

平成 30 年 6 月 19 日現在

機関番号：12608

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K18176

研究課題名(和文)「混雑ツイート」を用いた大地震時の混雑度推計のための言語表現と人間行動の関連

研究課題名(英文) Relationship between linguistic expression and human behavior for estimating crowd congestion degree after a large earthquake using Twitter data

研究代表者

沖 拓弥 (Oki, Takuya)

東京工業大学・環境・社会理工学院・助教

研究者番号：40712766

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：SNSの一つであるTwitterの即時性に着目し、混雑に直面したユーザによって投稿されたテキスト(ツイート=言語情報)や写真(画像情報)を利用して、鉄道駅などの施設や鉄道路線上、道路上の混雑度をリアルタイムに把握することの可能性について検討した。
具体的には、東日本大震災発生前後(2011年3月11日14時00分～12日14時00分)に投稿された日本語ツイートを例に、混雑状況に言及したツイートの抽出方法や、群集混雑の時間推移と空間分布を把握する上での有用性を示すとともに、Webアンケート調査を実施することで、実際の混雑度と、混雑の程度を表す言語表現や行動選択との間の関係を明らかにした。

研究成果の概要(英文)：Using the tweets and photographs posted by users confronted with the crowd congestion, the possibility of grasping the degree of congestion in the facilities (such as railway stations), on the railway lines, or on the roads in real time was discussed.
More specifically, using tweets written in Japanese posted before and after the Great East Japan Earthquake (for 24 hours since 2:00 p.m. on March 11, 2011) as a case study, the method to extract tweets related to crowded conditions and the usefulness of tweets for grasping the time transition and spatial distribution of crowd congestion were demonstrated. Furthermore, conducting a Web questionnaire survey, the relationships between the actual congestion degree and the linguistic expressions about the congestion level and the selected behaviors were clarified.

研究分野：建築計画，都市防災

キーワード：防災・減災 群集混雑度 混雑度評価点 アドレスマッチング Webアンケート Twitter 東日本大震災 時空間分布

1. 研究開始当初の背景

(1) 東日本大震災（東北地方太平洋沖地震）が発生した2011（平成23）年3月11日、東京都心部においては、公共交通機関の運行再開を待つ人々が鉄道駅構内や駅前広場に殺到したほか、通勤・通学先などの外出先から徒歩で帰宅を試みた人々が道路上にあふれ、至る所で重大な事故を引き起こしかねない危険な状況が発生した。

群集によるこうした局所的・突発的な混雑が、どこでどの程度発生しているかを早期に（あるいは事前に）把握し対応することは、群集事故防止の観点から非常に重要な課題であるといえる。

(2) 混雑度の把握を目的とした既存技術には、携帯電話の基地局データを利用する例（モバイル空間統計[株式会社NTTドコモ]）や、スマートフォン等の携帯端末でのアプリ使用時に得られるGPS情報を利用する例（混雑統計[株式会社ゼンリンデータコム]、混雑度レーダー[ヤフー株式会社]等）、あるいは、施設内や街頭に設置された防犯カメラの映像を利用する例（日本電気株式会社、株式会社日立製作所、等）がある。

前者は、広域の人口分布を把握できる長所を有するが、空間解像度（最小でも250m×250mメッシュ単位の集計値）やデータのリアルタイム性（早くても1時間程度のタイムラグが発生）に難点がある。一方、後者は、局所的な混雑状況をリアルタイムに把握できるものの、交通障害時等、広範囲にわたって混雑が発生している際に、多数の観測点から得られる情報を統合・管理し、混雑の全体像を把握することが難しい。

群集事故を防止するための警備・誘導を柔軟に行うためには、時々刻々と変動する施設内・道路上の混雑状況に関する情報を広範囲で収集可能である、新たな技術の開発が必要である。

(3) また、非常時に生じる混雑のメカニズムを分析し、混雑の抑制策について事前に検討しておくことは、将来の首都直下地震や南海トラフ地震に備える上でも重要である。しかし、大地震時にどのような混雑や混乱が発生し、そのとき人々がどのような行動をとるかを定量的に明らかにし、モデル化まで試みた既往研究はほとんど存在しない。これは、（東日本大震災時を含めて）対象事例がわずかであることに加え、実測調査や再現実験も困難であるためである。

したがって、従来とは異なる視点からのアプローチが必要であると考えられる。

2. 研究の目的

(1) 施設や道路の混雑度を、リアルタイムか

つ広範囲で把握する上では、SNS（ソーシャル・ネットワーキング・サービス）の一つであり、即時性・公開性に優れるTwitter（ツイッター）が有用である可能性がある。しかし、Twitterを群集混雑度の推定に用いるための方法について、詳細な検討を行っている研究はこれまでに存在しない。

そこで、多数のユーザによって投稿されたテキスト（ツイート＝言語情報）や写真（画像情報）の中から、混雑に直面し、眼前の混雑状況に関するものをどのように抽出し、大地震時の混雑度推計に対して、Twitter投稿情報がどの程度有用であるかを、実際のデータを用いて具体的に示すことを第一の研究目的とした。

(2) 大地震時にどのような混雑や混乱が発生し、その混雑や混乱が人々にどのような印象を与え、人々の行動にどのように影響するかを分析し、大地震時の混雑や混乱に関する基礎的な知見を得ることを第二の研究目的とした。特に、実際の混雑度（単位面積あたりの人数）と、混雑状況の表現に用いられるワード、および、人々が選択する行動との間の関係を具体的・定量的に示すことや、それらの個人差や地域差、性別や年齢ごとの傾向を明らかにすることを目的とした。

3. 研究の方法

(1) 東日本大震災発生前後（2011年3月11日14時00分～12日14時00分）にTwitterに投稿された日本語ツイートのうち、混雑状況を表現する際に使用されやすいと考えられる「混む（変化語も含む）」「混乱」「混雑」の3語を少なくとも1語含むツイート（以下、混雑ツイート）を抽出した。

(2) (1)で抽出した混雑ツイート数の時間推移を分析した。また、混雑ツイート中に共起する固有名詞・地名（特に鉄道駅名と鉄道路線名）をKH Coderにより抽出した上で、その頻度や時空間的特性を分析した。

(3) (1)で用いた混雑ツイートから、ランダムに1,000ツイートを選択した上で、そこから目視で混雑度表現（混雑状況を表現するために用いられている程度表現）を抽出し、その頻度等を分析した。

(4) Webアンケート調査の準備として、混雑ツイートに含まれるURL（2011年当時のTwitter向け主要画像共有サービスである「Twitpic」のURLに限定）のリンク先の画像全てを、ローカルストレージに保存する。さらに、保存した画像の中から、群集混雑が発生している状況を撮影した写真（以下、混雑画像）100枚（時間帯4区分×25枚）をランダムに抽出した。

(5) どの程度の群集混雑がどこで発生しているかを、SNS (Twitter) 上に投稿されるテキスト情報から推定する上での基礎的知見を得ることを目的に、民間調査会社 (株式会社マクロミル) の協力の下、Web アンケート調査を実施した (表 1)。

まず、被験者 1,236 名 (性別 2 区分×年齢 6 区分×103 名) に対して、個人ごとに混雑画像 4 枚 (各時間帯から 1 枚) を無作為に提示することで、被験者が群集混雑下にいる状況を仮想的に設定し、その状況における混雑度を表現するのに相応しいと考える言語表現を、上述の 52 種類の表現の中から選択してもらった。同時に、各表現から思い浮かべる混雑の程度 (10 段階評価) と、各表現の日常的使用度合い、および、画像で想定した程度の混雑状況下における行動選択についても尋ねた。さらに、被験者の属性 (年齢・性別・職業・居住地域等) や、SNS の利用状況、混雑情報に対する需要、情報の投稿・提供に対する考え方についても調査した。

表 1 アンケートの全調査項目

問1 属性についての質問	
(1) 性別	
(2) 年齢	
(3) 職業	
(4) 自宅の所在地 (町丁目まで)	*1
(5) 通勤・通学先の所在地 (町丁目まで)	*1
(6) 同居家族の人数 (5歳以下/6-12歳/13-64歳/65歳以上)	*1
問2 利用しているSNSについての質問	
利用しているSNS (全て)	*1
(1) 東日本大震災 [2011/3/11 (金)] 当日に利用したSNS	*2
(2) 2011/3/11 (金) 時点でのSNSの利用頻度	
(3) 2011/3/11 (金) 時点でのSNSの利用 (頻度) 頻度	
(4) 2011/3/11 (金) 時点でのSNSの投稿 (書き込み) 頻度	
問3 東日本大震災 [2011/3/11 (金)] 当日の行動についての質問	
(1) 地震発生時 (2011/3/11 (金) 14時46分) の居場所 (町丁目まで)	
(2) 地震発生時 (2011/3/11 (金) 14時46分) の状態	
(3) 地震発生時の居場所から移動を開始した日時	
(4) 自宅へ向かうために移動開始後、自宅に到着した日時	
(5) 自宅へ向かうための移動に利用した交通手段 (全て)	
問4 混雑状況にあることを確認した上での質問 (計4シーン、異なる写真を提示)	
(1) 提示された状況を下のようにSNSに投稿するか	
(2) 提示された状況下での混雑の程度を伝えるのにふさわしいと考える表現	
(3) 大地震発生時、帰宅経路上や目的地において、提示された状況と同程度の混雑の発生が (スマートフォンへの情報配信などで) あらかじめ判明していた場合の行動選択	
問5 混雑の程度を数値として用いられる表現についての質問	
(1) 思い浮かべる混雑の程度 (10段階評価)	
(2) 日常的に使用することのある表現 (全て)	
問6 混雑情報の拡散・提供に関する考え方についての質問	
(1) 混雑状況の把握に役立つとしたら、SNSへの混雑状況の投稿に協力するか (5段階評価)	*3
(2) スマートフォン等で混雑情報が提供されたとしたら、その情報は意思決定 (行動選択など) にどの程度役立つと思うか (5段階評価)	*3
(3) 混雑状況を投稿するとしたら、(携帯電話の基地局との通信データやGPS) 全地球測位システム (GPS) から得られる位置情報を利用するか (3段階評価)	*3
(4) 混雑に関する情報をどのような状態で入手できれば良い (入手したいと思うか) (全て)	

*1) 現在 (回答時点) と東日本大震災時点 (2011年3月11日 [金]) のそれぞれを尋ねた。*2) 閲覧したものと投稿したもののそれぞれを尋ねた。*3) 日常時と災害時のそれぞれを尋ねた。

(6) Web アンケート結果を集計することで、混雑度表現を定量化し、その有用性を検証するとともに、混雑度と人間行動の関連や、混雑度とSNSへの投稿意思の関連等について、基礎的な分析を行った。

4. 研究成果

(1) 混雑ツイート数の時間推移を図 1 に示してある。全体のツイート数は、本震の発生 (11日14時46分) 直後に、5倍以上に急増するものの、その後は次第に減少し、翌12日のツ

イート数はさほど多くないことがわかる。これに対し、混雑ツイートの数には、本震発生当日の夕方 18 時以降と、翌日朝 8 時以降に特に多くなるという特徴が見られた。全体のツイート数に占める混雑ツイートの割合は、11日よりも、むしろ12日の方が大きい。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いる。

以上の結果は、混雑状況の時間推移を把握する上での、混雑ツイートの有用性を示唆している。

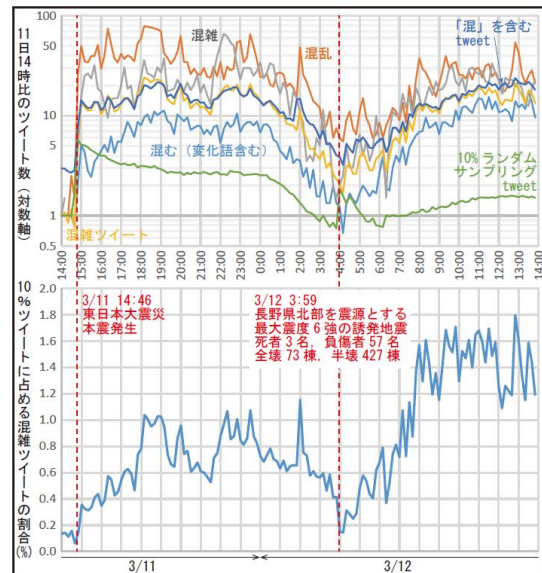


図 1 tweet 数の時間推移 (10 分単位)

(2) 混雑ツイート中に共起する固有名詞・地名の頻度を見たところ (表 2), 「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、位置情報の補充には適さないことが判明した。また、抽出した固有名詞・地名の空間分布を見ると、都心部の地下鉄や山手線、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ことがわかる (図 2)。さらに、混雑ツイート数が増加するタイミングは、各鉄道路線が運転を再開し、多くの人々が鉄道駅に殺到したとされる時刻と良く対応していることが見てとれる (図 3)。

本研究で用いた混雑ツイートの例では、位置情報 (ジオタグ) 付きのツイートは全体の約 0.2% 程度にすぎないため、ジオタグ付きツイートのみでは、混雑ツイートの空間的特徴を把握することは困難と言える。しかし、以上の結果は、ツイートに付属するジオタグ情報に頼ることなく、共起する固有名詞・地名を用いることで、混雑ツイートの時空間分布をある程度把握できる可能性を示唆している。東京都心部においては、公共交通機関の混雑状況を把握する上で、本手法が特に有効と考えられる。

表2 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	1.6	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。

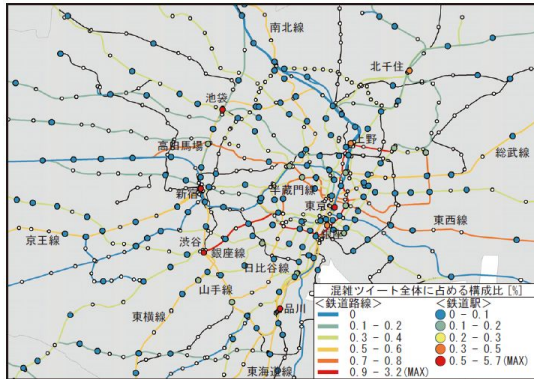


図2 混雑ツイート数の空間分布 (都心部周辺)

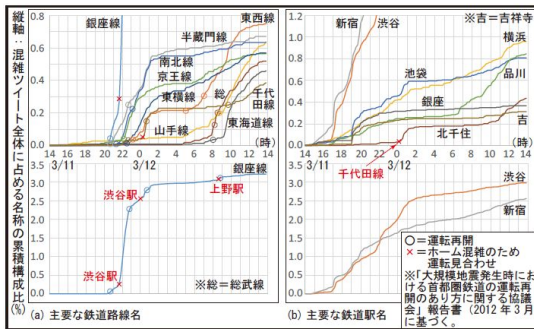


図3 主要な鉄道路線名・駅名の時間推移

(3) 混雑ツイートからランダムに選択した1,000 ツイートのうち、混雑度表現が含まれるツイートの割合は 37.5%であった。また、混雑「している」ことを伝える表現の方が、混雑「していない」ことを伝えたい表現に比べ、種類・登場頻度ともに多かった。

この結果は、現在の混雑がツイート投稿のトリガーとなりやすいことを示している。

(4) 群集混雑の状況の表現に用いられやすい程度表現 52 種類から、人々がそれぞれの

程度の混雑状況を思い浮かべるかを、「混雑度評価点」という指標を用いて定量化することができた(図4)。「混雑している」ことを表す表現と、「混雑していない」ことを表す表現の間には、評価点の平均値や頻度分布に差が見られた。

また、個人ごとに得られる各表現の評価点と、各混雑画像に対する選択表現に基づき、混雑画像100枚の混雑度を定量化したところ、混雑画像から判断される混雑度の高低と、混雑度評価点の順序が良好に対応していることを確認できた。

以上の結果は、混雑度推定におけるテキスト情報の有用性を示している。混雑ツイートの数は、ある時刻・地点における混雑状況に対する人々の関心の高さを示していると考えられ、実際の混雑の程度を把握する上では、ツイート数だけでなく、混雑度表現を利用することが望ましい。

本手法を用いれば、各地点の混雑度を推定するための携帯端末利用者を十分な数確保することが難しい状況下や、建物内部や地下空間などの電波が不安定な場所においても、混雑度を推定可能となることが期待される。

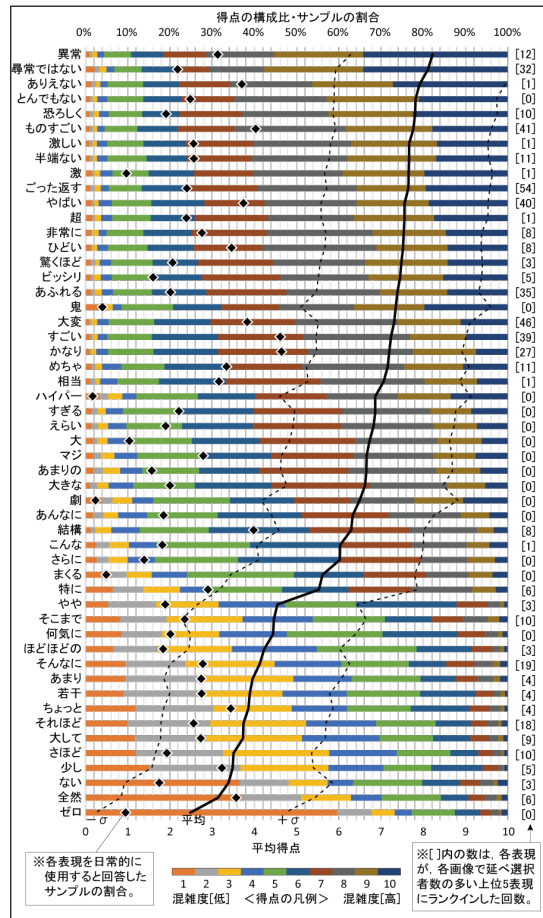


図4 各混雑度表現から思い浮かべる混雑の程度 (N=1,236)

(5) 提示した混雑画像の混雑度評価点が高いほど、「移動を中断し、その場にとどまる」

と回答した人の割合が高い傾向が見られた(図5)。ただし、この傾向は回答者の属性によって異なり、例えば女性は年齢が上がるほど「とどまる」割合が高くなるが、男性は逆の傾向が見られた(図6)。

これらの傾向を、混雑情報(テキストまたは混雑画像)が SNS に投稿された地点における、群集混雑度の推移予測に活用できる可能性がある。

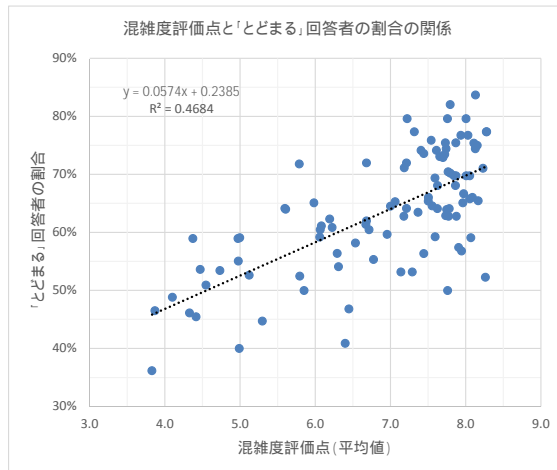


図5 混雑度評価点(平均値)と「とどまる」回答者の割合の関係(N=1,236)

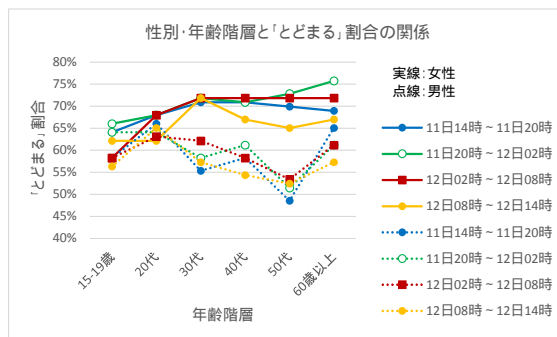


図6 性別・年齢階層と「とどまる」回答者の割合の関係(N=1,236)

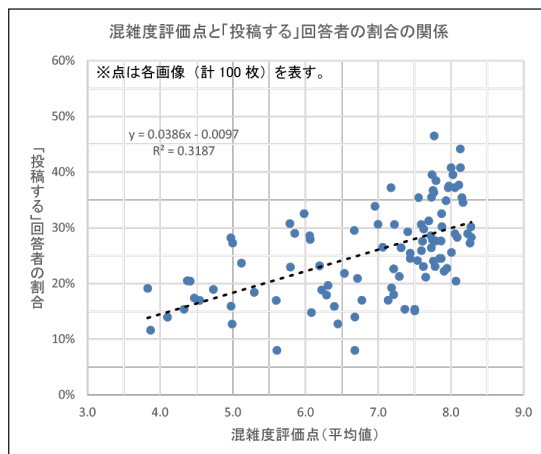


図7 混雑度評価点(平均値)と「投稿する」回答者の割合の関係(N=1,236)

(6) 混雑度評価点の高い混雑画像ほど、「状況を SNS に投稿したい」と回答した人の割合が増加した(図7)。すなわち、非日常的な混雑状況ほど SNS に投稿されやすいことを意味し、群集混雑発生を抽出する上での本手法の有用性を裏付けている。

(7) 研究期間全体を通じて実施した研究の成果を応用することで、Twitter API によりリアルタイムかつ自動的に抽出した「混雑ツイート」から、位置情報または本文中の鉄道駅名・鉄道路線名等の情報を用いて群集混雑発生地点を推定し、さらに、本文中の程度表現から群集混雑度を推定するという、一連の枠組みを構築できた。

現在、様々な状況(イベント時、交通障害時、降雪時等)におけるツイートデータを実際に収集しており、今後、これらのデータを用いて、群集混雑度の推定精度の検証をさらに行っていく予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計1件)

沖 拓弥, SNS 上のテキスト情報に基づく群集混雑度の定量化手法の検討: 東日本大震災時のツイートをを用いたアンケート調査結果の分析, 日本災害情報学会誌「災害情報」, 査読有, 第16巻, 第2号, 2018 (印刷中)

[学会発表](計3件)

Takuya Oki, Possibility of Using Tweets to Detect Crowd Congestion: A Case Study Using Tweets just before/after the Great East Japan Earthquake, Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2018), 査読有, 2018, 584-596, 2018年5月22日, ロチェスター(アメリカ)

沖 拓弥, SNS への投稿情報に含まれる混雑度表現の抽出と定量化手法: 東日本大震災発生前後の tweet を対象としたケーススタディ, 第30回人工知能学会全国大会, 査読無, 1I2-NFC-01-4in2, 2016, 1-3, <https://kaigi.org/jsai/webprogram/2016/pdf/887.pdf>, 2016年6月6日, 北九州国際会議場(福岡県北九州市)

沖 拓弥, 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, 情報処理学会第78回全国大会講演論文集, 査読無, 第2号, 2016, 5-6, id.nii.ac.jp/1001/0/0162502/, 2016年3月10日, 慶應義塾大学(神奈川県横浜市)

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

ホームページ等

T2R2 (Tokyo Tech Research Repository)

http://t2r2.star.titech.ac.jp/cgi-bin/researcherinfo.cgi?q_researcher_content_number=CTT100518739

6. 研究組織

(1) 研究代表者

沖 拓弥 (OKI, Takuya)

東京工業大学・環境・社会理工学院・助教
研究者番号：40712766

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし

(4) 研究協力者

該当なし