

反語を含むツイートの抽出手法

鈴木 凜次[†] 灘本 明代[†]

[†] 甲南大学 知能情報学部 〒 658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]s1571062@s.konan-u.ac.jp, ^{††}nadamoto@konan-u.ac.jp

あらまし ツイートには様々な疑問文が含まれている。疑問文が含まれているツイートの中にはツイートをしている人の訴えや意見を含むものがある。本研究ではこのような文章を反語と呼ぶ。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しアラートを出すことにより、このような炎上を未然に防ぐことを目的とする。具体的には、ツイート内の疑問文に着目し、(1) 疑問文に著者の意見を含むツイートと (2) ツイート中に意見を含まない疑問文がある。かつ、疑問文以外の文に意見または解説を含む。そして、ツイート内に感情の反転があるツイートの2つのツイートを反語を含むツイートと呼びその抽出手法を提案する。

キーワード Twitter, 反語, 意見マイニング, 感情マイニング

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日々様々なやり取りを交わしている。しかしながら、簡単に誰もが意見を発信できることにより、ユーザはトラブルに巻き込まれる可能性が増えている。炎上は不適切な言動のみならずユーザ間に齟齬を生じることにより起こる場合がある。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場合が多数あるが、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、ユーザ間に齟齬が生じ炎上する場合がある。例えば、「犬の飼い主が飼い犬を蹴っているが、こんな行為が犬のしつけといえるのか？」というツイートは、飼い主の行動が異常だったという情報を得ていない人には疑問文だけで意見が読み取れない。このツイートを読むユーザは疑問文に対して肯定的な意見、または否定的な意見を持っているなど様々な読み取り方ができる。この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。

一方、一般的な表現法としての反語がある。反語には本来の意味とは逆の意味を含ませるものすべてを指す。これは皮肉や揶揄、揚げ足取りなど多くの表現法を含み、反語を用いることで他者との会話の中で自分の伝えたい意見と文面上での意味との違いが生じて話の面白みにつながる。その反語がツイートに含まれる場合、ツイートを読むユーザが反語と理解して解釈し

なければ齟齬が生じる。しかしながら、すべての反語を対象とし、反語の自動抽出を行う場合、あまりに様々な表現法の文章が対象となる。そこで、本研究では反語を抽出するはじめての一步として、皮肉や揶揄を含む反語を対象としない。具体的には、ツイート内の疑問文と著者の意見に着目し、(1) 疑問文に著者の意見を含むツイートと (2) 著者の意見を含まない疑問文があり、かつツイート内に感情の反転があり、疑問部に外の文に意見を含むツイートの2つのツイートを反語を含むツイートと呼びその抽出手法を提案する。

以下、2章で関連研究について述べる。3章で、Twitter から反語を抽出する手法について述べ、そして4章で手法に対する実験、考察について述べる。最後に5章で、まとめと今後の課題について述べる。

2 関連研究

反語、意見抽出に関する研究は数多くされている。

磯野ら [1] は機械学習の観点から、Web 掲示板から皮肉の分類と自動検出する手法の提案をしている。本研究では、皮肉ではなく反語に着目している点、Twitter から反語を含むツイートを抽出しアラートを出すという点で異なる。

古舞ら [2] は意見文のベクトル化および意見文のベクトル化手法を用いた反論生成手法の提案を行っている。本研究では、反語という意見の齟齬を生じる表現によって読み取られる意見が様々になってしまうものに対しアラートを出し、炎上を防ぐという点で異なる。

藤村ら [3] は単語 n-gram の素性を用いることで文の構造を考慮した評判抽出手法を提案している。本研究では、意見抽出とともに疑問文である反語に対しアラートを出しユーザ間で齟齬による衝突を防ぐという点で異なる。

立石ら [4] は Web 文書集合から2つの辞書と共起パターンを用いることで意見抽出する手法の提案をしている。本研究では、Web 文書よりも砕けた表現である Twitter を用いる点、疑問

1 : <https://twitter.com/TwitterJP/status/923671036758958080>:Twitter のアクティブユーザ数

表 1 反語の例文一覧

例文
(a) この図書館は素晴らしい！不満な人などいるのか？
(b) この図書館は大きい！どれほど本があるのだろう？
(c) この図書館は立派だ。さぞ多くの本があるのだろう。

文の反語に注目し意見を抽出するという点で異なる。

飯田ら [5] はウェブ上のレビューやブログから意見抽出を目的とした機械学習による属性の付与を行うことで対となる評価値を抽出する手法を提案している。本研究では、意見とともに肯定・否定を感情の反転として考慮することで反語を抽出する点で異なる。

3 Twitter から反語の抽出

3.1 反語の定義

Twitter には疑問文を含むツイートが多数ある。一方、ツイートの中には他者に訴えかける言い方をするものや自分の意見を発信するものも多く存在している。疑問文かつ筆者の訴えや意見を含むツイートのことを本研究では反語と呼ぶ。また、一般的な反語としては多くのものが含まれる。文に疑問文を含むツイートや含まないツイート、ツイート内で肯定・否定を反転させるものや反転しないもの、また、皮肉や揶揄表現も反語に含まれるとされている。ここで反語の例文の一部を表 1 に示す。本研究では 1 の (a)(b) をとし、(c) の皮肉を含む文は対象としない。

実際に我々は人が反語であると判定したツイートを分析する事により、反語の特徴を抽出した。実験は、ランダムに抽出した 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。実験にはクラウドソーシングを用いた。その結果 1,750 の反語を抽出した。次にこの 1,750 の反語を手で分析し、以下の 3 つの特徴を決定した。

特徴 1： 疑問文を含む。

実験結果の反語にはすべて疑問文が含まれていた。これは、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含まれると考えられる。そこで、反語の特徴として、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、「労働力減少というい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 つの特徴から本研究で対象とする反語を以下の 2 つの定義からなる。尚、本研究で対象とする反語は定義 1 または定義 2 の物とする。

定義 1： 疑問文に意見を含む。

筆者の訴えや意見を強調するために疑問文中にも自身の意見を含むことがある。よって疑問文中に意見を含む場合、反語として抽出する。例えば、「好きになったら特別な 1 人になりたい独占したいって考えおかしいのかな？」や「そんなこ焼きが嫌いな大阪県民なんているのかなあ？」といったように著者の意見、訴えを疑問文に含んでいる。そして、その答えを相手に求めることによって読者に共感を与え強い説得力を生んでいる。

定義 2： ツイート中に意見を含まない疑問文がある。かつ、疑問文以外の文に意見または解説を含む。そして、ツイート内に感情の反転がある。

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、疑問文以外の文に意見または解説を含み、かつツイート内に感情の反転がある場合、反語とする。例えば「イチローのように大業をなす人は人のことを批判せず、常に前向きに自分と向き合っている人である。私はそんな人が大業をなすと考える。人のことばかり批判している人は心が狭くなり大業をなせない。なぜなら、イチローは野次に反応しないし、気にもならないはず。イチローは、外野の客席にいる酔っぱらいオヤジの野次に反応するだろうか？外野のアドバイスをまともに受けるだろうか？」である。この例文では「イチローは、外野の客席にいる酔っぱらいオヤジの野次に反応するだろうか？外野のアドバイスをまともに受けるだろうか？」の 2 つの疑問文には意見が含まれていない。しかしながら、この疑問文以外の文に「私はそんな人が大業をなすと考える。」にツイートした人の意見が含まれている。さらに、このツイートの前半部「大業をなす人は人のことを批判せず、常に前向きに自分と向き合っている人である。」は「批判」を否定し、「前向きに」「向き合う」によってポジティブである。それに対し後半部の「人のことばかり批判している人は心が狭くなり大業をなせない。」は「批判」「心が狭い」や「大業をなせない」ことよりネガティブである。よって、このツイートはツイート内でポジティブ、ネガティブの反転が起きている。このように、意見を含まない疑問文を含むツイートであっても、疑問文以外の所にツイートをしている人の意見が含まれ、かつツイート内で感情の反転が起きているツイートも反語と定義する。

本研究では、この 2 つの定義に基づき反語の抽出を行う。

3.2 反語の抽出手法

図 1 に反語抽出の流れを示す。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 には話し言葉のような砕けた表現が用いられるため、より一般的な疑問を表す記号や助詞を用いる。例えば、「明日は雨なの？」、「この作品の作者は何を考えて作ったのかな？」といったようなツイートである。そこで、疑問を表す記号の「？」

表 2 反語候補ツイート

ツイート
・ 一家に一台たこ焼き機！子供のころから食べてた！なんて大阪ではよく聞くけど、そんなに一緒に育ってきたようなたこ焼き。そんなたこ焼きが嫌いな大阪県民なんているのかなあ？ ・ 他人を羨む前に、その人が抱えている悲しみを知るべし幸せそうに見える人も、他人であるあなたの知らない多くの問題を抱えている。相手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だろうか？自分のダメな部分だけを見て、落ち込むことに意味はあるのだろうか？ ・ 好きって言われた言葉信じててもやっぱり続かなくて嫌われて……俺がヤキモチやくから悪いんだけど…好きになったら特別な 1 人になりたい独占したいって考えおかしいのかな？ ・ 「将来が不安でたまらない」と考えることによるメリットって、いったい何があるのか？私は、不安というのは、考えた時点で負けだと思っている。先の不安を考えて、いいことがあるのだろうか？何もない。なのに、なんでわざわざ楽しくないことを考える必要があるのだ。

表 3 実験結果

適合率	再現率	F 値
0.54	0.16	0.24

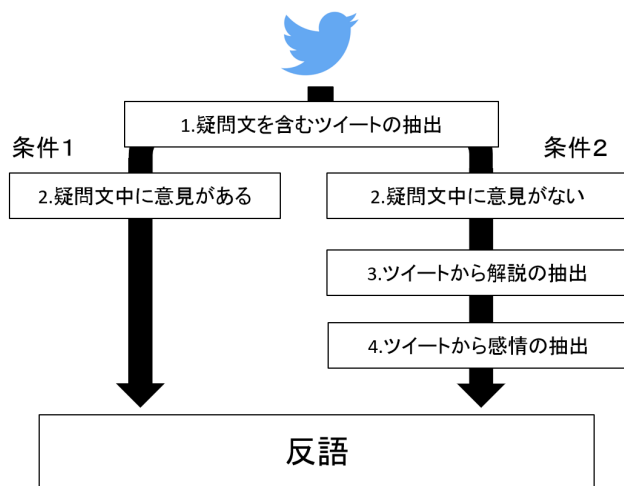


図 1 反語抽出のフロー

や疑問の終助詞「か」「かな」「かしら」を文末に含んでいる文を疑問文とする。

意見文の抽出

意見の抽出に NICT によって開発された意見（評価表現）抽出ツール [6] を用いる。本ツールは 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、その文に評価情報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。

しかしながら、実際に意見抽出ツールのみではすべての意見文を抽出することができなかった。そこで、「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

ここでいう解説文は著者の主観的な思考や意見だけでなく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

ツイートの感情抽出は様々な手法があるが、本研究では辞書を用いた感情抽出を行う。辞書は、山本ら [7] [8] が提案している感情語辞書を用いる。山本らの辞書は 10 次元の感情語辞書であるが、山本らが提案している 10 次元の感情をツイート用にポジティブ、ネガティブの 2 次元に次元削減した手法を用いる。これら辞書を用いて、文単位に単語のポジティブネガティブの各々の単語の総和の内、値の大きい感情をその文の感情とする。

4 実 験

提案手法により反語が抽出の精度を測る実験を行った。

4.1 実験手順

反語の定義の実験の時に用いた、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 件のツイートである。実験に用いたデータの一部を表 2 に示す。

4.2 実験結果と考察

適合率、再現率、F 値の結果を表 3 に示す。本研究では、反語に起因する炎上を防ぐことを目的としているため、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。本論文で抽出された反語の例、反語の正解データであるが本論文では抽出されなかった反語の例を表 4 と表 5 に示す。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。」「一人ひとり考え方も捉え方も違うとは思うよ。」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！」「あっさりチームを捨てたような人を信じられるのか？」といった筆者の主観的な思考が入って

表 4 抽出された反語ツイート

ツイート
・安定した職業につきなさい！周りの人はみんないいいます！安定した職業とはなんですか？公務員？大企業？ いくら大企業だって、100%潰れないとはいきれないし、僕の親の年金すら危うい日本財政をみたら公務員だって果たして安定なのか？ 確実に安定した職業などないのです！
・お酒は飲んでも呑まれるなっていうけど、飲み会の席って楽しいし、話してるとついついお酒が進んじゃって… 自分の不注意が原因なんだろうけど、1回は失敗あると思うんだよねwお酒で1回も失敗したことない人っているのかな？ いたらすごい自己管理ちゃんとしてるんだろうな…
・彼らが何かする訳があろうか。いや無い。(反語)むしろ向こうは「何かしてもらうことに慣れきってしまった人」である可能性が非常に高い。 だから今回もちょっとたしなめた振りをしてうまくやり過ごして許してもらおうとしてたんじゃないか。弱者の皮を被った強者はまことに質が悪い。
・人の気持ちを考えるって、そんなに難しいことだろうか。相手の立場になって考える。それができないほど人間って想像力ないかな。 なんか心無い言葉を見過ぎてうんざりしてる。一人ひとり考え方も捉え方も違うとは思うよ。でも、それを他人に投げつけるのは正解なのか？

表 5 抽出できなかったツイート

ツイート
・名古屋グランパスはベンゲルにアドバイザーになってもらうとか言っているが、いいのか？ 96年の初優勝がかかった一番大事な時期に、アーセナルのオファーを受けてあっさりチームを捨てたような人を信じられるのか？ 監督として戻って来て、今度こそ契約を完了させてくれるなら別だが……
・ある分野の第一人者、たとえば優秀なスポーツ選手や有名な芸能人が畑違いの事柄にまで首を突っ込んで、一般大衆に身の施し方をあれこれ指南しようとする風潮に、私は強い違和感を覚える。 我々が投資で大きな利益を上げられるからといって、世の中のあらゆる事柄に関して良い助言ができるだろうか？
・今後公務員数は確実に減る。公務員が減るということは、教員数が減る。教員数が減ると日本の教育レベルも全体低に落ちます。 教育レベルが落ちるとバカな子供が増えます。バカな子供が増えると犯罪が増えて治安が悪くなる。数年後の日本は果たして大丈夫なのか？
・作中最強はターちゃんの設定しているのがかなり上手く機能している。梁師範もペドロも安易に嘯ませ役にはなってない。 ターちゃんクローンという作中準最強キャラを出した後は一回ギャグに戻したのもクレバー。 ギャグ→バトル串団子→ギャグときちんと円環させたジャンプ漫画なんて他にあるだろうか。

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。また、本研究では反語を含むツイートの中にすべて意見が含まれるとしていたが、自分の意見を説得させる根拠として一般的な常識や現実にあったニュースを用いている場合があったため、考慮する必要があると分かった。今回の結果から適合率、再現率がともに低いことから、本研究において用いているそれぞれの手法のどれだけ反語に対して有効であるのか改めて確認することが必要だとわかった。そこで疑問文、意見（評価表現）抽出ツール、意見・解説に用いたルールのそれぞれに対して精度を算出し、改善点を洗い出しを行うべきだと思う。また、意見抽出ツールがどれほどTwitterの碎けた表現に対応できているかという点を考慮するべきだと分かった。

5 まとめと今後の課題

本研究では、ツイート内の疑問文に着目し、(1) 疑問文に著

者の意見を含むツイートと(2) ツイート中に意見を含まない疑問文がある。かつ、疑問文以外の文に意見または解説を含む。そして、ツイート内に感情の反転があるツイートを反語と定義し、反語の抽出手法の提案を行った。

具体的には、ツイートからの疑問文の抽出、意見抽出ツールを用いた意見抽出、及び感情語辞書を用いた感情分類を用いて反語の抽出手法の提案を行った。提案手法の精度を求める実験を行ったが、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。

謝 辞

本論文の一部はJSPS 科研費17K00430, 16K07973 及び、私学助成金(大学間連携研究補助金)の助成によるものである。ここに記して謹んで感謝の意を表する。

文 献

- [1] 磯野 史弥, 松吉 俊, 福本 文代, “Web 掲示板における皮肉の分類および自動検出”. 研究報告自然言語処理 (NL), 2013-NL-213, 7,1 - 8,2013-09-05.
- [2] 古舞 千暁, 有木 康雄, 滝口 哲也, “議論システムにおける賛成／反対意見の生成手法の検討”. SIG-SLUD, B5, 02, 82 - 83, 2018-10-15 .
- [3] 藤村 滋, 豊田 正史, 喜連川 優, “文の構造を考慮した評判抽出手法”. 電子情報通信学会第 16 回データ工学ワークショップ, 2005, 2005 , 2005 .
- [4] 立石 健二, 福島 俊一, 小林 のぞみ, 高橋 哲朗, 藤田 篤, 乾 健太郎, 松本 祐治, “Web 文書集合からの意見情報抽出と着眼点に基づく要約生成”. 情報処理学会研究報告情報学基礎 (FI), 2004 , 93(2004-FI-076) , 1 - 8 , 2004-09-16 .
- [5] 飯田 龍, 小林 のぞみ, 乾 健太郎, 松本 祐治, 立石 健二, 福島 俊一, “意見抽出を目的とした機械学習による属性-評価値対同定”. 情報処理学会研究報告自然言語処理 (NL), 2005 , 1(2004-NL-165) , 21 - 28 , 2005-01-11 .
- [6] T. Nakagawa, “意見（評価表現）抽出ツール”, 2011, <https://alaginrc.nict.go.jp/opinion/index.html> [Online; accessed 12-February-2019].
- [7] 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 .
- [8] Yuki Yamamoto, Tadahiko Kumamoto, Akiyo Nadamoto, “Followee Recommendation Based on Topic Extraction and Sentiment Analysis from Tweets”, International Conference on Information Integration and Web-based Applications & Services(iiWAS2015), pp. 215-225, 2015.