

氏 名	石原 靖弘
授与した学位	博 士
専攻分野の名称	工 学
学位授与番号	博甲第5410号
学位授与の日付	平成28年 9月30日
学位授与の要件	自然科学研究科産業創成工学専攻 (学位規則第4条第1項該当)
学位論文の題目	項の末尾情報を利用した日本語意味役割付与システムの高性能化
論文審査委員	講師 竹内 孔一 教授 尺長 健 教授 太田 学

学位論文内容の要旨

本論文では、項の末尾表現に注目した特微量を利用した日本語意味役割付与システムの高性能化について述べる。意味役割とは文中の述語と、その述語と係り関係にある句や節などの構成要素との意味的関係を表す。意味役割が同定できると構文として異なる表現も意味が同様であることがとらえられるようになる。構成要素は項と呼ばれ、例えば「本を買った」と「買った本」の2文において、「買う」の述語に対する項である「本」の意味的関係は同じである。

述語に対する項の意味役割を識別する研究は英語を対象に、構文情報をベースとした手法が多く発表されている。一方で日本語を対象にした研究では、表層格を意味的関係ラベルとする述語項構造解析が多く研究されている。しかしながら、付与対象の関係ラベルの種類数がより多い意味役割付与における有効性は明らかではない。

そこで、本研究では、最近構築された日本語意味役割付与データを対象に、項の末尾情報を利用した統計的学習モデルを基にした意味役割付与システムを構築し、提案する特微量の有効性を実験により明らかにする。意味役割付与コーパスとして国立国語研究所が構築した現代日本語書き言葉均衡コーパスに意味役割を付与したコーパスを利用する。日本語意味役割付与システムとして統計的学習モデルである Conditional Random Field (CRF) を利用する。日本語では項の末尾に出現する機能表現が意味役割と密接に関係することから本論文では項の末尾表現をとらえる複数の手法を提案する。

項の末尾表現を特徴化する手法として固定長 n-gram および可変長 n-gram を利用したモデルを提案する。可変長 n-gram モデルでは表層、基本形、読み、品詞といった複数の情報を扱うことができる Hierarchical Tag Context Tree (HTCT) を利用する。これにより可変長でかつ最適な末尾情報を利用した n-gram を意味役割付与の識別に利用することができる。

固定長モデル、および可変長モデルを基本として、先行研究である日本語述語項構造解析において提案された特微量との比較実験を行った結果、項の末尾情報を加えた提案手法が高い識別性能を示すことを明らかにした。また、両モデルの併用により識別性能がさらに向上することを明らかにした。さらに、大規模言語資源を用いた特微量を加えた場合の比較実験において、項の末尾情報を加えた提案手法がさらに高い識別性能を示すことを明らかにした。これらの実験では、項の末尾情報を利用したモデルの性能が利用していないモデルと比べて 5% 有意水準で高いことを確認した。以上の結果から本論文で提案する項の末尾情報を利用した提案手法は日本語意味役割付与システムの高性能化において有効であることを明らかにした。

論文審査結果の要旨

近年の人工知能技術の発展は、言葉を入力とするサービスの社会への普及を促進するとともに、文の意味をより正確に理解する意味処理技術への期待を高めている。特に日本語の場合、意味を正確に理解するためには文の構造化技術が重要である。意味役割付与は、この構造化技術の一つであり、文内の述語と項の意味的関係を同定することにより、日本語文書中に生じる多様な表現に対して、意味の識別を可能とする技術である。本論文では、意味役割付与の性能を向上させるために、項の末尾情報を利用する手法を提案している。これは、項の末尾に出現する機能表現が意味役割と密接に関係するという日本語の特性に着目したものである。項の末尾情報を取り入れた新たなアプローチを提案している点が評価できる。

提案手法は、項の末尾情報を統計的学习モデルである Conditional Random Field に特徴量として取り入れて利用する。項の末尾情報を取り入れる方法としては、2種類のモデル（固定長モデル、および、可変長モデル）を提案し比較している。ここで、固定長モデルは単語の表層、基本形、品詞、読みの全ての情報を n-gram として取り入れるモデルである。一方、可変長モデルは意味役割付与に有益な情報を Hierarchical Tag Context Tree を利用して階層構造の中から選択して取り入れるモデルである。日本語意味役割付与実験により、先行研究で提案された特徴量を利用した意味役割付与システムに比べて、各モデルを利用した意味役割付与システムが共に高い識別性能を示すことを明らかにしている。また、両モデルの併用により識別性能がさらに向上することを明らかにしている。さらに、大規模言語資源を用いた特徴量を加えた場合において、提案手法が有効であることを明らかにしている。なお、これらの実験では、項の末尾情報を利用したモデルの識別性能が利用しないモデルに比べて有意に高いことを確認している。

論文発表会では、以上の研究内容について適切な説明が行われ、質疑に対する応答も適切であった。本論文は、文の意味理解への応用が期待される日本語意味役割付与において、項の末尾情報を利用する手法を提案し、その有効性を定量的かつ多面的に明らかにしており、情報工学に寄与するところが大きい。よって、本論文は博士（工学）の学位論文に値すると認める。