

奨励金No.1493

膨大な位置情報付きツイートから地域課題は見つかるか？

倉田 陽平

東京都立大学都市環境学部観光科学科 准教授

Can we find local administrative issues from a huge amount of geo-tagged tweets?

Yohei Kurata,

Department of Tourism Science, Graduate School of Urban Environmental Sciences,
Tokyo Metropolitan University, Associate Professor



「わざわざジオタグを付けて不満を述べているツイートは地域課題を指摘しているのではないか」という仮説のもと、まず機械学習（BERT）によって独自の不満ツイート検出器を構築し、それをまず北海道内のジオタグが付いた膨大なツイートに適用し、北海道内のジオタグ付き不満ツイートを目視確認したところ、割合は小さいものの確かに地域課題を指摘するツイートが散見された。一方でジオタグをつけながら自宅や職場で日常的な愚痴を投稿しているツイートも多数見られた。そこで、日常的な愚痴ツイートの目視を回避すべく、あまりジオタグ位置が変化しない投稿者による投稿を無視し、対象地を南西諸島に変えて膨大なジオタグ付きツイートを検証したところ、確かに地域課題を指摘するツイートが効率的に見つけられた。また、手法の一般化のため、独自の不満ツイート検出器の利用を回避すべく、地域課題指摘ツイートと感情分析の結果との関連性を見たところ、地域課題指摘ツイートは感情分析でネガティブ判定を受けるものに多く含まれ、感情分析を使えば不満ツイート検出器を使わずとも地域課題指摘ツイートを効率的に検出できる可能性が示唆された。そこで、次は対象地を愛知県に変え、ジオタグ位置の変わらない投稿者による投稿の無視、ネガティブ判定のツイートのみの目視確認を実施したところ、効率的な地域課題の検出に成功した。

Based on the hypothesis that “geo-tagged complaining tweet may point out local administrative issues” we first built our original complaining-tweet-detector using machine learning technique (BERT), applied it to a large number of geotagged tweets in Hokkaido, and confirmed that some geotagged complaining tweets in Hokkaido were surely pointed out local administrative issues, although, their number was small. On the other hand, there were also many geotagged complaining tweets about daily complaints at home or at work. Therefore, in order to avoid seeing the daily complaining tweets, we should ignore the tweets posted by posters whose geotag positions were stable, and in the next step we targeted geotagged complaining tweets in Okinawa, and verified a large number of geotagged tweets. In addition, in order to generalize the method and avoid the use of our original complaining-tweet-detector, we looked at the relationship between complaining tweets and the results of sentiment analysis, and found that tweets about local issues were often included in those that were negative judged by sentiment analysis, suggesting that sentiment analysis could be used to efficiently detect tweets pointing out local issues instead of the complaining-tweet-detector. In the third step we targeted geotagged tweets in Aichi Prefecture, by ignoring tweets posted by stable users, and making use of sentiment analysis we succeeded in detecting local issues efficiently.

1. 「地域課題」の定義

本研究で言う「地域課題」とは各地の公共空間に生じている種々の課題を指すものとし、行政や

地域住民や公益企業（鉄道会社など）が対処を検討すべきものを言い、個別の商店に関する問題（混雑・品切れ）や解決や予防ができない広域的問題

（寒暖や悪天候）は含まないものとした。

以下に例として本研究の手法で実際に検出した地域課題を指摘するツイートを示す。

「北海道最南端の白神岬です。トンネル脇の空き地に無理矢理作った感が凄くて、あるのは石碑と仮設トイレくらい。あとは白神岬灯台が見える距離にありますが立ち入り禁止で近付けません。休憩目的ならば、ここに来るのはやめた方が良いでしょう。<https://t.co/u7Zk5v8gdu>」[1]

「もう1週間沖縄に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」[2]

「ここの交差点は信号がわかりにくい。交差点は1つなのに、2つあると勘違いして、右左折後に止まる車がいる。（@常普請橋 in 小牧市）<https://t.co/5K7iBJAcZu>」[3]

2. 一連の研究の流れ

本研究は、三つの学会発表研究[1-3]から構成され。北海道[1]、南西諸島[2]、愛知県[3]と検証対象地を変えつつ、そのたびに手法を更改しつつ研究が進められた。いずれの研究でも鈴木祥平氏（現日本大学国際共生学科常任講師）が2016～17年に収集した日本周辺ジオタグ付きツイート（3246万件）を利用し、GIS上で各検証対象地の陸域上のツイートの抽出処理（クリップ操作）を行ったうえでそこから地域課題を指摘するツイートの抽出を行った。

倉田・吉田（2023）[1]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投稿者が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）によって独自構築された不満ツイート判別器を適用し、911件の種々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。確かに地域課題を指摘するようなツイートは実在するもののその出現率は極めて小さい。一方で、自宅や職場と思しき地点のジオタグをつけて、日頃の愚痴を常習的に投稿しているツイートがジオタグ付き不満ツイートの圧倒的多数を占めることも確認された、このことから人々の指摘した地域課題を効率よく見つけるためには、「ジオタグ参照先が固定的な投稿者によるツイートは無視してジオタグが日々変動する投稿者のツイートのみを見れば良い」という効率化策が[1]において提案された。実際にこの効率化策が機能することが倉田・堀（2024）[2]において実証された。なお[2]では佐多岬以南の南西諸島内のジオタグが付けられたツイート236932件から、72件の地域課題を指摘するツイートを検出した。また[2]においては、[1]において独自の不満ツイート判別器が用いたことを課題視し、その代替としてより汎用性の高い感情分析（Sentiment Analysis）が採用できないかという視点から、検出された課題指摘ツイートに対し、ネガポジ判定を行ったところ、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ないため、感情分析でポジティブ判定されたツイートを目視確認対象から外す、という第二の効率化手法が提案された。さらに倉田・堀（2024）[3]では[1-2]が観光地の性格が強い地域で検証を行ったため、非地元民によるツイートが検証の主対象となっていたのではないかと反省から、観光地の性格が弱い愛知県の対象とし愛知県内の地点のジオタグが付与された1319092件のジオタグ付きツイートから1984件の地域課題を指摘するツイートを検出した。

なお南西諸島を対象とした倉田・堀（2024）[2]では北海道を対象にした[1]のデータの再分析結果から。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という実績から、ジオタグの緯度の標準偏差が0.7度未満

という基準を地域課題を投稿しそうにない動かない投稿者（常習的投稿者）とみなす基準に用いたが、投稿者のジオタグのばらつきの指標として経度ではなく緯度の標準偏差を採用したのは、その算出が表計算ソフト上で容易に行え、また経度差1度に相当する距離は地球上の南北で距離が変わってしまうが、緯度差1度に相当する距離は全地球上で一定だからである。この緯度標準偏差0.7度基準を[3]にて愛知県のデータに適用してみたところ、全投稿者が「動かない投稿者」に分類されてしまったことから、広い北海道で求めた基準値を愛知県に適用することに無理があったと判断し判定基準を改め、「動き回る投稿者」と「動かない投稿者」に二分するという単純な理屈から、「ジオタグの緯度の標準偏差」が全投稿者の中央値以上か未満かをもって、動き回る投稿者か否かを判別することとしたところ、実際にこれが有効に機能することが[3]において確認された。

さらに[3]では地域課題指摘ツイートの感情分析が行われ、[2]同様に地域課題を指摘するツイートには感情分析の結果がポジティブ判定のものが含まれず、そのほとんどがネガティブ判定であるため、地域課題指摘ツイートを見出す際にネガティブ判定のツイートのみを目視確認対象とするという効率化策が提案された。

3. 三検証対象地の比較とまとめ

北海道を対象にした倉田・吉田（2023）[1]で

は地域課題を指摘するツイートの出現率は0.0192%であった。次に南西諸島を対象とした倉田・堀（2024）[2]では地域課題を指摘するツイートの出現率は0.0431%であった。最後に愛知県を対象とした倉田・堀（2024）[3]ではその出現率は0.1504%と突出していた（表1）。これは愛知県に地域課題が集中しているという意味ではあるまい。その理由を考えると愛知県は車社会であるため、個々人のパーソナルスペース意識が強く、公共交通の混雑や温度管理や日陰不足に帯する不満指摘が発生しやすい土地柄ではないかという理由に行きついた。一連の研究を通して我々は合計で $143+72+1984=$ 合計 2199 件の地域課題指摘ツイートの文例集を得た。研究開始時には地域課題とは何かという定義も定まらない状態だったが、これだけのサンプルが得られれば、逆にこれらを教師データとして機械学習（BERT）を行い、そこで地域課題指摘ツイート判別器を構築し、それを日々投稿される膨大なツイートに適用すれば、リアルタイムで地域課題の抽出が実現できるのではないかと考えている。このような地域課題指摘のリアルタイム検出が実現すれば、一般市民が使い慣れたツイッターなどのSNS上で地域課題を呟くだけで、その指摘が埋もれること無く行政に刻々へと届けられる未来が拓かれる可能性がある。一連の検証により膨大なツイートの中には、地域課題を指摘する投稿は極めて希だが確かに存在しており、ツイートを地域の課題をモニタリングす

表1 北海道・南西諸島・愛知県の地域課題・観光関連課題の出現率の比較

		北海道 [1]	南西諸島 [2]	愛知県 [3]
検討ツイート数		365056	536932	1319092
地域課題	件数	143	72	1984
	出現率	0.0392%	0.0304%	0.1504%
観光関連 地域課題	件数	70	102	159
	出現率	0.0192%	0.0431%	0.0121%

るソーシャルセンサとして活用できる可能性が実証された。ここで課題となるのが膨大なツイートの中から極めて希な課題指摘ツイートをどのように発見することだが、一連の研究で二つの検出効率化手法、①ジオタグ参照先が変動しない投稿者の投稿は目視確認対象から外す、②感情分析によりポジティブ判定される投稿は目視確認対象から外す、の有効性が確認された。

4. 発表（研究成果の発表）

- [1] 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、観光情報学会全国大会第 19 回全国大会予稿集、pp. 26-27、江別.
- [2] 倉田陽平、堀健一郎（2024）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、情報処理学会第 86 回全国大会予稿集、横浜.
- [3] 倉田陽平、堀健一郎（2024）膨大なジオタグ付きツイートからの愛知県内の地域課題の効率的検出、観光情報学会第 25 回研究発表会予稿集、pp. 33-36、名古屋.