

時間粒度が粗密な大規模測位データを用いた 移動手段推定に関する研究

東浦 圭亮¹ 田村 直樹¹ 平野 流¹ 鷺田 健斗¹ 坂倉 波輝¹ 山口 公平¹ 浦野 健太¹ 米澤 拓郎¹
河口 信夫^{1,2}

概要: スマートフォンの普及が進むとともに、位置情報データの収集が容易となり、位置情報データを用いた研究が盛んに行われるようになった。そのなかでも、ユーザがどのような交通手段を用いて移動を行ったかを推定する移動手段推定は、多く研究されている分野の一つである。またスマートフォンを用いた位置情報の大規模な集積及び利活用を目的とした、大規模位置情報データプラットフォームが登場するようになった。このようなプラットフォームは容易な位置情報データの大規模分析を可能にする反面、位置情報データを収集する際に用いる SDK の特性やスマートフォンがおかれている環境によって、収集される位置情報に大幅な誤りを含む場合がある。また一般的な移動手段推定の研究で用いられるデータは、測位間隔が数秒程度で一定かつ密であることが多いが、大規模測位データではデバイスの消費電力的な観点から、取得される位置情報は測位間隔が数分程度と長いことが多く、測位間隔も一定ではない。移動手段推定に関する研究はこれまで数多く行われているが、このような性質を持つデータを対象とした移動手段推定の検討は十分に行われていない。そこで本研究は、時間粒度が粗密な大規模測位データを対象とした移動手段推定手法の提案を行う。提案手法は、データセットの前処理工程と移動手段推定工程の、大きく2つの工程に分かれる。前処理工程では、測位点間の重心に着目した独自のアルゴリズムによる誤差が顕著なデータの除去と、推定に用いる特徴量の付与を行う。移動手段推定工程では、速度の閾値判定による滞在/非滞在判定や移動手段の変化点検出などの処理を行った後、Random Forest による移動手段推定を行う。実世界で運用実績のある SDK により取得された位置情報データに対して提案手法を適用し、ユーザの移動手段を滞在・徒歩・電車・車の4種類に分類するタスクを行った結果、79.97%の精度での分類を達成した。

Transportation Mode Estimation for Spatio-Temporally Biased Large Scale Positioning Data

KEISUKE HIGASHIURA¹ NAOKI TAMURA¹ RUI HIRANO¹ KENTO WASHIDA¹
NAMIKI SAKAKURA¹ KOHEI YAMAGUCHI¹ KENTA URANO¹ TAKURO YONEZAWA¹
NOBUO KAWAGUCHI^{1,2}

1. はじめに

スマートフォンの普及が進み、スマートフォン内蔵のGPS受信機を用いたユーザの位置情報データ収集が容易になった。これに伴い、位置情報データの活用に関する研究や、位置情報を利用したサービスの開発が盛んに行われるようになった。位置情報技術は現在非常に注目されてい

る分野の一つであり、位置情報サービスの国内市場規模は数十兆円に及んでいる [1]。位置情報データに関する研究例としては、混雑情報分析 [2][3]、リソース分配最適化 [4]、行動分析 [5][6][7]、土地利用 [8][9] などがあり、多岐にわたる。その中でも、ユーザがどのような交通手段を用いて移動を行ったかを推定する移動手段推定 [10][11][12][13][14] は、これまでに数多くの研究が行われている分野の一つである。移動手段の推定によって、ユーザに対して乗換情報や渋滞情報などの移動手段に合わせた情報の提供が可能となる。また、ユーザが普段どのような移動手段を用いてい

¹ 名古屋大学大学院 工学研究科
Graduate School of Engineering, Nagoya University

² 名古屋大学 未来社会創造機構
Institutes of Innovation for Future Society, Nagoya University

るかという情報は、ユーザの社会的な属性（社会人・学生 etc.）を理解する助けとなり、この属性情報を用いてパーソナライズされた広告を発信できる。

さらに近年では、スマートフォンを用いた位置情報の大規模な集積及び利活用を目的とした、大規模位置情報データプラットフォームが登場している。プラットフォームの例としては、位置情報にユーザの性別や年代などの属性情報を紐付けたデータを収集・提供している KDDI Location Data[15] や、位置情報と性別・年代をはじめとする属性情報やペルソナデータを活用した人流マーケティングツールである agoop 社のマチレポ [16] などが挙げられる。このようなプラットフォームの登場によって、位置情報データの容易な大規模分析が可能となった。

しかし、このようなプラットフォームにおける位置情報活用において、先行研究で検討されてきた移動手段推定を適用するにはいくつかの課題が存在する。一つ目は、位置情報データの収集に用いる SDK の特性やスマートフォンがおかれている環境などの様々な要因によって、収集される位置情報に大幅な誤りを含んでいる点である。提案手法の精度検証で用いるデータで確認された位置情報の誤り例を図 1 に示す。被験者が辿った本来の移動ルートから数十キロ程度離れた場所で、誤った測位が発生している。この測位は、位置情報の補正に用いられる Wi-Fi アクセスポイントに紐付けられた位置情報の誤りが原因で発生したと考えられる。これまでの移動手段推定に関する研究では、このような大幅な位置情報誤差に対する考慮が十分に行われていない。二つ目は、デバイスの消費電力を抑える観点から、一般的な SDK で取得される位置情報はサンプリング間隔が数分程度の時間粒度が疎なデータであるケースが多い点である。提案手法の精度検証で用いるデータのサンプリング間隔の分布を図 2 に示す。サンプリング間隔が 300 ～400 秒の間に多く分布していることが分かる。またサンプリング間隔が一定ではなく、時間粒度が粗密なデータであることが分かる。これに対して多くの先行研究では、サンプリング間隔が数秒程度で一定かつ時間粒度が密な位置情報を対象としている。本研究で対象とする大規模測位データはサンプリング間隔が数百秒程度と広く、先行研究が対象としているデータと比べてユーザの位置の情報量が十数分の一以下である。このような性質を持つデータに対して適用可能な移動手段推定手法は、十分に検討されていない。

そこで本研究では、大幅な位置情報誤差を含み、時間粒度が粗密な大規模測位データに対して適用可能な移動手段推定手法の提案を行う。国土交通省が全国 70 都市の居住者を対象に実施した全国都市交通特性調査 [17] によれば、人が選択する移動手段は徒歩・電車・車の 3 種類で 80%以上を占めている。移動に用いられる交通手段の構成比を表 1 に示す。このような背景から本研究では、ユーザの移動

表 1 移動に用いられる交通手段別構成比（平成 27 年度全国都市交通特性調査集計結果より）

交通手段	割合 [%]
電車	16.5
バス	2.7
自動車（運転・同乗）	45.1
自動二輪車	2.4
自転車	13.8
徒歩	19.5

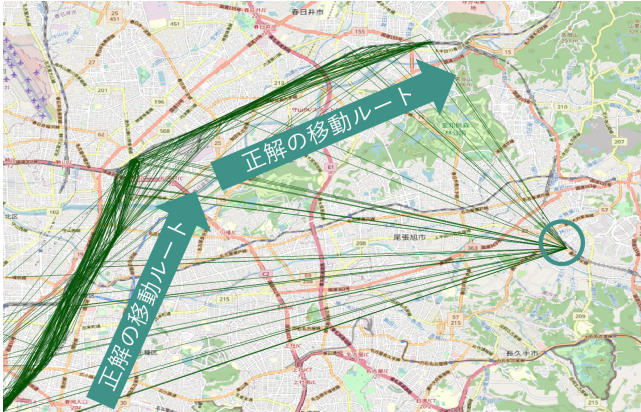


図 1 大幅な測定誤差を生じている位置情報の例

手段を徒歩・電車・車に滞在を加えた 4 種類に分類するタスクを考える。提案手法はデータセットの前処理工程と移動手段推定工程の、大きく 2 つの工程に分かれる。さらに、実際に運用実績のある SDK を用いて取得された位置情報データに対して本手法を適用し、精度の検証及び提案手法の有効性について議論を行う。本研究の貢献は、以下の通りである。

- 大幅な位置情報誤差を含み時間粒度が粗密な大規模測位データに対して、移動手段推定を高精度に実行可能な手法を提案した
- 誤差が顕著なデータの除去を実行する独自のアルゴリズムを提案した
- 実際に運用実績のある SDK によって取得されたデータを用いて、提案手法の有効性を検証した

2. 関連研究

移動手段の推定に関する研究は、これまでも多く行われてきた。本章では、本研究に関連する研究について説明し、本研究が対象とする問題への取り組みや解決方法が不十分なことを示す。

Zhu ら [10] は、GPS で記録される位置情報を用いて、ユーザの移動手段を徒歩・電車・車・バイクの 4 つに分類するタスクを行った。速度と加速度についての閾値処理によってユーザの移動手段が変化した地点を推定し、その地点間の移動手段を Random Forest によって分類するという

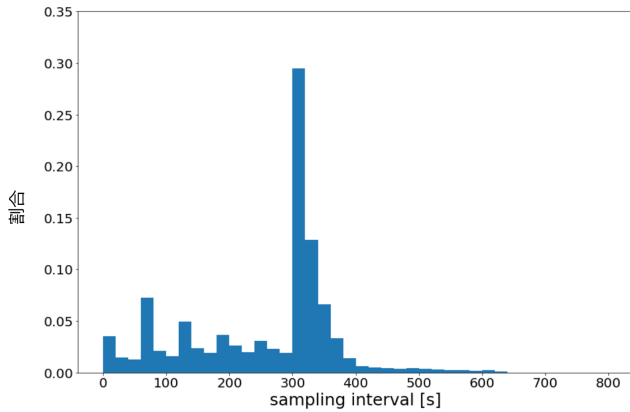


図 2 提案手法の検証で用いた大規模測位データのサンプリング間隔の分布

アプローチによって、Geolife GPS trajectory dataset[18]に含まれる移動履歴データを 82.85%の精度で分類した。大橋ら [11] は、スマートフォンに搭載されている GPS 受信機と加速度センサから得られる位置情報と加速度から、ユーザの移動を車とバイクの 2 つに分類するタスクを行った。車とバイクで走行する際に観測される振動の差異に着目し、加速度センサで計測される振動を特徴量として用いて、従来識別が困難であった車とバイクの移動を 80%以上の精度で分類した。しかしこれらの研究は、位置情報の測位間隔が数秒程度と短く、また測位間隔が一定であるデータを対象としており、本研究で扱う大規模測位データとは条件が大きく異なる。

本研究で対象とする大規模測位データのように、測位間隔が粗密で測定誤差を含むデータを対象とした研究も存在する。金村ら [14] は、測位間隔が粗密で測定誤差を含む位置情報と、駅や路線との距離などの地図情報を用いて、ユーザの移動手段を停止・歩く/走る・電車・車・バス・バイク・その他に分類した。金村らは、時間粒度が粗密なデータに対して有効な移動手段推定手法を開発するため、測位間隔 3 秒で収集したデータに対してリサンプリングを施し、時間粒度が粗密なデータセットを新たに作成した。リサンプリングは、測位間隔が平均 300 秒・分散 200 秒のガウス分布に従うように行った。また測定誤差として、各測位点に平均 20m・分散 30m のガウス分布に従う距離誤差を加えている。推定器には NLP 分野において LSTM と同精度を実現した GatedCNN を用いており、77.6%の精度での分類に成功した。しかし本研究で扱う大規模測位データには測定誤差が数十キロに及ぶ測位点が存在するなど、先行研究の考慮の範囲を超えている。

本研究では、このような先行研究における課題を克服し、大規模測位データに対して適用可能な移動手段推定手法を提案する。

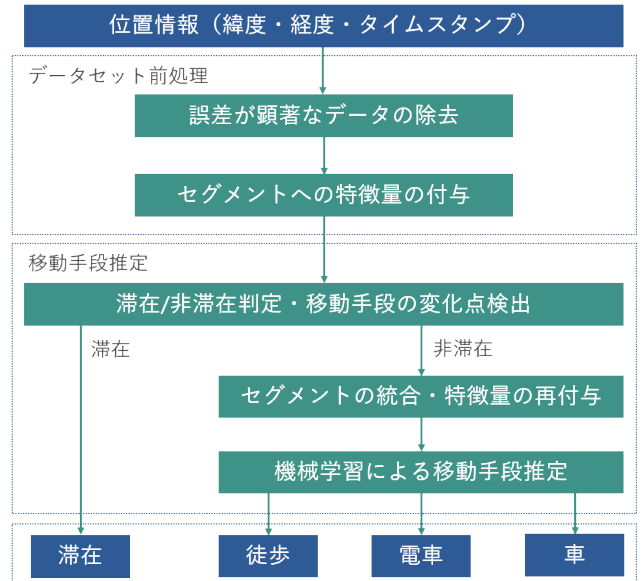


図 3 提案手法の概略

3. 提案手法

本章では、提案手法の手順と具体的な処理内容について説明する。提案手法は図 3 に示されるように、データセットの前処理工程と移動手段推定工程の、大きく 2 つの工程に分かれる。

データセットの前処理では、誤差が顕著なデータの除去と隣接する測位点間を結ぶ直線（以下セグメント）への特徴量付与を行う。誤差が顕著なデータの除去は、測位点間の重心に着目した独自のアルゴリズムで行う。その後、各セグメントに対して特徴量を付与する。本研究では付与する特徴量として、測位点に含まれる緯度経度の情報から算出できるものと、駅からの距離などの地理的な特徴量の、大きく分けて 2 つを採用する。

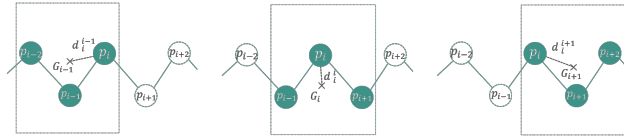
移動手段推定工程では最初に、速度の閾値判定による各セグメントの滞在/非滞在判定と移動手段の変化点検出を行う。その後、連続する非滞在セグメントを単一セグメントへ統合し、統合後のセグメントに対して特徴量の再付与を行う。最後に Random Forest を用いて、統合したセグメントに対して移動手段ラベルの付与を行う。

3.1 データセットの前処理

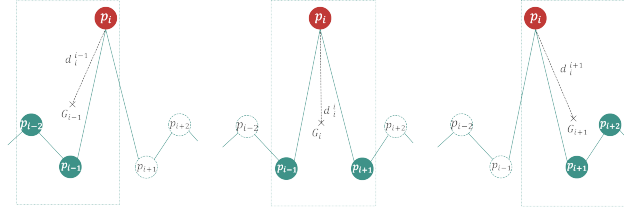
移動手段の推定にあたって、まずデータセットの前処理を行う。データセットの前処理は、誤差が顕著なデータの除去とセグメントへの特徴量付与の、2 ステップで構成される。以下で各ステップの詳細な説明を行う。

3.1.1 誤差が顕著なデータの除去

本項では、図 1 に示されるような、誤差が顕著なデータの除去処理について説明する。除去対象となるデータの検出は測位点間の重心に着目した独自のアルゴリズムによ



(a) 正常な測位データである場合



(b) 誤差が顕著な測位データである場合

図 4 測位点除去アルゴリズム ($n=1$ のとき)

で行う。アルゴリズムは以下の3ステップで成る。

- step1: 測位点 $p_i = (lat, lon)$ について、前後 n 点の範囲にある測位点群の重心 $\vec{G}_i = (lat, lon)$ を式 (1) によって算出する
- step2: \vec{p}_i と \vec{G}_k との距離 d_i^k [m] を求める
- step3: 条件式 (2) を満たす場合、 p_i を除去対象とする

$$\vec{G}_i = \frac{\sum_{k=i-n}^{i+n} \vec{p}_k}{2n+1} \quad (1)$$

$$\frac{\sum_{k=i-n}^{i+n} d_i^k}{2n+1} > d_{th} \quad (2)$$

このアルゴリズムは、短い時間の間に測位された点同士は位置的にも近くなり、測位点間の重心からの距離も近くなるという事実に基づいたものである。式中の n , d_{th} は実験的に決定される値であり、本研究では $n=1$, $d_{th}=4000$ [m] とした。 $n=1$ とした時の各変数の関係を図4に示す。図4(a)のように対象点が正常な測位データである場合、重心と対象点との距離は短く閾値を超えないため、除去対象とならない。それに対して図4(b)のように対象点が誤差が顕著な測位データである場合、重心と対象点との距離が長く閾値を超えるため、除去対象となる。

3.1.2 セグメントへの特徴量付与

本項では、セグメントに対しての特徴量付与について説明する。本研究で用いる特徴量は、位置情報に含まれる緯度経度から算出されるものと、POI情報から得られる地理的なものの、大きく分けて2つに分かれる。本研究では関連研究[14]を参考に特徴量を選定し、表2の通りとした。速度はセグメント長を始点から終点への移動時間で割った値、加速度は対象セグメントの一つ前にあるセグメントにおける速度からの変化量、方向転換角は対象セグメントが

表 2 提案手法で採用した特徴量一覧

位置情報から算出する特徴量	地理的特徴量
速度 [m/s]	最寄駅からの距離 [m] 最寄路線からの距離 [m]
対数速度 [$\log_{10}(\text{m/s})$]	
加速度 [m/s^2]	
セグメント長 [m]	
方向転換角 [$^\circ$]	

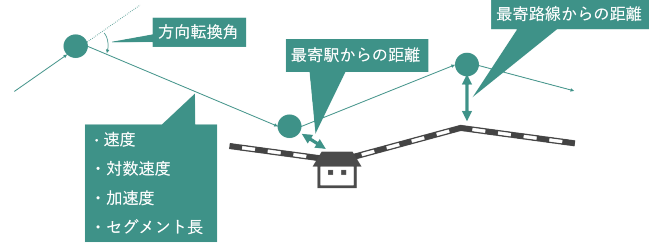


図 5 提案手法で採用した特徴量一覧

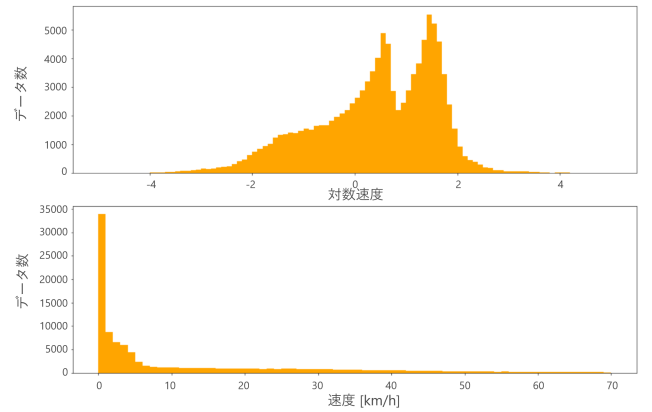


図 6 セグメントごとの速度と対数速度の分布

向いている方向 ($0 \sim 360^\circ$) と対象セグメントの一つ前にあるセグメントが向いている方向の差とした。また最寄駅からの距離及び最寄路線からの距離は、各セグメントの始点と終点における値の平均値とした。対数速度は、人の歩行の速度分布域とその他の乗り物 (電車・車) の速度分布域を分離し、分類を行いやすくするという[14]の知見に基づいて設定した。対数速度のグラフを見ると、人の歩行速度域である $1.0 \sim 3.1$ [m/s] ($10^0 \sim 10^{0.5}$) とその他の乗り物の速度域である 10 [m/s] (10^1) で分布の山を形成しており、それぞれを分離できていることが分かる。

3.2 移動手段推定

移動手段の推定は、滞在/非滞在推定・移動手段の変化点検出、セグメントの統合・特徴量の再付与、Random Forest による移動手段の分類の3ステップで構成される。以下で各ステップの詳細な説明を行う。

3.2.1 滞在/非滞在推定・移動手段の変化点検出

本項では各セグメントにおけるユーザの滞在/非滞滞在の判定方法、及び移動手段が変化した点の検出方法について

説明する.

滞在/非滞在の判定は、各セグメントの速度の閾値処理によって行う. 本手法では、セグメントの速度 v がある閾値 v_{th} より低ければ滞在, そうでなければ非滞在と判定する. すなわち,

$$\begin{cases} \text{滞在} & (\text{if } v < v_{th}) \\ \text{非滞在} & (\text{if } v \geq v_{th}) \end{cases} \quad (3)$$

とする. 滞在与判定されたセグメントには、移動手段の予測ラベルとして「滞在」を付与する. 本研究では、都市空間における人の一般的な歩行速度 [19] である 1.7 [m/s] より十分小さい, $v_{th} = 0.35$ [m/s] とした.

次に、移動手段の変化点の検出方法について説明する. 交通手段が切り替わる地点では、乗換動作が行われる関係上、ユーザの速度はゼロまたはゼロに近い値を取ると考えられる. これについては関連研究 [20] でも言及されている. したがって本研究では、図 7(a) に示されるように、滞在与非滞在が切り替わる地点を移動手段の変化点として扱う.

3.2.2 セグメントの統合・特徴量の再付与

本項では、非滞在と判定されたセグメントの移動手段を徒歩・電車・車の 3 種類に分類するための準備について説明する.

まず、図 7(b) に示されるように、非滞在が連続するセグメントを、単一のセグメントに統合する. 次に、統合前のセグメント群に付与されていた各特徴量の平均値を算出し、それらの値を統合後のセグメントに対して付与する. すなわち、セグメント群 $s_{st}, s_{st+1}, \dots, s_{end-1}, s_{end}$ を統合する時、セグメント s_i に付与されている n 個の特徴量を $(f_{i1}, f_{i2}, \dots, f_{in})$ とすると、統合後のセグメントに付与される特徴量 $F_{st,end,j} (1 \leq j \leq n)$ は,

$$F_{st,end,j} = \frac{\sum_{i=st}^{end} f_{ij}}{end - st + 1} \quad (4)$$

となる.

3.2.3 Random Forest による移動手段の分類

本項では、機械学習による移動手段の分類について説明する. 本手法では 3.2.2 項で再構成されたセグメントの特徴量を入力とし、予測ラベル「徒歩・電車・車」のいずれかを出力するモデルを構築する. 各特徴量の推定への寄与率を分析するため、本研究で用いる機械学習手法は [14] で用いられた SVM, Random Forest, GatedCNN のうち、Random Forest とした.

4. 評価

本章では、3 章で述べた提案手法の精度検証を行う. まず、提案手法による推定精度の検証に用いるデータセットの概要、推定に用いる機械学習モデル構築の詳細を説明す

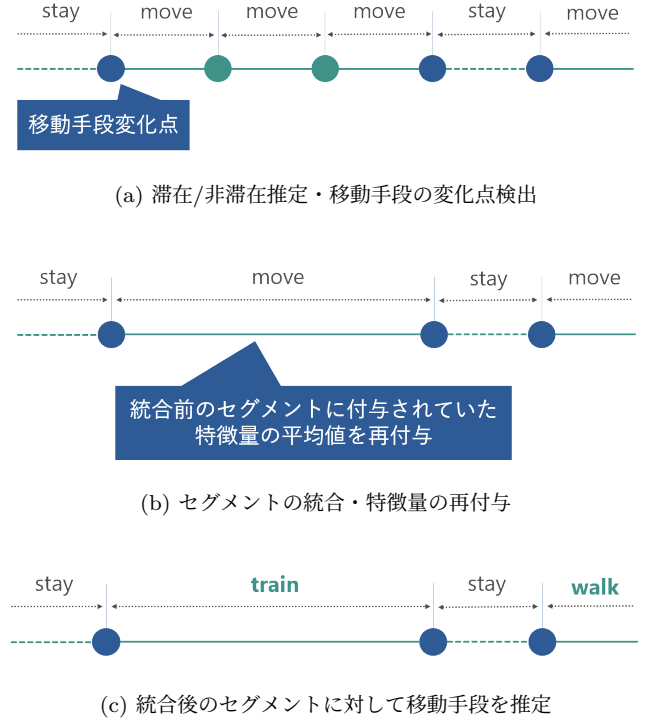


図 7 提案手法における移動手段推定の流れ

る. その後推定結果の提示と結果の分析を行う.

4.1 データセット及び機械学習モデル

提案手法の検証に用いるデータセットの概要を述べる. 今回使用したデータは、スマートフォン 350 台で収集された、のべ 1199 日分の移動履歴データである. これらのデータは、札幌・仙台・東京・愛知・大阪・広島・福岡の 7 都市で収集された. データの収集を行う被験者は、位置情報記録アプリケーションを立ち上げたスマートフォンを所持し、移動するルート・日時・交通手段が詳細に記された工程表に従って移動を行った. 計測に用いた移動経路は合計で 48 パターン用意した. 経路の設定は、線路沿いを複数の移動手段を用いて移動する経路を多く設けるなど、緯度経度の情報のみではどのような手段を使った移動かを判別しにくいように行った. 記録されるデータには以下のデータが含まれる.

- デバイス ID
- 測位時間を示すタイムスタンプ
- 緯度・経度

記録されたデータの例を表 3 に示す. 収集したデータに対する正解ラベルの付与は、タイムスタンプと工程表に記された時刻を照らし合わせて行った. 移動手段ごとの正解ラベル数を表 4 に示す.

機械学習モデルの構築では、移動経路 48 パターンのうち 31 パターンを学習データに、残りの 17 パターンをテストデータとして、学習及び精度検証を行った. 学習データ

表 3 検証に用いたデータの一例

ID	timestamp	latitude	longitude
001	2021-03-20 10:01:30	35.154626	136.966321
001	2021-03-20 10:06:41	35.163732	136.963379
001	2021-03-20 10:12:05	35.166445	136.953323
.	.	.	.
.	.	.	.
.	.	.	.

表 4 移動手段ごとの正解ラベル数一覧

移動手段	ラベル数 (統合前)	ラベル数 (統合後)
stay	2798	-
walk	10375	1498
train	34628	7482
car	7395	1375

とテストデータの分割は、7都市のデータがどちらにも含まれるように行った。また表4に示されるように、移動手段間にデータ量のばらつきが見られたため、学習を行う際にダウンサンプリングを施した。

4.2 分類結果

提案手法を用いて検証データを分類した結果を図8に示す。推定精度は式(5)で計算され、全体の推定精度は79.97%であった。また、各ラベルについての適合率及び再現率を表5に示す。推定精度に関する考察は4.3.2項で述べる。

$$\text{精度} = \frac{\text{予測と正解が一致したセグメント数}}{\text{総セグメント数}} \times 100 \quad (5)$$

4.3 考察

4.3.1 測位点除去アルゴリズム

3.1.1項で述べた測位点除去アルゴリズムをデータセットに適用した結果、データセット内の測位点112,929点中の約2.6%に相当する2,930点が除去された。除去前後のデータセットの変化例として、図1のデータにアルゴリズムを適用した結果を図9に示す。図1と比較して誤差が顕著なデータが減少しており、提案手法のアルゴリズムの有効性が示された。一方で除去に失敗したデータもいくつか確認でき、完全な除去には至らなかった。除去に失敗した理由としては、

- 閾値のチューニングが十分でなかった
- ユーザが速度の速い移動手段を用いており、かつ測位間隔が通常より疎だった場合、測位点が地図上の広範囲に広がってしまい、重心が正常な測位点の系列の方向に十分偏らなかった
- 異常な測位の連続発生により、異常な測位点の近くに

正解ラベル	stay	93.41 % (170)	4.95 % (9)	1.10% (2)	0.55% (1)
	walk	12.71% (582)	74.60% (3416)	1.83% (84)	10.85% (497)
	train	8.00% (771)	0.44% (42)	82.65% (7967)	8.92% (860)
	car	7.06% (196)	4.03% (112)	8.86% (246)	80.05% (2223)
		stay	walk	train	car
		予測ラベル			

図 8 テストデータの分類結果 (括弧内：データの絶対数)

表 5 ラベルごとの適合率・再現率

ラベル	適合率 [%]	再現率 [%]
stay	9.89	93.40
walk	95.45	74.60
train	95.95	82.65
car	62.08	80.05

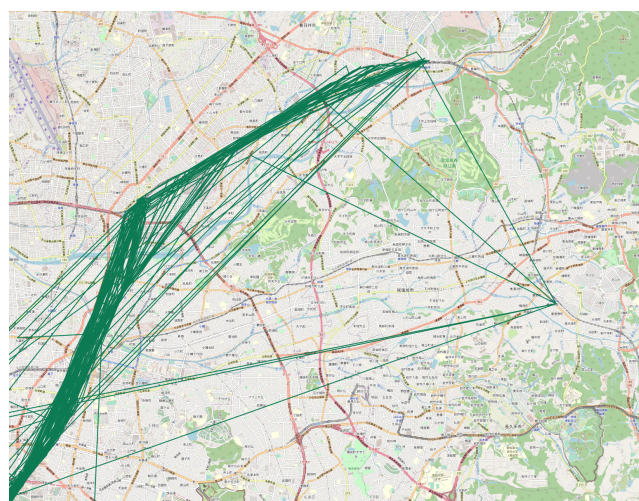


図 9 測位点除去アルゴリズムの適用結果例

重心が偏ってしまい、重心と異常な測位点との距離が近くなってしまった

などが考えられる。測位誤差の大きさは数十キロ以上に及ぶ大きなものから数百メートル～数キロの小さいものまで存在するため、重心からの距離の閾値処理のみで完全な測位点除去を行うのは難しい。改善策としては、測位点除去ルール追加が考えられる。例えば図1を見ると、誤差が顕著なデータの前後で移動軌跡の回転角が大きくなっており、この特徴量を測位点除去ルールとして取り入れることで精度が向上する可能性がある。

4.3.2 推定精度

全体の推定精度は79.97%であり、使用したデータの条

表 6 各特徴量の寄与率一覧

特徴量	寄与率 [%]
速度	29.9
対数速度	17.8
セグメント長	34.5
方向回転角	2.9
最寄駅からの距離	2.8
最寄路線からの距離	6.3

件が比較的類似している関連研究 [14] の 77.6% より 2% 高い推定精度を達成した。提案手法による移動手段推定が有効にはたらいっていることが示された。

その一方で表 5 を見ると、正解ラベルが stay であるデータの適合率が 10% を下回っており、移動手段推定がうまく行われていないケースも確認できた。すなわち、実際には正解ラベルが移動（徒歩・電車・車）であるが、予測ラベルを滞在としているケースが非常に多く見られる。これは、正解ラベルとして徒歩・電車・車が割り振られている場合でも、実際の被験者の振るまいに以下のようなパターンが含まれることが原因の一つとして考えられる。

- 歩行時に信号待ちで停止
- 駅での停車
- 信号待ちや渋滞での停車

工程表と照らし合わせて正解ラベルを付与する方法では、先述のようなユーザの多様な行動に対して適切なラベル付けを行うのは難しく、正解ラベルの付与方法の再検討が必要である。

また移動履歴データは時系列データの性質を持っているが、今回の推定手法では時系列的要素を十分に考慮していない。時系列を考慮できる機械学習手法や深層学習手法を取り入れることで、推定精度が向上する可能性がある。

4.3.3 各特徴量の寄与度

本研究では推定器として Random Forest を採用したため、推定における特徴量寄与率を算出できる。今回の推定における各特徴量の寄与率を表 6 に示す。速度とセグメント長が推定に大きく寄与している。地理的特徴量は 2 つ合わせて 10% 程度の寄与があったが、速度やセグメント長と比較すると影響は小さかった。これは 4.1 節でも述べたとおり、今回のデータセットは線路沿線を複数の移動手段で移動する工程を多く含んでおり、移動に違いによる地理的特徴量の変化が少なくなったためと考えられる。

5. 結論

本研究では、時間粒度が粗密な大規模測位データに適用可能な移動手段推定手法の提案を行った。

データセットの前処理では、測位点間の重心に着目した独自の測位点除去アルゴリズムによって、誤差が顕著なデータを除去した。このアルゴリズムを検証データに適用

した結果、誤差が顕著なデータの多くが除去され、提案手法の測位点除去アルゴリズムの有効性が示された。

移動手段の推定では速度の閾値処理と Random Forest によって、ユーザの移動手段を滞在・徒歩・電車・車の 4 種類に分類した。全体の推定精度は 79.97% であった。推定を誤ったケースとして、正解ラベルは徒歩・電車・車であるが、予測ラベルが滞在となったパターンが多く見られた。これはユーザの多様な動作によって、被験者が実際にとった動作と正解ラベルの間に発生した齟齬が原因の一つとして考えられ、「電車乗車中に駅で停車する」「車乗車中に信号待ちや渋滞で停車する」などの、今回付与した正解ラベルからは抽出できないユーザの多様な動作に対して、適切なラベリングを行う手法を検討する必要がある。また本研究のテーマである大規模測位データは時系列データであるが、今回の推定モデルでは時系列的な要素を十分考慮できていない。時系列を考慮した推定モデルへの改良によって、精度が向上する可能性がある。

謝辞

本研究の一部は、JST CREST(JPMJCR21F2)、NICT 委託研究の支援を受けています。また、データ提供にご協力いただきましたプログウォッチャー社に感謝いたします。

参考文献

- [1] 遠藤結城, 戸田浩之, 小池義昌. 移動手段推定のための表現学習を用いた GPS ログからの特徴抽出. 情報処理学会論文誌データベース (TOD), Vol. 8, No. 3, pp. 12–23, 2015.
- [2] Zihan Kan, Luliang Tang, Mei-Po Kwan, Chang Ren, Dong Liu, and Qingquan Li. Traffic congestion analysis at the turn level using taxis' GPS trajectory data. *Computers, Environment and Urban Systems*, Vol. 74, pp. 229–243, 2019.
- [3] Eleonora D'Andrea and Francesco Marcelloni. Detection of traffic congestion and incidents from GPS trace analysis. *Expert Systems with Applications*, Vol. 73, pp. 43–56, 2017.
- [4] Haoran Zhang, Xuan Song, Yin Long, Tianqi Xia, Kai Fang, Jianqin Zheng, Dou Huang, Ryosuke Shibasaki, and Yongtu Liang. Mobile phone GPS data in urban bicycle-sharing: Layout optimization and emissions reduction analysis. *Applied Energy*, Vol. 242, pp. 138–147, 2019.
- [5] Chao Chen, Chengwu Liao, Xuefeng Xie, Yasha Wang, and Junfeng Zhao. Trip2Vec: a deep embedding approach for clustering and profiling taxi trip purposes. *Personal and Ubiquitous Computing*, Vol. 23, pp. 1–14, 2019.
- [6] Xiaowei Hu, Shi An, and Jian Wang. Taxi driver's operation behavior and passengers' demand analysis based on GPS data. *Journal of Advanced Transportation*, Vol. 2018, pp. 1–11, 2018.
- [7] Abigail M. Kidd, Ashley D'Antonio, Christopher A. Monz, Kevin Heaslip, Derrick Taff, Peter Newman. A GPS-based classification of visitors' vehicular behavior in a protected area setting. *Journal of park and*

- recreation administration, Vol. 36, pp. 69–89, 2018.
- [8] 庄子和之, 青木俊介, 米澤拓郎, 河口信夫. 大規模ユーザの滞在情報に基づくエリアの特徴付けと COVID-19 による影響分析. 情報処理学会論文誌, Vol. 62, pp. 1644–1657, 2021.
 - [9] Gang Pan, Guande Qi, Zhaohui Wu, Daqing Zhang, and Shijian Li. Land-use classification using taxi GPS traces. IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 1, pp. 113–123, 2013.
 - [10] Qiuhui Zhu, Min Zhu, Mingzhao Li, Min Fu, Zhibiao Huang, Qihong Gan, and Zhenghao Zhou. Identifying transportation modes from raw GPS data. Communications in Computer and Information Science, Vol. 623, pp. 395–409, 2016.
 - [11] 大橋洋輝, 秋山高行, 山本正明, 佐藤暁子. 車両走行時の振動発生モデルに基づく移動手段識別方式. 情報処理学会論文誌, Vol. 56, pp. 23–34, 2015.
 - [12] Leon Stenneth, Ouri Wolfson, Philip Yu, and Bo Xu. Transportation mode detection using mobile phones and GIS information. GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems, pp. 54–63, 11 2011.
 - [13] Adel Bolbol, Tao Cheng, Ioannis Tsapakis, and James Haworth. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. Computers, Environment and Urban Systems, Vol. 36, No. 6, pp. 526–537, 2012.
 - [14] 金村卓郎, 西村拓哉, 倉島健, 戸田浩之. 疎な移動軌跡を対象とした移動手段推定. データ工学と情報マネジメントに関するフォーラム, 2018.
 - [15] KDDI Location Data. <https://k-locationdata.kddi.com/>.
 - [16] Agoop マチレポ. <https://www.agoop.co.jp/service/machirepo/>.
 - [17] 都市における人の動きとその変化～平成 27 年度全国都市交通特性調査集計結果より～. <https://www.mlit.go.jp/common/001223976.pdf>.
 - [18] Yu Zheng, Hao Fu, Xing Xie, Wei-Ying Ma, and Qian-nan Li. Geolife GPS trajectory dataset - User Guide, geolife gps trajectories 1.1 edition, 2011.
 - [19] 渡辺美穂, 羽藤英二. 移動軌跡に着目した都市空間の歩行速度分析. 都市計画論文集, Vol. 42.3, pp. 535–540, 2007.
 - [20] Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. Learning transportation mode from raw GPS data for geographic applications on the web. Proceeding of the 17th International Conference on World Wide Web 2008, WWW’08, pp. 247–256, 2008.