

滞在時間帯と経路情報を用いた 混合最大エントロピー逆強化学習に基づく早期目的地予測

小西達也^{1,a)} 下坂正倫^{2,b)}

概要：ユーザの位置履歴情報に基づく目的地の推定はユビキタスコンピューティングの中心的な応用課題の一つである。この問題に対して活動時間帯に基づくものや、逆強化学習に基づくものが提案されてきたが、前者は移動時の経路情報を加味しないため最終的な目的地推定の精度を上げにくい、後者は移動開始直後では経路情報を利用できないために移動開始直後の推定精度が低いという問題がある。本研究は、これら2つの枠組みを統合し、移動直後から一定水準の精度を保証しつつ、移動するにつれて目的地予測精度を向上させていく手法の実現を目的とする。具体的には、滞在場所と時間帯に関係する分布を混合比とする、最大エントロピー逆強化学習 (MEIRL) を用いた統計モデルを提案する。スマートフォンにより取得された位置履歴データを用いた実験により、提案手法が既存の目的地予測手法と比べて高い精度で目的地を予測できることを確認した。

キーワード：目的地予測, 逆強化学習, 経路情報, 滞在パターン

1. 緒論

近年のスマートデバイス普及により、ユーザの情報を収集することが容易になってきた。ユーザの日常生活を様々なデータから生活パターンを把握、推定することができ、ユーザの未来の行動を予測することが可能となる。特に、位置情報はユーザの状況、行動などのコンテキストを把握するのに最も一般的に使われるものであり、その商業的、経済的な利用価値の高さからもユビキタスコンピューティングの領域では研究が盛んに行われている。

その中でも、目的地予測、もしくは経路予測は行動データ利活用の中心的な応用課題の一つと言える。事故や混雑を回避するための迂回路の提示や目的地における予定や天気のお知らせ [1] は勿論のこと、途中経路におけるパーソナライズされた広告の表示、車であれば最適な給油、給電計画の提案 [2], [3], スマートフォンの基地局切替の最適化 [4] など応用の幅は広い。

目的地を予測するにあたり、まず考えられるのは出発地

点や出発時刻などの滞在パターンである。これに着目すると、例えば夕方に職場を出発すればその目的地は自宅である可能性が高い、すなわちその移動行動は帰宅であるというような予測を行うことができる。しかし、移動時の経路情報を加味しないため最終的な目的地推定の精度を上げにくいという問題がある。Manasseh ら [5] は、Decision Tree などの多クラス識別の枠組みで目的地推定に取り組み、Nadembega ら [6] は、多クラス識別モデルに加えて階層的なクラスタリングを用いて目的地予測精度の向上を図った。これらは特徴量として現在地やそれ以前の位置情報を組み込んではいないものの、経路を包括的に表現できるモデルではないため、予測精度に改善の余地がある。

一方で、移動パターンを観察することでその目的地を予測するという経路情報に着目したアプローチも考えられる。直感的には、自宅を出た瞬間には複数の候補から目的地が絞れなくても、例えば次に経由したのが最寄りの駅なのか、バス停なのかという情報がわかるだけでも、目的地を絞り込むのに役立つのは容易に推察できる。しかし今度は逆に、出発直後においては経路情報を用いることができないために、移動行動の初期において目的地推定精度に難がある。Xue ら [7] は、地点から地点への遷移確率を元に目的地に至る確率を計算する手法を取ったが、経路以外の情報を使っておらず出発地点での目的地予測精度に問題があった。一方 Ziebart ら [8] は、逆強化学習を目的地予測に

¹ 東京大学
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, JAPAN 113-8656

² 東京工業大学
Tokyo Institute of Technology, 2-12-1 Oookayama, Meguro-ku, Tokyo, JAPAN 152-8552

a) tkonishi@miubiq.cs.titech.ac.jp

b) simosaka@miubiq.cs.titech.ac.jp

応用できることを示したが、研究の主眼はあくまで与えられた目的地に対する経路予測であった。マルコフモデルに着目して位置と位置の間の遷移確率をモデル化する研究も多い [9], [10]。Huang ら [11] は、位置情報と行動ラベルの組み合わせに対して HMM を適用し、目的地に対して移動や行動のラベルが周期的であることを明確に示した。しかし、遷移のみに着目しているため、時刻などの人の行動に大きく影響を与える要素についてはまだ考慮の余地がある。Krumm ら [12], [13] は、目的地の事前確率と、最短経路と実際の運転経路から計算される距離の差や経過時間の尤度を組み合わせてモデル化を行っている。しかしこれらは目的地のみに依存した定式化であり、目的地へ向かう出発地からの経路情報を十分に考慮できていない問題点がある。

目的地予測に対する上述の 2 つのアプローチ、すなわち滞在パターンに着目した手法と経路情報に着目した手法に内在する問題点は、相互に補うことが可能である。本論文では、これら 2 つの枠組みを統合し、経路情報が少ない移動開始時には出発地や時刻などの滞在パターンを考慮し、移動が進むにつれて経路情報を加味していく。これにより、移動開始直後から予測精度を一定水準担保しつつ、移動するにつれて予測精度の向上が期待できる。本研究では、このように 2 つのアプローチを統合した早期目的地予測モデルの構築を目的とする。

本研究の主眼はあくまで滞在時間帯と経路情報を統合した早期目的地予測モデルの構築であるが、経路情報を取り扱う際、路線の乗り換えなど、移動手段を考慮した経路の候補を活用する枠組みも新たに提案する。Krumm らの研究 [13] では、ユーザが行ったことのない目的地も予測対象とする Open World な予測モデルを構築したことは画期的であったものの、自動車での移動を前提に置いたこのモデルでは、目的地への到達確率が既に行ったことのある目的地からの物理的な距離に依存しているという仮定を置いている。しかし特に都心などの交通網が発達した地域においては、人の日常生活における駅間、ひいては目的地の物理的な距離は意味を持たなくなるため、不十分な仮定と言える。そのため本研究では都心における生活の足である電車の経路に着目し、路線検索 API により取得された経路候補を活用することで Open World な予測モデルを構築するという新たな枠組みを提案する。

本研究の貢献は以下のようにまとめられる。まず、移動開始直後から一定の予測精度を保ち、移動につれて予測精度を向上させていく目的地予測モデルの構築である。次に、経路情報の扱いにおいて、路線検索 API から取得した移動経路の候補を活用する新たな枠組みの提案である。

以降、2 節では提案モデルの詳細について述べ、3 節では用いる特徴量の設計について記述する。4 節では、提案モデルの性能評価のためスマートフォンを用いて複数人、数週間に及ぶ実際の位置情報を収集し、そのデータセット

を利用して性能評価実験を行った。最後に 5 節で本研究の貢献と将来課題をまとめた。

2. 早期目的地予測モデルの構築

提案手法では、移動開始直後から一定水準の精度を保証しつつ、移動に従って目的地予測精度を向上させていく数理モデルを構築する。具体的には、最大エントロピー逆強化学習により算出される経路情報の尤度から、滞在時間帯を入力とする多クラス識別モデルによる目的地への到達確率を混合比として目的地の事後確率を計算する。

2.1 問題設定

提案手法においては目的地とそこに至る経路上の位置情報を離散化して扱うため、クラスタリングなどの前処理を施す必要がある。このとき、得られた位置情報データセットから、 L 個の位置クラスタが得られたとする。さらに、得られた離散化された位置情報から、予測する「目的地」の候補としていくつかの位置クラスタを抽出する。この抽出された位置クラスタを本研究では「停留点」と呼び、以下では K (ただし $K \leq L$) 個抽出されたとし、目的地を表現する確率変数を $y \in \{1, 2, \dots, K\}$ とおく。このうちどこに向かっているのかの予測を本研究における問題設定とする。位置情報の離散化、停留点の抽出に関する詳細は 4.1 節で示す。

また、本論文では離散化された L 個の位置ラベル $\{1, 2, \dots, L\}$ を遷移する毎にステップ t が 1 ずつ更新されるとし、ステップ t における位置ラベルを ℓ_t と表記する。経路情報以外の、滞在パターンを示す静的なコンテキスト情報も同様に \mathbf{x}_t と表記する。また、簡単のため $\ell_{1:t} \equiv \{\ell_1, \ell_2, \dots, \ell_t\}$, $\mathbf{x}_{1:t} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$ と表現する。

2.2 滞在時間帯による目的地確率を混合比とした

経路情報に基づく最大エントロピー逆強化学習モデル
上述のように、経路情報が少ない移動開始時には滞在パターンから目的地を予測しつつ、移動が進むにつれて経路情報を加味する数理モデルを構築し、ステップが小さいうちから精度の良い早期目的地予測の実現を目指す。ステップ t において目的地が k である確率を、

$$p(y = k | \mathbf{x}_{1:t}, \ell_{1:t}) = p(y = k, \ell_{1:t} | \mathbf{x}_{1:t}) \quad (1)$$

$$\propto p(y = k, \ell_{1:t} | \mathbf{x}_{1:t}) \quad (2)$$

$$\propto p(y = k | \mathbf{x}_{1:t}) p(\ell_{1:t} | y = k) \quad (3)$$

のように定義する。この式において、 $p(y = k | \mathbf{x}_{1:t})$ の部分が静的なコンテキスト情報、すなわち滞在時間帯に応じた目的地の分布を表し、 $p(\ell_{1:t} | y = k)$ が経路情報により計算される尤度を表す。また、以上の提案モデルにおいて予測される目的地 \hat{k} は、 $\hat{k} = \operatorname{argmax}_k p(y = k | \mathbf{x}_{1:t}, \ell_{1:t})$ により

求められる。 \hat{k} が複数ある場合には、ランダムにそれらの内 1 つを選択し予測結果とする。

2.2.1 滞在時間帯による目的地確率

滞在時間帯による目的地確率 $p(y = k | \mathbf{x}_{1:t})$ を多クラス識別手法により算出する。ユーザにより滞在パターンが異なることを考慮し、この多クラス識別器はユーザ毎に学習を行う。ここで停留点の抽出は全ユーザ共通で行っているため、特定のユーザの過去の行動履歴、すなわち訓練データセットに含まれない目的地をも予測結果に含める必要がある。本研究では、訓練データセットに含まれる停留点とその集合を $u \in U$ 、含まれない停留点とその集合を $v \in V$ としたとき、

$$p(y = u | \mathbf{x}_{1:t}) = (1 - \alpha) \frac{\exp(\mathbf{w}_u^T \mathbf{f}_s(\mathbf{x}_{1:t}))}{\sum_k^K \exp(\mathbf{w}_k^T \mathbf{f}_s(\mathbf{x}_{1:t}))} \quad (4)$$

$$p(y = v | \mathbf{x}_{1:t}) = \frac{\alpha}{|V|} \quad (5)$$

のように目的地確率を計算した。ただし、 $|V|$ は集合 V に含まれる停留点の数、 α は訓練データセットに含まれない停留点を選択する確率を表すモデルパラメータ、 \mathbf{f}_s は滞在時間帯に関する特徴量ベクトルであり、 \mathbf{w}_k は停留点 k に対する特徴量ベクトルの重みパラメータである。式 (4) では多クラス対数線形モデルをその学習法の一つである FOBOS[14] により最適化して用いた。また、正則化には l_2 正則化を用いた。上記では滞在時間帯に基づく目的地確率の算出には対数線形モデルを用いたが、時間帯に基づく滞在場所を分布で表現する手法 [15] の利用も可能である。

2.2.2 逆強化学習により算出される経路情報の尤度

経路情報による尤度 $p(\ell_{1:t} | y = k)$ を計算するにあたって、本研究では最大エントロピーマルコフ決定過程 [8] により定式化を行う。従って、2.2 節の式により、提案する目的地予測モデルは混合最大エントロピーマルコフ決定過程となる。

ここで、ステップ t から $t+1$ において位置ラベルが ℓ_t から ℓ_{t+1} に遷移したとき、マルコフ決定過程における状態を ℓ_t 、行動を ℓ_{t+1} と設定する。マルコフ決定過程においては状態 ℓ_t に対する報酬関数 $R(\ell_t)$ の設定が重要であるが、経路に関する報酬を正しく設定することは難しい。そのため、理想的な経路情報から最適報酬関数を設定するための枠組みとして、停留点 k に対する報酬関数が以下のようなパラメータ θ_k により表現されるものと仮定し、事前に取得した位置履歴データから θ_k を以下の規範により最適化する。詳細な手続きは [8] を参照のこと。

$$R(\ell_t | y = k) = \theta_k^T \mathbf{g}_r(\ell_t) \quad (6)$$

$$p(\ell_{1:T} | y = k) \propto \exp \left\{ \sum_{t=1}^T R(\ell_t | y = k) \right\} \quad (7)$$

$$\hat{\theta}_k = \operatorname{argmax}_{\theta} \left\{ \sum_{\ell_{1:T} \in \mathcal{D}} \ln p(\ell_{1:T} | y = k) - \Omega(\theta) \right\} \quad (8)$$

ただし $\Omega(\theta)$ は正則化項を表し、ここでは l_1 正則化すなわち $\Omega(\theta) = \gamma \|\theta\|_1, \gamma \geq 0$ とした。また \mathcal{D} は経路情報の訓練データセットである。これにより、停留点 k に対する、ステップ t までの実際の経路 $\ell_{1:t}$ の尤度は、

$$p(\ell_{1:t} | y = k) = \frac{1}{Z_k} \exp \left\{ \sum_{\tau=1}^t R(\ell_{\tau} | y = k) \right\} \quad (9)$$

と計算される。ここで Z_k は分配関数である。また、報酬関数の学習に用いる訓練データ、すなわち理想的な経路として、本研究では路線検索 API*1 で取得した経路駅の緯度、経度の点列を与えた。路線検索 API により理想的な経路を取得しているため、収集した経路データに履歴のない出発地、目的地であっても予測が可能である点が本研究の特徴である。

3. 特徴量設計

提案手法においては、式 (4)、式 (5) で表される多クラス対数線形モデルと、式 (9) で表される最大エントロピー逆強化学習という 2 つの学習器を用いる。それぞれに対して、以下のように特徴量 $\mathbf{f}_s, \mathbf{g}_r$ を設計した。

3.1 滞在時間帯に関する特徴量

式 (4) で用いる特徴量 $\mathbf{f}_s(\mathbf{x}_{1:t})$ について、その設計を述べる。本研究では滞在時間帯を表現する特徴量として、以下の 3 つの観点から特徴量 $\mathbf{f}_{\text{dep}} \in \mathbb{R}^L, \mathbf{f}_{\text{time}} \in \mathbb{R}^H, \mathbf{f}_{\text{week}} \in \mathbb{R}^2$ を設計し、 $\mathbf{f}_s(\mathbf{x}_{1:t}) = \mathbf{f}_{\text{dep}}(\mathbf{x}_{1:t}) \otimes \mathbf{f}_{\text{time}}(\mathbf{x}_{1:t}) \otimes \mathbf{f}_{\text{week}}(\mathbf{x}_{1:t}) \in \mathbb{R}^{2LH}$ とする。

出発地点 L 個の位置ラベルのうち、出発地点のラベルを ℓ_{dep} とする。このとき、出発地点を表現する特徴量 $\mathbf{f}_{\text{dep}}(\mathbf{x}_{1:t})$ は、

$$\mathbf{f}_{\text{dep}}(\mathbf{x}_{1:t}) = \left[f_{\text{dep}}^{(1)} f_{\text{dep}}^{(2)} \cdots f_{\text{dep}}^{(L)} \right] \quad (10)$$

$$f_{\text{dep}}^{(i)} = \begin{cases} 1 & (\text{if } i = \ell_{\text{dep}}) \\ 0 & (\text{otherwise}) \end{cases} \quad (11)$$

なる L 次元ベクトルの 1-of- L 表現である。

出発時刻 出発した時刻を h として、24 時間を H 分割した時間帯のそれぞれについて h との近さをガウシアンで平滑化したものを特徴量として用いる。具体的には、出発時刻を表現する特徴量 $\mathbf{f}_{\text{time}}(\mathbf{x}_{1:t})$ は、

$$\mathbf{f}_{\text{time}}(\mathbf{x}_{1:t}) = \left[f_{\text{time}}^{(1)} f_{\text{time}}^{(2)} \cdots f_{\text{time}}^{(H)} \right] \quad (12)$$

$$f_{\text{time}}^{(i)} = \frac{\exp \left\{ -\frac{(h - \min(i, H-i))^2}{2} \right\}}{\sqrt{2\pi}} \quad (13)$$

なる H 次元のベクトルである。

曜日 人の生活パターンが平日と休日とで異なることを仮

*1 駅すばあと Web サービス。 <https://ekiworld.net/service/sier/webservice/api.html>

定し、平日か休日かを考慮する特徴量を作成する。この特徴量 $f_{\text{week}}(\mathbf{x}_{1:t})$ は、

$$f_{\text{week}}(\mathbf{x}_{1:t}) = \begin{cases} [01] & (\text{if weekday}) \\ [10] & (\text{if weekend}) \end{cases} \quad (14)$$

なる2次元のベクトルである。

3.2 経路情報に関する特徴量

ある停留点に向かう際に取り得る経路が複数存在することを考える。それら複数の経路は、選択のされやすさ、そこから到達可能な停留点の数など、目的地を予測するにあたり重要な要素が異なる可能性があり、すべてを同様に扱うべきではない。そのため、ステップ t における位置 l_t を入力として、候補として得られる n 個の経路毎に抽出された特徴量を並べたベクトル $\mathbf{g}_r(l_t) \in \mathbb{R}^n$ を経路情報に関する特徴量として利用する。本研究で利用する路線検索APIでは出発駅と目的駅を指定した上で「早さ」「安さ」「乗換回数の少なさ」の基準によって経路が取得でき、それぞれの候補のうち最上位であったものを利用する。すなわち $\mathbf{g}_r(l_t) \in \mathbb{R}^3$ である。また、出発駅と目的駅は、それぞれ出発地と目的地の緯度、経度から最寄りの鉄道駅を取得し利用している。

特徴量として、「早さ」「安さ」「乗換回数の少なさ」の基準で取得された理想的な経路 $\tilde{\zeta} = [\tilde{\zeta}_1 \tilde{\zeta}_2 \tilde{\zeta}_3]$ のそれぞれに対して、与えられた位置 l_t がどれだけ距離的に近いのかを表現することを考える。ここで、 $d(\tilde{\zeta}_i, l_t)$, ($i = 1, 2, 3$) は理想的な経路 $\tilde{\zeta}_i$ 上の点を結んだ線と、実際の経路上の点 l_t の実世界上の最短距離を表すとする。これらを用いて、特徴量ベクトル $\mathbf{g}_r(l_t) \in \mathbb{R}^3$ を以下のように計算する。なお、 $\lfloor x \rfloor$ は x を越えない最大の整数を表す。

$$\mathbf{g}_r(l_t) = \left[g_r^{(1)} g_r^{(2)} g_r^{(3)} \right] \quad (15)$$

$$g_r^{(i)} = \begin{cases} X - \left\lfloor \frac{d(\tilde{\zeta}_i, l_t)}{D} \right\rfloor & (\text{if } d(\tilde{\zeta}_i, l_t) \leq XD) \\ 0 & (\text{otherwise}) \end{cases} \quad (16)$$

すなわち、理想的な経路 $\tilde{\zeta}_i$ と実際の経路上の点 l_t の距離 $d(\tilde{\zeta}_i, l_t)$ が D 未満であれば特徴量 $g_r^{(i)} = X$ 、 $2D$ 未満であれば $g_r^{(i)} = X - 1$ というように、 $\tilde{\zeta}_i$ と l_t が近いほど離散的に値が大きくなる特徴量である。

4. 性能評価実験

滞在パターンと経路情報を組み合わせた提案手法において、移動開始時、移動中ともに安定した目的地予測性能を達成できるか検証するための実験を行った。実験に用いたデータは被験者10名による実際の位置履歴情報である。

4.1 位置情報の収集

本研究の提案モデルの性能評価を行うための、日常的な

表 1 収集したユーザ毎の位置情報数

ユーザ番号	位置情報数	ユーザ番号	位置情報数
1	5445 個	5	2504 個
2	128 個	6	4780 個
3	3951 個	7	4001 個
4	34 個		

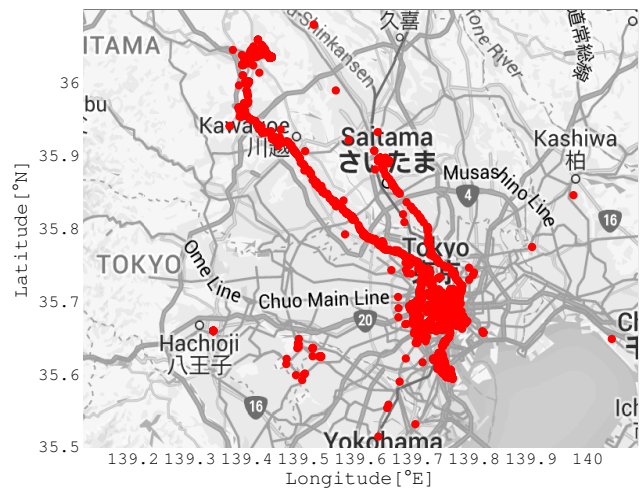


図 1 収集した位置情報 (20843 点)

生活行動データを取得しデータセットを構築した。データ収集期間は2015年5月11日～5月31日の21日間で、被験者は年齢、性別不問の成人10名(うち位置情報を取得できたのは7名)である。被験者の所持するAndroid端末に、データ収集用の専用アプリケーションをインストールし、位置情報を含む様々なデータを取得した。特に位置情報に関しては、GPSまたはネットワークの情報を利用して5分間隔で取得し、計20843点の位置情報を取得した。位置情報には、緯度・経度に加え取得した時刻も含まれる。ユーザ毎に収集された位置情報の数は表1の通りである。図1に、実際に収集された位置情報の一部地域を拡大してプロットした。

4.1.1 位置情報の離散化

本研究では位置情報の離散化には、凝集型階層クラスタリングの一つである群平均法 (UPGMA: Unweighted Pair Group Method with Arithmetic mean) [16]を用いた。UPGMAでは位置情報の密度を考慮してクラスタリングを行うため、外れ値に対して頑健であるという特徴を持つ [17]。例えば自宅や職場など位置情報が多く取得されると考えられる地点において、他の点の影響を受けにくく、直感的なクラスタリングができると期待される。

クラスタリングした結果、20843点の位置情報は99点に離散化された。図2に、クラスタリングされた位置の一部地域を拡大してプロットした。

4.1.2 停留点の抽出

上記のクラスタリングにより得られた離散化された位置情報から、個人毎の経路データが得られる。この経路デー

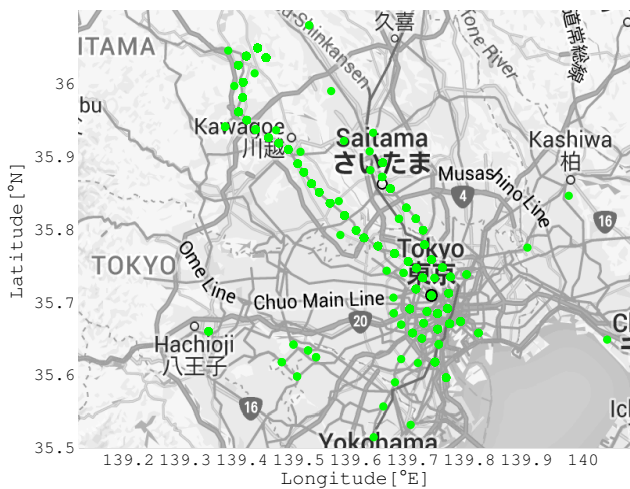


図 2 全ユーザの位置情報のクラスタリング結果 (99 点)

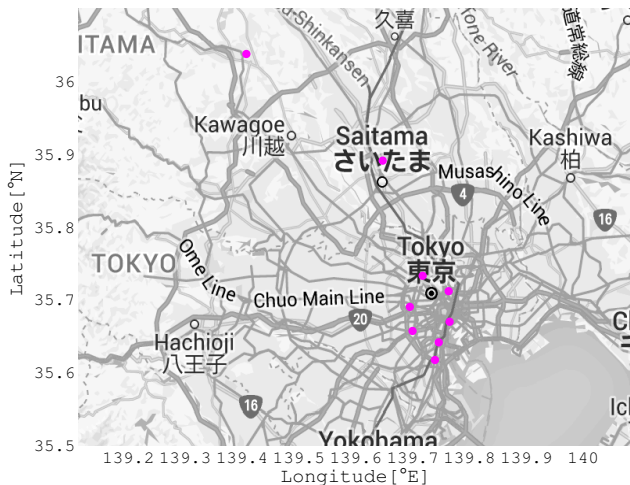


図 3 抽出された停留点 (9 点)

タから、ある一定時間以上滞在した地点を、目的地の候補として扱うべき「停留点」として抽出する。ここでは、20分以上滞在している点を停留点として計9点抽出した。また、図3に、抽出された停留点をプロットした。

この結果経路データが計245個抽出できた。さらに本研究では遷移による目的地確率の推移を観察するため、遷移が4回以上起きている経路データ、すなわち出発地点を含めて位置情報5点以上から構成される経路データのみをデータセットとした。また、旅行と思われる移動を除くため、関東以外の地点が含まれる経路データは除外した。この結果、最終的にデータセットに含まれる経路データは計20個であった。

また、出発地と目的地が同じテストデータ3つを図4にプロットした。必ずしも同じ位置ラベルを経由しているわけではないことがわかるが、これは位置情報を取得するタイミング(本研究では5分毎)によっては同一のラベルが取得されるとは限らないためである。このように位置情報をサンプリング、さらに離散化した結果、得られる経路上

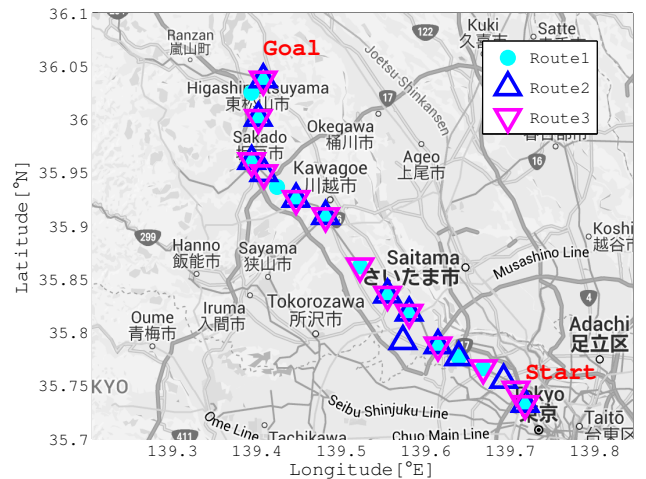


図 4 出発地と目的地が同じ3つの経路

の位置ラベルが厳密に等しくなくても目的地予測が可能であることも本実験で検証する。

4.2 比較手法

本研究では、出発時点からある一定以上の精度を保証しつつ、経路を経る毎に精度を向上させていく目的地予測モデルの構築を目的とする。そのため、比較手法には出発時点における情報のみを使った目的地推定モデルと、経路情報のみを使った目的地推定モデルを用いた。具体的には、提案モデルでも利用している、式(4)と式(5)で表現される多クラス対数線形モデルと、Ziebartら[8]の提案している逆強化学習によって学習された経路情報による予測モデル

$$p(y = k | l_{1:t}) \propto p(l_{1:t} | y = k) p(y = k) \quad (17)$$

である。ここで $p(y = k)$ は目的地の事前分布であり、ここでは $p(y = k) = 1/K$ とした。また、提案手法と同様に予測された目的地 \hat{k} は、 $\hat{k} = \underset{k}{\operatorname{argmax}} p(y = k)$ として求める。ただし \hat{k} が複数ある場合は、その内の1つをランダムに選択する。

4.3 実験結果

データセットに対して、Leave-one-out 交差検定 [18] で性能評価した結果を図5にプロットした。ここで、式(12)、式(13)に示した出発時間帯に関する特徴量パラメータ $H = 8$ 、式(16)に示した経路情報に関する特徴量パラメータ $X = 5, D = 100 \text{ m}$ 、式(4)、式(5)に示した訓練データセットに含まれない停留点を予測結果とする確率 $\alpha = 0.25$ を設定した。横軸は、目的地予測を行ったステップ t の、経路全体の長さ T に対する割合である。縦軸は予測の正解率を表し、値が大きいほど高性能である。

移動開始時の精度に着目すると、提案手法の予測精度は多クラス識別モデルのそれと等しい。すなわち、経路情報を利用できない移動開始時においては、滞在パターンによ

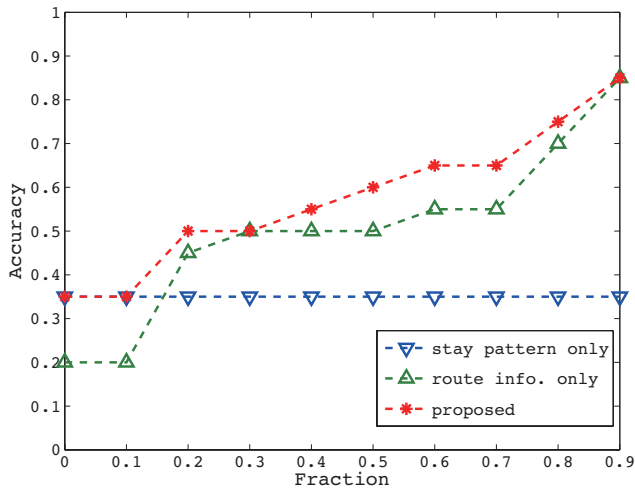


図 5 目的地予測精度

る予測を利用していることが分かる。次に、経路の経過につれて提案手法の精度が上昇していくことがわかる。これは経路情報を利用していくことで段々と目的地が絞られていくことを示している。また、滞在パターンをも考慮したことで、移動中の目的地予測精度において、経路情報のみから目的地予測したものよりも若干の精度向上が確認できる。

以上、実際の位置情報を用いた実験により、提案手法が多クラス識別モデルと経路情報とを統合的に利用し、双方の利点を上手くいかして精度良く目的地予測を行えることが定量的に確認された。

5. 結論

本研究では、滞在時間帯などの静的な情報と経路情報を組み合わせることにより、移動開始直後から一定水準の精度を保ちつつ、経路を経る毎に精度を向上させていく目的地予測モデルを構築した。提案モデルにおいて、特定の目的地 k に対する確率 $p(y = k | x_{1:t})$ と、経路情報から計算できる尤度 $p(\ell_{1:t} | y = k)$ はそれぞれ独立に計算できる。

前者に関しては、既存の多クラス識別モデルの枠組みがそのまま利用できるため、特徴量の設計を含め更なる精度向上の余地がある。後者の学習では、路線検索 API を利用することにより訓練データセットにない出発地と目的地に対しても尤度が計算できるという特徴がある。

実験の結果、既存研究の 2 つのアプローチの問題点を解決、すなわち移動開始直後の精度を向上しつつ移動につれて目的地予測精度を向上していく早期目的地予測が可能であることを確認した。特に、本研究では 5 分間隔という決して高頻度ではないサンプリングレートで取得された位置情報を、更にクラスタリングで離散化しているが、厳密に同じ離散化された位置ラベル列を経路データとして得られなくても、目的地が精度良く予測できることが確認できた。

本研究ではユーザの移動経路が電車の経路に深い関係が

あると仮定し、路線検索 API に出発地の最寄り駅と、目的地の最寄り駅を用いた。しかし車や徒歩、自転車による移動も日常生活における重要な移動手段であり、複数の交通手段を加味した検索結果を利用し、それに適合した特徴量の設計により本アプローチの適用範囲が広がることが期待される。

参考文献

- [1] Marmasse, Natalia et al.: Location-aware information delivery withcommotion, *Proc. of HUC*, pp. 157–171 (2000).
- [2] Horvitz, Eric et al.: Some help on the way: Opportunistic routing under uncertainty, *Proc. of UbiComp*, pp. 371–380 (2012).
- [3] Kobayashi, Yuichi et al.: A route search method for electric vehicles in consideration of range and locations of charging stations, *Proc. of IV*, pp. 920–925 (2011).
- [4] Pathirana, Pubudu N et al.: Mobility modelling and trajectory prediction for cellular networks with mobile base stations, *Proc. of MobiHoc*, pp. 213–221 (2003).
- [5] Manasseh, Christian et al.: Predicting driver destination using machine learning techniques, *Proc. of ITSC*, pp. 142–147 (2013).
- [6] Nadembega, Apollinaire et al.: A destination prediction model based on historical data, contextual knowledge and spatial conceptual maps, *Proc. of ICC*, pp. 1416–1420 (2012).
- [7] Xue, Andy Yuan et al.: Destination prediction by sub-trajectory synthesis and privacy protection against such prediction, *Proc. of ICDE*, pp. 254–265 (2013).
- [8] Ziebart, Brian D et al.: Maximum Entropy Inverse Reinforcement Learning., *Proc. of AAAI*, pp. 1433–1438 (2008).
- [9] Ashbrook, Daniel et al.: Learning significant locations and predicting user movement with GPS, *Proc. of ISWC*, pp. 101–108 (2002).
- [10] Hariharan, Ramaswamy et al.: Project Lachesis: parsing and modeling location histories, *GIS*, pp. 106–124 (2004).
- [11] Huang, Wenhao et al.: Hierarchical destination prediction based on GPS history, *Proc. of FSKD*, pp. 972–977 (2013).
- [12] Krumm, J.: Real time destination prediction based on efficient routes, Technical report, SAE (2006).
- [13] Krumm, John et al.: Predestination: Inferring destinations from partial trajectories, *UbiComp*, pp. 243–260 (2006).
- [14] John C. Duchi et al.: Efficient Online and Batch Learning Using Forward Backward Splitting, *JMLR*, pp. 2899–2934 (2009).
- [15] McInerney, James et al.: Modelling heterogeneous location habits in human populations for location prediction under data sparsity, *Proc. of UbiComp*, pp. 469–478 (2013).
- [16] Sokal, R. R.: A statistical method for evaluating systematic relationships, *UKSB*, Vol. 38, pp. 1409–1438 (1958).
- [17] Everitt, Brian S et al.: *Clustering analysis* (2001).
- [18] Kohavi, Ron et al.: A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proc. of IJCAI*, pp. 1137–1145 (1995).