

# 日本語テキストからの感情抽出

## Affect Extraction from Text in Japanese

菅原 久嗣\*<sup>1</sup>      アレナ ネビアロスカヤ\*<sup>1</sup>      石塚 満\*<sup>1</sup>  
 Hisashi Sugawara      Alena Neviarouskaya      Mitsuru Ishizuka

\*<sup>1</sup>東京大学大学院 情報理工学系研究科  
 Graduate School of Information Science and Technology, The University of Tokyo

Recently, many texts can be seen in Internet. These texts contains not only character strings but also affective signals of writers. This paper introduces an approach to extract an affect from texts. By this, texts are classified into 9 affect categories: "Joy", "Sadness", "Angry", "Disgust", "Fear", "Guilt", "Shame", "Interest", and "Surprise" with the affect score which shows strength of the affect.

### 1. はじめに

インターネットが一般的に広く浸透し、ニュースサイト・Weblog・電子メール等によって、大量のテキストを Web 上で閲覧できるようになった。しかしテキストに含まれる情報は文字列そのものだけでなく、書き手の感情のように文章を読解する事によって得られる情報も存在するはずである。書き手の感情をテキストから自動で抽出し、出力できるようになれば、書き手と読み手の意思疎通がよりスムーズになる。これは人型エージェントを例としたテキスト読み上げロボットの表情・ジェスチャー構成や、レビュー情報の評価への応用にもつながる。

本研究では、日本語テキストから感情を抽出する手法を提案する。

### 2. 提案手法

#### 2.1 概要

図 1 は本手法の全体像を示す。

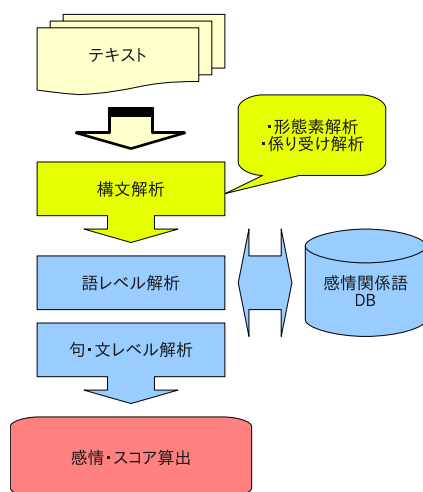


図 1 感情抽出フロー

このシステムは、オンラインコミュニケーションからテキスト感情を抽出する Neviarouskaya らの手法 [Alena 05] を元に行っている。Izard が提唱した、人間の表情形成による感情分類 [Izard 71] から、感情を「喜び」「哀しみ」「怒り」「嫌悪」「恐怖」「罪」「恥」「興味」「驚き」の 9 つに振り分け、その感情の強度を 0.0~1.0 のスコアによって表したものを感情抽出の結果とする。

まず与えられたテキストを構文解析（形態素解析・係り受け解析）し、文構造の情報を取得する。次に事前に用意した感情関係語データベースを参照して、感情に関する語が文中に含まれているかどうかを検証し、文に含まれている語それぞれに 9 次元のベクトルで表現した感情スコアを割り当てる。そして時制・人称・副詞・接続助詞等、文構造を考慮して句単位・文単位にスコアを計算し、最後に文全体の感情とそのスコアを提示する流れとなっている。

#### 2.2 感情関係語

感情スコアの算出に必要な「感情関係語」を WordNet-Affect [Strapparava 04] から収集し、日本語の感情語 DB を作成する。WordNet-Affect では、感情に関係する語に Emotion Category として Anxiety, Hope, Distress 等のラベルが付与されている。これらのラベルを、用意した 9 つの感情のうちどれに分類されるかを 3 人のアノテーターにその強度と共に 1 つ以上選択・判別してもらった。そして 2 人以上が選択した感情カテゴリを採用し、強度の平均をそのスコアとしてラベルに感情スコアのベクトルを割り当てた (表 1)。次に各単語に付与されたラベルから、それぞれの感情カテゴリで最も強度が高いものをその単語のスコアとし、9 次元のベクトルで感情スコアを算出した (表 2)。これをラベルが付与された全ての単語 (300 語) で検証し、得られたデータを感情関係語データベースとして構築する。

ラベル	感情カテゴリ	強度
Astonished	驚き	1.0
Liking	喜び	0.5
	興味	0.3
Blame	哀しみ	0.4
	怒り	0.1
	罪	0.8

表 1 WordNet-Affect における Emotion Category の例

連絡先: 菅原 久嗣, 東京大学大学院 情報理工学系研究科 石塚研究室, 〒113-8656 東京都文京区本郷 7-3-1, 03-5841-6775, sugawara@mi.ci.i.u-tokyo.ac.jp

感情関係語	ラベル	感情スコア
約束 (promise)	Hope[興味:0.1, 喜び:0.3]	興味:0.1 喜び:0.3
勇氣 (brave)	Pride[喜び:0.4] Admiration[喜び:0.6, 驚き:0.5]	喜び:0.6 驚き:0.5
衝突 (clash)	Anger[怒り:0.9] Anxiety[恐怖:0.4] Resentment[怒り:0.6]	怒り:0.9 恐怖:0.4

表 2 感情関係語のスコア算出例

### 2.3 語レベル解析

入力されたテキストの文中に感情関係語データベースに含まれている単語があった場合、それに感情スコアを代入する。

### 2.4 句レベル解析

次に、句単位の感情スコアを算出する。この場合それぞれの感情関係語のカテゴリで、最大のスコアを採用してその句の感情スコアとみなす。例えば形容詞 [怒り:0.8, 恐怖:0.1] と名詞 [怒り:0.4, 恐怖:0.5] で組み合わせることができる名詞句の感情スコアは、「怒り」「恐怖」それぞれの最大スコアを採用して [怒り:0.8, 恐怖:0.5] とする。また、この時感情スコアのうち最も強度が高いカテゴリが喜び・驚き・興味のいずれかに該当する場合、その句を”Positive”な句とし、それ以外を”Negative”な句とする。これを係り受けの順に、文が主語(名詞句) + 目的語(名詞句) + 述語(動詞句) という句の形になるまで繰り返す。

動詞が否定形になっている場合、感情スコアはゼロとする。

### 2.5 文レベル解析

主語・目的語・述語の感情スコアから、文全体の感情スコアを算出する。ここで、次のルールを定義する。

1. 述語 + 目的語から、述部の感情スコアを出す。”Positive”と”Negative”が一致している場合は句レベル解析と同様に感情カテゴリ別の最大のスコアを採用し、述部の感情スコアとする。一致していない場合、述語の感情スコアを述部の感情スコアとする。
2. 主語 + 述部から、文の感情スコアを出す。算出方法は 1. と同様。

こうして得られた、文の感情スコアの中で最も数値の高い感情カテゴリをその文の感情として出力する。数値も感情の強度として、共に出力する。

文が接続助詞で結ばれている場合は次のルールを適用してスコアを算出する。

1. ~ので・~たり、のように順接・並列の接続助詞が用いられている場合、それぞれの感情カテゴリのスコアの最大値を採用する。
2. ~だが・~けれど、のように逆接の接続助詞が用いられている場合、接続助詞より後ろの文の感情スコアを採用する。

## 3. 感情の抽出実験

### 3.1 実装と結果

この感情抽出手法を実装し、実際に感情スコアを算出した。入力テキストは読売新聞のニュース記事 [YOMIURI] から収集した。表 3 にその一部を示す。

### 3.2 実験の考察

実験では、感情を抽出できずスコアが出せない例が目立った。スコア算出に必要なデータベース内の感情関係語が充分でなく、語彙をもっと増やす必要がある。既に登録されている語の名詞形・動詞形の派生語や、同義語・類義語をデータベースに加えたり、相互情報量を用いて Web から感情関係語と共起の強い語を収集して新たにデータベースに加える等、感情関係語を充実させていくのが今後の課題である。また単語単位でなく、「気持ちが悪い」のように語が組み合わせられて意味を持つケースや、「経済成長率の低下」「失業率の低下」のように、係り受けする語によって語の感情的な意味が異なるケースもあるため、1つの語彙だけでなく語句の組み合わせも感情関係語としてデータベース化する事も視野に入れていく。

入力テキスト	感情スコア
午前 11 時、合格番号が発表されると、集まっていた受験生や父母たちから歓声があがり、合格者たちは抱き合って喜んだ。	喜び:0.9
練習中、ジュリのタックルをものに受け負傷したデコは応急処置を受け、足をひきずりながらラ・マシアを後にした。	哀しみ:0.7
英国の通信・放送分野の監督機関は 4 日までに、英 BBC 放送に大使、ラジオ番組で英有名男優らを侮辱するような発言があったとして、BBC への罰金としては過去最高額となる 15 万ポンドの支払いを命じた。	怒り:0.7

表 3 感情の抽出例

## 4. おわりに

本論文では、感情抽出手法の概要と実験結果について論じた。今後はデータベースの充実化を図り、抽出精度を向上させる事を重点的に進めていく。

また、さらに日本語文法での感情表現を精査し、複数の語句で感情的意味を持つケースを分析して、抽出手法に取り入れる事も検討していきたい。

## 参考文献

- [Alena 05] Nevirouskaya A., Prendinger H., Ishizuka M.: Textual Affect Sensing for Social and Expressive Online Communication, ACII2007, Springer LNCS 4738, pp.218-229.(2007)
- [Izard 71] Izard, C. E.: The Face of Emotion, New York:Appleton-Century-Crofts.(1971)
- [Strapparava 04] Strapparava, C., Valitutti, A.: an Affective Extension of WordNet, Proceedings of LREC'04, pp.1083-1086.(2004)
- [YOMIURI] <http://www.yomiuri.co.jp/>