

ツイートからの反語抽出手法

11571062 鈴木 凜次 (灘本研究室)

あらまし：ツイートには、様々な疑問文が含まれているが、その中にはツイートをしている人の訴えや意見を含む疑問文がある。本研究ではこのような疑問文を反語と呼ぶ。この反語を含むツイートをしている人の意見を理解せずにリプライや拡散をすることにより、齟齬が生じ炎上する場合がある。そこで、本研究ではこの反語を含むツイートを抽出する手法の提案を行う。

1. はじめに

Twitterには現在日本国内だけでも4,500万人を超える月間アクティブユーザ^[1]が存在しており、この膨大なユーザが様々な会話している。その中で不適切な言動やユーザ間の齟齬により炎上が起こる場合がある。ツイートをを行う際にツイートの言動に注意すると不適切な言動による炎上は未然に防げるが、ユーザ間の齟齬においてはユーザ同士の理解度の違いが関わるため個人で未然に防ぐ事は難しい。

また、ツイートには疑問文が含まれるが、その中にはツイートをしている人の訴えや意見を含む疑問文がある。本研究ではこのような疑問文を反語と呼ぶ。反語を含むツイートは多数ある。この反語を含むツイートをしている人の意見を理解せずにリプライや拡散をすることにより、齟齬が生じ炎上する事がある。例えば、「他人の幸せな部分だけを見て、羨ましがすることに意味はあるのか?」というツイートは、疑問文だけでユーザの意見は明示されていない。そしてこのツイートを読むユーザは肯定、否定など様々な読み取り方をする。ツイートの解釈の違いによりユーザ同士が衝突し炎上する場合がある。

そこで、反語に対してアラートを示すことにより、このように炎上を阻止できると考え、本研究では、ツイートから反語を抽出する手法の提案を行う。

2. Twitter から反語の抽出

2. 1. 反語の定義

反語の例文を表1に示す。我々はクラウドソーシングにより反語を人手で抽出し分析を行った。具体的には10,000のツイートを被験者10人が反語か反語でないかを判定した。その結果1,750の反語を抽出することが出来た。次にこの1,750の反語を人手で分析し、以下の3つの

特徴を抽出した。

特徴1. 疑問文を含む。

筆者の訴えや意見を強調するために用いられる反語には疑問文が必要になる。そこで、反語の性質として、ツイート内に疑問文が必ず存在している。

特徴2. 意見、解説を含む。

反語は筆者の訴えや意見を疑問文によって強調する表現方法のため、ツイートに筆者の意見や解説を含んでいる。意見の例としては、「労働力減少というみでも社会的損失だと思う。」といったような主観的な意見や訴えかける文である。解説の例としては、「大統領とはその国の顔となる存在」といったような事実に基づいた文や著者の思考の根拠になる文である。

特徴3. 文章中で感情の反転が起きている。

筆者の意見を強調する反語では、例えば、「みんなカレーが大好きだ！カレーが嫌いな人などいるのか?」といったようなカレーに対しポジティブな意見のあとにネガティブな疑問文が続くツイートがある。このように反語は、ツイート中でポジティブとネガティブの両方が存在することにより感情の反転が生じている。

上記の3つの特徴から反語を抽出するために以下の2つの条件を示す。

条件1. 疑問文に意見を含む。

筆者の意見を強調するために疑問文中に自身の意見を含む場合がある。よって疑問文に意見を含む場合、反語として定義する。

条件2. ツイート中に疑問文を含むが、疑問文には意見を含まない。かつ、ツイート内に感情の反転がある。そして、疑問文以外の文に意見または解説を含む。

ツイート内には疑問文を含むが、その中には意見を含まない場合、そのツイート内に感情の反転が

表1：一般的な反語の例

肯定・否定の反転あり	この図書館は素晴らしい！不満がある人なんているのか？
肯定・否定の反転なし	この図書館は大きい！どれほどの本があるのか？
皮肉など	この図書館は立派だ。さぞ多くの本があるんだろうな。

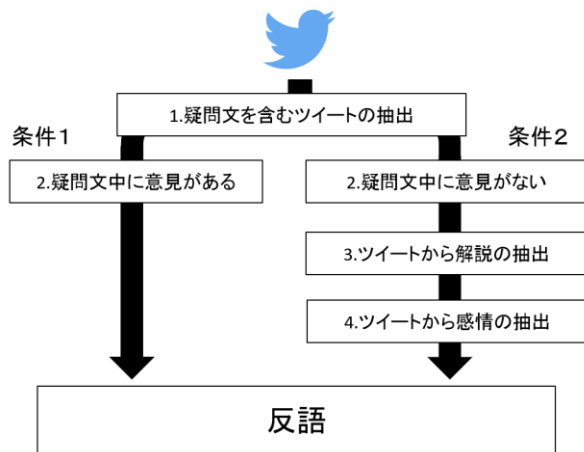


図 1：反語抽出の流れ

あり、かつ、疑問文以外の文に意見または解説を含む場合、反語とする。

本研究では、この 2 つの条件に基づき反語の抽出を行う。

2. 2. 反語の抽出手法

図 1 に反語の抽出の流れを示す。以下、この流れに沿って各々の手法について述べる。

疑問文の抽出手法

Twitter には、くだけた表現が用いられるため、例えば、「明日は雨なの？」といったように「？」を使うことが多い。又、「明日は雨か。」のように終助詞「か」を用いる場合も多い。そこで疑問を表す「？」や疑問の終助詞「か」、「かな」、「かしら」を含んでいる文を疑問文とする。

意見文の抽出手法

意見の抽出に NICT により開発された意見（評価表現）抽出ツール^[2]を用いる。このツールは 1 行につき 1 文が書かれたテキストファイルを入力とし、その文に意見が存在すると判定した場合、評価表現・評価タイプ・評価極性・評価保持者を出力する。今回意見抽出ツールにて意見が存在すると判定された文を意見を含む文とする。

さらに、「思う」、「感じる」、「考える」という著者の思いや考えていることを示すために用いる動詞を含む文も意見文として抽出する。

解説文の抽出手法

ここでいう解説文とは上記意見抽出手法で抽出できないツイートから著者の主観的な思考やツイートの根拠とされる部分を含む文である。例えば、「大統領とはその国の顔となるべき存在である。」、「この世の中には英語をしゃべらない人間だって大勢いる。」といった「とは」、「には」などの機能表現や強い使役を示す「べき」、理由を示す「だから」を含む文である。よって、これら「とは」、「には」、「べき」、「だから」を含む文を解説文とする。

感情抽出手法

ツイートの感情抽出には山本ら^[3]が提案している感情語辞書を用いる。本論文では、感情の反転を得るため感情語辞書をポジティブ、ネガテ

表 2：実験結果

適合率	0.54
再現率	0.16
F 値	0.24

ティブの 2 感情に分類し使用する。また感情抽出は文毎に行う。

3. 実験

提案手法により反語が抽出できたかどうかの実験を行った。

3. 1. 実験手順

無作為に抽出した 10,000 件のツイートから提案手法を用いて反語を抽出し、適合率、再現率、F 値をそれぞれ求める。ここで、正解データはクラウドソーシングで反語と判定された 1,750 件のツイートをを用いる。

3. 2. 実験結果と考察

適合率、再現率、F 値の結果を表 2 に示す。本研究では、反語に起因する炎上を防ぐはじめの一歩とするため、再現率が重要であるが、再現率が非常に低い結果となった。これは、意見、解説の抽出に用いているルールが簡易的なルールであったためと考えられる。例えば、「カレーは日本人の国民食になるほど親しまれていると思う。」という文ならば「思う」を含んでいるため提案手法のルールを用いて抽出されているが、「私はカレーが三日経ても平気！むしろ嬉しい！」といった筆者の主観的な思考が入っている文の場合に提案手法のルールでは反語と判定されない。そこで疑問文の抽出に用いたルールは多くの正解を含む疑問文を持つツイートの抽出ができていたため、意見、解説の抽出に用いているルールの改善を行うことで再現率、さらには適合率の向上につながるのではないかと考える。

4. まとめと今後の課題

本研究では、Twitter 上に存在する反語を抽出する手法を提案した。反語の抽出には感情分析と意見抽出ツールを用いた。結果として提案手法による反語の抽出の精度は良いとは言えなかった。今後の課題としては、提案手法の意見、解説の抽出を行う際の抽出手法の再考が必要である事がわかった。また、さらに反語の分析を行うことで文の構成を用いた反語の抽出方法の検討が考えられる。

参考文献

- [1] 「Twitter Japan のツイート: 月間利用者数」, <<https://twitter.com/TwitterJP/status/923671036758958080>>2019 年 1 月 21 日アクセス
- [2] 「意見（評価表現）抽出ツール」, <<https://alaginrc.nict.go.jp/opinion/index.html>>2019 年 1 月 17 日アクセス。
- [3] 山本 湧輝, 熊本 忠彦, 灘本 明代, “ツイートの感情の関係に基づく Twitter 感情軸の決定”. 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), F5-2, 2015.