

専攻	システム情報工学	学籍番号	893741	指導教官氏名	増山 繁 助教授
申請者氏名		山本 和英			中川 聖一 助教授

論 文 要 旨

論文題目	日本語文章における表層的機械処理の有効利用に関する研究
------	-----------------------------

(要旨 和文 1,200 字程度)

(1)

自然言語処理研究の一つの傾向として、処理の対象が単一の文から複数文へと徐々に移りつつあることが挙げられる。複数の文を対象にする談話処理の分野では、以前から文章の抄録／要約などの試みがなされてきたが、最近では、計算機及び機械可読文書の普及に伴うテキストの分類などの自動検索、あるいは対話処理などを対象にして、様々な研究が活発に行われている。本論文では今後ますます重要になると考えられる自然言語処理を対象にして研究活動を行ってきた。その中でも、従来ほとんど研究が行われていない段落分け、そして要約、文章の自動分類などの談話処理を取り上げ、検討を行った。また、対象言語としては日本語を取り扱った。

文章の段落分けは、段落のない文章(文の羅列)を入力とし、段落に分けた文章を出力するという処理である。本論文では、段落分けに必要な要素として接続的語句と単語間の類縁性の二点を取り上げ、最終的には両要素を併用して考慮することによって段落分けを試みた。この手法で雑誌記事、新聞コラムの文章を段落分けした結果についてアンケート調査を実施し、作成した手法の有効性を支持する結果を得た。

文章の要約作成については、文章中から重要な文を単に抽出する「抄録」のように前後の文に全く結束性のない文章が生成されることを避け、要約文として出力された文章が最小限の結束性を保つように文の抽出を行った。さらに抽出した文に含まれる連体修飾語の一部を削除することにより、一段の文短縮を試みた。また、これらのシステムを実際に計算機上に実現することが重要と考え、実在の論説記事(新聞社説)を対象にした要約システムを構築した。

さらに、従来ほとんど試みられていない、何らかのつながりを持った複数新聞記事の要約も試みた。ここでは、類似した複数文章の要約に特有の問題である類似記述部分の特定、およびその削除を試み、ある程度重複した記述の省略、同じ語に対する修飾語句の削減など

の点に着目することにより、文章の短縮化を試みた。

本研究はまた、文章の自動分類の研究にも取り組んだ。この研究では、従来数多く研究が行われているベクトルモデルの分類手法のうち、語に重みづけを行うこと (*term-weighting*) による手法を拡張し、語をいくつかのグループに分類した上で重みづけを行う手法を提案した。この手法によって新聞の10種類のコラムを分類する実験を行った結果、再現率、適合率ともに高い結果が得られた。