

GPSデータのクラスタリングによる 日常生活における場所の重要度の分析

平田 瑠¹ 原 直¹ 阿部 匡伸¹

概要: 本報告では, 詳細な見守りが可能な行動範囲の推定のために, 日常生活における場所の重要度を滞在地と経由地の観点に分けて検討し, ユーザが日常生活をする上で重要な場所を調査した. 日常生活の生活パターンを反映した場所の特徴量として, ユーザの行動パターンのクラスタリングを用いた共起クラスタ数の期待値を提案する. 1日あたりの滞在時間, 1日あたりの訪問回数, 共起クラスタ数の期待値で, 9名, 234日分のGPSデータを用いて, 全体の重要な場所の総数に対する割合を算出した. その結果, 滞在地については, 1日あたりの滞在時間が上位50%の場所の中の重要な滞在地の数において, 全体の重要な滞在地の数に対する割合が83.9%, 経由地については, 共起クラスタ数の期待値が上位50%の場所中の重要な経由地の数において, 全体の重要な経由地の総数に対する割合が76.7%と, それぞれ他の2つの特徴量よりも高い数値を示した. また, 3種類の特徴量を用いて実験協力者9名の重要な場所と重要でない場所の分類を行い, ROC曲線のAUCで評価したところ, 実験協力者9名の平均では, 重要な滞在地の分類において, 1日あたりの滞在時間に基づく分類が $AUC = 0.802$, 重要な経由地の分類では, 共起クラスタ数の期待値に基づく分類が $AUC = 0.734$ と, 3つの特徴量を用いた中で最も良い分類性能を示した.

Analysis of significance of places in daily life by clustering GPS data

RUI HIRATA¹ SUNAO HARA¹ MASANOBU ABE¹

1. はじめに

近年, わが国では高齢化が進み, それに伴い認知症患者数は増加傾向にある. 特に, 認知症による徘徊は大きな社会問題であり, 平成30年の行方不明者のうち, 19.2%が認知症が原因とされている[1]. また, 高齢者だけでなく, 9歳以下の子供の行方不明者数も増加傾向にある. そのため, 高齢者や子供を対象とした見守りに注目が集まっている.

高齢者や子供を見守る側にとって, 見守り対象者の外出に付き添う, 常に現在位置を監視する, といった見守りの方式は負担が大きい. そこで, 見守る側の負担を軽減するため, 位置情報システムを利用した自動で見守りを行うシステムが提供されている[2][3][4]. これらの見守りシステムでは, 見守り対象者の現在位置を取得し, 見守り対象者が事前に設定されたあるエリアに行くと, 見守る側に通知するが, このエリアを見守り対象者の日常的な行動に基づいて定めることができれば, 誤報や見逃しの少ない詳細な

見守りが実現できると考えられる.

本研究では, ユーザの日常生活において, 特に重要な場所に注目した, 詳細な見守りが可能な行動範囲の推定方式の確立を目指す. 日常生活における場所の観点として, 「滞在地」に加えて「経由地」という観点到に注目する. 「滞在地」は目的地として訪れる場所で, 「経由地」はそこを経由する場所である. よく利用するバス停や頻繁に通る交差点など, 日頃からそこを経由して別の目的地に行く場所であれば, ユーザの行動範囲に大きな影響を与えると考えられる. 一方で, このような経由地は, 個人の行動範囲への影響が小さい場所が, 滞在地よりも多く存在すると考えられる. さらに, 同一の経由地でも個人ごとに重要度が大きく異なり, 同一の滞在地における個人ごとの重要度の違いよりも大きいと考えられる. したがって, 経由地の中でも特に重要な経由地を精度よく見つけ出すことが, 個人の行動範囲の推定へ活用するために必要となる.

ユーザのGPSデータから訪れた場所を抽出し, ユーザの日常行動を分析する研究において, Matsuoら[5]による

¹ 岡山大学大学院ヘルスシステム統合科学研究科

ユーザの行動範囲の推定や、Ye ら [6] によるユーザ個人の生活パターンを発見する方式を提案されている。しかしながら、両者とも場所を「滞在地」の観点に注目してユーザの日常行動を分析しており、「経由地」の観点での分析は行われていない。

本報告では、「滞在地」と「経由地」の2つの観点から、ユーザの日常生活における重要な場所を調査した。ユーザの日常生活における生活パターンに注目した特徴量として、ユーザの行動パターンをクラスタリングにより分析して得られた共起クラスタ数の期待値を提案する。また、1日あたりの訪問回数と1日あたりの滞在時間、共起クラスタ数の期待値の3つの特徴量について、それらの値が上位50%の場所に含まれる重要な場所の割合を調査した。また、3つの特徴量を用いて、重要な場所と重要でない場所の分類実験を行った。

2. 提案方式

本研究では、日常生活における「場所」を「滞在地」と「経由地」の観点に分けて、重要度を検討する。「滞在地」は目的地として訪れる場所で、「経由地」はそこを経由する場所である。例えば、重要な経由地の例として、「信号交差点」「駐車場・駐輪場」「駅」などが考えられる。これらは、一般的な「滞在地」のように長時間滞在することは少ないが、ユーザの行動範囲を推定するにあたって、重要な場所であると考えられる。

本報告では、ユーザが訪れた場所について、上記の2種類の属性に関するユーザの日常生活における場所の重要度を調査し、重要な場所を抽出するために有効な場所の特徴量を検討する。場所の特徴量として、特に日常生活における生活パターンに注目した特徴量を提案する。日常生活といっても、学校や仕事に行く日もあれば、友達と遊ぶ日もあるように、複数のパターンが存在すると考え、ユーザの行動パターンのクラスタリングを用いた共起クラスタ数の期待値を提案する。

2.1 GPS データからの「場所」の抽出

本研究では、ユーザの行動を「場所から場所への移動の系列」ととらえ、GPS データの1点1点から得られる緯度経度座標と時刻情報から、ユーザの「場所」から「場所」への移動系列を抽出する。

本節では、本研究において重要度を検討する「場所」を、ユーザのGPS データから抽出し、「場所」から「場所」への移動系列を作成する方式について説明する。

本研究で注目したい経由地は、滞在時間が短いことが多いため、DBSCAN [7] [8] [9] のような密度推定のみに基づく分析方式よりは、SMoT (Stops and Moves of Trajectories) [10] やその派生方式 [11] [12] [13] のように、GPS データの遷移を捉えた分析方式の方が望ましいと考えられる。

2.1.1 時系列処理のための前処理

まず、ユーザのGPS データには測定時のデータ脱落があるため、すべて1秒間隔のデータ点となるようにリサンプリングする。各GPS データ点の緯度、経度を実測点の緯度、経度からそれぞれ補間することで、リサンプリングを実現している。なお、補間関数には、区分的3次エルミート内挿多項式 (Piecewise Cubic Hermite Interpolating Polynomial; PCHIP[14]) を使用している。

次に、1秒間隔のデータ点群から、各データ x_t における、速度 v_t (m/s)、加速度 a_t (m/s²)、回転角 ω_t (rad) を、以下の式により計算する。

$$v_t = \sqrt{dx_t^2 + dy_t^2} \quad (1)$$

$$a_t = \sqrt{(dx_{t+1} - dx_t)^2 + (dy_{t+1} - dy_t)^2} \quad (2)$$

$$\omega_t = \arctan(dx_{t+1}/dy_{t+1}) - \arctan(dx_t/dy_t) \quad (3)$$

ここで、 dx_t と dy_t は、データ点 x_t と x_{t-1} の、経度方向の変位と緯度方向の変位をメートル単位で表した値である。

2.1.2 時系列データから状態データへの変換

まず、各データ点 x_t には、滞在状態の候補を推定するためのスコア STOPSCORE_t を、式 (4) により、付与する。

$$\text{STOPSCORE}_t = \frac{1}{1 - \exp(10(v_t - 0.75))} \quad (4)$$

この関数は、0.2 m/s 以下でほぼ0となり、1.2 m/s 以上でほぼ1となるように設計したシグモイド曲線である。

次に、GPS データの計測誤差による計算のばらつきを抑制するため、 STOPSCORE_t に、孤立点除去のための3点のメディアンフィルタと、平滑化のための21点の移動平均フィルタを適用する。

そして、 STOPSCORE_t に基づいて、各データ点の滞在状態の候補 STATE を付与する。各データ点の STOPSCORE_t に対するしきい値判定をおこない、しきい値を超えた点を滞在状態候補、しきい値未満の点を移動状態候補とみなす。なお、本稿では、しきい値は0.5とした。

最後に、すべてのデータ点を走査し、移動状態候補が連続した点や滞在状態候補が連続した点を結合して、状態データ \tilde{S}_n を作成する。

2.1.3 状態データの補正

状態データ \tilde{S}_n には、以下の情報を付与する。これらの情報は、各状態データに含まれる時系列データ点群から計算することができる。

- STATE: 移動状態、あるいは、滞在状態
- CENTROID: 状態の代表地点の緯度と経度。状態内の全点の重心。
- TIME: 開始時刻と終了時刻
- DURATION: 状態の継続時間
- DIRECTDISTANCE: 状態の開始地点と終了地点の直線

距離

- CURVEDISTANCE: 状態内の各点を移動した場合の総移動距離

ただし、STATE は、前項の操作で推定された状態の候補である。実際には、GPS データの誤差のため、本来の行動とは異なる状態遷移を行っている場合がある。そこで、連続する状態を統合して補正した、状態データ S_m を構築する。

まず、以下の 5 つのフィルタに基づいて、各状態データのうち、誤った状態候補になっていると考えられる状態データを修正する。

- (1) CURVEDISTANCE が 1 m 以下の場合、直前の状態データと同じ STATE に変更する
- (2) STATE が移動状態の状態データについて、CURVEDISTANCE が 10 m 未満、かつ、DIRECTDISTANCE が CURVEDISTANCE の 1/10 以下の場合、滞在状態に変更する
- (3) STATE が移動状態の状態データについて、その前後の状態データがいずれも滞在状態であれば、その状態データの STATE を滞在状態に変更する。ただし、当該状態データの DURATION が 30 sec 以下、かつ、前後の状態データの DURATION の合計より少ない場合に限る
- (4) STATE が滞在状態の状態データについて、その前後の状態データがいずれも移動状態であれば、その状態データの STATE を移動状態に変更する。ただし、当該状態データの DURATION が 30 sec 以下、かつ、前後の状態データの DURATION の合計より少ない場合に限る
- (5) STATE が移動状態の状態データについて、その前後の状態データがいずれも滞在状態で、かつ、それらの CENTROID の座標の直線距離が 10 m 以内であれば、滞在状態に変更する。

最後に、同一の STATE が連続している状態データ $S_{n'-1}$ と $S_{n'}$ を結合する。このとき、各状態データの座標点や時間の情報は、統合後の状態データに合致するように再計算する。以上により、移動状態のデータ S_m と滞在状態のデータ S_s が交互に出現する、遷移系列 S_{sm} を得る。

2.1.4 GeoHex による量子化と「場所」の作成

前項の処理で得られた移動状態と滞在状態の遷移系列 S_{sm} から滞在状態のデータ S_s のみを抽出して、滞在状態のみの遷移系列 S_s を作り、各滞在状態データの CENTROID の緯度経度を GeoHex [15] により量子化する。これによって、各滞在状態データが CENTROID の情報としてただ 1 つの GeoHex を持つ、滞在状態のデータ $\langle S_s, H_h \rangle$ と、それらの遷移系列 $\langle S, H \rangle$ を得る。

次に、すべての滞在状態のデータ $\langle S_i, H_j \rangle$ の CENTROID H_j を GeoHex の集合 H として得る。そして、 H の各要

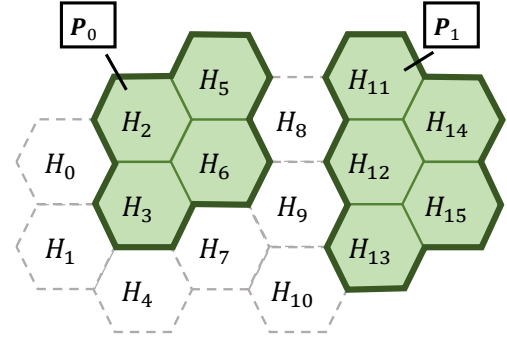


図 1 GeoHex 群から「場所」への変換

素 H_j について、 H の要素の中から隣接する GeoHex を検索し、図 1 のように、隣接したひとかたまりを「場所」として抽出する。このようにして、GPS データからユーザの訪れた「場所」の集合 $P = \{P_0, P_1, \dots\}$ を得る。このとき、各「場所」 P_j は、その「場所」を形成する GeoHex H_j の集合を持つ。図 1 の例であれば、

$$P_0 = \{H_2, H_3, H_5, H_6\}$$

$$P_1 = \{H_{11}, H_{12}, H_{13}, H_{14}, H_{15}\}$$

となる。なお、本稿では、量子化に用いる GeoHex のレベルを 11 (最長の対角線の長さが約 12 m) とした。

2.2 「場所」から「場所」への移動系列の作成

前項で得られた滞在状態の遷移系列 $\langle S, H \rangle$ と、「場所」の集合 P を用いて、「場所」から「場所」への移動系列 $\langle S, P \rangle$ を作成する。

各滞在状態データ $\langle S_s, H_h \rangle$ の H_j が、どの「場所」 P_k の要素なのかを調べ、 $\langle S_s, H_h \rangle$ を「場所」 P_k と対応付ける。例えば、

$$\langle S_1, H_2 \rangle \rightarrow \langle S_2, H_{11} \rangle \rightarrow \langle S_3, H_3 \rangle \rightarrow \langle S_4, H_6 \rangle$$

$$P_0 = \{H_2, H_3, H_5, H_6\}$$

$$P_1 = \{H_{11}, H_{12}, H_{13}, H_{14}, H_{15}\}$$

のような遷移系列と「場所」の集合があるとすると、 $\langle S_1, H_2 \rangle$ は P_0 、 $\langle S_2, H_{11} \rangle$ は P_1 、 $\langle S_3, H_3 \rangle$ は P_0 、 $\langle S_4, H_6 \rangle$ は P_0 に対応付けられ、

$$\langle S_1, P_0 \rangle \rightarrow \langle S_2, P_1 \rangle \rightarrow \langle S_3, P_0 \rangle \rightarrow \langle S_4, P_0 \rangle$$

のような「場所」から「場所」への移動系列が得られる。

2.3 「場所」の特徴量

本報告では、「場所」の特徴量として、1 日あたりの滞在時間、1 日あたりの訪問回数と、複数の生活パターンで共通して訪れる場所に注目した特徴量の 3 つを考える。

ユーザの日常生活に複数の生活パターンがあると仮定して、クラスタリングによってユーザの行動を分割し、複数

のクラスタで共通して訪れている場所に注目した特徴量を提案する．特徴量は，2.2 節の方式で得られた「場所」から「場所」への移動系列 $\langle S, P \rangle$ と，2.1.3 節で各状態データに与えられた情報のうち，状態の継続時間の DURATION を用いて算出する．

2.3.1 1 日あたりの滞在時間と訪問回数

2.2 節の方式で得られた「場所」から「場所」への移動系列 $\langle S, P \rangle$ を用いて，「場所 P_k 」ごとに 1 日あたりの滞在時間 D_k と訪問回数 C_k を算出する．

まず，「場所」から「場所」への移動系列 $\langle S, P \rangle$ を 1 日ごとに区切り， n 個に区切られた系列 $\langle S, P \rangle_d$ を得る．その後， $\langle S, P \rangle_d$ の各データ $\langle S_i, P_j \rangle_d$ について，連続して同じ P_j が出現していた場合は，その連続する 2 つのデータを結合している．

「場所」 P_k の 1 日あたりの滞在時間 D_k と訪問回数 C_k を以下の式で算出する．

$$D_{P_k} = \sum_{\langle S_i, P_j \rangle, j=k} \text{DURATION}(\langle S_i, P_j \rangle) \quad (5)$$

$$V_{P_k} = \sum_{\langle S_i, P_j \rangle, j=k} 1(\langle S_i, P_j \rangle) \quad (6)$$

ただし， 1 は 1 を返す関数である．

上記の手順を，例を用いて説明する．2 つの「場所」 $\{P_0, P_1\}$ があるとして，以下のような，1 日ごとの「場所」から「場所」への移動系列 $\langle S, P \rangle_d$ が得られたとする．

$$\langle S, P \rangle_1 : \langle S_1, P_0 \rangle \rightarrow \langle S_2, P_1 \rangle \rightarrow \langle S_3, P_0 \rangle$$

$$\langle S, P \rangle_2 : \langle S_4, P_1 \rangle \rightarrow \langle S_5, P_0 \rangle \rightarrow \langle S_6, P_1 \rangle$$

$$\langle S, P \rangle_3 : \langle S_7, P_1 \rangle \rightarrow \langle S_8, P_0 \rangle \rightarrow \langle S_9, P_1 \rangle$$

このとき，「場所」 P_0 と P_1 の滞在時間 D_{P_0} ， D_{P_1} は， $d_i = \text{DURATION}(\langle S_i, P_j \rangle)$ とすると，

$$D_{P_0} = d_1 + d_3 + d_5 + d_8$$

$$D_{P_1} = d_2 + d_4 + d_6 + d_7 + d_9$$

であり，「場所」 P_0 と P_1 の訪問回数 V_{P_0} ， V_{P_1} は $V_{P_0} = 4$ ， $V_{P_1} = 5$ となる．

2.3.2 生活パターンに注目した特徴量

ユーザの行動をクラスタリングし，複数のクラスタで共通して訪れている場所に注目した特徴量を抽出する．抽出には，節の方式で得られる 1 日ごとに分割された「場所」から「場所」への移動系列 $\langle S, P \rangle_d$ を用いる．

ある 1 日の行動は，Bag-of-places として表現する．これは，自然言語処理等の分野でしばしば用いられるベクトル空間モデル Bag-of-words と同等の概念である．すなわち，1 日の行動を，全「場所」の訪問回数のベクトルで表現する，ということである．場所のベクトル w は以下の式で表される．

$$w = (p_0, p_1, \dots, p_N)$$

ここで， N はユーザのデータ収集期間全体にわたる「場所」の異なり数である．

次に，1 日ごとに作成されたベクトル w の集合を，k-means 法により k 個のクラスタに分割する．クラスタ分割によって，1 つ 1 つのクラスタがユーザの生活パターンを表し，似た行動をしている日が同じクラスタに分類されることが考えられる．

次に，クラスタ C_i の Bag-of-places を計算する．これは，式 (7) のように，クラスタ C_i に属するベクトル w の和 $W^{(C_i)}$ を求めることで計算できる．

$$W^{(C_i)} = \sum_{w \in C_i} w \quad (7)$$

こうして，クラスタごとに各「場所」の訪問回数を得る．

その後，複数のクラスタで共通して訪れている「場所」は，複数の生活パターンで共通して訪れている「場所」だと考え，複数のクラスタで共通して訪れている「場所」に注目する．具体的には，「場所」ごとに，その「場所」の訪問回数が 1 回以上となっているクラスタの数を記録する．これを「場所」の共起クラスタ数と呼ぶ．

例えば，「場所」が 5 個あり $P = \{P_0, P_1, P_2, P_3, P_4\}$ ， $k = 5$ でクラスタリングしたとき，以下のようなクラスタごとの Bag-of-places を得られたとする．

$$W^{(C_0)} = (5, 2, 1, 0, 0)$$

$$W^{(C_1)} = (0, 4, 1, 3, 0)$$

$$W^{(C_2)} = (4, 0, 0, 0, 1)$$

$$W^{(C_3)} = (0, 1, 0, 0, 2)$$

$$W^{(C_4)} = (6, 2, 1, 0, 1)$$

この例における各「場所」 P_i の共起クラスタ数 c_i は $c_0 = 3$ ， $c_1 = 4$ ， $c_2 = 3$ ， $c_3 = 1$ ， $c_4 = 2$ となる．

ここで，k-means 法は初期値に依存するため，本報告では n 回のクラスタ分割を繰り返し，その結果の統計量に基づいて，特徴量を抽出する． n 回のクラスタ分割の繰り返しの都度，各「場所」 P_i の共起クラスタ数 c_i を計算し， $c_i = k$ となった回数 $n(c_i = k)$ を集計する．

最後に，各「場所」 P_i ごとに，共起クラスタ数の期待値 $E(P_i)$ を式 (8) で算出し，これを生活パターンに注目した場所の特徴量とする．

$$E(P_i) = \frac{1}{n} \sum_{k=1}^5 k \cdot n(c_i = k) \quad (8)$$

本報告では，k-means 法のクラスタ数を $k = 5$ とし， $n = 1000$ 回のクラスタ分割により特徴量抽出している．

3. 「場所」の重要度付与実験

3.1 実験の概要

ユーザにとっての場所の重要度をもつデータを収集した．

表 1 場所の重要度とその説明 (8 段階)

重要度	説明
7	生活圏としてとても重要な場所である
6	生活圏として重要な場所である
5	生活圏としてやや重要な場所である
4	生活圏に含まれる場所である
3	生活圏に含まれる場所であるがあまり重要ではない
2	生活圏に含まれる場所であるが重要ではない
1	生活圏ではない
×	滞在地/経由地ではない

表 2 実験協力者ごとの重要度が付与された場所の数

userA	userB	userC	userD	userE
120	100	49	77	143
userF	userG	userH	userI	合計
80	65	45	145	824

実験協力者は 9 名であり、以下の手順で抽出された各々が訪れた場所に対して情報を付けるよう依頼した。

(1) 2.1 節の方式により 2019 年 5 月 1 日から 12 月 20 日までの計 234 日間の GPS ログから訪れた場所を抽出

(2) 訪問回数が 2 回以上の場所を抽出

それぞれの場所に付与する情報は以下の通りである。

- 滞在地としての重要度
- 経由地としての重要度
- 場所の名前
- 場所の用途 (13 種類)

なお、滞在地としての重要度と経由地としての重要度は、表 1 のような基準を示して 8 段階で付けてもらった。また、場所の用途は、滞在地を想定した「居住地」、「勉強」、「バイト」、「普段の食事 (外食)」、「ショッピング」、「娯楽・趣味」、「ライフサービス」、「美容・ヘルスケア」の 8 種類と、経由地を想定した「信号待ち」、「駐車・駐輪」、「交通機関」の 3 種類、さらに、上記のどれにも属さない場所には「どれもでない」、その場所に行った記憶がなく、どんな場所かわからない場合は、「不明」の合計 13 種類のタグを重複を許して付けてもらった。

3.2 実験結果

重要度が付与された場所のうち、「不明」タグが付けられた場所を除くと、9 名で合計 824 のデータが得られた。表 2 に実験協力者ごとに得られた場所データの数を示す。

表 3 に、場所の重要度ごとの場所の数について、実験協力者 9 名分を集計したものを示す。一方の重要度が 7 で、もう一方の重要度が 0 となるような、滞在地としてのみ重要な場所や経由地としてのみ重要な場所があるだけでなく、また、滞在地としての重要度と経由地としての重要度が共に 7 である場所のように、滞在地としても経由地としても重要な場所があることもわかった。

表 3 重要度ごとの場所の数

滞在地\経由地	0	1	2	3	4	5	6	7	合計
0	9	48	44	34	21	30	14	27	227
1	89	86	65	13	5	3	4	3	268
2	25	9	8	9	7	5	8	7	78
3	26	13	2	5	3	4	1	5	59
4	20	2	1	9	5	5	5	10	57
5	24	8	3	6	2	1	1	9	54
6	10	12	3	1	1	1	2	3	33
7	18	2	1	10	6	1	1	9	48
合計	221	180	127	87	50	50	36	73	824

表 4 重要度の変換表

元の重要度	変換後の重要度
×	0
1, 2	1
3, 4	2
5, 6, 7	3

表 5 重要度 (4 段階) ごとの場所の数

滞在地\経由地	0	1	2	3	合計
0	9	92	55	71	227
1	114	168	34	30	346
2	46	18	22	30	116
3	52	29	26	28	135
合計	221	307	137	159	824

3.3 重要な場所の出現傾向の分析

本節では、重要度が付与された場所のデータを用いて、重要な場所の出現傾向を分析する。まず、分析のため、場所に付与された 8 段階の重要度を、表 4 により変換して 4 段階にした。4 段階の重要度ごとの場所の数について、実験協力者 9 名分を集計した表を表 5 に示す。

以降、滞在地としての重要度が 3 である場所を「重要な滞在地」、経由地としての重要度が 3 である場所を「重要な経由地」と呼ぶ。重要な滞在地は 135 個 (16.38%)、重要な経由地は 159 個 (19.30%) となった。また、重要度 (3, 0) の滞在地としてのみ重要な場所が 52 個 (6.31%)、重要度 (0, 3) の経由地としてのみ重要な場所が 71 個 (8.62%) ある。このことから、目的地として訪れる重要な場所だけでなく、経由するだけの重要な場所が存在することがわかった。

さらに、28 個 (3.40 %) ある重要度 (3, 3) の滞在地としても経由地としても重要な場所のように、全体の半数弱 (約 45 %) は滞在地、経由地の両方としての重要度を持つ場所であるため、1 つの場所に対して滞在地と経由地の両方の観点で重要度を考える必要があることが示唆された。

3.3.1 場所の用途に基づく重要な場所の出現傾向の分析

ここでは、場所の用途と場所の重要度の関係について分析する。場所の用途ごとに、重要な滞在地と経由地の数を表 6 に示す。

重要な滞在地に注目すると、まず、「普段の食事 (外食)」、「ショッピング」タグが付けられた場所は、重要な滞在地の数が最も多く、場所の総数も比較的多い。「バイト」タグが

表 6 場所の用途別の重要な滞在・地と経路・地の数

	居住地	勉強	バイト	普段の食事 (外食)	ショッピング	娯楽 趣味
場所の総数	46	63	19	104	117	131
重要な滞在・地の数	26	23	12	42	44	21
重要な経路・地の数	7	10	1	26	21	15
滞在・地と経路・地の 両方として重要な場所の数	4	2	1	15	15	7
	ライフサービス	美容・ヘルスケア	信号待ち	駐車・駐輪	交通機関	
場所の総数	40	12	248	71	66	
重要な滞在・地の数	10	4	1	16	7	
重要な経路・地の数	9	0	69	11	23	
滞在・地と経路・地の 両方として重要な場所の数	6	0	0	1	5	

付けられた場所は、場所の総数は少ないものの、それに対する重要な滞在・地の数の割合が大きい。このことから、これら 3 つのタグが付けられる場所は、日常生活において重要な滞在・地である可能性が高いと考えられる。

逆に、「信号待ち」タグが付けられた場所は、重要な滞在・地である割合が極めて低く、重要な滞在・地を抽出する際には、「信号待ち」タグが付けられた場所をできるだけ抽出しないようにする必要があると考えられる。

次に、重要な経路・地に注目する。「普段の食事 (外食)」, 「信号待ち」, 「交通機関」タグが付いた場所は、総数に対する重要な経路・地の数の割合が高いため、これらの場所は、日常生活における重要な経路・地である可能性が高いと考えられる。特に、「信号待ち」タグが付いた場所は、場所の総数も重要な経路・地の数も最大であるため、重要な経路・地を考えるにあたって特に注目すべきであると考えられる。

最後に、滞在・地と経路・地の両方として重要な場所に注目すると、「居住地」の他に、「普段の食事 (外食)」, 「ショッピング」, 「交通機関」タグが付けられた場所に多いことがわかる。岡山駅など都市部の大型の駅では、交通機関の利用の他に食事や買い物ができる店が多くあるため、このような場所は滞在・地と経路・地の両方として重要な場所になりやすいと考えられる。

3.3.2 場所の特徴量を用いた重要な場所の出現傾向の分析

1 日あたりの滞在時間、1 日あたりの訪問回数、共起クラスタ数の期待値の 3 つの特徴量に関して、重要な場所の出現傾向を調べた。まず、実験協力者ごとに、3 章の実験で得られた場所に各特徴量の値が大きい順に順位をつけ、それぞれの上位 50 % の場所を抽出した。次に、抽出された上位 50 % の場所の中に、重要な滞在・地と重要な経路・地が何個含まれているかを調べた。そして、重要な滞在・地の総数と重要な経路・地の総数に対する割合を算出した。

表 7 に、各特徴量の上位 50 % の場所に含まれる重要な滞在・地の数と全体の総数に対する割合を示す。全実験協力者の平均では、滞在時間に注目して抽出した場合が重要な滞在・地の割合が最も高く、滞在時間の上位 50 % の場所を見れば重要な滞在・地を 80 % 以上抽出できることがわかった。このことから、1 日あたりの滞在時間は、重要な滞在・地の特徴を最もよく捉えられていると考えられる。

表 8 に各特徴量の上位 50 % の場所に含まれる重要な経

表 7 特徴量の上位 50 % に含まれる重要な滞在・地の数と全体の総数に対する割合

	総数	特徴量		
		滞在時間	訪問回数	共起クラスタ数の期待値
user A	21	81.0	52.4	52.4
user B	9	88.9	55.6	44.4
user C	14	78.6	64.3	71.4
user D	11	90.9	90.9	81.8
user E	20	65.0	80.0	80.0
user F	18	94.4	77.8	88.9
user G	4	100.0	100.0	100.0
user H	17	70.6	52.9	58.8
user I	21	85.7	85.7	95.2
平均	-	83.9	73.3	74.9

表 8 特徴量の上位 50 % に含まれる重要な経路・地の数と全体の総数に対する割合

	総数	特徴量		
		滞在時間	訪問回数	共起クラスタ数の期待値
user A	46	41.3	65.2	69.6
user B	5	80.0	80.0	100.0
user C	6	50.0	66.7	50.0
user D	17	35.3	82.4	88.2
user E	9	33.3	66.7	66.7
user F	8	12.5	25.0	62.5
user G	1	100.0	100.0	100.0
user H	7	28.6	71.4	71.4
user I	61	59.0	70.5	82.0
平均	-	48.9	69.8	76.7

路・地の数と全体の総数に対する割合を示す。重要な経路・地の抽出では、共起クラスタ数の期待値に注目して抽出した場合が割合が最も高いことがわかった。以上から、共起クラスタ数の期待値は重要な経路・地の特徴を最もよく捉えられていると考えられる。

なお、user G は重要な滞在・地、経路・地の総数が他のユーザーと比べて少なく、どの特徴量に注目して抽出しても、網羅率が 100 % となった。

4. 場所の特徴量を用いた重要な場所の分類性能の比較

3 章の実験で得られた重要度が付与された場所のデータに対して、重要な場所とそうでない場所の分類を行った。分類には 2.3 節の 3 つの特徴量をそれぞれ 1 つずつ用いて行い、特徴量ごとの分類性能を評価し、比較した。

分類は、以下の手順で行った。まず、 N 個の重要度が付与された場所のデータを、1 つの特徴量に注目し、その特徴量が大きい順に順位を付ける。その後、上位 k 個の場所が重要な場所であると予測し、残りの $N - k$ 個が重要ではない場所であると予測し、以下の評価値を得る。

- True Positive (TP): 重要な場所を正しく重要な場所と予測できた数
- False Positive (FP): 重要でない場所を誤って重要な場所と予測した数
- False Negative (FN): 重要な場所を誤って重要でない

場所と予測した数

- True Negative (TN): 重要でない場所を正しく重要でない場所と予測できた数

これを k を 0 から N まで変化させて行い、計 $N + 1$ 個の TP, FP, FN, TN の組を得た。

4.1 分類性能の評価指標

特徴量ごとの分類性能の評価には ROC 曲線の AUC の値を用いる。ROC 曲線とは、FP rate (R_{FP}) を横軸に、TP rate (R_{TP}) を縦軸にとって描いた曲線である。これらは N 個のデータを分類して得られた評価値 TP, FP, FN, TN を用いて、以下のように求められる。

$$R_{FP} = \frac{FP}{FP + TN}$$

$$R_{TP} = \frac{TP}{TP + FN}$$

ROC 曲線の AUC の値は、ROC 曲線と横軸と直線 $R_{FP} = 1$ に囲まれた部分の面積であり、 $k = i$ で予測した場合の FP rate と TP rate をそれぞれ $R_{FP}^{(i)}$, $R_{TP}^{(i)}$ とすると、以下の式で計算される。

$$AUC = \sum_{i=1}^N R_{TP}^{(i)} (R_{FP}^{(i)} - R_{FP}^{(i-1)})$$

なお、AUC は 0 から 1 の値をとり、ランダムな分類であれば 0.5 の値をとる。

4.2 分類性能の比較

1 日あたりの滞在時間、1 日あたりの訪問回数、共起クラスタ数の期待値の 3 つの特徴量をそれぞれ用いて、実験協力者 9 名がそれぞれ訪れた場所に対して、重要な場所と重要でない場所の分類を行った。特徴量ごとの性能の比較には ROC 曲線の AUC の値を用いる。また、各特徴量の全ての実験協力者についての平均の性能指標として、横軸に FP rate、縦軸に各 FP rate に対する全実験協力者の TP rate の平均 (Average of TP rate) をとり、FP rate = 0.05 刻みで描いた曲線の AUC の値を用いる。user G は重要な滞在地、経由地の総数が他のユーザと比べて少ないため、平均の性能評価は user G を除く 8 名の実験協力者で行った。なお、本報告では、実験協力者が訪れた場所の中から、滞在地としての重要度と経由地としての重要度が共に 0 である場所を除いた場所に対して分類を行った。

4.2.1 重要な滞在地の分類

表 9 に重要な滞在地と重要でない滞在地を分類した結果の AUC を、その場合の平均の ROC 曲線を図 2 に示す。平均の ROC 曲線を見ると、FP rate = 0.25 付近まではどの特徴量で分類した場合も同等の性能で、それ以降 1 日あたりの滞在時間での分類が他よりも優れた性能を示した。平均の ROC 曲線の AUC の値を見ても、1 日あたりの滞在時間で分類した場合が最も高く、分類性能が最も良いという

表 9 重要な滞在地の分類性能 (AUC)

	user A	user B	user C	user D	user E
滞在時間	0.713	0.859	0.768	0.781	0.666
訪問回数	0.592	0.630	0.734	0.820	0.736
期待値	0.483	0.496	0.724	0.791	0.748
	user F	user G	user H	user I	平均
滞在時間	0.925	0.939	0.752	0.801	0.802
訪問回数	0.823	1.000	0.597	0.822	0.738
期待値	0.870	0.971	0.592	0.892	0.722

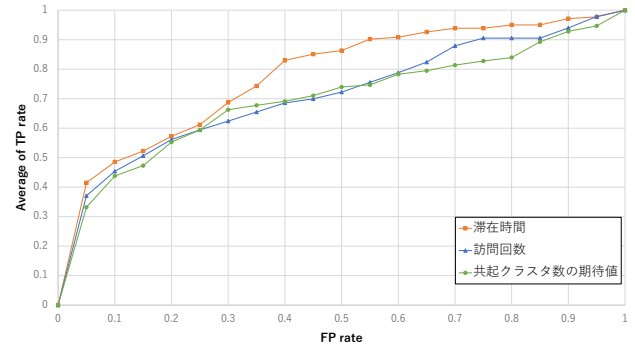


図 2 重要な滞在地の分離性能

表 10 重要な経由地の分類性能 (AUC)

	user A	user B	user C	user D	user E
滞在時間	0.414	0.750	0.581	0.382	0.471
訪問回数	0.635	0.808	0.613	0.725	0.569
期待値	0.751	0.825	0.599	0.791	0.615
	user F	user G	user H	user I	平均
滞在時間	0.136	0.922	0.414	0.622	0.494
訪問回数	0.342	0.875	0.635	0.759	0.656
期待値	0.562	0.938	0.703	0.833	0.734

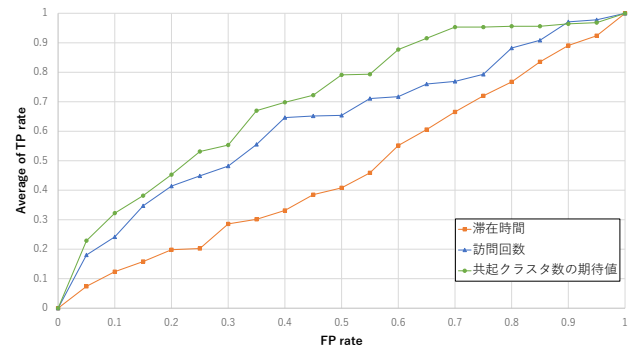


図 3 重要な経由地の分離性能

結果が得られた。3.3.2 節で特徴量ごとに重要な滞在地の網羅率を調べた時 (表 7) に、1 日あたりの滞在時間が重要な滞在地の特徴を最もよく捉えられていると考えられたので、妥当な結果と言える。

4.2.2 重要な経由地の分類

表 10 に重要な経由地と重要でない経由地を分類した結果の AUC を、その場合の平均の ROC 曲線を図 3 に示す。平均の ROC 曲線を見ると、共起クラスタ数の期待値に基づいて分類した場合の性能が常に高く、AUC の値も共起クラスタ数の期待値の場合の値が最も高いので、共起クラスタ数の期待値に基づく分類は、重要な経由地と重要でない

表 11 上位 50 % の場所における「普段の食事 (外食)」「信号待ち」,
「交通機関」タグの付いた重要な経由地の占める割合 (%)

	滞在時間	訪問回数	共起クラスタ数の期待値
user A	23.73	38.98	49.15
user B	8.00	8.00	10.00
user C	4.76	4.76	4.76
user D	13.16	31.58	34.21
user E	4.23	8.45	8.45
user F	2.56	7.69	10.26
user G	0.00	0.00	0.00
user H	4.55	18.18	18.18
user I	27.78	41.67	45.83

経由地の分類性能が最も高いという結果が得られた。3.3.2 節で特徴量ごとに重要な経由地の網羅率を調べた時 (表 7) の結果からも妥当な結果と言える。

共起クラスタ数の期待値による重要な経由地と重要でない経由地の分類が最も良い性能を示した理由として、3.3.1 節で場所の用途ごとに場所の重要度を調べた際 (表 6)、場所の総数に対する重要な経由地の割合が高かった「普段の食事 (外食)」、「信号待ち」、「交通機関」のタグが付けられた重要な経由地をうまく分類できたためと考えられる。表 11 に、実験協力者ごとに、3 つの特徴量それぞれの値が大きい場所の上位 50 % のうち、「普段の食事 (外食)」、「信号待ち」、「交通機関」タグが付いた重要な場所の占める割合を求めたものを示す。実験協力者 9 名中 5 名のデータで、共起クラスタ数の期待値が上位 50 % の場所に、3 つのタグの付いた重要な経由地の占める割合が最も高くなっている。このことから、共起クラスタ数の期待値は 3 つのタグが付けられた重要な場所の特徴をよく捉えられており、分類性能の向上につながったと考えられる。

なお、user G は重要な経由地の数が少なく、どの特徴量でも 0 % となっている。

5. まとめ

本報告では、詳細な見守りが可能な行動範囲の推定のために、日常生活における場所を滞在地と経由地に分けて、場所の重要度を検討した。その上で、実験協力者 9 名に日常生活をする上での場所の重要度の調査を行った。調査の結果、滞在地としてのみ重要な場所だけでなく、経由地としてのみ重要な場所も存在すること、また、滞在地、経由地の両方として重要な場所があることがわかった。

さらに、1 日あたりの滞在時間、訪問回数の他、生活パターンに注目した特徴量である共起クラスタ数の期待値を提案し、3 つの特徴量を用いて、日常生活をする上での重要な場所と重要でない場所の分類を行い、ROC 曲線とその AUC の値を用いて性能を評価した。結果として、重要な滞在地と重要でない滞在地を分類する場合は 1 日あたりの滞在時間が、重要な経由地と重要でない経由地を分類す

る場合は共起クラスタ数の期待値に注目した分類が最も良い性能を示した。

今後の課題として、ユーザの GPS データから得られる特徴量だけでなく、地理データから得られる情報を取り入れた新たな特徴量の他、既存の特徴量を組み合わせて分類に用いることを検討する。

6. 謝辞

本研究は JSPS 科研費 20K12079 の助成を受けて実施したものである。

参考文献

- [1] 警視庁, 平成 30 年における行方不明者の状況, 入手先 (<https://www.npa.go.jp/safetylife/seianki/fumei/H30yukuehumeisha.pdf>) (参照 2020-05-11).
- [2] 子どもの防犯・老人徘徊対策 — 定額制位置検索システムどこ・イルカ mini, 入手先 (<https://www.dokoiruka.jp/>) (参照 2020-05-11).
- [3] 【公式】子供用 GPS の「みもり」, 入手先 (<https://mimori-ai.jp/>) (参照 2020-05-11).
- [4] スマートフォン・パーソナルセキュリティシステム「e-遠隔見守り TM」, 入手先 (<http://www.e-sares.co.jp/pr-rw.html>) (参照 2020-05-11).
- [5] Y. Matsuo, S. Hara, and M. Abe, "Algorithm to estimate a living area based on connectivity of places with home," HCI International 2015 - Posters Extended Abstracts, Part II, CCIS 529, pp.570–576, Springer International Publishing, Aug. 2015.
- [6] Y. Ye, Y. Zheng, Y. Chen, J. Feng and X. Xie, "Mining Individual Life Pattern Based on Location History," Proceedings of the 10th IEEE International Conference on Mobile Data Management, pp. 1–10, Nov. 2009.
- [7] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp.226-231, Aug. 1996.
- [8] G. Schoier and G. Borroso, "Individual movements and geographical data mining. Clustering algorithms for highlighting hotspots in personal navigation routes," Proceedings of the International Conference on Computational Science and ITS Applications, pp. 454-465, June 2011.
- [9] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, "Discovering Personally Meaningful Places: An Interactive Clustering Approach," ACM Transactions on Information Systems, Vol. 25, No. 3, Article 12, July 2007.
- [10] L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macedo, B. Moelans and A. Vaisman, "A model for enriching trajectories with semantic geographical information," Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, No. 22, pp. 1-8, Nov. 2007.
- [11] A. T. Palma, V. Bogorny, B. Kuijpers and L. O. C. Alvares, "A Clustering-based Approach for Discovering Interesting Places in Trajectories," Proceedings of the 2008 ACM symposium on Applied computing, pp.863-868, Mar. 2008.
- [12] T. Luo, X. Zheng, G. Xu, K. Fu, and W. Ren, "An Improved DBSCAN Algorithm to Detect Stops in In-

- dividual Trajectories," ISPRS International Journal of Geo-Information, Vol. 6, Iss. 3, Mar. 2017.
- [13] L. Bermingham and I. Lee, "A probabilistic stop and move classifier for noisy GPS trajectories," Data Mining and Knowledge Discovery, Vol 32, Nov. 2018.
 - [14] Fritsch, F. N. and R. E. Carlson, "Monotone Piecewise Cubic Interpolation," SIAM Journal on Numerical Analysis, Vol. 17, pp.238–246, Apr. 1980.
 - [15] GeoHex, 入手先 <http://www.geohex.org/> (参照 2020-05-11)