

Web 文書からの主観的・客観的評価表現の抽出

Extracting Subjective and Objective Evaluative Expressions from the Web

中川 哲治†
Tetsuji Nakagawa

川田 拓也†
Takuya Kawada

乾 健太郎†
Kentaro Inui

黒橋 禎夫††
Sadao Kurohashi

1 はじめに

Web 上に存在する膨大な情報の中から、意見や評価に関するどのような情報があるかを俯瞰し集約する技術が近年重要となってきた。意見や評価を述べた情報にも様々な種類があるが、従来研究では、例えば商品のレビュー記事を対象にしたような、主観的な表現により直接著者の意見が表明されている評価情報を対象としているものが多い。一方で、「ガン抑制作用がある」というような客観的な記述ではあるが、肯定的(または否定的)な評価を暗に含んでいる情報を扱うことも有用であると考えられる。そこで本稿では、このような主観的または客観的な評価表現を Web 文書から抽出し分類することを考慮し、評価表現のタイプ分類を検討するとともに、評価表現の抽出と評価極性の分類を試みた結果を報告する。

2 評価情報と評価タイプ

本稿では、文中に存在する評価を表す基本的な単位を評価情報と呼ぶことにし、文中で実際にその評価が表明されている箇所を評価表現と呼ぶことにする(本稿で扱う評価情報に関する詳細については [6] を参照)。評価表現の例としては、「死刑制度に賛成はしないのですが、それに代わる終身刑などの制度が 必要なんだろうと考えます」や「美肌効果もある カテキン」(下線が評価表現を表す)のようなものがある。

評価表現には、賛成の態度を述べたもの、物の特長を述べたもの、良い状態や出来事を述べたものなど、様々な表現が存在する。特に本研究では主観的な評価だけではなく客観的な評価も扱っているため、評価表現をいくつかの種類にタイプ分けすることが有用であると思われる。そこで、次のような評価表現の分類を考えることにする:

当為 提言や助言、対策を述べたもの。

例。“～すべきだ”、“～しましょう”、“～するのがよい”

要望 希望や要求を述べたもの。

例。“～してほしい”、“～を求める”、“～を望まない”

感情+ / 感情- 自分自身の気持ちを述べたもの。

例。“好き”、“うれしい”、“嫌い”、“悲しい”

批評+ / 批評- 賛成 / 反対・称賛 / 批判等の態度の表明や、出来事や動作の主観的な記述。

例。“素晴らしい”、“つまらない”、“同意する”、“納得できない”

メリット+ / メリット- 利点や欠点、特長や課題について述べたもの。

例。“～できる”、“効果がない”、“うるさい”

採否+ / 採否- 積極的に利用したり選択する姿勢を述べたもの。

例。“利用する”、“導入する”、“採用する”、“推進する”

出来事+ / 出来事- 良い / 悪い出来事や状態を述べたもの。

例。“壊れた”、“受賞した”

これらの評価タイプの中で、「感情+ / -」「批評+ / -」や「メリット+ / -」の一部は主に主観的な評価表現となり、また「メリット+ / -」の一部と「採否+ / -」「出来事+ / -」

表1 評価タイプの分布

評価タイプ	個数
当為	217
要望	47
感情(+ / -)	47(26/21)
批評(+ / -)	616(304/312)
メリット(+ / -)	644(439/205)
採否(+ / -)	287(194/93)
出来事(+ / -)	59(32/27)

は客観的な記述により表現されることが多い評価情報となる。評価表現によっては、どのタイプに分類するかが曖昧な場合も存在するが、このように評価表現のタイプ进行分类することにより、評価情報タグ付きコーパス作成の際に何を評価表現とするかの規準が明確になり、また評価表現を文書中から抽出して利用する場合に評価タイプで分けることで抽出結果を整理して提示しやすくなると思われる。これらの評価タイプが付与されたコーパスを Web から収集した文書により作成したところ、各評価タイプの出現数は表1のようになった。「批評+ / -」や「メリット+ / -」が多くの割合を占め、「要望」「感情+ / -」「出来事+ / -」の評価表現は少なかった。

3 評価表現の抽出と分類

2節で定義した評価タイプのうち、一般的な評価表現として理解しやすい「感情+ / -」「批評+ / -」「メリット+ / -」の評価情報を対象として、その評価表現を文書中から自動抽出する実験と、評価表現の極性(肯定か否定か)を分類する実験を行った。

評価表現の抽出手法としては、条件付き確率場(CRF)により文中の各形態素に評価表現の開始(B)、中間(I)、評価表現以外(O)を表すタグを付与する方法 [1] を用いた。これは固有表現抽出等の情報抽出でよく用いられる方法であるが、ここで抽出対象としている評価情報は文中の任意の箇所出現する可能性があるため、このような手法を用いることにした。評価表現の抽出を行う際に、評価を表すためによく使用される単語の情報は非常に有用であると考えられる。そこで、小林ら [3] により作成された辞書と、東山ら [4] により作成された辞書を合わせて使用することにした。小林らの辞書は「快適」などのような、主観的な評価表現とその極性(肯定、否定、中立)を収集した辞書であり、東山らの辞書は「がん」や「健康」などのような、客観的な記述に用いられるがそれ自身望ましくない、または望ましくないような極性を有する名詞を収集した辞書である。辞書中に登録されている評価表現の数は14,779個である。CRFで使用した素性としては、前後2つまでの形態素の出現形、原形、品詞大分類、品詞細分類、評価表現辞書中での極性、を使用した。形態素解析にはJUMANを使用した。

評価極性の判定手法としては、Support Vector Machine(SVM)により与えられた評価表現の極性を二値分類により求める方法を用いた。評価極性の判定の際にも、前述の評価表現辞書を使用した。SVMで使用した素性としては、評価表現中の全ての形態素の出現形、原形、品詞大分類、品詞細分類、辞書に登録された評価極性の、unigram, bigram, KNP

† 情報通信研究機構, National Institute of Information and Communications Technology

†† 京都大学, Kyoto University

で係り受け解析した場合の親子関係, である。また, 評価表現中の肯定極性を持つ語の数と否定極性を持つ語の数の差も素性に使用した。SVM のカーネル関数は線形カーネルを利用し, コストパラメータ C の値は 1 とした。

実験に使用したコーパスは, 28 個のトピックについての文集合 (各トピック 200 文で, 計 1,275 個の評価表現を含む) から構成されるため, トピックごとにコーパスを分割して 28 分割の交差検定を行った。なお, 評価極性分類の実験では, コーパス中の正しい評価表現を入力として与え, その評価極性を分類するようにした。実験の結果を表 2 に示す。評価表現抽出については, Recall(正解データ中の評価表現数に対する正しく抽出された評価表現数の割合), Precision(システムが出力した評価表現数に対する正しく抽出された評価表現数の割合), F-measure(Recall と Precision の調和平均) により評価を行った。なお, 正しく評価表現が抽出されたかどうかの判定は, 評価表現の末尾 (主辞) の文節が, システムの出力と正解データとで一致しているかどうかにより行った。このようにした理由は, 評価表現の開始位置は人間でも判断が揺れることが多く, 正解データ作成の際にも明確な規準は設けておらず, 実際上は評価表現の主辞を正しく同定することが重要であると考えたためである。

評価表現抽出の精度については, Precision と比較して Recall が非常に低い結果となった。ただし, 多量の Web 文書から俯瞰的に情報を得るような応用を考えた場合, 対象となる文書は多く, 人間が見ることができる情報は限られているため, Recall よりも Precision がより重要であると思われる。コーパスの一部について, 2 人の作業者にタグ付けをしてもらい, 片方を正解, 片方をシステム出力として評価したところ, Recall, Precision, F-measure はそれぞれ 0.59, 0.68, 0.63 であったため, まだ抽出精度には向上の余地があると思われる。

評価極性分類の精度については 0.74 であった。なお, 作業者の間の一致率は 0.91 であり, 単純に評価表現中の肯定的な単語と否定的な単語の数の多数決で分類した場合の精度は 0.71 であった。

次に, 評価極性の分類に失敗した原因を調べた。始めに SVM により学習されたモデルのパラメータを調べたところ, 評価表現中の肯定的な単語と否定的な単語の数の差の素性に対する重みが大きく, 評価表現辞書の役割が非常に大きかった。そこで, 辞書に起因する問題を中心として, 分類誤りの原因を以下のタイプに分類した:

- 辞書に適切な単語が登録されていたが, その文脈中における単語の用法や単語の組み合わせにより評価表現の極性が変化したもの (極性変化)
例. “〈否定〉- するものではありません”, “〈がん細胞〉- を〈攻撃〉- して〈死滅〉- させます”
- 評価極性を判定する上で重要な語句が辞書中に含まれていないために失敗したと思われるもの (辞書カバレッジ)
- 評価極性を判定するには文脈が不足しているもの (文脈不足)
例. “私でもできる”
- 辞書中に不適切な語が含まれていたために失敗したと思われるもの (辞書誤り)
- 辞書に適切な単語が登録されており, 単語の極性は変化していないが, 誤ってしまったもの (分類誤り)
- コーパスの正解タグの誤り, 形態素解析等の失敗 (データ誤り)
- 辞書による処理が困難な反語等による評価表現 (その他)
例. “なんでそれが採決で賛成になるんだ”

複数の原因が考えられる場合, 最も大きな原因と思われるものに分類することにして, 200 個の誤りを調べた。表 3 に結果を示す。この結果から, 誤りの原因としては辞書のカバレッジと極性変化によるものが多く約半数の誤りがこれらの原因

表 2 評価情報抽出・評価極性判定精度

評価表現抽出 Recall	0.18
評価表現抽出 Precision	0.51
評価表現抽出 F-measure	0.25
評価極性判定精度	0.74

表 3 評価極性判定の誤り分類

原因	個数	(割合)
極性変化	54	(27%)
辞書カバレッジ	48	(24%)
文脈不足	30	(15%)
辞書誤り	22	(11%)
分類誤り	22	(11%)
データ誤り	19	(10%)
その他	5	(3%)

により占められていた。辞書のカバレッジについては, 多量の生テキストから評価表現辞書を自動作成する研究が盛んに行われているが [2], そのような辞書の自動獲得がこの誤りを改善するのに有用であると思われる。極性変化については, 極性を持つ語の文脈による極性変化を扱う方法などが提案されているが [5], そのような極性の変化を考慮するモデルを用いるように改善していくことが考えられる。他に, 文脈不足の問題についてはコーパスのタグ付け仕様の見直しなどが, 辞書誤りや分類誤りについては辞書の見直しや素性の改良等の対策が考えられるため, 今後行っていきたいと考える。

4 結論

本稿では評価表現をタイプ別に分類し, 評価の種類をより明確にすることをまず試みた。また特定の評価タイプについて, 評価表現の抽出と評価極性の分類を試みるとともに, 評価極性の分類誤りの原因を調査した。今後は, 評価表現辞書の拡充と, 評価表現辞書の利用方法の改良を進めることにより, 評価情報抽出の精度向上を図っていきたいと考える。

参考文献

- [1] Breck, E., Choi, Y. and Cardie, C.: Identifying expressions of opinion in context, *Proceedings of IJCAI 2007*, pp. 2683–2688 (2007).
- [2] Kaji, N. and Kitsuregawa, M.: Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents, *Proceedings of EMNLP/CoNLL 2007*, pp. 1075–1083 (2007).
- [3] Kobayashi, N., Inui, K. and Matsumoto, Y.: Opinion Mining from Web Documents: Extraction and Structurization, *Journal of the Japanese Society for Artificial Intelligence*, Vol. 22, No. 2, pp. 227–238 (2007).
- [4] 東山昌彦, 乾健太郎, 松本裕治: 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会発表論文集 (2008).
- [5] 池田大介, 高村大也, Ratinov, L.-A., 奥村学: 単語極性反転モデルによる評価文分類, 情報処理学会研究報告 2007-NL-180 2007-NL-180, pp. 43–48 (2007).
- [6] 川田拓也, 中川哲治, 森井律子, 宮森恒, 赤峯享, 乾健太郎, 黒橋禎夫, 木俣豊: Web テキストにおける評価情報の整理・分類およびタグ付きコーパスの構築, 言語処理学会第 14 回年次大会発表論文集 (2008).