論文内容の要旨

博士論文題目 Development of Pairwise Comparison-based Japanese Dependency Parsers and Application to Corpus Annotation
(候補の直接比較に基づく日本語係り受け解析器の開発とこれを用いたコーパスアノテーション支援)

氏名 岩立将和

(論文内容の要旨)

本論文は、高性能な係り受け解析モデルの実現に向けて3つのアプローチを提案している。高精度の単体解析モデルの開発、係り受け解析器の組み合わせによる精度向上の実現、並列構造解析器との組み合わせによる精度向上の実現である。

日本語係り受け解析においては、工藤らの相対的な係りやすさを考慮したモデルが、決定的解析モデルや文脈自由文法を用いた構文解析アルゴリズムに基づく手法を上回る精度を示している。工藤らのモデルではすべての係り先候補文節間の相対的な選択選好の強さを学習し、解析時には選択選好の強さのスコアを比較して係り先を選択する。これに対し本論文では、一度に比較対象とする係り先候補文節を二候補に限定し、選択選好を二つの候補同士の対戦からなるトーナメントで直接表現したモデル(トーナメントモデル)を提案している。SVMの出力する分離平面からの距離をスコアとみなし、このスコアと部分解析精度との関連を調査したところ、トーナメントモデルによる手法がより高性能であり、スコア順に係り受け関係を選択することで99%の正解率で60%の被覆率を達成できることを示した。

また、トーナメントモデルとは大きく性質の異なる日本語係り受け解析器であるKNPの解析結果をトーナメントモデルの素性として用いることで、両解析器を上回る精度を達成できることを示す。

係り受け解析器にとって解析が難しい問題の一つである並列構造に対して、dual decomposition アルゴリズムを用いて並列構造解析器の出力と係り受け解析器の出力を統合した。このアルゴリズムはスコアに基づいて最適解を選択する解析器を想定しているため、トーナメントモデルの長所を保持しつつ分離平面からの距離に基づいて係り受け関係にスコアを付与したモデルを用いた。

コーパス構築において人間による作業の労力を削減し効率的なアノテーション作業を実現するために、本稿では2つのアプローチをとった。まず、SVMのスコアを用いて係り受け解析器の能動学習を行うことで、人手によるアノテーションコストを削減しながら、高性能な係り受け解析器を構築できることが確認された。次に、係り受け構造と並列構造を別々にアノテーションするという状況で、並列構造が与えられると可能な係り受け構造が制約されることを利用して高精度の係り受け解析器を構築することで、係り受けアノテーションの効率をさらに向上させることを示した。

氏名	岩立将和
----	------

(論文審査結果の要旨)

平成24年1月23日に開催した公聴会の結果を参考に平成24年2月13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、研究活動を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。

岩立将和は、本博士論文において、日本語係り受け解析の解析精度向上とそのためコーパス構築に関する一連の研究を行い、次のような新しい解析手法や今後の発展につながる様々な提案を行った。

1. 日本語係り受け解析の係り受け曖昧性の問題に対処し、性能向上を達成するための方法として、競合する2文節を直接比較するトーナメントモデルの適用を行い、従来のどの手法よりも高性能な機械学習に基づく日本語係り受け解析システムを構築した。さらに、同手法を用いることにより、高い被覆率で極めて高い解析精度が保証される部分的係り受け解析が実現できることを示した。
2. 規則に基づく日本語係り受け解析システム KNP の解析結果を学習のための素性として用いることにより、トーナメントモデルに基づく係り受け解析の性能向上が達成可能であることを示した。
3. 日本語の統語解析の困難な問題である並列構造解析の問題に対処するため、並列構造範囲の同定と係り受け解析を Dual Decomposition アルゴリズムを用いて統合し、並列構造解析の精度向上の可能性を示した。
4. さらなる解析精度向上のためのコーパスアノテーション作業の効率を向上させるため、能動学習に基づく有効な事例選択、および、並列構造アノテーションと係り受けアノテーションを分離した方法を示し、効率的なアノテーション手法を示した。

日本語係り受け解析の性能向上と関連する様々な問題への対処方法を提案した本研究は、独創性が高く、しかも実用面でも有用であり、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。