

理工学専攻修士論文要旨

提出年度：2006年度

提出日：2007年1月30日

専修コース：知能情報コース

学生番号：35605079

学生氏名：杉村 和徳

研究指導教員：原田 実 教授

(論文題目)

照応解析に重点を置いた意味解析システム SAGE の高精度・高速化

(内容の要旨)

日本語文章では、読み手に容易に推測することのできる格要素（ゼロ代名詞と呼ばれる）は、文章上表現されないことが多い。また、同じ言葉の繰り返しが嫌われるために、指示代名詞が頻繁に使われる。このような表現（照応関係と呼ばれる）は、自然言語インターフェースを持つ応用システムにおいて文章を解析する上で大きな問題となる。原田研究室は2003年に、文章中のゼロ代名詞と指示代名詞の照応解析システム Anasys を開発している。しかし、一般的な文章への対応を考慮していない点や、解析精度において実利用するには十分なレベルではなかった。

そこで本研究では意味解析システム SAGE の精度向上に伴い、高精度な新照応解析システム Anasys を開発することを目的とした。本システムは、SAGE の一部として構成され、その手法として従来の先行詞を特定するための得点・属性を4種から7種に詳細化した。そして先行詞決定には、サポートベクターマシン (SVM) を用いた。

本システムは大きく(1)照応詞検出部と(2)先行詞解析部の2つに分かれる。

- (1) 照応詞検出部では文章中の各用言の深層格の EDR 共起辞書内における出現割合を計算することで、統計的に各用言の必須格を決定し、文章中の用言が意味解析上 agent 格も a-object 格も持たずに、必須格として agent か a-object を持つ場合、ゼロ代名詞として検出する。指示代名詞は意味解析システム SAGE の品詞解析情報を基に検出する。
- (2) 先行詞解析部は、4ステップで行われる。(i)照応詞を含む文とその前3文後1文を解析範囲とし、その範囲にある名詞を先行詞候補として抽出する。(ii)抽出された先行詞候補に、「概念類似度得点」、「語間距離得点(前方・後方)」、「特性得点」、「共起関係詞種別」「文内の文節位置得点」「固有名詞判定」の7つの得点・属性を計算する。(iii)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・属性を学習識別器サポートベクターマシン(SVM)を用いて総合化し先行詞であるかどうかを判定する。SVMを用いるにあたって必要となる学習データは、あらかじめゼロ代名詞78事例、指示代名詞112事例に対して、先行詞候補を手によって先行詞、非先行詞に判別したデータを用いる。(iv)最後に、SVMによって先行詞と判定された中で最も先行詞らしい文節を、照応詞に照応関係として補完する。なお、学習データの作成を支援するために専用の対話型システムを作成した。

評価実験では、指示代名詞の精度は適合率71.7%、再現率58.9%、ゼロ代名詞の精度は適合率74.4%、再現率55.2%となった。意味解析に基づく照応解析として高い精度を残し、意味解析情報に照応関係を付加する事で、質問応答システム Metis の精度向上に貢献したので本手法の有効性を示すことができたと言える。