

ソーシャルメディアとAIカメラを用いた 都市間交通需要の把握

原 祐輔¹・羽藤 英二²

¹正会員 東北大学准教授 大学院情報科学研究科 (〒 980-8579 仙台市青葉区荒巻字青葉 6-6-06)
E-mail: hara@tohoku.ac.jp (Corresponding author)

²正会員 東京大学教授 大学院工学系研究科 (〒 113-8656 東京都文京区本郷 7-3-1)
E-mail: hato@bin.t.u-tokyo.ac.jp

観光需要などの都市間交通需要は季節変動などの要素があることは知られているが、ソーシャルメディア上のつぶやきから、都市間交通需要を予測することができれば、新たな交通調査・交通需要予測手法として有用である。本研究では全国の地名を含むツイートを用いて、バスタ新宿のAIカメラを用いた高速バス利用者動向を予測するモデルを構築する。ケーススタディとして、大阪府行き・宮城県行き的高速バス利用者数の予測モデルを日々ツイートされる単語の頻度を主成分分析した9つの主成分得点を特徴量とした回帰モデルを構築し、実用性のある予測精度を達成することで本手法の実用可能性を示した。

Key Words: *intercity transportation, passenger demand, Twitter, AI camera, natural language processing*

1. はじめに

都市間交通の需要は出張のビジネストリップや旅行・観光需要、異動や引っ越しに伴う需要などの季節変動が存在することは知られているが、それに加えてイベント開催や災害発生に伴う突発的な需要も存在する。また、都市間交通需要の中でも航空機や新幹線・特急等の鉄道によるトリップと比べて、若年層が利用する高速バスは多様な料金や複数の事業者にまたがることもあり、その需要予測は困難である。しかしながら、全国的な都市間バスターミナルの整備や高速バス事業者間での車両や乗客の融通、鉄道未整備地域へのアクセシビリティの確保など、高速バス事業者の役割はいまだ大きく、リアルタイム性をもち精度の高い需要予測を行うことへの期待は大きい。

そこで、都市間交通需要の安価な調査・予測手法の構築のために、本研究ではソーシャルメディアとAIカメラに着目する。ソーシャルメディアとはインターネットを利用して誰でも手軽に情報を発信し、相互のやりとりができる双方向型メディアであり、代表的なものとしてFacebookやTwitter等のSNSを挙げることができる。本研究ではTwitterのツイート(つぶやき)に着目する。AIカメラとは、画像処理等のアルゴリズムを搭載したネットワークカメラを意味し、低コストで日々の人流や車両数を検知することができる。この2種のデータを組み合わせることで、安価な都市間交通需要の調査・予測手法を構築する。

本研究の問いは雑多な情報が含まれるソーシャルメディア上のツイート内から、都市間交通需要、特に高速バス利用に影響を与える要素を抽出し、予測モデルに利用可能な特徴量を生成できるかである。この問いを通して、ソーシャルメディアを用いた都市間交通需要の予測可能性について基礎的検討を行う。

本研究の貢献は以下の3つである。

- 各都道府県の地名を含むツイートの中で高頻度で用いられる単語と1日の高速バス利用者数と関連のある単語は異なり、必ずしも交通施設と直接関連がない単語が多く含まれることを示した。(4.(2))
- ソーシャルメディア上のツイートの時系列推移を元にした特徴量を用いて高速バス利用者数の回帰モデルを構築し、検証データにおいても約50-70(人/日)の誤差で予測できることを示した(5.(1))
- 大阪府と宮城県の予測モデルという2つのケーススタディを通して、交通施設に関連する話題、地域内の地名や商業施設名、イベント名の話、季節性に関連する話題の3つが共通に現れる要素であることを示唆した。(5.(2))

2. 既往研究のレビュー

日本国内における全国レベルの都市間流動を把握するためのデータとして、交通機関毎の輸送人員を集計した貨物・旅客地域流動調査(総流動データ)があり、出発地や目的地、トリップ目的、利用交通手段などを把

握するための純流動データとして、1990年から5年ごとに全国幹線旅客純流動調査が実施されている。同調査は、既存の3調査(航空旅客動態調査、幹線鉄道旅客流動実態調査、道路交通センサス)に補充調査として幹線バス旅客流動調査、幹線フェリー・旅客船旅客流動調査を実施し、それらを統合処理(異なる交通機関間の乗継ぎ処理)を行い整備したものである。

塚井・奥村¹⁾は全国の旅客流動に対して業務交通と情報通信パターンの関係性を明らかにする先駆的な研究を行っている。塚井・奥村²⁾では、情報交流行動と実際の交通行動の代替性・補完性を通信コストや交通コストの変化による交通量・通信量の変化率として表し、現在のサイバー・フィジカルシステムに通じる研究がなされている。全国幹線旅客純流動調査は一つのトリップの中で複数の交通機関や便を乗り継ぐ行動を把握できることが、他の事業者ごとの統計調査には見られない大きな特徴であるが、全体の需要を把握するためには拡大係数の設定が重要となる。奥村ら³⁾は旅行目的や個人属性に関しての補正を行うために、追加情報を考慮できる鉄道の拡大係数設定方法を提案している。山口ら⁴⁾は公共交通ネットワーク再編のために、都市間交通マクロ需要モデルを用いて都市間ネットワークのLOS弾力性を明らかにしている。

上記の研究のように、総流動データや全国幹線旅客旅客純流動調査のデータを用いて、マクロスケール・都市間レベルの交通需要予測を行うのが主流のアプローチである。このアプローチに対して、補完するアプローチとして、モバイル空間統計データのような人流・滞在データやソーシャルメディアを用いるアプローチがある。ここでは特にソーシャルメディアを用いたソーシャルセンシングに関する研究レビューを行う。

ソーシャルセンシングとは、本来センシング目的ではなかった様々なログデータ、特に人自身や人の行動に伴うログデータを用いて、人をセンサーのようにして扱う新しい研究分野である。例えば、モバイル空間統計は携帯電話基地局への通信という本来の目的を持つセンシングデータを用いて、エリア内の滞在人口の特定に用いられているという点で広義のソーシャルセンシングの一つと考えられる。他にも有名な例として、Googleでのインフルエンザの検索ログからインフルエンザの流行予測が可能であることを示した Ginsberg et al.⁵⁾やTwitter上での地震発生に関するツイートから地震の発生時刻や震源地が予測可能であることを示した Sakaki et al.⁶⁾を挙げることができる。ソーシャルセンシングの初期の研究では異常事象検知に用いられることが多く、その方法論の提案として、Kleinberg⁷⁾、Fung et al.⁸⁾の方法がしばしば用いられている。実空間における異常事象検出として、Aggarwal and Subbian⁹⁾はイ

ベント検出を、Scjulz et al.¹⁰⁾は交通状態推定を、Mai and Hranac¹¹⁾やPan et al.¹²⁾は交通状態推定や交通事故検出を行なっている。ソーシャルメディアからの情報抽出を用いた行動モデルの推定の研究として原¹³⁾、Hara¹⁴⁾は東日本大震災時の帰宅行動をTwitterのツイートデータのみから抽出し、帰宅・宿泊行動モデルの推定とその要因分析を行っている。

ソーシャルセンシングの二つ目のアプローチはトピックモデルを用いた潜在的意味解析である。これはソーシャルメディア上でのテキストデータを活用することを目指している。Yin et al.¹⁵⁾やHong et al.¹⁶⁾はジオトピックモデルと呼ばれるテキストデータと位置情報を用いて、特定のトピックの空間的な広がりをデータから明らかにしている。Chen et al.¹⁷⁾はLanguage Enhanced Hinge Loss Markov Random Fields という方法を使ってツイートから空間的な交通状態の推定を行なっている。原ら¹⁸⁾、Hara et al.¹⁹⁾はプローブカーの情報とツイートの情報の双方を用いて、交通状態推定とともに、渋滞の原因を同時に把握する手法を開発している。最初のアプローチである異常事象検出が、検出後に人手によって検出された異常事象を解釈する必要があったのに対し、二つ目の潜在的意味解析によるアプローチはそれらもデータオリエンテッドに行う点でより高度化されている。

本研究はTwitter上での様々なツイート・話題という一見都市間交通需要とは直接関連がないと思われるデータから、適切な情報抽出を行うことで、需要予測を試みる。これはTwitter上で表出される社会情勢や話題が間接的に実空間での交通行動や移動需要、その制約条件などを表しているという仮説に基づく。このアプローチを通して、ソーシャルメディアデータのもつ交通需要予測の可能性について検討を行う。

3. データ概要

本研究で用いるデータはTwitterのツイートデータとバスタ新宿に存在する各バスの乗車人数を把握するためのAIカメラデータである。バスタ新宿のAIカメラデータは行先別にバスの乗車数がわかるため、バスタ新宿発の高速バスを利用した都市間移動を把握することができる。ツイートに関しては東北大学情報科学研究科 乾研究室によって収集されたものである。

本研究では分析対象を2020年1月から2021年2月までの14ヶ月を対象とする。その間、収集された日本語ツイート数は約82億(8,218,659,843)ツイート、1日平均で約2000万ツイートである。1日あたりの日本語ツイート総数は正確には公開されていないが、6000-8000万ツイートと言われており、本研究で用いるデータは

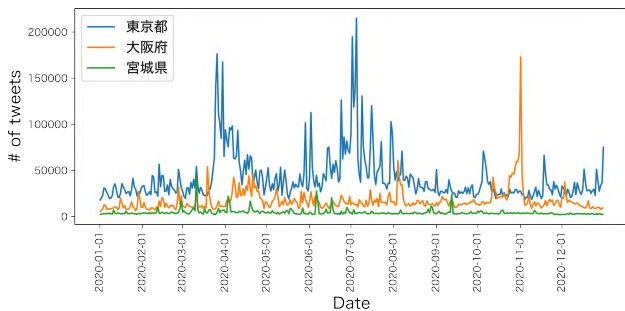


図-1 2020年の地名含みツイートの推移(東京都・大阪府・宮城県)

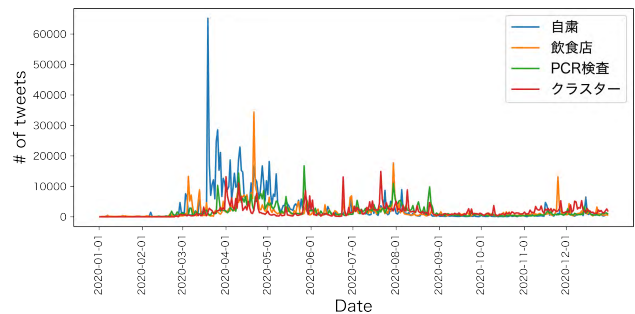


図-3 「自粛」「飲食店」「PCR検査」「クラスター」の出現頻度の推移

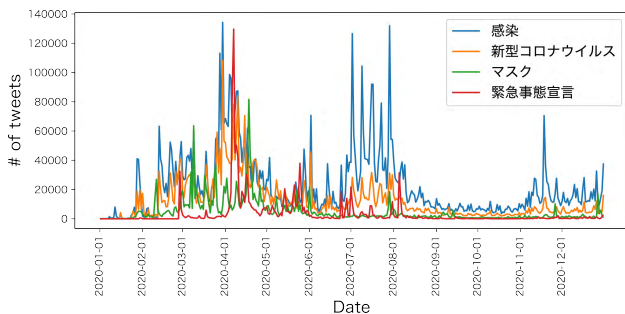


図-2 「感染」「新型コロナウイルス」「マスク」「緊急事態宣言」の出現頻度の推移

日本語ツイートの1/3から1/4が収集できていると考えられる。

Twitterのツイートデータは必ずしも地域や地理情報に関するツイートではないため、次に地名含みツイートを抽出した。本研究では、地名含みツイートの定義として「東京都」や「渋谷区」、「さいたま市」のような正式な都道府県名と市区町村名を含むツイートとした。これらのツイートには「渋谷」や「銀座」、「仙台」だけを含むツイートが含まれていないが、人名等との固有名詞との衝突があり、分析が煩雑になるため本研究では除外し、正式名称を含むツイートのみを分析対象とした。この地名含みツイートの精度については今後の課題とする。

本研究の地名含みツイートの定義を用いてツイートの抽出を行うと、期間中の総数は約1.1億(109,883,933)ツイート、1日平均で約25万ツイートであり、これを本研究の分析対象とする。これは本研究が収集した全ツイートの1.5%である。この地名含みツイートはある地域について言及していることのみを表し、必ずしもそのユーザーアカウントがその地域にいることや旅行を計画しているなどを表すわけではない。例えば、各地のニュースや災害に関する言及が含まれる。一例として、東京都の地名含みツイートと大阪府の地名含みツイートの推移を図1に示す。東京都の2020年3月30日前後のピークは「コメディアン志村けんさん死去70歳 新型コロナウイルスに感染して肺炎を発症し、入

院して治療を受けていたコメディアン志村けんさんが、29日夜、東京都内の病院で亡くなりました。」という訃報、7月5日前後のピークは「小池都知事が再選を確実に任期満了に伴う東京都知事選は午後8時に投票が締め切られた。投票率は午後7時半時点で37.32%で、前回は3ポイントほど下回っている。」という都知事選の話題により、東京都の地名含みツイートが増加している。同様に大阪府の11月1日前後のピークは「大阪都構想」住民投票 反対多数が確実」という大阪都構想の話題により、大阪府の地名含みツイートが増加している。一方で、東京や大阪と比べると人口規模が異なるため、宮城県に関する地名含みツイート数は少ないが、3月11日前後に東日本大震災に関連したツイートが多くなることが観測された。このようにソーシャルメディア上では政治の話題や有名人の結婚・訃報といった話題が多く言及されるため、ある地域の地名含みツイートが増加することそのものが必ずしも実際の交通行動を増加・減少させるわけではないことが示されている。

4. 基礎分析

(1) 形態素解析と頻出単語把握

データ概要に記載した地名含みツイートの形態素解析をMecab²⁰⁾及びmecab-ipadic-NEologd^{21) 22)23)}を用いて行った。その結果、分析期間中に1回以上現れた単語の種類は約170万(1,706,504)である。2020年に特徴的な頻出単語の推移を図2, 3, 4に示す。図2は「感染」「新型コロナウイルス」「マスク」「緊急事態宣言」の出現頻度の推移である。感染者数が増加し、2020年4月7日に7都府県に出された緊急事態宣言が出された、また新型コロナウイルス拡大初期にマスク不足に悩まされたことを読み取ることができる。図3は「自粛」「飲食店」「PCR検査」「クラスター」の出現頻度の推移である。2020年1回目の緊急事態宣言前の感染拡大期に、活動の自粛が大きく叫ばれていたことや大規模な感染者のクラスターが報道されるたびに、ソーシャル

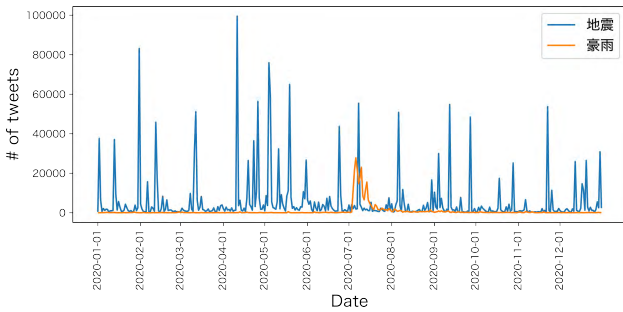


図-4 「地震」「豪雨」の出現頻度の推移



図-5 東京都の地名含みツイートのワードクラウド

メディア内で大きく言及されていることがわかる。図4は上記の新型コロナ関連の単語とは異なり、災害関連の単語である「地震」「豪雨」の出現頻度の推移である。地震に関しては、中型から大型の地震が発生するたびにピークが立ち上がり、実際の地震の発生と強く相関していることがわかる。これは既往研究⁶⁾の結果とも一致する。同様に、「豪雨」も熊本県を中心に九州や中部地方など日本各地で発生した令和2年7月豪雨時に多く言及される一方で、平常時にはほぼ言及されない単語であることがわかる。このように、多くの単語はその日の社会情勢や現実空間の出来事を部分的に示す一方で、その単語が表出される傾向は大きく異なることが示唆される。

次に各都道府県ごとの傾向を概観する。東京都、大阪府、宮城県の地名含みツイートをそれぞれの頻出単語を用いて作成したワードクラウドを図5, 6, 7に示す。2020年は新型コロナウイルスの感染拡大した特殊な年であり、東京・大阪ともにそれらに関連する単語が多いものの、頻出単語であったとしても東京・大阪・宮城で傾向が異なることが示されている。たとえば、大阪固有の話題として「イソジン」「ライブハウス」を挙げることができる。一方で、宮城県では「東日本大震災」の話題や「福島県」のような隣接する県名と共起していることがわかる。



図-6 大阪府の地名含みツイートのワードクラウド



図-7 宮城県の地名含みツイートのワードクラウド

(2) バスタ利用者との相関がある単語

次に、バスタ新宿のAIカメラのデータを用いる。バスタ新宿のAIカメラデータは行先別の乗客数が把握可能であるため、各日の都道府県別高速バス利用者を算出する。「大阪・USJ・京都」行きのバスのように複数の目的地があるバスについては、降車地の都道府県名の数で乗客を単純に按分する。このようにして作成した都道府県別日乗客数と都道府県の地名含みツイート内の単語の相関係数を計算し、その絶対値が大きい単語を図8, 9に示す。

まず、大阪について考察する。図6と図8を比較すると、単純に大阪に関連してつぶやかれていた頻出単語とは傾向が大きく異なることがわかる。上述の通り、単語の種類は約170万存在するため、分析期間中にバス乗車客数と発言された単語の間に相関があったとしても、偶然の相関である可能性は否めないが、興味深い特徴として「新大阪駅」「伊丹空港」「山陽新幹線」「国際空港」「大阪難波駅」「フェリーターミナル」といった交通施設との相関があることが挙げられる。見慣れない単語である「ノコン」「秘密結社」は悪ノ秘密結社というTRPGサークルが悪ノコンというイベントを実施していること、「BANZAI JAPAN」という女性アイドルグループや「私立慶南学院高等部」という男性ボーカ



図-8 大阪府行きバス乗客数と相関が高い単語



図-9 宮城県行きバス乗客数と相関が高い単語

ルユニットが大阪でイベントを実施していることが事後的に確認された。「NMB48 劇場」という単語もあることから、イベントに関連する単語が相関が高いことが示唆される。「テイクアウト」「ボトルビールテイクアウト」は高速バスの乗客数と負の相関があり、緊急事態宣言中および感染拡大期と大阪行きのバス乗客数が負の相関があることを間接的に示している。

宮城県も図 7 と図 9 を比較すると、同様の傾向を発見できる。単純な頻出単語では出てこなかった「仙台国際空港」という交通施設名や「舞台劇」というイベントを示唆する単語、「テイクアウト」や「デリバリー」「ステイホーム」など高速バスの乗客数と負の相関がある単語が見られた。このように表出される単語は必ずしも同じではないものの、交通施設やイベント、感染症拡大期を示す単語などが各地への高速バス乗客数と相関があることをこの分析結果は示している。

5. ソーシャルメディアによる高速バス乗客数の予測モデル

(1) 大阪・宮城モデルの構築

地名含みツイートからバスタ新宿から各都道府県へ向かう高速バスの 1 日の乗客数を予測するモデルを構

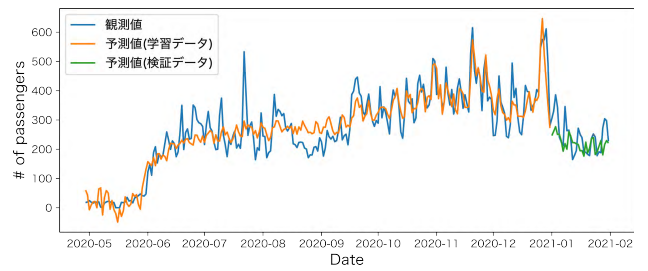


図-10 大阪府行きバス乗客数の観測値と予測値

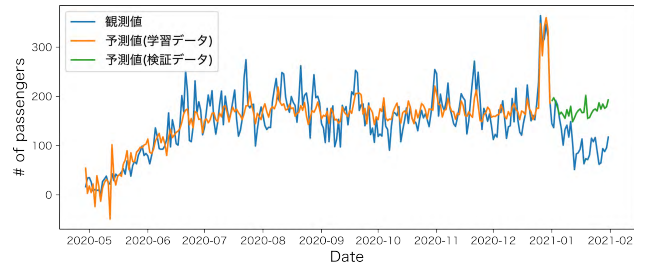


図-11 宮城県行きバス乗客数の観測値と予測値

築する。全ての単語を用いると、予測する日数に比べて単語種類が多く過学習が発生するため、各都道府県別バス乗客数との相関係数の絶対値が 0.25 以上の単語を抽出した。大阪府の場合、用いた単語数は 1309、宮城県の場合は用いた単語数は 562 である。さらに主成分分析を行い、次元削減した各主成分得点を特徴量として、日高速バス乗客数を予測する回帰モデルを構築した。用いた特徴量の数は 9 つの主成分得点と切片である。入手できたバスタ新宿のデータの関係から、2020 年 4 月 29 日から 2020 年 12 月 31 日までのデータで予測モデルを作り、2021 年 1 月 1 日から 31 日までの 1 ヶ月を検証データとして精度検証を行う。

予測モデルの性能を示す。大阪モデルの学習データの RMSE (Root Mean Square Error) は 51.94(人/日)、検証データの RMSE は 54.54(人/日)であり、宮城モデルの学習データの RMSE は 30.71(人/日)、検証データの RMSE は 72.79(人/日)である。観測された実際の日高速バス利用者数と予測結果を図 10, 11 に示す。大阪の予測モデルについては、学習データに含まれていない 2021 年 1 月以降も学習期間と同様の精度で予測できていることが示されている。宮城の予測モデルは検証期間に観測値よりもやや過大に予測しているものの、大きく予測を外してはいない。興味深い結果として、2020 年は平年と比較するとやや特殊な年であり、2020 年 5 月の乗降客数は非常に少なく、6 月頃から少しずつ回復する傾向がある。このような傾向をソーシャルメディアのツイートから適切に把握することができていることはソーシャルメディアによる交通需要予測の可能性を示唆しているといえよう。

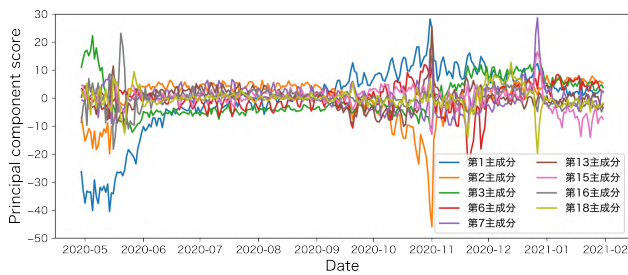


図-12 大阪府地名含みツイートの主成分

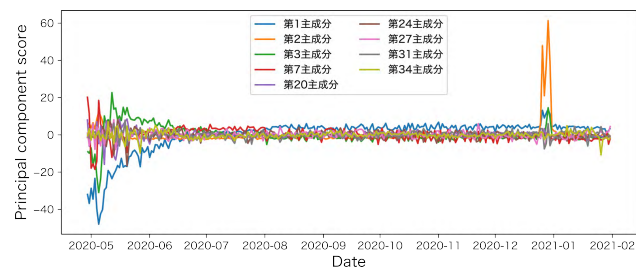


図-13 宮城県地名含みツイートの主成分

(2) ソーシャルメディアの主成分の考察

では、なぜソーシャルメディア上でのつぶやきから、都市間交通需要を把握することができるのか、その要因を各主成分を構成する要素から考察する。

大阪の回帰モデルに用いた9つの主成分を表示したのが図12である。青線で示される第1主成分は5月当初は少なく、11月頃から増加傾向にある。この主成分において正の因子負荷量が高いのは「新大阪駅」「国際空港」「東海道・山陽新幹線」「伊丹空港」「大阪難波駅」「御堂筋」「なんば」「石橋阪大前駅」「蛍池駅」といった交通施設や地域名である。オレンジ線で示される第2主成分において負の因子負荷量が高いのは「なくす」「都」「構想」「住民」「成長」「権限」「区」「街頭」「支配」「チラシ」「インフラ」「基礎自治体」「都市計画」「成長戦略」といった単語であり、これらは大阪都構想に関連する話題であることが示される。これは11月1日に大阪都構想の住民投票が行われた際に大きく負の値となっていることから明らかである。第3成分は「クラフトビア」「ボトルビールテイクアウト」「寒い朝」「クリスマス」「臨時休館」「年末年始」「さむい」「年賀状」「大型連休」「クリスマスマーケット」「歳末」「5月中旬」「寒さ」といった単語が正の因子負荷量が高い。この成分は季節性、特に連休や年末などの非日常時を表す成分と解釈できる。第6成分は「ラウンドワンスタジアム」「アパホテル」「千日前店」「心齋橋店」「個展」「相鉄フレッサイン」「大日本プロレス」「オタロード」「近鉄百貨店」「試合開始」「グランフロント大阪」「試合結果」「びっくりドンキー」「日本橋」「コメダ珈琲店」「ホテルモントレ」「ポケモンセンターオーサカ」といった単語の負の因子負荷量が高かった。これらはホテルや店舗名、イベントと関連する話題と解釈できる。第7成分は「四條畷市長選挙」「四條畷市民」「土井」「12月27日」などの12月27日に実施された四條畷市長選挙に関連する単語が多く見られた。この成分が高速バス乗客数を説明する要素になる解釈が難しく、年末年始など非正常性との擬似的な相関の可能性もある。第13成分は「スパイシー」「サバ」「マンガでわかる」「らーめん」「スープ」「手渡し」「発送」などの単語

の負の因子負荷量が高い。しかしながら、これらの成分が表す話題の解釈は難しく、また高速バス乗客数との関連も難しい。第15成分も「四條畷市長選挙」「土井」「四條畷市民」といった四條畷市長選挙に関連する単語の正の因子負荷量が高かったが、「名神高速道路」「天王山トンネル」「下り線」「合流地点」といった道路ネットワークに関連する単語の正の因子負荷量が高かった。第16成分は「黄金野」「渚栄」「北浜東」「石町」「渚内野」「百舌鳥西之」「土師町」といった大阪府内の町丁目名の正の因子負荷量が高く、大阪府内の細かな地名を表す話題と考察される。第18成分は「混雑度」「和泉砂川」「富田林西口」「学園前」「新石切」「尾崎駅」といった駅名や「波太神社」「林昌寺」「総持寺」「観心寺」といった寺社仏閣名の正の因子負荷量が高く、鉄道や寺社を表す話題と考察される。

同様に、宮城の回帰モデルに用いた9つの主成分を表示したのが図13である。5月当初に大きく減少している第1主成分において負の因子負荷量が高いのは「テイクアウト」「連続」「88人」「122人」「外出自粛」「休業」「不要不急の外出」「ステイホーム」といった新型コロナウイルスの拡大期に関連する話題である。第2主成分は「和歌山県知事」「菅首相」「鈴木康友」「湯崎英彦」「高浜虚子」「平尾雅子」「桐山建志」「ダルビッシュ」といった政治家名や人名と「門松」「12月28日」「エアコミケ」「今年1年」「年末年始」「年末」「新年」といった年末を表す単語が見られた。話題がうまく分離できていない可能性が高いが主成分得点からも明らかに年末を表す話題と考えられる。第3主成分は「不要不急」「給付金」「振込」「跨ぐ」「医療施設」「特別定額給付金」「給付金申請」「10万円」「買い求める」「マイナンバーカード」といった単語の正の因子負荷量が高い。これらは緊急事態宣言やそれに伴う特別定額給付金の話題であると考えられる。第7主成分は「学校」「パチンコ」「休校」「9月入学」「休業要請」「アベノマスク」「自粛要請」「百貨店」「仙台青葉まつり」といった単語の正の因子負荷量が高い。これらも新型コロナウイルスに関連する話題であるが、特に施設の休業やイベントの中止に関連する話題と考えられる。第20主成分は「特別定額給付金」「マイナポータル」「10万

円)「申請方法」などの正の因子負荷量が高く、第 3 主成分と類似する傾向を示している。第 24 主成分は「ドライブスルー」「お持ち帰り」「宅配サービス」「店頭販売」「テイクアウト」といった飲食業の形態に関する単語と「仙台市若林区若林」「青葉まつり」「みちのく伊達政宗歴史館」といった宮城県内の地名や施設名、イベント名の正の因子負荷量が高い。第 27 主成分は「博士峠」「つむじ倉滝」「倉滝」「青森県青森市」「宮城県大崎市古川米倉」「宮城県仙台市青葉区花京院」「横手」といった東北地方の地名の負の因子負荷量が高い。第 31 主成分は「下り」「長者原」「鶴巣」といった高速道路・東北自動車道に関連する単語、「気仙沼駅」「小牛田駅」といった駅名、「仙台国際空港」といった交通施設の負の因子負荷量が高い。第 34 主成分は「療養中」「退院」「新型コロナウイルス」「リハビリ」「フェイスシールド」といった新型コロナ感染症に関連する話題の負の因子負荷量が高い。

大阪モデルと宮城モデルの主成分を構成する単語の考察から、2 府県への高速バス乗車客数に関連する共通の項目を整理することができる。1 点目は交通施設(鉄道・道路・空港)に関するツイートである。2 点目は地域の地名や商業施設名、開催されたイベント名である。3 点目は年末年始などの季節性を表す話題である。これらの 3 点は高速バス乗客数に直接関連している話題と解釈できる。大阪モデルで特徴的であったのは政治的な話題の多さである。大阪都構想と四條畷市長選挙という 2 つの話題が予測モデルの特徴量として含まれていたが、これらが全国的にも見られる傾向であるのかどうかは、他の都道府県の首長選挙などと合わせて今後分析を行う必要があるだろう。一方で、宮城モデルでは新型コロナ感染症に関する話題が多く含まれている。これは図 1 で見られたように宮城県に関するツイートは大阪府と比較すると非常に小さく、全国的な話題である新型コロナ感染症の中に地域の話が隠れてしまっている可能性がある。この点についても他の都道府県に関するモデルを構築し、全国的な傾向を見る必要があると考えられる。

6. おわりに

本研究では、都市間交通需要をソーシャルメディア上のツイートのみから把握・予測する手法の開発の基礎的検討として、バス・新幹線の都道府県別高速バス乗客数をソーシャルメディア上のツイートから予測するモデルを構築した。各都道府県や市区町村名を含むツイートを地名含みツイートとして定義し、日本全国のツイートの傾向を分析した。ツイート数の増減は必ずしも実行動と関連せず、社会情勢を表すツイートや政

治的ツイートに大きく影響を受けていることを示した。

ケーススタディとして、大阪府行き的高速バス乗客数と宮城県行き的高速バス乗客数の予測モデルを構築し、それぞれ学習に用いていない検証期間の予測精度が 54.54(人/日)、72.79(人/日)と、特徴量が 9 つ程度の単純な回帰モデルであっても比較的高い精度でツイートから高速バス乗客数を予測できることを明らかにした。また、予測モデルに用いた主成分から、ソーシャルメディアには(1)地域内の交通施設に関連する話題、(2)地域内の地名や商業施設名、イベント名などの話題、(3)季節性に関連する話題が含まれているため、都市間交通需要を予測できる可能性が示唆された。

今後の課題として、全国的なケーススタディへの拡大や分析対象期間の拡大による一般性の確認、トピックモデル等を用いた話題の適切な把握、他の交通機関利用者数の予測可能性の検討が挙げられる。

謝辞： 本研究は国土交通省新道路技術会議「マルチスケールな拠点空間計画のための新たな行動モデル研究」の助成を受けたものです。また、本研究を進めるにあたって東北大学 乾健太郎教授、山口健史研究員からは Twitter データの提供と有益なコメントを頂きました。記して感謝致します。

参考文献

- 1) 塚井誠人 and 奥村誠: 代替性・補完性を考慮した業務交通と通信パターンの重力モデル, *都市計画論文集*, Vol.34, pp.85-90, 1999.
- 2) 塚井誠人 and 奥村誠: 情報伝達の複雑性を考慮した通信と交通の情報流量分担モデル, *土木学会論文集*, No.667, pp.113-121, 2001.
- 3) 奥村誠, 山口裕通, and 大窪和明: 全国幹線旅客純流動調査の鉄道サンプル拡大方法に関する研究, *土木学会論文集 D3 (土木計画学)*, Vol.67, No.5, pp.671.911-671.918, 2011.
- 4) 山口裕通 and 奥村誠: 都市間交通需要の los 弾力性に関する研究, *土木学会論文集 D3 (土木計画学)*, Vol.69, No.5, pp.1.629-1.638, 2013.
- 5) Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L.: Detecting influenza epidemics using search engine query data, *Nature*, Vol.457, No.7232, pp.1012-1014, 2009.
- 6) Sakaki, T., Okazaki, M., and Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors, *Proceedings of the 19th international conference on World wide web*, pp. 851-860, 2010.
- 7) Kleinberg, J.: Bursty and hierarchical structure in streams, *Data mining and knowledge discovery*, Vol.7, No.4, pp.373-397, 2003.
- 8) Fung, G. P. C., Yu, J. X., Yu, P. S., and Lu, H.: Parameter free bursty events detection in text streams, *Proceedings of the 31st international conference on Very large data bases*, pp. 181-192, Citeseer, 2005.
- 9) Aggarwal, C. C. and Subbian, K.: Event detection in social streams, *Proceedings of the 2012 SIAM international conference on data mining*, pp. 624-635, SIAM, 2012.
- 10) Schulz, A., Ristoski, P., and Paulheim, H.: I see a car crash:

- Real-time detection of small scale incidents in microblogs, Extended semantic web conference, pp. 22–33, Springer, 2013.
- 11) Mai, E. and Hranac, R.: Twitter interactions as a data source for transportation incidents, Proceedings of the Transportation Research Board 92nd Annual Meeting, 2013.
 - 12) Pan, B., Zheng, Y., Wilkie, D., and Shahabi, C.: Crowd sensing of traffic anomalies based on human mobility and social media, Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 344–353, 2013.
 - 13) 原祐輔: Twitter を用いた東日本大震災時の首都圏の帰宅意思決定分析, *自然言語処理*, Vol.20, No.3, pp.315–334, 2013.
 - 14) Hara, Y.: Behaviour analysis using tweet data and geo-tag data in a natural disaster, *Transportation Research Procedia*, Vol.11, pp.399–412, 2015.
 - 15) Yin, Z., Cao, L., Han, J., Zhai, C., and Huang, T.: Geographical topic discovery and comparison, Proceedings of the 20th international conference on World wide web, pp. 247–256, 2011.
 - 16) Hong, L., Ahmed, A., Gurumurthy, S., Smola, A. J., and Tsioutsoulklis, K.: Discovering geographical topics in the twitter stream, Proceedings of the 21st international conference on World Wide Web, pp. 769–778, 2012.
 - 17) Chen, P.-T., Chen, F., and Qian, Z.: Road traffic congestion monitoring in social media with hinge-loss markov random fields, 2014 IEEE international conference on data mining, pp. 80–89, IEEE, 2014.
 - 18) 原祐輔, 松田耕史, 川崎洋輔, 三谷卓摩, and 桑原雅夫: Probe と tweet を用いたマルチリソースによる潜在交通状態推定に関する研究, *土木計画学研究・講演集*, Vol.51, 2015.
 - 19) Hara, Y., Matsuda, K., Kuwahara, M., Okazaki, N., and Inui, K.: Estimating traffic states and identifying their causes and effects through probe and social media data analysis, Proceedings of the 21st HKSTS International Conference, 2016.
 - 20) Kudo, T.: Mecab: Yet another part-of-speech and morphological analyzer.
 - 21) Sato, T.: Neologism dictionary based on the language resources on the web for mecab, 2015.
 - 22) 佐藤敏紀, 橋本泰一, and 奥村学: 単語分かち書き用辞書生成システム neologd の運用 — 文書分類を例にして —, *自然言語処理研究会研究報告*, pp. NL–229–15, 情報処理学会, 2016.
 - 23) 佐藤敏紀, 橋本泰一, and 奥村学: 単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討, *言語処理学会第 23 回年次大会 (NLP2017)*, pp. NLP2017–B6–1, 言語処理学会, 2017.

(Received 2021. 9. 30)

(Accepted 2021. 9. 30)

UNDERSTANDING OF INTERCITY TRAVEL DEMAND BY USING SOCIAL
MEDIA AND AI CAMERA

Yusukue HARA and Eiji HATO