携帯電話位置情報データを用いた滞留施設・滞留目的の推定手法

大村碧[*1]・大佛俊泰[*2]・岸本まき[*2]・荻野光司[*3]

Estimating Stay Facilities and Purposes from Mobile Phone Location Data

Aoi OMURA[*1], Toshihiro OSARAGI[*2], Maki KISHIMOTO[*2], Koji OGINO[*3]

Abstract: Mobile phone location data can be collected over wide areas at high temporal resolution. However, such data primarily capture movement trajectories, while stay locations and activity purposes remain unknown. In this study, we first extract stay locations for each person from mobile phone location data and link each location to building-use information. We then develop a stay-purpose model using person-trip survey data with known purpose labels. The model achieves high performance for Work and School: recall and precision are above 80%. Key predictive factors include arrival time, age, and building usage. We apply the model to mobile phone location data to infer stay purposes. Our results show that even within the same building-use (e.g., hospitals), the stay locations with shorter dwell-times are typically classified as Clinic visits, whereas those with longer dwell-times are classified as Work. In other words, the model enables assigning different purposes to stay at the same type of facility.

Keywords: 携帯電話位置情報データ（mobile phone location data），滞留場所(stay locations)，滞留目的（stay purpose），パーソントリップ調査（person trip survey），ロジスティック回帰（logistic regression）

1. はじめに

近年，情報通信技術の普及と発展により，携帯電話位置情報ポイント型データ（移動軌跡データ）を用いて人々の移動や滞留を推定し，都市計画分野に応用する研究が多く行われている．例えば，小沢・田中（2023）は，移動軌跡データから滞留情報（滞留場所と滞留目的を含む情報）を抽出し，地域単位での集計量について分析している．しかし，ミクロな空間単位で個人の移動軌跡を分析する研究は少ない．

移動軌跡データは，広域におけるユーザの移動を把握できるものの，移動軌跡の取得頻度に粗密があり，位置座標の精度も不安定である．また，取得されたデータは単なる移動軌跡であり，ユーザの行動を解釈する上で必要な滞留施設や滞留目的などに関する情報を直接得ることは難しい．そこで，本研究では，移動軌跡データに滞留施設と滞留目的を付与する手法を提案する．

2. 滞留施設の推定方法

2.1 滞留地点の推定方法

本節では，移動軌跡データから滞留地点を抽出する方法を提案する．移動軌跡データは，ユーザに移動が見られない場合，再び動き出すまでデータ取得が一時停止されるという特徴を有する．そのため，実際には長時間滞留していても，取得点が少ない場合が多い．この特性により，ポイント密度に基づいてポイント群を抽出する手法である DBSCAN などを適用することは難しい（藤田ほか, 2012）．そのため，本稿では，以下のいずれかの条件 1，条件 2 に当てはまるものを滞留と判定する．

条件 1：複数のポイントが同一地域に一定時間以上集まっている状態を滞留と判定する．そのために，まず前時点のポイントからの乖離距離が 100m 以下であれば同一ポイント群に属するとみなす．そして，そのポイント群の合計滞留時間が閾値 τ を超える場合，滞留と判定し（図 1 (a) ①），そのポイント群を「滞留クラスタ」とする．

条件 2：データ取得が一時停止されるという特性により，条件 1 に該当しない単独のポイントであっても長時

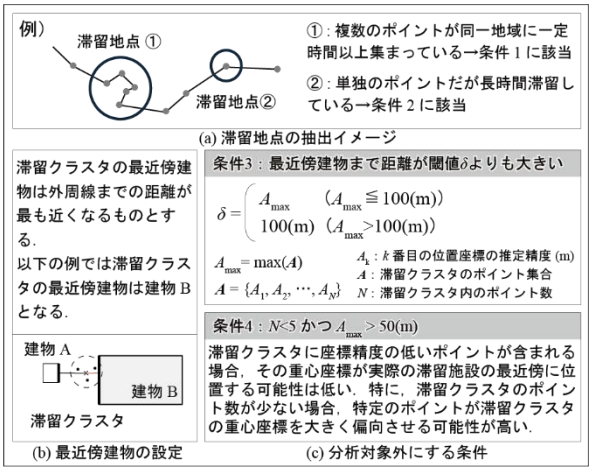


図 1 滞留場所の抽出条件

*1 学生会員 東京科学大学環境・社会理工学院（Institute of Science Tokyo）
〒152-8550 東京都目黒区大岡山 2-12-1 E-mail: omura.a.2b84@m.isct.ac.jp
*2 正会員 東京科学大学環境・社会理工学院（Institute of Science Tokyo）
*3 学生会員 東京科学大学環境・社会理工学院（Institute of Science Tokyo）

間滞留している場合がある．そこで，当該ポイントから次のポイントまでの取得間隔が閾値 τ を超え，かつユーザの移動速度 s_k が一定以下である場合，そのポイントを滞留と判定する（図 1 (a) ②）．

2.2 滞留施設の推定

周辺の建物との位置関係に基づき滞留施設を同定する方法を図 1(b)(c)に示してある．まず，滞留クラスタの重心座標の最近接の位置にある建物を特定し，滞留施設とする（図 1(b)）．この時，位置座標の精度を考慮し，条件 3，条件 4 のいずれかに該当する場合には，滞留施設の特定は困難（分析対象外）であるとみなす（図 1(c)）．

3. 滞留目的の推定方法

3.1 滞留目的推定モデルの位置付け

移動軌跡データから一定の条件を満たす滞留クラスタを抽出することで，滞留施設や到着時間・滞在時間などを推定することができる．しかし，それらの情報のみでは，滞留施設における滞留目的を推定することはできない．そこで，個人単位の滞留施設・滞留時間・滞留目的が記録された都市交通調査を用いて滞留目的推定モデルを構築し，これに移動軌跡データから得られる情報を外挿することで．滞留施設における滞留目的を付与することとする．

3.2 滞留目的推定モデルの構築

滞留目的は，来訪時刻や滞在時間といった時間的要因，年齢や性別などの個人属性，および滞在施設の用途といった場所的要因が複合的に関係していると考えられる．これら複数の要因を同時に考慮して滞留目的を確率的に推定するために，本稿では多クラスロジスティック回帰モデルを採用する．

多クラスロジスティック回帰は，一般に目的変数が複数のカテゴリをとる場合に適用され，説明変数との関係は線形和で表現される(図 2)．単純なモデル構造のため，各説明変数が滞留目的の判別にどの程度寄与しているかを推定パラメータから解釈できる点に特徴がある．この他に，ランダムフォレストやニューラルネットワークなどの手法も存在するが，パラメータ数が多くモデルが複雑でブラックボックス化しやすい．本稿では滞留目的を同定することと同時に，その同定要因を明らかにするため，解釈可能性の高いロジスティック回帰が適していると判断した．

目的変数と説明変数の概要を図 3 に示してある．モデル構築にあたり，滞留目的は，都市交通調査で採用される代表的な 7 区分に設定した．なお，施設内における滞留目的の把握を目的とするため，屋外での滞留は分析対象外とした．説明変数には，滞留時間割合や到着時刻分布といった連続的特徴量に加え，建物用途を表すダミー変数を用いた．多重共線性を回避するため，建物用途の一部を基準カテゴリとして除外した．

複数のクラス $k = 1,2,...,K$（各クラスの説明変数の数は d 個）に対して説明変数 ${}_j r_k^t = ({}_1 r_k^t, {}_2 r_k^t, ..., {}_d r_k^t)$ を用いて各クラスの確率を計算する．各クラス k に対して係数（重み）$\beta_{kj} = (\beta_{k1}, \beta_{k2}, ..., \beta_{kd})$ を持ち，クラスのスコア z_k を求め，これらのスコアを右のSoftmax関数に通して確率に変換する．求めた確率を用いてランダムに滞留目的を推定する．

$$z_k = \sum_{j=1}^{d} \beta_{kj} \, {}_j r_k^t + b_k \quad \cdots式 (1) \qquad p_k(x_i) = \frac{\exp(z_k)}{\sum_{l=1}^{K} \exp(z_l)} \quad \cdots式 (2)$$

z_k : 滞留目的 k のスコア　　β_{kj} : 滞留目的 k における説明変数 j ごとの係数（重み）
${}_j r_k^t$: 属性区分(時刻、性別年齢、トリップ番号、建物用途) t、滞留目的kにおける説明変数 j の値
d : 説明変数の総数　　b_k : 誤差項　　$p_k(x_i)$: データ i の滞留目的が k である確率
K : 滞留目的の総数

滞留目的ごとのデータ数の偏りを考慮するため，滞留目的ごとの補正係数 w_k を求める．

$$w_k = \frac{1}{\sqrt{n_k}} \quad \cdots式 (3)$$

k :移動目的
w_k :移動目的 k の補正係数
n_k : 移動目的 k のデータ数

最尤推定を目的として以下の $\overline{L(\boldsymbol{\beta_k})}'$ を最大化するパラメータ $\boldsymbol{\beta_k}$ を求める．

$$\overline{L(\boldsymbol{\beta_k})} = \frac{\sum_{i=1}^{N} w_k \log p_k(x_i)}{N} \quad \cdots式 (4) \qquad \overline{L(\boldsymbol{\beta_k})}' = \overline{L(\boldsymbol{\beta_k})} - \lambda \sum_{k,j} |\beta_{kj}| \quad \cdots式 (5)$$

k :移動目的　　$p_k(x_i)$: データ i の滞留目的が k である確率
λ :正則化強度（$\lambda > 0$）　　$\boldsymbol{\beta_k}$: 滞留目的 k の重みベクトル
N : データ数　　β_{kj} : 滞留目的 k に対応する説明変数 j の係数（重み）
w_k :移動目的 k の補正係数

具体的には確率的勾配上昇法でパラメータベクトル $\boldsymbol{\beta_k}$ を更新し，正解のクラスがなるべく高い確率になるようにパラメータを調整する．

$\beta_{kj} > 0$ のとき　$\boldsymbol{\beta_k} \leftarrow \boldsymbol{\beta_k} + \eta \left(\sum_{i=1}^{N} w_k \left[\mathrm{I}\left(y_i = \mathrm{k}\right) - p_k(x_i) \right] x_{ij} - \lambda \right)$ …式 (6)

$\beta_{kj} = 0$ のとき　$\boldsymbol{\beta_k} \leftarrow \boldsymbol{\beta_k} + \eta \left(\sum_{i=1}^{N} w_k \left[\mathrm{I}\left(y_i = \mathrm{k}\right) - p_k(x_i) \right] x_{ij} \right)$ …式 (7)

$\boldsymbol{\beta_k}$:滞留目的 k の重みベクトル　　x_{ij} : データ i の説明変数 j の値
β_{kj} : 滞留目的kに対応する説明変数 j の係数(重み)　　y_i : データiの正解クラス
η : 学習率(一回の更新でどれだけ変化させるか)　　$\mathrm{I}(y_i=k)$: データ i の正解クラス k なら1、違えば0
i : 各データ(全部で N 個)　　$p_k(x_i)$: データiをクラス k と予測した確率
w_k :移動目的 k の補正係数　　λ :正則化強度（$\lambda > 0$）

図 2 滞留目的の推定方法

	概要
目的変数	勤務，通学，買物等へ（買物/食事/社交），通院，送迎，帰宅，その他
説明変数 (特徴量)※	滞留時間割合，移動時間割合，到着時間割合，性別/年齢割合，トリップ順序割合
説明変数 (ダミー変数(建物用途))	住宅，文教厚生施設(公共施設，病院等)，商業系施設，工業系施設，宿泊施設，事務所系

※説明変数(特徴量)の定義と算出方法の例

＜説明変数の定義＞

滞留時間割合(${}_j r_k^t$) : 移動目的 k ごとの総数に占める特定の滞留時間 t (30分毎)のデータ数の割合
移動時間割合(${}_j r_k^t$) : 〃　特定の移動時間 t (30分毎)のデータ数の割合
到着時間割合(${}_j r_k^t$) : 〃　特定の到着時間 t (30分毎)のデータ数の割合
性別/年齢割合(${}_j r_k^t$) : 〃　特定の年齢性別 t (10歳毎)のデータ数の割合
トリップ順序割合(${}_j r_k^t$) : 〃　特定のトリップ番号 t※※のデータ数の割合
※※1日のうち1番最初のトリップ（ある地点からある地点への移動）を1、その次のトリップを2,...,1日の最後のトリップを n とする

＜算出方法例＞

滞留目的 k ,滞在時間 t におけるデータ数を ${}_t N_k^t$ とする．
滞留目的ごとの割合 ${}_t r_k^t$ は，右の式で求める．

$${}_t r_k^t = \frac{{}_t N_k^t}{\sum_t {}_t N_k^t}$$

他の説明変数も同様の方法で滞留目的ごとの割合を算出する．

図 3 目的変数と説明変数の概要

4. 滞留施設・目的の推定結果および分析事例

4.1 使用データの概要

石川県白山市および小松市を分析対象地域に設定し，対象地域内の居住者の滞留行動を分析した（表 1）．移動軌跡データは，株式会社 Agoop 社が提供するポイント移

動軌跡データを使用した．居住地域コードが石川県白山市および小松市であり，分析対象地域内で移動が観測されたユーザを分析対象とした．滞留施設の推定時に用いる建物データは，石川県都市計画基礎調査（2021 年度時点）を用いた（表 2）．建物データの建物用途は，滞留目的推定モデルの目的変数（7 区分）に合わせて再分類した（表 3）．

滞留目的推定モデルの構築で用いる都市交通調査データは，2018 年度の東京都市圏パーソントリップ調査データ（PT データ）を用いた．本稿では，分析対象地域である石川県白山市および小松市と地域特性が類似する地域の PT データのみを使用した．具体的には，①人口密度が 80 人/㎢以上 500 人/㎢以下，かつ，②昼夜間人口比率が 85%以上 115%以下である 89 市区町村のデータを分析対象とした（データ数 32,098 件）．

4.2 滞留施設推定手法の推定結果

本稿では滞留時間の閾値 τ は 10 分に設定し，滞留施設を推定した（図 4）．10 代男性と 40 代女性の同定結果を図 5 に示してある．両者とも病院を訪れた後に商業施設へ立ち寄る様子が確認できる．ただし，移動軌跡のみでは病院や商業施設における滞留目的が通院か勤務か，あるいは買い物かといった詳細までは知ることができない．

4.3 滞留目的推定モデルの推定結果

図 6 にパラメータ推定の際に用いたデータ数，設定した条件を示す．なお，パラメータ推定に用いる学習率 η および正則化強度 λ は，グリッドサーチに基づき決定した（η=0.00005，λ=0.0001）．

滞留目的推定モデルのパラメータの推定結果を図 7(a)(b)に示す．「勤務」や「通学」では，性別/年齢割合の影響が大きい．特に，「通学」は，概ね学齢期（20 歳以下）に集中する活動であることが強く反映されている．また，「買物等」は商業施設，「帰宅」は住宅に関する値が大きく，建物用途が目的推定に強く寄与していることが確認できる．学習・検証データの平均対数尤度の推移を図 7(c)に示してある．

表 1 対象地域の概要

	石川県白山市/小松市
人口	計215,000人
人口密度（人/km²）	146.25/286.26
昼夜間人口比率	97.95/101.12
都市区分	地方中核都市圏周辺都市（中心都市40万人以上）
GPSデータ固有ID数	37,157人
使用データ取得日	2023年 8月5,6日,11月18,19日,12月23,24,30,31日, 2024年 1月1日,2月4,6～8,10日,5月3~6日,8月 17,18,20~22,24,25,28,31日,9月1,3～5日

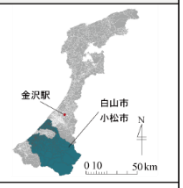
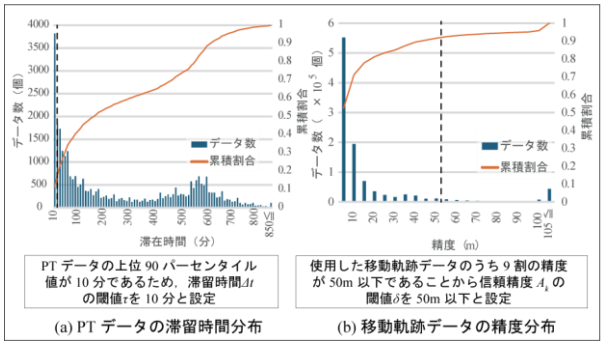


表 2 石川県都市計画基礎調査の概要

調査頻度	数年~10年ごとの行政単位での調査（自治体により異なる）
調査項目	建物ID，所在地，建物用途，構造，延床面積など
調査地域	日本全国（本稿では石川県19市町村）
調査方法	主に現地踏査（目視調査）
実施	国土交通省，地方公共団体

表 3 建物用途の読み替え

建物利用現況調査	用途読み替え
商業施設，商業系用途複合施設，娯楽施設，遊戯施設	商業系施設
住宅，共同住宅，店舗等併用住宅，店舗等併用共同住宅，作業所併用住宅	住宅
業務施設	事務所系
官公庁施設，文教厚生施設	文教厚生施設
運輸倉庫施設，重工業施設，軽工業施設，サービス工業施設，家内工業施設，農林漁業用施設，危険物貯蔵・処理施設	工業系施設
宿泊施設	宿泊施設
その他，不明	その他



(a) PT データの滞留時間分布　(b) 移動軌跡データの精度分布

図 4 滞留場所の抽出条件設定



(a) 10 代男性（平日）

(b) 40 代女性（平日）

※移動軌跡データではプライバシー保護のため，居住地付近の位置情報は秘匿される．そのため，ユーザの居住地は滞留施設として抽出されない．

図 5 滞留施設の抽出例

(a) 使用データ数

データセット名	データ数
学習データ	22,468
検証データ	6,419
テストデータ	3,211
合計	32,098

(b) 学習終了条件とグリッドサーチの導入

学習終了条件：一つ前の学習から平均対数尤度の変化が 0.0001 以下になった場合に学習を終了

※学習率と正則化強度の決定にはグリッドサーチを導入
候補値
(0.01,0.005,0.001,0.0005,0.0001,0.00005,0.00001,0.000005)
から全ての組み合わせを試して学習が安定して収束し，最も平均対数尤度が大きくなる組み合わせを採用．

図 6 使用データ数と学習終了条件



実際の目的＼予測結果	勤務	通学	買物等	通院	送迎	帰宅	その他
勤務	928	6	40	2	5	7	81
通学	10	263	0	0	3	0	17
買物等	3	0	536	7	4	6	14
通院	11	1	0	46	4	0	41
送迎	5	0	3	4	141	5	20
帰宅	1	0	0	0	2	253	8
その他	107	15	100	69	60	51	332

図 7 滞留目的推定モデルの推定結果

テストデータに対する滞留目的の推定精度（再現率・適合率）を図 7(d)に，混合行列を図 7(e)に示す．「通院」と「その他」は誤分類されやすい傾向が確認された．これは「通院」が他の滞留目的と比較してデータ数が少ないことや，「その他」には複数の滞留目的が含まれるため，行動の特徴が捉えにくいことが原因と考えられる．一方，「勤務」や「通学」は再現率と適合率が共に 80%を超えており，推定精度は高く良好である．

4.4 移動軌跡データでの分析事例

4.2 節で抽出した滞留施設に対して滞留目的推定モデルを適用し，滞留目的を推定した．10 代男性の例（図 4(a)）では，病院での滞在時間が短いため「通院」，商業施設での滞在時間が長いため「勤務」と同定された．すなわち，この男性は患者であり，商業施設の従業員と推察される．一方，40 代女性の例（図 4(b)）では病院での滞在時間が長いことから「勤務」，商業施設での滞在時間が短いため「買い物」であると同定された．すなわち，この女性は病院スタッフであり，商業施設の訪問客であると推察される．以上のように，同一建物用途における滞留であっても，滞留時間などの情報から異なる滞留目的が同定されており，一定の解釈が可能な結果を与えている．

5. おわりに

本稿では，移動軌跡データを用いて個人単位で滞留施設・滞留目的を推定する方法を提案した．具体的には，まず，移動軌跡データから滞留地点を抽出し，滞留施設を推定した．次に，パーソントリップ調査データを用いて，ロジスティック回帰による滞留目的推定モデルを構築した．このモデルを石川県白山市・小松市の移動軌跡データに適用することで，移動軌跡データから滞留施設・滞留目的の推定が可能であることを確認した．移動軌跡データのノイズ処理や滞留目的「その他」における推定精度の向上，利用圏分析などへの提案手法の適用は今後の課題である．

参考文献

小沢 啓次郎・田中 貴宏 (2023) 地方都市都心部の利便施設特性指標に基づくエリア区分と滞留の分布傾向に関する研究.「日本都市計画学会都市計画論文集」, **58**(3) , 1600 - 1607.

藤田 将成・伊藤 浩二・小林 稔・南 弘征・水田 正弘 (2012) GPS 移動履歴とスポット情報閲覧履歴の統合解析に基づいたスポット推薦手法の提案.「データ分析の理論と応用」, **2**(1), 53-68.