

機械学習を用いた X 上の炎上予測モデルの提案

家村 芽弥^{1,a)} 小玉 直樹² 齋藤 孝道^{2,3}

概要：インターネット技術の普及と発展に伴い、ソーシャルネットワーキングサービス（SNS）が広く利用され、情報発信や共有において中心的な役割を果たすようになった。この進展は情報の発信や共有を容易にする一方で、炎上や偽情報の拡散といった新たな課題をもたらしている。本論文では、炎上問題に対応するため炎上予測モデルの開発を目的とし、機械学習技術を用いて SNS 上の単一ポストが炎上するか否かを早期に予測する二値分類モデルを二段階アプローチで構築した。X で話題となったポストとアカウント情報をデータセットとして採用し、各ポストに対する炎上予測及びその予測結果に基づいて分析を行った。その結果、単一ポストが炎上する前に、炎上すると予測することができた。また、政治的な話題では、比較用の話題を使用した比較データセットと比べ、炎上すると予測したポストの割合が大きかった。

キーワード：ソーシャルメディア, X, Twitter, 炎上

Proposal for a Machine Learning-Based Flare-Up Prediction Model on Platform X

MEIYA IEMURA^{1,a)} NAOKI KODAMA² TAKAMICHI SAITO^{2,3}

Abstract:

With the widespread adoption and development of internet technology, social networking services (SNS) have become widely used, playing a central role in information dissemination and sharing. While this progress has facilitated the transmission and sharing of information, it has also introduced new challenges such as the spread of online firestorms and misinformation. This paper aims to address the problem of online firestorms by developing a firestorm prediction model. We constructed a binary classification model using machine learning techniques to predict whether a single post on SNS will ignite a firestorm early on. We adopted posts and account information that became topics of interest on platform X as our dataset and conducted predictions and analyses based on those predictions for each post. As a result, it was possible to predict a post would become a firestorm before it actually did. Additionally, political topics showed a higher proportion of predicted firestorms compared to a comparative dataset using different topics.

Keywords: Social Media, Twitter, X, social media firestorm

1. はじめに

インターネット技術の普及と発展により、ソーシャルネットワーキングサービス（SNS）が広く利用されるよう

になり、情報発信や情報共有において有益なツールとして中心的な役割を果たすようになった。しかし、その情報発信力や情報共有力を悪用する事例も度々報告されている。たとえば、Rice ら [1] によると、X 社は 2016 年の米国大統領選挙において、ロシアから X (旧 Twitter) を介して工作行為が行われたとのことである。

このような SNS の高い情報拡散力は、世論を特定の方向へ誘導する行為や社会的不和を招こうとする行為は「影響力工作」と呼ばれ、直ちに犯罪ではないため、行為者が

¹ 明治大学大学院
Graduate School of Meiji University
² 明治大学
Meiji University
³ レンジフォース株式会社
Rangeforce, Inc.
^{a)} ce235004@meiji.ac.jp

公開されることなく秘匿される。その結果、対処が困難であり、抑止されないことが多い。影響力工作を抑止したい側は、当該行為者を特定し、いわゆるアトリビューションを試みる。アトリビューションとは、「サイバー活動の背後に誰がいたのか、なぜそのようなことをしたのかを導く分析プロセス」である。

本論文では、X 上のポストが炎上する可能性を予測するために、拡散予測モデルと感情分析モデルを組み合わせた二段階アプローチを用いて、炎上予測モデルの作成を行った。

また、各ポストに対する炎上予測及びその予測結果に基づいて分析を行った。政治的な話題のデータセットと政治と無関係な企業の広告やニュースでトレンドになった比較用のデータセットに対して、作成した炎上予測モデルを適用した。その結果、単一ポストが炎上する前に、炎上すると予測することができた。また、炎上予測モデルを政治的な話題のデータセットと比較用の話題のデータセットに適用した結果、比較用の話題よりも政治的な話題の方が、炎上すると予測されたポストの割合が高かった。

2. 関連知識

2.1 炎上

総務省 [2] によると、炎上とは、「ウェブ上の特定の対象に対して批判が殺到し、収まりがつかなさそうな状態」、「特定の話題に関する議論の盛り上がり方が尋常ではなく、多くのブログや掲示板などでバッシングが行われる状態」とされている。本論文では、文献 [3][4] に基づき、「多くの人々が集中的に否定的な反応を示す現象」とする。

2.2 Transformer

Transformer[5] とは、2017 年に発表された深層学習モデルであり、自然言語などの時系列データを扱ってタスクを行うように設計されている。自然言語処理で広く用いられていた RNN（リカレントニューラルネットワーク）構造を使用せず、Attention を核にした構造を持つ。

2.3 RoBERTa

RoBERTa (Robustly Optimized BERT Approach) [6] とは、BERT (Bidirectional Encoder Representations from Transformers) モデルを改良した、Transformer ベースの深層学習モデルである。

2.4 影響力工作 (influence operations)

アメリカの RAND 研究所 [7] によると、影響力工作とは、「敵対者に対する競争上の優位性を追求するために、敵対者に関する戦術的情報の収集を行いプロパガンダの普及を行うこと」とされている。本論文では文献 [8] に基づき、影響力工作を「国家間など敵対する組織間での競争（戦い）

における情報戦の一種で、競争相手国の意思決定に影響を与え、ターゲットの行動の変容を促す一連の行為」とする。類似の概念・用語として、「積極工作 (active measures)」、「情報操作 (information manipulation)」、「プロパガンダ工作」、「フェイクニュース」があげられるが、ここでは包括的な概念として影響力工作を用いる。

3. 関連研究

3.1 ポストの拡散予測に関する研究

M.T. Uddin ら [9] は、テキストや画像の内容にアクセスできない状況でも、メタデータのみでコンテンツの人気度を正確に予測することを目指した。M.T. Uddin らは、投稿時間、投稿者のフォロワー数、過去のエンゲージメントなどのメタデータだけを用いて、ニュース記事の人気度を予測する手法を提案している。

この手法では、機械学習モデルを用いて、メタデータからコンテンツが「人気がある」か「人気がない」かを予測する二値分類モデルを開発した。結果として、メタデータのみでも高い予測精度が達成され、特に投稿時間や投稿者の影響力が予測に大きく寄与することが確認された。

3.2 炎上に関する研究

Kim ら [3] は、中国の Weibo とアメリカ合衆国の Twitter におけるオンライン炎上の違いを調査した。

6 か月間にわたるトレンドワードと関連ツイートの定量的コンテンツ分析を通じて、オンライン炎上と見なされる基準や対象が中国とアメリカ合衆国では違うことが判明した。アメリカ合衆国のユーザは、政府や政治に対して嘲笑や社会的・政治的意見を表明する傾向が強く、集団行動を積極的に行う。一方で、中国のユーザは企業やメディアをターゲットにすることが多く、社会的責任に関する炎上が一般的である。この研究は、文化や社会的背景がオンライン炎上の性質やターゲットに大きく影響することを明らかにした。

4. 炎上予測モデル

4.1 概要

本論文では、X 上のポストが炎上する可能性を予測するために、二段階アプローチを採用している。

最初の段階では、機械学習に基づく拡散予測モデルを用いて、単一のポストが広く拡散される可能性を評価する。このモデルは、Transformer を利用した二値分類モデルであり、フォロワー数、ツイート時の時間帯、アカウント作成日時など、複数の特徴量を組み合わせている。拡散が予測されたポストは次の段階へと進む。

第二の段階では、拡散すると予測されたポストに対して、RoBERTa ベースの感情分析モデルを用いた二値分類を適用し、ポストの感情的なトーンを「positive」と「negative」

に分類する。この分析により、ネガティブな感情を含む拡散されたポストを特定し、これらが炎上する可能性が高いと判定する。

4.2 拡散予測モデル

4.2.1 拡散の定義

Zhu ら [10] によると、拡散とは、「拡散は情報がどれだけ迅速かつ広範囲に共有されるかを示すプロセス」であるとしていいる。本論文では、「情報が投稿者から投稿という媒介を通じて受信者に伝達する行為」とする。

4.2.2 学習データセット

二値分類モデルの作成に際して使用した学習データセットは、拡散するとしてポストを 313,647 件、拡散しないとしてポストを 752,392 件の計 1,066,039 件を使用した。そのうちの 90 % を学習データ、10 %をテストデータとして使用した。

4.2.3 精度評価指標

拡散予測モデルは回帰分析アルゴリズムを用いて、「拡散する」、「拡散しない」の 2 つのクラスに分類する 2 値分類モデルである。判別結果と正解ラベルに基づき、表 1 のような混同行列を作成できる。

表 1: 混同行列 (拡散予測モデル)

| | | 正解ラベル | |
|--------------|-------|-------|------|
| | | 拡散しない | 拡散する |
| 二値分類モデルの判別結果 | 拡散しない | TP | FP |
| | 拡散する | FN | TN |

表 1 に基づき Accuracy, Precision, Recall, F1 値をそれぞれ算出して精度評価を行う。以下に Accuracy, Precision, Recall, F1 値を求める式を示す。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2(Precision \cdot Recall)}{Precision + Recall}$$

4.2.4 特徴量

特徴量については、取得したデータをそのまま使用した特徴量が 5 種類、本文やリポスト数から作成した特徴量を 5 種類使用した。表 2 と表 3 に一覧を示す。

表 2: アカウントとテキストデータの特徴量

| 特徴量 | 内容 |
|-------------------|-----------------------|
| author-created-at | アカウントが作成されてからの経過時間 |
| created-at | 投稿の投稿日時 |
| followers-count | アカウントがフォローしているアカウントの数 |
| following-count | アカウントをフォローしているアカウントの数 |
| tweet-count | アカウントが投稿した数 |

表 2 より、アカウントデータの特徴量は「author-created-at」、「followers-count」、「following-count」、「tweet-count」である。テキストデータの特徴量は「created-at」である。

表 3: 作成したデータの特徴量

| 特徴量 | 内容 |
|----------------|-----------------------|
| text-length | 投稿のテキストの長さ |
| day-of-week | 投稿した日時の曜日 |
| hour-of-day | 投稿時間帯 |
| hourlycount | 時間差分 (1 時間) |
| tenminutecount | 時間差分 (10 分間) |
| countsdata | 10、30、60、120 分後のツイート数 |

表 3 は、本文やリポスト数から作成した特徴量を示している。この表に含まれる各特徴量の内容は以下の通りである。

- text-length: 各投稿のテキストの長さを示している。この特徴量は、テキストの長さそのものを分析するために使用される。
- day-of-week: 投稿が行われた曜日を表す。この特徴量は、曜日ごとの投稿頻度や反応の違いを分析する際に使用される。
- hour-of-day: 投稿された時間帯を示す特徴量である。この特徴量は、時間帯による投稿の拡散パターンや反応を評価するために利用される。
- hourlycount: 投稿後 1 時間隔でのリツイート数の増減を記録した特徴量である。この特徴量は、投稿がどれほど迅速に拡散されたかを評価するための指標である。
- tenminutecount: 投稿後 10 分間隔でのリツイート数の増減を記録した特徴量である。この特徴量は、短時間での拡散速度を詳細に把握するために使用される。
- countsdata: 投稿後 10 分、30 分、60 分、および 120 分後のリツイート数を示している。時間経過に伴う投稿の拡散動向を分析するために使用される。

4.2.5 精度

本論文で作成した拡散予測モデルの二値分類モデルの混同行列を表 4 に示す。

表 4: 二値分類モデルの精度評価値 (拡散予測モデル)

| Acc | Pre | Rec | F1 |
|-------|-------|-------|-------|
| 0.999 | 0.971 | 0.825 | 0.892 |

モデルの評価結果は、正解率 (Accuracy) が 0.999, 精度 (Precision) が 0.971, 再現率 (Recall) が 0.825, そして F1 値が 0.892 という結果を示した。また、モデルの判定処理速度に関しては、170,176 件のデータを 3.145 秒で処理することができることが確認された。

4.3 感情分析モデル

4.3.1 学習データセット

二値分類モデルの作成に際して使用した学習データセットは、Positive のポストを 355,116 件、Negative からのツイートを 354,951 件の計 710,067 件を使用した。そのうちの 90 % を学習データ、10 % をテストデータとして使用した。また、正解データはクラウドサービスを利用して判定された結果に基づいている。

4.3.2 精度評価指標

感情分析モデルは、様々な特徴量を用いてポストが否定的なポストかどうかを予測する。作成したモデルは RoBERTa を用いて、「Positive」、「Negative」の 2 つのクラスに分類する 2 値分類モデルである。判別結果と正解ラベルに基づき、表 5 のような混同行列を作成できる。

表 5: 混同行列 (感情分析モデル)

| | | 正解ラベル | |
|--------------|----------|----------|----------|
| | | positive | negative |
| 二値分類モデルの判別結果 | Positive | TP | FP |
| | Negative | FN | TN |

表 1, 表 5 に基づき Accuracy, Precision, Recall, F1 値をそれぞれ算出して精度評価を行う。以下に Accuracy, Precision, Recall, F1 値を求める式を示す。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2(Precision \cdot Recall)}{Precision + Recall}$$

4.3.3 精度

本論文で作成した感情分析モデルの二値分類モデルの混同行列を表 6 を示す。

表 6: 二値分類モデルの精度評価値 (感情分析モデル)

| Acc | Pre | Rec | F1 |
|-------|-------|-------|-------|
| 0.728 | 0.616 | 0.664 | 0.719 |

モデルの評価結果は、正解率 (Accuracy) が 0.728, 精度 (Precision) が 0.616, 再現率 (Recall) が 0.664, そして F1 値が 0.719 という結果を示した。

5. 予測実験

本節では、4 章で作成したモデルを使った炎上予測実験の詳細について示す。ここでの予測実験とは、作成したモデルに対して、X でトレンドとなった話題のポストとポストを投稿したアカウントの情報を入力とし、予測を行うことである。出力される値は、炎上する可能性が高い、炎上する可能性が低いかのどちらかである。

5.1 予測実験で用いたデータセット

収集したデータセットの一覧を表 7 に示す。また、表 8 に各データセットの概要を示す。収集方法は、期間を指定し、キーワードでフィルタリングを行い収集を行った。「#ローソンのロールケーキ食いたい」、「au 障害」、「#台風 14 号」の 3 つは、政治的な話題との比較対象として使用した。以降、特に断りが無い限り、他のポストやアカウントとの直接的な相互作用なしに投稿されたポスト、すなわち、リポストやリプライではないポストをオリジナルポストと呼ぶ。

5.2 予測実験手法

この実験では、5.1 節のデータセットと 4 章で作成した炎上予測モデルを用いて、炎上予測を行った。この時、炎上予測を行うポストはオリジナルポストを対象として行った。結果の数と割合を表でまとめ、特定の話題がその他の話題と比較して、何かしらの異なる傾向が現れるかどうか調査した。

5.3 予測実験結果

予測実験の結果を表 9 に示す。炎上するポストの割合は、その話題のポスト全体を 100 とした時の割合を示す。炎上予測は、拡散すると予測されたデータを対象に行った。

表 9 から、政治的な話題のデータセットは、比較用の話題を使用したデータセットと比較して、炎上すると予測したポストの割合が高いことがわかる。特に、「G7 LGBT」で

表 7: データセット一覧

| キーワード | 収集期間 | ポスト数 | オリジナルポスト数 |
|------------------|-------------------------|---------|-----------|
| G7 汚染水 | 2023/01/05 ~ 2023/05/28 | 10,961 | 994 |
| G7 BRICs | 2023/01/01 ~ 2023/05/28 | 149,282 | 11,595 |
| G7 LGBT | 2023/01/01 ~ 2023/06/07 | 538,103 | 13,981 |
| 2022 年沖縄知事選挙 | 2022/08/31 ~ 2022/09/16 | 955,096 | 25,779 |
| #ローソンのロールケーキ食べたい | 2022/08/10 ~ 2022/08/25 | 342,115 | 871 |
| au 障害 | 2022/06/30 ~ 2022/07/06 | 70,603 | 10,433 |
| #台風 14 号 | 2022/09/12 ~ 2022/09/26 | 417,975 | 51,034 |

表 8: データセットの概要

| キーワード | 関連するイベント |
|------------------|-----------------------------------|
| G7 汚染水 | G7 と 福島原発処理水をめぐる意見の盛り上がり |
| G7 BRICs | G7 と BRICS の GDP をめぐる長期的な意見の盛り上がり |
| G7 LGBT | G7 と LGBT に関する意見の盛り上がり |
| 2022 年沖縄知事選挙 | 2022 年 9 月 11 日に投票が行われた沖縄知事選挙 |
| #ローソンのロールケーキ食べたい | 企業広告 |
| au 障害 | 2022 年 7 月に発生した KDDI の通信障害 |
| #台風 14 号 | 2022 年 9 月に小笠原近海で発生した台風 |

は、炎上するポストの割合が約 0.7 % と非常に高くなっている。また、政治的な話題である「G7 汚染水」「G7 BRICs」「G7 LGBT」では、拡散すると予測されたすべてのポストが炎上すると予測された。

5.4 考察

表 9 の炎上予測実験結果から、すべてのデータセットにおいて、炎上するポストの割合が 1 % 未満と非常に小さく、ごく少数のポストが大規模な拡散・炎上につながっていることが示唆される。特に、政治的な話題である「G7 汚染水」「G7 BRICs」「G7 LGBT」「2022 年沖縄知事選挙」では、炎上すると予測されたポストの割合が高かった。特に「G7 LGBT」では、他の政治系の話題のデータセットと比較用の話題を使用したデータセットに比べて、炎上すると予測されたポストの割合が高い。これより、社会的な感受性が高いため、ディスカッションが活発になりやすいことを示している可能性があると考えられる。

一方で、比較用の話題を使用したデータセットである「#ローソンのロールケーキ食べたい」「au 障害」「#台風 14 号」では、炎上すると予測したポストの割合が炎上データセットに比べて小さいことが確認された。「#ローソンのロールケーキ食べたい」では、ポスト数に対してオリジナルポストが少なく、リポストが多いデータセットであったにもかかわらず、拡散すると予測されたポストは 3 件のみで、炎上すると予測されたポストはなかった。政治的な話題のデータセットでは、ポスト数が多い場合、オリジナルポストも多くなり、炎上すると予測されたポストも増加

した。

このように、炎上予測モデルは政治的な話題において炎上の可能性を高く評価し、一般的な話題ではその可能性を低く評価する傾向が見られた。これは、モデルの感情分析と拡散予測の能力が適切に機能していることを示しており、特にセンシティブな話題に対して有効に働く可能性が高いと考えられる。

6. 比較実験

本節では、5 章で述べた炎上予測の結果に基づき、比較実験を行った。比較対象は表 7 に示した 7 つのデータセットである。

6.1 実験手法

この実験では、5 章の炎上予測実験結果に基づき、炎上すると予測されたポストがボットアカウントによるものか、人間のアカウントによるものかを、ボット判別モデルを用いて分析を行う。また、データセットごとに比較を行う。今回使用したボット判別モデルには、RandomForestRegressor を用いた二値分類モデルを採用している。

6.2 結果

拡散予測モデルによる予測結果とボットのアカウントによるポストの割合を調査した結果を、表 10 に示す。炎上予測モデルの結果を表 11 に示す。ボット拡散率は、ボットによるポストのうち、拡散すると予測されたポストの割合を示しており、その話題における拡散予測ポスト全体を 100 としたときの割合を示す。ボット炎上率は、ボットによるポストのうち、炎上すると予測されたポストの割合を

表 9: 炎上予測実験結果

| キーワード | 拡散すると予測 | 拡散しないと予測 | 炎上すると予測 | 炎上しないと予測 | 炎上するポストの割合 |
|------------------|---------|----------|---------|----------|------------|
| G7 汚染水 | 1 | 993 | 1 | 0 | 0.101 |
| G7 BRICs | 17 | 11,578 | 17 | 0 | 0.147 |
| G7 LGBT | 104 | 13,877 | 104 | 0 | 0.744 |
| 2022 年沖縄知事選挙 | 120 | 25,659 | 61 | 59 | 0.237 |
| #ローソンのロールケーキ食べたい | 3 | 868 | 0 | 3 | 0.000 |
| au 障害 | 6 | 10,429 | 0 | 6 | 0.000 |
| #台風 14 号 | 27 | 51,007 | 7 | 20 | 0.014 |

示しており、その話題における炎上予測ポスト全体を 100 としたときの割合を示す。

表 10 から、政治系の話題のデータセットでは、比較用の話題を使用したデータセットと比較して、ボットによるポストが拡散する割合が低いことがわかる。特に、「G7 汚染水」「G7 BRICs」においては、0%，約 5.9%と低くなっている。

表 10: 拡散予測モデルにおけるボットによるツイートの割合

| キーワード | 拡散すると予測 | ボットによるポスト数 | ボット拡散率 |
|------------------|---------|------------|---------|
| G7 汚染水 | 1 | 0 | 0.000 |
| G7 BRICs | 17 | 1 | 5.882 |
| G7 LGBT | 104 | 16 | 15.385 |
| 2022 年沖縄知事選挙 | 120 | 17 | 14.167 |
| #ローソンのロールケーキ食べたい | 3 | 3 | 100.000 |
| au 障害 | 6 | 1 | 16.667 |
| #台風 14 号 | 27 | 5 | 18.519 |

表 11 から、政治系の話題の「G7 LGBT」「2022 年沖縄知事選挙」は、炎上すると予測されたポストの割合が高くなっている。比較用の話題を使用したデータセットの「#台風 14 号」では、炎上すると予測されたポストの割合が 28.571 と高くなった。

表 11: 炎上予測モデルにおけるボットによるツイートの割合

| キーワード | 炎上すると予測 | ボットによるポスト数 | ボット炎上率 |
|------------------|---------|------------|--------|
| G7 汚染水 | 1 | 0 | 0.000 |
| G7 BRICs | 17 | 1 | 5.882 |
| G7 LGBT | 104 | 16 | 15.385 |
| 2022 年沖縄知事選挙 | 61 | 8 | 13.112 |
| #ローソンのロールケーキ食べたい | 0 | 0 | 0.000 |
| au 障害 | 0 | 0 | 0.000 |
| #台風 14 号 | 7 | 2 | 28.571 |

6.3 考察

表 10 と表 11 の結果から、政治的な話題においては、ボットによる活動が確認されたが、主に人間のユーザによる活動が中心であることが確認された。特に、政治系の話題の「G7 LGBT」や「2022 年沖縄知事選挙」では、一定数の

ボットによる投稿が見られたものの、全体としてはボットによる拡散や炎上の割合は比較的低く、主に人間のユーザによる活動が中心であることが確認された。

一方、比較用の話題を使用したデータセットにおいては、ボットが拡散に非常に大きな役割を果たしていることが明らかになった。特に、比較用の話題を使用したデータセットの「#ローソンのロールケーキ食べたい」では、拡散されたすべてのポストがボットによるものであり、ボット拡散率が 100%に達している。これは、ボットが意図的にこの話題を拡散している可能性を強く示唆している。

また、表 11 のボット炎上率に関しては、特に「#台風 14 号」や「G7 LGBT」などの話題で、ボットが炎上の一因となっていることが示されている。しかし、政治的な話題のデータセット全般では、ボット炎上率は相対的に低く、炎上は主に人間のユーザによる影響が大きいことがわかった。

これらの結果から、ボットの役割は話題の種類に依存している可能性があり、政治系の話題ではない比較用の話題のデータセットではボットが拡散に大きく寄与する一方、政治的な話題ではボットの影響が限定的であることが明らかになった。

7. まとめ

本論文では、X 上のポストが炎上する可能性を予測するために、拡散予測モデルと感情分析モデルを組み合わせた二段階アプローチを用いて、炎上予測モデルの作成を行った。拡散予測モデルのデータセットには 1,066,039 件のポストを使用し、F1 値は 0.892 であった。感情分析モデルのデータセットには 710,067 件のポストを使用し、F1 値は 0.719 であった。

作成した炎上予測モデルを政治的な話題のデータセットと比較用の話題を使用したデータセットに適用したところ、比較用の話題に比べて、政治的な話題のデータセットの方が炎上するポストの割合が高いことが確認された。今回使用した政治的な話題のデータセットと比較用の話題を使用したデータセットは、厳格なランダムサンプリングによって選ばれた多様な話題から成るデータセットを用いた。また、政治的な話題ではオリジナルポストが多く、それに伴って炎上の可能性も高まる一方、比較用の話題を使

用したデータセットでは炎上のリスクが低いことが確認された。

さらに、比較実験では、政治的な話題ではない比較用の話題のデータセットにおいてボットが拡散に大きく寄与しているのに対し、政治的な話題ではボットの影響が限定的であることが示された。

今後は、炎上予測モデルの精度向上や時系列に焦点を当てた分析、さらにデータセットを増やして炎上すると予測されたポストを投稿したアカウントの特徴分析を試みたい。

8. 研究倫理

本論文では、個人の特定を防ぐために、ポストの詳細やアカウントの具体名は省略して記述した。

謝辞

本研究はレンジフォース株式会社の支援により実施した。

参考文献

- [1] Damian J Ruck, Natalie M Rice, Joshua Borycz, and R Alexander Bentley. Internet research agency twitter activity predicted 2016 u.s. election polls. *First Monday*, Vol. 24, No. 7, Jun 2019.
- [2] 令和元年版 情報通信白書. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd114300.html>.
- [3] Yingru Ji Chen Xing Jiayu Gina Qu Sora Kim, Kang Hoon Sung. Online firestorms in social media: Comparative research between china weibo and usa twitter. *Public Relations Review*, Vol. 47, pp. 49–61, March 2021.
- [4] Bruno S. Frey Katja Rost, Lea Stahel. Digital social norm enforcement: Online firestorms in social media. *PLOS ONE*, Jun 2016.
- [5] Niki Parmar Jakob Uszkoreit Llion Jones Aidan N. Gomez Lukasz Kaiser Illia Polosukhin Ashish Vaswani, Noam Shazeer. Attention is all you need. *arXiv*, Jun 2017.
- [6] Yutaka Watanobe Md. Ashad Alam Md. Mostafizer Rahman, Ariful Islam Shiplu. Roberta-bilstm: A context-aware hybrid model for sentiment analysis. *arXiv*, Jun 2024.
- [7] RAND Corporation. Information operations. <https://www.rand.org/topics/information-operations.html>.
- [8] 一田和樹, 齋藤孝道, 他. ネット世論操作とデジタル影響工作—「見えざる手」を可視化する. 原書房, 2023.
- [9] Tanveer Ahsan Mohammed Shamsul Alam Department of Computer Science Md. Taufeeq Uddin, Muhammed Jamshed Alam Patwary and Bangladesh Engineering, International Islamic University Chittagong. Predicting the popularity of online news from content metadata. *IEEE Access*, October 2016.
- [10] Haseon Park a Xun Zhu, Youllee Kim. Do messages spread widely also diffuse fast? examining the effects of message characteristics on information diffusion. *Computers in Human Behavior*, Vol. 103, pp. 37–47, February 2020.