東海大学紀要情報通信学部
Vol.9,No.2,2016,pp.7-14

論文

ニュース投稿をリツイートしたユーザの投稿に着目した災害関連情報の自動収集に向けた一検討

眞中 絢美[*1]，児玉 史緒里[*2]，内田 理[*3]，石井 啓之[*4]，宇津 圭祐[*5]

A Study on Collection of Disaster-related Information by Focusing on Posts by Users who Retweeted News Posts

by

Ayami MANAKA[*1], Shiori KODAMA[*2], Osamu UCHIDA[*3], Hiroshi ISHII[*4] and Keisuke UTSU[*5]

(received on Sep.30, 2016 & accepted on Dec.28, 2016)

あらまし

東日本大震災（2011年）以降，災害時における情報共有の手段として，ソーシャルメディアの活用について注目されている．著者らは，災害時においてソーシャルメディアのひとつであるTwitterに投稿される情報を被災者支援に活用することを目的に検討を行っている．そして，災害に関するニュース投稿（告知投稿）のリツイート直後の投稿に着目し，ここから災害関連情報を抽出することについて検討している．先行研究においては，2015年の小笠原諸島西方沖地震発生時の投稿をサンプルに分析を行っている．そして本論文では，サンプルから告知投稿に関連する投稿を判別する方法について検討を行っている．

Abstract

Twitter is noticed as the useful social media for collecting disaster-related information especially after the Great East Japan Earthquake in 2011. We have studied the utilization of posts on Twitter which is one of social media, to support victim. Then, to obtain disaster-related information, we focus on tweets posted immediately after retweets of a disaster-related announcement post. In our precious study, we analyze “sample posts” that were tweeted as a result of the earthquake off the west coast of Ogasawara Islands in 2015.　In this paper, we study a method to collected disaster-related information from the sample posts.

キーワード：災害関連情報，ツイッター
Keywords: *Disaster-related information, Twitter, social media*

1. はじめに

災害発生時には，被害状況の収集が重要である．テレビやラジオはこの目的で広く利用されている．一方，これらは放送型のメディアであり，個人的な情報発信には使えない．しかし現在では，SNS（Social Networking Service）などのソーシャルメディアを利用することにより，被災者が単に情報を受け取るだけでなく，自らが情報を発信することが可能となっている．

ソーシャルメディアの代表例のひとつにTwitter[1)]がある．Twitterにおいては，140文字以内の短文を投稿することができ，写真等の画像を添付することも可能である．ユーザ数は約3億2,000万人（2015年9月30日現在）[2)]であり，日本にも多くのユーザが存在する．ユーザ数が多い理由の一つとして，容易な手順で情報を発信できることが考えられる．Twitterにおいて，特定のアカウントによる投稿が自身のホーム画面のタイムラインに表示されるようになる仕組みのことを“フォロー”といい，あるアカウントをフォローしているユーザのことを“フォロワー”という．また，他のアカウントの投稿を自分のフォロワーと共有する機能のことを“リツイート”という．

Twitterはユーザが思ったり感じたりしたことを気軽に投稿できるという特性から，即時性が高いメディアであると言える．この特性に着目し，投稿された膨大な量の情報の中から，有益な情報を抽出する試みが行われている．また，災害発生時における情報共有にも利用されている．例えば2011年3月11日に東日本大震災が発生した際，Twitterを通して自身の状況を知らせることで安否の確認に繋がったことや，停電などでテレビやラジオから情報を入手できない人々に向けて警報や避難場所などの情報が多数投稿，拡散された

*1 情報通信学研究科情報通信学専攻 修士課程
Graduate School of Information and Telecommunication Engineering, Course of Information and Telecommunication Engineering, Master’s Program
*2 情報通信学部通信ネットワーク工学科 学部生
School of Information and Telecommunication Engineering, Department of Communication and Network Engineering
*3 情報理工学部情報科学科 教授
School of Information Science and Technology, Department of Human and Information Science, Professor
*4 情報通信学部通信ネットワーク工学科 教授
School of Information and Telecommunication Engineering, Department of Communication and Network Engineering, Professor
*5 情報通信学部通信ネットワーク工学科 講師
School of Information and Telecommunication Engineering, Department of Communication and

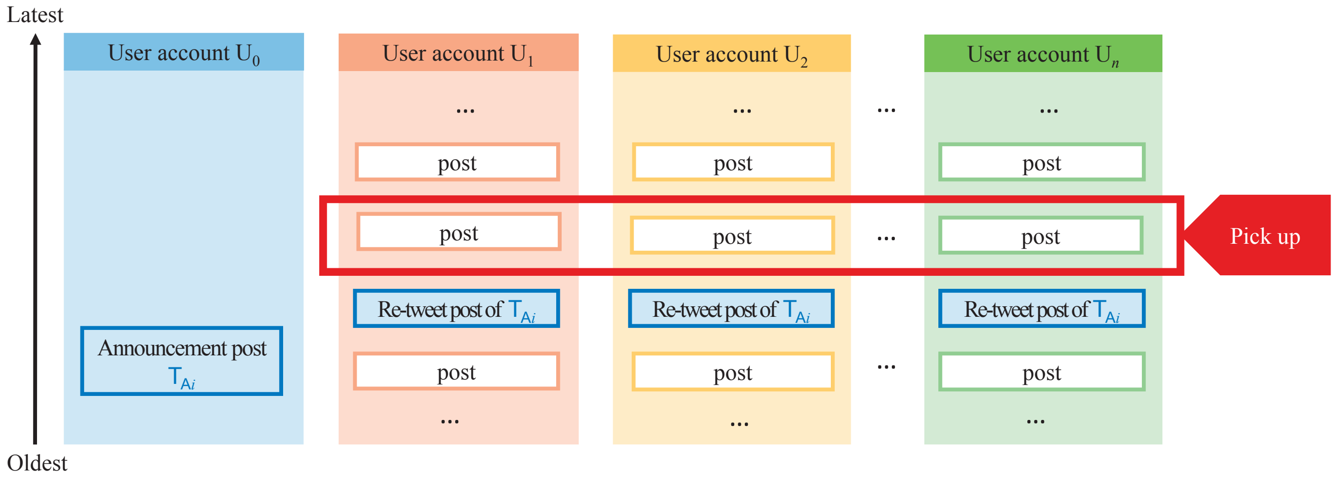

Fig 1.　Sample posts

事例が報告されている[3]．

また，Twitterの即時性の高さを活かし，投稿された情報を活用することについて注目されている．しかし，災害関連情報を収集するために，単純にキーワード検索を実施したとしても，検索結果は災害に関連しない情報（ノイズ）を多く含む結果となり適切ではない．筆者らは先行研究[4]において，災害関連情報の収集に関して，ニュース等の投稿を行っているアカウントによって，災害発生後にニュース等が投稿（本研究では告知投稿（Announcement post）と呼ぶ）された際，この告知投稿をリツイートしたアカウントにおけるリツイート直後の投稿に着目している．筆者らは，この投稿には，告知投稿に関連する情報や告知投稿の文面のみでは知りえない付加的な情報が含まれている，すなわち災害関連情報が含まれている可能性が高いと考えている．この仮説を検証するため，2015年5月30日に発生した小笠原諸島西方沖地震の発生後に投稿された投稿を収集し，分析を行い，高い確率で災害関連情報が含まれていることを示している．

本論文では，告知投稿に関連する投稿を機械的に収集することを目的とし，先行研究[4]で収集した投稿の中から，告知投稿の内容に関連する投稿を判別するための方法について検討している．本論文の構成は以下の通りである．2章では，Twitterに投稿された情報の活用に関する関連研究について紹介する．3章では，先行研究[4]において検討した，小笠原諸島西方沖地震の発生後の告知投稿のリツイート直後の投稿の収集と分析について述べている．4章では告知投稿の内容に関連する投稿を判別する方法について検討し，3章のサンプルをもとに評価している．最後に5章で本論文をまとめている．

2. Twitter 上の情報の活用に関する関連研究

Twitter をはじめとするソーシャルメディアから抽出したデータの活用に関してさまざまな検討が行われている．

Twitter においては，商品やイベントの告知投稿が商品メーカーやイベント主催者によって投稿されることが多く，これを見たユーザが告知投稿の内容について，関心を示す内容の投稿を行うことが見受けられる．また，告知投稿の文面のみでは知り得ない付加的な情報を別ユーザが投稿することも見受けられる．しかし，上記のように告知投稿に対して関連すると思われる投稿の多くは，必ずしも告知投稿と明示的に関連付けられているとは限らないため，他のユーザがこれらの投稿を見つけることは容易ではない．

これに対し塚本ら[5]は，Twitter において，告知投稿のリツイート直後の投稿から関連すると思われる投稿を自動で収集する手法を提案した．しかし，対象とする情報の種類については限定的であり，災害情報の抽出に関する検討は行われていない．

一方，Twitter の即時性の高さに着目し，非常時や災害時などの情報収集手段としての活用を試みた研究が近年多数報告されている．例えば後藤ら[6]は，災害に関する質問に対し Twitter から抽出された回答のリストを提示するシステムを開発した．著者ら[7]は，大規模災害時において帰宅困難者が一時的に滞在できる施設に関する情報を Twitter から抽出し，その避難情報をわかりやすくマップ上に提示するシステムを提案した．また，利用者のニーズに適合するような被災者支援情報を Twitter から抽出し提供するシステム[8]，Twitter の災害関連する投稿を複数のカテゴリーに分類し，地図上にマッピングする手法[9]，Twitter の災害に関連する投稿の有益性についてニューラルネットワークを用いて判定する手法を提案した[10]．

3. 告知投稿のリツイート後の投稿に関するサンプル収集と分析

筆者らは災害関連情報の抽出について，塚本らの検討[5]で述べられていた，告知投稿についてリツイートを行ったアカウントにおける，リツイート直後の投稿に着目することによって，収集することを検討している．塚本らの検討[5]では，映画の公開情報，サッカー

の試合情報，コンビニエンスストアの新商品情報，テレビ番組などに関する告知投稿を対象に検討されている．これに対し先行研究[4]では，塚本らの検討[5]では扱われていなかった災害情報に着目し，2015 年 5 月 30 日に発生した小笠原諸島西方沖地震に関する告知投稿を例に分析し検討を行っている．災害情報に関する告知投稿に対し，災害関連情報を高い確率で抽出できれば，筆者のグループが検討を進めている災害時における被災者支援システムの構築[7)8)9)10]において，より信頼性の高い情報を活用でき，有用性を高められると考えられる．

以下では，告知投稿として小笠原諸島西方沖地震に関するニュースを例にとり，各告知投稿に関するリツイート後の投稿が，災害関連情報であるかどうかを分類する．

3.1 サンプル投稿の収集

告知投稿に関するサンプル投稿を収集する．本稿では，小笠原諸島西方沖地震の発生の際に，ニュースを発信している公共性の高いアカウントである，Yahoo! ニュースや NHK ニュースのアカウントに着目する．そして，告知投稿の例として，災害に関するさまざまなニュース 10 件を対象とする．これを Table 1 の T_{A1} ～T_{A10} に示す（地震発生後，投稿された時刻の順である）．そして，ある告知投稿 T_{Ai}（$i = 1, 2, \cdots, 10$）のリツイートを行ったアカウント U_j（$j = 1, 2, \cdots$ n）における，T_{Ai} のリツイート直後の投稿を収集する（Fig. 1）．この投稿のことを，以下ではサンプル投稿と呼ぶ．収集には Web アプリケーション[11]を使用する．本 Web アプリケーション[11]では，以下のようにサンプル投稿が収集される．

- ニュースのリツイートを行った先着 100 ユーザが対象．
- リツイート後，10 分以内の投稿が対象
- リツイート後，3,200 件以上投稿したユーザは除外

収集されたサンプル投稿の件数について Table 1 に示している．尚，T_{A1} と T_{A10} に関しては，収集されたサンプル投稿の件数が 100 件を超えており，アプリケーションで収集できる件数の上限を超過している可能性がある（本来収集されるべきリツイート直後の投稿の数は Table 1 に示したサンプル投稿の数以上である可能性がある）．

3.2 判別方法

T_{A1}～T_{A10} の各告知投稿に対するサンプル投稿について，災害関連情報であるかどうか判別する．本研究では，以下のカテゴリーA および B のいずれかに該当するものを災害関連情報とみなす．

カテゴリーA：サンプル投稿はもとの告知投稿の内容に関連しているものである．

カテゴリーB：サンプル投稿はもとの告知投稿の内容からでは知り得ない，付加的な情報を含むものである．

例として，T_{A1} とこれに対するサンプル投稿の例を Fig. 2 に示す．Example 1 では，地震の震度を知り，その大きさについて驚きを表すような投稿となっている．Example 2 では，T_{A1} に掲載されていた地震発生時の震度分布図についての感想が述べられている．また，投稿にはリツイートに対しての感想であることを示す “>RT” や “rt” が含まれている．これらは，カテゴリーA に該当すると予想される．Example 3 は，T_{A1} が示す地震についての感想は述べておらず，地震発生によって生じた鉄道運行状況についての内容である．この投稿に記載された URL（Uniform Resource Locator）のリンク先には “東海道新幹線が上下線ともに運転見合わせ．” といった情報に加え，当時の運行情報を記載している URL が示されている．この例では，T_{A1} の文面のみでは知りえない付加的な情報が含まれている．具体的には，告知投稿の文面からは，新幹線の運転停止について知ることはできず，T_{A1} と Example 3 の両方を見ることによって，地震の発生により新幹線の運転停止が生じている事実を知ることができる．すなわち，カテゴリーB に該当すると予想される．

評価方法として，まず，研究室の関係者 7 名（学生 6 名，教員 1 名）により，サンプル投稿が以下のカテゴリーA, B に該当するかどうかを判断する．そして，各告知投稿について，すべてのサンプル投稿のうち，カテゴリーA，カテゴリーB のそれぞれに属する投稿であると，被験者 7 名のうち過半数の 4 名以上が回答したものの割合を求める．

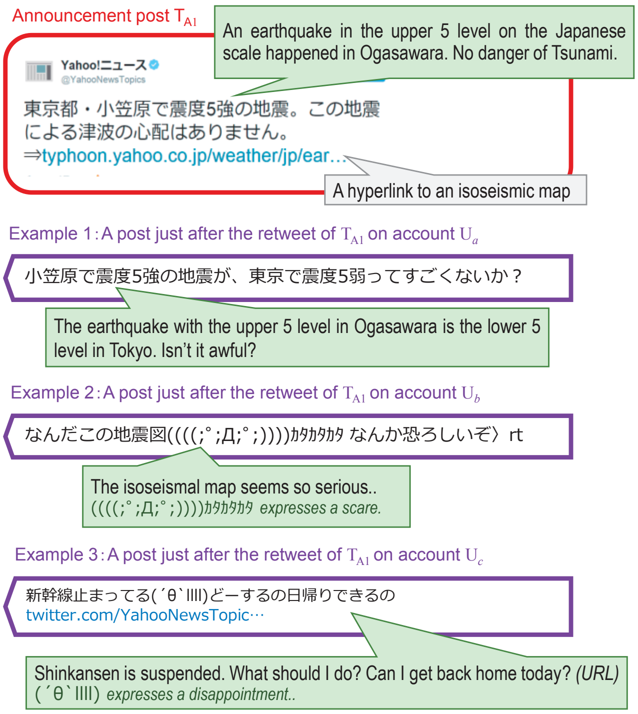

Fig. 2　Example posts after retweeting of the announcement posts

Table 1 Focused news tweets (T_{A1}～T_{A10})

	Account / posted date and time	Original tweet	Tweet translated in English
T_{A1} (132 samples)	Yahoo! News, @YahooNewsTopics / 20:32, 30 May, 2015	東京都・小笠原で震度 5 強の地震。この地震による津波の心配はありません。⇒ http://typhoon.yahoo.co.jp/weather/jp/earthquake/20150530202422.html	The earthquake had an intensity of upper 5 happened in Ogasawara, Tokyo. There is no danger of Tsunami. *(URL) (The original tweet includes the image of a seismic intensity map.)*
T_{A2} (69 samples)	NHK News @nhk_news / 20:40, 30 May, 2015	【速報 JUST IN 】東海道新幹線 停電で運転見合わせ http://nhk.jp/N4JV4Efx #nhk_news	[Breaking news] Tokaido Shinkansen (super express train) has been suspended due to blackout.
T_{A3} (54 samples)	Yahoo! News @YahooNewsTopics / 20:47, 30 may, 2015	小笠原と神奈川東部で震度 5 強、埼玉で震度 5 弱の揺れを観測。東海道新幹線は、地震による停電のため、東京駅と新大阪駅の間で上下線ともに運転を見合わせています。⇒ http://www3.nhk.or.jp/news/html/20150530/k10010097291000.html … 運行情報はこちら。⇒ http://transit.loco.yahoo.co.jp/traininfo/top	The earthquake measured an intensity of upper 5 in Ogasawara and southern Kanagawa; and measured that of lower 5 in Saitama. Tokaido Shinkansen has been suspended in both directions between Tokyo sta and Shin-Osaka sta due to blackout owing to the earthquake. *(URL)* Operation information is here. *(URL)*
T_{A4} (69 samples)	NHK News @nhk_news / 20:50, 30 May, 2015	小笠原諸島と神奈川県で震度 5 強 http://nhk.jp/N4JV4Eg0 #nhk_news	The earthquake had an intensity of upper 5 happened in Ogasawara. *(URL)*
T_{A5} (49 samples)	NHK Metropolitan @nhk_shutoken / 21:04, 30 May, 2015	【ニュース】＜埼玉と東京 600 世帯停電＞ 東京電力によりますと、さきほどの地震の影響で、埼玉県と東京都のあわせて 600 世帯で停電が起きているということです。東京電力で状況を詳しく調べる… http://www3.nhk.or.jp/shutoken-news/20150530/5184843.html #nhk	[News] <600 households in Saitama and Tokyo lost power> According to Tokyo Electric Power Co., total of 600 households in Saitama and Tokyo have lost the power due to the earthquake a short time ago.
T_{A6} (22 samples)	NHK Metropolitan @nhk_shutoken / 21:20, 30 May, 2015	【鉄道情報】JR 東日本によりますと今回の地震の影響で、東海道線と横須賀線、山手線、京浜東北線は、いずれも東京駅と品川駅の間で運転を見合わせています。 http://www3.nhk.or.jp/shutoken-news/20150530/5184873.html #nhk	[Railway information] According to East Japan Railway Company, Tokaido line, Yokosuka line, Yamanote line, and Keihin-tohoku line has suspended between Tokyo sta and Shinagawa sta. *(URL)*
T_{A7} (28 samples)	NHK Metropolitan @nhk_shutoken / 21:19, 30 May, 2015	【関東で震度 5 強 交通への影響】地震の影響で、JR 山手線などが運転見合わせ。現在の運行情報。 http://yahoo.jp/F0OMlN	[Earthquake of upper 5 in Kanto region, Traffic influence] Yamanote line and other lines of JR (Japan Railway Company) has suspended due to the earthquake. Current operation information. *(URL)*
T_{A8} (45 samples)	NHK News @nhk_news / 21:24, 30 May, 2015	【埼玉と東京 600 世帯で停電】東京電力によると、地震の影響で、埼玉県と東京都の合わせて 600 世帯で停電。 http://yahoo.jp/GmfDZv	[600 households in Saitama and Tokyo lost power] According to Tokyo Electric Power Co., total of 600 households in Saitama and Tokyo have lost the power due to the earthquake. *(URL)*
T_{A9} (13 samples)	NHK Metropolitan @nhk_shutoken / 22:02, 30 May, 2015	【地震情報】JR 東日本千葉支社によりますと、地震の影響で運転を見合わせていた総武線快速と武蔵野線は、いずれも午後 9 時 55 分、上下線とも全線で運転を再開しました。 http://www3.nhk.or.jp/shutoken-news/20150530/5184873.html #nhk	[Earthquake information] According to Chiba branch of East Japan Railway Company, Sobu line and Musashino line resumed in both directions at 9:55 pm. these lines had been suspended due to the earthquake. *(URL)*
T_{A10} (139 samples)	NHK News @nhk_news / 23:04, 30 May, 2015	【速報 JUST IN 】JR 山手線など 午後 11 時半ごろ再開見込み http://nhk.jp/N4JV4Egf #nhk_news	[Breaking News] Yamanote line and other lines of JR will be resumed around 11:30 p.m. *(URL)*

3.3 判別結果

Fig. 3 に，各告知投稿について，全サンプル投稿のうち，サンプル投稿がカテゴリーA，カテゴリーB のそれぞれまたはいずれかに属すると，被験者 7 名のうち過半数の 4 名が回答したものの割合を示す．

(i) カテゴリーA に属すると判別されたサンプル投稿の割合

T_{A1}～T_{A10} に対する，全サンプル投稿のうち被験者 7 名中 4 名以上がカテゴリーA に属すると判別したサンプル投稿の割合は，最高のものでは T_{A1} に対するサンプル投稿で 84.1%，最低のものでは T_{A2} に対するサンプル投稿で 32.4%であった．T_{A1}～T_{A10} の全ての平均では 51.6%であり，サンプル投稿のうちおよそ 2 件に 1 件は該当するものと考えられる．この結果より，地震に関する告知投稿についてリツイートを行ったユーザにおける，リツイート直後の投稿の内容が，告知投稿と関連している可能性が高いことが示された．

(ii) カテゴリーB に属すると判別されたサンプル投稿の割合

T_{A1}～T_{A10} に対する，全サンプル投稿のうち被験者 7 名中 4 名以上がカテゴリーB に属すると判別したサンプル投稿の割合は，最高のものでは T_{A1} に対するサンプル投稿で 27.3%，最低のものでは T_{A10} に対するサンプル投稿で 3.6%であった．T_{A1}～T_{A10} の全ての平均では 14.0%であった．この結果により，(i)の評価と比較して割合は低いものの，地震に関する告知投稿について，リツイートを行ったユーザのリツイート直後の投稿に着目することによって，告知投稿の文面のみでは知り得ない付加的な情報が得られる可能性があることが示された．

(iii) カテゴリーA およびカテゴリーB のいずれかに属すると判別されるサンプル投稿の割合

T_{A1}～T_{A10} に対する，全サンプル投稿のうち被験者 7 名中 4 名以上がカテゴリーA およびカテゴリーB のいずれかに属すると判別したサンプル投稿の割合は，最高のものでは T_{A1} に対するサンプル投稿で 84.1%，最低のものでは T_{A8} に対するサンプル投稿で 42.2%であった．T_{A1}～T_{A10} の全ての平均では 57.3%であった．この結果により，サンプル投稿のうちおよそ 2 件に 1 件が該当すると考えられ，告知投稿との関連性が高い，もしくは告知投稿の文面のみでは知り得ない付加的な情報が得られる可能性が高いことが示された．尚，(i)，(ii)の結果を合わせて考えると，告知投稿に対するサンプル投稿がカテゴリーA に属する割合が高くても，そのサンプル投稿がカテゴリーB に属している割合が高いとは限らない，すなわち，サンプル投稿の内容が告知投稿との関連があると判別されるものであったとしても，必ずしも告知投稿の文面のみでは知り得ない付加的な情報が得られるとは限らないと考えられる．

以上(i)，(ii)，(iii)の結果より，小笠原諸島西方沖地震に関する告知投稿について，告知投稿のリツイート直後の投稿に着目することによって，災害関連情報が得られる見通しを得た．

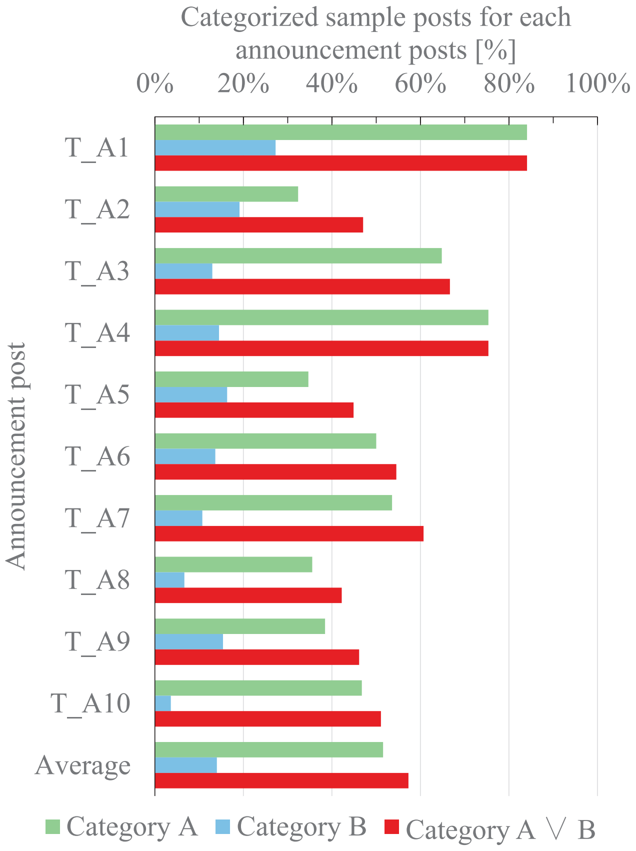

Fig.3 Evaluation result of the sample post for each announcement posts T_{Ai}

4. 告知投稿の内容に関連する投稿の判別

サンプル投稿の中から告知投稿の内容に関連する投稿（カテゴリーA に該当する）を機械的に収集することを目的に，サンプル投稿の判別方法について検討する．

4.1 判別方法

前述の通り，災害関連情報を収集するために，単純にキーワード検索を実施したとしても，検索結果はノイズを多く含む結果となり，適切ではない．ここで著者らは，カテゴリーA に該当する投稿には，告知投稿の文面に含まれている特徴的な語句（特徴語句）を含むと考え，これをもとに判別する以下の手順を提案する．まず，告知投稿のテキスト（外部にリンクするための URL，およびハッシュタグを除く）について，Yahoo!テキスト解析:キーフレーズ抽出 [12]を利用しキーフレーズの抽出を行う．ここで，抽出されたキーフレーズは重要度順に表示される．なお，重要度については本論文では取り扱わない．次に，抽出されたキーフレーズをさらに形態素解析ウェブアプリ UniDic-MeCab[13]を用いて形態素解析を行う．形態素解析を行

Table 2 Results of the morphological analysis

Announcement post		Result of the key phrase analysis (in English) {degree of importance}	Result of the morphological analysis
T_{A1}	東京都・小笠原で震度５強の地震。この地震による津波の心配はありません。	地震 (earthquake) {100}, 小笠原 (Ogasawara) {96}, 津波 (Tsunami) {68}, 震度５強 (earthquake intensity level upper 5) {61}, 東京都 (Tokyo Metropolitan) {59}, 心配 (anxiety){45}	地震, 小笠原, 津波, 震度, ５強 (*1), 東京, 都, 心配
T_{A2}	【速報 JUST IN 】東海道新幹線 停電で運転見合わせ	JUST IN {100}，東海道新幹線 (Tokaido-Shinkansen){95}，停電 (blackout) {48}，速報 (newsflash) {40}，運転 (operation) {36}	ＪＵＳＴ, ＩＮ, 東海道, 新幹線 (*1), 停電, 速報, 運転
T_{A3}	小笠原と神奈川東部で震度５強、埼玉で震度５弱の揺れを観測。東海道新幹線は、地震による停電のため、東京駅と新大阪駅の間で上下線ともに運転を見合わせています。運行情報はこちら。	東海道新幹線 {100}, 震度 {90}, 上下線とも (both up and down tracks) {80}, 神奈川東部 (eastern part of Kanagawa Prefecture) {72}, 小笠原 {67}, 新大阪駅 (Shin-Osaka station) {57}, 停電 {51}, 観測 (observe) {48}, 運行情報 (information of operation) {48}, 揺れ (quake) {46}, 震度５弱 (earthquake intensity level lower 5) {44}, 埼玉 (Saitama) {44}, 震度５強 {43}, 地震 {41}, 運転 {38}, 東京駅 (Tokyo station) {34}, 間 (between) {18}	東海道, 新幹線(*1), 震度, 上下線, 神奈川, 東部, 小笠原, 新大阪(*1), 駅, 停電, 観測, 運行, 情報, 揺れ, 震度, ５弱 (*1), 埼玉, 震度, ５強 (*1), 地震, 東京, 駅, 間
T_{A4}	小笠原諸島と神奈川県で震度５強	小笠原諸島 (Ogasawara islands) {100}, 神奈川県 (Kanagawa prefecture) {68}, 震度５強 {54}	小笠原, 諸島, 神奈川, 県, 震度, ５強(*1)
T_{A5}	【ニュース】＜埼玉と東京　６００世帯停電＞ 東京電力によりますと、さきほどの地震の影響で、埼玉県と東京都のあわせて６００世帯で停電が起きているということです。東京電力で状況を詳しく調べる…	東京電力 (Tokyo Electric Power Company) {100}, さきほど (just now) {62}, 埼玉県 (Saitama prefecture) {42 }, 600 世帯停電 (600 households blackout) {38}, ニュース (news) {35}, 地震 {34}, 東京都 {34}, 影響 (affect) {29}, 状況 (status) {27}	東京, 電力, さきほど, 埼玉, 県, ６００ 世帯(*2), 停電, ニュース, 地震, 東京, 都, 影響, 状況
T_{A6}	【鉄道情報】ＪＲ東日本によりますと今回の地震の影響で、東海道線と横須賀線、山手線、京浜東北線は、いずれも東京駅と品川駅の間で運転を見合わせています。	JR 東日本 (East Japan Railway Company) {100}, 横須賀線 (Yokosuka line) {53}, 京浜東北線 (Keihin-Tohoku line) {52}, 東海道線 (Tokaido line) {49}, 山手線 (Yamanote line) {38}, 品川駅 (Shinagawa station) {28}, 鉄道情報 (Railway information) {26}, 地震 {24}, 運転 {22}, 東京駅 {20}, 影響 {20}, 間 {11}	ＪＲ, 東, 日本, 横須賀, 線, 京浜, 東北, 線, 東海道, 線, 山手, 線, 品川, 駅, 鉄道, 情報, 地震, 運転, 東京, 駅, 影響, 間
T_{A7}	【関東で震度５強 交通への影響】地震の影響で、JR 山手線などが運転見合わせ。現在の運行情報。	JR 山手線 {100}, 影響 {99}, 震度５強交通 {90}, 運行情報 {81}, 地震 {69}, 関東 (Kanto) {69}, 運転 {64}	ＪＲ, 山手, 線, 影響, 震度, ５強 (*1), 交通, 運行, 情報, 地震, 関東, 運転
T_{A8}	【埼玉と東京 600 世帯で停電】東京電力によると、地震の影響で、埼玉県と東京都の合わせて 600 世帯で停電。	停電 {100}, 東京電力 {81}, 埼玉県 {58}, 地震 {47}, 東京都 {47}, 影響 {40}, 東京 600 世帯 {34}	停電, 東京, 電力, 埼玉, 県, 地震, 東京, 都, 影響, 東京, ６００世帯(*2)
T_{A9}	【地震情報】ＪＲ東日本千葉支社によりますと、地震の影響で運転を見合わせていた総武線快速と武蔵野線は、いずれも午後９時５５分、上下線とも全線で運転を再開しました。	JR 東日本千葉支社 (East Japan Railway Company Chiba) {100}, 武蔵野線 (Musashino line) {66}, 上下線 {64}, 総武線快速 (Sobu line rapid service) {64}, 地震 {55}, 運転 {50}, 全線 (whole line){49}, 地震情報 (earthquake information){33}, 影響 {27}, 午後 9 時 55 分 (9:55 am) {21}	ＪＲ, 東, 日本, 千葉, 支社, 武蔵野, 線, 上下, 線, 総武, 線, 快速, 地震, 運転, 全線, 地震, 情報, 影響, 午後, ９時 (*2), ５５分 (*2)
T_{A10}	【速報 JUST IN 】ＪＲ山手線など 午後１１時半ごろ再開見込み	午後 11 時半ごろ再開見込み (will be resumed at 11:30 pm) {100}, 速報 JUSTIN (newsflash just in){94}, JR 山手線 {89}	午後, １１時 (*2), 半, ごろ, 再開, 見込み, 速報, ＪＵＳＴＩＮ, ＪＲ, 山手線

って分割された語句を，本論文では特徴語句と呼ぶ．Table 2 に各告知投稿に対する，抽出したキーフレーズと，特徴語句について示す．

以下では，判別手順について告知投稿 T_{A1} を例に具体的に述べる．T_{A1} に含まれるキーフレーズは「地震，小笠原，津波，震度 5 強，東京都，心配」であった．このキーフレーズをさらに UniDic-MeCab [13]を用いて形態素解析すると，特徴語句は「地震，小笠原，津波，震度，5 強，東京，都，心配」となる．ここで，キーフレーズをそのまま用いるのではなく，キーフレーズをさらに形態素解析を行い特徴語句で判別を行う理由は以下のとおりである．キーフレーズを解析する際に，複合名詞として抽出されることがあるが，この複合名詞をもとに判別を行っても，告知投稿の内容

に関連すると判断できないことがあるからである．例えば，T_{A1} に対するサンプル投稿うち，「震度 5 強」のように複合名詞となっているキーフレーズを含むものは 1 件のみである．一方で，キーフレーズを形態素解析し「震度」と「5 強」に分割した場合，サンプル投稿のうち「震度」を含むものは 15 件，「5 強」を含むものは 1 件であり，より適切に判別できる．なお，キーフレーズについて形態素解析を実施する際，以下の取り扱いを行う．

- 半角英数字は全角文字として取り扱う．これは UniDic-MeCab[13] においては，半角英数字は記号として解析される仕様となっているため，適切に形態素解析を実施するためである．
- 名詞と接頭辞および接尾辞が連続している場合は連結する（Table 2 の(*1)が該当）．これを実施する理由として，T_{A1} のキーフレーズ「震度 5 強」を例に説明する．この語句は形態素解析を行うと，「5」（名詞-数詞）と「強」（接尾辞）のように分割される．しかし，この分割を行う前の語句の意味と，この分割を行った場合の語句の意味が，異なってしまい，結果として判別に用いることは適切ではないと考えられるためである．
- 連続する数字は結合し，結合された数字の直後が名詞の場合は結合する（Table 2 の(*2)が該当）．これを実施する理由として，T_{A9} のキーフレーズ「午後 9 時 55 分」を例に説明する．この語句は形態素解析を行うと，「午後（名詞）」，「９（名詞-数詞）」，「時（名詞）」，「５（名詞-数詞）」，「５（名詞-数詞）」，「分（名詞）」と分割される．ここで，「９」はこれに続く「時」と結合することにより意味をなすからである．また「５」「５」は数字の連続のため結合し「５５」となるが，これに続く「分」と結合することで意味をなすからである．

4.2 判別方法の評価

判別方法に関する評価について以下のように実施する．各告知投稿のサンプル投稿を 4.1 節で示した方法で判別を行い，判別結果より，適合率，再現率，F 値を以下により求める．

適合率（Precision）: 特徴語句を含むサンプル投稿のうち，特徴語句を含むかつカテゴリーA に該当する割合．

$$\text{適合率} = \frac{\text{特徴語句を含むかつカテゴリーAに該当するサンプル投稿の数}}{\text{特徴語句を含むサンプル投稿の数}} \quad \text{(a)}$$

再現率（Recall）: カテゴリーA に該当するサンプル投稿のうち，特徴語句を含むかつカテゴリーA に該当する割合．

$$\text{再現率} = \frac{\text{特徴語句を含むかつカテゴリーAに該当するサンプル投稿の数}}{\text{カテゴリーAに該当するサンプル投稿の数}} \quad \text{(b)}$$

F 値（F-measure）: 適合率と再現率の調和平均．

$$\text{F値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad \text{(c)}$$

判別方法の評価結果について Fig. 4 に示す．各告知投稿に対するサンプル投稿の適合率が最高となったものは，T_{A1}, T_{A3}, T_{A4}, T_{A7}, T_{A9} に対するサンプル投稿で 1.000 であった．最低となったものは，T_{A6} に対するサンプル投稿で 0.714 であった．T_{A1}～T_{A10} に対するすべての平均は 0.903 であった．各告知投稿に対するサンプル投稿の再現率が最高となったものは，T_{A9} に関するサンプル投稿で 0.800 であった．最低となったものは，T_{A4} に対するサンプル投稿で 0.192 であった．T_{A1}～T_{A10} に対するすべての平均は 0.529 であった．各告知投稿に対するサンプル投稿の F 値が最高となったものは，T_{A9} に関するサンプル投稿で 0.889 であった．最低となったものは，T_{A4} に対するサンプル投稿で 0.323 であった．T_{A1}～T_{A10} に対するすべての平均は 0.644 であった．

評価結果より，適合率が平均 0.903 となったことから，告知投稿の内容と関連したサンプル投稿を高い精度で判別できることを示した．一方で，再現率は平均 0.529，F 値は 0.644 にとどまった．本研究の目的は災害関連情報の収集であることから，告知投稿に関連する内容に該当しないサンプル投稿は含まれない（ノイズが含まれない）ことが望ましいため，精度が高い本方法が有用であると考えられる．なお，告知投稿によって，適合率が異なっているため，適合率が高い傾向にある複数の告知投稿に関するリツイート直後の投稿を収集することにより，効率的に告知投稿の内容に関連する投稿を収集できると考えられる．効率的に告知投稿に関連する内容のサンプル投稿を収集するため，どの告知投稿に注目すべきかについては今後の検討課題である．

5. まとめ

著者らは，ニュース等の投稿を行っているアカウントによって，災害発生後にニュース等が投稿（告知投稿）された際，この告知投稿をリツイートしたアカウントにおけるリツイート直後の投稿には，告知投稿に関連する情報や告知投稿の文面のみでは知りえない付加的な情報が含まれている可能性が高いと考えている．この仮説を検証するため，先行研究 [4] では 2015 年 5 月 30 日に発生した小笠原諸島西方沖地震の発生後に投稿された告知投稿に着目し，投稿の収集と分析を行った．その結果，サンプル投稿のうちおよそ 2 件に 1 件がこれに該当するという結果を得ている．本論文では，告知投稿の内容に関連する投稿（カテゴリーA に該当）を機械的に収集することを目的とし，その判別方法について検討した． そして，告知投稿をキーフレーズ解析し，さらに形態素解析を行った特徴語句を判別に用いる方法を提案した．評価として，先行研究 [4] で取り扱ったサンプル投稿を対象に，告知投稿の内容に関連する投稿の判別を行った．そして判別結果について，適合率，再現率，F 値を求めた．高い適合率で判別できることを示した．

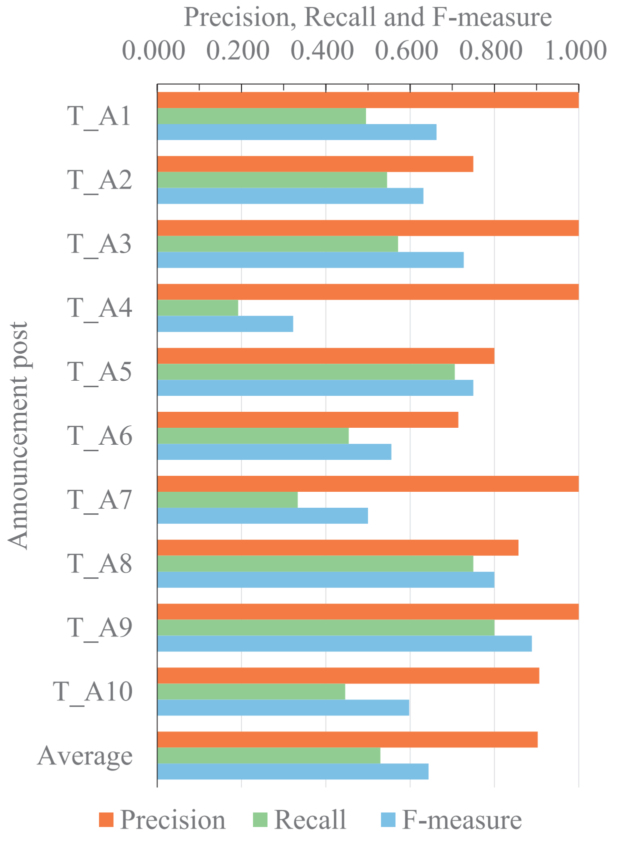

Fig. 4 Evaluation results for each announcement post

今後の検討課題は以下のとおりである．本論文では小笠原諸島西方沖地震発災後の告知投稿に着目して検討を行ったが，別の種類，規模の災害でも同様の結果が得られるか検討を行う予定である．また，効率的に告知投稿の内容に関連するサンプル投稿を収集するため，どの告知投稿に注目すべきかについて検討する．また，告知投稿の文面では知り得ない付加的な情報が含まれているサンプル投稿（カテゴリーBに該当する）を判別する方法について検討を行う．その後，自動収集システムの構築について検討を行う予定である．

謝辞

本研究は学校法人東海大学総合研究機構「研究奨励補助計画」の援助を受けて行ったものである．

参考文献

1) Twitter, http://twitter.com
2) Twitter, Incについて, https://about.twitter.com/ja/company
3) 佐々木智也，“拡大を続けるTwitterの震災における活躍と今後の展望–サービス開始から5年、コミュニケーションツールから社会インフラへ–”，AD STUDIES Vol.26 pp.20-24，http://www.yhmf.jp/pdf/activity/adstudies/vol_36_01_04.pdf
4) 児玉史緒里，眞中絢美，内田理，石井啓之，宇津圭祐，“告知投稿のリツイート直後のツイートに着目した災害関連情報の収集”，電気学会資料通信研究会 CMN-16-004，PP.19-23，2016年1月
5) 塚本悠馬，笹野遼平，高村大也，奥村学，“マイクロブログ上の告知投稿に対する非明示的な関連投稿の収集”，情報処理学会研究報告，Vol.2013-NL-214 No.14, pp.1-8, 2013
6) 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
7) 高畑洋貴，六瀬聡宏，榎本光，齋藤大樹，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指示情報の可視化”，言語処理学会第20回年次大会 発表論文集 pp.82-84，2014
8) 馴田俊平，六瀬聡宏，榎本光，齋藤大樹，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
9) 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, 2015
10) R. Kitajima, R. Kamimura, O. Uchida, F. Toriumi:oriumi:ida, F. Toriumi:r: “Classification and Mapping of Disaster Relevant Tweets for Providing Useful Infor on Soft Computing and Pattern Recognition, 2015
11) https://retweets.azurewebsites.net/?lang=ja
12) 無料WEB便利ツール：キーフレーズ抽出，http://tools.metro-bb.com/api/keyphrase
13) 形態素解析ウェブアプリ UniDic-MeCab, http://www4414uj.sakura.ne.jp/Yasanichi1/unicheck/