

## システム情報工学研究科修士論文概要

年 度	平成 26 年度	学位名		修士( 工学 )
専 攻	知能機能システム	専攻	著者氏名	王 駿琪
指導教員氏名		延原 肇		
論文題目				
種々のデータ圧縮手法に基づく分類器の設計及び性能解析とそれらの多言語 Tweets 分類への応用				
論文概要				
<p>近年言語の文法等に依存せず、どのような言語で記述されたテキストにも適用できるデータ圧縮法を利用した分類を、ソーシャルメディアの話題分類に応用する研究が注目されている。このような分類手法の枠組みでは、データ圧縮として様々な手法を採用することができるが、Tweets の話題分類にどの手法が適切であるかを明らかにされていない。そこで、本研究の第一の目的を、様々な圧縮手法のうち、Tweets の分類に最適な手法を明らかにすることに設定する。また、従来のテキスト分類手法では、その言語特有の文法構造を分類に利用しているため、対象言語によって自然言語解析器を変えなければならない。よって複数の言語を対象とするサービスに関して分類を行うことは難しい。この問題に対しても、言語固有の文法などを利用しないデータ圧縮手法が有効であると考えられるため、本研究の第二目的を、複数言語への適用可能性の検証に設定する。</p> <p>よって、本研究では、日本語、英語、スペイン語、フランス語とドイツ語の5ヶ国語の合計およそ80万件のTweets に対し、Deflate, Gzip, Bzip2と Snappy の4つの圧縮手法を用いて分類を行い、それら圧縮手法の特性の比較検討を行う。さらに、情報量の方面から考慮するために、1 次経験エントロピー法も分類手法として採用する。Twitter のハッシュタグを利用したテストデータセットを対象に分類実験を行い、分類精度と計算時間の観点から比較を行い、各圧縮手法の分類性能及び多言語 Tweets 分類での応用機能を示す。</p>				
審査日		平成 27 年 1 月 29 日		
審査員	(大学名 職名)	(学位)	(氏名)	
主査	筑波大学 准教授	博士(工学)	延原 肇	
副査	筑波大学 教授	工学博士	鬼沢 武久	
副査	筑波大学 教授	工学博士	丸山 勉	