Vol.2020-CSEC-91 No.16
Vol.2020-SPT-40 No.16
Vol.2020-EIP-90 No.16
2020/11/25

Twitterにおけるボット判別技術の現状と課題

田畑 唯斗[1]　嶋田 里聖[1]　利光 能直[2]　菊田 翼[2]　齋藤 孝道[1]

概要：Twitter などのソーシャルメディアにおいて，コンピュータプログラムによりコンテンツを自動で投稿するボットと呼ばれるアカウントがある．一般に，ボットはソーシャルメディア上でのニュース配信や広告配信などを目的に活用されている．しかし，一部のソーシャルメディア利用者は，フェイクニュースの拡散や情報操作などの行為にボットを利用している．このような行為を防ぐため，研究機関などにより様々なボット判別技術の提案がされているが，国内では言語の問題もあり海外の手法をそのまま利用できない．本論文では，国内外におけるボット判別技術の現状について調査し，国内でボットを判別する際の問題点について整理した．

キーワード：ボット，ボット判別，Twitter，トロール，フェイクニュース

Current state of bot detection technology in Twitter

Abstract: In social media such as Twitter, there are accounts called bots that automatically post content using a computer program. In general, bots are used for the purpose of distributing news and advertisements on social media. However, some social media users are using bots for activities such as spreading fake news and manipulating information. Various techniques have been proposed by research institutes and other organizations to prevent such behavior, but due to language barriers, foreign techniques cannot be used in Japan. In this paper, we surveyed the current state of bot detection technology in Japan and abroad, and summarized the problems involved in discriminating Bots in Japan.

Keywords: Bot，Bot detection，Twitter，Troll，Fake news

1. はじめに

Twitter などのソーシャルメディアの普及に伴い，誰でも手軽に情報を発信できるようになった．また，ソーシャルメディア上での発言がニュースに取り上げられたり，著名人が情報を発信することによって，人々の購買傾向や移動傾向などが変化することがある．そのため，ソーシャルメディアは社会に多大な影響を及ぼすとして様々な方面から注目されている．

ソーシャルメディア上でコンピュータプログラムによりコンテンツを自動で投稿するボットと呼ばれるアカウントがある．一般にボットは，ソーシャルメディア上でのニュース配信や広告配信などを目的に活用されている．しかし，一部のソーシャルメディア利用者は，フェイクニュースの拡散や炎上を発生させることを目的とした，一般的でない行為にボットを利用している．これらの行為により，世論がある方面に有利になるよう誘導される等の悪影響がもたらされる恐れがある．このような行為を防ぐため，様々な研究機関などにより，Twitter アカウントがボットかどうか判別する技術 (以降，ボット判別技術と呼ぶ) が提案されている．

本論文では，Twitter においてボットが一般的でない行為に利用された事例や国内外におけるボット判別技術の現状について調査した．また，ボット判別技術に関するツールや応用事例について併せて調査した．さらに，これらの調査を踏まえ，国内でボットを判別する際に考えられるいくつかの課題について整理した．

[1] 明治大学
Meiji University
[2] 明治大学大学院
Graduate School of Meiji University

2. 関連知識

2.1 ボット

ボットとは，コンピュータプログラムによって自動化さ

れたソフトウェアやシステムのことである．特に，ソーシャルメディアにおいて，投稿や拡散などの行為が自動化されたアカウントのことをソーシャルボットと呼ぶ．

Twitter の自動化ルール [1] によると，ソーシャルボットを用いて自動的に，同一内容を何度も投稿するスパム行為や無関係なリンクを含む投稿をすることなどは禁止されている．これらの行為や Twitter の規約 [2] に違反する行為を行うとアカウントが凍結される可能性がある．

本論文におけるボットとは，ソーシャルボットのことを指すとする．

Yang ら [3] によるとボットは以下の 3 種類に分類できる．

- **コンテンツを自動投稿するボット**
  事前にボット作成者が設定した時間に，決められたコンテンツを投稿する．また，投稿時のトレンドを自動で取得し，投稿するコンテンツを柔軟に変更するボットも存在する．例として，ニュース配信や広告配信用のボットが挙げられ，一般的な利用方法である．
- **人間の操作するアカウントになりすますボット**
  ユーザ名やユーザ ID が，ある人間の操作するアカウント（以降，人間のアカウントと呼ぶ）に非常に似通った偽のアカウントを作成し，ターゲットとなった人間のアカウントと同じような投稿を繰り返す．これにより，ソーシャルメディア運営側に削除されにくいボットを作成できる．
- **ボット同士で協調するボット**
  複数の他のボットとグループを構成し，特定の話題に関する投稿を大量に行う．これにより，本来注目されていない話題がトレンドに上がる可能性がある．

Twitter におけるこれらのボットは，公式の開発者アカウントや，Twitter 外部サービスなどを用いた様々な方法で作成することができる．

2.2 Twitter におけるメタデータ

Twitter においてメタデータとは，アカウントやツイートに関するデータのことである．例えば，アカウントのフォロー数，フォロワー数，総ツイート数，自己紹介文などが挙げられる．このメタデータは Twitter API[4] を用いて取得でき，ボット判別の際の特徴量として利用される．

2.3 トロール

トロールとは，意図的に挑発的な内容や無関係な話題の投稿を行い，場を荒らすアカウントのことである．これにより，意図的に人気の操作や世論の誘導を行うこともある．トロールは投稿が自動化されていない場合もあるので，厳密にはボットと異なる．また，トロールに関して有名な組織として，ロシアのインターネットリサーチエージェンシー（IRA）が挙げられる．IRA は「オリギノのトロール工場」とも呼ばれ，大量のトロールを操作している．

2.4 フェイクニュース

インターネットやその他メディアで配信される，事実と異なる情報のことである．ジョーク目的だけでなく，著名人や政治活動などの信頼に影響を与えることを目的に作成されることもある．

フェイクニュースに関する研究は，2016 年に行われた米国大統領選がきっかけで盛んになっている．例えば 2019 年，Bovet ら [5] は，フェイクニュースが 2016 年の米国大統領選にどれほど関与していたか調査した．Twitter 上で，この選挙の候補者に関連するアカウント 1,100 万件から収集した 1 億 7,100 万件のツイートを用意し，そのうち記事へのリンクを含む 3,000 万件のツイートを抽出した．このツイートを分析した結果，25%がフェイクニュースへのリンクであった．

2018 年，Shao ら [6] は，フェイクニュースの拡散にボットが関係しているかどうか調査した．2016〜2017 年の 10 ヶ月間，40 万種類の記事を広めた 1,400 万件のツイートを収集し分析を行った．その結果，記事が大規模に拡散される前の早い段階で，ボットが拡散を行なっていることを明らかにした．その際，フォロワーが多いアカウントをターゲットにリプライやメンションを行ったり，その記事の内容を信じやすい人を標的にしていた．

一方，Vosoughi[7] らの研究では，人間のアカウントによって，真実よりも嘘の方が何倍も早く拡散されると結論付けている．この研究では，2006 年〜2017 年の間に 300 万件のアカウントによって拡散された，125,000 種類の記事を分析した．その結果，ボットより人間のアカウントの方がフェイクニュースの拡散に影響を与えていた．また，ボットはフェイクニュースと正しいニュースどちらの拡散も行っていたが，その影響力は同程度であった．しかし，フェイクニュースの方が人間のアカウントに拡散されやすいため，正しいニュースより数倍も早く拡散された．

2.5 ソーシャルメディア上での大規模な情報拡散

ソーシャルメディア上で大規模な情報拡散が起こることで，人々の購買傾向や移動傾向などに影響が及ぶ可能性がある．その中で特に，批判的な意見などが話題にあがり，大規模に拡散される状況は，炎上と呼ばれている．原因となる記事やツイートの真偽に関わらず炎上する恐れがある．

例えば，2020 年 5 月に「#検察庁法改正案に抗議します」というハッシュタグを含むツイートが大規模に拡散された．このハッシュタグを含むツイートはリツイートを含め，2 日間で約 480 万件以上投稿され，Twitter でのトレンドに掲載された．

また，2018 年 9 月にスポーツ関連商品を扱う NIKE の 30 周年記念キャンペーンにおいて，批判が殺到し株価が低下した．「#NeverNike」というハッシュタグと NIKE スニーカーを燃やす画像を含んだツイートが多く投稿された．

このような炎上の背景に，マーケティングや情報操作などの目的がある可能性がある．しかし，炎上の具体的なメカニズムは解明されておらず，実際に意図的に情報拡散が引き起こされたかどうかは判明していない．

そのような背景において，様々な研究機関で意図的な情報拡散を分析するための研究が行われている．例えば 2020 年，Pacheco[8] らは，Twitter 上で大規模に拡散された話題において，意図的に協調しているアカウントグループを見つけるために調査すべきフレームワークを提案した．以下に提案された 5 つのフレームワークを示す．

- スクリーンネームの共有：アカウントのユーザ名である screen_name(@˜˜) を複数アカウントで再利用して共有しているグループはあるか
- 画像の共有：同一または類似している画像を大量に投稿しているグループはあるか
- ハッシュタグ：複数のツイートにおいて，同一のハッシュタグを付ける順番が非常に類似しているアカウントのグループはあるか
- 共通リツイート：同一ツイートをリツイートしているグループはあるか
- 同期化：同一の話題に関して，偽の情報が含まれるコンテンツを短い間隔で投稿しているグループはあるか

これらを調査することが，今後，意図的な炎上の検知する研究を前進させることに役立つ可能性がある．

3. ボットの利用事例

本節では，情報を操作するためにボットが用いられた事例を紹介する．

3.1 米国大統領選挙中の世論への影響

2019 年，Rice ら [9] は，Twitter 上でロシアのボット集団が，2016 年に行われた米国大統領選挙に与えた影響を調査した．ここでのボット集団とは，ロシアのインターネットリサーチエージェンシー（IRA）によって操作されているボットとトロールの集団を指す．Rice らはこの調査に，Twitter 社から公開されている IRA のボット集団によるツイートのアーカイブ [10] を用いた．このアーカイブには、IRA による 3,613 件のボットと，そのアカウントのツイート 900 万件が含まれている．

Rice らはこのアーカイブのうち，2016 年の米国大統領選挙中に発生した 770,005 件の英語のツイートを調査した．候補者に対してボット集団が大量にツイートやリツイートを行った前後における支持率の変化を比較した．その結果，ボット集団が投稿したツイートが大規模に拡散された後の支持率が上昇していることが判明した．Rice らは，ボット集団が投稿したツイートが 25,000 件リツイートされることで，支持率が約 1 パーセント増加したと予測した．

また，Rice らの調査によると，ボット集団が投稿した各ツイートを最初にリツイートしたアカウントのうち，91% が IRA によるボット集団と関係のない米国市民のアカウントであった．よって，IRA のボット集団が米国市民のコミュニティに参入した可能性があったと結論付けている．

3.2 ボットによる世論の統制

2016 年，Woolley[11] はボットが政治目的で利用された事例についてまとめた．以下は Woolley がまとめたうちから一部を抜粋したものである．

- 2015 年，Abokhodair ら [12] によると，シリア政府の支持者たちは Twitter 上でシリア内戦から人々の関心を逸らすための話題作りにボットを使用した．支持者達は，シリアやシリア内戦に関するハッシュタグを用いて他の話題を拡散した．
- 2012 年，Downes[13] によると，英国の政治家候補である Lee Jasper は， クロイドンノース補欠選挙において実際の支持率と異なる印象を世論に与えるため，ボットを用いてフォロワー数を増やした．
- 2012 年，Coldewey[14] によると，元米国大統領候補の Mitt Romney はボットを購入し，フォロワー数を増やした可能性があるとした．7 月 21 日から 24 時間で 117,000 人のフォロー数を追加で獲得しており，大部分がボットであった．
- 2011 年，York[15] によると，デモ活動によって批判されているシリア政府の支持者達は，抗議活動から興味を逸らすために，バーレーンの企業 EGHNA[16] が操作するボットを用いた．支持者達は，シリアやデモ活動に関するハッシュタグを用いて他の話題を拡散した．

Woolley は，これらの調査を踏まえ，今後の政治的場面における，ボットの動向を監視するべきだと結論付けた．

3.3 ボット，トロールによるポップカルチャーへの影響

Bay ら [17] によると，映画『Star Wars: The Last Jedi』に対して，ボットやトロールによる批判ツイートがあった．Bay らは 2017 年 12 月 13 日から 2018 年 7 月 20 日までに，この映画の監督が所有するアカウントのユーザ ID「@rianjohnson」に対する 1,273 件のツイートを調査した．そして，ツイートを投稿したアカウントのうち批判ツイートを投稿した 206 件のアカウントを調査した．

その結果，批判ツイートを投稿した 206 件のアカウントのうち，44 件のアカウントがボットやトロールであった．この調査から，ボットやトロールは政治的な話題だけでなく，ポップカルチャーの話題にも用いられていることが明らかとなった．

4. ボット判別技術の現状

本節では，Twitter におけるボット判別技術に関する研究について紹介する．また本論文では，4.4 節の研究に関して再現を行ったので，併せて説明する．

4.1 Online Human-Bot Interactions: Detection, Estimation, and Characterization

2017 年，Varol ら [18] はボット判別に用いる特徴量を以下の 6 カテゴリに分類した．

(1) ユーザベースの特徴
アカウントのメタデータから抽出できる特徴のこと．例えば，アカウントのフォロー数，フォロワー数，投稿したツイート数，プロフィールの画像や自己紹介文などがある．

(2) 関連ユーザの特徴
アカウントがリツイート，メンションをした，もしくはリツイート，メンションをされたいずれかのアカウントから抽出できる情報．例えば，使用言語や総ツイート数などがある．

(3) ネットワークの特徴
リツイート，メンション，ハッシュタグの 3 種類それぞれの関係性によって構築されたネットワークから抽出できる特徴．例えば，リツイートやメンションにおけるネットワークでは，アカウントをノードとし，関わるアカウントの数や，拡散された方向などがある．また，ハッシュタグにおけるネットワークでは，ハッシュタグをノードとし，同一ツイート中に含まれたハッシュタグから得られるそれらの結びつきなどがある．

(4) アカウントの活動時間の特徴
アカウントの活動間隔から抽出できる特徴のこと．ここでの活動とは，ツイート，リツイート，メンションが挙げられる．例えば，あるアカウントがツイートをしてから次のツイートが行われるまでの間隔などがある．

(5) テキストの特徴
ツイートのテキストから抽出できる特徴のこと．例えば，ツイートテキストの長さ，品詞などの言語的特徴や使用言語などがある．

(6) 感情の特徴
ツイートのテキストから抽出できる感情的特徴のこと．例えば，ツイートテキストの感情度合や絵文字の使用率などがある．

また，Varol らはこれらの特徴量を用いて機械学習を行い，ボット判別をした．機械学習とその評価において，そのアカウントがボットか人間かの情報が必要である．そこで Varol らは，ハニーポットアプローチ [19] を用いて収集したアカウントをボットとした．ここで，ハニーポットアプローチとは，一般ユーザはフォローしないと思われるハニーポットアカウントを用意し，このアカウントをフォローしたアカウントをボットとする手法である．

ハニーポットアプローチを用いて得られた 15,000 アカウントのボットとそれらによるツイート 260 万件，手動で人間のアカウントだと検証された 16,000 アカウントとそれらによるツイート 300 万件をデータセットとした．ツイートの収集は Twitter API を用いて行われた．そして Varol らは，機械学習アルゴリズムに Adaboost，ロジスティック回帰，ランダムフォレストの 3 種類を使用し，それぞれの精度を比較した．

その結果，ランダムフォレストを用いた学習によるボット判別が最も精度が高く，AUC 値が 0.95 であった．

また，Varol らは追加で，総ツイート数が 200 件以上，もしくは収集期間の 2015 年 10 月から 3 か月間で 90 件以上投稿したアカウントを Twitter API を用いて収集した．追加で収集した 3,000 件のアカウントに対して手動でボット判別を行い，データセットを作成した．

そして，ハニーポットアプローチによるデータセットで学習した判別モデルを，追加で判別した 3,000 件のデータセットで評価した．その結果，追加で判別したボット集団を一部検出することができなかったため，AUC 値が 0.85 に低下した．

そこで，Varol らは，追加で判別した 3,000 件のデータセットをハニーポットアプローチによるデータセットに加えて学習を行った．その結果，ハニーポットアプローチで検出されたボット集団，手動で判別したボット集団の両方を検出でき，ボット判別モデルの AUC 値が 0.94 に向上した．

4.2 Identifying Correlated Bots in Twitter

2016 年，Chavoshi ら [19] は，アカウント同士の関係性をもとにクラスタリングをし，ボットの判別を行った．具体的には，ユーザがリツイートやシェアをした後，そのツイートを別のユーザがリツイートするまでの間隔が 20 秒以内であるツイートが一時間で 40 件以上あるアカウントを相関性が高いアカウントとした．そして，相関性が高いアカウントが人間である可能性は極めて低いため，ボットと判別している．

また，教師なし学習を用いているため，正確な判別率を算出することはできないが，Chavoshi らがボットと判別したアカウントの 45%が Twitter 社によって停止された．

4.3 Deepbot: A Deep Neural Network based approach for Detecting Twitter Bots

2019 年，Luo ら [20] はアカウントのツイートテキストをもとに，ディープラーニングを用いて，ボット判別を行った．また，人間のアカウントとボットによるツイートをそれぞれ，学習に 144,000 件，評価に 62,000 件用いた．

このデータセットは，研究機関である PAN[21] が公開している，英語とスペイン語のアカウントで構成されたデータセット [22] である．

そして Luo らは，ツイートテキストの特徴抽出に GloVE を，ディープラーニングにおける損失関数に Binary Cross

Entropy，オプティマイザに Adam を用い，バッチサイズは 512 として学習を行った．その結果，ボット判別モデルの accuracy 値が 79.64，ROC 値が 87.04 であった．

4.4 Language-Agnostic Twitter-Bot Detection

本論文では，4.4.1 節のボット判別手法をもとに我々が再現を行ったので，4.4.2 節で説明する．

4.4.1 Knauth[23] による研究

2019 年，Knauth[23] はアカウントのメタデータから，言語に依存しない特徴量のみを利用し，ボット判別を行った．学習と評価において，ソーシャルメディアや Web における情報操作の調査を行っている MIB[24] が公開しているデータセット [25] を用いた．このデータセットは以下のアカウントタイプから構築されている 8,385 アカウントのデータセットである．

- 人間のアカウント：3,473 アカウント
- 政治家候補者のツイートをリツイートしたアカウント：991 アカウント
- 有料アプリのスパム：3,457 アカウント
- Amazon.com のスパム：464 アカウント

これらのアカウントタイプそれぞれからメタデータから抽出した，言語に依存しない特徴量を使用した．そして Knauth は，機械学習のアルゴリズムは Adaboost を用いて学習を行った．その結果，ボット判別モデルの accuracy 値が 0.988，AUC 値が 0.995 であった．

また，Knauth は特徴量の組み合わせを複数用いて機械学習を行うことで，ボット判別に重要な特徴量を算出した．その結果，「レーベンシュタイン距離を用いて算出したユーザ名とユーザ ID の類似度」，「ユーザ名の長さ」，「デフォルトのプロフィールを使っているか」などの特徴量が重要であった．

4.4.2 本論文における再現

本論文で、我々は Knauth と同様のボット判別手法を異なるデータセットに対して行った．ボット判別モデルの学習と評価に，Kaggle[26] で提供されているデータセット [27] を用いた．このデータセットはボット判別を目的に公開されている 2,797 アカウントのデータセットである．データセットのうち 2,553 アカウントが英語を使用しており，日本語を使用しているのは 11 アカウントであった．データセット中のアカウントの大多数が英語を用いているため，言語に依存しない特徴量として，表 1 に示す特徴量を用いた．

Knauth の研究と同様，我々が行った再現において，機械学習アルゴリズムに Adaboost を用い，データセットのうち 2,238 件を学習に，残り 559 件を評価に用いた．結果としてボット判別モデルの accuracy 値が 0.88 であった．

4.5 Contrast Pattern-Based Classification for Bot Detection on Twitter

2019 年，González ら [28] は，アカウント情報とそのアカウントのツイートテキストの感情分析を利用して，ボット判別を行った．González らは学習と評価において，4.4 節と同様の MIB[24] が公開しているデータセット [25] のうち，英語を使用しているツイートと，González らが Twitter API を用いて収集したメキシコの政治家 4 名のツイートを用いた．MIB が公開しているデータセットには英語で発言している政治家候補者アカウントがあるため，スペイン語で発言しているメキシコの政治家 4 名の選挙運動中のツイートを追加で収集した．

表 1　再現に用いた特徴量

特徴量名	特徴量の説明
location	場所情報を記載しているか
has_description	自己紹介文を載せているか
has_url	URL を載せているか
followers_count	フォロワー数
friends_count	フォロー数
listed_count	リスト数
favorites_count	「いいね」した数
verified	認証マークがついているか
status	最新ツイートがあるか
has_default_profile	デフォルトのプロフィールか
has_default_profile_image	デフォルトのプロフィール画像か
has_extended_profile	不明（現在は使われていない）

最終的に用いたデータセットは，人間のアカウントによるツイート 31,654 件とボットによるツイート 19,804 件であった．また，González らは機械学習に，ロジスティック回帰，Adaboost，ランダムフォレストなど 11 種類のアルゴリズムを使用し，それぞれの精度を比較した．

それぞれのアルゴリズムによる実験の結果，ランダムフォレストを用いた学習によるボット判別が最も精度が高く，AUC 値が 0.999 であった．

4.6 ソーシャルボットの検出：言語非依存性の特徴量とボット集団の定量化

2018 年，杉森ら [29] は，Twitter における日本語アカウント 17,902 件と英語アカウント 24,386 件を対象に，アカウントのメタデータから抽出した言語に依存しない 326 個の特徴量と機械学習を用いて，ボット判別を行った．データセットは Twitter API を用いて，2017/12/7〜2018/1/4 の期間に収集されたものである．また，収集された全てのアカウントに対してボットか人間かの情報を付与するために，ボット判別ツールの Botometer を利用した．なお，Botometer については 5.1.1 節で説明する．これにより，ボットと判別されたアカウントをボット，されなかったアカウントを人間のアカウントとした．以下にその結果の内訳を示す．

- 英語アカウント
  - ボット：12,193 アカウント
  - 人間：73,616 アカウント
- 日本語アカウント

– ボット：8,951 アカウント
– 人間：63,866 アカウント

また，ボット判別モデルを構築するために，人間のアカウントの数をボットの数に合わせるよう，無作為に選択した．これにより，ボットおよび人間のアカウントは，英語でそれぞれ 12,193 アカウント，日本語でそれぞれ 8,951 アカウントとなった．杉森らはこのデータセットに対し，学習に重要な特徴量を算出するために，日本語のアカウントのみを用いたボット判別モデル，英語のアカウントのみを用いたボット判別モデルの 2 つを構築した．その結果，双方の言語それぞれのボット判別モデルに共通して重要な 20 種類の特徴量が分かった．

そして杉森らは，算出した共通して重要な 20 種類の特徴量を用いて，日本語と英語両方のアカウントを用いたボット判別モデルを構築した．その結果，日本語と英語の両方のアカウントを用いたボット判別モデルの AUC 値が約 0.95 であった．

5. ボット判別の技術に関するツール

本節では，Web 上で提供されている「ボット判別を行うツール」と「ボット判別を応用したツール」について紹介した後，「ボット判別技術を応用した研究」について併せて紹介する．

5.1 ボット判別を行うツール

5.1.1 Botometer

Botometer[30] はインディアナ大学の Observatory on Social Media (OSoMe) が開発しているツールであり，Web 版と API 版の 2 種類がある．Web 版ではボットかどうかを 0 から 5 のスコア，API 版では 0 から 1 のスコアで判定する．このツールでは，0 は人間のアカウント，Web 版での 5 または API 版での 1 はボットに近いとしている．

Botometer 利用者は，判定したいアカウントのユーザ名である screen_name(@˜˜) を指定する．すると，Botometer はメタデータなどから特徴量の抽出を行う．そのあと，抽出した特徴量と，事前に数万件のデータで学習を行ったボット判別モデルを用いて，スコアを算出する．

また，このツールでは，判定したいアカウントのボット判別の結果だけでなく，判定したいアカウントがどのような種類のボットに近いかを同様に判定し，その結果を表示する．具体的には，判定したいアカウントが，Varol らが分類した [31] 以下の 6 種類のボットそれぞれに対し，どれくらい類似的な特徴を持つかを 0 から 5 のスコアで判定する．

- Astroturf：政治的発言を行うボット
- Fake follower：フォロワーを増やすために購入されたボット
- Financial：ハッシュタグ (#) に似たタグで，企業情報を表示するキャッシュタグ ($) を投稿するボット
- Self declared：Twitter で活動しているボットを閲覧できる botwiki.org[32] というサイトのボット
- Spammer：不特定多数のユーザに悪質なツイートやメッセージを送るなどの迷惑行為を行うスパムボット
- Other：ユーザから報告されたボット

API 版を用いて判別を行う場合は，Botometer が事前に学習を行った 2 種類のボット判別モデルから選択できる．英語に適応したボット判別モデルと，言語に依存しないボット判別モデルである．これにより言語に依存しない特徴量だけを用いたスコアを算出できる．

5.1.2 TweetBotOrNot

TweetBotOrNot[33] は，ミズーリ大学の Michael W. Kearney が開発しているツールであり，ボットかどうかを 0 から 1 の数値で判定する．このツールでは，0 が人間のアカウント，1 がボットに近いとしている．

TweetBotOrNot 利用者は，判定したいアカウントのユーザ名である screen_name(@˜˜) を指定する．すると，TweetBotOrNot はメタデータなどから特徴量の抽出を行う．その後，抽出した特徴量と学習を行ったボット判別モデルを用いて，スコアを算出する．

5.1.3 Debot

Debot[34] はニューメキシコ大学の Nikan Chavoshi らのチームが開発しているツールであり，ユーザ間の相関関係によってボットを検出する．

Debot 利用者は，ブラウザまたは API を用いて以下の情報を閲覧できる．

- 日付を入力し，その日に Debot によってボットと判別されたアカウント
- 特定の話題を入力し，その話題に関連する，Debot によってボットと判定されたアカウント
- アカウント名を入力し，そのアカウントが今まで何回 Debot によってボットと判別されたか

5.2 ボット判別を応用したツール

5.2.1 BotSlayer

BotSlayer[35][36] は 5.1.1 節と同様のインディアナ大学の OSoMe が開発しているツールである．

Twitter において，ある話題を拡散したアカウントが情報操作を行っているかどうかをスコアで判定するツールである．

ツール上に入力した話題に対して，Twitter 上で関連するアカウントが複数表示される．また，それぞれが情報操作を行ったかどうかを示す指標である BS Score を確認できる．さらに，表示されたアカウントがボットかどうか表示され，その話題にボットがどれほど関連しているか確認することができる．

5.2.2 Hoaxy

Hoaxy[37] はインディアナ大学の IUNI と CNetS が共同

で開発しているツールであり，Twitter 上の記事やトレンドに関するキーワードなどがどのように拡散されているかを視覚化することができる．このツールには以下の 2 つの機能がある．

(1) 記事検索機能
このツールに入力したキーワードをもとに，Twitter に投稿された関連記事を検索する．その検索結果から，記事が拡散された様子を可視化することができる．また，拡散された際に関連した記事，ツイート，アカウントを参照することができる．

(2) Twitter 検索機能
このツールに入力したニュース記事などのリンクをもとに，拡散の様子を追跡する．Twitter 検索機能では関連のあるツイート，アカウントを参照できる．

これらの機能によって得られた検索結果から，関連する上位 1,000 アカウントが可視化される．また，Botometer を用いた 1,000 アカウントそれぞれのボット判定結果が併せて表示される．

5.3 ボット判別技術を応用した研究

5.3.1 Towards a language independent Twitter bot detector

2019 年，Jonas ら [38] は，ツイートテキストが機械を用いて作成されたのか，人間が作成したのかを判別する研究を行った．データセットには，Twitter API で収集された，スウェーデン語，フィンランド語，英語によるツイートを各 5,000 件用いた．また学習に，テキスト中の URL の数や，アプリの種類やボットかどうかを示すデバイスタイプなどの言語に依存しない特徴量を用いた．以下に特徴量に用いた 6 種類のデバイスタイプを示す．

(1) モバイル：iPhone や Android からの投稿
(2) ウェブ：Mac や Web クライアントからの投稿
(3) アプリ：Instagram などのアプリからの投稿
(4) SMM：ツイートを自動投稿する Hootsuite，TweetDeck などのアプリからの投稿
(5) ボット：ボットによる投稿
(6) その他：上記のどれにも当てはまらない投稿

Jonas らは単一言語のみを学習した判別モデルを構築した．それぞれの学習に，その言語のツイートを 4,000 件，評価に 1,000 件を用いた．

結果として，英語のみを学習した判別モデルが最も評価が高く，accuracy 値が 0.992，Precision 値が 0.991，Recall 値が 0.973 であった．

しかし，単一言語のみを学習した判別モデルを用いて，異なる言語のツイートを判別すると accuracy 値が低下した．そこで，スウェーデン語とフィンランド語を用いて学習を行い，異なる言語である英語によるツイートで評価した．すると，Precision 値が 0.998，Recall 値が 0.920 であり，複数言語を用いて学習を行うことで，異なる言語に適応しやすくなると分かった．

6. 国内における現状と課題

本論文での調査により，日本語のアカウントを対象にしたボット判別の研究が少ないことがわかった．これらの根拠の一つとして，ボット判別に用いることができる日本語のアカウントのみで構成されたデータセットがほとんどないことが挙げられる．また，これらが主な原因で日本語のツイートテキストを特徴量として利用できるボット判別ツールがほとんど開発されていない，

ボット判別には，ツイートテキストに自然言語処理などを用いて得られる言語的特徴を，特徴量の一つとして扱う手法がある．しかし，言語的特徴を特徴量の一つとして用い，国外のアカウントで構成されたデータセットを学習すると，国外のツイートテキストの言語的特徴に対応したボット判別モデルが作成されてしまう．

また，日本語と国外の言語との間に，文法や文章表現の違いがある．そのため，作成したボット判別モデルを日本語のアカウントに利用した場合に有用でない可能性がある．さらに，日本語のアカウントのみで構成されたボットのデータセットがないので，日本語のアカウントのボット判別を正しく評価することができない．そのため，このモデルを日本語のアカウントに対するボット判別に用いた場合，正確にボット判別できるか評価を行うことができなく，有用であるか不明である．

また，国外のボット判別ツールは日本語のツイートテキストの言語的特徴に対応していないものが多く，国外のボット判別ツールを日本語のアカウントに対して用いた結果が有用であるかどうかわからない．

よって今後，国外のアカウントを用いて作成したモデルと国外のボット判別ツールを，日本語のアカウントのボット判別に利用した時の有用性を正しく評価するために，日本語のアカウントのみで構成されたボットのデータセットを構築する必要がある．これらの有用性が低かった場合，新たに日本語のツイートテキストの言語的特徴を利用したボット判別モデル及びボット判別ツールを作成するべきである．

7. 研究倫理

本論文では，調査した内容を悪用されないために，ボット判別技術やその他関連技術の具体的手順は省略して記述した．

8. まとめ

本論文では，ボットを一般的でない行為に用いた事例や，国内外におけるボット判別に関する現状について調査し，国内でボットを判別する際の課題について考察した．

その結果，Twitter におけるボット判別の手法が世界中

の研究機関によって多数提案されていることが分かった．また，ボット判別を行うツールなどが Web 上で公開されており，研究機関だけでなくソーシャルメディア利用者がボット判別をできるようになった．しかし，日本語のアカウントを対象としたボット判別の研究や，日本語のツイートテキストを利用したボット判別ツールの開発はあまり行われていないので，今後，研究及び開発を行う必要があると思われる．

参考文献

[1] 自動化ルール. https://help.twitter.com/ja/rules-and-policies/twitter-automation.
[2] The Twitter Rules. https://help.twitter.com/en/rules-and-policies/twitter-rules.
[3] Kai-Cheng Yang, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. Arming the public with AI to counter social bots. *CoRR*, Vol. abs/1901.00912, , 2019.
[4] Twitter API Documentation — Docs — Twitter Developer. https://developer.twitter.com/en/docs.
[5] Alexandre Bovet and Hernán A. Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature Communications*, Vol. 10, No. 1, p. 7, Jan 2019.
[6] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature Communications*, Vol. 9, No. 1, p. 4787, Nov 2018.
[7] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, Vol. 359, No. 6380, pp. 1146–1151, 2018.
[8] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. Uncovering coordinated networks on social media, 2020.
[9] 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.
[10] Information Operations - Twitter Transparency Center. https://transparency.twitter.com/en/reports/information-operations.html.
[11] Samuel C. Woolley. Automating power: Social bot interference in global politics. *First Monday*, Vol. 21, No. 4, Mar. 2016.
[12] Norah Abokhodair, Daisy Yoo, and David W. McDonald. Dissecting a social botnet. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work Social Computing - CSCW ’15, 2015.
[13] Jasper admits to using Twitter bots to drive election bid — Inside Croydon. https://insidecroydon.com/2012/11/26/jasper-admits-to-using-twitter-bots-to-drive-election-bid/.
[14] Romney Twitter account gets upsurge in fake followers, but from where? https://www.nbcnews.com/tech/tech-news/romney-twitter-account-gets-upsurge-fake-followers-where-flna928605.
[15] Syria’s Twitter spambots — Jillian C York — Opinion — The Guardian. https://www.theguardian.com/commentisfree/2011/apr/21/syria-twitter-spambots-pro-revolution.
[16] EGHNA — Drupal.org. https://www.drupal.org/eghna.
[17] Morten Bay. Weaponizing the haters: The last jedi and the strategic politicization of pop culture through social media manipulation., 10 2018.
[18] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. *CoRR*, Vol. abs/1703.03107, , 2017.
[19] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. Identifying correlated bots in twitter. 11 2016.
[20] Linhao Luo, Xiaofeng Zhang, Xiaofei Yang, and Weihuang Yang. Deepbot: A deep neural network based approach for detecting twitter bots. *IOP Conference Series: Materials Science and Engineering*, Vol. 719, p. 012063, jan 2020.
[21] PAN. https://pan.webis.de/.
[22] PAN19 Author Profiling: Bots and Gender Profiling — Zenodo. https://zenodo.org/record/3692340.
[23] Jürgen Knauth. Language-agnostic Twitter-bot detection. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pp. 550–558, Varna, Bulgaria, September 2019. INCOMA Ltd.
[24] My Information Bubble project. http://mib.projects.iit.cnr.it/index.html.
[25] MIB Datasets. http://mib.projects.iit.cnr.it/dataset.html.
[26] Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com/.
[27] detecting twitter bot data — Kaggle. https://www.kaggle.com/charvijain27/detecting-twitter-bot-data.
[28] O. Loyola-González, R. Monroy, J. Rodríguez, A. López-Cuevas, and J. I. Mata-Sánchez. Contrast pattern-based classification for bot detection on twitter. *IEEE Access*, Vol. 7, pp. 45800–45817, 2019.
[29] 杉森真樹, 笹原和俊, 時田恵一郎. 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 1P205–1P205, 2018.
[30] Botometer® by OSoMe. https://botometer.osome.iu.edu/.
[31] Kai-Cheng Yang, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, Vol. 1, No. 1, pp. 48–61, 2019.
[32] Welcome to Botwiki — Botwiki. https://botwiki.org/.
[33] Michael Kearney. tweetbotornot: Detecting twitter bots (web app: https://mikewk.shinyapps.io/botornot/), 06 2018.
[34] DeBot. https://www.cs.unm.edu/ chavoshi/debot/index.html.
[35] OSoMe: BotSlayer. https://osome.iuni.iu.edu/tools/botslayer/.
[36] Pik-Mai Hui, Kai-Cheng Yang, Christopher Torres-Lugo, Zachary Monroe, Marc McCarty, Benjamin D. Serrette, Valentin Pentchev, and Filippo Menczer. Botslayer: real-time detection of bot amplification on twitter. *Journal of Open Source Software*, Vol. 4, No. 42, p. 1706, 2019.
[37] Hoaxy® by OSoMe. https://hoaxy.iuni.iu.edu/.
[38] Jonas Lundberg, Jonas Nordqvist, and Mikko Laitinen. Towards a language independent twitter bot detector. In *Proceedings of 4th Conference of The Association Digital Humanities in the Nordic Countries : Copenhagen, March 6-8 2019*, Vol. 2364 of *CEUR Workshop Proceedings*, pp. 308–319. University of Copenhagen, 2019.