

論文内容の要旨

博士論文題目 Statistical Induction of Tree-Generating Grammars for Natural Language Parsing
(木生成文法モデルの統計的学習による構文解析)

氏名 進藤 裕之

(論文内容の要旨)

自然言語処理において、構文解析は重要な課題である。統計的アプローチによる構文解析では、あらかじめ構文木コーパスから文法モデルに基づく文法規則と確率を推定し、それらの規則群を用いて解析を行う。本論文では主に、確率的文法モデルの設計と、構文木コーパスから文法規則と確率を自動的に推定する手法について論じ、獲得された確率的文法規則を用いて高精度な構文解析器を実現することを目的とする。

近年、任意の大きさの部分木を文法規則とする木生成文法モデルの統計的推論方法が盛んに研究されており、自然言語処理の様々なタスクにおいて一定の成功を収めている。しかしながら、これらの文法モデルを構文解析に用いた場合、高度に最適化された最先端の構文解析器と比較して、精度が劣ることが知られている。

本論文では、まず、上記の文法モデルに木の挿入操作を導入して非連続な部分木パターンも扱えるように改善し、サンプリング法に基づいて本文法モデルの統計的推論を行う手法を提案する。また、実際の構文解析実験において、本手法が構文解析の精度向上に寄与することを示す。

次に、部分木を自動的に細分化することにより、文脈情報を捉えてデータに自動適応することのできる確率的文法モデルと、サンプリング法に基づく統計的推論方法を提案する。また、実際の構文解析実験において、本手法が従来の構文解析器よりも高精度を達成することを示す。

さらに、木生成文法モデルの統計的推論に用いるサンプリング法を効率化する新たな近似サンプリング法を提案する。本手法は、構文木コーパスに含まれる共通の部分木をまとめて更新することで、最適解の探索を効率的に行うことができる。また、実験により本手法の有効性を示す。

(論文審査結果の要旨)

平成 25 年 8 月 2 日に開催した公聴会の結果を参考に平成 25 年 9 月 4 日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動が続けていくための十分な素養を備えていることを示すものと認める。

進藤 裕之は、本博士論文において、確率的構文解析のための文法学習に関する新たな手法を提案した。本博士論文の貢献は次のようにまとめることができる。

1. 従来盛んに研究が行われている句構造文法や木接合文法ではなく、木の挿入操作をもつ木挿入文法をベースとし、構文木コーパスから文法規則の学習を行う方法を示した。この文法の採用により、基本的な文法規則と修飾要素などの随意的な文法規則を区別して文法規則が学習可能となり、よりコンパクトな文法規則学習が行えることを示した。さらに、本手法が実際の構文解析にいよても精度向上に寄与することを示した。
2. 文法規則中の句のラベルを細分化することにより文法規則が適用される文脈情報を自動的に適用学習する手法を提案した。また、マルコフ連鎖モンテカルロ法によるサンプリングを利用した効率的な学習法を提案した。これにより、提案法が既存手法の解析性能を大きく上回る構文解析を達成することを示した。
3. 木生成文法モデルの統計的推論を効率的に行うため、構文解析データ中に現れる共通の部分木まとめてブロック化し、ブロックに含まれる変数の同時分布からサンプリングを行い、その部分木ブロック化サンプラーを従来のマルコフ連鎖モンテカルロ法と組み合わせて交互に実行する手法を提案した。さらに、この手法により尤度の高い文法規則の学習が可能であることを示した。

確率文法に基づく様々な文法規則学習と統計モデル学習を提案し、解析性能が極めて高い構文解析を実現した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士（工学）の学位論文として価値あるものと認める。