

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

Experiment of Emotion Reasoning Focusing on Cause of Emotion from Diary Text

徳久雅人 村上仁一 池原悟

Masato Tokuhisa Jin'ichi Murakami Satoru Ikehara

鳥取大学 工学部 知能情報工学

Department of Information and Knowledge Engineering, Faculty of Engineering, Tottori University

1 はじめに

情緒推定のために情緒生起の原因を表す用言の知識ベース化が行われたが、その知識ベースの情緒推定の能力は評価が不十分であった[1]。本稿では、日記文を対象とした場合でのこの知識ベースの情緒推定の性能を評価する。

情緒推定の性能を評価する上で問題となる点は、(1)情緒推定の正解を決めること、および、(2)情緒推定の結果の解釈である。(1)に対して、本稿では[2]の情緒タグ付き日記コーパスを用いることとする。(2)に関して、情緒推定には、様々な言語的な情報を参照する必要があり、元々、用言の知識ベースだけでは人間と同レベルの情緒推定は実現できない。そこで、本稿では、情緒推定の誤り傾向を分析し、誤りの原因の所在を明らかにするとともに、さらに必要となる言語的な情報を考察する。

2 情緒生起の原因を表す用言の知識ベース

情緒生起の原因を表す用言の知識ベースとして[1]を使用する。この知識ベースは、日本語語彙大系[3]の結合価パターンと[4]の情緒生起の特徴ラベルを基にして作成されたもので、14,812パターンに、情緒属性が11,712セット対応付けられている。情緒属性は、情緒名、情緒主、情緒対象、原因で構成されている。情緒名は8種類の情緒《喜び》、《悲しみ》、《好ましい》、《嫌だ》、《驚き》、《期待》、《恐れ》、《怒り》が使われている。以下の例では、結合価パターンに(a)と(b)という2つのセットが対応付けられている。このようなパターンのマッチすることで、文が情緒生起の原因を表すかどうか判断できる：

- M1(3主体)がN2(3主体)と話が合う
- (a) 情緒名:《喜び》, 原因:《同意》
情緒主: M1, 情緒対象: N2
- (b) 情緒名:《喜び》, 原因:《同意》
情緒主: M1, 情緒対象: N2

3 情緒タグ付き日記コーパス

情緒推定のテストデータとして、情緒タグ付き日記コーパス[2]を使用する。この日記は、図書[5]の日本語部分を収録している。情感で章立てがされており、情緒の種類をカバーするように実験ができる。

本コーパスでは、節の単位で文が分割されており、その単位で日記の作者の情緒がタグで明示されている。情緒タグの付与は、情緒タグ付与の経験者6名と、初心者1名による。なお、コーパスの規模は以下のとおりである：

- 日記のストーリー: 268話
- 分割後の節: 1,642節
- 情緒タグ(経験者6名ぶん): 11,461個

4 情緒推定の実験

4.1 実験の目的と方法

本知識ベースの性能評価を目的とする。知識ベースを運用

する方法に対する評価は目的とはしない。したがって、知識ベースに定義された情報を手動で運用して情緒を推定する。

また、本知識ベースで情緒推定の対象となる文は、(1)情緒主が登場する文、(2)情緒生起の原因を表す文(情緒状態を表す文、情緒反応を表す文は除外)、(3)情緒主に起こった事態を表す文(否定文、疑問文、未来・願望の文、条件・仮定の文は除外)、であり、そのような文に対して実験を行う。

情緒推定の正解は、経験者6名のタグの多数決により決める。つまり「正解の情緒」とは、「日記から読み取られる作者の情緒」である。具体的には4名以上の一致かつ上位2位までを「正解の情緒」とする。それ未満は「少数意見の情緒」とする。

4.2 情緒推定の様子

日記の一例を示しながら、情緒推定の様子を説明する。

- 文1) 蒲郡のおばさんからみかんが1箱届いた。
文2) とても甘くて
文3) おいしかった。

文1)は「みかんが(日記の作者へ)届く」という事態であるので、実験対象の文である。文2)および文3)は、情緒主が登場しないので、対象外である。文1)への情緒タグは、《喜び》4名、《好ましい》2名、《なし》1名と付与されたので、「正解の情緒」は《喜び》、「少数意見の情緒」は《好ましい》と《なし》である。一方、「届く」のパターンより、「情緒名:《喜び》, 原因:《獲得》, 情緒主: N2, 情緒対象: M1」が得られ、正しく情緒が推定できる。

4.3 実験結果

実験対象となる文の中で、「正解の情緒」の得られた節は、215節であった。その範囲で、「正解の情緒」と「推定された情緒」の一致した数より、性能を測る。「一致率」は次式で求める：

$$(\text{一致率}) = \frac{(\text{一致した数}) \times 2}{(\text{正解の情緒数}) + (\text{推定された情緒数})} \times 100 (\%)$$

結果は表1となった。この結果より、次のことが言える：

- 本知識ベースは、初心者と同程度の性能である
- 経験者でさえ、情緒推定の一致率は高くはなく揺れがある

表1 情緒推定の一致率

推定者	一致率	一致数	正解数	推定数
知識ベース	48%	108	225	224
初心者	50%	112	225	221
経験者	52 ~ 61%	158 ~ 198	297 ~ 320	292 ~ 352

※ 経験者は、自身を除く5名の多数決で3名以上の一致のある情緒を正解とした

5 誤り事例の分析

まず、荒く誤り傾向を分析すると次のようになった：

- 結合価パターンがマッチしない (24節)
- 少数意見に一致 (33節)
- 曖昧性による不一致 (34節)
- 全くの不一致 (35節)

曖昧性による不一致とは、正解の情緒や推定された情緒が複

数存在する場合である。ここには部分的に一致するものがある:

(例1) あ的美容師に髪を台無しにされた。

正解の情緒:《怒り》,《嫌だ》, 少数意見:《悲しみ》
／推定された情緒:《悲しみ》,《嫌だ》,《怒り》

判定の基準を緩くする(少数意見も正解とする, 曖昧な場合は部分一致も認める)と, 215節のうち144節が正答となる。その残りについての誤り事例について言語表現上の特徴を分析し, 次のように分類した:

- 動作の内容を考慮する必要がある (21節)
- 主体的な表現を考慮する必要がある (10節)
- 文脈を考慮する必要がある (6節)
- パターンへの対応付けの誤り (5節)
- テンス・アスペクト等を考慮する必要がある (3節)

それぞれについて, 具体例をもって説明する。

5.1 動作の内容を考慮すべき場合

動作の内容が日常的な行為である場合がある:

(例2) 駅までタクシーを利用した。

正解の情緒:《なし》／推定された情緒:《好ましい》

(例3) 夕食の準備ができ、

正解の情緒:《なし》／推定された情緒:《期待》

強いて言えば推定された情緒のとおりかもしれない。動作の日常性や情緒の強さの検討が必要である。

格要素の名詞や修飾表現に着目すべき場合がある:

(例4) 家族で回転寿司へ行った。

正解の情緒:《好ましい》／推定された情緒:《なし》

(例5) ロマンチックな一夜を過ごした。

正解の情緒:《好ましい》／推定された情緒:《好ましい》,《嫌だ》
「(物理的な移動の)行く」には, 情緒属性が対応付けられていなかった。(例4)では, 「回転寿司」という場所や「家族」という同伴者から正解の情緒が推定されたものと思われる。関連研究[6],[7]のように名詞ごとの極性や共起関係を取り入れる必要がある。(例5)では修飾表現の極性で, 推定される情緒の曖昧性を解消するという方法が考えられる。

5.2 主体的な表現を考慮すべき場合

日本語の言語表現は, 詞と辞の関係として知られているように客体的表現と, 主体的表現に分けられる[8]。主体的表現には話者の情緒を読み取る手がかりがある。

(例6) 今日友達と行ってみた。

正解の情緒:《期待》／推定された情緒:《なし》

(例7) 壊してしまった。

正解の情緒:《悲しみ》／推定された情緒:《喜び》,《悲しみ》

(例6)では「てみた」より話者の試行的動作と読み取れるため,《期待》と判断された。(例7)では, 「てしまった」は残念な結果に対して使われる表現であるので《悲しみ》と判断された。しかし, 知識ベースでは, 「壊す」が意図的な表現で, その完了時において(完遂)による《喜び》が生じると考えられて登録されていた。この登録は誤りではないが, 結果として過剰推定となった。

(例8) やっと仕事が見つかった。

正解の情緒:《喜び》／推定された情緒:《喜び》,《驚き》

「やっと」には成立の難しいことが成立したという意味がある。偶然の発見というよりも, 努力の結果見つかったという解釈から,《驚き》の推定は正解ではない。

以上より, 情緒主が, 登場人物かつ話者の場合, 主体的な表現を用いて情緒の曖昧性を解消したり, 過剰な推定を抑制した

りすることが, 誤りへの対策として挙げられる。

5.3 文脈を考慮すべき場合

文脈とは, ここでは, 複数の文(節)の関係を指す。

(例9) 1日中、頭がガンガンしていた。休みをとりたい気持ちは山々だったけど、今が一番忙しい時期だし、同僚に迷惑をかけたくなかったので、とにかく仕事に行った。

下線部の正解の情緒:《嫌だ》／推定された情緒:《なし》

「仕事遂行」と「健康維持」の2つのゴールにおいて「(頭痛の際)仕事に行く」というプランは, コンフリクトを生む。この文では, 接続詞を参考にして, 関係が解析できるが, 一般には, 接続詞が明示されない場合がある。

5.4 テンス・アスペクト等を考慮すべき場合

用言の持つテンス・アスペクト的な側面や, 後続するテンス・アスペクトの表現の影響がある。

(例10) 去年の夏、日光に肌をさらし過ぎたかも・・・。

正解の情緒:《悲しみ》,《嫌だ》
／推定された情緒:《期待》,《恐れ》

「肌をさらす」には, その結果を見越して《期待》や《恐れ》が対応付けられている。知識ベースを構築する際, 現在の時制を基準にしたためである。日記文では, 「肌をさらした」と過去になっているため, その結果が既に得られており《恐れ》ではなく《悲しみ》や《嫌だ》が正解の情緒となっている。

6 おわりに

本稿では, 情緒生起の原因を表す用言の知識ベースを用いて, 日記文からの情緒推定を実験し, その誤り事例の分析により, 情緒推定に必要な言語的な情報を考察した。

本知識ベースは情緒分析の初心者程度の性能であり, 知識ベース自体の不具合ではないことが確認できた。

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] 田中努, 徳久雅人, 村上仁一, 池原悟: 情緒生起情報付き結合価パターン辞書の開発, 言語処理学会第12回年次大会発表論文集, pp.1151-1154, 2006.
- [2] 徳久雅人, 松浦大樹, 村上仁一, 池原悟: 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] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙大系, 岩波書店, 1997.
- [4] 徳久雅人, 岡田直之: パターン理解的手法に基づく知能エージェントの情緒生起, 情報処理学会論文誌, 39(8), pp.2440-2451, 1998.
- [5] 石原真弓: 英語で日記を書いてみる, 表現集編, ベレ出版, 2002.
- [6] 目良和也, 市村匠, 相沢輝昭, 山下利之: 語の好感度に基づく自然言語発話からの情緒生起手法, 人工知能学会論文誌, 17(3), pp.186-195, 2002.
- [7] Liu, H., Lieberman, H., Selker, T.: A Model of Textual Affect Sensing using Real-World Knowledge, IUI2003, pp.125-132, 2003.
- [8] 時枝誠記: 国語学原論, 岩波書店, 1941.