

Word2Vec を用いた感情表現の自動獲得に関する研究

水間大資† 1 瀧田孝康† 1 紫尾太朗† 1 田中智紘† 1

概要：近年、計算機の性能向上と搭載メモリの急激な増大により、以前までは不可能だった大規模なニューラルネットワークの実装が可能となり、それをきっかけとして、人工知能に関する関心が高まってきている。このニューラルネットワークを利用して、文章中の単語の前後関係から単語をベクトル化し、従来の自然言語処理では不可能だった単語どうしのベクトル演算を実現したシステムとして Word2Vec がある。本研究では、この Word2Vec を利用して、人間の感情を表現する単語から感情ベクトルを生成し、一般の文章の感情表現が可能となるかどうかを実験により検証した。感情表現には Plutchik の定めた 8 つの基本感情を用いた。インターネット上に公開されているいくつかの文章を使用して検証を行った結果、単にすべての単語の和を使った場合では、人間が文章から感じる感情とは異なる結果となることがわかった。

キーワード：機械学習

Study on automatic acquisition of feeling expressions using Word2Vec

DAISUKE MIZUMA† 1 TAKAYASU FUCHIDA† 1 TARO SHIBI† 1
TOMOHIRO TANAKA† 1

Abstract: In recent years, a sharp increase in computer performance and memory implementation of massive neural networks ago it was impossible becomes possible, it sparked a growing concern about the artificial intelligence. Vector processing of words using neural networks and vectored the words from the context of the words in the sentence, was impossible in traditional natural language processing system is Word2Vec. Whether or not generate emotion vectors from Word to express human emotions, use this Word2Vec in this study, General sentences emotion can be verified by experiment. Using your feelings by Plutchik eight basic emotions. We found that results is different from emotion if you simply used the sum of all of the words, man feels from writing and texts are published on the Internet using the validation results.

Keywords: Mechanical learning.

1. はじめに

近年、計算機の性能が向上し搭載メモリの増大により、以前では不可能だった複雑な計算が可能になった。それを利用した大規模なニューラルネットワークを利用した学習が注目されており、それをきっかけとして、人工知能への関心が高まっている。このニューラルネットワークを利用し、文章中の単語の前後関係から単語をベクトル化するシステムとして Word2Vec がある。この Word2Vec を用いることで従来の自然言語処理では不可能であった単語どうしのベクトル演算が可能になった。本研究ではこの Word2Vec を用いて人間の感情を表現する単語から感情ベクトルを生成し、一般の文章の感情取得が可能なのかを調べた。

2. 外部状況

Word2Vec を用いた外部研究としては” Word2Vec を用いた顔文字の感情分類” [2] のようなものがある。[2] の研究では顔文字を用いて感情分類をしているが、本研究では単語

総数 157,339,847、単語種類 269,399 から感情表現に使われている単語を抽出している。

3. 提案手法

本節では単語をベクトルで表し、文章のベクトルを単語のベクトルの和を得て 8 つの感情ベクトルとの cos 距離をパラメータとして得る。

3.1 提案するアルゴリズム

今回の感情表現には Plutchik の定めた 8 つの基本感情を利用した。この基本感情は「喜び」、「信頼」、「心配」、「驚き」、「悲しみ」、「嫌悪」、「怒り」、「予測」の 8 つである。今回の実験ではこの 8 つの感情ベクトルとの cos 距離をそれぞれパラメータとして求める。

感情ベクトルは図 1 のように基本感情の類義語のベクトルを足していったその合成ベクトルを正規化して単位ベクトルにしたものである。これを 8 つの感情ですべて求めた。

†1 鹿児島大学
Kagoshima University

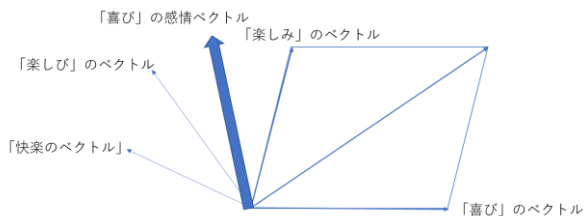


図 1. 「喜び」の感情ベクトルの取得

4. 実験結果

OS:Windows10Pro, CPU:intel core i3 3.2GHz, メモリ 4GB
の PC の Cygwin 上で実験を行った。

4.1 感情のグループのベクトルを求める

図 2 のように感情ベクトルを得た。8 つの基本感情の類義語のベクトルを加えていき、その合成ベクトルを単位ベクトルにしてその感情のグループのベクトルした。今回は 200 次元、CBOW で取得した。

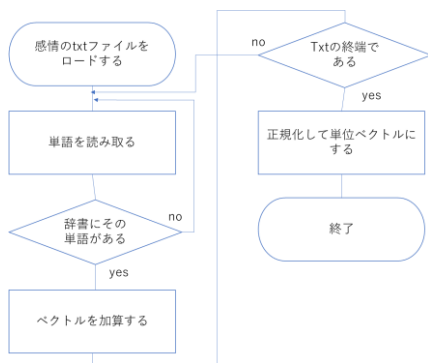


図 2.実験 4.1

4.2 文章から感情を推定する

図 3 のようにパラメータを得た。インターネット上に公開されている文章を分かち書きにし、単語ごとのベクトルを加えてそれらの和を取り、8 つの基本感情のグループとの cos 距離をパラメータとして取った。

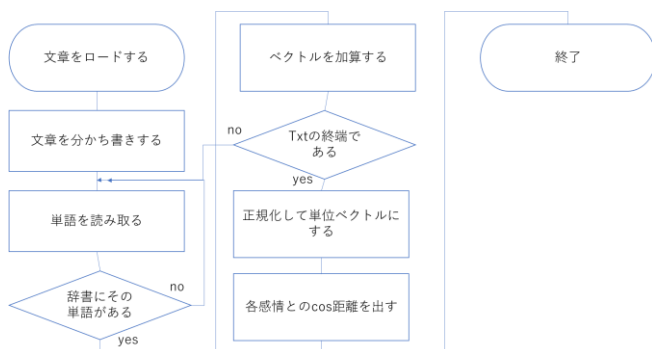


図 3.実験 4.2

「嬉しい話」のパラメータは図 4 のようになった。「嬉し

い」は喜びの類義語であるが、一番大きいパラメータを取ったのは「心配」であった

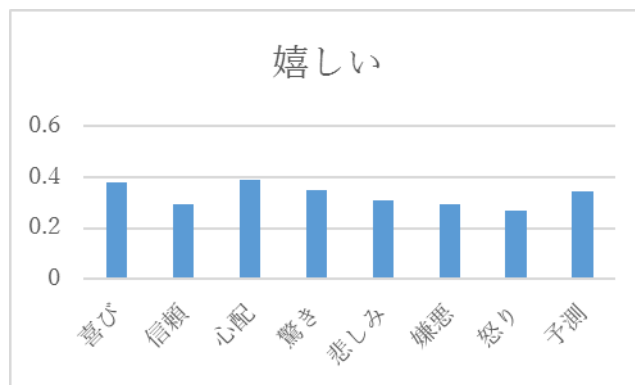


図 4.嬉しい話

「心配な話」のパラメータは図 5 のようになった。一番大きいパラメータを取ったのは「予測」であった。

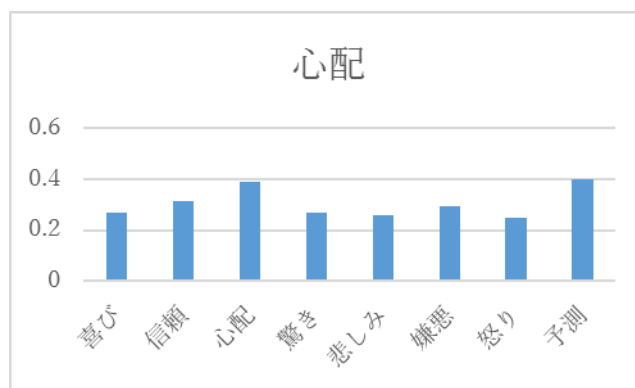


図 5.心配な話

「驚いた話」のパラメータは図 6 のようになった。一番大きいパラメータを取ったのは「心配」であった。

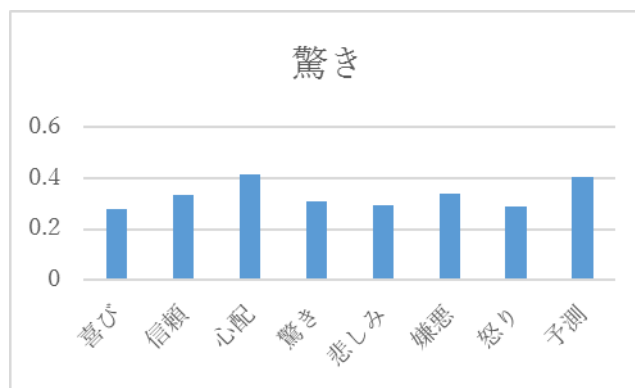


図 6.驚いた話

「悲しい話」のパラメータは図 7 のようになった。一番高い値を取ったのは「心配」であった。

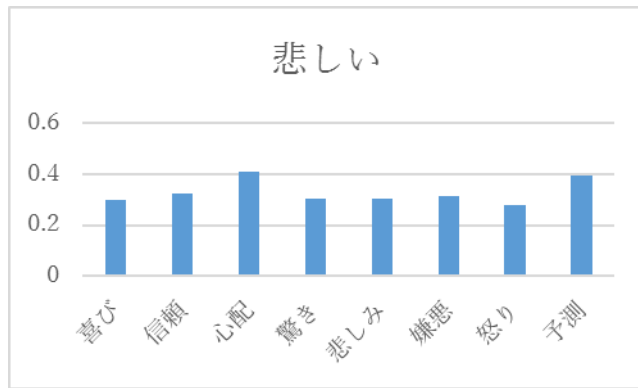


図 7.「悲しい話」

嫌な話は図 8 のようになった。一番高いパラメータを取ったのは「心配」であった。

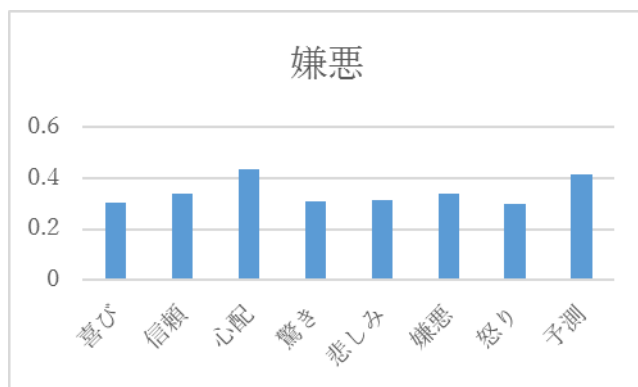


図 8.「嫌な話」

「怒った話」は図 9 のようになった。一番高いパラメータを取ったのは「予測」であった。

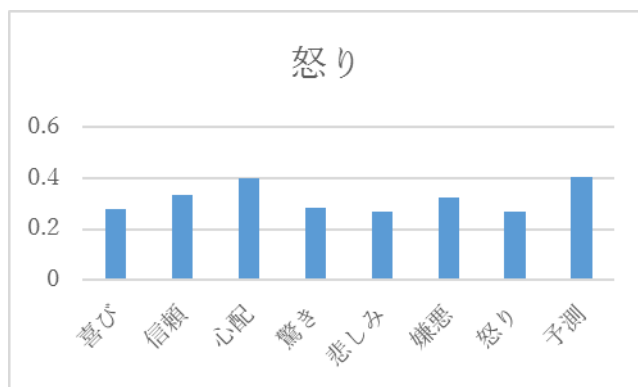


図 9.「怒った話」

5. 考察

実験 4.2 の結果を見ると単に単語のベクトルの和だけを求めた場合、人間が感じた感情と違う結果では得られることが分かった。また図 4~9 の形が似ているところより感情を表している以外の単語が結果に影響を及ぼしていると考え

える。

6. まとめと今後の課題

今回単語をベクトル空間に写像する Word2Vec を用いて文章から感情を推定しようとしたが、人間の感じ方とは違う結果が出てしまった。今回の実験では単語ベクトルの和と感情ベクトルとの $\cos$ 距離で評価をしたが、別の評価方法が求められ、それを探するのが今後の課題になる。また感情を表していない語句を取り除く一般語フィルターを作り、フィルターを通す前後での結果の差の評価も課題になる。

参考文献

- [1] Efficient estimation of word representations in vector space, T Mikolov, K Chen, G Corrado, J Dean - arXiv preprint arXiv:1301.3781, 2013 - arxiv.org
- [2] 黒崎優太, 高木友博, "Word2Vec を用いた顔文字の感情分類", 言語処理学会, 第 21 回年次大会 発表論文集, (2015 年 3 月)