

結合価パターン辞書からの情緒を明示する用言の知識ベース化

黒住亜紀子 徳久雅人 村上仁一 池原悟

鳥取大学 工学部 知能情報工学科

{s032023,tokuhisa,murakami,ikehara}@ike.tottori-u.ac.jp

1 はじめに

既に一般の辞書やソーラスには、用言が情緒を明示するかどうかという情報が書かれているものがある。しかしながら、情緒推定のための知識ベースとしては、情緒主や情緒対象などの情報が不足しており、そのまま情緒推定に使用することは出来ない。

そこで本稿では、情緒を明示する用言を対象に、情緒推定用の情報を追加し、情緒属性付き結合価パターン辞書を構築する。そして、日記テキストを対象に情緒属性付き結合価パターン辞書を用いた情緒推定実験を行い、辞書の性能を評価する。また実験結果から、今後の推定精度向上へ向けて考察を行う。

2 情緒属性付き結合価パターン辞書の作成

2.1 情緒属性付き結合価パターン辞書

本稿では、情緒を明示的に表す文の情緒推定を目指す。そこで、日本語語彙大系 [1] において、情緒を表していると考えられる「感情動作」、「感情状態」に属する用言の結合価パターンの中から、情緒を明示的に表すパターンを抽出して、「情緒主」、「情緒対象」、「情緒名」(本稿ではこれらを「情緒属性」と呼ぶ)の付与を行う。なお、辞書に付与する情緒は《喜び》,《悲しみ》,《好ましい》,《嫌だ》,《驚き》,《期待》,《恐れ》,《怒り》,《なし》の9種類とする。

図1に目標とするパターン辞書の一部を示す。「愛好する」という用言から、人物 $N1$ が $N2$ に対して《好ましい》という情緒を抱くことが分析できる。

日本語語彙大系に存在する項目
愛好する(あいこうする) (1) 31 感情動作 (状態 受身不可)
 $N1$ が $N2$ を愛好する $N1$ be fond of $N2$
[$N1$ (4人) $N2$ (1001 抽象物 1560 行為)]
追加項目 情緒主: $N1$, 情緒対象: $N2$, 情緒名: 《好ましい》

図1 目標とするパターン辞書の一部

2.2 用言パターンの分析

用言パターンは、以下の4つに大別する。

- 直接表現: 情緒を明示的に表す用言をここに分類する。さらに、情緒のみを表す(例: 愛好する)場合「単純」に、非情緒状態や他の情緒が混在する(例: 愛読する)場合「複雑」に分類する。
- 原因表現: 将来的に情緒が引き起こされる原因を表す用言をここに分類する。
- 反応表現: 情緒に大きく影響された行動や状態を表す用言をここに分類する。
- その他: 上記3つに属さない用言をここに分類する。大別した後、直接表現に対して情緒属性を付与する。なお、これらの分類を行う際には次の点に注意する。
 - 話者の視点: 情緒が話者自身の情緒であるのか、話者の信じる他人の情緒であるのかに注意する。
 - 相: どの時間軸の表現であるかに注意する。

2.3 用言パターンの分析例

分類と情緒属性付与の具体例を以下に示す。

直接表現の例 「 $N1$ が $N2$ を愛する」。情緒主が $N1$, 情緒対象が $N2$, 情緒は《好ましい》

原因表現の例 「 $N1$ が $N2$ を欺く」。 $N1$ が信頼を裏切ったことにより $N2$ が《怒り》を抱くと予測できる。

反応表現の例 「 $N1$ が $N2$ を笑う」。 $N1$ が $N2$ に対して持つ情緒の反応として「笑う」と考えられる。

2.4 分析結果

初めに1人目が感情に関する用言パターン 1,615 件を直接表現・原因表現・反応表現に分類し、情緒属性を追加する。次に2人目が話者の視点・相を考慮して校正する。この結果、60%(1,030/1,615)の用言が直接表現に分類された(表1)。このうち2者が共に直接表現と判断したものは882パターンであった。

さらに、分類の揺れを調べるため、直接表現の単純・複雑、および、非直接表現の原因・反応・その他の5つの分類項目から、それぞれ50パターンをランダムに選び、同意率を調べた。結果、どの項目も90%以上の同意率が得られた。

表1 感情状態/動作の用言の分類パターン数

分析者	直接表現	非直接表現
	(単純/複雑)	(原因/反応/その他)
1人目	1,034(939/95)	581(204/377/-)
2人目	1,030(916/114)	585(218/104/261)

3 日記テキストに対する情緒推定実験

3.1 実験の目的

情緒属性付き結合価パターン辞書の性能評価を目的とする。本辞書は情緒を明示的に表す用言パターンを情緒推定の対象とする。次の2つの項目で辞書を評価する。また、人手による推定の評価値と比較を行う。

情緒推定の精度： 情緒を明示する日記文に対する情緒推定の性能

辞書化の精度： 上記の日記文の中でも、パターンに適合した場合の情緒推定の性能

3.2 情緒タグ付き日記コーパス

情緒タグ付き日記コーパス [2] をテストデータとして、情緒推定実験を行う。本コーパスは、図書 [3] の日本語部分を収録している。[3] は情感ごとに章立てされており、網羅的な情緒に対して実験することができる。

本コーパスは、節の単位で文が分割されており、節ごとに経験者6名と初心者1名によって情緒タグが付与されている。コーパスの規模は以下の通りである。

- 日記のストーリー：268話
- 節数：1,642節
- 情緒タグ(経験者6名分)：計11,461個

本稿では情緒推定の正解を、経験者6名のタグの多数決で決定する。ただし、4名以上の一致かつ上位2位までを正解の情緒タグとする。

3.3 実験対象

次の条件を満たす節を実験対象として日記コーパスから抽出し、119節を得た。

- 形態素解析可能
- 述語部分が情緒を明示的に表現
- 述語数1
- 正解タグが決定可能

3.4 情緒推定の方法

情緒推定は、プログラムによるパターン照合によって行う。本辞書の結合価パターンを、パターン照合用に加工して照合に使用する。

まず、入力文の形態素解析結果に適合するパターンを検索する。その際、結合価パターンに定義されている変

数の名詞意味属性制約によって、適合パターンの絞り込みを行う。絞り込み後、回答候補となるパターンが複数ある場合は、人手で最適パターンを選択し、回答出力とする。1つのパターンに複数の情緒タグが付与されている場合は、付与されている全てのタグを回答情緒として出力する。

3.5 情緒推定の例

日記中の1節に対する情緒推定の例を示す。

入力文： 自然を満喫した。

正解タグ： 《喜び》, 《好ましい》

適合パターン： 「N1がN2を満喫する」

情緒主：N1, 情緒対象：N2, 情緒名：《喜び》

回答出力： 情緒主：話者, 情緒対象：自然,

情緒名：《喜び》

入力文にパターンが適合すると、パターンの持つ情緒属性が得られる。この例の場合、《好ましい》の出力はなかったが、情緒主、情緒対象、情緒名：《喜び》は正しく推定できている。

3.6 評価方法

評価はそれぞれ、再現率、適合率、F値、節単位の正解率で評価する。

$$\langle \text{再現率 } R \rangle = \frac{\langle \text{正解タグと一致した回答タグ数} \rangle}{\langle \text{正解タグの総数} \rangle}$$

$$\langle \text{適合率 } P \rangle = \frac{\langle \text{正解タグと一致した回答タグ数} \rangle}{\langle \text{回答タグの総数} \rangle}$$

$$\langle \text{F値} \rangle = \frac{2 \times R \times P}{R + P}$$

<節単位の正解率>

$$= \frac{\langle \text{回答タグが正解タグと1つ以上一致した節数} \rangle}{\langle \text{実験に使用した節数} \rangle}$$

[2] にならい、9分類系、5分類系、3分類系それぞれの評価値を算出する。各分類系は以下の通りである。

9分類系： 《喜び》, 《悲しみ》, 《好ましい》, 《嫌だ》, 《驚き》, 《期待》, 《恐れ》, 《怒り》, 《なし》

5分類系： 《P》, 《N》, 《A》, 《S》, 《なし》

3分類系： 《Positive》, 《Negative》, 《なし》

また、同じテストデータに対する人手の評価値との比較を行う。人手の評価値は、以下の条件で正解タグを決定して求める。

初心者： 経験者6名の多数決(4名以上一致かつ2位まで)で正解タグを決定

経験者： 自身を除く5名の多数決(3名以上一致かつ2位まで)で正解タグを決定

3.7 実験結果

3.7.1 情緒推定の精度

表 2 に、辞書を用いた情緒推定、初心者・経験者による情緒推定のそれぞれの評価値を示す。カッコ内の数字は分類系を示す。9 分類系での情緒推定の F 値が 54%、人手の F 値が 70~79% と、差が約 20% であった。119 節のテストデータのうち、84 節がパターンに適合し、35 節がパターンに適合しなかった。

表 2 情緒推定の精度の評価値

推定者(分類)	再現率	適合率	F 値	節単位の 正解率
辞書(9)	0.57	0.52	0.54	0.61
(5)	0.66	0.62	0.64	0.66
(3)	0.67	0.66	0.66	0.67
初心者(9)	0.71	0.7	0.71	0.78
(5)	0.90	0.85	0.87	0.92
(3)	0.96	0.96	0.96	0.97
経験者(9)	0.66~0.85	0.63~0.88	0.7~0.79	0.79~0.93
(5)	0.87~0.94	0.79~0.96	0.85~0.92	0.92~0.97
(3)	0.94~0.99	0.95~0.99	0.95~0.98	0.97~1.0

3.7.2 辞書化の精度

パターンに適合した 84 節の評価値を表 3 に示す。適合した節の場合、人手に近い推定精度となっており、辞書化は高い精度で行われていることが分かった。

表 3 辞書化の精度の評価値

推定者(分類)	再現率	適合率	F 値	節単位の 正解率
辞書(9)	0.80	0.68	0.74	0.86
(5)	0.93	0.85	0.89	0.94
(3)	0.94	0.92	0.93	0.95
初心者(9)	0.72	0.74	0.73	0.81
(5)	0.92	0.88	0.90	0.81
(3)	0.98	0.98	0.98	0.99
経験者(9)	0.66~0.87	0.66~0.89	0.71~0.80	0.80~0.94
(5)	0.89~0.97	0.83~0.99	0.88~0.95	0.92~0.99
(3)	0.94~0.99	0.95~0.99	0.95~0.98	0.96~1.0

4 情緒推定の誤り分析

情緒推定の誤り傾向を分析する。傾向は以下のようになった。入力文に対する正解タグと回答出力を例に、傾向別に分析する。

- パターンに適合したが情緒は不正解 (12 節)
- パターンに不適合 (35 節)
 - 辞書に類似した表現が存在 (8 節)
 - 辞書に類似した表現が存在しない (25 節)

4.1 パターンに適合したが情緒は不正解

例 1: 信じられなかった。

(正解)《驚き》/(回答出力)《期待》

例 2: 本当に良かった。

(正解)《喜び》/(回答出力)《好ましい》

例 1 は「N1 が N2 を信じる」(情緒名:《期待》)に適合した節である。否定のモダリティが付くことにより、情緒が変化している。日本語語彙大系では、一部のパターンが否定のモダリティを持っているが、情緒推定にとっては不足しているので、パターンを展開する必要があると考えられる。

例 2 は「N1 が N2 に良い」(情緒名:《好ましい》)に適合した節である。《喜び》と《好ましい》の情緒は、文脈によっても揺れがある。「良かった」の含まれる節は 8 節あり、そのうち 4 節は正解タグに《好ましい》が含まれるため、一概にパターンへの情緒属性付与の間違ひではないと考えられる。

4.2 パターンに適合しなかった節

4.2.1 辞書に類似した表現が存在

例 3: 腹立たしかった。

(正解)《怒り》/(回答出力)適合パターンなし

例 4: 度肝を抜かれたが、

(正解)《驚き》/(回答出力)適合パターンなし

例 3 は、辞書中に「N1 は腹が立つ」、「N1 が N2 に腹を立てる」、「N1 が N2 に立腹する」という類似したパターンがあるが、「腹立たしい」というパターンはない。言い回しによる微妙な違いを吸収する機構を使用する、または違いごとにパターンを増やす必要がある。

例 4 は、辞書中に「N1 が N2 の度胆を抜く」という類似パターンを持つが、受身形のパターンがない。日本語語彙大系では、一部のパターンが受身のモダリティを持っているが、情緒推定にとっては不足しているので、パターンを展開する必要があると考えられる。

4.2.2 辞書に類似した表現が存在しない

例 5: すごく幸せだった。

(正解)《喜び》,《好ましい》

/(回答出力)適合パターンなし

例 6: 午後は最悪だった。

(正解)《嫌だ》/(回答出力)適合パターンなし

例 7: とてものがっかりした。

(正解)《悲しみ》/(回答出力)適合パターンなし

例 8: ほっとした。

(正解)《好ましい》/(回答出力)適合パターンなし
例5,6の場合,日本語語彙大系において「N1が幸福だ」「N1が最悪だ」というパターンが「属性」に分類されている。本稿では,情緒を表す用言は「感情動作」,「感情状態」に分類されていると想定して辞書を作成したが,「属性」や「相対関係」に属する表現にも情緒を明示する用言パターンが見られるため,これらの分類も視野に入れる必要があると考えられる。

例7,8の「副詞+する」という表現,ほかに複合動詞などは,日本語語彙大系ではカバーされていない。

4.2.3 内容が情緒を表現

例9:今回のことで天災の恐ろしさを実感した。

(正解)《恐れ》/(回答出力)適合パターンなし

例9は,「実感する」という部分ではなく,「恐ろしさ」という内容が情緒を表している。この他に「感じる」,「思う」などは,内容部分が情緒を表現していることが多く,その場合はパターン部分のみでの情緒推定は出来ないため,内容部分からの情緒推定を行う必要があると考えられる。

5 考察

5.1 情緒生起の原因に着目した情緒推定との比較

[4]では,同じ日記コーパスに対して,情緒生起の原因を表現する用言の知識ベースを用いて情緒推定を行っている。正解タグ,人手評価を本稿と同条件で決定した際の一一致率(F値)は,知識ベースが48%,初心者が50%,経験者が52~61%であった。初心者と経験者の差が最大11%見られるが,本稿では初心者と経験者の差はほとんど見られず,直接表現においては人手による推定精度は安定していると考えられる。

5.2 客体的表現とゼロ代名詞に関する考察

実験に使用した日記コーパスは,日記を書いた話者の視点から書かれており,情緒主=話者である,という前提の元での情緒主の省略が多い。これに対して,作成した情緒属性付き結合価パターン辞書は「N1がN2を愛好する」のような客体的表現の中での関係(情緒主:N1,情緒対象:N2,情緒名:《好ましい》)を重視して情緒タグ付与を行ったため,主格や目的格が省略された文に対する情緒推定には問題がある。

例として,「N1はN2が厭だ」というパターンの場合を考える。辞書化は,「情緒主N1は情緒対象N2のことが《嫌だ》」という考えで行われる。しかし,日記コーパ

ス中にはこのパターンに適合する「N1は厭だ」という節が見られる。この節は情緒主が省略された形で,「情緒主=話者は情緒対象N1が《嫌だ》」という文である。例に上げた結合価パターンでは,「は格」が情緒主として定義されているが,情緒主の省略された文では,「は格」が情緒対象となっている。このように,「は格」,「が格」などの定義が,格要素の省略によって違ってくるため,正確に情緒対象を捉えることができない。

本稿では,日記コーパスに情緒対象の情報が付与されていないため,情緒対象の出力の評価を行っていない。しかしながら将来的には,日記コーパスに情緒対象の付与を行った後,再度の情緒推定実験を行って,格要素の省略に対してパターンがどのように変化するかを調査するとともに,正しい情緒属性の出力が出来るよう,パターンや推定機構の改良を行う必要がある。

5.3 今後の改良点

解析精度の向上に向けて,モダリティによる変化に情緒出力も対応させるためのパターンの展開が必要である。また,日本語語彙大系の「属性」,「相対関係」項目の辞書への追加を検討する。さらに,情緒主,情緒対象の推定精度の調査を行う必要がある。

6 おわりに

本稿では,用言パターンの分析により,情緒属性付き結合価パターン辞書(規模約1,000件)を作成した。そして,本辞書を用いた情緒推定実験を行った。結果,情緒推定精度の9分類系でのF値は54%であったが,辞書化は高精度で行われたことが分かった。また,誤り分析から,本辞書の改良点を示した。

謝辞

本研究は科研費(若手B:17700151)の下で行いました。

参考文献

- [1] 池原,宮崎,白井,横尾,中岩,小倉,大山,林:“日本語語彙大系”,岩波書店,1997.
- [2] 徳久,松浦,村上,池原:“情緒タグ付き日記コーパスの作成と評価”,第5回情報科学技術フォーラム,一般講演論文集,第2分冊,pp.235-236,2006.
- [3] 石原:“英語で日記を書いてみる表現集編”,ベレ出版,2002.
- [4] 徳久,村上,池原:“日記文における情緒生起の原因に着目した情緒推定の実験”,電子情報通信学会2007総合大会講演論文集,シンポジウム「円滑なコミュニケーションと感情」,2007(発表予定).