

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の可能性

The potential of removal for Noise on SNS during disaster considering multimodal information

Twitter に投稿された災害情報を効率的に取得するためには、災害に関連した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や救助要請とは直接関わらないノイズとなる情報を取り除くが必要になる。本研究では、画像とテキストという複数のモダリティに着目することで、単純なテキスト処理や、画像の参照だけでは除くことが困難なノイズとしての情報を明らかにする。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。

1. はじめに

本研究では、ソーシャルメディアから災害時の救助・被害を効率的に情報抽出するために、被害情報とは無関係なデータ=ノイズ=を除去するデータクレンジングについて検討する。

東日本大震災の際、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も増加している (NHK2018)。

ソーシャルメディアの一種であるソーシャル・ネットワークワーキング・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、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 の運用を行っている (内閣官房情報通信技術総合戦略室 2019)。例えば、2019 年台風 19 号の影響で長野県の千曲川が氾濫した際に、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 SNS 活用の具体例として挙げられる (NHK2019)。

しかし、災害時に SNS から情報収集を行うことには課題もある。専属の自治体職員がつきっきりで情報収集し、災害と関係がある投稿か判断を行い、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。上記の 2019 年台風 19 号の事例について、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 (NHK2019)。

また、NHK は Social Listening 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。ソルトは、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に機械的に「ふるい」にかけて上で、人が見ただけ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「スタッフや機材を更に増やし、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 (足利義則 2014)。

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害や救助要請の把握が可能になった一

方で、デマ¹の投稿や被災地以外からの投稿や拡散 (福長秀彦 2019)、一般ユーザによるマスメディア報道の再発信などのノイズ情報が引き起こす情報の埋没 (藤代裕之・松下光範・小笠原盛浩 2018) が問題として顕在化し、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。

特に災害時の SNS 上では、常時の投稿に加え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²が発生することが指摘されている

(Kleinberg2003, Toriumi・Sasaki・Shinoda・Kazama・Kurihara・Noda 2013)。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災害に関連しないノイズ情報もふんだんに含んだ 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、人的労力を削減しつつも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³。

こういった問題に対する一つの取り組みとして、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 (大竹清敬 2020)。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。

1.1 本研究の目的

つまり、SNS 上には多くのノイズとなる被害とは無関係な情報が多く存在するため、そもそも災害と関連したツイートなのか確認を行う必要がある。そのためには人的リソースを大量に確保する必要があり、信頼性を判断する必要がある。上述のように、災害時に被害や救援要請といった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、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、ノイズ情報を削除しデータをクレンジングを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。このようなシステムを実現する端緒として、SNSの一つである 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、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に着目し、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することで、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う。各クラスの特徴を算出し、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。

2. Twitter における救助要請・被災情報抽出

Twitter は、匿名性の高い SNS として知られ、2017 年の時点で 4500 万人ものユーザが利用している。“ツイート”と呼ばれる投稿には 140 字の上限があり、情報の受信が容易である。また、ボタン一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。

こうした特徴から、Twitter は災害時の救助・被害情報共有に使われることがある。Twitter Japan 社は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⁴。

また、Twitter Japan 社はウェザーニューズ社と連携し、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は、災害被害

や救助を求める声とは無関係な情報がノイズとして混在し、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、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ことが多い（図 1-3 参照）。こういったニュース映像の切り抜き投稿はあくまで二次的な被害報告であり、機械的に分類を行う上でノイズとなりえる。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。

理想的な災害情報へのアクセスは、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である。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。

3. 関連研究

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、Twitter の投稿からインフルエンザの流行を可視化したり(Aramaki・Maskawa・Morita 2011)、地震の発生を観測したり(Sakaki・Okazaki・Matsuo 2010)するなどのように、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(榊剛史・松尾豊 2012)。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など工夫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。

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（佐藤翔輔・今村文彦 2019）。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。

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用いて、観光に関するツイートの分類を行った（橋田修一・田村慶一・酒井達弘 2018）。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（鳥海不二夫・榎剛史 2017）。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、どのようなユーザー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。

また、Twitter のユーザー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、ユーザーのデモグラフィック属性を推定する手法を提案した（蔵内雄貴・山内俊郎・内山匡 2013）。匿名性の強いソーシャルメディアのユーザーのデモグラフィックを推定可能になれば、災害時のデマの抑制や、デマ伝播のメカニズム解析などに利用することも考えられる。

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ー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、ユーザー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⁵。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（森國泰平・吉田光男・岡部正幸・梅村恭司 2015）。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。

4. 調査方針と対象災害

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、ユーザー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、こと災害時において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、組み合わせで処理するマルチモーダルな技術が必要になってくる⁶。

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。

文章の特徴として、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)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。分類に紐付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。つまり、“被害”に分類されるツイート画像にはノイズとして他のツイートが多く存在する可能性がある。

そのため、人の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5章で、ステップ2については6章で各々述べる。

4.3. 対象の災害と収集データ

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。気象庁の発表によると、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⁷。

対象災害のツイートデータはTwitter API⁸を用いてキーワードをクエリとした収集を行った。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。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



図1 逆L字画像例⁹



図2 テレビ報道をカメラで撮影した画像の例¹⁰



図3 ニュース関連画像の例¹¹

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。

これらのツイートの中から画像・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。

5. ノイズとなる投稿に関する調査

まず、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。

5.1. 投稿画像の分類

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¹⁵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、被害情報の2次情報でありノイズと考えられる例と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。想定する画像例を図1から図3に示す。これらの例に挙げた画像は計算機による画像認識を行った場合、被害画像として認識される可能性があるため、ノイズとして扱う必要がある。

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1)
2. テレビのニュースをカメラで撮影した画像(図2)
3. ディア媒体を問わず、ニュースに関する画像(図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。分類担当者は画像一枚ずつ参照し、その他情報の一切を参照せず分類を行った。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1に記す。

3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。

表 1 人手による画像分類クラス（抜粋）

分類クラス	元となったツイート数
ニュース関連	1546
Twitter	1538
ゲーム	8230
テレビのニュース	1471
アニメ画像	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

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」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。

5.2. テキスト処理

ツイートテキストの傾向分析を行うため、4章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、3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)¹²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プロジェクトでストップワード¹³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して設定した。

表 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	被害	TF-IDF	逆L字	TF-IDF	防災グッズ	TF-IDF
チェイス	0.3242	前代未聞	0.1966	師団	0.4122	坂本	0.2052	早め	0.5147	セット	0.6452
傭兵	0.3007	安倍	0.1903	派遣	0.3264	八	0.1825	岐阜	0.3666	グッズ	0.3340
解説	0.2681	府県	0.1640	陸上	0.2264	筑後川	0.1754	放流	0.1693	楽天	0.3184
祭司	0.2384	東洋	0.1640	神瀬	0.2065	ダム	0.1611	ダム	0.1585	袋	0.2343
即死	0.1950	八	0.1575	実施	0.1883	増水	0.1414	日田	0.1580	用品	0.2254

表 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	救助	Cos類似度	被害	Cos類似度	避難	Cos類似度	風景	Cos類似度
被害	0.4728	避難	0.4206	被害	0.5291	天気図	0.9031	自衛隊	0.7939	風景	0.5341	地図	0.4818	被害	0.5341
避難	0.4519	風景	0.3697	天気	0.5234	地図	0.5234	被害	0.3499	地図	0.5291	テレビニュース	0.4519	避難	0.3759
地図	0.4254	テレビニュース	0.3173	天気図	0.5183	避難	0.3455	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
人	0.3173	被害	0.2904	避難	0.4818	風景	0.3294	SNS	0.2139	避難	0.4499	人	0.4206	人	0.3697

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。

$$TF_{i,j} = \frac{N_{i,j}}{X}$$

X はある文章 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、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。

$$IDF_i = \log_e \frac{N}{df_i} + 1$$

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。

$$TFIDF_{i,j} = TF_{i,j} \times IDF_i$$

本稿では、表 1 のクラス 23 種類をもとに、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 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“News23”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。

5.3. 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 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”などがあつた。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

表 4 細分化した“被害”ツイートにおける TF-IDF 上位 5 単語抜粋

被害-テレビニュース	TF-IDF	被害-ネットニュース	TF-IDF	被害-ユーザが撮影	TF-IDF	被害-ライブカメラ撮影	TF-IDF	被害-自治体発信	TF-IDF	被害-分類不可能	TF-IDF	被害-元ツイート消去	TF-IDF
熊本	0.361	熊本	0.395	川	0.291	氾濫	0.3950	災害	0.4000	峡	0.2617	熊本	0.2230
当時	0.311	豪雨	0.28	雨	0.271	水位	0.3402	駐屯	0.2539	ダム	0.2580	お願い	0.2088
園	0.285	大雨	0.202	氾濫	0.248	カメラ	0.3112	豪雨	0.2461	洪水	0.2051	地区	0.2036
寿	0.199	指示	0.182	熊本	0.235	ライブ	0.2847	派遣	0.2377	大雨	0.1864	坂本	0.1998
千寿	0.163	災害	0.179	大雨	0.223	川	0.2581	熊本	0.2256	中国	0.1764	連絡	0.1978

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、1枚の画像を複数のクラスへ分類した場合や1つのツイートに異なるクラスへ分類可能な画像を複数枚添付されていることで同一IDのツイートが他クラス間で存在している場合に、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。

また、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が困難であり、分類エラーを起こす可能性が高いことが示唆された。

6. 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、5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同士を5.2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。

6.1. 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4に示す。

6.2. 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

表 5 “被害” ツイート数とその割合

	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%

“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¹⁴，“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。

6.3. 細分化により得られた知見

6.2 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。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モデルを構築することで、抽出精度向上に繋がると考えられる。

6.4. 複数のモダリティを活用した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 5 に示す。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。

7. 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、投稿された画像を手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。

5 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。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。つまり，“被害”に関する画像以外はノイズとして扱うための特徴抽出が比較的容易であると言える。

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、誤分類を起こ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。

6 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ず、被災者自身が投稿した画像を機械的に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。

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ことが期待される。

8. おわりに

特に災害時の被害情報や救援情報における情報処理では、これらの情報の取りこぼしをすることは、人命救助の観点から望ましくない。また、信頼性判断のためには計算機による情報選別と人による情報選別の協調が重要になる。

本稿では、災害関連ツイートから被害や救助要請とは直接関わらないノイズ情報の除去を目的として、Twitter 上の災害情報に関する調査を行った。本研究では、画像とテキストという複数のモダリティに着目することで、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。これにより、災害関連ツイートから被害や救助要請とは直接関わらないノイズ情報を削除するデータクレンジングの可能性が示された。ただし、ノイズ情報の各モダリティの特徴をどのように反映して行くかは今後検討していく必要がある。

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ピンポイントな抽出を行うためには、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。

注

- (1) 総務省「最新関連デマ情報の出現情報」『平成23年版情報通信白書』1(4)を参照。“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、災害時のツイート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。
- (2) Kleinberg, J. (2003), 鳥海(2017) 参照。ソーシャルメディアにおけるバースト現象とは、特定の話題が大きく取り上げられ、著しく投稿が増える情報の急激な肥大化のことを指す。
- (3) 総務相消防庁「大規模災害時におけるソーシャル・ネットワークキング・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)」を参照。
https://www.fdma.go.jp/singi_kento/kento/items/sns04/shiryo02.pdf (Accessed November 28, 2021)
- (4) Twitter Japan「災害に備えるためのTwitter活用法」
https://blog.twitter.com/ja_ip/topics/company/2019/how-to-utilize-twitter-during-naturaldisaster (Accessed November 28, 2021)
- (5) 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」では2011年から2012年にかけて収集したツイートでは、位置情報付きツイートは約0.18%存在していた。昨今のデータリテラシーの浸透やプライバシーに対する意識変化などから、位置情報付きツイートの比率はさらに低くなっていることが考えられる。
- (6) NTTデータ「マルチモーダルAI技術が拓く未来」を参照。
<https://www.nttdata.com/jp/ja/data-insight/2019/1024/> (Accessed November 28, 2021)
- (7) 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)」を参照。
<https://www.data.jma.go.jp/obd/stats/data/bosai/report/2020/20200811/20200811.html> (Accessed November 28, 2021)
- (8) Twitter API: Twitter社はWebスクレ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APIの利用申請する必要がある点を留意しなければならない。<https://help.twitter.com/ja/rules-and-policies/twitter-api> (Accessed November 28, 2021)
- (9) Tweet例: 図1 逆L字

- (10) Tweet例: 図2 テレビ報道をカメラで撮影した画像例
<https://twitter.com/KANKAI79/status/1279148724766167041> (Accessed November 28, 2021)
- (11) Tweet例: 図3 何らかのメディアのニュース画像例
<https://twitter.com/TAKAYA05378399/status/1279181014711930880> (Accessed November 28, 2021)
- (12) MeCab: 形態素解析器
<https://taku910.github.io/mecab/> (Accessed November 28, 2021)
- (13) 大島裕明・中村聡史・田中克己, 2017, 「SlothLib: Webサーチ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 Letters』6(1): 113-116を参照。ストップワードとは、単語分析において、分析結果に影響を与えない単語を、予め排除しておく単語である。
- (14) 宋晨潔・藤代裕之, 2020, 「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(Web)』120(160): 18-23.
- (15) AI研「クラス分類」参照。本稿での「クラス」は、ある画像に何が写っているかを示す。
<https://ai-kenkyujo.com/term/classification/> (Accessed February 24, 2022)

参考文献

- 森野穰・安尾萌・松下光範・藤代裕之, 2021, 「Twitterに投稿された画像の分類に基づくツイート文の傾向分析—令和2年7月豪雨のツイートデータを対象に—」, 『第13回データ工学と情報マネジメントに関するフォーラム』, No.I25-1.
- 総務省消防庁, 2013, 『大規模災害時におけるソーシャル・ネットワークキング・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』.
- 内閣官房情報通信技術総合戦略室, 2019, 『災害対応における自治体調査(2019年度)』.
- NHK, 2018, 「大災害時の「助けて」ツイートは救助につながったか」『NHK政治マガジン』
<https://www.nhk.or.jp/politics/articles/feature/10863.html> (Accessed November 28, 2021)

- NHK, 2019, 「長野県台風十九号でツイッターの救助要請収集 約 50 件救助に」 『NHK 政治マガジン』
<https://www.nhk.or.jp/politics/articles/lastweek/25652.html> (Accessed November 28, 2021)
- 毎日新聞, 2021, 「教訓はどこに/5 災害救援 SNS の功罪」
<https://mainichi.jp/articles/20210111/ddm/010/040/006000c> (Accessed November 28, 2021)
- 足利義則, 2014, 「震災ビッグデータからソーシャルリスニングへ」 『放送メディア研究』 No.11, pp.290-293
http://www.nhk.or.jp/bunken/book/regular/media/media11/2_06.pdf (Accessed February 10, 2022)
- 福長秀彦, 2019, 「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』 2, pp.48-70.
- 藤代裕之・松下光範・小笠原盛浩, 2018, 「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』 6(2), pp.49-63.
- 大竹清敬, 2020, 「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～」
https://www.soumu.go.jp/main_content/000672984.pdf
(Accessed November 28, 2021)
- 榑剛史・松尾豊, 2012, 「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?—」, 『人工知能学会誌』, 27(1), pp.67-74.
- 鳥海不二夫・榑 剛史, 2017, 「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』 58(6), pp.1287-1299.
- 佐藤翔輔・今村文彦, 2019, 「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』 37(4), pp.383-396.
- 橋田修一・田村慶一・酒井達弘, 2018, 「曇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』 Jsa2018(32): 204-01.
- 蔵内雄貴・山内俊郎・内山匡, 2013, 「マルコフ確立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』, J96-D(6): pp.1503-1512.
- 鳥海不二夫・榑剛史・吉田光男, 2020, 「ソーシャルメディアを用いた新型コロナ禍における感情変化の分析」, 『人工知能学会誌』, 35(4): F-K45_1-7.
- 橋本康弘・岡瑞起, 2012, 「都市におけるジオタグ付きツイートの統計」, 『人工知能学会誌』, 27(4): pp.424-431.
- 森國泰平・吉田光男・岡部正幸・梅村恭司, 2015, 「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』, 8(4), pp.16-26.
- Kleinberg, J., 2003, Bursty and hierarchical structure in streams, *Data Mining and Knowledge Discovery*, 7 (4), pp.373-397.
- Toriumi, F., Sasaki, T., Shinoda, K., Kazama, K., Kurihara, S. and Noda, I., 2013, Information Sharing on Twitter During the 2011 Catastrophic Earthquake, *Proc.22nd International Conference on World Wide Web, WWW'13 Companion*, pp.1025-1028.
- Aramaki, E., Maskawa, S. and Morita, M., 2011, Twitter catches the flu: detecting influenza epidemics using twitter, In: *Proceedings of The Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, pp.1568-1576.
- Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, Earthquake shakes Twitter users: real-time event detection by social sensors. In: *Proceedings of the 19th international conference on World wide web, WWW'10*, pp.851-860, New York, NY, USA. ACM.
- Morino, Y., Megumi, Y., Matsushita, M. and Fujishiro, H., 2020, 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