

# 災害時における現地情報 Tweet 抽出手法と評価

湯沢 昭夫<sup>†</sup> 小林 亜樹<sup>††</sup>

<sup>†</sup> 工学院大学工学部情報通信工学科 〒163-8677 東京都新宿区西新宿 1-24-2  
<sup>††</sup> 工学院大学情報学部情報通信工学科 〒163-8677 東京都新宿区西新宿 1-24-2  
 E-mail: <sup>†</sup>c513110@ns.kogakuin.ac.jp, <sup>††</sup>taki@cc.kogakuin.ac.jp

あらまし Twitter などの SNS を通じて、災害時の被災地状況を知ることには価値がある。単純には災害名での検索が想定されるが、災害名を含むものの、被災地の現状が含まれないものも存在する。そのため、大量にある tweet の中から被災地の状況を含む tweet を見つけるのは困難である。本研究では、災害名と関連性の高い語を抽出して用いることで、被災地の現状を含む tweet を抽出する手法を提案する。まず、一定時間範囲内に投稿された災害名と地名を含む tweet を抽出し、これを用いて災害名と関連性の高い単語を抽出する。次に、災害名を含む tweet の投稿時刻よりも後に同一投稿者によって投稿された tweet 集合に対し、災害名と関連性の高い単語をクエリとした検索を行う。これにより、被災地の状況を含む tweet を抽出する。実際に起きた災害を対象に検証を行い、本手法の有効性を確認した。キーワード Twitter, SNS, 現地情報, 情報抽出

## 1. はじめに

全世界で約 3 億人が利用している [1]Twitter は、tweet と呼ばれる 140 文字の短いメッセージを投稿し、利用者同士でのコミュニケーションや情報共有するためのソーシャルネットワーキングサービスである。投稿の手軽さから、情報発信へのハードルが下がっており [5]、2012 年 10 月の時点で、全世界で 1 日 5 億を超える tweet が投稿されている [2]。この大量の tweet の中には、利用者の身近な出来事や災害情報といった多種多様な情報が存在していることが報告されている [6] [7]。

一方で、2011 年 3 月 11 日に発生した東日本大震災の際において、救難活動や被災状況の共有、家族の安否確認などに Twitter が活用された [9]。また、2016 年 4 月 14 日に発生した熊本地震の際においても、漏水箇所やデマ情報の打ち消しなどに活用され [3]、東日本大震災の時よりも災害に関する tweet が上回っていることが報告されている [4]。

しかし、こうした被災地の状況について言及している情報の多くは、人手によって収集が行われていたのが現状である。また、Twitter 上で検索を行う場合、これらの情報には検索クエリを含まない可能性も出てくる。そのため、検索クエリと関連性のある情報を収集することが出来ない可能性が高い。

そのため、Twitter を対象とした情報を検索、抽出する試みは多数存在する [7] [8] [10] [11]。例えば、新田ら [7] は、ユーザから与えられたクエリを観測対象とし、クエリと関連性の高い語は観測対象に関連する情報を含んでいると考え、それらを抽出する手法を提案した。検索クエリと関連性の高い語に着目している点で本研究と似ているが、災害を対象としていない点で、本研究とは異なる。また、Sakaki ら [8] は、地震が発生した際に投稿される tweet はリアルタイム性が高く、その tweet には地震に関連する内容が言及されていると考え、地震に関連する語を含む tweet の分析することにより、震源地や台風の進路を推定した。災害情報に着目しているが、本論文では災害の状況

について言及している情報も抽出対象としている点で、先行研究とは異なる。

本研究では、検索クエリと関連性の高い語を抽出し用いることで、災害時における被災地の状況について言及している情報を抽出する手法を提案する。

本論文の構成として、第 2 章で提案モデルについて説明する。第 3 章で提案モデルを実現するための試作システムについて説明をする。第 4 章で実験と評価を行う。第 5 章で実験結果についての考察を行い、最後に、第 6 章でまとめと今後の課題について述べる。

## 2. 提案手法

### 2.1 現地情報の定義

本研究における現地情報とは、災害時、被災地やその周囲の状況について言及している情報のことである。

例えば、災害の影響による物理的な被害状況、どこでどの程度揺れたか、どのような揺れ方をしたのか等の揺れに関連する情報、被災地やその周囲でしか知りえない情報。このような、災害による被害情報や個人の体験や見聞が現地情報である。

現地情報は、救援活動や被災地の災害の状況を把握する上で重要な情報であるが、人や見方によって異なる可能性がある。

例えば、「被災地の近くじゃないから揺れは弱かった」という内容の tweet があるとする。この tweet の投稿者は、被災地の近くには居ないが、揺れを感知できる範囲内に居ることが分かり、投稿者の現在地ではあまり揺れなかったことが読み取れる。しかし、文章内の「被災地の近くじゃない」とは、どの場所を指しているのか明確ではないため、ある人によっては、被災地周辺では揺れがなかったという情報が読み取れるから現地情報と判断し、どこを指しているか分からないから現地情報ではないと判断する人もいる。

このように、人や見方によって判断が異なるため、どこまでを現地情報として含むか定義するのは困難である。そのため、

上記のみに関わらず、被災地の状況について言及していると、人が判断した情報は全て現地情報とする。

## 2.2 概要

本研究では、後述のアルゴリズムに従って、tweet に対して自動的に4つの属性によるラベル付けを行い、そのうち、特定の組み合わせを現地情報と推定し抽出する手法を提案する。図1に本手法の全体像を示す。

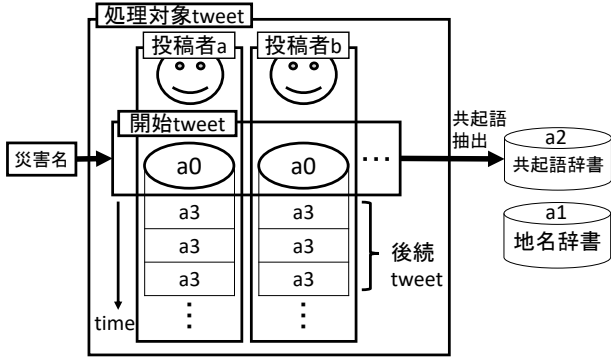


図1 提案手法の概要図

災害名とは、クエリとして与えられるキーワードであり、開始 tweet 集合とは、災害後の一定時間内に存在する災害名を tweet 本文中に含む tweet 集合である。

共起語辞書とは、開始 tweet 集合内で、災害名と共起している語（以降、共起語）による頻出上位語の集合である。また、これとは別途、地名辞書を持つ。地名辞書とは、都道府県名や市区町村名といった地名の集合である。

後続 tweet 集合とは、開始 tweet を投稿した同一の投稿者により投稿された tweet であり、開始 tweet を起点とし、開始 tweet よりも時間的に後ろに存在する一連の tweet 集合である。

処理対象 tweet 集合とは、開始 tweet 集合と後続 tweet 集合の和集合である。

これらの状態を前提条件として、処理対象 tweet 集合に対して、表1のような4つの属性によるラベル付けを行う。tweet 本文中に、災害名を含む tweet には「a<sub>0</sub>」、地名を含む tweet には「a<sub>1</sub>」、共起語を含む tweet には「a<sub>2</sub>」のラベルが付与される。また、後続 tweet には「a<sub>3</sub>」のラベルが付与される。

表1 属性対応表

属性名	説明
a <sub>0</sub>	災害名を tweet 本文中に含む
a <sub>1</sub>	地名を tweet 本文中に含む
a <sub>2</sub>	共起語を tweet 本文中に含む
a <sub>3</sub>	後続 tweet である

このとき、該当する属性が付与されたなら“1”，付与されていないなら“0”の2値変数とすると、tweet の特徴 A は表2のように (a<sub>0</sub>a<sub>1</sub>a<sub>2</sub>a<sub>3</sub>) の2進数表記できる。

本研究では、これら4つの属性の組み合わせによって、現地情報を含むと推定される tweet を抽出する。これを推定現地情報 tweet と呼ぶ。

表2 属性表記の例

処理対象 tweet \ 属性	a <sub>0</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>
	tweet1	0	0	0
tweet2	0	0	0	1
tweet3	1	0	1	0
⋮	⋮	⋮	⋮	⋮

次節では、推定現地情報 tweet の4属性の組み合わせについて説明を行う。

## 2.3 推定現地情報の基準

現地情報には、災害による被害情報や、個人の体験や見聞などの情報も含まれている。そのため、現地情報を含む tweet には、災害と関連性の高い語や、都道府県や市区町村といった特定の地域が特定できる地名が含まれていると考えられる。すなわち、a<sub>0</sub>a<sub>1</sub>a<sub>2</sub> の属性の組み合わせにより、現地情報を含む tweet を抽出できるのではないかと考えられる。

また、災害時に開始 tweet を投稿した投稿者は、被災者である可能性がある。開始 tweet を被災者による災害情報の第一報とすると、開始 tweet を起点とした後続 tweet は、第二報、第三報である可能性がある。そのため、後続 tweet には、現地情報を含む可能性がある。

したがって、処理対象 tweet 集合のうち、A = (11\*0)∪(\*111) の特徴を持つ tweet (\*は don't care) は現地情報を含む tweet と推定し、それらを抽出する手法を提案する。

## 3. 試作システム

### 3.1 概要

本研究では、A = (11\*0)∪(\*111) の特徴を持つ tweet は現地情報を含むと推定し、この特徴を持つ tweet を抽出する。図2に手順の流れの概要を示し、次節以降に詳細を述べる。

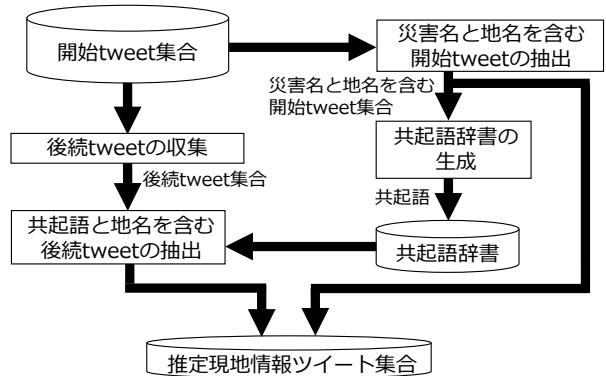


図2 試作システムの概要図

### 3.2 準備

災害後の一定時間範囲内で Twitter に投稿された tweet を、Twitter の Streaming API を用いて収集する。ただし、実装上は無償利用できる Twitter の sample API を利用したため、全体の1%である。また、リツイート・引用ツイートを除外した。

収集した tweet 集合を  $T$  とし、件数を  $i$  とするとき、各 tweet を  $\{t_1, \dots, t_i\}$  として表す。また、災害名  $q$  を設定する。

tweet から、その tweet の特徴  $A$  を得る関数として、

$$\text{attr}(\text{tweet})$$

を定義する。

### 3.3 開始 tweet の抽出

tweet 集合  $T$  を対象に、tweet 本文中に災害名  $q$  を含む tweet を抽出し、これを開始 tweet と呼ぶ。

得られた tweet 集合を

$$T_q = \{t_q | t_q \in T, \text{attr}(t_q) = (1 * * 0)\} \quad (1)$$

とする。

### 3.4 災害名と地名を含む開始 tweet の抽出

tweet 集合  $T_q$  を対象に、形態素解析器である MeCab<sup>(注1)</sup> で形態素解析を行う。このとき、MeCab の出力として品詞が「名詞」、品詞細分類 2 が「地域」と解析された語を地名  $l$  とし、tweet 本文中に地名  $l$  を含む tweet を抽出する。ただし、MeCab のシステム辞書として、mecab-ipadic-NEologd<sup>(注2)</sup> を用いる。

得られた tweet 集合を

$$T_{ql} = \{t_{ql} | t_{ql} \in T_q, \text{attr}(t_{ql}) = (11 * 0)\} \quad (2)$$

とする。

### 3.5 共起語辞書の生成

tweet 集合  $T_{ql}$  を対象に、形態素解析を行い、災害名  $q$  以外の、品詞が { 名詞, 動詞, 形容詞 } のいずれかに該当する語を共起語とし抽出する。そして、頻出上位  $M$  件の共起語を共起語集合とする。ただし、URL や “@” をはじめとする英字や記号で構成されるものは不要とし除去する。

得られた共起語集合を

$$C = \{c_1, \dots, c_M\} \quad (3)$$

とする。ただし、 $M$  は共起語の件数である。

### 3.6 後続 tweet の収集

同一の投稿者によって、開始 tweet の投稿時刻よりも、時間的に後に投稿された tweet を、Twitter の REST API を用いて収集する。これを後続 tweet と呼ぶ。

3.3 節で得たツイート  $t_q \in T_q$  の後続 tweet 集合を

$$\text{succT}(t_q, \text{cond.}) \quad (4)$$

とする。

条件 cond. は、

- (1) succT の要素数である tweet 件数が  $N$  件以内、
- (2) 時間  $[at1, at2]$  内に投稿された tweet、

の前者のみか、または両者を満たすかの 2 通りについて考える。ここで、

$$N = |\text{succT}(t_q, \text{cond.})| \quad (5)$$

を意味する。また、時刻  $at1$  は災害発生前の任意の時刻を、時刻  $at2$  は、災害発生後の任意の時刻であり、別に設定される。succT による後続 tweet の属性ベクトル  $A = (** * 1)$  である。

$\forall t_q \in T_q$  についての後続 tweet 集合の和集合を

$$S = \bigcup_{t_q \in T_q} \text{succT}(t_q, \text{cond.}) \quad (6)$$

とする。

### 3.7 共起語と地名を含む後続 tweet の抽出

後続 tweet 集合の和集合  $S$  を対象に、tweet 本文中に共起語  $c_M$  を含む tweet を抽出する。

得られた tweet 集合を

$$T_c = \{t_c | t_c \in S, \text{attr}(t_c) = (** 11)\} \quad (7)$$

とする。

tweet 集合  $T_c$  を対象に、MeCab で形態素解析を行い、MeCab のシステム辞書に登録されている地名  $l$  を含む tweet を抽出する。

得られた tweet 集合を

$$T_{cl} = \{t_{cl} | t_{cl} \in T_c, \text{attr}(t_{cl}) = (* 111)\} \quad (8)$$

とする。

### 3.8 推定現地情報 tweet 集合

本手法は、 $A = (11 * 0) \cup (* 111)$  の特徴を持つ tweet を推定現地情報 tweet として抽出する手法であるから、(2) 式と (8) 式より、抽出された tweet 集合  $T_{ql}$  および tweet 集合  $T_{cl}$  の和集合が目指す集合である。そのため、災害名と地名を含む tweet 集合  $T_{ql}$  と共起語と地名を含む tweet 集合  $T_{cl}$  の和集合を、推定現地情報 tweet 集合とする。

推定現地情報 tweet 集合を

$$T_{eds} = T_{ql} \cup T_{cl} \quad (9)$$

とする。

## 4. 評価実験

### 4.1 実験目的

提案手法の有効性を、Streaming API で得られた tweet 集合のうち、 $A = (1 * * 0)$  の特徴を持つ tweet を抽出した場合 (以降、比較手法) と、提案手法を用いて tweet を抽出した場合において取得した tweet を比較し検証する。

また、次章では実験結果や取得した tweet の内容についての考察を行う。

### 4.2 実験条件

2016 年 6 月 16 日 北海道函館市で起きた震度 6 弱の地震を対象とし、災害名  $q = \text{“地震”}$  とする。

Twitter の Streaming API を用いて、地震発生 1 分前の

(注1) : <http://taku910.github.io/mecab/>

(注2) : <https://github.com/neologd/mecab-ipadic-neologd>

14:21:00 から 15:59:59 の間に日本語を用いて投稿された、合計 29670 件の tweet を収集した。そのうち、 $A = (1 * * 0)$  の特徴を持つ、合計 984 件の開始 tweet を実験に用いる。ただし、リツイート・引用ツイートは除去した。

開始 tweet 集合に streaming API を使っているため、全 tweet 対象にはできないが、検証の目的にはこれらから辿れる一部のサンプルを用いていると理解すれば問題ない。

各パラメータとして、 $M = 10$  とした。succT は条件 cond. の (1), (2) の両方を満たすものとする。 $N = 20$ , at1 の時刻を 14:21:00, at2 の時刻を 15:59:59 とした。

### 4.3 現地情報の判定基準

提案手法と比較手法の精度を比較するために、提案手法および比較手法によって抽出した  $A = (1 * * 0) \cup (* 111)$  の特徴を持つ、合計 1181 件の tweet に対して、4 人の実験参加者に、現地情報を含むか否かの判定を行ってもらった。本実験では、3 人以上の実験参加者が現地情報を含むと判断した tweet 合計 71 件を正解とした。表 3 に結果を示す。

表 3 は、提案手法と比較手法それぞれにおける、抽出された tweet 合計 1181 件のうち、人手により現地情報を含むと判断された tweet 数 (正解), 1181 件に含まれる正解の割合 (割合) を示している。

各行は、1181 件のうち、 $k$  人が現地情報を含むと判断した正解件数を示している。ただし、 $k$  は現地情報を含むと判断した実験参加者の人数である。

表 3 現地情報を含むと判断した人数表

	4 人	3 人	2 人	1 人	0 人
正解	24	47	186	209	715
割合	0.0203	0.0398	0.1575	0.1770	0.6054

### 4.4 実験結果

実験で得られた結果を図 3, 表 4 に示す。

表 4 は、提案手法と比較手法それぞれにおける、抽出された tweet 数 (合計), うち人手により現地情報を含む正解とされた tweet 数 (正解), 同様に人手により不正解とされた tweet 数 (不正解), 合計に含まれる正解の割合 (割合) の 4 項目を示している。

各行が手法, あるいは抽出対象とした属性毎になっており, 提案手法によるものは最上行に示されている。提案手法の欄のうち, 残りの 4 行は, この提案手法で抽出対象とした tweet の持つ属性を本手法による最も細分化した状態で, その内訳を示したものである。

図 3 は, この表 4 の横軸を正解 tweet 数, 縦軸を正解割合としてプロットしたものである。青三角が比較手法であり, 赤丸が提案手法とその内訳によるものである。提案手法では, 表 4 における正解割合の高い順に 1~4 項目 (属性値) までを抽出すると想定した場合のプロットとなっている。したがって, 赤線右下の「上位 4 件」ラベルが当初提案手法による測定結果である。

この図 3 の縦軸は検索精度を表している。一方, 横軸は再現率に相当する軸として用意した。ただし, 本実験では一部の

tweet のみが分析対象であるため, 真の正解集合全体を得ることはできない。そのため, 実験データセット中に見いだされた正解 tweet 数を横軸にとっている。

この結果, 図 3 のグラフにおいては右上に行くほど良い結果であると言える。

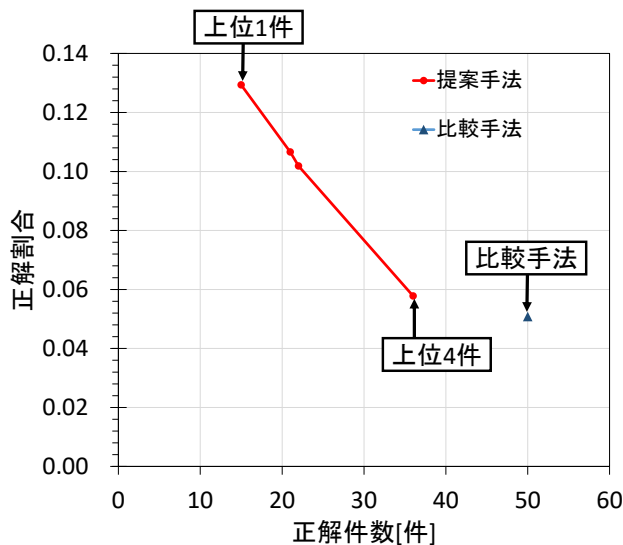


図 3 比較手法と提案手法の比較結果

表 4 各手法の抽出結果

	属性	合計	正解	不正解	割合
提案手法	$11 * 0 \cup * 111$	623	36	587	0.0578
	0111	116	15	101	0.1293
	1111	81	6	75	0.0741
	1100	19	1	18	0.0526
	1110	407	14	393	0.0344
比較手法	$1 * * 0$	984	50	934	0.0508

図 3, 表 4 より, 比較手法は, 正解件数において提案手法よりも高い値を示した。これは, 提案手法と比べて比較手法は, 地名や共起語といったフィルターを通さなかったためだと考えられる。提案手法は, 正解 tweet であるのにも関わらず, 4 属性の組み合わせに合致しない tweet は除外してしまうため, 正解件数が比較手法と比べて低いのではないかと考えられる。

しかし, 正解割合においては, 提案手法の方が高い値を示した。これは, 現地情報を抽出する上で, 後続 tweet がうまく作用したのではないかと考えられる。そのため, 各区分の正解割合が高くなったと考えられる。

このように, 提案手法の一定の有効性が確認できた。ただし, 他の災害でも同様の傾向かどうかは不明であるため, 別のデータでの実験は今後の課題である。

表 5 に, 災害名を“地震”としたときの頻出上位 10 件の共起語を示す。出現回数とは, 開始 tweet 集合  $T_{q_i}$  内で, 該当する共起語が抽出された回数のことを指す。また, 共起語が現地情報を抽出する際に寄与したかを確認するため,  $A = (* * 11)$  の特徴を持つ tweet の本文中に, 該当する共起語を tweet 本文中に含む tweet の合計件数, その正解, 不正解の件数, 正解割

表 5 共起語集合 (災害名  $q = \text{“地震”}$ )

順位	共起語	出現回数	品詞	$A = (**11)$			割合
				合計	正解	不正解	
1	北海道	278	名詞	74	2	72	0.0270
2	大丈夫	112	名詞	90	9	81	0.1000
3	函館	85	名詞	84	16	68	0.1905
4	あっ	50	動詞	29	1	28	0.0345
5	震度 6 弱	34	名詞	43	4	39	0.0930
6	揺れ	33	名詞, 動詞	79	24	55	0.3038
7	日本	32	名詞	21	0	21	0
8	心配	32	名詞	56	3	53	0.0536
9	ない	28	形容詞	198	9	189	0.0455
10	震度	25	名詞	181	15	166	0.0829

合を示す。

表 5 より、共起語として震源地である「北海道」や「函館」といった地名が抽出された。また、地震と関連性のある語として、「揺れ」や「震度」、「震度 6 弱」が抽出された。投稿者の感情を表現する語として、「心配」や「大丈夫」が抽出された。また、現地情報を抽出する際に、「日本」や「北海道」はあまり寄与しないことが、「揺れ」や「函館」は寄与することが確認できた。これより、共起語は現地情報を含む tweet を抽出する際にある程度寄与することが確認できた。今後の課題として、他の災害でも共起語が有効に働くか確認する必要がある。

表 6 に、本手法と比較手法で抽出された tweet の一部を示す。それぞれの tweet において、ID、属性、正解であるか否か、その tweet の本文を示す。また、tweet 本文中に出現する共起語を下線、地名を太字で示す。ID は各 tweet の通し番号の役割を果たす。各 tweet についての考察は次章で述べる。

## 5. 考察

### 5.1 共起語の改善点

表 6 より 6-1, 6-3 は、表 5 の共起語集合より、「ない」が含まれていたため抽出された tweet だと考えられる。「ない」は、様々な場面で日常的に使われる単語であるため、現地情報とは全く関係のない tweet を多く抽出するのではないかと考えられる。そのため、共起語を生成する際に、頻出単語を使うのではなく、TF-IDF 値を使うといった共起語辞書の生成の見直しの必要性が考えられる。

6-2 や 6-5 のような、「大丈夫」を含んでいて正解 tweet と判定されたものがある。6-2 と 6-5 の「大丈夫」を比較すると、「大丈夫」に続く付属の文字の「かー」、「でした」の違いが確認できる。「大丈夫かー」だと、心配しているように感じ取れる。「大丈夫でした」だと、心配されていた側の安否確認のような感じだと受け取れる。このことから、ある単語に続く付属の文字によって、文章の意味合いが変わってくると考えられる。そのため、同じ単語を含んでいても正解 tweet と不正解 tweet に分かれたのだと考えられる。単純に名詞、動詞、形容詞を取ってくるのではなく、その単語に続く付属の文字を活用することにより、現地情報をより多く抽出できるのではないかと考えられる。

### 5.2 辞書に登録されていない単語による抽出漏れ

2-1 は、「内浦湾」といった地理情報を含んでいるが、MeCab のシステム辞書にその地名が登録されていなかったため、地名を含まないと機械的に判断された tweet である。そのため、 $a_1$  の属性が“0”となっている。「内浦湾」で地域を特定することができると、MeCab の出力として品詞細分類 2 は“地域”が返ってきてほしいが、実際には、品詞細分類 2 は“固有名詞”と返ってきてしまう。

このように、辞書に登録されていない単語によって抽出漏れが発生するため、システム辞書の見直しや地域に特化した辞書の生成といった対策を検討する必要がある。

### 5.3 方言話者による tweet

3-1 では、tweet 本文に「わや」という言葉が含まれている。これは、北海道の方言で、めちゃくちゃなさま、手が付けられないさま、を意味しており、緊急地震速報で投稿者の周りのみんなが騒いでいることが伺える。3-1 は、現地情報を含まないと判断されたが、2-1 と同様に抽出漏れを防ぐために、方言にも対応できる辞書を用意していく必要が出てくるのではないかと考えられる。

### 5.4 tweet 内容の重複

4-1 のような、「地震」という単語のみを投稿する投稿者は多数いた。また、「地震あったの?」、「地震かー」といった内容の tweet や、4-2 のような地震速報 Bot も多数ある。これらの tweet は、伝えようとしている内容が同じであるため、ノイズになると考えられる。このように、内容が重複している tweet 集合から、代表として tweet を 1 件のみ抽出して、それ以外を除外するといったような、ノイズの除去が行えると考えられる。だが、その内容の重複をどう判断するのか、どうやって抽出するかは今後の課題である。

6-5 は、震源地やその周囲に住んでいる人たちを心配している tweet である。表 5 の共起語表より、「大丈夫」を含んでいるため抽出されたと考えられる。しかし、この心配している tweet は多数あり、内容も重複しているため、1-3 と同様に、内容が重複している tweet 集合から、tweet 1 件のみを抽出することで、ノイズの除去につながると考えられる。

### 5.5 表記揺れによる抽出漏れ

5-1 の投稿者は、青森の下北地域、もしくは下北半島のことを「下北」と省略している。仮に「下北半島」が地名として辞書に登録されている場合、「下北」では地名として認識されないため、現地情報を含む tweet を抽出する際の漏れとなってしまふ。それを防ぐために、このような表記ゆれに対応していくことは今後の課題である。

### 5.6 嘘ツイート問題

7-1 は、地震による影響で家の煙突が折れたらしいのだが、嘘か本当か見分けがつかない tweet もある。人ですら判断が出来ないため、デマ問題は未解決である。

## 6. おわりに

本論文では、tweet に対して自動的に 4 つの属性によるラベル付けを行い、そのうち、特定の組み合わせを現地情報と推定

表 6 抽出した tweet の内容

ID	$a_0$	$a_1$	$a_2$	$a_3$	正解	tweet 本文
1-1	1	1	1	0	○	STV で地震についてインタビューに答える読売センターの 函館 北部所長さん、「経験した中では大きい方だと思うが、被害は大きくなさそう」と。新幹線も止まってないみたい。その他、強い揺れを感じたが物が倒れたりはしていない。救急車の出動もなし。怪我人もなし、と。
1-2	0	1	1	1	○	五稜郭タワー、揺れたのは 10 秒ぐらい。観光客を避難誘導するまでもなく、動揺もそれほど見られなかった、とのこと。
1-3	0	1	1	1	○	夫のところ（札幌市北区）はちょっと揺れたらしい。中央区、というかうちは、揺れなかったが。
2-1	1	0	1	0	×	内浦湾で震度 6 弱、朝もあったけど、どうしてあそこで起きたのか。。。 \n かなり揺れたわ、緊急地震速報間に合わなかった
3-1	1	0	0	0	×	地震警報でみんなわや笑笑
4-1	1	0	0	0	×	地震
4-2	1	1	1	0	×	[第 1 報] 14:21:26 苫小牧沖 深さ 10km M6.4 最大震度 6 弱 #地震
5-1	0	0	1	1	○	私の方は下北だったけどそんな揺れなかったよ！ \n ありがとお麩っ！！
6-1	0	0	1	1	×	気づかないぞ……
6-2	0	1	1	1	○	@■■■■ ああああご心配ありがとうございます… !;;; \n もろ 函館 に住んでいますが全然大丈夫でした \n 余震に気をつけます \ (˘o˘) /
6-3	0	1	1	1	×	納得いかないドイツ語単語シリーズ。カメラマンが Fotograf. なんて Lehrer みたいに -er につけないんだらう…。
6-4	1	1	1	0	○	@■■■■ こんにちは !! \n 札幌も緊急地震速報入ったんですが、札幌の揺れはほぼ無かったです。 \n \n 函館の方も今のところあまり大きな被害があるような報道はされていないので……大丈夫なのかな？ \n 引き続き警戒が必要ですよね。 \n \n ありがとうございますっ !!'
6-5	0	1	1	1	×	北海道組 大丈夫 かー !?
7-1	1	0	0	0	×	地震で家の煙突折れた www

し抽出する手法を提案した。また、2016 年 6 月 16 日 北海道函館市で起きた震度 6 弱の地震を対象に実験を行い、本手法の有効性を確認した。

現地情報を含む tweet を抽出する際に、寄与しなかった共起語があった。これは、日常的に使われる単語であったため、現地情報を含まない tweet を多く抽出する結果となってしまった。共起語の生成方法については検討課題である。

今後の課題として、北海道地震に対してのみ実験を行ったため、他の災害でも同様の傾向かどうかは分からない。そのため、別のデータでの実験を今後行っていく必要がある。また、今回の 4 属性の分類カテゴリで、正解件数が多いにもかかわらず、割合が低いカテゴリについて細分化できないか、引き続き分析していく、

## 謝 辞

本研究の一部は科研費（26242013）の助成を受けたものである。

## 文 献

- [1] THE HUFFINGTON POST: Twitter が国内ユーザー数を初公表「増加率は世界一」、[http://www.huffingtonpost.jp/2016/02/18/twitter-japan\\_n\\_9260630.html](http://www.huffingtonpost.jp/2016/02/18/twitter-japan_n_9260630.html), (参照 2017-01-15)
- [2] CNET : Report: Twitter hits half a billion tweets a day, <https://www.cnet.com/news/report-twitter-hits-half-a-billion-tweets-a-day/>, (参照 2017-01-15)
- [3] 毎日新聞: 情報発信でツイッター活用 大西市長に聞く, [- mainichi.jp/articles/20161017/k00/00e/040/121000c, \(参照 2017-01-15\)
  - \[4\] 毎日新聞: ツイッター投稿、1 週間で 2 6 1 0 万件, <http://mainichi.jp/articles/20160519/k00/00m/040/059000c>, \(参照 2017-01-15\)
  - \[5\] 垂水 浩幸, “実世界インタフェースの新たな展開: 4. ソーシャルメディアと実世界”, 情報処理学会誌, Vol.51, No.7, pp.782-788 \(2010\).
  - \[6\] 山本 修平, 佐藤 哲司, “実生活ツイートに対する局面推定の精度向上に関する検討”, 情報処理学会論文誌, Vol.56, No.6, pp.1496-1506\(2015\).
  - \[7\] 新田 直子, 角谷 直人, 馬場口 登, “単語間の関係性の経時変化を考慮したマイクロブログからの実世界観測情報の抽出”, 日本データベース学会和文論文誌, Vol.13-J, No.1, pp.7-12\(2014\).
  - \[8\] Takeshi, S. Makoto, O. and Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors, Proc. 19th International Conference on World Wide Web \(WWW 2010\), pp.851-860,\(2010\).
  - \[9\] 河井 孝仁, 藤代 裕之, “東日本大震災の災害情報における Twitter の利用分析”, 広報研究 = Corporate communication studies, Vol.17, pp.118-128\(2013\).
  - \[10\] 山本 修平, 佐藤 哲司, “Twitter からの実生活情報の抽出法の提案”, DEIM Forum, C9-5, \(2012\).
  - \[11\] 斎藤 翔太, 伊川 洋平, 鈴木 秀幸, 村上 明子, “Twitter を用いた災害情報の早期発見”, 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol.114, No.81, pp.7-12\(2014\).](http://</a></li>
</ol>
</div>
<div data-bbox=)