

Web 上の客観的記述からの評価情報抽出に関する技術的検討

中川 哲治*, 宮森 恒*, 赤峯 享*, 乾 健太郎*, 黒橋 禎夫*†

* 情報通信研究機構, † 京都大学

{tnaka, miya, akamine, inui}@nict.go.jp, kuro@i.kyoto-u.ac.jp

1 はじめに

Web 上の文書をはじめとしたテキストから、文書の書き手等が表明している評価情報を抽出し分類するような意見情報処理は応用の広いタスクであり、近年盛んに行われている [4]。従来の評価情報を扱った研究では、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、「発がん作用を促進する」というような客観的な記述は否定的な評価を暗に含んでいると解釈できるが、このような評価情報を抽出することも重要である。しかしながら、肯定的または否定的な事実の記述も評価の一種とみなして抽出しようとした場合、多様な評価表現を取り扱う必要があるため、評価情報の抽出が困難となる。本稿では、客観的な記述により表明された意見を含む広範な評価表現の抽出を行うことを目指して、Web 上の文書から作成された評価情報タグ付きコーパスを使用し、抽出方法の検討を行うとともにその課題を考察する。

2 客観的記述に含まれる評価情報

「私はパンが好きだ」「この制度には疑問があります」などの文では、主観的な表現により直接的に評価が記述されている。一方で、「この食品は発がん作用を促進する」や「買って三日後に壊れた」などの文は、主観的な表現は含まれていない客観的な記述であるが、否定的な評価情報を含んでいるとみなすことができる。このような、一般的に望ましい、または望ましくない状態が客観的な記述により表現された評価表現は *evaluative factual* [2] とも呼ばれる。評価を表す主観的な表現が明示的には含まれていないため取扱いが困難であり、意見情報処理を対象とした既存の研究でも主観的な表現を中心に意見情報を扱っているものが多い [3]。しかしながら意見情報を処理する上で、このような非明示的な評価も含めた多様な評価情報を処理することは重要であると思われる。そこで本稿では、主観的な表現により記述された評価情報だけではなく、客観的な表現により記述された評価情報も対象として、

評価情報の抽出手法を検討することにする。

上記のような多様な評価情報の抽出に関する研究を行うため、評価情報の付与されたコーパスを作成した [6]。このコーパスは、「バイオエタノール」や「年金問題」などの 20 個のトピックに対して、各トピックについて 200 文ずつ、計 4,000 文を Web 上の文書から収集して作成されたコーパスである。このコーパスには、主観的な記述か客観的な記述かに関わらず、何らかの評価が含まれていると解釈できる文に対して、その評価の保持者、評価の対象、評価を表明している文中の箇所、評価の極性 (肯定か否定か) の情報が付与されている。また各文に対して、その文が与えられたトピックに関連しているか、その文が評価情報を含んでいるかを表す情報が付与されている (詳細は [6] を参照)。以後、このコーパスを利用して評価情報の抽出方法を検討していく。

客観的な記述による評価表現がどの程度存在するかを調べるために、上記のコーパス中で評価が含まれる文について、その評価表現がどのような手がかりにより認識できるかという分類を、コーパスを作成したのと同じ作業によりタグ付けしてもらった。ここでは、次の 3 とおりに分類をおこなった:

1. 「～べきだ」「～たい」のようなモダリティにより判断できる評価表現
例。「説明に努めるべきだ」
2. 主観的な単語 (「うれしい」「素晴らしい」など) の存在により判断できる評価表現
例。「この法案には賛成です」
3. 客観的な単語だけからなる評価表現
例。「劇症肺炎を発症し、死亡例がでた」

分類の結果を表 1 に示す。コーパス中の約 2 割の文に評価表現が存在し、そのうち主観的語句を含む評価表現が半数強を占めているが、客観的な語句だけからなる評価表現も 3 割程度と多くの割合を占めていることが分かる。モダリティや主観的な単語については、よく用いられる表現の数が限られるために、規則や辞書を整備することにより評価表現の扱いが比較的容易に行えることが期待できる。しかしながら、客観的な語

表 1 コーパス中の評価情報の割合

評価有無	文数 (割合)	内訳	
		モダリティ	
評価有り	875 (22%)	主観的	143 (16%)
		客観的	470 (54%)
			262 (30%)
評価無し	3125 (78%)		

句だけから構成される評価情報の場合、評価の表明に用いられる単語が広範にわたるため、これらの評価情報を抽出することはより難しい問題になることが考えられる。

3 評価情報抽出方法の検討

今回使用したコーパスには、評価情報の保持者や評価対象等の様々な情報が含まれているが、評価情報を抽出する場合にはまず与えられた文中に評価表現が存在するかどうかを判定することが重要となる。そこで本節では、評価情報の有無を判定するというタスクを考える。

この評価情報の有無判定のタスクでは、下記のとおり定義される指標を用いて、評価情報抽出精度の評価を行うことにする：

- 再現率＝システムが正しく評価表現有りとして判定した文の数／コーパス中の評価表現が存在する文の数
- 精度＝システムが正しく評価表現有りとして判定した文の数／システムが評価表現有りとして判定した文の数
- F 値＝ $2 \times \text{再現率} \times \text{精度} / (\text{再現率} + \text{精度})$

なお、このコーパスの一部の文(2トピック 400 文)に対しては二人の作業員によりタグ付けが行われているため、二名の作業員のタグ付け結果がどの程度一致するかその割合を計算することができる。片方の作業員の結果を正解、もう片方の作業員の結果をシステム出力として考えた場合の再現率、精度、F 値は、それぞれ 0.74、0.79、0.76 であったが、この値は機械処理を行う場合における抽出精度の上限の目安として考えることができる。

3.1 評価表現辞書に基づく評価表現有無の判定

文中に評価表現が存在するかどうか、また評価表現が存在した場合にその評価極性が肯定であるか否定であるかを判断する場合に、その文中に含まれる「快適」や「不満」などの評価を表す単語が重要な手がかりとなる。また、「発がん作用を促進する」というような客観的な記述に評価が含まれるかどうかを判断する際には、好ましい、または好ましくない性質を一般的に持つ名詞(この場合は「がん」)の存在が、非常に役に立つと思われる。そこで、極性を持つ単語から構成される評価表現辞書を用いることにより、文中から評価表

現を抽出する際の課題を調べることにする。これまで様々な評価表現辞書が作成されているが、ここでは小林ら [1] により作成された辞書と、東山ら [5] により作成された辞書を使用した。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、5,474 語からなる。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いか、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。この辞書の各単語には確信度のスコアが付与されているため、ここでは一定の閾値以上のスコアを持った 3,312 語を使用した。

上記の辞書を使用して、辞書に含まれる単語が与えられた文中に出現していた場合に評価表現がその文中に存在すると判定するシステムを考える。そのようなシステムにより評価表現を抽出した場合の抽出精度を表 2 に示す。この表において、辞書(主観)は小林らの辞書を用いた場合、辞書(主観+客観)は小林らの辞書と東山らの辞書を両方用いた場合の結果を表す。辞書(主観)を用いた場合と比較して、辞書(主観+客観)を用いた場合再現率は向上しているが、精度は悪化しており、全体として F 値は下がっている。またどちらの場合も、作業員間一致率の F 値と比べるとまだ低い値である。

評価が存在するにもかかわらず評価が無いと判定された文の数を、その評価表現によりモダリティ、主観的、客観的に分類した結果を表 2 に示す。この結果を表 1 に示された分布と比較すると、評価が存在するにも関わらず抽出できなかった評価情報の内訳として、客観的な記述のみによる評価情報の割合が多くなっており、客観的な記述に含まれる評価が抽出誤りに占める割合が比較的大きいことが分かる。

3.2 誤りの分析

この節では、3.1 節の辞書(主観+客観)を使った実験において、評価が有るにも関わらず無いと誤って判定した場合と、評価が無いにも関わらず有ると誤って判定した場合について、どのような事例が存在するかを分析した結果を説明する。

3.2.1 評価が有るにも関わらず無いと判定した例

評価を含んでいるにも関わらず、評価が無いと判定してしまった文(再現率を低下させた事例)は、辞書のカバレッジ不足を起因とするものが多く、誤りの半数程度を占めていた。不足していた単語としては、次のようなものがあった: 「いらぬ」「福音」「愛用」「許すな」「進まず」「不健康」「つぶれない」。このような辞書のカバレッジが原因の誤りを減らすには、辞書の

量をさらに増やしていく必要がある。ただし、

(1) バイオ燃料としては、トウモロコシなどから作る〈バイオエタノール〉が[商用化されている](+)

(2) 〈フライドポテト〉は[ファーストフードの基本なんだ](+)

(上記の例では、コーパス中にタグ付けされた評価を表す箇所を [] で、評価の対象を 〈 〉 で、評価極性を () で示している。以下同じ) などの表現は、評価を表さない文脈でも多く用いられるものであるため、単純に「商品化」や「基本」を辞書に登録すると精度を下げる恐れがある。また、

(3) アメリカ合衆国では、政府が国を挙げて〈バイオエタノール〉の[使用を進めています](+)

(4) [泥沼の様相] 見せる〈電子政府商戦〉(-)

(5) 〈消費税増税〉が[年金制度を救う、頼みの綱](+)

(6) 〈電子政府〉の[足元が揺らいでいる](-)

のように、個々の単語の持つ一般的な望ましさの度合を個別に見るだけでは判断が難しく、複数の単語の組合せを考慮しなければならない場合もあった。他に、

(7) 放射線療法と違い、〈化学療法〉は[全身に作用を及ぼします](-)

というような文脈にも依存して評価が含まれている場合や、

(8) 〈電子政府計画〉で、主要な行政手続きの 8 割でネットからの[利用率が 1% 未満であることが判明](-)

などのように客観的記述の内容が解釈ができなければ評価の有無が判断できない場合などが存在した。

3.2.2 評価が無いにも関わらず有ると判定した例

評価を含んでいないにも関わらず、評価が有ると判定してしまった文(精度を低下させた事例)の場合も、辞書中に不適切な単語が含まれていたことが原因で失敗した例が多かった。しかしながら、評価を含むことがある単語が使用されている場合でも、

(9) 今日から抗がん剤 TS1 を服用する

(10) カルボコン、メルファランなどがアルキル化剤の抗がん剤であり、血液系のガンに対して使用される

(11) 捨て子を救済するために設置した『このとりのゆりかご』

(12) 諸問題の解決、改善策提案を行うことを目的として設立した

(上記の例では、評価表現辞書に含まれた単語を下線で示している。以下同じ) などのように、評価を表す表現として使われていない場合がしばしばあった。「がん」のようなそれ自身望ましくない極性を持つと

思われる語が存在していても、「がんを治す」「がんに効く」というような使われかたをしていない場合、評価を持たないことがある。そのため、極性を持つ語どのような文脈で評価を構成するかを扱うためのモデルを考えることが今後必要である。

他に見られた誤りとして

(13) 年金制度に未来はあるのか

のように、条件や仮定や疑問を表す節に出現しており、直接評価を表してはいない場合があった。また今回使用したコーパスでは、誰かが何かに対して評価を表明している場合、つまり評価の保持者と評価の対象が明らかかな場合に限り評価表現としてタグ付けしている。そのため、

(14) リンパ癌を克服

のような全体としては望ましい事態を表している表現を抽出しても、評価情報のタグは付与されていないため誤りとなる。評価の保持者や評価の対象が存在しないといけないという制約は問題を難しくしているため、今後この仕様についてさらに検討していくことも考えられる。

3.3 機械学習を用いた評価情報有無の判定

評価情報の有無を判定する方法の一つとして、機械学習を用いて文中に評価情報が存在するかないかを分類することができる [8]。ここでは Support Vector Machine (SVM) を用いて、これまで説明してきた evaluative factual を含む評価情報の有無を判定することを試みる。

使用した素性としては、形態素解析システム JUMAN[7] により得られた文中の全ての形態素の出現形、原形、品詞大分類、品詞細分類の unigram と bigram を使用した。さらに評価表現辞書を利用する場合には、辞書中の単語が n 種類以上 ($n = 1, \dots, 10$) 出現したことを表す素性と、極性を持つ単語に対してその極性と前後の単語の出現形や品詞大分類との bigram を用いた。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、20 個のトピックの文集合(各トピック 200 文)から構成されるため、19 個のトピックで学習を行い残り 1 個のトピックでテストを行うことを 20 回繰り返して実験を行う。

実験の結果を表 3 に示す。この表において、辞書無しは辞書を利用しなかった場合、辞書(主観)は、小林らによる辞書のみを用いた場合、辞書(主観+客観)は小林らによる辞書と東山らによる辞書を両方用いた場合の結果である。辞書(主観)を利用した場合の抽出精度が最も高く、F 値で 0.50 であったが、表 2 に示され

表2 辞書に基づく評価表現有無の判定結果

使用した辞書	抽出精度			抽出できなかった評価情報の内訳 (全体の割合)		
	再現率	精度	F 値	モダリティ	主観的	客観的
辞書 (主観)	0.69	0.45	0.53	44(16%)	124(45%)	110(40%)
辞書 (主観+客観)	0.88	0.32	0.45	19(18%)	45(42%)	42(40%)
作業者間一致率	0.74	0.79	0.76			

表3 SVMによる評価情報有無の判定結果

使用した辞書	抽出精度			抽出できなかった評価情報の内訳 (全体の割合)		
	再現率	精度	F 値	モダリティ	主観的	客観的
辞書無し	0.42	0.60	0.48	72(14%)	258(51%)	178(35%)
辞書 (主観)	0.45	0.62	0.50	62(13%)	247(51%)	173(36%)
辞書 (主観+客観)	0.43	0.60	0.48	70(14%)	253(51%)	177(35%)
作業者間一致率	0.74	0.79	0.76			

ている辞書 (主観) に基づいた手法による結果よりも低い値であった。SVM を用いた場合、辞書に基づく方法と比較して再現率が低く、精度が高くなる傾向があった。評価表現が存在するにも関わらず、評価表現を抽出できなかった文をその手がかりをもとに分類した結果を表3に示す。表1と比較して、抽出に失敗した評価表現は客観的な単語から構成されるものがやや多いことが分かる。3.1節の辞書に基づく方法とは異なり、SVMを用いたこの評価情報抽出手法では、評価情報の抽出に失敗した場合にその原因を特定するのが容易では無いため誤りの分析は行っていないが、辞書やコーパスの拡充や手法の改良による改善の余地はまだ多くあると思われる。

4 結論

本稿では、従来から広く研究の対象とされてきた主観的な表現からの評価情報抽出だけではなく、客観的な記述に含まれる評価情報に注目して、その抽出方法を検討した。コーパスを用いた分析や実験の結果、客観的な記述による評価情報はコーパス中に存在する全ての評価情報の中で無視できない割合を占めていることが分かった。また既存の評価表現辞書を用いて評価情報の有無の判定を試みた結果から、抽出に失敗した事例を分析した。

今後の課題として、評価情報辞書を拡充するとともに、評価表現が文脈等により影響を受ける場合に対処するための適切なモデルの検討をすることが挙げられる。さらに評価情報有無の判別だけではなく、文中で評価を表現している範囲の同定、評価情報の保持者の同定、評価の対象の同定、および評価極性の同定を、客観的な記述に含まれる評価情報に対して行っていきたいと考える。

参考文献

- [1] 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).
- [2] Nigam, K. and Hurst, M.: Towards a Robust Metric of Opinion, *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications* (2004).
- [3] Wiebe, J., Wilson, T., Bruce, R., Bell, M. and Martin, M.: Learning Subjective Language, *Computational Linguistics*, Vol. 30, No. 3, pp. 277–308 (2004).
- [4] 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).
- [5] 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).
- [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).
- [7] 黒橋禎夫, 河原大輔: 日本語形態素解析システム JUMAN version 6.0, 京都大学大学院情報学研究科 (2007).
- [8] 関洋平, Evans, D. K., Chen, H.-H., Ku, L.-W., 神門典子: 意見分析タスク — 多言語テキストを対象とした意見情報抽出技術の評価 —, *情報処理学会研究報告 2008-FI-89 2008-NL-183*, pp. 51–58 (2008).