-0210.pdf>.
- [68] 山田直治, 磯田佳徳, 南正輝, 森川博之. B-15-23 gps 搭載携帯電話を用いた移動経路履歴に基づく訪問地予測方式 (b-15. モバイルマルチメディア通信, 一般セッション). 電子情報通信学会総合大会講演論文集, Vol. 2010, No. 1, p. 639, 2010-03-02.
- [69] 北村隆一. 変動についての試行的考察. 土木計画学研究・論文集, Vol. 20, pp. 1–15, 2003.
- [70] 力石真, 藤原章正, 張峻屹, KW Axhausen. 6週間の交通日誌データを用いた交通行動の変動特性の分析 活動発生と出発時刻を例に. 土木計画学研究・論文集, Vol. 26, No. 3, pp. 447–455, 2009.
- [71] 北村隆一, 藤井聡, 山本俊行. 離散時間パネル調査の調査期間, 調査間隔, 標本数の最適化. 土木学会論文集, Vol. 2001, No. 681, pp. 13–23, 2001.
- [72] 関本義秀, 中村敏和, 増田祐介, 金杉洋. 大規模な gps 情報をもとにした東京都市圏における震災時の行動分析. 土木計画学研究・講演集, Vol. 45, No. 249, 2012.
- [73] 生形嘉良, 関本義秀. Gps データによる観光実態の分析. 土木計画学研究・講演集, Vol. 45, No. 370, 2012.
- [74] 矢島美寛, 平野敏弘. 時空間大規模データに対する統計的解析法. 統計数理, Vol. 60, No. 1, pp. 57–71, 2012.
- [75] 浅原彰規, 寺本やえみ, 丸山貴志子, 柴崎亮介. 時空間ネットワーク解析を用いた主動線抽出. 情報処理学会論文誌, Vol. 54, No. 1, pp. 135–145, jan 2013.
- [76] Akinori Asahara, Kishiko Maruyama, Akiko Sato, and Kouichi Seto. Pedestrian-movement prediction based on mixed markov-chain model. In *Proceedings of the 19th ACM SIGSPATIAL GIS 2011*, pp. 25–33, New York, NY, USA, 2011. ACM.
- [77] 貞広幸雄. 空間集計データにおける面補間法の推定精度評価. 都市計画, Vol. 49, No. 2, pp. 75–81, jun 2000.
- [78] Jeffrey S Simonoff. *Smoothing methods in statistics*. Springer, 1996.
- [79] 田中耕市. 1990年代の東京23区における都市密度変化と土地利用転換: 事務所建築物と集合住宅を対象として. 地学雑誌, Vol. 117, No. 2, pp. 479–490, apr 2008.

- [80] 伊庭幸人, 汪金芳, 田栗正章, 手塚集, 樺島祥介, 上田修功. 統計科学のフロンティア 11 計算統計 I 確率計算の新しい手法. 岩波書店, Tokyo, 2006.
- [81] C. M. ビショップ. パターン認識と機械学習 (下) ベイズ理論による統計的予測, pp. 187–195. シュプリンガー・ジャパン, New York, 2008.
- [82] 国土地理院. 国土基盤地図情報 1 / 2 5 0 0 0. <http://www.gsi.go.jp/kiban/index.html>.
- [83] Walter L. Smith. Regenerative stochastic processes. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, Vol. 232, No. 1188, pp. 6–31, 1955.
- [84] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on*, Vol. 13, No. 2, pp. 260 – 269, April 1967.