

意見極性の系列に基づくレビュー記事の評価値推定に関する研究

堀内 雅人

近年、レビューサイトやSNSなどで個人が発信する、特定の商品やサービスに関する評判情報が増加し続けている。評判情報からの情報獲得において、情報のすべてを閲覧することは時間的負担のかかる作業である。一方で、一部の評判情報のみを閲覧した場合には、商品に対して偏った印象のみを獲得することになる。商品のレビューには、レビュアーによる評価値が付与されていることがある。評価値は、ユーザが商品への評価を概観したり、閲覧するレビューを選択する際に役立つ。しかし、レビュアーによる評価値の付与基準は曖昧であり、ユーザが混乱することも少なくない。

本研究は、レビュー記事の内容から評価値を推定することにより、ユーザのレビュー閲覧を支援することを目的とする。本研究では、レビュー記事を文単位に分割し、文ごとに評価極性を付与し、評価極性の系列からレビュー対象商品の評価値を推定する手法を提案する。

本研究で提案する評価値の付与手法の手順は、評価表現辞書の類義語による拡張、拡張した辞書による評価極性の付与、付与した極性による系列を用いた評価値の推定の順である。評価表現辞書の類義語による拡張では、日本語 WordNet を用いて評価表現辞書の表現数を拡張した。拡張した辞書による評価極性の付与では、文ごとに評価極性を付与し、辞書拡張前後での評価極性の付与数、付与精度などを検証した。系列を用いた評価値の推定では、評価極性の系列により重みが増加する推定式を導入し、レビューの評価値推定を行った。

提案法の実装評価にあたっては、実運用されているレビューサイトのレビューを用いた。拡張した辞書による評価極性の付与では、名詞編で約 15 倍、用言編で約 30 倍の表現に極性を与えることができた。また、拡張後の辞書を用いて文への極性付与を行い、実際のレビュー記事に対して、拡張前の約 3 倍の文に極性を付与でき、なおかつ極性付与の精度が 7.3 % 向上した。系列を考慮した評価値推定では、系列を考慮しない場合と比較して、平均二乗偏差で 0.2~0.43、近傍一致率で 3~11 % 推定精度が向上した。カテゴリ、商品ごとの比較においては、類似したカテゴリや商品では、類似していないカテゴリや商品と比べて、推定式のパラメータの差が小さかった。これは、人の主観が入りにくいカテゴリと、入りやすいカテゴリで、極性系列の傾向が異なっていることが要因と考えられる。

(指導教員 佐藤 哲司)