

論文内容の要旨

博士論文題目 Acquiring Causal Knowledge from Text Using Connective Markers
(接続標識に基づく文書集合からの因果関係知識獲得)

氏名 乾 孝司

(論文内容の要旨)

人間のような深い言語理解能力を工学的に実現することを妨げている問題のひとつとして、計算機で利用可能な大量の常識的知識をいかに構築するかという問題がある。本論文では、常識的知識のうち、人工知能の分野をはじめ、多くの学術領域において関心が向けられている因果関係に関する知識に着目し、大規模な電子化文書集合から因果関係知識を自動的に獲得する方法について論じる。

因果関係とは、何らかの依存関係がある出来事間の関係を指し、例えば、「大雨が降ると洪水が発生する」や「列車に乗るので切符を買う」などがある。従来より、因果関係は様々な観点に従った幾つかの種類に分類されてきた。本論文では、談話理解研究で設計された Allen のプランオペレータに関わる 4 種類の因果関係（“原因”，“効果”，“前提条件” および “手段”）を取り扱う。

知識獲得の基本的アプローチは、接続助詞（「ため」や「ので」）などに代表される接続標識を手掛かりとすることにある。ただし、どの接続標識を含む言語表現であっても、常に必然性の高い因果関係を表現する保証はない。また、同一の接続標識を含む言語表現から、上述した因果関係のうち、2 種類以上の因果関係知識が獲得できる可能性がある。そのため、接続標識以外の情報を利用することで獲得可能な因果関係を同定し、適切な因果関係知識を抽出する枠組みを検討する必要がある。

本論文では特に接続標識「ため」を含む複文に注目し、(1) 対象（「ため」を含む複文）中に獲得すべき因果関係はどの程度現れるか、(2) 対象中に現れた獲得すべき因果関係をどの程度の精度で同定でき、どの程度の量の知識が獲得できるか、(3) 獲得された因果関係知識がどのような応用に貢献できるか、について議論している。

まず、(1) では、言語テストを用いることにより、「ため」を含む複文から獲得すべき因果関係知識の比率を調査した。新聞記事から得た約 1000 件の「ため」を含む複文を対象とし、おおよそ 9 割の事例が獲得すべき因果関係を含んでいることを確認した。(2) では、「ため」を含む複文から 4 種類の因果関係知識の自動獲得課題を取り扱った。機械学習手法に基づく因果関係同定器を構築し、約 1000 件の未知データを用いたテストの結果、“原因結果”，“前提条件”，“手段” の各関係について、80 %の再現率で 95 %以上の獲得精度を達成した。また、“効果” 関係については、30 %の再現率で 90 %の獲得精度を達成した。新聞記事一年分の文書集合に本手法を適用した場合、27,000 件を超える因果関係知識が獲得できる見積もりを得た。(3) では、因果関係知識を利用して、出来事の望ましさに関する語彙的知識をブートストラップ的に獲得する手法を提案し、因果関係知識の有用性を示した。

氏名	乾 孝司
----	------

(論文審査結果の要旨)

平成16年1月20日に開催した公聴会の結果を参考に平成16年2月19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。

乾 孝司は、本博士論文において、大規模な言語データから因果関係に関する知識を自動獲得する研究を行い、そのための有用な手法を提案するとともに、その評価および利用について論じている。本研究の貢献は次のようにまとめることができる。

1. 言語データの中で「～ため」のような因果関係を表しうる接続表現を対象に、どのような種類の因果関係知識が抽出可能かどうかを調査し、その自動分類法を提案したこと。
2. 上記分類を自動化するために必要な文の意志性を判定する手法を提案し、高い精度で意志性の判定が可能であることを示したこと。
3. 上記手法を統合し、因果関係の下位分類である「原因」「前提」「手段」「効果」のそれぞれについて、自動獲得による再現率と精度を詳細に調査し、90%を超える高い精度で大量の因果関係知識を自動獲得することを可能であることを示したこと。
4. 抽出された因果関係知識を利用して、出来事の望ましさに関する語彙知識をブートストラップ的に獲得する手法を提案し、因果関係知識の有用性を示したこと。

因果関係知識の自動獲得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。