

Twitter からの消費者ニーズの抽出手法に関する研究

川島 崇秀

近年、著しく普及してきたソーシャルメディア上には、製品やサービスに関する口コミが日々大量に投稿されている。これらの口コミ情報は、企業のマーケティング活動等に有益であり、膨大な書き込みの中から企業にとって価値のある口コミを抽出する技術が求められている。ソーシャルメディアの中でも Twitter は近年顕著な普及を見せており、リアルタイム性の高さ・ユーザの多様性・投稿量の多さといった視点から分析対象として大きな注目を集めている。

Twitter を対象とした口コミ分析に関する代表的な研究としてセンチメント分析が知られている。センチメント分析とは、Twitter 上のレビューを肯定と否定の2つのカテゴリに分類することである。しかし、この分類では、投稿を感情と いう視点で肯定的か否定的かの2値に区分する為、消費者の顕在的なニーズを表す要望を含む投稿を、直接的に抽出することはできない。また、Twitter を対象とした従来の辞書ベースの要望抽出手法では、Twitter 上に投稿される多様な文章表現への対応が難しく、十分な分類性能を達成することができなかった。

本研究では、Twitter のような短文かつ省略の多い文章を対象に、ビジネス応用が期待される要望抽出を高精度に実現する手法を提案する。提案法の特徴は、これまで情報抽出の分野で有効性が確認されている教師あり機械学習のアルゴリズム SVM(Support Vector Machine)を要望抽出に適用する点にある。教師あり学習で必要となる教師データの収集は、人手による収集ルールの生成と機械学習を融合した半教師あり学習手法 Distant Supervision を適用することで低コストな学習データの収集を試みる。

評価実験では、ソーシャルゲームに関する投稿を対象とし、提案手法と従来法の分類精度の比較を行うことによって、本手法の有効性を検証した。教師データの収集時に使用したゲームタイトル名を使用する場合と、使用しない場合の2通りの方法でソーシャルゲームに関する投稿を各 1000 件ずつ収集し、人手でラベル付けを行ったものを評価用データとして用意した。評価データに対して、構築した分類器を用いて分類を行った結果、いずれの方法で収集した評価用データに対しても、適合率、再現率、F 値において提案手法が高い評価を示し、本手法の有効性が確認された。

(指導教員 佐藤 哲司)