

災害時におけるソーシャルメディアからの情報収集を目的とした ノイズ情報分離に関する調査

森野 穰*藤代 裕之*松下 光範*¹

Investigation on separation of noise information for information gathering from social media in disaster situations

Yutaka Morino*¹, Hiroyuki Fujishiro*² and Mitsunori Matsushita*¹

Abstract – In the crisis, there is a need for rapid understanding of the disaster situation and accurate information sharing. Social media play an important role in real-time information dissemination by victims. However, disaster-related information contains much noise, such as false and irrelevant information, making it difficult to sort by necessity. To address these issues, it is essential to construct an efficient information sharing interface that combines mechanical information processing and human confirmation. In this study, we analyze tweet data related to the July 2020 heavy rains and attempt to extract valuable and noise information for understanding the disaster situation, focusing primarily on the images in tweets. First, image classification will be conducted manually to classify the posted images. Next, we will perform textualization and clustering of images to examine the possibility of information separation using the CLIP model. Furthermore, based on the classification results, we will discuss the need for an information sharing interface to organize reliable information in a crisis efficiently. The results of this research are expected not only to improve the reliability of disaster information on SNS, but also to contribute to the design of information organization and sharing system between people and computers.

Keywords : Disaster information management, Information triage, Information reliability, Social information system, Twitter(X)

1. はじめに

地震や豪雨などの大規模災害時において、人命救助や被災状況を的確に把握するためには、迅速な情報収集といかに関係者間で情報共有するかが大きな課題である。これらの災害発生時には、行政やマスメディアによる情報発信は災害対応において重要な役割を果たしているが、他方で被災者自身がX（旧 Twitter）などのソーシャルメディアでリアルタイムに被災状況の発信を行うことも一般的となった^{[1][2]}。特にXは、速報性の高い情報を短文テキストに加えて画像を投稿することで一定の信頼性を担保させつつ、被災地の状況を早く広く伝達できるという特徴がある^[3]。こういったSNSの情報を自治体が活用することで実際に救助につながった事例も存在し、その有用性についても認識されつつある^[4]。しかし、SNSに投稿される情報は膨大であり、災害とは無関係なエンターテインメント関連の投稿^[5]、デマや誤情報の善意による拡散や被災地とは遠い地からの日常的な投稿といったノイズ情報も多く含まれている^{[6][7]}。令和6年1月1日に発生

した能登半島地震の際には、実際に救助を必要としたユーザからの投稿があったものの、インプレッション稼ぎと言われる行為や、大量に誤情報の複製・伝播が行われたことも確認されている^{[8][9]}。総務省の調査によると約21,000件の救助要請投稿があり、元の投稿をそのままコピーしたケースが散見された^[10]。今後は、こういった状況への対応がこれまで以上に重要になる。先の自治体による有効活用事例でも、膨大な情報の中からデマへの対応や信頼性を判断するために、専属の職員を多数配備し情報整理を行うことで対応していた。人命救助の観点から情報の取りこぼしはあってはならず、信頼性の観点から人目による情報確認は必要不可欠である。そのため、機械的な情報処理と人による情報処理の双方を活用する人—計算機間の協調が可能な情報処理システムを構築することで、人が見るべき情報へのアクセス支援が必要である。

そこで、本研究では、被災者の発信と思われる災害情報を統合して共有し、かつ投稿された情報の真偽や優先度を協調的に判定することで、質の高い災害情報を協調的に収集する情報トリージング^[11]の実現を目指す。この実現においては、被災地の実情を熟知する現場の人々と、被災地外からの情報提供や支援に関わる

*1: 関西大学大学院 総合情報学研究科

*2: 法政大学 社会学部 メディア社会学科

*1: Graduate School of Informatics, Kansai University

*2: Faculty of Sociology, Hosei University.

遠隔の人々と、計算機とが、互いに連携しながら情報を精査し補完できる仕組みを構築することが重要となる^[12]。そこで、その端緒としてソーシャルメディア上でどのような情報が流通しているか調査を行い、実際の被災者による発信と思われる情報と、災害とは無関係な情報との機械的な分離可能性について検討する。

本稿では、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」を対象として、X（旧 Twitter）上に投稿されたツイートデータの収集・分析を実施する。本稿では、特にツイートに添付された画像に着目し、災害状況の把握に有用な画像・無関係な画像の特徴を明らかにする。まず、人手作業によって分類した大量のツイート画像の特徴を概観し分類を行う。その上で、自然言語と画像の双方を扱うことが可能な CLIP モデルを用いて、画像のテキスト化を行いクラスタリングすることで、被災状況や救助要請に関する情報とノイズ情報をどの程度切り分けられるかを検証する。

2. 関連研究

特に大規模災害時の際には、日常的な投稿に加え、災害情報の拡散による情報のバースト現象も古くから報告されており^[13]、現在でもその状況は変わらない^[14]。日本だけでなく、中国で最も規模の大きいソーシャルメディア Weibo でも同様の状況が報告されている。He らは中国で発生した洪水の際、Weibo 上の救助要請投稿を分析した結果、救助を求める投稿が、無関係な情報に埋もれる、という問題を指摘している^[15]。こういった状況に対処するため、災害時における、ソーシャルメディア利用可能性について検討している研究も多い。国立研究開発法人情報通信研究機構 (NICT) は、旧 Twitter 上の投稿をリアルタイムで収集・分析するインタフェースを構築し、サービスとして提供していた（現在は停止中）^[16]。自然災害や物資の流通、交通事故や火災などに関連する投稿を表示し、その投稿と矛盾すると考えられる投稿を同時に表示することで、人目による真偽判断の手助けを行った。

宋らは、NHK への取材をもとにした調査で明らかとなった、実際に救助に繋がったとされる救助要請ツイートの特徴に着目し、それらの自動抽出を試みているものの、チャットボットやニュースサイトへのリンクが多く含まれていたことを課題として挙げている^[17]。Jain らは、複数の災害に関するツイートデータセットである CrisisMMD を利用し、OpenAI embedding モデルでツイートを分類することで従来の手法と比較して精度が向上することを示し、事前訓練された GPT モデルの重要性について論じている^[18]。Rezk らも CrisisMMD のデータセットを利用し、画像とテキストの両方に重要度を付与する Multimodal Channel Atten-

tion model を構築し、複数のモダリティを統合して分析することを提案している。これにより、テキストのみ/画像のみでの分類精度と比較して精度が向上することを示した^[19]。

また、災害時におけるソーシャルメディア利用には、デマやフェイクへの対応という大きな課題もある。事件発生時や発災時などといった危機的状況では、人は心理的な不安や感情的な理由により誤った情報を信じたり共有したりする行為を取りやすいことが明らかになっている^[20]。Plotnick らの論文では、災害時に人々がどのように偽情報を判断するのかについてのアンケートを行い、集められた 341 件の回答から、情報の内容そのものより人々が真偽を判定する文章の文法やその発信者の信頼性で判断することを明らかにした^[21]。しかし、発災時の情報発信では文法的に正しい文章を書く余裕がなかったり、発信者の信頼性を判断する根拠がなかったりする場合も少なくないと考えられるため、文法や発信者情報を前提とするのは限界があると推察される。川村らは、SNS 上の誤情報の拡散があった場合に、その自動抽出の可能性を示した^[22]。また、自治体へのヒアリングから、防災部局の人手不足や時間不足による情報収集・発信の困難さも課題として挙げられた。藤代は、偽・誤情報とフェイクニュースへの対策には、ネット空間に流通する情報をニュース・コンテンツ・広告のように適切に切り分けることが重要となると述べており、それらへの対策方針を示すことで、伝統メディアの役割の大きさを示唆している^[23]。

上述したように、災害時のソーシャルメディアを取り巻く課題は、情報の膨大化やデマなどへの対応が必要不可欠である。これまで、多くの研究で災害情報の分類や有益情報や偽・誤情報を計算機によって抽出する試みがなされてきた。これに対し本研究では、危機的状況にいる被災者情報の取りこぼしが人命に直接関わる点、最終的な信頼性の評価は人による判断が必要な点を考慮し、有益な情報を直接抽出するのではなく無関係な情報を排除するアプローチを取ることで、最終的な人目による判断を支援することを志向する。つまり、確実にノイズである情報の削除を繰り返し行っていくことで、質の高い災害情報を協調的に収集する情報トリアージシステムの実現を目指す。

3. 調査方針と対象災害

本章では、対象災害のデータおよび、調査方針について説明する。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。自身の被災状況を伝える際には画像を含めた投稿をすることが推奨されている点から、分析対象をツイートに添付された画像とした。まず、より実

表1 人手による画像分類クラス抜粋

分類クラス	画像数
SNS	244
Twitter	654
グッズ	222
ゲーム	8539
二次元画像	2515
デバイス緊急通知	1932
救助	628
テレビ	333
TVNews	504
逆L字	525
WebNews	653
新聞	491
地図	584
天気図	1104
防災グッズ	847
動物	3674
植物	1809
本	304
飲食物	1774
風景	2868
ユーザ撮影被害	2367

災害下の状況を考慮し、収集したツイート画像は人手で分類を行う。次に、自然言語と画像のマルチモーダルな処理が可能なCLIPを用いて、画像のテキスト化とクラスタリングを行うことで、人手による分類と計算機によるクラスタリングの差異を分析する。これにより、どのような情報が災害時に流通するか、またノイズとなる情報の分離は可能か調査を行う。

「令和2年7月豪雨」発生の際のX(旧Twitter)データはTwitter API¹を用いてキーワード「救助」および「避難」をクエリとした収集を行った。データ収集対象期間は、全国を通して最も降水量の多い期間であった令和2年7月4日～7日を含むように、7日から前後一週間を目処に令和2年7月1日～15日とした。「令和2年7月豪雨」のツイート総数は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、合計で

47,434件のツイートとなった。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。

4. データ処理と結果

4.1 豪雨災害画像の人手分類

本調査では、災害時のツイートに含まれる画像の分析手順を述べる。膨大な画像データを分類するため、分類作業は4人の大学生（以下、分類担当者）が行い、一枚の画像に対して分類担当者一人が割り振られた。その後、分類正当性を担保するため、著者による確認後、さらに誤分類データの排除を試みた。はじめに、これらの画像の分類を行う。まず、分類担当者に対して、著者によるランダムサンプリングで多数観測できた「ゲームに関する画像」や、単純な画像処理技術では細かな分類が難しいと想定される「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、「テレビのニュースをカメラで撮影した画像」、メディア媒体を問わずニュースに関連する画像などの提示を行った。分類担当者は、1つのツイートに紐づいている画像（最大4枚）を一枚ずつ目視し、各々の判断で分類を行った。その際、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、分類作業後の画像・動画の総数は292,466枚となった。その後、分類クラス名が表記揺れしていた場合や、同一概念のクラスが複数作成されていた場合を確認したため、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（以下、可読性を考慮して、この画像分類の結果を「分類クラス」と記す。）。次に、これらの分類クラスの中から、災害被害を顕著に示していると考えられる「被害」に分類された画像に対して、必要に応じてツイート本文も参照しながらより詳細な分類を行った。その結果、「被害」に分類された画像には、ユーザが実際に撮影したと考えられる被害画像の他、テレビニュースが報道したと考えられる被害画像やネットニュースが報道したと考えられる被害画像、災害地域のライブカメラが撮影したと考えられる被害画像などがノイズとして混在することが確認された。本研究では、ユーザが実際に撮影したと考えられる被害画像を「被害」分類クラスとして扱う。最後に、人手作業の誤分類を排除するため、分類担当者とは別の12名の大学生に対して、分類クラス内の画像が適切なクラスに割り振られているかの確認を依頼した。これらの分類結果の一部を表1に記す。本研究では表1で示した合計32,571枚の画像を対象に分析を行う。

1: <https://developer.x.com/ja/docs/x-api>(2024/01/22 確認)

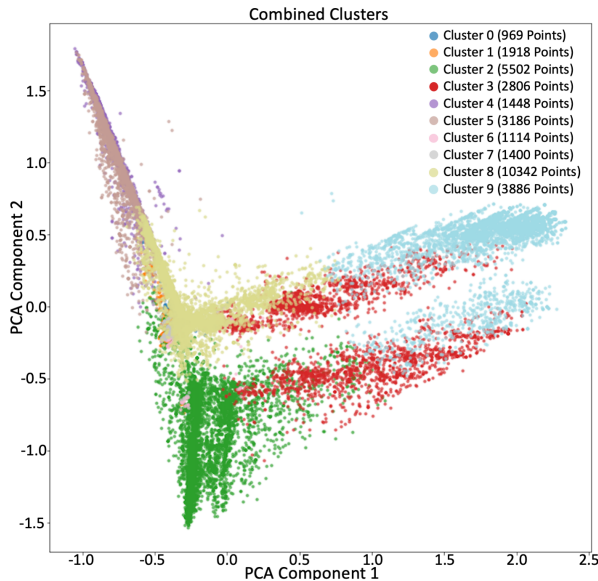


図1 CLIP 結果を用いたクラスタリング散布図

4.2 CLIP による言語化と分類

テキスト情報と画像情報の複数モダリティを扱うことができ、自然言語特徴を用いた画像分類が可能になる CLIP^[24](ver.1.0.) を用いて、分類した画像の image to text 処理を行う。ある画像を入力とし、その画像に写っている可能性の高いものを単語で出力する。その際に出力される単語は New General Service List²内にある単語とし、上位 10 件を抽出した。その CLIP 結果と、その単語の出現数を表 2 に示す。その後、CLIP で算出した確率値を全て 1 に正規化することで one-hot ベクトルとして扱うこととした。これは、CLIP で算出した確率値の分散を排除したうえで分析を行うためである。

次に、これらの結果から k-平均法を用いてクラスタリングを行う。その際、元々の特徴量からより低次元の特徴で表現する PCA 分析を用いて、累積寄与率が 0.9 を超えるように次元削除を行い、エルボー法とシルエットスコアを用いて最適なクラスタ数を算出した。PCA 分析の結果、次元数は 751 となり、最適クラスタ数は 10 となった。各クラスタに割り振られた画像数を表 3 に示す。また、PCA により次元削除されたクラスタリング結果を視覚的に理解するための散布図を図 1 に示す。各点は 1 画像データを表し、クラスタごとに色分けされている。クラスタ 3 や 9 は第 1 主成分（横軸）に広がっており、PCA Component1 が分散の主要な要因であると考えられる。クラスタ 5 や 2 は第 2 主成分（縦軸）に広がっており、PCA Component2 がクラスタ内のばらつきを捉えていることが確認できる。

5. 議論

5.1 人手による画像分類について

本研究では、大学生 4 名による人手分類を実施した後、分類クラスの統合・整理を行うことで分類を行った。これは、SNS にはデマや不確実な情報が出回る点・人命救助の観点から機械処理による取りこぼしを防ぐ必要がある点を鑑み、最終的な情報の確認には人手が必須条件であるためである。その結果、直接災害には関係のないエンターテインメント関連の情報も多く混在することも確認した。これは、ゲームのコンテンツとして“救助”を行ったり、ゲーム内の敵から“避難”するといった行為が見られるためだと考えられる。また、一部の画像では明確に区分することが困難であることも明らかになった。特に、「被害」クラスに関しては、テレビニュース、ネットニュース、ユーザ撮影、ライブカメラといった別クラスの画像が多く混在した。つまり、被災者が自ら撮影した画像とマスメディアが撮影した画像、さらにテレビニュースやネットニュースのスクリーンショット等が被害を示す画像だとして判断される場合があることを示している。どれも被害を示している情報ではあるものの、実際にユーザが置かれている状況やマスメディア等で報道されていない被害の把握のためには、ノイズ情報となる可能性が高い。特に、被災者自らが撮影したと思われる被害状況等を優先的に確認したい場合には、メディアによる二次的な報道画像や過去の報道画像をどのように排除するかが課題となる。このような混在は、災害時の情報の信頼性を評価する上で重要な要素となるため、例えば出典を識別するためのメタデータの作成や、より詳細な分類基準の作成、ノイズ情報の明確化および排除が必要になる。別メディアからの情報を適切に切り分けることは、情報の信頼性の低下を防ぐだけでなく、ソーシャルメディアとマスメディアの情報を相補的に利用者に届けることを可能にする。

本研究では、人手分類の誤りを排除するために別途 12 名の大学生による検証を実施したが、個々の分類担当者の主観的判断による影響は完全には排除できていない。今後、特定のクラス（例：被害・救助など）のみを精度高く選別する際には、複数名のアノテーション結果を統合するといったアンサンブル手法を活用することで、分類の精度向上が期待できる。また、被害の深刻度や救助の緊急性といった定量化が必要な局面では、人手による判断基準の明確化や客観的な評価指標の策定が不可欠である。

5.2 CLIP を用いた画像の text 化について

本研究では、CLIP を用いて画像に写っている可能性の高い単語を抽出し、上位 10 件の単語を用いた。結果を概観すると、一般的な単語（例：「translate」、

² <https://www.newgeneralservicelist.com/> (2024/1/22 確認)

災害時におけるソーシャルメディアからの情報収集を目的としたノイズ情報分離に関する調査

表 2 分類クラス別 CLIP 結果 出現数上位 5 単語抜粋

SNS	出現数	%	Twitter	出現数	%	グッズ	出現数	%	ゲーム	出現数	%	緊急通知	出現数	%	テレビ	出現数	%
translate	219	4.632	translate	592	6.254	toy	108	2.078	tournament	5799	2.527	translate	1843	3.395	episode	141	1.814
remind	88	1.808	account	404	3.306	collection	35	5.157	menu	4889	5.809	text	1111	1.215	broadcast	120	4.014
text	79	1.265	mention	355	1.580	shelf	27	2.156	rank	4463	4.473	application	1016	1.200	video	114	1.616
invite	67	2.134	report	257	1.316	bag	26	6.564	carry	3772	2.057	announcement	974	1.537	movie	98	3.261
conversation	63	5.773	reply	206	0.848	display	24	2.448	game	3640	1.205	report	922	1.098	translate	75	1.703

TVNews	出現数	%	WEBNews	出現数	%	飲食物	出現数	%	逆 L 字	出現数	%	植物	出現数	%	新聞	出現数	%
broadcast	361	3.470	translate	404	2.330	food	1172	3.411	broadcast	448	3.943	plant	1313	4.005	newspaper	481	10.429
situation	229	2.101	report	324	1.576	lunch	1031	5.014	situation	318	2.378	flower	933	6.897	article	479	13.066
news	208	2.462	article	306	4.316	meal	1017	8.238	news	296	2.475	spring	813	2.226	weekly	447	1.903
incident	199	2.351	news	182	1.996	dinner	889	4.242	flood	212	14.580	grow	764	5.029	publication	354	1.560
reporter	159	3.774	situation	169	1.558	eat	776	1.662	incident	209	2.486	bunch	724	1.709	edition	346	1.339

地図	出現数	%	天気図	出現数	%	動物	出現数	%	二次元画像	出現数	%	被害	出現数	%	風景	出現数	%
map	525	4.467	forecast	1064	18.671	cat	1531	2.496	translate	849	1.815	flood	2066	24.531	weather	529	4.585
nearby	401	4.656	weather	1051	25.621	lucky	1466	1.803	episode	472	2.079	situation	1254	1.099	sky	422	4.901
area	371	1.924	rain	952	7.168	spot	1222	3.223	story	380	1.049	severe	1153	1.173	rain	421	3.909
route	361	6.903	situation	942	1.162	sit	1098	1.599	trigger	338	1.149	stream	1005	4.318	outside	321	1.097
location	343	4.087	map	790	1.731	dog	855	2.790	index	321	1.695	river	977	3.349	wet	287	2.233

防災グッズ	出現数	%	本	出現数	%	救助	出現数	%
supply	503	3.059	translate	180	2.310	operation	441	5.869
commercial	319	3.374	publication	121	2.175	rescue	422	15.023
product	316	1.824	cover	111	4.051	incident	344	2.370
package	312	2.931	novel	107	1.587	army	295	3.099
consumer	285	2.496	book	104	3.986	military	287	2.807

表 3 CLIP 結果を用いたクラスターリング結果

	Cl:0	Cl:1	Cl:2	Cl:3	Cl:4	Cl:5	Cl:6	Cl:7	Cl:8	Cl:9	合計
グッズ	6	1	7	3	0	0	0	1	204	0	222
地図	0	0	25	0	14	15	0	0	530	0	584
SNS	0	0	205	1	2	21	1	0	14	0	244
飲食物	0	30	63	0	0	0	1066	0	615	0	1774
新聞	0	0	417	0	0	12	0	0	62	0	491
防犯グッズ	1	0	129	0	0	1	1	0	715	0	847
本	0	0	185	1	2	0	0	1	115	0	304
救助	0	0	41	0	0	299	0	0	288	0	628
TV	0	0	71	6	3	7	1	0	244	1	333
天気図	0	0	1	0	1064	1	0	0	38	0	1104
Twitter	0	0	570	9	12	10	0	0	53	0	654
被害	0	7	5	0	23	2042	0	0	290	0	2367
植物	3	1525	8	0	1	15	6	1	250	0	1809
TVNews	0	0	115	4	74	146	0	0	163	2	504
動物	929	30	33	0	3	26	33	1373	1247	0	3674
逆 L 字	1	0	78	3	75	211	0	1	153	3	525
WebNews	0	0	366	1	50	93	1	0	139	3	653
風景	5	315	59	3	100	274	3	4	2105	0	2868
2D 画像	18	8	832	18	22	8	2	17	1588	2	2515
緊急通知	0	0	1858	6	2	2	0	0	64	0	1932
ゲーム	6	2	434	2751	3	1	0	2	1465	3875	8539
合計	969	1918	5502	2806	1450	3184	1114	1400	10342	3886	32571

「report」, 「text」) が頻出しており, 画像の内容を具体的に反映していない可能性がある. このことは, 画像内に文字情報が入り込むことが多いこと, CLIP のモデルが災害時の画像に特化した学習を行っていないこと等が, そういった汎用的な単語を出力しやすいことに起因すると考えられる.

“SNS” や “Twitter” といった, ソーシャルメディアのスクリーンショットが多く含まれている分類クラスでは, 「translate」, 「text」といったテキスト情報を示す単語が多く含まれ, 「account」等のソーシャルメディアに起因する語が確認できた. ほぼテキスト情報のみで構成されている“緊急通知”では「translate」, 「text」, 「announcement」, 「report」から確認できるように, テキスト情報の画像であることを示している.

“ゲーム”, “グッズ”, “二次元画像”といった, エ

ンターテインメントに関する分類クラスでは, コンテンツ特有の単語が上位にきている. 最も画像数の多かった「ゲーム」分類クラスの上位単語には「tournament」, 「menu」, 「rank」, 「game」といった, ビデオゲームに関連する画像であることが特定可能な単語が多くを占めている. 特に「tournament」は全画像数 8,539 枚に対して 5,799 枚 (67.9%), 「menu」は 4,889 枚 (57.2%), 「game」は 3,640 枚 (42.4%) で出現していることから, 比較的特定が容易であると考えられる. ただし, これらの単語が特定のゲームでのみ出現する可能性もあるため, あらゆるジャンルのゲーム画像に対して比較していく必要がある. “グッズ”に関しても「toy」や「collection」, 「display」といった, グッズを飾っていることを示している単語が確認できた. 漫画やアニメキャラクターのイラストや, アニメ映像のスクリーンショット・漫画の切り抜き画像が多く含まれている“二次元画像”では, 「translate」が最上位なもの, 「episode」, 「story」といった, エンターテインメントや映像作品を示す単語が確認できた.

“テレビ”, “TV ニュース”, “逆 L 字”といった, テレビ番組やニュースをカメラで撮影した画像が多く含まれている分類クラスでは, 「broadcast」や「episode」, 「news」等の, テレビ映像やニュースに関係する語が確認できた. ただし, 災害被害発生中に放送映像の外側に別途付けられる L 字型の帯が映り込んでいる“逆 L 字”分類クラスに関しては, 逆 L 字自体を示す語は確認できなかった. また, “新聞”, “Web ニュース”, “TV ニュース”といった, ニュースに関する分類クラスでは, 「newspaper」, 「article」, 「news」, 「report」といったマスメディアの発信であることを示す単語が多く確認できた. “地図”, “天気図”といった Google

Map³や Apple 社の天気アプリ⁴, Yahoo!天気アプリ⁵などのスクリーンショット画像が多く含まれている分類クラスでは,「map」が共通して多く出現しているものの,“天気図”では「forecast」,「weather」が最上位と差別化されており,適切に分離できていることが確認できた。

特に,“被害”分類クラスの画像では,「flood」,「situation」といった単語が最上位に出力されており,具体的な被害の様態を示す単語が抽出されていた。「stream」や「river」といった河川を示す単語も上位に確認できることから,豪雨災害時の増水状況や洪水状況を示している可能性が高い。しかし,“被害”分類クラスと混交しやすいと考えられる“風景”分類クラスにおいて,「weather」や「sky」,「rain」といった単語が上位ではあったものの,第7位に「flood」,第9位に「river」,第16位に「stream」が出現したように,出現した単語自体は“被害”と類似していることが確認できた。つまり,画像のみを用いた分類では両者は混交しやすく,“被害”と“風景”画像の分離は困難であった可能性がある。

CLIPでの単語抽出は,抽出する単語が少なければ,より特徴的な単語を抽出しやすく,画像をより簡潔に表現することが可能になるが,クラスごとの差異が小さくなりクラス判別がかえって困難になることも考えられる。一方で,抽出する単語数が多ければ,より細かな分類が可能になるが,低確率な単語が含まれることでノイズ情報が混入する確率が高くなるという特徴がある。そのため,クラスタリングを行うにあたり十分な特徴抽出が可能かつクラスタ特徴が明確になるといった,適切な単語数を決定することが,精度向上に繋がると考えられる。今後,実際の単語数の決定は,異なる単語数を抽出した上でクラスタリング精度の比較を行うことが有効と考えられる。

5.3 クラスタの特徴

CLIPを用いて出力した出現単語を one-hot ベクトル化を行い,それを用いたクラスタリングの結果,クラスタ2(図1内の緑色点)およびクラスタ8(図1内の黄色点)には全ての分類画像種が含まれることが確認された。これらのクラスタには画像の多様性を原因とした,ノイズ情報が多くなっていることが考えられる。そのため,どのような単語がクラスタリングに寄与しているか確認するため,クラスタ2と8の頻出単語を抽出した。その結果を表4に示す。

“被害”分類画像が最も多く割り振られたクラスタ5(図1内の灰桜色)には,災害とは無関係な分類ク

表4 CLIP 結果を用いたクラスタリング結果

クラス2	出現数	クラス8	出現数
translate	5323	location	648
report	1899	episode	640
text	1634	supply	608
announcement	1565	map	584
application	1110	nearby	584
invite	1075	operation	531
story	974	commercial	499
article	970	video	484
weekly	875	broadcast	471
paragraph	867	route	456

ラスの画像がクラスタリングされることは少なかったことが示された。しかし,“被害”の他に“救助”,“TVnews”,“逆L字”,“風景”といった画像が多く含まれる結果となった。これは,マスメディア等で既に報道された情報と,一見“被害”と見間違えような“風景”の画像が誤ってクラスタリングされている可能性があることから,クラスタ5に分類された画像は人目での確認が必要不可欠な情報だと考えられる。

クラスタ2の頻出単語上位には「translate」,「report」,「text」などを確認した。特に「translate」は,なんらかの日本語文字情報が含まれる場合に抽出されていた。つまり,テキスト情報を含む画像が多く分類されている可能性が高く,テキストのみの画像が多かった“デバイスの緊急通知”がクラスタ2に多く分類されていることもそれを裏付けている。一方,クラスタ8に関しては,一位に出現した「location」は648回であり,第十位の「route」も456回と,分散が小さい傾向が見られ,特徴的な単語は確認できなかった。「location」,「map」,「route」といった地理的情報に関連する単語は頻出しているが,「location」は52.9%が,「map」は89.9%が,「route」は79.2%が“地図”分類クラスによるものだった。つまり,このクラスタ8は,“地図”に関連する単語が基準となりつつも,他分類クラスにおいて分類基準とは外れた画像や,地理的関連単語が出現した場合に割り振られることで,ノイズ情報が多く混交した可能性がある。今後,クラスタ2やクラスタ8のように多様な画像が含まれているクラスタに対して,そのクラスタのみで分類を行った,他クラスタリング手法を適応・比較したりすることで,その詳細を明らかにする必要がある。

また,クラスタ2及び8に多種の画像が分類されたのは,クラスタリングを行う際に CLIP 出現単語を one-hot ベクトルとして扱ったことに起因するとも考えられる。これは,CLIP 技術の結果に依存しすぎないための処置であったものの,出現単語の組み合わせ

³ [https://www.google.co.jp/maps\(2025/02/05 確認\)](https://www.google.co.jp/maps(2025/02/05 確認))

⁴ [https://apps.apple.com/jp/app/weather/id1069513131\(2025/02/05 確認\)](https://apps.apple.com/jp/app/weather/id1069513131(2025/02/05 確認))

⁵ [https://weather.yahoo.co.jp/weather/promo/app/\(2025/02/05 確認\)](https://weather.yahoo.co.jp/weather/promo/app/(2025/02/05 確認))

次第では適切にクラスタリングが困難だったことが考えられる。これは、5.2 節で上述したように、各画像を決定するにあたり十分な出力単語数の検討が必要不可欠であると考えられる。今後、クラスタの解釈性を高めるためには、各クラスタの代表的な画像を可視化した上で実際にどのような単語が出力されているか詳細に検証し、どのような単語が分類に与しているのか分析することが求められる。また、クラスタ間の類似度の計算や、近いクラスタの統合可能性について検討することで、より直感的な解釈が可能になると考えられる。

本研究の結果は、災害時にソーシャルメディア上で流通する情報の特性を検証し、ノイズ情報の分離可能性を検証することで、有益な情報へのアクセス支援の可能性を示している。CLIP を用いた画像のテキスト化では、災害時の投稿画像を分類クラス別に特徴づけた単語が確認できたものの、「translate」といった汎用性の高い単語が複数同時に出現している分類クラスや、単語分布が類似している分類クラスも確認された。これは画像のみで分類を行うことの限界点を示しているとも考えられる。CLIP の結果を用いたクラスタリングでは、大量に投稿されたエンターテインメント関連の画像や、植物、動物といった、災害とは全く無関係な情報の分離が可能であることを示唆している。これは、図 1 から観察できるように、クラスタリングにおける第一主成分が分散に大きな影響を与えているクラス 3 及びクラス 9（横軸に大きく配置されている）にはゲームに関する投稿が多く含まれていることから裏付けされる。ただし、クラス 2 及びクラス 8 のクラスタリングでは全分類クラスの画像が確認でき、one-hot ベクトル化による情報の単純化が影響している可能性があり課題も残る。本研究のノイズ分離アプローチは、災害時における情報の信頼性向上と、迅速な情報共有を促進するための基盤となる可能性を持つ。今後、クラス 2 及びクラス 8 において、分離しきれなかったノイズ情報の詳細化や、本調査で得られた結果を用いて、被災者や自治体/消防等が活用可能な、ノイズ情報を適切に分離することが可能な情報共有インタフェースの設計について検討を進める必要がある。

6. おわりに

本研究は、災害時の被害情報共有システムの構築を目指し、災害時に流通する情報について概観するとともに、それらの分類と災害被害情報とは無関係なノイズ情報の分離可能性について検討した。また、本研究は、ソーシャルメディア投稿のクラスタリング精度向上を目指すものではなく、クラスタリングの結果によって得られた各分類クラスの特徴を観察することを

主としており、これにより災害時に流通した画像情報の整理を行った。今後、ノイズ情報の詳細をより明確化し、投稿に対する信頼性等の評価を人が行う仕組みを発展させることで、人—計算機間の災害情報共有システムの構築が期待される。

利益相反の開示

本研究は JST RISTEX（課題番号 JPMJRS23L2）の支援を受け実施された。

参考文献

- [1] Takahashi, B., Tandoc, E. C. and Carmichael, C.: Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines, *Computers in Human Behavior*, Vol. 50, pp. 392–398 (2015).
- [2] Zou, L., Liao, D., Lam, N. S., Meyer, M. A., Gharaibeh, N. G., Cai, H., Zhou, B. and Li, D.: Social media for emergency rescue: An analysis of rescue requests on Twitter during Hurricane Harvey, *International Journal of Disaster Risk Reduction*, Vol. 85, p. 103513 (2023).
- [3] Zubiaga, A. and Ji, H.: Tweet, but verify: epistemic study of information verification on Twitter, *Social Network Analysis and Mining*, Vol. 4, No. 1, pp. 1–12 (2014).
- [4] JichitaiWorks: Encouraged residents through SNS during a typhoon and quickly and safely rescued about 50 people., JICHITAI WORKS INC. (online), available from (<https://jichitai.works/article/details/330>) (accessed 2025-02-02). (in Japanese).
- [5] Morino, Y., Matsushita, M. and Fujishiro, H.: Vocabulary cross-contamination between entertainment content and disaster-related social media posts, *2024 International Conference on Information and Communication Technologies for Disaster Management*, pp. 1–6 (2024). (in Japanese).
- [6] Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A. and Chakraborty, B.: A review on application of data mining techniques to combat natural disasters, *Ain Shams Engineering Journal*, Vol. 9, No. 3, pp. 365–378 (2018).
- [7] Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A. and Anderson, K.: Natural language processing to the rescue? extracting “situational awareness” tweets during mass emergency, *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 5, No. 1, pp. 385–392 (2011).
- [8] KyodoNews: Firefighters dispatched with false information immediately after the Noto earthquake; false rescue request on SNS, Kyodo News (online), available from (<https://nordot.app/1128295158484008985>) (accessed 2025-02-02). (in Japanese).
- [9] NHK: False information about the Noto Earthquake, many impression zombies from overseas, NHK (online), available from (<https://www3.nhk.or.jp/news/html/20240202/k10014341931000.html>) (accessed 2025-02-02). (in Japanese).

- [10] 総務省: 偽・誤情報に対する X の取り組みについて, デジタル空間における情報流通の健全性確保の在り方に関する検討会 (第 15 回) 資料 1523, (オンライン), 入手先 (https://www.soumu.go.jp/main_content/000938666.pdf).
- [11] Fujishiro, H., Matsushita, M. and Ogasawara, M.: Effective Use of Social Media in Large-Scale Disasters: The Applicability of Information Triage, *Social Informatics*, Vol. 6, No. 2, pp. 49–63 (2018). (in Japanese).
- [12] Yasuo, M., Morino, Y. and Matsushita, M.: Media Characteristics of Social Network Services in Collecting Disaster Information, *Proc. the 30th Conference on Interactive Information Access and Visualization Mining*, The Japanese Society for Artificial Intelligence, pp. 47–54 (2023). (in Japanese).
- [13] Kleinberg, J.: Bursty and hierarchical structure in streams, *Data mining and knowledge discovery*, Vol. 7, No. 4, pp. 373–397 (2003).
- [14] TheMainichi: Kumamoto Earthquake: There are 26.1million tweets a week, , available from (<https://mainichi.jp/articles/20160519/k00/00m/040/059000c>) (accessed 2025-02-02). (in Japanese).
- [15] He, C., Deng, Y., Yang, W. and Li, B.: "Help! Can You Hear Me?": Understanding How Help-Seeking Posts are Overwhelmed on Social Media during a Natural Disaster, *Proc. ACM Hum.-Comput. Interact.*, Vol. 6, No. CSCW2 (online), available from (<https://doi.org/10.1145/3555147>) (2022).
- [16] 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, *Proc. the 26th International Conference on Computational Linguistics: System Demonstrations*, pp. 263–267 (2016).
- [17] Song, C. and Fujishiro, H.: An examination of the features of t rescue-request tweets—with the case of torrential rainfall in July 2020, *IEICE Technical Report*, Vol. 120, No. 166, pp. 18–23 (2020). (in Japanese).
- [18] Jain, T., Gopalani, D. and Kumar Meena, Y.: Informative task classification with concatenated embeddings using deep learning on crisisMMD, *International Journal of Computers and Applications*, pp. 1–18 (2025).
- [19] Rezk, M., Elmadany, N., Hamad, R. K. and Badran, E. F.: Categorizing crises from social media feeds via multimodal channel attention, *IEEE Access*, Vol. 11 (2023).
- [20] Gupta, A., Lamba, H. and Kumaraguru, P.: \$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter, *2013 APWG eCrime Researchers Summit*, pp. 1–12 (2013).
- [21] Plotnick, L., Hiltz, S. R., Grandhi, S. A. and Dugdale, J.: Real or Fake? User Behavior and Attitudes Related to Determining the Veracity of Social Media Posts, *CoRR*, Vol. abs/1904.03989 (online), available from (<http://arxiv.org/abs/1904.03989>) (2019).
- [22] Kawamura, T. and Sasaki, Y.: Spreading Misinformation on SNS in the Event of a Great Disaster and Problems with Action by Local Government, Technical Report 2, Hokkaido Research Organization Building Research Department Northern Regional Building Research Institute (2022).
- [23] Fujishiro, H.: Direction of countermeasures against false/misinformation and fake news, *IPSJ SIG Technical Report*, Vol. EIP-104, No. 20 (online), available from (<http://arxiv.org/abs/1904.03989>) (2024). in Japanese.
- [24] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. and Sutskever, I.: Learning transferable visual models from natural language supervision, *International conference on machine learning*, pp. 8748–8763 (2021).

(2025 年 2 月 10 日受付)

著者紹介

森野 穰



2020 年関西大学総合情報学部卒業, 2022 年関西大学大学院総合情報学研究科博士課程前期課程修了, 関西大学大学院総合情報学研究科博士課程後期課程入学. 災害時の Twitter を対象にした, 人—計算機間の協調に関する研究に従事.

藤代 裕之



1996 年広島大学文学部哲学科卒業. 2007 年立教大学 21 世紀社会デザイン研究科博士前期課程修了. 徳島新聞, NTT レゾナントを経て, 2013 年法政大学社会学部准教授, 2020 年同教授. フェイクニュースなどソーシャルメディア研究に従事.

松下 光範



1995 年大阪大学大学院基礎工学研究科物理系専攻制御工学分野博士前期課程修了. 同年日本電信電話株式会社入社. 2008 年関西大学総合情報学部准教授. 2010 年同教授. 現在に至る. インタラクションデザインに関する研究に従事. 博士 (工学). 2003 年情報処理学会論文賞, 2007 年バーチャルリアリティ学会論文賞, 2013 年 Laval Virtual Award ほか各賞受賞. 電子情報通信学会, 情報処理学会, 人工知能学会, 芸術科学会, ACM 各会員.