

GPS データに基づく日常生活における 特別な行動の検出

小林 誠¹ 原 直² 阿部 匡伸¹

概要：本報告では、日常生活における特別な行動を検出する方式を検討する。日常生活における特別な行動とは、飲み会や、買い物など、日常生活とは少し違う行動のことを指す。このような行動をした日を、「日常の中の非日常」と呼ぶ。「日常の中の非日常」は旅行などのような「非日常」と比較して忘れやすい傾向があると考えられる。「日常の中の非日常」を検出することで、新たな気づきや発見がある思い出を振り返ることができる。本報告では、ライフログとして記録された GPS (Global Positioning System) データを用いて「日常の中の非日常」を検出する方法を提案した。提案手法は、1日ごとのデータを行動パターンが似ているクラスタに分割し、クラスタリングにおける外れ値を「日常の中の非日常」として検出するという方法である。検出精度の評価として、異常標本精度と正常標本精度が一致する点での分岐点精度による評価をおこなった。評価結果としては、参加者4名中2名の分岐点精度が80%以上となり、提案手法の有効性が示された。また、適合率の評価によりランダムに異常判定するよりも、提案手法により異常判定したほうが良い性能を示すことがわかった。

Study of detecting special behavior in daily life based on GPS data

MAKOTO KOBAYASHI¹ SUNAO HARA² MASANOBU ABE¹

1. はじめに

ライフログとは、人間の活動をデジタルデータとして残すこと、あるいは残した記録自体のことであり、様々な方法で活用されている [1]。本報告では、ライフログデータの1つである、GPS (Global Positioning System) を用いた位置情報のデータ (以下、GPS データ) に着目する。近年、スマートフォンなどのデバイスの普及で GPS が身近な存在となっている。GPS 機能を搭載したデバイスを持ち歩くだけで、ユーザの行動履歴が自動的に収集される。GPS データからは、緯度、経度、時刻の情報が得られるため、ユーザがいつ、どこにいたのかを把握することができる。また、GPS データは継続的に記録することで、ユーザの移動経路も把握することができるという特徴がある。この GPS データの特徴に着目し、旅行者の GPS データから観光ルートの分析、および都市の観光魅力度を算出する研究 [2] や、携帯電話の GPS データと操作ログを収集し、利

用者に対して推薦情報という形態で周辺店舗の情報提供をおこなうサービスの検討 [3] がある。

本報告では GPS データを用いた応用例として、思い出の振り返りを検討する。特に、日常生活における特別な行動に着目し、その振り返るべき日を特定する方法を提案する。

日常生活における特別な行動とは、朝家から学校に行き、夕方まで勉強し、帰宅するという日常生活がある中で、ある日の夜は友達と飲みに行ったというような、いつもと少し違う行動を指す。このような行動をしている日を「日常の中の非日常」と呼ぶ。これらの特別な行動は、ある期間内に不定期 (2週間に1回、月に1回、...) におこなわれることが多いため、1回1回の出来事を詳細に覚えていない、または、忘れてしまっている日が存在していると考えられる。そこで「日常の中の非日常」を検出することで、新たな気づきや発見があるような思い出を振り返ることができる。

「日常の中の非日常」を検出するための提案手法としては、まず1日ごとのデータを、行動パターンが似ている日のクラスタに分割する。そしてあるクラスタの外れ値とな

¹ 岡山大学 大学院ヘルスシステム統合科学研究科

² 岡山大学 大学院自然科学研究科

る日を検出することで、日常生活から外れた「日常の中の非日常」を見つけるという方法である。クラスタリングの手法としては混合正規分布モデルを使用し、クラスタ数はベイズ情報量規準 (Bayesian information criterion, BIC) に基づき決定する。外れ値の検出にはクラスタリングに使用した混合正規分布に対する負の対数尤度を用いる。

2. 研究背景

2.1 ライフログ

ライフログとは、人間の活動をデジタルデータとして残すこと、あるいは残した記録自体のことであり、例えば動画や写真、位置情報などが挙げられる。ライフログに関する研究としては、食事画像から食事のバランスを推定する研究 [4] や、レシートの情報をライフログとして蓄積することで、ユーザの生活面や健康面のサポートをおこなうサービスの検討 [5] がある。本報告では、ライフログとして記録される GPS データに着目し、思い出の振り返りを検討する。

2.2 思い出の振り返り

思い出を振り返るときに、振り返りたい思い出として例えば、日常とは全く違う行動をしている旅行や帰省などの「非日常」が挙げられる。しかしそれ以外にも、日常生活とは少し違うことをしている飲み会や買い物などの「日常の中の非日常」も振り返りたい思い出であると考えられる。ここで、非日常 (主に旅行) と日常の中の非日常 (飲み会、買い物など) を比較してみると、飲み会や買い物は、旅行ほど強く印象に残らない傾向があると考えられる。そこで、日常の中の非日常を検出し、その日に撮影した動画や写真、行動の履歴である GPS データを用いてその日を振り返ることができれば、自分でも詳細に覚えていない、または忘れてしまっているような思い出を振り返ることができると考えられる。

思い出振り返りの先行研究としては、小出らによる研究 [6]、林らによる研究 [7]、Namba らによる研究 [8]、が挙げられる。小出らによる研究では、旅行者の街歩き中の思い出作りの支援を検討しており、非日常である旅行に着目している。林らによる研究では、非日常である旅行を主な検索対象としており、その日を「特別な日」として検索する方法を提案している。また、Namba らによる研究では、GPS データと写真を用いて、検索された「特別な日」を振り返るシステムを検討している。林らによる研究、Namba らによる研究ではいずれも、「特別な日」を検索するとき、1日ごとの特別度を定義し、上位から検索するという方法を使用している。この方法の問題点としては、上位から何件目までが振り返るべき思い出であるかの基準がなく、検索数が少なければ、非日常だけの振り返りとなっていることが挙げられる。しかし、非日常だけではなく、日常の

中の非日常の中にも振り返るべき思い出があると考えられる。本報告では、先行研究で主な対象としている非日常ではなく、日常生活における特別な行動をとった「日常の中の非日常」を検出することを目的とする。

2.3 「日常の中の非日常」の定義

本報告では、ある1日を「日常」、「非日常」、「日常の中の非日常」の3種類のどれかに分類する。それぞれの特徴を例を挙げながら以下にまとめる。

2.3.1 日常

日常とは、ある行動パターンがあり、それが何日も繰り返されているような日であると考えられる。例えば以下のような行動パターンが考えられる。

- 家に1日中滞在
- 家から学校に行き、帰宅
- 家から学校に行き、バイトをした後、帰宅

ここに挙げた行動パターンを持つ日は、複数日存在すると考えられる。このような日を「日常」とする。バイトやサークル活動などの週1回くらいの頻度の活動も「日常」に含める。

2.3.2 非日常

非日常とは、旅行や帰省などで、生活するエリアが変わるような日を指す。例えば、県外に旅行に行き、1日を通して家に1回も滞在していないような日である。よって、本報告で、非日常と想定しているのは以下のような日である。

- 旅行
 - 帰省
- このとき、出発日や帰着日も非日常に含める。

2.3.3 日常の中の非日常

日常の中の非日常とは、日常と少し違うことをしている日、ある期間内に不定期 (2週間に1回、月に1回、...) に何かをしている日を指す。以下のような日を「日常の中の非日常」としている。

- ショッピング (服など)
- 飲み会
- 外食
- 趣味の活動 (ドライブ、サイクリングなど)
- 研究室のイベント

2.4 混合正規分布モデルによるクラスタリング

混合正規分布モデルとは、混合モデルの1種で、正規分布の線形重ね合わせで表されるモデルである。混合モデルを使うことによって、それぞれのクラスタのデータを表現するような確率分布を割り当てることが可能になる。

クラスタ数 K の混合モデルは一般的に式 (1) の形を取る。

$$p(\mathbf{x} | \Theta) = \sum_{k=1}^K \pi_k p_k(\mathbf{x} | \theta_k), \quad \pi_1 + \dots + \pi_K = 1 \quad (1)$$

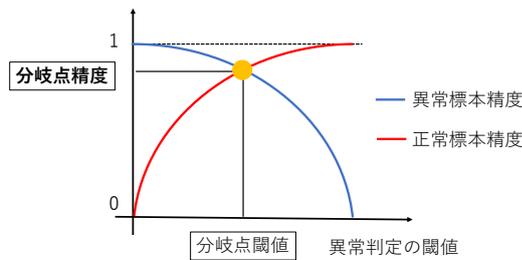


図 1 分岐点精度

ただし、 $p_k(x | \theta_k)$ は x について規格化された確率分布である。 θ_k は第 k 番目の確率分布に含まれるパラメータ、 Θ は右辺に現れる未知パラメータの集合で、 $\Theta = \{\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K\}$ である。和の中の第 k 項は第 k クラスと呼ばれる。

混合正規分布モデルのパラメータ Θ の推定は、期待値-最大化法を使用する。

2.5 ベイズ情報量規準

BIC を基準に混合正規分布のクラス数 K を決定する。ベイズ情報量規準 (Bayesian information criterion, BIC) はクラスタリングする際のクラス数の選択の目安として使われる指標である。BIC は式 (2) で定義される。

$$\text{BIC} = -2 \ln(\mathcal{L}) + K \ln(n) \quad (2)$$

\mathcal{L} は尤度、 K は推定するパラメータ数、 n はサンプル数である。一般に、BIC を最小にするモデルを最善の最尤モデルとみなすことができる。

2.6 異常検出の性能評価指標

異常検出の性能評価として、異常標本精度と正常標本精度が一致する点での精度である分岐点精度を用いた評価 [9] をおこなう。また、適合率による評価もおこなう。

2.6.1 分岐点精度

異常標本精度と正常標本精度の定義は以下の通りである。

$$(\text{異常標本精度}) = \frac{tp}{(tp + fn)} \quad (3)$$

$$(\text{正常標本精度}) = \frac{tn}{(fp + tn)} \quad (4)$$

異常標本精度は実際に異常標本であるデータのうち、異常標本であると判定されたデータの割合であり、正常標本精度は実際に正常標本であるデータのうち、正常標本であると判定されたデータの割合である。ここで、tp (true positive) は正しく異常標本と判定された数、fn (false negative) は誤って異常標本と判定された数、tn (true negative) は正しく正常標本と判定された数、fp (false positive) は誤って正常標本と判定された数である。

分岐点精度は図 1 に示すように、異常判定の閾値を変化させたときの異常標本精度、正常標本精度を描いたグラフ

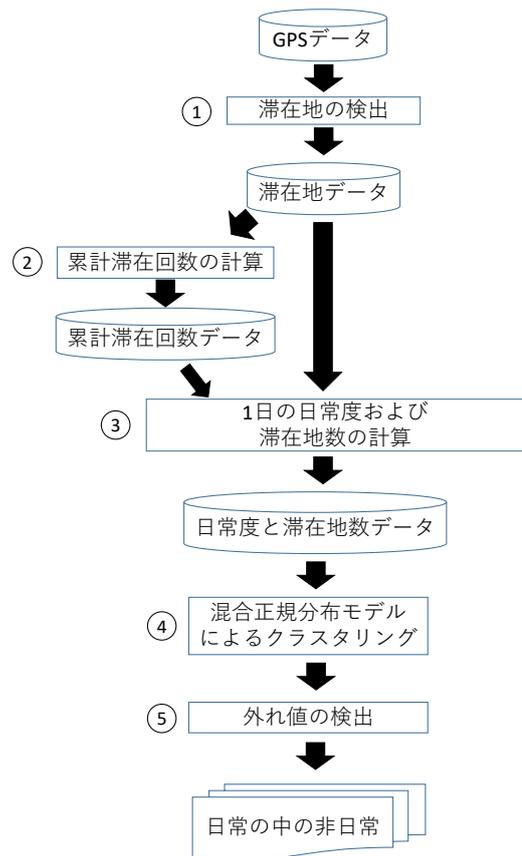


図 2 提案手法

の交点の y 座標の値となる。また交点の x 座標の値は、分岐点閾値と呼ばれる。

2.6.2 適合率

適合率 (Precision) は、式 (5) で定義される [10]。

$$\text{Precision} = \frac{tp}{(tp + fp)} \quad (5)$$

本報告では、異常標本であると判定されたデータの中で、正しく異常標本だと判定された割合を評価する指標として使用する。

3. 提案手法

提案手法の特徴は、クラスタリングにより 1 日ごとのデータを行動パターンが似ている日のクラスタに分割し、各クラスタの外れ値となる日を「日常の中の非日常」として検出するという点にある。クラスタリングをおこなうのは、日常生活の中にはいくつかの行動パターンが存在していると考えられるためである。提案手法の概要図を図 2 に示す。提案手法の流れは以下の通りである。

- (1) GPS データに基づく滞在地の検出
- (2) 滞在地に対する累計滞在回数の計算
- (3) 1 日の日常度および滞在地数の計算
- (4) 混合正規分布モデルによるクラスタリング
- (5) 外れ値を日常の中の非日常として検出

本章では、著者の 2018 年 4 月 24 日から 12 月 4 日まで

のデータを例として、提案手法の流れを説明する。なお、GPS データは、実験参加者に GPS ロガーを携帯してもらうことによって収集している。使用した GPS ロガーは、CanMore Electronics 社の GT-740FL (Sport LogBook) である。GPS データの取得間隔は 1 秒である。

3.1 GPS データに基づく滞在地の検出

収集された GPS データから滞在地を検出し、滞在地データを作成する。滞在地データは表 1 の形式で表される。

本報告では、林らによる研究 [7] で使用される滞在地検出のアルゴリズムを使用した。まず、GPS データから 5 分以上続けて留まっていた箇所の中心座標を停留点として検出し、停留点データを作成する。検出される停留点としては、同じ場所を意味する停留点も含まれる。そのため、停留点データ内で 2 地点間の距離が 100 m 以内である地点を同じ場所に行ったとみなし、その中点を 1 つの滞在地として検出する。さらにその中点との距離が 100 m 以内である地点も同じ滞在地として検出するという計算を繰り返し、滞在地データを作成する。

3.2 滞在地に対する累計滞在回数の計算

滞在地を検出した後に、ある 1 つの滞在地に対して GPS データ取得期間中の滞在回数を合計する。これは、同じ滞在地ラベルを持つ停留点の総数として計算する。累計滞在回数を計算することによって、表 2 のような滞在地ラベルごとの累計滞在回数のデータを得る。

3.3 1 日の日常度および滞在地数の計算

滞在地データと累計滞在回数データをもとにして、1 日の「日常度」と 1 日に滞在した滞在地の数を計算する。計算することで表 3 のような日ごとの日常度と滞在地数のデータを得る。

3.3.1 日常度

日常度は、ある 1 日で滞在したすべての滞在地の累計滞在回数の総和として計算する。例えば表 2 のように累計滞在回数が計算されているとする。ある 1 日の滞在地が、A-0 → A-1 → A-2 → A-0 である日の日常度は、累計滞在回数の総和 (200 + 100 + 50 + 200 = 550) となる。この計算を 1 日ごとにおこなう。

累計滞在回数が多くなる地点としては、家、学校、バイト先など日常的に訪れる地点が考えられる。また、飲み会などで初めて行くような場所は、累計滞在回数が 1 回となるため日常度への影響は小さくなる。そのため、日常度は累計滞在回数の多い地点に依存した数値になると考えられる。本報告では、ある程度行動パターンが似ている日は同程度の日常度になることを想定している。

3.3.2 滞在地の数

日常度の計算と同時に 1 日の滞在地数を数える。これ

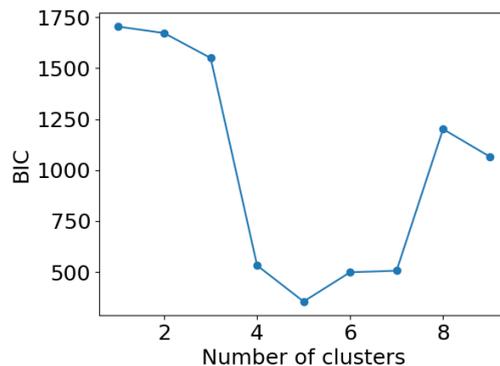


図 3 BIC の変化

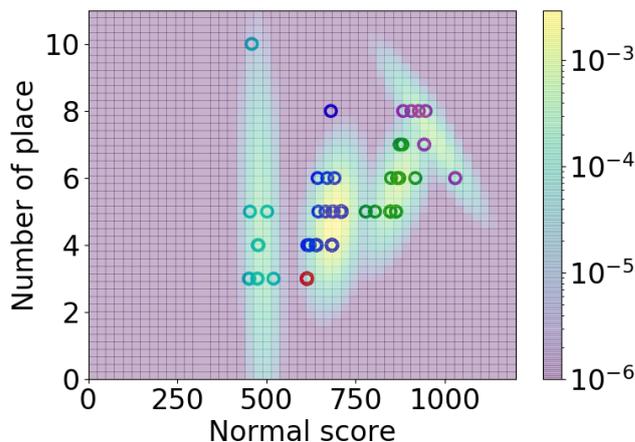


図 4 混合正規分布モデルによるクラスタリング。カラーバーは確率密度である。

は、日常度に加え、1 日の滞在地の数に着目し、外れ値の検出をおこなうためである。

3.4 混合正規分布モデルによるクラスタリング

日常度と 1 日の滞在地の数に基づいてクラスタリングをおこなう。クラスタリングの目的は、行動パターンが似ている日を 1 つのクラスターにまとめることである。クラスタリングの手法としては、2.4 節で説明した混合正規分布モデルを使用する。

$$p(\mathbf{x} | \Theta) = \sum_{k=1}^K \pi_k N(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (6)$$

ここで、 $\mathbf{x}_i = \begin{bmatrix} u_i & v_i \end{bmatrix}$ であり、 u_i は 1 日の日常度、 v_i は 1 日の滞在地の数である。

クラスター数は BIC を基準に選択する。本報告では、1 から 9 までクラスター数を変化させたとき、BIC が最も小さくなる際のクラスター数でクラスタリングをおこなう。図 3 は BIC の変化の例であり、この場合はクラスター数 5 を選択する。そして、クラスター数 5 でクラスタリングをおこなった結果の例が図 4 である。横軸は日常度で、縦軸は滞在地の数であり、クラスターごとに色分けされている。

表 1 滞在地データの例

停留点 ID	緯度	経度	停留開始時刻	停留時間	滞在地ラベル
1	34.689941	133.923204	2018-01-01 18:00:00	3600 秒	A-0
2	34.683636	133.920288	2018-01-01 19:30:00	5400 秒	A-1
3	34.689988	133.923271	2018-01-01 21:30:00	1800 秒	A-0
⋮	⋮	⋮	⋮	⋮	⋮

表 2 累計滞在回数データの例

滞在地ラベル	累計滞在回数
A-0	200
A-1	100
A-2	50
⋮	⋮

表 3 日ごとの日常度と滞在地数データの例

日付	日常度	滞在地数
2018-01-01	550	4
2018-01-02	800	6
⋮	⋮	⋮

なお、混合正規分布モデルによるクラスタリングをおこなう際には、Python の機械学習ライブラリ scikit-learn 0.19.1 の gmm を利用した。混合正規分布のパラメータの初期値は、k-means 法により設定される。

3.5 外れ値を日常の中の非日常として検出

正常のモデルからずれているデータを異常なデータとして検出する。そのため、データの「異常度」を定義し、正常からのずれの度合いを測る。本報告では、異常度として負の対数尤度を使用する。ある確率分布 $p(\mathbf{x})$ が与えられているときの観測値 \mathbf{x}' に対する異常度を式 (7) に示す。

$$\text{anomaly_score}(\mathbf{x}') = -\ln p(\mathbf{x}') \quad (7)$$

確率分布に混合正規分布を使用しているため、異常度は式 (8) になる。

$$\text{anomaly_score}(\mathbf{x}') = -\ln \left\{ \sum_{k=1}^K \hat{\pi}_k N(\mathbf{x}' | \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k) \right\} \quad (8)$$

異常度を 1 日ごとに計算した結果を、図 5 に示す。この図は横軸が日常度であり、縦軸が異常度をとる。この異常度が高い日を検索することによって、日常の中の非日常を検出するというのが提案手法である。

4. 評価実験

4.1 評価実験に用いたデータ

評価実験は、筆者が所属する研究室の学生 4 名を対象とした。GPS データは 2018 年 4 月 24 日から 12 月 4 日までのデータを使用した。実験参加者ごとのデータ日数とクラスタ数を表 4 に示す。

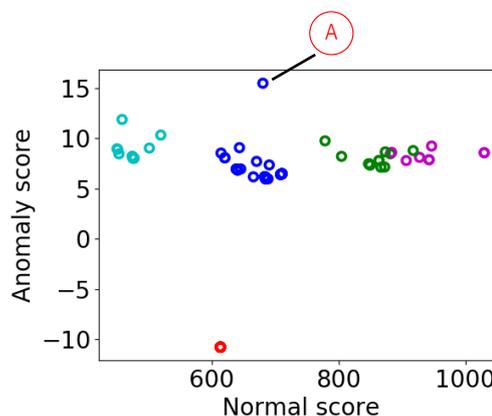


図 5 異常度の計算結果。図中の A はドライブに行った日であり、日常の中の非日常が外れ値として表されている。

表 4 実験参加者ごとのデータ日数およびクラスタ数

	データ日数	クラスタ数
参加者 A	119	7
参加者 B	78	8
参加者 C	99	6
参加者 D	144	7



図 6 滞在地の閲覧画面

4.2 実験参加者によるラベル付け

参加者によるラベル付けは、1 日ごとの滞在地を閲覧させることによっておこなう。本実験では、参加者の負担を減らすため、収集されたデータの中からいくつか抽出したデータに対して、ラベル付けさせた。抽出したデータは、クラスタ内でデータを異常度の高い順に並べたときの、上位から 4 件、中央から 4 件、下位から 4 件である。なお、クラスタ内のデータが 12 個に満たない場合は、クラスタ内のすべてのデータを抽出した。抽出したデータは日付順に並び替えたものを参加者にラベル付けさせた。図 6 に滞在地の閲覧画面を示す。使用した地図は、OpenStreetMap*1で

*1 <https://openstreetmap.jp/>

表 5 「日常」,「日常の中の非日常」データの個数と「日常の中の非日常」の割合

	日常	日常の中の非日常	日常の中の非日常の割合 (%)
参加者 A	42	12	22.2
参加者 B	40	20	33.3
参加者 C	25	19	43.2
参加者 D	63	15	19.2

表 6 評価結果

	分岐点閾値	分岐点精度 (%)	適合率 (%)
参加者 A	6.70	81.0	52.9
参加者 B	6.57	60.0	42.9
参加者 C	6.31	80.0	76.2
参加者 D	8.61	73.3	40.7

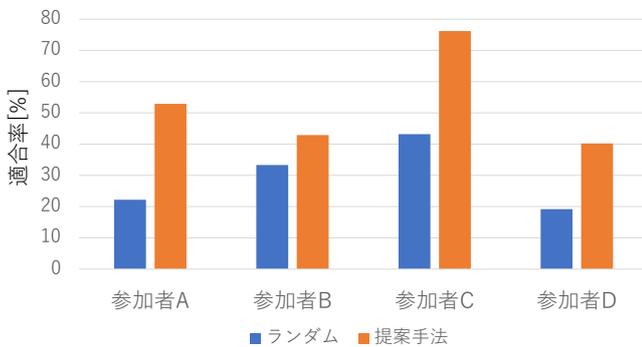


図 7 適合率の比較

ある。地図上には、1日の中で滞在した地点にマーカーが表示される。そして、マーカーをクリックすると、滞在を開始した日時、滞在を終えた日時、滞在時間が表示される。

ラベルは、日常のデータであれば「0」、日常の中の非日常のデータであれば「1」を付ける。提案手法で取り除けなかった非日常を発見した場合はラベルとして「2」を付けさせた。滞在した覚えのない場所が存在するような GPS データのエラーがある場合はラベルとして「3」を付けさせた。ラベル 2 と 3 はいずれも評価の対象外とした。参加者によりラベル付けされた「日常」、「日常の中の非日常」データの個数と「日常の中の非日常」の割合を表 5 に示す。

4.3 実験結果

分岐点精度と適合率により性能評価をおこなった結果を示す。参加者それぞれの分岐点閾値と分岐点精度と適合率を表 6 にまとめる。

表 6 より、参加者によって分岐点精度にばらつきはあるが、参加者 A と参加者 C は分岐点精度が 80% 以上であり、「日常の中の非日常」データの異常度が高い傾向にあることがわかる。また、参加者 C の適合率が最も高く、76.2% となっていることがわかる。

表 5 の「日常の中の非日常の割合」は、データをランダムに正常標本もしくは異常標本と判定した場合の適合率と

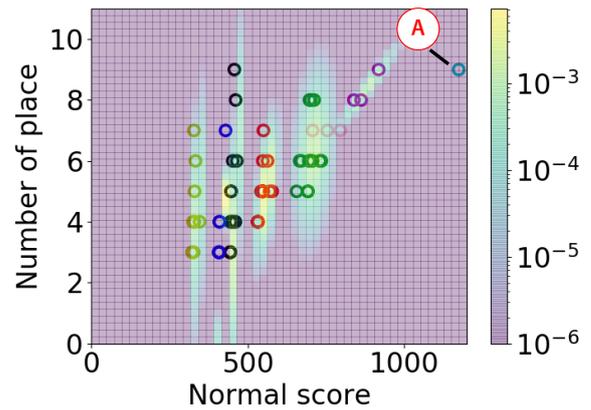


図 8 参加者 B のクラスタリング結果。カラーバーは確率密度である。図中の A が日常度が極端に大きいデータである。

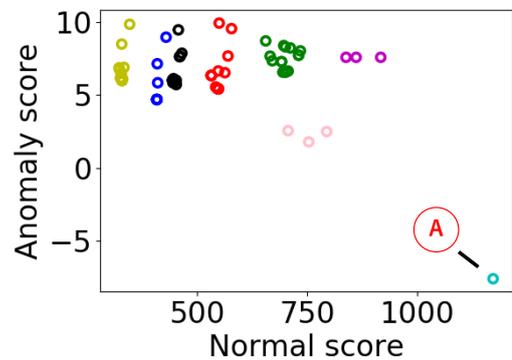


図 9 参加者 B の異常度の分布。図中の A の異常度が小さくなっている。

考えることができる。図 7 にランダムに異常判定した場合と提案手法により異常判定した場合の適合率の比較を示す。図 7 から、提案手法により異常判定した方がランダムに異常判定した場合よりも適合率が全体的に高くなっていることがわかる。特に、分岐点精度の高い参加者 A と参加者 C は適合率が 30 ポイント以上向上している。提案手法を用いて分岐点閾値を求め、その閾値に基づき異常判定することで、適合率が向上している。

4.4 考察

4.4.1 外れ値検出できないデータ

本報告の目的とする外れ値の検出ができていない例として、1つのデータで1つのクラスターを形成してしまうケースがある。図 8 に参加者 B の日ごとのデータをクラスタリングした結果、図 9 に参加者 B の異常度の分布を示す。

図 8 を見ると、他のデータと比較して日常度の大きなデータが 1 つあり、そのデータ 1 つだけで 1 つのクラスターを構成していることがわかる。検出したいデータに対してクラスターが 1 つ割り当てられてしまうため、外れ値の検出ができず、結果として図 9 のように異常度が極端に小さな値になってしまう。

表 4 より、参加者 B はデータ数が最も少なく、クラスター

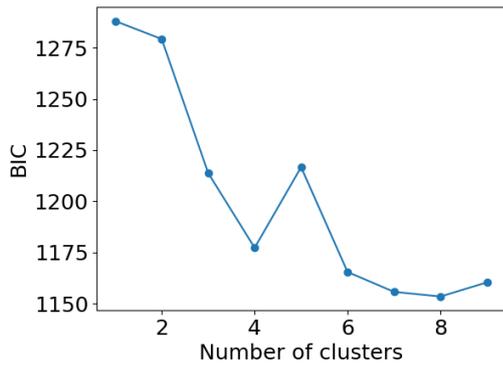


図 10 参加者 B の BIC の変化

表 7 クラスタごとの分岐点閾値と分岐点精度. クラスタ内が「日常」データや「日常の中の非日常」データだけである場合は、評価できないため *** としている.

番号	日常	日常の中の非日常	分岐点閾値	分岐点精度 (%)
0	0	2	***	***
1	12	0	***	***
2	7	5	5.17	80.0
3	9	2	9.42	100
4	4	0	***	***
5	2	0	***	***
6	8	3	6.49	75.0

数が最も大きくなっていることから、データ数に対してクラスタ数が多くなりすぎていることが原因の1つと考えられる。提案手法は、ある程度同じような行動をしている日をクラスタにまとめて、それらのクラスタから外れている日を検出するアルゴリズムである。しかし、クラスタ数が不要に多くなってしまうと、クラスタ内のデータ数が少なくなり、外れ値として検出されるべきデータにクラスタが割り当てられてしまい、検出できなくなることが考えられる。参加者 B の BIC の変化を図 10 に示す。本実験では、BIC の値が最も小さくなるクラスタ数 8 によりクラスタリングしたデータで実験をおこなったが、図 10 を見ると、クラスタ数 4, 6, 7 のときに BIC の値が比較的小さい値になっている。これらの値において、本実験よりもよい分岐点精度が得られる可能性が考えられる。

4.4.2 クラスタごとの分岐点閾値と分岐点精度

本報告では、全体のデータをいくつかのクラスタに分割しているため、それぞれのクラスタでの分岐点精度を評価することもできる。ここでは分岐点精度が最も高かった参加者 A に注目する。参加者 A の各クラスタの分岐点閾値と分岐点精度を、表 7 に示す。表 7 から分岐点精度が 100% であるクラスタがあることがわかる。分岐点精度が 100% であるということは、ある異常判定の閾値を設定すると、その閾値より異常度の小さなデータはすべて正常(日常)のデータであり、閾値より異常度の大きなデータは、全て異

常(日常の中の非日常)であることを示している。よって、本報告で目的としていた、外れ値として日常の中の非日常を検出することができていることがわかる。ただ、番号 0 に関しては日常の中の非日常のデータだけのクラスタとなっており、4.4.1 で示したような欠点も現れていることがわかる。

4.4.3 アンケートに基づく考察

評価実験をする際に、参加者に気づきや要望を書くように依頼した。ラベル付けはクラスタごとにおこなってもらったが、クラスタ内で「家 → スーパー → 家」の行動をしているデータが多かった等のコメントがあったため、クラスタリングにより、行動パターンが似ているクラスタを作成できていることがわかった。

また、地図上のマーカーの位置をズームしても滞在した店舗の名前などが表示されず、その場所で何をしていたのかが思い出せないという意見があった。今回使用した地図は、OpenStreetMap であったが、より詳細な情報が記載されている地図に差し替える必要があると考えられる。そして、現状ラベル付けをしてもらう際に見せるのは、滞在地のみとなっているが、それだけでは不十分であることもわかった。例えば、普段滞在している場所の近くで、いつもと違うことをしている場合にそれらが別の滞在地であると感じにくいという意見があった。今後は、移動経路も見せながらラベル付けしてもらうべきであると考えられる。

5. まとめ

本報告では、思い出の振り返りに関して旅行や帰省などの「非日常」ではなく、飲み会や買い物などの「日常の中の非日常」を検出する方法を提案した。提案手法は、まず 1 日ごとのデータを、行動パターンの似ている日のクラスタに分割する。そして、外れ値となる日を「日常の中の非日常」として検出するという方法である。

日常の中の非日常の検出精度を評価するために、4 名の実験参加者による評価実験をおこなった。各参加者によってラベル付けされたデータをもとに、分岐点精度による性能評価をおこなった。実験の結果から参加者 4 名中 2 名の分岐点精度が 80% 以上となり、提案手法の有効性が示された。また、クラスタごとに精度を評価すると分岐点精度が 100% となるクラスタがあり、外れ値として「日常の中の非日常」を検出できることがわかった。そして、適合率の評価から、ランダムに異常判定する場合より、分岐点閾値による異常判定の方が優れていることを確認した。

今後は、クラスタ数を変化させたときの分岐点精度を評価し、最適なクラスタ数の決定方法を検討する。また、本報告では日常度と 1 日の滞在地の数に着目したが、GPS データから抽出できる情報として、時間の情報がある。滞在時間を用いた異常度の定義を検討し、本報告で提案した手法との比較実験をおこなう。

参考文献

- [1] ゴードンベル, ジムゲメル, 飯泉恵美子 (翻訳), 人生の「すべて」をデジタルに記録する! ライフログのすすめ, 早川書房, Jan. 2010.
- [2] 長尾光悦, 川村秀憲, 山本雅人, 大内東他, “GPS ログからの周遊型観光行動情報の抽出,” 情報処理学会研究報告, vol.2005, no.78, pp.23–28, 2005.
- [3] 佐藤一夫, “ライフログ: 2. プライバシ保護を考慮したケータイ行動ログの利活用について,” 情報処理, vol.50, no.7, pp.598–602, 2009.
- [4] 北村圭吾, 山崎俊彦, 相澤清晴, “食事ログの取得と処理—画像処理による食事記録—,” 映像情報メディア学会誌, vol.63, no.3, pp.376–379, 2009.
- [5] 徳永清輝, まつ本真佑, 中村匡秀, “レシート蓄積による消費者向けライフログサービスの考察,” 電子情報通信学会技術研究報告, vol.110, no.282, pp.95–100, 2010.
- [6] 小出渉太, 木下雄一朗, 郷健太郎他, “街歩き旅行者の写真撮影促進にもとづく思い出形成支援,” 第76回全国大会講演論文集, vol.2014, no.1, pp.615–617, 2014.
- [7] 林啓吾, 原直, 阿部匡伸, “滞在地の特徴量を利用した「特別な日」検索方式,” 電子情報通信学会技術研究報告, vol.114, no.500, pp.89–94, 2015.
- [8] A. Namba, S. Hara, and M. Abe, “LiBS: lifelog browsing system to support sharing of memories,” Proceedings of UbiComp/ISWC 2016 Adjunct, pp.165–168, Sept. 2016.
- [9] 井手剛, 入門 機械学習による異常検知 R による実践ガイド, コロナ社, 2015.
- [10] C.D. Manning, P. Raghavan, and H. Schütze, “Introduction to Information Retrieval,” pp.142–143, Cambridge University Press, 2008.