

楽天市場の商品レビューを利用した 評価軸と評価表現辞書の同時構築手法

沼尾研究室 金兵裕太



～概要～

楽天市場やamazonといったショッピングサイトでは、一つ一つの商品に対して購入者のレビューが寄せられています。しかし商品レビューが大量にある場合、全てに目を通すことは困難です。また、レビューには評価の点数が付与されているケースが多いですが、これはユーザが独自に判断した評価であるため、レビュー内容を正しく反映しているとは限りません。

そこで本研究では、商品レビューの情報を集約することで、ユーザの商品選択を支援するシステムの構築を目的としています。



図1 提案システムを用いた商品比較のイメージ

～システムの実装方法～

提案システムでは、大量のレビューを「何について」「どのような」評価であるかに応じて分類します。これを実現するために必要となるのが評価表現辞書と評価軸です。

評価表現辞書とは、「良い」や「悪い」、「安い」といった対象を評価する語句（評価表現）を、肯定 or 否定の情報と共に登録した語彙集です。

評価軸は対象の評価指標を表したものです。例えば「掃除機」の場合、「価格」といった標準的なものの他に、「吸引力」や「音」に関する評価軸が存在すると考えられます。評価軸は、その軸を特徴付ける単語（特徴語）の集合として表現します。

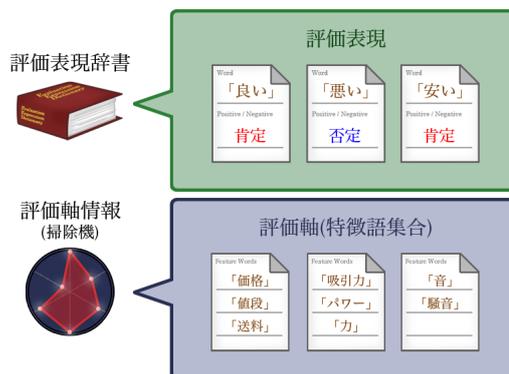


図2 評価表現辞書と評価軸

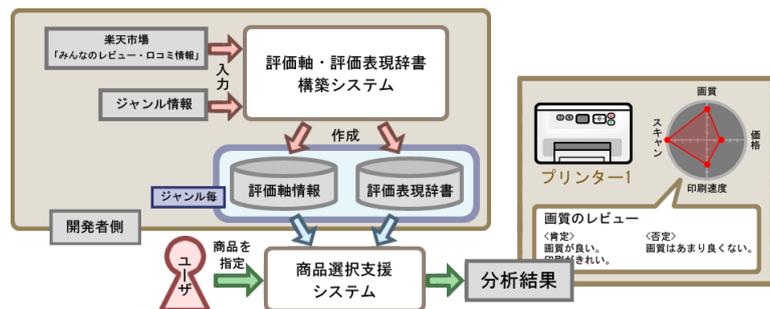


図3 システム全体の概要

評価表現辞書と評価軸情報は、商品レビューをコーパスとして学習を行うことで構築していきます。入力して与えるのは「プリンター」や「掃除機」といったジャンル毎のレビューです。これによって、各ジャンルに応じた評価表現辞書と評価軸を獲得することができます。

提案システムは学習によって構築したデータベースを利用して、ユーザに情報を提供します。

～評価軸と評価表現辞書の構築～

学習は「良い」「悪い」といった種表現をもとに行います。

評価表現辞書の構築には那須川らの手法[1]を用います。「安くて良い」といった表現から、「安い」は肯定表現であると推測します。

評価軸の構築では、特徴語の「抽出」と「分類」を行う必要があります。「特徴語は文章中で評価表現と共に出現する頻度が高い」という仮定をもとに、評価表現の周囲に現れる名詞を候補として収集します。その後、特徴語と評価表現の共起情報を用いたTF-IDFを特徴ベクトルとして、k-means クラスタリングによる分類を行います。この分類により意味的に類似した特徴語の集合、すなわち評価軸ができます。

～評価実験と考察～

楽天市場の商品レビューをコーパスとして評価表現辞書を構築し肯定・否定情報の適合率を計算すると、各ジャンルで85%以上の高い結果を得ました。また、楽天トラベルのレビューを用いた評価軸の構築を行い、関連研究[2]をもとに作成した正答データとの適合率を計算すると、半数以上の評価項目で適合率が80%を超える結果となりました。

以上の結果から評価軸と評価表現辞書の構築が一定以上の精度で行えている事が確認できました。今後はさらにレビュー内容を反映した評価軸の構築や、提案システムの実装が課題となります。

[1] 那須川哲哉, 金山博, “文脈一貫性を利用した極性付評価表現の語彙獲得”, 情報処理学会自然言語処理研究会(NL-162-16), pp. 109-116, 2004.

[2] ゲン ファム タン タオ, 岡部誠, 尾内紀夫, 林貴宏, 西岡悠平, 竹中孝真, 森正弥, “新たな弱教師付き型分類手法 Bautext”, 情報処理学会論文誌 Vol. 52 No. 1, pp. 269-283, 2011.