

理工学専攻修士論文要旨

提出年度：2005年度
提出日：2006/01/30
専修コース：知能情報コース
学生番号：35604179
学生氏名：面来 道彦
研究指導教員：原田 実 教授

(論文題目)

日本語文章要約システム ABISYS の属性・文脈を考慮した高精度化

(内容の要旨)

近年の情報化社会の中で、自動要約の研究が注目されてきている。これまでの要約研究では、重要文抽出型要約では、不必要な部分を抜き出してしまいうため、必要な部分を抽出できなく、重要語句抽出型では、表層的な情報に基づく研究がほとんどで、表層格で述語節の格を補うため、語意に基づく適切な格が抽出されず要約文が不自然になっているなどの問題がある。さらに、意味解析を行っていないため、語意の違いによる述語の区別がされず、述語ごとに適切な格を補うことができない。そこで、原田研究室では、文章を意味解析した結果の意味グラフを用いて要約文を生成するシステムとして日本語要約システム ABISYS を開発してきた。昨年までに開発を重ねてきた ABISYS2004 は精度の高い要約をするものであったが、重要語句の抽出、必須格の算出、要約文の日本語としての適切さの点で問題が見られた。

本研究では、重要語句選択のための得点として従来用いられていた反復、見解、位置、主題・焦点得点に加えて、文の役割を得点として与える文脈得点、筆者の意志を表すムードによって得点を与えるムード得点、要約文で重視される傾向にある固有表現に得点を与える固有表現得点の3得点を新たに追加した。また、日本語として正しい文章を出力するために要約文生成の中で、必須格の抽出方法の改善として、用言に対する必須格を求める計算方法を見直すとともに、この計算に用いる必須格辞書を再構築した。これらのように従来手法に、本研究での改善を加えて日本語要約システム ABISYS2005 の開発を行うこととした。

要約処理の具体的な方法として次の4ステップで行う。

- (1) 語句への得点では、体言に反復、主題・焦点、位置、見解、文脈、ムード、固有表現の7種類の得点を与える。
- (2) 重要語句得点の総合化として、前ステップで与えられた7種類の得点を学習識別器サポートベクターマシン(SVM)を用いて総合化し重要語句であるかどうかを判定する。なお、SVMを用いるにあたって必要となる学習データは、あらかじめ大量の重要語句候補に対して、人手によって重要語句となるかどうかを判別されたデータを用いる。
- (3) 要約文生成は3ステップで行われる。()重要語句から述語節までのパスを抽出する。()述語節においては、その用言の必須格を算出し、必須格としてつながっている語を抽出する。()必須格として体言節が抽出された場合は、特定の抽出ルールに従って隣接格を抽出する。
- (4) 補完処理として、括弧の処理や句読点など、要約文として生成されたものの中で不整合性があるものに関して補完を行う。

本システムによる要約と、人手による2つの要約、文の先頭から要約率分抜き出した要約の4つを内容のカバーと読みやすさの点で主観で順位付け評価したところ、文の読みやすさの点では平均2.54(他システムの平均3.11)、内容カバーの点では平均2.08(他システムの平均3.25)となった。また生成された要約文の日本語としての正しさは昨年度95%から97%と、従来のシステムより要約品質が向上した。また、第3者作成の要約を正解データとしてutility値を評価したところ、昨年度0.45から0.54と大きな精度向上が認められた。よって本システムの有効性を示すことができたと言える。