Vol.2017-IS-142 No.1
2017/12/2

Twitter におけるアカウント情報の特徴を利用したアカウント判別分析

我妻　拓哉[†1]　吉村　博幸[†2]

概要：近年，Twitter は最新情報やリアルタイムでの情報収集を目的として利用されている．しかし，その中には有害なリンクの投稿や同じツイートの繰り返し投稿等，他の利用者にとって迷惑となるスパム行為を一方的に繰り返す不正アカウントが増え続けており，一般ユーザーを装って利用者から情報収集が行われる危険性がある．そこで本研究では，フォロー返し率やフォロー返され率等のアカウント情報を利用して，正規アカウントと不正アカウントの特徴の差異に基づいた判別分析を行った．その結果，不正アカウントを検知して利用者から情報収集を未然に防止することが可能になった．

キーワード：Twitter，スパム，判別分析

Account discriminant analysis using characteristic of account information on Twitter

Takuya WAGATSUMA[†1]　Hiroyuki YOSHIMURA[†2]

Abstract: In recent years, Twitter has been widely used for the purpose of gathering up-to-date information and real-time information. However, false accounts, which post harmful links, repeat the same tweet, etc., do spamming acts regarded as a trouble for other users and are increasing. There is a danger that false accounts pretend to be general users, and collect information from the users. Therefore, in this research, discriminant analysis using the difference between the characteristics of the regular account and the false account, i.e., the follow-up rate and the follow-back rate, was performed. As a result, it was shown that it is possible to detect false accounts and to prevent the collection of user information from them beforehand.

Keywords: Twitter, Spam, Discriminant analysis

1. はじめに

近年，ソーシャルネットワークサービス（SNS）の利用者が増加している．なかでも，国内での Twitter 利用者は増加しており，2015 年 12 月時点で 1 カ月間に Twitter にログインした月間アクティブユーザー数は 3500 万人であった．世界全体では 3 億 2000 万人で，約 1 割が日本国内からのアクセスだった．Twitter Japan 設立時の 2011 年 3 月は 670 万人だったので 5.2 倍に増加し，この増加率は日本が世界トップだった[1]．日本での Twitter の利用目的は，情報の収集・発信・共有（メディア機能）に特化している傾向がある．一方，Twitter を利用し悪質な迷惑行為を行うアカウントも増加している．

スパムの被害について Twitter 社は公表しており，2010 年 1 月時点では 1 日に約 5000 万件，1 秒あたり約 600 件のツイートが発信されているが，その約 2%をスパムツイートが占めており，1 日約 100 万件がツイートされている[2]．

スパム被害には，ただスパム宣伝を受け取るだけでなく，宣伝ツイートを勝手にツイートされてしまうケースが増加している．この原因は，ユーザー自身が悪意のある連携アプリを認証してしまったためと思われる．つまり，連携したアプリに与えた Write 権限（連携アプリがユーザーのアカウントでツイートできる）を利用されてしまうためである．他方，Twitter 以外のサービスからパスワードが流出し，同じパスワードを Twitter でも使用していた場合にアカウントを本当に乗っ取られてしまうケース（リスト型攻撃）もある[3]．

Twitter のアカウント作成に必要なものは，以前までは Twitter ID とパスワードのみであったが，不正なアカウント作成を未然に防ぐために現在，電話番号とメールアドレスの登録が必要となっている．しかし，メールアドレスだけでもアカウント作成が可能な方法があるため，不正アカウントが減少しているとは言い難い．Twitter を管理する Twitter 社においても利用者によるスパム報告や Twitter ルールに違反したアカウントを凍結・削除する対策をしているが，Twitter は情報拡散機能に長けているため不正アカウントによる被害は多くの人に悪影響を及ぼす可能性がある．スパム被害を防ぐためには，利用者自身が不正アカウントを未然に見抜く必要がある．そのためには，利用者が不正アカウントかどうかを判別できる事が重要になってくる．

†1 千葉大学大学院工学研究科
Graduate of Engineering, Chiba University

†2 千葉大学大学院工学研究院
Graduate of Engineering, Chiba University

そこで本研究では，一般の利用者が容易に得ることのできる情報を判別項目とし，それらを組み合わせて正規アカウントと不正アカウントを判別する手法を提案する．

2. 関連研究

Twitter の不正アカウント判定に関するサービスと研究は数多く存在している．以下では，不正アカウント判定に関するサービスと研究について紹介する．

2.1 不正アカウント判定に関するサービス

Fake Follower Check[4]では，フォロワーが同一言語利用者か，100 日以上投稿しているか，フォロワーが 250 人以下であるかを判定基準として，Fake/Inactive/Good の三段階で評価するという利点がある．しかし，無料版においては，判定はパーセンテージのみであり，かつ日本語でのアプリ提供はなしという欠点がある．

一方，Botometer[5]では，使用方法，ネットワーク，感情，コンテンツ，友達を判定基準として，対象アカウントがボット（bot）である可能性を 100 点満点の数値ではじき出す．しかし，判定基準の詳細には No Data Available 表示が多くみられる．これは，日本語に対応していないためであり，Fake Follower Check と同様に日本語でのアプリ提供はない．

2.2 不正アカウント判定に関する研究

Chen らによる研究[6]では，機械学習させた分類器を作成してスパムツイートと非スパムツイートを検出する手法を提案している．具体的には，6 億件のツイートを収集し，トレンドマイクロの Web Reputation System を適用して 650 万件のスパムツイートを検出した．これらのツイートからスパムツイートと非スパムツイートを区別できる 12 種類の特徴量を抽出し，ランダムフォレスト，C4.5 決定木，ベイズネットワーク，Naïve Bayes，k 近傍法，及びサポートベクターマシンの 6 つの機械学習をして分類器を作成した．その分類器によって，10 日間毎日 10 万件のスパムツイートと 10 万件の非スパムツイートのデータセットに対してスパムツイート検出を行った．ランダムフォレストと C4.5 決定木は非スパムツイートに対する検出率は 90%以上の検出率を維持したが，スパムツイートに対しては最高値で 90%，最低値で 40%以下の値を推移し安定せず，不正確であったと報告されている．

また，和田らによる研究[7]では，文字 n-gram を用いた文体類似度，ツイートのクライアント，及び投稿時間を基にスパムツイートを検出する手法を提案している．これにより，スパマーによるアカウント乗っ取りによって発生する，正規アカウントから投稿されたスパムツイートの検知を行う．クライアントと投稿時間の関係を用いて，n-gram において n=2 のとき，文体非類似度に重み付けを行う手法で 82.8%の正答率を出している．

さらに，中村らの研究[8]ではまず，Twitter 社のスパム定義から 29 種の特徴を抽出する．そこからスパムユーザフィルターを開発し，機械学習を用いて分類器を作成後，Twitter 社にアカウント凍結・削除されていないスパムアカウントに対して 94.7%の割合で正しく判定している．

一方，岩井らの研究[9]では，上記の結果を踏まえた上で，スパム行為となりすまし行為の検知手法を提案している．Twitter 社の判断基準と独自考案を含めた 8 つのスパム判定項目を作成して，判別的中率 95.8%の精度を持つスパム判別手法を確立している．また，判別的中率 96.3%の精度を持つなりすまし判別手法を提案している．これら 2 つの検出手法を用いてスパムかなりすましかを判定し，スコアで算出するアプリケーション LookUpper の開発を行っている．

このように，従来手法では判別的中率 94.7%[8]や 95.8%[9]の精度でスパムアカウントが検出されているが，本研究では，従来手法より優れた判別的中率を持つ判別手法の考案を目標とする．

3. 判別項目について

3.1 スパムの定義

Twitter 社は，ユーザー名の不正確保や招待スパム，ユーザー名の売買，マルウェア，フィシング，スパム，以上６つの行為をアカウントの凍結条件としている．スパムの定義として，「“フォロー獲得”や“フォロー急増”をうたうサービスを利用または利用を助長する場合」や「誤解を招くようなアカウントの作成や反応を行っている場合」など，17 個の基準が設けられている[10]．本研究では，これらの基準を踏まえた上で，一般の利用者を「正規アカウント」，それ以外の逸脱した迷惑行為を行うアカウントを「不正アカウント」として定義する．

3.2 不正アカウントの特徴分析

3.2.1 アカウント情報の収集

本研究では一般の利用者が得ることのできる情報を判別項目とすることを目的としているので，whotwi[11]を用いてアカウントの情報を得る．whotwi は，Twitter 利用者を分析して，仲良しの人やツイート内容，ハッシュタグ，クライアント，時間帯，文字数などを分析できるサイトである．また，フォロー状況を分析して，フォローを返していない人，フォロー返しされていない人の一覧を見ることができる．これらの情報から，Twitter 社のアカウントの凍結条件と実際の不正アカウントと判明しているアカウントから特徴を抽出する．

3.2.2 判別項目の作成

各判別項目の詳細を表 1 に示す．不正アカウントの特徴からフォロー系，プロフィール系，ツイート系の３つに分類した．また，Twitter 社のスパム定義に則った判別項目は「Twitter 社の判別基準」，新たに考案した判別項目は「新たに考案した判別基準」として内訳に記した．以上を踏まえた上で，11 個の判別項目の詳細について以下に示す．

判別項目 1 はアカウントのフォロー数とフォロワー数を用いた判別項目である.不正アカウントの相互フォロー(互いに相手をフォローしている状態のこと）数について分析したところ，フォロー数に対する相互フォローの割合が少ないアカウントが多い．その割合を用いて不正アカウントかどうかを判別する.

判別項目 2 もアカウントのフォロー数とフォロワー数を用いた判別項目である．不正アカウントの相互フォロー数について分析したところ，フォロワー数に対する相互フォローの割合が少ないアカウントが多い．その割合を用いて不正アカウントかどうかを判別する.

判別項目 3 は Twitter 歴（Twitter の利用期間）を用いた判別項目である．不正アカウントは利用者によるスパム報告や Twitter ルールに違反したアカウントの凍結･削除されているため, 正規アカウントに比べて Twitter 歴が短いアカウントが多い．その日数を用いて不正アカウントかどうかを判別する.

判別項目 4 は 1 日のツイート回数を用いた判別項目である．不正アカウントは自動的にツイートを行う bot などを利用するため，1 日のツイート回数が正規アカウントよりも多い．その回数を用いた判別項目である.

判別項目 5 はひとりごと率を用いた判別項目である．ひとりごと率とはツイート数に対するメンション(特定の「@ユーザー名」を含むツイート）でないツイートの割合のことである．不正アカウントのツイートは有害なサイトへの誘導を目的としているため，リツイートやリプライではなく一方的にツイートする割合が多い．その割合を用いて不正アカウントかどうかを判別する.

判別項目 6 は平均文字数を用いた判別項目である．平均文字数は累計文字数をツイート回数で割ることで導き出される．不正アカウントは一方的にツイートし，ツイート回数も多いため平均文字数も多くなる．その文字数を用いて不正アカウントかどうかを判別する.

判別項目 7 は 1 日平均文字数を用いた判別項目である．1 日平均文字数は累計文字数を Twitter 歴で割ることで導き出される．不正アカウントは，Twitter 歴は短いが一方的にツイートするため累計文字数は多くなるので，1 日平均文字数は多くなる．その文字数を用いて不正アカウントかどうかを判別する.

判別項目 8 は平均ツイート間隔を用いた判別項目である.不正アカウントは自動的にツイートを行う bot などを利用するため，ツイートの平均間隔は短いことが多い．その数値を用いて不正アカウントかどうかを判別する.

判別項目 9 はリンク率を用いた判別項目である．リンク率とはツイート数に対するリンクを含むツイートの割合のことである．不正アカウントはアフィリエイトやマルウェアサイトなどの有害な URL を含むツイートが多く存在する.その割合を用いて不正アカウントかどうかを判別する.

判別項目 10 はメディア（画像と動画のこと）率を用いた判別項目である．メディア率とはツイート数に対する画像や動画を含むツイートの割合のことである．不正アカウントは画像や動画を載せることで，一般の利用者が有害な URL へ誘い出すツイートが多く存在する．その割合を用い

表 1 不正アカウントの判別項目の詳細
Table 1 Detail of discriminant item of false account.

判別項目	内容	内訳
判別項目 1 （フォロー系）	フォロー返し率 相互フォロー数/フォロワー数 ×100	Twitter 社の 判断基準
判別項目 2 （フォロー系）	フォロー返され率 相互フォロー数/フォロー数 ×100	Twitter 社の 判断基準
判別項目 3 （プロフィール系）	Twitter 歴	新たに考案し た判別基準
判別項目 4 （ツイート系）	1 日ツイート回数	新たに考案 した判別基準
判別項目 5 （ツイート系）	ひとりごと率	新たに考案 した判別基準
判別項目 6 （ツイート系）	平均文字数	新たに考案 した判別基準
判別項目 7 （ツイート系）	1 日平均文字数	新たに考案 した判別基準
判別項目 8 （ツイート系）	平均ツイート間隔	新たに考案 した判別基準
判別項目 9 （ツイート系）	リンク率 ツイートにリンクが含まれている割合	Twitter 社の判 断基準
判別項目 10 （ツイート系）	メディア率 ツイートにメディアが含まれている割合	Twitter 社の 判断基準
判別項目 11 （ツイート系）	リツイート率 ツイート数に対するリツイートのツイートの割合	新たに考案 した判別基準

表 2 各判別項目のスコア
Table 2 Score of each discriminant item.

スコア / 判別項目	正規	不正	全体
判別項目 1	0.86	0.30	0.58
判別項目 2	0.94	0.36	0.65
判別項目 3	0.00	0.20	0.10
判別項目 4	0.00	0.04	0.02
判別項目 5	0.18	0.62	0.40
判別項目 6	0.10	0.58	0.34
判別項目 7	0.24	0.12	0.18
判別項目 8	0.04	0.04	0.04
判別項目 9	0.08	0.30	0.19
判別項目 10	0.06	0.16	0.11
判別項目 11	0.00	0.06	0.03

て不正アカウントかどうかを判別する．

判別項目 11 はリツイート（ツイートを再投稿すること）率を用いた判別項目である．リツイート機能を使うと，ツイートをすべてのフォロワーにすばやく共有できる．不正アカウントはその機能を利用して，自分のツイートや他の不正アカウントのツイートを再投稿する割合が多い．その割合を用いて不正アカウントかどうかを判別する．

表 2 に各判別項目のスコアを，正規アカウント，不正アカウント，全アカウントごとに示す．スコアとは対象アカウント数に対して正しく判定されたアカウント数を示す．判別項目 2 における正規アカウントのスコアは 0.94 と高い値を示すが，不正アカウントは 0.36 と低い値を示す．そこで，これらの判別項目を組み合わせて判別分析を行う．

4. パラメーターフィッティングによる判別条件の決定

パラメーターフィッティングによる判別条件の決定方法を本節で示す．パラメーターフィッティングとは，実験データなどに対応した計算式があるとき，式の係数（パラメーター）を実験値と計算式が一致するように選択することである．本研究では，実験データをパラメーターデータとし，実験値はアカウントを正しく判別しているかどうかを指している．それぞれの判別項目を判別分析によって，使用する判別項目を決定する．使用する判別項目から条件を決定し，パラメーターを作成する．

4.1 判別分析について

判別分析とは，目的変数がカテゴリーデータ（群データ），説明変数が数量データの時に適用できる解析手法のことを言う．判別分析には，線形判別分析とマハラノビスの距離による判別分析の大きく分けて 2 種類ある[12]． 線形判別分析により，説明変数が 2 変数x， yの場合，2 群の境界となる式（1）を求めれば，式（1）の値の正負によりどちらの群に属するかを判別することができる．

$$z = ax + by + c \quad (a,\ b,\ c\text{は定数}) \quad \cdots\cdot(1)$$

ここで，a， bを判別係数という．

一方，マハラノビスの距離による判別分析とは，グループの重心までのマハラノビスの距離ともう一方のグループの重心までのマハラノビスの距離を求め，距離の短いほうのグループに属すると判別する方法のことを指す．なお，データの散らばりの程度を標本分散，標本共分散で測り，データの散らばりを考慮に入れた距離をマハラノビスの距離という．以下に説明変数が 2 変数x， yの場合を示す．

対象となる点 P を$(x,\ y)$とする．グループ A の平均値を$\overline{x_A}$，$\overline{y_A}$と標準偏差をs_{x_A}とs_{y_A}とし，グループ B の平均値を$\overline{x_B}$，$\overline{y_B}$と標準偏差をs_{x_B}とs_{y_B}とする．点 P からグループ A の重心までのマハラノビスの距離を式（2）によって求め，点 P からグループ B の重心までのマハラノビスの距離を式（3）によって求める．

$$D_A^2 = \begin{pmatrix} {s_{x_A}}^2 & s_{x_A y_A} \\ s_{x_A y_A} & {s_{y_A}}^2 \end{pmatrix}^{-1} \begin{pmatrix} x - \overline{x_A} \\ y - \overline{y_A} \end{pmatrix} \quad \cdots\cdot(2)$$

$$D_B^2 = \begin{pmatrix} {s_{x_B}}^2 & s_{x_B y_B} \\ s_{x_B y_A} & {s_{y_B}}^2 \end{pmatrix}^{-1} \begin{pmatrix} x - \overline{x_B} \\ y - \overline{y_B} \end{pmatrix} \quad \cdots\cdot(3)$$

そして，以下のルールに従い点 P を判別する．

$D_A^2 > D_B^2$　ならば点 P はグループ A に属する

$D_A^2 < D_B^2$　ならば点 P はグループ B に属する

4.2 使用する判別項目と条件の設定

4.2.1 使用する判別項目

表 1 の判別項目を用いて実際に正規アカウントと不正アカウントを判別できるか，判別分析を用いて検証する．本検証では，マイクロソフト社の Excel にある「分析ツール」を用いる．各変数について，目的変数には「正規アカウント」または「不正アカウント」の 2 群，説明変数には「判別項目 1」から「判別項目 11」の計 11 個を設定する．

適用対象として，正規アカウント 100 件と不正アカウント 100 件，合計 200 件のアカウントを用いる．これらは Twitter ルール[10]に従い，ユーザー名の不正確保に当たらない 6 ヶ月以内に Twitter の更新があるものを使用する．また，ツイート情報は最新のツイート 600 件以内を用いる．

まず，正規アカウント 50 件，不正アカウント 50 件の計 100 件のパラメーターデータを用いて 11 項目の判別項目を判別分析にかけ，その結果からパラメーターフィッティングを行う．次に，使用する判別項目から条件を設け，パラメーターを作成する．そして最後に，作成されたパラメーターを用いて，残り 100 件のテストデータを正規アカウントと不正アカウントに正しく判別できるか検証する．

4.2.2 使用する判別条件と結果

線形判別分析とマハラノビスの距離による判別分析を用いて，11 個の判別項目を組み合わせて，2 変数と 3 変数における判別的中率を出し，パラメーターフィッティングにより最適な組み合わせを導く．表 3 に，2 変数の場合における線形判別分析とマハラノビスの距離による判別分析のスコアを示した．この結果から，スコアの高い判別項目 1 と 3 を選び，3 変数の場合における線形判別分析とマハラノビスの距離による判別分析を行った．その結果，判別項目 3 と 7，1 と 2 と 3，1 と 3 と 4 を用いたマハラノビスの距離による判別分析が判別的中率 92.0%を超えた．そのため，これらをそれぞれ条件 1，2，3 とした．

さらに判別的中率を上げるために新たに条件 4 を加えた．条件 4 では，クライアント名，アカウント名，ユーザー名によって判別する．クライアントとは，Twitter のサービスを利用して独自機能を搭載するクライアントソフトウェア

表 3 判別分析のスコア（2 変数の場合）
Table 3 Score of discriminant analysis in 2 variables.

マハラノビス＼線形	1	2	3	4	5	6	7	8	9	10	11
1		0.88	0.86	0.80	0.84	0.81	0.82	0.79	0.80	0.80	0.79
2	0.87		0.86	0.83	0.77	0.78	0.84	0.79	0.76	0.75	0.78
3	0.89	0.87		0.85	0.85	0.84	0.85	0.87	0.63	0.85	0.85
4	0.83	0.85	0.91		0.79	0.81	0.54	0.68	0.59	0.72	0.68
5	0.82	0.76	0.86	0.85		0.77	0.77	0.64	0.71	0.74	0.70
6	0.79	0.78	0.89	0.83	0.77		0.74	0.50	0.75	0.72	0.75
7	0.85	0.85	0.92	0.74	0.84	0.83		0.28	0.77	0.75	0.29
8	0.78	0.79	0.86	0.84	0.64	0.79	0.77		0.70	0.71	0.57
9	0.79	0.77	0.83	0.81	0.74	0.72	0.84	0.73		0.68	0.67
10	0.80	0.77	0.85	0.75	0.66	0.70	0.77	0.71	0.68		0.70
11	0.78	0.79	0.85	0.71	0.79	0.73	0.70	0.69	0.50	0.68	

の総称を指す. 例として「Twitter for iPhone」や「twittbot.net」，「TweetDeck」などがあり，不正アカウントは自動ツイートを行う bot 機能を利用する傾向が高い．このため本研究では，bot や有害サイトを用いたクライアントの割合が 90%を超えている場合は不正アカウントとみなす．一方，この割合が 90%以下の場合はアカウント名とユーザー名で判別を行う．また，正規アカウントにおいても bot を使うアカウントが存在するため，誤判別を防ぐためにアカウント名とユーザー名での判別を追加する．正規アカウントにはない不正アカウントの特徴として，アカウント名とユーザー名に相互フォローや出会い，Twitter アカウント売買などを仄めかす文字や初期設定のままの文字を使う傾向がある．これらを条件 4 として用いる．

表 4 から，正規アカウントについては，条件 1，2，3，4 においてすべてスコア 1.00 であり，正規アカウント 50 件すべて正しく判別されることが分かる．一方，不正アカウントについてスコアは，条件 1, 2 において 0.84，条件 3 において 0.86，条件 4 において 0.50 であり，不正アカウント 50 件中それぞれ，42 件，43 件，25 件が不正アカウントと判別されることがわかる．

以下に解析フローの手順について示す．また，解析フローチャートを図 1 に示す．

手順 1：対象アカウントを条件 1 で判別する．そこで不正アカウントと判別されたアカウントを不正アカウントとする．

手順 2：手順 1 で正規アカウントとして判別されたアカウントに対して条件 2, 3 を適用し，両条件で不正アカウントと判別されたアカウントを不正アカウントとする．

手順 3：条件 1，2，3 すべてで正規アカウントとして判別されたものに対して，条件 4 を適用して判別させる．

条件 4 に反するアカウントは不正アカウントと判別する．条件 4 を追加した理由は，条件 1，2，3 だけでは不正アカウントを正規アカウントと誤判別されるためであり，これを防ぐためである．

表 4 パラメーターフィッティングの結果
（パラメーターデータの場合）
Table 4 Result of parameter fitting in parameter data.

説明変数	アカウント	スコア（件数）
条件 1	正規アカウント	1.00（50）
	不正アカウント	0.84（42）
条件 2	正規アカウント	1.00（50）
	不正アカウント	0.84（42）
条件 3	正規アカウント	1.00（50）
	不正アカウント	0.86（43）
条件 4	正規アカウント	1.00（50）
	不正アカウント	0.50（25）

条件については以下に示す

・条件 1：マハラノビスの距離による判別分析
（Twitter 歴&1 日平均文字数）

・条件 2：マハラノビスの距離による判別分析
（フォロー返し率&フォロー返され率&Twitter 歴）

・条件 3：マハラノビスの距離による判別分析
（フォロー返し率&1 日ツイート回数&Twitter 歴）

・条件 4：クライアント名，アカウント名，ユーザー名

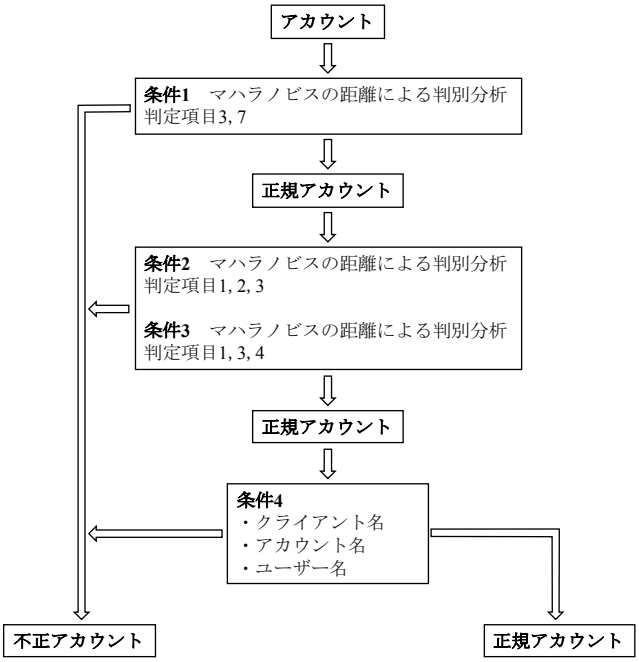

図 1 解析フロー
Figure 1 Analysis flow.

以上の解析フローを踏まえた結果を表 5 に示す．正規アカウントについては 50 件中 50 件を正規アカウントとして，不正アカウントについては 50 件中 46 件を不正アカウントとして判別することができた．つまり，全体のアカウント 100 件中 96 件を正しく判別でき，条件 1，2，3，4 を踏まえた判別的中率は 96.0%を示した．

なお，本論文において正規アカウントと誤判別された不正アカウントのツイート内容を確認したところ，同じツイート投稿を行っていることや有害サイトへのアクセスを助長していることが確認できた．

表 5 パラメーターフィッティングデータによる判別結果
Table 5 Result of parameter fitting data.

		スコア（件数）	
		正規アカウントである	不正アカウントである
使用したアカウント	正規アカウント（50）	1.00（50）	0.00（0）
	不正アカウント（50）	0.08（4）	0.92（46）
判別的中率		96.0%	

5. テストデータによる判別結果

残りの正規アカウント 50 件と不正アカウント 50 件の計 100 件をテストデータとして用いて，パラメーターフィッティングを行い，実際に正規アカウントと不正アカウントを正しく判別できるか検証した．その結果を表 6 に示す．

表 6 から，正規アカウントについては 50 件中 50 件を正規アカウントとして判別でき，不正アカウントについては 50 件中 49 件を不正アカウントとして判別することができたことがわかる．結果として，全体のアカウント 100 件中 99 件を正しく判別でき，条件 1，2，3，4 を踏まえた判別的中率は 99.0%であった．従来手法[9]では 95.8%であったのに対して，本手法では 3.20%精度が高くなった．

なお，正規アカウントと誤判別してしまった 1 件の不正アカウントについては，条件 1 から 3 すべて正規アカウントと判別されている．また，クライアントは Web Client を使用しているため，条件 4 においても正規アカウントと判別されている．さらに，ツイート内容を確認したところ，同じツイート投稿を行っていることや有害サイトへのアクセスを助長していることが確認できた．

表 6 テストデータによる判別結果
Table 6 Result of test data.

		スコア（件数）	
		正規アカウントである	不正アカウントである
使用したアカウント	正規アカウント（50）	1.00（50）	0.00（0）
	不正アカウント（50）	0.02（1）	0.98（49）
判別的中率		99.0%	

6. おわりに

本研究では，アカウント情報から不正アカウントの特徴抽出し，そこから判定項目を作成後，判別分析を用いて正規アカウントと不正アカウントを判別することを目標とした．本研究の判別手法によって正規アカウントと不正アカウントに対して，パラメーターデータにおいて 96.0%，テストデータにおいては従来手法[9]よりも高い 99.0%の精度で判別が可能になった．

今後は誤判別してしまったアカウントの傾向を分析し，ツイート本文の内容の特徴などによって新たな判別項目を導入することにより，判別手法のさらなる高精度化を検討する．また，判別分析において説明変数を 4 次以上に拡張させて条件数を減らすことにより，解析フローの単純化を検討する．さらに，Twitter 利用者が不正アカウントのフォローを未然に防ぐことができるようなアプリケーション開発を目標とする．これは，Twitter のみならず，他の SNS における不正アカウントの判別への応用も含めたアカウント判別分析へ貢献できると考えられる．

謝辞 本研究を進めるにあたりご指導頂いた吉村博幸准教授，並びに日常の議論を通じて多くの知識や示唆を頂いた吉村研究室の皆様に感謝します．

参考文献

[1] “Twitter が国内ユーザー数を初公表 「増加率は世界一」”. http://www.huffingtonpost.jp/2016/02/18/twitterjapan_n_9260630.html, (参照 2017−10−28).
[2] “Twitter を使ったスパムの状況について”. https://blog.twitter.com/official/ja_jp/archive1/ja/2010/twitter-1.Html, (参照 2017−10−28).
[3] “【最新】Twitter 乗っ取りの対処法，スパムを勝手にツイートする不審なアプリを確認して連携解除する方法【iPhone/Android/PC】”. http://appllio.com/twitter-app-revoke-by-smartphone-iphone-android-pc, (参照 2017−10−28).
[4] “Fake Follower Check”. https://fakers.statuspeople.com, (参照 2017−10−28).
[5] “Botometer by OSoMe”. https://botometer.iuni.iu.edu/, (参照 2017−10−28).
[6] Chen C., et al.. A Performance Evaluation of Machine Learning-Based Streaming Spam Tweets Detection, IEEE Transactions on Computational Social Systems, 2015, vol. 2, no. 3, pp. 65-76.
[7] 和田なぎさ，奥谷貴志，山名早人．Twitter におけるアカウント乗っ取りによるスパムツイートの検出, DEIM Forum 2013, 2013, 5-F5.
[8] 中村悠一，山田剛一，絹川博之．Twitter におけるスパムユーザフィルタの開発とその評価（マイクロブログ，D 分野: データベース），情報科学技術フォーラム講演論文集, 2012, vol. 11, no. 2, pp. 99-100.
[9] 岩井一樹，佐々木良一．Twitter のスパム検知機能となりすまし検知機能を強化するアプリケーション LookUpper の開発と評価，情報処理学会論文誌，2015, vol. 56, no. 9, pp. 1817-1825.
[10] “Twitter ルール”．https://support.twitter.com/articles/253501, (参照 2017−10−28).
[11] “whotwi”．http://ja.whotwi.com, (参照 2017−10−28).
[12] 涌井良幸，涌井貞美．実習 多変量解析入門〜Excel 演習からムリなくわかる，技術評論社, 2011.