5P-06

ツイートとリプライからなる木構造を用いた炎上度判定

内山史也 † 荒瀬由紀 ‡ 鳴海紘也 * 河原林健一 ¶
横浜市立横浜サイエンスフロンティア高等学校 † 大阪大学 ‡
東京大学 * 国立情報学研究所 ¶

1 はじめに

本論文では，あるツイートとそれに対するリプライを入力として，そのツイートの炎上の度合いを判定するモデルを提案する．ここでいう炎上とは，ツイートした本人が当初意図していなかったようなネガティブなリプライを多数誘発する現象であり，当人の心証を下げる可能性がある．

ツイートのテキストに着目した既存研究として，単語分散表現と SVM を用いて有害表現を判定する研究が行われている [1]．また，日本語評価極性辞書を用いて感情値を定量化し，時系列情報で炎上パターンをクラスタリングした研究も存在する [2]．一方、ツイートとそのリプライからなる木構造（リプライツリー）の分析により，リプライの付き方にはいくつかのパターンがあることがわかっている [3]．また，ユーザー間での言及情報をグラフとしてとらえ，リンク数の増加と連結成分の偏りなどを比較した研究も存在する [4]．

本研究では，これら，ツイートにおけるテキストの感情推定とグラフ理論によるアプローチを用いることで，炎上ツイートを判定することを目的とする．具体的には，炎上ツイートとそのリプライの感情値に逆転現象があると仮定し，リプライツリーの感情分析を行うことで炎上度合いを判定する指標を提案する。さらに木構造におけるノードの次数等の構造的パターンを利用して炎上度判定を行う指標を設計する．Twitter より収集した炎上ツイートと炎上が生じていない (非炎上) ツイートを用いた評価実験により、その予測精度を調査した．

2 実験

まず，Twitter より炎上・非炎上ツイートのリプライツリーを収集し，その特徴を観察した．

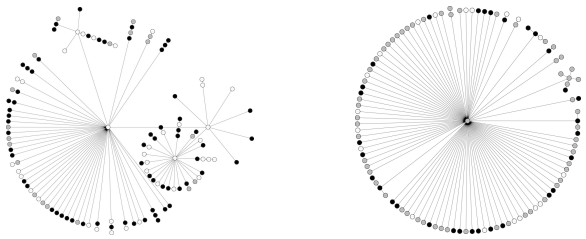
2.1 ツイートの収集

2020 年 12 月 21 日から 2021 年 1 月 1 日にかけて投稿された 7 件の炎上ツイートと 11 件の非炎上ツイートを収集した．炎上ツイート及び非炎上ツイートは特定のまとめサイト*1から収集した。現状ではリプライが元ツイートの意図に反しているかどうかを自動的に判定することは難しいため，手動で元ツイートを取得した．炎上ツイートとは，リプライ数が 50 以上のツイートで，上記のサイトに挙げられたものから選択されたツイートのことを示す．また，非炎上ツイートとは，リプライ数が 50 以上のツイートで，著者らによって炎上と確認される現象が発見されなかったツイートのことを示す．その後，各ツイートに付いたリプライを収集し，各ツイートをノード，言及関係をエッジとしたリプライツリーを構築した．ただし，Twitter API の制限によって全てのツイート間のエッジを獲得できないため，途中で元ツイートとの経路が切れている部分木については本研究の調査の対象外とした．

2.2 ツイートの感情推定

感情推定には Microsoft Azure の Text Analysis [5] を使用し, ツイート x の本文からポジティブ率 P_{Pos}, ニュートラル率 P_{Neu}, ネガティブ率 P_{Neg} を算出した $(0 \leq P \leq 1)$. 図1は，炎上と非炎上のリプライツリーの一例である．中心には炎上判定の対象となる元ツイート，その周辺には各リプライが丸で示してある．白ノードをポジティブ度の高いツイート，灰ノードをニュートラル度の高いツイート，黒ノードをネガティブ度の高いツイートとした．ただし，各色分けは $P_{\mathrm{Pos}}, P_{\mathrm{Neu}}, P_{\mathrm{Neg}}$ のうち最大値を取る種類の感情に割り振った．

この図から，炎上ツイートのリプライツリーは，(1) エッジで結ばれるノード同士で感情の変化が多く見られる，(2) 元ツイートに対するリプライの次数が大きくなる，(3) グラフの最大長が長くなるということが分かる．



(1) 炎上ツイートのリプライツリー例　(2) 非炎上ツイートのリプライツリー例

図1: 収集したリプライツリーの例

2.3 時間変化

本節では炎上したリプライツリーの特徴が時間的にどのように変化しているかを調べた．具体的には，収集できたリプライツリーに関して

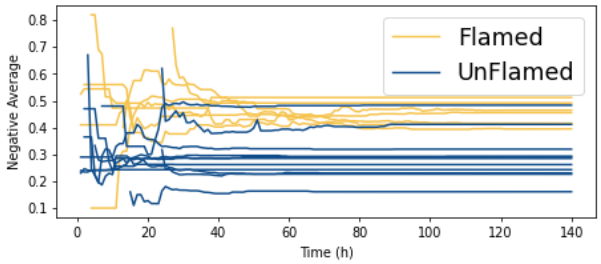
1. リプライのネガティブ率 P_{Neg} の平均
2. ツリーの最大高さ
3. ツリーの最大次数
4. ツリーの最大次数（根ノード除く）
5. 感情の逆転度（根ノードの P_{Neg} とリプライツイートの平均 P_{Neg} の差の絶対値）
6. 感情の変化回数（ツリーの各経路における感情の変化の回数）
7. 発展指数（葉ノードの数を，根ノードに直接繋がっているノード数で割った数値）

の時間ごとにおける変化を折れ線グラフで図2に表した．

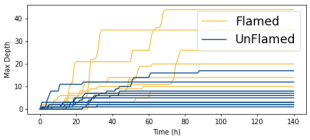
Flaming Tweet Detection based on Sentiments and Structures of Tweet-Reply Tree
Fumiya UCHIYAMA†, Yuki ARASE‡, Koya NARUMI* and Ken-ichi KAWARABAYASHI¶
† Yokohama Science Frontier High School
‡ Osaka University
* The University of Tokyo
¶ National Institute of Informatics

*1 http://www.twtimez.net/, https://tsuisoku.com/

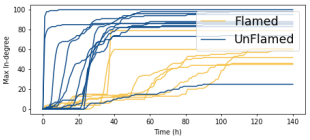
Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.



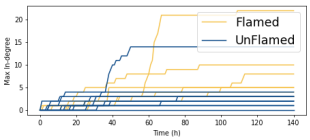
(1) リプライのネガティブ平均



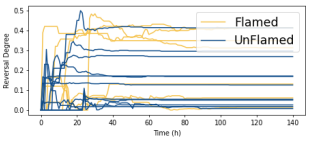
(2) ツリーの最大長



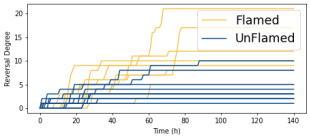
(3) 最大次数



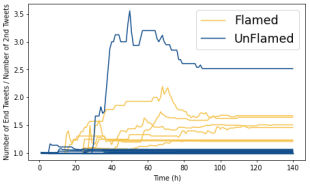
(4) 最大次数（根除く）



(5) 感情の逆転度



(6) 感情の変化回数



(7) 発展指数

図2: 時間による各指標の変化

図2の結果では，感情の逆転度と感情の変化回数については炎上と非炎上であまり有意な差が見られなかったのに対して，根ノードを含む最大次数では，ほとんどの非炎上ツイートの方がより早く次数が大きくなっていることが確認できた．また，ほとんどの非炎上ツイートの発展指数は 1.2 以下に収まっていることがグラフから確認できた．よって，多くの非炎上ツイートでは早くリプライが付き，炎上ツイートではリプライに対して時間的に遅れてさらに多くリプライがつく傾向があると判明した．実際の元ツイートやリプライの内容も含めて判断すると，これらは，元ツイートに対して反論するツイートに多くの共感が集まり，またそれに対して反論が付くなどといった炎上ツイートの持つ議論の発展性を示しているものと考えられる．

2.4 指標を用いた基礎的な分類手法

2.3節の観察に基づき，提案指標を用いた基礎的な分類モデルを作成した．根ノード以外のツイートの集合 $\{x|x \in X\}$ を入力として，経過時間を $t\,(0 \leq t)$ [h]，最大次数を $m\,(0 \leq m \leq |X|)$，発展指数を $d\,(1 \leq d \leq |X|-1)$ と置いて炎上度 $FI(X)$ を次式で定義した．

$$FI(X) = \begin{cases} 0 & (X = \emptyset) \\ \frac{dt}{1.2m} & (\text{otherwise}) \end{cases}$$

2.3節より，炎上ツイートでは時間的に遅れて最大次数 m が増加する傾向があることが分かったため，時間 t が経過してから m が大きくなると炎上の可能性が高くなるようにした．また，発展指数 d が 1.2 を超えると炎上ツイートの割合が高くなるため，1.2 を基準とした発展指数比を取り入れた．

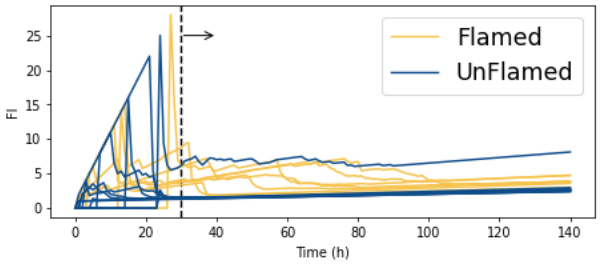


図3: 時間による FI 値の変化

図3が時間ごとにおける FI 値の推移である．図3では，ツリーの最大次数や発展次数が増えているときに FI 値が増加し，最大次数や発展次数の増加が収まると FI 値も減少していることがわかる．$t \geq 30$ の時，FI 値が 1.67 を超えるとその後炎上するツイートが多くなるため，ツイートをポストしてから 30 時間後に FI 値が 1.67 を超えているかどうかで炎上／非炎上を概ね分類することが可能であると期待される．

3 おわりに

本研究では，感情推定とグラフ的指標を用いた炎上ツイートの特徴の分析を行った．収集したツイートデータにおいて，炎上ツリーと非炎上ツリーでは最大次数や発展指数の推移の特徴に差があることが分かった．また，グラフ的要素の分析では，リプライツリーの最大次数や発展指数を用いることで，炎上する可能性が低いツイートとそうでないツイートを分類できる可能性を示した．

今後の課題として，今回調べた指標では分類困難な非炎上ツイートが存在したので，分析するデータ数を増やしたり，他の感情的指標と組み合わせることで分類性能を向上する予定である．また，リプライだけでなくリツイートに注目して炎上の特徴を調べることも重要だと考えられる．

謝辞 本研究は，国立研究開発法人科学技術振興機構グローバルサイエンスキャンパス（GSC）「情報科学の達人」育成官民協働プログラム（国立情報学研究所，情報処理学会，情報オリンピック日本委員会）の支援のもと実施したものである．

参考文献

[1] 三宅剛史，松本和幸，吉田稔，北研二: 分散表現を用いた有害表現判別に基づく炎上予測, 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会 (第 15 回) , 2017.

[2] 渡辺みずほ, 佐藤哲司: リプライのポジネガ極性を用いた Twitter 炎上の分類手法の提案, DEIM, 2020.

[3] Nishi, R., Takaguchi, T., Oka, K. et al. Reply trees in Twitter: data analysis and branching process models. Soc. Netw. Anal. Min., 2016.

[4] 小出明弘，斉藤和巳，大久保誠也，鳥海不二夫：Twitter の @-message で構成される成長ネットワークの分析, 第 74 回全国大会講演論文集,2012.

[5] Text Analytics API Documentation `https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/`

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.