

図3 本手法のデータフロー

表1 Google Places 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
ENDIFOR
    
```

図4 潜在位置推定プロセスのアルゴリズム

のように得られた複数の周辺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、抽出された複数の施設の種類から1つに絞りこみを行う作業を滞在施設の種類の判別プロセスにて行い、滞在施設の種類を推定するというものを行っている。以下では、各プロセスの詳細について述べる。

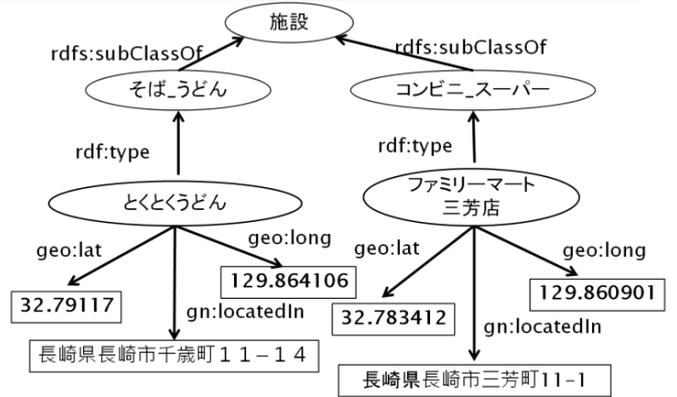


図5 長崎施設オントロジーの一部

2.1 潜在位置推定プロセス

本プロセスは、GPS データを入力とし、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ることや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4参照)。

本アルゴリズムでは、速度が3km/h以下のデータをトリガーとして、そこから次に速度が3km/h以上になるところを探し、それまでの経過時間を計算した。この時間が5分以上であれば滞在地点として出力した。この際出力する情報は緯度・経度とこの滞在中の緯度・経度データの標準偏差、滞在開始・終了時間とした。滞在所と推定された地点の緯度・経度に関してはその滞在中の緯度・経度データの平均をとることとした。

2.2 Google Places

Google Places 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、施設の種類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。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、検索をかけた。

ここで得られる施設の種類は PT 調査で用いられている施設の種類と異なるため、変換が必要である。ここでは、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個の施設の種類が用意されており、これをそれぞれ14個あるPT調査における施設の種類(表2参照)の何れかに対応させた。

2.3 施設オントロジー利用プロセス

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5に示す。長崎施設オントロジーでは、施設の種類を表す各クラスのインスタンスとしてそ

の施設の種類の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^{*4} からスクレイピングにより、施設の種類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、ここでの施設の種類の PT 調査における施設の種類の異なるため、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類の交換を行った。

施設オントロジーを利用して、Google Places 同様、緯度・経度と 2.2 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類の、住所を取得する SPARQL クエリを作成し、実行した。

2.4 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のうち、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類の追加するというを行っている。

2.5 滞在施設の種類の判別プロセス

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類の 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。

4 節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 2.4 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種類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。

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
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

- 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、施設の種類の種類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いため、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「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「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、比較的訪れる回数が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、施設の種類の種類全 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の施設の種類の候補から絞り込むことで、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

3.1 長崎市圏交通行動調査

本手法の評価を行うために、長崎市圏において PP 調査を行い、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、この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^{*5}を用いた。トリップチェーンは、Web ダイアリーを利用し収集した。Web ダイアリーとは、その日に滞在した施設と施設の種類の種類、交通手段、出発・到着時刻、目的な

*4 i タウンページ : <http://itp.ne.jp>

*5 プローブパーソンシステム : <http://www.transfield.co.jp/pp.html>

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。

3.2 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。

$$\begin{aligned} \text{正答率} &= \frac{\text{正しく検出出来た滞在地数}}{\text{提案手法が検出した滞在地数}} \\ \text{再現率} &= \frac{\text{正しく検出出来た滞在地数}}{\text{評価の対象とした滞在地数}} \end{aligned}$$

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を示す。滞在施設の種類の推定の結果を表 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、施設の種類の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・通学先におい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、滞在施設の種類の推定プロセスにおいて改善の余地があることを示す。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推定に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提案した。本手法は、PP 調査で取得した GPS データを入力し、地図情報基盤を用いて滞在した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。

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の推定への応用可能性の検証が挙げられる。

表 3 滞在施設の種類の推定の結果

	本手法によって検出されたか		再現率
	○	×	
対象の滞在地か	○	174	80
	×	146	N/A
正答率		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が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

参考文献

[石田 11] 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.

[Nohara 13] Koutaro Nohara, Keisuke Matsuda, Takuya Maruyama, “Development and Application of Smartphone-based Travel Survey Method”, 13th World Conference on Transport Research, 2013.

[西岡 11] 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.

[遠山 05] 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.