

## 膨大なジオタグ付きツイートから沖縄等の地域課題の効率的検出

倉田陽平<sup>†</sup>・堀健一朗<sup>†</sup>・三原洋太郎<sup>‡</sup>東京都立大学大学院都市環境科学研究科観光科学域<sup>†</sup>東京都立大学大学院都市環境学部観光科学科<sup>‡</sup>

## 1. 先行研究とその課題

倉田・吉田 (2023) は「わざわざジオタグを付けてつぶやかれている不満談ツイートは行政や地域住民と共有したいような地域課題を指摘しているツイートである」という仮説のもと、北海道内のジオタグが付けられたツイート 365056 件に対し、機械学習 (BERT) によって構築された不満談判別器を適用し、実際に、911 件の不満を述べたツイートを検出し、その中から 143 件の地域課題を指摘するツイートを検出した。この研究においては抽出された不満ツイートには自宅や職場などと思しき固定的な地点のジオタグを付けて暑いだの眠いだの日常の愚痴を日課のように述べただけのただの愚痴ツイートが膨大にあり、その中から地域課題指摘ツイートを目視で見つける作業のストレスが実感された。この不毛な目視作業を回避するアイデアとして倉田・吉田 (2023) はジオタグの位置が固定的な投稿者の投稿を排除することを提案した。また倉田・吉田 (2023) が用いた不満談分類器は彼らが学術研究用に提供された不満談コーパスを学習データにしており、第三者が再現できないという問題があった。

## 2. 本研究の目的

①先行研究で得られた「ジオタグが固定的な投稿者の投稿を排除すれば参考にならないツイートを排除できる」というアイデアの有効性を実証する。②独自構築が必要な不満談分類器を使わなくてもより一般的な感情分析ツールを援用すれば不満談ツイートを抽出できることを実証する。③先行研究では北海道を対象地として、膨大なジオタグ付きツイート中に行政施策の参考になりそうな地域課題指摘ツイートがあることを実証したが、今回は別の対象地 (南西諸島) で同様の実証を行う。以上三点を本研究の目的とする。なお南西諸島を対象地としたのは特定の緯度より南方という条件で抽出しやすく、北海道同様、観光客の往来が多く、観光関係の地域課題の指摘が期待されるためである。

## 3. ツイートデータの取得

本研究では当研究室修了生の鈴木祥平氏 (現東京工大メディア学部助教) が 2016~17 年に収集した日本周辺のジオタグ付きツイート (3246 万件) を利用した。これに対し鹿児島島の佐多岬 (北緯 31 度) 以南という条件で抽出を行ったところ 236932 件のツイートを得た。その空間分布を図 1 に示す。これを見ると沖縄県はもちろん、周辺国や洋上にも多数のツイートが分布していることが分かる。

洋上ツイートについては、電波の到達範囲を考えるとこれら船員や船客が洋上からツイートしたものである可能性は考えづらい。またそもそも洋上なのでその地点に関する地域課題に言及している可能性も考えづらい。また国外ツイートについては地域課題を指摘しているかどうか判断することが難しいので、本研究では検討対象外とした。そこで、GIS を用いてジオタグの位置が日本国土上のツイートのみを抽出したところ、ツイート数は 156684 件へと絞り込まれた。

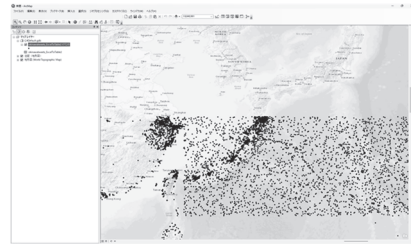


図 1 対象とした北緯 31 度以南のジオタグ付きツイートの空間分布

## 4. 大量投稿者の観察

ツイート利用研究で常々問題となるのが、コンピュータによる大量の機械的投稿 (いわゆるボット) である。そこで、まず投稿数が異常に多い (上位 1%) 30 投稿者のツイートについて観察したところ、气象台と思しき地点のジオタグを付与してひたすら天気状況を投稿し続ける気象ボットや、台風を中心位置をジオタグ参照してひたすら台風状況を報告し続ける台風ボットや、地震の震源地をジオタグ参照してひたすら報告し続ける地震ボットなどが見られた。前節で指摘した洋上ツイートはこの二者が原因だった。

## 5. ジオタグが固定的な投稿者のツイートの削除

先行研究で得られたアイデアをふまえ、自宅や職場と思しき位置のジオタグを付けて連日つぶやいている投稿中毒者の投稿は地域課題の把

Efficient Detection of Regional Issues in Okinawa from a vast amount of geo-tagged tweets

<sup>†</sup>Yohei Kurata, Kennichiro Hori, Dept. of Tourism Science, Graduate School of Urban Environment Sciences, Tokyo Metropolitan Univ.

<sup>‡</sup>Yotaro Mihara, Dept. of Tourism Science, School of Urban Environment Sciences, Tokyo Metropolitan Univ.

握において役立ちそうにないので、ジオタグが固定的な投稿者の投稿を削除することにした。ここで問題となるのが「ジオタグが固定的」の判定をどうするかという点である。一案としては、個々の投稿者につきジオタグの分布円を求め、その直径からジオタグが固定的かを判断すると言った手法が考えられるが、本研究では簡便さを重視し、エクセルのピボットテーブル機能で容易に求められる①経度の標準偏差と②緯度の標準偏差を候補に考え、後者を採用した。なぜなら経度差1度に相当する距離は地球の南北で変化するが、緯度差1度の距離は地球上で不変であるためである。また、先行研究で用いた北海道のツイートデータを再分析したところ、ジオタグの緯度の標準偏差が0.7度(78km相当)未満の投稿者はほとんど地域課題を指摘するようなツイートをしていなかった実績が判明した(図2)ので、今回のデータに対してもジオタグの緯度の標準偏差が0.7度未満の投稿者を検出したところおよそ7割の投稿者が該当し、彼らのツイートを削除したところ、総ツイート数は156684件から7203件に絞り込まれた。

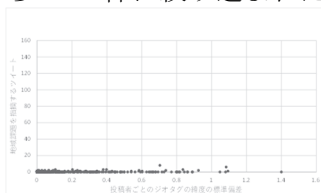


図2 先行研究データでの投稿者別のジオタグの緯度の標準偏差(横軸)と地域課題指摘ツイート数(縦軸)との関係

### 7. 目視による地域課題指摘ツイートの抽出

先の7203件のジオタグ先変動投稿者ツイートに対し二名の判定者により「地域課題を指摘するツイート」とさらに「観光に関する課題を指摘するツイート」の目視抽出を行った結果、前者は72件、後者は102件が抽出された。以下に具体例を示す。なお特に空港や港については暇を持て余す人々が盛んに書き込みを行うためか、課題指摘が多く見られた。また観光関係課題指摘ツイートとしては、案内不足やオーバーツーリズム関係の指摘が多く見られた。

**地域課題指摘ツイートの具体例:**「自衛隊機が並んでるのもみんな見えてるんだけど、見せてええもんなんか、こういうの(@国内線ターミナル 見学者デッキ(南))」「ネコ避けのためらしい。観光客にはウケるニャンコの影に住人の苦労があるようで。@座間味島」「大東犬ココいま目の前イノシシが駆け抜けてったこわかったあー」「もう1週間沖縄いるけど那覇市本当にバイク駐車場無い。警察の窓口かけても把握して

無いレベル。路面は石灰岩?混ぜた舗装ですぐ滑るし。いつかまたバイクで来る県じゃないな。」「くらげ注意報発令中(@宮之浜)」**観光に関する課題指摘ツイートの具体例:**「下田原城跡説明ないから分かんないw」「これまた久高島の秘境ビーチヤマガービーチ道無き道を進んで岩の間を這って通りねけんに行けないビーチ地元の人しかわからんわコリ…」「楽しみにしてた穴場の海ここ行くまでに険しい所があって朝降った大雨で地面ぐしゃぐしゃでジャングル化した」「久高島の秘境ビーチ、ウティ浜手作りハシゴを恐々降りていった…まじで怖かったwww@久高島」「夜は閑散とした…期待外れ@美浜アメリカンビレッジ」「非常に見晴らしの良さそうなちゅらさん展望台は終了しました空気読めない観光客のせいで土地を提供していた牧場が拒否だそうで」

### 8. 感情分析判定との対比

最後に感情分析の援用によってより簡便に地域課題指摘ツイート抽出できるようになるかを検討するために前記のジオタグ先変動投稿者ツイート7203件に対しNTT コムのCOTOHA APIを利用し感情分析を行い、文章のネガポジを機械判定した。この結果(表1)、ポジティブ判定のツイートには地域課題を指摘するようなツイートはほぼ含まれず(1096件中8件)、即ちポジティブ判定のツイートは目視対象としなくても構わなさそうであることが示唆された。一方でネガティブ判定のツイートには課題を指摘するようなツイートが多く含まれ(277件中28件)、注意して目視する必要性が示唆された。一方でネガティブツイートが即苦情というわけでもなく、「むし暑い」といった単なる愚痴や、「綺麗すぎてずるい」など遠回しな表現で賞賛しているツイートも散見された。このことから単語レベルの極性の積み重ねで文書のネガポジを判定する感情分析の援用では限界があり、ある程度文意にまで踏み込んで、苦情か否か判定できるBERTのような手法の必要性が感じられた。

表1 課題指摘ツイートと感情分析判定との関係

	全ツイート	地域課題指摘ツイート
Positive	1096	8
Negative	277	28
Neutral	5797	36
判定不能	33	0
総数	7203	72

**謝辞** 本研究は日立財団倉田奨励金の助成の下で実施された。**参考文献** 倉田陽平, 吉田伊武貴(2023)膨大なジオタグ付きツイートから見る道内各地の観光行政の課題 観光情報学会第19回全国大会予稿集, pp. 26-27.