

Sentiment Analysis & Politeness tools

Sentistrength

- developed based on **MySpace comments**
- ソフトウェア工学の研究で最も頻繁に使用されるツール
- 15のTwitter感情分析ツールで最も平均精度が高い
- Positive 1~ 5 , negative -1 ~ -5
- 分析テキスト全体にpositiveとnegativeの両方スコアリング
- $P+N > 0$: positive,
 $P+N < 0$: negative,
 $P = -N \ \&\& \ p < 4$: neutral
 $P = -N \ \&\& \ p \geq 4$: 除外 (両方の気持ち高すぎる)

Alchemy

- IBM
- **感情ラベルを返す**テキスト処理APIを提供
- 感情分析ツールを含む**テキスト処理APIを提供**
非常に短いテキスト (Tweet)
比較的長いテキスト (ニュース記事) を処理する
- スコアは $[-1, 1]$ の範囲内

NLTK

- 単純な一連の単語モデルを使用して、各テキストに対して3つの確率(否定的, 中立的, 肯定的)を返す.
- 中立的スコアが0.5以上なら中立
- 上記ではない場合、最も大きいスコアを持つ感情とする
- www.text-processing.comで提供されているAPIを使用

Stanford NLP

- テキストを解析し，より高度な文法分析を実行
→短いテキストでのbag of wordsモデルよりも優れているという研究結果あり
- テキストを文に**分割**(困難な場合がある)
- 0 : negative, 2 : neutral, 4 : positive
- ドキュメントレベルのスコア
 $-2 * \#0 - \#1 + \#3 + 2 * \#4$ (#0はスコア0のテキスト数)

Senti4SD

- The tool was built using questions, answers and comments from **StackOverflow** as its training data.

[6] Fabio Calefato, Filippo Lanubile, Federico Maiorano, and Nicole Novielli. 2017. Sentiment Polarity Detection for Software Development. *Empirical Software Engineering* (2017), 1–31.

SentiCR

- trained on a dataset that comprises of **code review comments from Gerrit** and two other datasets developed in prior work [6, 31] using **supervised learning techniques(教師あり)**.
- オラクルに基づいてテキストの感情ラベルを返す

[6] Fabio Calefato, Filippo Lanubile, Federico Maiorano, and Nicole Novielli. 2017. Sentiment Polarity Detection for Software Development. Empirical Software Engineering (2017), 1–31.

[31] Marco Ortu, Alessandro Murgia, Giuseppe Destefanis, Parastou Tourani, Roberto Tonelli, Michele Marchesi, and Bram Adams. 2016. The emotional side of software developers in JIRA. In 2016 IEEE/ACM 13th Working Conference on Mining Software Repositories (MSR). IEEE, 480–483.

Politeness tool

- based on "domain-independent lexical(語彙) and syntactic features operationalizing key components of politeness theory, such as indirection, deference, impersonalization and modality"
- 丁寧さ理論の重要な要素を操作可能にする**ドメインに依存しない**語彙および構造上の特徴に基づいて丁寧さを測定
- WikipediaやStack Exchangeで訓練
- 0-1の間のポライトネススコアを予測