

感情表現のモデル

中山 記男^{†‡} 江口 浩二^{‡†} 神門 典子^{‡†}

[†]総合研究大学院大学 情報学専攻 [‡]国立情報学研究所

E-mail: [†] norio@grad.nii.ac.jp, [‡] {eguchi, kando}@nii.ac.jp

1. はじめに

インターネット上での商取引やブログの増加により、特定の商品や出来事についての個人の主観的な態度（以下：態度）を表明したテキストが増加している。この態度の表明は、企業の商品開発へのフィードバック、特定の施策に対する国民の反応を知る手がかりなどとして、注目を集めている。さらに 2004 年春には AAAI でテキスト中の態度や感情の抽出に注目したシンポジウム[1]が開かれ、その成果が単行本として出版されることから注目の高さが伺える。本研究では、このようなテキストに現れた個人の態度の表明の中でも、特に、「うれしい」「かなしい」などの個人の感情を表す感情表現に着目し、その特性を理解するためのモデルを提案する。感情とは、ある対象に対する主体の気分や心の動きであり、感情表現は感情を表明した部分のテキストである。WEB 上の 82 の書評について人手で感情表現を抽出し、「態度」「主体」「対象」「理由」の 4 つの軸から構成される多次元モデルに基づいて分析をした。加えて他の態度の表明を表す表現との比較、感情表現の自動抽出への方策を考察した。

2. 関連研究

テキストから態度を抽出する研究は、態度の種類によりいくつかに分類される。

大塚[2]は自由記述型アンケートに記述された書き手の要求意図に着目している。要求意図の要素に着目し、その要素が現れるパターンを用いることで直接「～してほしい」などの表現で記述されていない要求意図の抽出に成功している。意図が表明されている部分を手作業で抽出し、新たな手がかりを発見することで意図要求のモデルを作成している。

立石ら[3]、小林ら[4]は、Web 上に記述された商品の評判に着目している。Riloff ら[5]は新聞に記述された主観的な態度に注目している。これら研究は「商品 A は評判 B」のようなパターンを元として、ブートストラップにより評判 B にあたる表現を抽出している。また、特に立石らは商品の属性として適切な評価軸を作成することで、商品の評判を図示している。Riloff らは抽出された表現を強い表現と弱い表現に分類している。これらの研究は態度の自動抽出をブートストラップ法にて実現している。

Liu[6]らは、常識知を用いてメールの本文テキスト中に含まれる感情を抽出し、あらかじめ決められた感情カテゴリーへ分類することでテキストの感情を理解することに注目している。田中ら[7]は、結合価パターンを用い、日記中の文の情緒推定を行うことで、辞書へ情緒生起情報付与をすることに注目している。感情語によって感情が記述されていなくとも、テキストには感情が存在するとの前提に基づいている。

本稿では、大塚や Liu ら、田中らと同様に、明示的に態度が表明されていなくても抽出できる手法を検討している。

3. 方法

Web 上の書評からなるコーパスに対し、感情表現の構成要素を定義し、その要素についてタグ付けを行い、感情表現のモデルについて検討した。

3.1. コーパス

コーパスは、杉田ら[8]が 2000 年 10 月に収集した書評に関するコミュニティに属する 332 の Web サイトである。本稿では、その中から 6 サイトを無作為に選択し、さらに各サイトから書評が記述されたファイルを 1 つずつ選択した書評サブコーパス 1 を実験に用いた（表 1 参照）。書評サブコーパス 1 の計 82 個の書評（うち 18 個は映画評）、計 1528 文に対し人手で感情表現のタグ付けを行った。6 サイトおよびその書き手は重複しない。ここで書き手とは、そのサイトでテキストを記述した人物である。書評は 1 冊の図書などの対象について連続して書かれたものを 1 つとした。文字数には記号とスペースを含む。

表 1 書評サブコーパス 1 の詳細

ファイル ID	書評数	文数	平均文字数	最大 / 最小文字数	文字数 (計)
274_20	17	498	676	1180 / 429	11493
103_72	13	295	537	824 / 480	6982
011_29	18	226	318	1153 / 18	5729
092_9	15	175	228	442 / 82	3432
154_61	1	119	1691	—	1691
156_30	18	215	295	630 / 100	5319
計	82	1528	—	—	34646

3.2. 実験

82 書評に対して以下の処理を行った。まず態度が表明されている部分に注目し、タグ付けした。1 つの態度の表明を抽出したら、続けて関連する構成要素を検討した。新たな要素が発見されるごとに、その要素をモデルに加えて既に分析済みのコーパスも新たなモデルでタグ付けをやり直した。この繰り返しによりタグ付けする要素を決定した。その結果、「態度」「主体」「対象」「理由」4 つの上位要素とその下位要素に着目した。図 1 にタグ付け作業の手順を示す。

以下、タグ付け作業の詳細とタグ付け要素について説明する。「態度」に関しては、感情だと判断された場合「感情」、感情以外であると判断された場合「感情以外」、感情かどうかの判断が難しい場合は「保留」とした。

「態度」以外の要素は、テキスト中には明示的に要素が記述されていないが、省略されていると理解できる場合は、「ゼロ（主体、対象）」とした。テキスト中に記述されてなく、かつ該当するものが無い場合は「不明」とした。加えて全てのタグ付けが終わった後に、タグ付けされた部分の属性に応じて分類を行っている。着目し、タグ付けした構成要素とその定義については表 2 にまとめた。

タグ付けと 4 つの要素の判断は一名の作業者が行った。タグ付けに際しては開始時と終了時の判断の差が発生しないよう、終了後にもう 1 度索見直しを行った。見直しでは、各要素を「主体は対象を（が）理由であ

るから（なので）態度と考えた」と言い換えて特に矛盾がないか確認した。

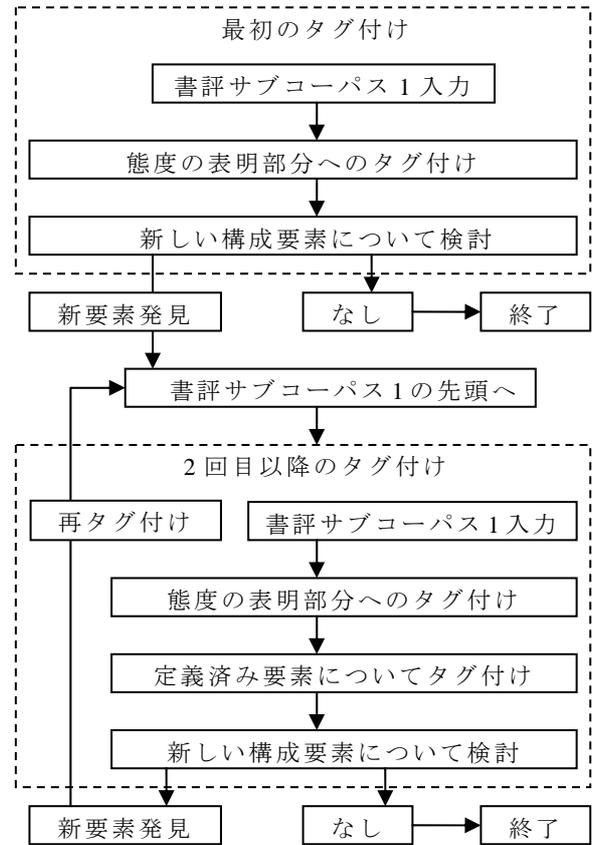


図 1 タグ付け作業の手順

表 2 着目した構成要素の定義

上位要素	上位要素の定義	下位要素	下位要素の定義	下位要素の分類
態度	態度の表明	感情	態度は感情である	
		感情以外	態度は感情ではない	
		判断保留	態度が感情であるかどうかの判断が難しいため保留	
		態度なし	態度は表明されていない	
主体	態度を表明した人物, モノ	主体記述	主体が記述されている	書き手, 登場人物, 第三者
		ゼロ主体	主体が省略されている	
		主体不明	主体が不明である	
対象	態度が向けられた人物, モノ	対象記述	対象が記述されている	登場人物, 場面, 様子, ストーリー, 第三者, 発言, 心象, 事実, 行動
		ゼロ対象	対象が省略されている	
		対象不明	対象が不明である	
理由	態度を表明した原因, 理由	理由記述	理由が記述されている	根拠, 原因, 理由
		理由不明	理由が不明である	

4. 結果

書評サブコーパス 1 にタグ付けした結果、「態度」がタグ付けされたのは全部で 420 箇所であった。この

420 箇所は、特定の対象に対する態度であり、文の内容に対するモダリティは含んでいない。420 箇所における上位要素の組み合わせを表 3 に示す。加えて、同

じく 420 箇所における下位要素の組み合わせにおいて、全 105 種類の組み合わせ中、組み合わせ出現数の多かったもの 17 種類を表 4 に示した。なお、下位要素の組み合わせについては「理由」のみ分類を行っていない状態で数を計算している。

4.1. 入れ子構造

表 3 における例 S5 では、「対象」である「サスペンフルな物語の構成」における「サスペンフル」も 3.2 節で示した定義に当てはまってしまう。同様にして、上位要素の中に入れ子構造でさらに上位要素が確認される事例があった。本稿では、入れ子構造が確認された場合、最も外側の態度のみをタグ付けした結果を示している。

4.2. モダリティに関する検討

今回のタグ付けでは、表 3 の例 S5 における「としかいえません」や S8 における「だけでしょか」、これには記号や顔文字などで表現される、文の内容に対するモダリティを考慮していない。今回タグ付けされた態度とこれらモダリティのパターンを比較することは、今後の課題である。

4.3. モデルの構成要素

本節では、実験にて着目したモデルの構成要素について分析する。表 3 で示すように、もっとも多くタグ付けされた上位概念の組み合わせは

[態度あり]+[主体あり]+[対象あり]+[理由不明]

で 251 件あった。「主体」については 420 件中 419 件でタグ付けされており、「対象」の 321 件、「理由」の 72 件に比べて多くタグ付けされていることがわかった。「理由」は上位要素ではもっとも少なかったが、72 件のうち 70 件が「対象」と同時にタグ付けされていた。「対象」の分類として書評の対象書籍そのもの以外では登場人物やストーリーについての記述が多かった。「理由」に関しては、「態度」が感情であるときよりも保留であるときに多く組み合わせられていることがわかった。

4.4. モデルに基づくコーパスの分析

本実験が分析したコーパスでは、書評のタイトルとして書籍の名前が記述され、その後のひとかたまりで書評が記述されていることが多かったため、「対象」が省略されている、すなわちゼロ対象である場合が多かった。また、「書き手」以外の「主体」がタグ付けされた例は全部で 17 件あったが、うち 16 件は登場人物の感情を記述したものであり、あらすじの説明である場合が多かった。

4.5. 考察

「主体」の下位要素である「主体記述」は全部で 8 件あり、そのうち 6 件が「書き手」と分類された。この 6 件全てで書き手が自身の態度が特殊、または読み手と一致しないかもしれないという意図が読み取れたことから、固有の意味を持つ表現であると考えられた。また「理由」は「対象」との組み合わせが多く、「主体」「対象」と同じ箇所にタグ付けされることもあり、他の要素との関係が強いと考えられた。「理由」と他の要素との関係は、さらなる分析が必要である。今回の実験は「態度」に対して他の要素をタグ付けしたため、今後は他の要素に対してのタグ付けとの比較も必要である。加えて、入れ子構造やモダリティ、「保留」とタグ付けた「態度」に対しての検討も課題である。

5. おわりに

本研究では、Web 上の書評に対し構成要素をタグ付けすることで、感情表現のモデルについて検討した。今後は、筆者が別報告にて提案した感情表現の自動抽出方法[9]などと組み合わせることで、自動抽出への応用を検討する。また、今回の実験が筆者だけによって行われているため、別の被験者による実験も行っていく予定である。

文 献

- [1] AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, Stanford, CA, USA, March 2004.
- [2] 大塚裕子, 自由記述アンケートと回答の意図抽出および自動分類に関する研究, 神戸大学大学院自然科学研究科博士論文, 2004.3.
- [3] 立石健二, 石黒義英, 福島俊一, インターネットからの評判情報検索, 人工知能学会誌, Vol.19, Num.3, pp.317-323, 2004.
- [4] Kobayashi, N., Inui, K., Matsumoto, Y., Collecting Evaluative Expressions for Opinion Extraction. Proceedings of the 1st International Joint Conference on Natural Language Processing (IJCNLP-04) pp.584-589, 2004.
- [5] Riloff E., Wiebe J., Wilson T., Learning Subjective Nouns using Extraction Pattern Bootstrapping, Proceedings of the Seventh Conference on Natural Language Learning (CoNLL-2003), Edmonton, Canada, May 2003.
- [6] Liu, H., Lieberman, H., Selker, T., A Model of Textual Affect Sensing using Real-World Knowledge. To Appear in Proceedings of IUI 2003, 2003.
- [7] 田中努, 徳久雅人, 村上仁一, 池原悟, 結合パターンへの情緒生起情報の付与, 言語処理学会第 10 回年次大会発表論文, pp.345-348, 2004.
- [8] 杉田茂樹, 江口浩二, 目録データベースと Web コンテンツの統合的利用方式, 情報処理学会研究報告 情報学基礎, Vol.2001, Num.20, pp.153-158, 2001.
- [9] 中山記男, 江口浩二, 神門典子, 感情表現の抽出手法に関する提案, 情報処理学会研究報告 自然言語処理 Vol. 2004 Num. 108 pp.13-18, 2004.11.

