

電子・情報工学専攻	学籍番号	031033	指導 教員	増山 繁
申請者 氏名	坂地 泰紀			石田 好輝

論文要旨 (博士)

論文題目	日本語文書からの意味的表現の抽出に関する研究
------	------------------------

(要旨 1,200 字程度)

文書内には、様々な実アプリケーションに役立つ多様な表現があり、それを抽出するための技術としてテキストマイニングが注目されている。テキストマイニングとは、テキストを対象としたデータマイニングである。テキストマイニングは、顧客からのアンケートの回答やコールセンターに寄せられる質問や意見を解析する技術として既に実用化されている。従来、テキストマイニング技術による獲得対象は、語がほとんどであったが、近年は語より情報を含んでいる表現という形で獲得する場合がある。例えば、「コスト」という名詞だけを獲得しても、後続する文字列が「を抑える」と「が上昇する」では意味が異なるため、このような場合には複数の語から構成される表現を獲得する必要がある。本論文では、日本語文書から特定の意味を持つ表現を自動的に抽出する手法を開発し、その有効性を検証する。獲得する表現の意味を考慮することで、多くのアプリケーションや解析に役立てることができる。

特に、因果関係は質問応答システムや因果ネットワークの構築に用いることができるため、最近、注目されている。例えば、因果ネットワークによって、ある主題に関する様々な因果関係を系統的に把握することができる。しかしながら、因果関係を人手で抽出するためには、非常に高いコストと時間がかかる。そこで、本論文では、因果関係を自動的に抽出する手法を提案する。本手法では、構文情報と因果関係を獲得するための手がかりとなる表現(手がかり表現)を用いて因果関係を抽出する。しかしながら、手がかり表現には、因果関係以外の意味を持つものがあるため、このような場合は、抽出対象から除外する必要がある。そこで、この問題に対応するために、半教師あり学習を用いたフィルタリング手法、すなわち、文が因果関係を含むか否かを判定する手法を開発した。また、本論文では、特許文書を対象としたテキストマイニングも行う。特許出願件数は年間 40 万件近くにもものぼり、一文書あたりの文章量も膨大であるため、出願動向調査に有用なパテントマップ(特許出願動向を可視化したもの)を手作業で作成するには多大な時間とコストを要するため、その作成に役立つ情報を自動的に抽出する技術が求められている。そこで、本研究ではパテントマップの作成に役立つ「直接的なユーザの便益に相当する表現」と「技術上の解決課題を示す表現」を自動的に抽出する手法を提案する。さらに、本研究では、新聞記事から景気動向を示す「根拠となる表現」を統計的手法を用いて自動的に抽出する手法を提案する。抽出された景気動向を示す「根拠となる表現」を景気が回復することを示す Positive 表現と、悪化することを示す Negative 表現に分類する手法も併せて提案する。