

審査の結果の要旨

氏名 松崎 拓也

本論文では、決定性の曖昧性解消器をスーパータギングに基づいて行う手法にさらにCFGフィルタリングという技術を適用して、これら2つの前処理技術を組み合わせ、それによって高効率なHPSG構文解析手法を提案したものである。提案方法が基本としているスーパータギングは、語彙化文法の構文解析を高速かつ精度よく行うために有効な方法として近年研究されているもので、入力文の各単語に対し少数の語彙項目をあらかじめ選択することで構文解析の際の探索空間を狭めるという手法である。しかしながら、スーパータギングを用いた既存の解析手法のうち多くのもでは、スーパータギングの際の誤りによる構文解析の失敗を防ぐためにチャートパーズングが用いられており、これが最も計算時間を要する部分となっており、その解決が望まれていた。スーパータギングの際の誤りによる構文解析の失敗はHPSG文法の場合にはさらに問題になる。これは次の2つの理由による。1つ目は、HPSG文法における語彙項目の数が一般に多数であることで、大規模解析済みコーパスから獲得した文法の場合に特に顕著な点で、このことが他の比較的少数の語彙項目を持つ文法に比べてHPSGにおけるスーパータギングを難しくしている。2つ目は、文法で用いられているデータ構造が複雑であることによって、チャートパーズングによるHPSG構文解析はそれ自体が一般に高コストとなることであり、このことが各単語に割り当てる語彙項目数の増加と相まって解析時間を急激に増大させる原因となる。

本論文では、従来のスーパータギングが有した上記の問題点を解決するために、それにCFGフィルタリングの手法を組み合わせ、さらに強力な前処理を行うことで、非常に高速な構文解析が可能であることを示している。CFGフィルタリングは、CFGによるHPSGの近似に基づく前処理手法である。CFGフィルタリングとスーパータギングを組み合わせることで、ほぼ確実に解析可能であるような語彙項目列を高速に列挙することが可能となる。具体的には、次のように構文解析を行う。まず、CFGフィルタリングを用いて、入力単語列に割り当てる語彙項目列のうち大域的に一貫性のあるものを、スーパータギングによるスコアの順に列挙する。そして、それらの語彙項目列を1つずつ曖昧性解消器へ入力し、最初に得られた構文木を出力とする。この手法においては、既存手法におけるチャートパーザを決定性の曖昧性解決器に置き換えることができ、非常に高速な構文解析を実現できる。曖昧性解消器は決定性のシフト還元型構文解析器として実装され、分類器を用いて構文解析動作を順次選択していくことで構文木をひとつ選び出す。この構文解析動作の際にも、近似CFGによる入力文上の構文森を利用する。大規模HPSG文法を用いた実験の結果から、提案手法は、同じ文法を用いた既存手法に比べ同程度の解析精度をもち、約10倍(20ミリ秒/文)の速度を達成することが明らかになった。これは本手法が非常に優れたものであることを示すものである。

このように本論文は、HPSG文法の構文解析法について新たな方法を与え、その有効性を実証したものであり、情報理工学のコンピュータ科学分野、特に自然言語処理において、顕著な貢献をするものである。よって本論文は博士(情報理工学)の学位請求論文として合格と認められる。