

2016/07/29 7:43

### 感情分析とは

- 文書の全体的な文脈の極性を判定し、話し手の主観的な態度を判別する
- その文章の意味は、ポジティブなのかネガティブなのか
- 用いる技術：自然言語処理、テキスト解析、計算言語学など
- ビジネスへの応用：マーケティングや顧客サービス解析など

### 利用事例

- [Yahoo!検索 リアルタイム検索](#)
- [Emotion i](#)
- [Emolyzr](#)

### やりたいこと

- ツイートを感情分析したい！
- つまり、ツイートテキストが、ポジティブなことを言っているのか、ネガティブなことを言っているのか判定したい

### どうやって実装するか(1)

- 感情語辞書を用いる
- 機械学習アルゴリズムを用いる
- 深層学習アルゴリズムを用いる

### 問題

- 訓練データがない！
- テキストデータはある（e.g. ['今日は蒸し暑いですね']）
- 極性を持ったテキストデータが少ない（e.g. ['今日は蒸し暑いですね', -0.6]）
- amazonレビューテキストを試みたがノイズが多い（レビューと関係ない文章、本の抜粋テキストの混入など）

### どうやって実装するか(2)

- (○) 感情語辞書を用いる
- (×) 機械学習アルゴリズムを用いる

- (×) 深層学習アルゴリズムを用いる

## 感情語とは何か

- 感情語とは、喜怒哀楽などの感情を表現している語
- 感情語の例：嬉しい、寂しい、がっかり
- 非感情語の例：は、です、しかし

## 程度副詞による極性値の増加

- 「とても嬉しい」は「嬉しい」よりもポジティブ度が高い
- 「とても」「少し」のような語を程度副詞という
- 程度副詞とは、係る語の状態や性質の度合いを表す副詞のこと
- 程度副詞には程度の値を設定し、係る語の極性値に影響を与える

## 程度副詞の例：「とても嬉しい」の場合

- 程度副詞 ['とても', 0.6] / 感情語 ['嬉しい', 0.4] の場合
- $(1+0.6) * 0.4 = 0.64$  となる
- ※程度副詞の0.6をそのままかけると感情語の値が元の値より小さくなってしまうので、1を加算する

## 否定の場合、極性値はゼロになる

- 「嬉しくない」という語の極性値をどう判定するか
- 「嬉しい」という感情語の極性が0.5だった場合、-0.5になるのか、0になるのか
- 「嬉しくない」とは、嬉しいことを否定しているが、哀しいことを意味するわけではない、と解釈する
- よって、否定の場合は極性値をゼロとみなす

## 用意するデータ

- 感情語
- 程度副詞
- 否定語

## データは分かった

- あらかじめ用意するデータは分かった
- テキストから感情を抽出する処理をどう実装するのか

## 3段階の感情抽出処理

- 語レベル解析
- 句レベル解析
- 文レベル解析

### 語レベル解析

- まずは単語ごとに極性を判定する
- 与えられたテキストを形態素解析し、単語に分割する
- 感情語を見つけ、感情語辞書から極性値を取得する
- 程度副詞を見つけ、程度副詞辞書から値を取得し、係る語を判定する
- 否定語を見つけ、係る語を判定する

### 句レベル解析

- 程度副詞があった場合、係る語の極性値に掛け、句としての極性を算出する
- 否定語があった場合、係る語の極性値をゼロとする

### 文レベル解析

以下の3つの処理を行う

- 文全体の感情スコア算出
- 時制の計算
- 接続助詞の計算

## 実装の全体像

データ

- 感情語
- 程度副詞
- 否定語

プログラム

- 語レベル解析
- 句レベル解析
- 文レベル解析