

ジオタグ付きツイートで用いられる名詞の空間分布

藤原 直哉¹, 桐村 喬², 平岡 喬之³

¹ 東北大学大学院情報科学研究科, ² 皇學館大学文学部コミュニケーション学科,

³ Asia Pacific Center for Theoretical Physics

連絡先: <fujiwara@se.is.tohoku.ac.jp> Web: <<https://sites.google.com/site/nfnetz/>>

- (1) **動機:** ある地域における我々の空間認知の特性は、そこで用いられる言葉の差異を通して特徴づけることができると考えられる。近年の SNS の発展により、例えばジオタグ付きツイートデータ(桐村, 2019)などを通して、このような分析が可能となりつつある。ある特定の単語が特定の地域で多くツイートされていれば、その単語はユーザーが認識している空間的単位と関連していると考えられる。本研究は、ジオタグ付きツイートの投稿本文中に現れる各名詞が他地域と比較して多く出現する空間を分析し、我々の空間認知の特性について考察することを目的とする。
- (2) **方法:** 本研究では京都市を含む矩形領域を対象とし、ツイッターの Public streams API を用いて取得された、同地域で 2014 年 1 月から 2018 年 3 月にかけて行われた約 740 万件のポイント単位のジオタグ付きツイートをを用いて解析を行った。ツイートは、Janome で形態素解析され、一般名詞と固有名詞が抽出され、抽出された名詞ごとに利用者数を様々なサイズの正方形グリッド単位で集計する。各グリッドでの各名詞の重要度を測る指標として、空間的 TF-IDF を導入する。これは単語の重要度の指標として自然言語処理で用いられる TF-IDF を空間に応用したものであり、各単語の各グリッドでの

重要度の指標である。空間的 TF-IDF が一定値以上の領域をディゾルブし、最大連結成分を、ある単語が特徴づける地域であると考え地図に表示する。

- (3) **結果:** 名詞には、対象領域では空間依存性が明確に見えなかったもの(「人」「笑」など)と、空間依存性が見えたもの(「三条」「二条城」「祇園」など)が得られ、後者では、多くの場合、最大連結成分はそれらの地名が指し示す実際の地域とおおむね同じ場所に出現した。異なるグリッドサイズで得られた主な地名の最大連結成分を図 1 に示す。「嵐山」や「二条城」、「下鴨神社」、「祇園」、「東寺」、「伏見」などは、いずれのグリッドサイズにおいても、最大連結成分は大きく変化しないが、「金閣寺」や「祇園」はグリッドサイズによっては実際の地域との大きく乖離することがある。

(4) **使用したデータ:**

・「ジオタグ付きツイートデータ(2014-2018 年)」

- (5) **謝辞:** 本研究は東大 CSIS 共同研究 No.918 の成果の一部として実施し、JSPS 科研費 18K11462 の助成を受けた。ここに記して謝意を表したい。

(6) **参考文献:**

桐村 喬編 (2019) 『ツイッターの空間分析』, 古今書院 (出版予定) .

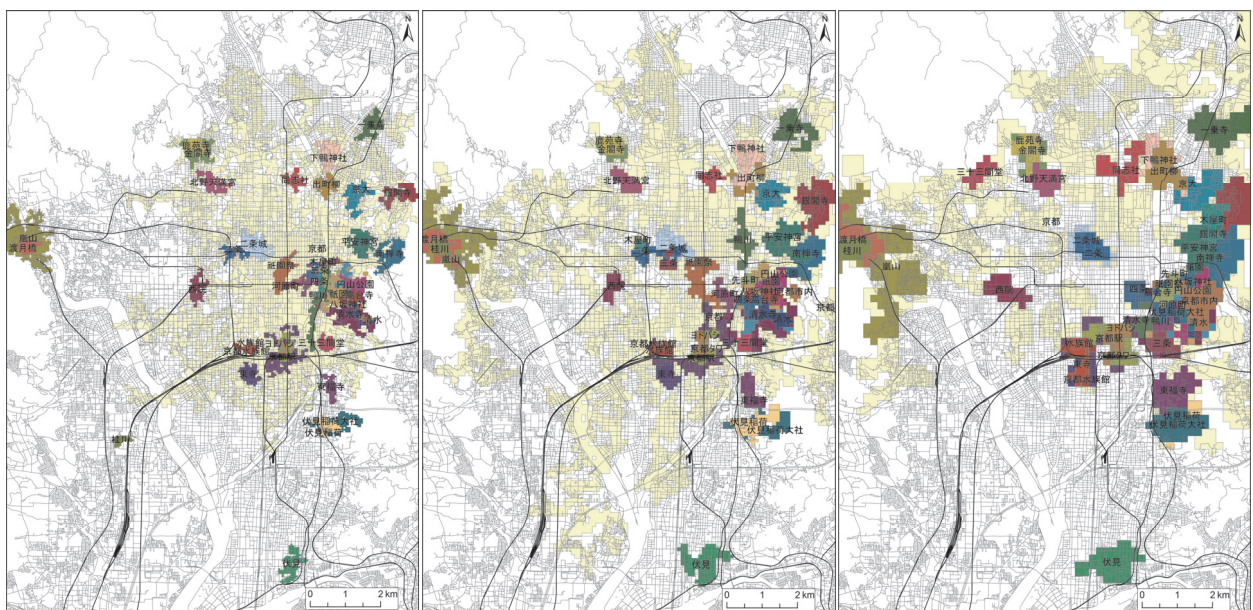


図 1: 主な地名の最大連結成分の空間分布(グリッドサイズは左: 50m/中: 100m/右: 200m)