

自然言語処理による文章の感情分析

3年1組29番 宮田 健吾

研究の狙い

私たちの身の回りにはあらゆる文章があふれています。そしてインターネットが生まれた現代では文章のあり方が変わってきています。昔は雑誌や新聞などのプロのライターが書いた文章しか読むことはできませんでしたが、現代ではブログや twitter、facebook などにより一般消費者が書く文章を目にする機会が増えました。これらの文章のことを **Consumer Generated Media (CGM)** と呼びます。CGM を分析すれば企業側が一般消費者の製品に対する評価や態度を調べ、製品の品質向上に役立つことができると考え、卒業研究のテーマにしました。

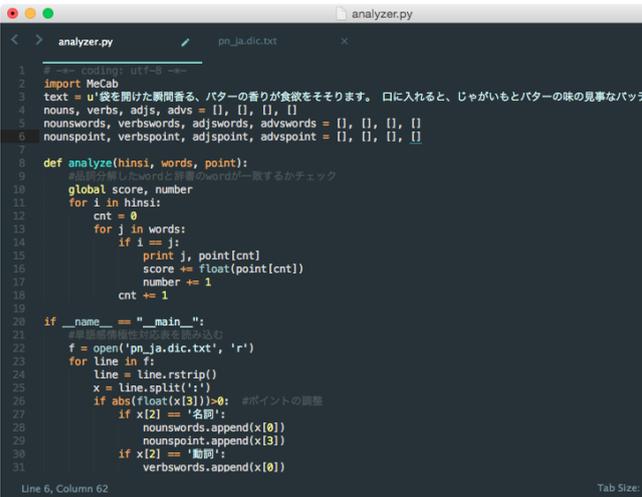
分析方法

自然言語処理は初めてということもあり、簡単な処理で文章の感情分析を行いました。今回は東工大の奥村・高村研究室で作られた単語感情極性対応表を使わせていただきました。この辞書の説明文を引用します。

日本語および英語の単語とその感情極性の対応表を、研究目的の利用に限り公開します。感情極性とは、その語が一般的に良い印象を持つか (positive) 悪い印象を持つか (negative) を表した二値属性です。例えば、「良い」、「美しい」などは positive な極性、「悪い」、「汚い」などは negative な極性を持ちます。

この辞書には単語と感情のポイントが掲載されています。1 から -1 までの段階でポジティブかネガティブかが分かります。この辞書は品詞別になっているため、MeCab と

いう形態素解析エンジンを使用しました。MeCab で文章を品詞分解した後、品詞別に単語感情極性対応表を照らし合わせ、ポジティブかネガティブかのスコアの平均値を出します。以下の画像は今回組んだプログラムの一部です。



```
1 # -*- coding: utf-8 -*-
2 import MeCab
3 text = u'役を離れた期間音響、バターの香りが食欲をそそります。口に入ると、じゃがいもとバターの味の見事なバター
4 nouns, verbs, adjs, advs = [], [], [], []
5 nounswords, verbswords, adjswords, advswords = [], [], [], []
6 nounspoint, verbspoint, adjspoint, advspoint = [], [], [], []
7
8 def analyze(hinsi, words, point):
9     #品詞分解したwordと辞書のwordが一致するかチェック
10    global score, number
11    for i in hinsi:
12        cnt = 0
13        for j in words:
14            if i == j:
15                print j, point[cnt]
16                score += float(point[cnt])
17                number += 1
18            cnt += 1
19
20 if __name__ == "__main__":
21     #辞書読み込み
22     #品詞分解した単語を解析
23     f = open('pn_js.dic.txt', 'r')
24     for line in f:
25         line = line.rstrip()
26         x = line.split(':')
27         if abs(float(x[3]))>0: #ポイントの調整
28             if x[2] == '名詞':
29                 nounswords.append(x[0])
30                 nounspoint.append(x[3])
31             if x[2] == '動詞':
32                 verbswords.append(x[0])
```

分析結果

このプログラムを用いて実際に感情分析をします。以下の文章を分析してみます。

とても素晴らしい卒業研究だ

卒業 -0.291601

研究 -0.653995

素晴らしい 0.998617

とても -0.169067

名詞2つと形容詞1つが辞書に引っかかりました。これらのポイントの合計は -0.0290115 になりました。ポジティブな文章のはずですが、ネガティブと判定されて

しました。卒業、研究、とてもという単語がネガティブと誤判定されたことが原因だと思われます。この実験で**単語感情極性対応表**自体にネガティブよりという問題が発見されました。(単語の1割がポジティブで残りの9割がネガティブ、商品**-0.812312**などの誤りが見受けられる)まだ実用には程遠いようです。暫定的な対処法としてはポイントの絶対値が0.9未満の単語は無視すると良いでしょう。

まとめ

今回の実装は非常に簡易的で精度が悪いため実用はできませんが、**単語感情極性対応表**の他に東北大の**日本語評価極性辞書**も使うことで精度を上げることができるかもしれません。また、**係り受け解析**という修飾被修飾関係を分析できる**CaboCha**というソフトウェアも利用すれば何がポジティブで何がネガティブなのかを明確化することができますと思います。もし自然言語処理による感情分析を用いた企業の製品の品質向上が実現したら、企業側は製品アンケートを実施する必要がなく、消費者がtwitterで感想をツイートすればいい時代が来るかもしれません。

本研究で利用したような**Mecab**や**単語感情極性対応表**など素晴らしいソフトウェア資産を誰でも利用できるように無料で公開されていることがすごいと思いました。自然言語処理のさらなる発展を願っています。

参考文献

高村大也, 乾孝司, 奥村学
"スピンモデルによる単語の感情極性抽出",
情報処理学会論文誌ジャーナル, Vol.47

No.02 pp. 627–637, 2006.

小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.3, pp.203-222, 2005.

東山昌彦, 乾健太郎, 松本裕治, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第14回年次大会論文集, pp.584-587, 2008.

MeCab: Yet Another Part-of-Speech and Morphological Analyzer
<http://taku910.github.io/mecab/>

Yet Another Japanese Dependency Structure Analyzer
<http://taku910.github.io/cabocha/>

単語感情極性対応表のサンプル

単語感情極性対応は以下のようなデータの集まりです。コロンで区切られています。

優れる: すぐれる: 動詞: 1
良い: よい: 形容詞: 0.999995
喜ぶ: よろこぶ: 動詞: 0.999979
褒める: ほめる: 動詞: 0.999979
めでたい: めでたい: 形容詞: 0.999645
賢い: かしこい: 形容詞: 0.999486
善い: いい: 形容詞: 0.999314
適す: てきす: 動詞: 0.999295
天晴: あっばれ: 名詞: 0.999267
祝う: いわう: 動詞: 0.999122