

ポイント型流動人口データを用いた徒歩移動者の推定方法に関する研究

A Study on The Method for Estimating Walking Data by Use of Point-Based GPS Data

○高木 尚哉^{*1}, 瀧澤 重志^{*1}
Naoya Takagi^{*1} and Atsushi Takizawa^{*1}

^{*1} 大阪市立大学大学院生活科学研究科

Graduate School of Human Life Science, Osaka City University.

キーワード：ポイント型流動人口データ; GIS; 徒歩推定; 二値分類; 地理的加重回帰モデル; 機械学習モデル

Keywords: Point-based GPS data; GIS; walking data estimation; binary classification; GWR; machine learning models.

1. はじめに

1.1. 人の移動に関するデータ

人の移動に関するデータとして、パーソントリップ調査（以降、PT 調査）に基づく統計的な精度を確保した静的なデータが都市計画や交通計画の分野で活用されている。国土交通省がまとめた近年の PT 調査についての現況¹⁾を以下に記す。PT 調査を行う対象範囲について、調査の効率性の視点から限られた地域のみを対象としてきたが範囲の拡大を行って地域のデータ利用ニーズに応える態勢がとられている。くわえて、標本率の設定により精度を担保する手段や調査項目の追加や意識調査など付帯調査の実施も進めてより詳細な分析が行えるようになってきている。しかしながら、PT 調査はサンプル調査であるため分担率の低い地方都市圏の公共交通利用者において誤差が大きくなる傾向にあることや、調査コストが多分に必要になること、調査の回答率が PT 調査の質に影響するなど課題もある。

また PT 調査は性質上、近距離の移動や短時間滞在の移動を把握することが困難である。一方で、道路断面の交通量を観測するスクリーンライン調査や交通系 IC カードを利用した駅の乗客数の調査、GPS を利用したプローブパーソン調査に加え、施設内の撮影データ、ビーコン等によるセンシングデータなど情報技術の進歩により新たなデータが取得できるようになっている。これら新しい様式のデータはパーソントリップデータにはない粒度での取得が可能となっており、ビッグデータとして様々な分野での活用が進められている。

1.2. GPS の利用による人の移動のデータの取得

その中で GPS データは当初、民生利用として航空機、船舶、測量、登山用など目印を持たない海上・山地において、座礁や遭難を免れる目的で現在地を把握することに主に活用されてきた。近年、個人でも GPS を搭載した端末の所有が広がり、GPS データの取得が容易にな

り、身近な生活シーンにおいても幅広く利用されるようになってきている。

このように精度良く、高頻度に位置情報が取得可能な GPS データを低コストで手に入れることができる一方で、都市部では精度の良いデータの取得に必要な衛星の補足数が高層建築の陰になることで不足するため、GPS データの位置誤差が 100m 以上となることが課題にあげられる。

また、GPS データは位置情報だけでなく、GPS を搭載した端末に付属するセンサーからの情報も合わせて取得することが可能であり、位置だけでなく移動速度や移動方向にも言及した分析が可能となっている。しかしながら、位置情報以外のセンシングデータや GPS データの提供元がオプションとして付加するユーザの属性情報と状態情報などのデータには多くの欠損が見られる仕様となっており、高度な分析が困難になっている。

1.3. 目的

本研究の最終的な目的は、PT 調査データに対し質の観点で課題があるものの大量にサンプルを取得することができる GPS データに着目し、ポイント型流動人口データと道路地図データベースを用いて人の分布を可視化し、人の分布とそこから取得できる他の情報を組み合わせ都市のウォーカビリティを評価することである。

筆者らは既報²⁾において、GPS データの測位誤差を考慮した人の分布を幾何的に可視化する方法を提案し、街路レベルでの人の存在確率を可視化することに一定の成果を見出した。しかし一方で、歩行者の移動と滞在とを正確に分類した分布を作成することが、移動に着目した都市のウォーカビリティの評価を進めるうえで課題となることを認識していた。この課題を解消するために、ポイント型流動人口データのオプションである推定移動手段の情報が利用できると考えたが、依然としてこのオプションは網羅性が低いことから、本報ではポイント型流

動人口データと道路地図データベースを用いて、徒歩で移動しているデータを推定する方法を提案し、街路レベルでのウォーカビリティ分析のデータ整備を行う。

2. ポイント型流動人口データの準備

2.1. ポイント型流動人口データの概要

本研究では、株式会社 Agoop（以下、Agoop）が提供するポイント型流動人口データ³⁾を使用する。ポイント型流動人口データは、Agoop の SDK を搭載したスマートフォン（Android と iOS）のアプリから収集した位置情報等データであり、位置情報や向き、速度などのセンサーデータから人の流れを点で把握することができる。今回利用するデータは、アプリからログを収集した直後の生データに対し、秘匿化・クレンジング・各種情報の付与が施されたものである。

表 1 に示すデ일리ID・年・月・日・緯度・経度・GPS 精度・移動速度・移動方向・100m メッシュ ID などの項目をもったデータが、テキスト形式で収録されている。デ일리ID はユーザ毎に付与される固有の ID であり、同日内において同一のデ일리ID をつなぐことで移動経路を把握することが可能になる。ログタイプカテゴリはログ取得時のアプリ状況を示し、その値に”background”がつくと、アプリを操作していない時に取得したログであることを示す。しかしこの項目は OS により取得タイミングが異なり、iOS ではアプリユーザが約 500m～1000m の大幅移動をした場合や滞在や出発をした場合にログを取得し、Android では一定の時間間隔でログを取得する。GPS 精度は GPS 測位の水平方向の精度を示すため正の緯度経度が取りうる範囲を円で表し、その半径をメートルで示す。この項目は、衛星の電波の遮断や反射が起こる場所（室内やビル街等）の測位では値が大きくなる。これは、真の観測地点が緯度経度から GPS 精度の範囲内において誤差があることを意味している。また、スマートフォンのセンサーから取得された移動速度やスマートフォンの磁力センサーで測定された移動方向のセンシングデータ、Agoop が独自に推定した交通手段を示す推定移動手段などが付加されている。

そして、ポイント型流動人口データは、Agoop が推計したユーザごとの推定居住エリア（100m メッシュ）内のデータが秘匿性の観点から除外されており、ログを取得するタイミングについても個々のスマートフォンの OS の仕様に依存するため断片的な情報となることが活用の前提となる。また、データには前述のとおり欠損している項目が存在しており、本報で着目する移動速度、移動方向、推定移動手段の有効値率（非 NULL 率）はそれぞれ順に 72.39%、53.34%、9.00%と集計されており、特に推定移動手段はほとんど欠損していることがわかっている。

2.2. 本報で使用するポイント型流動人口データ

本報で使用するポイント型流動人口データは、2019 年 1 月 7 日から 2019 年 8 月 18 日までの 21 日間に収集された大阪市のポイント型流動人口データを利用する。このデータから図 1 に示す大阪市地形図の道路区画ポリゴン上にマッピングされるデータを分析用として抽出し、その内訳を表 2 に示す。

表 1 ポイント型流動人口データのデータ項目

#	タイトル	csv項目	単位	有効値率
1	デ일리ID	dailyid	-	100.00%
2	永続ID	uuid	-	100.00%
3	年	year	年	100.00%
4	月	month	月	100.00%
5	日	day	日	100.00%
6	曜日	dayofweek	-	100.00%
7	時	hour	時	100.00%
8	分	minute	分	100.00%
9	緯度	latitude	度	100.00%
10	経度	longitude	度	100.00%
11	OS	os	-	100.00%
12	推定居住国名	home_countryname	-	100.00%
13	SIMのPLMN	plmn	-	99.82%
14	PLMN国名	plmn_countryname	-	99.82%
15	端末設定通貨	setting_currency	-	11.22%
16	端末設定言語	setting_language	-	100.00%
17	端末設定国名	setting_country	-	11.22%
18	ログタイプカテゴリ	logtype_category	-	11.38%
19	ログタイプサブカテゴリ	logtype_subcategory	-	11.30%
20	GPS精度	accuracy	m	100.00%
21	移動速度	speed	m/s	72.39%
22	推定移動速度フラグ	estimated_speed_flag	-	72.39%
23	移動方向	course	度	53.34%
24	推定移動方向フラグ	estimated_course_flag	-	53.34%
25	都道府県コード	prefcode	-	100.00%
26	市区町村コード	citycode	-	99.95%
27	100mメッシュID	mesh100mid	-	100.00%
28	推定居住地 都道府県コード	home_prefcode	-	99.95%
29	推定居住地 市区町村コード	home_citycode	-	99.95%
30	推定勤務地 都道府県コード	workplace_prefcode	-	99.82%
31	推定勤務地 市区町村コード	workplace_citycode	-	99.82%
32	推定移動手段	transportation_type	-	9.00%
33	性別	gender	-	49.69%

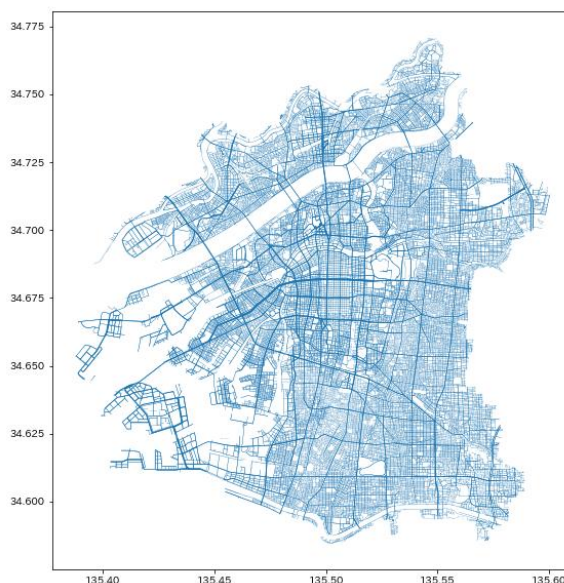


図 1 大阪市地形図の道路区画ポリゴン

3. ポイント型流動人口データによる徒歩移動者の推定方法

3.1. 徒歩移動者推定のデータセットの作成

本章ではポイント型流動人口データと道路地図データベースを用いて推定移動手段が未知のデータのうち徒歩移動者のデータを推定する方法を説明する。

まず表1で表されるポイント型流動人口データから推定移動手段が既知のものを対象に有効値のみを抽出したうえで特徴量エンジニアリングを行い、徒歩移動者を推定するためのデータセットを作成する。表3に推定に利用するデータの項目を示す。そして、このデータセットは推定移動手段が徒歩と徒歩以外のデータで件数が等しくなるように作成し543,870件となった。

表2 利用するポイント型流動人口データの内訳

#	日付	全データ件数	街路上のデータのうち移動手段が徒歩のもの	街路上のデータのうち移動手段が未知のもの	有効値率	
1	2019/01/07(月)	4,305,359	1,428,326	28,998	1,284,376	10.08%
2	2019/01/08(火)	4,298,515	1,417,896	30,716	1,264,247	10.84%
3	2019/01/17(木)	4,323,384	1,425,198	30,925	1,267,836	11.04%
4	2019/01/25(金)	4,756,121	1,588,545	33,924	1,419,189	10.66%
5	2019/01/26(土)	3,999,988	1,342,700	28,920	1,195,712	10.95%
6	2019/01/27(日)	3,686,261	1,198,324	24,885	1,070,422	10.67%
7	2019/01/30(水)	4,558,860	1,500,812	32,055	1,343,059	10.51%
8	2019/04/05(金)	4,710,580	1,608,217	40,949	1,386,935	13.76%
9	2019/04/15(月)	4,641,699	1,565,913	36,206	1,364,592	12.86%
10	2019/04/16(火)	4,583,714	1,550,576	35,668	1,353,695	12.70%
11	2019/04/17(水)	4,612,025	1,553,246	33,702	1,360,552	12.41%
12	2019/04/18(木)	4,703,127	1,594,453	34,510	1,396,856	12.39%
13	2019/04/20(土)	4,084,444	1,418,530	33,464	1,233,622	13.04%
14	2019/04/21(日)	3,817,764	1,294,110	28,204	1,131,402	12.57%
15	2019/07/29(月)	4,537,299	1,572,586	30,397	1,386,331	11.84%
16	2019/07/30(火)	4,527,648	1,578,719	30,587	1,390,122	11.95%
17	2019/07/31(水)	4,441,730	1,558,522	32,503	1,361,751	12.63%
18	2019/08/03(土)	3,838,054	1,354,393	29,331	1,176,792	13.11%
19	2019/08/04(日)	3,648,126	1,243,995	26,466	1,080,781	13.12%
20	2019/08/08(木)	4,595,886	1,610,142	31,780	1,413,324	12.22%
21	2019/08/09(金)	4,665,723	1,651,288	32,546	1,449,562	12.22%

表3 徒歩推定に利用するデータ項目

#	データ項目	型	備考
1	UNIX時間	連続値	年・月・日・時・分のデータから作成(秒は00)
2	曜日	カテゴリ値	-
3	緯度	連続値	-
4	経度	連続値	-
5	OS	カテゴリ値	One-Hot Encodingにより分解して学習させる
6	GPS精度	連続値	-
7	移動速度	連続値	-
8	移動方向	連続値	-
9	市区町村コード	カテゴリ値	One-Hot Encodingにより分解して学習させる
10	性別	カテゴリ値	One-Hot Encodingにより分解して学習させる
11	居住地が大区内	カテゴリ値	居住地が大区内の場合1
12	勤務地が大区内	カテゴリ値	勤務地が大区内の場合1
13	徒歩移動フラグ	カテゴリ値	推定移動手段が徒歩の場合1

3.2. 徒歩移動者の推定モデルの作成

本報では徒歩移動者の推定方法として、空間的異質性を考慮した地理的加重回帰（以下、GWR）と、一般に広く利用されている二値分類アルゴリズムを複数採用し、徒歩移動フラグを従属変数、それ以外の項目を特徴量として作成したモデルの精度を比較することで徒歩移動者の推定に最適な推定モデルを決定する。二値分類アルゴリズムには、非線形回帰分析の一つであるロジスティック

回帰・決定木をバギングベースで組み合わせたランダムフォレスト・決定木の勾配ブースティングのフレームワークであるLightGBMを採用する。

ここで、前節で作成したデータセットを学習データに80%、検証データに残りの20%と分割し、各モデルを一定の条件の下で作成するため同じデータを利用する。また、各モデルのハイパーパラメータはチューニングせずに作成する。

4. 徒歩移動者の推定モデルの検証

推定モデルの検証は、回帰モデルと分類モデルが混在するため混同行列で比較し、モデルの評価指標は参考とする。回帰モデルにおいては予測結果を標準化したのち0を閾値に二値化し結果を解釈する。

表4に検証データでの各モデルの推定精度を示す。ロジスティック回帰モデルの精度指標が最も悪くなっているが、ランダムフォレストとLightGBMがGWRよりも良い精度を示している。これは、ローカル空間においてポイント型流動人口データの振る舞いが多用であるためGWRでの推定が困難になっていると推察する。そして、二値分類モデルのROC曲線を示した図2を見ると、LightGBMのパフォーマンスが最も良いことが確認できる。

また、LightGBMとランダムフォレストの各指標がやや拮抗しているように見受けられる。特徴量の重要度を表5に示す。ここで、「os_1」や「市区町村コード_1」のように表現されている変数は表3に記載しているデータ項目をOne-Hot Encodingによりカテゴリの種類数に応じて作成されたダミー変数を意味し、機械学習において学習器が扱いやすいデータに変換したため発生している。例えば、元のデータセットにおいて、OSが”Android”の場合、os_1が1、os_2が0というように値が設定される。また、重要度の数値がモデル間で異なったスケールになっているが、これはアルゴリズムを実装したライブラリに依存している。特徴量の重要度において、ランダムフォレストは移動速度が特に高く、次いでGPS精度や、UNIX時間、緯度、100mメッシュID、経度が続くが移動速度と大きな差が発生している。これはランダムフォレストでは移動速度のみに依存したモデルになっていると考えられる。対してLightGBMにおいては傑出して高い特徴量が存在していないことから特定の特徴量に依存したモデルになっていないことに加え、緯度経度だけでなく100mメッシュIDも上位に位置していることから空間異質性を考慮した実質科学的に意味のあるモデルとして解釈できるのではないかと考える。LightGBMのモデルにおいて、ハイパーパラメータのチューニングを試みた結果を表4に、特徴量の重要度を表5に示す。チューニングをしていないLightGBMのモデ

ルと比較して、再現率を除く各指標の改善がみられる。適合率は約 2% 高くなっており、偽陽性の件数も少なくなっている。一方で特徴量の重要度を比較すると、

表 4 検証データでの各モデルの推定精度

#	指標	回帰モデル		二値分類モデル		
		地理的加重 回帰	ロジスティック 回帰 ※チューニングなし	ランダムフォ レスト ※チューニングなし	LightGBM ※チューニングなし	LightGBM ※チューニングあり
1	混同行列					
2	真陽性 (TP)	44,954	36,987	48,480	46,763	45,476
3	真陰性 (TN)	9,433	17,400	5,907	7,624	8,911
4	偽陽性 (FP)	26,825	35,665	28,879	25,561	22,826
5	偽陰性 (FN)	27,562	18,722	25,508	28,826	31,561
6	精度指標					
7	正解率	0.6666	0.5121	0.6801	0.6949	0.7082
8	適合率	0.6262	0.5090	0.6266	0.6465	0.6658
9	再現率	0.8265	0.6800	0.8913	0.8598	0.8361
10	F1値	0.7126	0.5822	0.7359	0.7381	0.7413
11	モデル評価指標					
12	R2	0.1477	-	-	-	-
13	RMSE	0.4615	-	-	-	-
14	Log Loss	-	0.6918	0.6155	0.5734	0.5599
15	AUC	-	0.5316	0.7255	0.7569	0.7744

表 5 ランダムフォレストと LightGBM の特徴量重要度の比較

#	ランダムフォレスト		LightGBM ※チューニングなし		LightGBM ※チューニングあり	
	特徴量	重要度	特徴量	重要度	特徴量	重要度
1	移動速度	0.808744	GPS精度	640	移動方向	19.830
2	GPS精度	0.054220	経度	623	経度	19.811
3	UNIX時間	0.037906	移動速度	537	UNIX時間	19.112
4	経度	0.021063	UNIX時間	536	緯度	18.041
5	100mメッシュID	0.020829	緯度	532	移動速度	13.166
6	緯度	0.014444	移動方向	347	GPS精度	13.144
7	勤務地が大阪府内	0.010072	100mメッシュID	282	100mメッシュID	9.709
8	居住地が大阪府内	0.008040	os_1	164	曜日	3.721
9	os_2	0.007602	曜日	73	os_1	1.544
10	os_1	0.006476	勤務地が大阪府内	51	居住地が大阪府内	1.173
11	移動方向	0.003645	居住地が大阪府内	46	勤務地が大阪府内	931
12	曜日	0.002484	市区町村コード_10	18	市区町村コード_1	363
13	市区町村コード_1	0.000621	市区町村コード_5	14	市区町村コード_6	360
14	市区町村コード_21	0.000440	市区町村コード_8	12	市区町村コード_3	357
15	市区町村コード_22	0.000390	市区町村コード_12	12	市区町村コード_10	304
16	市区町村コード_24	0.000336	市区町村コード_3	11	市区町村コード_7	302
17	市区町村コード_17	0.000335	市区町村コード_19	10	市区町村コード_4	297
18	市区町村コード_8	0.000298	市区町村コード_21	10	市区町村コード_8	295
19	市区町村コード_6	0.000238	市区町村コード_18	9	市区町村コード_19	282
20	市区町村コード_4	0.000176	市区町村コード_9	8	os_2	276
21	市区町村コード_15	0.000175	市区町村コード_7	7	市区町村コード_5	274
22	市区町村コード_11	0.000158	市区町村コード_2	6	市区町村コード_20	274
23	市区町村コード_5	0.000139	市区町村コード_14	6	市区町村コード_9	273
24	市区町村コード_14	0.000128	市区町村コード_24	6	市区町村コード_14	272
25	市区町村コード_9	0.000126	市区町村コード_6	5	市区町村コード_2	266
26	市区町村コード_2	0.000119	市区町村コード_13	5	市区町村コード_18	263
27	市区町村コード_20	0.000117	市区町村コード_17	5	市区町村コード_12	259
28	市区町村コード_23	0.000107	市区町村コード_22	5	市区町村コード_15	252
29	市区町村コード_10	0.000098	市区町村コード_23	5	市区町村コード_23	248
30	市区町村コード_19	0.000079	市区町村コード_16	4	市区町村コード_16	247
31	市区町村コード_3	0.000074	市区町村コード_1	3	市区町村コード_24	243
32	市区町村コード_12	0.000069	市区町村コード_15	3	市区町村コード_13	239
33	市区町村コード_18	0.000066	市区町村コード_4	2	市区町村コード_11	238
34	市区町村コード_16	0.000065	市区町村コード_20	2	市区町村コード_17	237
35	市区町村コード_7	0.000061	市区町村コード_11	1	市区町村コード_22	211
36	市区町村コード_13	0.000060	os_2	0	市区町村コード_21	206
37	性別_1	0.000000	性別_1	0	性別_1	0
38	性別_2	0.000000	性別_2	0	性別_2	0
39	性別_3	0.000000	性別_3	0	性別_3	0

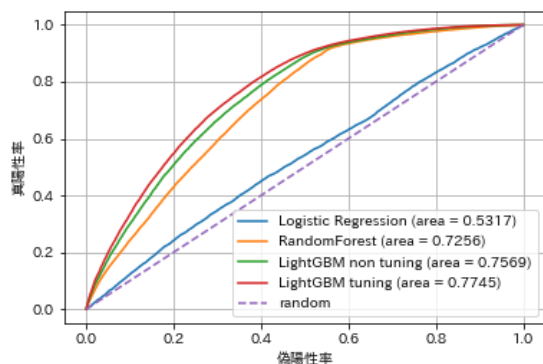


図 2 検証データでの二値分類モデルの ROC 曲線

チューニングなしのモデルでは GPS 精度が最も重要度が高いのに対し、チューニングありのモデルでは移動方向の重要度が高くなっている。同一地点でも移動方向に多様性があると考えられる徒歩移動の特性がより考慮された可能性がある。

5. まとめ

本報では、ポイント型流動人口データのうち、徒歩で移動しているデータの推定方法を報告した。空間的異質性の考慮が必要なデータにおいても、100m メッシュ ID など緯度経度だけでなく位置が判別できる変数を加えることで一定程度の推定が可能であることを確認できた。一方で、今回実施していない特徴量エンジニアリングや交差検証、ハイパーパラメータのチューニングを行い、モデルのあてはまりをよくするとともに、汎化性能を向上させる必要があると考える。

また、今回は精度比較と特徴量の重要度の観点から推定モデルの妥当性について行った考察を検証した。それに加え、モデルによる推定結果と徒歩を移動手段とするポイント型流動人口データとの地理的な関係についても検証を進めていく。

今後、徒歩を移動手段とするデータが集中する位置と既存のウォーカビリティ指標が高く出ている位置との地理的関係を比較することで、都市の街路レベルのウォーカビリティ分析を、これまででない高い空間解像度で行うための基盤データへと発展させていく予定である。

謝辞

本研究の一部は株式会社日建設計総合研究所、科学研究費補助金基盤研究(B)の補助を受けています。

【参考文献】

- 国土交通省「総合都市交通体系調査の事例集 2. パーソントリップ調査の企画・実施における事例」平成 30 年 6 月（最終閲覧日：2021 年 9 月 30 日）
https://www.mlit.go.jp/toshi/tosiko/toshi_tosiko_tk_000024.html
- 高木尚哉、瀧澤重志：ポイント型流動人口データを用いた歩行者の観測地点推定方法に関する研究、日本建築学会 情報システム技術委員会 第 43 回情報・システム・利用・技術シンポジウム論文集 報告 H16 2020 年 12 月
- 株式会社 Agoop「ポイント型流動人口データ仕様書 第 1.1 版」
- 一般社団法人 社会基盤情報流通推進協議会「【R01 年度】大阪市地形図（構造化データ_ESRI Shapefile）」G 空間情報センター（最終閲覧日：2021 年 9 月 30 日）
<https://www.geospatial.jp/ckan/dataset/r01-esri-shapefile>