

論文内容の要旨

博士論文題目 Studies on Efficient Parsing and Logic-based Inference based on Combinatory Categorical Grammar
(組み合わせ範疇文法に基づく効率の良い構文解析手法と論理推論に関する研究)

氏名 吉川将司

(論文内容の要旨)

組み合わせ範疇文法 (Combinatory Categorical Grammar, 以降 CCG) は、理論言語学において、様々な言語現象に対する適切な説明を与えることが可能である一方で、理論の数学的な性質から、それらの理論的知見をプログラムとして実装可能であるため、自然言語処理の文脈においても注目されている。本論文の目的は、CCG に基づく構文木を論理式に変換し推論を行う自然言語推論システムの開発と拡張を行うことである。本論文の貢献は具体的に、(1) 高速かつ高精度な CCG 構文解析手法、(2) CCG 解析の分野適応手法、(3) 論理推論システムにおける自然言語推論のための公理生成技術の開発である。

最初の貢献は、高速かつ高精度な CCG 構文解析器の開発である。語彙化文法である CCG においては、前終端記号であるカテゴリ (あるいはスーパータグ) が文の構造に関して多くの情報を保持しており、CCG に基づく構文解析では、スーパータグにより決定する統語関係を解決することが大部分を占める。本研究における解析手法ではこの特徴を活かし、CCG 木構造の確率をスーパータグの確率に基づいてモデリングし、それだけでは決まらない曖昧性は CCG 木から取り出される係り受け構造によって解消する。これにより、木構造の確率を局所的な要素の積として表せ、高速な A* 構文解析を用いて文に対する最尤な CCG 木を計算することが可能になる。実験では、日英語の CCG ツリーバンクにおける性能評価に加え、提案法の解析器を論理に基づく自然言語推論システムに統合することで、提案法の解析性能の含意関係認識タスクに対する影響も検証し、有用性を示した。

本研究では科学技術論文や対話を含む様々なドメインに対する CCG と論理推論に基づく自然言語処理システムの開発を目指す。そのためにはこれらの分野のテキストに対して頑健に CCG 解析ができる必要がある。本論文の 2 つ目の研究では、CCG 解析のための新たな分野適応技術を提案する。技術の要となるのは、比較的安価な係り受け構造に基づく言語リソースを用いた CCG ツリーバンクの自動生成である。提案法は、分野適応を行う解析器の構成に依らないものであり、近年の高性能な解析器にも適用可能である。実験では、生物医学論文、疑問文、対話テキスト、数学問題の 4 つの分野において提案法の有用性を検証した。中でも対話テキスト、数学問題において解析性能の著しい向上が見られた。

本研究最後の貢献は、論理推論システムに対するものである。論理推論システムに対して、知識ベースなどから得られる外部知識を導入する場合、知識データを増やすにつれ問題を解く速度が低下してしまうという緊張関係が存在する。そこで、

本研究では知識グラフ補完の技術を応用することによりこの緊張関係を緩和し、最高精度の論理に基づく含意関係認識システムの処理速度を大幅に改善可能であることを示す。さらに、本研究ではその仕組みを定理証明支援系の Coq のプラグインとして実装することにより、自然言語推論のための知識推論を含めた自動推論機構を構築する。実験において、提案法では処理速度を損なわずに知識データを追加することで含意関係認識の精度を向上させることができた。

氏名	吉川将司
----	------

(論文審査結果の要旨)

令和元年12月11日に開催した公聴会の結果を参考に令和2年2月7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。

吉川将司は、本博士論文において、自然言語文の解析性能の向上と分野適応、および、含意推論システムの高速化において顕著な研究成果を達成した。

本論文の貢献は以下のようにまとめることができる。

1. 組み合わせ範疇文法は自然言語文の構造に関して高い記述能力をもつが、解析速度の改善が重要な課題となっている。本論文では、単語に対するスーパータグの確率に基づくモデリングと係り受け構造における共起確率を局所的な情報として利用した高速なA*構文解析法を提案し、組み合わせ範疇文法に基づく高速な統語解析法を提案した。
2. 自然言語の解析では分野依存性が一つの問題であり、特に語の詳細な情報を利用する近年の統語解析においては、新たな分野のテキストに解析システムを適用するためには、該当分野のアノテーションデータを用意する必要があった。本論文では、係り受け構造を利用した言語データを利用し、組み合わせ範疇文法に基づくツリーバンクを自動生成する方法を提案し、安価な方法で新しい分野のアノテーションコーパスを作成する方法を示した。さらに、この方法を用いて、いくつかの分野において、解析性能の向上が達成可能であることを示した。
3. 含意推論に知識ベースなどの外部知識を利用するにあたり、知識グラフ補完の技術を応用することにより、高い精度をもちながら処理速度を大幅に改善する含意関係認識を可能にする方法を示した。この方法を定理証明支援システムのプラグインとして実装することにより、処理速度を損なうことなく、知識ベースの利用により含意関係認識の精度を向上できることを示した。

自然言語処理における統語解析および論理推論の高速化と高精度化法を提案した本研究は、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。