

東日本大震災時における首都圏の混雑状況に関する tweet の特徴

沖 拓弥[†]

東京工業大学大学院理工学研究科[†]

1. はじめに

東日本大震災時、首都圏では、地震による直接的な被害は軽微であったにもかかわらず、徒歩帰宅者や交通機関の運行再開を待つ人々によって、主要道路や一時滞在施設、鉄道ターミナル駅等は大きく混雑した。この反省を生かし、大地震時に生じる混雑の抑制策について検討することは、将来の首都直下地震や南海トラフ地震に備える上で重要である。しかし、東日本大震災時に、いつ、どこで、どの程度の混雑が生じたかを具体的に示している既往研究やデータは少ない。

本稿では、東日本大震災時における首都圏の混雑状況に関する tweet (混雑ツイート) から、共起する地名・固有名詞・時系列情報等を抽出し、混雑ツイートの頻度や時空間的特徴について基礎的な分析を試みる。

2. 混雑ツイートの定義

2011年3月11日14時から12日14時までの24時間に投稿された日本語 tweet のうち、10%をランダムに抽出した上で、まずは、混雑状況を表現する際に使用されやすいと予想される「混」という文字を含む tweet に着目した (図1(a))。さらに、KH Coderの「抽出語検索」機能を用いて、「混」という文字の用途を調べたところ、混雑に関連する可能性の高い語は、概ね、「混む」(変化語も含む)、「混乱」、「混雑」の3語に限定可能であることが判明した (図1(b))。そこで、本稿では、この3語を本文中に含む tweet を「混雑ツイート」と定義し、分析を行う。

3. 混雑ツイートの特徴

3.1 混雑ツイートの時間推移

tweet 数の時間推移を見ると (図2)、全体の tweet 数は本震の発生 (11日14時46分) 直後に5倍以上に急増し、その後は次第に減少する。これに対し、混雑ツイートの数には、本震発生当日

Characteristics of Tweets about Congestion Status in Tokyo Metropolitan Area Immediately after Great East Japan Earthquake

[†] Takuya Oki, Graduate School of Science and Engineering, Tokyo Institute of Technology

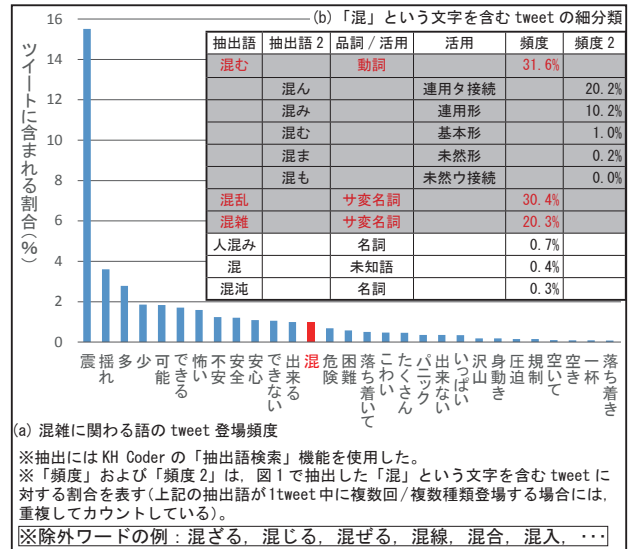


図1 混雑ツイートの抽出

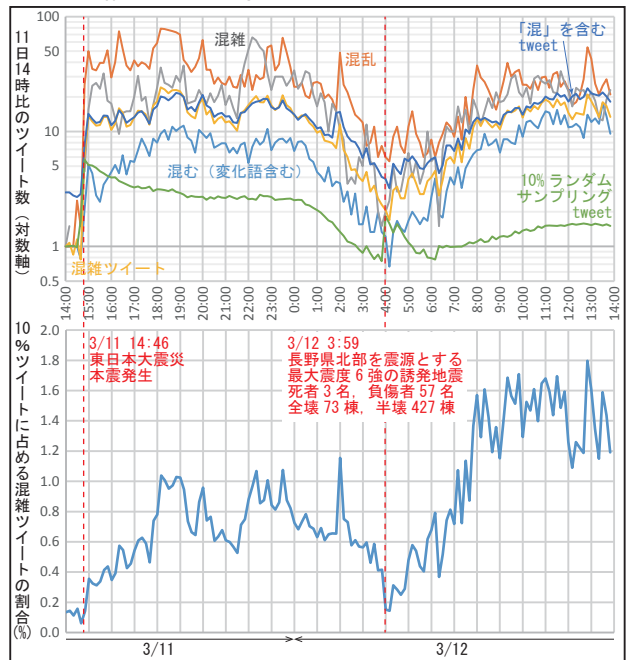


図2 tweet 数の時間推移 (10分単位)

の夕方18時以降と、翌日の8時以降に特に多くなるという特徴が見られる。これらの時間帯は、多数の人々が外出先から帰宅を試みたとされる時間帯と対応しており、混雑ツイートは、混雑状況の時間推移を良く表現できている可能性がある。

3.2 混雑ツイートの空間分布

ジオタグ (位置情報) 付きの tweet は全体の約0.2%程度であるため、混雑ツイートの空間的特

微の把握には不十分である。そこで、混雑ツイート中に含まれる固有名称・地名をKH Coderで抽出し、これを用いて位置情報を補完することを試みた(表1)。「混雑」という語は首都圏の鉄道駅・鉄道路線に関連する語と共起しやすいのに対し、「混乱」は広範囲の地域名等と共起しやすく、位置情報の補完には適さないと考えられる。

図3に、抽出した固有名称・地名(計929語)のうち、鉄道駅名・路線名データベース^{注)}に存在する588語の空間分布を示してある。鉄道路線では銀座線、東西線、半蔵門線等の地下鉄や山手線、鉄道駅では渋谷、新宿等の大規模鉄道ターミナル駅において、突出して混雑ツイート数が多い。

tweet数が増加するタイミングは、鉄道路線の運転再開時刻とよく対応しており、混雑状況の迅速な把握における混雑ツイートの有用性を示唆している(図4(a))。さらに、tweet数の増加の仕方に着目すると、銀座線は11日20時過ぎに急激に増加するが、その後は次第に少なくなっていることから、混雑は一時的であったと考えられる。一方、発災直後から翌日にかけて緩やかに増加し続けている新宿や渋谷、横浜等では、混雑状態が長時間持続していた可能性がある(図4(b))。

3.3 固有名称・地名補完における課題

混雑ツイートの多さは、ある時刻・地点における混雑状況に対する関心の高さを示していると考えられるが、混雑の程度を把握するためには、tweet本文の内容そのものをより詳細に分析する必要がある。また、混雑状況をさらに高密度に把握するためには、①tweetのサンプリングレートの値を上げることや、②KH Coderで抽出した固有名称・地名のうち、鉄道駅名・路線名以外の341語(36.7%)に含まれる、道路名(甲州街道、昭和通り等)やランドマーク名(マック、ディズニーランド、サンシャイン等)のような有益な情報を活用することが必要である。さらに、都道府県名から鉄道駅名まで様々な用途に用いられる「東京」等の地名や、全国に複数候補が存在する名称(「東西線」等)の扱いについても、今後検討を行う予定である。

4. まとめ

東日本大震災時の混雑ツイートに着目し、混雑状況の時間推移の把握に対する有効性や、tweet中に共起する固有名称・地名を用いることで、ジオタグ情報に頼ることなく、混雑ツイートの空間分布を把握できる可能性を示した。

表1 混雑ツイート中に含まれる固有名称・地名の一覧

順位	「混む」(変化語を含む) ユニーク抽出語数: 615			「混乱」 ユニーク抽出語数: 353			「混雑」 ユニーク抽出語数: 595		
	抽出語	全体	共起	抽出語	全体	共起	抽出語	全体	共起
1	日本	2.6	3.1	淡路	2.7	6.5	銀座線	3.2	11.3
2	東北	1.8	2.4	東京	5.7	6.1	東京	5.7	10.4
3	愛知	0.9	2.2	日本	2.6	3.5	渋谷	3.0	6.9
4	横浜	0.9	1.1	東北	1.8	1.8	新宿	2.5	4.6
5	京王線	0.6	0.8	千葉	1.1	1.7	半蔵門線	0.7	2.2
6	東西線	0.7	0.7	関東	0.8	1.1	池袋	0.8	2.1
7	山手線	0.6	0.7	宮城	0.6	1.2	東西線	0.7	1.8
8	東横線	0.5	0.7	仙台	0.5	0.8	南北線	0.6	1.6
9	パレ	0.3	0.7	茨城	0.4	0.8	品川	0.8	1.6
10	京浜東北線	0.4	0.5	福島	0.4	0.7	東海道線	0.5	1.4
11	総武線	0.5	0.5	太平洋	0.3	0.7	横浜	0.9	1.4
12	新宿線	0.3	0.4	田端	0.2	0.6	千代田線	0.4	1.3
13	マック	0.3	0.4	田端新	0.2	0.6	北千住	0.4	1.3
14	田園都市線	0.3	0.4	北	0.2	0.6	山手線	0.6	1.2
15	寄	0.2	0.4	気仙沼	0.3	0.4	上野	0.5	1.2
16	甲州	0.3	0.4	大阪	0.2	0.4	銀座	0.4	1.1
17	埼京線	0.2	0.4	東日本	0.3	0.4	有楽町線	0.3	1.1
18	井の頭線	0.3	0.3	新潟	0.3	0.4	荒川線	0.3	1.0
19	六本木	0.2	0.3	枝野	0.2	0.3	京王線	0.6	1.0
20	中央線	0.2	0.3	日	0.2	0.3	吉祥寺	0.3	0.9
21	秋葉原	0.2	0.3	中	0.2	0.3	三鷹	0.3	0.9
22	新潟	0.3	0.3	名古屋	0.2	0.3	日比谷線	0.3	0.9
23	目黒	0.2	0.2	近畿	0.1	0.2	東横線	0.5	0.9
24	朝	0.2	0.2	台東	0.1	0.2	築地	0.2	0.9
25	山手	0.1	0.2	関西	0.1	0.2	浅草線	0.3	0.8

※「混む」「混乱」「混雑」と共起する割合の高い順に、それぞれ抽出語を並べてある。
 ※「全体」は混雑ツイート全体に占める各抽出語の構成比(%)、「共起」は「混む」「混乱」「混雑」の各語を含むtweetに各抽出語が用いられる割合(%)を指す。
 ※首都圏に存在する鉄道駅名と考えられる抽出語を赤字で、鉄道路線名と考えられる抽出語を青字で強調して示してある。

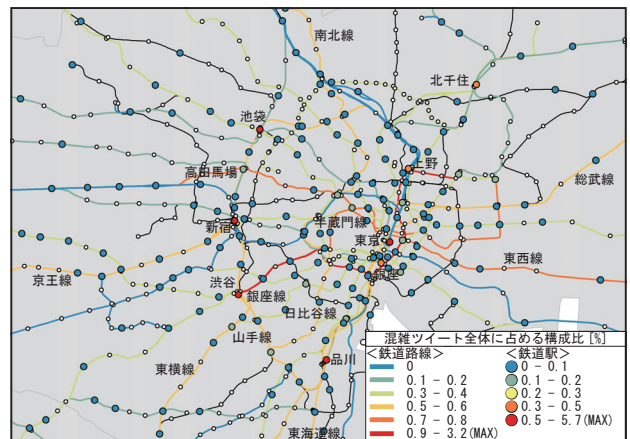


図3 混雑ツイート数の空間分布(都心部周辺) 注)

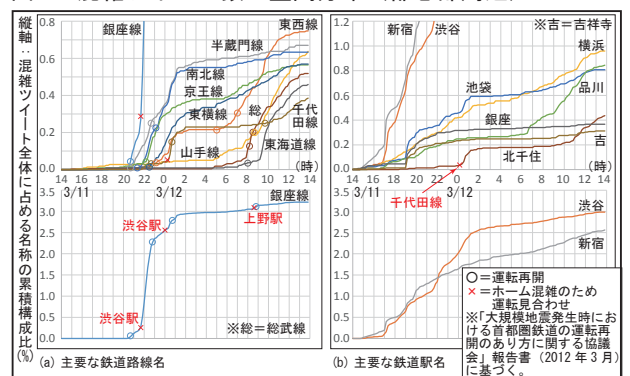


図4 主要な鉄道路線名・駅名の時間推移

注 鉄道駅名・路線名データベースおよび図3は、国土交通省「国土数値情報(鉄道データ)」(平成23年)をもとに筆者が加工した。

謝辞

本研究は、JSPS 科研費 15K18176 の助成を受けて行った研究成果の一部である。なお、本稿で使用している tweet データは、株式会社 NTT データから提供を受けたものである。