

物語の中の文における感情分類手法

A method to classify emotion in the sentences in novels.

吹越 弘崇†
Hiroataka Fukkoshi

杉本 富利†
Futoshi Sugimoto

1. はじめに

今日、音声合成技術の発達により、新聞やメール等の自動読み上げは非常に自然な合成音声で行うことが可能となった。しかし、感情表現が必要な朗読を自動で行うTTS(Text To Speech)システムは実現されていない。感情を表現する音声合成システム自体は森山らによって特定の感情ごとの音律データを準備することで実現されている[1]。このような技術を応用して感情付き自動朗読システムを構築するには、どの文をどの感情で読み上げるかを自然言語処理的に決定する必要がある。

本研究では、小型辞書並みの語彙を持つ、感情タグを付与した辞書を作成し、その辞書をもとに感情分類を行うことを試みる。

2. 感情分類手法

本システムで分類する感情はシュロスバークの円感情[2]から採用した「喜び」「怒り」「悲しみ」「驚き」「恐怖」「嫌悪」および「無感情」の7種である。分類手法は「主語と動詞の関係」から行うもの、「補語となる形容詞・副詞」から行うもの、慣用句から行うものの三手法を提案し、最終的な感情分類はこれらの三手法の結果を統合して行う。

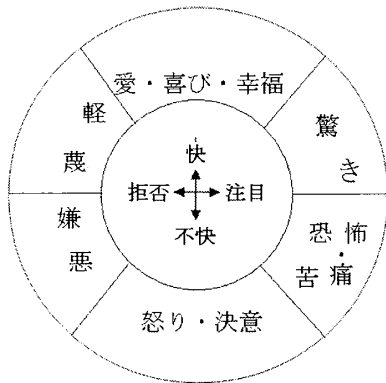


図1 シュロスバークの円感情

表1 「単文」への正規化手順

1. 全文の改行コードを削除する。
2. 全文の全角スペースを削除する。
3. 句点で単文を終了する。
4. 括弧および括弧内を削除する。
5. 鉤括弧で単文を開始する。
6. 鉤括弧で単文を終了する。
7. 句点-閉鉤括弧のつながりは単文とする。

まず、物語テキストを表1に示す手順で「単文」へ正規化する。次に、単文ごとに表層格解析を行う。表層格解析にはSynCha[3]を利用した。三手法全てで分類不能の場合は無感情とする。また、重文複文の場合に、複数の感情分類結果が得られた場合は、すべての感情値を総和する。

2.1 主語と動詞の関係に基づく感情分類

主語に「快・友好」「不快・敵対」「主・友好」「従・友好」「主」「従」という性質を与える。これは瀬谷[4]で提示されたものを使用した。この「主語の性質」と動詞との関係から感情を分類する。主語の性質は別に作成した辞書を参照する。主語となるものが文中にない場合は、無感情とする。

名詞性質辞書は503語の名詞とその性質をレコードに収録している。動詞辞書は分類語彙表[5]による類義語拡張を行い、15,612語の動詞に対し、各6種の性質との関係が収録されている。

表2 主語の性質辞書例

名詞	主	従	友好	敵対	快	不快	属性
ハエ	0	0	0	3	0	3	不快・敵対
桜	4	3	3	0	3	0	従・友好
犯罪	0	0	0	4	0	4	不快・敵対
生きがい	1	1	2	0	2	1	主・友好

表3 動詞動詞辞書例「睨む」

名詞の性質	動詞	喜び	怒り	驚き	悲しみ	恐怖	嫌悪	無感情
主	睨む	0	0	0	0	10	0	0
主・友好	睨む	0	0	0	0	10	0	0
従	睨む	0	0	0	0	0	10	0
従・友好	睨む	0	0	0	0	0	10	0
不快・敵対	睨む	0	0	0	0	10	0	0
快・友好	睨む	0	0	0	10	0	0	0

表4 形容詞・副詞辞書例

単語	喜び	怒り	悲しみ	驚き	恐怖	嫌悪	無感情
怖い	0	0	2	2	8	1	0
自由	8	0	1	0	2	0	1
-好き	9	0	0	1	0	0	1

表5 慣用句辞書例

慣用句	喜び	怒り	悲しみ	驚き	恐怖	嫌悪	無感情
悪態をつく	0	5	0	0	5	0	0
影で糸を引く	0	0	0	0	5	5	0
心が動く	0	0	0	5	0	0	5

† 東洋大学大学院工学研究科情報システム専攻

2.2 補語となる形容詞・副詞に基づく感情分類

表層格解析の結果から補語となる形容詞または副詞を選出し、形容詞辞書から感情の値を取り出す。形容詞辞書は990語の形容詞および副詞を収録している。

2.3 慣用句に基づく感情分類

文中に含まれる慣用句および連語と感情パラメータを記述した慣用句辞書を用意した。慣用句辞書は「日本語を使いさばく慣用句の辞典」[6]に掲載された1900語と実験用小説から発見した32語をあわせて1932語を収録している。慣用句には0から10の整数値で合計10になるように各感情に重みが割り振られている。本文中の慣用句を全文検索によって取り出して辞書に基づいて感情分類をおこなう。

$$E = \sum \left(\frac{C_{emo}}{C_{anc}} - \frac{S_{emo}}{S} \right)^2$$

E :誤差値

C_{emo} :アンケートで対象の感情を選択した人数

C_{anc} :総被験者数

S_{emo} :対象である感情の統合後における感情値

S :総合計感情値

3. 結果

被験者4名に、童話「白雪姫」[7]を読んでもらい、単文ごとに7種の感情のどれかに分類させ、感情ごとに分類した人数を数え、それを総被験者数4で割って、その値を一文の感情値とする。提案した手法による分類手法と被験者実験による結果の相違度を下の式によって計算した。計算結果は各手法ごとに表6解析結果表にまとめた。

解析に童話「白雪姫」を用いたのは、童話であり、表現が比較的容易であることと、作中にて喜怒哀楽の感情全てが登場するためである。

3.1 主語と動詞の関係に基づく感情分類結果

281文中191文で745回の解析を行った。うち感情分類を行うことができた解析が0回であった。すなわち、すべてが無感情に分類された。被験者によっても無感情と推定された文が多かった。平均誤差値が0.5465、最大誤差値が2.0000、最小誤差値が0.0000、誤差値の分散が0.3688であった。

3.2 形容詞・副詞に基づく行う感情分類結果

281文中191文で110回の解析を行った。平均誤差値が0.6113、最大誤差値が1.8200、最小誤差値が0.0850、誤差値の分散が0.0475であった。

表6 解析結果表

手法	検出数	平均	最大	最小	分散
主語と動詞	745	0.5465	2.0000	0.0000	0.3688
形容詞・副詞	110	0.6113	1.8200	0.0850	0.0475
慣用句	1	0.3125	0.3125	0.3125	0.0000
統合	856	0.6116	1.8200	0.0850	0.0474

3.3 慣用句に基づくから行う感情分類結果

281文中参照された慣用句は「悪態をつく」ひとつのみであった。この慣用句を含む文の誤差は0.3125であった。

4. 考察および今後の課題

全体として、解析の対象となった文の数が少なく、無感情と判定された文が多い。また、前後文脈を一切考慮に含まないため、特に情緒が強いと推測される「登場人物のセリフ」に対応ができない。

4.1 主語と動詞の関係に基づく感情分類についての課題点

主語の性質を辞書からの参照に頼るため、登場人物の固有名詞等に対応出来ない。また、同じ登場人物を指しているでも登場人物間の関係性から性質が変化することは十分にあり得る。解決策として「物語テキストにおけるキャラクタ関係図自動構築」[8]を利用することを考えている。また、辞書の語彙拡張を行うことを優先してきたため、各語の感情値における正確さの検証が不十分である。

4.2 形容詞・副詞に基づく感情分類についての課題点

辞書登録されている語が990語と少なく、類義語拡張等の余地がある。

4.3 慣用句に基づく感情分類についての課題点

辞書からの完全一致のみに頼っているため、発見性が低い。また、表記ゆれや誤用の対応が出来ない。解決策として[9]などを検索に組み込むことを考えている。また、辞書の拡張を優先してきたため、各慣用句の感情値における正確さの検証が不十分である。

4.4 統合性の課題点

今回は各手法の統合において、重み付け無しの総和を採用した。最適な重み付けの考案が必要である。

参考文献

- [1] 森山 剛, 森 真也, 小沢 慎治(2009):「韻律の部分空間を用いた感情音声合成」情報処理学会論文誌 Vol.50 No.3 pp1181-1191
- [2] Shlosberg, H(1952): "The Description of Facial Expressions in Terms of Two Dimension" Journal of Experimental Psychology, 44, pp.229-237
- [3] <http://cl.naist.jp/~ryu-i/syncha/> (Jan 2010).
- [4] 瀬谷 正敏(1977):「対人関係の真理」培風館
- [5] 国立国語研究所(2002):「分類語彙表」大日本図書出版
- [6] 現代言語研究会(2007):「日本語を使いさばく慣用句の辞典」あすろ出版
- [7] http://www.bauddha.net/snow_white/sanseido.html (Jul 2010)
- [8] 神代 大輔, 高村 大也, 奥村 学(2008):「物語テキストにおけるキャラクタ関係図自動構築」言語処理学会第14回年次大会発表論文集 pp380-383
- [9] 首藤 公昭, 田邊 利文, 高橋 雅仁(2009):「日本語の慣用的表現辞書について」情報処理学会研究報告 NL- Vol.2009-NL-191 No.3, pp1-7