

GPSデータを用いた観光客訪問スポットの自動抽出と分析手法に関する基礎的研究

小橋川 嘉樹¹・藤生 慎²・高山 純一³・中山 晶一朗⁴

¹正会員 金沢大学 自然科学研究科 (博士後期課程) 環境デザイン学専攻 (〒920-1192 石川県金沢市角間町) / 株式会社電通

E-mail: kobashikawa@stu.kanazawa-u.ac.jp

²正会員 金沢大学 理工研究域 環境デザイン学系 (〒920-1192 石川県金沢市角間町)

E-mail: fujju@se.kanazawa-u.ac.jp

³フェロー会員 金沢大学 理工研究域 環境デザイン学系

⁴正会員 金沢大学 理工研究域 環境デザイン学系

本研究は、GPSのログデータにより人の動きを可視化して地域計画に活用するため取得したデータの加工・集計を一貫して行う手法の開発・研究を目的としている。クルーズ客船で来訪した外国人観光客に携帯してもらったGPSロガーのデータの実データを対象に、汎用的なデータベースソフトウェアを使った分析フローを設計し、訪問した観光スポットの特定とその滞在時間の算出を行った。その結果、実態把握や課題抽出のためのデータとして活用するのに十分な品質の出力を得ることができた。さらにそのデータを対象に定量的・定性的な解釈を行った。その結果、今回被験者となった外国人観光客の典型的な観光行動の抽出を行うとともに、彼らの観光をより高密度で魅力的なものとしていくのに対して課題となりうる事項についての示唆を得ることができた。

Key Words : 活動分析, 時間利用, GIS, 観光・余暇行動, ビッグデータ

1. はじめに

(1) 背景

人々の回遊行動を把握するための手法として、GPSなどの位置情報データを用いたデータの収集が有望と考えられている。特に観光客は当該エリアに対する情報も少ないことから、記憶に頼った事後のアンケート調査で詳細を聴取することは難しい。そこで意識に頼らず行動により分析を行う必要があり、その観点からも位置情報データの活用への期待が高い。しかしこれまでのところ、位置情報データからその回遊行動を記述するためには人的リソースの活用による定性的な分析が主となってきた。

その背景には、これまで個人情報保護法とその周辺環境により個人を特定できる形での「トラッキング」可能なデータの取得に大きな制約があり、個別の研究や調査プロジェクトごとに完全にパーミッションを取得できる形でサンプルを収集する必要があることからサンプルが少数に限られてきたことの影響も考えられる。1件1件目視による検討を行うことが可能な件数であり、また対象も情緒的な「観光」「買い物回遊」などの行動が中心

であるため、人間の手による定性的な解釈で納得性の高い解釈が可能であったとも言えよう。

しかしIT技術の進歩によりいわゆるビッグデータが利用可能となることで到底人力では処理できない量のデータを対象とする必要が生じつつある。

一方でそこで利用が想定される位置情報データは、改正後の個人情報保護法で「識別非特定」情報に位置づけることで利用が可能となる場合がほとんどであると想定される。この場合は匿名化や集計処理による非特定化された加工済みのデータを基にして分析を行う必要が生じるため、現状の完全にパーミッションが取得できているローデータとは異なったアプローチが必要となることが想定される。

これら2つの観点から、位置情報データを「機械的に(人手を介さずに)集計する」手法へのニーズが高まることが予想される。

(2) 本研究の目的

位置情報データを用いて人々の行動特性を把握するには、

- ・緯度経度などの座標を、意味を持った場所（たとえば観光スポット）にマッチングする
- ・マッチングされた各時刻の地点情報を、誤差も加味しながら滞在・通過などを判定する
- ・目的に応じて、滞在・通過等の行動をとった人数、時間などを集計する

というプロセスが必要となる。今回はこれら一連のプロセスを自動かつ分析者の介入が極力少ない形で一貫して行う手法を試行し、その実現可能性を検証することとする。

さらに、出力された分析結果をもとに解釈を試み、位置情報データの分析からどのような知見を導出し得るかの検討を行う。

2. 既往研究

位置情報データによる行動分析についての研究は、多くが1990年代後半以降となっている。大森ら¹⁾はPHSのマイクロセルによる位置特定と日記式の2手法で高齢者の行動を調査し、その比較によりPHSデータの特性、特に欠落の可能性などを論じている。

藤田ら²⁾は小田原市において観光客の行動をGPS携帯電話により調査し立ち寄ったスポットの特定などを行っている。野村ら³⁾はさらにGIS上で移動速度や歩行者の密度を図示するための手法を鎌倉市におけるデータを使って提案している。しかしいずれも図示した結果を目視で分析することが前提となっている。

長尾ら⁴⁾、清水ら⁵⁾はいずれも北海道におけるレンタカーによる行動を対象に広域の回遊行動を分析している。長尾らは滞在・移動判定ロジックを検討してそれらのデータから観光地の魅力度の算出を、清水らは移動中の情報入手による予定変更の可能性の検討を行っている。

一方山本ら⁶⁾は新宿御苑内を小さいエリアに分割し、エリアごとの通過有無から統計的手法により行動パターンを分類し、それぞれの特徴を示している。Cantisら⁷⁾はシチリア島（イタリア）におけるクルーズ客を対象にGPSトラッキングデータから距離、時間などの諸指標を算出してクラスタリングを行い、クラスタごとの行動特性を分析している。

しかしいずれの研究も、観光スポットそれぞれについての特性を分析したりそれらの関係性を論じたりすることは十分ではない。また手法に関しても、図示による目視での解釈、計算量の多い統計解析手法、アンケート併用による属性との相関の検討など、ビッグデータとして蓄積された大量のデータを活用して都市内の行動を分析していくのに適した手法とは言い難い。

そこで本研究においては、データを入手してから観光スポットの特性を分析するのに必要なデータを出力する

まで自動化可能な一貫した分析フローを開発するとともに、その結果から観光スポットの特性を分析することを目標とする。

3. 研究の方法

今回は手法を検討するためパーミッション取得済みのデータで今後のビッグデータへの応用を視野に入れて妥当性を試行することとした。そのため2014年に実施済みの調査におけるGPSデータ40件を利用した。

(1) 位置情報データ収集

2014年度に金沢港に寄港したクルーズ客船の乗客に対し協力を依頼し、市内観光中にGPSロガーを装着してもらった。詳細は以下の通りである。

- ・調査日：2014年6月14日・7月2日・7月20日・9月21日
- ・対象人数：各日10名×4日＝計40名
- ・対象者抽出方法：下船口から駅前行き連絡バス乗り場への動線上にて対面にて協力を依頼
- ・GPSデータ取得間隔：3秒ごと

(2) マッチング用観光地データの作成

a) 観光スポットエリアの定義

まず初めに国土地理院が公開している基盤地図情報データから街区及び町丁目データのポリゴンを作成し、観光スポットに対応する街区または町丁目を定義した。

観光スポットの定義においては街区単位（住居表示未実施などで街区データが存在しない場合は町丁目単位）で当該エリア全体を含む最小範囲を地図上にて現地の状況を鑑みながら目視で判定している。

b) 滞在判定用エリアの定義

a)で定義した街区データはほとんどの場合道路に囲まれた領域であり、道路自体は含まれない。しかし特に金沢においては茶屋街や武家屋敷跡など町並みが観光スポットとなっている場所が多く、街歩きすなわち道路上での滞在を取得することが重要となる。そこでa)で定義した対象街区について周辺にバッファを設定することで、街区及び街区が面する道路を定義することとした。

なおこのバッファの幅については、対象エリアの道路幅等を勘案し、今回は6mを採用した。図-1にひがし茶屋街を例にこれらのステップを示す。また、図-2に今回分析用に定義した全スポット滞在判定範囲を示す。

(3) GPSデータのマッチング

a) GPSデータの取り込み

(1)で収集したGPSデータ計262,434点を取り込んだ。この中には調査前の動作テスト、回収後の電源切り忘れなどによる不要データも含まれているが、今後ビッグデー

タを用いて分析する場合にも同様に帰宅やホテルへのチ

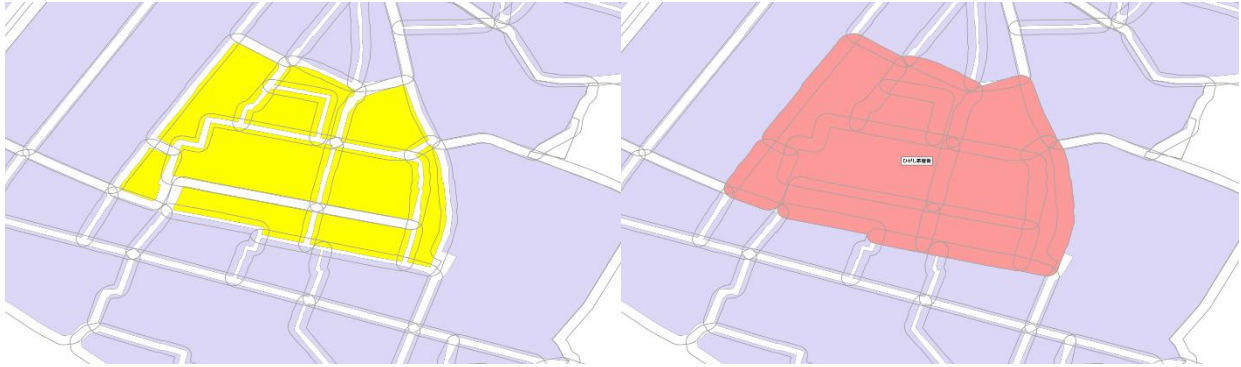


図-1 観光スポットエリアの定義（上）と滞在判定用ポリゴンの作成（下）（国土地理院の基盤地図情報を使用）

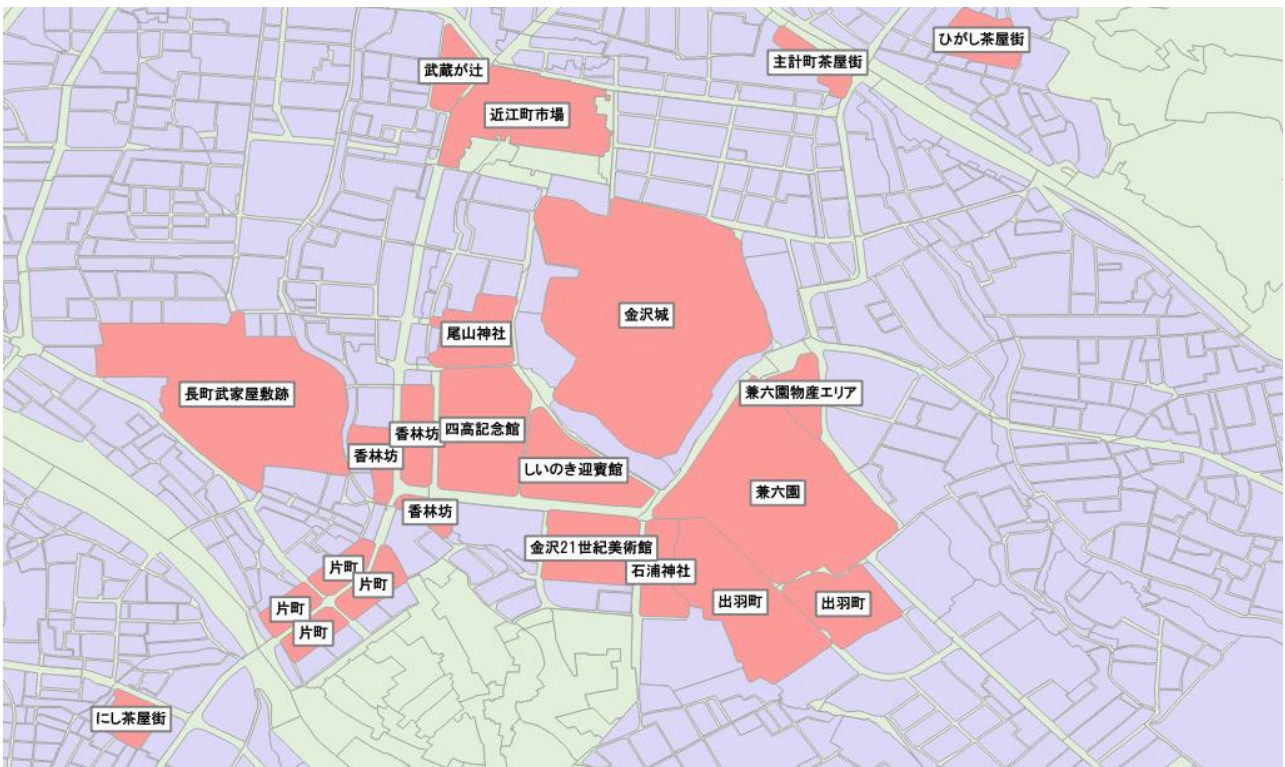


図-2 滞在判定用エリアの全体図 ※紙幅の関係で金沢駅東口・西口（この図の北北西にあたる）は省略した
（国土地理院の基盤地図情報を使用）

チェックイン後のように不適なデータは含まれる。そこで今回は不要データのカットも実際のフロー中で試行するため、いったん全データを取り込んだ。

b) GPSデータの所属エリア判定

ポイントが属する滞在判定用エリアを判定した。ここで、設定した滞在判定エリアのうち「片町」「四高記念館」「しいのき迎賓館」「出羽町」は所属判定されたポイントが生じなかったことから、以下の分析では除外している。

c) 各エリアへのチェックイン・チェックアウト判定

bで判定した各ポイントの所属エリアはバス・徒歩等で通過しただけのもの、GPSデータの誤差により実際に

は滞在していない点が記録されたものなども含まれている。そこで今回は5分を閾値としてこれを超えて同一エリアにとどまっている場合を滞在と判定している。

また同様の理由で、本来そのエリアにいたにも関わらずエリア外と判定される場合もある。この点については、5分以内に同じエリアに戻ってきた場合には引き続き滞在していたと判定することでクリアしている。

d) 不要データの除外

cで判定したチェックイン・チェックアウト時刻のうち全エリアを通して最も早いものと最も遅いもの間を有効活動時間とし、それ以外を分析対象から除外した。なおほとんどのサンプルについてこの有効活動時間は、

港との間の連絡バスが発着している金沢駅で開始・終了となっている。これによりaで述べたような分析対象外のデータを含んだ場合でも利用可能となる。

(4) 滞在情報の集計

a) 滞在時間の算出

(3)で滞在判定を行ったデータについて、各エリアの滞在有無、滞在時間等を算出した。さらにこれらについて主要な記述統計量も算出した。

b) 滞在についてのスポット間の関係性分析

あるエリアの滞在有無と他のエリアの滞在との関係性を分析した。今回のサンプルは40サンプルという少数のため、Fisherの正確確率検定により検討した。

(5) 分析環境

今回分析に使用した環境は次の通りである。

- ・分析環境：Microsoft SQL Server 2014 Standard
- ・表示環境：QGIS Desktop 2.8.2
- ・CPU：Intel Core i5 4690
- ・メモリ：32GB

なお今回はデータ量が小さいため、ディスクがパフォーマンスに与える影響は無視できる。

今回の地理的演算処理についてはGISではなく一般のデータベースソフトウェアSQL Serverの地理データ機能を活用し、複雑な集計や他のデータソースとの連結が容易なワークフロー環境を構築した。この他基盤地図情報やGPSロガーのXML形式データを処理するため、必要に応じてMicrosoft Visual C#によりプログラムを作成した。

4. 結果

(1) サンプル別の集計

各サンプルについて訪問スポット数を集計した結果を図-3に示す。訪問スポット数の平均は3.3か所、最頻値は3か所となった。なおここでは狭義の観光スポット（金沢城・兼六園・ひがし茶屋街・近江町市場・長町武家屋敷跡・石浦神社・尾山神社・金沢21世紀美術館）の7か所のみをカウントし、繁華街（武蔵が辻・香林坊）と金沢駅周辺（東口・西口）、及び滞在について他の観光スポットとの従属関係が想定される兼六園周辺の物産エリアは除外した。

次に、有効活動時間および7スポット合計の滞在時間を集計した結果を図-4に示す。有効活動時間は平均306.4分/標準偏差84.5分、観光スポット滞在時間は同じく144.3分/64.2分であった。またこの比率であるスポット滞在時間率は全体で47.1%であった。滞在時間率は、値が特に低い1サンプルを除くと25.5%~67.8%の間でまとまっている。このことから外れ値の閾値は20%程度が適

していると考えられる。なお、有効活動時間とスポット滞在時間率の間では相関はあまり見られないものの、有効活動時間が長くなると滞在時間率の分散が多少大きくなっている。

(2) 観光スポット別の集計

各観光スポットについて、訪問率を高い順に集計した結果を図-5に示す。また、そのうちより上位のスポットも全て訪問している割合も同時に示している。上位3つの金沢城・兼六園・ひがし茶屋街の3か所とも訪問しているサンプルが42.5%に達し、特にこれら3か所が金沢市内観光で人気が高いことが分かる。

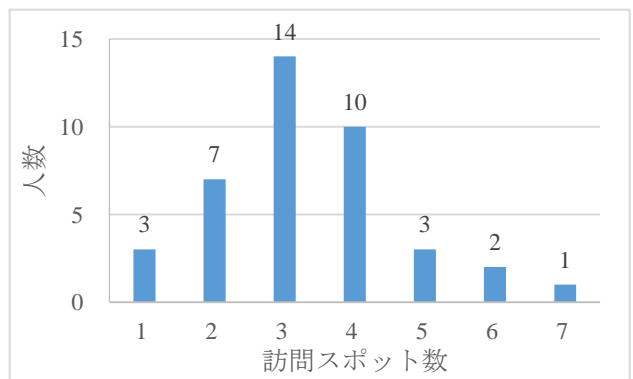


図-3 訪問スポット数別の人数 (n=40)

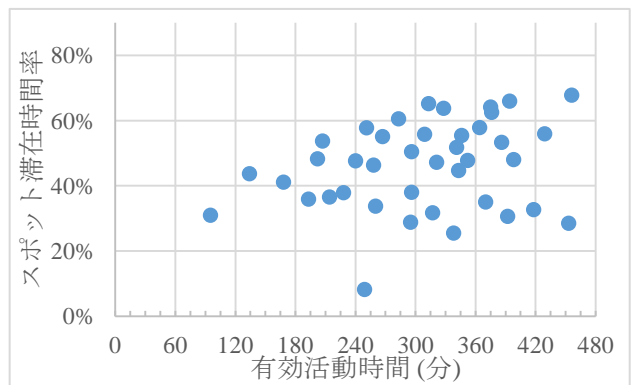


図-4 有効活動時間とスポット滞在時間率 (n=40)

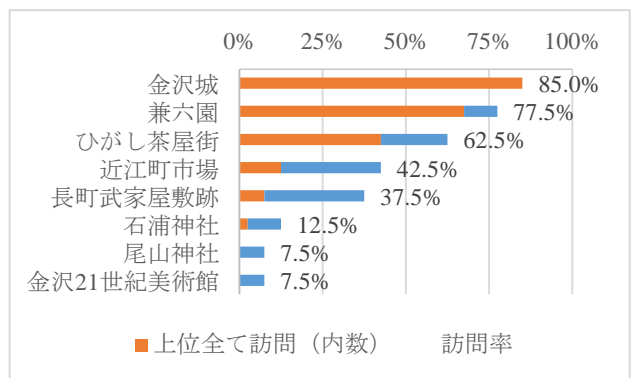


図-5 スポット別訪問率 (n=40)

次に、スポット別の滞在時間を表-1に示す。主要スポットかつ面積も広い上位5か所は滞在時間も長い。一方21世紀美術館は施設の性格を考えると短く、特別展等への入場はせず、屋外等の無料展示のみ観覧している可能性などが示唆される。

(3) 観光スポット相互の関係分析

各スポットについて重複訪問の状況をまとめたものを表-2に示す。まず近江町市場については近隣の長町武家屋敷跡、尾山神社、武蔵が辻が高い。長町武家屋敷跡は近江町市場に加えて香林坊も高い。この2か所と石浦神社も高いが、これはこの2エリア間と兼六園方面の移動手段はバスが一般的であり、その際に通過する広坂交差点の角に位置していることが可能性として考えられる。

一方長町武家屋敷跡と香林坊は相互に重複率が高いことからこれらの相関関係は高く、香林坊が長町武家屋敷跡への拠点となっていることがうかがえる。(実際にバス車内でも香林坊が長町武家屋敷跡への最寄りバス停としてアナウンスされている)

なお今回の分析では有意に重複訪問しにくい組み合わせは発生しておらず、この点からも主要3スポットを軸に追加の観光地を組み合わせるといった比較的画一的な周遊行動がメインであったと推察される。

5. 結果の検討

(1) サンプルについて

今回のサンプルでは平均3.3スポットを訪問しており、特に金沢城・兼六園・ひがし茶屋街の3大スポットへの

訪問が多かった。有効行動時間はクルーズ船の入港から出航までのスケジュールの兼ね合いもあって平均約5時間であった一方、回遊型で時間のかかる3~4スポットを回ることで、網羅的にスポットを訪問することは難しかったものと思われる。とはいえこの3大スポットではそれぞれ1時間前後費やしているサンプルが多く、あわたしさは感じられない。

ここで仮に旅の満足度が観光スポットでの滞在時間と相関を持つとした場合、満足度を上げる工夫としては有効活動時間を増やすこととスポット滞在時間率を上げるという2つのアプローチが考えられる。

前者については北陸新幹線の開通により対東京では最長で12時間14分の滞在が可能となっており、これらの交通手段の利用者はより多くのスポットを回る機会もあると考えられる。

表-1 スポット別滞在時間(n=40)

| | 訪問率 | 滞在時間(分) | | |
|-------------|-------|---------|------|-------|
| | | 平均 | 標準偏差 | 最大値 |
| 金沢城 | 85.0% | 46.3 | 24.4 | 104.3 |
| 兼六園 | 77.5% | 54.3 | 24.7 | 111.9 |
| ひがし茶屋街 | 62.5% | 33.2 | 24.7 | 94.1 |
| 近江町市場 | 42.5% | 38.7 | 22.4 | 90.7 |
| 長町武家屋敷跡 | 37.5% | 56.5 | 28.7 | 132.9 |
| 石浦神社 | 12.5% | 9.1 | 3.2 | 14.7 |
| 尾山神社 | 7.5% | 25.7 | 19.0 | 52.6 |
| 金沢 21 世紀美術館 | 7.5% | 19.6 | 17.0 | 43.6 |

表-2 スポット重複訪問率 (p値はFisherの正確確率検定による)

| 訪問スポット (%) | n | 重複訪問スポット | | | | | | | | | | | | | | | |
|-------------|----|----------|-------|--------|-------|-------|------|------|------|------|--------|-------|-------|------|------|-------|-------|
| | | 金沢城 | 兼六園 | ひがし茶屋街 | 近江町市場 | 武家屋敷跡 | 長町 | 石浦神社 | 尾山神社 | 美術館 | 金沢21世紀 | 物産エリア | 兼六園 | 香林坊 | 武蔵が辻 | 金沢駅東口 | 金沢駅西口 |
| 金沢城 | 34 | | 79.4 | 67.6 | 38.2 | 35.3 | 8.8 | 5.9 | 8.8 | 35.3 | 32.4 | 14.7 | 97.1 | 38.2 | | | |
| 兼六園 | 31 | 87.1 | | 61.3 | 35.5 | 41.9 | 16.1 | 6.5 | 9.7 | 38.7 | 32.3 | 12.9 | 100.0 | 45.2 | | | |
| ひがし茶屋街 | 25 | 92.0 | 76.0 | | 40.0 | 40.0 | 8.0 | 8.0 | 4.0 | 28.0 | 32.0 | 16.0 | 96.0 | 36.0 | | | |
| 近江町市場 | 17 | 76.5 | 64.7 | 58.8 | | 52.9 | 29.4 | 17.6 | 11.8 | 41.2 | 41.2 | 41.2 | 100.0 | 29.4 | | | |
| 長町武家屋敷跡 | 15 | 80.0 | 86.7 | 66.7 | 60.0 | | 26.7 | 13.3 | 13.3 | 26.7 | 80.0 | 20.0 | 100.0 | 46.7 | | | |
| 石浦神社 | 5 | 60.0 | 100.0 | 40.0 | 100.0 | 80.0 | | 0.0 | 40.0 | 20.0 | 60.0 | 40.0 | 100.0 | 60.0 | | | |
| 尾山神社 | 3 | 66.7 | 66.7 | 66.7 | 100.0 | 66.7 | 0.0 | | 0.0 | 66.7 | 33.3 | 33.3 | 100.0 | 0.0 | | | |
| 金沢 21 世紀美術館 | 3 | 100.0 | 100.0 | 33.3 | 66.7 | 66.7 | 66.7 | 0.0 | | 0.0 | 33.3 | 33.3 | 100.0 | 66.7 | | | |
| 兼六園物産エリア | 14 | 85.7 | 85.7 | 50.0 | 50.0 | 28.6 | 7.1 | 14.3 | 0.0 | | 21.4 | 14.3 | 100.0 | 21.4 | | | |
| 香林坊 | 13 | 84.6 | 76.9 | 61.5 | 53.8 | 92.3 | 23.1 | 7.7 | 7.7 | 23.1 | | 23.1 | 100.0 | 46.2 | | | |
| 武蔵が辻 | 7 | 71.4 | 57.1 | 57.1 | 100.0 | 42.9 | 28.6 | 14.3 | 14.3 | 28.6 | 42.9 | | 100.0 | 28.6 | | | |
| 金沢駅東口 | 39 | 84.6 | 79.5 | 61.5 | 43.6 | 38.5 | 12.8 | 7.7 | 7.7 | 35.9 | 33.3 | 17.9 | | 38.5 | | | |
| 金沢駅西口 | 15 | 86.7 | 93.3 | 60.0 | 33.3 | 46.7 | 20.0 | 0.0 | 13.3 | 20.0 | 40.0 | 13.3 | 100.0 | | | | |

凡例：重複訪問しやすい ■ p<0.01 ■ p<0.05 ■ p<0.1 重複訪問しにくい ■ p<0.01 ■ p<0.05 ■ p<0.1

一方後者についてスポット滞在時間率は47.1%に過ぎず、滞在中の半分の時間を食事や移動に費やしているこ

とになる。特に今回は外国人個人旅行者客ということと言語の問題もあり特に移動効率に影響が出ている可能性も

あり、この点についてはさまざまな情報提供により迷いにくい環境を整備するなどが有用となろう。

さらに付加的なアプローチとして、移動そのものへの価値の付加、例えば徒歩ルートの観光資源化やバスやバス停の魅力向上なども検討の余地がある。この点についてはスポット間の詳細な移動状況の分析により今後示唆を得ることも考えたい。

(2) 観光スポットについて

主要3スポットについては、いずれも長めの滞在時間となっていた。参考までに金沢市観光協会のモデルコースのうち最も基本的なものの所要時間と比較すると、金沢城（推奨約30分）・兼六園（同40分）では長め、近江町市場（同40分）、長町武家屋敷跡（同60分）はほぼ同等、ひがし茶屋街（同60分）・金沢21世紀美術館（同30分）は短めとなっている。特にひがし茶屋街の差は大きいことから、観光協会が想定しているような楽しみ方が実際には体験されていない可能性なども想定される。

一方スポット間の重複を見ると、香林坊・武蔵が辻の立地による相関は当然にしても、近江町市場と石浦神社という意外な組み合わせも抽出された。ここではバスのルートにその原因を求めることとしたが、GPSのデータを定量的に解釈することでこのような意外な組み合わせが発見されたことはデータ活用の意義の1つと言えよう。また今回は全員日帰りではあるものの平均5時間程度の有効活動時間が確保されていた。重複訪問に負の相関が出ていないことを考えると、この程度の時間が確保できれば主な観光地を回りきることができていると推察される。この点は有効活動時間が半日しか確保できないような旅程の場合には異なった結果となりうる。またマイナーな観光スポットも選択する余裕がある場合にはその選択行動に旅行者の関心や嗜好が現れる可能性があり、今後の研究の課題としたい。

6. 結論

本研究では、GPSロガーのローデータから立ち寄った観光スポットを特定しその滞在時間を算出し、個々の旅行者の行動の詳細分析と観光スポットの特性を分析した。将来の分析環境も考慮し分析ルールの単純化と作業フローの自動化を試みたが、出力された結果は現実に照らして考えても妥当性が高いものとなった。分析ルールのパラメーターや対象エリア設定の手法などさらなる検討と精緻化の余地が残されている部分もあるものの、全体の分析フローとしての有用性は検証できたと考えられる。特にスポットへの来訪有無に加えてその滞在時間も定量的に把握し詳細な分析ができる体制を整えることができたのは意義深い。

さらにそこから出力された実際のデータを用いて金沢における観光行動の特徴分析を行い、主要3スポットへの訪問率の高さや重複訪問状況を把握することができた。今回のサンプルは全体的にじっくりと見て回る行動が多かったが一部のスポットでは想定より短い滞在時間となっていることも発見され、課題の抽出へのGPSデータの活用可能性も検証することができた。

謝辞：本研究に多大なご協力をいただいたプリンセスクルーズ社および石川県の皆様にご協力いただいた。また滞在中の全行動のトラッキングという負担の大きい調査にも関わらず趣旨に賛同し快くご協力いただいた乗客の皆様にも厚く謝意を表する次第である。

参考文献

- 1) 大森宣暁, 室町泰徳, 原田昇, 太田勝敏: PHSの位置情報サービスを用いた高齢者の一週間の交通行動調査, 第19回交通工学研究発表会論文報告集, pp.113-116, 1999.12.
- 2) 藤田朗, 半明照三, 山田雅夫, 大内浩, 三宅理一: 7413 GPS 携帯電話を用いた回遊行動の調査分析: 小田原市中心市街地を事例として(GPS,都市計画), 学術講演梗概集. F-1, 都市計画, 建築経済・住宅問題 2003, pp.855-856, 2003.7.
- 3) 野村幸子, 岸本遅也: GPS・GISを用いた鎌倉市における観光客の歩行行動調査とアクティビティの分析(第4部 学術論文,情報化の視点からみた建築・都市のフロンティア), 総合論文誌 No.4, pp.72-77, 2006.2.
- 4) 長尾光悦, 川村秀憲, 山本雅人, 大内東: GPS ログからの周遊型観光行動情報の抽出(<特集>「ネットワークデータマイニング」「センサデータマイニング」), 電子情報通信学会技術研究報告. AI, 人工知能と知識処理 Vol.105, No.224, pp.23-28, 2005.7.
- 5) 清水哲夫, 伊藤龍秀, 村山秀敏, 田中倫之: 訪日韓国人のレンタカーによる観光周遊行動の分析, 第32回土木計画学研究発表会・講演集, 238, 2005, CD-ROM.
- 6) 山本 泰裕, 伊藤 弘, 小野 良平, 下村 彰男: 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(平成 18 年度日本造園学会全国大会研究発表論文集(24)), ランドスケープ研究: 日本造園学会誌: Journal of the Japanese Institute of Landscape Architecture Vol.69, No.5, pp.601-604, 2006.3.
- 7) Stefano De Cantisa, Mauro Ferranteb, Alon Kahanic and Noam Shovald: Cruise passengers' behavior at the destination: Investigation using GPS technology, Tourism Management Vol.52, pp.133-150, 2016.2.