

大規模災害時の情報提供を目的とした地名の曖昧性解消

六瀬 聡宏¹, 内田 理², 富田 誠³, 梶田 佳孝⁴, 山本 義郎⁵, 鳥海 不二夫⁶

¹ 東海大学大学院工学研究科情報理工学専攻
3bdrm018@mail.tokai-u.jp

² 東海大学情報理工学部情報科学科
o-uchida@tokai.ac.jp

³ 東海大学教養学部芸術学科デザイン学課程
study@tomita.me

⁴ 東海大学工学部土木学科
yokaji@tokai-u.jp

⁵ 東海大学理学部数学科
yama@tokai-u.jp

⁶ 東京大学大学院工学研究科システム創成学専攻
tori@sys.t.u-tokyo.ac.jp

1. はじめに

大規模災害が発生した際、被害を最小限に食い止めるには、災害発生後の迅速かつ確かな情報収集・伝達が重要である。例えば、東日本大震災が発生した際には、速報性の高い情報の受発信が行われるという特徴を有するTwitterが多数の被災者に利用されたことが判明しており[1]-[3]、Twitter利用者の約8割が情報収集に役立ったとの調査結果も報告されている[4]。また、災害発生時にTwitterを利活用する試みも既に多くの事例が知られている[5]-[7]。例えば、Micro Mappers[5]というプロジェクトでは、視覚的にわかりやすい災害情報地図を作成する手段の一つとしてTwitterが利用された。

しかし、Twitterのタイムラインを通して流通する情報は膨大であり、また重複する情報やノイズが多数存在しているため、被災者が自分の状況の適した情報を的確、かつ簡便に得ることは容易ではない。例えば、東日本大震災が発生した2011年3月11日には、約3,300万件のツイートが投稿された事が判明している[8]。

そのような背景から、我々は大規模災害時にTwitterから情報を収集・整理し、ユーザの属性や状況に応じて適切な情報を提供するシステムの構築を目指している[9]-[11]。現在は、ツイートを情報提供に適したカテゴリに分類し、地図上にマッピングするシステムを構築中である。本稿では、災害関連ツイートをマッピングする際に問題となる地名の曖昧性解消について検討した結果を報告する。

2. 関連研究

大規模災害時にツイートデータから地理情報を特定し、地図上に付与する試みは、これまでにも多数行われている。最も単純な方法としては、ツイートの付与されたジオタグ(経緯度)に基づく位置の特

定が考えられる。しかし、ジオタグが付与されているツイートは極めて少なく、橋本ら[12]の研究によると約0.18%であった。一方、Kitamoto[13]によれば、ツイート中に地名が含まれる割合は全体の12%であり、ジオタグが付与されたツイートよりも割合が多い。そのため、ツイート中の地名を抽出し、その位置を適切に特定することができれば、情報提供に有益であると考えられる。原らはツイートに含まれる地名(字・町名)を用いた有益な話題抽出と、地図上への展開を試みている[14]。また、災害時においてツイートが発信された位置を特定するために、ユーザプロフィールやユーザ間のリンク関係を利用しようとする試みもある[15][16]。

3. 災害関連ツイートの曖昧性解消

3.1. 東日本大震災時のツイート分類

我々の研究グループは、2011年3月7日0時から3月23日24時までの期間に収集された約4億の日本語ツイートを保有している。これは、TwitterAPIの制限などから、該当期間の日本語ツイートの8割程度を収集したものである。

我々は先行研究で、東日本大震災発生直後のツイート(2011年3月11日14時46分49秒~同日23時59分59秒)を対象として、複数の機械学習アルゴリズムを用いて「津波に関する情報」「避難(避難所・避難施設)に関する情報」「ライフライン(電気・ガス・水道など)に関する情報」「交通機関(鉄道・バス・飛行機)の運行状況に関する情報」「道路状況に関する情報」の計5カテゴリへの分類を試みた[17]。本研究では、これら各カテゴリ分類されたツイートの中で地名・施設名が一意に特定できない場合の曖昧性の解消を試みる

3.2. データセット

[17]で利用したデータセットの作成には以下の手順を踏んだ。まず、災害状況をツイートする際に多く用いられると予想されるキーワード（「地震」「津波」「避難」「停電」「運休」「通行止」など）によるフィルタリングでツイートを絞り込んだ。また、情報提供に有益なツイートであることを前提としているため、地震関連のハッシュタグ（#jisin, #jishin, #jishin_jp, #earthquake, #eqjp, #311jisin, #saigai のいずれか）を含む 50 文字以上のツイートのみを抽出し、先頭が RT, もしくは QT で始まるものは削除した。さらに、地図上にツイートをマッピングすることを目的としているため、国土数値情報ダウンロードサービス[18]を利用して作成した地名・施設名リスト上に存在する名称を含むツイートのみを抽出した。最後に、重複ツイートや不適切なツイートを手動で削除して残されたツイートをデータセットとして利用することとした。

このデータセットに対する正解カテゴリのラベルは著者が手動で付与した。なお、一つのツイートに対し複数のラベルを付与することも可能としている。

3.3. 地名・施設名データベースの作成

都道府県名と市区町村名、鉄道駅名、避難所や公共施設、学校の名称と、それぞれの経緯度情報からなる地名・施設名データベースを作成する。

都道府県名、および市区町村名は GeoNLP[19]が提供する地名辞書を利用した。都府県名、および市区町村名は末尾の「都・府・県・市・区・町・村」を含む表記と含まない表記の両方をデータベースに登録した（これは、東日本大震災時のツイートを目視で確認したところ、末尾の「都・府・県・市・区・町・村」を省略して記述しているケースが多数見られたことによる）。駅名は、駅データ.jp[20]のデータを利用し、自治体名同様、末尾の「駅」を除いた表記も辞書に追加した。避難所と公共施設、学校の名称は、国土数値情報サービスは[18]を利用した。Twitter では投稿する際、140 文字の制限があるため、地域における避難所の情報をツイートする際、学校名を略称で表記（例えば、「港小学校」のことを「港小」と表記）するツイートが多数見られた。これらの表記にも対応するために、小学校に関しては「○○小」、中学校に関しては「xx中」といった省略形の表記に置換した名称も辞書に登録することとした。

3.4. ツイート中の地名候補の取得

ツイート中の各地名の位置候補を、作成した地名・施設名データベースから取得する。ツイート中の地名・施設名は、地名・施設名データベースに登録された名称との単純マッチングで抽出する。なお、単純マッチングでは部分一致した文字列も候補として抽出されてしまう。そこで本研究では、候補として取得された文字列で最長一致するもののみを候補

として扱い、それ以外を候補から除外することとした。

3.5. 共起する地名の距離に基づく曖昧性解消

栗村ら[21]や落合ら[22]は、ソーシャルメディアに特有の特徴を利用して、Twitter を対象にした地名の曖昧性解消を試みている。

本研究ではツイートに 2 つ以上の地名が含まれていることを前提とし、平野ら[23]の手法を参考に、複数の地点候補をもつ曖昧な地名・施設名と共起する別の地名・施設名との距離を算出し、その距離が最短になる候補をツイートが指し示す地点であると決定する。これは、同一ツイート内で共起する地名・施設名は、距離が近いものであるという仮定にもとづいている。

地名・施設名データベースをもとに、対象ツイートに含まれる i 番目の地名・施設名を W_i ($1 \leq i \leq n$) と表記する (n は対象ツイートに含まれる地名・施設名の数)。また、 W_i に対する地点の候補数を $m(W_i)$ 、 W_i に対する j 番目の地点候補を $P_{i,j}$ ($1 \leq j \leq m(W_i)$) と表記する。もし W_i が曖昧性を有する、すなわち $m(W_i) > 1$ の場合、同一ツイートに含まれる他の全ての地名・施設名 W_k ($k \neq i$) の全ての地点候補 $P_{k,j}$ との距離を算出し、距離が最短となった $P_{i,j}$ を W_i の地点と決定する。下記のツイートを例に処理の手順を示す。

“東京駅近郊の被害状況。火事が多数発生。火の元、ガスは閉じましょう。横浜の商業施設で天井落下。水道の水漏れやエレベーターの停止などが起きている模様です”

ツイート中に含まれる「横浜」という地名は神奈川県や福井県、高知県、福岡県に存在するため、地図上にマッピングする地点の候補を複数持つ。そこで、共起している「東京」と「神奈川県の横浜」、「東京」と「福井県の横浜」といったように全ての候補に対して地理的な距離を算出する。この例では、「東京」との距離が最短となった神奈川県の「横浜」を、ツイートが指し示す地点と決定する。

4. 評価実験

4.1. 実験方法

地名の曖昧性解消の有用性を検証するため、東日本大震災が発生した 3 月 11 日のツイートを用いて実験を行う。本実験では、3.2.で述べたデータセットを利用し、「津波に関する情報」「避難（避難所・避難施設）に関する情報」「ライフライン（電気・ガス・水道など）に関する情報」「交通機関（鉄道・バス・飛行機）の運行状況に関する情報」「道路状況に関する情報」の各カテゴリごとに評価を行うこととした。今回は、各カテゴリごとに評価実験用のツイートを 30 件ランダムに抽出し、地名の曖昧性を

解消できるか否かを評価する。今回実験に使用したツイートの一部を表 1 に示す。

4.2. 実験結果と考察

実験結果を表 2 に示す。すべてのカテゴリに対して高い精度で曖昧性が解消できることを示しており、地名の曖昧性解消の有用性が検証できたと考えられる。実験結果の一部を表 3 に示す。

ツイート本文が比較的短く、含まれている地名が少ない場合には、多くのケースで曖昧性解消に成功した。また、地名を多く含むツイートの場合でも、避難所のように地名・施設名がひとつの都道府県や市区町村に限定されている場合には、各地点間の距離が短い場合最短の候補を選択する手法が有効に機能したと考えられる。

失敗したツイートに関しては、各カテゴリの特徴が影響していることがわかった。例えば、公共交通機関のツイートの場合は、情報を拡散する意図からひとつのツイートに多くの地名・施設名が含まれている場合が多く、そのような場合には距離的に離れた地名・施設名が共起し、提案手法が有効ではなくなる(表 3 の失敗ツイートの前者)。また、津波カテゴリで失敗したツイートも同様であり、例えば「熊野」という地名の曖昧性を解消しようとしたが、和歌山県に存在する熊野という地名に影響され、本来であれば「種子島の熊野」と判定されるべきものが「和歌山県の熊野」と判定されてしまった(表 3 の失敗ツイートの後者)。これらを踏まえると、地理的距離の他に、ツイートを投稿したユーザのプロフィールやフォロー・フォロワー関係を考慮するといったユーザの特徴に注目した手法を併用する必要があると考えられる。

5. まとめと今後の課題

本研究では、大規模災害時の情報提供を目的として、ツイート中に出現する地名・施設名の地点特定の曖昧性解消について検討した。本研究では少数のツイートをデータセットとして利用したが、今後はより大規模なデータセットを使用して精度を検証する必要がある。

なお、本研究では東日本大震災時のツイートを用いて地名の曖昧性解消を試みたが、台風や大雪、ゲリラ豪雨など、災害の種類や規模によってツイート内容が大きく異なることが予想され、本研究と同様の手法が地名の曖昧性解消に有効であるかを検証する必要がある。また、曖昧性が解消されたツイートを地図上にマッピングして被災者に提示するシステムを実装し、その有用性を検証する予定である。

謝辞

本研究の実施にあたり、ツイートデータの収集に

協力していただいた、クックパッド株式会社の兼山元太氏に感謝する。

なお、本研究は、平成 25 年度文部科学省「地(知)の拠点整備事業」補助金の助成を受けて実施した。

表 1 曖昧なツイート例

津波	<ul style="list-style-type: none"> ・<列車の運転見合わせについて>地震での津波の影響により、日豊本線杵築駅から宮崎駅間、国分駅から鹿児島中央駅、指宿枕崎線鹿児島中央駅から枕崎駅間、日南線南宮崎駅から志布志駅間列車の運転を見合わせしています ・四国・大分・宮崎・鹿児島沿岸部の方々は気をつけて下さいね
避難	<ul style="list-style-type: none"> ・帰宅困難な方向けに開放している施設▼【千代田区】厚生労働省合同庁舎 5 号館講堂(震が関)、文部科学省講堂(同)、農林水産省 7 階共用第 6 会議室・秘書課研修室(同)、昭和館(九段南)、気象庁講堂(大手町)▼【中央区】海上保安庁海洋情報部 1 階会議室(築地) ・北区滝野川第七小学校開放してました。田端と駒込の間を歩いてる方近いです
ライフライン	<ul style="list-style-type: none"> ・NHK 発：東京電力 相模原、川崎でも停電している地域がある模様。復旧のめどたたず ・NHK 発：【仙台市青葉区上杉山週小】避難所として、毛布が配られているが、停電のため真っ暗。館内ではラジオが流されています。すでにいっぱい中に入れない人がいる模様。 ・川崎は中原区は停電してませんが宮前区の方は停電していて信号なども消えているみたいです。
交通機関	<ul style="list-style-type: none"> ・【TBS ニュース 18:35 #jisin】震度 7 栗原市 仙台空港津波で冠水 女川町で住宅街が水没 町田市ディスカウントストアの駐車場崩壊 横浜市内建物の壁崩壊多し 現在の死者 29 人不明 37 人 ・新潟はまだ揺れています。ゆっくりとしているが、大きな横揺れ。JR 仙石線・松島の野蒜駅付近走行中の列車と連絡取れず。宮城では 10m を超える大津波が押し寄せている。安否が気遣われる。
道路	<ul style="list-style-type: none"> ・#jishin 【拡散希望】新宿周辺道路、青梅街道、大久保通り、小滝橋通りはどこも渋滞してまうす。出来る限り車両移動は控えてください。緊急車両が通行出来ません。また歩行者は信号無視したり道いっぱい広がって歩かない様に！ ・【長野県内の交通の影響】JR はほぼ全てで運転見合わせ、長野電鉄では徐行運転中、高速道路では佐久～群馬県藤岡まで通行止め

表 2 実験結果

カテゴリ	精度 [%]
津波	90.0
避難	93.3
ライフライン	93.3
交通機関	83.0
道路	86.7

表3 実験結果の一部

成功	<p>・ありがとうございます。タクシーも諦めるかなあ。 QT @FKU: 山手通り外回り, 初台付近, 歩く速度と大差ない渋滞 (道路カテゴリ)</p> <p>・【拡散希望】福島県相馬市の避難場所。中村第一小学校 中村字大手先 1 ,中村第二小学校 尾浜字細田 1 ,桜丘小学校 中村字桜ヶ丘 179 ,飯豊小学校 大曲字天神前 42 ,磯部小学校 磯部字上ノ台 467-2 ,日立木小学校 日下石字神明前 14 ,大野小学校 大坪字東畑 7#saigai (避難所カテゴリ)</p>
失敗	<p>・近鉄バスサイトより (追加情報) 高速バス情報 大阪 (一部京都) 発仙台, 山形, 福島, いわき, 水戸, 宇都宮, 新宿 (格安便), 横浜・東京, 熊谷, 小田原行。大阪行は仙台, 福島, 山形, いわき, 宇都宮, 新宿, 東京, 熊谷, 小田原, 富士吉田発が運休。(交通機関カテゴリ)</p> <p>・【津波観測情報】22時19分北海道浜中町 2m60cm 22時44分和歌山串本町 1m20cm 23時2分種子島熊野 1m30cm (津波カテゴリ)</p>

参考文献

- [1] 風間一洋, “Twitter における情報伝播”, 人工知能学会誌, Vol.27, No.1, pp.35-42, 2012.
- [2] 鳥海不二夫, 篠田孝祐, 栗原聡, 榊剛史, 風間一洋, 野田五十樹, “震災がもたらしたソーシャルメディアの変化”, JWEIN11, pp.41-46, 2011.
- [3] H. Wilensky, “Twitter as a Navigator for Stranded Commuters during the Great East Japan Earthquake”, Proc. of 11th International Conf. on Information Systems for Crisis Response and Management, pp.695-704, 2014
- [4] 株式会社 IMJ モバイル, “東北地方太平洋沖地震に伴う twitter, facebook 利用実態に関する調査”, http://www.imjp.co.jp/press/release/20110404_000581.html
- [5] Micro Mappers, <http://micromappers.com/>
- [6] 相田慎, 新堂安孝, 内山将夫, “「東日本大震災関連の救助要請情報抽出サイト」による救助活動支援”, 自然言語処理, Vol.20, No.3, pp.405-422, 2013.
- [7] 後藤淳, 大竹清敬, Stijn De Saeger, 橋本力, Julien Kloetzer, 川田拓也, 鳥澤健太郎, “質問応答に基づく対災害情報分析システム”, 自然言語処理, Vol.20, No.3, pp.367-404, 2013.
- [8] NEC ビッグロブ株式会社, “東日本大震災におけるツイッターの利用状況について”, <http://tr.twipple.jp/info/bunseki/20110427.html>
- [9] 六瀬聡宏, 長島俊, 内田理, 鳥海不二夫, “Twitter を用いた大規模災害時における情報提供システム”, 第 12 回情報科学技術フォーラム, O-055, 2013.
- [10] 高畑洋貴, 六瀬聡宏, 榎本光, 齊藤大樹, 近藤直人, 富田誠, 梶田佳孝, 山本義郎, 鳥海不二夫, 内田理, “大規模災害時における避難支援情報の可視化”, 言語処理学会第 20 回年次大会発表論文集, pp.82-84, 2014.
- [11] 馴田俊平, 六瀬聡宏, 榎本光, 齊藤大樹, 近藤直人, 富田誠, 梶田佳孝, 山本義郎, 鳥海不二夫, 内田理, “エリア限定型大規模災害時情報提供システム”, 言語処理学会第 20 回年次大会発表論文集, pp.67-69, 2014.
- [12] 橋本弘, 岡瑞起, “都市におけるジオタグ付きツイートの統計”, 人工知能学会誌”, Vol.27, No.4, pp.424-431, 2012.
- [13] A. Kitamoto, “Toponym-based Geotagging and Disambiguation for Social Media on Earthquake and Weather Events”, 10th International Conf. on Information Systems for Crisis Response and Management, 2013.
- [14] 原久美子, 木野泰伸, 鳥海不二夫, “字・町名をキーとした災害時 Twitter 情報の抽出と地図への展開”, JSAI2013, 2B4-NFC-02a-2, 2013.
- [15] 榊剛史, 原久美子, 吉田光男, 鳥海不二夫, 篠田孝祐, 栗原聡, 風間一洋, 野田五十樹, “災害情報基盤構築に向けたテキストデータからの地理情報抽出システム”, JSAI2014, 1H2-NFC-02a-4, 2014.
- [16] 榊剛史, 原久美子, 吉田光男, 鳥海不二夫, 篠田孝祐, 栗原聡, 風間一洋, 野田五十樹, “災害支援における Twitter 研究の可能性についてー既存研究のサーベイを通してー”, DRIS 第 2 回合同研究会, 2014.
- [17] 六瀬聡宏, 長島俊, 内田理, 鳥海不二夫, “Twitter を用いた大規模災害時における情報提供システム”, 第 13 回情報科学技術フォーラム, O-010, 2014.
- [18] 国土数値情報ダウンロードサービス, <http://nlftp.mlit.go.jp/ksj/>
- [19] GeoNLP <https://geonlp.ex.nii.ac.jp/>
- [20] 駅データ.jp <http://www.ekidata.jp/>
- [21] 落合桂一, 鳥居大祐, “時間変化する特徴語によるマイクロブログ地名曖昧性解消”, 情報処理学会論文誌, データベース, Vol.7, No.2, pp.51-60, 2014.
- [22] 栗村誉, 荒牧英治, 河原大輔, 柴田知秀, 黒橋禎夫, “ソーシャルメディアにおける空間的接近性と時間的一貫性を考慮した地名の曖昧性解消”, 情報処理学会研究報告, 自然言語処理研究会報告, 2014.
- [23] 平野徹, 松尾義博, 菊井玄一郎, “地理的距離と有名度をを用いた地名の曖昧性解消”, 情報処理学会第 70 回全国大会講演論文集, No. 2, pp. 85-86, 2008.