ソーシャルネットワーキングサービスにおける『不自然なフォロー行為』の分析

越森　壮一† 　荘司　慶行† 　Martin J. Dürst†

† 青山学院大学 理工学部 情報テクノロジー学科　〒 252-5258 神奈川県 相模原市中央区 淵野辺 5-10-1
E-mail: †koshimori@sw.it.aoyama.ac.jp, ††{shoji,duerst}@it.aoyama.ac.jp

あらまし　本研究では，ある二人のユーザ間のフォロー関係が不自然なものであるかを，ランダムフォレストにおける特徴の重要度から分析する手法を提案する．ソーシャルネットワーキングサービス上で唐突に他人からフォローされた際，そのフォロー行為に正当な理由があるか判断するのは困難である．そこで，ある二人のユーザを入力として与えると，その二者間のフォロー関係の有無を推定するランダムフォレスト分類器を作成する．この分類器に「フォロー関係は存在しないであろう」と推定されたにもかかわらず，実際にはフォロー関係があるユーザのペアについて，分類に大きく寄与した特徴を分析する．実際に Twitter で収集したデータセットを用いて分類と分析を行うことで，不自然なフォロー行為の特徴を明らかにする．

キーワード　Twitter，フォロー関係，ランダムフォレスト，特徴量分析

1　は じ め に

近年，スマートフォンの普及をうけて，老若男女問わず多くの人がソーシャル・ネットワーキング・サービス (SNS) を利用するようになっている．中でも特に利用者の多い Twi tter や Facebook， Instagram などのさまざまな SNS は，それぞれ異なる特色を持っている．SNS の中核の機能としてフォロー機能があり，これこそ『ソーシャルネットワーク』と呼ばれるゆえんである．ただ一言で『フォロー』といっても，サービスによって違いがある．たとえば，Facebook は実名で登録する特徴があり，Instagram は実生活と密着した写真や動画をアップロードする特徴がある．このため，Facebook や Instagram では現実世界で仲の良い友達でネットワークが形成されることが多い．一方，Twitter はフォローという行為を面識の無いアカウントでも気軽にでき，相手の発信している情報を簡単に見ることができる (ただし鍵つきのアカウントや，ブロックされている場合を除く)．Twitter のフォロー機能は知人友人とのコミュニケーション目的であったり，著名人の発信する情報へのアクセス目的であったり，さまざまな用途で使用されている．

SNS のフォロー機能は便利な反面，問題を引き起こす側面もある．例として，Twitter の事例を紹介する．Twitter の世界全体のユーザ数は 3 億 5,000 万人にものぼる．Twitter 社は 2 018 年 7 月上旬に他人になりすまして作成されたアカウント数千万件の大規模な削除を行った [1]．つまり，Twitter においてフォローし，されてきたアカウントのうち，おおよそ 20 パーセントは，有害なアカウントであったことが示されている．同時に，従来の問題報告のシステムでのスパム報告が『不審な内容またはスパムです』というもののみであったのを改善した．現在は，スパムの種類を 5 つの項目から選ぶことで，不正なアカウント対策への強化を行っている．Twitter 社がこのような不正なアカウントへの対策を随時行っていることからもわかるとおり，不正なフォロー関係は SNS の大きな問題の一つである．

ソーシャルネットワーキングサイトにおいて， 見ず知らずのユーザ同士でもフォロー等で気軽につながれることは大きな利点でもあるが，フォローされる側はブロックすべきスパムアカウントなどを見過ごしてしまう可能性がある．例えば，自分のアカウントが見ず知らずのだれかのアカウントから，唐突にフォローされた場合を考える．フォローしてきたアカウントのプロフィールや投稿内容を見て，自分との趣味などの共通点が見当たらない場合，なぜフォローしてきたのかわからない．このような場合，そのユーザがフォローしてきた理由を知ることはできず，ユーザは自分でそのフォローを受け入れるかの判断をする必要がある．そのユーザに対してどのような対応 (ブロックやフォローバック) するべきかの判断には手間がかかる．

そこで，本研究では，ソーシャルネットワークに潜んでいる不自然なフォロー関係の発見を目的とする．そのために,『あってもおかしくない自然なフォロー関係』と,『不自然で，存在しないはずのフォロー関係』について考える．本研究において，フォロー関係における不自然さを,『ある二人のユーザを第三者が見た際，その二人にフォロー関係があるかどうかを正しく判断できない状態』と定義する．第三者によるフォロー関係の推測と実際のフォロー関係の組み合わせを，図 1 に示す．この際,『誰が見てもその二人はフォロー関係にあると推測するのに，実際にはフォロー関係にない』,『誰が見てもフォロー関係はないと推測するのに，実際にはフォロー関係がある』，の 2 通りの不自然な関係が考えられる．前者の関係は，従来のフォロー推薦の研究で対象とされている，フォロー候補だと考えられる．Twitter ユーザ全員が，興味の近いすべての人や，有益な情報

1：ツイッター，偽フォロア削除へ　水増し防止，数千万件か
https://www.sankei.com/entertainments/news/180712/ent1807120003-n1.html

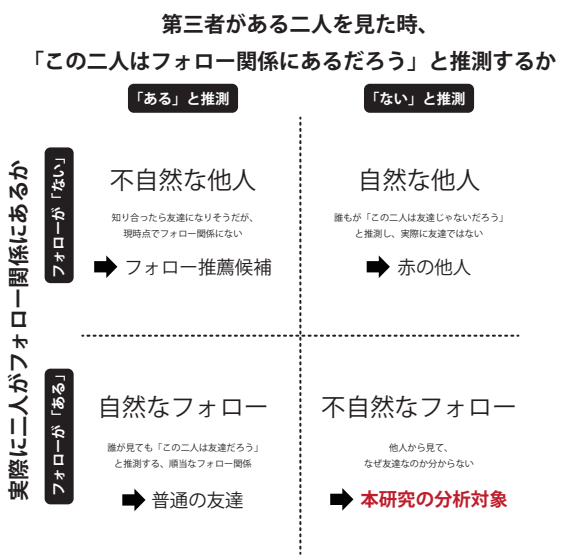

図 1　不自然なフォローと自然なフォローの位置づけ

をつぶやく有名アカウントのすべてをフォローするわけではない．一方で，後者のフォロー関係は，様々な可能性が考えられる．有名人でもない一般アカウントをむやみにフォローしているのはスパムアカウントやリンクファームの可能性もあり，また，ソーシャルネットワーク内には表れない現実世界での関係によって生じたフォロー関係の可能性もある．このような不自然なフォロー関係について，実際に収集し，分析する．

ここで，本研究では，フォロー関係の有無を推定する第三者の代わりに，機械学習による分類器を用いる．はじめに，フォロー関係の有無を分類するランダムフォレスト分類器を作成する．次に，存在しないであろうと分類器に判断されたにもかかわらず実際には存在するフォロー関係をもつユーザのペアを抽出する．最後に，抽出した誤分類ペアについて，分類時の特徴量の重要度を参考にしつつ，人手で分類し，ラベル付けしながら分析する．

本論文は本章を含め全 6 章で構成されている．本章では研究の背景と目的について述べた．第 2 章では関連研究として Twitter のフォロー形成や，アカウント推薦，スパム発見等について論じる．第 3 章では本研究で提案する手法を述べ，第 4 章では提案手法を実装し，実際の Twitter データセットを用いて評価実験を行う．第 5 章で実験を通して得られた結果に関して考察を行い，第 6 章でまとめと今後の展望について述べる．

2　関 連 研 究

関連研究として，Twitter のフォローネットワーク形成に関する研究や，Twitter ユーザの興味領域を考慮したアカウント推薦，スパムアカウントとなりすましアカウントの発見が挙げられる．

2.1　フォローネットワーク形成

ユーザが誰か他のユーザをフォローしたり，またフォローされる際には，フォローを行うにあたっての理由があると考えられる．ここではフォロー形成についての論文を紹介する．

小出ら [8] は Twitter において，ユーザのフォローがどのような目的で行われているかをフォローネットワークを分析することで議論している．ネットワーク内で中心的な役割を果たす高次数ユーザに着目して，フォローおよび被フォローの類似関係における特徴をブログやレビューサイトのネットワークと比較することで分析を行っている．また，高次数ユーザとそのフォロアのやり取りに着目して高次数ユーザの特徴を分析している．高次数ユーザを，フォロアへの一方的な情報発信のために利用する場合が多い実世界での著名人などのグループと，Twitter 上での活動，対話，活発さにより多くのユーザを獲得している実世界における著名人ではないが Twitter 上での有名人などのグループの 2 つのグループに分けることを可能にしている．前者はフォローの双方向性の低さ，フォロアとの対話数の少なさによりトピックごとに分割されていて，後者はグループ内のユーザの結びつきの強さが全体のトピックは明確ではないが，巨大なグループを構築した理由であると結論付けている．

北村 [9] はフォロー数とフォローアカウントの種類より，Twitter の利用パターンを検討している．フォローネットワークサイズとフォロー内容の関係を探索的に分析することにより，Twitter の利用に対する理解を深めることを目的としている．アンケートでどのようなアカウントをフォローしているのかを調査し，Twitter API を使い，フォロー数を求めた．これらをカテゴリカル因果分析を行い,『公式情報』,『未知の個人・その他』,『既知の個人』の 3 つの因子に分類している．さらにこれによって得られた 3 つの因子得点を目的変数とし，フォロー数を説明変数として Fractional polynomial 法により分析を行った結果,『フォロー数が非常に多い利用者では公式情報源のアカウントのフォローが少なくなる』,『フォロー数が非常に少ない場合を除けば，フォロー数が増えるほど未知の個人・その他のアカウントのフォローが多くなる』,『フォロー数が非常に少ない場合を除けば，フォロー数が増えるほど既知の個人のアカウントのフォローが多くなる』という 3 点を示唆している．これらの知見をもとに，本研究では，フォロー関係やフォロー数が不自然なフォロー行為を判断する上で有効であると考え，分類に用いる．

2.2　フォロー推薦

フォローするべきユーザを推薦する行為は，言い換えると『あってもおかしくないのに，存在しないフォロー関係』を発見する行為である．そこで,『あるはずのないフォロー関係』について議論する上で，ここではフォロー推薦についての論文を紹介する．

久米ら [4] は Twitter ユーザの興味領域と特徴語を用いた Twitter アカウント推薦を提案している．Twitter では，膨大な数のユーザの中から自分の趣味嗜好にあったユーザを探すのには手間がかかるという問題がある．従来のユーザ推薦手法は単語頻度・逆文書頻度のような指標で興味後を抽出し推薦を行っている．この手法では顔文字等の興味を示すキーワード以外のものが，興味領域を検出する際のノイズとなりうる．そこで各

キーワードをカテゴリごとに分け，どのカテゴリにも属さないキーワードを除外する．各カテゴリを嗜好情報とし，Twitter アカウント推薦の精度を高めている．

黒柳ら [5] は Twitter のリツイートに着目したユーザ推薦を提案している．同一ツイートをリツイートしたユーザはそのツイートに潜在的な興味を持っているのではないかと考えた．しかしリツイートだけでは興味だけではなく，面白半分など，興味を持っていない場合のリツイートも含まれている可能性があり，ユーザの興味を探るには不十分である．そのため，お互いのツイートの話題傾向を分析し，類似した傾向を持つユーザにも注目した．最終的に，同一のツイートをリツイートしたユーザかつ，お互いのツイートの話題傾向が類似しているユーザをの推薦を行った．

山本ら [6] はツイートの話題の類似度のみでなく，その話題に対するユーザの感情を考慮したフォロイー候補推薦を提案している．感情を 8 次元にわけ，話題が類似している，かつその話題に対するユーザの感情が類似しているユーザをフォロイー候補として推薦した．

これらの研究では，ユーザ同士の類似性に注目してフォロー推薦を行っている．これにならい，本研究でも，プロフィールや発言内容の類似度を分類時に特徴として用いる．

2.3 スパムアカウント，なりすましアカウントの発見

スパムアカウントについての研究は Twitter のみならずさまざまな分野で行われている．その中でも，Twitter におけるスパム発見は極めて活発に研究されている分野である [1]．Twitter はユーザ数が多く，その分スパム行為やなりすまし行為などの迷惑行為も多い．スパムの被害はただ宣伝広告を受け取るだけではなく，ユーザのアカウントを乗っ取り，ユーザ自身が知らないうちにスパムに加担しているというケースも少なからず存在する [2]．なりすましの被害では影響力や発言力の強さから，政治家 [3] や著名人 [4] のなりすましが特に多い．こういったなりすましは，一般の人と比べ影響力があるため，被害が大きくなる危険性がある．Twitter 社は認証バッジシステムを導入し，著名な人々のアカウントが本人であるかの指標を提示しているが，認証バッジが無い著名人のアカウントも存在している．そこで，発言内容に基づく手法や，プロフィールに注目した手法，フォロー関係に注目した手法など，様々なスパム発見手法が提案されている．

発言内容に基づく手法として，若井ら [7] はスパム検出を,『おすすめ』や『無料』など著者らが調査した際に多くのスパムアカウントが使用していた単語を含んだツイートを多数行っているかどうか，リンクつきのツイートを多数行っているか，自己紹介文などにより検出し，なりすまし検知を，ツイートを行う時間帯，スクリーンネームの規則性，使用するクライアント，いいね，リツイートされた数により検出し，ユーザに通知するアプリケーションを開発した．発言内容，フォロー関係に注目した手法として，Song ら [2] はツイートやユーザ同士のやり取り，非フォロアに送信されたツイートの比率などからスパムアカウント検出を行った．大量のスパムおよびスパム以外のユーザのツイートを収集し，分類器を作り，ツイートの分類により比較を行った．フォロー関係に注目した手法として，Wang [3] はフォロー関係に基づき疑わしきユーザの特定を行った．ベイズ分類アルゴリズムにより通常の行動と疑わしい行動を区別し，疑わしいユーザの特定をした．

これらの研究におけるスパムアカウントの持つ特徴は，本研究におけるランダムフォレストの特徴量を決定する上で参考にした．

3 提案手法

本研究では，ある二人のユーザ間のフォロー関係が不自然なものであるかを，ランダムフォレストにおける特徴の重要度から分析する手法を提案する．ある二人のユーザを入力として与えると，その二者間のフォロー関係の有無を推定する，ランダムフォレスト分類器を作成する．はじめに，分析に使用するデータを Twitter API を使用し，収集する．次に分析を行うために収集したデータの整形を行う．整形したデータを用いてランダムフォレスト分類器を学習し，特徴の重要度を算出する．フォローする確率が低いと分類器に判定されたにもかかわらず，実際にフォロー関係のあるパターンに注目し，人手で分析を行う．これにより，ユーザの不自然さを判断するためにどのような特徴が有効なのかを知ることができる．

3.1 Twitter データの収集

はじめに，Twitter API を用いて，分析に使うユーザのデータと，フォロー関係のデータを収集する．Twitter API は Twitter 社が公式に提供する API で，さまざまな情報を取得することができる．ここで取得するデータは『フォロウィー』,『プロフィール』,『タイムライン』である．これらの Twitter データをクローリングを行う際の注意点をまとめると，日本人ユーザを中心に集めるための仕組みの形成，重複したカウントの取得をしないための仕組みの形成，Twitter API の制限に注意することの 3 点である．

データを取得したアカウントのフォロウィーを次のクロール候補とし，順にデータを取ってゆく方式でクロールを行う．クロールの準備作業として，初期ノードのリストを作成する．シードとなるアカウントを手動で決定し，リストを作成した．

次に，作成した初期ノードのリストを元に，実際にクロールを行う．はじめに，リストから 1 件の Twitter id を取得し，タイムラインを取得する．次にプロフィール情報を取得する．プロフィール情報は JSON 形式で取得することができる．最後にユーザのフォロウィーの取得を行う．今回の実験では日本人ユーザを中心に取得するために，タイムライン，またはプロ

2：朝日新聞の Twitter アカウント、乗っ取られて詐欺の宣伝に使われる
https://www.zaikei.co.jp/article/20181113/477688.html
3：米ツイッター、プーチン露大統領になりすましのアカウント停止
https://jp.reuters.com/article/twitter-putin-idJPKCN1NY085
4：香取慎吾、佐藤二朗らも被害に　有名人のツイートに「そっくりな名前とアイコン」で宣伝をぶら下げる“新型なりすまし”が横行
http://nlab.itmedia.co.jp/nl/articles/1804/14/news036.html

フィールに 1 文字でも日本語が入っているユーザに限りフォロウィーを取得した．これらの手順をくりかえす事で，日本人を中心に，友達をクロールしてゆくクローラを作成した．

こうして収集したデータから，フォロー関係，ユーザプロフィール，ユーザタイムラインからなるデータセットを作成した．フォロー関係は (フォロー元 id，フォロー先 id) のタプルで表される．ユーザプロフィールは JSON 形式で，多くの属性と属性値のペアとして表される．この JSON には，ユーザごとに，プロフィールに設定された位置情報，位置情報つきツイートの機能を使っているか，プロフィール文，アカウント作成日，フォロイーの数，フォロアの数，いいねを行った数，リストに追加されている数，公式マークの有無，総ツイート数などが含まれる．タイムラインはあるユーザによるツイートの集合体として表される．各ツイートは JSON 形式で得られる．ツイートには，本文の他に，投稿時刻，位置情報，いいね数，リツイート数などが含まれる．

3.2 ランダムフォレストによる分類

ある二人のユーザ間のフォロー関係の有無を推定する，ランダムフォレスト分類器を作成する．はじめに，複数のフォロー関係に関する仮説に基づき，特徴量を決定する．その特徴量を用いてランダムフォレストで分類する．次に交差検証を行うことで分類の精度を確認する．最後にそれぞれ特徴量の重要度を計算し，どの特徴量が重要であったかを確認する．これにより，自然なフォロー関係の特徴を明らかにできる．

3.3 特 徴 量

ランダムフォレストによる分類に用いる特徴量を，フォロー関係に関する様々な仮説に基づき決定する．二者間のフォロー関係を推定する仮説を，フォローされた側の特徴，フォローする側の特徴，2 者の関係の特徴より考える．

フォローされた側の特徴について，はじめに,『そのユーザが有名人であればフォローされて当然である』と考えられるため，有名さを現す特徴として公式マークとフォロアの人数，リストに追加されている数を用いる．また,『プロフィールを詳細に書いているユーザや，古くからあるアカウントであれば，プロフィールなどの情報が不十分なアカウントや新規のアカウントより信頼性が高い』という仮説を立て，その特徴としてプロフィール文の文字数，プロフィール文に年齢，大学名 [5]，短期大学名 [6]，職業 [7]，都道府県，県庁所在地を含むか，位置情報つきでツイートを行っているか，場所の設定を行っているか，アカウント作成日を用いる．また,『bot であれば特定の単語に反応したり，定期的に自動でつぶやきを行うのでフォローされやすい』という仮説から，その特徴として，プロフィール文または名前に『bot』という単語を含むかを用いる．Twitter の『bot』と呼ばれるものは，機械による自動発言システムである．特定の時間に自動ツイートする bot や，特定のキーワードに反応する bot 等，さまざまな bot が存在する．

フォローする側の特徴について,『そのユーザがフォローを活発に行うユーザであればフォローする』と考え，その特徴としてフォロイーの人数を用いる．また,『相互フォローを積極的に行うユーザはフォロー先のユーザがどんなユーザでもフォローする』と考え，その特徴として，プロフィール文に『相互フォロー』という単語を含むかを用いる．また,『Twitter を活発に利用するユーザであればフォローする可能性が高い』と考え，その特徴として総ツイート数，いいねを行った数を用いる．

最後に，フォロアとフォロイーの 2 者の関係に関する特徴を用いる．例えば趣味嗜好が似ている，居住地が近い，など 2 者の特徴を比較することで初めて意味を持つ特徴になる場合があるため，2 者間の関係の特徴は重要である．ユーザのツイートに注目し,『2 者のツイートに類似性が見られた場合にはフォローする，またはフォローされる可能性が高い』という仮説を立てた．LDA (Latent Dirichlet Allocation) を用いて，ユーザのタイムラインを 15 次元のトピックに分類し，フォロアとフォロイーの発言内容のトピック分布の類似度を特徴として用いる．類似度の計算にはコサイン類似度を用いる．さらに,『2 者のプロフィールの自己紹介文に類似性が見られた場合にはフォローする，またはフォローされる可能性が高い』と考え，プロフィールについても同様に，LDA を用いて自己紹介文を 15 次元のトピックからなるベクトルにし，コサイン類似度を特徴として用いる．

これらの処理を行うことで，それぞれのユーザのペアについて，合計で 116 次元の特徴で表現可能になった．特長の一覧を表 1 に示す．

表 1　フォロー関係の有無の推定に用いた特徴量の一覧（合計 116 次元）

対象	注目する要素	特徴量
フォロー先 および フォロー元 （各 56 次元）	有名さ	公式マークの有無
		フォロア数
		リストに追加されている数
	信頼性	プロフィールの文字数
		自己紹介文内に 47 都道府県を含むか
		自己紹介文内に県庁所在地を含むか
		自己紹介文内に職業のいずれかを含むか
		自己紹介文内に大学名含むか
		自己紹介文内に短期大学名含むか
		自己紹介文内に年齢を含むか
		『位置情報付きでツイート』という機能を使っているか
		場所の設定を行っているか
		アカウント作成日
	bot	自己紹介文内または名前に『bot』という単語を含むか
	ツイート	トピック (15 次元)
	自己紹介文	トピック (15 次元)
	活発さ	フォロイー数
		総ツイート数
		いいねを行った数
	相互フォロー	プロフィールに『相互フォロー』という単語を含むか
2 者の関係 （4 次元）	興味の近さ	ツイートのトピック類似度
		自己紹介文のトピック類似度

3.4 分類と検証

整形された特徴量を用いて，ランダムフォレストで分類する．

5：Wikipedia 日本の大学一覧
https://ja.wikipedia.org/wiki/日本の大学一覧
6：Wikipedia 日本の短期大学一覧
https://ja.wikipedia.org/wiki/日本の短期大学一覧
7：Wikipedia 職業一覧
https://ja.wikipedia.org/wiki/職業一覧

任意のユーザ 2 名に関する特徴から,『ある 2 人のユーザ間にフォロー関係があるか』を正解ラベルとし学習する．ランダムフォレストは各決定木の学習に，ブートスラップサンプリングと呼ばれる，トレーニングサンプルから重複を許してランダムに選択する方法を用いて，多様な決定木を構築し，分類を行う．

次に，5 分割交差検証を行う．分割数については計算時間を考慮し，今回は 5 分割を選択した．学習データを同じサイズで 5 分割する．1 つを残した 4 つのグループを使い訓練し，残り 1 つのグループでテストを行う．これをすべてのグループがテストを行うまで繰り返す．それぞれの訓練セットとテストデータの分割結果に対して精度を計算し，その平均値を取ることにより精度を計算する．

3.5 特徴量の重要度の計算

最後に，どの特徴量が重要であったかを確認するために特徴量の重要度を求める．重要度とはその特徴が分類にどの程度寄与したかを表す．重要度はある特徴量で分岐する前と後での誤差の改善度で求めることができ，各決定木のジニ係数を使い求められる．ノード t におけるジニ係数 $E(t)$ は，$p(i|t)$ を特定のノード t においてクラス i に属するサンプルの割合とすると

$$E(t) = 1 - \sum_{i=1}^{c} p(i|t)^2 \tag{1}$$

と表される．ここですべての決定木におけるジニ係数の減少量の平均をとることで重要度を算出できる．N を決定木の本数，A_i を i 本目の決定木に学習に使われなかったテストデータを使用して正しく予想を行った時の正解率，$(B_m)_i$ をテストデータの特徴量 m となるデータをランダムに並び替え，i 本目の決定木に使用し，本来正しくないはずの予想を行った時の正解率とする．この時，実際に特徴量 m の重要度 I_m を求める式は，

$$I_m = \frac{1}{N} \sum_{i=1}^{N} (A_i - (B_m)_i) \tag{2}$$

となる．

こうして得られた特徴量は，あるユーザが別のユーザをフォローしているかの判別に有効な特徴である．言い換えると,『フォロー関係のあるペア』と『フォロー関係の無いペア』の違う点である．重要な特徴に注目することにより，自然または不自然なフォロー行為について分析できる．

4 評価実験

提案した手法について，実際のデータを用いて検証するため，プログラミング言語の Ruby と Python を用いて,『Twitter データ取得のためのクローラ』,『分類に用いるランダムフォレスト』,『ランダムフォレストの精度を確認するための交差検証』をそれぞれ実装し，評価する．

4.1 Twitter のクローリングによるデータセット作成

実験用のデータセットとして，2018 年 9 月から 2018 年 11 月にかけて Twitter より収集を行った．クロールしたデータの規模を表 2 に示す．収集したデータセットは 5 万件のユーザのプロフィールと，最新 200 ツイート，および 2,797 万ペアのフォロー関係である．2 つの API キーを用いて，ソーシャルグラフの別の箇所からクロールを開始した．クローラは Ruby を使い実装した．Ruby のライブラリである Twitter Gem を用いた．

クロールしたデータから実際に学習に用いるデータを作成する．分類器作成のために，実際にフォロー関係のあるユーザのペア約 2 万件とフォロー関係には無いユーザのペア約 2 万件を使用し，学習用データセットとしてまとめた．タイムラインと自己紹介文を 15 次元のトピックに分類する際，Python のライブラリ Gensim を用いて LDA で 15 次元のベクトルにした．Gensim に用いる文書データの前処理として，形態素解析器である MeCab を用いて名詞のみを抽出した．

実際のランダムフォレスト分類器の学習には，このフォロー関係のデータセットから，フォロー先・フォロー元の両者ともユーザプロフィールとタイムラインのデータセットに含まれた 2 万ペアを正例として用いた．負例として，同数のフォロー関係にないペアをランダムに抽出して用いた．

表 2 Twitter から収集したユーザ間のフォローに関するデータセット

データセット	件数
フォロー関係	27,979,261 件
ユーザプロフィール	51,160 件
タイムライン	51,160 件

4.2 ランダムフォレストによる分類器の作成

分類を行うために，ランダムフォレストを用いた分類器を作成した．ランダムフォレストは Python の機械学習ライブラリ scikit-learn [8]による実装を用いた．はじめに，ランダムフォレストのハイパーパラメータを決定した．ランダムフォレストに用いる決定木の本数である `n_estimators` は，1,000 に設定した．この値を大きくすると多数の決定木の平均を取ることができる．しかし，増やすことによる利益は徐々に減り，メモリの量や計算時間も増大する．乱数のシードである `random_state` は，結果を再現可能にするために，今回の実験では固定値に設定した．`max_depth` は決定木の深さの最大値を指定するものである．デフォルトの設定では，決定木の葉が純粋な状態，つまりデータが完全に分割されるまでノードが展開される．今回はデフォルトの値を用いた．`max_features` は決定木の考慮する特徴量の数を指定する．この値が大きすぎると決定木に使用する特徴量の数が，用意した特徴量の数と近くなり，似た決定木が増える．この値が小さすぎると，それぞれ異なった決定木が出来やすくなるが，データに適合しにくくなる．今回は，バランスをとるため，特徴量の数の平方根を使用した．

以上のハイパーパラメータを設定したランダムフォレストを scikit-learn の `fit()` メソッドで学習させた．

8 : scikit-learn
https://github.com/scikit-learn/scikit-learn

4.3 交差検証を行うプログラムの実装

最後に交差検証を自動的に行うプログラム実装し，ランダムフォレストの精度を確認する．scikit-learn に標準で含まれている `cross_val_score()` 関数を用いて 5 分割交差検証を行う．まず関数のパラメータを設定する．交差検証を行う推定器を `estimator` で指定する．今回は学習済みのランダムフォレスト分類器を指定した．正解ラベルデータを `y` で指定した．データの分割数を `cv` で指定する．今回は 5 分割交差検証を行うので 5 を指定する．

このようなパラメータ設定した交差検証プログラムを動かし，得られる 5 つの値の平均値をランダムフォレストの精度とする．

4.4 実 験 結 果

作成した分類器を 5 分割交差検証したところ精度 (accuracy) は 0.67 であった．ランダムフォレスト分類器を実装した際に，どの特徴が分類する上で重要であったかを，`feature_importances_` の値から求めた．重要度が 0 に近い特徴量を除いた上位 33 件の具体的な数値を表 3 に示す．最も重要度の高かったのはタイムラインのコサイン類似度であった．これは『2 者のツイートに類似性が見られた場合にはフォローする，またはフォローされる可能性が高い』という仮説に基づき，2 者のタイムラインのトピック分類の結果のコサイン類似度を算出したものである．次に重要度の高い特徴は，フォロー先のフォロア数であった．これは『そのユーザが有名人であればフォローされて当然である』という仮説に基づいた有名さを表す特徴の一つである．次に重要度の高い特徴は，フォロー先の総つぶやき数であった．これは『Twitter を活発に利用するユーザであればフォローする可能性が高い』という仮説に基づいた活発さを現す特徴の一つである．最後に，重要度が低かった特徴として，フォロー先の公式マークの有無などがあった．

5 考　　察

実データを用いた実験によって得られた結果から，重要度が高いと判定された特徴について議論し，分析のために誤分類の結果をグループ分けし，考察する．

5.1 各特徴の重要度に注目した考察

表 3 より，特徴量の重要度が大きく 3 つのグループに分けられていることがわかる．重要度の高いグループは，タイムラインの類似度からプロフィールのトピックのコサイン類似度，までの 8 つの特徴である．このグループに含まれる特徴の多くがフォロー先に関する特徴である．

最も分類上で有効だった特徴として，ユーザ間の類似度に関する特徴が 2 つとも重要度で上位（1 位，8 位）であった．フォロー関係にある 2 者の類似度はフォロー関係を推定する際にとても重要であることがわかる．この結果は，先行研究と合致する．多くの研究でトピックの近いユーザを推薦することが行われているが，確かにトピックの類似度はフォロー関係に大きく作用していることがわかった．

次に重要度の高い特徴量のグループは，フォロー元のフォロア数，フォロイー数，リストに追加されている数，総ツイート数，いいね数，プロフィールの文字数の 6 つの特徴であった．また，アカウント作成日の特徴では，フォロー先が 2010 年以降の古いアカウントであるという特徴や，場所の設定，位置情報付きツイートを行っているという特徴も 1 番重要度の高い特徴と比べると約 6 分の 1 程度の重要度であるが，分類には寄与することがわかった．

表 3　フォローの有無の推定において重要度の高かった重要度 (上位 33 件)

特徴	注目要素	重要度
タイムラインのコサイン類似度	ツイート	0.084
フォロー先のフォロア数	有名さ	0.083
フォロー先の総つぶやき数	活発さ	0.083
フォロー先がリストに追加されている数	有名さ	0.081
フォロー先がフォローしている数	活発さ	0.081
フォロー先がいいねを行った数	活発さ	0.079
フォロー先のプロフィールの文字数	信頼性	0.074
自己紹介文のコサイン類似度	自己紹介文	0.074
フォロー元のフォローしている数	活発さ	0.043
フォロー元がいいねを行った数	活発さ	0.040
フォロー元の総ツイート数	活発さ	0.040
フォロー元のフォロア数	有名さ	0.039
フォロー元のリストに追加されている数	有名さ	0.039
フォロー元の自己紹介文の文字数	信頼性	0.036
フォロー先が『位置情報付きでツイート』機能を使っているか	信頼性	0.013
フォロー先のアカウント作成日が 2010 年以前	信頼性	0.013
フォロー先が場所の設定を行っているか	信頼性	0.011
フォロー先の自己紹介文に職業のいずれかを含むか	信頼性	0.011
フォロー先の自己紹介文に短期大学名を含むか	信頼性	0.0085
フォロー先の公式マークの有無	有名さ	0.0076
フォロー元が『位置情報付きでツイート』機能を使っているか	信頼性	0.0075
フォロー先の自己紹介文に 47 都道府県を含むか	信頼性	0.0067
フォロー先のアカウント作成日が 2011 年	信頼性	0.0059
フォロー元が場所の設定を行っているか	信頼性	0.0055
フォロー先が県庁所在地を含むか	信頼性	0.0055
フォロー元の自己紹介文に職業のいずれかを含むか	信頼性	0.0049
フォロー元のアカウント作成日が 2010 年以前	信頼性	0.0045
フォロー先のアカウント作成日が 2012 年	信頼性	0.0041
フォロー元の自己紹介文に 47 都道府県を含むか	信頼性	0.0040
フォロー元の自己紹介文に県庁所在地を含むか	信頼性	0.0038
フォロー元の自己紹介文に短期大学名を含むか	信頼性	0.0037
フォロー先の自己紹介文に大学名を含むか	信頼性	0.0024
フォロー元の公式マークの有無	有名さ	0.0023

これらの結果より，フォロー元の情報よりも，フォロー先がどのような人物であるかという情報が重要であることがわかった．これにより，フォロー元がどのようなアカウントであるか，という情報よりも，フォロー先のアカウントが信頼性の高い人物であるかどうかが，人がフォローを行うかどうかに影響を与えるため，結果的にフォローの有無の推定において重要であったと考えられる．

5.2 フォローが無いと推定されたのに実際はあったユーザのペアの考察

分類結果について，100 件の誤分類を起こしたユーザのペアを集めた．これらの 100 件について人手で誤分類の原因を推測，ラベル付けし，いくつかのグループに人手で分類した．

まずフォロー先が企業のアカウントの際になぜ誤分類が多かったかを考える．誤分類された企業のアカウントは，宣伝を主体とするツイートが多く見られた．一般のユーザが宣伝のツ

表 4　100 件のフォロー関係は無いと推測されたが実際にはあったペアの内訳

アカウントの特徴	比率
フォロー先が鍵付きアカウント	14%
どちらかが活動停止しているアカウント	8%
フォロー先が企業のアカウント	6%
言語や言い回しが異なるアカウント	6%
フォロー関係がある理由が分からないアカウント	66%

イートを投稿することは基本的には無いため，タイムラインのトピックは離れている．一方で，このようなアカウントをフォローする理由として，セール情報などの情報収集目的でのフォローと，キャンペーン参加のためのフォローがある．SNS において，フォロー関係をともなう懸賞やキャンペーンは一般的である．フォローとリツイートで応募できるキャンペーンを行っている企業アカウントは数多く存在している．こうしたフォローをともなうキャンペーンを行うことにより，その分野，コンテンツに興味のあるユーザ層だけではなく，さまざまなフォロアを獲得できる．このことにより，本来であればフォローしないであろう，発言内容のトピックの異なる 2 者間にフォロー関係があるという現象が起きていると推測される．

ここで，キャンペーン終了後，フォロー解除が行われずにそのままフォロー関係が残っている場合も多いのではないかと考えた．この説を検証するため，実際にキャンペーンを行っている，または行ったことのあるアカウントのフォロア 50 件を抽出したデータセットを作成し，本実験で学習した分類器で分類を行った．分類の結果，正しくフォロー関係があると分類されたのは 24 件で，約 48 パーセントのユーザのペアはフォロー関係は無いと誤って分類された．純粋にその企業の情報が欲しいユーザが存在する可能性も高いが，このようなキャンペーンがユーザ間の興味の一致や類似性によらない，不自然なフォロー関係を生み出す一因であると考えられる．

これらの企業をフォローしている Twitter アカウントの中には，リツイートばかり行っているアカウントも存在した．こういったアカウントは，キャンペーン参加のためだけに作られたダミーアカウントであったり，フォロー数水増しのためのリンクファームであることが考えられる．これらのアカウントすべてがスパムアカウントというわけではないが，リツイートを自動化しているアカウントや，個人情報を盗むことを目的とした偽のキャンペーン情報を大量にリツイートしているアカウントも発見された．このような悪質なアカウント以外でも，リツイートをするために複数のアカウントを自動で生成する行為はスパムに値する．これらのことから，不自然なフォロー関係を抽出する事で，スパムアカウントによるフォローが発見できたと考えられる．

次に，タイムラインに共通点がなかったが，プロフィールに共通点があるが言い回しが違う，または言語が違う 2 者のフォロー関係について，なぜ誤分類されたかを考える．この理由は 2 つあると考えられる．1 つ目は，過去の趣味嗜好と現在の趣味嗜好が異なるのはありふれたことであるためである．過去，ある分野の情報収集や，コミュニティへの参加のためにフォローしたが，現在は別の分野について発言し，別のコミュニティに所属していることは珍しくない．新しい分野に興味が移っても，わざわざ人手でフォロー関係を解除するのは煩雑である．そのため，過去のフォロー関係を残しているアカウントが分類器に誤分類されたのではないかと考えられる．さらに，フォロー先が数年前から更新されていないアカウントも存在したため，そのようなアカウントのツイートはフォローを行っていてもタイムラインに流れてくることが無いため，特にフォロー関係を消す理由がない可能性が高くなると考えられる．今回の分析では最新 200 件のツイートを取得し分析を行っているため，それ以前にどのようなツイートをしていたかは考慮していない．そのため，自己紹介文が類似している場合を除き，ツイートから過去の趣味嗜好を知ることが難しく，誤分類につながった．

2 つ目は，言語の違いや，同義語により誤分類を起こした場合である．今回の実験と分析では日本人を対象とするようデータセットを作成した．しかし外国人をフォローしている日本人や，日本人であるが英語を使っている人は一定数存在する．その場合，本手法ではタイムライン，自己紹介文において，英語をはじめとする外国語と日本語を比較し分類をすることができないため，二者のタイムラインや自己紹介文に類似性がある場合でも誤分類を起こす．また自己紹介文を英語で書き，ツイートは日本語で行うアカウントとすべて日本語で書かれているアカウントにフォロー関係が存在した．人手で確認すると，自己紹介文に類似性が見られたが，タイムラインには類似性は見られなかった．このような場合も同様に誤分類を起こす．これらの誤分類は，不自然なフォロー関係である可能性もあるが，多くは分類時に言語や言葉の意味を考慮した類似度計算ができていなかったためだと考えられる．

最後に，分類器が誤分類を起こし，人手でもフォロー理由がわからなかったグループを考える．このグループの中でもフォロー先が鍵付きのアカウント，フォロー先が引用リツイートやウェブサイトを貼り付けた一言ツイートが多いアカウント，フォロー先の発信している情報に一貫性があるアカウント，リツイートが多めであまり情報発信をしないアカウント，または特に得られる情報が無いツイート (例えば『おなか減った』など) を頻繁に行うアカウントをフォローしているグループの 4 つに分けられる．まず，フォロー先が鍵付きのアカウントに関して，タイムラインからの情報がまったく無いため，なぜフォロー関係があるのか推測することが難しい．情報不足のため，このような関係を正しく学習できなかったためだと考えられる．

次にフォロー先が引用リツイート (自分のツイートに他人のツイートを参照として含めること) やウェブサイトを貼り付けた一言ツイートが多いアカウントの場合，そのアカウントのツイートしたい内容の本質は引用したツイートやウェブサイトであるが，文字列としてみた際には URL として含まれるため，タイムラインのトピック分類が適切に行われなかったため，分類器が誤分類したと考えられる．

次に，フォロー先が一貫した情報を発信しているアカウントの場合，フォロー元はたとえプロフィールやタイムラインに言

及がなくても，その情報を欲している場合がある．例えば災害情報やニュースアカウントなどである．今回の分析では毎日新聞のアカウントや，厚生労働省のアカウントなどがあった．一見フォローする理由がわからなくても，フォロー先が特定の情報を発信している場合は，一見して不自然なフォローのように見えるが，正当な理由のあるフォロー関係である．

最後に，あまり情報発信をしないアカウントをフォローしている場合，これはフォロー数水増しやリンクファームのアカウントであるか，もしくは現実での友達ではないかと考えられる．なぜなら 2 者に類似性が見られず，特に有益な情報を発信しているとは考えられないアカウントをフォローする可能性はきわめて低いからである．例外として，現実の友達であれば，情報を欲しているわけではないが，つながりとしてフォローをしておく，などの行為が考えられる．このような関係は Twitter 内だけに注目した場合，不自然なフォロー関係として判断されてもおかしくはない．現実では友達だが，Twitter では何も接点もないアカウントがフォロー関係を結んでいることは，SNS 内だけで見ると，不自然なフォロー関係と見なせると考える．

表 5　誤分類されたユーザのペアの例

フォロー元	フォロー先	2 者の特徴
個人	鍵付きアカウント	フォロー先が鍵付きのアカウントである．
個人	小説家	フォロー元の自己紹介文に『writer』と記述があるが， ツイートの内容は音楽に関するものがほとんど．
個人	ヨドバシカメラ	フォロー元は特に家電等について語ってはいない． しかし度々ヨドバシカメラアカウントはキャンペーンを行っている．
個人	ヨドバシカメラ	フォロー元は大量にリツイートを行っているアカウントである． さらに他にもキャンペーンを行っているアカウントの フォロー・リツイートを行っている．
個人	毎日新聞	2 者の共通点は見られない．
個人	個人	フォロー先が日本人だが，主に英語でつぶやいている．
個人	個人	フォロー先が 2011 年から活動停止している．2 者の共通点は見られない．

6 ま と め

本研究では，Twitter における『フォロー』という行為に着目し，ランダムフォレスト分類器により『不自然なフォロー行為』を発見し，分析を行った．2 者のタイムラインや，プロフィールの自己紹介文を用いたユーザの類似性や，アカウントのフォロアの数などから合計 116 次元の特徴量を作成し，ランダムフォレスト分類器で 2 者間のフォローの有無を分類した．約 4 万件のツイッターユーザの実データを用いた実験の結果，重要度の高い特徴は，タイムラインを用いたユーザの類似度，フォロー先のフォロア数，総ツイート数，リストに追加されている数，フォロイー数，いいね数，プロフィールの文字数，自己紹介文を用いたユーザの類似度であった．この分類器により本研究の目的であった不自然なフォロー行為の発見ができ，その特徴を明らかにできた．そして，分析結果から，不自然なフォロー行為の中でもキャンペーンを行っている企業は分類器に不自然とみなされやすい，などの知見が得られた．

今後の展望として，より正しい分析のため，分類の精度向上が必要である．言語の違いや，言い回しの違いによる誤分類を防ぐため，タイムラインやプロフィールの自己紹介文をいくつかのトピックに分類する際の精度を向上させる必要がある．さらに，フォロー先が鍵付きのアカウントを除く必要がある．加えて，フォロー関係の分析に役立ちそうな特徴量の追加も必要である．例えば，メンション機能を使ったやり取りに注目し，お互いの親密度などの特徴を追加する．また，今回は誤分類されたデータ 100 件を分析したが，より大規模な分析も必要だと考えられる．

本研究では，自然なフォローと不自然なフォローそれぞれの特徴が得られた．この分析結果を用いた応用として，不自然なフォローが発生した際に，不自然である理由と共に，警告してくれるアプリケーションが作成可能になると考えられる．これにより，フォローされた際に，スパムアカウント等をブロックする判断がしやすくなる．さらに自分が鍵付きアカウントである場合は，フォローを承認するか決めるための判断材料となる．こういったアプリケーションを作成することで，ソーシャルネットワークをより快適に，より安全にできると考えられる．

謝　　辞

本研究は JSPS 科研費 18K18161（代表:荘司慶行），18H03243（代表：田中克己）の助成を受けたものです．ここに記して謝意を表します．

文　　献

[1] Abdullah Talha Kabakus and Resul Kara. A Survey of Spam Detection Methods on Twitter. *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 3, 2017.

[2] Jonghyuk Song, Sangho Lee, and Jong Kim. Spam filtering in Twitter using sender-receiver relationship. In *Proceedings of International Workshop On Recent Advances In Intrusion Detection*, pp. 301–317. Springer, 2011.

[3] Alex Hai Wang. Don't follow me: Spam detection in Twitter. In *Proceedings of the 2010 International Conference on Security and cryptography (SECRYPT)*, pp. 1–10. IEEE, 2010.

[4] 久米雄介, 打矢隆弘, 内匠逸. 興味領域を考慮した Twitter ユーザ推薦手法の提案と評価. 研究報告知能システム (ICS), Vol. 2015, No. 1, pp. 1–8, 2015.

[5] 黒柳智士, 鈴木浩, 服部哲, 速水治夫. 同一ツイートをリツイートしたユーザ群の情報を用いたユーザ推薦アルゴリズムの提案. マルチメディア, 分散協調とモバイルシンポジウム 2013 論文集, Vol. 2013, pp. 80–83, 2013.

[6] 山本湧輝, 熊本忠彦, 灘本明代. 話題と感情の可視化に基づくフォロイー推薦. 第 8 回 Web とデータベースに関するフォーラム論文集, Vol. 2015, pp. 174–181, 2015.

[7] 若井一樹, 佐々木良一. Twitter のスパム検知機能となりすまし検知機能の開発と評価. 情報処理学会論文誌, Vol. 56, No. 9, pp. 1817–1825, 2015.

[8] 小出明弘, 斉藤和巳, 風間一洋, 鳥海不二夫. ネットワーク分析による Twitter ユーザのフォロー形成に関する一考察. 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 6, No. 2, pp. 164–173, 2013.

[9] 北村智. Twitter におけるフォロー数とフォローアカウントの種類の関係についての探索的分析. 東京経済大学コミュニケーション学会 コミュニケーション科学, Vol. 41, pp. 49–63, 2015.