

時空間的近接関係を考慮した 滞在行動の符号化に基づく 移動履歴分析手法の研究

塚原 元英¹・井上 亮²

¹非会員 東京電力ホールディングス株式会社
E-mail:tmkthrd@yahoo.co.jp

²正会員 東北大学准教授 大学院情報科学研究科 人間社会情報科学専攻
(〒980-8579 宮城県仙台市青葉区荒巻字青葉6-6-06)
E-mail:rinoue@tohoku.ac.jp

近年、広域の詳細な移動履歴を大量に取得できる環境が整い、このデータを活用した観光周遊行動把握を通して観光サービスの向上を図ることが期待されている。移動履歴から滞在場所・時間帯など滞在行動を抽出して作成した滞在行動系列データを用いて、周遊行動を分析する研究は行われているが、広域の周遊行動分析では不可欠の、類似度の高い滞在行動の集約方法は検討されていない。本研究は、データをコンパクトに圧縮する符号化方法であるハフマン符号化を援用し、滞在行動の時空間上類似度を考慮した分類法を提案する。この分類結果に基づく滞在行動の符号化を通して周遊行動を階層的に分類し、典型的な周遊行動の抽出を試みる。提案手法を沖縄本島の移動履歴に適用した結果、滞在・周遊行動は階層的に分類できたが、周遊行動が多様なため、典型的な周遊行動抽出には課題を有することを確認した。

Key Words : trajectory data, spatio-temporal clustering, Huffman coding.

1. はじめに

測位機能付き携帯情報端末の普及に伴い、位置情報を用いたサービスが広がりを見せている。現在地周辺や目的地への経路を確認できる地図サービスに加え、店舗の訪問を記録するチェックインサービス、ランニングなどの移動距離・時間など運動履歴を記録する健康情報サービス、実際の街を舞台に陣地を取り合う位置情報ゲームサービスなど、多くの企業が様々な位置情報サービスを提供している。この位置情報サービスの利用者の増加に伴い、多くの人の詳細な移動履歴を大量に収集できるようになっており、この豊かつ詳細な移動履歴データを用いた分析が注目を浴びている。データ利用先の一つは観光周遊行動の把握で、特に訪日外国人に向けた観光サービスの向上へと活用されることが期待されている¹⁾。

従来、観光周遊行動の分析は、アンケート調査や測位機能付き端末の貸与による調査に基づくものが主流で、調査件数や対象範囲、周遊行動観測の時空間解像度に限りがあった。しかし、移動履歴データを活用すると、大勢の旅行者が取った広域の周遊行動を、詳細な空間・時間解像度で分析することが可能である。

移動履歴データを利用して観光周遊行動の分析を目指

した研究は、これまで数多く行われている、しかし、詳しくは後述するが、広域で取られた詳細な移動履歴データから、時間・空間解像度の高い分析を行う方法については議論が十分に行われていないと筆者らは考える。

ここで、広域の対象領域内の周遊行動を、詳細な空間・時間解像度で分析する場合に生じる問題を指摘する。広域の対象領域には、旅行者が訪問・滞在しうる場所は数多く存在し、また、各旅行者の各滞在地の到着時刻・出発時刻も多様である。そのため、滞在場所・時間帯を基に滞在行動を分類し、類似の滞在行動を集約して周遊行動を表現しなければ、周遊行動の取り得るパターン数が膨大になるため、類似の周遊行動を取る旅行者を発見できない。詳細に記録された広域の移動履歴を元に周遊行動を分析するには、類似の滞在行動を集約するための分類方法の検討は重要な課題である。

さて、類似の滞在行動を集約する場合、その基準となる分析の解像度は、空間は地域単位から地点・施設単位まで、時間は午前・午後の区分から10分単位までなど、幅広い設定が考えられる。しかし、異なる設定によって一貫性のない滞在行動分類・符号化や周遊行動分類結果が得られる分析手法は、結果の解釈に支障を来し有用性が乏しい。分析手法は、滞在行動分類の空間・時間解像

度設定に対して周遊行動を階層的に分類できること、更に、分類結果の良し悪しを判断する合理的な基準を示せることが重要だと筆者らは考える。

そこで本研究は、情報符号化手法の一つで、コンパクト符号を構成するハフマン符号化に滞在行動分類の範を求め、分析の時間・空間解像度設定に対応して、滞在地点・開始/終了時刻からなる滞在行動を階層的に分類し、その分類結果に基づく滞在行動の符号化を利用した周遊行動分類手法について提案する。その後、沖縄本島で取得された旅行者の移動履歴データを用いて、提案手法の適用可能性を確認する。

2. 先行研究における滞在地点・滞在行動の分類

移動履歴データを利用して観光周遊行動の分析を目指す先行研究の多くは、それぞれの旅行者が訪れた場所やその時間帯や滞在時間などの「滞在行動」を移動履歴データから抽出して符号化し、周遊行動を滞在行動の符号系列と表した上で分析するアプローチを取っている。

このうち、後半の周遊行動分析に関しては、滞在行動の符号系列を分析する様々な手法が提案されている。滞在场所間の遷移確率を利用した典型的な周遊順路の抽出^{2,3}や、滞在地訪問の有無に基づく周遊行動の分類⁴、データマイニング手法の1つである系列パターンマイニング⁵を用いた周遊行動の部分系列の抽出^{6,7}、遺伝子配列の類似度評価手法⁸を適用した周遊行動の分類^{9,10,11,12}が提案されている。

一方、前半の移動履歴データから各旅行者の滞在场所・時間帯を抽出し、滞在行動を符号化する方法に関しては、これまで多くの検討は行われていない。大半の先行研究では、分析者が事前に地域区分を定め、旅行者の設定地域内への進入や一定時間以上の滞留をもって滞在与判定し滞在行動を符号化しているが、地域区分設定に関する議論は見られない。その理由は、公園⁴や動物園^{11,12}、世界遺産の旧市街⁸など、狭い領域内の周遊行動を分析する場合、旅行者が滞在する地点が限られ地域区分設定に自由度が少ないこと、また、市町村単位²や地域単位⁹などの空間単位に基づく分析を前提していることが原因と考えられる。

分析者の主観で滞在地域区分を定めず、移動履歴データを基に滞在行動を分類する研究として、Giannotti et al. (2007)⁹や西野ら (2008)⁷は、滞在地点密度の高い領域を見つけ、その領域内の滞在地点を集約して符号化する方法を提案している。しかし、広域を対象とした分析では滞在地点が分散し、高密度の領域が見つからないため、細分化された滞在地点しか設定できない。集約基準となる密度の閾値を、変更して調整することは可能だが、階層

的な滞在场所の分類を構成できるとは限らず、広域の周遊行動分析への適用には限界がある。

また、Zhang et al. (2009)³は、滞在场所の階層的分類手法を提案している。滞在地点密度に基づく滞在地点分類を行った後、各分類内で同様の分類を繰り返して階層構造を持った分類を作成する。しかし、各分類に使用する密度の閾値設定基準が示されておらず、合理的な分類を行う方法が明らかではない。

以上のように、移動履歴データに基づいて滞在行動を分類する方法は検討されているが、その分類基準の設定方法については十分に議論されていない。

3. 周遊行動・滞在行動の階層的分類手法の提案

本章では、データ量を圧縮する符号化手法の一つであるハフマン符号化を概説した後、ハフマン符号化を援用した滞在行動の階層的分類・符号化、および、その結果を利用した周遊行動の分類方法を提案する。

なお、本研究では、各個人の移動履歴データから停留点を探索し、{滞在地点座標, 滞在開始時刻, 滞在終了時刻}を抽出したデータを「滞在行動」と定義し、この階層的な分類を考える。

(1) ハフマン符号化

データ要素集合 $D = \{d_1, d_2, \dots, d_n\}$ で構成され、各データ要素の出現頻度割合は重み $W = \{w_1, w_2, \dots, w_n\}$ であるデータの符号化を考える。短い符号でデータを表現できれば、データ量が減少し処理速度の向上が期待できるため、可能な限り短い符号長となる符号化が望ましい。ここで、データを二進数の符号 $C(D, W) = \{c_1, c_2, \dots, c_n\}$ を用いて符号化するとしよう。

符号化によるデータ圧縮度の評価は、一般に平均符号長 $L(C)$ を用いる。 d_i に設定する符号 c_i の符号長 $|c_i|$ を l_i と記すと、平均符号長は式(1)で書ける。

$$L(C) = \sum_{i=1}^n w_i l_i \quad (1)$$

この平均符号長は、情報理論における平均情報量(エントロピー)との対応関係を有している。このデータが有する平均情報量は式(2)で表される。

$$H(D) = -\sum_{i=1}^n w_i \log_2 w_i \quad (2)$$

もし全データ要素 s_i に対し、符号長 $l_i = -\log_2 w_i$ の符号を設定できれば、平均符号長は平均情報量と一致し、データの情報量を最もコンパクトに表現した符号化となる。しかし、通常、全データ要素に対して $-\log_2 w_i$ は整数とならず、各データ要素を表現するにはより長い符号が必要となるため、

$$L(C) \geq H(D) \quad (3)$$

となり、平均情報量は平均符号長の下限を表す。

このように、データ要素 d_i に符号長が $-\log_2 w_i$ に近い符号を与えられると、データ量の少ないコンパクトな符号化を達成できる。ハフマン符号化は、以下に示す単純なアルゴリズムで、コンパクトな符号化を達成する。

まず、出現頻度の昇順にデータ要素を並べたリストを作成する。次に、頻度の小さい2要素を選びリストから削除し、選んだ2要素を集約した要素集合を作成する。その後、作成した要素集合をデータ要素リストに追加し、リストを出現頻度の昇順に並べる。この一連の操作を全要素が一つの集合に含まれるまで繰り返し、ハフマン木と呼ばれる二分木を構築する。ハフマン木のノードから出る左右の辺に符号0もしくは1を付け、葉ノードに対応したデータ要素に、根ノードからのパスに付けられた符号を設定したものが、ハフマン符号となる。

なお、データ要素に対応する符号を読むと、次の符号を読まずに元データに復号できる性質を持つ符号を、瞬時符号と呼ぶが、ハフマン符号は瞬時符号の中で平均符号長が最短の符号となることが知られている¹³⁾。

(2) 滞在行動の階層的分類手法の提案

滞在行動にハフマン符号化のアルゴリズムを適用すると、滞在行動が有する情報量を最もコンパクトに表現する基準で、滞在行動を階層的に分類し符号化できる。しかし、その単純な適用では、滞在行動の出現頻度のみが考慮され、滞在行動の時間的・空間的な類似性が分類に反映されないため、滞在行動分類として適切ではない。

そこで、滞在场所や時間帯の類似度が高い滞在行動のみを集約する制約を課した上で、ハフマン符号化の手順を用いた滞在行動分類法を提案する。

まず、滞在地点と滞在開始・終了時刻を基に滞在行動間の類似性を定義する。滞在行動 i, j の滞在地点間ユークリッド距離を d_{ij} 、滞在行動 i の開始時刻を st_i 、終了時刻を et_i 、空間距離と時刻差の重み付けパラメータを α ($0 \leq \alpha \leq 1$)と表し、類似度指標 r_{ij} を式(4)で定義する。

$$r_{ij} = \alpha d_{ij} + (1 - \alpha) \{|st_i - st_j| + |et_i - et_j|\} \quad (4)$$

次に、類似度指標を用いて相対近傍グラフを作成し、類似滞在行動の対を求める。相対近傍グラフは、式(5)の条件を満たす場合に、点 i, j が近傍関係にあると判断し、リンクを張るグラフである。

$$r_{ij} \leq \max\{r_{ik}, r_{jk}\} \quad \forall k \neq i, j \quad (5)$$

なお、相対近傍グラフは距離の公理を満たさない指標に対しても適用できるが、滞在行動間の類似性指標 r_{ij} は距離の公理を満たす。

本研究では、上記の手順から得られた類似滞在行動対のみを集約することとし、出現頻度の和が最小の類似滞在行動対を集約する操作を繰り返して作成した二分木を用いた、滞在行動の符号化を提案する。なお、出現頻度の和が最小の対が複数ある場合は、類似度指標最小の対

を優先して集約する。また、滞在行動集合間の類似性は、その要素の類似性で評価する。二分木の各ノードは滞在行動の分類を表し、根ノードから葉ノードに向かうと詳細な時間・空間解像度に基づく分類となる。

なお、この滞在行動分類は、空間距離と時刻差の重み付けパラメータ α に依存する。本研究は、滞在行動データが持つ情報を最もコンパクトに表現できる符号化が最適と考え、平均符号長が最小となる α を選ぶ。滞在行動の類似性を考慮した符号化による平均符号長は、出現頻度のみによるハフマン符号化による平均符号長よりも長い。この差が滞在行動の類似性評価に関する情報量を表すと解釈できる。

(3) 滞在行動の階層的分類に基づく周遊行動分類

滞在行動分類を表す二分木を用いて、時間・空間範囲が広いノードから順に下位階層ノードに分割し、分類毎に滞在行動を符号化して各旅行者の周遊行動を表すと、階層的に周遊行動を分類できる。図-1に例を示す。

図-1の(a)(b)は、(i)が滞在行動分類を示す二分木、(ii)が周遊行動分類を示す木である。太線で囲まれたノードは、各段階の分類を表す。

図-1(a)(i)は、滞在行動分類の根ノードを分割し、滞在行動を2分類した状態を表す。各旅行者の周遊行動をこの滞在行動分類に基づき符号化すると、3分類できる(図-1(a)(ii))。次に、滞在行動分類から最も時間・空間範囲が広い分類を選択し、その分類を1階層下のノードに移動して2分割する。図-1(b)(i)は滞在行動分類1を10・11に分割する例を表す。その分類に合わせて周遊行動分類を細分化する(図-1(b)(ii))と、滞在行動の階層的な分類と対応した、周遊行動の階層的な分類が得られる。

このように、時間・空間的範囲の大きな滞在行動分類から順に細分化し、各分類に基づき周遊行動を分類する

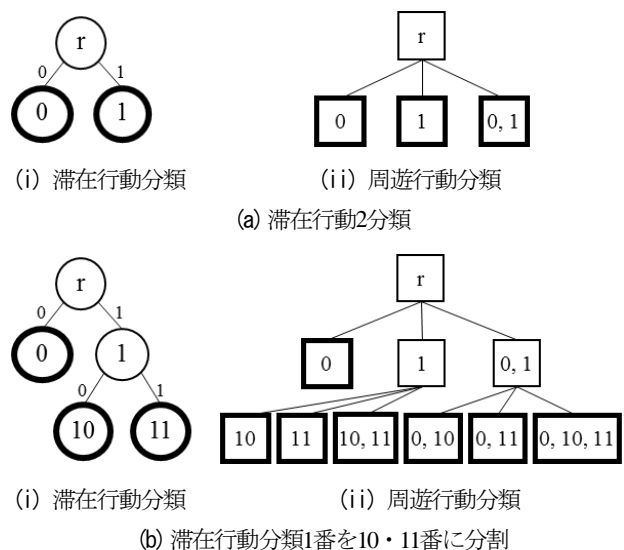


図-1 滞在行動の階層的な分類に基づく周遊行動分類例

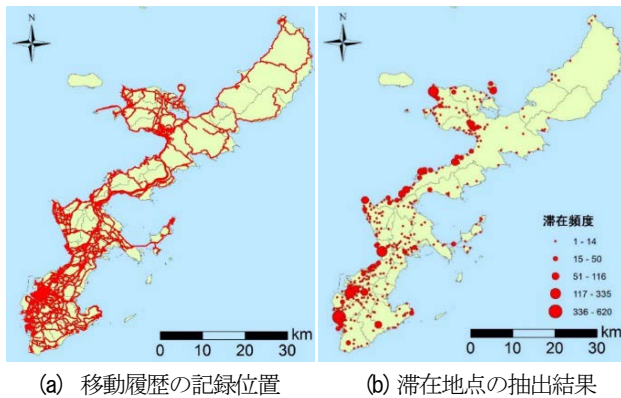


図-2 滞在地点の抽出

と、滞在行動分類と周遊行動分類が対応した階層的な分類を行うことができる。

4. 沖縄本島の観光周遊行動分析への適用例

(1) 分析データと滞在地点の抽出

日本道路交通情報センターの調査で取得された、沖縄本島のレンタカー移動履歴データを用いて、提案手法の適用可能性を検討する。

移動履歴データは、2014年8月29日から2014年12月1日に沖縄を訪れた614組の旅行者の2泊3日から5泊6日までの旅程の移動を車載GPSで取得し、1~5秒間隔の時刻・位置座標を記録している。図-2(a)は、記録された全ての位置を示す。本分析では、各旅行者の移動履歴を一日単位に分割し、一日の周遊行動を分析する。

提案手法の適用に先立ち、各旅行者の移動履歴データから滞在地点と滞在開始・終了時刻を抽出し、滞在行動の系列データを作成する。まず、本分析では15分以上・半径100m以内に留まる場合を停留と定義し、停留時間内の位置の重心を停留点とした。なお、日を超えた停留は停留開始日に含める。一日の停留点が3箇所を満たない旅行者のデータは以後の分析から除外し、4,167箇所の停留、823日の周遊行動データを得た。

この段階では、各旅行者の停留点は座標で示され観光地とは対応付けられておらず、異なる旅行者が同じ観光地を訪れたことを把握できない。そこで、同じ観光地での滞在を表すデータを作成するため、距離の近い停留点を集約する。停留点について全域最小木を作成し、距離300m以内のリンクで繋がれる一連の停留点を集約し、その重心を滞在地点の位置とする。図-2(b)は集約した滞在地点、計399地点を表し、円の中心が滞在地点位置、円の大きさは滞在頻度を表す。滞在頻度が高い地点は、那覇空港近くのレンタカー営業所周辺(頻度620)、那覇市中心部(頻度335)、北谷(頻度374)、美ら海水族館周辺(頻度198)である。

また、本分析では、滞在開始・終了時刻の分析最小単位を10分に設定した。この滞在地点・滞在時刻に関する設定に基づいて、滞在地点と滞在開始・終了時刻が同じものを同じ滞在行动として整理すると、滞在行动の総数は3,714である。

次節では、上記の手順で作成した1日単位の滞在行动系列を用いて、提案手法を用いた周遊行動分類を行う。

(2) 提案手法による滞在・周遊行動の分類

まず、分析に先立ち、空間距離と時刻差の重み付けパラメータ α の設定を行う。 α を0から1まで0.05単位で変化させ、提案手法を用いて滞在行动分類・符号化を行い、各設定での滞在行动の平均符号長を確認した(図-3)。 $\alpha=0.55$ のときに平均符号長が最小となったため、この値を用いて以後の分析を行う。

滞在行动分類の二分木を図-4に、この分類に基づく滞在行动分類数と周遊行動分類数の関係、および、滞在行动分類の最大頻度の関係を図-5に示す。滞在行动30分類の場合、823日分の周遊行動は692分類され、滞在行动15分類でも、周遊行動は388分類になる。このように、広域の観光周遊行動は多様であることが確認できる。移動履歴のデータ件数が大きいと特徴的な周遊行動を抽出できる可能性はあるが、本研究で使用した千件弱の移動履歴データでは、広域の周遊行動の分析を行う上で限界があることが確認された。

ここで、滞在行动分類結果を確認する。28分類時の二分木を図-6に示し、滞在行动分類結果の一部を表-1に、また、156分類の結果の一部を表-2に示す。滞在地点分布や到着・出発時刻分布を見ると、それぞれ特徴のある分類がされていることが確認できる。表-2の分類0100と分類0101は表-1の分類010から、表-2の分類001110と分類001111は表-1の分類0011から、それぞれ分割されたものである。例えば、分類010は午後の南部滞在を表すが、分類0100は夕方方の南部滞在、分類0101は正午付近に最南端部に滞在する行動を表しており、分類010の滞在行动をより詳細に分類した結果を表す。

次に、滞在行动分類・符号化を用いて周遊行動分類を行った例を確認する。滞在行动28分類に対応した分類を表-3に、滞在行动156分類に対応した分類を表-4に示す。表-4の{0101,0110}と{0100,0101,0110}は、表-3の{010,0110}と階層構造を有し、滞在地点分類010が分類0100と分類0101に細分化されたことによる結果である。

以上のように、提案手法は滞在行动の階層的な分類・符号化を基に、周遊行動を階層的に分類できることは確認できた。しかし、今回使用したデータでは、データ件数に対して周遊行動のパターンが多いため、周遊行動は細かく分類され、高い頻度で生じる周遊行動を発見できなかった。なお、本手法は、相対近傍グラフに基づく類似

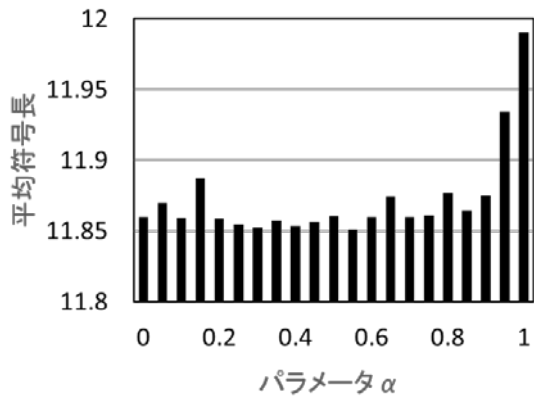


図-3 平均符号長の違い

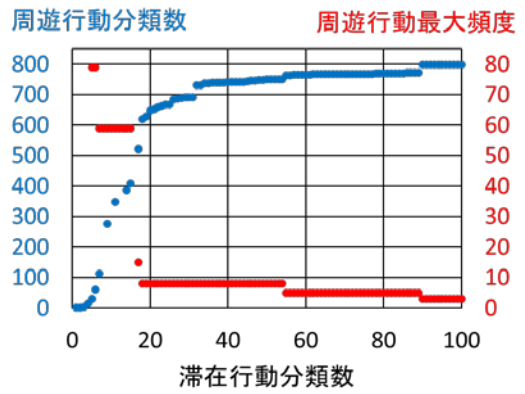


図-5 滞在行動分類数と周遊行動分類数の関係

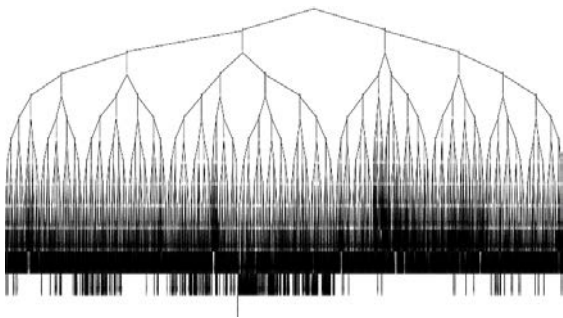


図-4 滞在行動分類の二分木 ($\alpha = 0.55$)

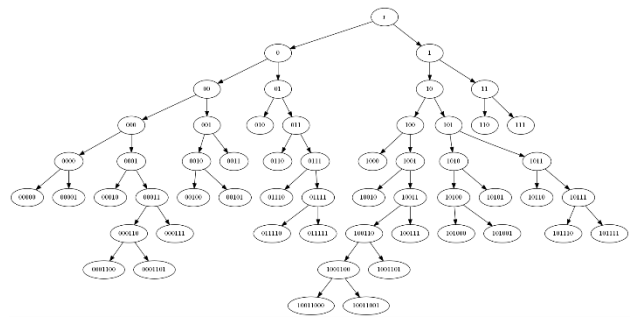


図-6 滞在行動分類 (28分類) の二分木 ($\alpha = 0.55$)

表-1 滞在行動の28分類 (一部)

符号	010	0110	0011	00100
空間分布				
頻度	414	229	215	101
到着時刻	13:48 (2:27)	11:18 (0:50)	16:06 (1:30)	10:57 (3:24)
出発時刻	15:16 (2:20)	12:02 (0:52)	17:05 (1:40)	12:26 (2:23)

表-2 滞在行動の156分類 (一部)

符号	0100	0101	001110	001111
空間分布				
頻度	210	204	56	56
到着時刻	15:50 (1:28)	11:41 (1:09)	17:10 (0:48)	17:23 (0:52)
出発時刻	17:18 (1:13)	13:10 (0:55)	18:13 (0:38)	18:46 (0:55)

表-3 周遊行動分類 頻度上位 5 位 (滞在行動 28 分類時)

滞在行動	頻度
010, 0110, 111	8
010, 111	6
0110, 111	5
010, 110, 111	5
010, 0110	5

表-4 周遊行動分類 頻度上位 (滞在行動 156 分類時)

滞在行動	頻度
0110, 1111	3
0101, 0110	3
頻度2 の分類 計15分類	
0100, 0101, 0110	2
0100, 0101, 111011	2
0100, 0101, 111001	2
...	

滞在行動対探索が $O(n^3)$ の計算量を要するが、比較的大規模なデータにも適用可能と考えられる。大規模なデータへの適用を通して、分析手法の有効性を検討する必要があると筆者らは考える。

5. 結論

本研究は、広域で取得された詳細な移動履歴データから周遊行動を分類分析における、滞在行動分類・符号化の重要性に着目し、時空間上類似性を考慮した滞在行動のコンパクトな符号化を通して滞在行動・周遊行動の階層的分類を行う方法を提案した。提案手法を、沖縄本島内のレンタカー移動履歴データに対して適用し、滞在行動の符号化結果を利用して周遊行動を階層的に分類できることを確認した。ただし、本分析では、周遊行動は細分化され、典型的な周遊行動を発見するには至らなかった。より大規模な移動履歴データへの適用を通して、提案手法の有効性を確認することが今後の課題である。

謝辞：分析に使用した旅行者の移動履歴データは、公益財団法人日本道路交通情報センターより提供いただいた。また、本研究は、JSPS 科研費 24241053 および 25249069 の助成を受けた。

参考文献

- 1) 例えば、観光庁. 2014. 携帯電話から得られる位置情報等を活用した訪日外国人動態調査 報告書. <http://www.mlit.go.jp/kankoch/shisaku/kanko-chi/gps.html> (2016年4月11日閲覧).
- 2) 長尾 光悦, 川村 秀憲, 山本 雅人, 大内 東. 2005. 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. ICS-140, No. 5, pp.23-28.
- 3) Zheng, Y., Zheng, L., Xie, X., and Wei, M. 2009. Mining interesting locations and travel sequence from GPS trajectories. *International conference on World Wide Web*, pp. 791-800.
- 4) 山本 泰裕, 伊藤 弘, 小野 良平, 下村 彰男. 2006. GPS を用いた新宿御苑における利用者の行動パターンに関する研究. *ランドスケープ研究*, Vol. 69, No. 5, pp.601-604.
- 5) Agrawal, R. and Srikant, R. 1995. Mining sequential patterns. *Proceedings of the Eleventh International Conference on Data Engineering*, pp. 3-14.
- 6) Giannotti, F., Nanni, M., Pinelli, F., and Pedreschi, D. 2007. Trajectory pattern mining. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 330-339.
- 7) 西野 正彬, 瀬古 俊一, 青木 政勝, 山田 智広, 武藤 伸洋, 阿部 匡伸. 2008. 滞在地遷移情報からの行動パターン抽出方式の検討. 情報処理学会研究報告ユビキタスコンピューティングシステム (UBI), Vol. 110, No. 2008-UBI-020, pp. 57-64.
- 8) Needleman, S.B. and Wunsch, C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, Vol. 48, No. 3, pp. 443-53.
- 9) Shoval, N. and Isaacson, M. 2007. Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American Geographers*, Vol. 97, pp. 282-297.
- 10) Shoval, N., McKercher, B., Birenboim, A., and Ng, E. 2015. The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. *Environment and Planning B*, Vol.42, No.1, pp.76-94.
- 11) 矢部 直人, 有馬 貴之, 岡村 祐, 角野 貴信. 2010. GPS を用いた間行動調査の課題と分析手法の検討. *観光科学研究*, No. 3, pp. 17-30.
- 12) 川瀬 純也, 伊藤 史子. 2015. 配列解析による上野動物園来園者の時空間行動類型化～閉じた空間内での回遊行動に着目して～. 地理情報システム学会講演論文集, D-7-2, CD-ROM.
- 13) 例えば, ジョーンズ, G.A., ジョーンズ, J. M. 2006. 情報理論と符号理論. 一樂 重雄, 河原 正治, 河原 雅子訳, シュプリンガー・ジャパン,

TRAJECTORY DATA ANALYSIS BASED ON THE CODING OF VISIT BEHAVIOR CONSIDERING SPACE-TIME PROXIMITY

Motohide TSUKAHARA and Ryo INOUE