

旅行者の移動履歴を利用した旅行情報のランキング手法

胡 玉龍[†] 趙 耕弘[†] 顔 洪[†] 宮原 徹^{††} 高山 敏典^{††}

森永 久之^{††} 出口 貴嗣^{††} 牛尼 剛聡[‡]

[†]九州大学大学院芸術工学府 〒815-0032 福岡県福岡市南区塩原 4-9-1

^{††} 株式会社ゼンリンデータコム 〒105-7421 東京都港区新橋 1-6-1 日本テレビタワー21F

[‡]九州大学大学院芸術工学研究院 〒815-0032 福岡県福岡市南区塩原 4-9-1

E-mail: [†]huyulong19880301@yahoo.co.jp, [†]cocojoe0220@gmail.com, [‡]ushiana@design.kyushu-u.ac.jp

あらまし 近年、インターネットの普及やコンピュータの処理速度の向上などに伴い、生成される大容量のデジタルデータ（ビッグデータ）の活用が注目されている。代表的なビッグデータの一つとして、人間の移動履歴データがある。本研究では、旅行者の移動履歴を観光情報のランキングに活用する手法を開発することを目的とする。現在、インターネット上には様々な観光情報の検索システムが存在しており、多くの観光情報の検索システムでは、口コミ等によって検索結果をランキングしている。しかし、観光情報の中には、適切な訪問時期があるにもかかわらず、インターネット上に存在する多くの観光情報の検索システムは、適切な時期を考慮したランキングを行っていないため、ランキングされた情報が必ずしもユーザにとって有益でない場合がある。この問題を解決するために、訪問する時期を考慮にした観光情報のランキング手法が重要であると考えた。我々は、この目的のため、九州エリアにおける大量的な旅行者の存在履歴を収集し基礎的な検討を行った。本論文では、その検討結果について説明する。

キーワード ビッグデータ, 観光エリア, 移動履歴, ランキング手法, 旅行情報

1. 背景

現在、様々な分野においてビッグデータの有効的な活用が注目されている[1]。具体的には、ビッグデータの分析により、ビジネスや教育など様々な分野において新しいサービスの出現や、既存サービスの高度化が期待されている。ビッグデータの種類として、様々な種類のデータが考えられるが、代表的なビッグデータの一つにヒューマンプローブデータがある。ヒューマンプローブデータとは、人間の活動にもとづいて取得可能なデータである。ヒューマンプローブデータには様々な種類のデータが考えられるが、代表的な例として、移動履歴データがある。現在、ほとんどの携帯電話やスマートフォン等の携帯端末にGPS機能が搭載されており、携帯端末経由でインターネットサービスを利用した際、ユーザの許諾を得た上でGPSデータを取得・蓄積し、実空間上の人の滞在パターンを取得することができるため、様々な応用が期待できる。

移動履歴データの活用が期待されている分野の一つに、観光産業がある。観光産業は成長産業として、大きな期待が寄せられている分野であり、日本では、国家プロジェクトと

して、観光の振興が実施されている。観光を振興するために、観光者に対して、適切な観光情報を提供することは、重要な課題である。現在、Webの普及に伴って、多くの観光者がWeb上で観光情報の取得を行っている。観光情報の取得を行うための代表的な手段の一つとして、観光情報検索システムがある。観光情報検索システムでは、主にユーザが興味をもつ地域やジャンルをクエリとして指定する事が多い。一般に、ユーザのクエリに対して、多数の検索結果が含まれるため、検索結果をユーザに対して適切であると推定される順序でランキングすることが重要である。現在、多くの観光情報検索システムにおいては、インターネット上でユーザの口コミや採点などを利用して観光地情報のランク付けを行っている。口コミや評価に基づいたランキングは有用であるが、観光情報の中には、時期に依存したものも多く存在している。従来の口コミに基づいたランキング手法では、時期を考慮していないため、ユーザが訪問する時期には不適切な情報が上位にランキングされてしまうことがあるという問題が存在する。この問題点を解決するために、本研究では、旅行者の携帯端末の移動履歴データを利用して、訪問時期を考慮した観光情報のランキングを行う手法を開発

することを目的とする。

旅行者の移動履歴データを利用して訪問時期を考慮した観光情報のランキングを行う手法を開発するために、我々は、九州地域を訪問した旅行者の移動履歴データを分析し、観光に関する特徴が抽出できるかを検討した。以下、その検討結果について述べる。

2. 関連研究

Web を利用した観光支援の研究はこれまでも多く行われてきた。例えば、栗山ら[3]は、複数の観光地と重要度、距離や時間を考慮した観光ルートを作成するシステムを提案している。しかし、観光地に関する時期については考慮していない。

一方、Web 等に存在する情報を利用して自動的に観光に有益な情報を抽出することを目的として、様々な研究が行われている[4,5]。例えば、近年、爆発的に普及した代表的なマイクロブログサービスである Twitter では、利用者がモバイル端末から投稿した際に、モバイル端末が自動的に現在地を計測し、その情報をツイートに付与することが可能である。このような位置情報ツイートから観光に有用な情報を自動的に取得する研究が行われている。藤坂ら[6]は、位置情報ツイートにより、特定の地域でイベントが発生している手法を提案している。一方、長谷川ら[7]は、旅行者の位置情報付きツイートに関して、時間的な連続性、空間的な連続性及び内容の関連度から、ツイートの組織化を行う手法を提案している。これらの研究では、位置情報のみならずテキスト情報を利用しているため、高度な意味的な解釈が可能である。しかし、モバイル端末から投稿されるツイートのすべてに関して位置情報が付いているわけではなく、それらから、ある地域の訪問者数などを正確に推定することは困難である。本研究で利用するデータは、テキスト情報は付加されていないが、任意の時刻におけるエリアごとの訪問者数を高精度で推定できる。したがって、位置情報ツイートでは取得が困難な特徴が抽出できると期待できる。

3. 本研究で利用するデータ

本研究で利用する行動履歴データは、NTT ドコモ社の携帯電話やスマートフォンに搭載された「ドコモ地図ナビ」の地図アプリにて、利用者の許諾を得た上でオートGPSをONにしている際に蓄積された位置情報である。具体的には、携帯電話やスマートフォンに搭載されたGPSによって、計測された利用者ごとの時刻、緯度、軽度の情報である。利用者のプライバシーを保護するために、個人を特定できる情報が記載されず、総体的かつ統計的に加工したデータを利用する。

データの集計方法としては、九州全域を10キロメートル四方のメッシュに分割し、それぞれのメッシュ内に存在する人数を集計したものである。集計の周期は3時間毎とする。

具体的には、0時、6時、9時、12時、15時、18時、21時から3時間後までの人数を集計したものである。データの期間としては、2012年1月1日0時から2012年12月21日21時までである。対象とするユーザは、九州エリア以外に自宅があると推定されるユーザである。これは、今回は旅行の振る舞いを抽出することを目的とするためである。データは約200万レコードから構成される。

4. アプローチ

従来の旅行情報のランキング手法では、ユーザのロコミや採点などを利用して情報のランク付けを行っている。代表的なサービスとして、4travel.jp が提供する「観光人気ランキング」がある。しかし、単純にロコミや採点は、ユーザが訪問した時期を考慮に入れていない。しかし、一般に、観光には適した時期というものが存在することが多く、時期によってユーザの満足度は変化すると考えられる。

そこで、本研究では、旅行者の移動履歴データを利用して、観光エリアに関する特徴を分析し、観光情報の検索結果のランキングの改善に活用することを目的とする。

5. 移動履歴を利用した観光エリアの分析

移動履歴から観光エリアの時期による特徴が取得可能であることを検証するために、いくつかの分析実験を行った。本章では、それらについて説明する。

5.1 毎月の存在者数による分析

まず、観光エリアごとに、一日の存在者数を合計した。その中で、九州における代表的な観光エリアを含むと考えられる代表的なメッシュを8箇所選択肢し、折れ線グラフとして表したものを以下に示す。なお、メッシュは、わかりやすさのために、そのメッシュに含まれる代表的な施設や地域の名称で名前を付けている。例えば、以下で「福岡空港」と名付けられているメッシュは、福岡空港だけでなく、博多駅等も含まれているエリアを表している。

このグラフから、すべてのエリア(メッシュ)において、年末年始、ゴールデンウィーク、お盆に存在人数が増加する局所的なピークが存在することが観測される。このことから、本研究で使用する移動履歴データが、実際の人の移動をパターン的大局的な特徴を反映していると予想される。

また、時期によって、観光エリアごとに変化のパターンが違ふ事がわかる。このことから、観光エリアは時期によって、存在者数が異なり、その変化のパターンは、観光エリアごとによって異なっていることがわかる。これは、観光エリアごとに適した時期が存在し、適した時期に多くの訪問者があることを表している可能性がある。例えば、「沖縄那覇」エリア(グラフ中ではオレンジ色の線で表現)は、一年を通じて多くの滞在者が存在しているが、6月頃は存在者数が少なくなることや、ゴールデンウィークやお盆の時期にも、他地域ほど顕著な増加を見せない。

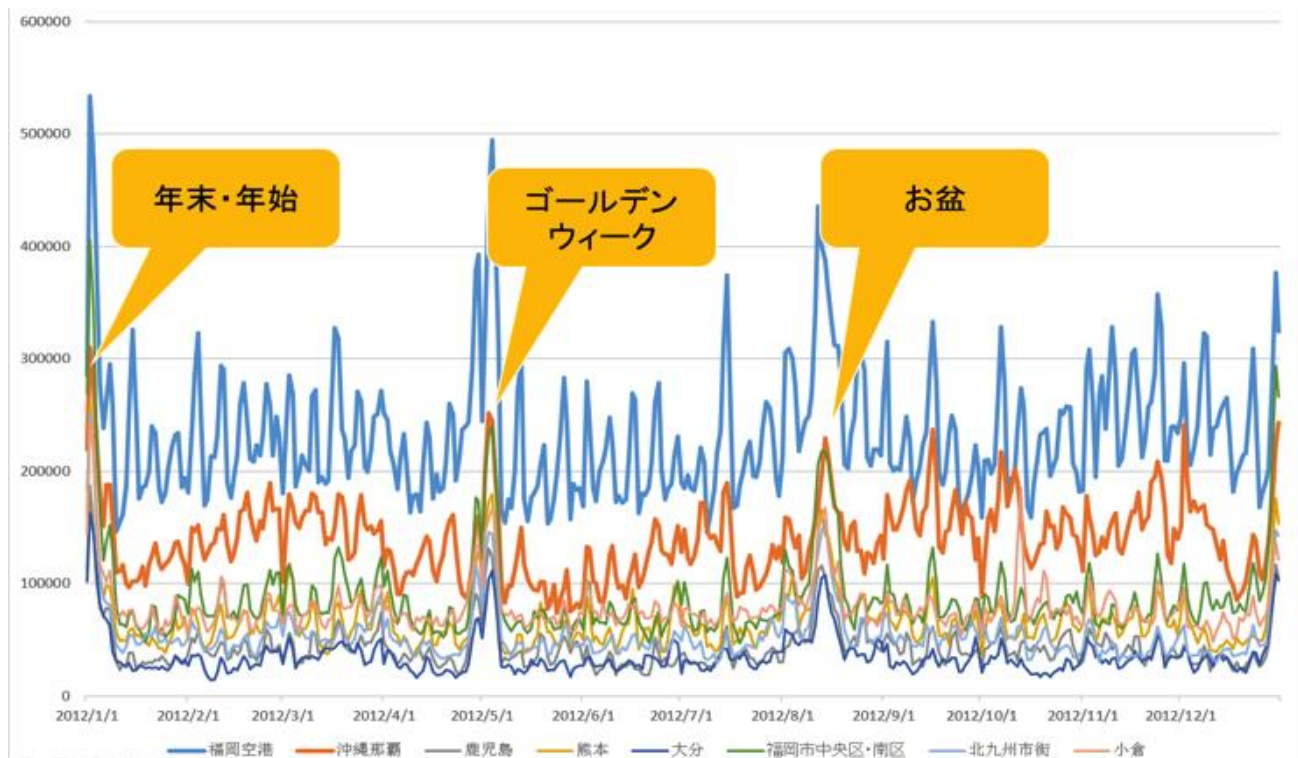


図 1:観光エリアの存在人数推移

このように、観光エリアごとに、変化のパターンが異なることを検証するために、各月の滞在者数の合計数が多い順に、上位 20 位を調査した。調査結果を以下の図 2 に示す。

この結果より、エリアによって、特徴的な順位の変化を観測できる。いくつかの特徴的な例を以下に示す。「宮崎」エリア(図中のオレンジのセル)は、通常は15位程度にランキングされるが、2月のみ9位にランキングされる。この理由としては、2月にはプロ野球のキャンプが宮崎で開催されるため、多くの訪問者があることが理由であると考えられる。

「石垣島」エリア(図中の青色のセル)は、7月と9月のみ、16位にランキングされる。しかし、他の月は上位20位にランキングされない。この理由としては、7月と9月は海水浴やマリンスポーツに適した時期であることが理由であると考えられる。

「別府」エリア(図中の緑色のセル)は3月、5月、11月、12月にのみランキングされる。これは、夏季には「別府」の滞在者数がそれ程多くないことがわかる。この理由としては「別府」は温泉で著名な観光地であり、気温が高い夏季に温泉を訪問したい人は少ないことが理由であると考えられる。

なお、すべての月で「下関・門司」エリアの滞在者数が最も多い。これは、鉄道、自動車によって、陸路で本州から九州を訪問する際には、このエリアを通過することになるからである。本研究で使用したデータは、人物が通過したのか滞

在したのかの区別を行っていないため、このような現象が発生すると考えられる。今後、通過した人物と一定時間滞在した人物とを、起動履歴により区別することにより、この問題は解決できると思われる。

5.2 月別の存在者割合による分析

月別の存在者数による分析では、基本的に都市の中心部や空港や駅など交通のハブとなるエリアが上位にランキングされる傾向があった。これは、これらの地域は様々な観光の要所となる部分であり、年間を通じて恒常的に多くの旅行者が訪問する。

しかし、観光エリアによっては、特定の期間に多くの観光者が訪問したとしても、絶対的な数が多くなければ、ランキングに反映されない。つまり、通常は、訪問者が少ないエリアにおいて、特定の期間だけ顕著に多くの訪問者がある場合、特に時期に依存した観光資源が存在していると考えられる。時期に依存した観光資源は、時期が限定されているため一般的な認知度が低い場合もあり、それらを発見してユーザに提示できれば、訪問を検討している利用者に有用であると考えられる。

順位	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司	下関・門司
2	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港	福岡空港
3	沖縄	沖縄	沖縄	沖縄	沖縄	沖縄	沖縄	沖縄	沖縄	沖縄	沖縄	沖縄
4	天神	天神	天神	天神	天神	小倉	小倉	天神	天神	小倉	天神	天神
5	小倉	小倉	小倉	小倉	小倉	天神	天神	小倉	小倉	天神	小倉	小倉
6	熊本	熊本	熊本	熊本	熊本	熊本	熊本	熊本	熊本	熊本	熊本	熊本
7	北九州1	北九州1	北九州1	北九州1	北九州1	北九州1	北九州1	北九州1	北九州1	北九州1	北九州1	北九州1
8	鹿児島	鹿児島	鹿児島	鹿児島	鹿児島	大分	鹿児島	鹿児島	鹿児島	鹿児島	鹿児島	鹿児島
9	鹿児島2	宮崎	大分	福岡西区	長崎市	長崎市	鹿児島2	大分	沖縄2	長崎市	大分	大分
10	大分	福岡西区	福岡西区	大分	大分	福岡西区	大分	長崎市	久留米	福岡西区	福岡西区	鹿児島2
11	福岡西区	久留米	久留米	長崎市	福岡西区	鹿児島	沖縄2	鹿児島2	福岡西区	沖縄2	長崎市	長崎市
12	久留米	大分	鹿児島2	鹿児島2	久留米	鹿児島2	福岡西区	福岡西区	大分	大分	鹿児島2	福岡西区
13	長崎市	鹿児島2	長崎市	久留米	鹿児島2	沖縄2	久留米	久留米	長崎市	沖縄3	宮崎	久留米
14	宮崎	長崎市	宮崎	宮崎	宮崎	久留米	長崎市	宮崎	宜野湾	ハウスステンボス	久留米	宮崎
15	シーガイア	宜野湾	宜野湾	能古島	能古島	宜野湾	沖縄3	能古島	沖縄3	久留米	沖縄2	佐賀市
16	北九州2	沖縄2	佐賀市	沖縄2	大分2	北九州空港	石垣島	佐世保	石垣島	鹿児島2	大分2	能古島
17	佐世保	シーガイア	沖縄2	熊本2	別府	シーガイア	能古島	沖縄2	宮崎	宮崎	長崎港	宜野湾
18	熊本2	沖縄3	別府	北九州2	熊本2	能古島	熊本2	佐賀市	鹿児島2	長崎港	沖縄3	北九州2
19	熊本3	熊本3	シーガイア	大分2	佐賀市	沖縄3	佐賀市	シーガイア	北九州2	北九州空港	能古島	シーガイア
20	佐賀市	熊本2	熊本3	北九州空港	北九州空港	宮崎	長崎港	長崎港	長崎港	宜野湾	別府	別府

図 2: 月別の上位 20 位の観光エリアのランキング

上記の考察に基づき、観光エリア毎に、各月の存在者数の割合を計算し、ランキングに応用する。エリア a の、 m 月の存在者率 $r(a, m)$ を以下のように定義する。

$$r(a, m) = \frac{v(a, m)}{\sum_{i=1}^{12} v(a, i)}$$

ここで、 $v(a, m)$ はエリア a の m 月の総存在者数を表す。

2012 年の各月において、存在者率が最も高かったエリアを表 1 に示す。

月	エリア (メッシュコード)	エリアが含む地域	推定される理由
1	493012	雲仙	
2	473132	日南	プロ野球キャンプ
3	473153	宮崎	
4	493007	山都町	通潤橋放水
5	453044	屋久島	観光シーズン
6	493150	小国町	
7	453033	屋久島 (海岸)	海ガメ産卵
8	473025	桜島	
9	492967	武雄	
10	493110	南阿蘇	カントリーゴールド
11	473102	串間	プロ野球キャンプ
12	492967	武雄	

表 1: 各月の存在者率 1 位のエリア

表 1 において、エリアは標準地域メッシュ[8]の 2 次メッシュのメッシュコードを用いて表している。また、それぞれのエリアに含まれる代表的な地域を示している。

表 1 に示した時期に、それぞれのエリアの存在者率が高かった理由を個別に調査した。すべてのエリアについてその理由は明らかにすることはできなかったが、いくつかのエリアに関しては、予想される理由を発見できた。

2 月には宮崎県の日南エリアでは、埼玉西武ライオンズ、広島東洋カープのプロ野球キャンプが開催される。図 3 に 1 月～3 月の日南エリアの存在者数の推移を示す。図から、3 月中は恒常的に多くの人が存在していることがわかる。

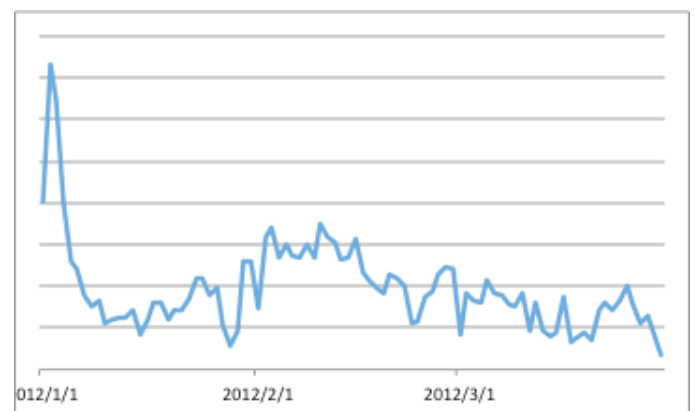


図 3: 1 月～3 月の日南エリアの存在者数の推移

また、4 月には、熊本県の日田エリアでの存在者数が多く存在する。図 4 に、4 月を中心に前後 1 ヶ月の日田エリア

の存在者数の推移を示す。このグラフから、4月の28日、29日、30日に存在者が顕著に多いことがわかる。これらの日は、通潤橋と呼ばれる橋から放水が行われたためであると考えられる。

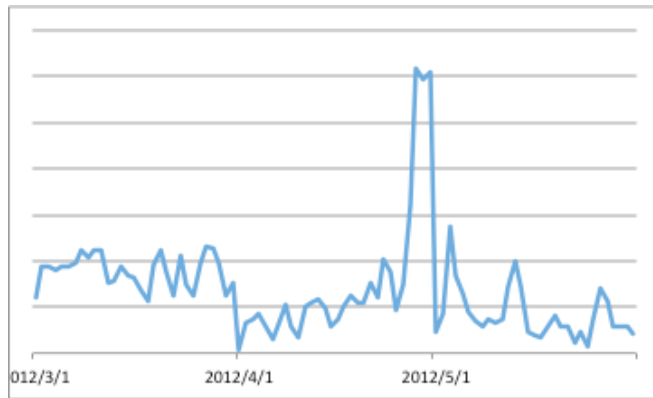


図 4: 3 月～5 月の日田エリアの存在者数の推移

また、5 月は屋久島の観光にもっともよい季節であると考えられており、またゴールデンウィークの観光シーズンであるため、多くの訪問者があったと考えられる。一方、7 月にも屋久島の海岸エリアの訪問者が多い。この地域は、ウミガメの産卵と孵化で有名な地域であり、7 月下旬に産卵と孵化が見られる可能性がある時期であり、多くの訪問者が存在したと予想される。

さらに、10 月に阿蘇エリアでの存在人数が増大しているのは、この地域で、10 月下旬にカントリーゴールドと呼ばれる、日本最大のカントリーミュージックのフェスティバルが開催されていることが原因であると考えられる。

このように、存在者の割合を利用すると、地域で固有のイベントなどが検出できるため、時期を考慮した観光情報のランキングに有用であると考えられる。

6. おわりに

本論文では、携帯端末の移動履歴から取得される旅行者の移動履歴を利用することによって、時期を考慮した旅行情報のランキングを実現するための基礎的な検討を行った。今後、今回の検討で得られた知見に基づいて、時期を考慮した観光情報のランキング手法を開発する予定である。

参 考 文 献

- [1] Intel IT Center, Big Data in the Cloud: Converging Technologies, <http://www.intel.com/content/dam/www/public/us/en/documents/product-briefs/big-data-cloud-technologies-brief.pdf>, 2013.
- [2] ビクターマイヤー＝ショーンベルガー, ケネス・クキエ, 斉藤栄一郎 (訳): ビッグデータの正体, 2013.
- [3] 栗山 恭嘉, 花野 博司, 澤 悠太, 村田 佳洋, 柴田 直樹, 安本 慶一, 伊藤 実, 藤原 礼征, ”多数の観光候補地から効率良い観光スケジュールを自動的に作成・提案するシステム P-Tour の Google Maps を利用した設計と実装”, 情報処理学会研究報告: マルチメディア通信と分散処理研究会報告, vol. 2008, No 21, pp.261-266, 2008.
- [4] Hidetsugu Nanba, Aya Ishino and Toshiyuki Takezawa, Chapter 6 Automatic Compilation of Travel Information from Texts: A Survey, Theory and Applications for Advanced Text Mining, Sakurai, S. (Eds.) INTECH, 2012.
- [5] 難波英嗣, 観光情報の自動編纂, 知能と情報, Vol. 26, No. 1, 2014.
- [6] 藤坂達也, 李龍, 角谷和俊, 地域イベント発見および特性検証のための実空間マイクロブログを用いたユーザ移動パターン分析システム, 情報処理学会全国大会講演論文集, Vol. 72, No.1, pp. 845-846 2010.
- [7] 長谷川馨亮, 馬強, 吉川正俊, 行動の時空間連続性を考慮した旅行ツイートの組織化, 第 5 階データ工学と情報マネジメントに関するフォーラム, 2013.
- [8] <http://ja.wikipedia.org/wiki/地域メッシュ>