

(続紙 1)

京都大学	博士 (情報 学)	氏名	佐藤 翔輔
論文題目	社会現象としての災害・危機の理解を目的にした 言語資料解析システムの開発		
(論文内容の要旨)			
<p>本論文は、社会現象としての災害・危機の理解の支援に資することを目的として、災害の社会的側面についてCGDC (Customer Generated Digital Contents) としてデジタルテキスト化された記述・記録を効果的に解析するためのテキストマイニングシステム (TRENDREADER : TR, トレンドリーダー) について開発・検討を行なったものである。</p> <p>近年では、個人や組織の応急対応や復旧・復興を対象とする社会現象としての災害の研究が盛んに取り込まれるようになってきている。社会現象としての災害に関する研究は、自然現象の研究に比べて大きな遅れをとっている。昨今では災害の社会現象としての側面についての記述や記録されるテキストデータが電子的に生成・編集・管理されており、社会現象としての災害を系統的に観測・収集できる環境が整備されつつある。このようなデータを系統的に解析するシステムの開発は、社会現象としての災害の理解を深める上で必要不可欠である。</p> <p>本論文では、以上のような問題意識のもと、社会現象としての災害・危機の理解を支援するための言語資料を解析するシステムの開発を行なった。具体的な内容は次のとおりである。</p>			
1) ドキュメント順序を考慮したキーワード自動抽出手法の開発			
<p>社会現象としての災害・危機を理解する言語資料解析システムを開発するに先立ち、テキストマイニングの基礎となる単語の重み付け手法をシステムのコア技術として開発した。災害・危機に関するドキュメントは何らかの順序を有しており、この順序の変化に伴うドキュメントの内容を捉えることで、対象となる言語資料を効果的に理解することができると考え、既存の代表的な手法の拡張と統計解析手法の応用によってドキュメント順序を考慮したキーワード自動抽出手法を開発した。</p>			
2) 時系列的に増加するコーパスを対象とした時系列変化を捉える分析手法 (トレンド解析システム) の開発			
<p>ウェブニュースやブログのように、時系列的に増加するコーパスから対象の時系列変化を捉えることを目的にしたトレンド解析システムを開発した。キーワードを対象にした手法として、特定のキーワードの重みの時系列変化を可視化するふるまい分析とキーワードの将来動向をリアルタイムで予測する手法を開発した。ドキュメントを対象にした手法として、時系列的に増加するコーパスから、一定の時間間隔の中で重みの高いキーワードを多く含むドキュメントを注目すべきドキュメントとして自動的に提示する手法を開発した。トレンド解析システムとして開発した手法を統合し、国内外で発生した災害・危機のウェブ情報を収集・解析し、解析結果を公開・配信する状況認識のためのシステムを開発した。</p>			
3) ある1時点で形成されるコーパスを対象した典型的な特徴を捉える分析手法 (リーダー解析システム) の開発			

社会調査における自由回答のように、ある1時点で形成されるコーパスから対象の典型的な特徴を捉えることを目的にしたリーダー解析システムを開発した。キーワードを対象にした手法として、代表的なキーワードのグルーピングを行ない、自由回答群の内容の内訳を把握する手法を開発した。ドキュメントを対象にした分析手法として、基準変数にした軸の典型性の高いドキュメント、知的能力・言語能力の高い典型的なドキュメントを自動的に抽出する手法を開発した。

(論文審査の結果の要旨)

本論文では、災害・危機の社会現象の側面について記述・記録されたデジタルテキストデータを効果的に解析するテキストマイニングシステムの開発が試みられた。システムのコア技術となるドキュメント順序を考慮してキーワード抽出するという従来にない手法が開発・提案し、ウェブニュースなどの時系列的に増加するコーパスを対象とするトレンド解析システムと、社会調査における自由回答などのある1時点で形成されるコーパスを対象とするリーダー解析システムが開発している。

本論文の研究によって得られた成果は、次のようにまとめられる。

- 1) システムを開発するに先立ち、テキストマイニングの基礎となる単語の重み付け手法をシステムのコア技術としてドキュメント順序を考慮したキーワード自動抽出手法を開発した。災害・危機に関するコーパスに開発手法が適用したところ、開発手法は既存の代表的な手法に比べてより中頻度の単語に高い重みを与えることが示され、言語情報学的に妥当性の高いキーワード抽出が行えることが明らかになった。2004年新潟県中越地震のウェブニュースに対して配信日時を基準変数としたところ、それぞれの時間断面でフェーズ固有の単語が高い重みを示した。また、抽出されたキーワードは、災害のフェーズや採用した基準変数と関連性の高い単語であり、開発手法は災害科学的にも妥当性の高いキーワード抽出が行えることが確認した。
- 2) トrend解析システムでは、キーワードを対象にした分析手法として、特定のキーワードの重みの時系列変化を可視化するふるまい分析とキーワードの将来動向をリアルタイムで予測する手法が開発した。前者は、複数の災害のウェブニュースについて同一のキーワードの重みの時系列変化を比較し、個々の災害事象が活発になるタイミングの共通性や特異性を客観的に明らかにする手法としての有効性が確認した。後者はでは、2007年新潟県中越沖地震のウェブニュースの事例でキーワードの動向を予測できる可能性を明らかにした。ドキュメントを対象にした分析手法として、時系列的に増加するコーパスから、一定の時間間隔の中で重みの高いキーワードを多く含むドキュメントを注目すべきドキュメントとして自動的に提示する手法が開発した。災害・危機について記述されたYahooニュース及びRelief Web中の記事から注目すべき記事や情報を抽出し、国内の危機管理の実務者や研究者等に配信したところ、事案の現況把握やふりかえりに活用されており、一定の有用性を示した。
- 3) リーダー解析システムにおいては、キーワードを対象にした分析手法として、注目キーワードのクラスター分析を行ない、自由回答群の内容の内訳を把握する分析手法の有用性について検証した。リーダー解析におけるドキュメントの分析として、基準変数にした軸の典型性の高い自由回答と、知的能力・言語能力の高い典型的な自由回答を自動的に抽出する機能を開発した。新潟県や柏崎市が実施した域内の被災者を対象にした質問紙調査で得られた自由回答に対して、年齢、生活復興感、被害額などの個人・世帯属性を基準変数にしたところ、それぞれの基準変数の観点から特徴的な回答を客観的に抽出できることを確認した。

以上、本論文の研究は、ドキュメント順序を考慮するという新しいキーワード自動抽出手法を提案したうえで、これを用いた社会現象としての災害・危機の理解を支援するシステムへの発展・応用がなされ、実践的に減災や危機管理に大きく貢献するものと考えられる。

よって、本論文は、博士（情報学）の学位論文として価値あるものと認める。また、平成23年2月9日に論文内容とそれに関連した試問を行なった結果、合格と認めた。