

論文内容の要旨

博士論文題目 Relation Extraction: Perspective from Weakly Supervised Methods
(関係抽出:弱教師あり手法の観点から)

氏名 Phi Van Thuy

(論文内容の要旨)

関係抽出は、文章中で言及されている対象の間の意味関係を認識および抽出するタスクである。教師付き学習に基づくこれまでの関係抽出システムは、関係毎のラベル付きの大規模な訓練データが必要であった。ただし、実際には、ほとんどの関係抽出タスクには、教師ありの訓練データが存在しない。

本研究では、主に2つの弱教師ありのアプローチに焦点を当てる。すなわち、ブートストラッピング、および、遠距離教師付きの (distantly supervised) 関係抽出方法である。これら両手法は、教師付き学習において、ラベル付きサンプルを取得するコストを削減するための方法である。本研究では、最初に、ブートストラップにおけるシードサンプルの選択の自動化、および、遠距離教師付き関係抽出におけるノイズの削減というサブタスクに取り組む。この取り組みは、それらの問題をランキング問題として定式化した初めての研究であり、これら両サブタスクに適用可能な手法を提案する。実験により、提案された方法は、これらのサブタスクに対して、従来のベースラインシステムよりも優れた性能を達成することがわかった。

本論文の後半では、遠距離教師付き学習、すなわち、知識ベースに格納された関係知識を実文章と照合することによって自動的に訓練データを生成する弱教師あり学習法に焦点を当てる。本論文で提案する新たな手法は、双方向ゲート型再帰ユニットモデルと階層的注意機構を組み合わせた新しいニューラルモデルであり、関係抽出に適している。新たに追加したピースワイズアテンションと呼ばれる注意機構は、セグメントレベルの表現によって構築され、遠距離教師付き学習の性能を大幅に向上させることを示す。さらに、遠距離学習によって自動的に収集されたサンプル集合の中で最も正例として確からしいデータを限られた文脈の中で推定する文脈的推定法を提案する。実験により、我々が提案した方法は、ベンチマークデータセットに対して、最新のベースラインよりも優れていることがわかった。

氏名	Phi Van Thuy
----	--------------

(論文審査結果の要旨)

令和元年7月24日に開催した公聴会の結果を参考に令和元年8月22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。

Phi Van Thuy は、本博士論文において、弱教師あり学習を前提とした関係抽出に関する研究を行い、ブートストラップ法におけるシード選択、および、遠距離教師付き学習におけるノイズ削減について統一的な視点を提供し、その性質を探求した。また、遠距離教師付き学習のための新たなニューラルモデルを提案し、従来法を上回る性能を達成することを実験により示した。本論文の貢献は以下のようにまとめることができる。

1. 関係抽出タスクにおける弱教師付き学習であるブートストラップ法および遠距離教師付き学習について統一的な議論を行い、前者におけるシード選択および後者におけるノイズ削減のためのランキングに基づく手法を提案した。
2. シード選択およびノイズ削減に関し、提案手法が従来手法を上回ることを示した。
3. 関係抽出に対する遠距離教師付き学習のため、両方向再帰ニューラルネットワーク上にピースワイズアテンションと呼ぶ注意機構をもつ新たな手法を提案し、従来手法を上回る認識性能が可能であることを示した。
4. 遠距離学習によって自動的に収集されたサンプル集合の中で最も正例として確からしいデータを限られた文脈の中で推定する文脈的推定法を提案し、実験により提案手法が、ベンチマークデータセットに対して、最新のベースラインよりも優れていることを示した。

対象間の関係抽出に対し、特に弱教師あり学習の設定で、シード選択やノイズ除去に関する手法や、新たな関係認識手法を提案した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士(工学)の学位論文として価値あるものと認める。