

係り受け関係と感情誘発語を用いた感情生起表現の抽出手法の提案

伴昌彦[†], 内田理[‡], 菊池浩明[‡], 中西祥八郎[‡]

[†] 東海大学大学院 工学研究科, [‡] 東海大学 情報理工学部
e-mail: masa@shime.ds.u-tokai.ac.jp

1. はじめに

テキストに表れる感情表現には「彼の元気な顔を見た」のように“その事柄によって書き手は感情が引き起こされた”と読み手が推測できる表現がある。このような表現は人間の経験的なことに起因するため、コンピュータで処理することは難しい。このため、「彼の元気な顔を見た」のような表現の解析を行うためにはまず、表現そのものをテキストから適切に抽出することが必要不可欠である。

本稿では「元気な顔を見る」のような表現を「感情生起表現」、「元気」のような語を『感情誘発語』とし、この『感情誘発語』を用いてテキストから感情生起表現を抽出する手法を提案する。

2. 感情生起表現の抽出について

2.1 感情生起表現の定義

本稿では「感情生起表現」を“その表現の示す事柄から、書き手は感情を引き起こされたと読み手が推測できる表現”とする。例 2.1 の (1) は単に書き手が自分の感情を表している表現であるため、感情生起表現にはならない。一方、例 2.2 では「元気」や「残酷」という語から書き手は感情を引き起こされたと読み手が推測できる表現であるため、感情生起表現となる。

例 2.1 感情生起表現にはならない表現例

- (1) ～で、昨日は嬉しかったし、～
- (2) さっき弟の顔を見た。

例 2.2 感情生起表現になる表現例

- (1) ～だったため、彼の元気な顔を見たら～
- (2) どうやらあの人は残酷なことをされて～

2.2 感情誘発語の定義

例 2.1 の (2) 「弟の顔を見る」と例 2.2 の (2) 「元気な顔を見る」は表現としては類似しているが、例 2.1 の (2) は感情生起表現ではない。これは「弟」という語の語意には、感情の意味合いが含まれていないことから判断できる。一方、「元気」や「残酷」の語意には感情の意味合いが含まれているため、例 2.2 は感情生起表現と判断できる。このことから、『感情誘発語』とはその語意に感情の意味合いが含まれる語とする。

2.3 感情語データベースの作成

本稿では感情生起表現の候補表現の抽出や『感情誘発語』の収集に「感情語」を用いる。この「感情語」は感情表現辞典[1]と現代形容詞用法辞典[2]から収集する。

感情表現辞典には「喜ぶ〈喜〉」「悲しい〈哀〉」などの約 2,000 個の感情表現が「喜・怒・哀・怖・恥・好・厭・昂・安・驚」の 10 種に分けられ収録されている。

現代形容詞用法辞典には 1,010 語の形容詞の語義毎に「イメージ値」が付与されている。イメージ値とはこの辞典で「その形容詞に対して日本人が抱く共通の評価」とされており、+3～-3 の七段階に区分された値である。

この二つの辞典から著者らの主観により選別した語を「感情語」として収集する。但し、感情表現辞典から感情語を収集する際にはその語が分類されていた感情を付与させておく。つまり、感情表現辞典からは感情のラベルが付与された語が品詞に限らず収集され、現代形容詞用法辞典からは感情のラベルが付与されていない形容詞（この辞典において）が収集される。また、著者らの選別により「感情語」とはならない語は感情表現辞典では「透く〈喜〉」や「落下〈厭〉」、現代形容詞用法辞典では「水っぽい」や「丸まっちい」などのような語である。

このようにして収集した「感情語」を総じて「感情語データベース」とする。また、異なり数で感情表現辞典から 1,309 語、現代形容詞用法辞典から 613 語の計 1,922 語を感情語データベースに登録した。

3. 感情生起表現の候補表現の抽出

感情生起表現の候補となる表現（以下、候補表現）はテキストの文中にある「感情語 β 」を基に抽出する。処理としては、まず「感情語 β 」を含む一文を抽出する。次に抽出した一文から候補表現の末尾となる文節と先頭となる文節（以下、末尾文節、先頭文節）を決定する。そして、先頭文節と末尾文節の間の連続した文節列（先頭文節と末尾文節を含む）を候補表現として抽出する。「感情語 β 」とは「感情語データベース」で感情のラベルが付与された、品詞が「動詞、形容詞、形容動詞」の語である。また、本稿で対象とする「感情」は「喜・怒・哀・怖・厭・驚」の 6 種であるため、「感情語 β 」の感情もこの 6 種のみとする。

3.1 末尾文節の決定

候補表現の末尾文節になる文節は主辞形態素が「動詞、形容詞、サ変名詞」のいずれかであり、“主辞が感情語 β の文節”と次の条件 (1)～(3) に示すいずれかの係り受け関係にある文節とする。尚、“主辞が「動詞、形容詞、サ変名詞」の文節”を「文節_{pre}」、 “主辞が感情語 β の文節”を「文節_{emo}」と表現する。

- (1) 文節_{emo}を係り先とする文節_{pre}
- (2) 文節_{emo}の係り先が“主辞が「思う、感じる」などの文節”である時に、同じくこの文節を係り先とする文節_{pre}
- (3) 文節_{emo}（形容詞、形容動詞）が係り先の文節を連体修飾している時、同じくこの文節を係り先とする文節_{pre}

図1に各条件の例を示す。図1では網掛け部分が末尾文節になる。尚、構文解析にはCaboCha[12]を用いた。

例3.1 (1)の例

～だったが、趣味で **意気統合できた** ことは **嬉しかった**。〈喜〉
 文節_{pre} (動詞) 文節_{emo} (感情語β)

例3.2 (2)の例

意気のない **行動を** **残念に** 思う。〈厭〉
 文節_{pre} (サ変名詞) 文節_{emo} (感情語β)

例3.3 (3)の例

希望を **失った** **哀しい** 出来事であった。〈哀〉
 文節_{pre} (動詞) 文節_{emo} (感情語β)

図1. 末尾文節の決定例

図1の例3.1のように、文節_{pre}から文節_{emo}の間に「～こと」など“自立語を含まない文節”がある場合は、この文節の係り先の文節を調べる。また、条件に該当する文節_{pre}が複数個ある場合は、文末に一番近い文節_{pre}を末尾文節とする。

末尾文節の条件を満たす文節がなければ、候補表現の抽出は行わない。また、「感情語β+ない」など否定表現である時なども候補表現の抽出は行わない。

3.2 先頭文節の決定

先頭文節は、3.1節で決定した末尾文節から文頭の間にある接続詞や逆接表現など、“「節」の区切りを表すような表現を含む文節”の**次の文節**とする。以下に条件の一部を示す。

- (1) 主辞形態素が接続詞である文節
- (2) 「私は」「彼は」などの人称表現を含む文節
- (3) 「だが」「でも」など逆接表現を含む文節
- (4) 読点を含む文節
- (5) 主辞形態素が動詞であり、「助詞-接続助詞：～て」を含まなく、連体修飾節でもない文節
- (6) 末尾文節以降の文節を係り先とする文節

ここで、条件を満たす文節がなければ文頭の文節を先頭文節とする。

3.3 候補表現の抽出

3.1節、3.2節で決定した「末尾文節」と「先頭文節」の間の連続した文節列(先頭文節、末尾文節を含む)を感情生起表現の候補表現として抽出する。この時、末尾文節の次が“自立語を含まない文節”であれば、この文節も含めて候補表現として抽出する。例3.4では網掛け部分を候補表現として抽出する。

例3.4 候補表現の抽出例

～だったが、**趣味で意気統合できたこと**は嬉しかった。〈喜〉
 逆接表現 先頭文節 末尾文節 感情語

4. 感情誘発語辞書の構築

本稿では3章で抽出した候補表現が感情生起表現であるかの判定に『感情誘発語』を用いる。感情誘発語は2.2

節で“その語意に感情の意味合いが含まれる語”とした。このことから、感情誘発語は語の意味が示されている国語辞典を用いて収集する。国語辞典は三省堂現代国語辞典[3]、新辞林[4]、岩波国語辞典[5]、広辞苑[6]を用いる。一つの国語辞典を用いるよりも、複数の国語辞典を用いた方がより多くの感情誘発語が収集できるためである。

また、感情誘発語には「感情値」を付与する。感情値とは収集した感情誘発語の語意に感情の意味合いが含まれる度合いを示した値($0 < x(h) \leq 1$)である。

尚、感情値を付与した感情誘発語を総じて『感情誘発語辞書』とする。

4.1 候補語の抽出と感情値の付与

感情誘発語の収集のために「感情語データベース」の全ての語に「感情、心、気持ち」の3語を加えたものをseed語とする。このseed語は国語辞典から感情誘発語を収集する基になる語である。まず、seed語から感情誘発語の候補語(以下、候補語)を収集し、感情値($0 < x(h) \leq 1$)を付与する。この処理は4つの国語辞典毎に独立して行う。また、このseed語の感情値は1とする。

4.1.1 候補語の収集

候補語はseed語を定義文中に含む見出し語とする。例4.1では【趣味】の定義文中にseed語の「感興」が含まれているため、【趣味】を候補語として収集する。

例4.1 seed語を定義文中に含む見出し語の例

【趣味】**感興**をさそう状態。(広辞苑)

但し、見出し語がJICST科学技術用語シソーラス[8]から抽出した“科学技術用語”である場合は感情誘発語になる可能性が低いため、候補語として収集しない。

4.1.2 候補語への感情値の付与

収集した候補語(見出し語)に定義文中のseed語から感情値を付与する。この感情値は定義文中のseed語の見出し語への“関連度”とseed語の感情値から計算する。

国語辞典では、見出し語の意味解釈に特に必要な語は定義文の文末近くに集まり易くまた、定義文の文末の語は見出し語の上位概念であることが多い。つまり、定義文の文末に近い語ほど、見出し語への関連度は大きいと言える。このような考えを基に、定義文中のseed語の“見出し語への関連度”と“感情値”から候補語 h の語義 m^i 毎の感情値 x^m を求める計算式を式(1)に示す。また、感情値 x^m は最大で1とする。

$$x^m = \begin{cases} \max_j \sum_k x_j^i(s_k) \cdot y_j^i(s_k), & d_j^i(s_k, r_k) > 0 \\ \max_j \sum_k x_j^i(s_k), & d_j^i(s_k, r_k) = 0 \end{cases} \quad (1)$$

ここで、 $y_j^i(s_k)$ は次式で求める。

$$y_j^i(s_k) = \alpha \left(d_j^i(0, r_k) - 1 \right) \cdot \alpha \left(d_j^i(s_k, r_k) - 1 \right)$$

ここで、 $x_j^i(s_k)$ は語義 m_i の j 番目の定義文中の k 番目のseed語の感情値を表す。 $d_j^i(s_k, r_k)$ は (i, j) の中の係り受け木

から生成した seed 語 s_k を含む部分木の「根」 r_k から seed 語 s_k までの距離である。 r_k の値は seed 語 s_k が変わると変化する。部分木は図 2 のように、定義文の係り受け木の根（心持ち）の子にあたる節（「しようという」「積極的な」）の数だけ生成する（部分木 1, 2）。生成する部分木はこの節の先祖と子孫にあたる全ての節とする。部分木中の数字はその部分木における節の出現順番である。また、 $d_j^i(s_k, r_k)$ に関しては図 2 において、 s_1, s_2 がそれぞれ「何か」「積極的」とするとする。この時、 s_1, s_2 から生成される部分木はそれぞれ部分木 1 と部分木 2 になる。この時、部分木 1 における $d_1^1(s_1, r_1) = 4-1=3$ となり、部分木 2 における $d_1^1(s_2, r_2) = 2-1=1$ となる。この時、「～状態。」「～事柄。」など“定義文特有の文末表現を含む文節の節”がある場合はこの一つ前の節を r_k とする。また、“自立語を含まない文節”は出現順番に含めない。また、 α は定数 ($0 < \alpha < 1$) である。

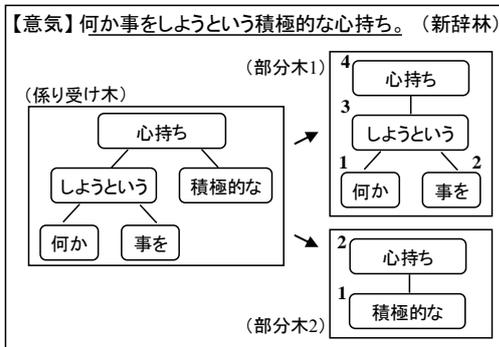


図 2. 部分木の生成例

さらに、ある語義の見出し語への関連度をその語義の定義順番によって求めることが出来る。国語辞典では見出し語が多義である場合、最初に定義されている語義ほどその見出し語の一般的な語義である。つまり、最初に定義されている語義ほど、見出し語への関連度は大きいと言える。このような考えを基に、ある語義の候補語への関連度 w^{m_i} を求める計算式を式(2)に示す。

$$w^{m_i} = \frac{(M_h - i) + 1}{\sum_i i} \quad (2)$$

ここで、 M_h は候補語 h の語義数、 m_i は候補語 h の i 番目の語義である。式(2)では最初に定義されている語義から関連度が線形的に減少する。また、語義の関連度 w^{m_i} の総和は 1 となる。

式(1)(2)から、候補語の感情値 $x(h)$ を求める計算式は(3)式となる。また、感情値 $x(h)$ は最大で 1 となる。

$$x(h) = \sum_i w^{m_i} \cdot x^{m_i} \quad (3)$$

4.2 候補語の再帰的収集

4.1 節で収集された候補語には『感情誘発語』になる語が多く含まれる。このため、4.1 節で収集された候補語を「新たな seed 語」として再び 4.1.1 節と 4.1.2 節の処理を行うことで「新たな候補語」を収集することができる。しかし、4.1 節で収集された候補語には『感情誘発語』にはならない語も含まれている。このため全ての語を「新

たな seed 語」にすると、『感情誘発語』にはならない「新たな候補語」が多く抽出される。これにより、4.1 節で収集された候補語から「新たな seed 語」となる語を選択する。そして、4.1.1 節と 4.1.2 節の処理を「新たな seed 語」で行い、「新たな候補語」を収集する。このような処理を「新たな候補語」が収集されなくなるまで再帰的に行うことで、より多くの感情誘発語の収集を行う。

4.2.1 新たな seed 語の選別

以下の条件を満たす候補語を「新たな seed 語」とする。

(1) 候補語の表記に漢字が含まれている場合

感情値が閾値 δ_1 以上であり、“候補語の表記に含まれる「漢字の感情値」の平均値 $\chi(h)$ ” が閾値 δ_2 以上の候補語

(2) 候補語の表記に漢字が含まれていない場合

感情値が閾値 δ_3 以上の候補語

「漢字の感情値」とは、4.1 節と同様の手法により漢語辞典[9][10]から求めた漢字一字の感情値のことである。これを用いて“候補語の表記に含まれる「漢字の感情値」の平均値 $\chi(h)$ ” を求める計算式を式 (4) に示す。

$$\chi(h) = \frac{\sum_i x(c_i)}{N_h} \quad (4)$$

ここで、 $x(c_i)$ は候補語の表記に含まれる漢字 c_i の感情値、 N_h は候補語の表記に含まれる漢字の字数である。例として「趣味」の漢字の感情値の平均値を求める。「趣味」は漢字の字数 $N_h=2$ であり、「趣」と「味」の感情値をそれぞれ $c_1=0.49$ 、 $c_2=0.43$ とすると、 $\chi(h)=0.46$ となる。

“候補語の表記に漢字が含まれていない場合”とは平仮名や片仮名のみで表記される候補語であり、このような候補語は上記のような判定ができないため、 δ_3 の閾値を δ_1 より高い値に設定する。

4.3 候補語から感情誘発語辞書への登録

次の条件(1)(2)に該当する候補語を感情誘発語に決定し、感情誘発語辞書に登録する。

(1) 複数の国語辞書から共通に収集された候補語

この場合、この感情誘発語の感情値は収集された辞書毎の感情値の平均値とする。

(2) 感情値が閾値 δ_3 以上であり、“候補語の表記に含まれる「漢字の感情値」の平均値”が閾値 δ_2 以上である候補語

4.4 感情誘発語の語彙数の拡大

4.3 節までは国語辞典のみを用いた感情誘発語の収集であったため、この手法だけでは抽出しきれない感情誘発語も多数ある。そこで、日本語語彙大系[10]（以下、語彙大系）から感情誘発語と同一の意味属性の語を収集することによって、感情誘発語の語彙数を拡大させる。

4.3 節までで収集した感情誘発語の中から“感情値が閾値 δ_1 以上である感情誘発語の表記”と“語彙大系の体言の意味属性名の表記”が同一である場合に、その意味属性の語（以下、類語）を全て感情誘発語辞書に登録す

る。類語の感情値 $x(h)$ は式(5)によって計算する。

$$x(h) = \frac{\sum_i^{L_h} x_{h,i}}{N_h} \quad (5)$$

ここで、 N_h は類語の意味属性数、 L_h は“類語が持つ意味属性名の表記”と“感情値が閾値 δ_1 以上の感情誘発語の表記”の同一数であり、 $x_{h,i}$ はその感情値である。

5. 候補表現からの感情生起表現の判定

3章で抽出した候補表現が感情生起表現であるかどうかを4章で構築した感情誘発語辞書を用いて判定する。判定は、感情誘発語辞書の感情値を用いて行う。つまり、感情値が閾値 ε 以上の感情誘発語を一語でも表現中に含む候補表現を感情生起表現と判定する。この時、その感情生起表現が表す感情は3章の末尾文節の決定に用いた「感情語 β 」の感情とする。例 5.1 では、「趣味で意気統合できたこと」という候補表現中に感情値 0.69 の感情誘発語「趣味」が含まれている。この時、 $\varepsilon < 0.69$ であれば、この候補表現は感情生起表現として判定される。また、末尾文節の決定に用いた感情語 β 「嬉しい」の感情〈喜〉をこの感情生起表現の示す感情とする。

例 5.1 候補表現からの感情生起表現の判定例

趣味で 意気統合できた こと
感情値：0.69

→ 感情生起表現：趣味で意気統合できたこと 〈喜〉

6. 実験と結果

感情誘発語辞書では、seed 語 1,925 語を基に4つの国語辞典から再帰的に感情誘発語を収集した結果、感情誘発語は異なり数で 11,348 語収集された。また、再帰の平均回数数は 7.5 回であった。語彙大系からは異なり数でさらに 6,000 語収集された。よって感情誘発語辞書には異なり数で 19,273 語が登録された。

感情生起表現の抽出実験には国会会議録[11]（約 870MB 約 6 百万文）を用いる。このテキストから「喜・怒・哀・怖・厭・驚」の6種の感情毎に感情生起表現を抽出し、そこから感情毎に無作為に 100 表現抜き出し、その表現を著者らが次の4つに分類した。

- (i) 示された感情の感情生起表現である
- (ii) 感情生起表現ではない
- (iii) 感情生起表現ではあるが、他の感情を表す
- (iv) 判定が困難

また、各処理における閾値及び定数は $\alpha=0.85$, $\delta_1=0.4$, $\delta_2=0.05$, $\delta_3=0.6$, $\varepsilon=0.25$ とした。結果を Table 1. に示す。

Table 1. 抽出した感情生起表現の分類

	i	ii	iii	iv
喜	56 表現	27 表現	6 表現	11 表現
怒	47 表現	37 表現	8 表現	8 表現
哀	61 表現	21 表現	9 表現	9 表現
怖	55 表現	25 表現	12 表現	8 表現
厭	54 表現	27 表現	8 表現	11 表現
驚	28 表現	27 表現	29 表現	16 表現

例 6.1 に抽出された感情生起表現とその抽出に用いた感情誘発語の国語辞典での収集過程の例を示す。括弧の値は感情誘発語の感情値である。

例 6.1 感情生起表現と感情誘発語の抽出過程の例

〈喜〉 実現へ向けて大きく**前進**をした (前進 0.49)

[望ましい→進歩→前進] (岩波国語辞典)

〈怒〉 合意された**協定**も守らない (協定 0.37)

[厳しい→嚴重→協定] (岩波国語辞典)

〈哀〉 家も**被害**を受ける (被害 0.64)

[危ない→危険→危害→被害] (新辞林)

〈怖〉 新たな社会**病理**が生み出されること (病理 0.35)

[病理→病氣→苦痛] (広辞苑)

〈厭〉 **不正**な取引が含まれていた (不正 0.43)

[悪い→不正] (三省堂現代国語辞典)

〈驚〉 知らない人が**結構**いるんだなと (結構 0.26)

[素晴らしい→結構] (新辞林)

7. 考察

抽出した表現には指示詞や語の省略によって感情生起表現であるかの判定が正確に行えなかった表現が2割近くあり、これには照応解析を取り入れて対応する必要がある。また、例 6.1 のように著者らが意図した通りの感情誘発語で表現の判定ができた場合以外に、誤った語で感情生起表現と判定されるケースも2割近くあったため、感情誘発語辞書の整備を行う必要がある。

今回の実験で抽出された感情生起表現は総計で約 8,000 個であり、あまり多くはなかった。これは表現の抽出に用いた「感情語 β 」の語彙数 (317 語) が少ないことが原因の一つに考えられる。しかし、今回構築した感情誘発語辞書の感情誘発語を「感情語 β 」として、テキストから感情生起表現を抽出することで、今回、抽出した感情生起表現よりもさらに多くの感情生起表現の抽出が行えることが期待できる。

8. おわりに

本稿では感情誘発語を用いた感情生起表現の抽出手法を提案し、感情生起表現の判定に感情誘発語を用いる有効性が確認された。今後の展望は感情誘発語辞書の整備や、ブログなどからの大規模な感情生起表現の抽出、さらには抽出した感情生起表現を用いて他のテキストの感情表現を解析する手法の考案などである。

参考文献

- [1] 中村明：「感情表現辞典」, 東京堂出版 (1993)
- [2] 飛田, 浅田：「現代形容詞用法辞典」, 東京堂出版 (1994)
- [3] 市川孝他：「三省堂現代国語辞典」, 三省堂 (1998)
- [4] 松村明他：「新辞林」, 三省堂 (1998)
- [5] 西尾実他：「岩波国語辞典 第六版」, 岩波書店 (2000)
- [6] 新村出：「広辞苑 第五版」, 岩波書店 (1998)
- [7] 「JICST 科学技術用語シソーラス 2004 年度版」, 科学技術振興事業団 (1999)
- [8] 山口, 武田：「岩波新漢語辞典 第二版」, 岩波書店 (2000)
- [9] 藤堂他：「漢字源」, 学習研究社 (2001)
- [10] 池原, 宮崎他：「日本語語彙大系」, 岩波書店 (1999)
- [11] 国会会議録検索システム, <http://kokkai.ndl.go.jp/>
- [12] CaboCha, <http://www.chasen.org/~taku/software/cabocha/>