

不特定分野の商品レビューを対象とした評価情報の自動認識

河野 一志* 町田 翔* 村松 拓実* 小林 美穂* 延澤 志保*

*東京都市大学知識工学部情報科学科

1 はじめに

商品レビューは商品に対する感想や評価が記載されている文章であり、商品の購入を考えているユーザにとっては有用な情報を含む。しかし、商品レビューの数は膨大で、すべてを読むには時間と労力が必要になる。そこで、効率よく評価に関する情報を獲得するため、評価の要約や抽出をする研究が行なわれている [1]。評価の要約や抽出には、商品の機能や特徴が用いられるが、それらは商品の分野によって異なるため、あらかじめ商品の分野を指定する必要がある。

本稿では、品詞情報と係り受け情報を利用することで、商品レビューに含まれている評価情報を、分野を指定せずに自動認識する手法を提案する。

2 関連研究

小林らは、意見抽出のための評価表現の収集を行なっている [2]。小林らの研究では、意見のある対象もしくは対象のある側面に対する、記述者の主観的な評価を表す記述として、意見を<対象, 属性, 評価表現>の要素からなると定義している。対象とは、評価の対象となる商品名やサービスであり、属性は意見の焦点となっている対象の部分や性質、評価表現は対象もしくは属性の質や量に関しての値、または記述者の好悪を示す心的な態度を示している [2]。小林らは、<対象>の<属性>{が/は/も/に/を}<評価表現>といったパターンを複数用いて、ブートストラップに基づき評価表現と属性を相互に獲得することで、評価表現の収集を行なっている。しかし、初期辞書としてあらかじめいくつかの表現を手で与えているため、指定した分野にしか対応していない。

平山らは、評価表現の抽出に「形容詞」、「名詞、ナイ形容詞語幹」、「名詞、形容動詞語幹」、「動詞(係り先の次の語が形容詞の場合)」という品詞パターンを用いている [3]。この品詞パターンを用いることで、商品の分野を指定せずに評価表現の抽出が可能になるため、本稿ではこれらの品詞パターンを用いる。

3 提案手法

3.1 提案手法の概要

本稿では、不特定分野の商品レビューを対象とした評価情報を自動で認識して表示する手法を提案する。本手法では、評価情報を商品の対象、属性、評価表現の組からなる [2] と定義する。例として「このリンゴの味はとても美味しい。」という文では、対象は「リンゴ」、属性は「味」、評価表現は「美味しい」とし、これらの組を評価情報としてコンピュータが自動で認識する。

3.2 提案システムの流れ

本稿で提案する評価情報を自動認識するシステムの流れを図1に示す。入力の商品レビュー文章で、出力は人が商品の評価を一目で判断できるよう、商品レビュー文章の評価情報に色を付けハイライト表示したものとします。

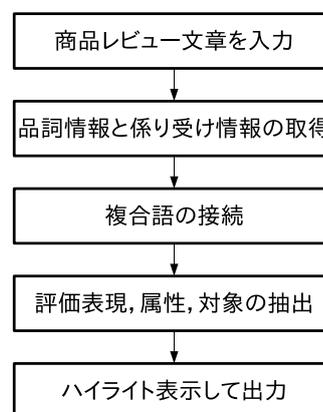


図1: 評価情報を自動認識するシステムの流れ

3.3 品詞情報と係り受け情報の取得

入力した文章から評価情報抽出に利用する品詞情報を取得するため、形態素解析ツール MeCab¹ を用いて形態素解析を行ない、形態素の品詞情報を取得する。また、係り受け解析ツール CaboCha² を用いて係り受け解析を行ない、係り元や係り先といった係り受け情報も取得する。係り受け解析を行なうために入力文章は句点と改行で区切り、文単位で処理を行なう。

¹形態素解析ツール MeCab, <http://taku910.github.io/mecab/>

²係り受け解析ツール CaboCha, <https://taku910.github.io/cabocha/>

Automatic Recognition of Evaluation Information for Product Reviews in Unspecified Domains.

Kazushi Kouno*, Sho Machida*, Takumi Muramatsu*, Miho Kobayashi*, and Shiho H. Nobesawa*.

* Faculty of Knowledge Engineering, Tokyo City University

3.4 複合語の接続

本手法で抽出する評価表現や属性に「使いやすい」、「大きさ」、「問題ない」、「よかった」、といった表現が出現する場合があるが、複数の形態素に分かれているため、本研究ではこれらを複合語として1つに接続する処理を行なう。本手法で扱う複合語は、評価表現と属性に関するものを接続して扱うため、同じ文節内に出現する形態素を以下の品詞のパターンで接続する。

- 動詞 + 形容詞 形容詞
- * + 名詞, 接尾 名詞
- * + 形容詞, 接尾 形容詞
- * + 動詞, 接尾 動詞
- 接頭詞, 名詞接続 + 名詞 名詞
- 名詞, ナイ形容詞語幹 + {ない} 名詞
- * + 助動詞{た} *

3.5 評価表現の抽出

評価表現を抽出するため、表1の品詞パターン [3] に当てはまる品詞の語を評価表現として抽出する。また、

表 1: 評価表現抽出のための品詞パターンと例 [3]

品詞	例
形容詞	良い, 安い, 美味しい
名詞, ナイ形容詞語幹	問題, 仕方, とんでも
名詞, 形容動詞語幹	静か, 強力, 面倒

品詞パターンだけでは抽出できない評価表現を取得するため、評価値表現辞書¹に記載されている語句も評価表現として扱う。これにより、副詞や名詞で評価表現に成りうる語を評価表現として扱うことができる。例として、名詞の「最高」や副詞の「まあまあ」などが挙げられる。

3.6 属性、対象の抽出

対象と属性の抽出を行なう。対象の品詞は名詞、属性の品詞は名詞または動詞とする。対象と属性は共起パターン [2] から係り受けを考慮した以下のパターンで抽出する。

- <評価表現> <対象>
- <評価表現> <属性>
- <属性> <評価表現>
- <対象> <評価表現>
- <対象>{の} <属性>

矢印は係り受けを表しており、係り元 係り先という関係を示す。

¹ 評価値表現辞書 http://www.syncha.org/evaluative_expressions.html

4 実験

実験対象は、Yahoo!ショッピング²の掃除機の商品レビュー文章 11 件 (総単語数 1,020 個) と、豚の角煮の商品レビュー文章 35 件 (総単語数 1,427 個) である。この2つの商品レビュー文章を入力とする。評価は、本手法で抽出した評価情報と、評価情報を人手で作成した正解データを比較し、本手法で抽出できた評価情報に対する適合率、再現率、F 値を求める。

掃除機と豚の角煮の商品レビュー文章から評価情報を認識してハイライトした結果の一部を図2に示す。また、評価の結果を表2に示す。適合率から、評価情報を80%以上認識できたと分かり、評価情報を自動認識できたことが分かる。

味は若干濃いめ。思ったより量が少なく少しかかり。半額クーポンがついてたのでそれでまあ妥当な金額かと思いましたが。脂身は同時に購入した他ショップより少なく、身は固めでした。もう少し脂身があってもよいかと。3個買いましたが、おまけをつけていただきました。おまけはまだよく見ていませんが、うれしく思いました。やっぱり100gでこの定価は高いかと。

図 2: 評価情報をハイライトした結果の一部

表 2: 評価情報の適合率、再現率、F 値

	適合率	再現率	F 値
掃除機	93.7%	79.1%	0.858
豚の角煮	82.9%	80.8%	0.818

5 まとめ

本研究では、商品の分野によって異なる評価情報を、不特定分野の商品レビューから品詞情報と係り受け情報を利用して自動認識しハイライト表示を行なった。掃除機と豚の角煮の商品レビュー文章で評価実験を行なった結果、ともに適合率は82%以上、再現率は79%以上、F 値は0.8を超える結果となり、不特定分野の商品レビューから評価情報を自動認識できた。

参考文献

- [1] 乾孝司, 奥村学, “テキストを対象とした評価情報の分析に関する研究動向,” 自然言語処理, Vol.13, No.3, pp.201-241, 2006.
- [2] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “意見抽出のための評価表現の収集,” 自然言語処理, Vol.12, No.3, pp.203-222, 2005.
- [3] 平山拓央, 湯本高行, 新居学, 高橋豊, “属性評価モデルに基づく商品評価の抽出と提示,” DEIM Forum, F2-5, pp.1-7, 2011.