論文内容の要旨

博士論文題目 An Integrated Robust Parsing using Multiple Knowledge Sources
(複数の知識源を統合的に用いた頑健な自然言語処理)

氏名 今一 修

(論文内容の要旨)

自然言語処理システムを実用的なものにするためには、文法的に適格な文だけではなく、文法的に不適格な文も処理できるようなシステムを構築する必要がある。これまでに文法的に不適格文を処理するための様々な手法が提案されているが、それらの多くは、統語的な情報だけを利用したものや、意味的な情報だけを利用したものなど、一部の不適格性を対象にしたものがほとんどであった。より柔軟に文法的に不適格文を処理するためには、統語・意味・文脈情報を統合的に利用することが必要である。これらの情報は言語情報に属するものであるが、近年では、大規模コーパスの普及により、文法知識や語彙知識を統計的に獲得することが可能になっている。本論文では、統語・意味・文脈情報などの言語情報だけではなくコーパスから獲得される統計情報も利用する統合的な解析手法を提案する。

第3章では、統語・意味・文脈情報を統合的に利用して文法的に不適格文を処理する手法について述べる。ここでは新しい概念としてコストと報酬が導入される。コストは部分解析木に含まれる不適格性の割合を示し、報酬は部分解析木が解析成功に貢献する割合を示している。この手法は、二段階モデルに基づいており、通常の解析が失敗した場合にのみ、不適格文処理を行なう。何らかの不適格性が原因で入力文の解析に失敗した場合、それまでに生成された部分解析木の中から失敗を回復するために有効な部分解析木をコストと報酬のトレードオフを考慮して選択する。この手法には4つのモジュールが含まれているが、それらはすべて、コストと報酬の二つの指標によって制御される。

第4章では、統計情報を言語情報(文法制約)と組み合わせて解析を行なう手法について述べる。統計情報としては、確率付き係り受け情報を用いる。この手法では、統計情報と文法制約を知識レベルでは独立に記述しておき、解析を行なうレベルでそれらを統合的に利用している。そのため、知識レベルでの高いモジュール性が実現でき、システムの保守・管理の点でも自由度が高くなる。また、係り受け情報を利用することにより、もっともらしい解析結果から優先的に生成することができ、パーザの効率化も実現できる。

第5章では、第4章で述べた手法を第3章で述べた手法と融合し、言語情報と統計情報を統合的に利用した文法的に不適格文処理について述べる。第3章で述べた手法は二段階モデルに基づいており、通常の解析が失敗した場合に、不適格文処理を実行している。それに対して、本章では、二段階モデルを一段階モデルに再構成し、効率的に文法的に不適格文の処理を行なう手法を提案する。一段階モデルを実現するために、第4章で用いたコストと報酬を第5章で用いた統計情報と統合して一つの指標にしている。この統合した指標に基づいて解析を行なうことによって、統計的にもっともらしく、かつ、不適格性の少ない構造から優先的に処理していくことができる。

氏名	今一修
----	-----

(論文審査結果の要旨)

平成9年12月26日に開催した公聴会の結果を参考に平成10年2月17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。

今一修君は、本博士論文において、解析システムが想定する文法的に逸脱した文をも柔軟に処理するため、統語、意味、文脈など種々の情報を統合的に利用する枠組みの提案を行った。本研究は次のようにまとめることができる。

- 文法的に逸脱した文の処理は、従来は、制約を緩めた文法を記述して文法違反の文をも解析できるようにする、あるいは、文の解析に失敗した場合に、誤りの原因を推定しつつ制約を緩和する、などの方法が取られてきた。本研究では、単一化文法を利用し、解析処理の失敗を単一化の失敗として統一的に捉え、単一化の失敗時に処理を破棄するのではなく、失敗の度合いに応じてコストを発生することにより、文法的な逸脱に統一的な測度を与えた。さらに、コストを発生した解析結果には、コスト以外に全体的な解析へ貢献すると思われる尺度を句の大きさや種類によって報酬として定義した。このコストと報酬のバランスによって、最適な解析誤り修正を行う枠組みを提案した。本提案は、従来の手法の多くの欠点を克服しており、独創性の高い成果である。
- 前項の解析手法は単一化文法に基づいており、また、文法的に不適格な構造を含む場合は、適格文の解析処理が失敗した時にのみ起動される。この手法の効率上の問題を解決するため、統計的構文解析と上記の手法との統合の方法を提案している。統計的な方法だけでは曖昧性の解消能力は不十分であり、誤った結果を出力する危険性があるが、文法的な制約とコストに基づく制約緩和の方法と融合することにより、統計的にはもっともらしく、不適格性が少ない解釈を最優先で出力する方法を実現している。従来、独立して研究されて来た統計に基づく言語解析と制約に基づく言語解析の手法を融合し、さらに、制約違反の概念にまで対象を拡張した。統計情報と言語的制約の新しい統合法の方向性を示した研究として高く評価できる。

以上のように、本研究は、文法的に不適格な文を解析するためのモジュール性の高い統合環境を提案しており、今後の実用化へ向けた自然言語処理への新しい方向性を示している。情報科学、言語処理の分野において高い貢献があると評価する。