

時空間上の近傍関係に基づく滞在の符号化を利用した移動履歴分類の提案

井上 亮・塚原元英

Trajectory Classification Based on Coding of Stays

Considering Spatio-Temporal Proximity

Ryo INOUE and Motohide TSUKAHARA

Abstract: 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じた観光サービス向上が期待されている。移動履歴から滞在場所・時間帯など滞在を抽出したデータを用いて、周遊分析を行う研究は行われているが、広域の周遊分析では不可欠の、類似度の高い滞在を集約し分析する方法は検討されていない。本研究は、データの情報量を表すエントロピーに最も近いコンパクトな符号化を作成するハフマン符号に着目し、滞在の時空間上の類似度を考慮した分類を提案し、この結果を用いて典型的な周遊の抽出を試みる。提案手法を沖縄本島の移動履歴に適用した結果、滞在・周遊は階層的に分類できたが、周遊行動の多様さのため、典型的な周遊の把握には課題を有することを確認した。

Keywords: 移動履歴 (trajectory), 分類 (classification), ハフマン符号 (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、詳細かつ豊富な移動履歴を収集可能な環境が整っており、その移動履歴を活用した分析が注目を浴びている。応用先の一つは観光周遊行動の把握で、例えば、訪日外国人に対する観光サービスの向上への活用などが期待されている（例えば、観光庁、2014）。

移動履歴に基づく周遊分析については、これまでに多く検討されてきたが、広域かつ詳細な移動履歴を用いた高い時空間解像度を有する分析を行う方法は議論が十分にされていない。広領域には、旅行者が訪問しうる場所は多く存在し、各場所の滞在時間帯もまた多様である。そのため、滞在場所や時間帯を基に滞在を分類し、類似の滞在を集約しなければ、取り得る周遊パターン数は膨大となり、類似の周遊行動を見つけられないという課題を有する。

さて、滞在分類の解像度は、空間は地域単位から地点・施設単位まで、時間は午前・午後の区分から分単位までなど、様々な設定が考えられる。しかし、異なる設定によって一貫性のない滞在や周遊の分類結果が得られる手法は、結果の解釈に支障を来すため、実用性に欠けるため、分析の解像度設定に対

して階層的な分類結果をもたらす手法が望ましいであろう。また、分類結果の妥当性を判断できる合理的な基準を示すことも重要である。

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分類の範を求める、分析の時間・空間解像度設定に対応して、滞在地点・開始/終了時刻で表される「滞在」を階層的に分類し、その分類に基づいて周遊行動を分類する手法を提案する。その後、沖縄本島の移動履歴を用いて、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分類

移動履歴を基に周遊分析を行う先行研究の多くは、まず旅行者の滞在を抽出して符号化した滞在系列を用いて周遊を分析するアプローチを取る。

後半の滞在系列分析については、多くの手法が提案されている。滞在場所間の遷移確率に基づく典型的周遊順路の抽出(長尾ら, 2005; Zheng et al., 2009)や、滞在地訪問の有無に基づく周遊分類(山本ら, 2006), 系列パターンマイニングによる周遊の部分系列抽出(Giannotti et al., 2007; 西野ら, 2008), 遺伝子配列の類似度評価手法(Needleman and Wunsch, 1970)による周遊分類(Shoval and Isaacson, 2007; 矢部ら, 2010; Shoval et al., 2015; 川瀬・伊藤, 2015)が提案されている。

しかし、前半の、移動履歴から滞在を抽出して符

井上 亮 〒980-8579 仙台市青葉区荒巻字青葉 6-6-06

東北大学 大学院情報科学研究科

Phone: 022-795-7478 E-mail: rinoue@tohoku.ac.jp

号化する方法については十分に検討されていない。大半の研究は、事前設定の地域区分を基に滞在を判定し、地域区分設定の是非は議論していない。その理由は、公園(山本ら, 2006)や動物園(矢部ら, 2010; 川瀬・伊藤, 2015), 世界遺産の旧市街(Shoval and Isaacson, 2007)など比較的狭い領域内の周遊を対象とし、旅行者の滞在地点が限られ地域区分設定に自由度が少ないとことや、市町村単位(長尾ら, 2005)や地域単位(Shoval et al., 2015)などの空間単位に基づく分析を前提していることにあると考えられる。

Giannotti et al. (2007) や西野ら(2008)は、滞在地点密度が高い領域を探索し地域区分設定を行う方法を提案している。しかし、広域の分析では滞在地点が分散するため、細分化された地域区分設定となる。また、Zheng et al. (2009)は、滞在地点密度に基づく分類を繰り返して滞在の階層的分類を行う手法を提案しているが、分類に使用する密度の閾値設定基準は示されていない。

3. 周遊・滞在の階層的分類の提案

本研究では、各個人の移動履歴から停留点を探索し、滞在地点座標、到着・出発時刻を抽出したデータを「滞在」とし、その分類を通して周遊行動を分析する。その際、データの情報量を表すエントロピー(平均情報量)に最も近いコンパクト符号を構成できるハフマン符号化(Huffman, 1952)を援用して滞在を階層的に分類する方法を提案し、得られた滞在符号系列を用いた周遊分類を検討する。

3.1 滞在の階層的分類

滞在データに対しハフマン符号化を適用すると、データの平均情報量に最も近い、コンパクトな符号を構成するという基準で、滞在を二分木で階層的に分類できる。しかし、単純な適用では、滞在の出現頻度のみが考慮され、滞在の時空間上の類似度が評価されないため、適切な滞在分類とは言えない。そこで、滞在場所や滞在時間帯の類似度が高い滞在のみを集約する制約を課した分類を提案する。

まず、滞在地点と到着・出発時刻を基に滞在の類似性を定義する。滞在 i, j の地点間ユークリッド距離を d_{ij} 、滞在 i の到着・出発時刻をそれぞれ at_i, dt_i 、空間距離と時刻差の重みパラメータを α ($0 < \alpha < 1$) と表し、類似度指標 p_{ij} を式(1)で定義する。

$$p_{ij} = \alpha d_{ij} + (1 - \alpha) \{ |at_i - at_j| + |dt_i - dt_j| \} \quad (1)$$

次に、この類似度指標を用いて相対近傍グラフを作成し、類似滞在対を得る。

$$p_{ij} \leq \max\{p_{ik}, p_{jk}\} \quad \forall k \neq i, j \quad (2)$$

本研究は、頻度の和が最小の類似滞在対を集約する操作を繰り返して二分木を作成する過程を通して、滞在の分類・符号化を行うことを提案する。なお、頻度の和が最小の対が複数ある場合は、類似度指標の昇順に集約する。また、滞在集合間の類似性は、その要素の類似性で評価する。二分木の各ノードは滞在分類を表し、根から葉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、分類結果はパラメータ α に依存する。本研究は、滞在データの平均情報量に近い分類・符号化が最適だと考え、平均符号長最小となる α を選ぶ。この平均符号長は、頻度のみ考慮したハフマン符号よりも長いが、この差は滞在の類似性考慮により追加された情報量を表すと解釈できる。

3.2 滞在の階層的分類に基づく周遊の階層的分類

滞在分類の二分木を用い、カバーする時空間範囲が広い滞在分類から順に分割して周遊を表すと、周遊を階層分類できる(図-1)。図-1の(a)(b)は、(i)が滞在分類を表す二分木、(ii)が周遊分類を表す木である。太線で囲まれたノードは、各段階の分類を表す。

滞在2分類(図-1(a)(i))は、周遊を3分類する(図-1(a)(ii))。次に、時空間範囲が最も広い滞在分類を2分割する(図-1(b)(i))と、滞在の階層分類と対応した周遊の階層分類が得られる(図-1(b)(ii))。

時間・空間的範囲の大きな滞在分類から順に細分化し、各分類に基づき周遊分類を行うと、滞在分類と周遊分類が対応した階層的分類ができる。

4. 沖縄本島の観光周遊分析への適用例

4.1 データと滞在地点の抽出

日本道路交通情報センターが取得した、沖縄本島

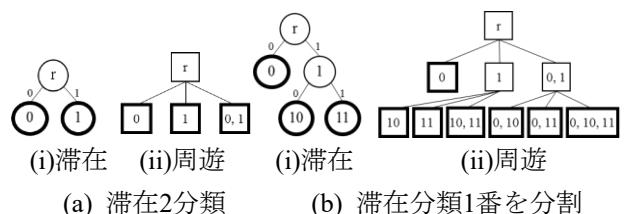


図-1 滞在の階層的分類に基づく周遊分類

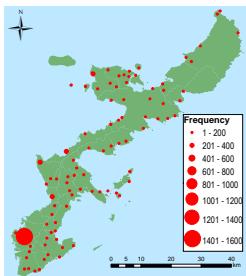


図-2 滞在地点抽出

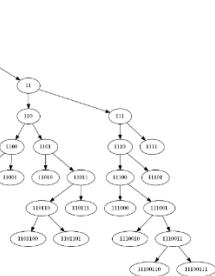


図-3 滞在 20 分類時の二分木

の 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 日の旅程のレンタカー移動履歴を用い、提案手法の適用可能性を検討する。なお、データは車載 GPS による 1 ~5 秒間隔の時刻・位置座標を記録する。本分析では、各旅行者の移動履歴を 1 日単位に分析する。

はじめに、各旅行者の移動履歴から滞在地点と到着・出発時刻を抽出し、滞在系列を作成する。まず、30 分以上・半径 1km 以内の停留を抽出し、その時間内の位置重心を停留点とした。なお、日をまたぐ停留は停留開始日に含めた。1 日に 2 箇所以上の停留点がある周遊を分析対象とし、3,917 箇所の停留、992 日の周遊を得た。この段階では、各旅行者の停

留点は座標で示されており、異なる旅行者による同じ観光地の訪問を把握できないため、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 1km 以内の辺で繋がれる一連の停留点を集約し、その重心を滞在地点とする。図-2 は集約した滞在 90 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。また、到着・出発時刻の分析単位を 1 時間単位に設定した。上記設定に基づき、滞在地点と開始・終了時刻が同一の滞在を集約すると、滞在の総数は 1,001 であった。

4.2 提案手法による滞在・周遊の分類

空間距離と時刻差の重みパラメータ α は、 $\alpha=0.05$ 刻みで変化させ滞在の平均符号長を確認し、 $\alpha=0.1$ で平均符号長が最小となったため、この値を用いる。

滞在 20 分類では、992 日の周遊は 515 分類された。20 分類時の二分木を図-3 に示し、滞在分類の頻度上位 4 位までを表-1 に示す。広域の観光周遊は多様であることが確認できる。移動履歴の件数が大きいと特徴的な周遊を抽出できる可能性はあるが、本研究で使用した 1000 件弱の移動履歴では、広域周遊の分析を行う上で限界があることが確認された。

表-1 滞在 20 分類時の頻度上位 5 位の滞在分類

分類	01	100	000	1010
位置				
頻度	810	560	456	247
到着時刻 (標準偏差)	13:48 (3:23)	14:01 (2:23)	7:55 (4:22)	17:45 (1:57)
出発時刻 (標準偏差)	15:44 (3:45)	15:57 (2:02)	11:16 (1:53)	20:19 (2:23)

表-2 滞在40分類時の滞在「01」の下部分類

分類	011	0100	01010	01011
位置				
頻度	386	208	110	106
到着時刻 (標準偏差)	11:14 (0:48)	17:17 (3:31)	13:29 (0:46)	16:36 (1:59)
出発時刻 (標準偏差)	12:41 (0:41)	20:13 (2:47)	14:44 (0:51)	19:06 (2:07)

表-3 周遊分類 頻度上位 5 位 (滞在 20 分類時)

滞在	頻度
01, 100	58
000, 01	47
000, 01, 100	28
000, 100	24
01, 100, 1010	22

表-4 周遊分類 頻度上位 5 位 (滞在 40 分類時)

滞在	頻度
000, 100	24
000, 011	22
011, 100	19
0100, 011, 100	11
000, 0100, 100	11

表-1 の地点や到着・出発時刻の分布を見ると、南部・中部における滞在が、主に時間帯によって分類されたことが確認できる。表-2 は、表-1 の分類 01 を 4 分類した結果を示しており、より詳細な滞在分類が得られている。

次に、周遊分類について、滞在 20 分類時の上位 5 位を表-3 に、滞在 40 分類時の結果を表-4 に示す。表-4 の {011, 100} {0100, 011, 100} は、表-3 の {01, 100} と階層構造を有し、より細く分類されている様子が確認できる。

以上のように、提案手法は滞在の階層的分類を基に、周遊を階層的に分類できることは確認できた。しかし、今回分析に用いたデータでは、周遊データ件数に対して、抽出された周遊パターン数が極めて多いため、周遊が細かく分類され、高い頻度で生じる特徴的な周遊を発見することはできなかった。

5. おわりに

本研究は、広域で取得された詳細な移動履歴から周遊分類を行う分析における、滞在分類の重要性に着目し、滞在の時空間上の類似性を考慮したハフマン符号化を援用した符号化を通して、滞在・周遊の階層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・周遊の階層的分類が可能であることを確認した。ただし、本分析では、周遊は細分化され、典型的な周遊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

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・25249069 の助成を受けた。

参考文献

- 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*, 330–339.
- Huffman A. (1952) A method for the construction of minimum-redundancy codes. *Proceedings of the I.R.E.*, 1098–1102.
- 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*, **48**(3): 443–53.
- 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*, **97**: 282–297.
- 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*, **42**(1): 76–94.
- 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*, 791–800.
- 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/kankochou/shisaku/kankochi/gps.html> (2016 年 4 月 11 日閲覧).
- 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.
- 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, **ICS-140**(5), 23–28.
- 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), **110(2008-UBI-020)**: 57–64.
- 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, **69**(5): 601–604.
- 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, (3): 17-30.