

複数のモダリティを考慮した
災害関連ツイートのノイズ除去に
関する研究

総合情報学研究科
知識情報学専攻

インタラクションデザインの理論と実践

20M7120

森野 穰

論文要旨

1 はじめに

災害が発生した際、ソーシャルメディアには様々な災害情報が投稿される。特に Twitter 上ではその速報性と拡散力の強さから、自身の被害状況や救助・支援を求める投稿が見られる (図 1 参照)。そういった投稿を自治体は独自に収集することで、実際に被災者の救援に繋げていた。一方で、災害時の Twitter には、テレビ報道の拡散や個人の感想、デマなどへの反応といった被害情報や救援要請に直接関わらない投稿が多く含まれているため、被害情報や救援情報に関わるツイートを的確に抽出することは困難であると指摘されている [1]。Twitter に投稿された災害時の被害情報・救援要請情報を活用するためには、投稿情報の信頼性を担保する必要がある、最終的には人手による投稿の確認が必須となっている。災害時という人的リソースが欠如した状態で Twitter に投稿された災害情報を効率的に確認し活用するためには、膨大な SNS 情報の中から災害被害とは無関係な情報をノイズとして扱うことで取り除き、確認すべき情報の絶対量を減らすことが望ましい。そこで本研究は、SNS 情報の肥大化に焦点を当て、膨大な情報の中から必要な情報をピンポイントで抽出するというアプローチではなく、データをクレンジングすることで必要性が高いと推定される情報を残すというアプローチを採用。その端緒として、Twitter の災害関連単語を含む投稿から災害被害に関する投稿と災害被害とは無関係な投稿に関する調査を行った。調査では、投稿されたテキスト情報と画像という複数のモダリティに着目し、災害時に投稿されるツイートのうち画像を伴うツイートから、画像の内容にもとづいて人手でクラス分類したうえで、各クラスのツイートの文章特徴を明らかにする。これにより、Twitter に投稿された災害情報を効率的に取得するためのデータクレンジングの可能性について考察する。

2 収集データの分類と調査方針

本稿では令和 2 年 7 月豪雨を対象にデータ収集を行った。対象期間は令和 2 年 7 月 1 日～15 日とし、Twitter API を用いてキーワードをクエリとした。キーワード“救助”で得られたツイートは 110,261 件、“避難”で得られたツイートは 370,531 件となった。このツイートの中から何らかの画像・動画が含まれているツイートは、“救助”ツイートには 18,197 件、“避難”ツイートには 34,777 件存在した。調査の前提として、災害時の Twitter に投稿された画像および動画を人手で参照し、どのようなものが写っている画像か調査し分類する。人手によって分類された画像から機械的に求めたテキスト特徴を抽出することで、ノイズとなる投稿の抽出が可能になるか調査を行う。テキスト特徴の抽出は分類クラス別で TF-IDF 値の算出を行い、分類クラス別の画像群を代表する特徴的な単語の抽出を行う。その後、算出した TF-IDF 値をもとに文章間の類似度を示すコサイン類似度を算出すること分類に紐付いたツイート群間の類似性を調査する (ステップ 1)。次に、特に“被害”に分類される画像群はステップ 1 では分類しきれない画像と考えられるため、人手で判断した被害画像に分類した画像群をさらに細分化し、どのような特徴を持つのか調査を行う (ステップ 2)。

3 調査結果

ステップ 1 の調査の結果、例えば“ゲーム”に分類された画像群では、特定のゲームの固有表現単語が TF-IDF 値上位に算出された。“TV ニュース”で上位に算出された単語は“News23”で報道された単語が大部分を占め、繰り返しツイートされていたことが確認できた。付随した画像は同様に“News23”で報道されたニュースの切り抜き画像であった。コサイン類似度が高かった語



図 1: 被害報告ツイートの一例

表 1: “被害” ツイート数とその割合（一部抜粋）

| | ツイート数 | 被害画像ツイート割合 (%) | 画像ツイート割合 (%) | ツイート総数割合 (%) |
|-------------|-------|----------------|--------------|--------------|
| 被害-自治体発信 | 95 | 2.297% | 0.200% | 0.020% |
| 被害-分類不可能 | 153 | 3.700% | 0.323% | 0.032% |
| 被害-元ツイート消去 | 101 | 2.443% | 0.213% | 0.021% |
| 被害-テレビニュース | 66 | 1.596% | 0.139% | 0.014% |
| 被害-ネットニュース | 204 | 4.933% | 0.430% | 0.043% |
| 被害-ユーザ撮影 | 2689 | 65.030% | 5.669% | 0.564% |
| 被害-ライブカメラ撮影 | 209 | 5.054% | 0.441% | 0.044% |

群として、“救助”と“自衛隊”，“地図”と“天気図”と“天気”などがあつた。また，分類した画像クラスの文章間で類似度が高かつた例は，“テレビニュース”と“被害”，“避難”が挙げられた。この結果は，一般的な報道と被害情報や避難情報が混交する可能性が考えられる。特に“被害”は複数のクラスと類似度が高く算出され，災害情報全体を象徴する語が算出されている可能性が示された。

ステップ2の調査の結果，“被害”に分類されたツイートは，“テレビのニュースで報道された被害の画像”，“ネットニュースで報道された被害の画像”，“ユーザが撮影した被害の画像”，“ライブカメラで撮影された被害の画像”，“自治体が発信した被害の画像”，“分類不可能だった被害の画像”，“元ツイートが消されていた被害の画像”などに細分化された（表1参照）。“テレビのニュースで報道された被害の画像”で上位に算出された単語の“園”，“寿”，“千寿”はいずれも熊本県球磨郡の特別養護老人ホーム“千寿園”に関するニュースであつた。自治体や政治家によって投稿されるツイートには“豪雨”や“雨”などによる被害を表す単語に加えて，“災害”に伴う自衛隊の“派遣”状況を伝える単語が確認された。被災者の救助要請や被害の状況報告が最も反映されていると想定される“ユーザが撮影した被害の画像”のTF-IDFにおいて，特出した値や単語は認められなかつた。

4 おわりに

“ゲーム”や“テレビのニュース”などに分類されたツイートは単純なテキスト処理を行うだけでそのテキスト特徴を抽出することが可能であり，ノイズとして扱うべき情報であるといえる。災害時に必要になるであろう“被害”に分類された画像に対して元ツイートのテキストを参照することで，画像とツイートの詳細な状況などを確認するといった質的調査を行った結果，“被害”画像をさらに細分化して分類できることが確認できた。一方で，被災者の救助要請や被害の状況報告が最も反映されていると想定される“ユーザが撮影した被害の画像”のTF-IDFにおいて，特出した値や単語は認められなかつた。災害時におけるTwitter上の画像を人手による分類を行ったことで，害被害や救助要請とは無関係なツイートが多く存在し，無関係なツイートを可能な限りフィルタリングできた場合，全体の0.564%ほどのツイートまで情報削減できることが明らかとなつた。

参考文献

- [1] 藤代 裕之, 松下 光範, 小笠原 盛浩: 大規模災害時におけるソーシャルメディアの活用—情報トリアージの適用可能性, 社会情報学, Vol. 6, No. 2, pp. 49-63 (2018).

目次

| | | |
|----------|------------------------------------|-----------|
| 1 | 序論 | 1 |
| 1.1 | 本研究の背景 | 1 |
| 1.2 | 災害時における SNS 利用状況と実例 | 1 |
| 1.3 | 災害時における SNS 利用の課題 | 4 |
| 1.4 | 本研究の目的 | 5 |
| 2 | Twitter における救助要請・被災情報抽出 | 6 |
| 2.1 | 災害時の Twitter と課題 | 6 |
| 2.2 | Twitter を利用した理想的な災害情報へのアクセス | 7 |
| 3 | 関連研究 | 8 |
| 3.1 | Twitter をソーシャルセンサとして扱う研究 | 8 |
| 3.2 | Twitter 情報の分類を行った研究 | 8 |
| 3.3 | Twitter の位置情報を利用した研究 | 9 |
| 3.4 | Twitter のユーザに着目した研究 | 9 |
| 3.5 | 災害情報収集システムに関する研究 | 10 |
| 3.6 | 複数のモダリティに着目した研究 | 11 |
| 4 | ツイートデータにおける各モダリティ特徴に関する調査方針 | 13 |
| 4.1 | 調査の概要 | 13 |
| 4.2 | 調査手続き | 13 |
| 5 | 対象の災害と収集データ | 15 |
| 5.1 | 令和2年7月豪雨 | 15 |
| 5.2 | 収集したツイートデータ | 16 |
| 6 | ノイズとなる投稿に関する調査 | 18 |
| 6.1 | 投稿画像の分類手順 | 18 |
| 6.2 | 投稿画像の分類結果 | 20 |
| 6.3 | テキスト処理 | 20 |
| 6.4 | コサイン類似度の算出 | 22 |
| 7 | 被害投稿の特徴に関する調査 | 24 |
| 7.1 | 被害画像の分類と細分化 | 24 |
| 7.2 | 細分化された分類のテキスト特徴 | 26 |
| 7.3 | 細分化により得られた知見 | 26 |
| 7.4 | フィルタリング後の想定ツイート数 | 28 |
| 8 | 議論 | 29 |

1 序論

本研究では、自治体によるソーシャルメディアを利用した災害時の救助・被害の円滑な情報収集活動を支援するために、被害情報とは無関係なデータを除去するデータクレンジングについて検討する。本章では本研究に至った背景や災害時におけるソーシャルメディアの活用例を挙げ、ソーシャルメディア活用の課題を概観する。

1.1 本研究の背景

東日本大震災の際、多くの被災者がソーシャルメディアに自身の被災状況を発信したことが知られている [1][12]。そのソーシャルメディアの投稿が救助活動につながった事例に注目が集まった。こうした事例が広く知られるようになったこともあり、災害時に自身の被害状況を報告する投稿や、救助や支援を求める投稿をソーシャルメディアで発信するユーザーも増加している¹。ソーシャルメディアの一種であるソーシャル・ネットワーキング・サービス (以下、SNS と記す) を災害時に利用する人の増加に伴い、自治体の対応も変化しつつある。内閣官房情報通信技術総合戦略室の 2019 年度における災害対応における SNS 活用状況に関する自治体調査によると、1,145 もの市町村が SNS を活用した情報発信を行っている。同様に、計 958 市町村が SNS を活用した情報収集に関心を持っており、既に 107 の市町村が実災害下で SNS を運用し、情報収集を行っている [30]。一般に広く利用されている SNS として、Facebook²、Line³、Twitter⁴、Instagram⁵などが挙げられる。自治体が情報発信に利用した SNS のうち最も利用市町村が多い SNS は Facebook となり 918 市町村、次点が Twitter となり 696 市町村、続いて Line となり 185 市町村、複数利用した市町村は 559 となっている。自治体が情報収集に利用した SNS のうち最も利用市町村が多い SNS は Line となり 68 市町村、次点が Twitter となり 64 市町村、続いて Facebook となり 43 市町村、その他 SNS を利用した市町村は 24 となっている (重複利用あり)。

1.2 災害時における SNS 利用状況と実例

本節では、代表的な SNS を例に、災害時において SNS がどのように利用されているか概観する。

1.2.1 災害時における Facebook 利用の実例

Facebook は実名での利用が強く推奨されている SNS の 1 つである。ユーザーによる投稿はその実名性から一定の信頼性が担保されるが、他の SNS と比べてクローズドなコミュニティが形成されるという特徴がある。

自治体による Facebook 利用の実例としては、避難指示、避難所開設や給水などの被災者に直結する災害関連情報の発信などが挙げられる。大分県日田市は平成 28 年の熊本地震発

¹<https://www.nhk.or.jp/politics/articles/feature/10863.html> (2021/12/18 確認)

²<https://www.facebook.com/> (2021/12/18 確認)

³<https://twitter.com/> (2021/12/18 確認)

⁴<https://twitter.com/> (2021/12/18 確認)

⁵<https://www.instagram.com/> (2021/12/18 確認)

【平常時】



【熊本地震発生時】



図 1.1: 大分県日田市 Facebook トップページデザインのデザイン変更 (文献 [29] より図引用)



図 1.2: Facebook の災害情報時センター (MetaPlatform 社のウェブサイト⁸より図引用)

生⁶Facebook の自治体の公式アカウントのページデザインを変更し、地震関連情報の発信に利用した [29]。Facebook のトップページの変更例を図 1.1 に示す。

Facebook は Meta Platforms 社⁷ (2021 年に社名を Facebook から Meta Platform に変更) によって運営され、独自に災害時情報センター⁸を設けている。この災害時情報センターではユーザが自身の安否、Facebook 上でつながっている友達の安否の報告が可能な他、災害の影響を受けた地域にいる人の安否が確認可能である。携帯端末で災害情報センターを利用した際の画面を図 1.2 に示す。

1.2.2 災害時における LINE 利用の実例

日本で最も普及している SNS である LINE は Facebook や Twitter と比べて、1 対 1 での双方向の情報伝達に長けている。そのため、家族や友人間での安否のやり取りなどが行わ

⁶<https://www.jma.go.jp/jma/press/1605/14a/kaisetsu201605141000.pdf> (2022/2/8 確認)

⁷<https://about.facebook.com/ja/> (2022/2/8 確認)

⁸<https://about.fb.com/ja/news/2014/10/safetycheck/> (2022/2/9 確認)



図 1.3: 千葉県による LINE 運用（防災・減災における Line 活用⁸より図引用）

れ、Facebook と同様にクローズドなコミュニティといえる。

自治体による Facebook 利用の実例としては、自治体の公式アカウントによる被災者への情報発信や、LINE の基本的なメッセージ機能を用いたユーザ対チャットボットによる双方向のコミュニケーションが挙げられる。令和元年台風 15 号の際、千葉県は実際に AI チャットボット（現在はサービス運用を終了）を活用した情報発信を行った⁹。実際に千葉県によって運用された Line の画面を図 1.3 に示す。当サービスは被災者からの問い合わせに対応する形で情報提供を行っていた。

1.2.3 災害時における Twitter 利用の実例

Twitter は匿名で利用されることが多い SNS である。1 対多のコミュニケーションが行われ、Facebook や Line と比べるとオープンなコミュニティが形成される。自治体による Twitter 利用の実例としては、公式アカウントによる情報発信¹⁰やユーザと実際にやり取りすることで被害状況を把握していたことが挙げられる。自治体による Twitter による情報発信例を図 1.4 に示す。

東日本大震災の際に投稿された一件のツイートが東京都知事の目に留まったことから、自衛隊が派遣され救助が行われた。2019 年台風 19 号の影響で長野県の千曲川が氾濫した際

⁹<https://linegov.com/service/disaster.php> (2022/2/8 確認)

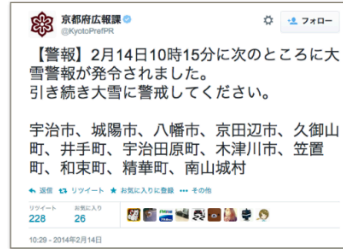
¹⁰https://www.kantei.go.jp/jp/singi/it2/senmon_bunka/bousai_SNS_kentoukai/dai2/shiryo_3_5.pdf (2022/2/9 確認)

1. 防災情報

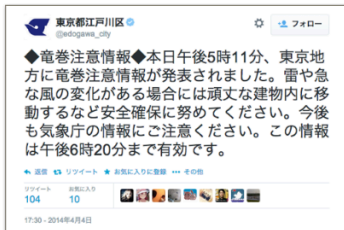
地震



大雨、大雪などの気象警報



竜巻

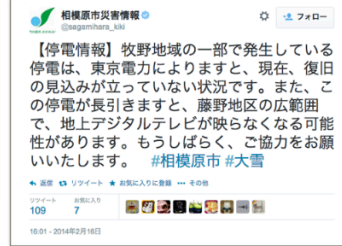


2. 交通情報



3. ユーティリティ関連の障害情報

停電



通信障害

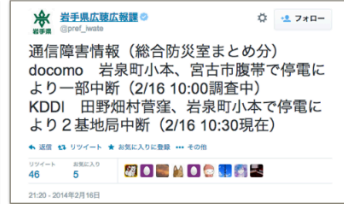


図 1.4: 自治体による Twitter による情報発信 (自治体における災害時の情報発信と収集に向けて⁸より図引用)

に、長野県は独自に Twitter に投稿された救助要請を収集し約 50 件の救助につなげていたことが自治体の Twitter 活用の具体例として挙げられる¹¹。

1.3 災害時における SNS 利用の課題

SNS を活用した情報発信や情報収集が盛んに行われるようになり、被災者の被害状況や救助要請が把握可能になった一方で、災害時に SNS から情報収集を行うことには課題もある。例えば、デマ¹²情報の拡散はその課題の 1 つである。東日本大震災では「コスモ石油のガスタンクが爆発した直後に有害物質を含んだ雨が降る」という間違った情報が拡散 [20] され、2016 年熊本地震の際 Twitter には「おいふざけんな、地震のせいであちの近くの動物園からライオン放たれたんだが熊本」という投稿や「イオンモール熊本で火災が発生した」という投稿が拡散された。こういったデマの拡散について、災害時という「不安を引き起こす社会的状況」が影響し [37]、「ほとんどのユーザがツイートによるデマ発生当初はそのツイートが間違っていることには気づかずに善意で拡散してしまう」と指摘されている [36][34]。デマの投稿や被災地以外からの投稿や拡散、SNS 上の情報に対する正誤判断の難しさが浮き彫りとなっている。

さらに、特に災害時の SNS 上では災害に関する情報共有や拡散が行われることで情報のバースト現象が発生し [5][11]、一般ユーザによるマスメディア報道の再発信などが引き起

¹¹<https://www.nhk.or.jp/politics/articles/lastweek/25652.html> (2021/12/18 確認)

¹²“デマ”は政治的な目的で意図的に流す虚偽の情報のことではあるが、災害時の SNS に関しては誤った情報についてデマとの言葉が使われており本稿でも同様とする。<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/html/nc143c00.html> (2021/12/18 確認)

こす情報の埋没 [35][9] が問題として顕在化している。時間的・人的リソースが限られた状況下において、バースト現象が発生した SNS の膨大な情報を吟味し、埋没している確度の高い情報のみを収集するには、多大な人的労力が必要となる。そのため、これらの情報を活用した救助活動を円滑に行うためには、機械的に情報を選別することでフィルタリングを行い、効率的な情報収集を可能にする仕組みが必要である [26]。

上記の 2019 年台風 19 号の事例では、専属の自治体職員がかかりきりで情報収集を行い、直接被災者とやり取りする必要があるなど人的な負担が大きい。長野県危機管理防災課は「より大きな水害や地震の際には、被害が広範囲に及ぶ中で要請の情報をどう正確に収集するのか課題も多いので、今後、ツイッターの活用をしっかりと議論していきたい」と話している⁹。

東日本大震災を機に、NHK は Social Listning Team (SoLT: ソルト) と呼ばれる SNS の情報収集・分析専門のチームを立ち上げ、Twitter などのソーシャルメディアから災害情報や事故情報などの収集を行い報道につなげているが、独自に収集する SNS の情報は「内容の信憑性や重要度を機械的に判断するには精度の問題がある。結局は大量のツイートを機械的に「ふるい」にかけたうえで、人が見た目で判断するシンプルな方式が最も効率的だ」とし、ソルトの活動を次の大災害時に活かすためには「スタッフや機材をさらに増やし、検索と分析のノウハウを蓄積する必要がある」と述べている [13][14]。

1.4 本研究の目的

上述のように、災害時に本当に必要な情報を収集するためには、情報選別を機械的に行ったのちに選別された情報を人手で確認し、投稿されている情報の信頼性を担保することが必要不可欠である。ここで、機械的な情報選別の際に留意しなければならないのは、どれだけ機械的に情報を抽出しても、最終的にその情報を確認するのは人であるという点である。そのため、人が判断する以前の機械的な情報抽出の際において、被災情報の「取りこぼし」があってはならない。災害情報を活用する上で最も必要なシステムの要件は、「提示されている情報の量に問わず、必要な情報が遺漏なく存在すること」であると言える。そのため本研究では、SNS 情報の肥大化に焦点を当て、膨大な情報の中から必要な情報をピンポイントで抽出するというアプローチではなく、データをクレンジングすることで必要性が高いと推定される情報を残すというアプローチを採る。適切なデータクレンジングが可能になれば、扱う SNS 情報の絶対量が減ることで救助要請や被害情報などの必要性が高い情報の抽出精度が上がるが見込まれる。このようなシステムを実現する端緒として、SNS の 1 つである Twitter の災害関連単語を含む投稿から災害被害に関する投稿と災害被害とは無関係な投稿に関する調査を行った。調査では、画像とテキストという複数のモダリティに着目し、災害時に投稿されるツイートのうち画像を伴うツイートから、画像の内容にもとづいてクラス分類したうえで、各クラスのツイートの文章特徴を明らかにし、それにもとづいて災害と無関係なツイートを除外できるか検討する。

2 Twitterにおける救助要請・被災情報抽出

本研究では、情報収集の対象としてTwitterに着目する。Twitterは、上述したように匿名性の高いSNSとして知られ、日本国内では2017年の時点で4500万人ものユーザが利用している¹。一般にツイートと呼ばれる投稿には140字の上限があり、情報の送受信が容易であるため速報性が高いという特徴がある。また、ボタン1つで他人の投稿を再発信できるリツイートという機能があるため、気軽に拡散される傾向がある。ツイートには文字に加えて合計4枚の画像もしくは動画を添付することができる。

2.1 災害時のTwitterと課題

Twitterの拡散力や速報性が高いという特徴から、Twitterは災害時の救助・被害情報共有に利用されることが期待されている。Twitter Japan社²も災害時の利用を想定した情報発信の指針を示している。災害時の救援要請ツイートの投稿する際には「#救助」ハッシュタグとともに要請内容、写真、住所または位置情報など、具体的かつ正確な情報を付けること、及び救助が完了したら報告ツイートをするとともに救助要請ツイートを削除することを推奨している[27]。TwitterJapan社による指針の発信例を図2.1に示す。

また、Twitter Japan社はウェザーニューズ社³と連携し、被害状況を知らせる際は「#減災レポート」という専用のハッシュタグを投稿本文中に含め、撮影した写真と位置情報を追加して投稿するよう呼びかけている。災害時に特に有効であるとされる位置情報の追加は、メディアリテラシーの観点などからユーザ自ら位置情報の取得を拒否している場合にはあらかじめ位置情報の取得を許可するプロセスが必要になる。加えて、「#救助」ハッシュタグが付与されたツイートや位置情報が追加されているツイートには、災害被害や救助を求める声とは無関係な情報も混在し、その絶対数も少ない。そのため、これらの情報のみを用いて情報を収集することは現実的ではない。

救助を求める際の投稿と被害を報告するような投稿のどちらにも共通するのは具体的な要請内容と写真が追加される傾向があることである。しかし、災害時のTwitterに投稿される画像は、被災したユーザ個人の被災状況の発信や救助を求めるための写真だけでなく、スマートフォンのカメラを用いて撮影されたテレビ画面やニュース映像の切り抜きなどが添付され、メディアの報道の総括や感想が述べられる場合が多い。また、被害とは無関係な画像が投稿される場合も多く存在する。そのため、自治体が救助要請となる投稿や災害被害に関する画像が添付された投稿を収集する際は、こういった投稿がノイズとなりえる。また、先述した文字数制限により計算機による単純なテキスト分類は困難となっており、投稿される画像はその雑多性から計算機による単純な画像内の物体検出や画像の分類は困難である。

¹<https://twitter.com/TwitterJP/status/923671036758958080> (2020/12/18 確認)

²<https://about.twitter.com/en/who-we-are/our-company> (2020/12/18 確認)

³<https://jp.weathernews.com/> (2020/12/18 確認)

救援要請ツイートの例



- ① 救援要請の内容を具体的に書きましょう
- ② 救助を待っている場所の正確な住所がわかる場合は具体的に書きましょう。
- ③ #救助 ハッシュタグをつけましょう。
- ④ 写真を添えて状況がわかるようにしましょう。
可能な際には「Twitterカメラ」を利用して写真を撮ると正確な位置情報が追加できます。今いる場所の正確な住所がわからないときにぜひ活用しましょう。

注意：救助された後は混乱を避けるためにもツイートを削除してください。

図 2.1: TwitterJapan 社が推奨する救援要請ツイートの例（文献 [27] より図引用）

2.2 Twitter を利用した理想的な災害情報へのアクセス

理想的な災害情報へのアクセスには、ハッシュタグや位置情報の有無にかかわらずユーザ個人の被災状況を迅速に把握できることが必要になる。これらの情報に迅速にアクセスするには、システムが機械的に無関係な情報を除き、それで得られた信頼性の高い情報を人手で確認することにより各個人の状況を判断することが必要不可欠になる。人が被害と無関係な投稿を判断する際、投稿されるテキストおよび投稿される画像を複合的に判断する。機械的に被害と無関係な投稿を検する際にも、これらのテキストおよび画像の双方に着目することで、より多くの無関係な投稿の排除を行うことが可能になる。そのため、本稿では、SNS 上の災害被害情報と誤認されるような投稿の分析を行い、ノイズとして扱うべきデータを明らかにすることでデータのクレンジングを試みる。

3 関連研究

本章では、Twitter に投稿された災害情報を扱った先行研究の例を挙げ、災害時における Twitter 利用の可能性とその問題について述べる。

3.1 Twitter をソーシャルセンサとして扱う研究

ソーシャルメディアを、実世界を観測するためのセンサのように利用するための研究が行われている。例えば、Kwak らは SNS の情報伝播とその拡散に関する定量的な調査 [6] を行い、McCormick らは SNS で得られる情報を用いて逆説的に人口統計情報の推定 [7] を行っている。Aramak らは Twitter の投稿からインフルエンザの流行を可視化し [2] 実世界で観測された事象と Twitter 空間で観測された事象とのギャップの調査 [15] を行った。Sakaki らは Twitter 情報を用いることで地震の発生を観測したり [10]、國友らは土砂災害の前兆・発生状況把握を試みる [21] など、ソーシャルメディアの特性を活かすことで「今実世界で何が起きているのか」が観測可能であることを示している。榊らは、ソーシャルメディアを情報収集にすることの意義を、ユーザが非常に多く、観測可能な事象に対する柔軟性が高いことであるとし、センサ自体が人間のため高度な知的処理が可能である反面、信頼性・可用性を得ることは難しいと述べている [23]。この問題を解決するためには、イベントやユーザの状態をモデル化し、必要なツイートを選択的に抽出するといった工夫をするなど、ユーザの投稿する文章や画像を分析することが必要不可欠となる。

3.2 Twitter 情報の分類を行った研究

佐藤らは、Twitter Japan 社が発信した救助要請を行う際の記載方法の一つである「#救助」というハッシュタグが付与されたツイートに着目した分類を人手で行った [24]。2018 年西日本豪雨災害と 2017 年 7 月九州北部豪雨災害のツイートを比較することで実際にツイートされる「#救助」を用いた救助要請ツイートの実態を定量的に明らかにした。その結果、2つの比較対象災害ツイートにおいて実際に救助要請をしていたツイートはわずかであり、ニュース記事や被災者以外からの善意による投稿が多くを占めていた。実際に「#救助」が付与されたツイートの内訳を図 3.1 に示す。

こういった文章の分類を機械的に行う方法も研究されている。橋田らは、ある画像に写っているものを機械的に検出する際に一般に利用される技術である畳み込みニューラルネットワークをテキスト分類に適用することで、観光に関するツイートの分類を行った [32]。投稿された文章に観光地に対する意見が含まれるか、含まれないかの判定を行い、ポジティブネガティブ判定をすることで分類を行っている。鳥海らは、災害時にも頻繁に発生するソーシャルメディアのバースト現象に着目した分類を行った [28]。この研究では投稿が急激に増えるバーストが発生した際、どのようなトピックが含まれているかのトピック分類や、どのようなユーザによって拡散されたのかのコミュニティ分類を行うことで、バースト現象の詳細を明らかにしている。

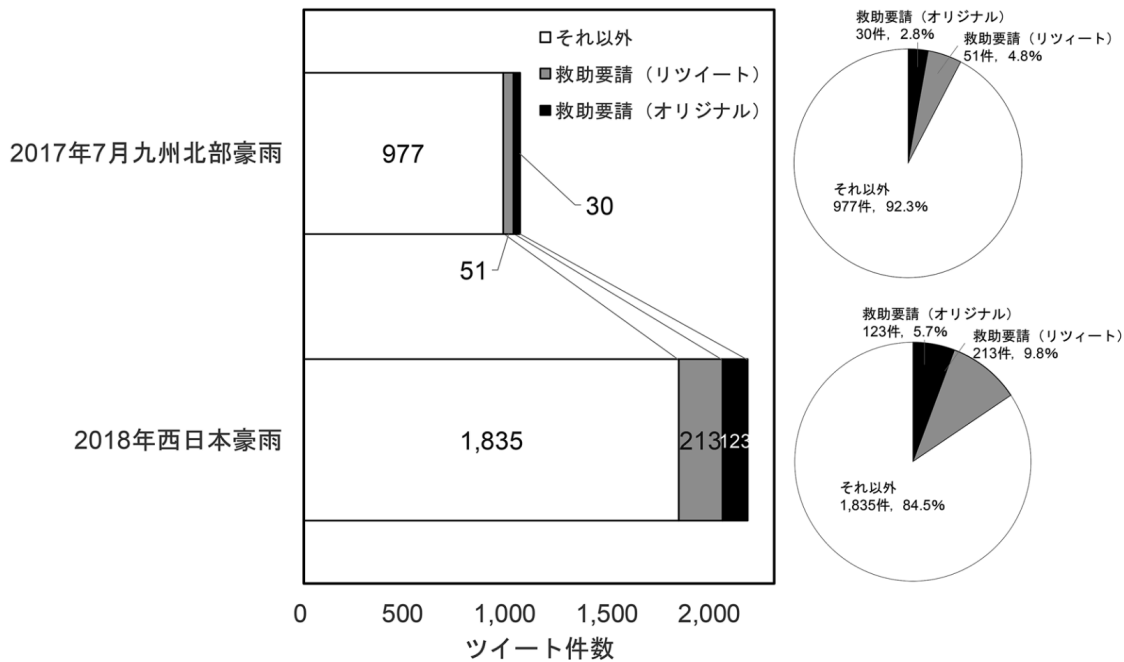


図 3.1: 「# 救助」ツイートの内訳 (大ラベル) (2017 年 7 月九州北部豪雨災害と 2018 年西日本豪雨災害) (文献 [24] より図引用)

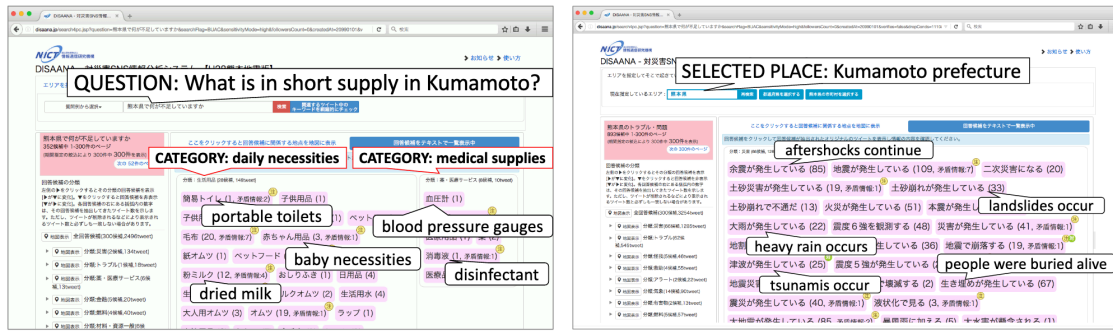
3.3 Twitter の位置情報を利用した研究

Twitter は、位置情報を追加してツイートすることができるため災害時に特に有効であるとされている。しかし、メディアリテラシーの観点などからユーザが自らの位置情報付与を拒否している場合は、予め設定から位置情報を許可するプロセスが必要になる。災害時においても、ユーザの位置情報付きのツイートは少なく、わずか 0.18% ほどとなっている [33]。そこで森國らはツイートの投稿位置を位置情報を利用せずに推定するため、ツイート中のノイズとなる単語のフィルタリング手法を提案している [38]。元々位置情報が付与されていたツイートをエリア別での単語出現頻度を学習し、位置情報が付与されていないツイートの投稿位置の推定を行っている。

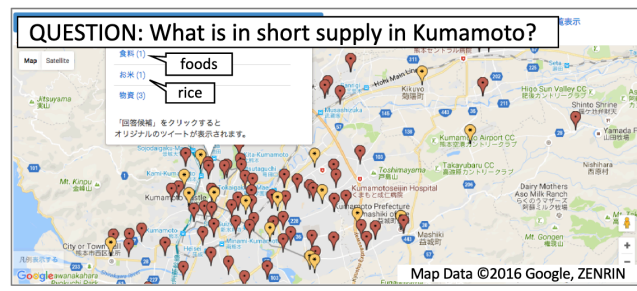
3.4 Twitter のユーザに着目した研究

Twitter のユーザに着目した研究も行われている。蔵内らは、投稿される文章には多分にノイズが含まれていることを前提として、ユーザのデモグラフィック属性を推定する手法を提案した [22]。Eisenstein らは各ユーザに着目し、投稿の記述から地名などの地理的特徴を持つ単語からユーザの位置情報を生成する手法を提案している [3]。伊藤らはユーザの Twitter 上のプロフィール文章とツイート文章と同一人物の Blog 上のプロフィールから、性別や年齢、職業などのユーザ属性を推定するための推定機を構築した [17]。

匿名性の強いソーシャルメディアのユーザのデモグラフィックが高精度で推定可能にな



(a) Results from the QA mode. (b) Results from the problem-listing mode.



(c) Results from the QA mode located on a map.

図 3.2: Example screenshots of DISAANA (文献 [8] より図引用)

れば、ユーザによる位置情報の記載がなくとも被災位置の特定や、災害時のデマ伝播のメカニズム解析やデマの抑制などに利用することも考えられる。

3.5 災害情報収集システムに関する研究

総務省所管の国立研究開発法人情報通信研究機構 (NICT) は災害情報収集システム「DISAANA」を開発している [19][8]。DISAANA は Twitter の投稿をリアルタイムで分析することで、どこで何が起きているのかを抽出するシステムである。このシステムは 5W1H を含む問いかけ文によるキーワード検索機能や、地点をクエリとした情報収集機能を提供しており、災害発生時の情報ソースとして SNS を活用することで、時々刻々と変化する状況に対して速報性の高い情報を提供できるという特徴を持つ。しかし、Twitter には匿名アカウントで何でも発信できるという特性があるため、散見される無責任なデマや無関係な情報のフィルタリングは困難である。この問題に対して、DISAANA は提示する情報と反対の趣旨の情報を提示することで、システム利用者自身に確認を促している。NICT はこのシステムの問題として「対災害 SNS 情報分析システム DISAANA は、入力された質問等に対して、質問中の語句やその同義と見なせる表現で Twitter 情報 (日本語による全ての投稿の 10% 程度) を検索し、機械的に回答候補を抽出します。従いまして、そもそも不正確な Twitter 情報や、自動分析の誤りによって、デマや誤り、その他不適切な回答候補が出力される場合もございます。」¹としている。

¹<https://disaana.jp/rtime/disclaimer4pc.jsp?ref=1> (2022/2/7 確認)

3.6 複数のモダリティに着目した研究

本研究では、画像とそれに対応したテキストを併用することで、救援情報の抽出精度を向上させることを目指している。このような、異なるモダリティのデータを併用する試みの類例として、石橋は病理情報を対象としたデータ検索システムを提案している [16]。このシステムは人手でラベリングした病理標本画像を特徴量化し、類似画像を認識している。病理診断書から得られたテキストとそれに付随する画像を紐付け、紐付けられた画像の特徴量を利用することで、対象の病理画像と病理情報の検索を可能にするデータベースを構築している。石橋らによる統計データ解析の手順フローを図 3.3 に示す。

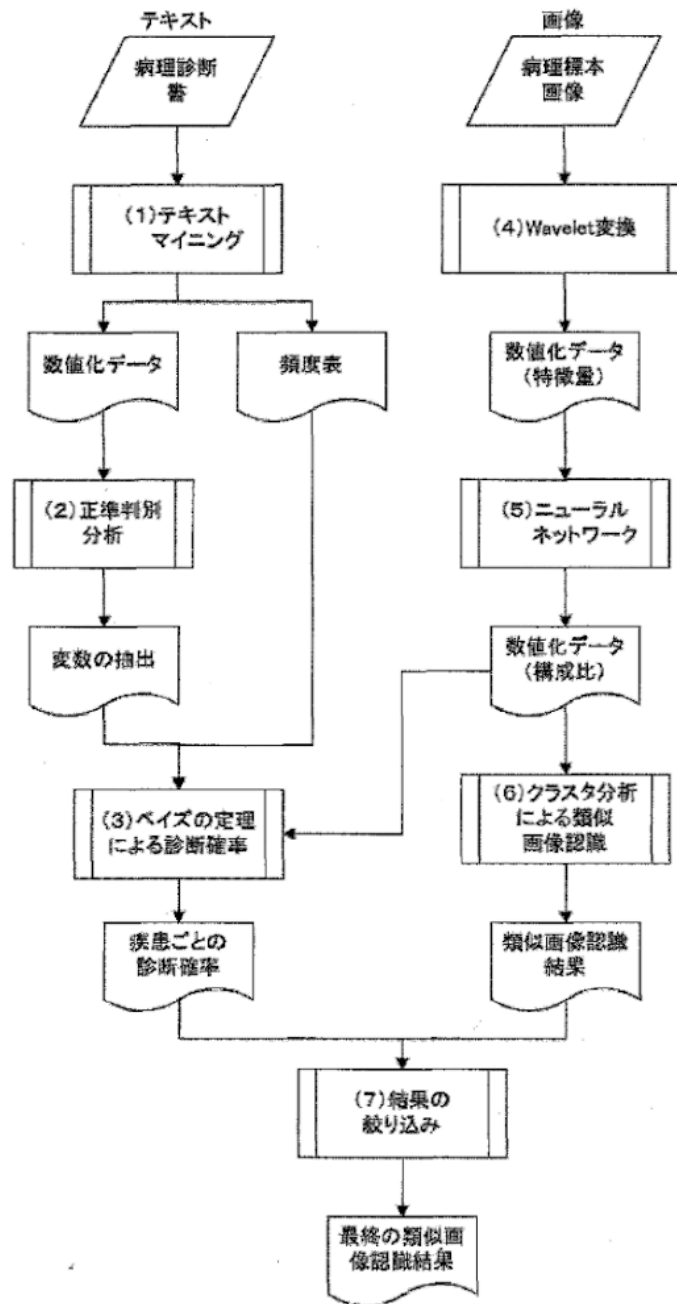


図 3.3: 石橋らによる統計的データ解析の手順 (文献 [16] より図引用)

4 ツイートデータにおける各モダリティ特徴に関する調査方針

4.1 調査の概要

Twitterには画像や映像、音声、テキストなど複数のモダリティの情報が含まれている。常時であれば、こういった複数のモダリティが存在することで、ユーザ間のコミュニケーションが促されSNSの発展につながるが、災害時にはこれらの膨大な情報がノイズとして認識される場合がある。これらのツイートデータを災害情報の収集源として扱うためには、文章トピックやユーザ属性、画像の分類や位置情報などを多面的に考慮していく必要がある。本研究は、複数のモダリティを含む情報を対象に、モダリティごとの特性を相補的に利用することで、効率的に救援情報や被害情報を抽出することを志向している。そのためには、多面的な情報を入力とし、組み合わせるマルチモーダルな技術が必要になってくる¹。マルチモーダルな技術を扱うためには、それぞれのモダリティを含む情報を各々定量的に特徴化する必要がある。Twitterに投稿された画像の特徴とテキストの特徴を定量的に特徴量化するためには、各モダリティの情報が分類されている必要がある。前提として、Twitterには災害や被害とは無関係の画像が多く投稿されていることが想定される。そのため、収集したツイートにどのような画像が添付されているかの調査を行い、その特徴を明らかにする必要がある。その上で、それらの画像はどのような文章と同時に投稿されているかを調査し、各投稿の文章特徴を明らかにすることで、どのような情報として扱うべきか考察する。

4.2 調査手続き

まず、災害時のTwitterに投稿された画像および動画の調査を行う。全ての画像・動画を人手で参照し、どのようなものが写っている画像か調査し分類する。次に、先で行った画像の分類をもとに、画像が添付された投稿の文章特徴の調査を行う。文章の特徴は、ある文中におけるその単語の重要度を示すTF-IDF(Term Frequency-Inverse Document Frequency)[4]値を算出する。TF-IDF値を算出することで、分類別の画像群を代表する特徴的な単語の抽出が可能になる。ここで、本稿の分類手法では、複数のクラスに対して同一IDのツイートが分類されるケースがある。これは、同一画像が複数のクラスに分類される場合、同一ツイート内で複数の画像が添付され各々が違うクラスへの分類が見込まれる場合に発生する。これらの同一IDのツイートがTF-IDF値算出の際に影響を与えている可能性を考慮し、最後に画像の分類に紐づいたツイート群間の類似性について調査する。分類に紐づいたツイート群間の類似性は、算出したTF-IDF値をもとに文章間の類似度を示すコサイン類似度を算出することで調査する。TF-IDFによって算出した特徴的な単語が、分類した画像群間でどの程度類似しているか明らかになることが期待される。

ただし、ノイズとなる投稿に関する調査で行うコサイン類似度算出の懸念として、特に“被害”に分類された画像群は、他クラスと混交している可能性がある。そのため、人の判

¹<https://www.nttdata.com/jp/ja/data-insight/2019/1024/> (2021/12/28 確認)

断した被害画像に分類した画像群をさらに細分化し、どのような特徴を持つのか調査を行うことで、被害画像の機械的な抽出可能性について考察する。以下の手順で調査を行う。

まず、人手によって分類された画像から機械的に求めたテキスト特徴を抽出することで、ノイズとなる投稿の抽出が可能になるか調査を行う（ステップ1）。次に、ステップ1では分類しきれない画像に対し人手でツイート本文と画像を確認することで、どのような分類が可能になるか調査を行う（ステップ2）。ステップ1については6章で、ステップ2については7章で各々述べる。

5 対象の災害と収集データ

本章では、本研究の対象災害の概要とその被害について述べ、災害における実情を概観する。それらの被害や状況をふまえ、ツイートデータの収集を行った。

5.1 令和2年7月豪雨

本研究では令和2年7月3日から31日にかけて発生した「令和2年7月豪雨」を分析対象とする。国土交通省気象庁「令和2年7月豪雨令和2年(2020年)7月3日～7月31日(速報)¹」の発表によると、九州地方や岐阜県周辺で記録的な大雨となり、球磨川や筑後川、最上川といった大河川での氾濫が相次ぐ他、土砂災害によって甚大な被害をもたらした。その被害例を図5.1²に示す。当災害では414名が緊急消防援助隊の活動により救助された³。

同期間中、最も総降水量が多かった雨量観測地は、長野県木曾郡王滝村で総降水量は2135.5mmになり、高知県の安芸郡馬路村で観測され総降水量は2032.5mmとなった。九州南部・九州北部・東海・東北地方の多くの地点で時間別降水量が観測史上最も高い値となったことでも知られている。

内閣府「令和2年7月豪雨による被害状況等について」[31]の発表によると、特に熊本県では豪雨被害による死者が多く65名となり、全国でも計84名の死者を出すこととなった。さらに行方不明者2名、重軽傷者67名となり、総人的被害者数は計153名となった。住居被害も極めて多く、全壊した家屋は1,621棟、半壊した家屋は4,504棟、一部破損に留まった家屋は3,503棟にも上る。浸水被害は、床上浸水した家屋は1,681棟、床下浸水した家屋は5,290棟となり、住居被害は合計16,599棟となった。

ライフラインにも大きな被害を受け、東北電力・中部電力は一部障害が発生し、九州電力は鹿児島県・宮崎県・大分県で障害が発生したほか、熊本県では約290戸停電するなど電力供給も滞った。水道被害に関しては、15都道府県にわたり37,653戸(各都道府県における最大断水戸数を合算)が被害を受けた。通信被害に関しては、NTT東日本⁴・NTT西日本⁵・NTTコミュニケーションズ⁶・KDDI⁷・ソフトバンク⁸の提供する固定電話が被害を受けた他、NTTドコモ⁹・KDDI・ソフトバンクが提供する携帯電話にも一部支障が発生した。地上波放送被害に関しては、大分県日田市では停電を原因とした放送被害が発生し約240世帯に影響した。鹿児島県霧島市では落雷を原因とした放送被害が発生し約100世帯に影響した。また、大分放送による大分県日田市の地上波ラジオは停電を原因とした放送被害が発生し、約30,540世帯に影響を与えた。

土砂災害被害も頻発し熊本県で226件、全国で961件の土砂災害が発生した。その被害

¹<https://www.data.jma.go.jp/obd/stats/data/bosai/report/2020/20200811/20200811.html> (2021/12/28 確認)

²https://www.mlit.go.jp/river/shinngikai_blog/hazard_risk/dai05kai/dai05kai_sankosiry02.pdf (2022/2/8 確認)

³https://www.fdma.go.jp/disaster/info/items/201203_oame54.pdf (2022/2/8 確認)

⁴<https://www.ntt-east.co.jp/> (2022/2/8 確認)

⁵<https://www.ntt-west.co.jp/> (2022/2/8 確認)

⁶<https://www.ntt.com/index.html> (2022/2/8 確認)

⁷<https://www.kddi.com/> (2022/2/8 確認)

⁸<https://www.softbank.jp/> (2022/2/8 確認)

⁹<https://www.nttdocomo.co.jp/> (2022/2/8 確認)

表 5.1: 収集したツイート総数(キーワード別)

| 収集キーワード | ツイート数 | 画像付きツイート数 |
|---------------|--------|-----------|
| “救助” ツイート | 110261 | 18197 |
| “避難” ツイート | 370531 | 34777 |
| 合計ツイート(重複削除後) | 476827 | 47434 |

の例を図 5.2 に示す。

当災害において、国管理の 8 河川、都道府県管理の 194 河川で氾濫が確認された。また、国が管理する球磨川では 2 箇所が決壊する事態となった。その被害の例を図 5.3 に示す。

5.2 収集したツイートデータ

Twitter 社は Web スクレイピングによるツイートデータの取得を禁止している。そのため、ツイートデータを収集する場合は、予めプログラミングコードなどを提出の上 TwitterAPI の利用申請を行う必要がある点を留意し、対象災害のツイートデータは Twitter API¹⁰ を用いてキーワードをクエリとした収集を行った。収集キーワードは「救助」および「避難」を設定した。データ収集対象期間は令和 2 年 7 月 1 日～15 日に設定した。全国を通して最も降水量の多い期間が令和 2 年 7 月 4 日から 7 日であったため、その後一週間を目処に収集を行った。

本研究で扱ったツイートの総数は 476,827 件であった(表 5.1 参照)。そのうちキーワード“救助”で得られたツイートは 110,261 件、“避難”で得られたツイートは 370,531 件、“救助”と“避難”両方のキーワードが含まれ重複したツイートは 3,965 件となった。このツイートの中から何らかの画像・動画が含まれているツイートは、キーワード“救助”で得られたツイートには 18,197 件、キーワード“避難”で得られたツイートには 34,777 件、“救助”と“避難”両方のキーワードが含まれ重複したツイートは 5,540 件となり、合計で 47,434 件のツイートとなった。つまり単純なキーワード検索を行った場合、476,827 件のツイートが結果としてヒットする。画像・動画が含まれているツイートに絞った場合 47,434 件のツイートが結果としてヒットする。これらのツイートの中から画像および動画を抽出した結果、合計 94,111 枚の画像・動画となった。なお、画像・動画はツイートの ID と紐付いている。

¹⁰<https://help.twitter.com/ja/rules-and-policies/twitter-api>(2021/12/28 確認)

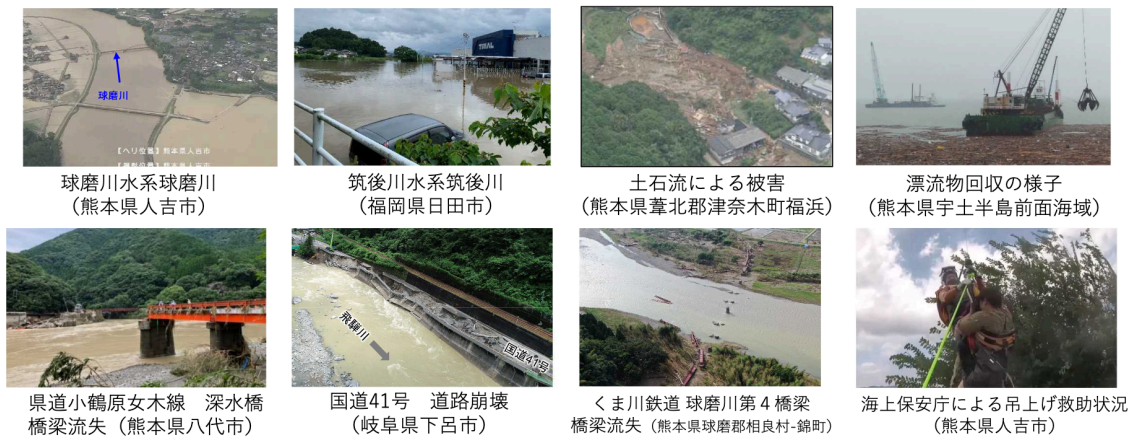


図 5.1: 令和 2 年 7 月豪雨による被害例（令和 2 年 7 月豪雨による被害と対応²より図引用）



図 5.2: 令和 2 年 7 月豪雨による土砂災害例（令和 2 年 7 月豪雨による被害と対応²より図引用）

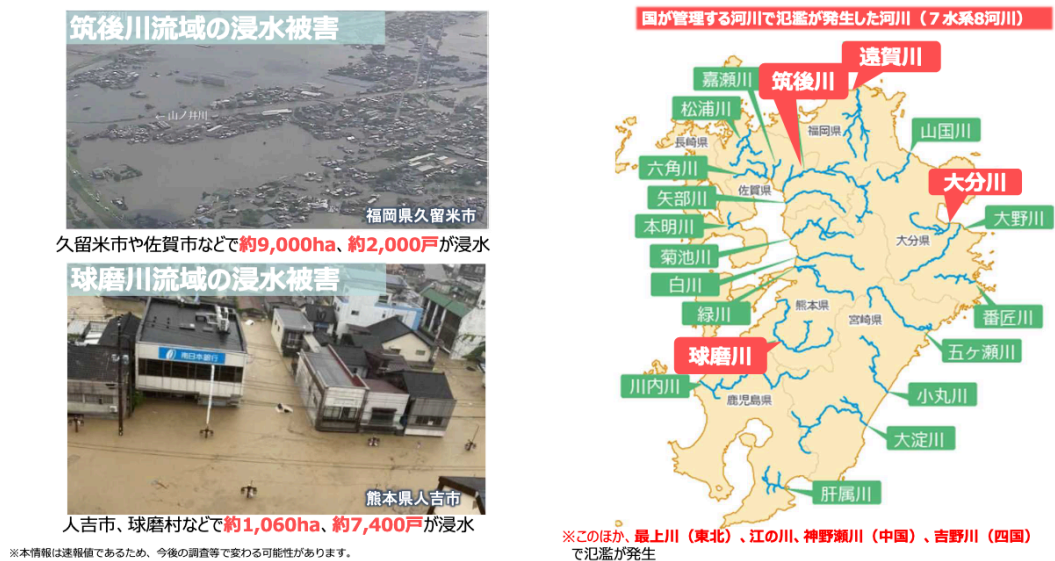


図 5.3: 令和 2 年 7 月豪雨による河川による浸水被害（令和 2 年 7 月豪雨による被害と対応²より図引用）

6 ノイズとなる投稿に関する調査

本章では、Twitter で発信される情報の特徴を明らかにするため、ツイートに含まれる画像の分析を行う。

6.1 投稿画像の分類手順

Twitter では一投稿あたり 4 枚まで画像を添付することが可能であり、投稿によってはそれぞれの画像が異なる内容を示すケースがある。その際は、それぞれ異なる画像クラスとして個別に分類を行った。分類にあたっての観点は以下の通りである。

- (1) 災害に関連するか
- (2) メインに写っているものが何か
- (3) ニュースに関連するか

以下に分類の手順を述べる。はじめに、分類作業を行う 4 人の大学生（以下、分類担当者）に対して、ニュースに関連する画像例、および目視によるサンプリングにより多数観測できたゲームに関連する画像例として、以下の 4 つのクラスを例示した。ニュースメディアに関する想定した画像例を図 6.1 から図 6.3¹²³に示す。

- (1) 逆 L 字と呼ばれる災害被害発生中のニュース速報などが放送される際に付けられるようなものが映り込んでいる画像（図 6.1）
- (2) テレビのニュースをカメラで撮影した画像（図 6.2）
- (3) メディア媒体を問わず、ニュースに関する画像（図 6.3）
- (4) ゲーム画面やアニメ・漫画などの画像

分類担当者は、各人の判断のもと上記以外のクラスを作成して分類を行った。とある画像が複数の分類クラスに該当すると判断した際は、同一画像を該当するすべての分類クラスに複製することとした。その結果、分類担当者が実際に分類した画像・動画の総数は 292,466 となった。

その後、クラスの名称に表記揺れが起きたクラスや同一概念のクラスが複数作成される可能性があるため、分類担当者に聴取を行った上で、同一の分類クラスであると判断できるものを著者の判断により統合した。その分類結果の一部を表 6.1 に記す。4 章で収集したツイートおよび画像収集時にリツイートを排除していなかったため、同一のツイート ID を持つ画像が多く確認できた。また、同一の画像でも異なるツイート ID を持つ画像も確認できた。これらの画像を人手で分類した結果、分類担当者によっては同一の画像を異なるクラスに分類するケースが多く存在した。

¹<https://twitter.com/KANKAI79/status/1279148724766167041>（現在は投稿削除済み）

²<https://twitter.com/TAKAYA05378399/status/1279181014711930880>（2022/2/8 確認）

³<https://twitter.com/amasehimika147/status/1150073316863959041>（2022/2/8 確認）

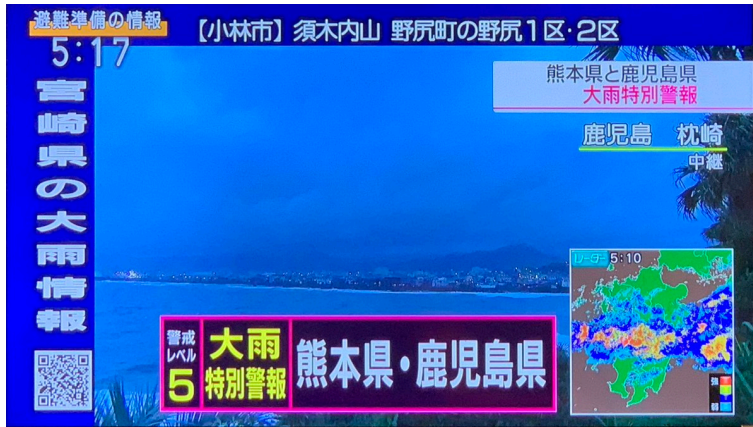


図 6.1: 逆L時が確認できる画像例¹



図 6.2: テレビの報道をカメラで撮影した画像例²

九州 非常に危険 線状降水帯発生中
2019年7月13日 21時43分 9wRk1p

九州、危険な雨の降り方。発達した雨雲が同じような場所にかかり続ける、線状降水帯が発生中。熊本、鹿児島に警戒レベル4相当の土砂災害警戒情報発表。

同じ場所に滝のような雨

13日(土)20時30分現在、熊本県から鹿児島県にかけて、線状に活発な雨雲が連なる線状降水帯が発生しています。熊本県から鹿児島県の県境付近では、夕方から同じような場所で1時間に50ミリ以上の滝のような雨が降り続けています。1時間雨量は、熊本県球磨郡山江村(山江)で73.0ミリ、鹿児島県出水市(出水)では、3時間で125.5ミリの大雨となっています。(1時間雨量、3時間雨量ともに13日20時30分まで)

土砂災害警戒情報発表中

熊本県や鹿児島県には、警戒レベル4にあたる土砂災害警戒情報が発表されています(13日20時30分現在)。同じような地域で、非常に激しい雨が長く続くため、土砂災害の危険が高まっています。安全な場所へ避難、夜間のためかえって避難が危ない場合は、自宅が2階建ての場合は2階へ避難。崖が近い場所は崖から離れた部屋へ移動するなど早め早めの行動をお願いします。

図 6.3: ニュースに関する画像例³

表 6.1: 人手による画像分類クラス抜粋

| クラス | 基となったツイート数 |
|----------|------------|
| Any ニュース | 1546 |
| Twitter | 1538 |
| ゲーム | 8230 |
| テレビのニュース | 1471 |
| 二次元画像 | 2115 |
| 人 | 2389 |
| 地図 | 1205 |
| 救助 | 691 |
| 被害 | 4293 |
| 避難 | 3826 |
| 風景 | 3168 |
| 飲食物 | 1445 |
| 天気 | 842 |
| 防災グッズ | 618 |
| 自衛隊 | 439 |
| 逆L字 | 324 |

6.2 投稿画像の分類結果

最も多くの画像が分類されたクラスは“ゲーム”であった。“ゲーム”に分類される画像群には「Identity V 第五人格⁴」と呼ばれるゲームに関する画像や動画が多く確認できた。この原因として、ゲーム内で一般的に利用される「救助」という語が、検索結果として収集されたことが考えられる。次点で多かったのは“被害”のクラスに分類された画像であった。“被害”に分類された画像には、個人が撮影したと思われる被災画像の他に、メディアの報道で流れた被災画像、ウェブニュースのサムネイルに設定された被災画像、日本各地に設置されたライブカメラからの被災画像などが確認できた。

6.3 テキスト処理

ツイートテキストの傾向分析を行うため、6.1 節で得られた各画像クラスを一群として TF-IDF 値を算出する。はじめに、4 章で取得したツイートデータから、他者のツイートを再発信するリツイートと呼ばれるものを排除した。また、ツイート本文に含まれる URL を排除した。その後、全角英数字を全て半角英数字に変換した。次に、形態素解析器 MeCab (ver.0.966.2)⁵を用いて形態素解析を行い、単語ごとに分割した。最後に、それらの単語から名詞のみを抽出した。その際、SlothLib プロジェクト [18] でストップワードに設定された単語と、ツイート取得の際に設定したキーワードである救助・避難をストップワードと

⁴<https://www.identityvgame.com/jp/> (2020/12/18 確認)

⁵<https://taku910.github.io/mecab/> (2020/12/18 確認)

して設定した。分析はTF-IDFを用いる。Term Frequency (TF)とは、ひとつの文章中におけるある単語の出現頻度を指す。その単語の出現回数が多ければ多いほどこの数値は大きくなる。算出方法は、以下の通りである。

$$TF_{i,j} = \frac{N_{i,j}}{X} \quad (6.1)$$

X はとある文章 j に出現する単語の総出現回数であり、 $N_{i,j}$ はとある文章 j における単語 i の出現回数を示している。

Inverse Document Frequency (IDF)とは、とある単語が、いくつの文章で出現したかを示す指標であり、特定の文章でしか出現しない場合に数値は大きくなる。算出方法は以下の通りである。

$$IDF_i = \log_e \frac{N}{df_i} + 1 \quad (6.2)$$

N は総文章数を示す。 df_i は単語 i を含む文章数である。以上のTF値、IDF値をもとにTF-IDF値を算出する。算出方法は以下の通りである。

$$TF-IDF_{i,j} = TF_{i,j} \cdot IDF_i \quad (6.3)$$

本稿では、表6.1のクラスを含む23種類のクラスをもとに、以上の式からTF-IDF値を算出した。算出の際、各画像分類クラスのツイートテキストに出現する単語をTerm、各画像分類クラスのツイートテキスト群をDocumentとして扱う。算出の際、全ての文章のうち8割を超えるクラスで出現した単語はクラスの特徴として不適であるため、それらの語を除外した。算出したTF-IDFの結果の一部を表6.2に示す。この分析によって、各画像分類クラスのツイートテキストの傾向から、画像を特徴づける単語の傾向を見つけることが可能になる。

TF-IDF値算出の際、全てのクラスのうち8割を超えるクラスで出現した単語を除外したことで、各々のクラスで固有の単語が上位に算出された。“ゲーム”でTF-IDF値が高く上位に算出された単語は、前節で挙げたゲーム固有の表現であった。“TVニュース”で上位に算出された単語は、“【2018年7月】避難勧告対象101万4930世帯・232万1947人。大雨特別警報が11府県にわたって発令される前代未聞の状況のなか宴会を楽しむ安倍さん／東洋経済オンライン”というツイートが大部分を占め、繰り返しツイートされていたことが確認できた。このツイートはBotに依るもので、付随した画像は“News23⁶”で報道されたニュースの切り抜き画像であった。この原因として、本来、同一クラス内であればツイートIDの重複により、同内容のツイートは排除されるが、繰り返しツイートされることで異なるツイートIDが付与されていたため重複したと考えられる。“被害”で上位に算出された単語である“坂本”および“八”は熊本県にある八代市坂本町の一部であった。“逆L字”では7月8日に岐阜県で大雨特別警報が発表された際のツイートが収集されており、早めの避難を呼びかけるツイートが確認できた。

⁶<https://www.tbs.co.jp/news23/> (2022/01/13 確認)

表 6.2: TF-IDF 上位 5 単語抜粋

| ゲーム | TF-IDF | テレビニュース | TF-IDF | 救助 | TF-IDF |
|------|--------|---------|--------|----|--------|
| チェイス | 0.3242 | 前代未聞 | 0.1966 | 師団 | 0.4122 |
| 傭兵 | 0.3007 | 安倍 | 0.1903 | 派遣 | 0.3462 |
| 解説 | 0.2681 | 府県 | 0.1640 | 陸上 | 0.2264 |
| 祭司 | 0.2384 | 東洋 | 0.1640 | 神瀬 | 0.2065 |
| 即死 | 0.1950 | 八 | 0.1575 | 実施 | 0.1883 |

| 被害 | TF-IDF | 逆L字 | TF-IDF | 防災グッズ | TF-IDF |
|-----|--------|-----|--------|-------|--------|
| 坂本 | 0.2052 | 早め | 0.5147 | セット | 0.6452 |
| 八 | 0.1825 | 岐阜 | 0.3666 | グッズ | 0.3340 |
| 筑後川 | 0.1754 | 放流 | 0.1693 | 楽天 | 0.3184 |
| ダム | 0.1611 | ダム | 0.1585 | 袋 | 0.2343 |
| 増水 | 0.1414 | 日田 | 0.1580 | 用品 | 0.2254 |

6.4 コサイン類似度の算出

同一IDがTF-IDF算出の際に影響を与えている可能性を考慮し、ツイート群間のコサイン類似度の算出を行った。コサイン類似度は、ベクトル同士の角度の近さを表現することができる。これを文書特徴の一つであるTF-IDFを用いて算出することで、文書同士の類似度を表現することができる。類似度が高く算出される傾向の高かったクラスを表6.3に示す。

類似度が高かった語群として、“救助”と“自衛隊”，“地図”と“天気図”と“天気”などがあった。また，“テレビニュース”は，“被害”や“避難”と類似度が高く，一般的な報道と被害情報や避難情報が混交する可能性が考えられる。“被害”では複数のクラスと類似度が高く算出され，災害情報全体を象徴する語である可能性が示された。

また，当初の懸念通り，同一IDのツイートが他クラス間で存在している場合，クラス間のコサイン類似度が高く算出される傾向にあった。分類担当者への聴取から，画像を分類するクラスの判断が困難だった場合，本来分類すべきクラスではなく，コサイン類似度が高く算出される傾向のあったクラスへ割り振るケースを確認した。

そのため，同じ画像でも分類担当者によっては，違うクラスに割り振られた場合も存在する。この結果から，類似度が高く算出される傾向のあるクラス同士の画像は，人間の判断による画像分類においても分類エラーを起こす可能性が高いことが示唆された。

表 6.3: コサイン類似度が高い傾向のあるクラス 抜粋

| テレビニュース | Cos 類似度 | 人 | Cos 類似度 | 地図 | Cos 類似度 | 天気 | Cos 類似度 |
|---------|---------|---------|---------|-----|---------|-----|---------|
| 被害 | 0.4728 | 避難 | 0.4206 | 被害 | 0.5291 | 天気図 | 0.9031 |
| 避難 | 0.4519 | 風景 | 0.3697 | 天気 | 0.5234 | 地図 | 0.5234 |
| 地図 | 0.4254 | テレビニュース | 0.3173 | 天気図 | 0.5183 | 避難 | 0.3455 |
| 人 | 0.3173 | 被害 | 0.2904 | 避難 | 0.4818 | 風景 | 0.3294 |

| 救助 | Cos 類似度 | 被害 | Cos 類似度 | 避難 | Cos 類似度 | 風景 | Cos 類似度 |
|---------|---------|---------|---------|---------|---------|----|---------|
| 自衛隊 | 0.7939 | 風景 | 0.5341 | 地図 | 0.4818 | 被害 | 0.5341 |
| 被害 | 0.3499 | 地図 | 0.5291 | テレビニュース | 0.4519 | 避難 | 0.3759 |
| テレビニュース | 0.3141 | テレビニュース | 0.4728 | 被害 | 0.4499 | 地図 | 0.3728 |
| SNS | 0.2139 | 避難 | 0.4499 | 人 | 0.4206 | 人 | 0.3697 |

7 被害投稿の特徴に関する調査

4.2節での懸念通り，調査1の結果から“被害”では複数のクラスと類似度が高く算出される傾向があった．他の分類クラスのツイートが比較的多く混交している可能性を考慮し，再度人手による画像分類を行うことで，細分化の可能性に関する調査を行う．これは，6章で行った画像特徴に着目した分類と，TF-IDFの特徴では分類できなかった画像群を，画像とテキストを人目で確認することで調査する．

まず，“被害”に分類された画像を再度確認する．何を意図した画像か判断が難しいため，画像の元となったツイートを参照することで，その画像の出典やツイートの意図まで確認する．それにより，画像を見ただけでは判断が難しい分類を行うことが可能になる．

その後，分類担当者は各自の判断で被害画像のみを細分化して分類を行う．最後に，細分化した“被害”に分類されたツイート文間で6.3章と同様の手順でTF-IDF値の算出を行う．“被害”に関する画像を細分化し，各クラスの持つ文章的な特徴を明らかにする．

7.1 被害画像の分類と細分化

“被害”に分類されたツイートには，ヒューマンエラーを起こしていた画像を除くと“テレビのニュースで報道された被害の画像”（66ツイート），“ネットニュースで報道された被害の画像”（204ツイート），“ユーザが撮影した被害の画像”（2689ツイート），“ライブカメラで撮影された被害の画像”（209ツイート），“自治体が発信した被害の画像”（95ツイート），“分類不可能だった被害の画像”（153ツイート），“元ツイートが消されていた被害の画像”（101ツイート）に分類された．“分類不可能だった被害の画像”は，中文や英文でツイートされていた場合，日本国以外で発生した災害について言及されている場合（e.g., 2019年9月28日韓国南東部での石油タンカーの爆発，2020年7月6日に始まった中国江西省での洪水災害），元ツイートを辿っても画像の撮影状況や画像の出典が不明な場合に分類されている．“自治体が発信した被害の画像”は地方自治体によって運営されているアカウントや政治家のアカウントから投稿された被害の画像が分類されている．再分類の結果と再分類されたツイートのみで算出したTF-IDF値の上位5単語を表7.1に示す．

表 7.1: 細分化した“被害”ツイートにおける TF-IDF 上位 5 単語抜粋

| テレビニュース | TF-IDF | ネットニュース | TF-IDF | ユーザが撮影 | TF-IDF | ライブカメラ撮影 | TF-IDF |
|---------|--------|---------|--------|--------|--------|----------|--------|
| 熊本 | 0.3606 | 熊本 | 0.3952 | 川 | 0.2912 | 氾濫 | 0.3950 |
| 当時 | 0.3114 | 豪雨 | 0.2803 | 雨 | 0.2708 | 水位 | 0.3402 |
| 園 | 0.2854 | 大雨 | 0.2021 | 氾濫 | 0.2482 | カメラ | 0.3112 |
| 寿 | 0.1986 | 指示 | 0.1823 | 熊本 | 0.2354 | ライブ | 0.2847 |
| 千寿 | 0.1631 | 災害 | 0.1792 | 大雨 | 0.2232 | 川 | 0.2581 |

| 自治体発信 | TF-IDF | 分類不可能 | TF-IDF | 元ツイート消去 | TF-IDF |
|-------|--------|-------|--------|---------|--------|
| 災害 | 0.4000 | 峡 | 0.2617 | 熊本 | 0.2230 |
| 駐屯 | 0.2539 | ダム | 0.2580 | お願い | 0.2088 |
| 豪雨 | 0.2461 | 洪水 | 0.2051 | 地区 | 0.2036 |
| 派遣 | 0.2377 | 大雨 | 0.1864 | 坂本 | 0.1998 |
| 熊本 | 0.2256 | 中国 | 0.1764 | 連絡 | 0.1978 |

7.2 細分化された分類のテキスト特徴

“テレビのニュースで報道された被害の画像”で上位に算出された単語である“園”，“寿”，“千寿”はいずれも熊本県球磨郡の特別養護老人ホーム“千寿園”に関するニュースであった。また、TF-IDF 値の第6位は“取材” (0.1565)であったが、その他の分類において“取材”のTF-IDFは最高で0.0233となった。

“ライブカメラで撮影された被害の画像”で上位に算出された単語はいずれも、河川の氾濫に関連する単語であった。これはライブカメラの設置場所が主に河川の観測所にあるためだと考えられる。

“自治体が発信した被害の画像”の最上位に算出された“災害”は，“ネットニュースで報道された被害の画像”では第5位に確認できたものの、その他の分類におけるTF-IDF値は比較的低い傾向が確認できた。自治体や政治家によって投稿されるツイートには“豪雨”や“雨”などによる被害であることに加えて，“災害”による被害であることをより強調している。また，“災害”に伴う自衛隊の“派遣”状況を伝える投稿が多く確認された。

“分類不可能だった被害の画像”で上位に算出された“峡”，および“ダム”は中国の長江中流にある“三峡ダム”を指していた。分類不可能とする判断基準の一つが中国語による投稿であった。

“元ツイートが消されていた被害の画像”で上位に算出された“お願い”は，“「救助を」お願いします”などといった救助を要請する場合に多く使われていた。救助要請ツイートは具体的な地名を入れるという特徴があり [25]，“地区”や“坂本”といった単語はその救助要請ツイートの特徴を反映したものだと考えられる。

一方で、被災者の救助要請や被害の状況報告が最も反映されていると想定される“ユーザが撮影した被害の画像”のTF-IDFにおいて、特出した値や単語は認められなかった。

7.3 細分化により得られた知見

7.2節で得られた結果から、キーワードによるツイート抽出には、共通の文字を扱う中国語のツイートも抽出されるため、こうした無関係な中文によるツイートを除去する必要がある。

“元ツイートが消されていた被害の画像”には、具体的な地名が書かれており、Twitter Japan社が想定している救助要請の形式を保ったツイートや、実際に救助されたケースなどが確認できる可能性が高い。例としてその一部を表7.2に示す。この101件のツイートは今後定性的に調査する必要がある。

“被害”に分類されたツイートの細分化が可能であった一方で、被災者の救助要請や被害の状況報告が最も反映されていると想定される“ユーザが撮影した被害の画像”のTF-IDFにおいて、特出した値や単語は認められなかった。つまり，“被害”に関する投稿には特徴となる単語は見受けられず、投稿された文章のみを用いた被害投稿の抽出は困難である。

ユーザが撮影した被害画像の投稿を抽出するためには、テキスト特徴に加えて画像特徴といった異なるモダリティの特徴を考慮した分類モデルを構築することで、抽出精度向上に繋がるか、今後確認する。

表 7.2: 消されたツイート本文より一部抜粋 (@○○○は特定のユーザに向けた発信)

彼女の実家含めその地区、まだ救助が来てないみたいです。小さい子供も、お年寄りもたくさん取り残されています。電波も悪いみたいです。道路も壊れているので、ヘリで救助をお願いします！ #熊本県球磨郡球磨村渡糸原地区水篠地区 #救助

@○○○ いつもありがとうございます。こちらの地区も朝からずっと孤立し取り残されています。少しでも早く救助の方をお願いします。 #熊本県球磨郡球磨村渡糸原地区水篠地区 #救助

豪雨での救助活動。ランドアースに関わって、ガイドさん達を知っている者として本当に誇らしいです。会社は二階まで浸水し、甚大な被害を受けました。これ以上被害が出ないことを願います。会社で飼っていた猫ちゃん、生きていました。良かった…

友達から連絡がきました。 #救助 #救助希望 人数は10数名だそうです！

@○○○ 芦北町吉尾地区は未だに孤立状態です 山に接した家が多く、土砂崩れの危険が非常に高いです 毎回の様に災害時道路が寸断されますが、今回すぐに復旧は無理と思います (写真添付) 昨日から水道、電気止まっています 高齢者が多い地区です 90代の祖母が居ます 早急にヘリでの救助をお願いしたいです！

一刻も早い救助をお願いします @○○○ @○○○ @○○○ @○○○ @ #救助 #救助要請 #熊本 #八代市 #坂本町 #鎌瀬 坂本町鎌瀬地区と 鎌瀬地区 孤立状態 高齢者や持病守ってる方が多くいる 早めの救出と食料支援 SOS

お借りした画像で葉木橋から肥薩線側を見たものです。祖母が土砂崩れに巻き込まれてるかもしれないです。家は動いていませんが土砂が入り込んでいる可能性が高いです。昨日からずっと連絡が取れません。現地の間では無いので

救助要請をしていいのかわからずテレビを見るばかりです。 #八代市坂本町

#救助 #救助要請

イラクのIS(ダーイッシュ)より 今年5月の中頃に出された動画から ・3つの爆弾を使用しての連続攻撃1つ目で部族の人間を狙い、2つ目3つ目で駆けつけた警官と救助に来た人間を狙い爆破 少しずつではあるが、イラク・シリアでの活動が目立ち始めている

@○○○○○ 芦北町 吉尾 市居原地区 本日救助ヘリは来たものの、1人ずつの救助との事で、結局誰も避難していません 水や電気の無い中、近隣で助け合い備蓄の食料と水で凌いで道の復旧を待っている様ですが、写真を見たところすぐに復旧するとは思えません 土砂崩れも心配ですが 至急支援物資をお願い致します！

うちの祖父母も球磨村神瀬乙で昨日の朝から連絡取れてません。公民館に避難したようですが、詳細不明です。早く救助をお願いします。食料も足りているのか心配です。

一刻も早い救助をお願いします @○○○ @○○○ @○○○ @○○○ @○○○ #救助 #救助要請 #熊本 熊本県球磨郡球磨村神瀬乙 769 神瀬多武除地区 昨日の朝から祖父母と連絡とれてない状態 道路も寸断されて、孤立 報道もされてない地域

一刻も早い救助をお願いします @○○○ @○○○ @○○○ @○○○ @○○○ @○○○ #救助 #救助要請 #熊本 #熊本 #八代市坂本町葉木佐瀬野地区 坂本町葉木佐瀬野地区に祖母と連絡が取れない 土砂崩れも発生しているので巻き込まれてる可能性が高い 現地の情報がない。SOS 要請

@○○○ お願いします。坂本荒瀬にまだ親が取り残され、孤立しています。福祉館に避難はしたようですが、連絡が一昨日の夕方から取れません。小さな避難所は、まだ、救助がおこなわれていません。着のみ着のままギリギリで避難しているので水もないはず。各集落で高台に点在して避難しているはず。

@○○○ @○○○ よろしくをお願いします。八代市坂本にある坂本地域福祉センターに避難した両親と全く連絡が取れません。食料も電気もなく、どうしているか本当に不安です。あちこちの避難所に連絡しましたが、どこにもいないようです。救助を待っていると思われます。

表 7.3: “被害” ツイート数とその割合（一部抜粋）

| | ツイート数 | 被害画像ツイート割合 (%) | 画像ツイート割合 (%) | ツイート総数割合 (%) |
|---------------|--------|----------------|--------------|--------------|
| 被害-自治体発信 | 95 | 2.297% | 0.200% | 0.020% |
| 被害-分類不可能 | 153 | 3.700% | 0.323% | 0.032% |
| 被害-元ツイート消去 | 101 | 2.443% | 0.213% | 0.021% |
| 被害-テレビニュース | 66 | 1.596% | 0.139% | 0.014% |
| 被害-ネットニュース | 204 | 4.933% | 0.430% | 0.043% |
| 被害-ユーザ撮影 | 2689 | 65.030% | 5.669% | 0.564% |
| 被害-ライブカメラ撮影 | 209 | 5.054% | 0.441% | 0.044% |
| “被害” 画像ツイート総数 | 4135 | 100.000% | 8.717% | 0.867% |
| 画像ツイート総数 | 47434 | | 100.000% | 9.948% |
| ツイート総数 | 476827 | | | 100.000% |

7.4 フィルタリング後の想定ツイート数

これまでに分類した“被害”に該当するツイート数を結果として表 7.3 に示す。“救助”および“避難”でツイート検索を行った場合、そのツイート数は 476,827 件となる。その中から何らかの画像や動画が添付されているツイートに絞って検索を行った場合、そのツイート数は 47,434 件となる。画像や動画が添付されているツイートから、“被害”に関するツイートに絞られた場合、4,135 件となる。“被害”からユーザ自身によって撮影されたと想定される“ユーザ撮影”ツイートに絞った場合、2,689 件となる。

8 議論

本稿では、令和2年7月豪雨の期間に投稿されたツイートを収集し、投稿された画像を人手で分類した分類クラス基にツイートテキストの特徴を調査した。その結果がもたらす価値や課題点について述べる。

6章で行った画像分類の結果、投稿される画像の傾向や、誤分類が発生しやすい分類クラスが明らかになった。その画像分類クラスを一つの文書とみなしてツイートテキスト群を作成し、分類クラスの文書間の TF-IDF を算出した結果、各分類クラスで特徴的な単語が確認できた。つまり、画像特徴のみに着目した分類クラス別の TF-IDF という単純なテキスト特徴抽出によって“被害”に関する画像以外ではある一定の特徴抽出が可能であった。

しかし、ある一定のツイートがノイズとして混合し、それらが大きく数値に影響したケースも確認できた。算出した TF-IDF を基に文書間のコサイン類似度を算出した結果、誤分類を起ししやすいクラス間が明らかになり、同一のツイートを複数クラスへ分類する可能性が示唆された。

7章で行った、災害時に必要になるであろう“被害”に分類された画像に対して元ツイートのテキストを参照することで、画像とツイートの詳細な状況などを確認するといった質的調査を行った結果、“被害”画像をさらに細分化して分類できることが確認できた。

一方で、“被害”を細分化したクラスの一つである“ユーザが被害を報告するツイート”には特徴的な単語は見られなかった。被災者自身が投稿した画像をピンポイントで抽出することは困難である。そのため、被災者自身の被害投稿をピンポイントで抽出するのではなく、無関係な投稿や情報をノイズとして扱い除去することで、災害時における迅速な情報収集を目指す必要がある。

7.4節で示したツイート数から、理想的なツイートの最終出力は、ユーザ自身によって撮影されたと想定される“ユーザ撮影”の2,689件となる。これらのツイートにアクセスするためには、単純なキーワード検索でヒットした476,827件から、ツイートの本文および添付画像を考慮してノイズとなる情報を除去することでフィルタリングを行っていく必要がある。“被害”とは無関係と想定されるツイートを排除しフィルタリングを行った場合、出力されるツイートは4,135件に絞られる。さらにそれらの4,135件からより高精度なフィルタリングが可能になった場合、ユーザ自身が被災・被害を示すツイートは2,689件となる。災害関連ツイートからノイズを除去した場合この2,689件のツイートが最終的に出力結果となることが想定される。

これまでの結果から、災害時の Twitter 上には、災害被害や救助要請とは無関係なツイートが多く存在することが明らかである。災害時における Twitter 上の画像を人手による分類を行ったことで、どのような情報が実際に投稿されているか明らかとなった。TF-IDF という単純なテキスト処理のみで一定のテキスト特徴を抽出が可能であったことは、時間的リソースが欠如している災害時において、必要な情報までのアクセスの容易さに繋がる。画像とテキストという複数のモダリティを考慮することで、より高精度な分類が可能になり、ノイズとして扱うべき情報のフィルタリングが容易になる。

本研究では、豪雨災害に着目し Twitter 上に投稿された画像および本文を対象に調査を行った。本研究で算出した TF-IDF 値やコサイン類似度はその特性上、他の災害発生時の

データを適用した際に同様の結果になるとは限らない。つまり，令和2年7月豪雨災害における特徴的な単語であり，災害全般における特徴的な単語とは言えない。本手法を，他の災害に適用させるためには，6章から7章で行った調査の結果と，他の災害が発生した際の画像分類クラスや単語を調査し比較する必要がある。

今後，災害時に特化した画像分類モデルを構築することで計算機による雑多な画像群の分類が可能になれば，単純なテキスト処理を行うだけで，被害情報や救援要請の抽出精度向上が見込まれる。本研究で行った人手による分類をもとに，機械的に画像特徴が抽出できる可能性がある。

9 おわりに

SNSを災害情報収集ツールとして活用するためには、最終的に人が判断することが必要になるが、人が判断する以前の機械的な情報抽出の際において被災情報の「取りこぼし」があってはならない。そのため本稿では、複数のモダリティを考慮した情報抽出およびノイズの除去を目的として、SNS上の災害情報に関する調査を行い、画像特徴とテキスト特徴の双方のモダリティを考慮することで、災害被害情報の分類細分化が可能であることを明らかにした。さらに、複数のモダリティに着目したフィルタリングを行った場合、“被害”に該当するツイートは全体の0.867%まで絞られる可能性を示した。全体の0.564%のツイートとなるユーザによって報告される実被害情報にアクセスするためには、より高度なフィルタリングを行う必要がある。

ただし、本研究は豪雨災害を対象とした調査を行ったため、これらの結果が他の災害が発生した際に適用されるかは、本研究で得られた分類結果から、他災害が発生した際のツイートデータと共通するであろう分類クラスや単語を明らかにする必要がある。

今後、実被害とは無関係な投稿をノイズとして扱うことが可能になれば、情報廃棄のアプローチを取ることで、より必要な情報へのアクセシビリティ向上が期待される。被害に関する投稿画像の分類モデルを構築し、より高度なフィルタリングを行うためには、マルチモーダルなアプローチが必要になると考えられる。テキスト特徴と画像特徴の複数モダリティを考慮したモデル構築が求められる。

謝辞

本研究は科研費基盤研究 (c) 特設 (科研番号: JP18KT0100) の支援を受けた。記して謝意を表す。

本研究を纏めるにあたり、関西大学総合情報学部総合情報学科の松下光範教授には、御指導、御鞭撻賜りましたこと厚く御礼申し上げます。4年間を通して学会発表やその他学外発表の機会を存分にいただきました。

法政大学社会学部藤代裕之教授には、社会情報学的なアプローチやメディアのあり方など、多くのことを教えていただきました。研究を通して貴重な意見をいただきました。また、小生が本研究テーマを選択したきっかけをいただきました。深く感謝申し上げます。

関西大学総合情報学部総合情報学科の山西良典准教授には、表現者としての在り方から論文の読み方まで、様々なことに御意見、御指摘いただきました。また、氏のウェブサイトで利用されているアイキャッチ画像の制作、NICOGRAPH2020およびコミック工学研究会の運営サポートという貴重な機会をいただきました。心より感謝いたします。

明治大学総合数理学部先端メディアサイエンス学科の中村聡史教授には、合同研究発表会や交流会など、多くの議論の場をいただきました。心より感謝いたします。

研究生活を送る上でお世話になりました松下研究室第7期生、第8期生、第9期生の皆様には御礼申し上げます。研究室に配属されたものの右も左も分からず迷走していた小生に、研究への取り組み方を御教示いただき、また議論の場では多くのご意見もいただきました。

松下研究室第10期の皆様には、同期生として、良き友人として、そして2年間ないしは4年間を共に過ごした戦友として、感謝の気持ちでいっぱいです。丹波でのプロジェクトをはじめ、研究室運営、学会発表など、様々な場面でお世話になりました。

最後になりましたが、大学院進学に理解を示し、学生生活を支えていただいた両親、長姉、次姉に感謝の意を記して謝辞といたします。

参考文献

- [1] Acar, A. and Muraki, Y.: 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).
- [2] Aramaki, E., Maskawa, S. and Morita, M.: Twitter catches the flu: detecting influenza epidemics using Twitter, *Proceedings of the 2011 Conference on empirical methods in natural language processing*, pp. 1568–1576 (2011).
- [3] Eisenstein, J., BrendanO' Connor, Smith, N. A., Xing, E.: A latent variable model for geographic lexical variation, *Proceedings of the 2010 conference on empirical methods in natural language processing*, pp. 1277–1287 (2010).
- [4] Jones, K. S.: Index term weighting, *Information storage and retrieval*, Vol. 9, No. 11, pp. 619–633 (1973).
- [5] Kleinberg, J.: Bursty and hierarchical structure in streams, *Data mining and knowledge discovery*, Vol. 7, No. 4, pp. 373–397 (2003).
- [6] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a social network or a news media?, *Proceedings of the 19th international conference on World wide web*, pp. 591–600 (2010).
- [7] McCormick, T. H., Lee, H., Cesare, N., Shojaie, A. and Spiro, E. S.: Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing, *Sociological Methods & Research*, Vol. 46, pp. 390 – 421 (2017).
- [8] Mizuno, J., Tanaka, M., Ohtake, K., Oh, J.-H., Kloetzer, J., Hashimoto, C. and Torisawa, K.: WISDOM X, DISAANA and D-SUMM: Large-scale NLP systems for analyzing textual big data, *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pp. 263–267 (2016).
- [9] Morino, Y., Megumi, Y., Matsushita, M. and Fujishiro, H.: Investigating the influence of Web-Media in disaster situations by analyzing diffusion of tweets, *The 34th Annual Conference of the Japanese Society for Artificial Intelligence*, The Japanese Society for Artificial Intelligence, pp. 3G5ES103–3G5ES103 (2020).
- [10] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors, *Proceedings of the 19th international conference on World wide web*, pp. 851–860 (2010).
- [11] Toriumi, F., Sakaki, T., Shinoda, K., Kazama, K., Kurihara, S. and Noda, I.: Information sharing on Twitter during the 2011 catastrophic earthquake, *Proceedings of the 22nd International Conference on World Wide Web*, pp. 1025–1028 (2013).

- [12] Wilensky, H.: Twitter as a navigator for stranded commuters during the great east Japan earthquake., *Proceedings of 11th International Conference on Information Systems for Crisis Response and Management*, pp. 695–704 (2014).
- [13] 足立義則: 震災ビッグデータからソーシャルリスニングへ. http://www.nhk.or.jp/bunken/book/regular/media/media11/2_06.pdf.
- [14] 足立義則: NHK「SoLT」によるファクトチェックの現況 (2017). <https://dot.asahi.com/galac/2017042400078.html>.
- [15] 荒牧英治, 若宮翔子: ツイート数と現実の統計量との差異に関する検討, *統計数理*, Vol. 64, No. 2, pp. 233–246 (2016).
- [16] 石橋雄一, 原敦子, 岡安勲, 栗原考次: 画像特微量とテキスト情報による画像検索可能な病理情報データベースの構築, *計算機統計学*, Vol. 24, No. 1, pp. 3–21 (2011).
- [17] 伊藤淳, 西田京介, 星出高秀, 戸田浩之, 内山匡: Twitter と Blog の共通ユーザプロフィールを利用した Twitter ユーザ属性推定, *研究報告情報基礎とアクセス技術 (IFAT)*, Vol. 2013, No. 4, pp. 1–8 (2013).
- [18] 大島裕明, 中村聡史, 田中克己: SlothLib: Web 検索研究のためのプログラミングライブラリ, *日本データベース学会*, Vol. 6, pp. 113–116 (2007).
- [19] 大竹清敬: 災害時における DISAANA、D-SUMM の活用～DISAANA・D-SUMM と防災チャットボット SOCDA～(2020). https://www.soumu.go.jp/main_content/000672984.pdf.
- [20] 盛浩小笠原, 浩誉川島, 裕之藤代: マスメディア報道は Twitter 上の災害時流言を抑制できたか? : 2011 年東日本大震災におけるコスモ石油流言の定性的分析, *関西大学社会学部紀要*, Vol. 49, No. 2, pp. 121–140 (2018).
- [21] 國友優, 神山嬢子: Twitter 情報を活用した土砂災害の前兆・発生状況把握の可能性, *土木技術資料=Civil engineering journal: 土木技術の総合情報誌*, Vol. 57, No. 9, pp. 18–21 (2015).
- [22] 蔵内雄貴, 内山俊郎, 内山匡: マルコフ確率場を用いたソーシャルネットワークからのユーザ属性推定, *電子情報通信学会論文誌 D*, Vol. 96, No. 6, pp. 1503–1512 (2013).
- [23] 榊剛史, 松尾豊: ソーシャルセンサとしての Twitter: ソーシャルセンサは物理センサを凌駕するか?(<特集>Twitter とソーシャルメディア), *人工知能学会誌*, Vol. 27, No. 1, pp. 67–74 (2012).
- [24] 佐藤翔輔, 今村文彦: 2018 年西日本豪雨災害における「#救助」ツイートの実態:2017 年 7 月九州北部豪雨災害との比較分析, *自然災害科学*, Vol. 37, No. 4, pp. 383–396 (2019).
- [25] 宋晨潔, 藤代裕之: 救助要請ツイートの特徴の検証—令和 2 年 7 月豪雨を対象に—, *電子情報通信学会技術研究報告; 信学技報*, Vol. 120, No. 166, pp. 18–23 (2020).

- [26] 総務省消防庁: 大規模災害時におけるソーシャル・ネットワーキング・サービスによる緊急通報の活用可能性に関する検討会報告書(案) (2013). https://www.fdma.go.jp/singi_kento/kento/items/sns04/shiryo02.pdf.
- [27] TwitterJapan: 災害に備えるための Twitter 活用法 (2019). https://blog.twitter.com/ja_jp/topics/company/2019/how-to-utilize-twitter-during-naturaldisaster.
- [28] 鳥海不二夫, 榊剛史: バースト現象におけるトピック分析, 情報処理学会論文誌, Vol. 58, No. 6, pp. 1287–1299 (2017).
- [29] 内閣官房情報通信技術総合戦略室: 災害対応における SNS 活用ガイドブック (2017). http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/pdf/h2903guidebook.pdf.
- [30] 内閣官房情報通信技術総合戦略室: 災害対応における自治体調査 2019 年度 (2019). https://www.kantei.go.jp/jp/singi/it2/senmon_bunka/pdf/2019SNSjtititai_chousa.pdf.
- [31] 内閣府: 令和 2 年 7 月豪雨による被害状況等について (2021). http://www.bousai.go.jp/updates/r2_07ooame/pdf/r20703_ooame_40.pdf.
- [32] 橋田修一, 田村慶一, 酒井達弘: 畳み込みニューラルネットワークを用いた観光ツイートの分類手法, 人工知能学会全国大会論文集第 32 回全国大会 (2018), 一般社団法人人工知能学会, pp. 20401–20401 (2018).
- [33] 橋本康弘, 岡瑞起: 都市におけるジオタグ付きツイートの統計 (<特集>人と環境に見る高次元データフローの生成と解析), 人工知能, Vol. 27, No. 4, pp. 424–431 (2012).
- [34] 福長秀彦: 「北海道胆振東部地震」と流言の拡散 SNS 時代の拡散抑制を考える, 放送研究と調査, Vol. 69, No. 2, pp. 48–70 (2019).
- [35] 藤代裕之, 松下光範, 小笠原盛浩: 大規模災害時におけるソーシャルメディアの活用—情報トリアージの適用可能性, 社会情報学, Vol. 6, No. 2, pp. 49–63 (2018).
- [36] 梅島彩奈, 宮部真衣, 荒牧英治, 灘本明代: 災害時 Twitter におけるデマとデマ訂正 RT の傾向, 研究報告 情報基礎とアクセス技術 (IFAT), Vol. 2011, No. 4, pp. 1–6 (2011).
- [37] 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 = The journal of the Institute of Electronics, Information and Communication Engineers, Vol. 95, No. 3, pp. 219–223 (2012).
- [38] 森國泰平, 吉田光男, 岡部正幸, 梅村恭司: ツイート投稿位置推定のための単語フィルタリング手法, 情報処理学会論文誌データベース (TOD), Vol. 8, No. 4, pp. 16–26 (2015).