

論文の内容の要旨

論文題目「テキスト情報を含む大規模データの分析と可視化に関する研究」

学位申請者 船山 貴光

キーワード：テキストマイニング データの可視化 トピックモデル 自己組織化マップ
対応分析

近年ビッグデータのブームもあり、オープンデータなど様々なデータが整備され、それらのデータを活用できる環境となった。また、データが蓄積され大規模データとなっている。本論文では、テキスト情報に着目し、大規模なテキストデータの解析と可視化法について提案をする。また、テキスト情報を地理情報システム(GIS)での活用についても示す。

第1章では、データ活用の現状と課題の紹介とテキストデータの解析方法であるテキストマイニングについて紹介する。

第2章では、学術論文データベースに登録されているタイトルとアブストラクトからなるテキストデータを用いてトピックモデルにより各論文の研究領域の推定を行った。さらに自己組織化マップを用いてトピックモデルの結果の可視化方法を提案する。

第3章では、災害時のTwitter情報とWebサービスを活用した減災のための地理情報システムのプロトコルを提案する。Twitter情報には、位置情報を示すジオタグが付いてない場合が多いため、Tweetのテキスト内の地域名を抽出し、その地名の緯度経度情報をWebサービスにより取得し、地図上に可視化した。

第4章では、災害時のTwitterデータに対して対応分析による可視化を用いて有益な情報の抽出を行う。また、対応分析による可視化を経時的に比較するための可視化方法を提案する。

第5章では、大規模データの可視化を行う。この章では、気象データとマーケティングデータについての可視化を取り上げる。また、D3.jsを用いたインタラクティブな可視化も行っ

た。

本論文では、テキスト情報の解析と可視化を行った。トピックモデルと自己組織化マップを用いた可視化により、潜在的なトピックを持つテキストデータを用いた状況の把握と各トピックの比較をすることができた。Twitter 情報を用いた減災のための地理情報システムの開発では、気象情報を Web サービスにより取得し地図上に可視化したが、Web サービスを活用することでその他の情報も収集することができる。それにより目的に応じた情報の選択により応用が可能である。