

災害情報収集におけるエンタテインメントコンテンツによる コンタミネーションの調査

森野 穰^{1,a)} 松下 光範^{1,b)}

概要：本研究の目的は、災害時の Twitter を利用した情報収集における、エンタテインメントコンテンツのコンタミネーションについて調査し、その影響を防ぐ方法について明らかにすることである。災害時において、自治体は Twitter を活用することで、被害状況や救助要請の把握につなげている。しかし、Twitter には災害時であっても多くのエンタテインメントコンテンツに関わる情報が投稿され、災害情報収集の妨げになっているという懸念がある。エンタテインメントに関連する投稿は、そのコンテンツやコミュニティの活性化につながっている一方で、災害時という特殊状況下においては、自治体による災害情報収集の障害となり得る。我々は、このような状態をエンタテインメントコンテンツによるコンタミネーションと定義し、その分離を目指す。その端緒として本稿では、災害時のツイートの中からエンタテインメントコンテンツに関連する投稿を分析し、1) どのようなコンテンツがコンタミネイトし易いのか、2) それは分離可能か、の 2 点について検討する。

キーワード：災害情報収集、ノイズ情報の除去、情報トリアージ

1. はじめに

東日本大震災の発災時、輻輳などで繋がりにくくなった電話通信回線の代わりにソーシャルメディアによる情報発信・共有が活用され注目を集めたことを契機に、その災害時の救助活動や災害支援への活用が期待されている。すでに自治体は実災害時に、不特定多数への情報発信が可能であり拡散力が高い Twitter などの Social Networking Service (以下、SNS) を用いて情報収集・発信を行っており、被害状況の確認や救助要請の把握といった事例が報告されている。例えば 2019 年台風 19 号の事例では、長野県自治体が Twitter の投稿を通じて 50 件の救助に繋げていたことが報告されている [1]。こうした利点がある一方で、SNS による情報収集は専属の職員を配備する必要があり、人的負担が大きいという問題が指摘されている [1]。これは、デマや事実と異なる投稿が存在することで情報の信頼性の担保が難しい [3] ことや、被災地以外からの無関係な投稿による情報過多が発生し、人目による情報確認を行うことが求められるためである。

この情報過多の原因の一つに、SNS の日常的な利用の多

様性が挙げられる。SNS 上では様々な情報のやり取りが日常的に行われている。特にエンタテインメントと SNS の親和性は高く、通常時における Twitter 投稿の 6 割はエンタテインメント情報であると言われている [2]。実際、災害時においても趣味や娯楽に関する情報が多く投稿されており、自治体による情報収集の際に、迅速な対応を困難にしている。SNS からの災害情報の収集では、災害に関するキーワードやハッシュタグ (例えば「救助」「被害」「支援」など) が用いられるが、それらはゲームやアニメなどでも頻繁に用いられる語彙であるため、選択的な情報収集を難しくしている。災害時の救助や支援を求めるという行動は「ユーザが要求している」という点でゲーム内行動の構造と類似しており、自治体が実際に行っているようなキーワードマッチングでの情報収集では分離することが困難である。

災害時のソーシャルメディアにおいて、エンタテインメントコンテンツの投稿を抑制するのではなく、それらを適切に切り離すことが可能になれば、自治体が災害情報を収集する際の人的負担軽減に繋がると考えられる。我々は、災害時のソーシャルメディアにおいて、災害被害とは無関係なエンタテインメントコンテンツが多く投稿されている状況を「エンタテインメントによるコンタミネーション」と定義し、災害情報との分離による解消が可能かを検討する。

¹ 関西大学
Kansai University, 2-1-1 Ryozenji, Takatsuki, Osaka 569-1095, Japan

a) k790414@kansai-u.ac.jp

b) m_mat@kansai-u.ac.jp

表 1 SVM の精度

	1	2	3	4	5	6	7	8	9
precision	0.743	0.721	0.921	0.729	0.943	0.707	0.87	0.907	0.843
recall	0.889	0.878	0.772	0.953	0.579	0.961	0.847	0.858	0.901
f1-score	0.809	0.792	0.840	0.826	0.717	0.815	0.859	0.882	0.871

2. 対象のデータと分析手法

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」を分析対象とする。TwitterAPIを用いてキーワード「救助」「避難」をクエリとして収集した476827件のツイートの中から、画像が添付されているツイート(47434件)を抽出した。このツイートデータに対し画像とテキストの両方を人目で確認した上で、あらかじめクラス分類を行ったデータを利用する[4]。対象のクラスは“1:二次元画像”(1571件)、“2:被害”(4135件)、“3:ゲーム”(6855件)、“4:ユーザによって撮影された被害”(2689件)、“5:グッズ”(746件)、“6:TVニュース”(1202件)、“7:逆L字”(325件)、“8:救助”(641件)、“9:ライブカメラ”(209件)の9クラスとした。これらのツイートを教師データとして、Support Vector Machine(以下、SVM)を用いて機械的にカテゴリ分類を行うことで、災害情報にコンタミネーションしたコンテンツ情報の分離が可能か検討する。

3. SVMによるクラス分類

災害時において、迅速な情報収集を行うためのモデルを膨大なツイートを用いて構築することは、時間的制約の観点から現実的ではない。そこで本研究は、他の機械学習モデルと比べて比較的データ量が少なくても結果が出やすいSVMを用いて分類を行う。各データ数の正規化は、最もツイート数の少ないライブカメラクラスに合わせ各200件とし、合計で1800ツイートを教師データとして利用する。分類を行うために教師データを各カテゴリで学習データ7割、テストデータ3割に分割し、ツイート本文に対しMeCab(Ver.0.996)を用いて形態素解析を行い、名詞を抽出した。抽出された語を特徴量としてBag-of-Words法によるベクトル化を行った。5分割交差検定を行い、GridSearchによりベストパラメータを決定した。SVMのコストパラメータとガンマ値の候補を0.01, 0.1, 1.0とし、カーネルの候補にはlinearとrbfを用いた。各パラメータの事前検証の結果、コストパラメータに1.0、ガンマ値に0.1、カーネルはrbfを利用するのが最適値となった。ベストパラメータを用いた場合の各々のクラスのprecision, recall, f1-scoreを表1に示す。全体の正解率(Accuracy)は0.820となった。Accuracyの算出方法は以下のとおりである。

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

4. 結果と考察

災害時において、コンタミネーションが発生するエンタ

テインメントとして、ゲームやイラスト・漫画(二次元画像)やグッズに関する情報があるため、これらを分析対象として選定した。特にゲームに関するツイートは最も多く、SVMによる判定においてもTPが全分類中最も高い129件、FNが11件となった。二次元画像に関しては、TPが104件、FNが36件となった。グッズに関係するツイートのTPは132件と、今回の分類において最も高い数値となったが、FPが96件と非常に多い結果となった。ユーザが撮影した被害の画像に関するツイートのTPは102件、FNは34件(うち、14件はグッズ)となり、FPは5件(全て被害と混合)していた。被害に関するツイート(実被災ユーザではない)のTPは101件、FNは39件(うち11件はグッズ)となった。

ゲームに関する情報はTPやprecisionの値が高いことから、分離できる可能性が高いと言える。ただし、グッズに関する情報は、TPは高いものの、他のクラスのツイートをグッズと分類される傾向が高く、最もコンタミネーションが起きる可能性が高い情報と言える。これは、グッズというコンテンツが、アニメコンテンツや救助カーのおもちゃ、ライブタオルなど、多岐に渡ったためだとも考えられ、より詳細な分類が必要と考えられる。災害時において、人命救助の観点から、被害情報や救助要請は確実に収集される必要がある、クラス4:ユーザが撮影した被害の画像に関するツイートのRecallが高いことが求められるが、precisionやaccuracyに関しては必ずしも1.0である必要はなくFPおよびTNが増えたとしても、情報の削減にはつながる。つまり、クラス4におけるTNおよびFNの件数を減らすことが、被害情報の遺漏を防ぐためには必要となる。

5. おわりに

本稿では、自治体による災害時の情報収集をより効率的にするため、SNSにおけるエンタテインメントによるコンタミネーションに着目し、その調査と分離可能性について検討した。

参考文献

- [1] 日本放送協会: 長野県ツイッターの救助要請収集で約50件救助に, <https://www.nhk.or.jp/politics/articles/lastweek/25652.html> (2019).
- [2] NECビッグロブ株式会社: 東日本大震災におけるツイッターの利用状況について~新たな情報摂取・共有スタイルの定着~, <https://www.biglobe.co.jp/pressroom/release/2011/04/27-1> (2011).
- [3] 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).
- [4] 森野 穰, 安尾 萌, 松下光範, 藤代裕之: 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, I-25.1 (2021).