

感情分析に基づくデマ検知アルゴリズムの開発 Development of Rumor Detection Algorithms based on Sentiment Analysis

QU LIJING[†] 成 凱[‡]
Lijing QU Kai CHENG

1. はじめに

SNS上で自由に発信された情報の中で、人々に役立つ情報がありながら、政治的・利益的な意図を待って拡散されるデマ情報も含まれており、社会問題となって対策が求められている。デマとはデマゴギー (demagogy) の略で、元々は政治的な目的で相手を誹謗し、相手に不利な世論を作り出すように流す虚偽の情報である。現在では、社会情勢が不安な時などに発生して、人心を惑わすような憶測や事実誤認による情報や単なる悪口や根拠のないウワサ話、流言飛語もデマの範囲になっている。デマは特定の状況 (不安、怒り等) を狙って発信するのは特徴であり、特定の雰囲気 (パニック状態、恐怖等) を作って、急速に強い感情を誘発するのも顕著な特徴となる。

デマ対策の一環として、デマ検知を素早く行う必要がある。デマは誇張した叙述で民衆を恐怖させ、焦慮させ、感情的な策略を用いて読者の感情を煽動する特徴を持っているため、感情分析によってデマ検知を効果的に行えると期待される。

感情分析 (Sentiment Analysis) は意見マイニング (Opinion Mining) とも呼ばれ、テキストから意見、感情、態度を理解するために使われ、商品販売、医療、健康、政治、金融等への応用が広がってきた。また、近年の研究では映画興行収入と株価の予測、政治選挙における民衆の意見分析などが行われている。本稿では、Opinionはある対象 (意見対象) に対する個人的判断や見解であり、その意見によって暗示される潜在的な整理された感情の傾向は Sentiment と定義する。

デマ検知に関する既存の研究では、機械学習を採用して、結果はデータセットに依存する場合が多い。また、従来の感情分析は感情辞書法と機械学習法が分かれるが、類似感情値をきちんと抽出できないことや、文脈によって異なる単語は誤って判定される問題点が存在し、デマ検知への応用はまだ成熟ではないところが知られている。

そこで、本研究では、感情の極性と強度をに基づく感情スコアを特徴量の一つとしてデマ検知アルゴリズムを開発する。

2. 関連研究

機械学習を用いてデマ情報の分類・推定に関する研究が現在主要な手法となる。Maoら[1]はデータの感情傾向性、拡散過程、ユーザ情報から特徴を抽出し、深層特徴を用いてアンサンブル分類機を作成し、デマを検知する実験を行った。Kwonら[2]はデマ情報に周期性、構造性と言語性

という三つの特徴から分析し、決定木、ランダムフォレストとSVM分類機でデマを検知する実験を行なった。

感情辞書法に基づく感情分析の方法では辞書を通して感情極性と感情強度を抽出する。須田ら[3]は感情極性辞書を用いた収集したツイートの感情極性値を算出し、特徴量の一つとしてデマ識別実験を行なった。牛込ら[4]はSNS上での流言の分類に関する研究では、データを処理する際に、日本語評価極性辞書を含む表現の一部とツイッターデータの頻度解析から辞書データを作成し、また極性辞書に含まれない表現を追加した。Shouら[5]は「食品と養生」、「医学と健康」という二つの領域のデマ情報における感情辞書を基づいてテキスト感情スコアの計算ルールを提案した。しかし、文法が複雑な文章を誤って判定される場合がある。

機械学習法に基づく感情分析の方法は書かれた文章の意見を判断することを指す。一連の単語から判断し、分類することとなる。Bhutaniら[6]はデマ情報とニュース情報の感情極性特徴を学習し、作成した分類機はテストデータの分類を予測した。

3. 提案手法

特徴抽出において単語の相関性が支えられる word2vec を適用し、似たような単語が似たようにエンコードされる表現が得られる。感情極性値より、程度副詞と否定詞を考慮したテキスト全体的な感情スコアを計算し、感情辞書を構築する。また、感情分析の結果をデマ検知に適用し、デマ検知の精度を高める。

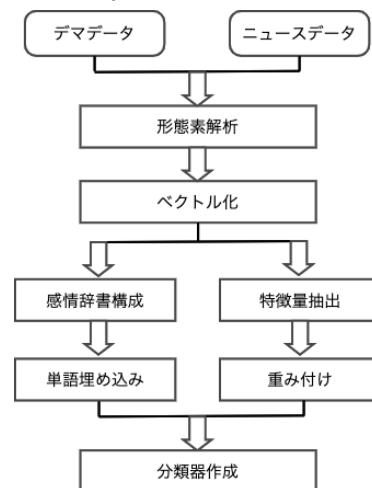


図1 提案手法の流れ

提案手法の流れは図1のように示す。データセットから前処理を行い、感情辞書を構築するにより、感情スコアを計算し、感情スコアとベクトルを特徴量として機械学習で訓練し、デマかデマでないかの分類を行う。

[†]九州産業大学情報科学研究科 Graduate School of Information Science, Kyushu Sangyo University

[‡]九州産業大学理工学部 Faculty of Science and Engineering, Kyushu Sangyo University

3.1 テキストの前処理

本研究の研究対象はテキストデータとなる。データの前処理は三つのプロセスがある。

- (1) テキストの正規化：元のデータセットからストップワード辞書より不要な句読点や emoji、URL などを削除し、新しいデータファイルに格納する。
- (2) 形態素解析：本研究で扱うデータはすべて中国語テキストなので、中国語形態素解析ツール jieba を適用し、前述で正規化したデータを形態素解析する。
- (3) ベクトル化：デマデータは HowNet 辞書に含まれる感情詞と高出現頻度の感情詞から感情辞書を構築する。Word2vec を適用し単語埋め込みでベクトル化を行う。ニュースデータは Bog of words を適用し形態素解析したデータをベクトル化する。TF-IDF を適用し重み付け処理を行う。

3.2 感情スコアの計算

前節で構築した感情辞書から各テキストの感情値を計算し、具体的には以下のように示す。

for 単語 word in words:

If 単語 word in 感情辞書 dic_sentiment:

 ポジティブスコア score_pos +=1

 for w in words [a: j]:

 if w in 程度副詞辞書 dic_degree:

 score_pos *= degree

 if (num(否定詞)%2!=0):

 score_pos *= -1

その中、上述ポジティブスコアの算出例によって、ネガティブスコアも計算できる。

3.3 分類器の作成

ベクトルと感情スコアはデータの特徴量として機械学習で訓練し、分類器を作成する。

4. 予備実験

分類器の作成を行う前に、感情スコアは特徴量としてデマ検知への有効性を検証する。予備実験では、トレーニングデータセットのデマデータとニュースデータを利用して実験を行なった。それぞれのポジティブスコア (Positive Score) とネガティブスコア (Negative Score) より全体感情スコアを計算する。

$$\text{Score} = \text{Positive Score} - \text{Negative Score}$$

4.1 データセット

トレーニングデータは Liu ら[7]の中国ソーシャルメディアにおけるデマの意味解析に関する研究にて収集した 31669 件デマデータと清華大学自然言語処理実験室によって作成したデータセットの「時政」カタログの新聞データ 30000 件を利用する。

テストデータは Song ら[8]の検知モデル CED (Credible Early Detection) の開発に関する研究で Weibo 偽情報報告プラットフォームから収集されたデータのうち、デマデータ 1538 件とデマでないデータ 1849 件を対象とする。

4.2 実験結果

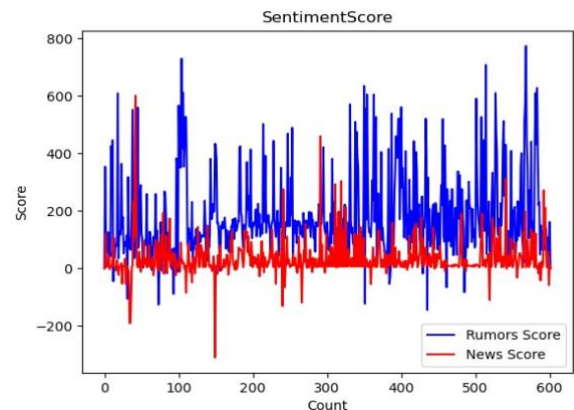


図2 デマとニュースデータの感情スコア

図2で示したように、デマデータの感情の起伏はニュースデータより激しいことがわかる。ニュースデータのスコアは主に 200 Score 以下になり、デマデータのスコアは 200 Score 以上の場合が多い。デマデータの Positive Score が高い理由は拡散したいデマデータは正面的な印象を与えられ、正義感を感じられる言葉がたくさんあると考えられる。それに対して、ニュースデータは感情的な言葉を使う場合が少ない。

5. 終わりに

本研究ではデマ情報の特徴を考察し、感情的な策略を用いて読者の感情を煽動する特徴によって、感情分析からデマ対策への応用について検討する。予備実験では感情スコアはデマ検知への有効性を検証する実験を行った。今後の予定は提案手法の適用性を検討し、感情辞書の構築により、計算した感情スコアと単語埋め込みベクトルを使用し、機械学習により分類機を作成することである。

参考文献

- [1] Mao, E., et al, "Research on detecting micro-blog rumors based on deep features and ensemble classifier", Application Research of Computers 11: 3369-3373(2016).
- [2] Kwon, Sejeong, et al, "Prominent features of rumor propagation in online social media.", 2013 IEEE 13th international conference on data mining, p. 1103-1108. (2013).
- [3] 須田剛裕, et al, "震災時におけるツイッターのトレンドワードと拡散情報を利用したデマ推定の一考察", 第75回全国大会講演論文集 2013.1: 99-100(2013).
- [4] 牛込龍太郎, et al, "ベジアンネットワークによる SNS 上での流言の分類", IEICE Conferences Archives. The Institute of Electronics, Information and Communication Engineers, 2017.
- [5] Huanrong, Shou, Deng Shuqing, and Xu Jian, "Detecting online rumors with sentiment analysis.", Data Analysis and Knowledge Discovery 1.7: 44-51(2017).
- [6] Bhutani, Bhavika, et al, "Fake news detection using sentiment analysis", 2019 Twelfth International Conference on Contemporary Computing (IC3). IEEE, 2019.
- [7] Liu, Zhiyuan, et al. "Statistical and semantic analysis of rumors in Chinese social media." Scientia Sinica information is 45.12: 1536-1546(2015).
- [8] Song, Changhe and Tu, Cunchao and Yang, Cheng and Liu, Zhiyuan and Sun, Maosong, "CED: Credible Early Detection of Social Media Rumors". arXiv preprint arXiv:1811.04175(2018)