

マイクロブログにおける感情極性値算出の検討

瀧 晃一[†] 尾崎 広之[†] 須鎗 弘樹[†]
[†] 千葉大学大学院融合科学研究科

1. はじめに

近年, Kahaneman がノーベル経済学賞を受賞したことにより, 人の感情や心理が関係する行動経済学が注目されるようになった. 一方で, 現在若者を中心に広く普及している Twitter や Facebook をはじめとする SNS を利用して, 人々の感情を読み取る研究が行われている. そこで本研究では, Twitter のデータを用いて各単語に独自の感情値を割り当て, 感情極性辞書の作成を試みる.

2. 関連研究

2011 年, Bollen らは気分プロフィール検査(POMS)に用いられる 72 の形容詞と関連して用いられること多い語句を集め, Google-Profile of Mood States(GPOMS)[1]を作成した. また, 2006 年に高村らが 53000 単語に対して, 1~ -1 の PN(ポジティブ/ネガティブ)値を割り当てた感情極性対応表[2]を作成した. しかし収録されているポジティブな単語の割合が約 1 割となっており, ツイートに対して用いたネガティブ値の高い結果しか得られない.

3. 研究手法

本研究の手法の流れを以下に示す.

1 ツイート取得

TwitterAPI を用いてツイートを取得した. 検索単語はポジティブ(嬉しい, 幸せ), ネガティブ(悲しい, 寂しい)に設定した.

2 形態素解析を行い各共起単語を取得

「Yahoo! 日本語形態素解析 API」を用いて取得したツイート本文の形容詞・副詞・形容動詞・感動詞・動詞・名詞である単語を抽出し, その単語の基本形, 読み, 品詞, 共起数を保存する. こうして仮の辞書を作成する.

3 各単語の累計共起数算出

各検索単語ごとの仮の辞書をポジティブ・ネガティブそれぞれ合算して, 各単語の基本形, 読み, 品詞, 共起数をまとめた辞書を作成する.

4 感情値を算出し辞書作成

各辞書の各単語ごとに感情値 (カイ二乗値: χ^2) を算出する.

5 ツイートの感情値算出

作成した辞書を元に形態素解析を行って一致した全ての単語の感情値を合計する.

4. 研究結果と考察

実際のツイートに対して本研究の手法で感情値を算出し

た. 使用したツイートはサッカー W 杯, 日本対コートジボワール戦の試合 1 時間前から試合 1 時間後までのツイートとし, 結果を図 1 に示す.

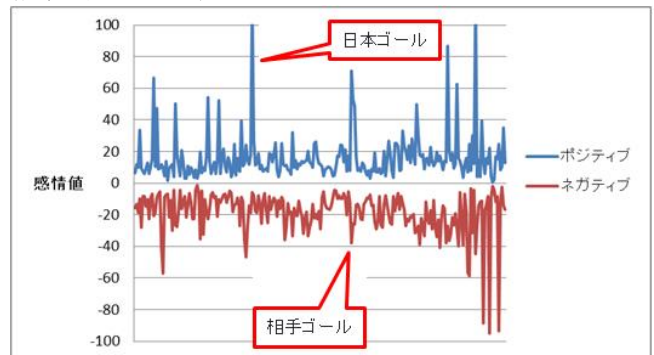


図1. #daihyo の感情値時系列変化

図 1 が示すように, ゴール時のポジティブな感情を抽出できたが, 失点時のネガティブな感情を明確には抽出できなかった. また, 日本が負けた試合であったため, 試合後はネガティブな値が大きかった. この値はツイートの文字数が影響していると考えられる. 1 ツイート平均の文字数が試合中は少なく, 試合前後は多かった. 文字数の変化を図 2 に示す.

1 ツイートの平均文字数

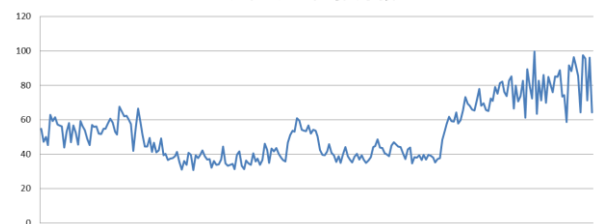


図 2: 1 ツイート平均の文字数時系列変化

5. まとめ

Twitter を用いて感情極性辞書を作成し, 感情極性値を算出した. 今後の課題として Twitter には特有表現などが含まれるため, 言葉の特性を掴む必要がある. そして多くの単語に感情値を割り当てるため, さらに多くのツイートを元に辞書を作成する必要がある. また対象とするツイートの種類により, 異なる辞書を用意する必要がある.

参考文献

[1] J. Bollen, H. Mao and X.-J. Zeng,

“Twitter mood predicts the stock market”. , J. Comp. Sci., vol. 2, no. 1, pp. 1-8, 2011.

[2] 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, vol. 47 no. 02 pp. 627-637, 2006