

審査の結果の要旨

氏 名 金山 博

本論文は、「From Opinion to Knowledge – Extraction of Sentiments and Demands (意見から知識への転換 – 評価と要望の抽出 –)」と題し、6章より構成される。製品・サービス・企業などに対して寄せられるユーザの意見は貴重な情報源であり、最新の評価や要望をすばやく大量に分析し意志決定に結び付けることで、戦略的な製品開発や販売が展開できる。このような目的に基づき本論文では、自然言語文で表現されたユーザの意見を自動的に解析し、活用するための言語処理手法について論じている。

第1章は「Introduction」と題し、研究の目的を明らかにしている。まず、評価や要望の分析における課題として、中核となる意味構造の定義、自然言語文の意味構造を同定するための言語解析手法の開発、分野固有の知識や語彙への対応の3点をあげている。また、これらの課題への取組みにおいて本研究では、コンポーネント化による汎用的な言語解析基盤の構築、統計的手法に加えて人手による知識の編集を可能にする可読性の高い知識表現の提案、実社会への応用に直ちに結びつく評価手法の確立、の3点を目標としたことを述べている。

第2章は「Background」と題し、評判分析の従来手法を総説するとともに、論文中で要素技術として用いる構文解析、教師なし学習、構造化されていない情報を管理するためのアーキテクチャであるUIMA (Unstructured Information Management Architecture) 上でのテキストマイニングについて説明している。

第3章は「Clause-level Evaluation Detection using a Tree Transfer Model」と題し、節単位での言語解析に基づく評価抽出手法について論じている。まず、評価の意味構造を表現するための評価フレームと呼ぶ知識表現を設計し、次に自然言語文から評価フレームへの変換に機械翻訳手法を適用することを提案している。これにより、動詞の格フレーム解析や語義のあいまい性解消など機械翻訳手法の従来研究成果を低コストで導入することが可能になる。実験の結果、表層的な単語マッチングを用いる場合と比較して、機械翻訳手法を用いる場合には大幅な抽出精度の改善が得られることを示している。

第4章は「Unsupervised Lexicon Induction for Domain-oriented Sentiment Analysis」と題し、評価抽出のための辞書を教師なし学習により獲得する手法を論じている。評判分析では、評価対象に対するユーザの態度が肯定的であるか否定的であるかの極性の判断が重要であるが、分析に必要となる語彙やそれに対する極性は評価対象の領域に強く依存する。そこで、本研究による提案手法では、特定の領域のコーパスの中から評価辞書に登録される語彙候補を自動抽出し、統計的検定による適切な候補を選択する。実験を通して、複数の文章にまたがる大域的な極性の一致度を用いることで辞書の被覆率を高めるとともに、閾値の設定に統計検定を用いることで人手によるパラメタ調整を介さず高い抽出精度が達成できることを示している。

第5章は「Demand Detection and Pattern Induction」と題し、ユーザが製品やサービスに求めている要望を抽出する問題に取り組んでいる。従来の評判分析が極性の判定に焦点をあてていたのに対して、多くのユーザから寄せられる意見文の中から具体的な要望対象を抽出するという課題に新たに取り組むものである。コーパス全体を分析することにより、要望を表す体言句を抽出するための構文パターンを教師なし学習により獲得する手法を提案し、高い抽出精度が得られることを示している。

第6章は「Conclusion and Future Work」と題し、第3章から第5章の研究を総括して研究の貢献を明確にするとともに、意見分析に関する今後の研究の方向性について論じている。

以上を要するに、本研究では、自然言語文で表現されたユーザの評価や要望の抽出法を論じ、新規性が高い手法や枠組み新たに提案して、有効性を示している。ユーザの意見を知識として活用するための実用性にも優れた基盤技術を提示した点で、当該分野への貢献も大きい。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。