

論文内容の要旨

博士論文題目 Event Relation Acquisition from Large Text Corpora
(大規模テキストからの事態間関係知識の獲得)

氏名 阿部 修也

(論文内容の要旨)

テキスト中の含意関係や因果関係を理解することが、質問応答、情報抽出、複数文章要約などの自然言語処理の応用に役立つと知られている。これを実現するためには、例えば、動詞「洗う」と動詞句「きれいになる」が、何かを洗うという行為の結果としてその何かきれいになるという因果関係である、といったような知識が必要である。本論文では、事態と事態の間にある関係を大規模にかつ機械的に獲得するために Extended Espresso と Two-phase method をという2つの手法を提案し、これらを用いた実験結果を示す。

Extended Espresso は、大規模コーパスから事態表現間の意味的関係の知識の獲得を目的として、実体間関係獲得手法として提案された Espresso を事態間関係に適用できるように拡張した手法である。この拡張は主に次の2つの点からなる。(1) 知識獲得のために事態表現を定義。(2) 事態間関係に適合するように共起パターンのテンプレートを拡張。日本語 Web コーパスを用いて実験したところ、(a) 事態間関係獲得に有用な共起パターンが多数存在し、パターンの学習が有効であることがわかった。また行為-効果関係については5億文 Web コーパスから少なくとも5000種類の事態対を約66%の精度で獲得することができた。

Two-phase method は、行為-効果関係、行為-手段関係のような事態間の関係を大規模コーパスから自動的に獲得するための手法である。共起パターンを利用する手法では事態を表現する述語間で共有される項を認識することが難しいという問題がある。これに対処するために、述語間で共有される名詞(アンカー)を用いて共有項を獲得し、共起パターンを用いて獲得した所与の関係を満たす述語対と共有項を組み合わせ、共有項と共に事態間関係を獲得することを提案する。このとき2種類の異なるアンカーを用いることで、精度を保ったまま再現率を向上できることを確認した。

氏名	阿部 修也
----	-------

(論文審査結果の要旨)

平成21年12月25日に開催した公聴会の結果を参考に平成22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、研究活動を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。

阿部 修也は、本博士論文において、文章中に記述された事象間の因果関係のような意味関係を理解するための基盤資源として、特定の意味関係をもつ事態の対を自動的に獲得する手法について研究し、2つ手法を提案することによって、大規模なテキストから、特定の意味関係をもつ事態対を自動的に獲得する手法、および、述語間で共有される名詞を利用することにより共通の項とともに事態間関係を獲得する手法を提案した。

本論文の貢献は、次のようにまとめることができる。

1. 実態間関係抽出のためのブートストラッピング法である Espresso アルゴリズムを事態間関係抽出に適用できるように拡張し、高い精度を維持しつつ、カバレッジの広い事態間関係の自動抽出が可能であることを示した。
2. 事態間関係抽出に関する重要な問題として、対となる事態で共通に参照される項を同定する問題がある。現実の文では、共通の項が常に存在するとは限らず、たとえ事態間関係が抽出できても、共通項の情報を入手することが容易でない場合がある。この問題を解決するために、述語対の同定と共通項同定を別処理として行う2段階法を提案した。前者では、事態間抽出のためのパターンに基づく手法を用いて述語対を抽出し、後者では、それぞれの述語と共起するアンカーとしての名詞の情報から共通項を同定する方法を示した。
3. 上記の共通項抽出において、アンカーとして用いる名詞の同定のために、具体事例に基づく方法とタイプに基づく方法という2つの手法を提案し、高い精度で共通項同定が可能であることを示した。

以上の通り、事象間関係知識の自動獲得に関する手法を提案した本研究は、独創性が高くかつ実用的であり、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。