

ディープラーニングを用いた暗示的怒りの自動検出手法

Automatic Detection Methods of Repressed Anger using Deep Learning

酒井 優介 藤ノ木 太郎 安藤 雅洋 湯川 高志
Yusuke Sakai Taro Fujinoki Masahiro Ando Yukawa Takashi

長岡技術科学大学
Nagaoka University of Technology

Japanese people tend to repress the expression of strong emotions. Conventional emotion detection systems don't detect repressed anger with sufficient accuracy. Documents containing repressed anger are unusual to express anger with all sentences. Therefore the author proposed and evaluated methods to detect emotions for each sentence and detect anger of the document by deep learning based on the emotion type and order.

The accuracy rate for the a method which firstly classifies repressed or expressed anger and then classifies document into detail kind of anger improved by 10 percent points compared with the classifier which inputs the whole sentence, but a sufficient method cannot be established because there are types of anger that are never detected. However, since anger could be detected by the type and order of emotions contained in the document, it was suggested that the method of using the sentence as input to the system is effective.

1. はじめに

これまで日本語の文書から明示的な感情を検出することは自然言語処理の分野で精力的に研究されてきた。しかし日本人は感情の表出を抑制する傾向にあり、怒りを表現する際にはそれが特に顕著である。そのため、日本語の文書から怒りを検出するには明示的な怒りだけではなく感情が抑制された暗示的表現による怒りも含めた検出が求められる。そこで、暗示的表現による怒りの自動検出技術の確立を目指し、本稿では暗示的表現による怒りが含まれる文書の特徴に基きディープニューラルネットワークを用いて怒り文書を検出する手法を提案し評価する。

2. 怒りの種類

木野ら [1] によると、怒りは“表情・口調”，“遠回し”，“嫌味”，“理性的説得”，“感情的攻撃”，“無視”，“いつもどおり”の7種類に分類される。小畑ら [2] は木野らの分類を元に文書からは検出が困難な“表情・口調”と“いつもどおり”を除いた怒りに“怒っていない”を追加して分類する手法を提案している。本検討においてもこの分類を用いる。このうち“嫌味・皮肉”，“遠回し”が暗示的表現による怒りである。

感情的攻撃 怒りに任せ、相手を責め立てる。
感情的説得 感情的だが相手の非を述べている。
理性的説得 理性的に相手の非を伝える。
嫌味・皮肉 嫌味や皮肉を言う。
遠回し さり気なく遠回しに伝える。
怒っていない 怒っていない

3. 本稿における着眼点

暗示的怒りが含まれる文書の多くは全体で怒りを表現していることは少なく、実際に起こったことを説明する文が大半を占め、一部に“嫌味・皮肉”や“遠回し”な表現が含まれていることが多い。そこで図1に示すように、文書全体ではなく個々の文を分類し、文書中出现する複数の怒りの種類から文書全体の怒りを決定することで、より正しい検出が期待できる。文をConvolutional Neural Network(CNN)を用いた分類器により

分類することで、大半が事象の説明で、一部のみ暗示的表現により怒りを表しているような文書でも、小畑らの全文を対象に分類する手法に比べ性能の向上が期待できる。

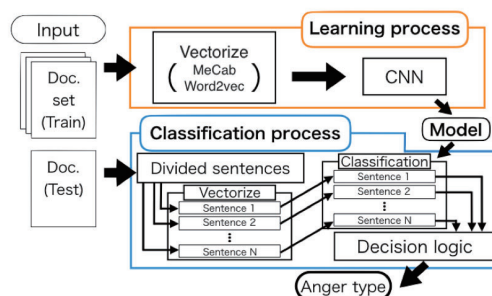


図1: 提案する分類手法の概要

4. 暗示的表現による怒りの検出手法

文書を個々の文に分け文それぞれの怒りを検出し、それらを統合することで文書で表現されている怒りを決定する3つの手法を提案する。それらを実験的に評価し、比較する。

4.1 文書中出现する怒りの種類に基づいた検出手法

図1の“decision logic”として図2に示すロジックを用いる。“感情的攻撃”もしくは“感情的説得”が文書中に一つでも含まれていれば、明示的な怒りと判断する。そのうち、“感情的説得”，“理性的説得”が含まれる文書は説得している文があるため、“感情的説得”とする。この感情的説得に当てはまらないものは感情的に「怒鳴りちらしている」だけとして“感情的攻撃”に分類する。

先に述べたように、暗示的表現による怒りは文書の一部に現れることが多い。そのため“嫌味・皮肉”，“遠回し”のいずれかが含まれている場合はそれをその文書の怒りとし，“嫌味・皮肉”，“遠回し”の両方が文書に含まれる場合，“嫌味・皮肉”は遠回しな表現とは言い難いため“嫌味・皮肉”とする。上記以外の文書で“理性的説得”が含まれる文書はその分類とし、

いずれの条件も該当しない文書を“怒っていない”とする。

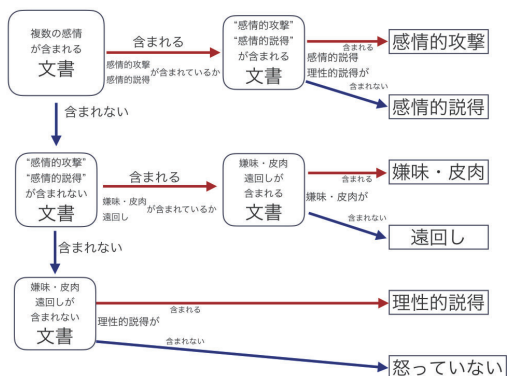


図 2: 文書全体の怒りの決定手法

4.2 明示的怒りを SVM により分離する手法

明示的怒りに比べ暗示的怒りの特徴は小さく怒っていない文書のようにみえる。そこで図3のように、はっきりとわかる特徴を持つ“明示的怒り”とそれ以外の二つにサポートベクターマシン (SVM) により分けてから、詳細な怒りの種類を検出する。

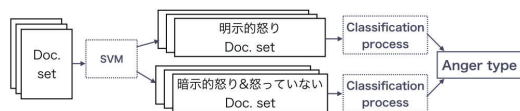


図 3: 2段階に分けた検出システム概要

前節の図2では1つ目の条件である“感情的攻撃”・“感情的説得”が含まれているかによって明示的に怒りとそれ以外に分けている。本節の提案手法ではSVMにより感情的な怒り文書を分離しているため“decision logic”には図2の第1条件を除いたものを用いる。

4.3 文毎の感情から文書の怒り決定ロジックにディープラーニングを用いた手法

暗示的怒りを表現している文書では全体で怒りを表現しているわけではなく、実際に起こった事象を説明する文の後に怒りを表現していることが多いという特徴がある。そこで、図4に示すように文書を表現している感情の順序にも着目するため Recurrent Neural Network(RNN) を用いた分類器を用いて文書全体の怒りを決定する。

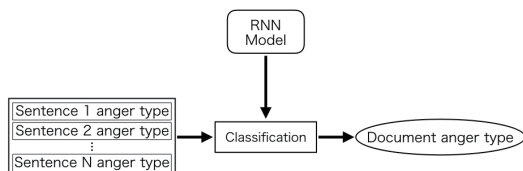


図 4: 感情の出現順序に着目した検出システム概要

5. 実験的評価及び考察

3種類の提案手法を実験的に評価した。怒りの種類毎のF値および精度を表1に示す。表の“全文入力”は、文書を文に分けずにシステムの入力としたもの、3種類の手法の数字は節番号に対応している。

手法4.1では、全文を入力とする小畑らの手法と比べて全体の精度は変わらなかったが、明示的怒りのみに着目すると向上した。しかし、“感情的説得”、“怒っていない”が1件も検出されなかった。明示的怒りに比べ、暗示的怒りの特徴が小さくこれらを同時に検出することは困難であると言える。

手法4.2の精度は0.38となり提案手法中最も良い結果が得られた。しかし、“感情的攻撃”、“怒っていない”が1件も検出されない結果となった。検出されない怒りの種類があることから、文書中にある怒りを表現している文があるかどうかにより文書全体の怒りを決定する手法に問題があると言える。

手法4.3は精度が0.18で従来手法よりも悪い結果となった。追実験として、各文の分類として正解を付与して、文書全体の怒りを決定するRNN部分のみを評価すると、0.49の精度が得られた。このことから、個々の文の感情分類精度を上げることができれば、本手法を用いて良い性能を達成することが期待できる。

表 1: 各手法における怒りごとのF値

	全文入力	手法 4.1	手法 4.2	手法 4.3
感情的攻撃	0.27	0.57	0.00	0.00
感情的説得	0.24	0.42	0.57	0.00
理性的説得	0.29	0.00	0.28	0.07
嫌味・皮肉	0.27	0.27	0.44	0.00
遠回し	0.00	0.10	0.46	0.09
起こっていない	0.42	0.00	0.00	0.32
精度	0.28	0.28	0.38	0.18

6. まとめ

本項では暗示的怒りを表現する文書の特徴に基づいた怒り検出手法の複数提案し、その精度を評価・比較をした。提案したどの手法も、文書を構成する個々の文の怒り感情を分類し、それらを統合して文書の怒り分類とする。文の怒り感情分類にはディープニューラルネットを用いている。

文書中に現れる怒りの種類に着目した手法では明示的怒りの検出精度は向上したが、暗示的怒りの検出精度は大きくは変わらない結果となった。

2段階に分けた怒り検出手法では、文書全文を入力とする手法に比べ性能が10ポイント向上し、本項での提案手法のうち最も良い結果が得られた。しかし全く検出されない怒りがあったことから暗示的怒り検出手法が確立できたとは言えない。

感情の出現順序や頻度に着目した検出手法では精度が0.18と低い性能となった。各文の感情分類として正解を付与し、文書全体の怒りを判別する部分のみを評価したところ精度は0.49となったことから、暗示的怒りの検出のために、文書中に現れる感情の順序、頻度も活用すべきであることが示唆された。

参考文献

- [1] 木野和代. 日本人の怒りの表出方法とその対人的影響. 心理学研究, 第70巻第6号, pp. 494-502, 2000.
- [2] 小畑仁志, 湯川高志. CRFを用いたクレームデータからの内包的怒りの検出. 平成28年度電気情報通信学会信越支部大会講演論文集, No. 6B-3, 2016.
- [3] R. Plutchik. *THE EMOTIONS*. University Press of America, 1991.