

氏名(本籍)	さだ みつ く がつ 貞 光 九 月 (福 岡 県)		
学位の種類	博 士 (工 学)		
学位記番号	博 甲 第 4947 号		
学位授与年月日	平成 21 年 3 月 25 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	統計的トピックモデルとその評価文書分類への応用に関する研究		
主査	筑波大学教授	博士(工学)	山本 幹 雄
副査	筑波大学教授	理学博士	大 保 信 夫
副査	筑波大学教授	工学博士	平 井 有 三
副査	筑波大学准教授	博士(工学)	宇津呂 武 仁
副査	筑波大学准教授	博士(工学)	滝 沢 穂 高

論 文 の 内 容 の 要 旨

本論文では文書全体のトピック情報を利用する新しい統計的言語モデル（以下、トピックモデルと呼ぶ）を提案し、さらにトピック情報を利用した評価文書分類手法の高精度化を行っている。

新しいトピックモデルとして混合ディリクレモデルを提案している。これは単語の確率分布としてそれぞれのトピックに対応するディリクレ分布を複数個仮定することで、トピック毎の単語出現確率の変動を直接捉えるモデルである。ただし、混合ディリクレモデルは過適応を引き起こしやすいため、階層ベイズモデルを用いることで過適応を緩和する推定手法も同時に提案している。実際のテキストデータを用いた実験によって、混合ディリクレモデルが従来のトピックモデルよりも予測力の観点において性能が高いことを明らかにしている。

応用研究として、混合ディリクレモデルに限らず従来のトピックモデルを含めて、意見分析の一分野である評価文書分類への適用を行っている。評価文書と呼ばれるレビュー等にはそのレビュー対象のカテゴリによって評価表現が異なる場合が存在するため、その変動をトピックモデルによって自動的に捉えることがアイデアの中心である。単純なトピックモデルの適用を行った場合、肯定極性・否定極性において別個にトピックが学習される問題を、極性間に共通のトピックを導入することで解消している。評価実験では Amazon における大量のレビューデータを用いて、従来のナイーブベイズ分類器や SVM 等の分類器に対し、提案手法の精度が有意に高いことを明らかにしている。

審 査 の 結 果 の 要 旨

トピック情報を利用した統計的言語モデルの基礎的な理論研究とそれを Web 上の評価文書分類に応用した論文である。前半の理論研究では、自然言語処理の従来の文脈モデルから自然に導かれる強力なトピックモデルを提案し、自然言語処理で特に重要な単語予測力の点で他の一般的なトピックモデルよりも高性能であることを実験的に示している。後半のトピックモデルを利用した評価文書分類については、前半で自ら提

案しているトピックモデルではあまり高い性能が出なかったものの、提案している一般的なトピックモデルを評価文書分類システムへ組み込む手法には新規性があり、性能も高いことが実験的に示されている。評価文書分類に対するトピックモデルの有効性を明らかにした点で学術的な貢献は大きいと判断できる。

よって、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。