

機械学習を用いたソーシャルボットの検出と分析： 英語と日本語の比較

杉森真樹[†] 笹原和俊^{‡,*} 時田恵一郎[‡]

[†] 名古屋大学情報文化学部 [‡] 名古屋大学大学院情報学研究科 ^{*} JST さきがけ

概要

ソーシャルメディアの普及に伴い、「ソーシャルボット」と呼ばれる自動発言プログラムが増加し、その中には個人や集団に様々な悪影響を与えるものもある。そのため、ソーシャルボットの性質や影響、検出手法について様々な研究が行われている。ソーシャルメディアのユーザーは世界中に存在しており、多様な言語が利用されているため、ソーシャルボットの検出を行う汎用的な方法を開発するためには、言語に依存しない特徴や振る舞いに注目する必要がある。しかし、現在のところソーシャルボットの検出について、言語の違いに注目した研究はほとんど行われていない。本研究では Twitter の投稿データを対象に機械学習を用いて分類モデルを構築し、英語と日本語の Twitter ボットのアカウントについて比較、分析を行った上で、言語に依存しないユニバーサルな検出手法について検討を行う。データ分析の結果、Twitter ボットの言語に依存しない特徴量や振る舞いを特定し、ユニバーサルな分類モデルの構築につながる知見が得られた。

1 はじめに

近年、Twitter や Facebook といったソーシャルメディアが広く普及し、単なるコミュニケーションのためのツールにとどまらず、研究やビジネス、政治活動などに利用され、社会において様々な役割を果たすようになっている。このような背景のもと、コンピュータープログラムによって自動化された「ソーシャルボット」と呼ばれる存在によって引き起こされる様々な悪影響が大きな問題となっている [14]。ソーシャルボットが引き起こす問題には次のようなものがある。

- スパムの拡散

従来のスパムと同様に、一般のユーザーに対して、有害な Web ページへのリンクをクリックさせ、特定の Web ページにアクセスさせたり、マルウェアに感染させることを目的とする。

- 人気の操作

ソーシャルボットによる大量の投稿によってソーシャルメディアのトレンド (話題となっているキーワード) に影響を与え、一般のユーザーの関心を操作できる可能性がある。また、特定のアクティビティを集中的に行うことで、特定のユーザーや商品がソーシャルメディア上で人気、影響力があるように見せかけることができ、マーケティングにも悪影響を及ぼす可能性がある。

- 意見の操作

ソーシャルボットによるフェイクニュースなどの拡散により、特定の対象に関する人々の印象を操作しようとする可能性がある。具体的な例としては、選挙運動が挙げられ、2016 年のアメリカ合衆国大統領

領選挙では、ソーシャルボットによってフェイクニュースなどの発信が活発に行われており、選挙に影響を及ぼしていた可能性がある [3]。

- 市場の操作

ソーシャルボットによるフェイクニュースの拡散により、現実の社会の市場の動向が操作される可能性がある。具体的な例としては、株式市場が挙げられる。2013 年には、テロに関するフェイクニュースによって、一時的にダウ工業株 30 種平均が数分間で約 140 ドル急落している。また、ほとんど資産価値のない会社の株価を高騰させて利益を得ようとするソーシャルボットの動きが観察されている [14]。

このような様々な悪影響をもたらすソーシャルボットを検出し、ユーザーに注意を喚起し、ソーシャルメディアから排除するためには、ボットの特性や検出手法、情報の拡散過程などについて様々な研究を行っていく必要がある。

2 関連研究

ソーシャルメディアの普及、発展に伴い、ボットがもたらす悪影響はソーシャルメディア上だけでなく、実社会にまで及んでおり、様々な先行研究が行われている。この章では、ソーシャルボット検出に関連する先行研究について解説する。

(1) 単体のアカウントを対象とした観察的手法

Varol らは、Twitter の英語アカウントを対象として、1150 個の特徴量と機械学習を用いて、AUC 値が 0.94 という高い精度の分類モデルを構築した。彼らは、この分類モデルを応用して、単体のアカウントがボットであるかどうかのスコアを [0 ~ 1] の範囲で算出する“Botometer”というシステムを開発し、公開している [13, 17, 19]。Chu らは、機械学習を用いて、人間とボットの 2 クラスに加えて、半自動化されているアカウントを「サイボーグ」と定義して、人間、ボット、サイボーグの 3 クラスの分類に取り組んでいる [11, 12]。

(2) 複数のアカウント（ボットの集団）を対象とした観察的手法

Abokhodair らは、シリア内戦に関する情報を拡散しているボットの集団を対象として、ボットの集団の構造や成長過程、ソーシャルメディア上でのシリア内戦に関する議論に与えた影響などを調査した [2]。Chavoshi らは、アカウント同士の過度に同期的な動作に注目して、投稿時間の一致などに基づく学習用のデータを必要としないボット検出手法を提案するとともに、“DeBot”というボット検出システムを構築している [7-10]。

(3) 単体のアカウントを対象とした実験的手法

Freitas らは、異なる特徴と戦略を持つ合計 120 個のボットアカウントを作成し、これらのボットが Twitter 上のコミュニティに侵入する過程や結果について調査を行なった。その結果、単純なメカニズムを採用しているボットであってもコミュニティに侵入することができ、必ずしも複雑なメカニズムを必要としないことを示した [15]。

(4) 複数のアカウント（ボットの集団）を対象とした実験的手法

Boshmaf らは、ボットの集団による大規模な活動を想定して、102 個のボットアカウントと 1 個のマスターアカウントからなるボットの集団のプロトタイプを構築し Facebook 上で稼働させることで、ボットの集団がもたらす影響とソーシャルメディアの脆弱性を調査した [4, 5]。

ソーシャルボットの性質、検出については様々な先行研究が行われている。しかし、ソーシャルメディアには世界中に多くのユーザーが存在しており、多様な言語が利用されているにもかかわらず、ボット検出手法に関する研究において言語の違いに注目した研究はほとんど行われていない。このような背景のもと、本研究では、Varol らによる Botometer に関する研究 [19] を主に参考として、機械学習を利用したボットの検出に関する研究を行う。特に、Twitter を対象として英語と日本語それぞれの言語において、ボット検出に関連する重要な特徴量について比較、分析を行う。さらに、言語に依存しないユニバーサルなボット検出手法について検討を行う。

3 データ

本研究では、単体の Twitter アカウントが持つ様々な特徴量と教師ありの機械学習を用いて、分類モデルを構築する。分類モデルを訓練するために、各アカウントから連続した一定量のツイートおよびそれらのメタデータから算出した特徴ベクトルと人間かボットかを事前に判別した正解ラベルが、学習用のデータとして必要となる。先述した Botometer のスコアによる分類が正しいと仮定し、正解ラベルの作成には Botometer を利用した。また、日本語アカウントの分類には、言語的な特徴を無視したスコアを利用した。

データを収集するにあたり、英語または日本語で投稿を行なった Twitter アカウントの最新 200 件の投稿を収集し、特徴ベクトルと正解ラベルを算出した。最新の投稿 200 件を対象とするのは、API 制限のためであり、先行研究においても最新の 200 件以降の投稿は対象とされていない。

2017/12/7 ~ 2018/1/4 の期間、英語についてはボット 12193 件と人間 73616 件、日本語についてはボット 8951 件と人間 63866 件のアカウントのデータを収集した。

英語については、ボット 12193 件と無作為抽出した人間 12193 件の合計 24386 件を用いて、英語アカウントに対する分類モデルを構築した。

日本語については、ボット 8951 件と無作為抽出した人間 8951 件の合計 17902 件を用いて、日本語アカウントに対する分類モデルを構築した。

4 結果

4.1 分類モデル

本研究では先行研究 [11, 12, 19] で利用されている特徴量と独自の特徴量を利用する。具体的には、次頁のテーブルに示した「コンテンツ」、「メタデータ」、「タイミング」、「ネットワーク」の 4 つの特徴クラス、326 個の特徴量を利用する。分類モデルの構築には “scikit-learn” [1, 18] を利用し、機械学習のアルゴリズムはランダムフォレストを採用する [6]。これは、先行研究の中でランダムフォレストを利用した場合の精度が最も高いという結果が得られていることに加えて [16, 19]、特徴量の重要度を導出することができるためである。英語と日本語それぞれの言語について、アカウントが人間であるかボットであるかを判別する分類モデルを特徴クラスごとに構築する。この分類モデルの精度を確認するために、一般的な指標である AUC を測定する。

- 1 コンテンツクラス
 - 1-1 品詞の出現頻度*1*2
 - 1-2 品詞の割合*1*2
 - 1-3 単語の総数*1
 - 1-4 単語の種類の数*1
 - 1-5 単語のエントロピー*1
- 2 メタデータクラス
 - 2-1 スクリーンネーム、登録名の長さ
 - 2-2 スクリーンネームに含まれる数字の個数
 - 2-3 タイムオフセット
 - 2-4 アカウントの日齢
 - 2-5 自己紹介文の長さ
 - 2-6 自己紹介文中に URL を含むか
 - 2-7 フォロー数
 - 2-8 フォロワー数
 - 2-9 お気に入り数
 - 2-10 プロファイルがデフォルトのままか
 - 2-11 プロファイルイメージがデフォルトのままか
 - 2-12 投稿数 (総数、1 時間あたりの数)
 - 2-13 通常のツイート数 (最新 200 件における総数、1 時間あたりの数)
 - 2-14 リツイート数 (最新 200 件における総数、1 時間あたりの数)
 - 2-15 メンション数 (最新 200 件における総数、1 時間あたりの数)
 - 2-16 リプライ数 (最新 200 件における総数、1 時間あたりの数)
 - 2-17 リツイートされた数 (最新 200 件における総数、1 時間あたりの数)
 - 2-18 リツイートされた数 (最新 200 件)*1
 - 2-19 お気に入りされた数 (最新 200 件)*1
 - 2-20 メンションした数 (最新 200 件)*1
 - 2-21 URL 数 (最新 200 件)*1
 - 2-22 ハッシュタグ数 (最新 200 件)*1
 - 2-23 メディア数 (画像数 + 動画数) (最新 200 件)*1
 - 2-24 画像数 (最新 200 件)*1
 - 2-25 動画数 (最新 200 件)*1
 - 2-26 投稿元に関する指標 (最新 200 件)*1*3
 - 2-27 フォロー比
 - 2-28 フォロワー比
 - 2-29 リストされている数
- 3 タイミングクラス
 - 3-1 通常のツイートの投稿時間間隔 (分)*1
 - 3-2 リツイートの投稿時間間隔 (分)*1
 - 3-3 リプライの投稿時間間隔 (分)*1
 - 3-4 メンションの投稿時間間隔 (分)*1
 - 3-5 投稿の時間間隔 (分)*1
- 4 ネットワーククラス (ハッシュタグ共起ネットワーク)
 - 4-1 ノード数
 - 4-2 エッジ数
 - 4-3 ネットワークの密度
 - 4-4 ネットワークの強度*1
 - 4-5 クラスタリング係数*1

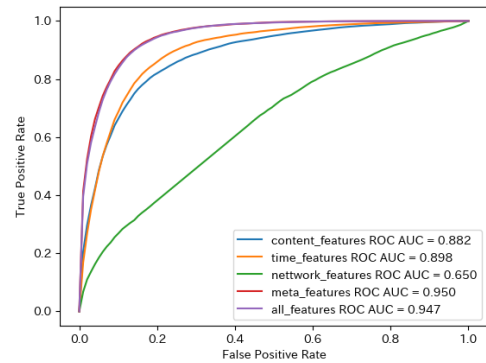


図 1-a : 英語アカウントの分類モデルの精度

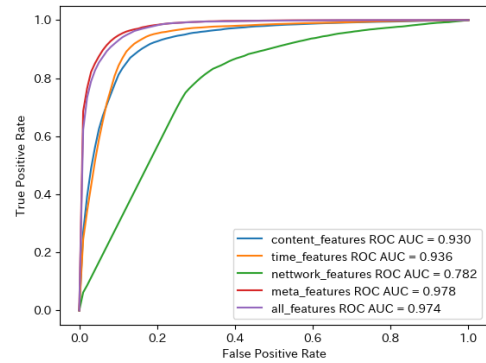


図 1-b : 日本語アカウントの分類モデルの精度

*1 [最小値, 最大値, 中央値, 平均, 標準偏差, 歪度, 尖度, エントロピー] を算出し、特徴量とする。

*2 品詞は、英語 [名詞, 動詞, 形容詞, 副詞, 助動詞, 限定詞, 間投詞, Wh-, 代名詞]、日本語 [名詞, 動詞, 形容詞, 副詞, 助動詞, 助詞, 接続詞, 感動詞, 接頭詞] を考慮する。

*3 投稿元に関する指標については、各投稿において、自動化が行われていないと思われる投稿元の値を 1、自動化が行われていると思われる投稿元の値を -1、それ以外の値を 0 とする。

図 1-a と b を見ると、英語と日本語の双方のアカウントにおいて、メタデータクラスが最も高い精度を示していることが分かる。コンテンツクラスとタイミングクラスがこれに続いており、高い精度を示している。一方で、ネットワーククラスについては、ある程度の精度を示しているが、他のクラスと比較すると精度が低いことが分かる。

異なる言語同士ではコンテンツクラスの特徴量に大きな違いが生じることが予想され、ユニバーサルな分類モデルで利用することはほとんど期待できない。しかし、コンテンツクラスの特徴量よりもメタデータクラスとタイミングクラスの特徴量が高い精度を示していることから、言語に依存しないユニバーサルな分類モデルにおいて、コンテンツクラスの特徴量を考慮せずとも、高い精度を示すことは可能であると考えられる。

4.2 重要な特徴量

主に 2 つの手法によって、英語と日本語それぞれの分類モデルにおいて重要な特徴量の選択を行う。

(1) 分類モデルにおける特徴量の重要度の測定

ランダムフォレストを用いた分類モデルにおける特徴量の重要度を、決定木ごとにその特徴量によって減少する基準値（ジニ不純度）の平均として算出する。

(2) 人間とボットの特徴量の分布が異なるかどうかの判定

人間のアカウントとボットアカウントの特徴量の分布に対して、コルモゴロフ・スミルノフ検定（KS 検定）を行い、人間のアカウントとボットアカウントで特徴量の分布が異なっていることを確かめる。

重要な特徴量を特定する手順としては、分類モデルにおいて重要度が高い特徴量を 50 個選択し、その中で人間のアカウントとボットアカウントで分布が異なっていることが確認できたものを重要な特徴量とする。

英語のアカウントにおいて重要な特徴量	日本語のアカウントにおいて重要な特徴量
<ul style="list-style-type: none">・ コンテンツクラス<ul style="list-style-type: none">– 名詞の頻度の最小値– 名詞の割合の最小値– 単語の総数の標準偏差、エントロピー– 単語の種類数の最小値、標準偏差、エントロピー– 単語のエントロピーの標準偏差・ メタデータクラス<ul style="list-style-type: none">– フォロー数– フォロワー数– お気に入り数– リツイートの総数、1 時間あたりの数– リプライの総数、1 時間あたりの数– 被リツイート数の総数、1 時間あたりの数– 被リツイート数の最大値、平均、標準偏差、エントロピー– 被お気に入り数の平均、標準偏差、エントロピー– メンション数の標準偏差– URL 数の中央値、平均– ハッシュタグ数の平均、歪度、エントロピー– リスト数・ タイミングクラス<ul style="list-style-type: none">– 通常のツイートの投稿間隔の最大値、標準偏差– リツイートの投稿間隔の最大値、中央値、平均、標準偏差、エントロピー– リプライの投稿間隔の最大値、中央値、平均、標準偏差、歪度、尖度、エントロピー– 全投稿の投稿間隔のエントロピー	<ul style="list-style-type: none">・ コンテンツクラス<ul style="list-style-type: none">– 助動詞の割合の標準偏差– 感動詞の割合の平均、標準偏差– 単語のエントロピーの標準偏差・ メタデータクラス<ul style="list-style-type: none">– フォロー数– フォロワー数– お気に入り数– リツイートの総数、1 時間あたりの数– リプライの総数、1 時間あたりの数– 被リツイート数の総数、1 時間あたりの数– 被リツイート数の最大値、平均、標準偏差– 被お気に入り数の平均– メンション数の最小値、平均、標準偏差、尖度– URL 数の平均– ハッシュタグ数の平均– メディア数の平均– 画像数の最大値、標準偏差、エントロピー– 投稿元に関する指標の最大値、平均・ タイミングクラス<ul style="list-style-type: none">– 通常のツイートの投稿間隔の最大値– リツイートの投稿間隔の最大値、中央値、平均、標準偏差、エントロピー– リプライの投稿間隔の最大値、中央値、平均、標準偏差、歪度、尖度、エントロピー– 全投稿の投稿間隔の最小値、エントロピー

英語と日本語について言語に依存しない分類モデルに利用可能な特徴量を特定するために、双方の言語の分類モデルに共通して重要な特徴量の特定と分析を行う。

双方の言語の分類モデルで共通して重要な特徴量について、KS 検定によって、2 つの言語を統合した場合でも人間とボットの分布が異なっていることが確認できた特徴量を、言語に依存しない重要な特徴量とする。以降では、言語に依存しない重要な特徴量の一部を解説する。

・ リプライ

リプライは他のアカウントとの会話に類似した役割を担うアクティビティである。リプライに関する特徴量は他のアカウントとの交流の頻度を調べる上で重要であると考えられる。

・ リツイート

リツイートは情報を拡散するためのアクティビティであり、リツイートに関する特徴量は情報を拡散させようとする程度の強さや他のアカウントのアクティビティにどれだけ注意を払っているかを調べる上で重要であると考えられる。

リプライ、リツイートに関する特徴量から、人間の多くがこれらのアクティビティを介してソーシャルメディア上で他者と交流していることがわかる。一方で、ボットはこれらのアクティビティを利用してソーシャルメディア上で他者と交流することに過度に積極的か、または過度に消極的であることがわかる。

このような言語によらず人間とボットで分布が大きく異なっている特徴量は、人間とボットの分類を行う上で重要なものであり、本研究ではこれらの言語に依存しない重要な特徴量に注目している。

また、特徴量の組の空間における分布を示した図 4 を見ると、ボットアカウントの分布が異常に集中している領域があることがわかる。この異常に集中した分布はボットの集団の存在を示唆していると考えられる。実際に確認を行ったところ、日本の芸能、ゴシップニュースを拡散させようとしているボットの集団の存在が確認された。

本研究は、ボットの集団ではなく単体のボットを対象とする機械学習を利用した検出手法に注目していたが、特徴量の組の分布に注目することでボットの集団の検出が可能であった。この新たなボットの集団の検出手法の可能性については、今後の研究において検討を行う必要がある。

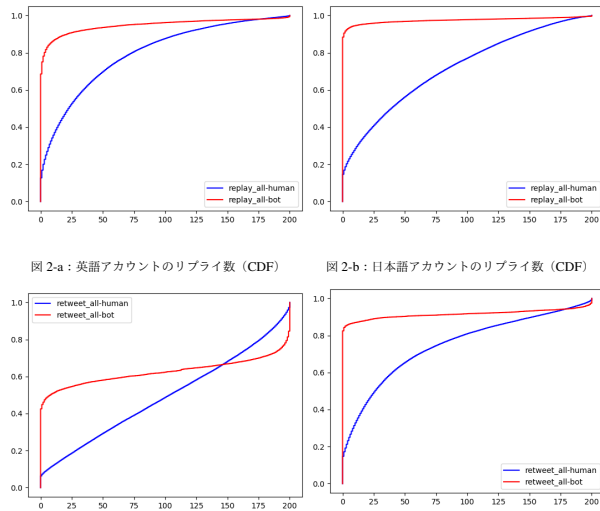


図 2-a : 英語アカウントのリプライ数 (CDF)

図 2-b : 日本語アカウントのリプライ数 (CDF)

図 3-a : 英語アカウントのリツイート数 (CDF)

図 3-b : 日本語アカウントのリツイート数 (CDF)

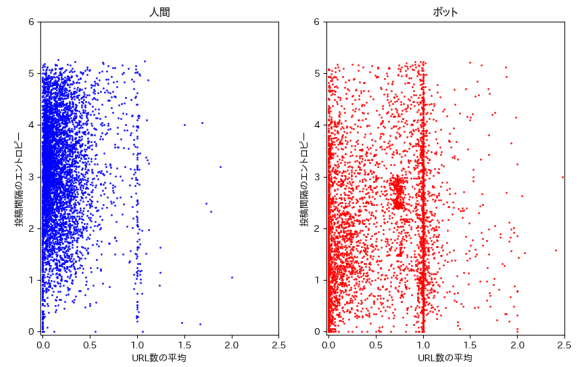


図 4 : URL 数の平均と投稿間隔のエントロピー (散布図)

4.3 言語に依存しない分類モデル

言語に依存しないユニバーサルな分類モデルの構築の可能性について検討を行うために、英語と日本語について、言語に依存しない重要な特徴量を用いて言語に依存しない分類モデルを構築する。また、構築した分類モデルが学習用のデータ以外にも有用であるか検証を行う。

英語と日本語について言語に依存しない分類モデルの構築には、言語に依存しない重要な特徴量の中から、後に行う分類モデルの検証において、精度を大きく低下させていた一部の特徴量や、共通したデータやアクティビティに基づく特徴量の中で重要度が低い一部の特徴量を除いた 20 個の特徴量を用いる。

- コンテンツクラス
 - 単語のエントロピーの標準偏差
- メタデータクラス
 - リプライの総数
 - リツイートの総数
 - 被リツイート数の総数
 - 被リツイート数の最大値、平均、標準偏差
 - 被お気に入り数の平均
 - メンション数の標準偏差
 - URL 数の平均
 - ハッシュタグ数の平均
- タイミングクラス
 - リツイートの時間間隔の最大値、中央値、平均、エントロピー
 - リプライの時間間隔の最大値、中央値、平均、エントロピー
 - 全投稿の時間間隔のエントロピー

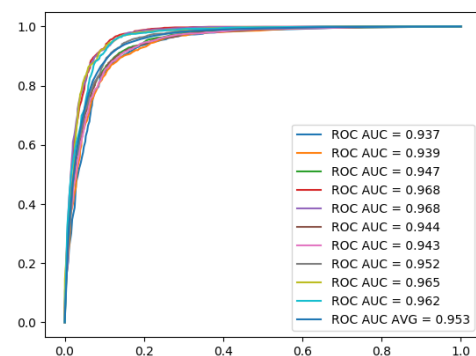


図 5 : 分類モデルの精度 (AUC)

この分類モデルの構築には、本研究で収集した全ての学習用のデータから、英語と日本語それぞれについて無作為抽出した、ボット 8000 件、人間 8000 件の計 32000 件を用いる。精度については、クロスバリデーション

ョンを用いて、平均 AUC を測定する。この結果、20 個の重要な特徴量のみを利用した分類モデルは、平均 AUC の値が約 0.95 と先行研究や全 326 個の特徴量を考慮した言語ごとの分類モデルとほぼ同等の分類精度を示している。

分類モデルが学習用のデータ以外のデータに対しても、同様の精度で分類を行えるかを確認するために、検証用のデータとして、英語と日本語それぞれについて手動で正解ラベルを作成した、ボット 250 件、人間 250 件ずつ、計 1000 件のデータを用意する。なお、手動による分類は正しいと仮定する。

検証の結果、1000 件のうち、約 95% のアカウントを正しく分類できることを確認した。この結果から、英語と日本語について言語に依存しない分類モデルを構築することができたと結論づけることができる。

このように、英語と日本語について言語に依存しない分類モデルを構築することができたことから、他の言語についても同様の研究を行うことで、言語に依存しないユニバーサルな分類モデルの構築は可能であると考えられる。

5 まとめ

ソーシャルメディアには世界中に多くのユーザーが存在し、多様な言語が利用されているにもかかわらず、ソーシャルボットの検出手法について、言語の違いに注目した先行研究が行われていなかった。本研究では、英語と日本語のソーシャルボットを対象として比較、分析を行うとともに、言語に依存しないユニバーサルな分類モデルの構築について検討を行った。結果として、英語と日本語を対象とした分類モデルにおいて、言語に依存しない重要な特徴量を特定し、それらの元になっている、英語と日本語においてボットが有する言語に依存しない特徴的なデータや振る舞いの存在を確認した。また、これらの特徴量を利用して、英語と日本語の双方の言語に対して、精度の高い分類モデルを構築することで、言語に依存しないユニバーサルな分類モデルを構築できる可能性を示した。今後は、英語と日本語以外の言語について同様の研究を行うことで、言語に依存しないユニバーサルな分類モデルの構築に取り組むとともに、ボットの集団を対象とした検出手法の構築について検討する。

謝辞

本研究は JSPS 科研費 JP16K16112 および JST さきがけ JPMJPR16D6 の助成を受けたものです。

参考文献

- [1] Scikit-learn. <http://scikit-learn.org/stable/index.html> <https://github.com/scikit-learn>.
- [2] Norah Abokhodair, Daisy Yoo, and David W. McDonald. Dissecting a social botnet: Growth, content and influence in twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 2016.
- [3] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 u.s. presidential election online discussion. In *First Monday*, 2016.
- [4] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. The socialbot network: When bots socialize for fame and money. In *Proceedings of the 27th Annual Computer Security Applications Conference*, 2011.
- [5] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. Design and analysis of a social botnet. In *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 2013.
- [6] Leo Breiman. Random forests. In *Machine Learning*, 2001.

- [7] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. Debot: Twitter bot detection via warped correlation. In *IEEE 16th International Conference on Data Mining*, 2016.
- [8] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. Identifying correlated bots in twitter. In *International Conference on Social Informatics*, 2016.
- [9] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. On-demand bot detection and archival system. In *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017.
- [10] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. Temporal patterns in bot activities. In *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017.
- [11] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. Who is tweeting on twitter: Human, bot, or cyborg. In *Proceedings of the 26th Annual Computer Security Applications Conference*, 2010.
- [12] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. Detecting automation of twitter accounts: are you a human, bot, or cyborg? In *IEEE Transactions on Dependable and Secure Computing*, 2012.
- [13] Clayton A. Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016.
- [14] Emilio Ferrara, Onur Varol, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. In *Communications of the ACM*, 2016.
- [15] Carlos A. Freitas, Fabricio Benevenuto, Saptarshi Ghosh, and Adriano Veloso. Reverse engineering socialbot infiltration strategies in twitter. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2015.
- [16] Kyumin Lee, Brian David Eoff, and James Caverlee. Seven months with the devils: A long-term study of content polluters on twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [17] OSoMe. Botometer. <https://botometer.iuni.iu.edu/>.
- [18] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine and learning in python. In *Journal of Machine Learning Research Scikit-learn: Machine and Learning in Python*, 2011.
- [19] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, 2017.