一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
LOIS2016-86(2017-03)

Twitterにおける感情分析を用いた炎上の検出と分析

高橋 直樹† 　檜垣 泰彦‡

†‡千葉大学工学研究科 〒263-8522 千葉県千葉市稲毛区弥生町 1-33
E-mail: †afna3267@chiba-u.jp, ‡higaki.yasuhiko@faculty.chiba-u.jp

あらまし　近年 SNS，特に Twitter において投稿に対して非難や誹謗，中傷のコメントが殺到する炎上という現象が発生している．本研究では炎上の定義に則り，Twitter での投稿（ツイート）に対するコメント（リプライ）に着目し，リプライの感情を分析することで炎上を検出することを目的とした．本手法では単語が持つ極性値の，修飾語の影響による変化や単語位置による変化を考慮した感情分析を行った．実験により，炎上ツイートにリプライを送るユーザは非フォロワーが多く，この属性とリプライの感情を検出条件とすることで高い精度で炎上ツイートを検出できることが確認できた．

キーワード　炎上，感情分析，自然言語処理

Flaming Detection and Analysis using Emotion Analysis on Twitter

Naoki TAKAHASHI† 　Yasuhiko Higaki‡

†‡Graduate School of Engineering, Chiba University　1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 263-8522 Japan
E-mail: †afna3267@chiba-u.jp, ‡higaki.yasuhiko@faculty.chiba-u.jp

Abstract　In recent years, a phenomenon called flaming in which post is flooded with criticisms and condemnations has occurred many times, in Social Networking Service (SNS), especially Twitter. The purpose of this study is to detect flaming tweets by analyzing emotions of reply. This emotion analysis method takes into account changes in polarity value by influence words and position of word in sentence. Experiment on flame detection and analysis of the reply show that users sending reply to flaming tweets have many non-followers, so by using this attribute and emotion of the reply as the detection condition, it is possible to detect the flaming with high accuracy.

Keywords　Flaming, Emotion Analysis, Natural Language Processing

1. はじめに

近年，Twitter[1]や Facebook[2]，Instagram[3]などのソーシャルネットワークサービス（SNS）の日本でのユーザ数は増加し続けており，国内インターネットユーザーにおける SNS の普及率は 2016 年末には 69.3%になると見込まれている[1]. SNS をユーザ同士のコミュニケーションのツールとしてだけでなく，企業や自治体が情報を発信する場として利用する動きが盛んになっており，今後も SNS 利用者数は増加していくことが予想される．しかし一方で，投稿に対して想定を大幅に超えた非難や批判，誹謗・中傷のコメントが殺到する「炎上」という現象[2]が SNS，特に Twitter において多数発生しており，社会問題の一つとなっている．

Twitter での炎上に関する研究はツイート内容とその世評との違いによる炎上事例について炎上予測を行うもの[3]や，投稿する際にツイート内の単語の持つ極性を特徴量として SVM で学習したモデルを用いた炎上予測を行うもの[4]あるが，炎上している投稿（ツイート）に対するコメント（リプライ）が持つ感情に着目した研究は行われていない．そこで本研究では，炎上の定義である「投稿に対して非難や批判，誹謗・中傷のコメントが殺到すること」[2]に則り，感情分析により非難や中傷などのリプライを判別し，そうしたリプライが殺到しているかどうかで炎上を検出することを目指した．

感情分析手法には三和ら[5]の提案したツイートの感情分析手法を元に，課題として挙げられていた，部分を否定する表現がツイート全体の否定となってしまう点や顔文字に対応できていない点の改善を図った．

2. 提案手法

2.1. 概要

本手法の炎上検出までの流れを図 1 に示す．ツイート及びそれに対するすべてのリプライを取得する．その際に投稿内容だけでなく，お気に入り数やリツイート数，フォロワー数など Twitter 特有の情報も取得す

[1] Twitter, https://twitter.com/
[2] Facebook, https://www.facebook.com/
[3] Instagram, https://www.instagram.com

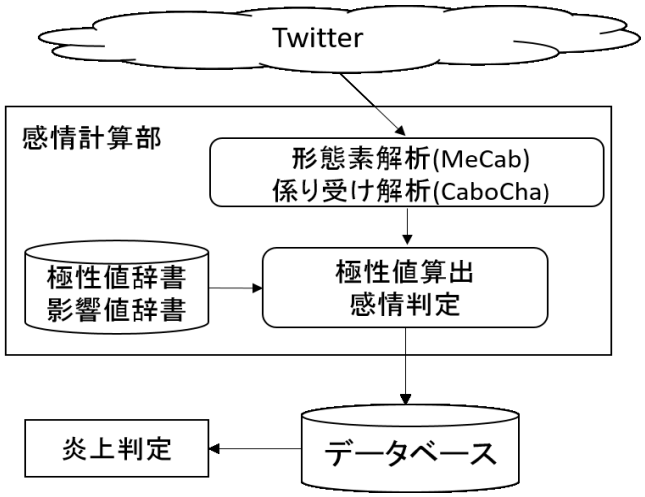

図 1. 炎上検出の概要

る．次に感情計算部において提案する感情分析手法を用いて各リプライの感情を判定する．この感情とは，ポジティブ（肯定的・楽観的・積極的・好意的）からネガティブ（否定的・悲観的・消極的・敵対）の一つの軸であるとする．こうして得られた感情やデータから炎上の判定を行う．

2.2. 感情分析

リプライが持つ感情を分析するため，リプライの文章の極性値を計算する.計算に利用する形態素と辞書，および極性値の計算手法について述べる．本手法では前述した三和ら[5]の極性値算出手法をもとに，係り受けや顔文字を考慮した極性値の算出を行った．

2.2.1. 事前準備

Twitter での投稿内容に使われる表現として，ユーザ名・ハッシュタグ・URL がある．ユーザ名は，ユーザ ID とは別にユーザが自分で決めることが出来る名前であり，半角英数字およびアンダーバー(_)を用いた合計 15 文字以内の文字列である．ハッシュタグとは Twitter においてツイートのタグとして用いられる文字列である．シャープ(#)と英数字や日本語で構成される．URL は Web ページへのリンクや画像・動画を投稿する際にツイート内に現れる．リプライの極性値を計算するにあたり，これらは極性値を持つとは考えにくい．したがってリプライの文章中からこれらに該当する部分を取り除く．

表 1. ユーザ名，ハッシュタグ，URL の除去例

元の文章	@you おはよう！ http://test.com/test1 #テスト
除去後の文章	おはよう!

2.2.2. 辞書の作成

・極性値辞書

本手法では日本語評価極性辞書（用言編）[6]，日本語評価極性辞書（名詞編）[7]をベースとして極性値辞書を作成した．日本語評価極性辞書は単語にポジティブ，ネガティブ，どちらでもないの 3 種類に，評価極性がそれぞれ 1.0，-1.0，0 で付与されているものである．この辞書を元に，顔文字を追加した極性値辞書を作成した．顔文字はその投稿の感情を表す表現の一つと考えられるため，顔文字を感情語と見なすこととした．Twitter API によりランダムで取得したツ約 10 万件のツイートの文章中において正規表現「(¥(．*?¥))」にマッチしたものを顔文字とし，その中でも 10 回以上出現した 164 個の顔文字を人手で極性値を付与し極性値辞書に追加した．追加した顔文字とその極性値の例を表 2 に示す．

表 2. 追加した顔文字の例

顔文字	極性値	顔文字	極性値
(*^_^*)	1.0	(≧▽≦)	1.0
(-_-)	-1.0	(´；ヮ；｀)	-1.0
(・ω・)	0	(°_°)	0

・影響値辞書

「とても」や「少し」などの程度や量を表す語や「ない」などの否定を表す語によって修飾される単語の極性値が変化することを考慮し影響値辞書を作成した．極性値を持つのは名詞と用言であることから，影響語となるのは主に副詞であると考えた．そこで影響語の選定に当たっては田和[8]の程度副詞の分類を参考に，また副詞以外にも影響語となり得るいくつかの語を人手で選定し 4 段階の影響語に分類した．分類例を表 3 に示す．

表 3. 影響語の分類例

影響度合い	単語例	影響値
極大	とても，本当に　など	2.0
大	かなり，だいぶ など	1.5
小	少し，ちょっと など	0.5
打消し	ない，ぬ	-1.0

さらに，文中にしばしば現れる記号の中でも表 4 に示す 3 つを新たに極性値辞書または感情値辞書に追加した．

感嘆符「!」は文の強調の意図で利用されることが多いため，文の感情値を強める作用があると考えた．疑問符「?」は文の内容に対する疑問を表すため，文の感情を弱める作用があると考えた．音符「♪」は「嬉し

表 4. 辞書に追加した記号

記号	追加した辞書	極性値または影響値
感嘆符「!」	影響値辞書	1.5
疑問符「?」	影響値辞書	0.5
音符「♪」	極性値辞書	1.0

い」や「楽しい」などポジティブな感情を表す際に利用されることが多いため，ポジティブな極性を持つ感情語であると考えた．

2.2.3. 形態素解析・係り受け解析

本手法では形態素解析に MeCab，係り受け解析に CaboCha を利用する．形態素解析に利用する辞書は IPA 辞書とともに，mecab-ipadic-NEologd と呼ばれる辞書を併用した．これは IPA 辞書を拡張したもので，Web 上の様々な言語資源から得た新語や固有表現を追加しており，Twitter 上で利用されやすい流行語やネット用語などに対応できると考えられるからである．

2.2.4. 極性値の算出

文章中の全単語数を length, 単語番号を i, j(i, j=1, 2, …, length)とする．単語極性値 b_i は極性値辞書に含まれていれば対応した極性値，含まれていなければ 0 である．また，単語影響値 e_{ij} は，i 番目の単語に対する j 番目の単語の影響値であり，影響語辞書に含まれていれば影響語辞書に対応した影響値，含まれていなければ 1.0 である．式中のパラメータ α，β，μ については，α=1，β=0.1，μ=1 とした．

既存手法では影響値を文全体に掛け合わせているが，本手法では影響語は文全体ではなく係り受け先の単語に影響値を掛け合わせる．文章中の i 番目の単語に対する影響値 e_i を，全単語の i 番目に対する影響値の積で表す．

$$e_i = \prod_{j}^{length} e_{ij} \tag{1}$$

また，日本語の文において，文の極性値を決定付け得る動詞や形容詞などの述語が文末に近い位置に出現しやすいことを考慮するため，文末に近いほど単語の持つ単語極性値 b_i に対して重みを加える．これを補整単語極性値 b_i として求める．

$$b'_i = e_i \times b_i \times exp\left(-\left(\frac{i}{length} - \mu\right)^2\right) \tag{2}$$

リプライ文中の各単語が持つ極性値を合計し，さらにポジティブな極性を持つ単語数を c_p，ネガティブな単語数を c_n として，極性値を持つ単語数で調整することで，リプライの極性値 a を求める．また，s は文に感嘆符「!」または疑問符「?」が存在する場合それぞれの影響値，存在しない場合は 1.0 である．

$$a = \frac{\sum_i^{length} b'_i}{\alpha c + \beta\left(length - (c_p + c_n)\right)} \times s \tag{3}$$

$$c = \begin{cases} 1, & (c_p = c_n \text{のとき}) \\ |c_p - c_n|, & (c_p \neq c_n \text{のとき}) \end{cases} \tag{4}$$

2.2.5. 感情の判定

求めたリプライの極性値 a によって，表 5 のように感情を割り当てる．

表 5. 感情の判定

感情	極性値
ポジティブ	$a \geq 0.6$
少しポジティブ	$0.2 \leq a < 0.6$
どちらでもない	$-0.2 < a < 0.2$
少しネガティブ	$-0.6 < a \leq -0.2$
ネガティブ	$a \leq -0.6$

以上が感情分析の手法である．

2.2.6. 実験と評価

前述した感情分析手法の評価を行うため，実験を行った．Twitter API を利用し無作為に 177 件のツイートを取得した．1 人当たり 30 件，24 人の学生にこれらのツイートが持っていると考えられる極性値を「ポジティブ」「少しポジティブ」「どちらでもない」「少しネガティブ」「ネガティブ」の 5 段階で評価してもらった．得られた結果から各ツイートの平均極性値 aavr をそのツイートの極性値とし，表 5 を参考に 5 段階の感情を振り分け，これを正解データセットとした．

また，同じツイート群に対して提案手法を用いてツイート極性値の算出を行う．結果から先ほどと同様に感情を判定し，正解データと同じ感情となれば正解とした．同様の辞書およびデータを用いて三和ら[5]の既存手法で感情分析を行った結果と合わせて表 6 に示す．

表 6. 各手法の正解率

既存手法	提案手法
0.542	0.683

提案手法は既存手法と比較して約 14%正解率が高い結果となった．これは，係り受け解析を行うことで影響値のかかる単語を正確に判別し単語単位での影響を考慮した点や，顔文字を極性値辞書に追加した点などによって，既存手法の課題を解決することが出来たからであると考えられる．

2.3. 炎上の検出

感情分析によって判断された感情を用いて炎上検出を行う．本手法では，炎上の定義に則りリプライの内ネガティブなものがポジティブなものより多いツイートを炎上ツイートとして検出することとする．

3. 実験 1

3.1. 対象データ

炎上検出実験を行うための対象となるデータを収拾した．

ネットニュースで炎上とされているツイートを炎上ツイートとして 7 件，Twitter より「リプライが 10 件以上ある」という条件を満たしたツイートを非炎上ツイートとして 93 件，合計 100 件と，またそれぞれに対するリプライをデータベースに保存した．

3.2. 結果

炎上と判断されたのは 13 件で，その内の 10 件が誤りであった．本手法による炎上ツイートの検出精度は表 7 のようになり，あまり良い精度とは言えなかった．

表 7．提案手法の精度

	適合率	再現率	F 値
提案手法	0.232	0.429	0.202
提案手法（改善後）	0.333	0.429	0.375

3.3. 感情分析精度の改善

感情分析の精度の改善を図るため，

・文を複数持つ文章の考慮

・極性値辞書および影響値辞書への新たな単語の追加．

を行った．

炎上ツイートへのリプライの文章には複数の文が含まれたものが多かった．現在の手法は文章内の文の数を考慮しておらず，文章中のすべての文を一つとして扱うこととなるため，式(2)より，各単語の極性値が小さくなってしまう．また二つの文とした際に文頭となる単語が，一つ目の文の単語より重みが大きくなってしまう場合があるため，期待する結果とは違ってしまう可能性が考えられる．そこで，文毎に極性値を算出し，最後に全ての文の極性値を積算したものが文章全体の極性値とする．ただし，文章においても，文章の末尾に近いほうがより重みが大きいと考え，式(2)を参考に文の極性値を補整した．

文の区切りの判断には文末で用いられることが多い，記号「。」「!」「?」「♪」や顔文字を用いた．

リプライ文章中の文の数を line，文の番号を k(k=1, 2, …, line)，k 番目の文の極性値を a_k，位置による補整を受けた文の極性値を a'_k とすると文章全体の極性値 A は次のようになる．

$$a'_k = a_k \times exp\left(-\left(\frac{i}{line} - \mu\right)^2\right)$$

$$A = \sum_{k=1}^{line} a_k$$

また，極性値辞書および影響値辞書へ新しい語を追加するため，本実験に利用したデータを利用してテキストを形態素解析し，リプライに存在する単語の基本形およびその出現回数を取得した．その中で極性値辞書・影響値辞書に登録されていない，品詞が「名詞」「動詞」「形容詞」「副詞」「助動詞」「接続詞」「接頭詞」「フィラー」であり，出現回数が 5 回以上の単語の中から人手でポジティブ語・ネガティブ語・影響語を選び，極性値と影響値を付与してそれぞれの辞書に追加した．追加した単語はポジティブ語が 10 語，ネガティブ語が 29 語，影響語が 2 語であった．

再び同様の手法で炎上検出を行ったところ，炎上と判断されたのは 9 件で，そのうちの 6 件が誤りであった．結果を表 7 に示す．

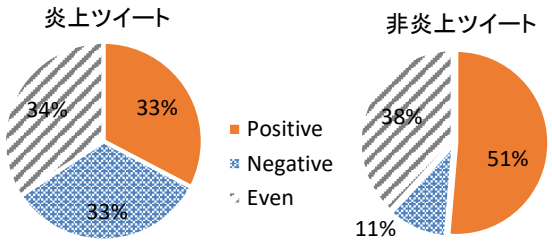

図 2．炎上/非炎上ツイートに対するリプライの感情

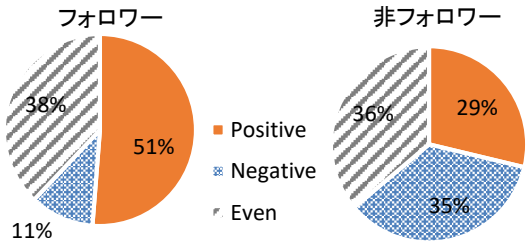

図 3．全ツイートにおけるフォロワー/非フォロワーからのリプライの感情

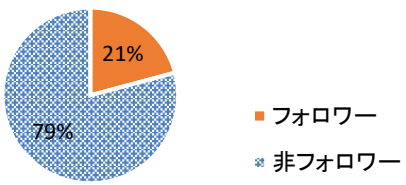

図 4．炎上ツイートに対するネガティブなリプライにおけるフォロワーの割合

3.4. 分析

3.4.1. グラフの利用

さらに検出精度の向上を図るため炎上ツイートへのリプライが持つ感情についての特徴の分析を行った．

図 2 は炎上/非炎上ツイートに対するリプライの感情の割合である．炎上ツイートの方がネガティブなリプライの割合が大きいことがわかる．

また，図 3 は全ツイートに対するフォロワー/非フォロワーからのリプライの感情の割合を示す．比べて分かるように非フォロワーからのリプライの方がネガティブなリプライの割合が多い．これよりフォローという行為がユーザに対して好意的な印象を抱いていることを示していると考えられる．一方で，図 4 から炎上ツイートにおけるネガティブなリプライの約 8 割が非フォロワーからのものであることが分かる．つまり，炎上ツイートにおいて批判や誹謗・中傷などのリプライを送っているのは非フォロワーが中心であると言える．これは，炎上したツイートは SNS やまとめサイトを通じて拡散されるが，それを目にして炎上を知ったユーザが便乗してリプライを送っていることが考えられる．

したがって，リプライを送ったユーザがフォロワーか非フォロワーか，という属性が炎上検出において重要であると考えた．

3.4.2. 決定木の利用

炎上検出に有効な条件を見つけるため， Twitter から得られる属性，さらにそれらを応用した属性を元に分析を行う．

表 8 の属性を用いてデータマイニングツールの Weka とそれに内蔵されている決定木の学習アルゴリズムである J48 を利用して炎上検出モデルを生成した．生成された決定木を図 5 に示す．

また，10 分割交差検定法により適合率と再現率，F 値を求めたところ表 9 のようになった．

非フォロワーネガティブリプライ率というネガティブリプライを送った非フォロワー数の割合が 0.111 より大きいか小さいかで判断できることとなった．10 分割交差検定法により精度を求めた結果，初期条件と比べて高い精度となった．この属性にはリプライの感情と非フォロワーの 2 つの属性が含まれており，炎上検出にはこれらが有効であると考えられる．

4. 実験 2

決定木で得られた条件を用いて再び新たなデータを用いて炎上検出を行った．4 日間，リプライが 20 件以上あるツイートとそのリプライを自動で取得したところ，合計 307 件のツイートを取得し、そのうち 2 件

表 8．決定木分類に用いる属性

属性	説明
フォロワー数	ツイートユーザのフォロワー数． 有名な人ほどその数は増える．
ポジティブリプライ率	ポジティブなリプライ数を全リプライ数で割ったもの．リプライにおけるポジティブなリプライの割合．
ネガティブリプライ率	ネガティブなリプライ数を全リプライ数で割ったもの．リプライにおけるネガティブなリプライの割合を表す．
味方率	リプライユーザにおけるフォロワーの割合．
非フォロワーポジティブリプライ率	リプライにおけるポジティブなリプライをした非フォロワーの割合．
非フォロワーネガティブリプライ率	リプライにおけるネガティブなリプライをした非フォロワーの割合．
正規化お気に入り数	ツイートのお気に入り数をそのユーザの平均お気に入り数で割ったもの．普段と比較してどの程度共感を得ているかの割合を表す．
正規化リツイート数	ツイートのリツイート数をそのユーザの平均リツイート数で割ったもの．普段と比較してどの程度注目されているかの割合を表す．
炎上/非炎上	ツイートが炎上しているかどうかを示す．決定木はこの属性を目的関数として分類を行う．

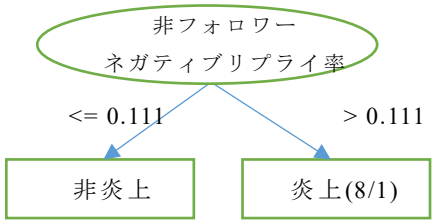

図 5．生成された決定木

表 9．決定木の分類精度

	適合率	再現率	F 値
炎上	0.778	1.00	0.875
非炎上	1.00	0.978	0.989

が炎上として検出された．一つは Web サイトのメンテナンスの謝罪ツイートであり，もう一つは有名人の謝罪ツイートであった．また，この二つ目のツイートは実際にネットニュースで炎上とされており，この条件を用いることで炎上ツイートが検出できることが確認できた．

5. おわりに

本研究では，既存手法の問題を解決した文章の感情分析手法を提案し，炎上の定義に従ったリプライの感情分析による炎上の検出を行った．検出精度はあまり良くなかったが，ツイートおよびリプライを分析することで得られた属性の内，「感情」と「非フォロワー」の二つに関連した属性を条件とした属性を条件に用いることで炎上ツイートを自動で取得することが出来た．したがって，炎上ツイートの検出にはこれらの二つの属性を用いることが重要である．

さらに精度を良くするには，炎上ツイートに対するリプライに現れる「煽り」や「皮肉」への対応が必要である．

文　献

[1] ICT 総研，“2016 年度 SNS 利用動向に関する調査 2016/08/16，http://ictr.co.jp/report/20160816.html, 2017．1．6 訪問

[2] 田代光輝，“ブログ炎上:学びとコンピュータハンドブック」，”東京電機大学出版局，pp. 68-72，2008

[3] 岩崎祐貴，“CGM における炎上の分析とその応用，”人工知能学会論文誌 30(1)，pp．152-160，2015

[4] 大西真輝，澤井裕一郎，駒井雅之，“ツイート炎上抑制のための包括的システムの構築，”人工知能学会全国大会論文集 29，pp．1-4，2015

[5] 三和未佐希，立間淳司，青野雅樹，“単語位置と強弱表現に着目したツイートの感情分析，”情報科学技術フォーラム講演論文集 13(2)，227-228，2014

[6] 小林のぞみ，乾健太郎，松本裕治，立石健二，福島俊一，“意見抽出のための評価表現の収集，”自然言語処理，Vol．12，No．3，pp．203-222，2005

[7] 東山昌彦，乾健太郎，松本裕治，“述語の選択選好性に着目した名詞評価極性の獲得，”言語処理学会第 14 回年次大会論文集，pp．584-587，2008．

[8] 田和真紀子，“程度副詞の評価性をめぐって，”宇都宮大学教育学部紀要，第 1 部 61，pp．25-36，2011