

感情コーパスからの感情語の自動抽出

Automatic Extraction of Emotional Words from Emotion Corpus

筒井 奈々†
Nana Tsutsui

篠山 学‡
Manabu Sasayama

1. はじめに

文から書き手の感情を推定する研究はいくつかある。松尾ら[1]や松本ら[2]は、感情を推定するために、感情語を用いている。感情語とは、感情を表す単語である。感情語を感情ごとに集めた感情語辞書を作成し、感情推定に用いている。この感情語辞書の作成は人手で行っており、手間がかかる。また、推定する感情の種類が変われば再び感情語辞書を作成しなければならない。本研究では、 χ^2 値を用いて感情コーパスから自動で感情語の抽出を行い、感情語辞書を構築する手法を提案する。

本研究では、「喜び」、「楽しみ」、「期待」、「感謝」、「好き」、「安心」、「憧れ」、「悲しみ」、「驚き」、「不安」、「怒り」、「焦り」、「嫌悪」の 13 種類の感情を使用する。

2. 感情語の自動抽出の概要

図 1 に本稿で提案する感情語の自動抽出の流れを示す。まず、感情コーパス内の各感情文を形態素解析し、自立語のみを抽出する(図 1-①)。感情コーパスは、文を各感情へ人手で分類し、集めたものである。抽出の対象とする自立語の品詞は形容詞、副詞、感動詞である。名詞や動詞は個人によって感じる感情が異なるため除外する。助詞や助動詞など非自立語は感情を持たないと思われるため除外する。次に抽出した自立語の χ^2 値を計算す

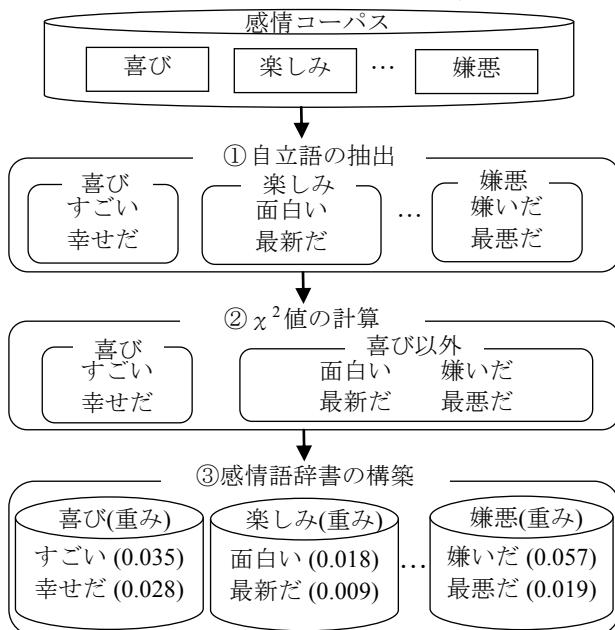


図 1. 感情語の自動抽出の流れ

† 香川高等専門学校詫間キャンパス電子情報通信専攻

‡ 香川高等専門学校詫間キャンパス情報工学科

る(図 1-②)。 χ^2 値とは、ある集合に対する複数の分割がどの程度一致しているかを示す指標である。 χ^2 値が大きいほど、ある感情に関連する単語である。 χ^2 値は、

$$\chi^2 = \frac{N \times (AD - BC)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)}$$

で求める[3]。ある単語 X を含むある感情 E の文数を A、単語 X を含む感情 E 以外の文数を B、単語 X を含まない感情 E の文数を C、単語 X を含まない感情 E 以外の文数を D、全文数を N とする。計算した χ^2 値の結果を元に感情語を自動抽出し、感情語辞書を構築する(図 1-③)。

3. χ^2 値を用いた感情語辞書の作成

χ^2 値を用いて感情語辞書を作成する。 χ^2 値が 6.635 以上の単語のみを感情語辞書に登録する。これは有意差を 1% とするためである。作成に用いた感情コーパスは合計 5283 文、抽出した自立語は合計 1142 語である。1142 語を対象に χ^2 値を用いて感情語辞書を作成した。 χ^2 値を用いて作成した 13 感情のうち嫌悪の感情語に対応する χ^2 値とともに表 1 に示す。次に χ^2 値から重みを求める。 χ^2 値は各感情コーパス内の文の数に影響される。そのため、感情推定を行う際、異なる感情語辞書の χ^2 値は比較することが出来ないからである。本研究では、ある感情 E の感情語辞書内の単語 X の重みを

$$\text{重み} = \frac{\text{単語 X の } \chi^2 \text{ 値}}{\text{感情 E の感情語辞書内の全単語の } \chi^2 \text{ 値}}$$

と定義する。嫌悪の感情語辞書と対応する重みを表 2 に示す。また、提案手法による感情語辞書に登録された感情語の数を表 3 の A に示す。表中の A は「提案手法による各感情語辞書に登録されている感情語数」を示す。提案手法による感情語辞書に登録された感情語数は合計 471 語である。

4. 評価実験

4.1 実験目的

提案手法による感情語辞書が感情推定に有効であることを確認するために評価実験を行う。まず提案手法による感情語辞書を用いて感情推定を行う。次に、既存手法を用いて感情推定を行う。これら 2 種類の感情推定した結果を比較し、提案手法による感情語辞書を評価する。

比較に用いる感情推定手法には、松本ら[2]の手法を使用する。松本らは人手によって作成した感情語辞書に対し tf・idf 法を用いて各感情語に重みを付与している。付与された重みを用いて感情推定を行う。

4.2 実験条件

感情推定を行う文には、感情コーパス内の感情文 5283 文からランダムに選んだ合計 1757 文を用いる。既存手法で用いる人手による感情語辞書を本研究に関係のない学生 1 名が書籍から作成した。作成された感情語辞書に登録された感情語の数を表 3 の B に示す。表中の B は「人手

表1. 感情語と χ^2 値

単語	χ^2 値
大嫌いだ	374.227
いやだ	277.579
嫌いだ	133.424
ばかばかしい	106.719
苦手だ	80.024
⋮	⋮
弱い	7.607

表2. 嫌悪の感情語辞書

単語	重み
大嫌いだ	0.161
いやだ	0.119
嫌いだ	0.057
ばかばかしい	0.046
苦手だ	0.034
⋮	⋮
弱い	0.003

による各感情語辞書に登録されている感情語数」を示す。人手により作成された感情語辞書に登録された感情語は292語である。

4.3 実験方法

提案手法による感情語辞書を用いた感情推定の手順を説明する。感情推定を行う文を形態素解析する。解析結果として得られた単語から感情語辞書を用いて感情語のみを取り出す。取り出した感情語には感情の重みが付与されている。各感情ごとに重みの合計を求め、値が最大となる感情を文の感情推定結果とする。

人手により作成された感情語辞書を用いた感情推定の手順を説明する。感情推定を行う文を形態素解析する。解析結果として得られた単語から感情語辞書を用いて感情語のみを取り出す。取り出した感情語には感情の重みが付与されている。各感情ごとに重みの合計を求め、値が最大となる感情を文の感情推定結果とする。

得られた2種類の感情推定した結果を比較し評価する。感情推定した結果と感情推定文を抽出した元のコーパスの感情が一致すれば正解とする。一致しない場合や感情推定文に感情語が1語も存在しない場合は不正解とする。

4.4 実験結果

提案手法による感情語辞書を用いた感情推定の正解率と人手により作成した感情語辞書を用いた感情推定の正解率を表4に示す。表中のAは「提案手法による感情語辞書を用いて感情推定を行った正解率」、Bは「人手により作成した感情語辞書を用いて感情推定を行った正解率」を示す。

表3. 各感情語辞書に登録されている感情語数

感情	A	B
喜び	53	46
楽しみ	27	32
期待	27	25
安心	38	17
感謝	17	8
好き	27	12
憧れ	41	15
驚き	33	9
焦り	32	22
不安	24	27
悲しみ	49	45
怒り	48	19
嫌悪	55	15

表4. 実験結果

感情	A(%)	B(%)	感情判定した文数
喜び	57.7	58.4	142
楽しみ	33.3	39.7	78
期待	15.1	18.8	488
安心	20.3	10.1	59
感謝	72.2	91.6	36
好き	41.0	39.2	56
憧れ	65.2	39.1	46
驚き	29.4	19.6	51
焦り	18.1	24.1	116
不安	9.8	18.3	273
悲しみ	30.0	20.6	233
怒り	23.2	18.9	116
嫌悪	53.9	19.0	63

13種類の感情のうち、7種類の感情（安心、好き、憧れ、驚き、悲しみ、怒り、嫌悪）は提案手法による感情語辞書の方が人手により作成された感情語辞書より正解率が上がった。

5. 考察

13種類の感情のうち7種類の感情で既存の手法より正解率を上げることが出来た。提案手法による正解率が良かった要因として、人手で作成された感情語辞書には登録していなかった感情語が登録できたことが挙げられる。提案手法による感情語辞書には安心の「ほっと」や悲しみの「ぺこぺこ」、驚きの「えっ」、「わっ」など人手で作成された辞書には登録されていない擬態語を登録することができた。

13種類の感情のうち6種類の感情では正解率を上げることが出来なかった。提案手法による正解率が悪かった要因として、人手で作成された感情語辞書には、名詞や動詞、接尾辞などが含まれていたことが挙げられる。感謝の「感謝」(名詞)、「お礼」(名詞)や不安の「心配」(名詞)などは人手で作成された感情語辞書にのみ登録されていた。「数学」や「テニス」、「歩く」など名詞や動詞の感情は個人の嗜好に左右されると考えたため、本研究では含まなかった。

6. おわりに

χ^2 値を用いて感情コーパスから感情語の自動抽出を行い、感情語辞書の構築を行う手法を提案した。13種類の感情のうち7種類の感情でよい正解率を出すことができた。

今後は名詞や動詞などの単語も含めて感情語の自動抽出を行う。しかし、名詞や動詞の単語には個人の嗜好によって感情が変化する単語が含まれる。そのような単語を除外する手法を考える。

参考文献

- [1] 松尾朋子,篠山学,“顔文字を考慮した対話テキストの感情推定に関する研究”,香川高等専門学校研究紀要,pp151-153(2010)
- [2] 松本和幸,“会話文からの話者感情推定に関する研究”,徳島大学大学院工学研究科博士論文(2008)
- [3] Yuming Yang, JanO. Pedersen, “A Comparative Study on Feature Selection in Text Categorization”, Proceedings of the Twentieth International Conference on Machine Learning (1997)