

駅-まち空間の電波強度分布に基づく 自己軌跡特定問題

後藤 祥孝¹・井澤 佳那子²・羽藤 英二³

¹非会員 東京大学大学院 工学系研究科社会基盤学専攻 (〒113-8656 東京都文京区本郷 7-3-1)
E-mail: goto@bin.t.u-tokyo.ac.jp

²学生会員 東京大学大学院 工学系研究科社会基盤学専攻 (〒113-8656 東京都文京区本郷 7-3-1)
E-mail: izawa@bin.t.u-tokyo.ac.jp

³正会員 東京大学大学院 工学系研究科社会基盤学専攻 (〒113-8656 東京都文京区本郷 7-3-1)
E-mail: hato0816@gmail.com

本研究では都市空間内において旅行者、特に歩行者が移動する際の経路選択行動を扱う。従前、経路選択行動は GPS を用いて取得される位置座標軌跡により移動経路の特定を行った上で、経路選択モデルなどを用いて行動分析や政策評価を行ってきたものの、GPS データは誤差が屋内やストリートジャングルにおいて増幅することや 2 次元座標での情報のみしか得られないという問題を抱えていた。本研究は端末から得られる Wi-Fi 受信電波強度の記録データを利用した 3 次元ネットワークにおける移動経路特定および経路選択モデルのパラメータ推定を行うことにより、歩行者の詳細な経路選択行動に関して考察することを目的とする。

Key Words : Route choice model, 3-D network, Recursive logit model, Wi-Fi data, Machine learning

1. はじめに

急速な都市の発展により複雑化した駅まちでは、歩行者が安全かつ快適に過ごすことができる空間を形成するためにその詳細な行動把握が求められている。GPS と携帯通信機器を用いたプローブ型の観測手法が発展し、車や人の微細な行動が捉えることが可能となってきたものの、GPS では屋内空間の観測が難しく、その詳細な行動把握が容易ではないために都市計画上有効な施策検討が困難であった。しかし、近年スマートフォンの普及により、GPS 以外にも Wi-Fi、加速度、磁気など多様なデータが取得できるようになり、従前明らかにすることが出来なかった駅構内といったより詳細なスケールの行動データの観測が可能となってきている。こうした行動データを利用し、個人の移動軌跡や滞在状況を明らかにすることによって、歩道拡幅やプログラム配置といった都市計画的施策の、個人の回遊行動や時間の消費、経済効果に関する知見を得ることが都市計画上求められている。

こうした背景のもと、交通行動の観測とモデリング、特に経路選択モデルはネットワーク上の交通行動の変化を政策変数によって評価できることから都市・交通分野における最重要手法の 1 つとされてきた。プローブパーソン (PP) 調査は GPS や加速度センサによる観測と web ダイアリによる行動文脈記録により時空間上での行動データが得られるため、歩行者の詳細な挙動を

解明することができる調査として近年おおいに活用された。しかしながら、その調査費用は莫大であり大量のデータを得ることが難しく、GPS は 2 次元位置情報であるため 3 次元ネットワークで構成される実空間では詳細な行動軌跡を把握することは難しい。一方、Wi-Fi や Bluetooth による観測はプライバシー上の問題を有するものの、長期間の大規模データを取得することができる上、3 次元ネットワーク上で数 10cm～数 m 単位での位置特定が可能であり今日急速に着目されている。

Wi-Fi データは 3 次元位置情報を持ち、2 次元位置情報を有する GPS の使用が念頭に置かれていた従来の経路選択行動の観測手法を適用することは難しい。本研究では、Wi-Fi データを用いた歩行者の 3 次元ネットワーク上の行動分析を行うために、経路選択行動の観測を 3 次元ネットワークに拡張するとともに機械学習の手法を取り入れた位置特定手法の提案を試みる。Wi-Fi データを使用した経路特定及び経路選択モデルの推定を通じて、3 次元ネットワーク上の経路選択行動を考察することが本研究の目的である。

2. 既往研究

(1) 軌跡特定手法

マップマッチングは一般に観測した位置データの観測誤差を補いネットワーク上で移動経路を特定する手法である。分析のために GPS データを経路データに変

換する手法についてこれまでに様々なマップマッチングアルゴリズム⁵⁾⁷⁾⁶⁾が提案されており、Wi-Fi データを経路データに変換する手法 (Danalet et al., 2014)³⁾も近年徐々に提案されてきている。しかしながら、マップマッチングアルゴリズムを適用することでリンクベースの詳細な走行経路を特定することができる。しかし詳細な経路データを使用する場合マップマッチングに膨大な計算時間を要し、その後の分析に用いるデータ量も膨大になるという欠点を有している。

屋内における歩行者を対象とするセンシング研究には、屋内位置推定、歩行者流トラッキング、群衆密度推定といったものがある。屋内位置推定は、一人の歩行者に着目した屋内におけるリアルタイムな位置座標の推定を行う。歩行者流トラッキングでは、環境の側に着目し環境内の歩行者の通過数や複数の歩行者に対する移動軌跡の特定を目的としており、群衆密度推定では空間内における歩行者の密度推定を行い、混雑状況の把握を目的としているため、経路選択行動の分析には適さない。

機械学習は教師データに基づいてデータ集合の持つ規則性を自動的に見つけ出すアルゴリズムであり、得られた規則性に基づいて新たなデータに対する予測を行うことが可能である (Shafique & Hato, 2015¹³⁾)。機械学習は教師データの形式によって教師あり学習と教師無し学習に大別される。教師あり学習では、教師データは入力ベクトル x と出力ベクトル t の組 $(x; t)$ からなる。入力ベクトル x は観測値や事前に決定される変数であり、出力ベクトル t は入力ベクトル x により影響を受ける変数である。

特に教師あり学習の特徴として、概して教師データがより多いほど未知データに対しての出力の精度はより高くなるというものがある。現在は多様なセンサが搭載されて様々なデータの収集が可能であるスマートフォンのような端末が一般に対して広く普及していることもあり、多様かつ大量のデータの収集が可能となってきている。こうした背景の中で、大量のデータ集合を扱い規則性を自動的に見つけ出す手法である機械学習の重要度がより高まっていると言えるだろう。また、教師あり学習では座標系上での位置が既知である Wi-Fi データを教師データとして利用することで未知データの位置を推定することができる。教師データの多寡が推定精度を左右するものの、ネットワーク上の離散的なノード、リンクといった情報にマッチングする際には、分類木を用いることで座標を介さずにデータを扱うことができるため、未知数の多い観測データを使用したマップマッチングに機械学習手法を導入することには妥当性もある。

教師あり学習の 1 つであるランダムフォレスト⁸⁾⁹⁾

は、複数の決定木を組み合わせることによって分類・回帰を行う機械学習アルゴリズムであり、学習データからランダムサンプリングによりサブサンプルを生成して各決定木に当てはめること、学習データの説明変数が複数存在するときにランダムに選択された変数を用いて決定木でのノードにおける分割を行うことの 2 つのランダム性を有しているため、マップマッチングを行う手法として適合性があると考えられる。以下に、フローを示す。

d 次元の入力データ $v = (x_1, x_2, \dots, x_n)$ 学習データ集合を $S = v$ 、生成する決定木の数を T とする。 t 番目 ($1 < t < T$) の決定木の根ノードにランダムサンプリングされた学習データ集合 $S^t_0 \in S$ が与えられたとき、ノードにおける最適な分割関数を求めて子ノードにデータを分割し、子ノードでも分割関数を決定することを繰り返して決定木を生成する。ここで分割関数とは、決定木のノードにおいて学習データを分ける関数であり $h(v; \theta)$ という形で書ける。 θ は分割を決定するパラメータであり、入力 v からランダムに特徴量を抽出するフィルタ、分割基準の幾何特徴と閾値を決定するパラメータとして働く。ノード i の学習に用いるデータを S_i とし、分割された左右の子ノードの学習データ S^L_i, S^R_i とする。ノードにおける分割関数は θ の最適パラメータ θ^* によって決定されるため次式から求める。

$$\theta^* = \arg \max I(S_i, S^L_i, S^R_i, \theta) \quad (1)$$

目的関数 I はエントロピーと情報利得の式によって定義される。集合 S 内のデータにクラスラベル $c \in C$ が対応しているときエントロピーは次式のように表される。なお、 $p(c)$ は集合 S 内のデータのクラス cc に分類される割合である。

$$H(s) = - \sum_{c \in C} p(c) \log(p(c)) \quad (2)$$

S 内が単一のクラスの時 $H(S) = 0$ 、 S 内のデータが全クラスで等確率の時 $H(S)$ は最大となる。また、ノードの分割によって得られる情報利得は次式のように表される。

$$I = H(S) - \sum_{i \in L, R} \frac{S^i}{S} H(S^i) \quad (3)$$

式 (3) の第 1 項は分割前エントロピー、第 2 項は分割後のエントロピーを意味している。 I が大きいほど分割により得られる利得が大きくなるため、 I が最大となるようなパラメータを決定する。以上のような分割の繰り返しによりノードに割り当てられたデータが一定値以下になったところで木の生成が終了しクラスラベル $c \in C$ に対して、確率 $p_t(c|v)$ が決定される。全ての木の生成が終了すると相加平均 $p_t(c|v) = \frac{1}{T} \sum p_t(c|v)$ によってランダムフォレストによるクラス識別を行う。

(2) 経路選択モデル

ネットワーク上の個人の行動を記述するために様々なモデルが研究されてきた。経路選択モデルはネットワーク上での個人の移動経路の選択を扱う。歩行者行動の分析では、交通機関選択のような一般的な離散選択モデルとの違いとして選択肢集合の複雑さがあげられ、選択肢を明示的に列挙する方法と選択肢列挙を必要としない選択確率の導出という 2 つのアプローチで経路選択問題が扱われてきた。本研究は、駅舎を含む駅まちにおける回遊行動の分析を目的とするため、逐次的な経路選択行動を記述できる経路非列挙型の経路選択モデルを適用する。

Fosgerau ら (2013)¹⁰⁾ により提案された Recursive Logit(RL) モデルはマルコフ連鎖の考え方を応用し、遷移確率行列を動的離散選択モデルにより記述することで経路列挙型の経路選択モデルと整合的な行動を記述することを可能としている。状態遷移の際に目的地までの期待効用を考慮することで、目的地からの再帰的な構造で確率を求めることにより、経路列挙を不要とする経路選択モデルとして提案された。状態 a から a' に遷移する際の経路選択確率は直接的効用 $v(a'|a)$ と期待効用の評価関数 $V(a')$ と期待効用に対する時間割引率を用いて次式で表される。

$$P(a'|a) = \frac{\exp\left(\frac{1}{\mu}(v(a'|a) + \beta V^d(a))\right)}{\sum_{a' \in A(k)} \exp\left(\frac{1}{\mu}(v(a'|a) + \beta V^d(a'))\right)} \quad (4)$$

3. モデルのフレームワーク

(1) Wi-Fi データの特徴

Wi-Fi 受信端末はネットワーク上に設置された Wi-Fi 基地局から発信される電波によって通信を可能としており、本研究ではそうした Wi-Fi 受信記録をネットワーク上の位置特定に利用する。本研究で観測データとして与えられるのは、端末が受信した基地局 (複数も許容) の識別 ID とそこからの電波強度である。時系列データを用いる場合は、時刻、基地局 ID、電波強度の列を位置特定に用いる。以下ではこの形式のデータを Wi-Fi ログデータと呼ぶ。電波強度が基地局との距離を変数として減衰する関数に従うとすれば、電波強度データは基地局との距離に変換することができる。従前の行動データの取得や行動分析に用いられてきた GPS によるデータは、2 次元座標上での絶対的な位置を表現するのに対し、Wi-Fi ログデータの場合は、基地局との距離という相対的な位置関係の情報しか得ることができないため移動経路特定を行う上では相対的な位置情報を絶対的な位置情報へと変換する必要がある。

(2) ランダムフォレストによる滞在リンク予測

データ取得位置が既知である Wi-Fi ログデータを教師データとし、新たな Wi-Fi ログデータの入力に対して位置を出力する機械学習の手法を用いることにより、移動経路をネットワーク上のリンクの時系列データとして特定する。機械学習手法を用いることで、現実のネットワークに存在する建物構造や周囲の人の移動により複雑になる Wi-Fi 基地局からの電波強度の誤差の記述を避けることが可能となると考えられる。

Wi-Fi ログデータを利用する場合、入力データベクトルは $x = (m_1, m_2, \dots, m_N)$ と書ける。ここで m_n は n 番目の基地局の電波強度を表す。また、出力となる分類は Wi-Fi ログデータを取得したときの滞在リンク $a \in A$ となる。ランダムフォレストを用いた学習では、教師データは $X = ((x_1, a_1), (x_2, a_2), \dots, (x_D, a_D))$ と書ける。予測を行う入力データ集合 D は (t, x_t) と表される。ここで t は Wi-Fi ログデータの取得時刻である。時刻 t の Wi-Fi ログデータ x_t を、教師データにより学習を行ったランダムフォレストに入力することにより、リンク a_t が出力され、時系列の Wi-Fi ログデータに対して、時系列の予測リンクデータが得られる。電波強度の誤差分散の情報を明示的に扱わないため、Wi-Fi の電波強度に影響を与える壁、天井との距離や周囲の建物といったネットワーク上の環境条件との関係を扱うことを避けることが可能である。

(3) 予測滞在リンクを用いた経路特定

ランダムフォレストによる出力は各時刻における滞在リンクであるため、幾何的マップマッチング手法である point to curve と同様に、リンクを順に結ぶことによってネットワーク上の経路とすることができる。最も単純な方法であるが、予測リンクが実際には接続していないことや、データに外れ値が存在し大きく外れた予測リンクを出力するなどの、予測誤差を吸収できないという場合が発生する。

さらに、ログデータの取得頻度が 1 秒ごとなど高頻度である場合には、各リンクを通過する間に複数の Wi-Fi ログデータが取得されることが考えられ、各時刻での滞在リンク予測においては Wi-Fi ログデータの外れ値に対して頑健性がないという欠点がある。そこで出力時系列リンクデータを一定の時間幅ごとに切り分けて時間構造化し、時間幅ごとの滞在リンク予測を行うことを提案する。

トリップに対して時系列順に並ぶ Wi-Fi ログデータ列 $\hat{m} = (\hat{m}_1, \dots, \hat{m}_n, \dots, \hat{m}_N)$ が得られたと仮定する。まず、一定の幅を持つ時間列 $(1, \dots, t, \dots, T)$ を定義する。ここで全ての $\bar{t} = (t^-, t^+)$ に対して $t = t^+ - t^-$ は一定である。時間 t 内に含まれるログデータの列を

表-1 調査手順

1.	各調査日、渋谷駅に近づいた時点で GALAXY NEXUS と Wi-Fi ログ収集アプリケーションを立ち上げる.
2.	データ送信が行われていることを確認した上でアプリを起動した状態で調査対象地点へ移動する.
3.	調査対象地点に到着後、以下の操作を行う. 3-1. 端末状態の安定を確保するため 30 秒程度静止させる. 3-2. 記録開始時刻、調査地点を記録する. 3-3. 3 分間端末の状態を一定に保つ. 3-4. 記録終了時刻、調査地点を記録する.
4.	次の調査地点へと移動し、3 に戻る.

表-2 テストデータ概要

取得日時	2016/11/1
Wi-Fi ログデータ数	458 件
生データ基地局数	220 種類
クリーニング後基地局種類数	219 種類

$\hat{m}^t = (\hat{m}_i^t, \dots, \hat{m}_j^t, \dots, \hat{m}_J^t)$ と定義することにより、 $\hat{m}$ を $(\hat{m}^1, \dots, \hat{m}^t, \dots, \hat{m}^T)$ のように時間 t ごとのログデータ列に分解する. ある時間内 t 内に含まれる全ての観測点が同じリンクに属するという仮定に基づいて $t = 1$ から順番にリンクの確定を行う. 時間 t 内の観測データ数を J とすると、各観測データに対してランダムフォレストによる予測を行うことにより、 J 個の予測リンクデータが出力される. J 個の Wi-Fi ログデータの予測リンクについて最も多く出力されたリンクを時間 t における通過リンクとして確定させる.

4. ケーススタディ

(1) 調査概要

a) 調査方法

本研究での実証的分析で用いたデータの概要について説明する.

Wi-Fi ログデータの収集においては一般にも利用されているスマートフォンであるサムスン電子製 GALAXY NEXUS SC-04D を用いる. 端末上で、Wi-Fi ログデータ収集用アプリケーションを起動しデータの収集を行う. データは約 1 秒間隔で取得され取得時刻、基地局の SSID、BSSID、受信電波強度が記録されていく. SSID は任意に設定できる基地局識別名であり同一 ID の基地局も存在する. BSSID は基地局の MAC アドレスをそのまま利用した識別 ID で、個々の基地局に固有で割り振られるため基地局の識別が可能である. 本研究では、データ取得時刻、BSSID、受信電波強度を利用する.

調査日は 2016 年 11 月 17 日, 11 月 24 日, 12 月 1 日, 12 月 8 日の全 4 回である. 表?? の手順に沿って各日で調査を行った.

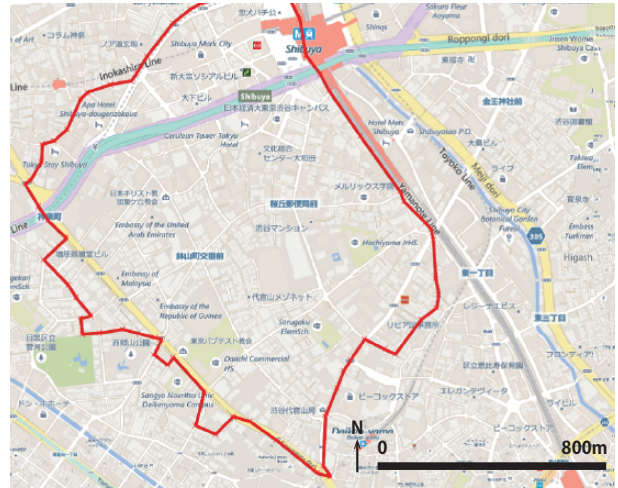


図-1 対象地域：渋谷桜丘地区

b) 経路データ

渋谷駅、東急百貨店東横店、渋谷マークシティにより構成される駅施設内に絞った場合のリンク予測と、図 1 で赤線に囲まれた渋谷駅と周辺市街地を含むネットワーク上でのリンク予測結果を示す. ネットワークデータは交差点をノード、街路をリンクとして作成し、歩行者経路選択行動を分析するため、特に片側 2 車線以上の道路の場合は道路の両岸にリンクを設け、横断歩道や歩道橋のリンクを作成した.

予測に用いた教師データを表?? に示す. 教師データは対象範囲内の全リンクに対して、それぞれリンク通過時刻を記録しながら Wi-Fi ログデータを取得することによって作成した. Wi-Fi ログデータは約 1 秒間隔で取得される. したがって、教師データ取得においてはリンク上を移動しながら連続的にデータが収集され、テストデータに関しても同様に約 1 秒間隔で Wi-Fi ログデータを取得が取得される. このとき、教師データと同様にリンク通過時刻の記録を行うことで、予測精度の確認可能となるようにデータの取得を行った.

予測は教師データをそのまま学習した場合、データクリーニング後教師データとテストデータで共通で観測された基地局のみを抽出した教師データを学習した場合とで行った. テストデータは下表 2 の通りである. 表 34 に予測結果を示した. 各時刻におけるランダムフォレストの予測精度は決してよいとは言えないが、生データをそのまま学習に用いるのに比較して、共通基地局のみを取り出して学習を行うことで精度は大きく向上した. 個々の予測結果と実状態を比較した場合の誤りの形式は表 5 のように整理される.

表-3 ランダムフォレストによる位置予測結果比較

	正答率
データクリーニング無し	34.3%
データクリーニング後	52.4%

表-4 教師データ概要

	全データ	建物内データ
取得日時	2016/9/11, 9/25, 9/28,10/1	2016/9/28,10/1
ログデータ数	20686 件	4979 件
基地局種類数	15737 種類	2032 種類

表-5 リンク予測失敗の例

隣接リンク	108
異なるフロア	28
その他	82
合計	218

表-6 経路特定手順

Step.1	各時刻における Wi-Fi ログデータからランダムフォレストにより予測リンクを出力する。
Step.2	全時刻の予測リンクの集合をトリップにおける滞在可能リンク集合とする。
Step.3	滞在可能リンクからなるネットワークにおいて端点となっているノードを列挙し、端点ノード集合とする。
Step.4	端点ノードから、滞在可能リンク集合からなるネットワークの全ノードに対する最短経路を探索する。
Step.5	端点ノード集合内の各ノードについての Step.4 により得られた最短経路集合を構成するリンクを滞在可能リンク集合に加える。
Step.6	滞在可能リンク集合から構成されるネットワークをトリップにおける滞在可能ネットワークとし全体ネットワークから抽出する。

(2) 経路選択モデル

a) 定式化

本節では、渋谷駅一桜丘地区 (図 1) においてネットワーク条件を説明変数とする経路選択モデルを RL モデルにより記述する。本研究におけるリンク遷移の直接効用は以下の式で定める。

$$v(i|j) = \beta_1 LL_i + \beta_2 ST_i + \beta_3 ES_i + \beta_4 SW_i \quad (5)$$

ただし i, j はリンク、 LL_i はリンク i の長さ、 ST_i はリンク i の階段ダミー変数、 ES_i はリンク i のエスカレーターダミー変数、 SW_i はリンク i の歩行者専用リンクダミー変数、 β は各説明変数のパラメータである。

b) 移動経路の特定

ランダムフォレストによる時刻ごとの滞在リンク予測結果を用いて、移動経路特定手法及び経路選択モデルのパラメータ推定手法を提案する。1 トリップについてランダムフォレストによる各時刻の予測結果から、そのトリップにおける滞在可能ネットワークを表 6 のようにして出力する。

この操作により各トリップについて、ランダムフォレストによるリンク予測結果から滞在可能ネットワー

表-7 経路特定と経路選択モデルパラメータ推定の統合手法

Step.1	経路選択モデルパラメータの初期値 θ_0 を設定する。
Step.2	トリップごとに得られた滞在可能ネットワークにおいて目的地までの RL 型リンク遷移確率行列を θ_0 を用いて計算する。
Step.3	各トリップで始点から終点に到達するまでリンク遷移を繰り返し、リンク遷移結果を出力する。
Step.4	全トリップについて Step.2,3 を行い、全体の経路選択結果として出力する。
Step.5	得られた経路選択結果を用いて、全体ネットワークにおいて RL 型経路選択モデルのパラメータを推定し、推定結果 θ を得る。
Step.6	閾値 ξ を設定し、 $ \theta - \theta_0 < \xi$ を満たせば θ をパラメータとして採用する。そうでなければ、 $\theta B \theta_0$ として Step.2 に戻る。

クが確定的に得られる。Step.2 において予測リンク集合を滞在可能リンク集合として設定するということは、ランダムフォレストにより出力された全てのリンクは、通過した可能性のあるリンクとして扱われるということを意味し、Step.4 において端点ノードからの最短経路探索は、通過した可能性のあるリンクとした 1 つの滞在可能リンクからは他のすべての滞在可能リンクに対して接続していたという考え方に基づいている。滞在可能ネットワークの抽出はランダムフォレストによる予測リンク集合をベースとして行っているため、最短経路探索というネットワーク情報のみに基づくリンクの列挙を行いながらも、あくまでも Wi-Fi ログデータを用いた観測に基づいたネットワークの限定を行っているといえるだろう。

次にトリップごとの滞在可能ネットワークを用いた経路特定と経路選択モデルパラメータ推定の統合手法を提案する。

表 7 の操作により、ランダムフォレストを用いたリンク予測結果を用いたネットワーク集合を与え、RL による経路選択確率の計算と遷移確率行列に従ったシミュレーション計算を行うことで、観測結果を反映した経路選択結果集合からの経路選択結果の特定を可能にした。

c) 推定結果

ランダムフォレストによる経路特定精度の検証のために、テストデータ作成時に記録されたリンク遷移を経路選択結果の用いた場合の推定を行った。推定結果は以下の表 8 である。リンク長、階段ダミー変数、エスカレーターダミー変数のパラメータがそれぞれ負に作用している。リンク長については、リンク長が長くなるほど経路効用が小さくなるということを意味しており、一般的な経路選択行動の際の規範と合致している。また、階段ダミー変数よりもエスカレーターダミー変数の方がより強く負に作用している。一般には自ら歩行する必要のないエスカレーターの方が階段よりも効用が大きくなると考えられるが、本研究において行動データ収集を行ったのは 20 代の男女 3 名であるため、

表-8 推定結果

説明変数	パラメータ値	t 値
リンク長 [10m]	-4.86	-5817.77
階段ダミー	-0.62	-95.22
エスカレーターダミー	-1.71	-0.65
歩行者専用リンクダミー	2.07	442.84
サンプル数	88	
初期尤度	-150.57	
最終尤度	-125.08	
尤度比	0.169	

階段を使うことにより生じる労力に対する価値認識よりも、階段を使う方がエスカレーターに乗るよりもより短い時間で移動が可能であるという価値認識の方が大きかったのではないかと解釈できる。最後に、歩行者専用リンクダミー変数が正に作用していることは、歩道を含む歩行者専用となっている道路の方が車道と歩道の境界がない道路よりも選択されやすいということの意味している。歩者分離が行われている道路のが自動車に対して感じる危険の度合いがより小さく、歩きやすいという解釈することができる。

5. おわりに

本研究では、Wi-Fi データを使用して機械学習による観測手法の提案と実データによる検証を行った。次に機械学習による位置予測精度が低いことを前提とした、位置予測結果を用いた経路と経路選択パラメータの同時推定を行う手法の提案を行った。位置予測結果から個々のトリップに対して滞在可能ネットワークを抽出し、滞在可能ネットワーク上でのリンク遷移確率にしたがった経路を出力するため、位置予測という観測とモデルによる行動原理の双方にしたがった経路特定手法といえるだろう。提案手法においては経路出力時のパラメータと出力経路から推定されるパラメータとが一致するような収束計算を行うことにより、経路特定結果とパラメータ推定結果の整合性を取った。

しかしながら、本研究は多くの課題を残している。Wi-Fi ログデータを安定して取得できていないために機械学習の精度が低くなっているということが考えられ、位置推定における予測精度の向上に向けては、安定的な教師データの作成が課題となる。しかしながら、安定した教師データを得るためにはネットワーク規模に対する教師データの取得数が深くかかわっており、データ収集コストとのトレードオフの関係にある。コストを抑えたデータ収集方法や、不安定な情報を教師データから除去するためのデータクリーニング手法の検討が必要であろう。一方で基地局情報の不安定性にはデータ取得時の周囲の環境の情報が含まれているともいえる。データクリーニング後の安定した Wi-Fi データに

よる歩行者の行動観測に合わせて、クリーニングによって除去されるデータについての分析を進めることにより、時刻ごとに出力される Wi-Fi データから位置のみではなく、端末周囲の歩行者密度や店舗密度に関する情報を得ることができると考えられる。

最後に、移動経路と経路選択パラメータの同時推定手法に関して、実データを用いた推定計算と精度検証を行う必要がある。その上で提案手法の、滞在可能リンク決定の段階でランダムフォレストによる出力割合による滞在可能性閾値の設定や、端点ノードから最短経路探索をするノードを決定する際の距離の閾値設定によりランダムフォレストによる予測誤差分散を考慮するなど、機械学習の予測精度に合わせたモデルの調整を行う必要があると考えている。

参考文献

- 1) White, C. E., Bernstein, D., and Kornhauser, A. L.(2000), "Some map matching algorithms for personal navigation assistants", *Transportation Research Part C: Emerging Technologies*, Vol.8, No.1, pp.91-108.
- 2) Bierlaire, M., Chen, J., and Newman, J. (2013), "A probabilistic map matching method for smartphone GPS data", *Transportation Research Part C: Emerging Technologies*, Vol.26, pp.78-98.
- 3) Danalet, A., Farooq, B., and Bierlaire, M. (2014), "A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures", *Transportation Research Part C: Emerging Technologies*, Vol.44, pp.146-170.
- 4) 佐藤智美, 小宮山哲, 下田雅彦, 劉渤江, and 横田一正 (2014), "Bluetooth の電波強度を用いた位置推定方式の検討", *In DEIM Forum*, pp.1-6.
- 5) 朝倉康夫, 羽藤英二, 大藤武彦, 田名部淳 (2000), "PHS による位置情報を用いた交通行動調査手法", 土木学会論文集, No.663, IV-48, pp. 95-104.
- 6) 薄井智貴, 三輪富生, 山本俊行, 森川高行 (2009), "歩行者プローブデータ多面的活用のためのデータクリーニング手法に関する研究", 土木計画学研究・講演集, Vol.39, CD-ROM.
- 7) 三谷卓摩 (2005), "プローブパーソン型交通情報配信システムの適用可能性に関する研究", 博士課程学位論文, 愛媛大学.
- 8) Breiman, L., (2001). "Random forests", *Machine learning*, Vol.45, No.1, pp.5-32.
- 9) 波部斉 (2012), "ランダムフォレスト", 情報処理学会研究報告.
- 10) Fosgerau, M., Frejinger, E., Karlstrom, A. (2013), "A link based network route choice model with unrestricted choice set", *Transportation Research Part B: Methodological*, Vol.56, pp.70-80.
- 11) 佐佐木綱 (1965), "吸収マルコフ連鎖による交通流配分理論", 土木学会論文集, Vol.121, pp.28-32.
- 12) 大山雄己, 羽藤英二 (2016), "移動軌跡情報に基づく時間構造化ネットワーク上の交通配分", 土木計画学研究・講演集, Vol.53, CD-ROM.
- 13) Shafique, M. A., and Hato, E. (2016). "Travel mode detection with varying smartphone data collection frequencies", *Sensors*, Vol.16, No.5, pp.716.