DEIM Forum 2023 4a-9-4

# オンライン議論の過熱と感情的投稿に関するダイナミクス

小林　将大†　　矢田竣太郎†　　若宮　翔子†　　荒牧　英治†

† 奈良先端科学技術大学院大学　〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: †{kobayashi.masahiro.kl0,s-yada,wakamiya,aramaki}@is.naist.jp

**あらまし**　オンライン議論における悪質な投稿は円滑な議論の進行や参加者の心理に悪影響を及ぼすことが知られており，これを防止することを目的とした研究が進められている．本稿では以上のような現状を踏まえ，オンライン議論におけるネガティブな感情の投稿が議論スレッドの過熱に及ぼす影響について分析を行う．まずオンライン議論の過熱の定義を行う．次にネガティブな感情とスレッドの過熱に関する 3 つの仮説をたてた．そして，Wikipedia から収集した議論スレッドデータの投稿に感情をラベリングして構築したデータセットを用いて仮説の検証を行った．検証の結果，2 つの仮説を支持する形でポジティブな投稿とネガティブな投稿がその後の議論の過熱に影響を及ぼすことが確認された．

**キーワード**　行動モデリング，情動伝染，Wikipedia，オンライン議論，SNS，感情分析，時系列分析

## 1　はじめに

オンラインにおける人々の交流は，SNS の社会への浸透とともにますます盛んになっている．それとともに，ヘイトスピーチの投稿やフェイクニュースの拡散をはじめとするオンラインでの交流から生じる社会的問題への対処が大きな課題となっている．その 1 つに挙げられるのが，オンラインでの不特定多数の人物が参加する議論 (以下このような議論を「オンライン議論」と呼ぶ) における，煽りや暴言をはじめとした参加者の精神的健康状態を悪化させうる [1], [2]「不健全な会話」[3] である．本稿ではオンライン議論における過熱現象をそのような「不健全な会話」の一種と捉える．この現象の発生を予防する技術を確立するために，過熱が発生するメカニズムを言語的特徴や投稿の感情から探ることが必要である．そのために以下に示す手法で過熱現象の分析を行う．分析の材料には Wikipedia 日本語版の「Wikipedia：井戸端」[1] と呼ばれるコミュニティページ (以下井戸端と称する) の投稿を用いる．

まず過熱現象を「単位時間当たりの投稿数 (＝投稿速度) の増加」「ネガティブな投稿の割合の増加」の 2 つの要素から定義する．これら 2 つの値が高いほど過熱が強まっているものとしネガティブな投稿と議論の過熱現象の間に成り立つ関係について考える．これまでの研究で,「ネガティブな流れのスレッドのほうがネガティブな気分になりやすい」 [4],「ネガティブな投稿はより視覚的な注意を引きやすい」[5]，さらに「ネガティブな言葉を含むほうが投稿がリツイートされる可能性が高くなる [6] といったネガティブな投稿による影響が報告されている．これらの研究からネガティブな投稿に関して次のような仮説を提唱する．

**仮説 1**　（ネガティブな投稿は）さらにネガティブな投稿を引き起こす

1：https://ja.wikipedia.org/wiki/Wikipedia: 井戸端

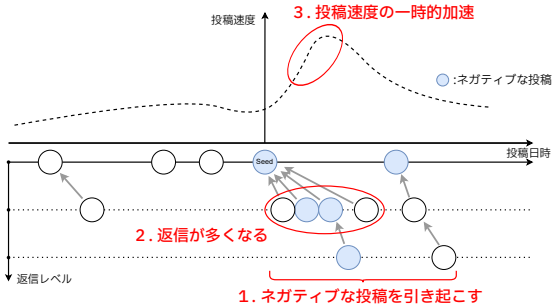


図 1: ネガティブな投稿に関する 3 つの仮説

**仮説 2**　（ネガティブな投稿は）他の感情の投稿に比べて返信が多くなる

**仮説 3**　（ネガティブな投稿は）スレッドの投稿速度を加速させる

本稿ではこれら 3 つの仮説を検証し，副次的にそれ以外の感情を持った投稿の効果についても検証を行う．3 節では感情ラベルを付与した投稿のデータセットを構築する手法について説明する．4 節では用語を定義し，またデータセットを用いた検証実験の方法について説明する．5 節では実験から得られた結果を述べる．

## 2　関連研究

オンライン上のネガティブな言動が閲覧者の心理状態や行動に与える影響は，すでに心理学や情報科学などの分野から着目されている．心理学的な観点からオンライン議論の感情的投稿に対する反応の分析を行った研究には，Syrjämäki らによる脳波測定やアンケートを用いたもの [4] がある．また Kohout ら [5] は感情的投稿に対する閲覧者の反応の分析をアイトラッキングを用いて行っている．本稿と同じく Web 上のリソースを用いた研究には Jiménez-Zafra らによる感情的ツイートの拡

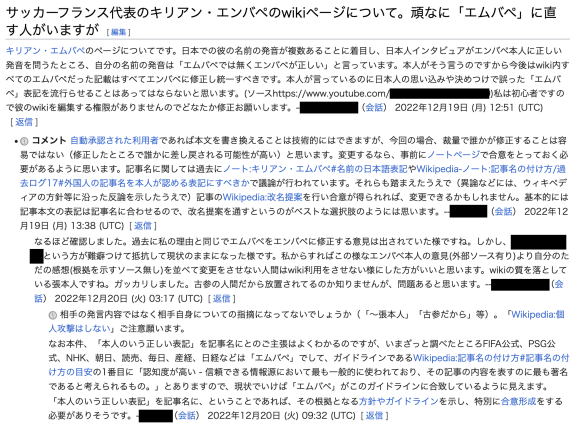


図 2: 井戸端のスレッドの一例．投稿に対して返信を行うとインデントが発生する．ユーザー名などテキストの一部は隠している．

表 1: 感情ラベルの割合

| | POS | m_POS | NEUT | m_NEG | NEG | Failed |
|---|---|---|---|---|---|---|
| 投稿数 | 5702 | 1029 | 3214 | 1496 | 10496 | 30725 |
| 割合 (%) | 10.83 | 1.95 | 3.10 | 2.84 | 19.93 | 58.07 |

散研究 [6] がある．本稿で提唱する仮説はこれらの研究を土台としている．

オンライン議論における過熱についての過去の研究には，筆者らによる過熱と言語の関連性の研究 [7] がある．この研究では過熱現象を時系列に関する複数のパラメータを用いて「過熱区間の存在」という形で定義している．対して本稿では過熱現象の「度合い」に着目するため，その定義に投稿速度とネガティブな投稿の数を用い，何らかのパラメータの値で過熱の発生の境界を線引きすることは行わない．

最新のオンライン議論に関する時系列的研究として挙げられるものには，Horawalavithana らによる初期投稿からその後の投稿の内容に加えて投稿者，投稿日時，返信対象の 3 種の情報を予測生成するオンラインディスカッションスレッドの生成の研究 [8] がある．また議論に限らないオンラインでの不健全な会話に関する時系列的研究には，Sahnan らによる Twitter の会話ツリーのヘイトスコアの時系列予測モデルの研究 [9] がある．Sahnan らの研究は本稿と同じく，ヘイトスピーチを含む投稿の「予防」を焦点にタスクを設定している．

## 3 データセット

本研究では，図 2 に示すような井戸端の投稿を分析に用いる．投稿のテキストに対して図 3 に示すように所属スレッド，投稿の順番，インデントレベル，投稿日時および感情ラベルの情報を与えることでデータセットを構築する．

投稿テキストの取得に際して，MediaWiki API [2] から井戸端のスレッドの HTML データを取得する．井戸端のデータはコ

2：https://www.mediawiki.org/wiki/API:Main_page

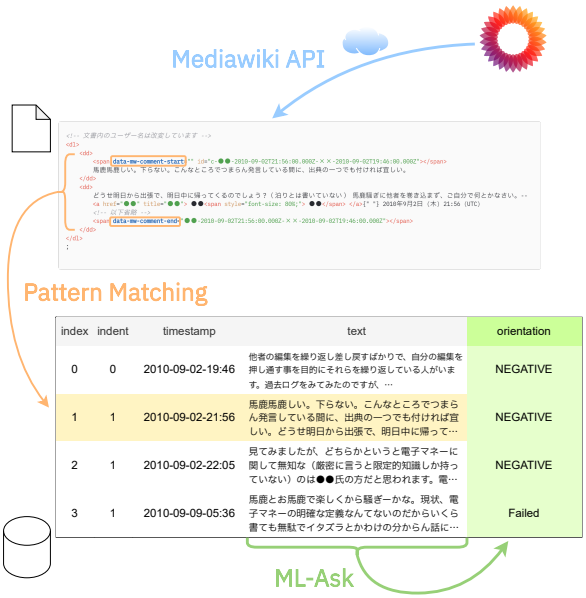


図 3: データセットの構築フロー

メント単位でなくページ全体のマークアップ文書 [3]であるため，データセットを構築するためには投稿を何らかの方法でページから抽出する必要がある．先行研究 [7] では Wikipedia マークアップパーサを用いて投稿を分割したが，本稿では HTML 文書に対するパターンマッチングを用いて投稿の抽出を行う．投稿の順番とインデントレベルはテキスト抽出の際に HTML 内の情報から同時に取得する．投稿日時は署名テキストからパターンマッチングを用いて取得する．

投稿テキストの感情分析には ML-Ask [10] の Python 実装として提供されている pymlask[4]を用いる．pymlask が使用する形態素解析器の辞書には neologd[5]を用いる．ML-Ask は Russell の円環モデル [11] をもとにしている．円環モデルは valence と呼ばれる感情の快不快を表す軸と arousal と呼ばれる感情の覚醒度を表す軸の 2 次元から成り立つモデルである．pymlask が感情解析結果として出力するデータにはこのモデルの valence に対応する orientation と arousal に対応する activation の 2 種類のラベルが含まれる．本稿で用いるデータはこのうちの orientation ラベルであり，本稿で述べる投稿の「感情」とは orientation ラベルのことを指す．pymlask が出力する orientation ラベルは POSITIVE，mostly_POSITIVE，NEUTRAL，mostly_NEGATIVE，NEGATIVE の 5 種類であり，文字数が多すぎるなど何らかの理由で pymlask が orientation ラベルを付与することができなかった投稿に対しては一律に Failed ラベルを与える．

データセットに含まれる井戸端のスレッドの数は 2005 年 4 月から 2022 年 10 月までに掲載されたものから抽出された 4832 件であり，抽出に成功した延べ投稿数は 52662 件である．このうちタイムスタンプの取得に失敗した 60 件の投稿は分析に投

3：Wikipedia マークアップと HTML の 2 種類が入手可能．
4：https://github.com/ikegami-yukino/pymlask
5：https://github.com/neologd/mecab-ipadic-neologd

稿日時データを用いる実験から除外している．各ラベルの付与割合は表 1 に示す.

## 4 検 証 実 験

### 4.1 設　　定

以下に本稿における用語の諸定義を行う.

**ネガティブ (ポジティブ) な投稿**　本稿における「ネガティブな投稿」とは ML-Ask が “NEGATIVE”とラベル付けした投稿とする．ポジティブな投稿についても同様の定義とする.

**Seed 投稿**　検証実験ではすべての投稿についてその投稿の時点を基準とし，その前後の投稿情報を求める．この検証の基準となる投稿のことを Seed 投稿と呼ぶ.

**Seed 感情**　Seed 投稿のテキストに対して ML-Ask が与えた orientation ラベルを Seed 感情と呼ぶ.

**投稿速度**　「ある時間区間における投稿速度 (件/日)」とは区間内の投稿件数を $x$(件)，区間幅を $d$(時間) としたときに $24x/d$ で表される量のことをいう．実験によって時間区間を閉区間とする場合と半開区間とする場合があるが，ここでは区間幅を一律に「大きい端点の値から小さい端点の値を引いた値」とする.

**周辺の投稿速度**　「窓幅を $w$(時間) としたときの時点 $t$ の周辺の投稿速度」とは，閉区間 $[t-0.5w, t+0.5w]$ における投稿速度のことをいう.

以上の定義に従い，仮説の検証実験を手法を述べるとともに必要なハイパーパラメータの設定値を示す.

### 4.2 投稿の感情に関する検証

「ネガティブな投稿はさらにネガティブな投稿を引き起こす」という**仮説 1** を検証する．これは定義した用語を用いて「Seed 感情が NEGATIVE である場合，その後の投稿の感情も NEGATIVE になる」と表現できる．検証実験の手順は以下のようにする．まず「その後」の基準となるハイパーパラメータの窓幅 $W$（時間）を定め，Seed 投稿の時点から $W$ 時間後までの投稿を感情ごとにカウントする．カウントから得られたその後の投稿数の分布を Seed 感情およびその後の感情ごとに求める．ただし Seed 感情が Failed のものは除外する．次に「2 つの Seed 感情間でその後のある感情を持った投稿の数に差が出るか」を統計的仮説検定により検証する．Seed 投稿後の投稿数の分布は明らかに正規分布でない [6]ため，検証にはマン・ホイットニーの U 検定 [12] を用いる．検証対象となる「その後の投稿の感情」は POSITIVE，NEUTRAL，NEGATIVE の 3 種類とし，検定で比較する 2 群 $G_1, G_2$ は次のように定める：

Failed を除く 5 つの Seed 感情から 1 つを選び，その後の投稿数の分布を $G_1$ とする．それ以外の 4 つの Seed 感情の後の投稿数の分布を $G_2$ とする．検定は有意水準 5%の片側検定とし，$G_1 > G_2$ とする仮説が採択可能である場合には投稿数が有意に多いものとして,「$G_1$ の Seed 感情にその後の特定の感情的投稿が他の Seed 感情と比べて多くなる効果がある」とする．逆に $G_1 < G_2$ とする仮説が採択可能である場合には「$G_1$ の Seed 感情にその後の特定の感情的投稿が他の Seed 感情と比べて少なくなる効果がある」とする．この U 検定を窓幅それぞれ $W = 6, 24, 48, 168, 1344$（すなわち 6 時間，24 時間，48 時間，7 日間，8 週間）の 5 つに設定して行い，特定の感情を持つ投稿の増減効果を持つ Seed 感情を確認する.

6：非正規性の確認は QQ プロットを用いて行った.

### 4.3 返信数に関する検証

「ネガティブな投稿への返信は多くなる」という**仮説 2** を検証する．これは定義した用語を用いて「Seed 感情が NEGATIVE である場合，Seed 投稿への返信が多くなる」と表現できる．検証実験では Seed 投稿への返信数の分布を Seed 感情ごとに算出する．返信は Wikipedia のトークに実装された返信機能を用いて返信された投稿のみを対象とする．返信数の算出には図 2 からわかるように「返信を行うとインデントが Seed 投稿よりも 1 段階右になる」ことを利用している．返信数の分布の有意差を測るため，U 検定を行う．検定は有意水準 5%の片側検定とし，比較する 2 群 $G_1, G_2$ の選択は 4.2 項と同様にする．そして $G_1 > G_2$ とする仮説が採択可能である場合は「返信数が多くなる効果がある」とする．逆に $G_1 < G_2$ とする仮説が採択可能である場合は「返信数が少なくなる効果がある」とする.

### 4.4 投稿速度に関する検証

「ネガティブな投稿はスレッドの投稿速度を一時的に加速させる」という**仮説 3** を検証する．これは定義した用語を用いて「NEGATIVE な Seed 投稿を行うとスレッドの投稿速度が一時的に上昇する」と表現できる．検証には 2 種類の実験を行う.

#### 4.4.1 周辺の投稿速度に関する実験

この実験では Seed 投稿の時点に加えてその前後 $2k$ 個の時点でのそれぞれの周辺の投稿速度を求め，これらの分布を Seed 感情間ごとに算出する．分布の平均値および 95%信頼区間をプロットすることにより，Seed 投稿の投稿前後における投稿速度の分布の推移を感情間で比較する．投稿速度 $[v_{-k}, \cdots, v_k]$ の分布は以下に述べる手法で求める．最初に周辺の時点数 $k$，ステップ $T$（時間），窓幅 $W$（時間）をハイパーパラメータとする．$t_0$ の前後それぞれ $k$ 個の時点は

- 前：$(t_0 - kT), (t_0 - (k-1)T), \cdots, (t_0 - T)$
- 後：$(t_0 + T), \cdots, (t_0 + (k-1)T), (t_0 + kT)$

とする．窓幅を $W$ として各時点の周辺の投稿速度を求め，それらを

- 前：$v_{-k}, v_{-k+1}, \cdots, v_{-1}$
- 後：$v_1, \cdots, v_{k-1}, v_k$

とする.

ハイパーパラメータの値は $kT$ をそれぞれ $kT = 6, 48, 168$ に設定して検証を行う．すなわち Seed 投稿の時点に加えて，その

(a) ±6 時間
(b) ±48 時間
(c) ±7 日間

の区間内 $k$ 個の時点の周辺の投稿速度の分布を求める．そのときのハイパーパラメータ $[k, T, W]$ の具体的な値は

(a) $[k, T, W] = [12, 0.5, 3]$
(b) $[k, T, W] = [16, 3, 24]$
(c) $[k, T, W] = [14, 12, 24]$

と定める．[7]

**4.4.2** 投稿速度の差分に関する実験

この実験では Seed 投稿の前後でそれぞれ投稿速度を求め、これらの差 $\Delta v$ のその分布を Seed 感情ごとに算出する．算出した分布の平均値および 95%信頼区間をプロットすることにより Seed 投稿前後での投稿速度の変化 Seed 感情ごとに比較する．$\Delta v$ は以下に述べる手法で求める．まず Seed 投稿の投稿時点を $t_0$ とし，前後の窓幅を $W$(時間) と定める，$W$ はハイパーパラメータである．次に「Seed 投稿の前の投稿速度」を半開区間 $[t_0 - W, t_0)$ における投稿速度 $v_{\text{before}}$ とする．同様に「Seed 投稿の後の投稿速度」を $(t_0, t_0 + W]$ における投稿速度 $v_{\text{after}}$ とする．2 つの投稿速度の差を $\Delta v = v_{\text{after}} - v_{\text{before}}$ として求める．この手法でハイパーパラメータ $W$ を

(a) $W = 0.5, 1.0, \cdots, 5.5, 6.0$
(b) $W = 3, 6, \cdots, 45, 48$
(c) $W = 12, 24, \cdots, 156, 168$
(d) $W = 168, 252, \cdots, 1260, 1344$

の等差 4 パターンで動かし，パターンごとの $\Delta v$ の分布を得る．これはすなわち上から順に

(a) 前後 6 時間まで 30 分ごと
(b) 前後 48 時間まで 3 時間ごと
(c) 前後 7 日間まで 12 時間ごと
(d) 前後 8 週間まで 0.5 週間ごと

に $\Delta v$ の分布を得ることを意味する．

## 5 結果と考察

### 5.1 投稿の感情に関する検証

図 4 は 4.2 項で述べた U 検定の結果を示す．縦 5 行の (a)-(e) はそれぞれ窓幅 $W$ を表し，横 3 列の POSITIVE, NEUTRAL, NEGATIVE は検定で投稿数を比較した Seed 投稿後の感情を表す．また図内の各ヒートマップは U 検定で得られた p 値と検定結果の有意性を示す．ヒートマップ内の 2 行「多い」と「少ない」はそれぞれ 2 種類の検定 ($G_1 > G_2$：特定の投稿が多くなる効果の検定と $G_1 < G_2$：特定の投稿が少なくなる効果の検定) を示し，5 列は検定を行った Seed 感情を表す．各セルにはそこから得られた p 値を小数点下 3 桁で丸めたものを掲載している．赤いセルは有意水準 5%で効果が見られることを意味する．

まず 1 列目の「POSITIVE な投稿」についての検定結果を見る．POSTIVE な投稿の増加効果がみられる Seed 感情は POSITIVE 以外の 4 種類である．そのうち mostly_POSITIVE は (b) から (d) までと，全般的に POSTIVE な投稿が比較的多くなる効果が見られた．NEUTRAL は (c) から (e) と長い窓幅での効果が見られた．一方で POSITIVE な投稿が少なくなる効果は NEGATIVE な Seed 投稿に見られ，これはすべての窓幅で確認された．

次に 2 列目の「NEUTRAL な投稿」についての検定結果を見る．NEUTRAL な投稿数が多くなる効果は Seed 感情が mostly_POSITIVE かつ窓幅が (a) の場合を除き，mostly_POSITIVE, NEUTRAL, mostly_NEGATIVE のすべての窓幅で確認された．また少なくなる効果は Seed 感情が POSITVE の場合に確認され，うち前者の効果はすべての窓幅で見られた．

最後に 3 列目の「NEGATIVE な投稿」についての検定結果を見る．NEGATIVE な投稿が多くなる効果は NEUTRAL, mostly_NEGATIVE, NEGATIVE の 3 つの Seed 感情で見られ，そのうち NEGATIVE な Seed 投稿による効果はすべての窓幅で確認された．mostly_NEGATIVE は (a) を除く窓幅で確認された．NEUTRAL による効果は 1 列目と類似しており，比較的長い窓幅で効果が見られた．対して少なくなる効果は POSTIVE な Seed 感情のみで，これはすべての窓幅で確認された．このことから投稿の感情に関する検証の実験結果は**仮説 1** を支持するような内容であると考えられる．

表 2: Seed 投稿への返信数

(a) 全返信

| Seed 感情 | POS | m_POS | NEUT | m_NEG | NEG | Failed |
|---|---|---|---|---|---|---|
| データ数 | 5702 | 1029 | 3214 | 1496 | 10496 | 30725 |
| 平均 | 0.213 | 0.234 | 0.260 | 0.264 | 0.240 | 0.221 |
| 標準偏差 | 0.734 | 0.702 | 0.880 | 1.246 | 1.016 | 0.801 |

(b) 1 件以上

| Seed 感情 | POS | m_POS | NEUT | m_NEG | NEG | Failed |
|---|---|---|---|---|---|---|
| データ数 | 751 | 144 | 462 | 214 | 1443 | 4120 |
| 平均 | 1.614 | 1.674 | 1.812 | 1.850 | 1.750 | 1.645 |
| 標準偏差 | 1.355 | 1.057 | 1.606 | 2.819 | 2.206 | 1.562 |

### 5.2 返信数に関する検証

表 2 に Seed 投稿への返信数の分布を示す．表 2a と表 2b を比較するとわかるように，いずれの Seed 感情でも返信数が 0 件である割合が非常に大きい．そのため表 2b に返信数が 1 件以上あった Seed 投稿への返信数を別途示す．表 2a を見ると，mostly_NEGATIVE と NEUTRAL への返信数がやや多く，mostly_POSITIVE への返信がやや少なく，Failed と

7 : $T$ に比べて $W$ が大きすぎると離れた時点の影響を受けすぎてしまうため (a) の $T = 0.5$ の場合のみ $W = 3$ とする．


図 4: Seed 投稿後の感情的投稿数の有意差の U 検定の結果．縦 5 行 (a)-(e) と横 3 列はそれぞれ窓幅と Seed 投稿後の感情を表す．赤いセルは有意水準 5%で帰無仮説を棄却可能であることを意味する．

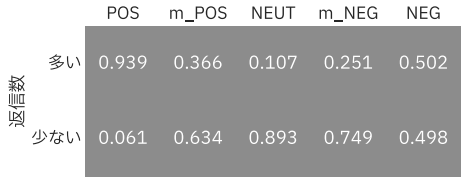


図 5: Seed 投稿への返信数の有意差の検定結果

POSITIVE への返信が少なくなる傾向が見られた．これは表 2b でも同様の結果となっている．また表 5 には各 Seed 感情についてのその返信数の差についての 5%の有意水準の U 検定の結果を示す．検定の結果として，$G_2$ に比べて有意に大きい (小さい) といえるような差がある $G_1$ が存在するという仮説を支持する p 値 (赤いセル) は，いずれの Seed 感情でも見られなかった．すなわち**仮説 2** を支持するような結果は返信数に関する検証から得ることができなかった．

### 5.3 投稿速度に関する検証

#### 5.3.1 周辺の投稿速度に関する実験

図 6 の縦 3 行 (a)-(c) はそれぞれハイパーパラメータを 4.4.1 項の (a)-(c) に設定したときの投稿速度の分布を示す．プロット内の各点はその時点における分布の平均値であり，線の周囲の薄い色で塗られた領域は分布の 95%信頼区間を意味する．1 列目には Seed 感情のラベルのうち POSITIVE，NEUTRAL，NEGATIVE の場合の分布を，2 列目には残りの mostly_POSITIVE，mostly_NEGATIVE，Failed の場合の分布をプロットしている．3 列目の灰色で図示された All は全投稿についての投稿速度の分布である．

まず (a) を見ると前後 0-2 時間ほどでは Seed 投稿が NEGATIVE と Failed の場合にやや高い数値になっており，それ以外の感情の場合に低い数値になっていることがわかる．低い数値を示した Seed 投稿の中では，POSITIVE が特に低い傾向になっている．mostly_POSITIVE や mostly_NEGATIVE は 95%信頼区間が大きいことから数値のばらつきが大きくなっている．前後 2-4 時間に着目すると，前 2-4 時間では POSITIVE な Seed 投稿の場合に他と比べてやや投稿速度が低くなる傾向が見られる．後 2-4 時間では NEGATIVE な投稿が他と比べて高い投稿速度となっている．Failed の場合の投稿速度はこの区間では All と近い傾向を見せている．前後 4-6 時間でもこの NEGATIVE の後の投稿速度が高く，POSITIVE の前後の投稿速度が低いという傾向は維持されていることがわかる．

次に (b) を見ると 1 列目の投稿速度はおおむね Seed 投稿が NEGATIVE な場合の速度が一番大きく，その次に NEUTRAL 場合の速度，さらにその下に POSTIVE な場合の速度となっており，この傾向は前 12 時間-後 24 時間ほどで見られる．それより外の区間では Seed 投稿が NEUTRAL の場合の投稿速度が NEGATIVE の場合の投稿速度を上回る結果となっている．対して 2 列目では “mostly_” 系の Seed 投稿の前後 6 時間の投稿速度の平均値のばらつきが 1 行目と比べて緩和されていることがわかる．これは窓幅をより大きな 24 時間としたことが原因と考えられる．Failed は前後ともに 15 時間より外の範囲で

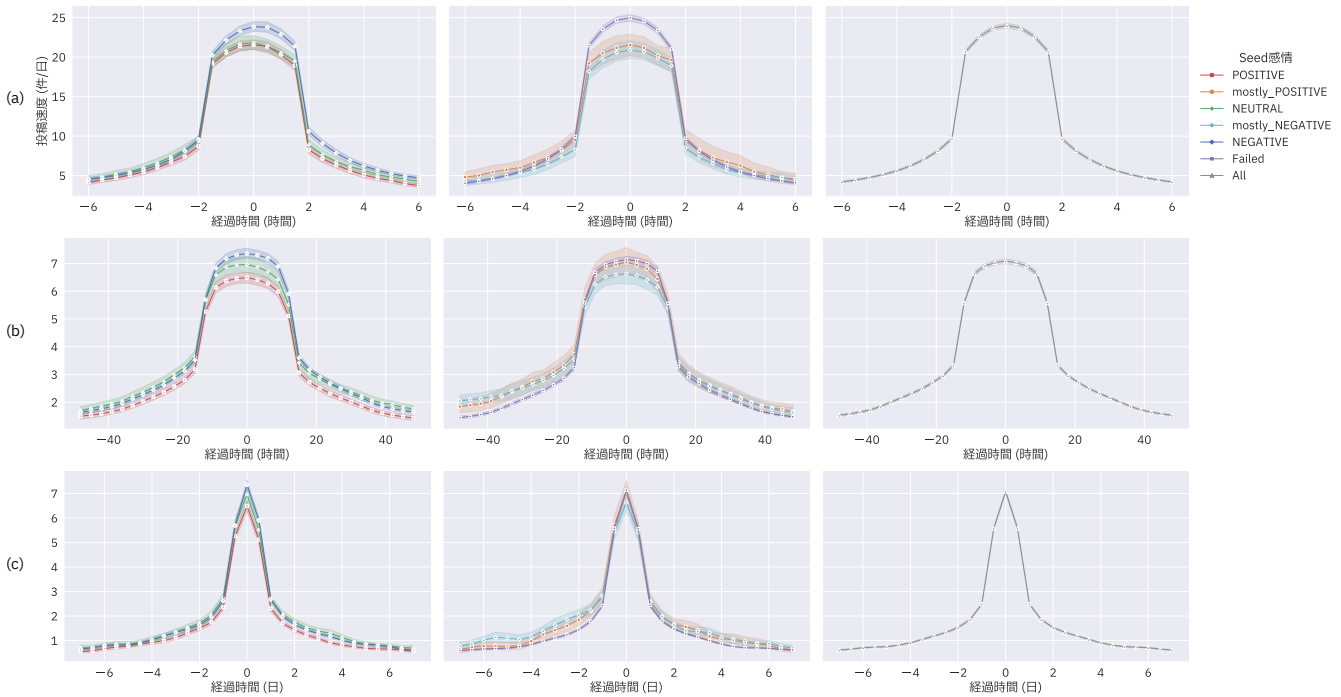


図 6: Seed 投稿の前後の速度．上から投稿の (a) 前後 6 時間 (b) 前後 48 時間 (c) 前後 7 日間までの分布を示す．横 3 列にそれぞれ Seed 感情を分けて掲載している．

低くなる傾向が見られる．後 20-40 時間あたりでは Seed 投稿が “mostly_” 系の場合に投稿速度が高い値となっている．ただし “mostly_” 系の傾向はデータの少なさによりばらつきが大きくなっていることを (a) と同様に念頭に入れる必要がある．

最後に (c) を見ると 1 列目では投稿時点から 5 日後あたりまで Seed 投稿が POSITIVE の場合の投稿速度が低い傾向が続いていることがわかる．この傾向は 6 日後あたりではほとんどなくなっている．2 列目では 2 日前から 6 日前あたりまで Seed 投稿が mostly_NEGATIVE の場合の投稿速度がやや高い傾向を示しているが，これも (a)(b) と同様にデータの少なさが影響している可能性がある．Failed は (b) で述べた速度が低くなる傾向が 1 週間前後まで起こっていることがわかる．以上を踏まえると，(a)(b) の結果より 24 時間後あたりまで Seed 投稿が NEGATIVE な場合に比較的投稿速度が高い水準になる傾向が見られたことから，**仮説 3** を支持することができると考えられる．

**5. 3. 2**　投稿速度の差分に関する実験

図 7 の縦 4 行 (a)-(d) にそれぞれ 4. 4. 2 項に挙げた $W$ の (a)-(d)4 パターンを示す．左列には Seed 感情が POSITIVE, NEUTRAL，NEGATIVE の $\Delta v$ の分布を載せ，右列にはそれ以外の $\Delta v$ の分布を載せている．平均値と信頼区間の表記は 5. 3. 1 と同様である．

まず (a) に着目すると，窓幅 1 時間あたりでは Failed と POSITIVE を除くすべての Seed 投稿の場合に $\Delta v$ の平均が正の値となっていることがわかる．特に “mostly_” 系の感情は平均値が高いが，信頼区間の幅もかなり広くなっている．窓幅を 6 時間ほどまで広げると速度差分の分布は “mostly_” 系，NEUTRAL，Failed の場合が 0 に漸近している．対して NEGATIVE な Seed 投稿のみ $\Delta v$ が正の値を保っており，逆に POSITIVE の場合は負の値を保っている．次に (b) に着目すると，左列では上述した「Seed 感情が NEGATIVE の $\Delta v$ が正の値をとり，NEUTRAL と POSITIVE の $\Delta v$ が負の値をとる」傾向が，0 への漸近をしつつ持続している．右列では “mostly_” 系感情の $\Delta v$ がやや 0 から外れて低くなる傾向があり，48 時間前後では-0.2 程度の平均となっている．

最後に (c) と (d) に着目すると，左列では NEGATIVE の $\Delta v$ が 0 より大きく,NEUTRAL の $\Delta v$ とPOSTIVE の $\Delta v$ が 0 より小さい関係がおおむねの区間で見られることがわかる．ただし $\Delta v$ の分布自体はいずれの Seed 感情でも窓幅を大きくするにつれて 0 に収束する挙動を見せており，Seed 感情間の速度差分にはっきりした差が見られるのは窓幅 3 週間あたりまである．右列では Failed と mostly_POSITIVE の $\Delta v$ がほぼ横ばいで平均が 0 に近い値の分布となっており，mostly_NEGATIVE の $\Delta v$ は平均が 0 に近づかずに窓幅 8 週間あたりまで負の値を保った形になっている．以上を踏まえると，窓幅 1-3 日辺りまでは特に NEGATIVE な感情の速度差分が大きくなる傾向が見られたことから数日単位では前後で投稿速度が上がっていることがわかる．よって 5. 3. 1 項と同じく検証 3 の結果からは**仮説 3** を支持することができると考えられる．

## 5. 4　考　　察

POSITIVE な Seed 投稿が POSITIVE でない投稿の数や投稿速度を比較的長期的に減らしている現象には締めくくりの投稿が影響していることが考えられる．議論が過熱することなく解決した場合，議論の出題者などが感謝の言葉を残し，それ以

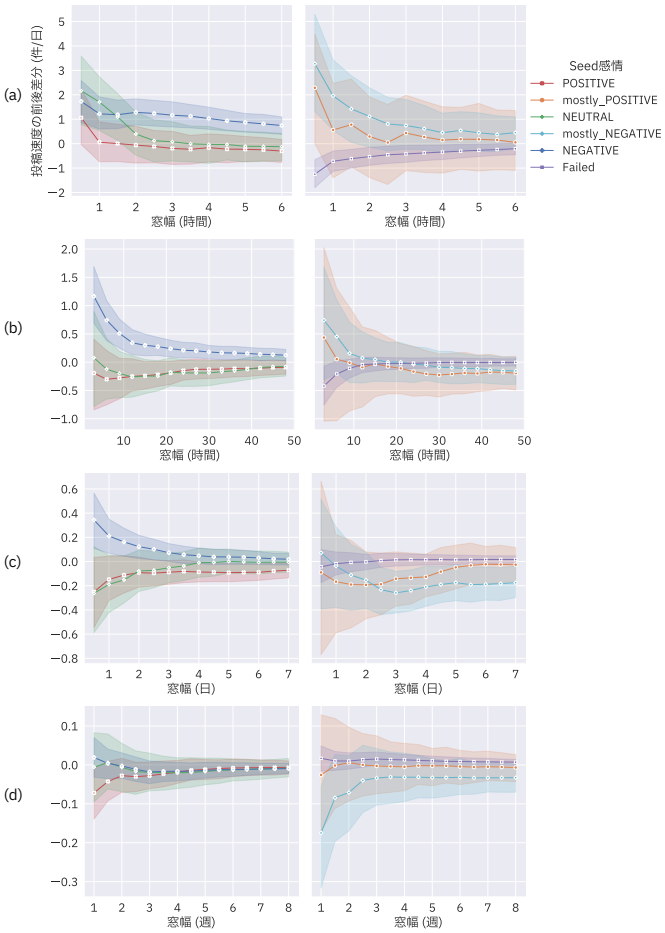


図 7: Seed 投稿の前後での速度差分 ($\Delta v$)．上から窓幅が (a)1-6 時間 (b)3-48 時間 (c)1-7 日間 (d)1-8 週間のときの分布を示す．横 2 列にそれぞれ Seed 感情を分けて掲載している．

降スレッドの書き込みが途絶えるケースがよく見られるためである．POSITIVE な Seed 投稿に予想されやすい「POSITIVE な投稿が多くなる効果」が今回の検証で見られなかったのもこの現象に関連する可能性がある．また，その後の投稿の数を長期的に増やす効果が見られないのは，図 1 からもわかるように元々NEGATIVE な投稿が多くなりやすい井戸端の特性との関連が推察される．

NEGATIVE な Seed 投稿に関しては仮説 1 のとおり，さらに NEGATIVE な投稿を呼び POSITIVE な投稿を減らすという現象がみられた．これは自然に推測できる仮説であるが，それをデータ面から裏付けることができた．また投稿速度の短時間の上昇も仮説 3 を裏付ける形で確認することができた．これは投稿の山が発生することを想定した言語的特徴の分析における過熱モデルとも合致しており，これをより具体的に示すことができた．一方で仮説 2「他の感情より返信が多くなる」に関しては，上述の結果にもかかわらず有意な結果を得ることができなかった．これについては返信数が 0 件の投稿が大半を占めていたことから，返信機能を使わないことで返信先を構造的に明確にしない返信が多く存在していることが推察される．

NEUTRAL な Seed 感情については 5.1 項で示したように，NEUTRAL な投稿が多くなる効果が見られた．投稿速度については短期的にも長期的にも減少する効果が見られた．このことから過熱しない議論は NEUTRAL な投稿が続くものを一種の基準として考えることができる．ただし，NEUTRAL と判定される投稿の中にも，潜在的に過熱を引き起こすものが存在しうる．これについてはより詳細な言語的特徴の解析が必要となる．“mostly_” 系の感情についてはデータ数が他の感情に比べて少なく，データ内のばらつきが全体的に大きい結果となった．よりはっきりとした結果を出すためにはデータ数の拡充が求められる．

## 6 おわりに

本稿ではオンライン議論において特定の感情を持った投稿がその後の投稿に及ぼす影響を調査すべく検証を行った．まず Wikipedia における日本語の議論ページにおける各投稿に対して Russell の円環モデルにもとづく感情ラベルを付与し，データセットを構築した．そして既存研究をもとに感情的投稿に関するいくつかの仮説を設定し，それにもとづき実験を行うことで仮説の是非を検証した．結果として，ネガティブな感情とはその後の投稿の感情を悪化させ，投稿の速度を一時的に上昇させることが既存の成果とも合致する形で明らかになった．またポジティブな投稿にはその逆の効果があることも明らかになった．今後の課題としてはデータセットの拡充や他の感情分析ライブラリ，言語，プラットフォームを用いた検証が挙げられる．また Russell の円環モデルのうち本稿の検証に用いていない arousal 軸についても，議論の投稿構造に何らかの影響があるか，引き続き調査を行いたい．

## 謝　辞

本研究は，JST，未来社会創造事業，JPMJMI21J2 および JSPS 科研費 JP19H01118 の支援を受けたものである．

## 文　献

[1] Adam G Zimmerman and Gabriel J Ybarra. Online aggression: The influences of anonymity and social modeling. *Psychology of Popular Media Culture*, Vol. 5, No. 2, p. 181, 2016.

[2] 木村昌紀, 余語真夫, 大坊郁夫. 日本語版情動伝染尺度 (the Emotional Contagion Scale) の作成. 対人社会心理学研究, Vol. 7, pp. 31–39, 2007.

[3] Ilan Price, Jordan Gifford-Moore, Jory Flemming, Saul Musker, Maayan Roichman, Guillaume Sylvain, Nithum Thain, Lucas Dixon, and Jeffrey Sorensen. Six Attributes of Unhealthy Conversations. In *Proceedings of the Fourth Workshop on Online Abuse and Harms*, pp. 114–124, Online, November 2020. Association for Computational Linguistics.

[4] Aleksi H Syrjämäki, Mirja Ilves, Poika Isokoski, Joel Kiskola, Anna Rantasila, Thomas Olsson, Gary Bente, and Veikko Surakka. Emotionally toned online discussions evoke subjectively experienced emotional responses. *Journal of Media Psychology: Theories, Methods, and Applications*, 2022.

[5] Susann Kohout, Sanne Kruikemeier, and Bert N. Bakker. May i have your attention, please? an eye tracking study

on emotional social media comments. *Computers in Human Behavior*, Vol. 139, p. 107495, 2023.

[6] Salud María Jiménez-Zafra, Antonio José Sáez-Castillo, Antonio Conde-Sánchez, and María Teresa Martín-Valdivia. How do sentiments affect virality on twitter? *Royal Society Open Science*, Vol. 8, No. 4, p. 201756, 2021.

[7] 小林将大, 矢田竣太郎, 若宮翔子, 荒牧英治. オンライン議論の過熱の言語的誘因分析. 人工知能学会全国大会論文集, Vol. JSAI2022, p. 2H5OS11a04, 2022.

[8] Sameera Horawalavithana, Nazim Choudhury, John Skvoretz, and Adriana Iamnitchi. Online discussion threads as conversation pools: predicting the growth of discussion threads on reddit. *Computational and Mathematical Organization Theory*, Vol. 28, No. 2, pp. 112–140, 2022.

[9] Dhruv Sahnan, Snehil Dahiya, Vasu Goel, Anil Bandhakavi, and Tanmoy Chakraborty. Better prevent than react: Deep stratified learning to predict hate intensity of twitter reply chains. In *2021 IEEE International Conference on Data Mining (ICDM)*, pp. 549–558, 2021.

[10] Rafal Rzepka Michal Ptaszynski, Pawel Dybala and Kenji Araki. Affecting corpora: Experiments with automatic affect annotation system-a case study of the 2channel forum. In *PACLING-09*, pp. 223–228, 2009.

[11] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, Vol. 39, No. 6, p. 1161, 1980.

[12] H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, Vol. 18, No. 1, pp. 50 – 60, 1947.