

審査の結果の要旨

氏名 ラオクンラット ナッスダ

本論文は、「Exploiting Non-Local Information in Relation Extraction from Documents (和題：文書からの関係抽出における非局所的情報の利用)」と題し、全 8 章から成る。計算機によってテキスト中に記述されているエンティティ間の関係を抽出する処理は、教師付き機械学習の問題として考えることが可能である。本論文は、関係抽出のための機械学習モデルに、実用的な計算コストで非局所的情報を組み込む手法を提案し、その有効性を実験的に検証した論文である。

第 1 章は「Introduction」と題し、自然言語処理技術の応用と必要性を述べ、本研究が対象とするタスクである関係抽出 (relation extraction) 処理を解説している。さらに、関係抽出処理において非局所情報 (nonlocal information) を利用するための既存の手法の長所と短所を整理し、本研究で stacked learning と呼ばれる機械学習手法をとる理由を説明している。

第 2 章は「Machine Learning」と題し、教師付機械学習による関係抽出の手法を説明している。特に、本研究で非局所的情報を利用するために用いる機械学習の枠組みである stacked learning についてその詳細を説明している。

第 3 章は、「Temporal relation extraction」と題し、本論文では関係抽出処理のなかでも、特に時間関係認識 (temporal relation extraction) と呼ばれるタスクに着目することを説明し、その学習・評価に用いるコーパスについて解説している。また、時間関係認識における先行研究を紹介し、局所的な情報だけでは、出力される関係の間に矛盾が生じ得ることを説明し、それが非局所的な情報を機械学習モデルに組み込むことによって解消されうることを説明している。

第 4 章は、「Local approach to temporal relation extraction」と題し、局所的な情報のみ、すなわち各時間関係の分類を独立に行う手法を述べている。分類の際に用いる特徴量 (素性) として、深い構文解析結果から得られる述語・項構造や句構造上での経路情報などの特徴量を利用することを提案し、従来手

法よりも高い精度が得られることを実験的に示している。

第5章は、「**Non-local approach to temporal relation extraction**」と題し、本研究の中心となる、非局所的情報を利用する手法を解説している。まず、イベントや時間表現の間に定義されている時間関係がスパースであることを述べ、それらを直接用いるだけでは非局所的情報を効果的に利用できないことを説明している。その問題に対処するため、複数の時間関係を組み合わせることで新たな時間関係を生成し、密な時間関係グラフが構築できることを説明している。またそれによって、非局所的情報の利用が時間関係抽出の精度向上につながることを実験的に明らかにしている。

第6章は、「**Using probability values as real-valued features**」と題し、**stacked learning** における最初のフェイズの分類結果の信頼性に関する情報を第2フェイズの分類の際に考慮する手法を述べている。具体的には、最初のフェイズのロジスティック回帰モデルによって得られる確率値を、第2フェイズの分類の際の特徴量として利用することを提案し、実験により確率値の利用が精度向上に寄与することを述べている。

第7章は、「**Discussion**」と題し、実験結果の分析を行っている。特に、深い構文解析によって得られる述語・項構造に基く特徴量を利用することが大きな精度向上につながる理由と、特定の種類の時間関係の分類が正しく行えないことの原因について詳細に考察している。

第8章は、「**Conclusion**」と題し、本論文を総括し、今後の課題について述べている。

以上これを要するに、本論文は自然言語文書の意味理解における重要な処理のひとつであるイベント・時間表現の間の関係認識処理において、非局所的情報を利用することで、実用的な計算コストでその精度向上が可能であることを明らかにしたものであり、情報処理工学に貢献するところが少なくない。

よって本論文は博士（工学）の学位請求論文として合格と認められる。