

## 混合マルコフモデルに基づく歩行者動線解析方式

浅原 彰規<sup>†1</sup> 丸山 貴志子<sup>†1</sup> 佐藤 暁子<sup>†1</sup>

本研究は歩行者の持つ測位装置から得られる測位データに意味づけする動線解析の研究であり、本稿では集積された多数の動線を混合マルコフモデルでモデル化する方式を提案する。これにより、歩行者の行動パターンの代表例提示、歩行者の次の行動予測など、様々な用途に活用できる。本研究では展示会会場の歩行者行動シミュレーションから得られた来訪展示履歴データに対し提案方式を適用し、歩行者が次に訪れる展示を予測する実験をしたところ、74.1%の精度で予測できることが確認できた。

### Pedestrian Trajectory Analysis Based on Mixed Markov Model

AKINORI ASAHARA,<sup>†1</sup> KISHIKO MARUYAMA<sup>†1</sup>  
and AKIKO SATO<sup>†1</sup>

This paper describes about a trajectory analysis which extracts information about trajectory; we propose a modeling method by Mixed Markov Model for stored trajectory data. The proposed method is available for representative trajectory extraction, or prediction of user's action, or other various applications. We applied the proposed method to exhibition visiting histories generated by a pedestrian action simulator. Then we obtained that prediction rate was 74.1%.

#### 1. はじめに

##### 1.1 背景

本研究は測位装置で取得された測位データから意味を抽出する動線解析技術の研究であ

る。ここでいう測位装置とは装置の現在位置を計測し、座標値として得る機能を持った装置を指す。測位装置に関する近年の動向として、屋外での GPS 測位に加え、無線 LAN<sup>1)</sup> や屋内設置型の GPS 発信器<sup>2)</sup> など屋内でも利用できる多種の測位方式の研究開発がさかんに進められている<sup>3)</sup>。これらの技術により、将来的にはあらゆる時と場所において、携帯端末の位置を検出できる環境の整備が進むと想定される。本稿では屋内外の制限なく位置を測定し、その測位データを利用する位置情報システムを屋内外空間情報システムと呼ぶ。

屋内外空間情報システムには、従来にはなかった新たなアプリケーションが存在する。たとえば、屋外にいる歩行者をビル内の会議室まで案内する詳細なナビゲーションサービスや、LBS (Location Based Services) と呼ばれるユーザ位置を利用した情報検索の屋内適用などである<sup>4)</sup>。

本研究では、この屋内外空間情報システムが屋内外によらず歩行者の位置情報が取得することに着目し、歩行者の位置を常時計測しつつ、その位置情報から新たな意味を抽出することによるアプリケーションの創出を目指す。たとえば、商業施設などの歩行者の平均的行動パターン抽出や歩行者行動予測の提示<sup>5)</sup>、例外行動の検出<sup>6)</sup>、当該施設の店舗設備の稼働率定量化、作業員配置の推薦などができれば、施設管理者による意志決定を支援することができる。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<sup>7)</sup> も考えられる。また、歩行者の移動履歴に基づき歩行者の状況を推測し、それに適応した情報を配信する機能も提供できる。

図 1 に当該アプリケーションのために想定されるシステムの構成を示す。本システムは、主に歩行者が測位機能付きの携帯端末を持ち歩くことを前提とする。携帯端末を保有する歩行者の位置は、携帯端末の GPS 受信機能や測位基地局によって測定され、測位データとしてサービスアプリケーションサーバに送付される。各種サービスアプリケーションサーバはこの測位データを使用し、地理空間情報サーバに地図情報を要求して、必要な地図情報を得て動作する。本過程において得られる測位データは動線解析エンジンにも送られ、その後、動線解析処理が実行される。動線解析エンジンは、アプリケーションサーバから要求があれば解析結果を配信するようになっており、動線解析の結果を利用したサービスが実現する。

本研究の目的は、数 100 m ~ 数 km 四方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の測位データを対象とした動線解析エンジンの実現である。今回、歩行者の移動履歴をもとに歩行者の行動を表現する混合マルコフモデルを生成することにより、典型的行動パターンの抽出や歩行者行動の予測を行う方式を提案する。また、展示会場の歩

<sup>†1</sup> 株式会社日立製作所  
Hitachi, Ltd.

行者行動シミュレーションから得られる来訪展示履歴データに提案方式を適用し、その有効性を検証する。

## 2. 関連研究

多数の位置情報から新たな意味を見出す技術は空間データマイニングと呼ばれ、多くの先行研究が存在する。豊嶋ら<sup>8)</sup>はRFIDを用いて歩行者のいた箇所の位置IDを取得し、その履歴から主な動線(主動線)を抽出する技術について報告した。この研究では、頻度の多い動線を抽出することに成功しているが、始時点間の一貫した主動線は抽出できていない。Leeら<sup>9)</sup>は座標値列の構成する線分間に距離を定義し、当該距離に基づく動線のクラスタリングを行うとともに、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。なお、動線のクラスタリングや主動線の抽出に関しては、空間データベースの1分野として動線形状に基づく方法が研究されている<sup>10)-12)</sup>。

また、蓄積した動線から動線以外の情報を得る技術も報告されている。神田ら<sup>7)</sup>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ち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。松尾ら<sup>13)</sup>は赤外線センサで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、たと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。青木ら<sup>5)</sup>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)で表現することにより、当該人物の行動を記号化する研究について報告している。この研究では、

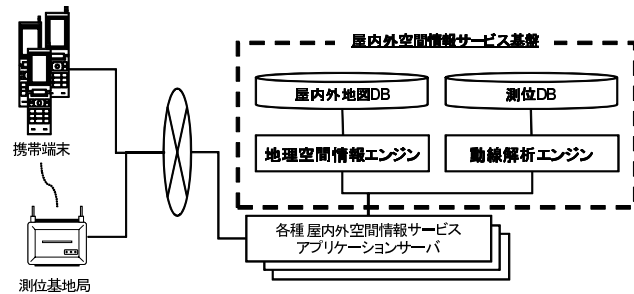


図 1 屋内外空間情報システム

Fig. 1 Indoor and outdoor spatial information system.

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。

なお、カメラ<sup>6),14)</sup>やレーザレンジファインダ<sup>15)</sup>などを用いた例もある。これらの方式は、カメラやレーザの視野に入っていた移動体を追跡することを目的とし、視野内を通過する移動体の動線形状類似度に基づき動線を分類する。したがって、長時間広範囲の歩行者動線の解析には適さない。

## 3. 動線解析処理の構成

動線解析とは、測位データに対して意味づけを行う処理である。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むデータであり、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。また、時系列測位データとは、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。

動線解析では、意味づけ対象の動線をサンプリングした時系列測位データの処理により、動線が意味づけられる。図2にこの処理の全体構成を示す。以下では、各処理について述べる。

### 3.1 前処理

図2(a)の前処理とは、意味づけが実施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<sup>16)</sup>、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。

### 3.2 状態判別処理

図2(b)の状態判別では、動線区間に対し区間中の移動体の状態を示す状態ラベルが付与される。ここで動線区間とは、動線の一部区間を切り取ったものを指す。動線区間と歩行者の行動などの情報を関連づけるには、動線区間がどのような状態を表現しているか判定する

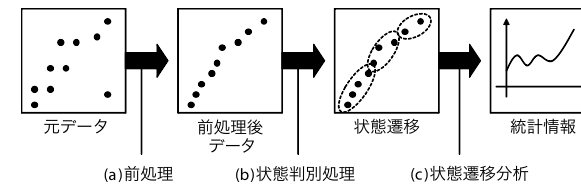


図 2 動線解析処理の全体構成

Fig. 2 Overview of the trajectory analysis.

必要がある．状態判別処理では，動線区間のサンプリングである時系列測位データを用い，移動中や滞留中など動線区間での状態を判別し，それに対応する状態ラベルを付与する．

### 3.3 状態遷移分析

図 2(c) の状態遷移分析は，動線を状態遷移へ変換した後に実施される統計処理である．前述の状態判別処理によって，動線区間が状態ラベルへ変換されるため，動線全体は状態間の遷移と見なすことができる．この状態遷移は一種の記号列であるので，記号列に対する統計処理が適用できる．たとえば，状態遷移をマルコフモデルであると見なしてモデルパラメータの推定を行ったり，記号列マッチングにより動線間の類似度を判定したりすることができる．

## 4. 状態遷移分析の重要性

### 4.1 主動線抽出の例における課題

動線解析の応用例として，主動線抽出処理を図 3 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 3(a) 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，主動線抽出が必要となる．

主動線抽出は，以下の 2 つのステップにて構成される．最初に図 3(a) の元データを図 3(b) の太線に囲まれた類似する動線の集合に分割するステップが実行される．一般的に，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，それらを生成する処理をクラスタリングと呼ぶ．次に，図 3(c) に示すように，各クラスタを代表する動線を生成するステップが実行される．この結果，各移動パターンがどの条件で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

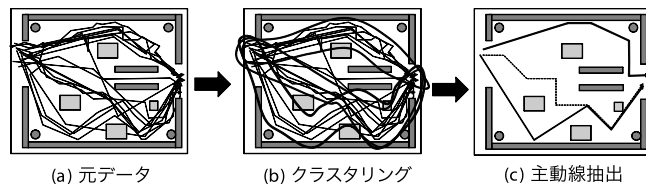


図 3 主動線抽出のプロセス  
Fig. 3 Processes for representative trajectories extraction.

目的に合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<sup>12)</sup>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，主動線の抽出が可能であるとも考えられる．たと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 4 に公園を散策する歩行者行動の例を示す．この場合，図 4(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 4(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かかわらず，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．

### 4.2 状態遷移分析による課題解決

本研究では図 2 の手順によってこの問題を解決する．まず，図 2(a) で測位の誤差などを低減させ，図 2(b) で移動や滞留など状態遷移を意味する記号列に変換したうえで，図 2(c) に相当する処理として，記号列に対するクラスタリングを実行する．これにより，たと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，大域的な歩行者行動の分類が可能となる．

図 2 のステップのうち，(a)，(b) は測位手段の影響が大きく，実測データを集めなければ評価が困難である．また，測位手段によっては (a)，(b) の処理は単純になることも多い．

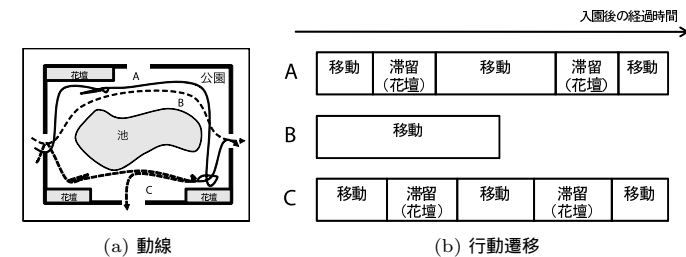


図 4 公園における歩行者行動の例  
Fig. 4 An Example of pedestrian actions in a park.

たとえば、公園内に配置された RFID 受信機を用い、発信器の接近を検知して測位とする場合、受信可能域への滞留時間の長さで状態を推定することになる。したがって、実用の環境によって適切な (a), (b) は異なる。一方, (c) は比較的独立に方式を検討することができる。そこで、今回は状態遷移記号列が得られたとして, (c) の状態遷移記号列の扱いについて検討する。

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。たとえば、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との間の距離が何 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。たとえば、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。ほかにも、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 1 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ときに各データがどの確率的事象によって生成されたかを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。混合確率分布で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。

状態遷移列を表す混合確率分布は歩行者行動の特徴を表現しており、種々の応用が可能である。たと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、確率分布やクラスターの変化から、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。ほかにも、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、それに関する推薦を行うことも考えられる。

以上のように、動線を状態遷移に変換し、確率モデルで表現することにより、種々の応用が見込まれる。そこで、本研究では図 2(c) の状態遷移分析として、多数の状態遷移記号列からそれらを適切に表現する確率モデルを得ることを目標とした。

## 5. 混合マルコフモデルによる歩行者行動のモデル化

### 5.1 歩行者行動の確率モデルの比較

歩行者行動の確率モデルをグラフィカルモデルで表現したものを図 5 (a) 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適用して得られる記号列である。歩行者は外部から観測不能な内部的状態変数  $z_i$  (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、観測不能な内部の状態  $z_i$  を内部状態と呼ぶ。(a) のモデルでは、空間状態  $d_{i+1}$  は  $d_i$  と  $z_{i+1}$  に依存する確率  $P(d_{i+1}|d_i, z_{i+1})$  に基づく確率過程により決定される。一般的に内部状態  $z_i$  のような観測不能な変数は隠れ変数と呼ばれ、図 5 (b) の HMM などの隠れ変数を取り入れたモデルに対し、観測可能な量から隠れ変数を推定する方式<sup>17)</sup> がよく知られている。この HMM の事前学習により、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<sup>5),6)</sup>。図 5 (a) のモデルでは、歩行者の次の空間状態は直前の空間状態とその時点での内部状態に依存する。一方、図 5 (b) 隠れマルコフモデルは、歩行者の空間状態は直前の空間状態に依存しない。しかし実際には、歩行者は空間的な隣接地点にしか行けないという制約のもとで行動しており、直前の空間状態の歩行者行動への影響は大きい。また、たとえば屋外へ向かおうとした歩行者が近い方の出口へ向かう、歩行者がたまたま目についた店舗へ入るなど、歩行者の現在地が意志決定に影響するのは自明である。その反面、(a) は複雑なモデルであり、モデル推定のために多数のデータが必要となる。

そこで本研究では、(a) のモデルを内部状態が時間変化しないと近似して得られる (c) の混合マルコフモデル<sup>18)</sup> の適用を提案する。混合マルコフモデルでは、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、以降の空間状態間遷移は直前の空間状態によってのみ定まる。つまり、混合マルコフモデルは複数のマルコフモデルを重ね合

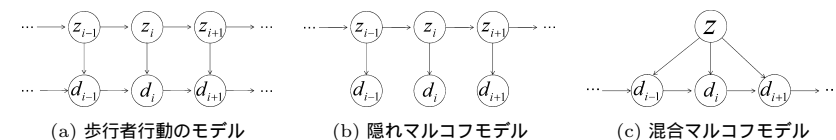


図 5 確率モデルの比較  
Fig. 5 Contrast between probabilistic models.

わせたモデルである。従来、混合マルコフモデルは人物の選択行動を表現するのに用いられていたが、上述のように空間的な制約を表現するのに適したモデルであるといえる。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対照的であり、内部状態がほとんど変化しない場合や、歩行者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、混合マルコフモデルによる近似が適する。また、混合マルコフモデルでは推定すべき内部状態は 1 つしかない。この唯一の内部状態は時間によらない特性を意味するため、内部状態が所属クラスタを意味すると見なせる。なお、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。

混合マルコフモデルで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あてはめる必要がある。本研究の対象は、数 100 m ~ 数 km 四方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<sup>19)</sup>。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 1 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り返し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すべて同一状態となる。この変更にともない、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すべ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。

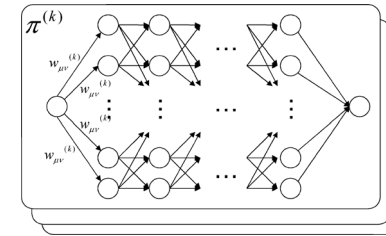


図 6 混合マルコフモデルの概念図  
Fig. 6 Abstract of mixed Markov model.

## 5.2 混合マルコフモデルの確率分布

提案方式では、最尤法を用いて状態遷移があてはまる混合マルコフモデル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式である。尤度とは確率モデルが観測データを説明する度合いであり、確率分布に観測データを代入して得られる。そこで、混合マルコフモデルの確率分布を求める。空間状態遷移の記号列を  $d_n \in D$ 、ただし  $d_n$  ( $n = 0, \dots, N$ ) とすると、混合マルコフモデルの確率分布は、

$$P(d|\{\pi^{(k)}\}, \{\omega_{\mu\nu}^{(k)}\}) = \sum_k^K \pi^{(k)} \psi(d|\{\omega_{\mu\nu}^{(k)}\}) \quad (1)$$

となる。ここで、 $\{\pi^{(k)}\}$  は各モデルの混合比率、 $\psi$  はマルコフモデル確率分布を、 $\{\omega_{\mu\nu}^{(k)}\}$  は  $k$  番目のモデルの状態  $\mu$  から  $\nu$  への遷移確率を表す。ここでいうマルコフモデルの確率分布は、たとえば、 $d$  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。このモデルの概念図を図 6 に示す。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。

## 5.3 EM 法によるパラメータ推定

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では、パラメータ推定と隠れ変数の推定には互いの結果が必要となってしまう。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、隠れ変数がこの期待値であると仮定して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、この困難を避ける。

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

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 7 EM 法による混合マルコフモデルのパラメータ推定  
 Fig. 7 Steps of the EM method.

$$L(D) = \sum_n \log P(d_n|\{\pi^{(k)}\}, \{\omega_{\mu\nu}^{(k)}\}) = \sum_n \log \left( \sum_k \pi^{(k)} \psi(d_n|\{\omega_{\mu\nu}^{(k)}\}) \right) \quad (2)$$

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になる。そこで Lagrange の未定定数法を適用すると、

$$\pi^{(k)} = \frac{1}{N} \sum_n \frac{\pi^{(k)}\psi(d_n|\{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')}\psi(d_n|\{\omega_{\mu\nu}^{(k')}\})} \equiv \frac{1}{N} \sum_n \gamma^{(k)}(d_n) \quad (3)$$

$$\omega_{\mu\nu}^{(k)} = \frac{\sum_n c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)}{\sum_{n,\nu} c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)} \quad (4)$$

が得られる。ただし、状態遷移列  $d_n$  において状態  $\mu$  から状態  $\nu$  への遷移が起こった回数を  $c(d_n, \mu \rightarrow \nu)$  とした。この  $\gamma^{(k)}(d_n)$  は混合モデルにおける各モデルの重みを表しており、負担率と呼ばれる。この結果は負担率に基づき分配された頻度分布が  $\omega_{\mu\nu}^{(k)}$  となることを示している。つまり、 $d_n$  内に  $\mu \rightarrow \nu$  の遷移が 1 度あるとその遷移の重みが  $\gamma^{(k)}(d_n)$  だけ増加する。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 7 のようになる。

## 5.4 確率モデルの活用

本処理によって得られた確率モデルは、様々な用途に適用できる。たとえば、状態遷移データ  $d_n$  は当該データの生成モデルによりパターン分類でき、それにより状態遷移の元となる動線も分類される。 $d_n$  の生成モデルを表す変数  $z_{nk}$  ( $k$  番目のモデルから生成されたデータは  $z_{nk} = 1$ 。それ以外は  $z_{nk} = 0$ ) は、 $d_n$  に対する事後確率最大化により推定できる。したがって、 $d_n$  は  $z_{nk} = 0$  の事後分布

$$P(z_{nk}|d_n) = \gamma^{(k)}(d_n) \quad (5)$$

つまり負担率が最大のモデルから生成されたと推定できる。また、各クラスタの生成モデルに最も適合する状態遷移の元となる動線を各クラスタの主動線とできる。そのため、クラスタ  $k$  のモデルに対する状態遷移  $d_n$  の対数尤度

$$L^{(k)}(d_n) = \log \psi(d_n|\{\omega_{\mu\nu}^{(k)}\}) \quad (6)$$

が最大となる  $d_n$  に対応する動線を主動線と見なせる。また、歩行者行動の予測も同様に行うことができ、 $d_n$  の  $m$  番目の空間状態を  $d_{n,m}$  とすると、

$$P(z_{nk}, d_{n,m+1}|d_{n,m}) \propto \sum_n \gamma^{(k)}(\{d_{n,m}\}) \psi(\{d_{n,m+1}\}|\{\omega_{\mu\nu}^{(k)}\}) \quad (7)$$

が最大となる  $d_{n,m+1}$  を計算すればよい。

## 6. 実 験

### 6.1 実験の目的と条件

本実験の目的は、混合マルコフモデル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歩行者の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、歩行者状態遷移から混合マルコフモデルを生成し、そのモデルに基づく歩行者行動の予測精度を算出する。これにより、当該確率モデルが歩行者行動をどの程度再現できるかが評価できる。

表 1 に本実験に用いたデータを示す。本来、実測の歩行者動線から得られた状態遷移を用いるのが望ましいが、実測環境を整えるのは容易ではないため、今回は群衆シミュレータ<sup>20),21)</sup>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を用いた。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をモデル化しており、歩行者 1 人 1 人の行動が特定のルールと乱数に基づき決定される。このシミュレータには歩行者が滞留した展示物

表 1 実験データの概要  
Table 1 Summary of experiment data.

項番	項目名	内容
1	データ内容	群集シミュレータ <sup>20),21)</sup> による来訪展示の履歴
2	サンプル数	1,337 件 × 10 セット
3	想定会場	技術展示会 (展示数 150 地点, 丸め後 65 地点)
4	想定時間	11 時 ~ 12 時の 1 時間

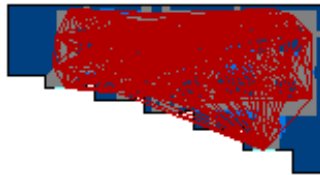


図 8 1 セット分の評価データ  
Fig. 8 One experiment data set.

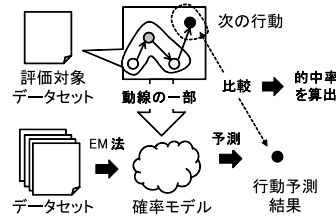
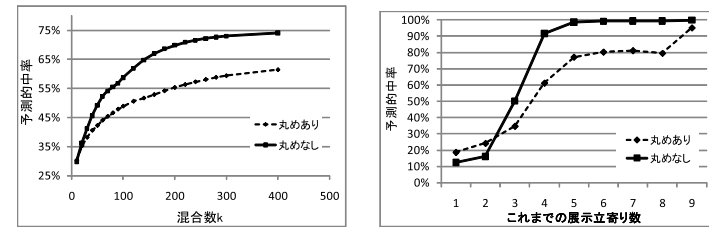


図 9 実験手順  
Fig. 9 Steps for the experiment.

の記録を出力する機能があるので、歩行者の滞留した展示の履歴と展示の位置を関連づけ、歩行者の状態遷移と見なすこととした。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、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。このうち、1 セット分の評価データを一例として図 8 に示す。ただし、このシミュレータで得られるのは展示への来訪記録である。もし用いるデータが位置情報であるなら、隣接する展示と区別がつかないこともありうる。そこで、展示物の来訪履歴に対し、隣接する展示を同一視するという処理も実施した。以下では、展示物の来訪履歴をそのまま用いたものを丸めなし、隣接展示を同一視したデータを丸めありと表記する。なお、本展示会の展示物は 150 展示を想定しているが、丸め処理を実施した後は展示数は 65 に減少した。

本評価実験の評価手順を図 9 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



(a) クラスタ数ごとの予測精度 (b) 展示立寄り履歴数ごとの予測精度

図 10 行動予測の精度

Fig. 10 Accuracy of the action prediction.

セットから 1 つ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。たと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。評価対象セットのすべ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 セットすべてを 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。

今回、比較のため、MM (Markov Model; マルコフモデル) による予測と HMM による予測に関しても同様の評価を行った。MM による予測では、歩行者の現在の状態のみから次の状態を予測する。ある状態 A を通る状態遷移が  $C_A$  件、うち A の次が状態 B であった状態遷移が  $C_{AB}$  件あった場合、遷移確率  $\frac{C_{AB}}{C_A}$  が最大の状態 B をもって予測結果とする。一方、HMM による予測では図 5 (b) の確率モデルを用い、最大確率となる次状態を予測結果とする。なお、HMM では隠れ状態の数を決める必要があるが、今回は 1 から 200 の範囲で最も精度が高くなるパラメータを採用した。また、前述のとおり 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。MM は地点間の相関関係のみを扱い、HMM は歩行者特性の時間変化のみを扱うモデルである。この精度比較により、地点間の相関と歩行者特性の両方を扱う提案方式の有効性が示せると考えられる。

### 6.2 結果と考察

図 10 (a) に、モデルの混合数 K を変えながら評価した結果のグラフを示す。横軸は K、縦軸は的中率すなわち予測精度を意味する。今回の実験では  $K = 400$  程度までしか調査していないが、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おり、丸めありの場合で 61.4%、なしの場合で 74.1%の精度に収束しつつある様子が分かる。したがって、シ

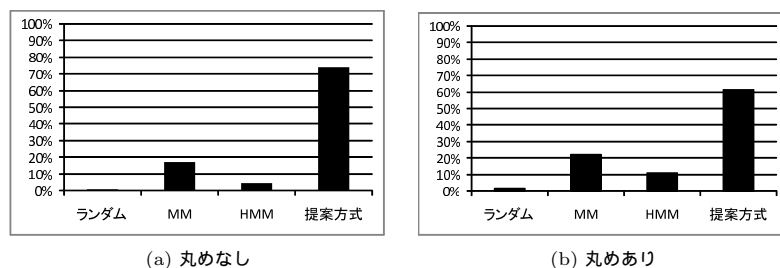


図 11 行動予測精度の比較  
Fig. 11 Contrast of the action prediction accuracies.

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クラスタを用いなければならぬことになる。

図 10 (b) に、 $K = 400$  の場合で、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。図からも分かるように  $n$  が大きいほど予測精度が改善されており、丸めなしに関しては  $n = 4$  の場合で 98.6% という高い精度で行動が予測できている。また、 $n$  が 1 または 2 のときは丸めありの方が精度が高いが、以降は丸めなしの精度の方が高い。この結果は、予測に利用できる履歴情報が多いほど精度が高いことを意味している。丸めありの方が移動しうる候補が少ないが、履歴を用いることにより、精度が向上し、精度の向上率の高い丸めなしの方が逆転したと考えられる。

もう 1 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。混合マルコフモデルにはステップ数の概念は存在しないため、歩行者が退場するタイミングは予測困難と考えられるが、実際には丸めありで 95.1%、丸めなしで 99.9% が的中した。丸めありの場合、最後のステップだけが極端に精度が向上している点から、本ミュレーションの退場する直前の行動だけが非常に特徴的であったようである。実際、この群集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。

最後に、従来方式と提案方式の精度比較を図 11 に示す。(a) は丸めなし、(b) は丸めあ

りの予測精度である。グラフ中のランダムという項目は、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、MM の項目は MM による予測精度、HMM の項目は HMM による予測精度、提案方式の項目は混合マルコフモデルによる予測精度をそれぞれ表す。丸めなしの場合、ランダムは候補数 150 なので  $\frac{1}{150} \approx 0.7\%$  となり、MM は 16.9%、HMM は 10.7% であった。提案方式は 74.1% であったので、従来方式よりも十分高精度であったといえる。丸めありの場合、ランダムは候補数 65 なので  $\frac{1}{65} \approx 1.5\%$ 、MM は 21.8%、HMM は 10.7% であった。提案方式の 61.4% は丸めなしの場合よりも低いですが、従来方式よりは高精度であった。

予測精度の比較により、丸め処理に対する混合マルコフモデルの特性が分かる。一般的には候補数が少ないほど予測精度は高くなるはずである。実際、従来方式では丸めありの方が高い精度が得られている。ところが、提案方式ではむしろ丸めありの方が精度が低い結果となった。これは、隣接する展示のいずれに来訪したかによって次の行動が異なることを意味する。この振舞いはシミュレーション上の人物行動に由来する。歩行者の行動を扱うモデルでは、位置だけでなく何をしていたかを正確に判断することも行動予測にとっては重要であると考えられ、この傾向は実際の歩行者の場合でも同様と推測される。今回、乱数に基づくシミュレーションという、ばらつきの強いデータを用いたため、必要なパターン数が大きくなったと考えられる。実測では偏りがあると考えられるため、今後、実測のデータを用いて検証する必要がある。

以上の結果から、混合マルコフモデルという単純なモデルであっても、混合数が十分大きければシミュレーション中の歩行者行動をモデル化できるといえる。ここで仮に、展示会場における次の来訪展示を推薦するアプリケーションを想定する。来訪者が 8 つの展示を訪れるとして、確率  $P$  で行動予測するシステムが推薦を行った場合、1 度も的中しない来訪者を 1% 未満にするには  $(1 - P)^8 < 0.01$  を満たさなければならず、 $P$  の精度は 43.8% 以上が必要となる。この精度は従来方式では達成できていないが、提案方式は達成できている。すなわち、シミュレーションでは混合マルコフモデルに基づく歩行者行動の分析は有効であったといえ、実際にも有効である見込みが得られたといえる。

## 7. おわりに

本研究では、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。シミュレーションで生成されたデータに提案方式を適用し、シミュレーション上の歩行



者行動予測を試みたところ、次の行動予測精度は全体で 74.1%、4 つ以上の行動履歴があれば 98.6% の確率で予測できていた。これにより、提案方式によるモデル化は適切であったと考えられる。

今回の実験ではシミュレーションによって生成された仮想的なデータを用いたが、実測データでは計測のエラーや歩行者の例外的行動などがありそのまま提案方式を適用することが妥当であるとは限らない。今後は、実測データを収集して提案方式の妥当性を評価することにより新たな課題を明確化し、動線解析を実用化するためのさらなる改善を推進する予定である。

### 参 考 文 献

- 1) クウジット株式会社: PlaceEngine<sup>®</sup>. <http://www.placeengine.com/>
- 2) 日立製作所: ニュースリリース: 2月18日(2009). <http://www.hitachi.co.jp/New/cnews/month/2009/02/0218.html>
- 3) 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).
- 4) 株式会社 ROAGroup: 日本の位置情報サービス(LBS)の動向と展望(2007).
- 5) 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).
- 6) 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).
- 7) 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).
- 8) 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).
- 9) Lee, J.-G., Han, J. and Whang, K.-Y.: Trajectory clustering: A partition-and-group framework, *Proc. 2007 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, ACM, pp.593-604 (2007).
- 10) Asahara, A., Sato, A. and Maruyama, K.: Evaluation of Trajectory Clustering Based on Information Criteria for Human Activity Analysis, *IEEE International Conference on Mobile Data Management*, pp.329-337 (2009).
- 11) Chudova, D., Gaffney, S., Mjolsness, E. and Smyth, P.: Translation-invariant mixture models for curve clustering, *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, ACM, pp.79-88 (2003).
- 12) Trajcevski, G., Ding, H., Scheuermann, P., Tamassia, R. and Vaccaro, D.:

Dynamics-aware similarity of moving objects trajectories, *Proc. 15th Annual ACM International Symposium on Advances in Geographic Information Systems*, New York, NY, USA, ACM, pp.1-8 (2007).

- 13) 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 (2007).
- 14) Nascimento, J., Figueiredo, M. and Marques, J.S.: Segmentation and Classification of Human Activities, *Workshop on Human Activity Recognition and Modelling* (2005).
- 15) 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, Vol.127, No.4, pp.537-545 (2007).
- 16) 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, Vol.17, pp.75-79 (2008).
- 17) Bishop, C.M.: *Pattern Recognition and Machine Learning*, Springer, New York (2006).
- 18) Poulsen, C.S.: Mixed Markov and latent Markov modelling applied to brand choice behaviour, *Intern. Journal of Research in Marketing*, Vol.7, pp.5-19 (1990).
- 19) 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.348-359 (2007).
- 20) Hamada, T., Hagiwara, T., Teramoto, T., Morishita, S., Umetsu, M. and Ohgama, M.: A Tree-dimensional Pedestrian-flow Simulation for High-rising Buildings, *ACRI2008*, Vol.LNCS 5191, pp.417-424, Springer (2008).
- 21) 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).

### 付 録

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、以下のとおりである。

$$L(D) = \sum_n \log P(d_n | \{\pi^{(k)}\}, \{\omega_{\mu\nu}^{(k)}\}) = \sum_n \log \left( \sum_k \pi^{(k)} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) \right) \quad (8)$$

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 の未定数として  $\pi^{(k)}$  の制約を表す  $\alpha$ , 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 .

$$L'(D) = L(D) + \alpha \left( \sum_k \pi^{(k)} - 1 \right) + \sum_{k,\mu} \beta_\mu^{(k)} \left( \sum_\nu \omega_{\mu\nu}^{(k)} - 1 \right) \quad (9)$$

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

$$\frac{\partial L(D)}{\partial \pi^{(k)}} = \sum_n \frac{\psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})} + \alpha = 0 \quad (10)$$

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 . この結果を式 (10) に代入して両辺に  $\pi^{(k)}$  をかけると

$$\pi^{(k)} = \frac{1}{N} \sum_n \frac{\pi^{(k)} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})} \equiv \frac{1}{N} \sum_n \gamma^{(k)}(d_n) \quad (11)$$

が得られる . 次に, 式 (9) に関し  $\omega_{\mu\nu}^{(k)}$  の微分係数が 0 となる条件の方程式は,

$$\frac{\partial L(D)}{\partial \omega_{\mu\nu}^{(k)}} = \sum_n \frac{\pi^{(k)} \frac{\partial}{\partial \omega_{\mu\nu}^{(k)}} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\})}{\sum_{k'} \pi^{(k')} \psi(d_n | \{\omega_{\mu\nu}^{(k')}\})} + \beta_\mu^{(k)} = 0 \quad (12)$$

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  $\mu$  から状態  $\nu$  への遷移が起こった回数を  $c(d_n, \mu \rightarrow \nu)$  とすると,

$$\psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) = \prod_{\mu,\nu} \omega_{\mu\nu}^{(k)c(d_n, \mu \rightarrow \nu)} \quad (13)$$

$$\frac{\partial}{\partial \omega_{\mu\nu}^{(k)}} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) = \frac{c(d_n, \mu \rightarrow \nu)}{\omega_{\mu\nu}^{(k)}} \psi(d_n | \{\omega_{\mu\nu}^{(k)}\}) \quad (14)$$

これを用いて式 (12) を書き換え, 両辺に  $\omega_{\mu\nu}^{(k)}$  をかけて  $\nu$  について和をとると,

$$\beta_\mu^{(k)} = - \sum_{n,\nu} c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n) \quad (15)$$

が得られる . 式 (12) にこれを代入し  $\omega_{\mu\nu}^{(k)}$  について解くと

$$\omega_{\mu\nu}^{(k)} = \frac{\sum_n c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)}{\sum_{n,\nu} c(d_n, \mu \rightarrow \nu) \gamma^{(k)}(d_n)} \quad (16)$$

となる .

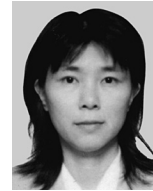
(平成 22 年 3 月 26 日受付)

(平成 22 年 10 月 4 日採録)



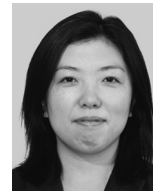
浅原 彰規

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 .



丸山貴志子 (正会員)

1988 年お茶の水女子大学大学院理学研究科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 .