

食卓日記データを用いた書き手の感情推定に関する分析

北嶋弓月[†] 大竹恒平[‡] 生田目崇[§]

中央大学大学院[†] 東海大学[‡] 中央大学[§]

1. 背景

近年、生活者のニーズや嗜好などを把握することを目的に、生活者が記述した文章を対象とした自然言語処理解析が様々な分野において行われている。自然言語処理解析の代表的な分析の一つとして、文章中出现する単語や文脈を用いた感情推定が挙げられ、ユーザが文章を記載した際の心情の把握に繋がり、マーケティング活動において有益な知見となりうる。多くの文章には書き手の複雑な感情が含まれており、従来のポジティブ・ネガティブという単純な2極の感情だけではなく、例えば喜怒哀楽のように、より複雑な感情を考慮することも生活者理解のための重要な情報となりうる。先行研究として、武内ら[1]は中村[2]による感情表現辞書を元に、ニューラルネットワークによって一般単語の感情の値を推定した。こうした方法により、単語だけではなく実際に記述された文章の感情を推定することが可能であると考えられる。

本研究では、食生活に関する約3万件の日誌データを用いて、文書構造から文献[2]の10感情（喜、怒、哀、怖、恥、好、厭、昂、安、驚）の推定を試みる。具体的には、感情表現辞書を作成し、書く文章に含まれる感情語を抽出する。さらに、各感情語の感情値から、各文章の感情値を算出し、書き手の感情を推定する。

2. 使用データ

本分析では、不特定モニタの食卓日誌データを用いて分析を行った。データ期間は1990年～2014年である。使用データ内には場面における食材選択行動に関係した文章が記述されている。また、実際に用いた食材、料理についての記載も存在する（表1）。

表1. データ例

日付	季節	文章	材料
1990/1/1	冬	今日はお正月...	おせち, ...

3. 感情辞書の作成と感情値の付与

本分析では、先行研究を参考に感情辞書の作成を行い、各感情語に対する感情値を求めた。

まず、感情表現辞書に記載されている10感情の感情語（表2）を元に、各感情語の類似単語を抽出し、感情辞書に追加した。具体的にはまず、日本語版 Wikipedia の文章中単語を用い、Word2vec [3]による単語の分散表現によって各単語の値を算出した。次に感情語と類似度の高い単語上位30語を感情語ごとに抽出し、これを新たな感情語として感情辞書に追加した。なお、本分析では、学習済みのデータを用いた。

表2. 感情語の例

感情	感情語例
喜	目出度い, 喜び, 愉悦, ...
怒	怒り, 憤怒, むかむか, ...
哀	悲観, 泣く, 冷たい, ...
怖	怖い, 恐れ, 震える, ...
恥	恥じらう, 羞恥, 赤らめる, ...
好	高鳴る, 大好き, 惹かれる, ...
厭	不快, 悔しい, 迷い, ...
昂	ときめく, 上がる, 高ぶる, ...
安	安心, のんびり, 楽, ...
驚	驚く, 仰天, 動揺, ...

次に、作成した感情辞書の各感情語に対して感情値を付与した。感情値とは、その感情語が10感情をそれぞれの程度所有しているかの度合を表す値である。感情値は最大1、最小0とし、感情表現辞書内の単語については当てはまる感情に1を付与した。また、追加した Wikipedia による類似語については Word2vec による各感情語との類似度を付与した。感情表現辞書に記載されている感情語2,063語と追加した感情語13,463語により、最終的な感情辞書の感情語数は15,526語となった。

4. 文章ごとの感情の推定

次に、作成した感情辞書を用いて各文章の感情値の推定を行った。ある文章 T 内のある感情値 T_s は以下のように算出されるものとする。

$$T_s = \frac{\sum_1^n v}{N}$$

ここで、分母 N は文章 T 内に含まれる感情語数を表し、分子 $\sum_1^n v$ は文章 T 内に含まれる感情語中の、感情 S に対する感情値 v の総和を表す。文章の感情推定の結果例として、図1に「幼稚園の運動会のお弁当は何にしようかな。遠足など行事が続いたので考えてしまう。子供の意見をとりに来て好きな物をいれよう。」という文章に対する感情値を表した。この結果から、この例文では、「喜、好、昂」の3感情が推定され、

Analysis of Human Emotion Estimation Using Food Diary Data
Yuzuki KITJIMA[†] Kohei OTAKE[‡] Takashi NAMATAME[§]
Graduate school of Chuo University[†] Tokai University[‡]
Chuo University[§]

文章が全体的にポジティブと考えられる感情を持っていることがわかる。

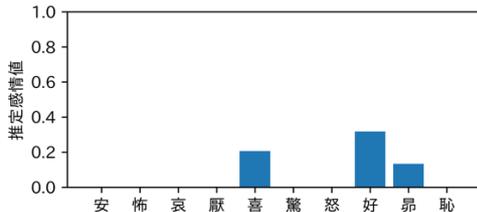


図1. 推定した感情値の分布例

5. 推定した感情の検証

本研究では、感情辞書から推定した文章の感情が妥当かの精度を評価するために、20代前半の大学生・大学院生 23 人を対象にアンケートを行った。アンケートには使用データ内で実際に記述されている 30 の文章に対して 10 感情の中で当てはまると思う感情を最大 3 つまで選択してもらった。アンケートの結果を正解データとし、モデルにより推定された感情とアンケートの正解データが一致した場合の正解率は 0.63 であった。30 文章の中で最も正解率の高かった文章は「朝、御飯をたきすぎてしまう。あまり長くジャーに入れておくと黄色っぽい御飯になりそうだ。お昼は、これをなんとかしなくては。」であり、最も低かった文章は「土曜なのに午前中も午後も用事が出来てしまった。一度家に帰って家で昼食をとるつもりだったが、用事が長びき外食することにした。」であった。

6. 考察

図 2, 3 はそれぞれ 5 節の検証の中で最も正解率の高かった文章と最も低かった文章について、推定感情値分布（左）と、アンケート結果による各感情の選択率（右）を比較した図である。

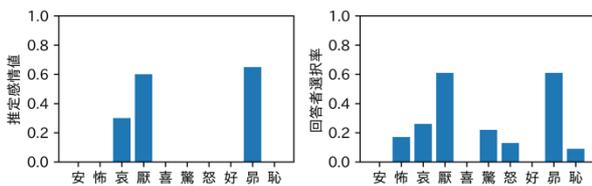


図2. 正答率の高い文章の分布

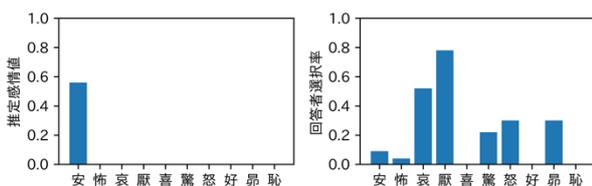


図3. 正答率の低い文章の分布

5 節の結果より、30 文章かつ感情選択肢が 10 感情であることを考慮すると、正解率は低くはないと考えられる。実際に、図 2 の正解率が高かった文章では、アンケート結果上位 3 感情と感情辞

書によって推定された 3 感情がどちらも「昂, 厭, 哀」と一致しており、人間が文章から読み取る感情をモデルが推定することができている。

一方、図 3 ではアンケート結果で「厭」の選択率が最も高かったのに対し感情推定によって推定された感情は「安」であった。実際に、各文章の感情推定で用いられた感情語を確認すると、図 2 の文章では「何とかする」「～ぼい」と特徴を捉えられているのに対し、図 3 の文章では「帰る」のみで推定を行っていた。このことから、複数の感情語が含まれている文章については推定を行うことが可能だが、単語間の係り受け関係といった複雑な構造を持つ文章の感情推定はうまくいかない傾向にあると推察できる。

また、文章全体の傾向として、「喜, 好, 昂, 厭」と推定された文章は正解になりやすい傾向にあるのに対して、「恥, 驚, 怖」が感情として推定された文章は正解しにくい傾向であった。これは、元々の言葉が持つ感情と日常で使われている言葉の感情が異なることが影響しているのではないかと考えられる。

7. まとめと今後の展望

本分析では、喜怒哀楽だけではなく、人間のより複雑な感情の推定が可能であるという仮説から生活日誌データを用い、感情辞書を作成した後、データの各文章が所有する 10 感情の感情値の推定を目的とし、分析を行った。

分析では、各文章の感情値推定を行った。また、検証としてアンケートを取り、アンケート結果を正解データとし、推定した感情との比較を行った。分析の結果、アンケート結果に対する感情推定の正解率は 0.63 であった。アンケートと推定した感情の比較から、「喜, 好, 昂」の感情は正解率が高い反面、「恥, 驚, 怖」の感情は正解率が低い傾向にあった。

今後の課題・展望として、本分析で学習データとして用いた日本語版 Wikipedia は日記のような私的な感情を含んで書かれた文章ではないため、データの特性を反映している日記やブログなどから学習を行うことでモデルの改善につながるのではないかと考える。また、食材選択行動のデータ特性を活かし、料理単位、材料単位での感情推定を行うことで感情による料理選択行動を明らかにできると考えられる。

参考文献

[1] 武内達哉, 萩原将文, “単語の持つ感情推定法の提案と単語感情辞書の構築”, 日本感性工学会論文誌, 2019
 [2] 中村明, 感情表現辞典, 東京堂出版, pp. 5-20, 1993
 [3] T. Mikolov, K. Chen, G. Corrado and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” *Ninth International Conference on Learning Representations*, 12 pages, 2013