

審査の結果の要旨

氏 名 平 川 秀 樹

自然言語文解析システムにおいては、入力文に対して既知の言語知識を適用して最も適切な解釈を出力することが目的となる。一般に、自然言語文解析システムは、形態素、構文、意味、文脈の各層において有効となる言語知識を扱う必要があり、各層の知識記述の方法として、句構造木、依存木、述語論理式など様々な文解釈記述枠組みが提案されている。これらの枠組みは、知識記述の質や能力、すなわちシステムの性能を規定する。したがって、どのような枠組みを採用するかは自然言語文解析システム設計上重要な課題の1つである。また、言語知識は、計算機処理の視点から見て、入力文に対して不可能な解釈を排除する「制約知識」と可能な解釈の優先順序付けを行う「選好知識」の2種類に分類される。制約知識は自然言語の各層で生じる解釈の組み合わせ爆発の抑制に有効であるが、その過度の適用は正解解釈の枝刈りによる性能劣化を引き起こしてしまう。一方、自然言語の各層の選好知識は、相異なる解釈を支持することもあり、全体として最適な解を得るためには各層の知識を統合評価する必要がある。このように、多層の知識、制約/選好知識の統合をいかに行うかは、実用的な自然言語文解析システム設計において解決すべき重要な課題である。

本論文は「Preference Dependency Grammar(PDG): Sentence Analysis Method Based on Integrated Multilevel Preference and Constraint (邦訳: 選好依存文法: 多層の選好/制約知識を統合した文解析方式)」と題し、英文で書かれ、上記の自然言語文解析システム設計における課題を解決する方法に関して論じたものであり、下記に示す8章からなる。

第1章(Introduction)では、自然言語文解析システム研究における背景、特に句構造文法と依存構造文法、ならびに、その解析手法(句構造解析と依存構造解析)について概観し、問題点を指摘し、さらに本論文で展開する解決策の方針を示している。

第2章(Sentence Analysis Model and the PDG design)では、上記の問題点を解決するための多レベルの言語知識を統合する文解析モデルについて検討し、PDGのシステムの概要について述べている。

第3章(Packed Shared Data Structures)は、PDGの圧縮共有データ構造とその構成法、特に、構文層の2つのデータ構造である句構造森と依存森について述べている。依存森は、依存木の集合を表す新たなデータ構造であり、依存グラフ、制約マトリックス、選好マトリックスから構成される。その最大の特徴は、本章で証明した句構造森とのマッピングにおける完全性と健全性、すなわち句構造木と依存木が相互に対応を持つことである。これによって、句構造と依存構造レベルを持つ多レベル圧縮共有データ結合モデルを実現可能となった。

第4章(Optimum Solution Search)では、依存森として統一的に表現された構文構造の集合から最適な依存構造を探索すること、すなわちスコア付き依存グラフから依存木の整合性を規定する制約条件を満足する最大のスコアを持つ木の探索を行う探索アルゴリズムを提案している。具体的には、分枝限定法に基づくグ

ラフ分枝アルゴリズムを提案し、被覆制約、単一役割制約、非交差制約、多重格禁止制約などの制約を充足する最適依存木を 99.8%の文で得ることを実証するとともに、組み合わせ爆発に対応する枝刈り手法の性能も優れていることを実験的に示した。

第 5 章(Scoring)では、多層の選好知識を依存森の選好知識表現に統合する方法を提案した。

第 6 章(Evaluation)では、ここまでで提案してきたアルゴリズムを英語技術文 12.5 万文を用いて実験的に評価している。まず、907 個の規則からなる英語基本文法による構文解析器を実装し、選好知識を統合することによって、選好知識のない場合に比べて、構文構造のアークの正解率で 10.9%を改善した 88.3%の正解率を達成した。

第 7 章(Future Work)では、今後の課題として、実システムの開発、意味処理への展開、解析・生成の双方向モデルの検討の 3 点をあげている。

第 8 章(Conclusion)は、本論文のまとめである。

以上を要するに、本論文は多層の選好／制約知識を統合した自然言語文の処理に関して、依存森と統語森の対応付けを証明し、その両者における選好／制約を活用する効率の良い解析アルゴリズムを提案するとともに、実際のシステムを構築し性能の改善を評価実験によって示したもので、計算機科学の発展に寄与するところが大きい。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。