形態素解析と日本語評価極性辞書を用いた文章のネガポジ判定

4424　豊田 誠弥　（指導教員　白濱 成希）

1. はじめに

インターネットが高度に発達し情報社会となっている今日ではインターネットニュースやSNSなどで多くの記事が溢れている．それらの文章を自動で感情分析しカテゴライズするのを目的として，本研究ではまず文章がポジティブかネガティブかを判断するネガポジ判定を行う．

実験ではまず文章に対してMeCabを用いた形態素解析を行い，日本語評価極性辞書と照らし合わせ一致する単語ごとにスコアを付け評価を試みた．

1. 原理
   1. 形態素解析について

形態素解析とは，文法的な情報の注記のない自然言語のテキストデータから対象言語の文法や，辞書と呼ばれる単語の品詞等の情報に基づき，形態素(おおまかにいえば言語で意味を持つ最小単位)の列に分割し，それぞれの形態素の品詞等を判別する作業である．

本実験の形態素解析で用いたMeCabは，工藤拓氏によって開発されたOSSの形態素解析エンジンである．言語，辞書，コーパスに依存しない汎用的な設計を基本方針としている．また平均的にChaSen，Juman，KAKASIより高速に動作する．

MeCabを用いて「本日はいい天気ですね」という文章を形態素解析すると，以下図1のようになる．

図1 MeCabを用いた形態素解析

本日 ホンジツ 本日 名詞-副詞可能

は ハ は 助詞-係助詞

いい イイ いい 形容詞-自立 形容詞・イイ 基本形

天気 テンキ 天気 名詞-一般

です デス です 助動詞 特殊・デス 基本形

ね ネ ね 助詞-終助詞

2.2.ネガポジ判定について

ネガポジ判定とは，主に人の発言や発想などが，前向き(ポジティブ)か後ろ向き(ネガティブ)かを判定することで，感情分析と呼ばれる技術の一種であるということができる．

実験では東北大学乾・鈴木研究室が公開リソースとしている日本語評価極性辞書を用いてネガポジ判定を行う．日本語評価極性辞書には名詞編と用言編があり，それぞれ図2,図3のような形式になっている．名詞編のn,p,eなどのアルファベットがそれぞれnegative(ネガ)，positive(ポジ)，そしてneutral(ニュートラル)となっておりニュートラルはネガポジどちらとも取れない中間という扱いになる．

図2 評価極性辞書の抜粋(名詞編)

加害者 n   〜である・になる（状態）客観

加虐  n   〜する（行為）

加護  p   〜する（行為）他人

加工性 p   〜がある・高まる（存在・性質）

加工品 e   〜である・になる（状態）客観

加算  e   〜する（行為）

図3 評価極性辞書の抜粋(用言編)

ネガ（評価） 厭わしい

ネガ（評価） 怨めしい

ネガ（評価） 縁 が ない

ポジ（評価） 若い

ポジ（評価） 取り扱い やすい

ポジ（評価） 取り消す こと が できる

1. 実験方法

MeCabで形態素解析した判定したい文章中の単語と，取り込んだ日本語評価極性辞書のデータが一致する個数によりネガポジ度をスコアとしてカウントするプログラムをPythonにて作成した．スコアはポジが+1，ニュートラルが0，ネガを-1とした．Pythonは3.7.1を使った．なお，用言データを用いる実装ができず構造のある文章に対する判定が不正確であったため，実験では名詞のみで判定を行った．

判定する文章はサンプルとして以下の2つを用いた．

1. 太宰治「走れメロス」序文  
   “メロスは激怒した。(中略)低声で、わずか答えた。”
2. 小学二年生の読書感想文「友だちっていいな」
3. 実験結果

「走れメロス」をネガポジ判定する際に辞書と一致した部分と最終的なスコアを出力した結果が以下図4である．

図4 「走れメロス」ネガポジ判定結果

[-1, '激怒']

[-1, '暴虐']

[-1, '邪悪']

[0, '敏感']

[-1, '内気']

[0, '人']

[0, '律']

[0, '気']

[0, '事']

[0, 'ぶらぶら']

[0, '今']

[0, 'つもり']

[1, '楽しみ']

[0, '様子']

[0, '落ちる']

[-1, '不安']

[0, '皆']

[0, '歌']

[1, '賑やか']

[0, '質問']

[0, '質問']

[0, '質問']

ネガポジ度：-3

次に，小学二年生が書いた読書感想文を判定した結果が以下図5である．

図5 小学二年生の読書感想文のネガポジ判定結果

[1, '名']

[0, '入れ歯']

[1, 'すてき']

[1, '本当']

[1, 'すてき']

[0, 'お話']

[0, '入れ歯']

[0, 'ほか']

[-1, 'かわいそう']

[0, '人']

[1, '名人']

[1, 'たっぷり']

[1, '名人']

[0, '今']

[0, '入れ歯']

[0, '問']

[0, '人']

[1, 'しあわせ']

[1, '大切']

ネガポジ度：8

5.考察

サンプル二つとも，おおよそ文章から予想される通りにネガポジが判定されていた．結果の出力をみると，サンプルが長文でも名詞かつ辞書に載っている単語となると一致する数は少ないことがわかる．より豊富な辞書を用意する他，機械学習などで語彙を自動で増やすシステムの構築なども精度向上に繋がると思われる．

シンプルに一致する名詞の数でスコアを付けているため，「天気が悪くなったが傘を持っていたため助かった」というような，文脈を踏まえた判定ができていないことが用言データを用いた判定の実装も含めて課題となる．また「死ぬ」と「近所迷惑」が同じスコアであるため同じ極性でも重みづけを付けることも必要になる．

1. おわりに

今回の実験で，形態素解析などの自然言語処理の基礎やPythonのコーディングを学ぶことができた．現状の成果が入門書レベルの内容であるため，今後は研究として差別化できることをやってみたいと感じた．

参 考 文 献

1. MeCab: Yet Another Part-of-Speech and Morphological Analyzer，<http://taku910.github.io/mecab/#news>
2. 東北大学 乾・岡崎研究室，日本語評価極性辞書， <http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary#b019bc19>
3. PythonでMeCabの出力をリスト化するモジュール(mecab-python)，<https://qiita.com/menon/items/2b5ad487a98882289567>
4. 太宰治，走れメロス，青空文庫，[https://www.aozora.gr.jp/cards/000035/files/156714913.html](https://www.aozora.gr.jp/cards/000035/files/1567_14913.html)
5. 下重諒人，友だちっていいな，<http://www.dokusyokansoubun.jp/text63rd/stei.html>