

疎な移動軌跡を対象とした移動手段推定

金村 卓郎[†] 西村 拓哉[‡] 倉島 健[‡] 戸田 浩之^{‡‡}

[†] 東北大学電気通信研究所 〒980-8577 宮城県仙台市青葉区片平 2-1-1

^{‡,‡‡} 日本電信電話株式会社 NTT サービスエボリューション研究所 〒239-9847 神奈川県横須賀市光の丘 1-1

E-mail: [†] kanemura@tohoku.riec.ac.jp,

[‡] {takuya.nishimura.fk, takeshi.kurashima.uf}@hco.ntt.co.jp, ^{‡‡} hirotoda@acm.org

あらまし 歩行・バス・電車などの移動手段はユーザーの行動内容を理解する上で重要な情報である。従来から、スマートフォンなどから取得される移動軌跡を用いて移動手段を推定する研究が行われてきた。従来手法では、移動手段推定を行う際には測位間隔が数秒から数十秒程度といった測位頻度が密な移動軌跡を入力としていた。しかし、このような密な移動軌跡は、位置情報ゲーム等の一部のアプリケーションを除いては、デバイスのバッテリー消費の観点から実際に収集するのは難しい。そこで本研究では、位置情報収集による電力消費を考慮したスマートフォンアプリにおいて記録される、測位間隔が数分程度の移動軌跡を対象とした移動手段推定に取り組む。実データの分析に基づく特徴量の前処理と地図情報の考慮による特徴量の追加と、これらを用いる推定モデルを提案する。関東在住の30人から収集された3週間分の移動軌跡を用いて3つの設定のもとで実験を行い、7種類の移動手段の分類問題に対して、いずれの設定でも提案手法は従来手法よりも10%以上高い精度を示した。

キーワード 移動手段推定, 移動軌跡, GIS, Gated Convolutional Network, Random Forest, SVM

1. はじめに

近年、スマートフォンや様々なセンサ[1][2], SNSなどの普及により、GPSやWi-Fiのアクセスポイント[3][4]などを利用して、ユーザーの行動履歴を収集する様々な手法が普及してきている。このように得られたユーザーの行動に関する情報を解析することは行動パターンや状況の推定を可能にするため多くの研究が行われている[5][6][7]。特にユーザーの歩行・列車・車などの移動手段を図1に示すような移動軌跡から推定する移動手段推定のタスクは、ユーザーのデバイスやアプリケーションに対しカスタマイズされた情報や広告を発信することを可能とする[8]。

移動手段推定のタスクは従来から研究が行われている。従来手法[9]では数秒間隔の測定点からなる移動軌跡を数分程度の長さの移動軌跡に分割し、分割された移動軌跡それぞれに対し移動手段推定の推定を行っている。このような従来手法においては、移動軌跡内に数秒間隔の大量の測位点が存在することを仮定しているが、GPSを使用する際には一般的なバッテリーの消費量を考えれば測位間隔は5分程度であることが多い[10]。さらに、実際の測位点は測位間隔にズレが生じる場合や、GPSの飛びなどの地理的な誤差が生じる場合も多いため[11]、従来手法はうまく働かない場合もある。また、他の手法と比較して疎な移動軌跡を対象とした手法も存在するが[13]、これも測位間隔を1分と仮定しており、これよりも測位間隔が広い場合の有効性は不明である。

図1に本研究で扱う移動軌跡の一例を示す。図1.1では道に沿って徒歩で移動している部分と電車の部分

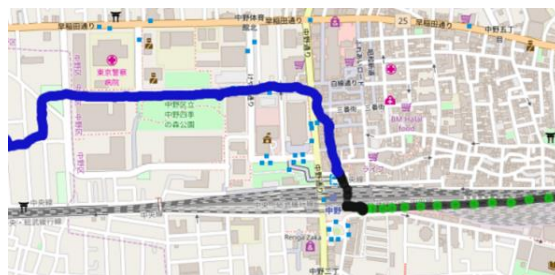


図 1.1 測位間隔が 3 秒の移動軌跡



図 1.2 測位間隔が 5 分の移動軌跡

図1. 本研究で扱う移動軌跡の一例を示す。青い線は徒歩で移動していることを示し、緑は電車に乗っていることを示す。図1.1は測位間隔が3秒に1回の移動軌跡を、図1.2は図1.1の移動軌跡を5分おきにサンプリングし直した移動軌跡であり、黒い点が測位点を示す。リサンプリングした移動軌跡では、ユーザーの細かい移動軌跡が失われていることが確認できる。本研究では図1.2のような移動軌跡を入力として高精度に移動手段を推定することに取り組む。

© OpenStreetMap Contributors

では移動速度や方向転換の角度などに大きな違いが確認できるが、図1.2では測位間隔が疎になった結果その差異が確認できなくなる。このように速度や方向転換率を特徴量として推定する従来手法は移動軌跡が密であればこそ可能であり、疎な移動軌跡は従来と同じ

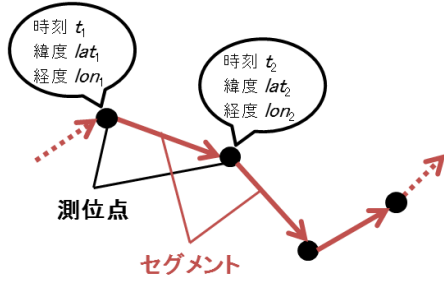


図2. 本研究における測位点とセグメントを示す. 測位点はGPSなどで取得された地点を示し[測位時刻, 緯度, 経度]からなる情報. 移動軌跡は測位点の集合を指す. セグメントは, 時間的に隣接した2つの測位点の間の部分を示す.

手法では移動手段の推定が難しい.

本研究では, 現実に収集可能である疎な移動軌跡からの移動手段推定手法の実現を目指し, 疎な移動軌跡の基礎分析に基づく特徴量の前処理と, 駅や線路といった地図情報の考慮を伴う推定モデルを提案する. 速度などの特徴量に対して前処理を適用することによって, 疎な移動軌跡からでも移動手段の分類に有益な情報を取得し分類器の入力とすることが可能になる. また, 地図情報の考慮によって, 図 1.2 に示すような移動軌跡からであっても駅や線路の位置を参考に電車と徒歩の移動手段を見分けることが可能になる.

本研究の貢献を以下に示す.

- 実用的な場面で取得されうる移動軌跡に対し, 既存手法を用いた移動手段推定を用いた際の問題点の分析を行った.
- 問題点の分析に基づく前処理と地図情報の考慮による特徴量設計と, 推定対象の前後も考慮することによって高精度な推定を可能にする推定モデルを提案した.
- 移動手段の7クラス分類タスクにおいて, 現実的に収集される移動軌跡を考慮した3つのデータセットを人工的に生成して実験を行い, いずれの設定においてもベースライン手法と比較して 10%以上の精度改善を達成した. また, 3つのデータセットでの精度を比較することで, 提案手法が時間的なノイズや地理的なノイズに影響を受けることも示した.

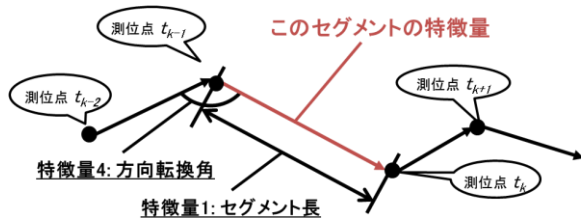
2. 先行研究

本章では疎な測位間隔のデータから移動手段の推定を行う際に参考にした先行研究について紹介する. 移動手段の推定に関する研究は1章で解説した通り, ユーザーにいろいろなサービスを提供することを可能にするため様々なアプローチの研究が行われている. Zheng ら[9][11]は, まず移動軌跡の測位点を連続した測位点間で速度および加速度の算出を行い, これらの数値が大きく変化したタイミングで移動手段が変わっ

表1. 先行研究における特徴量

カテゴリ	特徴量
測位点集合推定 Basic feature[9]	測位点集合の長さ 速度平均 速度期待値 速度分散 上位3つの速度 上位3つの加速度
測位点集合推定 Advanced feature[12]	速度変化率 停止率 方向変化率
セグメント 直接推定 Feature[14]	セグメント長 速度 加速度 方向転換率

ていると仮定し, 移動手段の変化点ごとに測位点集合を分割した後, 分割した測位点集合に対し移動手段の推定を行う手法を提案した. ここで Zheng らは移動手段の推定に用いる特徴量についても提案を行っており, 表1に示すような移動軌跡の速度平均・加速度平均からなる Basic feature をもとに推定を行った報告をしている[9]. その後, 速度変化率などの複雑な特徴量 advanced feature を設計し, Basic feature と Advanced feature を組み合わせることで精度が向上したことを報告した[12]. しかし, 本研究においては疎な測位間隔における移動手段推定を目的としているため, 測位点集合で分割する時間が測位点の測定間隔と同程度なため, 移動手段ごとに測位点集合を分割するのは困難である上に, 例えば平均速度や方向転換率および平均加速度などの特徴量は測位点集合内に多くの測位点を必要とするため本研究の要件を満たさない. また, Bolbal ら[13]は, ある隣接する測位点間の移動手段の推定に際して, 特徴量の抽出をその測位点とさらにその前後の測位点から行った. 図3に示したように特徴量にセグメント長・速度・加速度・方向転換率を用いて, 測位間隔を1分としたロンドン市内における移動手段の6種類(バス・車・自転車・列車・地下鉄・徒歩)分類タスクにおいて88%の精度での分類を達成した. この手法は疎な移動軌跡を対象としている点で本研究の問題設定に近いが, 想定している測位間隔が1分と実的に取得される移動軌跡に比べて短く, 測位間隔が5分程度の移動軌跡に対しても有効であるかどうかは不明である. 次に, 疎な測位間隔における移動手段推定に有用と考えられる, 地図情報を考慮した移動手段



特徴量2: 速度 $\Rightarrow$ 距離をセグメントの時間長で割ったもの
 特徴量3: 加速度 $\Rightarrow$ 直前のセグメントからの速度の変化量

図3. 本研究における各セグメントの特徴量として用いる4つの値を示す。これは[14]で提案されたものと同一である。

推定手法を紹介する。Stenneth ら[14]はシカゴ市内のバスのリアルタイム情報・バス停の位置情報・鉄道情報からセグメントに含まれる各測位点についてその測位点から最近傍のバスまでの距離を算出しセグメント全体での平均の距離を算出した値やセグメント内の各測位点の最近傍路までの距離の平均値などを特徴量として取り入れることで地図情報のない場合とある場合で分類器によっては20%近い精度の向上が確認できたことを示している。本研究でもこの手法を参考に、地図情報を考慮した移動手段推定手法を提案する。

3. 用語定義とデータセット

本章では本研究における移動手段推定を行う際の用語定義について行った後、本研究で用いたデータセットについて説明する。

3.1. 用語定義

本稿で用いる用語を以下のように定義する。図2にその内容を示す。

1. 測位点：GPSなどで取得された位置情報であり、[測位時刻，緯度，経度]の3つ組で示される情報
2. 移動軌跡：連続した測位点の集合
3. セグメント：時間的に隣接した2つの測位点の部分

3.2. データセット

本研究で用いた移動軌跡は、スマートフォンのGPS記録機能を用いて収集した首都圏周辺に在住する30人分の3週間分の移動軌跡である。収集した際の測位間隔は3秒に1回である。これをオリジナルの移動軌跡と呼ぶ。また、オリジナルの移動軌跡は各測位点に対して移動手段の正解ラベルが実際に移動していた本人の記録によって割り当てられている。移動手段ラベルは「stop」「walk/run」「train」「car」「bus」「bike」「other」の7種類である。オリジナルの移動軌跡の例は図1.1に示した通りである。ここからリサンプリングを行い、測位間隔が疎な移動軌跡を人工的に複数生成した。以下に生成したデータセットを示す。

1. 測位間隔が5分の移動軌跡：オリジナルの移動軌跡から、測位点を5分間隔でリサンプリングしたもの。図1.2にその一例を示す。移動手段ラベ

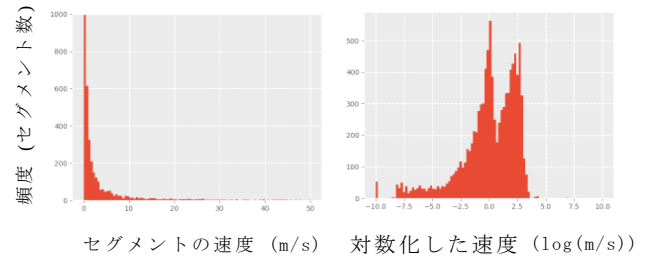


図4. セグメントにおける速度の分布を示す。横軸に速度，縦軸に頻度を示している。左は通常の分布であり，右は横軸に対数を取ったグラフとなっている。

ルはセグメントに対して割り当てられており，当該セグメントに対応するオリジナルの測位点に最も多く割り振られていた移動手段ラベルを割り当てている。

2. 測位間隔が不定の移動軌跡：オリジナルの移動軌跡から，測位点を平均300秒，分散200秒のガウス分布の間隔でランダムにリサンプリングしたもの。移動手段ラベルは測位間隔が5分の移動軌跡と同様に割り当てた。
3. 測位間隔が不定で測位誤差の大きい移動軌跡：測位間隔が不定の移動軌跡と同様にリサンプリングを行い，さらに測位位置に対して平均20m，分散30mのガウス分布に従う誤差を与えたもの。移動手段ラベルは測位間隔が5分の移動軌跡と同様に割り当てた。

3つのデータセットの生成手順はいずれも現実に収集可能な移動軌跡を想定して設定したものである。GPSログの記録はバッテリー消費量が大きいため，測位間隔は5分間隔程度に設定されることが多い[10]。また，デバイスの設定によっては一定間隔での測位ではなく不定な間隔での測位になることもあることも知られている。更に地理的な測位誤差が大きく生じる場合もある[11]。こういった様々な環境を想定した3つのデータセットを用いて分析を行う。

4. 提案手法

まず3.2で生成した測位間隔が5分の移動軌跡について基礎分析を行い，それに基づいた特徴量の従来手法からの改良について述べる。次に，本研究で用いる地図情報を用いた特徴量について述べる。最後に，これらで得られた特徴量を用いて移動手段の推定を行う分類器について述べる。

4.1. 疎な移動軌跡の基礎分析と特徴量設計

本研究では，推定対象となる各セグメントの特徴量として，Bolbal ら[13]の提案した，図3に示すようなセグメント長，方向転換角と速度，加速度の4つの特徴量を基本的には用いる。なお，速度を測位点間の経過時間で割った値，加速度を現在注目しているセグメ

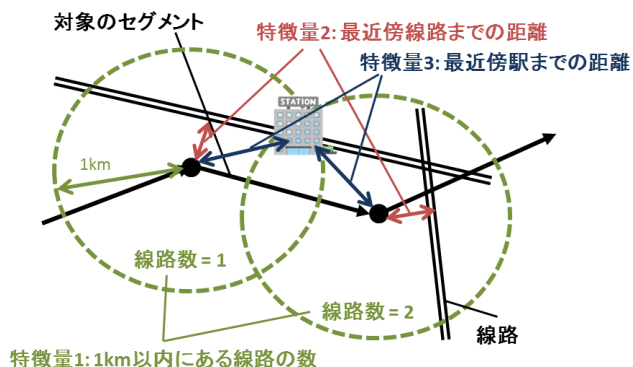


図5. 地図情報から取得する3つの特徴量を示す。

ントと前のセグメントの速度の差とする。

速度について、測位間隔が5分の移動軌跡について分析を行ったところ、特徴量の対数化を行うことで推定精度が向上されると期待されることが分かった。図4の左側のグラフに示す通り速度の分布においては分布が0近傍に立つことが確認できる。これらは停止している状態を示唆していると考えられる。一方で、乗り物に乗っているのか徒歩で移動しているのかの判別は通常は速度分布からは難しい。しかし、図4の右側に示されるように、対数速度分布においては、人間の歩行速度程度である約1m/sにピークをもつ分布と約12m/s(～40km/h)をピークに持つ分布を確認することができ、乗り物での移動と徒歩での移動がよく分離されることが確認できる。これに基づき、本研究では速度特徴量をそのまま用いるのではなく対数化の前処理を行ってから用いる。また、それと併せてそれぞれの特徴量の値が0から1に収まるように正規化を行う。これは特に後述するGated CNNのような深層学習を用いた分類器を用いる際に有効である[15]。

4.2. 地図情報を用いた特徴量

Stenneth ら[14]が示したように、地図情報は移動手段推定の精度向上に寄与することが知られている。本研究でも、地図情報を用いて各セグメントの特徴量を取得する。地図情報の情報源としてはボランティアが整備するフリーの地図データ OSM (OpenStreetMap) [16]を用いた。OSMは交差点・駅をノードで、道路・線路をエッジとするグラフで保存されている。

本研究では、地図情報を用いた特徴量を線路と駅の位置情報から算出する。下記に駅・線路情報をもとに算出する3つの特徴量を示す。

1. 半径1km以内の線路数
2. 最近傍線路までの距離
3. 最近傍駅までの距離

まず1.の半径1km以内の線路数は、セグメントを形成する測位点2点のそれぞれ半径1km以内の線路の数である。近辺に線路数が多ければ大規模な駅である可能性が高いということから、駅付近にいる際にその駅

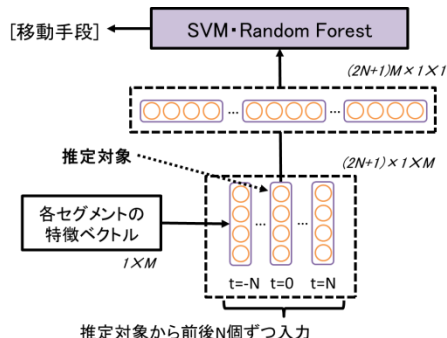


図6. SVM および Random Forest の入出力。M が各セグメントの特徴ベクトルの次元数、 $(2N+1)$ が window size を示す。

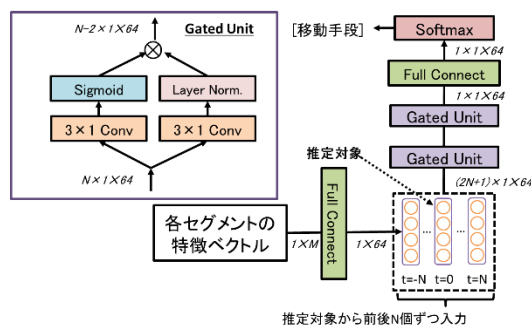


図7. Gated Convolutional Neural Network での推定の際の入出力。M が各セグメントの特徴ベクトルの次元数、 $(2N+1)$ が window size を示す。

の規模を示唆する。小さな駅と大規模な駅付近では人の行動の取り方[17]は異なるため駅の規模を判別するのは移動手段判定には有効と考えられる。

次に、2.の最近接線路までの距離は、セグメントを形成する測位点2点のそれぞれに対する線路の最近傍距離である。距離が近ければ電車に乗っている可能性が高い。したがって、線路情報を特徴量として電車という移動手段を判別する精度が上がることを期待される。

最後の3.の駅情報についてはセグメントを形成する2点に対して最近接の駅までの距離を特徴量として与える。例えば最近接の駅がとても近い際には駅で電車を待っていることやショッピングをしている可能性が高いため停止している可能性が高くなることが考えられる。

4.3. 分類器

本研究では分類器として SVM, RF (Random Forest), Gated CNN を用いる。本節では各モデルへの入出力と、モデルの内容について述べる。

まず、SVM と RF を用いた推定を行う際の入出力について述べる。Bolbal ら[13]の提案した手法を参考に、推定対象のセグメントの特徴量に加え前後のセグメントの特徴量も入力として用いて推定を行う。特に、SVM を分類器として用いる場合、これは Bolbal らの提案手法と同じである。また、セグメントの移動手段

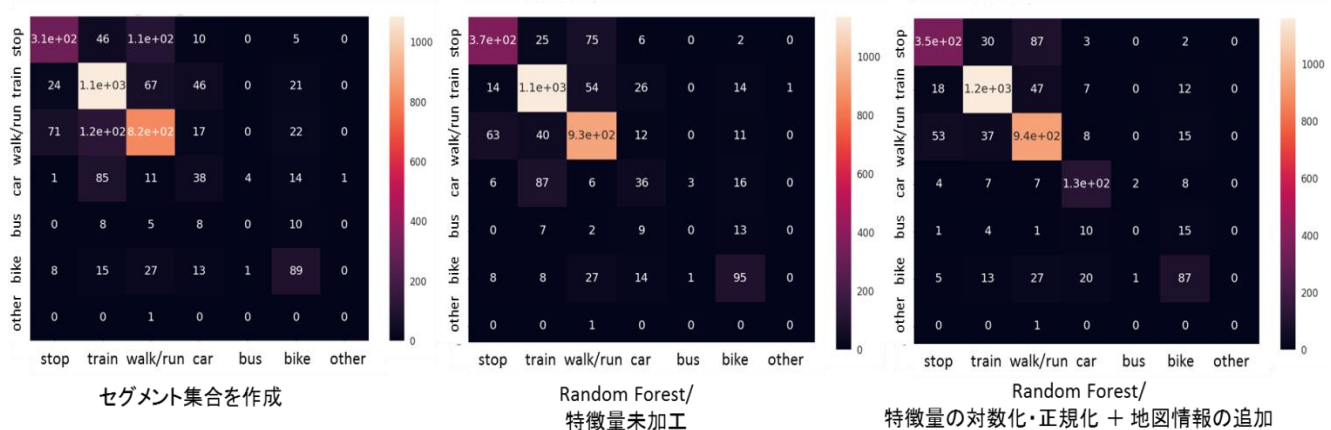


図8. 測位間隔が5分の移動軌跡に対する移動手段推定の混同行列. 列の要素が分類器の予測した移動手段を示し, 行の要素が実際の移動手段を示している. したがって混同行列の対角成分が大きいほど分類器の精度はよい. 一番左に示すベースライン手法と比較して, 真ん中の手法では精度が全体的に向上していることが確かめられる. また, 右に示す地図情報を追加した提案手法では特に「train」と「car」の分類性能が向上していることが確認できる.

表3. 測位間隔が5分の移動軌跡に対する移動手段推定の推定精度

測位間隔がランダム	特徴量未加工	対数化 + 正規化	地図情報の追加
セグメント集合を作成	0.752	-	-
SVM	0.463	0.803	0.822
Random Forest	0.822	0.823	0.856
Gated CNN	0.783	0.797	0.840

の推定を行う際に推定対象の前後のセグメントの特徴量も予測に用いる. 推定に用いるセグメントの数を window size と定める. SVM または Random Forest に与える特徴量の次元はセグメント1つの持つ特徴量次元と window size の積から与えられる. これらの入出力関係を図6に示す.

次に Gated CNN について説明する. Dauphin ら[18]によって提案された Gated convolutional networks (Gated CNN) は, Convolutional Neural Network[19]の畳み込み層に対し通常の畳み込みを行った要素と畳み込みを行った後にシグモイド関数を作用させた要素のテンソル積をとることで自然言語処理のタスクにおいて LSTM[20]と同程度の精度を実現した手法である. 本研究は移動軌跡が時系列データであるため Gated CNN が有効なモデルであると考え, これを用いた移動手段推定を提案する. window size 分のセグメントの特徴量を時間方向に畳み込みを行い移動手段の推定を行う. こ

こで Gated CNN で作成したモデルに対して各層に Drop out[21]と同層のニューロン間で正規化を行う Layer Normalization[22]を適用することで過学習を防止する.

5. 評価実験

5.1. 実験設定

本章では特徴量の対数化・正規化の前処理と, 地図情報の考慮が精度に与える効果を検証しながら, 4.3 で示した各分類器の性能をベースラインと併せて, 3つのデータセットを用いて比較する.

以下の4手法の推定精度を比較する.

1. セグメント集合作成
2. 未加工のセグメント特徴量
3. セグメント特徴量の対数化・正規化
4. 地図を用いた特徴量の追加

まずベースライン手法となる1のセグメント集合作成は, Zheng らの手法[12]である. セグメントを5つずつまとめてそれぞれセグメント集合とし, 表1に示した Basic feature と Advanced feature をセグメント集合ごとに算出し推定を行う. Zheng らは SVM を分類器として用いていたが, 予備実験の結果 Random Forest の方が高い精度を示したため本研究では Random Forest を用いた.

2の未加工のセグメント特徴量では Bolbal らの先行研究[11]をもとにセグメントの特徴量として測位点間のセグメント長(m), 速度(m/s), 加速度(m/s^2), 方向転換角(rad)を用いて推定を行う. window size の幅はいずれも7を採用し SVM, RF, Gated CNN を用いて推定を行った. SVM を用いた推定は Bolbal らの提案手法と同一である.

3のセグメント特徴量の対数化・正規化では方向転換角を正規化しセグメントのセグメント長, 速度, 加速

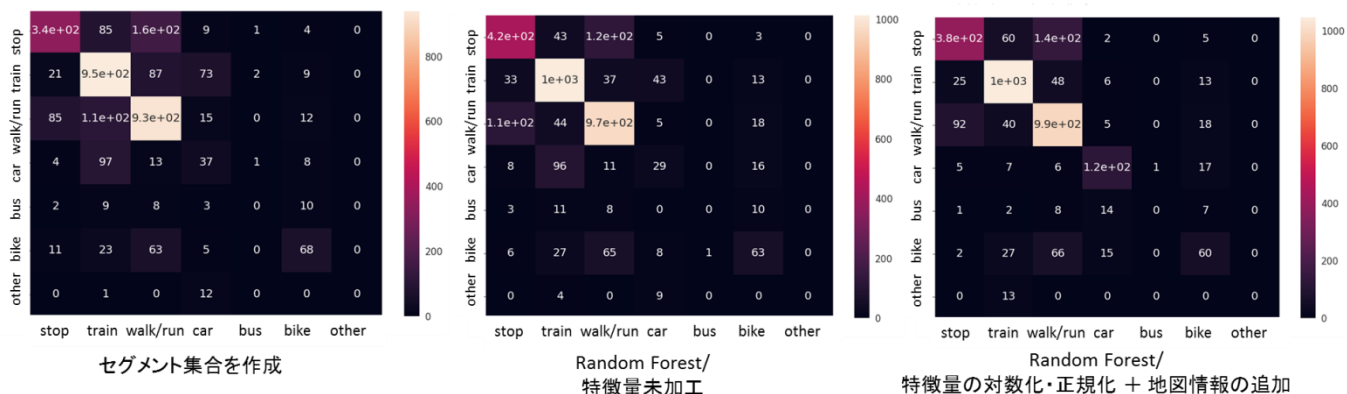


図9. 測位間隔が不定の移動軌跡に対する移動手段推定の混同行列.

表4. 測位間隔が不定の移動軌跡に対する移動手段推定の推定精度

測位間隔がランダム	特徴量未加工	対数化 + 正規化	地図情報の追加
セグメント集合を作成	0.711	-	-
SVM	0.416	0.752	0.763
Random Forest	0.766	0.764	0.800
Gated CNN	0.674	0.761	0.795

度に対して対数化・正規化し SVM, RF, Gated CNN を用いて推定を行った.

最後の4の地図を用いた特徴量の追加は前章の4.2で示した OSM を利用した特徴量を追加し SVM, RF, Gated CNN での推定を行った.

データセットは3.2で示した3つを用いてそれぞれ実験を行った. 30人のユーザーそれぞれの移動軌跡について, 前半7割を訓練用に, 後半3割を検証用に用いて学習と精度検証を行った. 測位間隔が5分の移動軌跡では, 訓練用データと検証用データそれぞれに含まれるセグメントの数は6641と3109だった.

精度評価は混同行列 (CM: Confusion matrix) と精度から行う. 混同行列は下記で与えられる.

$$CM = (cm_{ij})_{\substack{1 \leq i \leq 7 \\ 1 \leq j \leq 7}}$$

ここで, 混同行列の要素 cm_{ij} は分類器が「i」番目のクラスと判定したもののうち, 実際が「j」番目のクラス

であった件数で与えられる. また, 精度 (Accuracy) は全てのセグメント数に対して正解のセグメント件数の割合から推定の精度を定義した.

$$Accuracy = \frac{\text{正解したセグメントの数}}{\text{全てのセグメントの数}}$$

5.2. 実験1) 測位間隔が5分の移動軌跡を用いた実験結果

5分間隔の移動軌跡の推定結果において実施した1~4の実験についての結果と混同行列を表3と図8にそれぞれまとめた. 表3に示したとおりセグメント集合を作成したときの精度が75.2%であった. 実験条件2の未加工のセグメント特徴量を用いた測定ではSVM, Gated CNNに比較してRFが82.2%と最も高い精度を示した. この混同行列を確認したところセグメント特徴量に対し対数化・正規化を施したRFでは精度の向上は確認されなかったが, SVMに対しては精度の向上が確認できた. 混同行列にも示されるように, これらのセグメント直接推定においては「walk/run」「stop」「train」の分類に対してはよくできていることが確認できるが「train」と「car」の分類が不十分であることが分かる. 地図を用いた特徴量を追加したところ「train」と「car」の分類精度が向上した. 疎な移動軌跡の移動手段推定においては85.6%の精度を実現し, 既存手法に対して10%程度精度の向上を確認した.

5.3. 実験2) 測位間隔が不定の移動軌跡を用いた実験結果

表4および図9に示した, 測位間隔にのみノイズのある移動軌跡の推定結果は, 測位間隔が5分の移動軌跡に対する移動手段推定と比較していずれの手法においても5%程度の精度の減少が確認できるが, 手法間の精度の関係は変わらず, 提案手法がこの実験設定で

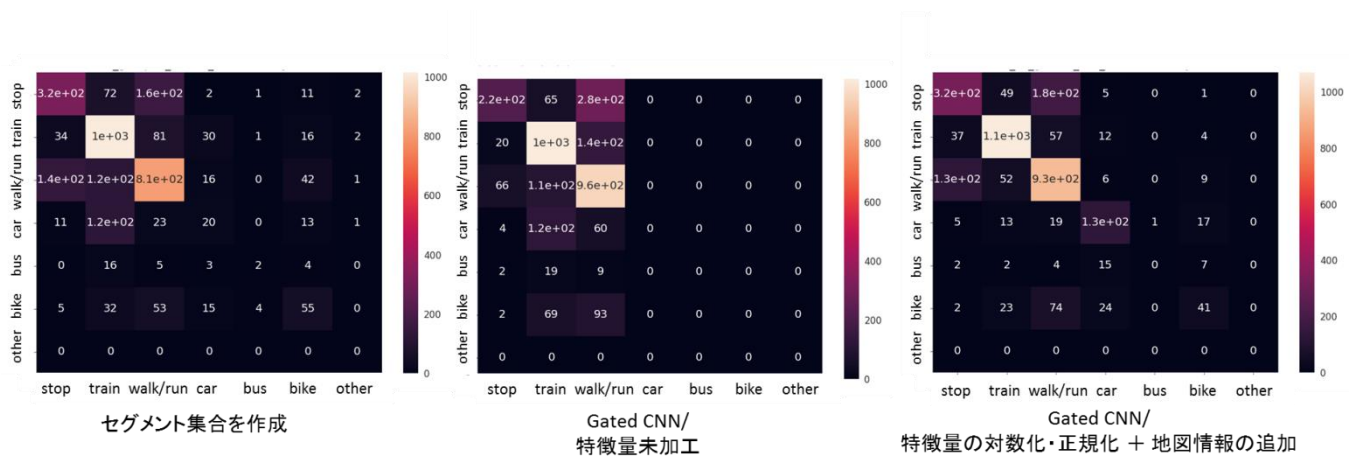


図10.測位間隔が不定で測位誤差の大きい移動軌跡に対する移動手段推定の混同行列.

表5. 測位間隔が不定で測位誤差の大きい移動軌跡に対する移動手段推定の推定精度

測位間隔がランダム	特徴量未加工	対数化 + 正規化	地図情報の追加
セグメント集合を作成	0.681	-	-
SVM	0.376	0.742	0.752
Random Forest	0.731	0.735	0.769
Gated CNN	0.705	0.739	0.776

も有効であると確かめられた.

5.4. 実験3) 測位間隔が不定で測位誤差の大きい移動軌跡を用いた実験結果

表5および図10に示した通り, 測位間隔不定の移動軌跡の場合, 測位間隔が5分の移動軌跡に対する移動手段推定と比較していずれの手法においても精度が7%程度減少することが確認できる. また, 最も良い精度を示したのは加工したセグメント特徴量に対し地図を用いた特徴量を追加しGated CNNで推定を行った場合であり, このとき77.6%の精度を示した.

5.5. 議論

3つの実験設定のいずれにおいても, 提案手法がベースライン手法よりも10%前後高い推定精度を示した. また特徴量の対数化と正規化による前処理, 地図情報の追加のどちらも精度向上に対して寄与していることが確かめられた. 一方で, 移動軌跡の測位間隔や測位誤差などのノイズを大きくしていくと, 提案手法も精

度が下がっていくことも確かめられた. これは今回用いた特徴量や分類器がいずれも時間的なノイズと地理的なノイズの双方から影響を受けることを示している.

いずれのデータセットでの実験においても, RFによる推定では量的特徴量の尺度を無視して順序として扱うため線形尺度に対しても対数尺度に対してもロバストとなり, 未加工の特徴量に示した値と近い結果を示している. 一方で, SVMやGated CNNでは特徴量の前処理が精度の向上に寄与していることが確かめられた.

図8, 図9, 図10に示されるように, 地図情報を考慮しない場合, 「train」と「car」の分類精度が低かった. 速度や加速度などの移動軌跡から取得できる特徴量ではこの二つを見分けることが難しいことを示している. 一方で地図情報を追加すると顕著に分類性能が向上していることが確かめられた.

6. まとめと今後の課題

本研究は, 実用的な場面で取得されうる移動軌跡に対し移動手段推定を行うことに取り組んだ. 生成した疎な移動軌跡の分析に基づく特徴量の前処理と地図情報を考慮した特徴量の追加と, それらを入力として用いる複数の分類器を提案した. 首都圏に在住する30人から取得した移動軌跡から人工的に生成した3つのデータセットを用いて精度の検証を行い, 既存のベースライン手法に比較していずれのデータセットでも10%前後の精度改善を達成した. この結果は, 提案手法が様々なノイズに対して頑健な手法であり実用的な場面での移動手段推定手法に適することを示している.

本研究では特徴量の加工として対数化と正規化を行い, 地図を用いた特徴量としては駅情報と線路情報を取り入れた. しかし, より実用的かつ汎用的な移動手段推定を考える際には, 地域における特徴量の考

案・そしてデータセットの蓄積が必要である。例えば、本条件では首都圏であったため地図を用いた特徴量では電車に乗っているかの判別に有用な線路・駅情報を取り入れたが、鉄道が発達していないエリアではバス停などのバスに関する特徴量が重要となることも考えられる。さらに、本研究で顕著な差を確認できなかった学習モデルの改良も必要となる。今後は今回得た知見を元により有益な知見を得ることに取り組む。

参 考 文 献

- [1] Helaoui, R., Riboni, D. and Stuckenschmidt, H.: A probabilistic ontological framework for the recognition of multilevel human activities, *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM, pp. 345–354 (2013).
- [2] Ouyang, R. W., Srivastava, A., Prabakar, P., Roy Choudhury, R., Addicott, M. and McClernon, F. J.: If you see something, swipe towards it: crowdsourced event localization using smartphones, *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM, pp. 23–32 (2013).
- [3] Saha, S., Chaudhuri, K., Sanghi, D. and Bhagwat, P.: Location determination of a mobile device using IEEE 802.11 b access point signals, *Wireless Communications and Networking*, 2003. WCNC 2003. 2003 IEEE, Vol. 3, IEEE, pp. 1987–1992 (2003).
- [4] Yang, D., Zheng, D., Yu, Z. and Yu, Z.: Fine-grained preference-aware location search leveraging crowdsourced digital footprints from LBSNs, *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM, pp. 479–488 (2013).
- [5] Diala, Nabouls., Marco, Fiore., Stephane, Ribot., and Razvan, Stanica. : Large-Scale Mobile Traffic Analysis: A Survey, *IEEE Communications Surveys & Tutorials*, pp. 124–161 (2015).
- [6] K, Nishida., H, Toda., K, Yoshida: Extracting Arbitrary-shaped Stay Regions from Geospatial Trajectories with Outliers and Missing Points, IWCTS'15 Proceedings of the 8th ACM SIGSPATIAL, November 03 doi:10.1145/2834882.2834884 (2015)
- [7] K, Yoshida., K, Nishida., H, Toda., A, Kawanobe: Deep_Feature_Extraction_from_Trajectories_for_Transportation_Mode_Estimation, Springer, doi:https://doi.org/10.1007/978-3-319-31750-2_5(2016)
- [8] JR 東日本アプリ <http://www.jreast-app.jp/>
- [9] Zheng, Y., Liu, L., Wang, L. and Xie, X.: Learning transportation mode from raw gps data for geographic applications on the web, *Proceedings of the 17th international conference on World Wide Web*, ACM, pp. 247–256 (2008).
- [10] NTT docomo イマドコサーチ機能 : <https://www.nttdocomo.co.jp/service/imadoco/usage/smt/keys.html>
- [11] MATSUSHITA, T. and TANAKA, T. :Study on Improving the Accuracy of Satellite Measurement in Urban Areas, *Transactions of the Society of Instrument and Control Engineers*, Volume 48 (2012) Issue 2 Pages 73-77
- [12] Zheng, Y., Chen, Y., Li, Q., Xie, X. and Ma, W.-Y.: Understanding transportation modes based on GPS data for web applications, *ACM Transactions on the Web (TWEB)*, Vol. 4, No. 1, p. 1 (2010)
- [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*, Volume 36, Issue 6, November 2012, Pages 526-537(2012)
- [14] Leon Stenneth., Ouri Wolfson., Philip, S., Yu and Bo, Xu.:Transportation mode detection using mobile phones and GIS information, *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Pages 54-63 (2011)
- [15] Coats, Adam; Ng, Andrew; Lee, Honglak.: An analysis of single-layer networks in unsupervised feature learning, *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 215-223, 2011
- [16] OpenStreetMap: <https://openstreetmap.jp/>
- [17] Nakamura, T., : A Study on the Relationship between Land Use around Railway Stations and the Railway Station Passengers, *Journal of the City Planning Institute of Japan*, Vol.50 No.3, October, 2015
- [18] Yann N. Dauphin, Angela Fan, Michael Auli, David Grangier: Language Modeling with Gated Convolutional Networks, cs,abs/1612.08083, 2017.
- [19] Alex, Krizhevsky., Ilya, Sutskever. and Geoffrey E. Hinton:ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems* 25 (2012)
- [20] Jozefowicz, Rafal, Vinyals, Oriol, Schuster, Mike, Shazeer, Noam, and Wu, Yonghui. Exploring the limits of language modeling. arXiv preprint arXiv:1602.02410, 2016.
- [21] Nitish, Srivastava., Geoffrey, E Hinton., Alex, Krizhevsky., Ilya Sutskever and Ruslan Salakhutdinov :Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research* 15 1929-1958 (2014)
- [22] Jimmy Lei Ba, Jamie Ryan Kiros and Geoffrey E. Hinton: Layer Normalization, arXiv preprint arXiv:1607.06450