

氏名(本籍)	豊田哲也(北海道)
学位の種類	博士(工学)
学位記番号	博甲第6342号
学位授与年月日	平成24年11月30日
学位授与の要件	学位規則第4条第1項該当
審査研究科	システム情報工学研究科
学位論文題目	大規模テキストクラスタリングのための概念抽象化に基づくネットワークと自己組織化マップ

主査	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	白川友紀
副査	筑波大学准教授	博士(工学)	古賀弘樹
副査	筑波大学准教授	博士(工学)	矢野博明
副査	筑波大学講師	博士(工学)	延原肇

論文の内容の要旨

インターネット、情報発信の多様な環境の普及および整備により、我々が得ることのできる情報量は爆発的に増加している。インターネットニュース、専門分野のデータベース、電子掲示板や個人単位でのブログなど、これら多様なテキスト情報の量は人間の情報処理能力を大幅に超えており、今後も情報環境側と人間の処理能力の間のギャップがますます増大することが予想される。本論文では、このような膨大なテキストの中から目的の情報を取得する方策として、大規模テキストデータの集合を閲覧可能にする可視化システムの開発と、これらを実現するためのテキストクラスタリング手法を提案している。具体的には、(1) 膨大なテキストデータを隣接構造(リンク構造)によって表現、解析する方法の提案、およびそれを実装したシステムの開発、(2) 膨大なテキストデータを距離構造によって配置するための自己組織化マップ(SOM)の提案、(3) 集合知の観点からの大規模データの抽象化、を行っている。

本論文は6章から構成されている。第2章では大規模法令データを各法令のリンク構造を利用しネットワーク化することで、非専門家でも法令全体を俯瞰できるシステムの構築を行う。第3章では、テキスト情報を取り扱うためのSOMの実装を行い、大規模データへ拡張した場合にも対応できるように、学習過程の高速化を提案している。第4章では、オンラインニュース等のデータの追加を考慮したSOMをWikipediaによる概念抽象化手法を組み込むことで構築し、追加学習や抽象化した際のラベリングの問題の解決を示している。さらに第5章では、ユーザインタラクションを取り入れたSOMに基づく2次元マップ型検索システムを実装し例示している。最終章では本論文を総括し、今後の課題等に言及している。

審査の結果の要旨

本論文では、大規模テキストデータと人間の処理能力のギャップを埋める技術として、リンク構造および距離構造に基づいた視覚化手法を提案している。法令データおよびWebニュースを例にして、テキスト間の接続構造を利用した視覚化および解析手法を示し、また距離構造を用いた自己組織化マップによる視覚化

およびその高速化・抽象化を提案している。さらに法令データベースおよび Web 上のオンラインニュースを用いた評価実験を通して、各提案手法の有用性を示している。実用の際の学習時間の長さなどの問題点は残っているものの、本論文で得られた知見は大規模テキストのクラスタリングおよび視覚化の分野に重要な指針を与え、さらに本論文は次世代のインタラクティブな検索エンジンの基礎技術として有用な学術資料となり、博士論文に値するものと考えられる。

平 24 年 9 月 26 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。