

GPS データに対する配列解析の援用

矢部直人

Application of Sequence Alignment Analysis to GPS Tracking Data

Naoto YABE

Abstract: In this paper, sequence alignment analysis that is mainly developed in bioinformatics is applied for GPS tracking data. The GPS data have been recorded from tourist visiting Tama Zoological Park in Tokyo, Japan. The zoo is divided into 31 zone and tourist's walking trajectories are coded into activity sequences that comprised of the alphabet of the zone. The ClustalTXY software is used in this study because the software can compute temporal similarities in addition to activity similarities. The results show that including temporal similarities can avoid local optimum solution and can be effective for clustering GPS tracking data.

Keywords: GPS (global positioning system), 配列解析 (sequence alignment analysis), ClustalTXY (ClustalTXY), 観光行動 (tourist activity), 多摩動物公園 (Tama Zoological Park)

1. はじめに

観光地内部における観光客の行動については、観光施設の立地決定などに役立つことから、従来から調査がなされてきた(橋本, 1991). 近年では、GPS や WiFi などの位置を測定する技術が普及し、行動調査を行いやすい環境が整ってきている(Shoval and Isaacson, 2007a). そのため、特に観光地や屋外の観光施設における GPS を用いた行動調査が多数行われている(古谷, 2006, 有馬, 2010). しかし、GPS を用いた行動調査は多数行われているものの、GPS データに対する分析方法は未だ発展途上にあるといえよう (矢部ほか, 2010).

そのような中で注目に値するのが、もともと生物

学において遺伝子配列を分析する手法であった配列解析を GPS データに援用し、行動パターンの分類を行った事例である(Shoval and Isaacson, 2007b). 配列解析を援用することで、従来は属性の面から観光客の分類を行っていたのが、行動の面から観光客を分類することも可能になる. 特に配列解析の特徴は、訪れた場所の順序によって行動を分類できる点にあり、従来の訪問場所の単純集計による行動分類とは異なり、観光行動の順序を考慮することができる.

しかしながら配列解析を GPS データに対して援用した事例は少ないため、どのような問題があるのかが明らかになっていない. そこで本研究では、配列解析を GPS データの分析に援用するに際しての問題点を明らかにし、解決策を検討することを目的とする.

以下、2 では配列解析の分析手法の説明を行い、

矢部直人 〒192-0397 東京都八王子市南大沢 1-1

首都大学東京都市環境学部自然・文化ツーリズムコース

Phone: 042-677-1111

E-mail: yabebeya@tmu.ac.jp

3ではGPSデータに対して適用した結果を示す。4ではGPSデータに配列解析を援用する際の問題点とその解決策を検討し、今後の課題を展望する。

2. 配列解析の分析手法とソフトウェア

ここでは、(Wilson, 1998; Wilson, 2006; Wilson, 2008)を参考に、配列解析の手法について述べる。

2.1 配列解析

配列解析の基本は、二つの配列を要素ごとに対応づけて並べる操作(ペアワイズアラインメント)を行い、類似度を求めていく手法である。配列の類似度は要素のペア一つ一つに定義される類似度の和で与えられる。要素はそのまま対応づけられるだけでなく、要素の挿入・欠失を扱うためギャップを対応させることも多い。配列要素 a と b を対応づける類似度は以下の(1)式で求められる。

$$d(a,b) \quad (1)$$

最も単純な類似度は、 $a=b$ のとき 1, $a \neq b$ のとき 0 とするスコアである。

配列解析の手順は、まずサンプル全ての配列同士のアラインメント(ペアワイズアラインメント)を行う。その後、類似度の高い配列同士を順番にアラインメントしていき、サンプル全体の系統樹を作成するマルチプルアラインメントを実行する。

2.2 ソフトウェア

本研究で用いる ClustalTX Y は、時間地理学の分野で研究対象となる活動日誌の分析のために開発され、ペアワイズアラインメントからマルチプルアラインメントまでを一貫して処理することができるソフトウェアである。また、活動の類似性のみではなく、以下の(2)式のように活動の空間的な類似性を含めた類似性を考慮することができる。

$$q(a,b) = u * d(a,b) + v * E(a,b) \quad (2)$$

ここで、 $E(a,b)$ は要素の空間的な類似性を表し、 u と v はそれぞれ活動の類似性と空間的な類似性のウ

ェイトであり、 u と v の合計は 1 となる。

3. GPS データへの配列解析の援用

3.1 調査の概要

多摩動物公園で行った調査は、来園者にGPSロガーを配布して歩行軌跡を記録する調査である。調査に用いたGPSロガーは、Qstarz社のBlack Gold 1300であり、上空が開けていることなど最適な条件のもとでは水平誤差が3mの精度を持つ。調査には39台のGPSロガーを用い、1秒間隔で歩行軌跡を記録した。

多摩動物公園の出入り口は正門の一角所だけであるため、GPSロガーは正門で配布し、正門で回収した。調査日時は、2009年7月9日(木)~12日(日)の4日間、調査時間はいずれも開園~閉園までの時間である。

7月9日~12日の4日間では、合計247の有効なGPSログを取得することができた。調査日の天候はいずれも晴れまたは曇りであり、気象庁のアメダス観測所(八王子)のデータによると、各日の最高気温は30.5°C(9日)、31.6°C(10日)、27.3°C(11日)、28.6°C(12日)となっている。本研究では、土日のデータ($N=109$)のみを用いる。

取得したGPSデータには加速度および速度の両方で補正を行った。具体的には、加速度 $-1.5 \sim 1 \text{ m/sec}^2$ の範囲外のデータを異常値とみなして除去した上で、速度60 km/h以上の地点を除去した。さらに、動物園内の来園者が通行可能な通路から3m以上離れた地点も除去した。

3.2 行動配列の作成

多摩動物公園を、地形や動物展示・休憩施設の位置などを考慮して、全部で31のゾーンに分割した(図1)。そして来園者の歩行軌跡を記録したGPSデータから、1分間隔でゾーンの位置をサンプリングし、行動配列を作成した(図2)。

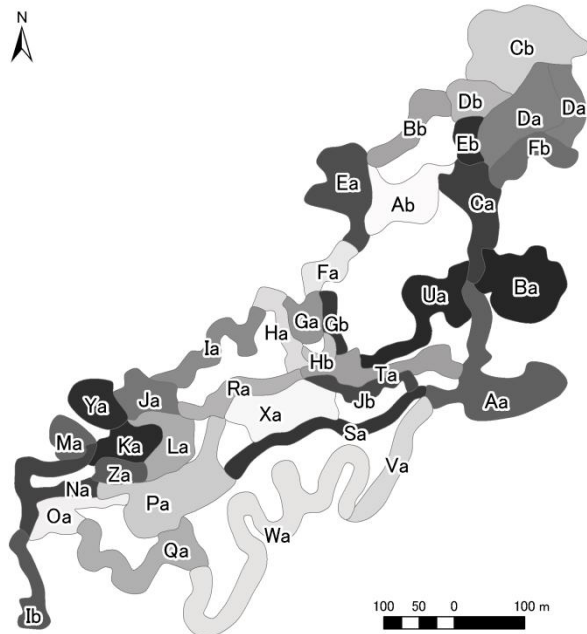


図-1 多摩動物公園の 31 ゾーン

>38

AaAaTaJbJbJbJbJbXaXaXaXaXaXaXaXaXaRaPaPaPaPaPaPaPaQaQaQaQa

>9

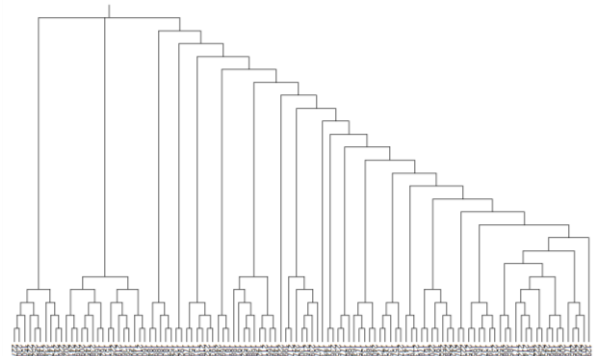
AaAaTaTaTaTaTaTaUaRaPaZaNaNaPaPaPaPaNaNaNaNaNaNaNaNaNaNaOaOaOaOaOaOa

図-2 行動配列の例

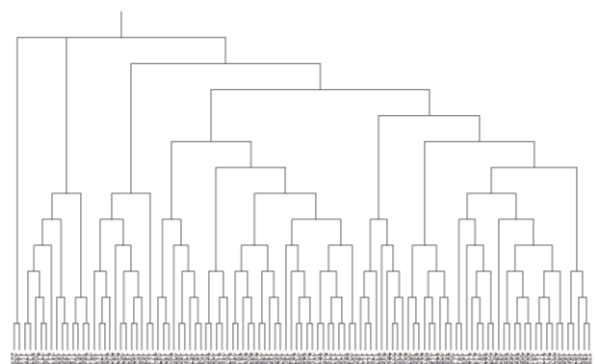
また、ClustalTXY を使用するに際して、空間的な類似性は行動配列で表されるため、時間的な類似性を示す値を XY データとして付加した。具体的には、開園時刻から 20 分間隔の時間を X 軸の値に入力し、Y 軸の値についてはすべて 0 を入力した。

3.2 行動配列の作成

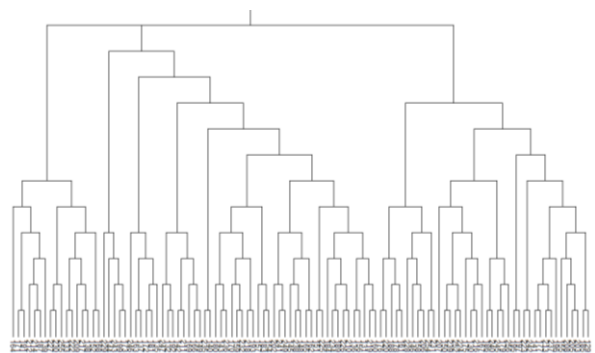
時間的な類似性を考慮しない分析の結果では、来園者の行動分類は必ずしもうまくいかないことが分かる。系統樹をみると、全てのサンプルが統合されるまでに多大なステップを要している（図 3a）。そのため、適切なグループ分けをするのが非常に困難である。このような結果になった原因は、休憩施設の滞在時間が長いいため、同じ休憩施設に滞在している来園者同士の行動の類似性が高いと判断されてしまうためである。これでは局所的な行動の類似



(a) 時間的類似性を含まない場合



(b) 時間的類似性を含む場合 ($u = 0.5, v = 0.5$)



(c) 時間的類似性を含む場合 ($u = 0.1, v = 0.9$)

図-3 マルチプルアラインメントの結果

性で来園者を分類することになってしまい、必ずしも園内全体の行動を反映した分類とはならない。

一方、時間的な類似性を加えた分析結果は、比較的うまく分類できることが分かった。活動の類似性のウェイトを 0.5、時間的類似性のウェイトを 0.5 とした場合、来園者は大まかに 3 ないしは 5 グループに分けることができる（図 3b）。さらに時間的類

似性のウェイトを 0.9 に高めた場合、来園者の行動は 3 グループに分けることができる (図 3c)。この 3 グループの行動は解釈が容易であり、グループごとの特徴も明確である。それぞれのグループの行動の特徴を述べると、第一グループは、昆虫館を訪れるグループである。第二グループは、園内を時計回りに回るグループ。第三グループは、園内を反時計回りに回るグループである。

4. おわりに

本研究では、GPS データの分析に配列解析を援用し、その問題点と解決策の検討を行った。その結果、時間的な類似性を考慮せず、活動の類似性のみで分析を行った場合には、グループ分けがうまくいかないことが判明した。その原因は、滞在時間が長い箇所が存在した場合、その局所的な類似部分にアライメントされてしまうためである。そのため、滞在時間が長い箇所の前後の行動が全く異なる場合でも、比較的類似した行動を取っているものとして分類されてしまう。そのため、系統樹によるグループ分けも困難であり、グループごとの行動の解釈も難しい。

それに対して、活動の時間的な類似性を考慮した分析では、来園者の行動をうまく分類できることが判明した。これは、活動の類似性のみでなく、時間的な類似性を考慮することが GPS で記録した行動の分類には有効であることを示している。今後は事例を蓄積し、本研究で得られた知見の検証を行う必要がある。また、ゾーン数や、活動のサンプリング間隔が分析結果に与える影響についても考察を加える必要があろう。

謝辞

Clarke Wilson 氏には配列解析ソフトウェア ClustalTXY をご提供いただいた。ここに記して感謝

いたします。本研究には、文部科学省科学研究費補助金若手研究 (B) (研究課題「配列解析を援用した原宿・表参道におけるストリートファッションの形成に関する研究」研究課題番号: 21720303) の一部を使用した。

参考文献

- 有馬貴之 (2010): 動物園来園者の空間利用とその属性—上野動物園と多摩動物公園の比較, 地理学評論, 83, 353-374.
- 橋本俊哉 (1991): 観光者の「徒歩行動」に関する研究—今後の課題と方向性—, 観光研究, 4, 11-20.
- 古谷知之 (2006): 携帯型位置情報端末を用いた観光行動動態の時空間データマイニング—箱根地域を事例として—, 都市計画論文集, 41, 1-6.
- 矢部直人・有馬貴之・岡村 祐・角野貴信 (2010): GPS を用いた観光行動調査の課題と分析手法の検討, 観光科学研究, 3, 17-30.
- Shoval, N. and M. Isaacson., 2007a. Tracking tourists in the digital age. *Annals of Tourism Research*, 34, 141-159.
- Shoval, N. and M. Isaacson., 2007b. Sequence Alignment as a Method for Human Activity Analysis in Space and Time. *Annals of the Association of American Geographers*, 97, 282-297.
- Wilson, W. C., 1998. Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A*, 30, 1017-1038.
- Wilson, C., 2006. Reliability of sequence-alignment analysis of social processes: Monte Carlo tests of ClustalG software. *Environment and Planning A*, 38, 187-204.
- Wilson, C., 2008. Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation*, 35, 485-499.