

ライフログによる時間帯別移動先予測に適する エリア粒度の検討

藪 圭輔¹ 榎堀 優¹ 間瀬 健二¹

概要：ライフログを用いたユーザの移動予測において、人間の行動パターンが時刻によって変化する性質を利用し、位置情報と時刻情報を併用する方法が提案されている。時刻情報を併用して移動予測をする場合、ユーザの移動先として予測する場所（エリア）の粒度を細かくすると、同一データ量ではエリア間あたりの移動回数が少なくなり学習に利用するデータ数が十分に確保できず、予測精度が低下する。これは特にブートストラップ時のデータ総量が少ないときに問題となる。本研究ではブートストラップ期間を想定した約1週間のライフログで学習した場合において、エリアの粒度を半径5m, 10m, 15mと段階的に変化させて移動予測確率の推定の尤もらしさを評価し、どの粒度であれば、時間帯を考慮した移動予測が時間帯を考慮しない予測よりも高い精度を示し、かつ、移動予測に適した特性を持つか検討した。結果、エリアの粒度が半径10m以上であれば、時間帯を考慮した移動予測確率推定の精度が時間帯を考慮しない場合と比べて高く、かつ、予測確率が0.5以上となる移動予測における予測スコアが0.8を超えたため、時間帯を考慮した予測が有効であると分かった。半径10mの円の面積は1住宅当たりの敷地面積に近いことから、1週間程度のライフログにおいては時間帯を考慮した移動先予測手法は建物間移動レベルの個人移動予測で有効である可能性が示唆された。

Investigation of Area Particle Size Suitable for Lifelog Based Mobility Prediction using Period of Time

YABU KEISUKE¹ ENOKIBORI YU¹ MASE KENJI¹

1. はじめに

近年のデジタルストレージの容量の増加やデバイスの小型化などにより個人がより手軽に様々な行動のログ（ライフログ）を取得・蓄積できるようになり、それを使ってユーザの行動や嗜好を分析する研究が多く行われている。その中に、ライフログを用いた個人移動予測がある。

ライフログを用いた個人移動予測に、ライフログ中の位置の遷移情報を学習する手法がある [1]。また、人間の行動パターンは時刻によって変化するため、時間帯別に予測を実施するものがある [2]。例えば、勤め人であれば、平日の朝は家から会社へ行く確率が高い。しかし、夜に会社からの帰宅後は、家から居酒屋に行く確率が高いかもしれない。

一方で、時間帯に分け、かつ、ユーザの移動先として予

測する場所（以下、エリア）の粒度を細かくすると、エリア間あたりの移動回数が少なくなり、学習に利用するデータ数の確保が難しくなる。ライフログを対象とするならば、半年や1年のログを学習することは可能だが、データ量が少ない時期に利用できないようでは普及の妨げになる。

そこで、本研究ではブートストラップ期間を想定した1週間程度のライフログを想定し、エリアの粒度を段階的に変化させ、半径5m, 10m, 15mとして移動予測確率の推定の尤もらしさを評価し、どの粒度までであれば時間帯を考慮した移動予測の方が時間帯を考慮しない予測よりも高い精度を示し、かつ、移動予測に適した特性を持つかを検討した。

以降の本稿の構成は次のとおりである。第2章で移動予測の関連研究について述べる。第3章でユーザの位置を推定するために使用したエリア識別器について述べる。第4章でユーザが存在したエリアと時間情報から移動予測する

¹ 名古屋大学 大学院 情報科学研究科
Graduate School of Information Science, Nagoya University

移動予測器について述べる。第5章で移動予測器で生成した移動予測モデルの移動予測確率推定の尤もらしさをエリア粒度別に評価した実験について述べる。第6章で本稿のまとめと今後の課題について述べる。

2. 関連研究

位置情報のみを用いた移動予測として、AshbrookらはGPSデータからユーザが滞在した場所を抽出し、移動遷移モデルにマルコフモデルを使用する手法を提案した[1]。Liaoらはダイナミックベイジアンネットワークを用いることで、移手段と移動予測を同時に行う手法を提案した[3]。

自分の位置情報と他の情報を併せた移動予測として、西野らは、GPSとスケジューラログを併用する手法を提案した[4]。スケジューラデータを用いたことによりユーザが普段と異なる行動パターンをとった場合でも未来の位置を予測できる可能性を示した。Domenicoらはユーザ自身の位置情報とその人の友人や、同じような嗜好をもつソーシャルグループの位置情報を考慮する移動予測手法を提案し、高い精度を得た[5]。しかし、これらの方法は併用する情報に汎用性がなく、ユーザによっては利用できない可能性がある。

汎用的な情報と位置情報を併用した移動予測に、西野らの手法がある[2]。西野らは位置情報と時刻情報のみを併用することで移動予測精度の向上に成功した。しかし、エリア粒度と学習データ量の関係は明らかにされていない。

本研究では、ユーザにとって利便性が高いと考えられる位置情報に併せて時間情報を用いた移動予測を対象に、ブートストラップ期間の1週間のライフログを想定した場合のエリア粒度と予測精度の関係を検討した。

3. エリア識別器の構築

本研究は、ライフログによる時間帯別移動先予測に適するエリア粒度を検討するため、移動先として予測する場所の粒度を任意に決定できる必要がある。そこで閾値を変化させることで任意の半径を設定できるエリア識別器と、それを用いたエリアの段階的構築を採用した。以下、それぞれについて詳しく述べる。

3.1 エリア識別器と粒度別閾値

エリア識別器の本体は、著者らが[6]で提案したものである。本識別器は図1に示すように、灰色で示した電波強度分布の重なった面積を閾値として、同一エリアかどうかを2値判別する。この閾値を変化させることで、同一エリアと判断する半径を制御できる。

本研究では、ユーザが一定期間にわたり静止していたならば、その場所は意味がある場所(エリア)であり、位置識別を実施する価値がある場所として学習した。エリア識

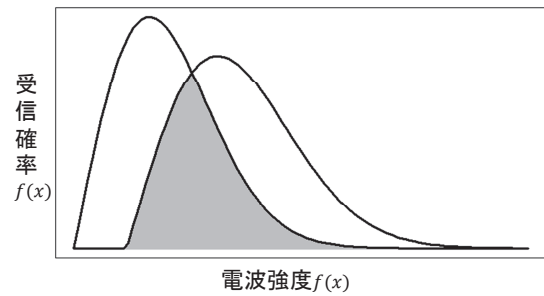


図1 2つの分布の一致部分

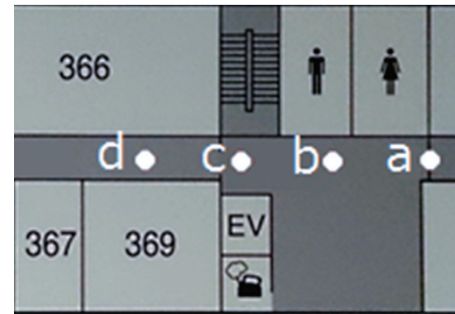


図2 データ観測点

別器は各エリアごとに一つ構築する。

図2に、エリア粒度別の閾値を決定した予備実験の概要を示す。Wi-Fi電波を5m間隔で取得し、a地点の電波ログで学習した識別器が、どの距離の地点の電波ログを同一エリアと判別するかどうかで閾値を決定した。

予備実験に使用するWi-Fi電波の収集にはサムスン電子製のNexus Sを使用して収集した。サンプリング周期を3秒とし、30分のデータを収集した。データの収集は図2の白丸で示す観測点aからdの計4箇所で行った。観測点付近はアクセスポイントの数が多く、常に10種類以上のWi-Fi電波を観測できた。

本研究では5m, 10m, 15mのエリア粒度を設定し、それぞれの閾値を以下の手順で求めた。

粒度が5mとなる閾値として、観測点aで取得されたデータを用いてエリア識別器を生成し、面積一致度の閾値を0.05刻みで変化させて観測点aと観測点bが同じエリアと識別されない最小の閾値を採用した。粒度が10mとなる閾値として、観測点aのデータで生成した識別器で、観測点a, bが同じエリアとされ、観測点cが別エリアと識別される最小の閾値を採用した。粒度が15mとなる閾値として、観測点a, b, cは同エリアと識別されるが観測点dでは別エリアと識別される最小の閾値を採用した。採用した閾値は、5m, 10m, 15mのそれぞれで、0.50, 0.30, 0.10となった。

3.2 エリア識別器群の逐次構築

本稿では、前節で示した任意の粒度で同一エリアかどう

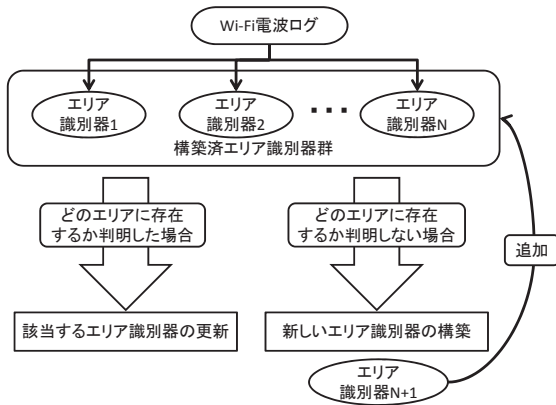


図 3 エリア識別器群の構築方法の概要図

かを判定できる識別器を使って、エリアを逐次的に構築した。エリアの構築は一種のクラスタリングであり、必要に応じてエリアの数を増加させる。これは、取得する Wi-Fi 電波ログに、緯度経度や建物内での位置といった教師信号が付加されていないためである。

本研究では、まず一つの静止期間の Wi-Fi 電波ログから一つのエリア識別器を生成する。詳しいエリア識別器の生成方法は [6] を参照されたい。逐次構築の概要を図 3 に示す。静止期間の Wi-Fi 電波ログを構築済みの全ての識別器で識別し、どのエリアで取得された Wi-Fi 電波ログであるか識別を試みる、該当するエリアが見つかった場合はエリアを担当する識別器を更新する。該当するエリアが存在しなかった場合は新たなエリア識別器を構築する。

本手順を 3.1 節で得たエリア粒度の閾値毎に実施し、異なるエリア粒度の識別器群をそれぞれ構築する。

4. 移動予測器の構築

本研究で用いた移動予測器はユーザがエリアを出発した時点で移動予測をする。移動予測器は出発するエリアと時刻を入力として扱う拡張マルコフモデルである。例を図 4 に示す。ただし、移動予測確率が 0 になる有向辺は省略してある。なお、ユーザの存在するエリアは、第 3 章で述べたエリア識別器によって判断する。移動予測確率の計算は学習時にユーザが訪れたすべてのエリアに対して行われる。

4.1 拡張マルコフモデル

本研究では位置情報によって作られたマルコフモデルに時間の概念を追加した拡張モデルを構築した。各ノードはエリアを表している。マルコフモデルは一つのノードに対して出発点と到着点の一つずつ与えられているが、拡張モデルは出発点を複数持っており、出発する時間帯ごとに移動確率を算出する。図 4 は 3 つの時間帯 (t_1, t_2, t_3) を設定した例である。ノード内の時間帯は各エリアから出発する時間帯を表しており、各有向辺はある時間帯に出発エ

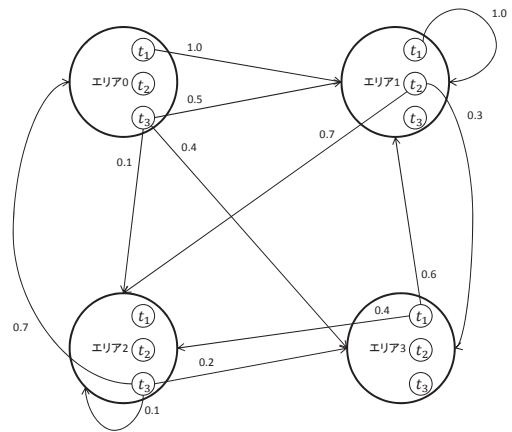


図 4 拡張モデルの例
移動予測確率が 0 になる有向辺は省略

アから到着エリアに移動する確率を表している。

生成した移動モデルの定義式を述べる。移動モデル $G = (A, T, D, E)$ はエリアの集合 A と時間帯の集合 T 、エリアと時間帯に基づく出発点 $D = A \times T$ 、有向辺の集合 $E = D \times A$ からなる。出発エリアを $a_1 \in A$ 、時間帯 $t \in T$ とすると出発点 $d = (a_1, t) \in D$ となる。到着点 $a_2 \in A$ とすると、全ての有向辺 $e = (d, a_2) \in E$ は移動確率 $p(d, a_2)$ を持つ。なお、時間帯 T の要素数を 1 にした場合はマルコフモデルと同一のモデルとなる。

4.2 エリア間移動確率の学習

エリア間の移動確率は出発エリアと到着エリア間の移動回数を元に学習する。時間帯 t にエリア a_1 を出発してエリア a_2 に到着する確率 $p((a_1, t), a_2)$ は式 1 となる。

$$p((a_1, t), a_2) = \frac{\text{時間帯 } t \text{ に } a_1 \text{ から } a_2 \text{ に移動した回数}}{\text{時間帯 } t \text{ に } a_1 \text{ から出発した回数}} \quad (1)$$

ただし、ある時間帯 t にエリア a_1 から出発した回数が 0 の場合、 $p((a_1, t), a_2) = 0/0$ となり移動確率が定義できなくなるため、移動確率を 0 として扱った。

5. エリア粒度別予測確率推定実験

5.1 移動予測に使用したデータ

実験に使用する Wi-Fi 電波ログはサムスン電子製の Galaxy Nexus を使用して収集した。被験者は 11 名で行い、被験者は全員大学生であり、男性が 10 名、女性が 1 名である。11 名の被験者へ Galaxy Nexus を携帯して生活するよう指示し、その間に収集した Wi-Fi 電波ログを使用した。サンプリング周期は 3 秒であり、10 日分のデータを収集した。

5.2 実験内容

実験では時間帯を 3 つの方法で設定して移動予測器を構

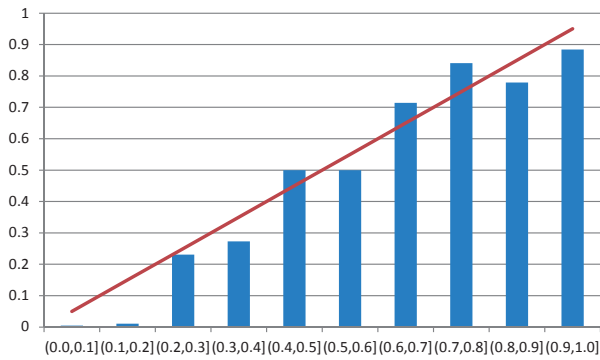


図 5 理想確率と実移動割合の例

築し、それぞれの移動予測モデルを評価して比較した。設定した3つの時間帯をそれぞれ、時間帯なし、共通時間帯、ユーザ別時間帯とする。時間帯なしの移動予測モデルは通常のマルコフモデルとなる。共通時間帯は全ユーザで同一の時間帯を設定した方法であり、朝(4:00~12:00)、昼(12:00~20:00)、夜(20:00~翌4:00)の3つの時間帯を設定した。ユーザ別時間帯はユーザごとに時間帯を自動構築する方法であり、ユーザごとの生活パターンを考慮した方法である。自動構築の詳細な方法は5.3節で述べる。

本実験では5.1節で述べた10日分のデータの内、最初の8日分を学習データとして扱い、最後の2日分をテストデータとして扱った。実験では予測確率を0.1刻みの10段階のビンに分け、各ビンごとに、予測回数、実移動回数、実移動割合を求め、予測スコアを算出した。例えば、ある1度の予測でエリア0からエリア1~4への移動確率が、それぞれ0.150, 0.150, 0.300, 0.400となった場合、(0.1,0.2]のビンの予測回数を2増加させ、(0.2,0.3]、(0.3,0.4]の各ビンの予測回数を1増加させる。その後、実際に移動したエリア、ここでは仮にエリア1であった場合、そのエリアへの移動確率のビンの実移動回数を1増加させる。全ての試行が終わった後に、各ビンの実移動回数を予測回数で割り、実移動割合を算出した。構築したモデルが理想的であった場合、例えば、予測確率が(0.0,0.1]のビンにおける実移動割合は0.050となる。その理想状態にどれだけ近いかを判断する指標として、式2に示す、予測スコアSを導入した、

$$S = 1 - |\text{実移動割合} - \text{理想移動割合}| \quad (2)$$

実移動割合は図5に示した直線のように予測確率が増すごとに増加していくものである。予測スコアSは直線で表される理想成功割合と、棒グラフで示される実際の実移動割合の差を元に算出され、値が1に近いほど理想移動割合と実移動割合の差が小さいものとなる。これは、構築された拡張マルコフモデルが、どれだけ実際のユーザの行動を反映できているかの指標となる。各被験者ごとの予測結果から予測スコアを算出するには被験者1人あたりの予測回数が十分でなかったため、予測スコアは全被験者の実移動

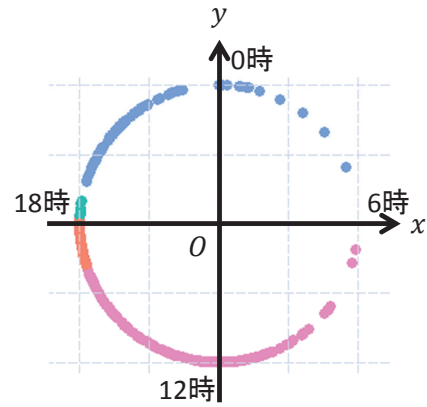


図 6 単位円上の x-means

回数と予測回数から算出した。また、評価に用いる予測スコアSは予測回数が少ないと十分に信頼できないと仮定し、予測回数が10回以下のピンは以降の議論からは除外した。

5.3 ユーザ別時間帯の自動構築

ユーザの行動パターン数はすべてのユーザに対して一定ではないと考えられる。そのため、ユーザごとに最適なクラスタ数を推定することで、時間帯別のエリア間移動数が増加して、予測精度が向上すると考えられる。本研究では、自動的にクラスタ数が決定される x-means を用いた [7]. x-means はデータを k-means により2分割ずつしていく方法である。これを2分割したクラスタが分割する前のクラスタよりも評価値が改善されなくなるまで続けることで入力データをクラスタリングする。本研究では、エリアを出発した時間をクラスタリングすることで、ユーザ別の時間帯を構築した。

本手法で用いた x-means は石岡らが提案した手法をベースとし、0時と24時に同じ意味を持たせるため、図6のようにデータを単位円上に並べ、データ間の角度差を距離とした。出発時刻hに対する極座標のパラメータθは式3で表せる。

$$\theta = f(h) = \frac{\pi(6-h)}{12} \quad (0 \leq h < 24) \quad (3)$$

ある2つの出発時間 h_1, h_2 の距離 $l(0 \leq l \leq \pi)$ は式4で求められる。

$$l = |f(h_1) - f(h_2)| \quad (4)$$

5.4 実験結果

5m 粒度における予測スコアの結果を図7に、10m 粒度における予測スコアの結果を図8に、15m 粒度における予測スコアの結果を図9に示す。グラフの色が灰色になっている予測スコアは移動回数が10回以下となったため評価には使用しなかったスコアを示している。

各エリア粒度、各時間帯での予測回数が11回以上とな

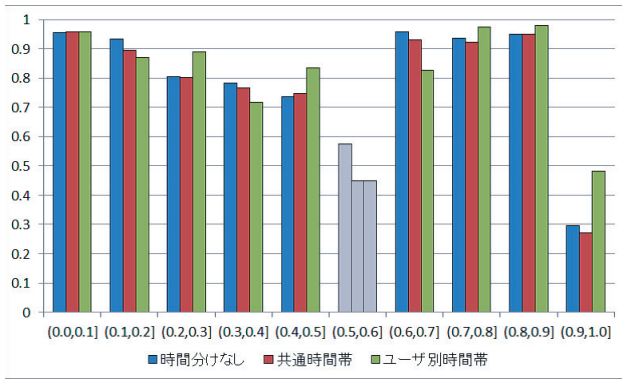


図 7 5 m 粒度での予測スコア

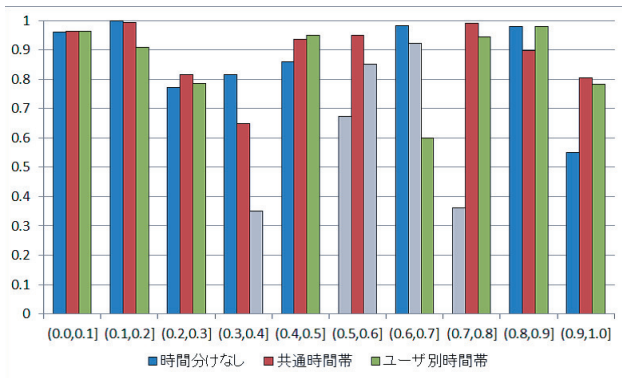


図 8 10 m 粒度での予測スコア

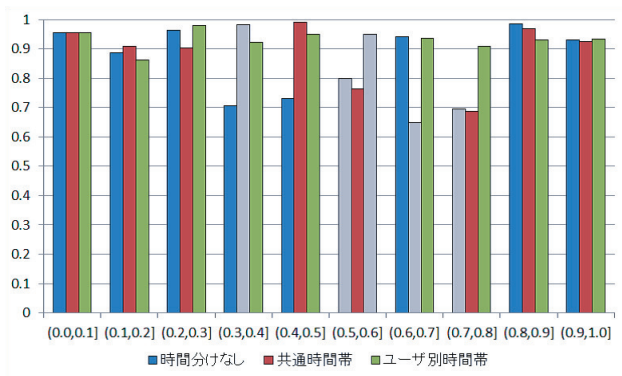


図 9 15 m 粒度での予測スコア

表 1 エリア粒度別の予測スコアの結果

エリア粒度	ユーザ別時間帯	共通時間帯	時間帯なし
半径 5m	0.836 ± 0.149	0.804 ± 0.203	0.817 ± 0.201
半径 10m	0.865 ± 0.123	0.889 ± 0.107	0.865 ± 0.144
半径 15m	0.931 ± 0.031	0.888 ± 0.099	0.888 ± 0.101

るビンの予測スコアの平均をとった結果を表 1 に示す。エリアの粒度が半径 5m,15m のときはユーザ別時間帯での予測スコアが一番高く、半径 10m では共通時間帯での予測スコアが一番高かった。

5.5 考察

5.5.1 半径 5m 粒度の結果の考察

エリア粒度が 5m のときの共通時間帯と時間帯なしの予測スコアを比較すると、(0.4, 0.5] のビンで共通時間帯のスコアの方が高くなっていること以外は時間帯なしの予測スコアの方が高くなっていた。共通時間帯では朝、昼、夜の 3 つの時間帯に分けたことで移動予測に使える学習データ数が十分に確保できなかったために予測スコアが低下してしまったと考えられる。また (0.9, 1.0] のビンのスコアが他と比べて極端に低いことが分かる。(0.9, 1.0] に分類される移動予測について詳しく調べたところ、移動予測毎の学習に使われたデータ数が多くの場合で 2 個以下であり、適切な学習ができていない可能性が高い。構築した移動予測モデルを移動先予測アプリケーションで用いる場合、「N%以上の予測確率が出た場合に予測先として提示する」といった使い方が考えられるため、予測確率が上位のビンで予測スコアが低いことは、移動先予測に著しく不適である。

ユーザ別時間帯と時間帯なしの比較ではユーザ別時間帯の方が予測スコアが高くなり、特に (0.9, 1.0] のビンで予測スコアに約 1.6 倍の差があった。これは、ユーザ別時間帯では無駄な予測の割合が抑制されたために予測スコアの向上に寄与していると考えられる。しかしながら、時間帯無しと比べて予測スコアがよいとしても、その値は 0.5 に満たず、低いスコアとなった。

エリア粒度が 5m のときはユーザ別時間帯の予測スコアが最も高かった。ユーザ別時間帯は時間帯の数が 5.545 ± 1.508 個となり、学習データ不足によってスコアが低下すると予想されたが、(0.9, 1.0] のビンに分類される移動予測で無駄な予測の割合が抑制されたために最も高いスコアとなった。しかし、前述のとおり、(0.9, 1.0] のビンの予測スコアは 0.5 に満たず、移動先予測アプリケーションでの利用には不適である。

5.5.2 半径 10m 粒度の結果の考察

エリア粒度が 10m のときの共通時間帯と時間帯なしの予測スコアを比較すると、共通時間帯のスコアの方が高くなった、これは、エリアの粒度が粗くなったことで、時間帯を朝、昼、夜の 3 つの時間帯に分けても学習データが十分に確保できたためだと考えられる。また、(0.9, 1.0] のビンのスコアが半径 5m 粒度の時に比べて高くなっている。詳しく調べると、無駄な予測の割合が大きく抑制されていたためであった。

ユーザ別時間帯と時間帯なしの比較ではどちらのスコアもほぼ同様となった。これは、ユーザ別時間帯の時間帯の数では学習データが十分に確保できなかった欠点と、学習データを行動パターンを考慮して時間帯に分けた利点が相殺されたためと考えられる。

エリア粒度が 10m のときは共通時間帯の予測スコアが最も高く、予測確率が 0.5 以上となるビンで予測スコアが

0.8を超えているため、移動先予測アプリケーションでの利用は可能であると考えられる。

5.5.3 半径 15m 粒度の結果の考察

エリア粒度が 15m のときの共通時間帯と時間帯なしの予測スコアはほぼ同様となった。これは、今回設定した共通の時間帯が半径 15m での行動パターンにあっていなかったためと考えられる。エリア粒度が粗くなったことによって予測を的中させやすくなったことと、行動パターンにあわない時間帯を設定したことによって、エリア粒度が 10m の時と比べ、予測スコアが上昇しなかったと考えられる。そのため、時間帯なしの予測スコアとほぼ同様となったと考えられる。

ユーザ別時間帯と時間帯なしの予測スコアを比較するとユーザ別時間帯のスコアの方が高かった。これは、エリアの粒度が粗くなったことで、ユーザ別時間帯の時間帯の数でも十分に学習データが確保されたためだと考えられる。

エリア粒度が 15m のときはユーザ別時間帯の予測スコアが最も高く、予測確率が 0.5 以上となるビンで予測スコアが 0.9 を超えているため、移動先予測アプリケーションでの利用は可能であると考えられる。

5.5.4 考察のまとめ

結果から、エリアの粒度が半径 10m 以上で時間帯を考慮した予測が有効である可能性が示唆された。特に半径 15m 程度でユーザ別時間帯を用いた移動予測モデルが一番精度が高いが、半径 10m 程度の粒度からは共通時間帯を用いた移動予測モデルが一番精度が高いと考えられる。半径 10m の円の面積は、1 住宅あたりの敷地面積の全国平均である 292m²[8] に近いため、時間帯を考慮した移動先予測手法は建物間移動レベルの個人移動予測に有効である可能性が示唆された。

6. おわりに

人間の行動パターンを考慮して、位置情報と他の情報を併せて移動予測をした研究は多く、特に時間情報を用いた移動予測は多くのデバイスで利用可能な点で利便性が高い移動予測手法だといえる。しかし、この手法はエリアの粒度を考慮した研究が十分になされておらず、エリアの粒度を細かくしすぎるとエリア間あたりの移動回数が減り、学習に使えるデータ数が減るために予測精度が低下する可能性があった。

本研究では、マルコフモデルを用いた位置情報のみで移動予測をする手法、マルコフモデルを拡張して全ユーザで共通の時間帯を用いてエリアを出発する時間帯を考慮した手法およびユーザごとに時間帯を自動構築してエリアを出発する時間帯を考慮する方法のそれぞれで移動予測モデルの確率の推定の尤もらしさを評価し、どの粒度であれば、時間帯を考慮した移動予測が時間帯を考慮しない予測よりも高い精度を示し、かつ、移動予測に適した特性を持つか検

討した。結果、エリアの粒度を 5m とした場合では、ユーザごとに時間帯を自動構築した方法が予測スコアが最も高かったが、最も確率が高い (0.9, 1.0] のビンを高い精度で予測できていないことから、半径 5m 粒度では時間帯を考慮した移動予測が有効であるとはいえないと考えられる。

エリアの粒度を半径 10m 以上にした場合で、時間帯を考慮した移動予測の方が時間帯を考慮しない移動予測よりも高いスコアとなった。半径 10m では共通時間帯を用いた場合の予測スコアが最も高く、予測確率が 0.5 以上となるビンで予測スコアが 0.8 以上となり、移動先予測アプリケーションでの利用に適していると考えられる。半径 15m ではユーザ別に時間帯を自動構築した場合の予測スコアが最も高く、予測確率が 0.5 以上となるビンで予測スコアが 0.9 以上となり、移動先予測アプリケーションでの利用に適していると考えられる。従って、時間帯を考慮した移動予測は建物間の移動を予測するといった使い方に対して有効である可能性が示された。

今回の実験では、ユーザの移動を時刻のみを使用して予測したが、人間の行動は曜日や月日でも大きく変わると考えられる。従って、時刻と併せて曜日や月日も用いてユーザの行動をパターン化した移動予測を今後の課題としたい。

謝辞 本研究は JSPS 科研費 24700062 の助成を受けたものである。

参考文献

- [1] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, Vol. 7, No. 5, pp. 275–286, 2003.
- [2] 西野正彬, 山田智広, 瀬古俊一, 茂木学, 武藤伸洋, 阿部匡伸. 時刻情報を含む特徴滞在パターンを用いた行動予測方式の検討. 電子情報通信学会総合大会講演論文集, Vol. 2009, No. 1, p. 125, 2009.
- [3] Lin Liao, Donald J Patterson, Dieter Fox, and Henry Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, Vol. 171, No. 5, pp. 311–331, 2007.
- [4] 西野正彬, 中村幸博, 八木貴史, 武藤伸洋, 阿部匡伸. ダイナミックベイジアンネットワークを用いたあいまいな表現を含むスケジューラデータと gps データからの行動予測. 電子情報通信学会技術研究報告. LOIS, ライフインテリジェンスとオフィス情報システム, Vol. 110, No. 141, pp. 29–34, 2010.
- [5] Manlio De Domenico, Antonio Lima, and Mirco Musolesi. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing*, Vol. 9, No. 6, pp. 798–807, 2013.
- [6] 藪圭輔, 榎堀優, 間瀬健二. レイリー分布を用いた wi-fi 電波強度分布に基づく少数標本時位置推定. 情報処理学会研究報告. UBI, [ユビキタスコンピューティングシステム], Vol. 2014, No. 23, pp. 1–8, 2014.
- [7] 石岡恒憲. x-means 法改良の一提案: k-means 法の逐次繰り返しとクラスターの再併合. 計算機統計学, Vol. 18, No. 1, pp. 3–13, 2006.
- [8] 総務省 統計局. 社会生活統計指標- 都道府県の指標- 2013. <http://www.e-stat.go.jp/SG1/estat/>.