

3つの感情軸による小説を読んだときに感じる感情分析

島山 侑也[†] 黄 潤和[‡]

法政大学大学院 情報科学研究科[†] 法政大学 情報科学部[‡]

1. まえがき

小説の文章から感情を読み取ることが出来れば、ただ読むだけではなく感情を付与する本の読み聞かせロボットの構築に利用することが出来る。そこで、本研究では、小説の文章から読み取れる感情の推定システムを提案する。本システムでは、「楽しい⇔悲しい」、「嬉しい⇔怒り」、「のどか⇔緊迫」の3つの感情軸を用いることでより人間の感情に近い結果が得られるようにしている。本システムを用いて算出された感情値と被験者実験によって得られた感情値の比較を行い、本システムが有効かどうかを検証する。また、この2種類の感情値の結果から、読み手の性格を考慮した読み聞かせロボットの構築をする場合にはどのようなことが必要かも考える。

2. 感情辞書

熊本らによる新聞記事を対象とし、人々が感じる印象を抽出する手法[1]や松本らによるニュース記事の信憑性判断を支援する手法[2]では、感情辞書[1][2]が用いられている。感情辞書は、5年分の読売新聞記事データベース(2002年版~2006年版)を解析し、記事に現れる各単語が記事の印象に及ぼす影響を数値化したもの[1]であり、「楽しい⇔悲しい」、「嬉しい⇔怒り」、「のどか⇔緊迫」の3つの感情軸を用いて分析が行われる。本研究では、インターネット上の電子図書館である青空文庫に登録されている短編小説計146冊を用いて、感情辞書の構築を行う。

感情辞書の作成方法であるが、熊本らによる手法[1]を参考にする。まず最初に、各感情軸の左側の感情を表す感情語群と右側の感情を表す感情語群を定義し、どちらかの感情語群に含まれる感情語が1語以上含まれている小説データを抽出し、感情語の数をそれぞれの感情語群ごとに数える。今回の実験で用いる各感情軸を表す感情語群は表1に表す。このとき、左側の感情を表す感情語群に属する感情語の数が右側の感情を表す感情語群に属する感情語より多く含まれている小説集合を S_L (数を N_L 個)とし、少ない小説集合を S_R (数を N_R 個)とする。次に、それぞれの小説集合から助詞、連体詞、指示詞以外の全ての単語を抽出し、単語ごとに出現回数を数える。このとき、ある単語 ω の小説集合 S_L において出現した小説の数を $N_L(\omega)$ 、小説集合 S_R において出現した小説の数を $N_R(\omega)$ とすると、それぞれの条件付確率は以下の式(1)で表される。

$$P_L(\omega) = \frac{N_L(\omega)}{N_L}, \quad P_R(\omega) = \frac{N_R(\omega)}{N_R} \quad (1)$$

この式(1)により求められた $P_L(\omega)$ と $P_R(\omega)$ を用いて単語 ω の感情値 $v(\omega)$ を式(2)にて表す。なお、 $weight_L$ や $weight_R$ は重みであり、該当する感情語が含まれる小説の数が多ければ多いほど大きくなるようになっている。

$$v(\omega) = \frac{P_L(\omega) * weight_L}{P_L(\omega) * weight_L + P_R(\omega) * weight_R} \quad (2)$$

$$weight_L = \log_{10} N_L$$

$$weight_R = \log_{10} N_R$$

これらの方法で単語 ω の感情軸における感情語として感情辞書に登録する。

表1 各感情軸を表す感情語群.

感情軸	感情語
楽しい ⇔悲しい	楽しい, 楽しむ, 楽しみだ, 楽しげだ 悲しい, 悲しむ, 悲しみだ, 悲しげだ
うれしい ⇔怒り	うれしい, 喜ばしい, 喜ぶ 怒る, 憤る, 激怒する
のどか ⇔緊迫	のどかだ, 和やかだ, 素朴だ, 安心だ 緊迫する, 不気味だ, 不安だ, 恐れる

3. 実験

本研究では、感情辞書を用いて算出された感情値と人々が感じる感情値の2つの感情値の比較を行い、今回提案した手法が有効かどうかを検証する。

3.1. 実験方法

最初に、人々が感じる感情値のデータを得るために、大学生60人による被験者実験を行った。本実験では、前後の文章によって同じ文章でも人々が感じる感情も変わるだろうという推測から、長さの異なる3パターンの文章を用意した。このとき、計2~3文で構成されている文章をパターン(1)、パターン(1)の文章が含まれている段落をパターン(2)、パターン(2)の段落とその前後の段落の計3~4段落で構成されている文章をパターン(3)としている。文章は、各パターンに応じた長さの文章を6種類の小説からそれぞれ抜粋したものを使用した。そして、被験者60人を各パターン20名ずつに分け、被験者にはこれら6種類の文章を読んでもらい、評価してもらった。評価方法は、3つの感情軸である「楽しい⇔悲しい」、「うれしい⇔怒り」、「のどか⇔緊迫」のそれぞれに対し、「左側の感情を…(1)感じる、(2)割と感じる、(3)やや感じる、(4)両方の感情を感じない、右側の感情を…(5)やや感じる、(6)割と感じる、(7)感じる」の7段階評価である。なお、パターン(1)、(2)では全体を読んだ評価、パターン(3)では各段落ごとに評価をしてもらっている。以上の結果によって得られた数値の平均を各パターンごとに算出し、これら

Emotion Analysis When Reading a Novel Based on 3-axes Emotion Model

[†]Yuya Hatakeyama, [‡]Runhe Huang

[†]Graduate School of C.I.S, Hosei University

[‡]Faculty of C.I.S, Hosei University

を人々が感じる感情値とする。

次に、小説データの感情値を算出する。まず、小説データを形態素解析システム Juman を用いて単語に分解する。このとき、Juman によって出力された結果をそのまま用いずに、熊本らによって定められたルール[1]に従って変換作業を行う。そして、感情辞書から各単語の感情値を感情軸ごとに取り出していき、感情軸ごとの平均値を計算し、それを小説データの感情値とする。なお、このとき得られる感情値は感情軸の左側の感情が強いほど 0、右側の感情が強いほど 1 に近づくようになっており、被験者実験では、感情軸の左側の感情が強いほど 1、右側の感情が強いほど 7 に近づくようになっているため、以下の式(3) [1]を用いて感情辞書で得られた感情値を変換し、被験者実験と同じスケールとなるように換算する。

$$\text{換算値} = (1 - \text{算出値}) * 6 + 1 \quad (3)$$

3.2. 実験結果と考察

表 2 が、感情辞書で得られた感情値と被験者実験のパターン(1)の比較、そして表 3 が感情辞書で得られた感情値と被験者実験のパターン(2), (3)で得られた感情値の比較を表したものである。軸 1 は「楽しい⇔悲しい」、軸 2 は「うれしい⇔怒り」、軸 3 は「のどか⇔緊迫」のそれぞれの感情軸を表しているとする。

感情値が 1 以上 4 未満の場合は感情軸の左側の感情、4 超え 7 以下の場合は右側の感情を感じ、4 の場合は両方の感情を感じないとすると、表 2 より、パターン(1)は 18 項目中 13 項目、表 3 より、パターン(2)は 17 項目、パターン(3)は 16 項目にて被験者が感じた感情と同様の感情を読み取れることが分かった。

パターン(1)では文章(2)において全て違う感情を読み取る結果となった。また、文章(5)の「のどか⇔緊迫」の項目はパターン(2)では正しく読み取れていたが、パターン(3)では読み取れていなかった。これは、文章(5)が長くなったことによって、その前後の文章の影響を受けたからであると考えられる。このことから、前後の文章の有無は非常に大事であることが分かるため、文章から感情を読み取る場合にあまり短すぎても良くないということが分かった。全ての実験で違う感情となったのは、文章(3)の「楽しい⇔悲しい」の項目である。このようになった理由は、文章から判断し辛くどちらを選んでいいかわからないような文章だったからだと考えられる。被験者実験でも楽しいと選んだ人がやや多いがどちらにするか迷っていたことがうかがえた。

被験者実験による感情値の最小値から最大値までの幅が広いのに対して感情辞書によって算出された感情値の幅は狭い。本システムでは、ある程度の感情を読み取ることが出来ると分かったが、その強さまでを正確に表すことは出来なかった。

4. むすび

本研究では、小説の文章から読み取れる感情の推定システムを提案し、有効かどうかを検証した。読み聞かせロボットの活用化のためには、読み手の性格を考慮する必要がある。今回の実験では、被験者に大学生を用いたため、今後は様々なタイプの人物の読み手に対応するため、様々なタイプの人物による実験が必要であると考えており、今後の課題とする。

表 2 感情辞書とパターン(1)の比較.

文章	感情値算出元	軸 1	軸 2	軸 3
(1)	感情辞書	5.11	4.56	4.04
	パターン(1)	6.15	4.75	5.70
(2)	感情辞書	5.36	5.31	5.23
	パターン(1)	3.65	3.75	3.90
(3)	感情辞書	4.33	3.11	2.96
	パターン(1)	2.65	2.60	1.35
(4)	感情辞書	3.53	3.12	3.78
	パターン(1)	1.60	1.80	2.75
(5)	感情辞書	2.94	2.67	4.19
	パターン(1)	2.00	1.50	2.25
(6)	感情辞書	4.61	4.65	5.75
	パターン(1)	5.85	5.45	6.30

表 3 感情辞書とパターン(2), (3)の比較.

文章	感情値算出元	軸 1	軸 2	軸 3
(1)	感情辞書	5.11	4.79	4.01
	パターン(2)	6.60	5.15	5.40
	パターン(3)	6.30	5.50	5.40
(2)	感情辞書	5.28	5.06	5.34
	パターン(2)	5.00	4.20	4.05
	パターン(3)	5.40	4.45	4.60
(3)	感情辞書	4.52	3.00	3.03
	パターン(2)	3.60	3.15	1.80
	パターン(3)	3.05	3.10	2.25
(4)	感情辞書	3.44	3.14	3.62
	パターン(2)	1.70	2.05	2.15
	パターン(3)	2.05	2.15	2.10
(5)	感情辞書	2.99	2.75	3.62
	パターン(2)	2.70	1.65	2.95
	パターン(3)	2.70	1.85	4.20
(6)	感情辞書	4.80	4.20	5.60
	パターン(2)	5.65	6.05	5.95
	パターン(3)	5.85	5.70	6.15

文 献

- [1] 熊本忠彦, 河合由起子, 田中 克己, “新聞記事を対象とするテキスト印象マイニング手法の設計と評価”, 電子情報通信学会論文誌, Vol.J94-D, No.3, pp.540-548, 2011.
- [2] 松本好史, 張建偉, 河合由起子, 中島伸介, 熊本忠彦, 田中克己, “センチメント分析に基づくニュース記事の信憑性判断支援”, 第 3 回データ工学と情報マネジメントに関するフォーラム(DEIM 2011), 2011.