

## 感情語辞書と感情分析システム EEAS

小林孝充<sup>†</sup>、石井解人<sup>†</sup>、恵谷菜央<sup>†</sup>、近藤友啓<sup>†</sup>、安達由洋<sup>†</sup>

東洋大学総合情報学部総合情報学科<sup>†</sup>

### 1. はじめに

近年、自然言語処理技術を用いて SNS 上の世論の動向や、Web 上の口コミあるいはレビューなどを分析する研究が盛んになりつつある。そして SNS や Web 上の大量の自由記述文を自動感情分析するには、網羅的な日本語感情語辞書の作成と高速な感情分析アルゴリズムの開発が不可欠である。

本研究では、各語に 10 カテゴリの感情の有無を表す {1, 0} の 2 値を要素とするベクトルを対応させた感情語データから感情語辞書を生成する感情語辞書管理システムと、感情語辞書を使用して高速に文の感情分析をするシステムを実現した。生成した感情語辞書は 6,660 語の見出し語を持っている。また、本感情分析システムは自由記述文 10,000 文を 1 秒程度で分析することができる。

### 2. これまでの感情語処理

コンピュータ処理を考慮に入れた日本語感情語辞書に関する研究は 2010 年頃から盛んになり始めた [1, 2]。文献 [3] は、感情語を 10 種の感情カテゴリ {喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚} に分類して収録した書籍である。文献 [1] では、文献 [3] に収録された感情語から 2,100 語を感情語辞書に登録して感情分析するシステムを実装している。文献 [4] では、文献 [3] の 10 感情カテゴリを用いて、新型コロナウイルスの蔓延が社会に与えた影響を SNS のツイート内容から読み取り分析した研究結果を報告している。また、語を positive と negative に分類する評価極性辞書 [5] とその辞書に基づいて文の極性を評価するシステム [6] が報告されている。

### 3. 感情語処理システム

本研究では、人手で作成した感情語データから感情語分散表現辞書を作成する感情語辞書管理システム (Emotional-word Dictionary Management System, EDMS) と、作成された辞書を用いて高速に自由記述文の感情分析をする感情解析システム (Emotional Expression Analysis System, EEAS) を開発した。

#### 3. 1 EDMS

EDMS は、文献 [3, 7, 8] および我々が収集した感情語データから感情語辞書を作成する。感情語データは、感情語とその語が属する 10 感情カテゴリ {喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚} を表す 10 次元のベクトルの組から構成される。例えば、感情語 “喜ぶ” に対し

Emotional-word dictionary and emotional expression analysis system

<sup>†</sup> Takamitsu KOBAYASHI, Kaito ISHII, Nao ETANI, Tomohiro KONDO, Yoshihiro ADACHI • Toyo Univesity

て、感情ベクトルは [1, 0, 0, 0, 0, 0, 0, 0, 0, 0] と表現される。

EDMS の感情語辞書生成手順を以下に示す：

- 手順 1：人手で感情ベクトルを付加した感情語 Excel データを入力
- 手順 2：感情語を形態素解析器 MeCab で解析して、単語の原形のリストで表現
- 手順 3：2 で求めた単語の原形リストを連結して辞書の見出し語を生成  
(単語の原形列を用いることにより、活用形や語形変化に柔軟に対応)
- 手順 4：見出し語と感情ベクトルを組みにして感情語辞書 (Emotional Word Dictionary, EWD) Excel ファイルに出力

上記手順により、原形 1 語からなる見出し語 3,432 語、原形 2 語からなる見出し語 1,492 語、原形 3 語からなる見出し語 1,116 語、原形 4 語からなる見出し語 426 語、原形 5 語からなる見出し語 151 語、原形 6 語からなる見出し語 34 語、原形 7 語からなる見出し語 8 語、原形 9 語からなる見出し語 1 語の合計 6,660 語の見出し語を持つ EWD を作成した。

#### 3. 2 EDMS の処理例

- 手順 1：感情語データ
- 感情語                      感情ベクトル  
“胸を弾ませる”      [1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
- 手順 2：感情語原形リスト  
[ ‘胸’, ‘ ’ を’, ‘ ’ 弾む’, ‘ ’ せる’ ]
- 手順 3：原形リストの連結  
“胸を弾むせる”
- 手順 4：  
見出し語                      感情ベクトル  
“胸を弾むせる”      [1, 0, 0, 0, 0, 0, 0, 0, 0, 0]

#### 3. 3 EEAS

EEAS は Python の辞書オブジェクト表現を用いて EWD に基づき高速に自由記述文の感情を分析する。

EEAS の感情分析手順を以下に示す：

- 手順 1：EWD を読み込み辞書オブジェクトに格納  
(見出し語をキー、感情ベクトルをバリューとする)
- 手順 2：分析対象の自由記述文を形態素解析して単語の原形のリストで表現  
[以下の手順 3 と 4 を、n を 9 から 1 まで減少させながら繰り返す]
- 手順 3：2 で得た文の単語原形リストを先頭から最後まで順に n 語スライスし、切り出した部分リ

ストを連結して EWD の見出し語と照合する。  
照合した見出し語に対応する感情ベクトルを抽出

手順 4：見出し語と照合した感情語の後ろに否定語（‘ない’、‘ではない’ など）が続くかリストの照合で調べる。否定語が続く場合は感情カテゴリベクトルを修正

手順 5：入力された文から抽出された感情ベクトルのリストを合成（各対応する成分ごとに最大値を取る）

手順 6：5 で得られた感情ベクトルから感情カテゴリ語に変換

手順 7：5 で得られた感情ベクトルから極性値に変換

#### 4 EEAS 処理例

分析対象文：“私は胸を弾ませました。”

手順 2：原形のリスト

[ ‘私’ , ‘は’ , ‘胸’ , ‘を’ , ‘弾む’ , ‘せる’ , ‘ます’ , ‘た’ ]

手順 3：6 語スライスの連結で見出し語“胸を弾むせる”と照合して、感情ベクトル [1, 0, 0, 0, 0, 0, 0, 0, 0] を抽出

手順 4：後ろに否定語は無い

手順 5：感情ベクトル [1, 0, 0, 0, 0, 0, 0, 0, 0] を合成

手順 6：感情ベクトルから感情カテゴリ語‘喜’に変換

手順 7：感情カベクトルから極性値 positive に変換

#### 4. 性能評価

TSUKUBA コーパス [9] を用いて、EEAS の極性分析の精度と計算速度を評価した。

##### 4. 1 精度評価

EEAS の精度：

正解率 0.7037037037037037

精度 [0.86191537 0.30107527 0.8215575]

再現率 [0.52796726 0.38956522 0.87335526]

F1 値 [0.65482234 0.33965125 0.8466649]

Oseti の精度：

正解率 0.6079182630906769

精度 [0.63157895 0.21656051 0.77630822]

再現率 [0.52387449 0.29565217 0.74013158]

F1 値 [0.57270694 0.25 0.75778838]

精度、再現率および F1 値は[negative, neutral, positive]の順で表示している。EEAS では、感情ベクトルから極性値を計算しており、Oseti と比べて約 9.5%正解率が上がっている。

##### 4. 2 速度評価

図 1 に、TSUKUBA コーパス [9] に対する感情ベクトルの計算速度の測定結果を示す。測定環境は Intel Core i7-1068NG7 CPU @2.30.GHz、16 GB memory である。EEAS は 10,000 文の感情ベクトルを約 0.96 秒で計算する。

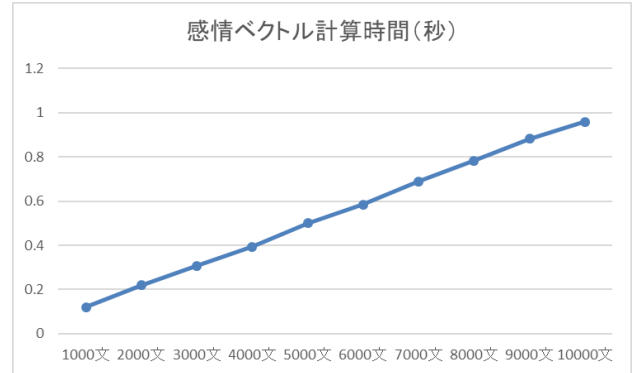


図 1：感情ベクトル計算時間

#### 5. まとめ

本研究では、精度よく高速に自由記述文の感情分析するための感情語辞書管理システム EDMS と、感情分析システム EEAS を開発した。

今後の課題として、より多くの分野の感情分析に利用できるように語彙を増やすことが挙げられる。なお、見出し語の検索にハッシュを使っているため感情語語彙を増やしても分析処理時間は増加しない。また、10 感情カテゴリは文献 [3] に準じているが、感情語は時代の変化や、SNS などのデジタル文化の隆盛とともに変化を続けており、新しい感情カテゴリの追加も大きな課題である。

#### 謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより楽天株式会社から提供を受けた「楽天データセット」を利用した。深謝する。

#### 参考文献

- [1] Michal Ptaszynski I : ML-Ask, <http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/repository/mlask.html>、(2020/07/15 アクセス)
- [2] 長岡技術科学大学旧自然言語処理研究室：日本語感情表現辞書、<http://www.jnlp.org/SNOW/D18> (2020/07/15 アクセス)
- [3] 中村明：感情表現辞書、東京堂出版、1993.
- [4] 鳥海不二夫、榊剛史、吉田光男，“ソーシャルメディアを用いた新型コロナ禍における感情変化の分析,” 2020.
- [5] 乾・鈴木研究室、日本語評価極性辞書、<http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary> (2021/01/06 アクセス)
- [6] 池上有希乃、<https://pypi.org/project/oseti/> (2021/01/06 アクセス)
- [7] 学研プラス、感情ことば選び辞典、2017.
- [8] 学研プラス、和の感情ことば選び辞典、2019.
- [9] 楽天株式会社 (2014): 筑波大学文単位評価極性タグ付きコーパス. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.6>