

## 人はどこで都市の様相について述べるのか?: 位置情報とマイクロブログにおける言及地名のずれに関する分析

西澤 元希<sup>†</sup> 荒牧 英治<sup>‡</sup> 若宮 翔子<sup>‡</sup> 宮部 真衣<sup>†</sup>

<sup>†</sup>和歌山大学

<sup>‡</sup>奈良先端科学技術大学院大学

### 1 はじめに

我々は普段生活する都市の中で、街の雰囲気や環境音、出来事など様々な情報を取得している。このような都市において把握する事柄(本研究では「様相」と呼ぶ)は、まちづくりや観光、住まい探しなどの様々な場面で価値がある。例えば、森永らが提案しているランドマークによるナビゲーションシステム [1] は、人が記憶しやすい地物を抽出して経路案内に利用しており、都市の様相を活用した事例の一つである。

従来は、都市で感じたことを人間が言葉で表現することで、様相の断片を利用可能な情報(様相記録)として収集していた [2]。様相記録の収集は、主に歩行実験やイベントの実施により行われるため、収集の負荷が大きい。近年、スマートフォンやマイクロブログが普及し、マイクロブログ上に様相記録が蓄積されつつある。マイクロブログを利用することにより、様相記録の収集負荷を軽減できると考えられる。しかしマイクロブログ上に投稿される情報は、東京にいながら大阪のことについて述べるといった、様相を把握した場所と様相記録を投稿した場所が異なる場合がある。

そこで本研究では、マイクロブログに投稿されるツイートの位置情報(発信地)と、ツイートにおいて言及されている地名のずれについて分析を行う。

### 2 分析の概要

#### 2.1 分析データ

今回は、代表的なマイクロブログの一つである Twitter における投稿(ツイート)を分析対象とする。本研究では位置情報付きのツイートデータを用いて、ツイートに含まれる地名とそのツイートの位置情報の違いの有無を分析する。

検証する地名として、Wikipedia の日本の観光地一覧のページ<sup>\*1</sup>を参考にし、440 件の観光地を選定した。これらの観光地名を含むツイートを、位置情報付きツイ

ト<sup>\*2</sup>から 1 つの地名につき 1,000 件を無作為に抽出した。観光地名を含むツイートが 1,000 件に満たないものに関しては、取得できたものすべてを用いることとした。ツイートが取得できなかった観光地を除いた 390 件の観光地(計 287,267 ツイート)を分析対象とする。

次に、観光地名を含むツイートの位置情報から、各ツイートが発信された地点の都道府県名を取得した。位置情報から都道府県名を取得する際は、逆ジオコーディングサービス<sup>\*3</sup>を利用した。なお、サービスの仕様上、都道府県を取得できなかったツイートは分析対象から除外した。取得したツイートの発信地と、言及されている観光地の都道府県名を比較し、マイクロブログからの様相記録の収集について検討する。

#### 2.2 分析内容

本研究では以下の 2 点について分析する。

項目 (1) 地名を含むツイートのうち、その場所で発信されるものはどれくらいあるのか?

項目 (2) 地名と発信地が一致しない場合、そのずれにはどのような原因があるのか?

項目 1 の分析により、マイクロブログから様相記録がどの程度収集可能か明らかにする。また項目 2 の分析により、地名と発信地が一致しない様相記録の抽出可能性を考察する。

### 3 分析結果と考察

#### 3.1 言及地名と発信地の不一致率の検証

ツイート内で言及されている観光地(以下、「言及地名」と)、ツイートの発信地が異なるツイート(以下、「不一致ツイート」)が観光地毎にどの程度存在するかを調査した。観光地 390 件(287,267 ツイート)中の不一致ツイート数は 81,524 件であり、ツイート全体の不一致率は 28.4 %であった。図 1 に観光地毎の不一致率の分布を示す。縦軸が度数、横軸が不一致率を示す。図 1 より、不一致率が 10 %以下のものが大部分を占めるが、不一致率が 50 %を超えるものも 20.5 % (390 件中 81 件)存在した。したがって、多くのツイートについては地名と発信地が一致しており、様相記録として

Where People Describe Information about City?: Analysis of Difference between Location Information and Place Names in Microblog Texts  
Motoki Nishizawa<sup>†</sup> Eiji Aramaki<sup>‡</sup> Shoko Wakamiya<sup>‡</sup> Mai Miyabe<sup>†</sup>

<sup>†</sup>Wakayama University

<sup>‡</sup>Nara Institute of Science and Technology

<sup>\*1</sup><https://ja.wikipedia.org/wiki/日本の観光地一覧>

<sup>\*2</sup>2011 年 7 月 15 日から約 1 年間収集されたものであり、ツイート数は 248,117,904 件である。

<sup>\*3</sup><http://www.finds.jp/rgeocode/index.html.ja>

表1: 言及地名と発信地のずれの原因と事例

ずれの原因	ツイート(太字部分は言及地名)	地名/発信地
(1) 地名として使われており、過去の話をしている	人吉は気候も仕事も、そして人も最高やね。橋の真ん中で車止まったら、普通、助けてくれへんよ。名も名乗らず去って行ったおっちゃん。ありがとう。俺が女やったら、惚れてまう。自分が男でよかった( ^~;;	人吉(熊本県)/福岡県
(2) 地名として使われており、未来の話をしている	明日は球磨、人吉へ。ちょっと遠いけど大好きな街。なんかワクワクするw	人吉(熊本県)/福岡県
(3) 区切りに失敗している	1人吉 吉祥寺行かなきゃ!	人吉(熊本県)/東京都
(4) 人名、店名等である	でも、めだかちゃんが、一時的かも知れませんが嫌いになりました。ひとまず俺も人吉善吉君の味方でいたいと思いました。	人吉(熊本県)/東京都
(5) RT, 引用 RT 部分に含まれる	お兄ちゃんも元タインドア派だから、へろへろだよ d( ) RT @****: @**** しかし、夏目友人帳の聖地巡礼で人吉に行って、それから沖縄に一日で行くとは恐ろしいハードスケジュール…自分じゃ絶対に無理だわw w	人吉(熊本県)/沖縄県

表2: 不一致原因と該当ツイート数

	原因(1)	原因(2)	原因(3)	原因(4)	原因(5)	原因(6)	合計
ツイート数(件)	2,522(3.1%)	2,780(3.4%)	19,171(23.5%)	11,573(14.2%)	467(0.6%)	45,210(55.3%)	81,724

※複数の原因によりずれが生じている場合があるため、合計ツイート数は不一致ツイート数 81,524 件より多くなっている。

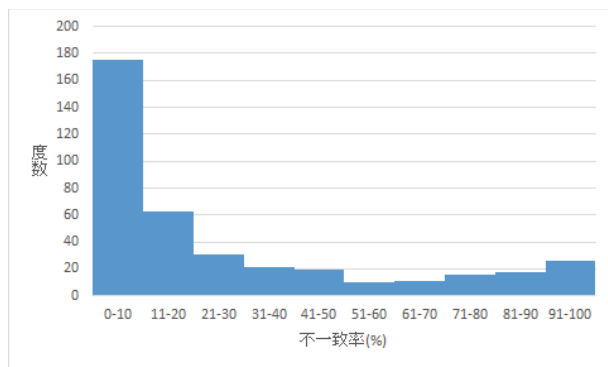


図1: 不一致率の分布

収集できると考えられる。ただし、不一致ツイートが多い地名も存在するため、マイクロブログから様相記録を収集する際には、不一致ツイートについても考慮する必要がある。

### 3.2 言及地名と発信地のずれの原因

本節では、ツイート内での言及地名と、ツイートの発信地が異なる原因について分析する。不一致ツイートの一部を確認した結果、不一致ツイートに含まれる言及地名には以下のような特徴がみられた。原因とそれぞれのツイート例を表1に示す。

- (1) 地名として使われており、過去の話をしている
- (2) 地名として使われており、未来の話をしている
- (3) 区切りに失敗している
- (4) 人名、店名等である
- (5) RT, 引用 RT 部分に含まれる
- (6) その他

また、これら6つの原因によるずれがどの程度生じているのかを、人手で分類した。分類結果を表2に示す。不一致ツイートの中でも、原因(1)に該当するツイ

トは様相として使えると考えられる。原因(2)~(6)については様相記録としては取り扱えないため、該当ツイートを除外する方法を検討する必要がある。例えば(3)に関しては、形態素解析などを用いて除外可能であると考えられる。さらに原因(2), (4), (5)に関しては「RT」や「明日」などの表現の抽出や前後文脈を見ることにより、除外できる可能性がある。ただし、原因(6)には、同一表記の地名が、異なる都道府県に存在するなどの事例が含まれるため、これらへの対応方法も今後検討する必要がある。

## 4 おわりに

本研究ではより多様性のある様相記録の収集を可能にするために、ツイートの位置情報と、そのツイート内で言及されている地名のずれについて分析を行った。具体的には、観光地を含むツイートを抽出し、ツイートの発信地と観光地の都道府県名が異なるツイートを分析し、様相記録として使用できるのかを検討した。

分析の結果、(1)約70%のツイートは発信地と地名が一致しており、様相記録として収集できること、(2)発信地と地名が一致しないものでも過去の経験は様相記録として利用できるため、他の不一致原因に関するツイートの除外による経験に基づく様相記録抽出方法を検討する必要がある、ということ进行を明らかにした。

今後は、言及地名と位置情報が一致しない様相記録の自動抽出方法を検討する。

## 参考文献

- [1] 森永寛紀, 若宮翔子, 谷山友規ほか: 点と線と面のランドマークによる道に迷いにくいナビゲーション・システムとその評価, 情報処理学会論文誌, Vol.57, No.4, pp.1227-1238(2016).
- [2] 北雄介: 経路歩行実験に基づく都市の様相の分析とモデル化に関する研究, 京大大学学位論文(2012).