

## 206 帰納的学習を用いた意味解析における変換規則獲得手法の改良

嶋家正樹 荒木健治 棚内香次  
北海道大学大学院工学研究科

## 1 はじめに

用例ベースでの自然言語処理は用例の作成に多大な労力がかかる。また、既存のツールの中にもどのような文に対しても精度が高いようなものは存在しない。そこで我々はこれまで帰納的学習を用いた形態素解析、構文解析を行ってきた[1][2]。本稿では帰納的学習を用いた意味解析を提案する。これに形態素解析、構文解析を組み合わせることにより帰納的学習を用いた総合的な自然言語処理システムが完成する。また今後は本研究で利用可能となる意味表現を用いた更なる自然言語処理も考えられる。

このような意味解析システムを完成させるために、複数の構文と意味の組み合わせからそれぞれに同等の表現が現れたときにそれを規則として抽出する実験を行った[3]。しかしコーパスの表現の多彩さから同等であると言える表現の獲得は難しく、獲得された規則は質も量も意味解析に耐えうるものではなかった。

そこで本稿では語の情報は考えずに構文表現の形式のみからルールの獲得を行う。これにより規則数不足の問題を解決すると同時に抽象度の低い規則も獲得することにより解析精度の低下もふせぐことができる。

## 2 意味表現と構文表現

本研究では EDR 日本語コーパス[4]を実験用のデータとして用いているため構文表現と意味表現もそれにならったものを採用している。EDR 日本語コーパスは決して良質のものとは言いきれないが、手軽に入手できる膨大な量を持つコーパスとして本研究では有効であると考えた。

## 3 処理

## 3.1 システムの概要

システムの概要は次のようになる。まず入力された構文解析結果は変換部において既存の規則を用いて意味表現に変換される。この時正しい意味表現も与えられ、変換結果の正誤を確認する。既存のルールで変換できなければ、構文解析結果と正しい意味表現の組からルール獲得部で新たなルールを獲得する。以下でそれぞれの処理について説明する。

## 3.2 規則の獲得

規則の獲得は次の手順で行われる。

1. 入力された構文表現、意味表現をルール候補となる区切りに分割する。
2. 意味表現中にある語を取り出し、それが全て含まれている構文表現を見つけ、その中でもっとも長さが短いものを意味表現に対応付ける。
3. それぞれの表現中に共通して現れる語を変数として変換規則とする。

## 3.3 変換

既存の規則を用いて構文解析結果を意味表現に変換する。この時、階層構造で表される構文解析結果は大きな枠組みから順次変換される。

## 4 実験

EDR 日本語コーパスの構文解析結果とその意味表現の組 622 を用いて規則獲得の実験を行った。

実験の結果 622 の変換規則の獲得に成功した。しかし獲得されたルールは複雑なものが多く見られるため、ルールの一般化が必要であると考えられる。

## 5 おわりに

本稿では変換規則の数が少ないという問題は解消することができた。今後は今回獲得された規則の一般化を行ってより実用的な変換規則を獲得する。その後、それを用いて解析、評価実験を行う予定である。

## 参考文献

- [1] 荒木健治, 棚内香次, “帰納的学習による語の獲得および確実性を用いた語の認識”, 信学論(D-II), vol.J75-D-II, No.7, pp.1213-1221(1992).
- [2] Y. Masatomi, K. Araki, K. Tochinal, “A Syntactic Analysis Method Using Inductive Learning” Proc. of A SC 2000, pp.447-451, Bannf, Canada (2000).
- [3] 嶋家正樹, 荒木健治, 棚内香次, “構文解析結果とその意味表現からの帰納的学習を用いた意味解析規則の獲得” 情報処理北海道シンポジウム 2000, pp.51-53(2000).
- [4] 日本電子化辞書研究所, “EDR 電子化辞書仕様説明書”(1995).