

マイクロブログを用いた旅行の進行状況に伴う行動変化の分析

高橋 一希[†] 加藤 大受^{†,††} 遠藤 雅樹^{†††} 莊司 慶行^{††††} 廣田 雅春^{†††††}

石川 博[†]

† 首都大学東京大学院 システムデザイン研究科 〒191-0065 東京都日野市旭が丘 6-6
 †† ウイングアーク1st株式会社 〒150-0031 東京都渋谷区桜丘町 20-1 渋谷インフォスタワー
 ††† 職業能力開発総合大学校 基盤ものづくり系 〒187-0035 東京都小平市小川西町 2-32-1
 †††† 大分工業高等専門学校 情報工学科 〒870-0152 大分県大分市大字牧 1666
 ††††† 首都大学東京 システムデザイン学部 〒191-0065 東京都日野市旭が丘 6-6
 E-mail: †takahashi-kazuki@ed.tmu.ac.jp, ††kato.d@wingarc.com, †††endou@uitec.ac.jp,
 ††††{y_shoji,ishikawa-hiroshi}@tmu.ac.jp, †††††m-hirota@oita-ct.ac.jp

あらまし 本論文では、Twitter の位置情報が付与された投稿を用いて、旅行の際に、その進行状況に応じて行動の種類が変化することを検証し、分析する。旅行には、様々な種類が存在する。中でも、観光地であるだけでなく、交通の拠点でもある東京都を含む旅行の種類は、複雑である。旅行の目的は一つとは限らず、旅行の際の行動は、旅行の進行状況や、旅行者の属性によって異なると考えられる。位置情報を用いて、ユーザの東京都を含む旅行を取得し、各投稿の行動を推定する。それらの結果を用いて、ユーザの行動の違いを検証する。

キーワード Twitter, 位置情報, 観光調査

1. はじめに

国土交通省観光庁の旅行・観光消費動向調査^(注1)によると、2015年の日本の旅行消費額全体において、日本人の日本国内旅行の消費額は20.9兆円、訪日外国人の日本旅行の消費額は3.3兆円となっており、消費額全体で日本人の消費額が大半を占めている。また、日本人国内旅行の一人あたりの旅行単価は前年度に比べて増加している。今後、日本人国内旅行者数を増加させるためには、国内旅行を活性化させることも重要な課題の1つである。日本人国内旅行消費額のうち、宿泊旅行消費額は15兆8,120億円で、前年比13.8%増、日帰り旅行消費額は4兆5,970億円で、前年比1.5%増となった。しかし、日本人国内述べ旅行者数は増加傾向にあり、宿泊旅行者数は前年度より増加している。一方、日帰り旅行者数は前年度より減少している。このことから、同じ国内旅行においても、宿泊旅行者数は多く、消費行動の割合も高いと考えられる。そこで、消費行動の促進につなげるために、国内の旅行者を日帰りの旅行者と、宿泊旅行者に分けてその行動を分析する。さらに、宿泊旅行者において、その日数に伴う行動の変化について分析する。

観光を促進するための調査として、旅行者に対するアンケート調査が多く行なわれている。アンケート調査は、しばしば、都道府県や自治体ごとに行われる。その利点として、アンケート結果から、その地域にしかない観光地やイベントに対する感想を得られることや、信頼性が高いことなどが挙げられる。しかし、アンケート調査の欠点として、準備や調査にかかる時間や人員

に対するコストが大きいこと、データ量に限りがあること、対象や設問をアンケートごとに設定しなければならないため、他の調査との比較が難しいことが挙げられる。そこで、アンケート調査の欠点を補うため、マイクロブログなどのソーシャルメディアから観光情報を抽出する研究が盛んである[1], [2], [3]。マイクロブログの投稿には、ユーザのその場の意見や感想が反映されているものが多く存在している。しかし、データ量が多い一方で、データとして整形されていないため、必要な情報を抽出する必要がある。そこで、マイクロブログのひとつであるTwitterの位置情報が付与された投稿を用いて、観光情報を抽出する研究が多くなされている。Twitterを用いることによって、コストを抑えて分析することが可能であり、対象とする場所や時間を、分析に応じて柔軟に変化させることができる。

また、旅行の内容や形態は日々変化し続けている。日本旅行業協会の調査^(注2)によると、日本の年齢別人口構成の変化に伴い、団体旅行の需要が減少し、個人旅行の需要が増加していることが示されている。また、じゃらん宿泊旅行調査2016^(注3)によると、旅行の同行形態において、一人旅の割合が2004年の調査以来11年連続で増加している。ソーシャルメディアを用いた分析は低コストで実現可能であるため、このような日々変化する旅行を把握するためにも有用である。

本研究では、日本人国内旅行を分析するために、東京都以外から東京都への旅行者の行動について分析する。東京都は日本の首都であり、国内の主要な観光地であることに加え、交通の

(注1)：国土交通省観光庁 旅行・観光消費動向調査 <http://www.mlit.go.jp/kankochou/siryou/toukei/shouhidoukou.html>

(注2)：日本旅行業協会「更なる国内旅行にむけて」 https://www.jata-net.or.jp/membership/info-japan/research/03_1st.html

(注3)：じゃらん宿泊旅行調査2016 http://jrc.jalan.net/jrc/files/research/jalasyuku_20160726.pdf

拠点としても大きな価値を持っている。2015年に北海道新幹線が開業^(注4)したことや、2016年に北陸新幹線が開業^(注5)したこともあり、国内の交通はますます便利になっている。また、東京都には、日本で旅客数が最も多い羽田空港^(注6)が存在するため、東京都は、交通の拠点として利用されていると考えられる。そこで、訪れる旅行者が多く、多様な使われ方をしている東京都は、国内旅行を分析する際のモデルケースとして適切である。

筆者らは、先行研究[4]として、東京都を訪れる観光客の中でも、東京都を旅行の主目的とした観光客と、経由地として付随的に東京都を訪れる人とは、行動が異なるという仮説のもと、分析を行った。Twitterから、東京都での投稿を含む、位置情報が付与されたツイート群を抽出し、それらを東京主目的旅行と東京経由旅行に分類した。分類結果から、東京主目的旅行と東京経由旅行とは、訪れるスポットに違いがあることが明らかになった。しかし、この先行研究では、ツイート本文を考慮せず、投稿に付与された位置情報のみを用いていたため、ユーザの行動の種類までは特定することができなかった。

そこで、本研究では、東京都を対象に、Twitterのユーザから旅行者と旅行期間を抽出し、主目的旅行と経由旅行に分類する。その後、国土交通省観光庁の旅行・観光消費動向調査をもとに、「ビジネス」、「観光」、「購買」、「食事」の4種類を定義し、各ツイートの行動を分類する。分類結果を用いて、旅行日数ごとに、旅行の進行状況に伴う行動の割合の変化を分析し、考察する。旅行の進行状況を考慮した分析を行うことによって、旅行における傾向を把握し、どのような旅行のパターンが存在するのか推測することができると考えられる。

本論文の構成は以下の通りである。2章では、関連研究について述べる。3章では、東京を対象とした旅行を分析するための行動の分類手法と、それらを用いた分析手法について述べる。4章では、旅行の進行状況に伴う、旅行中の行動の変化について分析した結果と、その考察を述べる。5章では、本研究のまとめを述べる。

2. 関連研究

本章では、関連研究について述べる。Web上の情報を用いて、観光情報を抽出する研究は、近年盛んに行なわれている。例えば、SNSに投稿された情報から旅行の感想や口コミを抽出する研究[2],[3],[5]や、旅行の内容や目的を抽出する研究[1],[6]、観光ルートの抽出や推薦をする研究[7],[8],[9]などがある。Luoらの研究[2]では、ソーシャルネットワークサービスにおける旅行関連の口コミが既存の社会関係に依存していることを示し、その通信ネットワークが高い中心性を持つことを明らかにしている。若木らの研究[1]では、目的を「対象」、「理由」、「行

動」の3つから構成されると定義し、旅行の口コミサイトの投稿を分析し、旅行の目的を抽出している。中島らの研究[7]では、観光に関連するツイートを分析し、手がかり語や品詞の特徴から、「食事」、「景観」、「行動」の3つに分類し、観光ルートを推薦する手法を提案している。倉田らの研究[9]では、Flickrに投稿されたジオタグ付き写真を収集し、写真撮影密度を標高とみため、その仮想地形の主要尾根線を抽出することで、観光ルートを抽出している。しかし、旅行者が訪れる観光スポットは、その地域を観光の主目的として訪れているか、旅行の経由地点として利用しているかによっても異なると考えられる。そこで、本研究では、旅行をその地域を観光の主目的として利用する主目的旅行と、経由地として利用する経由旅行に分類し、それぞれ分析を行う。

また、観光情報の分析に応用可能な、投稿者の行動を分析する研究として、Fujisakaらの研究[10]では、マクロブログユーザの投稿履歴から地理的パターンを抽出し、移動の基本モデルを提案している。田中らの研究[11]では、マイクロブログユーザの習慣行動に着目し、習慣行動モデルを構築することで、推定対象時間のユーザの行動を、「睡眠中」、「出勤中」、「勤務中」、「食事中」、「帰宅中」と「その他」の6種類に分類している。大原らの研究[12]では、マイクロブログユーザの定常時の時間的な行動パターンや興味・嗜好などについての分析を行い、行動や興味の変化に関するコンテンツの意外情報を抽出するためのシステムを作成している。これらの研究では、ユーザの行動の種類や変化について分析を行っているが、旅行の際の行動は習慣的なものではなく、定常時と異なるため、本研究では旅行中のユーザを対象とする。

3. 提案手法

本章では、本研究の提案手法について述べる。はじめにTwitterから旅行行程を表す一連のツイート群を取得し、それらの旅行を主目的旅行と経由旅行に分類する。その後、それぞれの旅行ツイート群の各ツイートに対し、行動を推定し、旅行を分析する。

3.1 旅行ツイートの抽出

本節では、分析に用いる旅行ツイートの抽出方法について述べる。はじめに、Twitterから、分析対象となるユーザの候補を抽出する。分析対象の都道府県で一度でもツイートしたユーザを分析対象の候補として抽出する。

次に、対象候補として抽出されたユーザの居住地をそれぞれ推定する。ユーザの投稿に付与された位置情報をもとに、ユーザが最も多くツイートした都道府県を、そのユーザの居住地と設定する。このとき、対象候補のユーザのうち、居住地が分析対象の都道府県であるユーザは分析対象外とし、残りのユーザを分析対象とする。

続いて、ユーザのツイートの位置情報の変化をユーザの移動とみなし、旅行行程を表す一連のツイート群を取得する。本論文では、それらのツイート群を旅行ツイートと呼ぶ。ユーザの居住地での投稿の一つ後に居住地以外で投稿されたツイートから、その後の居住地での投稿の一つ前の投稿までをそのユーザ

(注4)：北海道新幹線新幹線駅等を起点とする交通ネットワーク
http://www.pref.hokkaido.lg.jp/ss/stk/s-network_20160108.pdf

(注5)：北陸新幹線の概要
<http://www.ccis-toyama.or.jp/toyama/toukei/shinkansen/pdf/1-2.pdf>

(注6)：http://www.jacinc.jp/db/pdf/dom_ranking50_2427.pdf

の旅行ツイートとして抽出する。分析では、分析対象の都道府県での投稿を含む旅行ツイートを対象とする。

3.2 旅行ツイートの分類

本節では、抽出した旅行ツイートを主目的旅行と経由旅行に分類する方法について述べる。抽出した旅行ツイートに付与された位置情報をもとに、旅行内で最も多く投稿された都道府県とその投稿数を算出する。同じように旅行内で2番目に多く投稿された都道府県とその投稿数を算出する。今回は、旅行における主目的が複数存在するものを排除するため、旅行内で最も多く投稿された都道府県の投稿数が旅行内で2番目に多く投稿された都道府県の投稿数の2倍より多い旅行ツイートのみを用いる。旅行内で最も多く投稿された都道府県が分析対象の都道府県である旅行ツイートを主目的旅行とし、分析対象の都道府県以外の旅行ツイートを経由旅行とする。

3.3 行動の推定

本節では、旅行ツイート内の各投稿における行動の推定方法について述べる。各ツイートにおける行動を、野沢ら [6] の手法を用いて、分類する。分類のクラスは、「ビジネス」、「観光」、「購買」、「食事」の4つのクラスを対象とする。しかし、ユーザの投稿には、これらのクラスのどれにも属さないものが存在するため、分類の前処理として「分類対象外」のフィルタリングを行う。また、「分類対象外」のフィルタリングの後にも、どれにも属さないものが存在するため、「その他」というクラスを定義し、合計5クラスでマルチラベル SVM [13] を用いて分類する。

前処理として、行動の推定において、旅行ツイートに含まれる投稿のうち、Foursquare^(注7) のような、サービスに付随して投稿されたツイートを排除する。また、リプライ、リツイートを排除する。さらに、それ以外のツイート本文から、URL 部分と記号やマルチバイトの絵文字を削除する。その後、MeCab^(注8) を用いて形態素解析し、名詞、動詞、形容詞のみを抽出し、tf-idf を用いてベクトル化する。MeCab の辞書には、ツイートの本文中の新語に対応させるために ipadic を拡張した mecab-ipadic-NEologd^(注9) を利用した。さらに、LSA (Latent Semantic Analysis) [14] を用いて次元削減を行い、「分類対象外」をフィルタリングした後、「ビジネス」、「観光」、「購買」、「食事」、「その他」の5つのクラスに分類する。分析には、「ビジネス」、「観光」、「購買」、「食事」の4クラスのみを用いる。

3.4 日数ごとの分析手法

本節では、旅行ツイートを日数ごとに分析する手法について述べる。一つの旅行ツイートの中で、投稿日時が最も早いものを旅行の開始とみなし、その投稿日を旅行の初日とする。また、投稿日時が最も遅いものを旅行の終了とみなし、その投稿日を旅行の最終日とする。初日から最終日までの期間をもとに、 n 日旅行を定義する。ある n 日旅行をしているユーザ群を $U(n)$ とし、 $U(n)$ に属するユーザ u の m 日目のツイートを $T(u, m)$

と表す。あるカテゴリに属する n 日旅行の m 日目のツイート群を $C(T(u, m), c)$ と表す。このとき、 c はカテゴリのラベル、すなわち「ビジネス」、「観光」、「購買」、「食事」である。 n 日旅行の m 日目のあるカテゴリでの投稿の割合を

$$P(u, m, c) = \frac{\sum |C(T(u, m), c)|}{T(u, m)} \quad (1)$$

で表す。分析には、1日旅行から5日旅行までを用いる。

4. 実験

本章では、提案手法を用いて分析した結果とその考察について述べる。今回は、旅行の進行状況に伴う行動の評価において、旅行の種類による評価と、旅行者の居住地による評価の2種類を行った。旅行の種類による評価では、提案手法による分類結果を用いて、東京主目的旅行と東京経由旅行における違いを分析した。旅行者の居住地による評価では、旅行者の居住地が関東地方である東京旅行と、旅行者の居住地が九州地方である東京旅行に分けて、その違いを分析した。

4.1 データセット

はじめに、本論文で用いたデータセットについて述べる。本論文では、2015年5月1日から2016年1月31日までの間に投稿された位置情報付きツイートより、東京都で一度以上ツイートしたユーザを取得し、それらのユーザの旅行ツイートを取得した。Twitter からのツイートの取得には、Twitter Streaming API^(注10) を用いた。取得したユーザ数は、67,786 人であった。全てのユーザの旅行ツイートの総数は、192,604 件であり、全ての旅行ツイート内のツイート数は、1,275,564 件であった。また、取得した旅行を提案手法によって、東京主目的旅行と東京経由旅行に分類した。東京主目的旅行の件数は、138,962 件であり、東京主目的旅行に含まれるツイート数は、615,859 件であった。東京経由旅行の件数は、10,633 件であり、東京経由旅行に含まれるツイート数は、180,162 件であった。

4.2 行動の分類の性能評価

次に、取得したツイートの行動の分類結果について述べる。取得したツイートのうち、前処理で排除されなかったツイート数は、699,121 件であった。分類には、Python の機械学習パッケージである scikit-learn [15] を用いた。行動の分類のためのトレーニングデータには、野沢らが人手で作成した、4,849 件のトレーニングデータを用いた。トレーニングデータのラベルは、「ビジネス」が 702 件、「観光」が 1,073 件、「購買」が 659 件、「食事」が 928 件、「その他」が 1,745 件であった。

各クラスへの分類の前処理にあたる、「分類対象外」フィルタリングの分類において、5 交差検定を行った性能の評価を表 1 に示す。「分類対象外」フィルタリングの精度 (Accuracy) は、0.75 であった。「分類対象外」フィルタリングの F 値は、0.77 であった。

次に、行動の分類において、5 交差検定を行った性能の評価を表 2 に示す。行動の分類の精度は、0.44 であった。行動の分類の F 値は、0.56 であった。

(注7) : <https://ja.foursquare.com/>

(注8) : <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

(注9) : <https://github.com/neologd/mecab-ipadic-neologd>

(注10) : <https://dev.twitter.com/overview/documentation>

表 1 「分類対象外」フィルタリングの性能評価

属性名	適合率	再現率	F 値
分類対象	0.81	0.86	0.83
分類対象外	0.68	0.61	0.65
avg / total	0.77	0.77	0.77

表 2 行動の分類の性能評価

ラベル名	適合率	再現率	F 値
ビジネス	0.76	0.48	0.59
観光	0.52	0.29	0.37
購買	0.70	0.54	0.61
食事	0.77	0.45	0.57
その他	0.71	0.64	0.67
avg / total	0.69	0.48	0.56

表 3 東京主目的旅行と東京経由旅行の行動の割合

	東京主目的旅行	東京経由旅行
ビジネス	0.131	0.168
観光	0.264	0.135
購買	0.182	0.268
食事	0.423	0.429

これらの分類器を用いて、取得したツイートの行動をそれぞれ分類した後、その分類結果を用いて分析を行った。

4.3 旅行の種類による行動の評価

本節では、東京主目的旅行と東京経由旅行それぞれの行動の割合の実験結果について述べる。それぞれの行動の割合を表 3 に示す。また、旅行日数ごとの旅行の進行状況による行動の割合の変化を表 4 に示す。表 3 において、東京主目的旅行、東京経由旅行ともに、「食事」の割合が最も大きくなった。また、東京主目的旅行で二番目に割合が大きかったものは「観光」であり、東京経由旅行で二番目に割合が大きかったものは「購買」であった。

4.4 旅行者の居住地による行動の評価

本節では、居住地が関東地方であるユーザの東京旅行と居住地在九州地方であるユーザの東京旅行それぞれの行動の割合の実験結果について述べる。この際、主目的旅行、経由旅行についての旅行の種類は考慮しない。それぞれの行動の割合を表 5 に示す。また、旅行日数ごとの旅行の進行状況による行動の割合の変化を表 6 に示す。表 5 において、関東地方居住者旅行、九州地方居住者旅行ともに、「食事」の割合が最も大きくなった。また、関東地方居住者旅行で二番目に割合が大きかったものは「観光」であり、九州地方居住者旅行で二番目に割合が大きかったものは「購買」であった。

5. 考 察

本章では、旅行の種類による行動の評価、旅行者の居住地による行動の評価それぞれの結果に対する考察を述べる。

5.1 旅行の種類による行動の評価に対する考察

本節では、旅行の種類による行動の評価に対する考察を述べる。表 3 より、「観光」の割合は、東京主目的旅行のほうが大きい。また、「ビジネス」、「購買」の割合は、東京経由旅行のほう

が大きい。「食事」の割合は、大きな差は表れなかった。

表 3 において、「観光」の割合が東京主目的旅行の方が大きい原因として、旅行者は旅行の主目的となる観光地で観光に関する投稿を多くしていると考えられる。また、Twitter の性質として、興味・関心があるものに対する投稿が多く、興味・関心がないものに対しては投稿をしないという点が挙げられる。観光には、興味の対象となるものが存在するため、その場での感想を投稿しやすい。この性質により、ユーザが旅行の中で興味・関心の高い状態にある主目的において、「観光」の割合が大きくなったと考えられる。

表 3 において、「ビジネス」、「購買」の割合が東京経由旅行のほうが大きい原因としては、経由旅行においては、観光の主目的が他に存在するため、「観光」に関する投稿が少ないためである可能性がある。また、東京経由旅行において、ユーザの訪れるスポットとして駅や空港などの交通の拠点となる地域が多いことが、筆者らの先行研究 [4] で明らかになっている。駅や空港などのスポットは、お土産を買うための売店などが多く存在しているため、「購買」の割合が高くなった可能性がある。

次に、旅行の進行状況に伴う行動の変化に対する考察を述べる。表 4 より、3 日旅行から 5 日旅行において、「観光」は、初日と最終日に割合が低い傾向にあることがわかる。これは、3 日以上宿泊旅行において、旅行の中日に観光を行う旅行者が多いためであると考えられる。「ビジネス」は、初日と最終日に割合が高い傾向にあることがわかる。これは、旅行の前後にビジネスを行う旅行者が多いためであると考えられる。また、「購買」は、初日と最終日に割合が高い傾向にあることがわかる。これは、初日や最終日に駅や空港などで、買い物をする旅行者が多いためであると考えられる。「食事」は、全体的に割合の変化が小さい。これは、食事は生活の中で必須なものであるため、旅行日数によって投稿量の変化が少ないためであると考えられる。また、東京経由旅行の方が、旅行の進行状況に伴う行動の変化が大きい傾向にあった。

5.2 旅行者の居住地による行動の評価に対する考察

本節では、旅行者の居住地による行動の評価に対する考察を述べる。表 5 より、「観光」の割合は、関東地方居住者旅行のほうが大きい。また、「ビジネス」の割合は、九州地方居住者旅行のほうが大きい。「購買」、「食事」の割合は、大きな差は表れなかった。

表 5 において、「観光」の割合が関東地方居住者旅行の方が大きく、「ビジネス」の割合が九州地方居住者旅行のほうが大きい原因として、九州地方居住者の方がビジネス目的で東京を訪れることが多いことが予想される。

次に、旅行の進行状況に伴う行動の変化に対する考察を述べる。表 6 より、関東地方居住者旅行は、旅行の進行状況に伴う行動の変化が小さい傾向にあった。九州地方居住者旅行は、旅行の種類による行動の評価と同様に、旅行の初日と最終日に「ビジネス」、「購買」の割合が大きい傾向にあり、「観光」の割合が小さい傾向にあることがわかる。また、旅行の種類による行動の評価と比べ、割合の変化が大きい傾向にあった。関東地方は、九州地方と比べて東京までの距離が短いため、関東地方

表 4 主目的旅行と経由旅行の行動の割合の変化

		東京主目的旅行					東京経由旅行				
		1日目	2日目	3日目	4日目	5日目	1日目	2日目	3日目	4日目	5日目
1日旅行	ビジネス	0.128					0.124				
	観光	0.258					0.106				
	購買	0.187					0.352				
	食事	0.426					0.418				
2日旅行	ビジネス	0.137	0.137				0.15	0.169			
	観光	0.28	0.256				0.109	0.133			
	購買	0.187	0.195				0.28	0.292			
	食事	0.396	0.413				0.461	0.406			
3日旅行	ビジネス	0.141	0.113	0.148			0.179	0.15	0.216		
	観光	0.252	0.315	0.258			0.129	0.34	0.128		
	購買	0.176	0.154	0.172			0.224	0.14	0.236		
	食事	0.431	0.418	0.422			0.468	0.37	0.42		
4日旅行	ビジネス	0.138	0.109	0.117	0.147		0.17	0.161	0.156	0.234	
	観光	0.283	0.3	0.303	0.245		0.108	0.194	0.188	0.133	
	購買	0.172	0.158	0.146	0.165		0.256	0.194	0.25	0.237	
	食事	0.407	0.433	0.434	0.443		0.466	0.452	0.406	0.396	
5日旅行	ビジネス	0.15	0.108	0.117	0.092	0.129	0.225	0.119	0.148	0.2	0.209
	観光	0.23	0.281	0.292	0.304	0.241	0.112	0.286	0.296	0.3	0.128
	購買	0.168	0.141	0.155	0.135	0.182	0.219	0.286	0.222	0.233	0.227
	食事	0.451	0.47	0.436	0.469	0.448	0.444	0.31	0.333	0.267	0.436

表 5 関東地方居住者旅行と九州地方居住者旅行の行動の割合

	関東地方居住者旅行	九州地方居住者旅行
ビジネス	0.128	0.183
観光	0.243	0.186
購買	0.209	0.205
食事	0.42	0.426

居住者は日常的に東京へ行くことが多いと考えられる。このことから、関東地方居住者旅行において、旅行の進行状況に伴う行動の変化が小さかったと予想される。

6. まとめ

本研究では、Twitter の位置情報を用いて、東京を含む旅行を収集し、各投稿の行動を分類することによって、日数ごとに行動の変化を分析した。旅行の種類や旅行者の居住地を考慮した旅行の進行状況によって、行動の割合が変化することを確認した。その結果、東京主目的旅行において「観光」の割合が高く、東京経由旅行において「ビジネス」、「購買」の割合が高いことがわかった。また、旅行者の居住地が関東地方である旅行において「観光」の割合が高く、旅行者の居住地が九州地方である旅行において「ビジネス」の割合が高いことがわかった。さらに、3日旅行から5日旅行において、初日と最終日に「ビジネス」、「購買」の割合が大きく、「観光」の割合が小さい傾向にあることがわかった。

今後の課題として、分類性能を向上させることが挙げられる。また、今回は投稿の日付のみに着目し、時間帯を考慮していない。加えて、周期性を考慮していないため、通勤や通学などで、日常的に東京を訪れているユーザを旅行者と同様に扱っている。

旅行における投稿の時間帯や周期性を考慮することによって、より詳細な分析を行うことができると考えられる。

謝 辞

本研究は、首都大学東京傾斜的研究(全学分)学長裁量枠戦略的研究プロジェクト戦略的研究支援枠「ソーシャルビッグデータの分析・応用のための学術基盤の研究」及び JSPS 科研費 16K00157, 16K16158 による

文 献

- [1] 若木裕美, 有賀康顕, 中田康太, 藤井寛子, 住田一男, 鈴木優. 口コミ情報からの目的情報抽出. FIT2011 (第 10 回情報科学技術フォーラム), 2011.
- [2] Qiuju Luo and Dixi Zhong. Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tourism Management*, Vol. 46, pp. 274–282, 2015.
- [3] Anita Wenger. 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.
- [4] 高橋一希, 加藤大受, 遠藤雅樹, 廣田雅春, 莊司慶行, 石川博. 経由地としての東京 -マイクロブログを用いた旅行者の東京都内での行動分析-. 第 13 回観光情報学会全国大会, 2016.
- [5] Soojin Choi, Xinran Y Lehto, and Alastair M Morrison. Destination image representation on the web: Content analysis of macau travel related websites. *Tourism Management*, Vol. 28, No. 1, pp. 118–129, 2007.
- [6] Yuya Nozawa, Masaki Endo, Yo Ehara, Masaharu Hirota, Syohei Yokoyama, and Hiroshi Ishikawa. Inferring tourist behavior and purposes of a twitter user. *Artificial Intelligence for Tourism (AI4Tourism)*, 2016.
- [7] 中嶋勇人, 新妻弘崇, 太田学. 位置情報付きツイートを利用した観光ルート推薦. 研究報告データベースシステム (DBS), Vol.

表 6 関東地方居住者旅行と九州地方居住者旅行の行動の割合の変化

		関東地方居住者旅行					九州地方居住者旅行				
		1日目	2日目	3日目	4日目	5日目	1日目	2日目	3日目	4日目	5日目
1日旅行	ビジネス	0.123					0.208				
	観光	0.243					0.219				
	購買	0.207					0.163				
	食事	0.427					0.41				
2日旅行	ビジネス	0.142	0.142				0.179	0.2			
	観光	0.233	0.226				0.209	0.148			
	購買	0.241	0.244				0.203	0.216			
	食事	0.385	0.387				0.408	0.436			
3日旅行	ビジネス	0.136	0.127	0.157			0.218	0.098	0.204		
	観光	0.242	0.277	0.246			0.132	0.355	0.145		
	購買	0.187	0.176	0.19			0.241	0.135	0.223		
	食事	0.435	0.42	0.406			0.409	0.412	0.428		
4日旅行	ビジネス	0.129	0.122	0.131	0.148		0.23	0.079	0.075	0.198	
	観光	0.245	0.249	0.238	0.252		0.111	0.214	0.317	0.101	
	購買	0.196	0.173	0.221	0.182		0.246	0.193	0.168	0.238	
	食事	0.43	0.455	0.41	0.418		0.413	0.514	0.441	0.463	
5日旅行	ビジネス	0.149	0.095	0.116	0.091	0.127	0.234	0.154	0.075	0.141	0.18
	観光	0.219	0.248	0.287	0.303	0.251	0.129	0.346	0.254	0.328	0.1
	購買	0.175	0.212	0.182	0.146	0.192	0.242	0.135	0.209	0.188	0.25
	食事	0.456	0.445	0.414	0.461	0.43	0.395	0.365	0.463	0.344	0.47

2013, No. 28, pp. 1-6, 2013.

- [8] 新井晃平, 新妻弘崇, 太田学. Twitter を利用した観光ルート推薦の一手法. 第7回データ工学と情報マネジメントに関するフォーラム (DEIM2015), G7-6, pp. 1-8, 2015.
- [9] 倉田陽平, 首都大学東京. 大量写真データをもとにした観光地内の主要観光ルート網の自動抽出に向けて. 観光情報学会第8回研究発表会, pp. 49-52, 2013.
- [10] Tatsuya Fujisaka, Ryong Lee, and Kazutoshi Sumiya. Discovery of user behavior patterns from geo-tagged microblogs. In *ICUIMC '10 Proceedings of the 4th International Conference on Ubiquitous Information Management and Communication Article No. 36*, p. 36. ACM, 2010.
- [11] 田中成典, 中村健二, 寺口敏生, 中本聖也, 加藤諒. マイクロブログから抽出したユーザの習慣に基づく行動推定に関する研究. 情報処理学会論文誌データベース (TOD), Vol. 6, No. 3, pp. 73-89, 2013.
- [12] 大原啓詳, 灘本明代. Twitter 上のあるユーザの意外な情報抽出手法の提案. 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.
- [13] Vladimir N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, Vol. 10, No. 5, pp. 988-999, 1999.
- [14] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, Vol. 41, No. 6, p. 391, 1990.
- [15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830, 2011.