

# 琉球大学学術リポジトリ

## 自然災害テキスト情報源の主要記事抽出に対する教師なし分類化に関する研究

メタデータ	言語: 出版者: 琉球大学 公開日: 2015-04-14 キーワード (Ja): キーワード (En): 作成者: Enrique, Gutierrez Carlos, エンリケ, グティエレッツ カルロス メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/20.500.12000/30654">http://hdl.handle.net/20.500.12000/30654</a>

2015年 1月 30日

琉球大学大学院  
理工学研究科長 殿

論文審査委員

主査 氏名 アシャリフ モハマッド レザー

副査 氏名 山下 勝己

副査 氏名 名嘉村 盛和



### 学位（博士）論文審査及び最終試験の終了報告書

学位（博士）の申請に対し、学位論文の審査及び最終試験を終了したので、下記のとおり報告します。

記

申請者	専攻名 総合知能工学専攻 氏名 Gutierrez Carlos Enrique学籍番号 128673B	
指導教員名	アシャリフ モハマッド レザー	
成績評価	学位論文 <input checked="" type="radio"/> 合格 <input type="radio"/> 不合格	最終試験 <input checked="" type="radio"/> 合格 <input type="radio"/> 不合格
論文題目	Unsupervised Classification for Main Features Extraction in Natural Disaster Text Sources (自然災害テキスト情報源の主要記事抽出に対する教師なし分類化に関する研究)	
審査要旨（2000字以内）		
<p>自然災害はいつでも、災害が発生した地域および住民に、大きな衝撃を与えてきた。そのために、種々の規模の災害に対応することができる災害計画等を、過去の災害事例を参考にして、より確実でより有効なものへと改善する必要がある。一方、自然災害後および災害復興期間中においては、ニュースおよびソーシャルメディアによる災害に関する種々の情報が提供されることから、Web2.0が非常に注目されている。Web2.0は、ウェブサイト上の情報や機能を、外部のサイトやソフトウェアなどから参照したり呼び出したりすることができる。また、利用者や他の事業者が、ソフトウェアやウェブサービス</p>		

(次頁へ続く)

## 審査要旨

を組み合わせることで新たなコンテンツにできることから、自然災害計画を作成する際の有効なツールとなる。

本研究の目的は、自然災害期間に報告された大量のテキストデータから、主要な特徴およびパターンを抽出する教師なしモデルを提案することであり、また、将来、自然災害等が起こる際には、そのモデルが重要な役割を担えるようにすることにある。

まず、多量なデータを主成分分析手法（Principal Component Analysis : PCA）により処理することにより、高次元データの低次元化を図ると共に特徴抽出を図る。また、時間の経過に伴う関連記事の動向を、感度の高い主成分を用いることにより表現する。

次に、記事の空間的・時間的な特徴を、自己組織化マップ（Self Organizing Maps : SOM）を用いることにより抽出する。K平均法を用いてクラスタリングを行うと共に、時間的依存性を調査する。また、線形予測モデルを用いることにより、記事における最重要要因を見つけ出し、記事の動向を調査する。なお、提案手法の有効性については、実データを用いて解析すると共に、種々の角度から検証している。

次に、大量のテキストから、突発事故の主要因および関連規則を抽出するために、ランダムフォレスト手法（Random Forest : RF）を用いる。主要因抽出問題は、データマイニング問題として捉えることができ、膨大な量のデータを迅速に処理する必要がある。RF手法では、大量のデータを幾つかの小サブセットに分離し、個々の決定木に割り当てることにより、短時間で突発事故の主要因および関連規則を抽出することができる。また、パーティクルフィルタ（Particle Filtering : PF）を用いて、一連のニュースにおける関連用語を追跡することにより、時間推移に対する発生確率を推定することができる。

最後に、ニュース記事などの時系列データを対象に潜在的な意味を考慮したトピック追跡を行う手法として、潜在的ディリクレ配分法（Latent Dirichlet Allocation : LDA）が知られている。一般に、ツイッターは、1分間あたり数百のコメントを受け取るが、災害時には、瞬時に膨大なコメント数に跳ね上がる。ここでは、自然災害時のツイートに基づいて、LDAモデルにより学習させることにより、提案手法の有効性を検証している。

本論文は、自然災害期間に報告された大量のテキストデータから、主要な特徴およびパターンを抽出する教師なしモデルについて研究したものであり、また、将来、自然災害等が起こる際の有効なモデル作成法を提案すると共に、その有効性を実データから検証していることから、学術的な意義は非常に大きいものといえる。また学力確認の結果、上記の者は専門的分野および関連分野の十分な知識を有することが判明した。

以上のことより、学位論文審査員一同は、グティエレス・カルロス・エンリケ君を博士（工学）の学位を授与するにふさわしいものとし、学位論文及び学力確認をそれぞれ合格と認める。