

地理的・時間的観点を考慮した ジオタグ付きツイートの偏在性及び遍在性の可視化

遠山 由自^{†,a} 廣田 雅春^{‡,b} 石川 博^{‡,c} 横山 昌平^{†,d}

[†] 静岡大学大学院 総合科学技術研究科 [‡] 大分工業高等専門学校 情報工学科

[‡] 首都大学東京システムデザイン学部情報通信システムコース

a) gs15038@s.inf.shizuoka.ac.jp b) m-hirota@oita-ct.ac.jp

c) ishikawa-hiroshi@sd.tmu.ac.jp d) yokoyama@inf.shizuoka.ac.jp

概要 ジオタグ付きツイートにより、Twitter に投稿されている情報は地理的または時間的観点において、局所的または普遍的な分布を持っている。例えば、レストランに関するツイートでは、地理的観点において普遍的にツイートされるものとして全国チェーンのファミリーレストランがある。反対に局所的にツイートされるものとして、地域ごとに存在するご当地レストランがある。しかし、既存の検索システムでは、ある単語が持つこのような分布を把握することは困難である。そこで本研究では、クエリと共起している名詞をツイートから取得し、その共起名詞の分布を地理的及び時間的観点から可視化することで、共起名詞の中から局所的単語と普遍的単語を発見するシステムを構築する。

キーワード Twitter, ジオタグ, 局所性

1 はじめに

近年、Twitter¹ では、ジオタグ付きツイートの投稿が増加している。これにより、Twitter に投稿されている情報は地理的または時間的観点において、偏在性または遍在性を有する。

偏在性を有する情報とは、地理的観点においては地ビールのように特定の地域でツイートされる情報を指し、また、時間的観点においては、花粉症のように特定の季節にツイートされる情報を指す。一方で、遍在性を有する情報とは、地理的観点においては大手メーカーのビールのように地域を問わずツイートされる情報を指し、また、時間的観点においては、一般的な風邪などのように期間を問わずツイートされる情報を指す。

Twitter に投稿されている情報はこのような局所性および普遍性を持っているが、既存の検索システムでは、これらの情報の分布を正確に把握することは困難である。また、既存の情報の局所的な分布を抽出する研究として、長谷川ら [1]、三木ら [2] の研究があり、これらの研究では局所的な情報として地域特徴語を抽出しているが、それらが地理的及び時間的観点からどのように分布しているかは考慮していない。

そこで、本研究では、クエリと共起している名詞をツイートから取得し、その共起名詞の分布を地理的及び時間的観点から可視化することで、共起名詞の中から局所的単語と普遍的単語を発見するシステムを構築する。我々の従来の研究では [3]、地理的観点のみから分布の可視化を行ったが、本研究では時間的観点も考慮し可視



図1 可視化画面

化を行った。本研究により、既存の検索エンジンでは把握が困難である、クエリが持つ局所的または普遍的な分布が把握でき、マーケティングや観光支援に貢献できると考えている。

2 提案システム

本節では、クエリに対する分布の可視化を行うまでのシステムの流れと可視化画面について説明する。

2.1 システムの流れ

まず、ユーザが可視化画面でシステムに対して任意のクエリを入力する。次に、サーバ側でそのクエリを受け取り、一週間ごとにそのクエリに対する共起名詞の取得を行う。共起名詞の取得後、都道府県ごとにクエリと共起名詞の共起度計算を行い、共起度が高い順にソートする。ここまでの処理を一週間ごとに繰り返し、週ごとの各都道府県の共起名詞とその共起度のデータをデータベースに格納する。

2.2 可視化画面

図1に本システムの可視化画面を示す。可視化画面では、期間のスクロールバーからある期間を選択すると、その期間で共起名詞が存在する都道府県が黒く塗られ

Copyright is held by the author(s).

The article has been published without reviewing.

¹<http://twitter.com/>

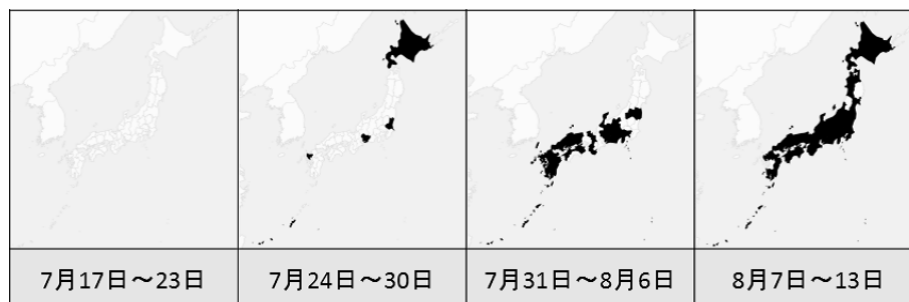


図2 「台風」が共起名詞として存在する都道府県

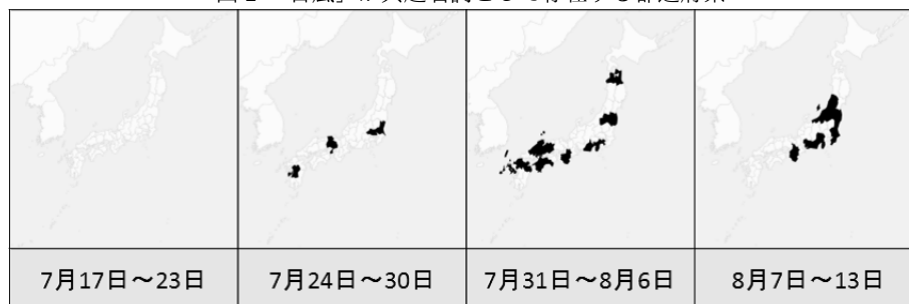


図3 「大雨」が共起名詞として存在する都道府県

る。次に、閾値のスクロールバーを変化させると、ある共起名詞が存在している県の数が増減する場合、地図上にその単語がプロットされる。これによって、共起名詞が地理的観点において局所的か普遍的か判断することができる。

地図上である県を選択すると、選択された県の全期間における共起名詞がワードクラウドとして表示される。このワードクラウドでは、期間を問わずその県で共起度が高い単語ほど大きく表示されるようになっている。そして、ワードクラウドの単語を選択すると、その単語に関する共起度推移グラフが表示される。共起度推移グラフとは、選択された単語の共起度が都道府県ごとにどのように推移するのかを表したものである。このグラフによって、選択された共起名詞がいつどこで流行っているのか確認することができる。

3 実験

本実験では、あるキーワードに対する共起名詞を地理的及び時間的観点から可視化した際に、共起名詞中に局所的または普遍的な単語が存在するか検証した。今回はキーワードとして「天気」を選択し、この単語を本システムに入力し、共起名詞の中に局所的または普遍的な単語が存在するか調査した。

まず、「天気」の共起名詞の中に「台風」という単語が存在した。図2に、2014年7月17日から8月13日にかけて一週間ごとに「台風」という単語が存在する県を示す。図2より、「台風」という単語が存在する県が西から東に向かって徐々に拡大していることが分かる。また、「大雨」という単語も共起名詞の中に存在し、その分布を図3に示す。図3より、「台風」と同様に西から東に向かって「大雨」という単語が存在する県が推移していることが分かる。これらのことから「台風」や「大雨」と

いう単語は地理的観点においては普遍的に様々な都道府県に存在しているが、時間的観点においてはある時期にだけ局所的に存在しているということが分かった。よって、共起名詞を地理的及び時間的観点から可視化した際に、共起名詞中に局所的または普遍的な単語が存在することが確認された。さらに、これらの単語が出現した期間と同様の時期である2014年7月27日から8月11日にかけて大型台風である台風11号が実際に上陸しており[4]、共起名詞を用いてある情報に対する分布を可視化することで現実と近い結果を得られることが判明した。

4 おわりに

本研究では、ジオタグ付きツイートをを用いて、ユーザが入力したクエリに対する共起名詞の分布を地理的及び時間的観点から可視化することで、共起名詞の中から局所的単語と普遍的単語を発見するシステムを構築した。実験では、あるキーワードに対する共起名詞を地理的及び時間的観点から可視化すると、共起名詞の中に局所的または普遍的な単語が存在していることが確認できた。今後の課題としては、可視化結果を高速で表示させるため計算アルゴリズムの改良や共起名詞中のノイズ単語の除去を行っていく。

参考文献

- [1] 長谷川 馨亮, 馬 強, 吉川 正俊: Twitterからの地域特徴語辞書の構築とその観光情報検索への応用, 第6回データ工学と情報マネジメントに関するフォーラム, 2014.
- [2] 三木 翔平, 新田 直子, 馬場口 登: 単語の地理的局所性の経時変化を考慮したツイートの発信位置推定, 第6回データ工学と情報マネジメントに関するフォーラム, 2014.
- [3] 遠山 由自, 廣田 雅春, 石川 博, 横山 昌平: ソーシャルメディア上に投影された情報の偏在性及び遍在性の可視化, 第6回 Wi2 研究会, 2015.
- [4] <https://ja.wikipedia.org/wiki/2014年の台風>