

論文審査の結果の要旨

氏名 鳥澤 健太郎

インターネットの爆発的な普及に伴い、膨大な量の自然言語で書かれたテキストに対して翻訳、情報抽出などの知的な処理を計算機によって施すことへの要求が高まってきている。一方で、従来より行われてきた自然言語に関する研究すなわち言語学においても数学的な定式化が重要視されるようになってきた。そのような自然言語の数学的定式化をテキストの計算機による知的処理において利用することは自然な発想であるが、自然言語の多様性を扱える数学的定式化は、計算機上での容易な実現を許さず、テキストを解析する速度、解析の精度など困難な問題を生じさせてきた。本論文に報告されている研究は、言語学的及び数学的に妥当な文法枠組みである主辞駆動句構造文法 (Head-driven Phrase Structure Grammar) に注目し、テキストの知的処理すなわち自然言語処理における枠組みの実用化を狙ったものである。より具体的には、主辞駆動句構造文法を用いてテキストを解析するアルゴリズムすなわち構文解析の、新しい発想による高速なアルゴリズムを提案し、新聞などのテキストを用いて高速化を実証している。さらに、高精度でテキストを解析する際に必要となる単語の意味的分類をテキストから自動的に学習するアルゴリズムを提案し、やはり新聞などを用いてその妥当性を検証している。

本論文は全7章からなり、第1章で研究の動機について、第2章では背景について述べた後、第3章、第4章において主辞駆動句構造文法の計算機科学的により妥当な定式化、記述方法を与えている。第5章では、それらの定式化にのっとり高速な構文解析アルゴリズムを提案、その高速化を実証、さらには、そのアルゴリズムの正当性を証明している。第6章では、単語の分類をテキストから自動的に学習するアルゴリズムを提案し、その性質を検討して、実験によりその妥当性を実証している。

第3章では、文法記述の構成要素となる確定節プログラムの定式化を従来提案されてきたものとは異なった手法で行っている。この定式化により、確定節プログラムの計算結果における単調性及び遅延評価の正当性が導かれる。これらの性質は第5章で提案される高速な構文解析アルゴリズムを可能とし、計算機科学的に意義深いといえる。

第4章では、第3章で定式化された確定節プログラムを用いて、言語学者による主辞駆動句構造文法の定式化を計算機科学的により取り扱い易いものへと翻訳している。さらには、その文法を用いた構文解析の定義を与えている。これにより確定節プログラムが主辞駆動句構造文法を記述するのに十分な柔軟性を持っていることが示され、また、構文解析の計算機科学的な性質についての詳細な検討がなされている。

第5章では、主辞駆動句構造文法のための高速な構文解析アルゴリズムが提案されている。この手法は、与えられた主辞駆動句構造文法をより単純な文法記述形式である文脈自由文法に自動的に変換し、構文解析の際に必要な試行錯誤のプロセスを減らすことによって高速化を図る。文脈自由文法への変換は文法記述の部分評価と考えることができ、文法の内に含まれている確定節プログラムの部分評価として捉えられる。本章ではこの観点からアルゴリズムの性質を詳細に検討し、その正当性を証明している。従来、主辞駆動句構造文法による構文解析はその速度が遅いことから実用化できないとされてきたが、本章で提案されたアルゴリズムは、新聞などに関して十分実用的な速度を達成しており、実用的観点から見ても意義深い結果であるといえる。

第6章では、意味的な語彙分類の学習手法が提案されている。意味的な語彙分類は構文解析や、表層の文を、知的な自然言語処理において必要な意味表現と呼ばれるデータ構造に写像する際に必要となる。本章では、語彙分類をテキストから自動的に学習するアルゴリズムを提案、新聞から語彙分類を自動学習させる実験をおこない、主辞駆動句構造文法を用いた統計的構文解析での精度向上、被験者の直感との整合性によってその評価をおこなっている。さらには提案するアルゴリズムが「極限における同定」という学習結果の収束に関する性質を満たすことが示されている。このようなアルゴリズムで同時に現実のテキストに適用できるものはこれまで提案されたことがなく、計算論的学習理論における前進といえる。

第7章では、本論文の結論と今後の展望に着いて述べている。主辞駆動句構造文法の構文解析手法に関しては、他の解析手法との組み合わせによりさらなる速度向上が期待できること、また、語彙分類の学習に関しては、構文解析などの技法と組み合わせることにより、その適用範囲を広げることが重要であることなどを述べている。

以上、本論文では、主辞駆動句構造文法のための新しい方式による構文解析アルゴリズム、及び、語彙の自動分類技法を提案、それらの有効性を実証している。その後、この方式はドイツや米国の有力な研究グループでも使われるようになり国際的にも影響力の大きい成果となっている。また提案された解析アルゴリズム、自動的語彙分類技法の計算機科学的な性質も明らかにされており、計算機科学に対する貢献も大きい。よって、本論文は博士(理)を与えるに十分な内容を有している。

なお、本論文第5章は、辻井潤一氏、西田健二氏、及び宮尾祐介氏との共同研究であるが、論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断される。