

長崎県域における行動ログデータの分析に 基づく滞在施設属性の推定

渡部 康祐¹・菅原 優²・森 賢二³・門間 俊幸⁴・松本 修一⁵・山口高平⁶

¹非会員 日本工営株式会社福岡支店第1技術部社会システムグループ
(〒812-0007 福岡県博多区東比恵1-2-12R&Fセンタービル5F)

E-mail: a5669@n-koei.co.jp

²非会員 慶応義塾大学理工学部管理工学科
(〒223-8522 神奈川県横浜市港北区日吉3-14-1)

E-mail: y.sugawara.z7@gmail.com

³非会員 国土交通省九州地方整備局福岡国道事務所調査課
(〒813-0043 福岡県福岡市東区名島3丁目24-10)

E-mail: mori-k8912@qsr.mlit.go.jp

⁴非会員 兵庫県県土整備部土木局道路街路課
(〒650-8567 兵庫県神戸市中央区下山手通5-10-1 (1号館10階))

⁵正会員 文教大学情報学部情報社会学科
(〒253-8550 神奈川県茅ヶ崎市行谷1100)

E-mail: shuichi@shonan.bunkyo.ac.jp

⁶非会員 慶応義塾大学理工学部管理工学科
(〒223-8522 神奈川県横浜市港北区日吉3-14-1)

E-mail: yamaguti@ae.keio.ac.jp

紙媒体が主流で行われてきた交通行動調査は、調査日や調査期間に限られるとともに多くの人的、金銭的リソースがかかっていた。近年、GPS付き携帯端末や位置情報アプリサービスの普及に伴い、情報通信技術を用いた交通行動調査も実施され始めている。本研究では、長崎県内における道路利用者等を対象に、スマートフォンから取得されたGPSデータから取得される行動ログデータを用いて、利用者の行動特性、立ち寄り特性等の把握分析手法について検討を行った。各利用者の滞在地点を抽出するために、Google Places API、施設オントロジー、行動ログデータに基づく規則性や時間的順序関係を扱う推論機構を用いて滞在施設属性を推定した。これにより、行動特性把握、交通行動調査の高度化・効率化に資する手法を検討した。

Key Words : ITS, TDM, Park & Ride, Smartphone, Smartphone Application, GPS, Location System

1. はじめに

現在、多くの大規模交通行動調査が実施されている。しかし、個人情報保護法の施行などにより、個人を特定できるデータの大規模な収集、解析が困難になってきており、「交通 IC カードや ITS を活用した安価で高精度な需要の連続的把握手法、交通行動メカニズムや交通サービスの評価構造の把握が必要」¹⁾と指摘されている。そのような中、交通行動調査に対する IT の活用も徐々に見られている。しかし、これら新たな交通行動調査手法に対して、体系的な整理がなされていないという問題がある。このような現状から、IT の活用により交通行動調査を高度化し、簡便かつ安価で高精度なものにすることの重要性が増してきている。

紙媒体による交通行動調査の一つとして、パーソントリップ調査（以下、「PT調査」と記す）が行われている¹⁾。PT調査は、人の動きを把握する調査で、交通行動の起終点、出発・到着時間帯、利用交通手段、滞在施設、施設の種類、行動目的などを知ることができる。これらは人の移動、すなわち交通行動を知るデータとなるため、道路交通計画、交通対策の基礎資料として重要視されているが、既往の交通行動調査は、被験者が作業をしなければならず、「記入の煩わしさ」、「記入漏れによるデータ欠損」、「データの電子化コスト」などが問題点として挙げられる。これらは、記入事項が多く、紙ベースで調査を行っていることに起因している。

PT調査の問題点を解決するために、近年、携帯電



図-1 PT調査自動化に向けた検討項目

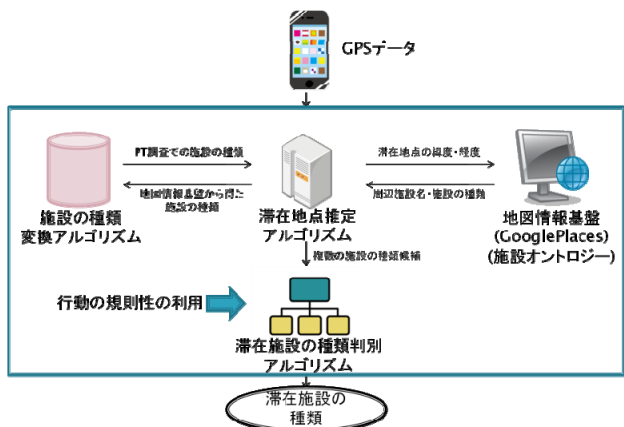


図-2 滞在施設の種類の推定手法の概要

話やスマートフォンなどのGPS端末およびWebを用いて、人の行動を記録する調査であるプローブパーソン調査（以下、「PP調査」と記す）など、PT調査の自動化が試みられている。しかしながら、PP調査では、経路情報はある程度把握できるものの、滞在施設、行動目的といった事項については、やはりユーザーの手動入力を前提としており、これらの自動化、効率化が求められる。

「PT調査を自動化する」ことを目標とした場合、検討項目は主に六つ存在すると考えられる（図-1参照）。その中で、GPSデータの取得²⁾、交通モード推定³⁾、滞在位置推定⁴⁾には、先行研究が存在する。そこで、本論文では、図-1において太枠で示した「滞在施設の種類の推定」に特化し、交通行動に関するGPSデータを利用した滞在施設の種類の推定手法の構築を試みた。また、長崎市圏において交通行動調査を実施（以下、「長崎市圏交通行動調査」と記す）し、収集データに構築した手法を適用し、その有用性について評価を試みた。

震災ビッグデータなどの大規模GPSデータの利活用は、道路事業においても注目されているところであり、当該データにより既存の交通実態データでは把握できない道路利用者の詳細な移動実態把握が可能になると期待されている。しかし現状では、多くの課題が存在し、様々な研究が行われており、今後広がりを見せていく必要がある。

本論文が対象とするPT調査の自動化は、大規模GPSデータの利活用の一例として位置づけられる。また、調査負荷やコストが膨大にかかる交通行動調査の効率化・精緻化に繋がることが期待される。

2. 滞在施設の種類の推定手法

図-2に本研究で構築した手法の概要を示す。本手法では、滞在地点推定アルゴリズム、地図情報基盤、施設の種類の推定アルゴリズム、滞在施設の種類の推定

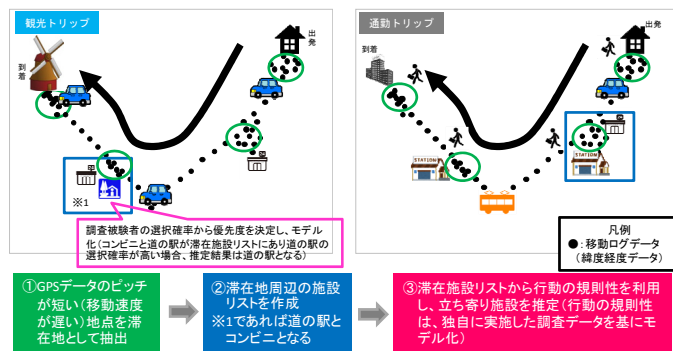


図-3 滞在施設の種類の推定手法の模式図

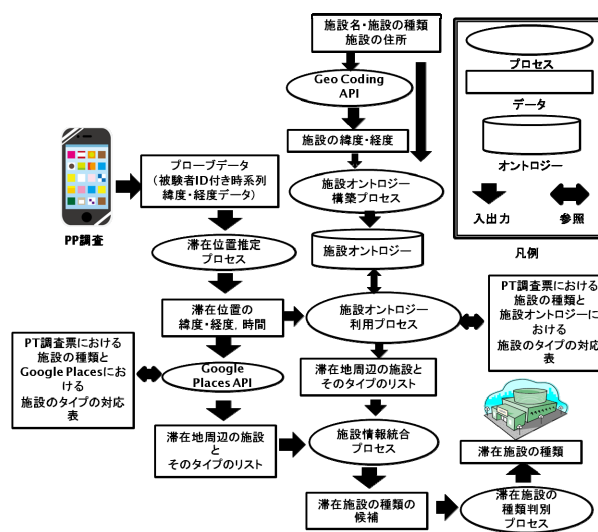


図-4 本手法の推定フロー

アルゴリズムの4つの部分が互いに情報をやりとりすることにより、GPSデータから滞在施設の種類の推定している。

また、本手法の模式図を図-3、推定フローを図-4に示す。長崎市圏交通行動調査から得られたGPSデータから、次節で後述する滞在位置推定プロセスにてGPSデータの速度に注目して閾値を設定し、滞在位置を推定する。その後、推定された滞在位置の緯度・経度をGoogle Places API（以下、「Google Places」と記す）と施設オントロジーに入力し、その周辺の施設情報を取得する（滞在地周辺の施設リストを作成）。Google PlacesはGoogleが無償で提供するWeb APIである。その際、表-1の対応表を用いてGoogle Places等から得られた施設の種類の推定をPT調査で用いられている施設の種類の推定に変換する。

このようにして得られた複数の周辺の施設情報を施設情報統合プロセスで統合し、施設の種類のみを抽出された複数の施設の種類の推定から1つに絞りこみを行う作業を滞在施設の種類の推定プロセス（長崎市圏交通行動調査結果を基に行動の規則性をモデル化）にて行い、滞在施設の種類の推定するというを行っている。次頁で、各プロセスの詳細について述べる。

表-1 GooglePlacesとPT調査の対応表の一部

GooglePlacesでの種類	PT調査での種類
accounting	事務所・会社
airport	その他
amusement_park	その他の商業施設
aquarium	その他の商業施設
art_gallery	公共施設
atm	その他
bakery	その他の商業施設
bank	銀行
bar	飲食店
beauty_salon	その他

```

FOR カウンタ = 1 to GPSデータの個数 DO
  IF そのデータの時速 <= 3 km/h THEN
    データ記憶用配列に緯度・経度と時刻を記憶
  ELSE
    IF データ記憶用配列の長さ >= 2 THEN
      IF データ記憶用配列の最後の時刻 - データ記憶用配列の最初の時刻 >= 5分 THEN
        滞在地点[i].滞在開始時間 = データ記憶用配列の最初の時刻
        滞在地点[i].平均緯度・経度 = 平均緯度・経度を計算
        滞在地点[i].距離の標準偏差 = 平均緯度・経度と各点を用いて距離の標準偏差を計算
        滞在地点[i].滞在終了時間 = データ記憶用配列の最後の時刻
        i++
      ENDIF
    ENDIF
    データ記憶用配列の初期化
  ENDIF
ENDFOR

```

図-5 滞在位置推定プロセスのアルゴリズム

(1) 滞在位置推定プロセス

本プロセスは、GPSデータを入力し、滞在したと推定される地点の緯度・経度の平均値と標準偏差、滞在開始・終了時間を出力する。

滞在地点を推定するにあたり、本研究では、滞在を「時速3km以下の状態が5分以上続いた状態」と定義した。これは徒歩の平均時速が4km/h程度であると言われていたりコンビニへの立寄りなど比較的短時間の滞在もPT調査の対象となることなどを加味して決定した。この定義に基づき、GPSデータから滞在地点を抽出するアルゴリズムを作成した（図-5参照）。滞在位置推定プロセスのアルゴリズムを図-6にまとめる。

(2) Google Places API

滞在地周辺の施設情報の取得のため、本手法ではGoogle Places APIを適用した。本APIは、緯度・経度と検索すべき半径を与えることにより、その緯度・経度から入力した半径内にあるGoogle Placesに登録されている施設の名前、住所、施設のタイプなどの施設情報を出力するものである。本手法で利用している施設情報の大部分をこのGoogle Placesにより取得している。

本手法では緯度・経度は各滞在地点の緯度・経度、半径は以下のように算出した。半径の基本値として50mを設定し、基本値に滞在地点の緯度・経度を算出した際の各点から滞在地点の緯度・経度までの距離の標準偏差を加える。算出された半径と各滞在地点の緯度・経度を入力として、Google Placesから得られた周

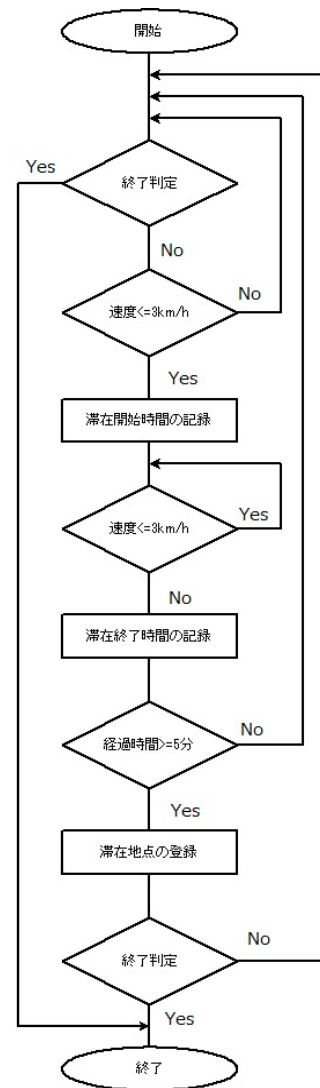


図-6 滞在位置推定プロセスのフローチャート

辺施設数が5つ以下であった場合には再度基本値を100mとして同様に半径を再計算し、検索をかけ直した。

ここで得られる施設の種類の種類はPT調査で用いられている施設の種類の異なるため変換が必要である。ここでは、2章4節で詳しく述べる長崎市圏交通行動調査というPT調査における施設の種類の種類の変換に用いる対応表の一部を表-1に示す。

Google Placesでは96個の施設のタイプが用意されており、これをそれぞれ14個あるPT調査における施設の種類の種類（表-2参照）の何れかに対応させた。

(3) 施設オントロジー利用プロセス

本節では、施設オントロジーとその要素技術であるRDF(Resource Description Framework)などを説明する。RDFとは、リソースを記述するための統一された枠組みであり、RDFでは、Subject（主語）、Predicate（述語）、Object（目的語）の三つ組（トリプルと呼ばれ

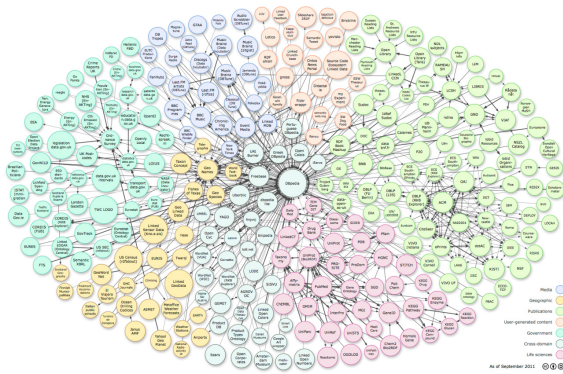


図7 LODクラウド

る)で物事の関係性を記述する。近年、各Webサイトで公開されているデータベース(政府、科学、写真、音楽など)をRDF化して連携するLinked Open Data(以下、「LOD」と記す)と呼ばれる活動が欧米を中心に急速に進んでいる。参考として、2011年9月19日時点でWeb上に公開されているデータをリンクさせ、一つの大きなクラウドとするプロジェクト「Linking Open Data」⁵⁾のLODクラウドの概念図を図-7に示す。

インターネットで公開されているもののみならず、様々なデータセットをRDF化して連携したデータのことをLinked Data(以下、「LD」と記す)と呼ぶ。LDを利用して今まで関連していなかったような敷居の高い領域や組織を横断することにより、コンテンツ間やサイト間でのデータ統合の実現が期待されている。LDを作成される際にオントロジーが用いられることも多い。オントロジーとは「概念の明示的仕様」であり、RDFにより言葉を階層構造と言葉間のネットワークで表現するデータモデルである⁶⁾。

本研究では、Google Placesでの施設の不足を補うために施設オントロジーを構築した。長崎市圏交通行動調査に対応するために構築した長崎施設オントロジーの一部を図-8に示す。長崎施設オントロジーでは、施設の種別を表す各クラスのインスタンスとしてその施設の種別に属する施設を定義した。各インスタンスは住所と緯度・経度を属性として持つ。なお、本研究では、iタウンページからスクレイピングにより、施設の種別とそれに属する施設、施設の住所と緯度・経度を抽出し、長崎施設オントロジーを自動構築した。また、Google Places同様、ここでの施設の種別とPT調査における施設の種別が異なるため、対応表を作成し、施設の種類の整合をとるために施設の種別の変換を行った。

施設オントロジーを利用して、Google Places同様、緯度・経度と2章2節で述べた方法により算出した半径を入力として、緯度・経度を中心として半径内に存在する施設の名称と施設の種別、住所を取得するSPA-

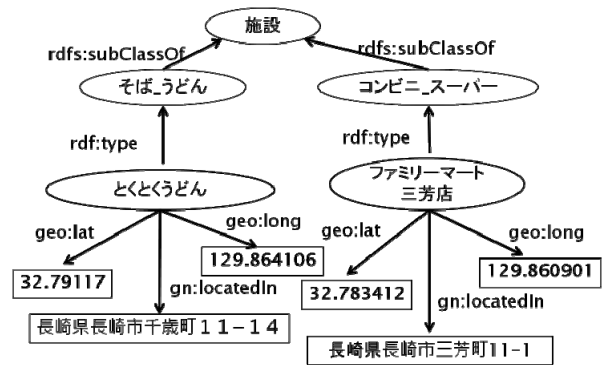


図-8 長崎施設オントロジーの一部

RQL クエリを作成し、実行した。

(4) 施設情報統合抽出プロセス

本プロセスは、各地図情報基盤から得られた周辺施設情報のうち施設の種類のみに注目し、複数の施設の種類の重複を除く機能を有する。

PT調査で使われる施設の種類には、住宅や自宅、勤務・通学先のように地図情報基盤から得られないものも存在する。住宅はGoogle Placesにはほぼ登録されておらず、住宅の住所を取得しオントロジー化するのは困難である。また、自宅や勤務・通学先は被験者によって異なるため、地図情報基盤から得られる共通的な情報のみではそれらの推定はできない。以上の理由により、施設の種類のうち自宅と勤務・通学先、住宅は本プロセスで追加を行う。

住宅については、周辺の施設数が少ない際に追加を行う。これは住宅街には地図情報基盤に情報がある商業施設などが比較的少ないことを利用している。また、自宅や勤務・通学先については長時間滞在に注目し、追加を行っている。具体的には、抽出された滞在地点を、距離の近いものでグループ化し、それらのグループの中で滞在時間が2時間以上である長期滞在をした累積回数を数える。長期滞在の累積回数が最も多いものを自宅、それ以外で多いものを勤務・通学先のグループというラベル付けを行う。各滞在地点がそれらのグループに属するときには対応する施設の種別を追加するというを行っている。

(5) 滞在施設の種別判別プロセス

本プロセスでは、施設情報統合プロセスでまとめられた複数の施設の種別を1つに絞り込む作業を行う。ここで絞り込む際にはトリップ(ある地点から目的地までの移動)の連続的集合データ(トリップチェーン)を分析し、得られた優先度を用いて絞り込みを行う。ここでは、長崎市圏交通行動調査での施設の種類の優先度を例として説明する。

表-2 長崎市圏交通行動調査における施設の種類の優先順位

優先順位	施設の種類の優先順位
1	自宅
2	勤務・通学先
3	住宅
4	デパート・モール
5	スーパー・コンビニ
6	飲食店
7	その他の商業施設
8	学校
9	病院
10	事務所・会社
11	公共施設
12	銀行
13	ホテル・旅館
14	その他

3章で後述する長崎市圏交通行動調査で利用された施設の種類の優先順位を表-2に示す。記載順は上に記載されるほど優先順位が高いことを表す。

自宅、勤務・通学先は2章4節で述べた規則を基に推定されているため、信頼度が高いと判断した。

住宅に関しては、他のものよりも優先度を低くしてしまうと住宅だとされることが著しく少なくなってしまうため、優先度を高く設定した。また、その他については特にデパート・モールからその他の商業施設までは施設の種類の候補に入る頻度が下にいくにつれて高くなる、つまりデパート・モールは希少性が高く、それ以降徐々に希少性が低くなる傾向があったこと、希少性の高い施設の種類の候補に入っている際にはそれが実際に滞在した施設の種類の候補であったことが多かったことから上記のように設定した。

この様な単純な優先度のみでは対応しきれない部分があったため、以下の2つの例外的なルールを作成した。

- 候補に住宅と学校の両者がある場合には、学校を住宅よりも優先とする
- 自宅または勤務・通学先の直後の滞在地点では、その他の商業施設をスーパー・コンビニよりも優先とする

1つめのルールは、優先度のみで施設の種類の絞り込みを行うと、施設の種類の候補が学校であるときでも住宅とされてしまうことが多いため設けたものである。学校はそれ自体の面積が広いので、学校内に滞在している際には検索半径内に地図情報基盤に登録されている施設が少ないことが多い。この場合、周辺施設が少ないために、施設の候補に住宅が加えられる。そして先ほど述べた優先度の大小を比較すると、住宅の優先度が高いため、施設の種類の候補が「住宅」とされてしまうのである。また、学校自体の優先度を高めてしまうと、学校付近の繁華街で買い物のために商業施設を利用する、または、食事のために飲食店を訪れるといった行

動まで施設の種類の候補が「学校」とされてしまうため、このような例外ルールを設けた。

2つめのルールは、比較的訪れる回数の多いその他の商業施設を正しく推定するため設けたものである。Google Placesでは、施設の種類の候補全96個のうち40個がその他の商業施設に対応づけられている等、その他の商業施設は含まれる施設の種類の幅が広い。そのため、滞在している回数も多いが、施設の候補に含まれる頻度も極めて高い。そのため、優先度を下手に高くしてしまうと、推定の失敗回数が増えてしまう。トリップチェーンの分析から、自宅や勤務・通学先を訪れた直後に利用することが多い傾向が判明したため、それを反映するためにこのようなルールを作成した。

以上の優先度とルールを用いて複数施設の種類の候補から絞りこむことで、施設の種類の推定を行っている。

施設情報統合プロセスで同一のグループであるとされた滞在が連続する場合、同一の滞在であると仮定し、そのデータを結合する。

3. 行動ログデータの収集と評価

(1) 長崎市圏交通行動調査の実施

本手法の評価を行うために、長崎市圏においてPP調査を行った。モニターとして、長崎市民10名を対象として10日間調査を行い、計100人・日のデータを取得した。ここで取得したデータはスマートフォンのGPSデータとトリップチェーンである。

GPSデータの取得に際し、本調査ではトランスフィールド社のプローブパーソンシステムを用いた。トリップチェーンは、Webダイアリを利用し収集した。Webダイアリとは、その日に滞在した施設と施設の種類の候補、交通手段、出発・到着時刻、目的などPT調査で取得できる情報をPC上で記入するシステムである。

このGPSデータから、100人・日分の滞在施設の種類の候補を推定し、トリップチェーンをもとに正誤判定し、推定の精度を検証した。また、検証の際に、携帯端末のバッテリー切れ等により、GPSデータが著しく欠損しているものは対象外とした。

以下に移動ログと推定結果（立ち寄り施設の推定結果）のサンプルを示す（図-9）

(2) 精度の検証と考察

精度の検証にあたって、正答率と再現率という指標を用いた。以下にその算出方法を示す。

- 正答率=正しく検出出来た滞在地数/提案手法が検出した滞在地数
- 再現率=正しく検出出来た滞在地数/評価の対象とした滞在地数

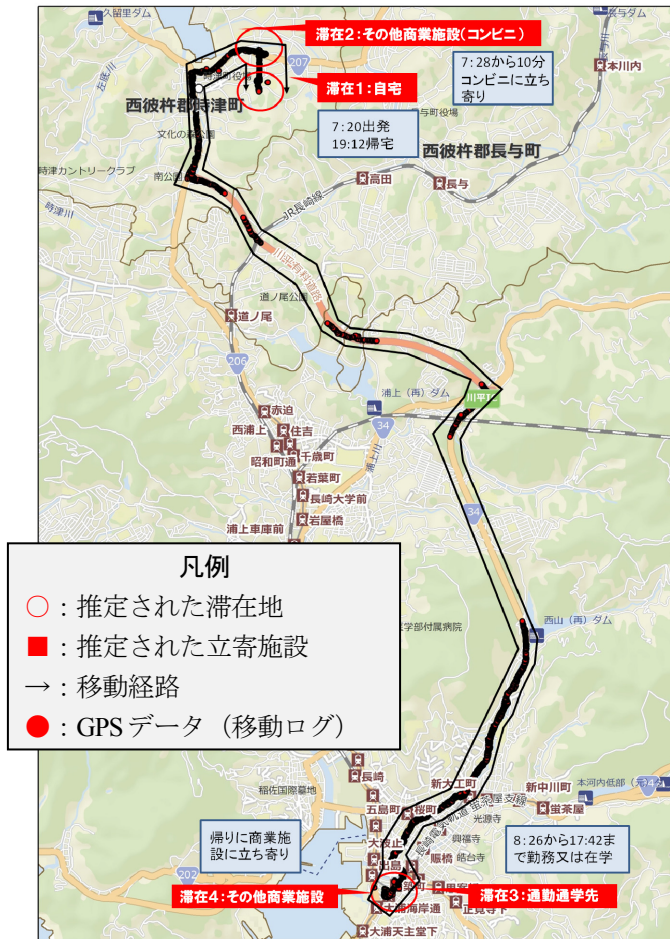


図-9 移動ログと推定結果のサンプル

正答率は、提案手法が推定した結果の中でどのくらい正解できているかを表す指標で、滞在でないものを提案手法が滞在と認識することによって数値が悪化し得る指標である。再現率は実際に滞在した施設のうちどれだけ正解できているかを表す指標で、実際の滞在施設のうちどれだけ網羅できているかを示す。滞在施設の種類の推定結果を表-3に示す。

検証の結果、今回の調査においては正答率が54.4%、再現率は68.5%という数値が得られた。また、各施設の種類の再現率と施設の種類の失敗状況を表-4および表-5にまとめる。なお表-4の失敗状況の項目は表-5と対応しており、数字が表-5における失敗コード、括弧内の数字がその失敗総数を示す。

表-4および表-5を見ると、施設の種類によって推定の精度が異なることが分かる。自宅や勤務・通学先では高い再現率であるが、その他の施設の種類の再現率は全体の再現率に及ばない。つまり、本手法においては、自宅と勤務・通学先に関しては高い精度で推定出来るが、他についてはまだ改善が必要である。また、失敗状況としては「候補に存在したが絞り込みで除外された」というものが最も多く、滞在施設の種類の判別プロセスにおいて改善の余地があることを示している。

表-3 滞在施設の種類の推定結果

	推定結果 (推定された施設数) 合計施設数=254	再現率
		正答率
対象の滞在地	174	68.5%
対象の滞在地以外	146	54.4%

表-4 各施設の種類の推定の失敗状況と内訳

コード番号	施設の種類の	総数	再現率	失敗状況
1	デパート・モール	20	65.0%	1(4),5(2),8(1)
2	スーパー・コンビニ	14	35.7%	1(1),2(6),3(1),5(1)
3	飲食店	8	12.5%	2(5),3(1),10(1)
4	その他の商業施設	27	40.7%	1(3),2(8),11(1),12(4)
5	住宅	6	66.7%	1(1),4(1)
6	学校	7	57.1%	2(2),9(1)
7	病院	4	0.0%	1(2),2(1),5(1)
8	事務所・会社	0		
9	公共施設	4	25.0%	2(2),5(1)
10	銀行	0		
11	自宅	100	88.0%	1(6),5(4),6(2)
12	勤務・通学先	60	78.3%	1(4),4(1),5(1),6(7)
13	ホテル・旅館	0		
14	その他	4	0.0%	2(3),3(1)

表-5 失敗コードと失敗状況の対応

失敗コード	失敗状況	件数
1	代表地点が遠い	21
2	候補に存在したが絞り込みで除外された	27
3	候補に存在したが勤務・通学先に含まれてしまった	3
4	住宅が認識されなかった	2
5	滞在が抽出されなかった	10
6	候補に存在したが自宅、勤務・通学先と認識されなかった	9
7	別の滞在がグループ化により合体されてしまった	0
8	内部の店はリストに存在するがその店自体が存在せず	1
9	該当施設敷地内に取れている敷地が広すぎて候補に入らず	1
10	候補に存在したが住宅とされてしまった	1
11	前後の滞在に統合されてしまった	1
12	候補に存在したが勤務・通学先とされてしまった	4

これらの結果から、本手法の優先度と2つのルールに基づく推定では、自宅や勤務通学先以外を目的地とする複雑な交通行動を推定しきれないので、今後トリップチェーンの分析を重ねてルールを増やす必要がある。また、現状推定に用いている緯度・経度と時間のみでは精度に限界があることも考えられるので、交通モードなど他の情報によって推定の精度が向上し得るか検証の余地がある。実際にPT調査の自動化を考えるために、今回用いたルールや優先度を他地域や他の施設の種類の応用した際の精度や工数を検討する必要がある。

4. おわりに

本研究では、交通行動調査のケーススタディとしてPT調査を取り上げ、PT調査の自動化に向けたプロセスの1つである滞在施設の種類の推定手法を構築した。

本手法は、PP調査で取得したGPSデータを入力とし、地図情報基盤を用いて滞在施設の種類の推定を行うものである。また、長崎市圏交通行動調査で取得したGPSデータを用いて検証を行ったところ、68.5%の精度で滞在施設の種類の推定することが出来た。

本手法の確立により走行経路やODとリンクした立

ち寄り施設の特定が可能となりPT調査の効率化のみならず,道路事業において必要となる道路利用者の詳細な移動実態の把握が可能になると考える.

今後の課題として,交通モード推定などの他の検討項目との連携や他地域や他の施設の種類への応用可能性の検証や精度向上による実務への適用性の向上が考えられるため,引き続き課題の解決に向けて検討を推進していく.

謝辞: 本研究を行うに際し, 文教大学阿部秀尚専任講師, 青山学院大学森田武史助教より多大なご助言等を得ました. ここに, あらためて感謝の意を表します.

参考文献

1) 石田東生, “危機にある大規模交通調査”, 交通工学, Vol.46, No.2, pp.1-2, 2011.

- 2) Koutaro Nohara, Keisuke Matsuda, Takuya Maruyama, “Development and Application of Smartphone-based Travel Survey Method”, 13th World Conference on Transport Research, 2013.
- 3) 西岡 康, 安東 直紀, 谷口 栄一, “スマートフォンを用いた交通モード判定モデルに関する研究”, 土木計画学研究・講演集, CD-ROM, 2011.
- 4) 遠山 緑生, 服部 隆志, 荻野 達也, “携帯電話の測位機能を用いた有意位置の学習”, 情報処理学会論文誌, Vol.46 No.12 p2915-2924, 2005.
- 5) Linking Open Data, <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData> [2014, April 24]
- 6) 溝口理一郎, 人工知能学会:オントロジー工学(知の科学), オーム社, 2005.

(2014. 4. 25 受付)