

商品レビューを対象とした有用性の定義と判別

佐々木 優衣

近年、Web を利用して情報収集を行うだけでなく、自ら情報を発信するユーザが増加している。その一例として、商品を対象としたレビューが挙げられる。レビューは、ユーザが商品の購入やサービスの選択を検討する際の重要な情報源となっている。しかし、Web 上に存在するレビューの数は膨大であり、その全てを閲覧し、有用な情報を判断した上で利用することは困難である。また、評価の根拠に乏しいレビューや、業者によるレビューなどがあることから、有用性の高いレビューばかりではないという問題がある。

本研究では、機械学習を用いた有用なレビューの判別により、オンラインショッピングにおけるユーザの購入の意思決定を支援することに取り組む。まず、「炊飯器」「掃除機」「洗濯機」「布団乾燥機」「パソコン」のカテゴリの中からレビューを抽出し、被験者 2 名による有用性の判定を行うと同時に、被験者間の判定の一貫性を調査した。その結果、「レビュー中に評価の根拠がある」「レビューの投稿者が商品の購入者である」「レビューの文章がしっかりしている」という特徴を満たすレビューが有用なレビューであると判定されることがわかった。また、 κ 値は 0.757 であり、被験者間で判定方針が安定していることを示すことができた。したがって、本研究では、有用なレビューを「評価の根拠や評価対象となる商品について詳しく述べられており、文章がしっかりとしている」レビューであると定義した。

次に、実験データを用いて、有用性判別システムの実装を行った。実験データは、「炊飯器」と「掃除機」のカテゴリの中からレビューを抽出し、有用なレビューの定義に基づいて人手で判定したレビュー 1,000 件を用いた。本研究では、レビューが有用か否かを、Support Vector Machine(SVM) を用いて判別する。各レビューの内容を定量的に判別の素性として分類器に与えるために、レビュー中に出現する全形態素を個々の素性ベクトルとして抽出し、「形態素 ID : 出現回数」の形で分類器に学習させた。

分類器を 10 分割交差検定によって評価したところ、「助詞」を用いた場合に F 値が 0.86 で最大となった。「形容詞」のみを素性として与えた場合には、F 値が 0.04 と著しく低くなった。しかし、「形容詞」だけでなく、「助詞」「名詞」「動詞」と組み合わせて用いることにより、F 値が最大で 0.83 となり、精度が向上することがわかった。「助詞」が効果的に用いられている文章は、文中の格関係が同定されやすく、明瞭な文章である。したがって、商品や商品の機能についての評価を表す「形容詞」とともに、格関係を明確にする「助詞」、評価の対象となる商品名や機能を表す「名詞」や、商品に関する動作を表す「動詞」を素性として用いることによって、分類器の性能が向上すると考えられる。この結果は、本研究で定義した有用なレビューの特徴である、レビューの「根拠」や「詳しさ」および「明瞭さ」を支持するものである。

今後の課題として、形態素的特徴のみだけでなく、レビューの対象となる商品説明文などから判別の素性を得ることによって、分類器の精度を向上させることが挙げられる。

(指導教員 関 洋平)