

東北環境再生モニタリングマップにおける 震災時ツイート提示機能の実装

加藤勝何^{†1} 内田理^{†1} 寺田一美^{†1} 福江潔也^{†1} 長幸平^{†1} 佐藤俊明^{†2}

概要: 2011年3月11日に発生した東日本大震災は、東北地方一帯に甚大な被害をもたらした。我々は“衛星観測と現地調査による被災地の環境再生モニタリングと地球環境教育の実践プロジェクト”に取り組み、東日本大震災で津波被害が甚大であった宮城県沿岸域の復興状況調査を行っている。現地写真や衛星データ、土壌塩分濃度などのデータの収集を長期的に続けており、先行研究ではそれを整理した形で Web マップに提示するシステム TERM MAP (Tohoku Environment Recovery Monitoring Map: 東北環境再生モニタリングマップ) を構築した。しかし、東日本大震災発生当時の状況や状態を示すデータが不足していたため、本研究では震災発生後に投稿されたツイートデータに着目することとした。これまでも災害時にどのようなツイートが投稿されたかを分析する研究は多数行われてきたが、東日本大震災当時の被災者や被災地の状況を現在の復興状況と共にわかりやすく地図上に提示する試みは見受けられない。本研究では、先行研究で構築した TERM MAP 上に東日本大震災発生後二週間のツイートを提示する機能を実装した。これにより、震災当時の被害状況と現在の復興状況を関連付けて確認することが可能となった。

キーワード: 東日本大震災, Twitter, アーカイブ, Web マップ

Implementation of a Function to Present Tweets Posted at the time of the Disaster on Tohoku Environment Recovery Monitoring Map

SHOKA KATO^{†1} OSAMU UCHIDA^{†1} KAZUMI TERADA^{†1}
KIYONARI FUKUE^{†1} KOHEI CHO^{†1} TOSHIAKI SATO^{†2}

Abstract: The Great East Japan Earthquake that occurred on March 11, 2011 brought serious damage in the Tohoku district in Japan. Our research group is working on a project such as “A project on the recovery monitoring of the destructed environment in the disaster area by satellite data analysis and field work and the practice of earth environmental education” and investigating the progress of recovery of coastal area of Miyagi prefecture where the damage of the tsunami was enormous by the earthquake. We are continuing to collect data such as on-site photos, satellite data, soil salt concentration etc. for a long time, and we constructed TERM MAP (Tohoku Environment Recovery Monitoring Map), a system that presents the survey data on a map in an organized form. However, there was insufficient data showing the situation at the time of the Great East Japan Earthquake. Therefore, we pay attention to the tweet data posted at that time. Though several studies have been conducted to analyze what kind of tweets were posted in the time of a disaster, no attempt has been made to extract by tweets and present the situation of the victims and affected areas at the time of the Great East Japan Earthquake on the map with the progress of recovery in an easy-to-understand manner. Then, in this study, we implement a function to display tweet data for two weeks after the Great East Japan Earthquake on TERM MAP. This function makes it possible to confirm the damage situation at the time of the disaster and the current recovery situation.

Keywords: Great East Japan Earthquake, Twitter, archive, Web Map

1. はじめに

我々は“衛星観測と現地調査による被災地の環境再生モニタリングと地球環境教育の実践プロジェクト” [1]に取り組み、東日本大震災で津波被害が甚大であった宮城県沿岸域の復興状況調査を行っている。現地写真や衛星データ、塩分濃度などのデータの収集を長期的に続けており、先行研究ではそれを整理した形で Web マップに提示するシステム TERM MAP (Tohoku Environment Recovery Monitoring Map: 東北環境再生モニタリングマップ) を構築した [2]。しかし、東日本大震災発生当時の状況や状態を示すデータ

が不足していたため、本研究では震災発生後に投稿されたツイートデータに着目することとした。これまでも災害時にどのようなツイートが投稿されたかを分析する研究 [3]-[13]は多数行われてきたが、東日本大震災当時の被災者や被災地の状況を現在の復興状況と共にわかりやすく地図上に提示する試みは見受けられない。そこで本研究では、先行研究で構築した TERM MAP 上に東日本大震災発生後二週間のツイートを提示する機能を実装した。

2. プロジェクトの概要

本プロジェクト [1] では、現地調査による被災地のデータ

^{†1} 東海大学
Tokai University

^{†2} (株)パスコ
PASCO CORPORATION

収集を行っている。現地調査は2012年の10月より開始し、2017年までの5年間実施する計画であり、これまでに8回の調査を行った。調査は、津波被害の甚大であった石巻市、女川町、気仙沼、仙台市、名取市、東松島市、南三陸町の7つの地域を対象としている。調査項目は写真撮影、衛星画像解析、土壌の塩分濃度測定、養殖場の水質調査の4項目である。また、調査データの発信は Web サイト”TERM MAP”[2]を構築することによって実現する。

3. 震災アーカイブ

本研究は、震災アーカイブの構築 [14]-[21]と捉えることもできる。関連する研究・既存サービスとしては、東北大学が主導する“みちのく震録伝” [14]-[16]や、国立国会図書館による“ひなぎく” [17][18]、Google社の“未来へのキオク” [20]などが挙げられる。これらは、東日本大震災に関するデータを多く蓄積しているが、震災当時投稿されたツイート群から被災者や被災地の状況や状態を示すものを抽出し、現在の復興状況と共にわかりやすく地図上に提示する試みは見受けられない。

4. TERM MAP

Webサイトの表示例を図1に示す。WebサイトはGoogle Maps APIを用いて、Webマップベースで構築する。マップ上へ調査データを表示することにより、地理的な位置とデータとの関係の把握を容易にする。また、本Webサイトでは外部データの取り込みも行っている。本稿では、いくつかのデータ提示例を紹介する。

4.1 写真提示

本プロジェクトでは、現地調査により半年に一度の頻度で定点撮影を行っている。これまでの調査によって収集した写真の総数が5000枚を超え、人手での整理が困難となったため、これらを自動で整理しWebマップを用いて提示するシステムの構築を行った[2]。写真の提示例を図2に示す。

4.2 塩分濃度可視化

現地調査によって測定した塩分濃度をWebマップとHighcharts[14]を用いて可視化した。マップ上にプロットされるピンを選択することによって、マップ下部にグラフが表示される。塩分濃度グラフの提示例を図3に示す。

4.3 外部データ表示

本研究では、国土数値情報[22]と国土交通省都市局『復興支援調査アーカイブ』[23]により提供されているデータを可視化している。可視化しているデータは、海岸線データ、河川データ、農業地域データ、津波浸水区域データの4つである。図4に海岸線データと津波浸水区域データを可視化した例を示す。

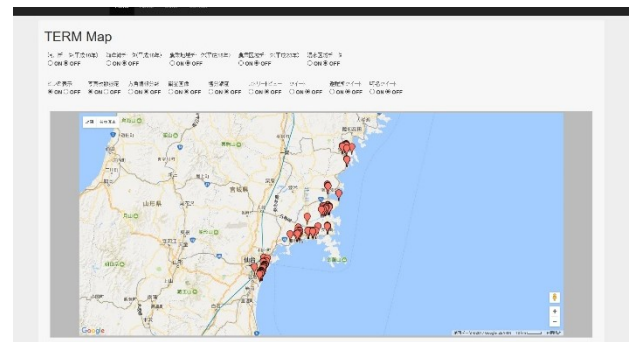


図1 TERM MAP 表示例

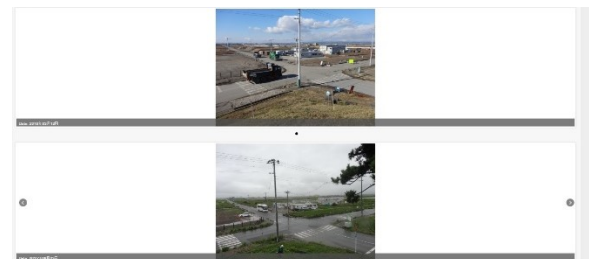


図2 写真の提示例

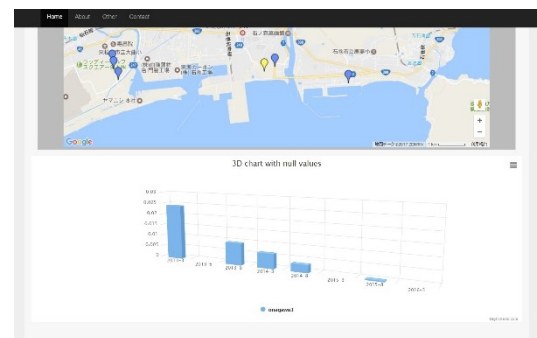


図3 塩分濃度の可視化例



図4 外部データの提示例

5. ツイート提示機能の実装

我々の研究グループは、2011年3月7日0時から3月24日24時までの期間に収集した日本語の約4億ツイート(TwitterAPIの制限により全体の8割り程度)を保有している。これらのツイートのうち、3月11日0時0分から

ら3月24日23時59分までツイート群から、被災地の状況や状態のわかるものを表示対象として抽出する。

5.1 ツイート抽出

まず、繰り返し同じ情報が出現し、ノイズとなることを防ぐため、処理対象の14日間の全てのツイートからリツイートを除く。次に、調査対象の地域である「石巻」「女川」「気仙沼」「名取」「東松島」「南三陸」をキーワードとして、これらの単語が含まれるツイートを抽出する。このとき、仙台に関しては津波被害のあった地域以外のツイートが圧倒的に多く見受けられたため、抽出対象から除外した。また、このキーワード抽出とは別に、リツイートを除外したデータから避難所名と大字名をキーワードとしてのツイートの抽出を行った。GeoNLP[25]から提供されている避難所名情報と大字名情報をもとに、本研究で扱う地域に関する辞書を作成する。このとき、「女川小学校」といった避難所名は、省略形が多く使われることを想定して「女川小」へ変換を行った。本研究の調査対象地域である7つの市町内に存在する避難所と大字名を抽出したところ避難所の辞書は829件、大字の辞書は、1227件となった。これを用いて対象地域に存在する避難所と大字に関するツイートを抽出する。この時点で、市町名、避難所名、大字名のキーワードを使ってツイートの抽出を行ったため、ツイートのデータセットが3つとなる。

5.2 キーワードによるフィルタリング

3つのツイート群に対して、キーワードを用いてフィルタリングを行う。この処理は、表示非対象のツイートに含まれる可能性の高い単語を含むツイートを除外する。これにはフィルタリング対象の単語を用いて処理を行う。フィルタリング対象として扱うためのキーワードとなる単語を表1に示す。これらの単語を1つでも含んでいるツイートは、この時点で除外する。

5.3 表示対象ツイートの抽出

フィルタリングをした時点でも、多くの表示非対象ツイートが残っており、3つのデータを合わせたツイート数の合計は約50万程度となる。そこで、一部のツイートを対象に表示対象ツイートと表示非対象ツイートを人手で分類する。分類したツイートの例を表2に示す。これらのツイートを学習データとして、機械学習を行って自動分類することにより、自動的に表示対象ツイートを抽出する。分類に用いる素性はMeCabにより抽出した名詞とした。このとき、代名詞、数詞、アルファベットのみ名詞は除外した。ただし、漢数字や全角記号などの一部の例外は取り切れていないものがある。分類器の生成に使用した素性数は1570であり、分類器はSVMを用いることとした。また、正常に分類されるか確認するために5分割交差検定を行った。入力データ数を表3、結果の平均値を表4に示す。表示対象ツイートのみを高精度で抽出することはできなかったが、

表示非対象のツイートを多く除去できることを確認した。3つのツイート群から抽出したツイート数を図5に示す。

5.4 地理情報付与

GeoNLPの避難所辞書・大字辞書には、避難所名と大字名の他に座標情報も含んでいる。そのため、ツイート1件ごとの本文に対して、避難所ツイートでは避難所名、大字ツイートには大字名で総当りを行い、マッチングした場合は、該当する座標情報をツイートに対して付与する。また、1つのツイート中で複数マッチングする場合は、その1つのツイートに対して、複数の座標情報が付与されるように処理を行った。そのため、1つのツイートが複数の箇所にマッピングされる可能性がある。

表1 フィルタリングキーワード

原発, 原子力, 放射, 福島, 東北電力, 東電, 東京電力, 線量, 安否, 母, 父, 弟, 兄, 妹, 姉, 家族, 親, 友, さん, 歳, 姪, 消息, 探し, 捜し, 枝野, 募金, ユニセフ, bot, BOT, 心配, 冥福, 大丈夫, 祈る, 実家, 心配, moi, casting, PersonFinder, 同級生
--

表2 表示対象・非対称のツイートの例

(a) 表示対象

女川総合体育館。やっとなソフトバンクきたー http://twitpic.com/4cm1lq
お返事遅くなりました。女川のいところからは、節電のため限られた時間しか電話ができず、です。電気は昨日復旧、仙台-女川間はガソリンさえあれば一般車でも行き来できるとのこと。しかし、町は「まるでターミネーターの映画のセットのよう」とのことです。
女川町総合体育館（宮城・女川町）／寒空の洗濯お湯が好評 http://dlvr.it/LGB3t

(b) 表示非対象

女川第一小学校の教頭が武藤にしか見えない
僕あ1年半前の2009年に女川を訪れたことがあります。主たる目的がJR石巻線の乗潰しにあったので、女川滞在時間は1時間あったかどうかなのですが、女川駅前から港にかけての街並みは観てるのですよ。 http://bit.ly/fwcQpD
今ZEROでインタビューされてた先生 俺の中学の時の先生だわ… 女川に転勤してたんだな、てか生きてたんだな…

表 3 学習データ数

表示対象ツイート	表示非対象ツイート
94	151

表 4 SVM の平均値

	適合率	再現率	F 値
平均値	0.61	0.41	0.48

6. ツイート提示

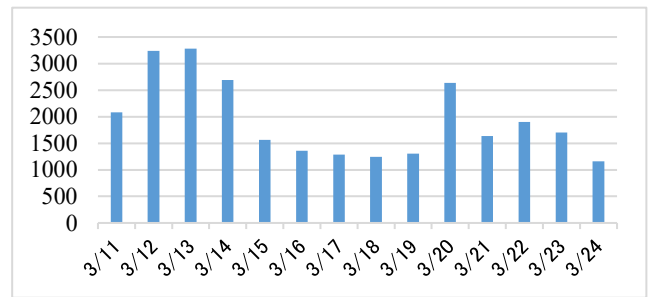
市町名によって抽出したツイートは、範囲が大きいためポリゴンレイヤーによって、6 の市町に範囲を重畳表示する。避難所名によって抽出したツイートと大字名によって抽出したツイートは、ピンをマップ上へプロットする。ツイートマッピングの例を図 6 に示す。いずれかを選択することによって、図 7 に示すようにツイートが表示される。このとき、日付の選択はマップ上部に設置したラジオボタンによって行う。

震災当時のツイート提示機能の実装により(若干のノイズが混じってはいるものの)当時の各地の状況を読み取ることができるようになった。例えば図 7(a)の例では、2011 年 3 月 15 日の昼の時点では、七郷小学校付近で毛布や水などの物資が足りていないが、同日 17 時には炊き出しが行われていることがわかる。大字名によるツイートの抽出は、図 7(b)のように地名の曖昧性の影響で、他の地域や人名に関するものが間違えて提示されてしまう問題点が残された。また、ニュースの引用といった同じ内容のツイートが多数投稿され、これらが同時に提示されてしまうケースがあった。そこで、同内容のツイートの集約などを検討する必要がある。

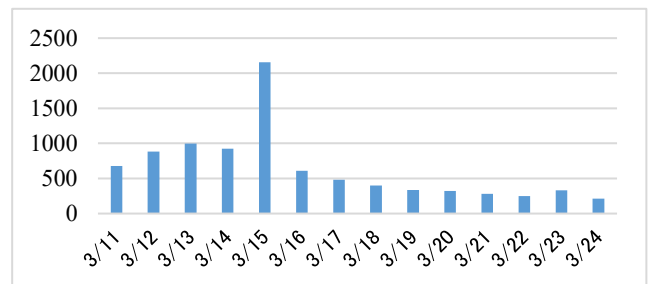
7. まとめ

本研究では、東北環境再生モニタリングマップにツイート震災時のツイート提示機能を実装した。これにより、被災当時の被災地の状況を地理情報と組み合わせて閲覧できる仕組みを実現した。今後は、地名の曖昧性を解消したツイートの正確なマッピングや、現在投稿されている被災地に関するツイートのマッピングなどについて検討を行う予定である。

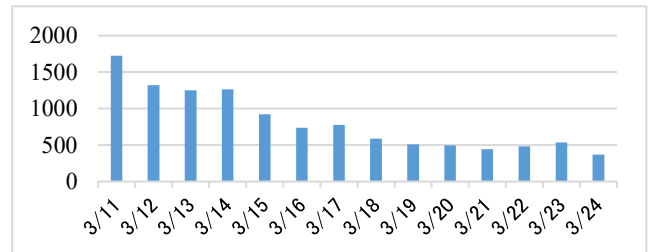
謝辞 本研究は JSPS 科研費・基盤研究(B)24300272、及び東海大学総合研究機構・プロジェクト研究・強化育成ステージ「災害・環境変動監視を目的としたグローバル・モニタリング・システムの構築による安全・安心な社会への貢献」の助成を受けて実施した。



(a) 6 市町の合計ツイート数



(b) 避難所ツイート数



(c) 大字ツイート数

図 5 3 月 11 日から 3 月 24 日のツイート数



図 6 ツイートマッピングの例

参考文献

- [1] 長幸平, 福江潔也, 内田理, 寺田和美, “衛星観測と現地調査による被災地の環境再生モニタリングと地球環境教育の実践について”, 日本写真測量学会平成 24 年度秋季学術講演会, 2012.
- [2] 加藤勝何, 内田理, 寺田一美, 福江潔也, 長幸平, 佐藤俊明, “TERM MAP: 東北環境再生モニタリングマップの構築”, 電子情報通信学会 基礎・境界ソサエティ大会, A-19-10, 2016.
- [3] 吉次由美, “東日本大震災に見る大災害時のソーシャルメディアの役割～ツイッターを中心に～”, 放送研究と調査, Vol.61, No.7, pp.16-23, 2011.

地点推定ワード: 七郷小 店長 〆 一期一会 〆 @MIYAGI_ken --- 2011年03月15日12時15分58秒 【緊急】@kinki0226: 仙台の七郷小学校にいますまだ、いまだに電気、なし、水分すらもなく、毛布すらありません (@karemarilive http://t.co/4rrWku9)” 浅村照也 @UDON01 --- 2011年03月15日13時01分22秒 @nobufogohara かなり逼迫した状況のようです。【緊急】@kinki0226: 仙台の七郷小学校にいますまだ、いまだに電気、なし、水分すらもなく、毛布すらありません ゆうゆ @jyo56 --- 2011年03月15日17時26分12秒 避難所七郷小学校、炊き出しをする人々に笑顔がありました
--

(a) 避難所ツイート

地点推定ワード: 本郷 きわちゃん @kiwakohaaaaan --- 2011年03月24日09時20分54秒 @aaaaarice そうなの? (◡_◡)でも清水もぎりぎりまで学校いるよ! 本郷13:42のねば間に合う! バナナ(“V”) @banana_macaron --- 2011年03月24日10時16分02秒 明日本郷(に)旅立つ院生にお別れの挨拶をしに学校へ行くか。本郷とかいづつでも会えるやん(・3・)アルエー
--

(b) 大字ツイート

図7 ツイートの表示例

災アーカイブプロジェクト「みちのく震録伝」, 電子情報通信学会誌, Vol.95, No.10, pp.921-926, 2012.

- [16] みちのく震録伝, <http://shinrokuden.irides.tohoku.ac.jp>
- [17] 河合美穂, “国立国会図書館における東日本大震災アーカイブの取組み”, 情報知識学会誌, Vol.22, No.4, pp.291-297, 2012.
- [18] ひなぎく NDL 東日本大震災アーカイブ, <http://kn.ndl.go.jp>
- [19] 岡本真, “震災アーカイブスの現状と課題”, 情報知識学会誌, Vol.22, No.4, pp.308-315, 2012.
- [20] 未来へのキオク, <https://www.miraiikioku.com>
- [21] 3がつ11にちをわすれないためにセンター, <http://recorder311.smt.jp>
- [22] 国土数値情報ダウンロードサービス, <http://nlftp.mlit.go.jp/ksj>
- [23] 国土交通省都市局『復興支援調査アーカイブ』, <http://fukkou.csis.u-tokyo.ac.jp>
- [24] Highcharts, <http://www.highcharts.com/>
- [25] GeoNLP, <https://geonlp.ex.nii.ac.jp>

- [4] A. Acar, Y. Muraki, “Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster”, International Journal of Web Based Communities, Vol.7, No.3, pp.392-402, 2011.
- [5] 三浦麻子, “東日本大震災とオンラインコミュニケーションの社会心理学 -そのときツイッターでは何が起こったか-”, 電子情報通信学会誌, Vol.95, No.3, pp.219-223, 2012.
- [6] B. D. M. Peary, R. Shaw, Y. Takeuchi, “Utilization of Social Media in the East Japan Earthquake and Tsunami and its Effectiveness”, Journal of Natural Disaster Science, Vol.34, No.1, pp.3-18, 2012.
- [7] S. Doan, B.-K. H Vo, N. Collier, “An Analysis of Twitter Messages in the 2011 Tohoku Earthquake”, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol.91, pp.58-66, 2012.
- [8] 篠田孝祐, 榊剛史, 鳥海不二夫, 風間一洋, 栗原聡, 野田五十樹, 松尾豊, “東日本大震災時における Twitter の活用状況とコミュニケーション構造の分析, 知能と情報, Vol.25, No.1, pp.598-608, 2013.
- [9] 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”, 2013 年度人工知能学会全国大会, 2B4-NFC-02a-2, 2013.
- [10] H. Wilensky, “Twitter as a Navigator for Stranded Commuters during the Great East Japan Earthquake”, Proc. 11th International Conference on Information Systems for Crisis Response and Management, pp.695-704, 2014.
- [11] O. Uchida, T. Rokuse, M. Tomita, Y. Kajita, Y. Yamamoto, F. Toriumi, B. Semaan, S. Robertson, M. Miller, “Classification and Mapping of Disaster Relevant Tweets for Providing Useful Information for Victims During Disasters”, IIEEJ Transactions on Image Electronics and Visual Computing, Vol.3, No.2, pp.224-232, Dec. 2015.
- [12] R. Kitajima, R. Kamimura, O. Uchida, F. Toriumi, “Potential Information Maximization: Potentiality-Driven Information Maximization and Its Application to Tweets Classification and Interpretation”, International Journal of Computer Information Systems and Industrial Management Applications, Vol.8, pp.42-51, Jan. 2016.
- [13] R. Kitajima, R. Kamimura, O. Uchida, F. Toriumi, “Identifying Important Tweets by Considering the Potentiality of Neurons”, IEICE Transactions on Fundamentals, Vol.E99-A, No.8, pp.1555-1559, Aug. 2016.
- [14] 今村文彦, 佐藤翔輔, 柴山明寛, “みちのく震録伝:産学官民の力を結集して東日本大震災のアーカイブに挑む”, 情報管理, Vol.55, No.4, pp.241-252, 2012.
- [15] 今村文彦, 佐藤翔輔, 柴山明寛, “震災経験を後世に伝える震