

SNS上の文章を用いた感情辞書構築の有効性検証 Emotion Dictionary Construction Using SNS Text

大澤 泰我[†] マッキン ケネスジェームス[†] 永井 保夫[†]
Taiga Ohsawa Kenneth J Mackin Yasuo Nagai

1. 序論

コミュニケーションロボットの需要が高まる中,人と対話システムの自然なコミュニケーションの実現が求められている.人と対話システムの自然な対話の実現において,感情情報は必要不可欠である.本研究では,深層学習を利用してテキストから感情を判断している[7][8][9].データセットは,SNS上から顔文字付きの文章データを収集し,その顔文字を正解ラベルとしている.このデータセットを学習することにより感情推定モデルを作成している.本研究で利用している顔文字をもとに集められたデータセットにおいて,文章データの正解ラベルである感情表現が,人の判断と一致しているのかが不明確という問題がある.そこで,感情を含む文章を実際に人が分類し,文章データの正解ラベルと人が判断した感情が一致しているかの検証を行った.

2. 関連研究

Ptaszynski[1]らによる感情的要素/感情表現分析システム ML-Ask は,約2,100語の辞書によるパターンマッチングにより求められた「喜,怒,哀,怖,恥,好,厭,昂,安,驚」の10クラスの感情モデルを採用している.Ptaszynskiらの感情的要素/感情表現分析システムは,感情表現辞書を使って感情を判断しており,感情推定精度は57.6%となっている.感情表現辞書は,感情表現辞典[2]を元に,3人の専門家にアンケートを取り,平均をとって定義された感情語や感情をあらわす単語が収録されている.

山本[3]らは,オリジナルの辞書を使い感情を判断しており,感情判断精度は55.0%となっている.オリジナルの辞書は,学生20人のアンケートを元に,文章毎の感情が定義されている.

山下[4]らは Plutchik の感情の輪[5]における基本感情を感情モデルとして用いている.この感情モデルは8つの基本感情があり,各感情がペアとなって互いに反対の感情を持つという特性がある.

本研究では,この Plutchik の感情の輪を元に8つの感情分類(喜び,信頼,恐れ,驚き,悲しみ,嫌悪,怒り,期待)を,提案システムの感情モデルとして採用する.

3. 感情判断における感情辞書構築

本研究は文章から感情を判断する研究を行っている.文章から感情を判断するためには,Ptaszynskiらの感情表現辞書や山本らのオリジナル辞書などのように,感情を含んだ単語や文章がどの感情に当たるのかを定義したものが必要となる.我々は,SNSから顔文字付きの文章を集め,顔文字に対応する感情表現を正解ラベルとし,顔文字を除いた文章データとともに記録したデータセットを感情辞書としている.

感情判断では,この感情辞書を元に深層学習を用いて,作成された感情推定モデルに基づき感情表現が求められる.感情辞書の品質は,感情推定モデルの精度に直接現れるため,感情辞書の構築は本研究にとって非常に重要である.

我々は, Twitter APIを利用してTwitterの文章を取得し,表1に示される顔文字付きの文章を取得している.

感情辞書は,取得したTwitterの文章の顔文字から求められる感情表現と,顔文字を除去したTwitterの文章のペアから構築される.

表1 感情表現の決定に利用した顔文字

喜び	(≧▽≦) (^^)
信頼	ε-(´▽`*) (;´?´)
恐れ	ｶﾞｸｶﾞｸ(((;´Д´)))ﾌﾞﾙﾌﾞﾙ Σ(°△°)
驚き	Σ(=´ω´=;)マヅ!? Σ(´Д´)ﾌﾌ!?
悲しみ	(/._.)ゞｸｼｸ (ρ´n´)ｸｽﾝ
嫌悪	(´_`´)ﾌｰﾝ (/._.)ﾊｱ…
怒り	むっ… (´_`´)# ムムム…((o(´_`´);)
期待	((o(´▽´o))わくわく ♪((O(ω▽ω)O))♪ﾄﾞｷﾄﾞｷ

感情辞書構築では,顔文字をもとに決定された感情表現が人の判断により決定された感情表現と一致しているのかが不明確という問題がある.

本論文では,取得されたTwitterの文章の顔文字を元に集められた感情表現と,実際に顔文字を除去したTwitter文章から人手により決定された感情表現が一致するかを検証する.

4. 感情辞書の有効性検証

感情判断において, Twitter文章の顔文字をもとに決定された感情表現(正解ラベル)が人の判断により決定された感情表現と一致しているのかが不明確という問題がある.そのため本論文では,感情辞書に収録されている文章データから実際に人手により分類・決定された感情表現と感情辞書の文章データの感情表現(正解ラベル)とが一致しているかを検証する.

4.1 検証手法

検証に使う文章データは,感情辞書から各感情表現につき2件ずつランダムに文章を選ぶ.

選ばれた文章を人手により感情表現し決定してもらうため,正解ラベルを伏せた状態で,アンケート調査を行った.

表2の文章データ16件をアンケートに記載し,それぞれの文章がどの感情表現に当たるのかを決定してもらい,集計した.集計したアンケートの回答である感情表現と各文章データに対応する感情表現(正解ラベル)を比較し,一致数を求める.一致数はアンケートの回答が,文章データに対応する正解ラベルと一致した数である.検証に使う文章データ(各感情表現につき2件ずつ選ばれた文章データ)

とその文章に対応する感情表現(正解ラベル)を表2に示す。

表2 感情辞書からランダムに選んだ検証用データ

正解ラベル	文章データ
喜び	タグ反応ありがとうございます
喜び	同い年の人少ないので私も嬉しいです!
信頼	午前中に用事があるけどお昼配信には間に合うよだからホッとしています。
信頼	とにかくエアコンが動いてくれてホッとしている
恐れ	どんどん狩られるんだらうなあ...
恐れ	今日は日本語能力試験を受けます
驚き	なんか電車1時間位遅れてるんやけどマジ?
驚き	今日色々あって遅刻で学校行くんよ
悲しみ	見に行きたかったー!
悲しみ	もっと韓国語出来たらたくさん話せるのに
嫌悪	今はわざわざ、赤の他人の赤ちゃん写真を延々とスクロールして見られるくらい
嫌悪	みんなに生きる希望がある事を知らしめています!
怒り	このままだと気持ちも身体も壊れてしまうわ
怒り	電車でおっさんに意地悪された。
期待	今日はみんなにとってどんな日だったかなあ~
期待	頑張ってお楽しみにしてる~!

4.2 検証結果

アンケートの集計による検証結果を表3に示す。表3の回答数は、感情辞書から各感情表現が決定された2件の文章に対する回答数である。表3の一致数はアンケートの回答(人手により決定された感情表現)が、表2の文章データに対応する感情表現(正解ラベル)と一致した数である。表3の一致率は回答数と一致数の比率を示す。

アンケートの文章数は16文章であり、アンケート回答者は25人、合計の回答件数は370件となっている。

今回の検証では、一致数が回答数の半分以上である場合(人が半分以上判断可能な場合)は、文章から人手により決定された感情表現と顔文字から決定された感情表現(正解ラベル)が一致していると考えことにする。

表3 検証結果

	回答数	一致数	一致率
喜び	46	40	87.0%
信頼	47	8	17.0%
恐れ	46	23	50.0%
驚き	46	15	32.6%
悲しみ	46	31	67.4%
嫌悪	46	2	4.3%
怒り	47	12	25.5%
期待	46	35	76.1%

表3の検証結果より、感情辞書の喜び、恐れ、悲しみ、期待それぞれの感情表現を示す文章データは、半分以上の人が文章データを正しい感情表現と判断可能であるということがわかった。つまり、顔文字より決定される喜び、恐れ、悲しみ、期待の感情表現は、人手により決定される感情表現と一致しており、顔文字からの感情表現決定手法は有効であることが示された。したがって、Twitter文章からの顔文字を用いた感情辞書の構築は可能である。

一方、顔文字から決定される信頼、驚き、嫌悪、怒りの感情表現(正解ラベル)は、人手により決定された感情表現と一致しないことがわかった。

その理由はTwitter文章から決定される感情表現は、感情表現によってばらつきが出ているためである。そのため、Twitterの文章データから感情表現を決定する際に使う顔文字の選択が重要になる。つまり、感情辞書構築の際、信頼、驚き、嫌悪、怒り、それぞれの感情表現を決定する際の顔文字を見直すことで、より確度が高い感情辞書の構築が可能と考える。

5. 結論

本研究は、人との自然なコミュニケーションを行う対話システムに必要な感情判断器の開発を目指している。

本論文で扱う、顔文字を元にした感情辞書において、文章データの正解ラベルである感情表現が、人手の判断による感情表現と一致しているのかの検証が不十分であった。そこで、本論文では、感情辞書の感情表現(正解ラベル)と実際に人手により決定された感情表現が一致しているか検証した。

その結果、感情辞書の文章データから決定された喜び、恐れ、悲しみ、期待の感情表現は、人手により決定された感情表現と一致しているため、Twitterからの顔文字を用いた感情辞書の構築は有効であることがわかった。しかし、感情辞書の信頼、驚き、嫌悪、怒りの感情表現は、人手により決定された感情表現と一致しないため、顔文字を見直す必要がある。

今後は、顔文字を変更することで、顔文字から決定された感情表現が人手により決定された感情表現と一致するような感情辞書の構築を検討する予定である。

参考文献

- [1] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka, Kenji Araki. A system for affect analysis of utterances in Japanese supported with web mining. 日本知能情報ファジイ学会誌, vol.21, No.2, pp.194-213 (2009)
- [2] 中村 明, 感情表現辞典. 東京堂出版, (1993).
- [3] 山本裕司, 吉村枝里子, 土屋誠司, 渡部広一. 会話文章からの話者の意図・感情判断. 情報処理学会研究報告, vol.2010-ICS-158, No.2 (2010)
- [4] 山下紗苗, 上泰, 奥村紀之. 日本語文章からの感情推定難易度判定. 人工知能会第33回全国大会, 4M3-J-9-03 (2019)
- [5] Robert Plutchik. "The nature of emotions". American Scientist. Vol.89. Iss.4, pp. 344-350 (2001)
- [6] 盛舒峰, 渡辺裕. コミックのセリフの感情分析. 電子通信学会基礎・境界ソサイエティ/NOLTAソサイエティ大会, A-10-18 (2018)
- [7] 大澤泰我, ケネス マッキン, 永井保夫. 対話システムのための感情判断器の検討. 情報処理学会第82回全国大会. 1S-06, (2020).
- [8] 大澤泰我, ケネス マッキン, 永井保夫. 文脈から感情を分析するための感情判断器の検討. 第19回情報科学技術フォーラム. F-026, (2020).
- [9] 大澤泰我, ケネス マッキン, 永井保夫. 文からの感情判断におけるベクトル化手法の検討. 情報処理学会第82回全国大会. 5Q-01, (2021).