

トピックを用いた地域ごとの言語による反応の差異の分析

光井 孝志[†] 廣田 雅春^{††} 加藤 大受^{†††} 荒木 徹也^{††††} 遠藤 雅樹^{†††††}

石川 博^{†††††}

[†] 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

^{††} 岡山理科大学 総合情報学部 〒700-0005 岡山県岡山市北区理大町 1-1

^{†††} ウイングアーク 1 s t 株式会社 〒150-0031 東京都渋谷区桜丘町 20-1 渋谷インフォスター

^{††††} 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

^{†††††} 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

E-mail: [†]mitsui-takashi@ed.tmu.ac.jp, ^{††}hirota@mis.ous.ac.jp, ^{†††}kato.d@wingarc.com,

^{†††††}{araki,ishikawa-hiroshi}@tmu.ac.jp, ^{†††††}tendou@uitec.ac.jp,

あらまし 近年, Twitter などのマイクロブログから観光情報を抽出する研究が多く行われている. 日本人や外国人といったユーザ属性によって, 地域ごとの反応や興味対象に差異があると考えられるが, 国籍や言語などのユーザ属性に着目し, 観光情報を分析する研究は発展途上である. 要因の 1 つとして, ユーザの使用言語が異なる場合に, 単一の尺度を用いた比較や分析が困難であることがあげられる. そこで, 本研究では, 日本人と外国人の地域ごとのユーザの反応の差異を分析するために, 異なる言語のツイートを同じ尺度で扱うための手法を提案する. 具体的には, 英語のツイートを日本語に翻訳し, LDA を用いてトピックを抽出する. そして, 日本語と英語のツイートについて, 地域ごとのトピック分布を作成し, その差異から, 日本人と外国人の反応が異なる地域を抽出し, 分析を行う.

キーワード Twitter, 地域特徴, LDA, ユーザ属性

1. はじめに

訪日外国人旅行者数は年々増加傾向にあり, 2016 年には過去最多の約 2,400 万人に及んだ^(注1). 訪日外国人旅行者による消費行動は日本の経済成長の活性化に大きな影響を及ぼしており, 訪日外国人に向けた観光産業の重要性は一層増している. そのため, 政府は観光を成長戦略の大きな柱として, 2020 年に行われる東京オリンピックまでには訪日外国人旅行者数の目標を 4,000 万人とし, 今後訪日旅行者のさらなる増加を目指している^(注2).

訪日外国人観光客の増加のために重要なことの一つとして, 観光客の期待に沿うような観光情報や観光プランを提供することがあげられる. 観光プランを提供する上で, 実際に観光地を訪れた人の反応や感想を知る必要があり, そういった観光情報を抽出し分析することが非常に重要である.

訪問先に対する反応を知るための情報源として, SNS をはじめとする Web 上の情報を用いることが多い. そこで, ブログを用いて訪日外国人を対象とした観光に関する研究 [1] が近年盛んに行われている. 本論文では, 代表的なマイクロブログの一つである Twitter^(注3) を利用する. Twitter ではツイートと呼ばれる 140 字以内の短文を投稿し, 共有することでユーザ間でコミュニケーションをとることができる. ツイートは一度に

投稿できる文字数が少ないため, スマートフォンなどの携帯端末を用いて手軽に現在の行動や感想などを投稿できる特徴がある. また, ツイートにはジオタグと呼ばれる緯度経度情報からなる所在地データを付与することができる. そのため, ジオタグ付きツイートの位置情報や本文を用いて, 訪問地へのイメージの分析や観光ルートの推薦 [2], 地域ごとの単語に注目する研究 [3] など, 観光分野に関連する研究は盛んに行われている. このような研究で抽出される, 観光地のイメージや観光地におけるユーザの行動や感想などの観光情報には, 日本人と外国人のように, ユーザの属性によって違いがあると考えられ, ユーザ属性を考慮することは重要である. そのため, 日本人や外国人といった国籍に関するユーザ属性の違いに注目し, 位置情報を用いて観光情報の差異を抽出した研究 [4] が行われているが, 国籍別のツイート内容を考慮した分析は行われていない. 国籍に着目し, 位置情報とテキストを用いて観光を分析をする研究は発展途上である. それは, Twitter ユーザの国籍によって, 投稿される言語が異なるため, 投稿されたツイートを同一の尺度で比較することが困難であることが要因の一つとして挙げられる. 日本人と外国人のツイートを使用言語に依存せずに同一の尺度で比較することができれば, それぞれの訪問先に対する反応の差異を抽出することができると考えられる. さらに, これを用いることで, 訪日外国人に向けた観光プランの提案などに応用することができると考えられる.

そこで, 本論文では, Twitter に投稿された異なる言語を同一の尺度で比較する手法を提案し, 日本人と外国人の地域ごとの反応の差異の分析を行う. はじめに, ツイート本文とユーザ

(注1): https://www.jnto.go.jp/jpn/statistics/data_info_listing/pdf/170117_monthly.pdf

(注2): http://www.mlit.go.jp/kankochonews02_000307.html

(注3): <https://twitter.com/>

言語設定を用いて、ツイートの使用言語を判別し、日本人のツイートと外国人のツイートに分類する。次に、それぞれの投稿内容を同一の尺度で扱えるようにするため、外国人のツイートを日本語に翻訳する。その後、ツイート集合に対し Latent Dirichlet Allocation (LDA) [5] を用いて分析地域のトピックモデルを作成する。その後、分析地域をメッシュで分割し、日本人と外国人それぞれの投稿の位置情報をもとにツイートをメッシュに振り分け、それぞれのメッシュごとにトピック分布を算出する。日本人と外国人それぞれのメッシュごとのトピック分布の差異から、日本人と外国人の訪問地に対する反応の差異を抽出する。

本論文は、以下の構成に従う。2 章では、本研究の関連研究について述べる。3 章では、日本人と外国人のツイートを同一の尺度とし、地域ごとの反応の差異を分析する手法について述べる。4 章では、予備実験を行う。5 章では、分析と考察を行う。6 章では本論文のまとめと今後の課題を述べる。

2. 関連研究

本章では関連研究として、外国人観光客を対象とした研究とツイートに対して LDA を用いたトピック分類から分析を行う研究について述べる。

2.1 外国人観光客を対象とした研究

訪日外国人観光客数の増加により、外国人観光客を対象とした研究は増えてきている。村上ら [1] は、海外の英語で書かれたブログから、頻出度の高い単語を分析することにより、外国人観光客から見た日本の観光地 (東京、北海道、石川県) イメージを明らかにした。また、それぞれのブログの記事に目を通すことにより、テキストマイニングの有用性を示した。Choi ら [6] は、マカオに対するイメージを Web 上の旅行マガジン、旅行ブログ、観光ガイドサイトからテキスト解析を行うことで抽出した。さらに、サイトごとのマカオに対するイメージの比較を行った。

佐伯ら [4] は、ユーザに付与された言語情報と時間情報を用いて、ユーザを日本人、在日外国人、訪日外国人に分類した。さらに、位置情報を用いて、分類した属性ごとにツイートを投稿したユーザ数の多いエリアを分析することにより、属性による訪れる観光地への頻度の違いを抽出した。本研究でも国籍のユーザ属性の違いに着目して、観光情報の差異を抽出する。しかし、佐伯らの研究では位置情報を用いて国籍の違いから観光情報の抽出をしているが、国籍別のツイート内容を考慮した分析は行えていない。そのため、本研究とは位置情報だけでなくテキストを用いて分析を行う点で異なる。

大久保らは [7] は、旅行ガイドブックと口コミサイトを用いて観光地イメージに関する研究を行っている。旅行ガイドブックは訪日前の観光地に対しての期待を形成し、口コミサイトは実際に観光地を訪れた際の評価が含まれていると仮定して、言語比較解析することにより、訪日外国人の観光地への期待と評価のギャップを出身国別に抽出している。Stepchenkova ら [8] は、ロシアとアメリカの観光に関する web サイトの情報から、ロシア人とアメリカ人から見たロシアの観光地に対してのイメージ

違いを分析した。これらの研究のように本研究でもユーザの国籍に着目し、テキストを用いて観光地への反応の差異を抽出を試みているが、異なる言語で投稿されたテキストを同じ尺度で比べるという点で異なる。

2.2 Twitter から LDA を用いてトピック抽出を行う研究

本研究では LDA によりツイートをトピックに分類し分析を行う。その際に以下の研究を参考にした。

渡邉ら [9] は、観光地を検索すると、その観光地と類似した観光地を可視化して表示するシステムを開発した。Twitter と Web 情報から観光地に関するテキストを抽出し、LDA により作成したトピックモデルを用いて類似した観光地を発見した。古賀ら [10] は、LDA を用いてユーザシステムの提案をしている。各ユーザのツイートに対し LDA によりトピックモデルを作成し、KL ダイバージェンスを用いて類似ユーザの提案を行なっている。堂前ら [11] は、ツイートの中のトピックには、地域に偏りがあるものと共通で現れるものがあるという仮定のもと、LDA から得たトピックを利用し、Twitter ユーザの生活に関わる地域を推定している。

3. 提案手法

本章では、訪問先への反応の差異を分析するために、日本人と外国人の異なる言語で投稿されたツイートを同一の尺度で扱い、比較する手法について述べる。本手法の大まかな手順を以下に示す。

- (1) 収集したツイートを日本人ツイートと外国人ツイートに分類
- (2) 外国人ツイートの本文を日本語に翻訳し、全ツイートのコーパスを作成
- (3) コーパスをもとに、東京都におけるトピックモデルの作成
- (4) 東京都をメッシュに分割し、日本人と外国人それぞれのトピック分布をメッシュごとに作成
- (5) 日本人と外国人のトピック分布の差異をメッシュごとに分析し、訪問地に対する反応の違いを抽出

3.1 日本人ツイートと外国人ツイートに分類

本節では、ツイートに対しユーザが付与した使用言語情報とツイート本文の言語判定により、分析に用いる日本人ツイートと外国人ツイートの抽出手法について述べる。

はじめに、東京都内で投稿された位置情報付きツイートを収集する。ツイートの収集には、Twitter Streaming API^(注4)を利用した。

次に、ユーザのプロフィール情報に含まれている使用言語情報を用いて、ユーザを日本語ユーザと英語ユーザに分類する。使用言語情報が日本語のユーザを日本語ユーザとし、英語のユーザを英語ユーザとする。今回は、日本語と英語以外の言語に設定されているユーザは使用しない。次に、分類したユーザの投稿のうち、分析に用いるツイートを抽出する。Twitter の

(注4) : <https://developer.twitter.com/en/docs>

ユーザのプロフィール情報に含まれている使用言語設定は、登録時にアクセスしたページの言語によって自動的に設定されるため、日本語設定のユーザの中にも日本人ではないユーザが存在している。そこで、収集したツイートの本文に対して Language-Detection^(注5) を適用することで言語判定を行う。このとき、前処理として、Instagram や Foursquare などの SNS と連携しているツイートやリプライ、リツイートを除いた。また、ツイート本文から、URL 部分と絵文字や記号を除いたのち、言語判定を行った。今回は、日本語ユーザの投稿のうち、言語判定の結果が日本語だったものを日本人ツイート、英語ユーザの投稿のうち、言語判定の結果が英語だったものを外国人ツイートと定義し、分析に用いる。使用言語情報と言語判定が一致しない投稿に関しては使用しない。

3.2 異なる言語のツイートを考慮したコーパスの作成

本節では、抽出した日本人ツイートと外国人ツイートを合わせたツイート集合を1つの文書とし、コーパスと呼ばれるテキスト集合を作成する手法について述べる。

トピックモデルを作成する前段階として、コーパスを作成する必要がある。この際に日本語と英語を同じ尺度で扱えるように考慮しなければならない。そこで、Google Translate API^(注6) を用いて英語のテキストを日本語に翻訳する。そうすることで、異なる言語のツイート集合を日本語のツイート集合として扱い、全ツイートを1つの文書とすることができる。

次に、各ツイートの本文に対して MeCab^(注7) を用いて形態素解析を行い、ツイート本文を単語単位に分解し、単語を全て基本形にする。MeCab の辞書データには、新語・固有表現に強く語彙数も多い mecab-ipadic-NEologd^(注8) を使用する。ユーザの訪れた地域に対する反応を抽出するためには、ユーザの反応と結びつかない品詞の単語は極力除去する必要がある。そこで、形態素解析の結果から名詞・動詞・形容詞と判定された品詞の単語のみ利用する。また、ストップワード^(注9) を用いてどの文章にも出てくるような単語を除去する。

品詞のフィルタリングを施した各ツイートを bag-of-words によるベクトルとして表現する。1ツイートを1文とし、全ツイートの集合を1つの文書とする。この文書を以降の分析において、コーパスとして利用する。

3.3 東京都のトピックモデル作成

本節では、作成したコーパスを用いて、東京都におけるトピックモデルを作成する手法について述べる。

トピックモデルの抽出には LDA を利用する。LDA は、一つの文書が複数のトピックからなると仮定した確率的トピックモデルである。LDA を利用することで、一つの文書に対してのトピックモデルと各トピックを構成する単語の分布を算出できる。また、トピックモデルから各文書に対してのトピック分布を得

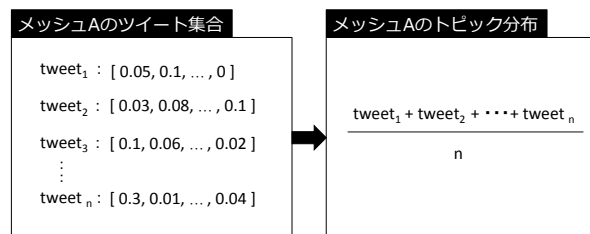


図1 メッシュにおけるトピック分布作成の例

ることができる。トピック分布とは、それぞれのトピックが出現する確率である。そこで、作成したコーパスに対して LDA を適用することで、東京都全体におけるトピックモデルを作成する。

3.4 メッシュごとのトピック分布作成

本節では、東京都をメッシュで分割し、日本人ツイートと外国人ツイートそれぞれにおけるトピック分布を、メッシュごとに作成する手法について述べる。メッシュごとのトピック分布を日本人ツイートと外国人ツイートそれぞれで作成するのは、それらの分布をメッシュごとに比較し、差異を抽出するためである。

はじめに、作成したトピックモデルを用いて、各ツイートのトピック分布を算出する。トピック分布は、対象とするツイートでそれぞれのトピックが出現する確率分布のベクトルで表される。次に、国土交通省が公開している3次メッシュコード^(注10)に基づき、東京都を約1キロメートル四方のメッシュに分割する。ツイートに付与されている位置情報を用いて、そのツイートがどのメッシュに含まれているかを判別し、分割したメッシュごとツイートを振り分ける。また、メッシュ内に割り振られたツイートのユニークユーザ数が閾値よりも低ければ、そのメッシュは利用しないこととする。

最後に、各メッシュに含まれる日本人ツイートと外国人ツイートそれぞれのトピック分布の平均値を算出する。閾値を超えたすべてのメッシュに対して、トピック分布の平均値を算出し、それをメッシュのトピック分布とする。図1に、メッシュにおけるトピック分布の作成の例を示す。

3.5 可視化・分析

本節では、メッシュごとに日本人ツイートと外国人ツイートそれぞれにおけるトピック分布の類似度を求め、算出された類似度の大きさにより色分けをすることで可視化し、訪問地に対する反応の差異の分析をする手法について述べる。

まず、メッシュごとに日本人ツイートと外国人ツイートそれぞれのトピック分布同士の類似度を求める。トピック分布はトピックの生成確率分布であるため、確率分布間の類似度を求める手法が必要である。確率分布間の類似度を求める手法として、比較対象のどちらかの確率分布の値に0が存在しない条件を持つ Kullback-Leibler divergence (KL 情報量) があり、

(注5) : <https://code.google.com/archive/p/language-detection/>

(注6) : <https://translate.google.co.jp/m/translate?hl=ja>

(注7) : <http://taku910.github.io/mecab/>

(注8) : <https://github.com/neologd/mecab-ipadic-neologd>

(注9) : <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

(注10) : http://nlftp.mlit.go.jp/ksj/old/old_data_mesh.html

$$D_{KL}(P, Q) = \sum_{i=1}^T p(i) \log \frac{p(i)}{q(i)} \quad (1)$$

と表せる。ただし、P と Q を日本人ツイートと外国人ツイートそれぞれのあるメッシュにおけるトピック分布、T をトピック数、i をトピック番号、p(i) と q(i) をそれぞれのメッシュにおけるトピック i の生成確率とする。KL 情報量は非対称な関数である。そのため、2 つの平均的な確率分布までの KL 情報量の平均を求めることで、比較の対称性を満たす Jensen-Shannon divergence (JS 情報量) があり、

$$D_{JS}(P, Q) = \frac{1}{2}D_{KL}(P, M) + \frac{1}{2}D_{KL}(Q, M) \quad (2)$$

$$M = \frac{1}{2}(P + Q) \quad (3)$$

と表せる。本研究では、比較の対象性を満たすため JS 情報量を用いて日本人と外国人の各メッシュ同士の類似度を求める。また、JS 情報量は 2 つの確率分布間の類似度が低いほど値が大きくなる。メッシュごとに求めた類似度の高さにより、メッシュの色を塗り分けることで可視化をおこなう。可視化した結果から類似度の低いメッシュについて分析をおこなう。

4. 予備実験

本章では、予備実験について述べる。

4.1 トピックモデルの検証

本節では、作成したトピックモデルが正しく作成されているかの検証について述べる。提案手法によってトピックモデルを作成したが、これをツイートに適用して抽出したトピックのうち、日本人ツイートにしか存在しないトピックや、外国人ツイートにしか存在しないトピックがある場合、日本人と外国人を同じ尺度で比較できない可能性が高い。そこで、そのようなトピックの偏りがなければ検証を行った。

まず、ツイートに付与されたトピック分布から、確率が最も大きなトピックをそのツイートのトピックとし、ツイートそれぞれにトピックを与える。次に、ツイートに付与されたトピックを集計し、トピックに対するツイートの割合をヒストグラムで表す。図 2 に、日本人ツイートと外国人ツイートのトピックの割合を表したヒストグラムを示す。横軸はトピック番号を表し、縦軸及びヒストグラムの色は、日本人と外国人それぞれのツイートの割合を表す。

図 2 は、ある特定のトピックに分布が偏っていないことを示している。また、日本人のツイートの割合が高く外国人のツイートの割合が低いようなトピックはない。また、外国人のツイートの割合が高く、日本人のツイートの割合が低いトピックもないことがわかる。このことから、日本人ツイートにのみ表れるトピックや、外国人ツイートにのみ現れるトピックは存在せず、トピック分布に偏りがなければ検証できた。

5. 分析実験

本章では、3 章で述べた手法を用いて東京都を分割したメッシュごとに日本人ツイートと外国人ツイートのトピック分布を算出し、トピック分布の類似度に着目した分析 (5.3) を行う。

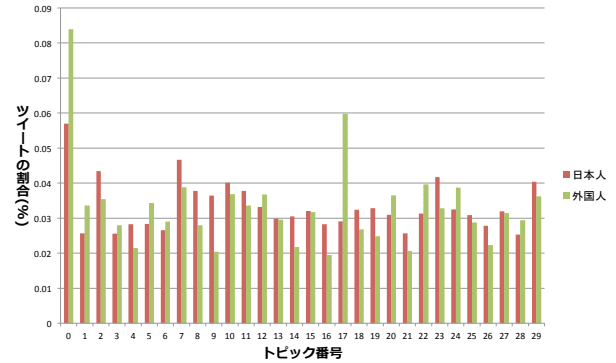


図 2 日本人と外国人のトピックに対するツイートの割合

表 1 ツイートから算出された主なトピック

トピック	トピック上位単語
10	旅行 ゲーム 建物 アニメ フード
18	走る 泣く サイクリング 路地 散歩
21	花 愛 写真撮影 女の子 美しい サクラ

5.1 データセット

2017 年 1 月から 2017 年 3 月の間に東京都内で投稿されたジオタグ付きツイート 513,384 件を用いた。3 章で述べた前処理に加え、bot ツイートなどを省くために期間中に 100 件以上投稿しているユーザのツイートは除去した。3.1 節の結果、日本人のツイートは 121,602 件、外国人のツイートは 18,828 件となった。

5.2 実装

LDA の実装には、python のライブラリである gensim^(注11)を用いた。また、LDA のトピック数は 30 とした。トピック分布を作成するメッシュのユニークユーザ数の閾値を、外国人は 10 人、日本人は 20 人とした。JS 情報量導出の際には、python のライブラリである SciPy^(注12)を用いた。

5.3 トピック分布の類似度に着目した分析

東京都を分割したメッシュにおける日本人と外国人のトピック分布の類似度により色分けを行なった可視化結果を、図 3 と図 4 に示す。図 3、図 4 において、日本人と外国人のトピック分布の類似度が高いメッシュは色が薄く、類似度が低ければ色が濃くなっている。

まず、図 3 の高尾山を含むメッシュに注目すると、メッシュの色は濃くなっており類似度が低い。このメッシュで投稿されたツイートを見てみると、外国人のツイートには、景色の綺麗さなど、高尾山の景色に関するツイートが多かったのに対し、日本人のツイートには景色を言及したツイートは少なく、高尾山に登った達成感や高尾山で食べた物の感想を述べるツイート

(注11) : [urlhttps://radimrehurek.com/gensim/](https://radimrehurek.com/gensim/)

(注12) : [urlhttps://www.scipy.org](https://www.scipy.org)

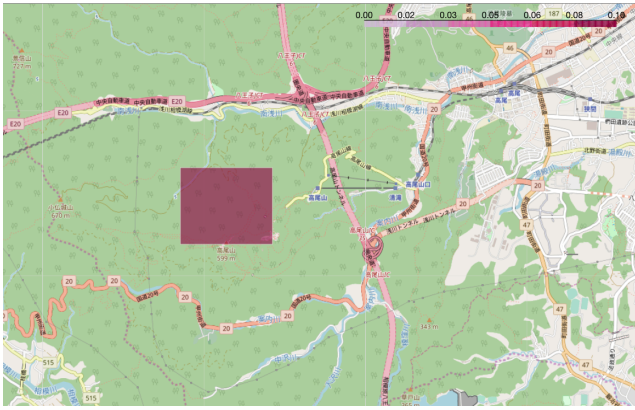


図3 高尾山を含むメッシュの類似度の大きさにより色分けを行なった可視化結果

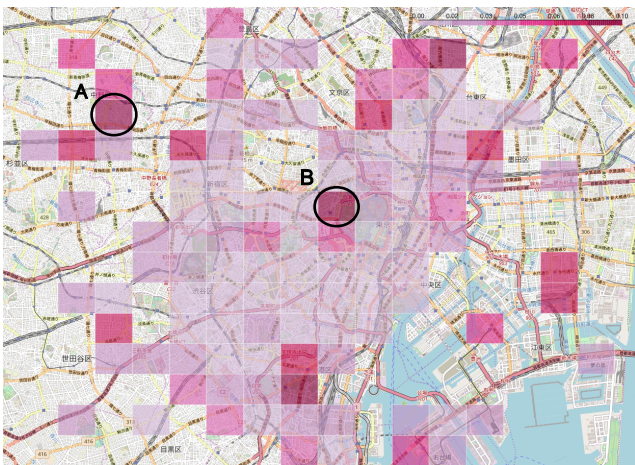


図4 東京都におけるメッシュごとの類似度の大きさにより色分けを行なった可視化結果

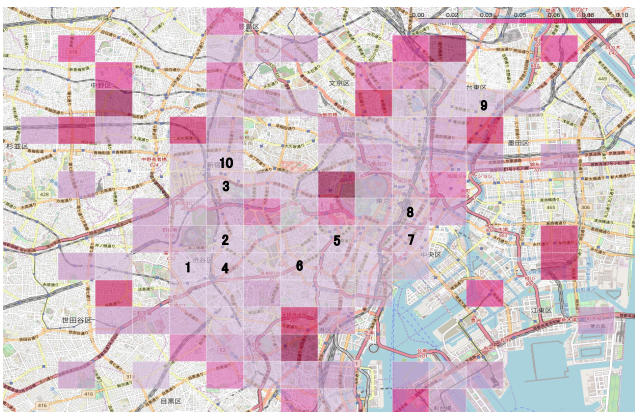


図5 訪問者数上位10箇所のメッシュ

などが大多数を占めていた。そのため、メッシュの類似度の低さは、この高尾山に対する反応の違いから抽出されたと考えられる。

次に、図4のAで囲まれている中野ブロードウェイを含む

メッシュも色が濃く類似度が低くなっている。日本人と外国人のAで囲まれているメッシュにおけるトピック分布の差をとると、表1より、トピック10の旅行、ゲーム、アニメ、などの単語を含むトピックが最も差が大きくなった。また、このトピックはトピック分布において外国人の方が数値高い。このことから、中野ブロードウェイでは外国人は日本人よりもゲームやアニメに対する反応が大きいことがわかる。また、外国人が旅行の際、アニメやゲームを目的に中野に来る人が多いことが考えられる。

次に、図4のBで囲まれている皇居周辺を含むメッシュに注目する。Bのメッシュも色が濃くなっており類似度が低くなっている。Bのメッシュにおける外国人のトピック分布に注目すると、表1より、トピック21の写真撮影、花、サクラ、美しいなどの単語を含むトピックの数値が高かった。また、Bのメッシュにおける日本人のトピック分布に注目すると、表1より、トピック18のランニング、散歩、サイクリングなどを含むトピックや食事を含むトピックの確率が高かった。このことから、皇居周辺において、外国人は花や写真撮影など風景に関することに興味があると考えられる。対して、日本人はランニングや、サイクリングなど運動に関することに興味があると考えられる。最後に、メッシュ内のユニークユーザ数を集計し、人数が多い上位10箇所の可視化結果のメッシュを確認したところ、図5のように全てのメッシュが最も薄い色になっている。これらのメッシュは、渋谷駅や新宿駅、池袋駅といった交通の要所となる場所であるためにユーザ数が多くなっていると推測される。しかし、これらの場所はあくまで経由地点であり、目的地として訪れているユーザは少ない。そのため、国籍に関わらず、投稿場所に関連しないツイートが多く投稿されているためであると考えられる。また、人数が多ければ投稿場所に関係のない日常生活で投稿されるようなツイートの割合が高くなってしまうため、類似度が高くなったと考えられる。

6. まとめ

本研究では、英語のツイートを日本語に翻訳することで、異なる言語のテキストを同じ尺度で扱う手法を提案した。そして、提案手法を位置情報付きツイートに適用し、場所ごとのトピック分布を作成した。日本人と外国人の訪れる場所に対する反応には差異があるという仮定のもと、作成した日本人と外国人のトピック分布の類似度を求めることで分析を行った。その結果、国籍の違いから、皇居周辺など訪問先でのツイートにおける反応の差異があることを確認した。

今後の課題として、人数の多いメッシュは投稿場所よらないツイートの割合が多くなってしまう、反応の差異を抽出する事が出来なかったため、在住者と訪問者を分類し、分析を行う必要があると考える。そのためには、1つのツイートだけではなく、1ユーザの全ツイートに着目して、ユーザの全投稿に対して分析地域で投稿された割合から在住者と訪問者を分類する事ができると考えられる。また、外国人でも在日外国人と訪日外国人の違いで訪問地の反応に差異はあると考えられるが、本研究では考慮することができなかった。また、手法の評価実験を

行うことができなかったため、手法の有効性を検討する必要があると考えられる。

今後の展望として、本研究で用いた Twitter などの SNS だけでなく、日本人と外国人のブログなども分析に用いることが挙げられる。これによって、観光地に対しての反応の差異をより的確に捉えることができると考えられる。また、英語だけでなく他の言語でも応用し、複数の国籍での反応の差異を分析することを検討している。

謝 辞

本研究は、首都大学東京傾斜的研究費（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] 村上嘉代子, 川村秀憲. 外国人から見た日本旅行: 英語ブログからの観光イメージ分析. 人工知能学会誌, Vol. 26, No. 3, pp. 286–293, 2011.
- [2] 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. 2013, No. 28, pp. 1–6, 2013.
- [3] 土田崇仁, 加藤大受, 莊司慶行, 廣田雅春, 石川博. マイクロブログにおける投稿場所を考慮した語の関連性の分析. 第 9 回データ工学と情報マネジメントに関するフォーラム, P2-6, 2017.
- [4] 佐伯圭介, 遠藤雅樹, 廣田雅春, 倉田陽平, 横山昌平, 石川博. 外国人 twitter ユーザの観光訪問先の属性別分析. 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 (DEIM Forum 2015), C4-3, 2015.
- [5] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.
- [6] 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.
- [7] 大久保立樹, 室町泰徳. 旅行ガイドブックと口コミの言語解析による訪日外国人の観光地イメージに関する研究. 都市計画論文集, Vol. 49, No. 3, pp. 573–578, 2014.
- [8] Svetlana Stepchenkova and Alastair M Morrison. The destination image of russia: From the online induced perspective. *Tourism Management*, Vol. 27, No. 5, pp. 943–956, 2006.
- [9] 渡邊小百合, 吉野孝. 観光地間の類似性を基にした向上点発見のための観光情報可視化システム. マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, Vol. 2016, pp. 1357–1362, 2016.
- [10] 古賀裕之, 谷口忠大. 潜在トピックに着目した twitter 上のユーザ推薦システムの構築. ヒューマンインタフェースシンポジウム, pp. 867–872, 2010.
- [11] 堂前友貴, 関洋平. 地域に偏りのあるトピックを用いた twitter ユーザの生活に関わる地域推定. 研究報告データベースシステム (DBS), Vol. 2013, No. 8, pp. 1–6, 2013.
- [12] Eric Jones, Travis Oliphant, and Pearu Peterson. SciPy: open source scientific tools for Python. <http://www.scipy.org/>, 2014.