

旅行ブログエントリの属性に基づいた旅行者の行動分析

藤井 一輝[†] 難波 英嗣[†] 竹澤 寿幸[†] 石野 亜耶[‡]

[†] 広島市立大学大学院情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

[‡] 広島経済大学ビジネス情報学科 〒731-0138 広島県広島市安佐南区祇園 5-37-1

E-mail: [†] {fujii, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp, [‡] ay-ishino@hue.ac.jp

あらまし 訪日外国人旅行者数は増加傾向にあり、今後も旅行者数を増やしていくためには、観光地のマーケティングは必要である。そこで本研究では、訪日外国人旅行者が記述した旅行ブログエントリを利用して、旅行者や各都市の特徴を明らかにすることを目指す。旅行ブログエントリは、旅行者の実体験や感想が記述されており、有益な旅行情報源と言える。本研究では、はじめに、旅行ブログエントリを投稿したブロガーの属性(「年齢」と「使用言語」と各旅行ブログエントリの属性(観光タイプ)を自動判定する。そして、判定された属性に基づいて分析することにより、旅行者や各都市の特徴を明らかにし、観光施策の展開を支援することを目指す。

キーワード 訪日外国人旅行者、情報抽出、観光分析

1. はじめに

日本では、観光を基幹産業と位置づけた多様な取り組みを積極的に行われており、2013年には訪日外国人旅行者数 1,000 万人を超えた。今後も多くの訪日外国人旅行者を招くためには、さらなる観光地のマーケティングが必要である。そこで本研究では、日本国内における観光地のマーケティングを行うため、訪日外国人旅行者により、体験記や感想が記述された旅行ブログエントリに着目する。観光庁が行った「日本を訪れた旅行者の出発前の旅行情報源」に関する調査では、インターネット媒体である「個人のブログ」は 24.1%、ロコミである「自国の親戚・知人」は 18.6%と報告されている。このように、今まで主流であったロコミに変わり、旅行ブログエントリは旅行情報の収集や交換する場として主流になりつつあり、有益な情報源であると言える。

以上の背景より、本研究では、訪日外国人旅行者が記述した旅行ブログエントリを利用して分析を行う。近年、旅行ブログエントリを利用して旅行者や訪問地を分析し、観光行政の支援に役立てる試みがいくつか存在している。例えば、Wenger[1]は、オーストリアに関する旅行ブログエントリから、旅行者の分析を行っている。その結果、女性の旅行者は、男性の旅行者に比べて食事に興味があることが明らかになった。しかし、これらの研究では、すべての分析を人手により行っている。そのため、大量の旅行ブログエントリを利用した分析が困難であるという問題点を抱えている。

そこで本研究では、旅行ブログエントリを投稿した旅行者であるブロガーの属性「性別」や「使用言語」、各旅行ブログエントリの属性「観光タイプ」を自動判定する手法を用いる。これらの属性を自動判定することにより、大量の旅行ブログエントリを利用して分析を行うことができる。そして、自動判定された属性に

基づいて分析を行うことにより、例えば、広島の旅行ブログエントリにおいて、「使用言語がフランス語であるブロガーが多い」かつ「観光タイプより、景観を目的とした内容の旅行ブログエントリが多い」と明らかとなれば、広島平和記念資料館などの観光施設までの公共交通機関にフランス語の案内板などの設置が必要である事がわかる。このように、属性に基づいた分析により、旅行者や各都市の特徴を明らかにし、観光施策の展開を手助けすることを目指す。

本研究の構成は以下の通りである。2 節では関連研究を紹介する。3 節では旅行ブログエントリからの属性の自動判定について、4 節では属性の自動判定における実験と考察、5 節では属性に基づいた分析について述べる。そして、6 節では本稿のまとめについて述べる。

2. 関連研究

観光施策を展開するためには、観光地のマーケティングが必要不可欠であり、一般的にアンケート調査が用いられてきた。アンケート調査により旅行者の分析を行っている研究として、林ら[2]の研究がある。林らは、関西空港国際線出発ロビーにて、出国待ちの旅行者 1,014 名を対象に、訪問国や旅行日数、観光目的などの項目に対してアンケート調査を行い、旅行動機を明らかにすることを試みた。また、Xia ら[3]はフィリッピン島を訪れた旅行者 464 名を対象に性別、観光地、居住地などの項目のアンケートを行い、決定木を用いて旅行者の行動分析を行っている。Jonsson ら[4]も同様にアンケート調査により、163 名の性別や旅行の動機などの情報を収集し、分析を行っている。しかし、アンケート調査は、人手により行われており、時間や労力といったコストが掛かってしまう問題点がある。そこで、本研究ではソーシャルメディアの 1 つである

旅行ブログエントリを用いて分析を行う。旅行ブログエントリでは、旅行者の実体験や感想が率直に記述されており、これらを分析することは旅行者の観点に基づいた分析であると言える。

ソーシャルメディアを用いて分析を行っている研究では、藤原ら[5]の研究がある。藤原らは、YouTube、Twitter、ブログを用いて、お祭りやイルミネーションといった観光イベントに関する分析を行っている。分析結果では、イルミネーションが混雑する時間帯や状況、環境を明らかにし、観光イベントの活性化につながる知見を得ている。また、佐伯ら[6]は、外国人に向けたプロモーションの支援を行うため、様々な言語で記述された Twitter を利用している。記述された言語圏の違いから分析を行い、「原宿はドイツ語圏の旅行者には人気だが、韓国語圏の旅行者には人気でない」など、言語圏によって、特徴の違いがあることを明らかにしている。そのため本研究でも、旅行ブログエントリに記述されている使用言語の情報をを用いて分析を行い、使用言語による特徴を明らかにしていく。

近年では、旅行ブログエントリを対象に観光分析を行っている研究が多数存在している。旅行ブログエントリは、作成や読み手とのコミュニケーションが容易であり、匿名性も確保されやすいことから、旅行や観光の記録や感想、意見を表現する手段として扱われている。そのため、旅行ブログエントリは、ブLOGGER独自の観点で訪問地の評価をしていると言える。また、観光マーケティングにおいて、従来の人伝えの口コミと旅行ブログエントリを比較した場合、全体的な信頼性は同一ではないが、旅行ブログエントリは、観光マーケティングや観光消費者の促進の手助けとして活用することが可能である[7][8]。Li ら[9]は、中国のポータルサイトの旅行ブログエントリを用いて台湾の観光客から見た中国のイメージ分析を行っており、「景観」や「買い物」、「宿泊」などのカテゴリごとに分類し、旅行ブログエントリの記述された内容を分析している。分析により、温泉に関する旅行ブログエントリは非常に少ないが、それについて書かれている記事全てが好印象であり、今後、温泉を観光スポットとして力を入れていく必要であることがわかった。同様に神田ら[10]は、世界遺産登録された「石身銀山遺跡とその文化的景観（島根県大田市）」に関する旅行ブログエントリを対象とし、記述内容を人手により「見る」や「食べる」、「泊まる」などのカテゴリに分類して分析している。その結果、「食べる」において、出雲そばや海の幸といった単語が頻繁に出現しており、地域の伝統的な食材に力を入れる必要があることが分かった。このように、旅行ブログエントリを利用した分析では、有益

かつ実用的な知見が得られている。旅行ブログエントリを分析する際に、「景観」や「買い物」、「宿泊」などのカテゴリに分類して分析する点で、本研究と非常に類似している。しかし、本研究では、これらの観光タイプの分類は従来研究のように人手ではなく、自動的に分類して分析を行う点で異なる。本研究と非常に類似している研究では、Wenger[1]や村上ら[11]の研究が挙げられる。Wenger は、旅行ブログエントリから、性別や年齢といった属性を人手により収集して分析しており、属性を収集する点で本研究と類似している。また、村上らの研究では、訪日旅行者から見た日本について分析するため、海外の旅行ブログのポータルサイトである Travel Blog¹を利用している点で類似している。しかし、本研究では、属性を自動判定することによって、人手で困難であった大量の旅行ブログエントリを対象に、分析することが可能となる。

3. 旅行ブログエントリからの属性の自動判定

本研究では、観光行政の支援に繋がる知見を得るため、大量の旅行ブログエントリを利用して、旅行者や各都市における特徴を明らかにすることを目的としている。例えば、外国人向けの観光案内板やパンフレットを作成する際、旅行者の使用言語を明らかにすることにより、観光施策の展開に役立てることができる。また、性別により、観光目的がどのように異なるかを明らかにすることにより、男性と女性それぞれに対して新たな取り組みを行うことが可能となる。そのため本研究では、まず大量の旅行ブログエントリから、分析の対象となる属性の自動判定を行う。自動判定する属性は以下の3種類である。

- **性別**

ブLOGGERの性別に関する情報であり、ブLOGGERが記述した旅行ブログエントリの集合から判定する。

- **使用言語**

ブLOGGERが使用する言語に関する情報であり、ブLOGGERが記述した旅行ブログエントリの集合から判定する。

- **観光タイプ**

旅行ブログエントリに記述されている観光目的に関する情報であり、旅行ブログエントリから観光の主な目的となる5種類の目的へ判定する。

以上の3種類の属性を自動的に判定し、判定された属性に基づいて分析を行っていく。これにより、人手では困難であった大量の旅行ブログエントリを利用した分析が可能となる。

これらの属性を判定するにあたり、本研究では、旅行ブログエントリが登録されている Travel Blog を利

¹ <https://www.travelblog.org/>

用する。Travel Blog では、旅行ブログエントリを投稿する際に、訪問地に関する情報をあらかじめ決めて投稿する仕様となっている。そのため、訪問地の情報を容易に取得できる。また、Travel Blog には、ブロガーのプロフィールページが設けられており、自由に記述することができる。プロフィールページには、性別などの情報が記述されているが、それらの情報を載せているブロガーはごくわずかであり、無記入のブロガーも少なくない。そのため本研究では、旅行ブログエントリに着目して属性の判定を行う。

属性の自動判定については、3.1 節では性別の判定、3.2 節では使用言語の判定、3.3 節では観光タイプの判定について、それぞれ説明する。

3.1. 性別の自動判定

本研究では、ブロガーの基本情報である性別に基づいて分析を行う。ブロガーの性別による訪問地の違いを明らかにすることにより、男性と女性のどちらにプロモーションを行えばよいかの判断が容易になる。例えば、ある訪問地において、男性の旅行者が少ないと分かれば、男性に対してのプロモーションが必要であることがわかる。

そこで本研究では、ブロガーの性別の判定を行うため、2つの手法を用いる。1つ目の手法は、Ikeda ら [12]が提案した半教師有り学習(SSL: Semi-Supervised Learning)による手法を用いる。Ikeda らは、ブロガーごとにライティングスタイルがあると仮定している。例えば、男性のブログではアクティブな活動が多く記述され、女性のブログではコスメなどの美容に関する話題が多いかもしれない。こういったライティングスタイルのようなブロガーの特徴を教師無しブログから得て、ブログ分類を行っている。つまり、教師無しブログが多いほど、様々な側面から評価したブログの特徴を得ることができ、教師有りブログが少量でも、大量の教師無しブログを用いることにより、教師有りブログの数を補える学習が可能となる。Ikeda らは、教師有りブログの数が少ない条件下でブロガーの性別の判定実験を行っている。その結果、教師有りブログのみの学習では正解率約 0.760、一方、提案手法である半教師有り学習を用いた手法では正解率約 0.890 と高い結果を得ている。そのため本研究でも、教師有りブログの特徴を教師無しブログから得た半教師有り学習を用いて性別の判定を行う。

2つ目の手法は、単語の出現頻度(TF : Term Frequency)を用いた手法である。人手により、男性と女性に判定された旅行ブログエントリから、TF 値を算出し、男性と女性のそれぞれに頻繁に出現する単語を収集し、収集された単語を機械学習の素性として、性別

の判定を行う。収集された単語として、男性の場合、人の名前や地名などの固有表現が多く収集された。女性の場合、「pm」や「am」などの時間を表す単語が多く収集された。

3.2. 使用言語の自動判定

Travel Blog に投稿されている旅行ブログエントリでは、主に英語によるものが多いが、フランス語やドイツ語など様々な言語で記述された旅行ブログエントリも存在する。そこで、旅行ブログエントリに使用された言語に着目し、ブロガーの使用言語による、訪問地の傾向を明らかにする。これにより各訪問地において、旅行者の使用言語の傾向が分かれば、観光案内の説明板やパンフレットの作成の取り組みなどに対して、より具体的な施策を支援することができる。

本研究では、ブロガーがどこの国の言語を使用しているかを判定するため、Java のライブラリである言語判定器(langdetect²)を用いる。ある言語で記述されたテキストに対して、53 言語について 99%以上の精度で使用言語の推定が可能であり、入力をテキストデータとし、使用言語とその確率を出力する。本研究では、使用言語の判定精度を検討するため、2つの手法により、ブロガーの使用言語の判定を行う。1つ目の手法は、langdetect により得られた確率が最も高い使用言語をブロガーの使用言語とする手法(Top)である。具体的な流れを図 1 に示す。まず、ブロガーが投稿した旅行ブログエントリに対して langdetect を使用する。ブロガーが投稿した旅行ブログエントリの集合から、langdetect により得られた使用言語とその確率の平均を求める。そして、平均の確率が最も高かった使用言語をブロガーの使用言語とする手法である。図 1 の場合、ブロガーの使用言語は英語となる。

2つ目の手法は、閾値を設けた手法(Threshold)である。具体的には、各言語の平均の確率に対して、閾値を設ける手法である。この手法では、ブロガーは複数の言語を使用することとなる。例えば、図 1 の場合、

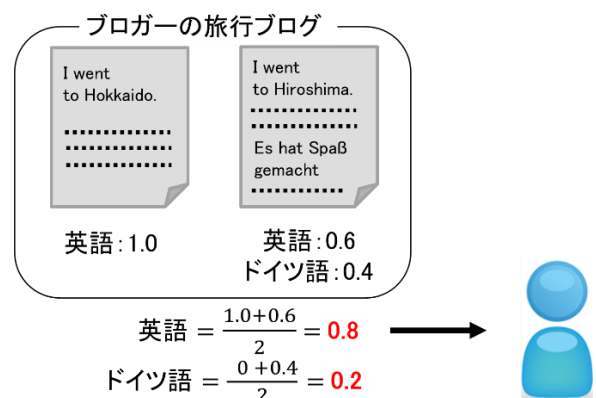


図 1: ブロガーの使用言語の判定

² <https://code.google.com/p/language-detection/>

閾値を 0.1 以上と設定した際、プログラマーの使用言語は英語とドイツ語となる。

3.3. 観光タイプの自動判定

同じ訪問地であっても、旅行者によって訪れた目的は様々である。ある旅行者は食事を目的に訪れ、また、ある旅行者は景観を目的に訪れている。旅行者が訪れた目的を明らかにすることにより、各訪問地の特徴を活かした観光施策を行うことができると考える。

そこで、訪問地へ訪れた目的を判定するため、藤井ら[13]の手法を用いて、旅行ブログエントリを観光の主な目的となる「買う」、「食べる」、「体験する」、「泊まる」、「見る」の 5 種類の観光タイプへ自動判定する。観光の目的を明らかにすることにより、各訪問地における特徴を捉え、訪問地の特徴を活かした施策を展開する支援につながる。観光の目的となる観光タイプの判定基準は、表 1 に示す通りである。藤井らは、情報利得(IG: Information Gain)とドメイン適応(DA: Domain Adaptation)を用いた手法により表 2 の結果を得ている。本研究でも、この手法を用いて、旅行ブログエントリを 5 種類の観光タイプへ分類することより、観光タイプの自動判定を行う。

表 1: 観光タイプの判定基準

タイプ	判定基準
買う	お土産に関する情報。
食べる	飲食に関する情報。
体験する	〇〇体験やスキューバダイビングなど、自分の体を使って楽しめる物についての情報が記載されている
泊まる	宿泊施設に関する情報。
見る	観光名所などの見て楽しめる物やイベントについての情報。
その他	上記の観光タイプに該当しない情報が記載されている場合。

表 2: 観光タイプの判定結果

	精度	再現率
買う	0.250	0.094
食べる	0.810	0.473
体験する	0.741	0.295
泊まる	0.410	0.149
見る	0.773	0.672
平均	0.597	0.327

4. 実験

本研究では、提案した手法の有効性を確認するため、性別の判定と使用言語の判定の 2 種類の実験を行った。

実験の詳細については、それぞれ、4.1 節、4.2 節で述べる。また、観光タイプの判定については、藤井ら[13]の手法を用いて判定を行うが、藤井らと同様の実験条件であるため、本節での実験は行わない。

本実験で使用する旅行ブログエントリでは、3 節で述べた Travel Blog を用いた。Travel Blog では、自由記述によるプログラマーのプロフィールページが設けられており、性別などの正解データの作成には、プロフィールページを用いて行った。また、性別の判定実験では、英語により記述された旅行ブログエントリのみを対象に行った。

4.1. 性別の判定実験

4.1.1. 実験条件

【実験に用いるデータ】男性 77 人と女性 151 人の合計 228 人のプログラマーを使用した。上記のデータに対し、人手により性別判定を行った結果を実験に使用した。

【機械学習と評価尺度】機械学習を用いて性別の判定を行った。機械学習には TinySVM³を用いた。線形カーネルを使用し、2 分割交差検定を行った。評価尺度には、正解率を使用した。

【実験手法】提案手法の有効性を確かめるため、以下に示す提案手法について実験を行った。

- **Baseline**: 全てのプログラマーを女性と判定した場合。
- **SSL**: 教師無しブログから教師有りのブログの特徴を捉える半教師有り学習を用いた手法。
- **TF**: 男性・女性のそれぞれに頻繁に出現した単語を素性として与える。
- **SSL+TF**: 半教師有り学習の手法(SSL)に頻繁に出現する単語(TF)を素性として与える。

4.1.2. 実験結果と考察

提案手法により、得られた実験結果を表 3 に示す。半教師有り学習と単語の出現頻度を用いた SSL+TF 手法では、全ての手法の中で最も高い正解率を得た。

表 3: 性別の判定結果

手法	正解率
Baseline 手法	0.662 (151/228)
SSL 手法	0.667 (152/228)
TF 手法	0.776 (177/228)
SSL+TF 手法	0.877 (195/228)

性別の判定実験での SSL 手法、TF 手法、SSL+TF 手法の実験結果について考察を行う。SSL 手法について、Ikeda らも同様に性別の判定実験を行っていたが、その際の正解率は約 0.890 であった。しかし、本実験では、期待していた結果を得ることが出来なかった。この原

³ <http://chasen.org/~taku/software/TinySVM/>

因として、対象としたデータの違いによるものであると考える。Ikedaらが実験の対象としたブログエントリは、スポーツや政治などの様々なトピックを含んでおり、このトピックの傾向なども有益な情報である。一方、本研究では、旅行や観光に限定したブログエントリを対象に性別の判定を行った。そのため、ブログエントリの特徴を正確に捉えることができず、性別判定の正解率が向上しなかったと考えられる。そのため、TF手法では、旅行ブログエントリのみを対象とし、男性と女性に頻繁に出現する単語を性別判定の素性として用いているため、正解率の向上が見られたと考える。

4.2. 使用言語の判定実験

4.2.1. 実験条件

【実験に用いるデータ】ブロガー109人に対し、人手によりブロガーの使用言語の判定を行った結果を実験に用いた。人手により判定された使用言語とブロガー数を表4に示す。ただし、1人に対して複数の言語を使用するブロガーも存在する。

表4: 人手による使用言語とブロガー数

使用言語	人数	使用言語	人数
英語	83	ポルトガル語	1
ドイツ語	10	スウェーデン語	1
スペイン語	9	アフリカーンス語	1
オランダ語	9	ハンガリー語	1
フランス語	6	フィンランド語	1
デンマーク語	5	スロベニア語	1
イタリア語	2	ルーマニア語	1
日本語	2		

【評価尺度】評価尺度には、精度と再現率を用い、本研究では再現率よりも精度を重視する。判定精度が低い場合、正確な分析が行えないためである。また、旅行ブログエントリは、日々作成され膨大に存在するため、対象とする旅行ブログエントリを増やすことにより、再現率の低さを補うことができる。

【実験手法】既存のライブラリ langdetect を用いて、使用言語の判定を行っていくが、判定精度を検討するため、以下の手法により実験を行った。

- **Baseline** : langdetect により判定された全ての使用言語をブロガーの使用言語とした場合。
- **Top** : langdetect により判定された使用言語において、最も高い確率を持つ言語をブロガーの使用言語とした場合。
- **Threshold** : langdetect により判定された使用言語において、その使用言語の確率に閾値を用いた場合。なお、閾値は予備実験により決定した。

4.2.2. 実験結果と考察

提案手法により得られた使用言語の判定結果を表5に示す。Baseline手法の精度に比べ、Top手法とThreshold手法の精度は、それぞれ、0.472ポイント、0.387ポイント向上した。特に、Top手法では、最も高い精度を得ることが出来た。

表5: 使用言語の判定結果

手法	精度	再現率
Baseline 手法	0.500	0.925
Top 手法	0.972	0.797
Threshold 手法	0.887	0.887

使用言語の判定結果について考察を行う。3.2節で述べたが、langdetectの言語判定の精度は99%である。しかし、Baseline手法であるlangdetectにより判定された全ての使用言語をブロガーの使用言語とした場合の精度は0.500であった。この理由は、文章量に関係している。langdetectでは、ある程度の長さの文章に対して、高精度で判定することができる。しかし、本研究で対象としたデータでは、写真をメインとした短い文書の旅行ブログエントリも少なくない。上記の理由により、Baseline手法の精度は0.500であった。一方、Top手法やThreshold手法では、高い精度を得ることができた。本研究では、最も高い精度を得たTop手法を用いて、使用言語の判定を行う。Top手法では、1人のブロガーに1つの使用言語のみしか対応つけることができないが、正確な分析を行うため、Top手法により使用言語の判定を行う。

5. 属性に基づいた分析

本研究は、日本に訪れた外国人旅行者や日本の各都市の特徴を明らかにするため、日本国内の旅行や観光について記述された旅行ブログエントリ7,490件と旅行者であるブロガー1,302人を対象に分析を行う。

はじめに、上記のデータを対象とし、属性の自動判定を行う。そして、自動判定された属性に基づいて分析を行い、外国人旅行者や各都市の特徴を明らかにしていく。詳細については、それぞれ、5.1節と5.2節で述べる。

5.1. 属性の自動判定

属性に基づいた分析を行うため、旅行ブログエントリから性別、使用言語、観光タイプの各属性を自動判定する。性別と使用言語の自動判定では、4節より有効性が確認できた各手法を用い、観光タイプの自動判定では、藤井らの手法を用いて行う。各属性における、属性判定の対象としたデータと判定された件数について述べていく。

【性別の自動判定】ブロガー1,302 人を対象に性別の自動判定を行った。自動判定された結果を表 6 に示す。

表 6: 自動判定によるブロガーの性別

性別	ブロガー数 (比率)
男性	513 (0.339)
女性	789 (0.661)

【使用言語の自動判定】ブロガー1,302 人を対象に使用言語の自動判定を行った。自動判定された結果を表 7 に示す。ただし、表 7 は、英語を除いた上位 5 件についての結果である。本研究で対象としている Travel Blog では、英語で記述するブロガーが多い。そのため、使用言語に基づいた分析では、使用言語が英語以外の旅行ブログエントリを対象に分析を行う。

表 7: 自動判定によるブロガーの使用言語

使用言語	ブロガー数
フランス語	22
ドイツ語	13
オランダ語	10
スペイン語	7
フィンランド語	7

【観光タイプの自動判定】旅行ブログエントリ 7,490 件を対象に観光タイプの自動判定を行った。自動判定された結果を表 8 に示す。ただし、複数の観光タイプに判定された旅行ブログエントリも存在する。表 8 より、観光タイプ「買う」に判定された旅行ブログエントリは 1 件であった。そのため、観光タイプに基づいた分析では、観光タイプ「買う」を除いて行う。

表 8: 自動判定による旅行ブログエントリの観光タイプ

観光タイプ	旅行ブログエントリ数
買う	1
食べる	1,134
体験する	315
泊まる	319
見る	3,213

5.2. 属性に基づいた分析

5.1 節により、自動判定された属性に基づいて分析することにより、訪日外国人旅行者や日本国内の各都市の特徴を明らかにしていく。

【使用言語に基づいた分析】使用言語に基づいて、旅

行者が訪れた都市について分析を行った。最も様々な使用言語の旅行者に訪れられている都市は首都である東京であり、18 カ国の使用言語の旅行者 68 人により訪れられている。しかし、東京周辺の都市である千葉や神奈川では、それぞれ 3 カ国の使用言語の旅行者 6 人、8 カ国の使用言語の旅行者 10 人が訪れており、東京と比べて訪れた旅行者が少ないがわかる。このため、東京周辺の都市では、東京に訪れた多様な使用言語の旅行者に対するプロモーションが必要である。一方で、京都や広島では、多様な使用言語の旅行者が訪れている。京都では 13 カ国の使用言語の旅行者 31 人、広島では 10 カ国の使用言語の旅行者 21 人が訪れている。そのため、使用言語と観光タイプに基づいた分析にて、広島と京都について分析を行う。

【観光タイプに基づいた分析】観光タイプに基づいて、旅行者の訪れた目的について分析を行った。各観光タイプの割合が高い上位 5 つの都市を表 9 に示す。また、観光タイプ「体験する」と「泊まる」のヒートマップをそれぞれ、図 2 と図 3 に示す。

表 9: 各観光タイプの割合が高い上位 5 都市

順位	食べる	体験する	泊まる	見る
1	宮崎	新潟	群馬	奈良
2	佐賀	静岡	佐賀	広島
3	埼玉	鹿児島	滋賀	香川
4	宮城	山梨	大分	福井
5	岐阜	長野	熊本	岡山

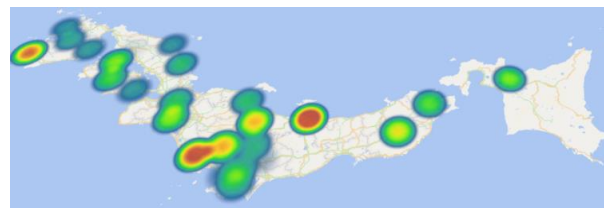


図 2: 観光タイプ「体験する」のヒートマップ



図 3: 観光タイプ「泊まる」のヒートマップ

ヒートマップでは、色が赤い都市ほど、その観光タイプを目的に訪れている旅行者の割合が高いことを示す。表 9 より、観光タイプ「食べる」では、主要都市から外れた地方の都市が上位であり、観光タイプ「見

る」では、世界遺産が存在する奈良や広島が上位であった。表9と図2より、観光タイプ「体験する」において、新潟では、スキーなどのウィンタースポーツを楽しめ、静岡では、富士山へ登山することができる。また、山梨では、様々なギネス記録を更新する遊園地がある。また、表9と図3より、関東や関西圏には多くの旅行者が泊まることを目的に訪れているが、九州地方でも多くの旅行者が泊まることを目的に訪れていることが明らかとなった。

【性別と観光タイプに基づいた分析】日本全体と愛媛における性別による観光タイプの割合を示した円グラフを図4に示す。図4(a)に示すように、日本全体における男性と女性の観光タイプでは、大きな違いは見られなかった。しかし、都市に着目すると、男性と女性により、訪れた目的が異なることが分かる。図4(b)は、愛媛県における男性と女性の観光タイプの割合であるが、男性の旅行者は食べることを目的に訪れており、一方、女性の旅行者は温泉などの体験することや宿泊することを目的に訪れている傾向にある。このように、旅行者の性別によって、同じ都市に訪れたとしても目的が異なる。そのため、観光施策を展開するにあたり、男性と女性それぞれ別のプロモーションが必要であることがわかった。例えば、図4(b)に示した愛媛県の場合だと、温泉や旅館に女性が喜ぶアメニティグッズを揃えるなどの取り組みにより、旅行者の増加に繋がると考えられる。

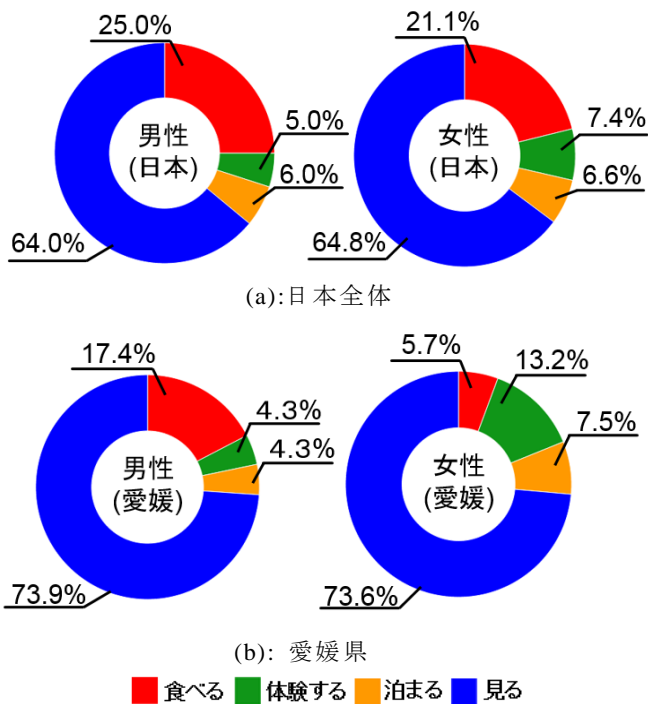


図4: 性別による観光タイプの割合

【使用言語と観光タイプに基づいた分析】使用言語に基づいた分析により、京都と広島に多様な使用言語の旅行者が訪れていることが分かった。そこで、京都と広島に着目して、使用言語と観光タイプに基づいた分析を行う。表9より、広島では、観光タイプ「見る」の占める割合が高い。また、京都でも同様に、観光タイプ「見る」の占める割合が高く、どちらの都市も、見ることを目的に訪れている旅行者が多い。ここで、使用言語と観光タイプ「見る」に関するグラフを図5に示す。

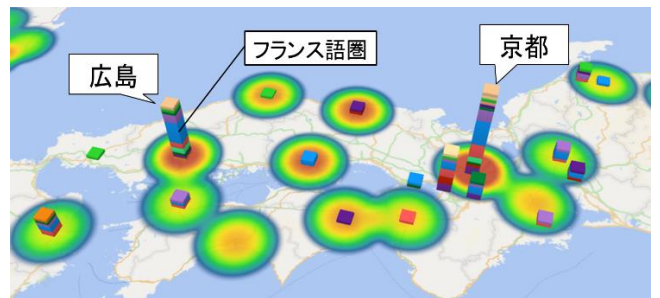


図5: 使用言語と観光タイプ「見る」に基づいた分析

棒グラフは旅行者の使用言語を表しており、ヒートマップは観光タイプ「見る」である。図5より、これら2つの都市には、どちらも世界遺産を保有しており、見ることを目的に多くの旅行者が訪れていることが分かる。さらに、他の都市に比べ、様々な使用言語の旅行者が訪れている。つまり、世界遺産がある都市では、多様な使用言語の旅行者が訪れると言える。また、旅行者の使用言語を表している棒グラフに着目すると、使用言語をフランス語圏とする旅行者の数が多いため、使用言語をフランス語圏とする旅行を対象に、各都市の割合について調べた。フランス語圏の旅行者が訪れた各都市の割合を図6に示す。

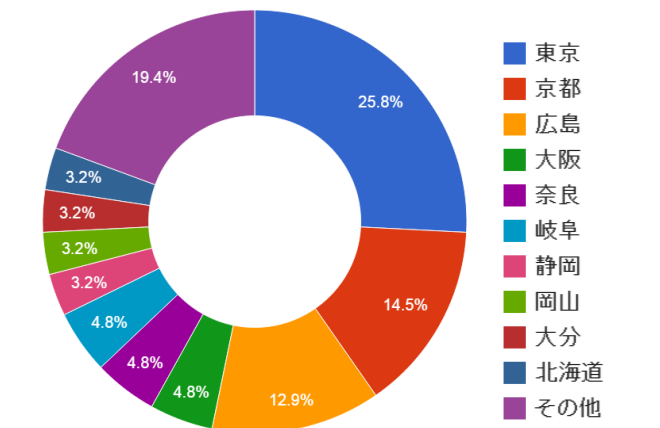


図6: フランス語圏の旅行者が訪れた各都市の割合

図 6 より、京都と広島には、フランス語圏の旅行者のうち、約 3 割の旅行者が訪れている。また、観光タイプ「見る」の占める割合が最も高かった奈良にも世界遺産があり、京都と広島に続いて多くの旅行者が訪れている。このことから、フランス語圏の旅行者は、世界遺産がある都市に多く訪れる傾向にあると言える。今後、世界遺産を保有している都市では、フランス語の説明板の設置などの取り組みにより、さらなる旅行者の増加に繋がると考えられる。

6. おわりに

本研究では、旅行ブログエントリを利用して、訪日外国人旅行者や各都市の特徴を明らかにした。そのために、まず分析の対象となる属性「性別」や「使用言語」、「観光タイプ」の自動判定を行った。性別の自動判定では、半教師有り学習と単語の出現頻度を用いた手法を組み合わせた SSL+TF 手法により、正解率 0.877 を得た。使用言語の自動判定では、言語判定器により最も高い確率を持つ言語をブロガーの使用言語とする Top 手法により、精度 0.972、再現率 0.797 を得ることができた。また、観光タイプの自動判定では、藤井らと同様の実験条件で行っており、精度 0.597、再現率 0.327 である。次に、日本国内に関する旅行ブログエントリを対象に、属性の自動判定を行い、判定された属性に基づいて分析を行った。その結果、日本の各都市に訪れる旅行者の目的を捉えることができ、各都市の特徴を明らかにした。さらに、性別によって訪れる目的が異なることがわかった。また、フランス語圏の旅行者は、世界遺産がある都市に多く訪れていることなど、訪日外国人旅行者や日本国内における各都市の特徴を明らかにした。

今後の課題として、分析に用いる属性を増やすことが必要である。年齢や居住地、旅行形態、旅行ブログエントリの投稿時期など自動的に判定し[14][15]、新たな属性に基づいて分析することにより、本研究では得ることが出来なかった知見を得ることができると考える。本研究では、属性に基づいて分析を行ったが、旅行ブログエントリの内容を用いることによって[16][17]、より具体的な分析が可能になると考える。また、本研究により得られた分析結果について、観光庁などの調査により得られた旅行者の統計と比較を行い、それぞれの違いを調べる必要がある。

参 考 文 献

- [1] Wenger, A.: Analysis of Travel Bloggers' Characteristics and their Communication about Austria as a Tourism Destination, *Journal of Vacation Marketing*, Vol. 14, No. 2, pp.169-176, 2008.
- [2] 林 幸史, 藤原 武弘: 訪問地域、旅行形態、年齢別にみた日本人海外旅行者の観光動機, 実験社会心理学研究 日本グループ・ダイナミクス学会, Vol.48, No.1, pp.17-31, 2008.
- [3] Xia, J., Ciesielski, V. and Arrowsmith, C.: Data Mining of Tourists' Spatio-temporal Movement Patterns ---A Case Study on Phillip Island, *Proc. of the 8th International Conference on GeoComputation*, pp.1-5, 2005.
- [4] Jonnson, C. and Devonish, D.: Dose Nationality, Gender, and Age Affect Travel Motivation? A Case of Visitors to the Caribbean Island of Barbados, *Journal of Travel and Tourism Marketing*, Vol.25, No.3-4, pp.398-408. 2008.
- [5] 藤原 泰士, 難波 英嗣, 竹澤 寿幸: ソーシャルメディアの分析によるイベント開催支援, 第 6 回データ工学と情報マネジメントに関するフォーラム (DEIM2014), 2014.
- [6] 佐伯 圭介, 村山 敬祐, 遠藤 雅樹, 横山 昌平, 石川 博: 外国人向け観光プロモーション支援のためのソーシャルメディア分析手法の研究, 第 6 回データ工学と情報マネジメントに関するフォーラム (DEIM2014), 2014.
- [7] Mack, R. W., Blose, J. E. and Pan, B.: Believe it or not: Credibility of Blogs in Tourism, *Journal of Vacation Marketing*, Vol.14, No.2, pp.133-144, 2008.
- [8] Akehurst, G.: User Generated Content: the Use of Blogs for Tourism Organizations and Tourism Consumers, *Journal of Service Business*, Vol.3, No.1, pp.51-61, 2009.
- [9] Li, Y.R. and Wang, Y.Y.: Exploring the Destination Image of Chinese Tourists to Taiwan by Word-of-Mouth on Web, *Proc. of World Academy of Science Engineering and Technology* Vol.7, pp.977-981, 2013.
- [10] 神田 佑亮, 藤原 章正, 張 峻屹: ブログ情報を用いた観光行動と満足度の分析に関する一考察, *土木計画学研究*, 公演集, Vol.39, 2009.
- [11] 村上 嘉代子, 川村 秀憲: 外国人からみた日本旅行 -英語ブログからの観光イメージ分析-, *人工知能学会誌*, Vol.14, No.2, pp.169-176, 2011.
- [12] Ikeda, D., Takamura, H and Okumura, M.: Semi-Supervised Learning for Blog Classification, *Proc. of the 23rd AAAI Conference on Artificial Intelligence*, pp.1156-1161. 2008.
- [13] 藤井 一輝, 石野 亜耶, 藤原 泰士, 前田 剛, 難波 英嗣, 竹澤 寿幸: 多言語旅行ブログエントリを用いた観光情報提示システム, 第 6 回データ工学と情報マネジメントに関するフォーラム (DEIM2014), 2014.
- [14] Yasuda, N., Hirao, T., Suzuki, J. and Isozaki, H.: Identifying Bloggers' Residential Areas, *Proc. AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, pp.231-236, 2006.
- [15] Schler, J., Koppel, M., Argamon, S. and Pennebaker, J.: Effects of Age and Gender on Blogging, *Proc. AAAI Symposium on Computational Approaches for Analyzing Weblogs*, pp.199-205, 2006.
- [16] Li, X. and Wang, Y.: Depicting Image of China as a Tourism Destination: A Travel Blog Approach, *Proc. of the 16th Graduate Students and Graduate Education Conference in Hospitality and Tourism*, 2011.
- [17] Murakami, K., Kawamura, H. and Suzuki, K.: Analyzing the Impact of Earthquake on Tourism' Activity: Travel Blog Mining, *Journal of Advancements in Computing Technology*, Vol.4, No.22, 2012.