

氏名	しらまつ しゅん 白松 俊
学位(専攻分野)	博士(情報学)
学位記番号	情博第288号
学位授与の日付	平成20年3月24日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	情報学研究科知能情報学専攻
学位論文題目	Salience-based Modeling of Discourse Context (顕現性に基づく談話文脈のモデル化)
論文調査委員	(主査) 教授 奥乃 博 教授 河原達也 教授 黒橋 禎夫

論文内容の要旨

本論文は、議論支援や会話などの高度な談話処理システムに不可欠な談話文脈の計算論的モデルを確立するために、顕現性に基づく談話文脈のモデル化に関する研究をまとめたものである。談話参加者の注目対象が発話ごとに変化する談話文脈に対して、談話実体への注目度である談話の顕現性の動的変化を定量的に扱うアプローチが述べられている。

第1章は序論で、計算論的モデルの観点から談話文脈のモデル化について述べ、従来手法の問題点を明らかにしている。

第2章では、談話文脈のモデル化の研究動向、談話文脈、顕現性、情報提示、ゲーム理論的語用論に関する従来研究について概観し、従来研究において議論されてきた顕現性への多様な影響要因の統合が問題点になっていることを指摘し、その問題点を解決するために2つの課題を提示している。第1の課題は、顕現性の定量的な定式化であり、第2の課題は、顕現性に基づいた参照結束性と文脈類似度の定式化である。

第3章では、本論文で扱うための言語データの形式であるGDA (Global Data Annotation) について述べ、それを扱うツールキットGDASDKの開発について述べている。GDASDKは、GDA処理用API、および、それに基づく自動タギングシステム、GDA簡易ビューワ等から構成される。

第4章では、第1の課題に対して、様々な影響要因を1つの尺度に統合するために、後続発話で継続的に参照される確率である「参照確率」という基準を提案している。さらに、談話文脈の動的変化を扱うために、新近性効果の減衰曲線を表す窓関数を導入している。

第5章では、第2の課題を解決するために、参照確率とゲーム理論に基づいて参照結束性の定式化をしている。従来、参照結束性は標準的理論であったセンタリング理論で説明されていたのに対して、ゲーム理論等で一般的に使用される期待効用原理で説明できるという仮説を立て、その仮説の言語一般性を検証するために、パラメタの統計的定義を行っている。

第6章では、第4章、第5章で定式化したモデルの実証について述べている。まず、話し言葉と新聞記事のコーパスを用いて、参照確率による顕現性の定式化の検証実験を行い、特に話し言葉で効果的に文脈の動的変化を扱えることができることを示している。また、日本語、英語の新聞記事コーパスを用いて、ゲーム理論に基づく参照結束性のモデルの検証実験を行い、参照結束性が期待効用原理で説明できるという仮説の言語一般性を示唆する結果を示している。

第7章では、第4～6章で提案した談話文脈のモデルを応用した情報提示プロトタイプシステムについて述べている。具体的には、談話文脈の流れを可視化する顕現性グラフ表示システムと、談話文脈に応じた関連情報提示システムである。また、関連情報提示のため、参照確率ベクトルを用いた文脈類似度を定義している。

第8章では本研究のまとめを行い、情報提示システムの今後の展開について述べている。特に、パブリックインボルブメント(住民の行政参画プロセス)における文脈共有支援への応用についての初期的な結果と今後の展望について述べている。

第9章は、結論である。

論文審査の結果の要旨

本論文は、顕現性に基づく談話文脈のモデル化に関する研究をまとめたものである。得られた主な成果は次の通りである。

1. 談話文脈を工学的に扱うために、顕現性に基づく談話文脈のモデル化に取り組み、コーパスを用いた顕現性の確率的な定式化を行い、さらに、新近性効果の減衰曲線を表す窓関数を導入し、談話文脈の動的変化を表現した。本表現法により、文脈の時系列性、定量性が取り扱えることを示すとともに、コーパスを用いて、話し言葉で特に有効であることを実証した。
2. 先行文脈からの繋がりよさを示す参照結束性を扱うために、参照確率とゲーム理論に基づいて参照結束性の定式化を行い、さらに、参照結束性のルールをゲーム理論の期待効用原理で説明可能という仮説を提示し、複数言語で検証するためにパラメタの統計的定義を行い、日本語と英語の新聞記事コーパスで検証実験を行い、仮説が両言語で成立することを示し、同仮説の言語一般性を示唆する結果を得た。
3. 提案した談話文脈のモデル化を応用したプロトタイプシステムとして、談話文脈の流れを可視化する顕現性グラフ表示システムと、参照確率ベクトルを用いた文脈類似度を定義し、それに基づいて関連情報を提示するシステムを試作し、本手法の有用性を示した。
4. パブリックインボルブメント（住民の行政参画プロセス）への適用可能性を検討するために、淀川コーパスに対して顕現性グラフ表示システムを応用し、文脈共有支援に有効であるという初期的な結果を得た。
5. 提案するモデル化では、コーパスにGDAタグが付与されていることが前提である。そのため、GDASDKツールキットを開発するとともに、実際GDASDKを使用して、話し言葉コーパスや淀川コーパスというプレインテキストに対する実証実験を行い、同ツールキットの有用性を示した。

以上本論文は、談話文脈のモデル化に対する工学的な定式化を行い、ツールキットを作成し、談話文脈の流れを定量化する技術を確立するとともに、参照確率、参照結束性に基づいた情報提示システムのプロトタイプを試作することにより、その効果を具体的に示しており、学術上、實際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成20年2月21日実施した論文内容とそれに関連した試問の結果合格と認めた。