

# SO-scoreとパターン辞書を用いた情緒推定

竹本 翔\*1 徳久 雅人\*2 村田 真樹\*2

\*1 鳥取大学 工学部 知能情報工学科

\*2 鳥取大学 大学院 工学研究科 情報エレクトロニクス専攻

{s082030, tokuhisa, murata}@ike.tottori-u.ac.jp

## 1 はじめに

情緒推定は、テキストの自動分析などへの応用が期待されている。情緒推定の手法には、用言の語義を定義した情緒属性付き結合価パターン辞書を用いた手法がある[1]。しかしその手法では、修飾語句を含む格要素への対処が未実装である。例えば、「おいしいご飯を食べる。」も「まずいご飯を食べる。」も《好ましい》と推定されてしまう。そこで本稿では修飾語句の評価極性を「好評」、「不評」および「なし」の3分類で捉え、それらを情緒推定に反映させるという手法の改良を行う。

## 2 パターン辞書を用いた情緒推定の手法

情緒属性付き結合価パターン辞書のレコードの1つを表1に示す。日本語パターンと意味属性制約は用言の語義を表す。それ以外の5項目は情緒属性を表し、そのうち判断条件は格要素同士の関係を、情緒原因は情緒が生起する原因を、情緒名は生起する情緒の候補を表す。情緒名には《喜び》,《好ましい》,《期待》,《悲しみ》,《恐れ》,《嫌だ》,《怒り》,《驚き》および《なし》の9分類を使用している。

表 1: 情緒属性付き結合価パターン辞書の一部

◆日本語パターン: $N1$ は $N2$ を 食べる
◆意味属性制約: $N1$ (人 動物) $N2$ (食料 生物)
◆判断条件: 生理・近 ( $N1, N2$ ) ◆情緒原因: 〈五感 (味覚)〉
◆情緒主: $N1$ ◆情緒対象: $N2$ ◆情緒名: 《好ましい》

従来手法[1]では、まずこの辞書に入力文を照合する。その際、入力文に対し、日本語パターンが適合し、かつ

意味属性制約が充足するレコードを検索する。次にそのレコードの判断条件の真偽を判定する。この時、判断条件の真偽は引数に代入される値(主に格要素)によって決まる。最後に、判断条件が真ならば、情緒名などの情緒属性を出力し、偽ならば情緒《なし》を出力する。

例を挙げると、「私はご飯を食べる。」という入力文の場合、表1の日本語パターンに適合し、かつ意味属性制約として“私”が(人 動物)を充足し、“ご飯”が(食料 生物)を充足するので、情緒属性が得られる。得られた判断条件「生理・近 ( $N1, N2$ )」は、格要素同士の関係が生理状態に関係するもので、かつ「接近」の関係であることを意味する。具体的な格要素を含めた判断条件「生理・近(私, ご飯)」はこの条件を満たすため、情緒名候補である《好ましい》を出力するというものである。

しかし、「私はまずいご飯を食べる。」という入力文の場合、先程の例と同じように情緒名が推定されてしまう。名詞変数  $N2$  に“ご飯”のみがマッチし、“まずい”という修飾語句がマッチしていないため、“まずい”が“ご飯”に与える影響が無視されたからである。このように、従来手法には、修飾語句の意味に合った情緒名が得られないという問題がある。

## 3 手法の改良

従来手法では、判断条件に代入する値に名詞変数  $N$  を用いているが、本稿の手法では、修飾語句を含むことのできる名詞句変数  $NP$  を用いる。そして、名詞句の評価極性に基づき判断条件の真偽を判定するように改良する。

### 3.1 Turney らの評価極性

まず名詞句に評価極性を付与する手法として Turney らの手法を用いる。Turney らは、次式を用い、コーパスから語句  $t$  の評価極性を算出した [2]。ここで、 $p(a, b)$  はコーパス内において語句  $a$  と語句  $b$  が同一文で共起する確率を、 $p(a)$  は語句  $a$  を含む文がコーパス内に出現する確率をそれぞれ表す。本稿では、 $SO\text{-}score$  を算出するためのコーパスに 2 年分のブログ文コーパスを用いた。ただし、このコーパスから極性の反転として扱われる接続表現 [3] を含む文は予め除いておく。

$$SO\text{-}score(t) = PMI(t, \text{“好評表現”}) - PMI(t, \text{“不評表現”})$$

$$PMI(a, b) = \log_2 \frac{p(a, b)}{p(a)p(b)}$$

$SO\text{-}score$  の値が正の場合、 $t$  が好評極性、負の場合、 $t$  が不評極性と解釈する。 $SO\text{-}score$  が 0 の場合は「極性なし」とし、また、ブログ文コーパスに出現した共起頻度が 3 以下の語句も極性なしとする。なお、本稿では好評表現に“良い”，不評表現に“悪い”を用いた。

### 3.2 名詞句の評価極性の算出

名詞句には多くのパターンがあるため、本稿ではブログ文で最も出現する上位のパターンのみを扱う。まず修飾語句に係ることで名詞句になる名詞句パターンを作成し、それらをブログ文コーパスに照合して名詞句を抽出する。その抽出結果から、最も抽出数の多い名詞句パターンを選択する。本稿ではその上位 3 パターンを扱うことにした。(1) 名詞 - の - 被修飾語句、(2) 形容詞 - 被修飾語句、(3) 形容動詞 - 被修飾語句の 3 つである。

これらの名詞句の評価極性を Turney らの方法で求める際、コーパス内に同一の名詞句が存在しないとき、共起頻度が得られず、評価極性が算出できない。そこで、本稿では、まず各名詞句をいくつかのパーツに分解する。次に、それぞれのパーツ毎に  $SO\text{-}score$  を算出し、その総和を名詞句全体の  $SO\text{-}score$  とする。ただし、最長一致により評価極性を算出するために、表 2 の分解方法のうち、上にあるものを優先する。

### 3.3 評価極性を用いた判断条件の判定

判断条件の第二引数  $NP_y$  の評価極性から、判断条件の真偽を判定する。表 3 に 2 つの判断条件について真理

表 2: 名詞句の分解パターンと例

名詞句パターン	分解方法	例
名詞 - の - 被修飾語句	分解なし	新年の挨拶
	名詞の + の被修飾語句	新年の + の挨拶
	名詞の + 被修飾語句	新年の + 挨拶
	名詞 + の被修飾語句	新年 + の挨拶
	名詞 + 被修飾語句	新年 + 挨拶
形容詞 - 被修飾語句	分解なし	美味しいおせち
	形容詞 + 被修飾語句	美味しい + おせち
形容動詞 - 被修飾語句	分解なし	雑なおせち
	形容動詞 + 被修飾語句	雑な + おせち

値表を示す。他の判断条件「目標実現・近、目標実現・離、対人関係・近、対人関係・離」についても同様である。

表 3: 判断条件「生理・近」、「生理・離」の真理値表

$NP_y$	生理・近 ( $NP_x, NP_y$ )	生理・離 ( $NP_x, NP_y$ )
好評	$T$	$F$
不評	$F$	$T$

例として、「私は美味しいご飯を食べる。」という入力文があったとする。この文は表 1 のレコードにマッチする。この時、従来手法では具体的な格要素が含まれる判断条件として「生理・近 (私, ご飯)」が得られるが、提案手法では  $NP_y$  として“美味しいご飯”が用いられるため、「生理・近 (私, 美味しいご飯)」が得られる。“美味しいご飯”の  $SO\text{-}score$  の値は正であり、好評極性が付与される。そして表 3 を参照すると  $T$  が成り立つため、情緒名《好ましい》を出力する。 $F$  となる場合は情緒名なしを出力する (情緒名の出力を抑制する)。

## 4 実験

### 4.1 実験条件

$SO\text{-}score$  の算出に用いたブログ文とは別のブログ文をテスト文とし、まずは、それらに人手で情緒タグを付与しておく。テスト文に自動で情緒推定を行った結果とタグを比較することで、精度 (一致数/出力数)、再現率 (一致数/正解数) を調べる。ここで、出力数は出力された情緒名の数 (複数出力有り) を表し、正解数は正解情緒名 (人手で付与した情緒タグ) の数を表し、一致数は人手で付与した情緒と、出力された情緒との一致数を表す。テスト文にはブログ文 3,000 文のうち、以下の条件を満たす 270 文を用いる。

- 結合価パターンにマッチし、情緒対象を含む
- 動詞の時制が過去と現在で、否定語を含まない
- 極性の反転と扱われる接続表現 [3] を含まない

## 4.2 実験結果

従来手法と提案手法を比較した結果を表 4 に示す。

表 4: 実験結果 (9 分類系)

	精度 (一致数/出力数)	再現率 (一致数/正解数)
従来手法	22.83 % (92/403)	34.07 % (92/270)
提案手法	22.89 % (92/402)	34.07 % (92/270)

## 4.3 出力が抑制された例

「平岩駅近くの温泉宿に泊まりました。」という文では表 5 と表 6 のレコードにマッチする。それにより、具体的な格要素の含まれる判断条件として「目標実現・近 ( $\phi$ , 近くの温泉宿)」と「目標実現・離 ( $\phi$ , 近くの温泉宿)」の 2 つが得られる。情緒対象 “近くの温泉宿” に対して評価極性を算出すると, “近くの” の *SO-score* は 0.42, “温泉宿” の *SO-score* は 2.25 であり, それらの和である 2.67 が “近くの温泉宿” の *SO-score* となる。正の値であるため, “近くの温泉宿” の評価極性は好評極性であると考え。評価極性が得られたので表 3 を参照すると, 目標実現・近 ( $\phi$ , 近くの温泉宿) では *T* が, 目標実現・離 ( $\phi$ , 近くの温泉宿) では *F* が成り立つ。よって表 6 のレコードの情緒属性の出力を抑制し, 表 5 のレコードの情緒名《好ましい》のみが出力された。

表 5: 情緒属性付き結合価パターン辞書の一部

- ◆日本語パターン:  $N1$  は  $N2$  に 泊まる
- ◆意味属性制約:  $N1$ (人)  $N2$ (場 場所 住居)
- ◆判断条件: 目標実現・近 ( $N1$ ,  $N2$ ) ◆情緒原因: 〈自由〉
- ◆情緒主:  $N1$  ◆情緒対象:  $N2$  ◆情緒名: 《好ましい》

表 6: 情緒属性付き結合価パターン辞書の一部

- ◆日本語パターン:  $N1$  は  $N2$  に 泊まる
- ◆意味属性制約:  $N1$ (人)  $N2$ (場 場所 住居)
- ◆判断条件: 目標実現・離 ( $N1$ ,  $N2$ ) ◆情緒原因: 〈制限〉
- ◆情緒主:  $N1$  ◆情緒対象:  $N2$  ◆情緒名: 《嫌だ》

## 4.4 誤り分析

表 4 の再現率において一致しなかった 184 文のうちランダムに 50 文を抽出して誤り分析を行った。誤りの内訳を表 7 に示す。抑制が行われても情緒が一致しなかったものを以下に示す。

表 7: 出力誤りの内訳

#	理由	件数
1	意味属性制約を充足していない	21
2	意味属性制約を充足する	
2.1	抑制なしで情緒不一致	27
2.2	抑制ありで情緒不一致	2

### 4.4.1 抑制なしで情緒不一致だったもの

「犬の散歩を兼ねて、いつものように周囲を散策する。」というテスト文は表 9 のレコードにマッチする。

表 8: 情緒属性付き結合価パターン辞書の一部

- ◆日本語パターン:  $N1$  が/は  $N2$  を/は 散策する
- ◆意味属性制約:  $N1$ (人)  $N2$ (場 場所)
- ◆判断条件: 目標実現・近 ( $N1$ ,  $N2$ ) ◆情緒原因: 〈成行き〉
- ◆情緒主:  $N1$  ◆情緒対象:  $N2$  ◆情緒名: 《期待》

その結果, 目標実現・近 ( $\phi$ , 周囲) が得られる。この文には提案手法が適用されず, 情緒名《期待》が出力される。人手で付与していたのは情緒名《好ましい》であったので, 情緒名《期待》とは一致しなかった。

### 4.4.2 抑制ありで情緒不一致だったもの

「縄文時代の遺跡を 2 箇所見学してきました。」というテスト文は表 9, 表 10 のレコードにマッチする。

表 9: 情緒属性付き結合価パターン辞書の一部

- ◆日本語パターン:  $N1$  は  $N2$  を 見学する
- ◆意味属性制約:  $N1$ (主体)  $N2$ (組織 施設)
- ◆判断条件: 目標実現・近 ( $N1$ ,  $N2$ ) ◆情緒原因: 〈教訓〉
- ◆情緒主:  $N1$  ◆情緒対象:  $N2$  ◆情緒名: 《好ましい》

表 10: 情緒属性付き結合価パターン辞書の一部

- ◆日本語パターン:  $N1$  は  $N2$  を 見学する
- ◆意味属性制約:  $N1$ (主体)  $N2$ (組織 施設)
- ◆判断条件: 目標実現・離 ( $N1$ ,  $N2$ ) ◆情緒原因: 〈迷信〉
- ◆情緒主:  $N1$  ◆情緒対象:  $N2$  ◆情緒名: 《嫌だ》

その結果, 目標実現・近 ( $\phi$ , 縄文時代の遺跡) と目標実現・離 ( $\phi$ , 縄文時代の遺跡) が得られるので, “縄文時代の遺跡” の *SO-score* を算出する。その結果, “縄文時代” の *SO-score* が  $-1.00$  で, “の遺跡” の *SO-score* が 0.48 となり, “縄文時代の遺跡” の *SO-score* は  $-0.52$  となる。そのため評価極性が不評価極性になり, 表 3 よ

り、表9の情緒名《好ましい》が抑制された。人手で付与していたのは情緒名《喜び》であったので、抑制が行われたが一致しなかった。

## 5 考察

提案手法により文の意味に合わない情緒名を抑制した数が少なかったが、わずかに性能の向上が見られた。これはブログ文の中で修飾される情緒対象が少なかったことと、*SO-score*による評価極性が不適切なものが含まれたことが原因である。

### 5.1 名詞句の分解方法について

表2の分解方法のうち、名詞句分解後の語句で*SO-score*を算出した結果を表11に示す。各名詞句はブログコーパスから「AのB」のパターンで抽出した語句のうち、頻度が上位のものから選択した。

表 11: A の B 型名詞句の *SO-score*

名詞句 (A-の-B)	分解なし	分解後の <i>SO-score</i>			
	A の B	A	A の	B	の B
感謝-の-気持ち	1.05	1.04	1.00	-1.35	-0.40
感謝-の-言葉	1.71	1.04	1.00	-0.43	-0.10
心-の-叫び	-0.68	0.10	-0.34	-0.80	-0.95
心-の-支え	0.78	0.10	-0.34	0.07	-0.24
大荒れ-の-天気	0.00	-0.19	0.36	0.60	-0.38
太陽-の-光	1.12	0.70	0.94	0.50	0.65
新年-の-ご挨拶	2.58	2.23	4.00	1.70	1.24
元旦-の-朝	1.58	1.40	2.08	-0.24	-0.18

表11より、分解なしの「AのB」が最も同意できる値になっていると考えられる。表2において、「Aの」や「のB」等の「の」がつく分解語句を優先して用いているのは、他の単語に繋がるという情報を含んでいると考えたからである。「Aの」+「のB」の値と「A」+「B」の値では、「Aの」+「のB」の値の方が同意できる値になっていると考えられる。

名詞句パターンの「名詞 - の - 被修飾語句」は、今回「AのB」という形であったが、「AのBのC」のように「の」が増えることによって名詞句はさらに多くの意味を得る。そのような語句が出現した場合、本手法では「の」一つ分までしか考慮していないため、文の意味に合った評価極性が得られない可能性もある。本手法の

精度を上げるには、修飾語句を多く含み、かつ文が整っているコーパスを用いる必要性が考えられる。

### 5.2 5分類系と3分類系での実験

本稿では主な手法として情緒9分類系で実験を行ったが、5分類系、3分類系で同様の実験を行った結果を表12、表13に示す。なお、5分類系は *Positive*(好ましい, 期待, 喜び), *Negative*(悲しみ, 恐れ), *Angry*(怒り, 嫌だ), *Surprise*(驚き) および, なし(なし)の5分類であり、3分類系は *Positive*(好ましい, 喜び, 期待), *Negative*(悲しみ, 嫌だ, 恐れ, 怒り) および, なし(驚き, なし)の3分類である。かつこ内は対応する9分類系の情緒名である。

表 12: 実験結果 (5分類系)

	精度 (一致数/出力数)	再現率 (一致数/正解数)
従来手法	45.43 % (154/339)	57.04 % (154/270)
提案手法	45.56 % (154/338)	57.04 % (154/270)

表 13: 実験結果 (3分類系)

	精度 (一致数/出力数)	再現率 (一致数/正解数)
従来手法	51.59 % (162/314)	60.00 % (162/270)
提案手法	51.76 % (162/313)	60.00 % (162/270)

## 6 おわりに

本稿では、判断条件で扱われている名詞を名詞句として扱うことで、修飾語句の意味を含めた情緒推定を行った。しかし、本手法はブログ文にはあまり適しておらず、*SO-score*による評価極性の算出においても、安定した判別結果が得られないため、まずは評価極性の割り出し方について再検討する必要性が考えられる。

**謝辞** 本研究は、科学研究費補助金(若手研究(B): 22700100)のもとで行いました。

## 参考文献

- [1] 野口和樹, 滝川晃司, 徳久雅人: “情緒属性付き結合価パターン辞書により格要素の評価極性を考慮した情緒推定”, 電子情報通信学会技術研究報告, 思考と言語, TL2011-32, pp.63-68, 2011.
- [2] Turney, P.D.: “Thumbs Up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews.”, Proc. of ACL, pp.417-424, 2002.
- [3] 中道龍三, 徳久雅人, 村上仁一, 池原悟: “情緒推定の手がかりとなる接続表現の収集”, 電子情報通信学会技術研究報告, 思考と言語, TL2008-44, pp.1-6, 2008.