

論文内容の要旨

博士論文題目 Opinion Mining from Web documents : Extraction and Structurization
(Web 文書を対象とした意見情報の抽出と構造化)

氏名 小林のぞみ

(論文内容の要旨)

CGM(Consumer Generated Media) やブログジャーナリズムのような新語の出現に象徴されるように、個人がネット上に発信する意見情報の重要性が広く認知されるようになった。これに伴い、個人の意見を Web 上のテキストから収集し、分析する技術に関する研究が活発になっている。本研究の目標は、Web 文書からこのような消費者の意見のある構造で抽出する技術を開発することである。本論文は、この意見抽出のタスクに対して、意見をどう構造化するか、および、構造化された意見をどのように抽出するか、という2つの問題に取り組んでいる。

一つ目の問題に対し、我々はコーパス分析に基づいて意見を4つの要素、Opinion holder, Subject, Aspect, Evaluation から成るユニットとして定義する。二つ目の問題に対して、意見抽出のタスクを (a)subject/aspect と evaluation の関係を抽出する問題、(b)aspect 間の階層関係を抽出する問題、の二つの部分問題に分けて取り組む。本論文では、まず aspect と evaluation の候補となる表現を半自動的に収集する枠組みを提案する。人手による表現の収集と提案する半自動収集手法を用いた場合とで比較実験を行い、人手よりも効率的に表現が収集できることを確認した。次に aspect と evaluation の関係を作成した辞書を使用して抽出する手法を提案する。これまでの意見抽出タスクでは、簡単なパターンベースの手法しか用いられていなかったが、aspect の表現が省略される場合も多く、パタンのみでは不十分だという問題があった。与えられた evaluation に対してその aspect を当てる問題は、日本語における照応解析のタスクと似ていることを指摘し、照応解析で用いられている機械学習ベースの手法を aspect-evaluation の関係抽出タスクに適用する。レビュー記事を対象に実験を行ったところ、パターンベースの手法と比較して、精度を20%以上向上させることができた。最後に、aspect に該当する表現が evaluation の表現よりもドメイン依存性が強く、網羅性の高い辞書を仮定することは容易ではないことから、aspect の辞書を仮定せずに aspect-evaluation の関係を抽出する手法を提案する。提案する手法は、関係が出現する統語的なパタンの手がかりと出現する文脈に関わらない統計的な情報を組み合わせるものである。Blog 記事を対象に実験を行い、統語的なパタンを使用したモデルが、ベースラインモデルと比較して10%近く精度と再現率を改善できることを示した。また、この手法を aspect 間関係抽出にも導入することで従来の統計ベースの手法と比較して精度、再現率を共に改善できることを示した。

氏名	小林のぞみ
----	-------

(論文審査結果の要旨)

平成18年12月28日に開催した公聴会の結果を参考に平成19年2月13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。

小林のぞみは、本博士論文において、Weblogなどに記述された人々の意見を抽出するために、意見の構造化に関する提案と意見抽出の自動化に関する新たな手法を提案した。本研究の貢献は次のようにまとめることができる。

- 意見に関するコーパスの分析により、意見情報を、評価者、評価対象、属性/部分(アスペクト)、評価表現の4つの要素からなるユニットとして表現することを提案したこと。
- 分野に依存した大規模コーパスから評価表現の候補を半自動抽出する方法を提案し、人手による抽出に比べてはるかに効率的な評価表現抽出が可能であることを示したこと。
- 意見抽出のタスクを、アスペクトと評価表現の対の抽出、および、アスペクト間の階層関係を抽出する2つの問題に分離し、それぞれの問題を解くための方法を機械学習に基づく手法として提案し、従来手法に比べてはるかに精度の高い意見情報抽出が可能であることを示したこと。
- 複数の分野のコーパスを用いて、提案手法が分野依存性が低い汎用の手法であることを示したこと。

このような問題の形式化と一連の解析手法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、自然言語処理および意見情報抽出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。