

論文 / 著書情報
 Article / Book Information

題目(和文)	制約付き独立話題分析に関する研究
Title(English)	Constrained Independent Topic Analysis
著者(和文)	西垣貴央
Author(English)	Takahiro Nishigaki
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10541号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:新田 克己,寺野 隆雄,渡邊 澄夫,小野 功,石井 秀明,小野田 崇
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10541号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名		西垣 貴央	
		氏名	職名		氏名	職名
論文審査 審査員	主査	新田 克己	教授	審査員	石井 秀明	准教授
	審査員	寺野 隆雄	教授		小野田 崇	青山学院・教授
		渡辺 澄夫	教授			
		小野 功	准教授			

論文審査の要旨 (2000 字程度)

本論文は「制約付き独立話題分析に関する研究」と題し、全6章からなる。独立話題分析におけるふたつの課題の解決方法について検討を行ったものである。

第1章「序論」では、膨大な量の文書データから、有益な情報を発見・抽出するための技術であるテキストマイニングの課題の一つの話題抽出に関する研究の背景および目的を述べている。話題抽出の方法として、確率的生成モデルに着目して話題を抽出する方法であるトピックモデルの研究事例と、話題間の関係に着目して話題を抽出する方法の研究事例について述べている。本論文では特に、抽出された話題間の独立性に着目しており、独立性の高い話題を抽出することの利点や、その独立性の高い話題を抽出する方法として独立話題分析があることが述べられている。

第2章「独立話題分析」では、独立性の高い話題を抽出する方法である独立話題分析の概念やアルゴリズムについて述べている。本章では検討対象とする独立話題分析のふたつの課題について報告している。第1の課題は、独立話題分析では話題間の独立性にのみ着目しているため、抽出した話題がユーザの求める話題と異なる場合が存在するという点である。また、その理由についてベンチマークデータを用いて述べている。第2の課題は、独立話題分析では、数が増加していくデータに独立話題分析を適用するためには時間的にも空間的にも非常に困難であるという点である。また、その理由についても図を用いて詳しく述べられている。

第3章「関連研究」では、ユーザ制約を取り入れて話題抽出を行う方法や、データが逐次的に増加していく場合における話題抽出を行う方法の関連研究をいくつか紹介している。

ユーザ制約を取り入れて話題抽出を行う関連研究でのユーザ制約には、制約付きクラスタリングの研究で多く使用されている Must Link 制約や Cannot Link 制約が存在すること、またこれらの制約がクラスタリング手法だけでなく数多くの話題抽出手法にも取り入れられていることが紹介されている。さらに、データが逐次的に増加する場合における話題抽出の関連研究についても話題抽出手法だけでなく、主成分分析や独立成分分析においても研究が行われていることが紹介されており、その必要性が述べられている。また、これらの関連研究では独立性の高い話題を抽出する独立話題分析においては、ユーザ制約を取り入れる方法およびデータが逐次増加していく場合における方法の研究が行われていないことも指摘している。

第4章「ユーザ制約付き独立話題分析」では、独立話題分析の課題の一つである、抽出した独立性の高い話題がユーザの望む話題と異なる場合の解決方法として、ユーザ制約付き独立話題分

析を提案している。ユーザ制約として2個の話題を1個の話題に統合する Merge Link 制約と、1個の話題を2個の話題に分離する Separate Link 制約を新たに定義している。また、従来のユーザ制約との違いについても述べている。さらに、提案しているユーザ制約付き独立話題分析で抽出した話題が、制約を満たしてかつ独立性の高い話題であるということを示すために、複数のベンチマークデータに適用して、制約の評価と独立性の評価をそれぞれ行い、その有効性を実験的に示している。

第5章「データ追加に基づく独立話題分析」では、独立話題分析の課題の一つである、データが逐次的に増加していく場合において独立話題分析を適用するのは困難であるという課題の解決方法として、データ追加に基づく独立話題分析を提案している。データ追加に基づく独立話題分析とは、初期データから独立性の高い話題を求め、その話題をデータが追加される度に独立性に基づいて更新していく手法であることが述べられている。さらに、そのデータ追加に基づく独立話題分析をベンチマークデータに適用し抽出した独立性の高い話題の分析を行うことで、その手法の有効性を実験的に示している。また、文書データが無限に増大する場合、その増大に応じて独立な話題の数の推定をするというような研究はこれまで行われていないことを述べるとともに、今後検討していく必要があることを述べている。

第6章「結論」では、本論文のまとめと、本論文で提案した制約付き独立話題分析の今後の課題について述べられている。課題として、ユーザ制約の簡略化および文書データが追加された時の話題の数の変化に対応を行う等が挙げられている。

以上を要するに、本論文は、膨大な量の文書データから独立性の高い話題を求める独立話題分析の課題であった、ユーザ制約を取り入れかつ独立性の高い話題を求める方法や、データが逐次増加する場合において独立性の高い話題を求める方法を提案しており、工学的に貢献するところが大きい。よって、博士（工学）の学位を授与するに十分な価値があるものと認められる。

注意：「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。