

位置情報付き SNS メッセージ分布に対する事件影響の評価

中渡瀬 秀一†

† 国立情報学研究所

水野 貴之‡

‡ 総合研究大学院大学

1 はじめに

本研究では位置情報付きの SNS メッセージ統計を利用して事件・事故・災害などが発生した場所の人々がそれらに対して与える注目度の評価手法を検討する。SNS による位置情報付きメッセージの発信はすでに広く普及しており、このデータを利用してローカルな事象 [1] や観光情報、地域のトピックなど場所に関連付けられた情報の抽出や編集の研究も近年盛んに行われている。

我々が注目するのは事件などが発生した時に増加するメッセージを通じて人々の事件への注目度合いを測ることである。この際、問題となるのは平常時のメッセージ密度レベルが時間や場所によって異っているため単純にその密度だけで比較できないことである。例えば同等な事件が発生しても人口密度が高い場所では低い場所より反響は大きい。したがってこれら時間・空間的な差異に関わらずメッセージ量に対する事件の影響を評価することが課題となる。

本研究では SNS データとして Twitter で発信された位置情報 (Geotag) 付きメッセージ (ツイート) を用いる。事件としては襲撃、爆破や殺傷など治安上問題となる事件を想定しており、これらの事件が発生した世界各地でのツイートデータを分析対象として収集した。本稿では例として 2015 年のフランスにおけるシャルリー・エブド襲撃事件* を用いて説明を進める。

図 1 に示すのはこの事件の発生地点 (東経 2.326 度, 北緯 48.860 度, 図中の赤点) を含む半径約 60km 圏内の Tweet の分布 (事件日とそれ以外の日) である。事件日にこのエリアでは通常に比べ約 2 割ツイートの増加が観測されている。また図 2 は事件発生地点の拡大図である。比較図により事件日にはツイート分布域の拡大が認められるが、この変化が通常の分布変動から逸脱しているのかを判断するのは困難である。そこでより定量的に分析し影響度を評価する方法を検討する。

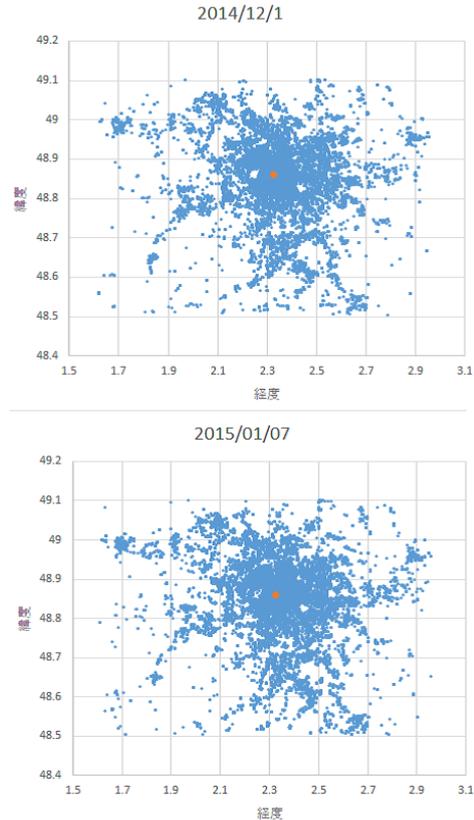


図 1: シャルリー・エブド襲撃事件前後における Tweet の空間分布の比較 (上: 事件以前の日, 下: 事件日)

2 影響の評価手法

評価を行うためツイート量を空間/時間的に規格化することを考える。説明を単純化するため以後分布は時間と発生地点からの距離の 2 次元で表現する。図 3(A) に 2015 年 1 月における 20km 圏内の分布を示す。事件現場付近では事件当日のツイート増加量が周囲に比べ著しいが元日付近も同様である。また空間的な分布は時間によらず現場から離れるほど低下傾向にある。

このような時間 (空間) に影響しない空間 (時間) 的分布を規格化によって次のように除去する。まず t 日目に観測されたある区域 d (d km 以上 $d+1$ km 未満圏内) のツイートの密度を $x_{t,d}$ とする。このとき同じ区域内で時間に関して規格化した $x_{t,d}$ の空間的規格化値 $y_{t,d}$ は $31x_{t,d} / \sum_{k=1}^{31} x_{k,d}$ となる。時間についても同様に考え $y_{t,d}$ に対する時間的規格化値 $z_{t,d}$ は $20y_{t,d} / \sum_{k=0}^{19} y_{t,k}$ となる。

Evaluating Influence of the incident on Geotagged Tweet Distribution

†Hidekazu NAKAWATASE ‡Takayuki MIZUNO

†National Institute of Informatics

‡The Graduate University for Advanced Studies

*2015 年 1 月 7 日 11 時 30 分 (CTE), フランスのパリ 11 区にあるシャルリー・エブド本社が武装襲撃され警官, 編集長ら 12 人が殺害された事件

次に前述した規格化によって、観測されたツイート分布を空間的規格化したものが図3(B), これをさらに時間的規格化したものが図3(C)である。(B)では区域ごとにツイート水準が規格化され通常時の水準が揃えられた結果、事件日(と後の抗議デモ日)のツイート増加が明確化した。時間方向に休日のツイート量が増加するパターンが残されている。一方(C)ではそのパターンも規格化によって除かれている。

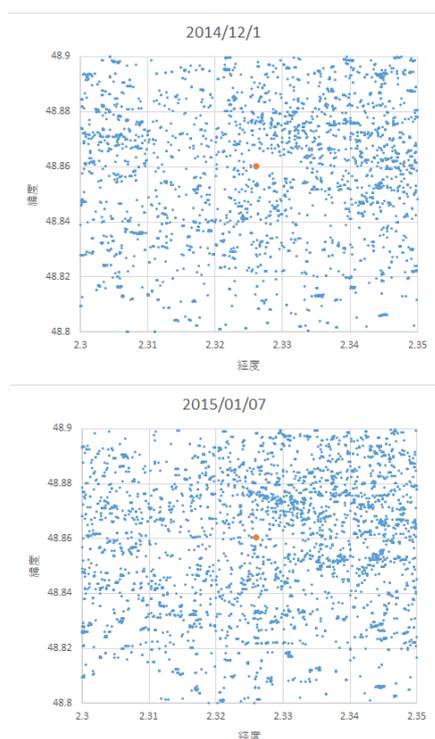


図2: 同拡大図(上: 事件以前の日, 下: 事件日)

3 手法の検証

この規格化による影響評価では平常時のツイート量水準に対して事件時のそれを比較する。その結果判明する平常時より多い部分は事件に関するツイートに相当する。実際に手法の検証を行うためにはツイートデータ(この例で日量約5万)の内容をチェックして事件に関連するツイートの量を調査する必要がある。その作業を現在進めている。

4 まとめ

本稿では事件がジオタグ付きツイートの時間空間分布に対して与える影響の評価法を検討した。事件に影響されたツイート量は時間・空間領域において減衰する特徴を持つと仮定し、平常時の分布に含まれる時間(空間)に依存しない空間(時間)固有の分布を除外する評価方法を提案した。これを実際の事件を例に説明した。今後、事件関連のツイート量を調査して精度を検証する予定である。

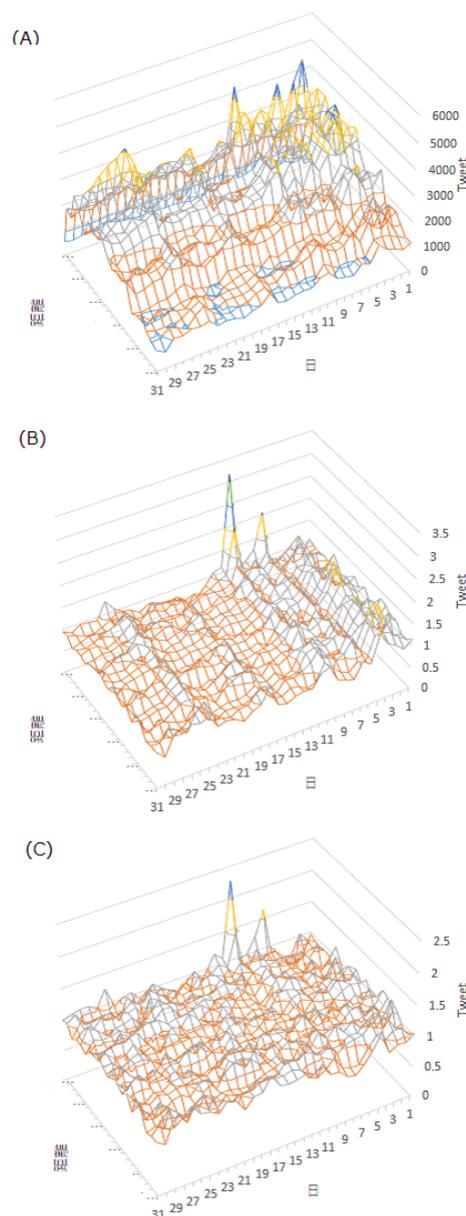


図3: Tweet分布の規格化(A: 観測値, B: 空間的規格化, C: 空間的・時間的規格化)

謝辞

本研究の一部は科研費15KT0052の助成を受けたものである。

参考文献

[1] Tatsuhiro Sakai and Keiichi Tamura. Identifying bursty areas of emergency topics in geotagged tweets using density-based spatiotemporal clustering algorithm. *Proc. of 2014 IEEE 7th Int. Workshop on Computational Intelligence and Applications*, pp. 95–100, 2014.