

Twitter からの震災時の行動経路および交通情報の自動抽出と可視化

石野亜耶・小田原周平・難波秀嗣・竹澤寿幸

Extracting Transportation Information and Traffic Problems from Tweets During a Disaster

Aya ISHINO, Shuhei ODAWARA, Hidetsugu NANBA and Toshiyuki TAKEZAWA

Abstract: 震災時に、安全な避難経路を確保するためには、他の被災者がどのように避難したのかといった行動経路に関する情報、また交通上のトラブルがどこで発生しているのかといった交通情報が重要である。そこで本研究では、震災時に Twitter 上で発信されたツイートから、機械学習を用いて被災者の行動経路および交通情報を自動抽出する手法を提案する。また自動抽出した情報を、地図上にマッピングし提示するシステムを構築する。

Keywords: 震災 (disaster), 情報抽出 (information extraction), 避難経路 (evacuation routes)

1. はじめに

2011 年 3 月 11 日 14 時 46 分に発生した東日本大震災では、日本で観測史上最大の Mw (モーメントマグニチュード) 9.0 を記録し、岩手県から千葉県までの 8 県におよぶ広域で震度 6 弱以上の強い地震動が発生した。東北日本の太平洋側で 10 メートルを超す大津波が襲来、東京電力福島第一原子力発電所の事故にとまなう大規模な避難が行われるなど、我々が経験したことのない大規模かつ複合的な大災害となった。震災時に、安全な避難経路を確保するためには、他の被災者がどのように避難したのかといった行動経路に関する情報、また交通上のトラブルがどこで発生しているのかといった交通情報が重要である。そこで本研究では、震災時に

Twitter 上で発信されたツイートから、機械学習を用いて被災者の行動経路および交通情報を自動抽出する手法を提案する。また、得られた行動経路を地図上にマッピングする事により、集約してユーザに提示し、閲覧できるシステムを開発する。このような情報を提示することで、被災者や、輸送物資を配送する援助者へ、利用可能な経路情報などの有益な情報を提供できることが期待できる。

本論文の構成は以下の通りである。2 節ではシステム動作例、3 節では提案手法、4 節では実験結果について述べ、5 節で本稿をまとめる。

2. システム動作例

本研究で構築したシステムについて、その動作例を紹介する。図-1 は行動経路提示システムの画面である。行動経路として、ユーザの移動元 (図中①) から移動先 (図中②) に向けて矢印が表示されており、移動手段に応じたアイコンが付与されている。この場合は、移動手段として車が用いられているた

石野亜耶 〒731-3194 広島市安佐南区大塚東 3-4-1

広島市立大学大学院 情報科学研究科

Phone: 082-830-1584

E-mail: ishino@ls.info.hiroshima-cu.ac.jp

め、車のアイコンが表示されている。また、アイコンをクリックすると、詳細な情報（図中④）を閲覧できるようになっている。

図-2 は交通情報提示システムの画面である。交通情報として、トラブル発生を開始地点（図中①）から終了地点（図中②）へ向けて矢印が表示されている。アイコン（図中③）をクリックすると、詳細な情報を閲覧することができる。本研究では、震災時に発信されたツイートから、震災時の行動経路および交通情報を抽出する手法を提案する。

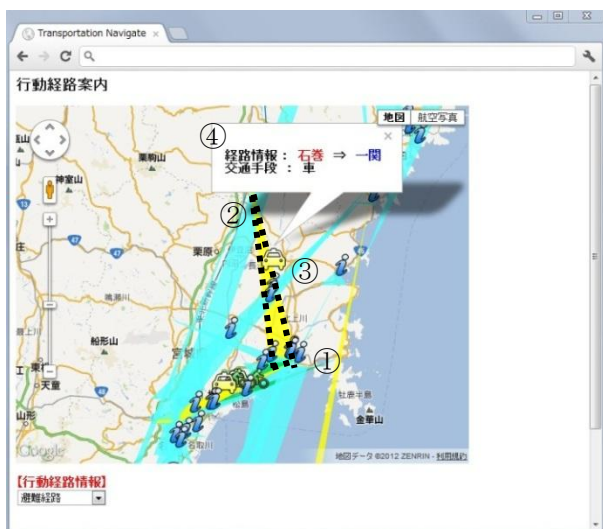


図-1 行動経路提示システム



図-2 交通情報提示システム

3. Twitter からの震災時の行動経路および交通情報の自動抽出

本研究では、東日本大震災の発生時に発信されたツイートから、行動経路と交通情報を自動的に抽出する手法を提案する。3.1 節では、行動経路の自動抽出手法、3.2 節、3.3 節では、交通情報の自動抽出手法について説明を行う。

3.1 行動経路の自動抽出

本研究では、行動経路として、移動元、移動先、移動手段を機械学習により抽出するために、以下の3種類のタグを定義する。図-3 がタグを付与した例である。

- FROM: 移動元
- TO: 移動先
- METHOD: 移動手段

義弟の安否確認が取れました。〈FROM〉石巻〈/FROM〉から〈METHOD〉徒歩〈/METHOD〉で〈TO〉仙台市内〈/TO〉の家まで帰って来たそうです。

図-3 行動経路にタグを付与した例

本研究では、機械学習としてCRFを使用した。CRF基本手法は与えられた文に含まれる語を分類するのに使用した。予備実験の結果から、窓幅は6と定めた。CRFには、素性として、単語、品詞、括弧（「, 『』など）、表1に示す手掛かり語を与える。

3.2 交通情報ツイートの検出

本研究では、震災時に発信されたツイートから、交通上のトラブル関連する情報を、交通情報として抽出する手法を提案する。しかし、交通情報を含むツイート（交通情報ツイート）の件数は、ごく少数であった。そのため、本研究では、まず震災時に発信されたツイートから、交通情報ツイートを検出し、その交通情報ツイートから、交通情報を抽出する。交通情報ツイートの検出手法については本節、交通情報検出手法については3.3 節で説明を行う。

表-1 行動経路の抽出に使用した手掛かり語

| タグ | 手掛かり語 | 件数 |
|------------|-------------------------------|------|
| FROM | “を出発” など FROM タグの後に使用される単語 | 5 |
| FROM TO | “学校” など避難場所になる場所に使用される単語 | 23 |
| | “方向” など FROM タグや TO タグに含まない単語 | 7 |
| | ANPI NLP で提供されている駅名 | 8619 |
| | ANPI NLP で提供されている東北地方の地名 | 1755 |
| | ANPI NLP で提供されている東北地方の学校名 | 806 |
| | ANPI NLP で提供されている電車の路線名 | 569 |
| TO | “へ避難” など TO タグの後に使用される単語 | 30 |
| METHOD | “で行く” など METHOD タグの後に使用される単語 | 19 |
| | “自動車” など移動手段になる単語 | 37 |

表-2 交通情報の抽出に使用した手掛かり語

| タグ | 手掛かり語 | 件数 |
|-------------|-----------------------------------|------|
| LINE | “号線” など道路関連の単語 | 23 |
| | ANPI NLP で提供されている電車の路線名 | 569 |
| | Wikipedia.から収集したバイパスの名前 | 1301 |
| | 高速道路の名前 | 60 |
| | 有料道路の名前 | 181 |
| LINE LOC | “通行止め” など交通上のトラブルが発生した際によく使われる単語 | 51 |
| | “通行可能” など交通上のトラブルが解消された際によく使われる単語 | 19 |
| LOC | ANPI NLP で提供されている駅名 | 8619 |
| | ANPI NLP で提供されている東北地方の地名 | 1755 |
| | “駅” など LOC タグによく含まれる単語 | 10 |

交通情報ツイートの検出には、機械学習として SVM を使用する。交通情報ツイートには、“通行通行止め”や“停止”などの語がよく含まれる。そのため、以下に示す手掛かり語を機械学習に使用する。

- ・ “通行止め”や“停止”など、交通上のトラブルが発生した際に、よく使用される語 (19 件)
- ・ “自動車道”や“インターチェンジ”など、道路に関連する語 (13 件)
- ・ “新幹線”や“地下鉄”など、交通手段に関連する語 (9 件)

3.3 交通情報の抽出

本節では、交通情報ツイートから交通情報を抽出

する手法について説明を行う。本研究では、交通情報として、トラブルが発生している道路や路線、区間を機械学習により抽出するため、以下の2種類のタグを定義する。図-4 がタグを付与した例である。

- LINE: トラブルが発生している道路や路線
- LOC: 交通上のトラブルの発生区間

地震で<LINE>中央自動車道</LINE>も<LOC>上野原</LOC>—<LOC>勝沼インターチェンジ</LOC>間などが通行止め。

図-4 交通情報ツイートのタグを付与した例

本研究では、機械学習として CRF を使用した。CRF 基本手法は与えられた文に含まれる語を分類するのに使用した。予備実験の結果から、窓幅は 6 と定めた。CRF には、素性として、単語、品詞、括弧(「」,『』など)、表 2 に示す手掛かり語を与える。

4. 実験

4.1 行動経路の自動抽出

本実験では、震災時に発信されたツイートとして、ANPI NLP で提供されているデータを使用した。(Neubig, G. et al., 2011) ANPI NLP で提供されている震災情報に関連するツイート 1303 件に対し、FROM, TO, METHOD タグを人手で付与したデータを機械学習に用いた。タグは訓練用に 728 件、評価用に 208 件付与された。機械学習には CRF++ (<http://www.chasen.org/~taku/software/CRF++>) を用いた。精度と再現率を用いて評価を行った。

実験の結果を表 3 に示す。精度に比べ、再現率は低い結果となった。再現率低下の原因は、主に地名に関する手掛かり語の不足であった。今後発生する震災に備え、各地の知名の詳細な辞書を予め構築しておく必要がある。

表-3 行動経路にタグを人手で付与した結果

| タグ | 精度 (%) | 再現率 (%) |
|--------|--------|---------|
| FROM | 78.4 | 40.8 |
| TO | 76.3 | 59.2 |
| METHOD | 91.7 | 64.7 |
| 平均 | 78.2 | 53.4 |

4.2 交通情報ツイートの検出

ANPI NLP で提供されている震災情報に関連する Twitter のデータ 1750 件に対し、危険情報ツイートかどうかを人手で付与したデータを、機械学習に用いた。人手で判定した結果、1750 件中、危険情報ツイートは 166 件であった。

実験の結果、精度 77.7%, 再現率 70.7%で、危険情報ツイートの判定を行うことができた。

4.2 交通情報の自動抽出

交通情報ツイート 166 件に対し、LINE, LOC タグを人手で付与したデータを、機械学習に用いた。タグは訓練用に 302 件、評価用に 106 件付与された。機械学習には CRF++を用いた。また、精度と再現率を用いて評価を行った。

実験の結果を表 4 に示す。実験の結果、高い精度を得ることができた。再現率低下の原因について述べる。タグを自動的に付与できなかったツイートには、通行止めの情報と、通行止め解除の情報が同時に記述されたものが多数あった。再現率を改善するためには、どの区間が通行止めになっているのかを解析するために、係り受けなどの情報を追加することが考えられる。

表-4 行動経路にタグを人手で付与した結果

| タグ | 精度 (%) | 再現率 (%) |
|------|--------|---------|
| LINE | 89.7 | 68.4 |
| LOC | 85.0 | 50.8 |
| 平均 | 87.0 | 57.1 |

5. まとめ

本研究では、東日本大震災の発生時に発信されたツイートから、機械学習を用いて、被災者の行動経路および交通情報を自動抽出する手法を提案した。また自動で抽出した情報を、地図上にマッピングし提示するシステムを構築した。

参考文献

Neubig, G., Matsubayashi Y., Hagiwara M. and Murakami K., 2011. Safety Information Mining – What can NLP do in a disaster –. Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP).