

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)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

専攻： 知能システム科学 専攻
Department of
学生氏名： 西垣 貴央
Student's Name

申請学位(専攻分野)： 博士 (工学)
Academic Degree Requested Doctor of
指導教員(主)： 新田 克己
Academic Advisor(main)
指導教員(副)：
Academic Advisor(sub)

要旨(和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

本論文は「制約付き独立話題分析に関する研究」と題し、文書データから話題間の独立性が最も高くなるような話題を抽出する独立話題分析における二つの課題の解決方法について検討を行った。

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である話題抽出に関する研究の背景および目的を述べた。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紹介した。本論文では特に、話題の独立性に着目し、独立性の高い話題を抽出することの利点や、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存在するという点などを述べた。

第 2 章「独立話題分析」では、話題間の独立性が最も高くなるような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紹介した。また、検討対象とする独立話題分析の二つの課題について述べた。第一の課題として、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を、ベンチマークデータを用いて具体的に述べた。第二の課題として、独立話題分析では、数が逐次的に増加していくデータに対して独立話題分析を適用する場合、時間的にも空間的にも非常に困難であるという点を、図を用いてその理由とともに詳細に述べた。

第 3 章「関連研究」では、第 2 章で述べた独立話題分析の二つの課題である、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を述べ、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を紹介した。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を紹介し、その必要性を述べた。また、本章で紹介した関連研究の中には、独立話題分析において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独立話題分析を適用する方法の研究が行われていないことを指摘した。

第 4 章「ユーザ制約付き独立話題分析」では、独立話題分析の第一の課題である、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た。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、それらの制約が従来のユーザ制約である Must Link 制約や Cannot Link 制約との違いについても述べた。さらに、提案した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ことを、複数のベンチマークデータに適用することで実験的に示した。

第 5 章「データ追加に基づく独立話題分析」では、独立話題分析の第二の課題である、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た。提案したデータ追加に基づく独立話題分析では、初期データから独立性の高い話題を求め、その話題を新たにデータが追加される度に更新していく方法であることを述べた。さらに、その提案したデータ追加に基づく独立話題分析で得られる話題を更新することで、全てのデータを使用して得られる話題と同じ話題を得ることができたのかを、複数のベンチマークデータに適用することで実験的に示した。また、提案するデータ追加に基づく独立話題分析は、従来の独立話題分析を増加するデータに適用する場合よりも計算量が小さくなることを述べた。

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た。課題として、ユーザ制約の Separate Link 制約の簡略化および文書データが追加される時の話題数が増減する場合の対応を検討することなどを挙げた。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、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。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

(博士課程)
Doctoral Program

論文要旨

THESIS SUMMARY

専攻：	知能システム科学	専攻
Department of		
学生氏名：	西垣 貴央	
Student's Name		

申請学位 (専攻分野)：	博士	(工学)
Academic Degree Requested	Doctor of	
指導教員 (主)：	新田 克己	
Academic Advisor(main)		
指導教員 (副)：		
Academic Advisor(sub)		

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

This paper is titled “Constrained Independent Topic Analysis” , and discusses two issues of Independent Topic Analysis.

The first chapter discusses the background and motive of this paper. As a topic extraction method, there are methods of extracting a topic focusing on a stochastic generative model and the relationship between topics. Especially in this paper, we discussed Independent Topic Analysis focusing on independence between topics.

The second chapter describes the Independent Topic Analysis. In the independent topic analysis, it is possible to extract the most independent topics. There are two issues with this method. The first issue is that extracted topics may differ from the ones wanted by user. The second issue is that applying Independent Topic Analysis to the increasing number of document data is a hard because temporal and spatial cost is large.

The third chapter states the related work about two issues. Related research concerning topic extraction with user constrained and topic extraction in incremental data are stated. And methods considering constrained topic extraction and incremental topic extraction are introduction. In the case of Independent Topic Analysis, there is no method applying to user constrained or incremental data, so methods using user constrained and methods applying incremental data are necessary.

The fourth chapter describes the proposed method that independent topic analysis with user constrained. Merge Link constraints and Separate Link constraints are newly defined as user constrained. Using benchmark data, the paper assesses obtained results to show the effectiveness of proposed method.

The fifth chapter describes the proposed method that independent topic in increasing data. Evaluation results using benchmark data show that the proposed method is able to extract the same topic as the topic extracted using all the data, and difficult to extract the same topic when the number of topics is small and the number of data included in each topic varies greatly.

The sixth chapter concludes the paper.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、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。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).