

Twitter におけるアカウント判別分析を用いた パラメーター作成と判別精度

我妻 拓哉^{†1} 吉村 博幸^{†2}

概要: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 利用者は増加している一方、Twitter を利用して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。

我々の前回の研究では、はじめに、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、使用する判別項目から判別条件 1, 2, 3, 4 を決定し、パラメーターを作成した。次に、残りの正規アカウント 50 件、不正アカウント 50 件の計 100 件をテストデータとした。これらのデータに対して、上で決定した判別条件を用いて、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どうか検証した。その結果、パラメーターデータに対して判別の中率 96.0%、テストデータに対して判別の中率 99.0%を示した。

今回の研究では、従来手法より優れた手法を提案するとともに、前回の研究からデータ依存性による判別の中率の偏りの改善と条件数を減らすことによる解析フローの単純化を目的とした。使用する判別項目から判別条件を決定してパラメーター作成時にデータ分割を行い、判別の中率の平均と標準偏差を算出して検証した。パラメーターフィッティングを行った結果、判別条件 1 と 4 を用いた場合に最も良い結果が得られ、パラメーターデータに対して判別の中率 $98.41 \pm 0.65\%$ 、テストデータに対して判別の中率 $98.41 \pm 0.74\%$ を示した。

前回の研究と比較したところ、パラメーターデータに対する判別の中率は $2.41 \pm 0.65\%$ 上昇し、テストデータに対して $0.59 \pm 0.74\%$ 減少したが、従来手法による判別の中率 95.8% より $2.61 \pm 0.74\%$ 高くなった。今回の研究により、データ依存性による判別の中率の偏りが改善されるとともに、条件数を減らすことにより解析フローの単純化が可能となった。

キーワード: Twitter, スпам, 判別分析

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、2017年10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4500万人であった。2016年9月(4000万人)から1年1ヶ月で500万人増加したことになる。ちなみに、Twitter Japan 設立時の2011年3月は670万人だったので約6.7倍に増加したことになる[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[2]。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]。また、近年報告されているスパム被害には、実在の人物になりすました偽アカウントを作り、“フォロワーを売る”というビジネス[4]やツイートにグロテスクな画像をリプライで送りつける迷惑行為が横行している[5]。

Twitter を管理する Twitter 社において、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、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。

2. 関連研究

中村らの研究[6]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94.70%の割合で正しく判定している。

一方、岩井らの研究[7]では、上記の結果を踏まえた上で、

^{†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

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、そのうち4つの判別項目を用いて判別的中率95.80%を持つスパム判別手法を確立している。また、判別の中率96.30%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。

筆者の前回の研究[8]では、パラメータデータにおける判別の中率96.00%、テストデータにおける判別の中率99.00%を示したが、これら2つの判別の中率には差があり、この差はアカウント情報によるデータ依存性が原因と考えられた。今回はアカウント情報によるデータ依存性の解消かつ従来手法の判別の中率94.70%[6]や95.80%[7]より優れた判別の中率を持つ判別手法の考案を目標とする。

3. 不正アカウントの定義

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、フィッシング、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、「アカウントへの反応(フォロワー、リツイート、いいねなど)を購入、販売または作為的に誇張しようとした場合」や「“フォロワー”、“リツイート”、“いいね”の獲得などをうたう第三者によるサービスまたはアプリケーションを利用したり、利用を助長したりした場合」など、14個の基準が設けられている[9]。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。特に、以下の項目に該当するアカウントを本研究では不正アカウントと定義する。

- ・有害、虚偽的または悪意のあるリンク(例: アフィリエイト、マルウェアサイト、クリックジャッキングページなどへのリンク)をツイートしている場合
- ・アカウントへの反応(フォロワー、リツイート、いいねなど)を購入、販売または作為的に誇張しようとしている場合
- ・フォロワー、リツイート、いいねの獲得などをうたう第三者によるサービスまたはアプリケーションの利用している場合
- ・1つのアカウントで重複した内容を複数投稿したりする場合、また重複またはよく似たアカウントを作成する場合

4. 不正アカウントの特徴分析から判別項目作成

4.1 アカウントの情報収集

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[10]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出来る。これらの情報から、Twitter社のアカウントの凍結条件と実際に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。

4.2 判別項目作成

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。

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社の判別基準
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/フォロワー数× 100	Twitter社の判別基準
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社の判別基準
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社の判別基準
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 割合	新たに考案した判別基準

4.3 判別分析について

本研究では正規アカウントと不正アカウントを判別する上で判別分析を用いる。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[11][12][13].

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.

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

ここで, a, b を判別係数という(図1参照).

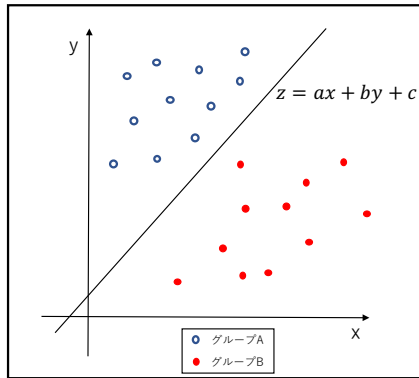


図1 線形判別分析

Figure 1 Linear discriminant analysis.

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 \bar{x}_A, \bar{y}_A , 標準偏差を s_{x_A}, s_{y_A} とし, グループBの平均値を \bar{x}_B, \bar{y}_B , 標準偏差を s_{x_B}, s_{y_B} とする. また, $s_{x_A y_A} =$

$$\frac{1}{n} \sum_{i=1}^n (x_{A_i} - \bar{x}_A)(y_{A_i} - \bar{y}_A), \quad s_{x_B y_B} = \frac{1}{n} \sum_{i=1}^n (x_{B_i} - \bar{x}_B)(y_{B_i} - \bar{y}_B)$$

は共分散とする. 点 P からグループAの重心までのマハラノビスの距離 D_A を式(2)によって求め, 点 P からグ

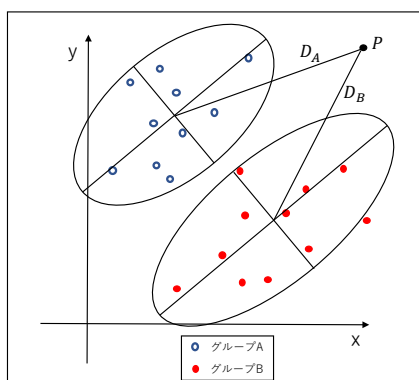


図2 マハラノビスの距離による判別分析

Figure 2 Discriminant analysis by Mahalanobis distance.

ループBの重心までのマハラノビスの距離 D_B を式(3)によって求める(図2参照).

$$D_A^2 = \begin{pmatrix} s_{x_A}^2 & s_{x_A y_A} \\ s_{x_A y_A} & s_{y_A}^2 \end{pmatrix}^{-1} \begin{pmatrix} x - \bar{x}_A \\ y - \bar{y}_A \end{pmatrix} \cdot \dots \cdot (2)$$

$$D_B^2 = \begin{pmatrix} s_{x_B}^2 & s_{x_B y_B} \\ s_{x_B y_B} & s_{y_B}^2 \end{pmatrix}^{-1} \begin{pmatrix} x - \bar{x}_B \\ y - \bar{y}_B \end{pmatrix} \cdot \dots \cdot (3)$$

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

5. パラメーターフィッティングによる判別条件の決定

5.1 使用する判別項目と条件の設定

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, 2変数と3変数におけるスコアを求め, 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. 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. なお, スコアとは判別分析を用いて, 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.

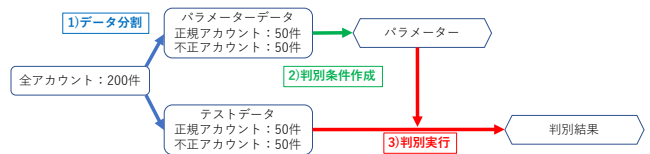


図3 判別手法

Figure 3 Discriminant method.

- 1) 正規アカウント50件と不正アカウント50件の計100件をパラメーターデータとする.
- 2) パラメーターデータを各判別項目の判別分析にかけ, スコアを算出する. 使用する判別項目から判別条件を設け, パラメーターを作成する.
- 3) テストデータに対し, 判別分析の判別率を算出する.

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スコア0.9200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.

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4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

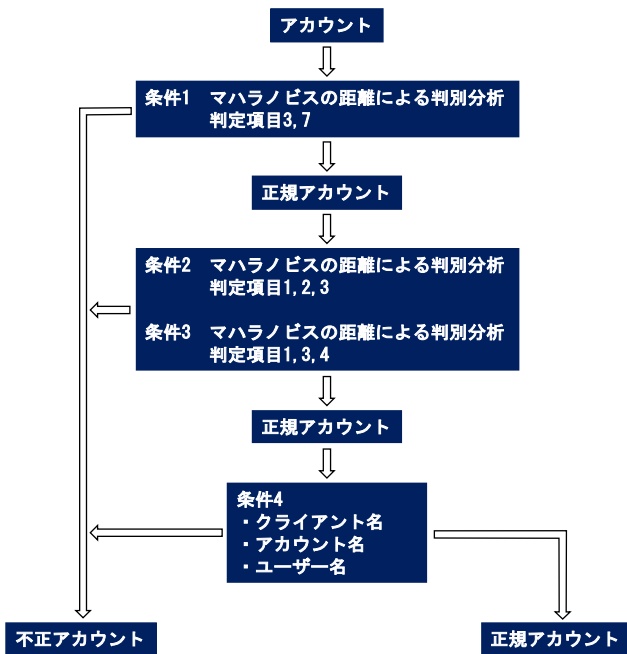


図4 解析フロー
 Figure 4 Analysis flow.

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。

5.2 パラメーターフィッティングによる判別結果

以上の図 4 の解析フローによるパラメーターデータの結果を表 2 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 を踏まえた判別の中率は 96.00% を示した。

表 2 パラメーターデータに対する判別結果

Table 2 Result of parameter 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)	1.000 (50)	0.000 (0)
	不正アカウント (50)	0.0800 (4)	0.9200 (46)
判別の中率		96.00%	

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、図 4 の解析フローによる結果を表 3 に示す。表 3 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 を踏まえた判別の中率は 99.00% であった。従来手法[8]では 95.80% であったのに対して、本手法では 3.20% 判別の中率が上がった。

表 3 テストデータに対する判別結果

Table 3 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)	1.000 (50)	0.000 (0)
	不正アカウント (50)	0.020 (1)	0.9800 (49)
判別の中率		99.00%	

6. 解析フローの再設定

筆者の前回の研究手法[8]ではパラメーターデータにおける判別の中率 96.00%、テストデータにおける判別の中率 99.00% を示したが、これら 2 つの判別の中率には差があり、この差はアカウント情報によるデータ依存性が原因だと考えられる。また、従来手法[6]と比べると判別の中率を算出する過程が複雑である。そのため本節では、アカウント情報によるデータ依存性の解消と解析手法の単純化について検証する。

6.1 データ分割

判別の中率のアカウント情報によるデータ依存性を防ぐため、データ分割を行い、判別手法を繰り返すことにより判別の中率の平均と標準偏差を算出する (図 5)。

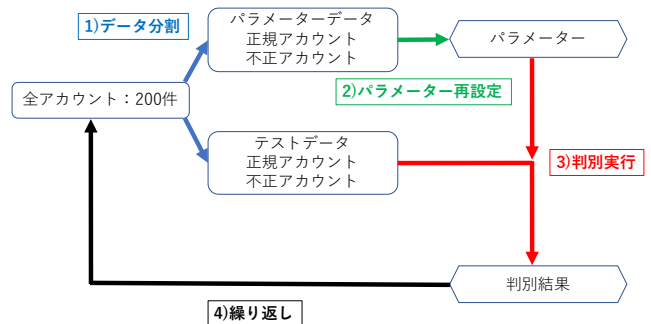


図 5 判別手法

Figure 5 Discriminant method.

- 1) 正規アカウント 100 件と不正アカウント 100 件の計 200 件をパラメーターデータとテストデータにデータ分割する。
- 2) パラメーターデータを各判別条件の判別分析にかけ、スコアを算出し、パラメーターを再設定する。
- 3) テストデータに対し、パラメーターを用いて、スコアを算出する。
- 4) この繰り返しを行うことで、スコアの平均と標準偏差を算出する。

データ分割は以下の 3 パターンで構成されており、

- i) パラメーターデータを正規アカウント 40 件、不正アカウント 40 件の計 80 件。テストデータを正規アカウント 40 件、不正アカウント 40 件の計 80 件
- ii) パラメーターデータを正規アカウント 50 件、不正アカウント 50 件の計 100 件。テストデータを正規アカウント 40 件、不正アカウント 40 件の計 80 件、
- iii) パラメーターデータを正規アカウント 60 件、不正アカウント 60 件の計 120 件。テストデータを正規アカウント 40 件、不正アカウント 40 件の計 80 件とする。

また、i)～iii)のそれぞれにおいてアカウントをランダムに選出し、4.2 節で決定した条件 1～4 における判別の中率を算出する。この試行を各 36 回計 108 回行い、スコアの平均と標準偏差を算出する (表 4 参照)。

表 4 データ分割
 Table 4 Data division.

データ分割	i	ii	iii
パラメーターデータアカウント数 (件)	80 (正規 : 不正 40 : 40)	100 (正規 : 不正 50 : 50)	120 (正規 : 不正 60 : 60)
テストデータアカウント数 (件)	80 (正規 : 不正 40 : 40)	80 (正規 : 不正 40 : 40)	80 (正規 : 不正 40 : 40)
試行回数	36	36	36

6.2 条件の再設定

まず、パラメーターデータに対して条件 1～4 を様々組み合わせ、スコアの平均と標準偏差を算出した (表 5)。条件 1, 4 を用いた場合の判別の中率が他の条件を組み合わせた場合の判別の中率より最も高かった。

表 6 に条件 1, 4 を組み合わせた場合のパラメーターデータに対する判別結果を示す。スコアの平均と標準偏差は

表 5 スコアの平均と標準偏差

Table 5 Score of average and standard deviation.

	条件 1		条件 1, 4		条件 2	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
パラメーターデータ	0.9638	0.0100	0.9841	0.0065	0.9486	0.0121
	条件 2, 4		条件 3		条件 3, 4	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
パラメーターデータ	0.9793	0.0056	0.9533	0.0127	0.9791	0.0063
	条件 1, 2, 4		条件 1, 3, 4			
	平均	標準偏差	平均	標準偏差		
パラメーターデータ	0.9826	0.0058	0.9825	0.0061		
	条件 1, 2, 3, 4		条件 2, 3, 4			
	平均	標準偏差	平均	標準偏差		
パラメーターデータ	0.9817	0.0059	0.9802	0.0058		

正規アカウントに対して、 1.000 ± 0.000 を示し、不正アカウントに対して、 0.9693 ± 0.0106 を示した。つまり、全体のアカウントに対して判別の中率は $98.41 \pm 0.65\%$ を示した。なお、条件 1, 4 を用いて解析フローの再設定を行った。

表 6 パラメーターデータに対する判別の中率の平均と標準偏差
 Table 6 Discriminant hit ratio of average and standard deviation in parameter data.

		スコア	
		平均	標準偏差
パラメーターデータ	正規アカウント	1.000	0.000
	不正アカウント	0.9693	0.0106
	全体	0.9841	0.0065
判別の中率の平均と標準偏差		98.41±0.65%	

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 6 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: 条件 1 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。

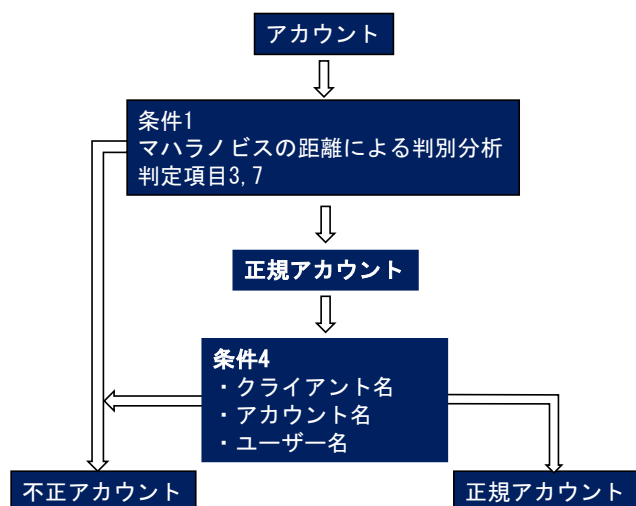


図 6 解析フロー
Figure 6 Analysis flow.

6.3 テストデータにおける判別結果

次に、テストデータに対して図の解析フローを踏まえた判別結果を表 7 に示す。スコアの平均と標準偏差は正規アカウントに対して、 1.000 ± 0.000 を示し、不正アカウントに対して、 0.9697 ± 0.0129 を示した。つまり、全体のアカウントに対して条件 1, 4 を踏まえた判別率の中率は $98.41 \pm 0.74\%$ を示した。

表 7 テストデータに対する判別率の中率の平均と標準偏差
Table 7 Discriminant hit ratio of average and standard deviation in test data

		スコア	
		平均	標準偏差
パラメーター データ	正規アカウント	1.000	0.000
	不正アカウント	0.9697	0.0129
	全体	0.9841	0.0074
判別率の中率の平均と標準偏差		98.41±0.74%	

以上をまとめると、解析フローの再設定を行った結果、パラメーターデータに対する判別率の中率は $98.41 \pm 0.65\%$ を、テストデータに対する判別率の中率は $98.41 \pm 0.74\%$ を示した。これらの判別結果に差はほとんどなく、従来手法[8]の 95.80% よりも $2.61 \pm 0.74\%$ 高い判別率の中率が得られた。

また、図 6 の解析フローは図 4 の解析フローに比べると条件が 2 つ減り、かつ手順が 1 つ減ったことにより、目標としていた解析手法の単純化が可能になった。

7. まとめ

本研究では、一般の利用者が容易に得ることのできるアカウント情報から不正アカウントの特徴を分析した。それらの特徴分析結果を踏まえた上で、判別項目を作成し、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検証した。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対して判別率の中率 $98.41 \pm 0.65\%$ 、テストデータに対して従来手法[7]の 95.80% よりも $2.61 \pm 0.74\%$ 高い判別率の中率 $98.41 \pm 0.74\%$ を示した。

今後の課題として、今回対象とした Twitter ユーザーは 200 件のみでありデータ数が少ないので、今後アカウント数を増やして解析する必要がある。

また、誤判別した不正アカウントの多くは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。そのため、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して、判別手法のさらなる高精度化を検討する必要がある。

さらに、アカウント情報取得の自動化と判別分析の自動化を実現することがあげられる。これにより、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として実用化できると考えられる。

参考文献

- [1] “ツイッター：国内利用者 4 5 0 0 万人に”。
<https://mainichi.jp/articles/20171027/k00/00e/040/370000c>, (参照 2018-02-09).
- [2] “Twitter の利用目的は「趣味の情報収集」 / LINE ユーザーの 6 割が毎日利用”。<https://markezine.jp/article/detail/20452>, (参照 2018-02-09).
- [3] “【最新】Twitter 乗っ取りの対処法, スпам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://applio.com/twitter-app-revoke-by-smartphone-iphone-android-pc>, (参照 2018-02-09).
- [4] “Twitter, 偽のフォロワーを売るビジネスに「対処する」 New York Times が「フォロワー工場」ビジネスを報じたことから”。
<http://nlab.itmedia.co.jp/nl/articles/1801/30/news079.html>, (参照 2018-02-09).
- [5] “Twitter で不適切画像をリプライする迷惑行為横行 Twitter 「対策行う」とコメント”。
<http://nlab.itmedia.co.jp/nl/articles/1801/15/news066.html>, (参照 2018-02-09).
- [6] 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.
- [7] 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.
- [8] 我妻拓哉, 吉村博幸. 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, 情報処理学会 情報システムと社会環境研究報告, 2017, 2017-IS-142(1), pp. 1-6.
- [9] “Twitter ルール”。<https://support.twitter.com/articles/253501>, (参照 2018-02-09).
- [10] “whatwi?”. <http://ja.whatwi.com>, (参照 2018-02-09).
- [11] 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.
- [12] 小西貞則. 多変量解析入門—線形から非線形へ, 岩波書店, 2010.
- [13] 内田治. すぐ分かる EXCEL による多変量解析 [第 2 版], 東京図書, 2000.