

# 言語分析を用いた誹謗中傷検出の試み

## — 深層学習を用いて —

2018SE061 野々垣 雄哉

指導教員：野呂 昌満

### 1. はじめに

SNS (ソーシャルネットワーキングサービス)の普及率の増加により、Web 上で誹謗中傷等の問題が深刻化している。

この問題を解決するために自然言語分析が用いられており、自然言語分析の精度向上が誹謗中傷の検出精度の向上につながる。

Daniel らはクラス分類の方法に着目して実験を行い、自然言語分析の精度向上を目指した。

本研究では、Daniel らの研究では言及されなかった形態素解析の精度に着目して実験を行い、形態素解析の精度が変化した場合に、日本語を対象とする感情分析においてどれだけの精度の変化が現れるのかということ明らかにする。

本研究の技術課題は以下の通りである。

1. 形態素解析エンジンの定義
2. 感情分析システムの作成
3. 妥当性の検証

本研究の着想により、標榜中傷検出を行う前段階として日本語を対象とする自然言語処理において形態素解析の精度の重要性が明確になる。

### 2. 関連技術

本研究で用いた関連技術を以下に示す。

1. MeCab
2. Django-transformer
3. ptorch
4. chABSA-dataset
5. fastText

上記の1から5の詳細を以下に示す。

MeCab とは形態素解析エンジンの一つである。[1]言語、辞書、コーパス(データベース化されている言語資料)に依存しない汎用的な設計方針を採用しており、C 言語、C++、Java、python 等、数多くの言語で使用することが可能である。設定することで様々な辞書を用いることが可能なため、日本語を対象とする形態素解析エンジンの中では最もよく使用されている。

Django-transformer とは、siny が作成した感情分析アプリケーションである。[2]

Python に向けた機械学習ライブラリである Pytorch を用いて

作成されている。

データセットには chABSA-dataset が用いられており、モデルには日本語学習済み fastText モデルが採用されている。

PyTorch とは、Python のオープンソース機械学習ライブラリである。[3]

Facebook の人工知能研究グループにより初期開発され、現在では python の機械学習ライブラリとして高い人気を誇る。また、Pytorch は自然言語処理と相性のいい Define by Runを採用している。

chABSA-dataset とは、TISインテックグループの TIS 株式会社が無償公開している機械学習で感情解析を行うためのデータセットである。[4]

単語のネガティブ度がラベル付けされており、ネガポジ分析において有用なデータセットとなっている。

fastText とは自然言語処理を高速、高精度化するために有効な機械学習ライブラリである。[5]

同様のアルゴリズムとしては Word2vec などが挙げられるが、Word2Vec を含む他のアルゴリズムに比較して、動作が軽く速いのが特徴である。[6]

### 3. 実験の概要

本研究で行う実験では、形態素解析の精度を向上させた際、感情分析の精度にどのような影響が出るのかを明らかにする。

形態素解析の精度の変更を形態素解析エンジンに搭載されているシステム辞書を変更することで実現する。

この実験を行うことで感情分析における形態素解析の重要性が明らかになる。

### 4. 技術課題の解決

#### 4.1. 形態素解析エンジンの定義

形態素解析を行うために適切な形態素解析エンジンを選択する。

本研究では、日本語の分析を行うことを前提としているため日本語の解析に特化した形態素解析エンジンかつシステム辞書を任意に変更可能な MeCab を用いる。

用いるシステム辞書はデフォルト辞書および 2020/9/10 更新の mecab-ipadic-NEologd とする。[7]

#### 4.2. 感情分析システムの作成

siny の作成した Django-transformer をもとに感情分析システムを作成する。

形態素解析エンジンには MeCab を用い、データセットには chABSA-dataset を用いる。このとき MeCab のシステム辞書をデフォルトの辞書と mecab-ipadic-NEologd とし、実験を行う。

用いるシステムの構造を以下の図 1 に示す。

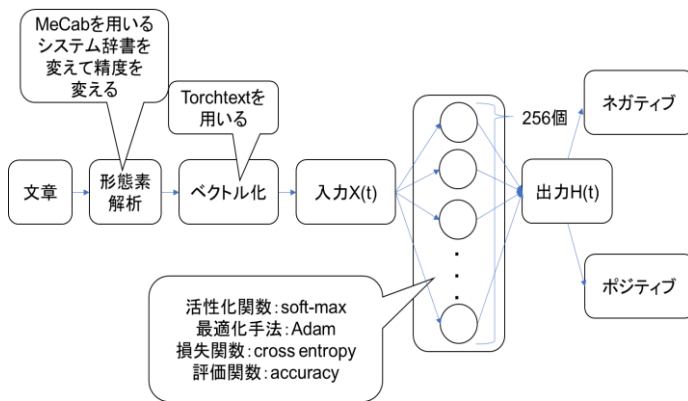


図 3.1: 作成システムの概略

図 1 の概略を以下に示す。

1. 文章を形態素解析
2. 1 の結果を Torchtext を用いてベクトル化
3. 2 をベクトル  $X(t)$  として入力
4. ベクトル  $H(t)$  として文章がネガティブかポジティブかを出力

### 4.3. 妥当性の検証

妥当性の検証を行うために Django-transformer を用いた感情分析システムを実装し、実験を行う。

これまでの技術課題とその解決方法をもとに MeCab のデフォルトのシステム辞書を用いた感情分析システムと mecab-ipadic-NEologd を用いたシステムの両方を設計し、これを実装したシステムを作成する。

本研究では正解ラベル付きデータセットを用いて実験を行っており、推測ラベルが正解ラベルとどれだけ一致しているかどうかを指標として妥当性の検証を行う。

### 5. 評価および考察

デフォルト辞書と mecab-ipadic-NEologd のそれぞれを用いた実験結果を以下に示す。

表 4.1: 実験結果

使用辞書	感情分析の精度
デフォルト	0.8042
mecab-ipadic-NEologd	0.8588

上記の精度は正解ラベルと推測ラベルがどれだけ一致しているかを表している。

表 1 の結果からデフォルトのシステム辞書を用いた場合と比較し、mecab-ipadic-NEologd を使用し実験を行うと感情分析の精度が 0.05 ほど向上した。

これより形態素解析の精度が向上すれば自然言語処理の精度全体も向上することが明らかになった。

このような結果になった要因としては形態素解析の精度が向上したことにより、正しい品詞を検出できるようになったことが挙

げられる。形態素解析を行うことで、文章の捉え方により生まれる自然言語の曖昧性を一意に定義づけることが可能になる。

このことから形態素解析の精度を向上させることにより、システムが正しい品詞を受け取ることで感情分析の精度向上につながったと考える。

### 6. おわりに

本研究では日本語の感情分析における形態素解析の精度の重要性を明らかにすることを目的とした。

今後の課題としては、用いるデータセットを誹謗中傷に特化したものへと変更し、実際に twitter などの SNS から収集した誹謗中傷データを対象として誹謗中傷かどうかの判別を行っていくことに加え、形態素解析のみでなく構文解析を事前に行うことで実験の精度に変化が生じるかということが挙げられる。

それと同時に、本研究では形態素解析に用いる辞書をデフォルトのシステム辞書と mecab-ipadic-NEologd に限定したが、この 2 種類以外に公開されているその他の適切なユーザ辞書を用いることで自然言語分析の精度がどれほど変化するかということも挙げられる。

本研究では実験を行わなかったが、今回用いた Django-transformer では日本語を対象とする自然言語分析の精度の向上を目指すために、中山光樹が作成した日本語学習済み fasttext モデルを用いており、これを作成するために学習に使用したデータは wikipedia2017/01/01 とある。ここで使用するデータ数を増やして学習を行い、Django-transformer と組み合わせることで感情分析の精度の向上を見込めると考えた。[8]

### 参考文献

- [1] 工藤拓: MeCab <https://github.com/taku910/mecab>
- [2] Siny: Django-transformer <https://github.com/sinjob/django-transformer>
- [3] Pytorch <https://pytorch.org/>
- [4] chABSA-dataset <https://github.com/chakki-works/chABSA-dataset>
- [5] fastText <https://fasttext.cc/>
- [6] Word2Vec <https://code.google.com/archive/p/word2vec/>
- [7] mecab-ipadic-NEologd <https://github.com/neologd/mecab-ipadic-neologd>
- [8] 中山光樹: 日本語学習済み fasttext モデル <https://qiita.com/Hironsan/items/513b9f93752ecee9e670>
- [9] Daniel Dichiou and Irina Rancea.: Using Machine Learning Algorithms for Author Profiling In Social Media [https://pan.webis.de/downloads/publications/papers/dichiou\\_2016.pdf](https://pan.webis.de/downloads/publications/papers/dichiou_2016.pdf) (2016).