

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

浅原 彰規^{1,a)} 佐藤 暁子¹ 丸山 貴志子¹

受付日 2011年3月30日, 採録日 2011年10月3日

概要: 本研究では, ある領域を訪れた多数の過去の歩行者の測位データから新たな歩行者の行き先を予測する方式として, 混合自己回帰隠れマルコフモデルを用いた方式を提案する. 本方式は従来の混合マルコフモデルに基づく方式に歩行者の内部状態の時間変化を加味した方式である. 提案方式の評価のため, 商業施設来店者の測位データに提案方式を適用し, 来店者が次に行く地点を予測する実験を行った. その結果, 従来 70%程度であった予測精度が, 提案方式では最大で 80%以上になることを確認した.

キーワード: 動線解析, 地理空間情報, 空間データマイニング, 屋内空間情報

Autoregressive Hidden Markov Model for Pedestrian-movements Prediction

AKINORI ASAHARA^{1,a)} AKIKO SATO¹ KISHIKO MARUYAMA¹

Received: March 30, 2011, Accepted: October 3, 2011

Abstract: In order to predict a visitor's movements by past visitors' trajectories, a method using Autoregressive hidden Markov model is proposed in this research. The proposed method is an improved conventional method to take into account temporal transition of pedestrians' internal states. To evaluate the method, an experiment was performed by using actual tracking-data of visitors in a shopping mall. As the result, a prediction rate, which marked 70% by the conventional method, is improved to 80% in maximum by the proposed method.

Keywords: trajectory analysis, geospatial information, spacial data mining, indoor spacial data

1. はじめに

1.1 背景と研究の目的

本研究では, ある領域を訪れた過去の不特定多数の歩行者の位置情報を用いて, 新たにそこを訪れた歩行者が次に行くであろう地点を予測する方式を提案する.

特定の歩行者の位置情報を得るために, 歩行者の現在位置を計測し座標値に変換する機能を持つ測位装置を用いる. 従来の屋外で用いられる GPS (Global Positioning System) 測位に加え, 無線 LAN を用いた測位 [1], [2] や屋内設置型の GPS 発信器 [3] による測位, さらにカメラやレーザなどを用いた測位など屋内でも利用できる測位方式

の研究開発がさかんに進められている [4]. 本研究では, 測位装置によって得られる歩行者の位置と, 位置が計測された時刻を対応付けたデータを測位データと呼ぶ.

測位データは, 近年になってさまざまなサービスに用いられるようになってきた. たとえば, 位置に基づき情報を配信するサービスや経路案内を行うサービスが知られている [5]. このような測位データを用いたサービスが普及するに従って, 歩行者の測位データを得る機会は増大している. このように取得された測位データを蓄積すると, そこから非常に多くの情報を取り出せると考えられる. たとえば, 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, 1人の測位データを蓄積し続ければその人の特徴が, それぞれ含まれている. 蓄積された測位データからこれらの情報を引き出すことにより, さまざまな価値を得られると考え

¹ 日立製作所
Hitachi, Ltd., Central Research Laboratory, Kokubunji,
Tokyo 185-8601, Japan

^{a)} akinori.asahara.bq@hitachi.com

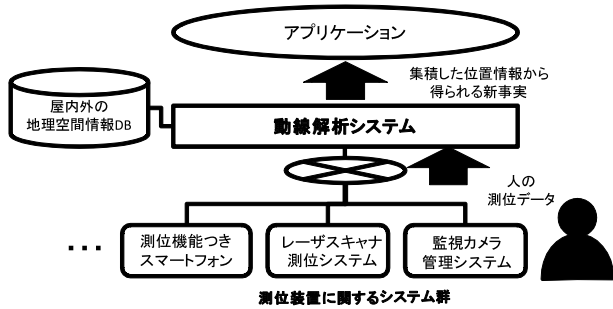


図 1 動線解析を用いたサービスシステム

Fig. 1 Service system with a trajectory analysis.

られる。このような、蓄えられた測位データの集合から新たな情報を引き出す技術を、本研究では動線解析と呼ぶ。ここで、動線とは歩行者が移動した連続的な位置の軌跡を表し、時系列に得られる測位データは動線をサンプリングした結果であるといえる。

図 1 に、動線解析を用いたサービスシステムの構成を示す。本システムでは、測位装置は順次、人の測位データを送付する。動線解析システムは、測位データを蓄積し、地図情報と組み合わせる新たな意味を抽出する。本システムによると、たとえば商業施設の施設管理者向けに、当該施設内の歩行者の平均行動パターン [6] や歩行者行動予測の提示 [7]、例外行動の検出 [8]、当該施設の店舗設備の稼働率定量化、店員配置の推薦など、意志決定を支援する情報提示ができる。また、その他の機器と連動することにより、たと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 [9] などにも利用できる。

今回、特定の領域内の歩行者の測位データを対象とし、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法について検討した。ここでいう状態遷移とは、たとえば「ある地点に滞留してから次のある地点へ移動する」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。ある領域に訪れた歩行者の状態遷移を確率モデルによって表現することにより、当該領域に訪れる多数の歩行者の移動の傾向を表現できる。歩行者の行動を表現するのにより適切なモデルを選べば、より正確に歩行者の移動を予測できるため、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る。本研究では、時間経過による状態の変化を導入したモデルとして、混合自己回帰隠れマルコフモデルを提案する。また、確率モデルの妥当性を評価する基準として行き先の予測性能を採用し、実測された歩行者の測位データを用いた実験的な評価を行った。

1.2 関連研究

歩行者の時系列測位データを用いた行き先予測の先行研究としては、歩行者の状態遷移を混合マルコフモデルで表現する方式が提案されている [10]。混合マルコフモデルは

高い精度で歩行者の行き先を予測できるが、時間的な内部状態の変化を扱わないモデルであるために困難を生じていた。たとえば、行きと帰りの区別がつかず、入場直後に出入口付近にいる歩行者を帰宅間際と誤認してしまい、結果として歩行者の回遊時間が長くなると予測精度が低下するという問題があった。

一方、本研究とは異なる対象の測位データに関する先行研究もある。田中らはカーナビゲーションシステムの時系列測位データを収集蓄積し、車両の行き先(目的地)を予測する方式について発表している [11], [12]。また、山田らによる携帯電話で収集した時系列測位データに基づく行動予測の研究も知られている [13]。これらの研究は、対象を特定個人に限定し、その人の時系列測位データ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研究である。それに対し、本研究では特定個人の時系列測位データを収集する期間はごく短くてもよい代わりに、多人数の時系列測位データを収集するようにした点が異なる。

本研究とは異なる構成のトラッキングに関する先行研究として、カメラやレーザレンジファインダなどを用いた例もある [14]。これらの方式は、通常、カメラやレーザの視野に入っている人物などの追跡のための方式であり、本研究のように特定施設内での歩行者の動線から意味を抽出するというようなものではない。

2. 歩行者の行き先予測とその課題

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、表 1 に示すように、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。この 1 人の歩行者の測位データを時間順に並べた時系列測位データは、クラスタリングなどの処理によって、離散的な状態遷移列へと変換できる。たとえば、文献 [9] では k-means 法により空間領域を分割し、滞留箇所に関する状態遷移が得られている。このようにして得られた離散的な時系列状態遷移列に対して統計処理を加えることにより、空間の利用状況を把握するだけでなく、典型的な行動パターン抽出するなど新たな情報を抽出できる。

本研究では、この状態遷移列を適切に近似できる確率モデルを生成することを目的とする。ここでいう確率モデルとは、分析対象となる歩行者の状態遷移を表現したモデルであり、歩行者の状態遷移確率の算定ができるモデルである。この確率モデルにより、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、たと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、

表 1 測位データの構成
Table 1 Structure of positioning-data.

| 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. |

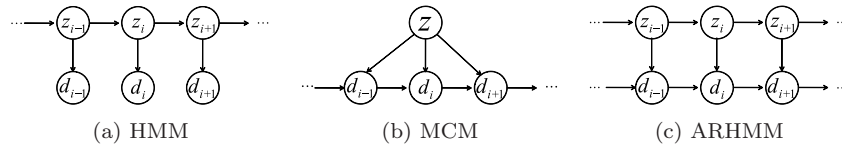


図 2 歩行者行動を表現する確率モデル

Fig. 2 Probabilistic models to represent pedestrian behaviors.

歩行者がおかれた状況に基づき行動を決定し、行動に移すまでのプロセスを表すのに適した確率モデルを選択しなければならない。そのため、各状態の実空間上の距離や経過時間、周囲にある物体など、実空間の特質を適切に取り込んだモデルが必要となる。考えられる歩行者行動のグラフィカルモデルを図 2 に示す。グラフ中の歩行者の観測可能な状態系列 $\{d_i\}$ 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記号列である。歩行者は、外部から観測できない内部的な状態変数 z_i (たとえば、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する。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状態 z_i を内部状態と呼ぶ。

一般的に内部状態 z_i のような観測できない変数は隠れ変数などと呼ばれ、観測可能な量から隠れ変数を推定する方式が知られている [15]。このような隠れ変数を取り入れたモデルとして、図 2(a) に示した隠れマルコフモデル (HMM; Hidden Markov Model) が知られている。HMM では隠れ変数の遷移のみに従属して観測される状態列が決まる。この HMM を用いた動線解析の例はいくつか知られており、事前の学習を行うことにより例外行動の検出が可能であるなどの報告がなされている [7], [8], [16], [17]。HMM は歩行者の空間状態は直前の空間状態に依存しないモデルであるが、実際には、直前の空間状態による歩行者の行動への影響は大きい。なぜなら、歩行者は空間的に隣接している地点にしか行けないという制約があるからである。また、歩行者が近い方の出口から出る、あるいは歩行者がたまたま目についた店舗へ入るなど、歩行者の現在地は意志決定にも影響する。

そこで、混合マルコフモデル (MCM; Markov Chain Mixture) [18] を適用した例 [10] がある。このモデルの構造を図 2 (b) に示す。MCM 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。MCM 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、空間状態間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。したがって、複数のマルコフモデルを重ね合わせたモデルであると見なせる。MCM 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が計算できる。歩行者の内部状態がほとんど変化しないと仮定できる場合や、歩行者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、MCM による近似は妥当である。しかし、内部状態の時間的変化を取り入れるには適さないという欠点がある。

そこで図 2 (c) のモデルを用いることを考える。このモデルでは、空間状態 d_{i+1} は d_i と z_i に依存する確率 $P(d_{i+1}|d_i, z_i)$ に基づく確率過程により決定される。このモデルは、自己回帰隠れマルコフモデル (ARHMM; AutoRegressive Hidden Markov Model) と呼ばれるモデルの一例になっている [15]。行き先の予測精度を高めるには、この ARHMM 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なる。以降では、この問題点について議論し、この問題点を回避した混合自己回帰隠れマルコフモデルを用いることを提案する。

3. 時間変化を取り入れた歩行者行動モデルの提案

3.1 自己回帰隠れマルコフモデル

ARHMM は、HMM の一種と見なせる。ARHMM のパラメータは大きく分けて 3 種類ある。1 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

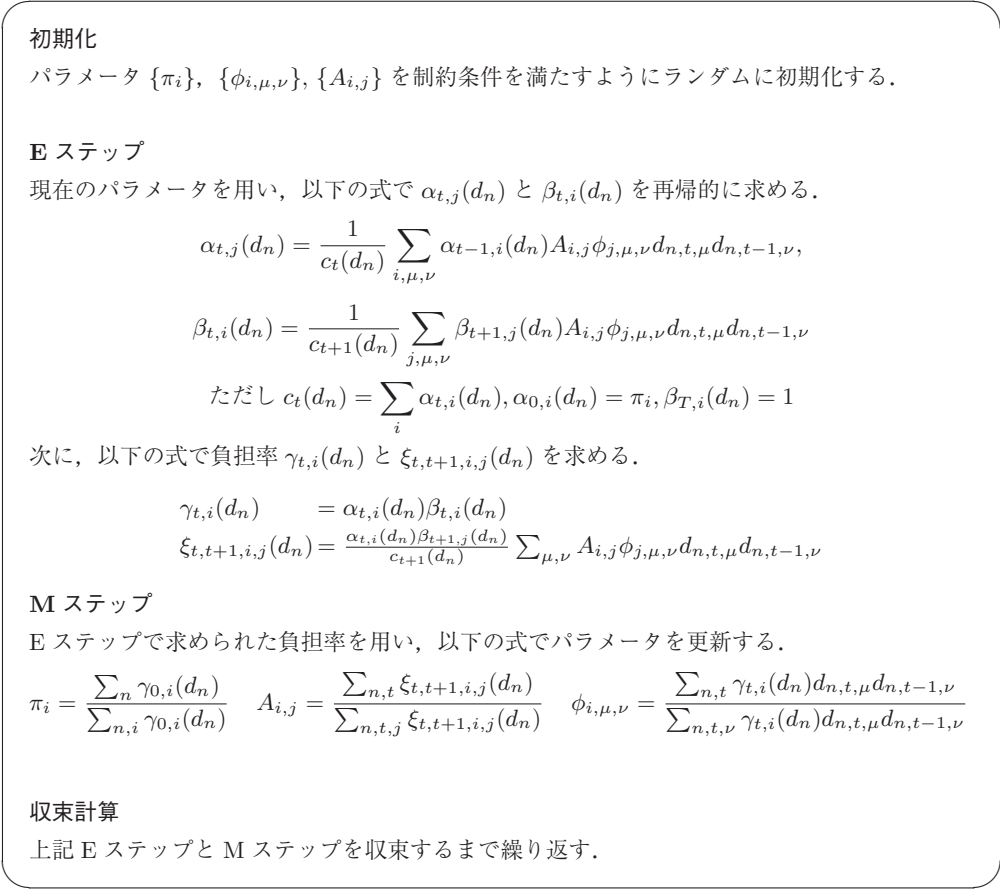


図 3 自己回帰混合マルコフモデルのパラメータ推定アルゴリズム

Fig. 3 Parameter estimation algorithm for ARHMM.

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 1 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, 時間経過に従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, 繰返し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, 1 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 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ルンク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 ステップを繰り返す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. 特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させ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しなければならない。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。

3.2 混合された自己回帰隠れマルコフモデル

ARHMM では非常に多くの内部状態を仮定しなければならず、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る。そこで、本研究では、内部状態間の遷移範囲に制限を設けることにより、計算するパラメータ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。しかし、歩行者行動を適切に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ともなう性能劣化はさほど大きくないはずである。それどころかサンプル数が限られた状況では、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。いい換えると、ARHMM を用いても 0 に近くなるはずのパラメータをあらかじめ 0 と固定し計算から除外することにより、性能を落とさずに計算すべきパラメータ数が削減できると考えられる。

図 4(a) に ARHMM 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ており、たとえば時刻 t で状態 1 のときに、時刻 $t+1$ で状態 2 に遷移しうることを時刻 t 状態 1 の円から時刻 $t+1$ 状態 2 の円への矢印で示している。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ことになる。本研究では、図 4(b) のように内部状態をいくつかのグループに分け、時間変化では当該グループ内にしか移動しないように制限する。この制限は、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いくつかの内部状態をとる、という仮定に基づく。この仮定は、MCM による人物行動のモデル化で用いたのと同質である。さらに簡単のため、グループの大きさを均一と仮定する。これを本研究では混合自己回帰隠れマルコフモデル (MARHMM; Mixed AutoRegressive Hidden

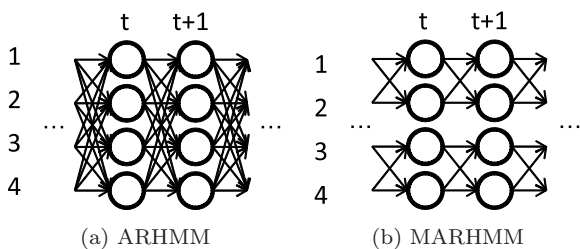


図 4 隠れ変数の状態遷移
Fig. 4 State-transitions of hidden variables.

Markov Model) と呼ぶ。MARHMM は、グループの数だけ ARHMM を混合させた形になっているので、以降ではグループの数を混合数と呼ぶ。

図 5 は ARHMM と MARHMM 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 ARHMM では M^2 個のパラメータが必要であるのに対し、(b) の MARHMM 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。混合数を L とすると (b) のパラメータ数は $L(M/L)^2$ 個になるので、MARHMM 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、MARHMM は ARHMM の一部のパラメータを 0 に固定した特別な形であると見なせる。この 0 に固定されるパラメータは、内部状態数と混合数によって定まる。したがって、ARHMM ではあらかじめ与えられるハイパーパラメータとして内部状態数があるが、MARHMM ではそれに加えて混合数もハイパーパラメータとなる。なお、内部状態数を 1 にすると図 5(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、MARHMM は MCM と等価になる。いい換えれば、MARHMM は MCM と ARHMM の中間的な性質のモデルといえる。

ここで、図 3 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 $A_{i,j} = 0$ となることが分かる。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、同様につねに $\xi_{t,t+1,i,j}(d_n) = 0$ となる $\xi_{t,t+1,i,j}(d_n)$ も存在するが、このようなつねに 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。

| | 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} | ... |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

(a) ARHMM

| | 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} | ... |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

(b) MARHMM

図 5 状態遷移行列 A_{ij} の比較
Fig. 5 Contrast of state-transition matrix A_{ij} .

なお、ARHMM 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、

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

のように求められる。

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、ARHMM は MARHMM や MCM 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、実用上は利用できるサンプル数が限られ、パラメータを決定するのが困難なため、十分な性能が得られないことが考えられる。たとえば歩行者の行動がグループごとに制限されていてすべての内部状態をとるわけではない場合について考える。パラメータ数に対して十分な数のサンプルを利用できる場合、それを用いて図 3 のパラメータ推定を行えば、自動的にパラメータの多くは 0 付近になり、高い予測性能が得られるはずである。ところが、パラメータ数に対して利用できるサンプル数が十分に多くない場合、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、一般に過学習と呼ばれる現象を引き起こす可能性がある。したがって、サンプル数が限られた状況ではパラメータ数を減らす必要があるが、単に内部状態数を減らしてパラメータを少なくしたのでは、モデルの表現力が低下して予測精度も低下してしまうという問題が発生してしまう。

MARHMM 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 1 つである。もし、MARHMM 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、0 付近になるべきパラメータを 0 に固定したのであれば、パラメータ数に対する相対的なサンプル数が多くなるため、結果として MARHMM の方が ARHMM よりも高い性能を得られる可能性があるのである。MARHMM 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、MCM は MARHMM 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。

4. 実験による評価

4.1 実験の目的と条件

前述のとおり、MARHMM が歩行者の行動を適切に表現するモデルであれば、ARHMM と MARHMM は類似した結果になると考えられる。逆に、MARHMM が歩行者の行動を適切に表現しないのであれば、性能を急激に劣化させてしまう恐れもある。この提案モデルの有効性は、実際に適用する問題に依存する。そこで、モデルの表現力を

表 2 実験の概要

Table 2 Abstract of the experiment.

| 項番 | 項目名 | 内容 |
|----|-------|---------------------------|
| 1 | データ源 | e 空間 Kansai 実験 [19] |
| 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 |

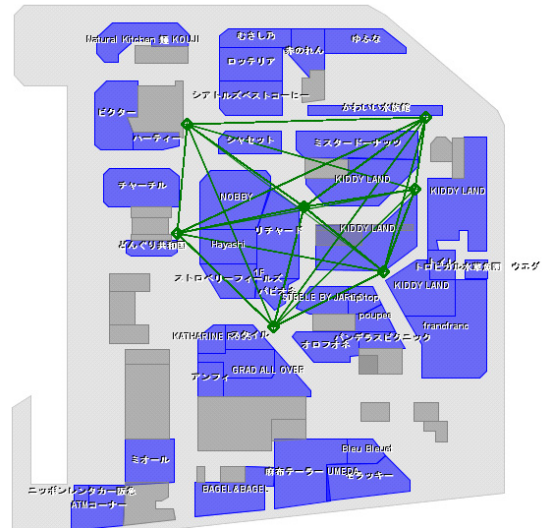


図 6 1 セット分の評価データ

Fig. 6 One dataset for the evaluation.

比較するため、実験で収集された測位データを用い行動予測性能を比較評価する。

表 2 に測位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。実験では本商業施設への来店者から参加者を募り、屋内 GPS [20] の受信機能を持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、端末には、受信した位置情報と時間帯ごとにあらかじめ定められたコンテンツを提示した。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った。なお、コンテンツを提示されるか否かはランダムに決まるようになっていた。最後に回収された端末からコンテンツ提示の履歴や位置情報を含むログを取り出し、参加者の行動を確認した。これにより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、この実証実験で収集された測位データを分析の対象とした。図 6 に 1 セット分の評価データを示す。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時系列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データとしている。

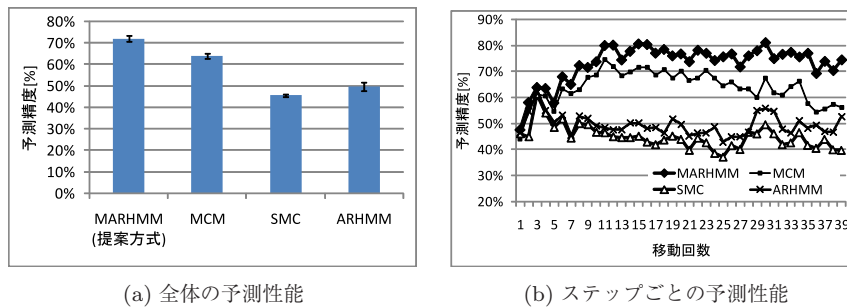


図 7 予測性能

Fig. 7 Prediction accuracies.

したがって、得られる測位データは屋内 GPS 電波の到達範囲にいる、という離散的な結果になる。そのため、図 6 に示された時系列測位データも離散的になっており、結果として多くの動線が重なってしまい把握が困難な図になっている。なお、計測は 5 秒に 1 回とした。前処理では、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など、計測誤差とおぼしき信号を除外した。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見なして処理を実行した。このことは、時系列測位データを「屋内 GPS 電波の到達範囲にいる」という状態の遷移に置き換えたことに相当し、たとえば GPS などによる連続的な座標の測位データを用いたとしても、空間領域を分割しどの領域にいるかという離散化をすることにより同様の結果が得られると考えられる。

今回、上記の時系列測位データ群に対し確率モデルを構築し、それを用いた行動予測の的中精度を評価した。今回用いた評価方法は 10 分割 Cross-Validation と呼ばれるものであり、測位データの時系列群をランダムに 10 分割し、うち 9 つをもとに確率モデルを構築し、残る 1 つに対して行動予測を行った。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。また、従来方式として、先行研究 [10] で用いられた混合のない単一のマルコフ連鎖モデル (SMC; Single Markov Chain)、混合マルコフモデル (MCM) に加え、自己回帰隠れマルコフモデル (ARHMM) を用いた行動予測精度を用い、提案方式である混合自己回帰隠れマルコフモデル (MARHMM) による行動予測精度と比較した。なお、HMM による予測性能は上記先行研究と同様に精度 10% 以下と著しく低い結果が得られたため、比較対象としては用いないこととした。

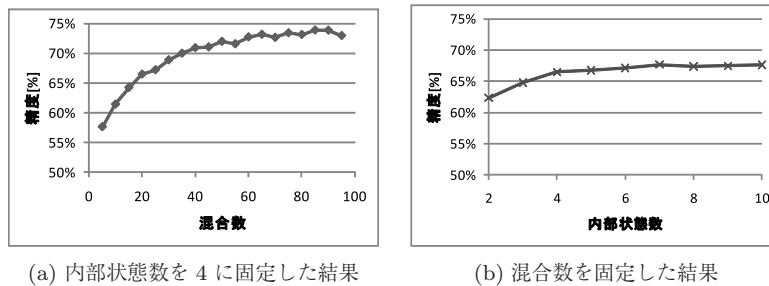
4.2 結果と考察

図 7 に予測性能を示す。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いることになり、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、各モデルの混合数や内部状態数は、実際に予測精度が最も高くなったものを用いており、MCM では混合数を 190、ARHMM では内部状態の数を 60 とし、提案方式の混合数は 95、それぞれの内部状態数は 4 としている。

図 7(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 MCM は 64.0% であった。また、ARHMM は 50.0% と SMC より高いものの、MCM 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 7(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。ARHMM も SMC と類似の振舞いをしており、ステップが進んでも SMC ほどには精度が低下していない。これは、モデルに時間変化を取り入れた効果であると考えられる。また、MCM 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 ARHMM 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 MCM 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、ARHMM と SMC の関係と同様、時間変化をモデルに取り込んだ効果であり、ステップ数が進むに従ってその効果が表れていると考えられる。

図 8 に提案方式のパラメータと精度との関係を示す。図 8(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 8(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分かる。また、内部状態の数も 4~6 くらいは必要と考えられる。



(a) 内部状態数を 4 に固定した結果 (b) 混合数を固定した結果

図 8 内部状態数と混合数の影響比較

Fig. 8 Accuracy contrasts by internal state number and model number.

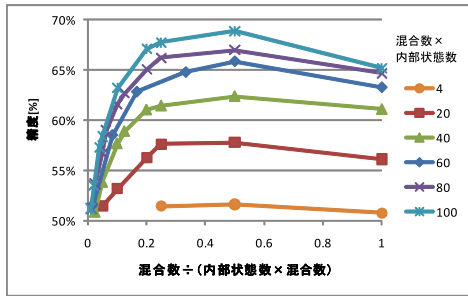


図 9 内部状態数 × 混合数を固定した場合の精度比較

Fig. 9 Accuracies when internal state number × model number are fixed.

図 9 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。たとえば、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 MARHMM の精度を示している。図 5 で示したとおり、混合数が 1、つまりグラフ左端の MARHMM は単なる ARHMM であり、混合数が大きな MARHMM 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 MARHMM は MCM に相当する。すなわち、図 9 のグラフは、グラフの左端を ARHMM、右端を MCM として、0 に固定するパラメータ数を変化させながら MARHMM 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 ARHMM 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、過学習が原因と考えられる。ARHMM 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 3,600、内部状態数 100 では 10,000 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ただか数十回程度であり、決定するパラメータ数にくらべて学習サンプル数が少ない。ところが MARHMM の場合、内部状態数 100 の ARHMM に相当する内部状態数 2、混合数 50 の MARHMM でもパラメータ数は 500 程度である。これは

内部状態数 100 の ARHMM のパラメータ数、10,000 程度にくらべて非常に少ない。つまり MARHMM では、決め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 ARHMM 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 ARHMM はパラメータ数 500 程度の MARHMM よりも精度が低い。これは、ARHMM の内部状態数、つまり図 5 の表の大きさを減らすのにくらべて、MARHMM 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 章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、ARHMM 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 MCM は、MARHMM にくらべて精度が低い。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、MARHMM 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 MCM (混合数 60) では 30.8 秒、MARHMM (混合数 30) では 166.2 秒、ARHMM (混合数 1) では 537.0 秒であった。また同様の環境において SMC のパラメータ推定に要した時間は 0.9 秒であった。この計測においては、実行環境として Xeon 3.2GHz を 2 つ搭載しメモリ 3GB、OS が WindowsXP 64 bit 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ともな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 MARHMM が有効であることが分かった。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 MARHMM 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。

5. おわりに

本研究では、歩行者の行動をモデル化する方式として、MARHMM を用いた方式を提案した。本方式は、従来の MCM を用いた方式に比べて、歩行者の時間経過にともなう行動変容を適切に扱うことができた。また、ARHMM で問題となる主記憶領域や処理時間、サンプル数の増大も回避することができた。本方式の効果について、実際の商業施設における歩行者の測位データを用いて、行き先予測の性能により評価したところ、従来の MCM では 64% の精度であったところを、72% の精度で予測できることが分かった。さらに、時間経過後の予測精度は、従来では 70% 程度であったところを提案方式では 80% を超える精度が実現できた。これにより、歩行者の位置情報を収集できる環境を整えれば、そこから価値を抽出して新たなサービスに活用できると考えられる。今回の実験は、特定の商業施設内の限られた測位データに対する評価であるが、対象の領域が広域に、歩行者が大規模になった場合でも同様の方式が適用できるかの検証は、今後の課題である。

謝辞 本研究は経済産業省「平成 21 年度 IT とサービスの融合による新市場創出促進事業」の(株)電通関西支社殿との共同自主事業成果をもとにしています。本成果を用いるにあたってご協力いただいた(株)電通関西支社の山川茂孝氏、志村彰洋氏に深く感謝の意を表します。

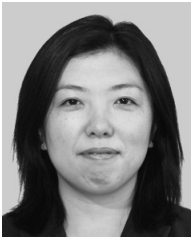
参考文献

- [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 (テーマ: モバイルコンピューティング・一般), 情報処理学会研究報告 MBL [モバイルコンピューティングとユビキタス通信研究会研究報告], No.90, pp.25-31 (2005).
- [3] Manandhar, D., Kawaguchi, S., Uchida, M., Ishii, M. and Tomohiro, H.: IMES for mobile users. Social implementation and experiments based on existing cellular phones for seamless positioning, *Proc. Int. Symposium on GPS/GNSS 2008* (2008).
- [4] 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).
- [5] 株式会社 ROAGroup: 日本の位置情報サービス (LBS) の動向と展望 (2007).
- [6] 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 (2007), 入手先 (<http://www-06.ibm.com/jp/press/20071012001.html>).
- [7] 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).
- [8] 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).
- [9] 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).
- [10] 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).
- [11] 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).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] Bishop, C.M.: *Pattern Recognition and Machine Learning*, Springer, New York (2006).
- [16] Nguyen, N.T., Phung, D.Q., Venkatesh, S. and Bui, H.: Learning and Detecting Activities from Movement Trajectories Using the Hierarchical Hidden Markov Models, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.955-960 (2005).
- [17] Yang, J., Xu, Y. and Chen, C.: Human action learning via hidden Markov model, *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol.27, No.1, pp.34-44 (1997).
- [18] Dias, J.G. and Vermunt, J.K.: 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).
- [19] 株式会社電通: 2010 年 2 月 10 日ニュースリリース (2010), 入手先 (<http://www.dentsu.co.jp/news/release/2010/pdf/2010012-0210.pdf>).
- [20] 日立製作所: ニュースリリース: 2 月 18 日 (2009), 入手先 (<http://www.hitachi.co.jp/New/cnews/month/2009/02/0218.html>).



浅原 彰規

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



佐藤 暁子 (正会員)

1998年東京大学大学院工学系研究科修士課程修了。同年(株)日立製作所入所, ICカードシステムの研究開発を担当。現在, 中央研究所主任研究員として位置情報サービスの研究に従事。



丸山 貴志子 (正会員)

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