

## 論文の内容の要旨

論文題目	ベイズアプローチを用いたアカデミックライティング支援システム
学位 申請者	宇都 雅輝

アカデミック・ライティングでは、情報を読者に効果的に伝達できるように文章を執筆することが重要である。このためには、論理構造を意識した文章執筆が不可欠である。特に、アカデミック・ライティングでは、1.「研究背景」や「先行研究の問題点」、「本論文の目的」といった論文要素の系列で定義される文章全体の大局的な構造（本論文では、このような論理構造を「論文構成」と呼ぶ）、2.「主張」とそれを支持する「データ」や「論拠」などで構成される「論証」、が重要である。しかし、これらの作業は、特に初心者にとって難度が高いことが知られている。そこで、本論文では、アカデミック・ライティングにおける論文構成と論証の構築を支援するシステムの開発を目的とする。

まず、論文構成の構築を支援するシステムを開発した。論文構成の構築を支援するシステムはこれまでも多数開発されてきた。それらのシステムの多くは、完成した論文を入力とし、その論文構成を画一的に形式化するように支援を与える。しかし、論文構成は、一般的な形式に準拠しつつ、執筆者の意図に合わせて柔軟に構築される必要があるため、既存システムには以下のような問題点が残る。

1. いったん論文を完成させて入力するバッチ処理にのみ対応しており、オンライン処理に対応していないため、論文を構成する過程が支援されない。
2. 論文構成の画一的な形式化のみを支援しており、多様な利用者の意図を表現することが難しい。

これらの問題を解決するために、本論文では、確率的アプローチを用いた論文構成構築支援システムを開発する。ここでは、「論文構成」を情報理論における情報源からの出力符号系列とみなしたメタファとしてとらえ、論文構成の構築過程を定式化する。具体的には、論文構成を論文要素カテゴリの系列データとし、それが $m$ 重マルコフ情報源に従うと仮定する。多重度の推定法としては、従来から、BIC(Bayesian

Information Criteria)やAIC(Akaike Information Criteria)といった統計的アプローチが知られているが、これらの手法では、本論文で扱う論文構成データのようなデータ長が短い場合の学習において、多重度を正しく推定できないことが多い。そこで、本論文では、情報論的アプローチにおいて最も高精度なベイズ符号(Bayes Code)により、データ長の短い論文構成データから多重度を推定することを目標とする。

しかし、この推定法を用いても、データ長が極端に短い場合、ベイズ符号語長が多重度の増加に従い単調減少し、多重度を正しく推定できない場合がある。そこで、本論文では、ベイズ符号語長が単調減少する場合の推定補正法を提案し、シミュレーション実験により補正手法が多重度の推定精度を向上させることを示す。この補正法を用いて過去の優良論文100件から予測精度の高い $m$ 重マルコフ情報源を高精度に推定する。さらに、推定されたマルコフ情報源に基づき論文構成の構築過程を逐次的にナビゲーションするシステムを開発した。

次に、論証の構築を支援するシステムを開発している。これまでも論証の構築を支援するシステムは多数開発されてきた。しかし、これらのシステムの多くでは、Toulminモデルと呼ばれる論証スキーマに論証を当てはめて可視化することで論証の構築を支援している。Toulminモデルとは、論証を「主張」、「データ」、「論拠」、「裏付け」、「限定」、「論駁」の6つの小単位に分類し、それぞれの関係を有向グラフとして形式化したものであり、論証の基準として多様な分野で利用され、その妥当性が示されている。しかし、論証の主目的である「主張」の正当化のためには、論証のToulminモデルへの当てはまりの良さよりも、文章間の因果の強さ、すなわち「論証の強さ」を重視した論証の推敲が重要である。また、一般の論証では、複数の「論拠」や「データ」、「裏付け」から主張を導くことが多く、論証の構成が複雑化すると、以下の問題が生じると考えられる。

1. 「論証の強さ」を全ての文章間について評価することが困難である。
2. 論証中の各文章がどの程度正当化できているかの推定が難しい。
3. 「主張」の正当化に対して各文章がどのように影響しているかを把握することが困難である。

以上の問題を解決するために、本論文では、Toulminモデルのベイジアンネットワーク表現を用いた論証構築支援システムを開発した。ここでは、論証中の文章を確率変数、文章間の因果を条件付確率で表すことで、Toulminモデルに当てはめて構築された論証をベイジアンネットワークとして定式化している。確率は主観確率とし、確率値に対応付けられた6段階のカテゴリを用いてユーザが評価する。本システムでは、このベイジアンネットワーク表現を用いて、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度、という論証の特性を表す3つの指標を算出し、それらの推定結果に応じて論証改訂のためのアドバイスをフィードバックする。

また、シミュレーション及び被験者実験により、それぞれのシステムの有効性を示した。

最後に、本論文で得られた主な研究成果を総括し、本論文をまとめるとともに本論文の課題について述べている。

## 論文審査の結果の要旨

学位申請者氏名	宇都 雅輝
審査委員主査	植野 真臣
委員	大須賀 昭彦
委員	長岡 浩司
委員	田中 健次
委員	田原 康之

宇都氏の学位論文では、アカデミックライティングにおける論文構成と論証の構築を支援するシステムを開発し報告している。

第1章では、アカデミック・ライティングの重要性和初心者にとって困難な点を挙げ、本論で扱う問題をまとめている。

第2章では、論文構成の構築を支援するシステムについて提案している。論文構成の構築を支援するシステムはこれまでも多数開発されてきているが、それらのシステムの多くは、完成した論文を入力とし、その論文構成を画一的に形式化するように支援を与えるものがほとんどであると指摘されている。

また、論文構成は一般的な形式に準拠しつつ執筆者の意図に合わせて柔軟に構築される必要があるため、既存システムには以下のような問題点が残ると本論文は指摘している。

1. いったん論文を完成させて入力するバッチ処理にのみ対応しており、オンライン処理に対応していないため、論文を構成する過程が支援されない。
2. 論文構成の画一的な形式化のみを支援しており、多様な利用者の意図を表現することが難しい。

この問題を解決するために、まず、確率的アプローチを用いた論文構成構築支援システムを開発している。そこでは、「論文構成」を情報理論における情報源からの出力符号系列とみなしたメタファとしてとらえ、論文構成の構築過程を定式化している。具体的には、論文構成を論文要素カテゴリの系列データとし、それが $m$ 重マルコフ情報源に従うと仮定している。多重度の推定法としては、従来から、BIC(Bayesian Information Criteria)やAIC(Akaike Information Criteria)といった統計的アプローチが知られているが、これらの手法では、本論で扱う論文構成データのようなデータ長が短い場合の学習において、多重度を正しく推定できないことが多いことを問題点として挙げている。そこで、本論文では、情報論的アプローチにおいて最も高精度なベイズ符号(Bayes Code)により、データ長の短い論文構成データから多重度を推定することを目標としている。

また、本論文では、この推定法を用いても、データ長が極端に短い場合、ベイズ符号語長が多重度の増加に従い単調減少し、多重度を正しく推定できない場合があることを指摘している。そこで、ベイズ符号語長が単調減少する場合の推定補正法を提案し、シミュレーション実験により補正手法が多重度の推定精度を向上させることを示している。実際にこの補正法を用いて過去の優良論文100件から予測精度の高い $m$ 重マルコフ情報源を高精度に推定している。さらに、推定されたマルコフ情報源に基づき論文構成の構築過程を逐次的にナビゲーションするシステムを開発している。

第3章では、論証の構築を支援するシステムを開発している。これまでも論証の構築を支援するシステムは多数開発されてきているが、これらのシステムの多くでは、Toulminモデルと呼ばれる論証スキーマに論証を当てはめて可視化することで論証の構築を支援するのみであると本論文は指摘している。Toulminモデルとは、論証を「主張」、「データ」、「論拠」、「裏付け」、「限定」、「論駁」の6つの小単位に分類し、それぞれの関係を有向グラフとして形式化したものであり、論証の基準として多様な分野で利用され、その妥当性が示されている。しかし、本論文の主張は、論証の主目的である「主張」の正当化のためには、論証のToulminモデルへの当てはまりの良さよりも、文章間の因果の強さ、すなわち「論証の強さ」を重視した論証の推敲が重要であるというものである。

また、本論文では、論証の強さを評価する場合、一般の論証では、複数の「論拠」や「データ」、「裏付け」から主張を導くことが多く、論証の構成が複雑化すると、以下の問題が生じると考えている。

1. 「論証の強さ」を全ての文章間について評価することが困難である。
2. 論証中の各文章がどの程度正当化できているかの推定が難しい。
3. 「主張」の正当化に対して各文章がどのように影響しているかを把握することが困難である。

以上の問題を解決するために、本論文では、Toulminモデルのベイジアンネットワーク表現を用いた論証構築支援システムを開発している。そこでは、論証中の文章を確率変数、文章間の因果を条件付確率で表すことで、Toulminモデルに当てはめて構築された論証をベイジアンネットワークとして定式化している。確率は主観確率とし、確率値に対応付けられた6段階のカテゴリを用いてユーザが評価するというものである。本システムでは、このベイジアンネットワーク表現を用いて、1. 論証の強さ、2. 文章の正当性、3. 主張への影響度、という論証の特性を表す3つの指標を算出し、それらの推定結果に応じて論証改訂のためのアドバイスをフィードバックする。また、2章と3章では、シミュレーション及び被験者実験により、それぞれのシステムの有効性を示している。

最後に、第4章において、本論文で得られた主な研究成果を総括し、本論文をまとめるとともに本論文の課題について述べている。

本論での提案は、アカデミックライティングに限定されず、Toulminモデルに基づく論証支援研究として一般化できる成果を出しており、議論モデルや知識創造モデルなどへの拡張も可能である。

よって、本論文は博士（工学）の学位請求論文として十分な価値を有するものと認める。