

京都大学	博士 (工学)	氏名	石 咏雪
論文題目	ANALYZING AND CATEGORIZING FLOOD DISASTER-RELATED TWEETS FOR EMERGENCY RESPONSE (危機対応を目的とした洪水災害関連ツイートの分析と分類)		
<p>(論文内容の要旨)</p> <p>本論文は、洪水災害時の危機対応に活用することを目的として、洪水に関連するソーシャルメディアの情報を分析するとともに、その内容に応じて多数のツイート情報を自動的に分類する方法を提案したものである。2015年関東東北豪雨による鬼怒川の災害を対象にした実証的な研究をまとめており、以下の6章から構成されている。</p> <p>第1章は序論であり、ソーシャルメディアの防災活用について、既往の研究を取りまとめるとともに、それらを踏まえて本研究の方向性を示している。まず、防災マネジメントサイクルにおける事前準備、危機対応、復旧・復興フェーズのうち、特に水害の危機対応の重要性に着目し、実時間情報としてのソーシャルメディアの活用の可能性と意義について検討している。すなわち、それらの情報が、水害の状況把握に用いられる可能性と、被災地の住民がより現実的な危機意識を持つために用いられる可能性があることに言及したうえで、ツイッターなどを用いた既往研究を取りまとめている。そのうえで、水害直後に発信されたツイートを機械学習の分類に適するような形で整理する方法を提示し、複数の機械学習法で水害時のツイートを分類することを主な目的として掲げている。</p> <p>第2章では、鬼怒川洪水直後に発信された100,000件以上のツイート情報を調べて、そのうち被災地やその周辺域から発信されたと思われる有益なツイートを104件(以降の章では170件)抽出している。抽出方法には、ツイッターのApplication Programming Interface (API)、情報通信研究機構が開発した対災害SNS情報分析システム(DISAANA)、ツイッターの高度検索を利用している。このサンプル分析を通して、水害時のツイートの内容で有益なものは、1)浸水状況に関する内容、2)被害状況に関する内容、3)救助や支援に関する内容に大きく分類されることを示している。またサンプル分析の結果から、これらのツイートの中には、鬼怒川、浸水、水深、破堤など頻出する特徴的なキーワードが含まれていることを見出し、これらのキーワードを利用した効率的なツイートの分類法が構築できれば、危機対応に活用できる可能性があることを示している。</p> <p>第3章では、上記の方法で抽出した170のツイートを対象に、対応分析と判別分析を実施している。まず、170のツイートを、その内容に応じて上記の1)~3)に分類して教師データを作成した上で、各分類のツイートが、どのようなキーワードを含むかを分析している。次に、様々なキーワードのうち、例えば破堤と決壊のように類似のキーワードを統合し、12のキーワード群に整理し直すことで、各ツイートが12次元のベクトル(各要素は1か0)で表現されるよう情報を集約する方法を提案した。対応分析の結果、どのようなキーワード群が同じツイート内に現われる傾向にあるか、またどのようなキーワード群が現われる場合に、上記の3つに分類される傾向にあるかを明らかにしている。さらに、判別分析の結果、57%の正解率で事前に定めたカテゴリーに分類されることを示している。</p> <p>第4章では、判別分析の他にも様々なテキスト分類の方法でツイートの分類を行い、各方法の分類精度や速度などについて検討を加えている。具体的には、判別分析の他にSupport Vector Machine, Naive Bayes, K Nearest Neighbours, Decision Tree, Random Forest, Logistic Regressionの6つの方法の特徴を整理するとともに、上記170のツイートを教師データと検証データに分けて、どの分類方法が適するかを調べた。その結</p>			

京都大学	博士 (工学)	氏名	石 咏雪
<p>果、Naive Bayes を除く 5 つの方法は同程度の精度となり、概ね 60 %以上の正解率が確保されていることが示された。また 170 のツイートを含めてテストデータとして使用して、別途入手した 7,269 の鬼怒川洪水に関連するツイートの分類を試みた結果、特に Support Vector Machine と Decision Tree の方法が有効であることを明らかにした。</p> <p>第 5 章では、実際に水害現象が進行し、関連するツイートがなされるようになった段階では、教師データとなり得るリアルタイムの情報が少ないことに着目し、半教師付き学習を試行し、上記の教師付き学習の方法と比較している。具体的には、事前に分類する教師の数を変化させて半教師付き学習を実施し、第 4 章で検討した SVM の手法と比較している。その結果、15 個のツイートを分類した上で半教師付き学習を実施することにより、教師付きと同等か場合によってはそれ以上の分類結果が得られることを明らかにした。</p> <p>第 6 章は結論であり、本論文で得られた成果について要約している。</p>			

氏名	石 咏雪
----	------

(論文審査の結果の要旨)

本論文は、洪水災害時の危機対応に活用することを目的として、災害直後に発信されたツイッターの情報を分析するとともに、その内容に応じて自動的に多数のツイッターを分類する方法を提案したものである。2015年関東東北豪雨による鬼怒川の災害を対象にした分析より、次のような成果を得ている。

(1) 水害時のツイートの内容は、浸水状況に関する内容、被害状況に関する内容、救助や支援に関する内容に大きく分類される。抽出したツイートの中には、写真付きで浸水の状況を説明するものもあり、別途航空写真などから推定した浸水範囲とツイートの情報がよく対応していること、またリモートセンシングでは抽出しきれない浸水の状況が、ソーシャルメディア上でツイートされていることが分かった。さらに、これらのツイートの中には、鬼怒川、浸水、水深、破堤など頻出する特徴的なキーワードが含まれており、これらのキーワードを利用した効率的なツイートの分類法が構築できれば、危機対応にも活用できる可能性が示唆された。

(2) ツイートに含まれる多数の単語から類似するものを選び、水害の分析に適する12個のキーワード群に集約する方法を提示した。そのうえで、それらのキーワード群を説明変数として、上記の目的に応じて分類するための判別分析を行った。交差検証の結果は、57%の正解率で事前に定めたカテゴリーにツイートが分類されることが分かった。また同様のデータセットで対応分析を行うことによって、例えば浸水関連情報のツイートには、浸水、水深、場所、交通のキーワード群を含む場合が多いことなどが明らかになった。

(3) 判別分析の他にも6つの分類手法を比較検討した結果、Naive Bayesを除く5つの方法は同程度の精度となり、概ね60%程度の正解率が確保されていることが示された。また170のツイートを全て教師データとして、別途入手した7,269のツイートの分類を試みて、分類の結果をRecall Rateと呼ばれる再現率で評価した結果、SVMとDecision Treeによる分類方法が有効であることが分かった。

(4) 半教師付き学習は、特に事前に教師として準備できるサンプル数が少ない場合に有効であることが分かった。本論文では15個のツイートを分類した上で半教師付き学習を実施することにより、教師付きと同等か場合によってはそれ以上の分類結果が得られることを明らかにした。

以上のように、本論文は、洪水災害時の危機対応にソーシャルメディアを活用することを目標として、鬼怒川洪水時に発信されたツイートを分析、分類する方法を検討した内容であり、学術上、實際上寄与するところが少なくない。よって本論文は博士(工学)の学位論文として価値あるものと認める。また、平成31年1月21日、論文内容とそれに関連した事項について試問を行い、申請者が博士後期課程学位取得基準を満たしていることを確認し、合格と認めた。

なお、本論文は、京都大学学位規程第14条第2項に該当するものと判断し、公表に際しては、当該論文の全文に代えてその内容を要約したものとすることを認める。