

回帰分析による Twitter 上のユーザ行動と プロフィール画像に関する調査

富永 登夢^a 土方 嘉徳^b

大阪大学大学院 基礎工学研究科

a) tominaga@hlab.sys.es.osaka-u.ac.jp b) hijkata@sys.es.osaka-u.ac.jp

概要 Twitter 上で、プロフィール画像はユーザの視覚的シンボルと見なされる。また、Twitter 上のユーザ行動は、自己表現と捉えられる。本研究では、日本人の Twitter ユーザを対象に、ユーザ行動とプロフィール画像の関係について調査と分析を行う。我々は、対象ユーザ群のプロフィール画像を 13 種類に分類し、カテゴリ間でのユーザ行動の違いをロジスティック回帰分析により調査した。その結果、ユーザ行動とプロフィール画像のカテゴリについて、いくつか有意な相関関係を発見した（“本人複数”カテゴリ（複数人が映っている写真を用いるユーザ群）に属するユーザは、返信機能を多用する）。さらに、我々は、ユーザ行動を基に、対象ユーザのプロフィール画像のカテゴリを予測するロジスティック回帰モデルを構築した。構築されたモデルの性能を評価したところ、“本人複数”カテゴリに対するモデルが最も良い結果を示した（F 値：0.789）。

キーワード Twitter, プロフィール画像, ユーザ行動, ロジスティック回帰

1 はじめに

Twitter は世界で幅広く利用されるソーシャルメディアの 1 つである。他のソーシャルメディアと異なり、投稿における文字数制限や、相手の許可を必要としないフォロー関係の構築が可能といった特徴を有する。これらの特徴を考慮して、ユーザのツイートやフォロー関係からユーザのモデル化を試みる研究 [1, 12] やユーザ間での情報拡散を分析する研究 [8, 10] がなされてきた。

我々は、ツイートの内容やフォロー関係だけでなく、画像に関する情報もまたユーザの特徴に影響を与えるだろうと想定している。本研究では特に、ユーザプロフィールに記載される写真や画像、すなわちプロフィール画像に注目する。これらは、ユーザプロフィールにおいて設定されるが、タイムラインにも表示されるため、ツイートの投稿者を特定する役割を果たす。プロフィール画像は、Twitter 上で自分が誰であるかを視覚的に表現する写真であるため、ユーザの内面的特徴と相関すると思われる。これは過去の研究 [9, 4] においても議論されている。本研究において、内面的特徴とは、心理的傾向や嗜好、利用目的などのユーザの内的特性とする。プロフィール画像と同様に、Twitter 上でのユーザ行動もまたユーザの内面的特徴の影響を受けると考えられる。例えば、他のユーザとのコミュニケーションを好むユーザは返信機能を多く使い、自身を宣伝する目的を持つユーザは URL を多用し、より詳細な情報を発信している可能性がある。また、近年の研究により、ユーザの心理的傾向はユーザ行動に影響を与えることが示されている [3, 11]。つまり我々は、Twitter 上のプロフィール画

像とユーザ行動が、ユーザの内面的特徴を媒介として、互いに相関することを着想した。

しかしながら、プロフィール画像とユーザ行動の関係はこれまでに明らかにされてこなかった。本研究では、この関係を調査するため、Twitter ユーザのプロフィール画像を人手で 13 種類に分類した。この分類は、プロフィール画像に写る対象に基づいて行われている。我々は、カテゴリ間でユーザ行動の傾向の違いを回帰分析により明らかにした。その結果、ユーザ行動とプロフィール画像のカテゴリについて、いくつか有意な相関を発見した（“本人複数”カテゴリ（複数人が映っている写真を用いるユーザ群）に属するユーザは、返信機能を多用する）。さらに、我々は、ユーザ行動を基に、対象ユーザがあるカテゴリに属するか否かを予測するロジスティック回帰モデルを構築した。全てのカテゴリの中で、“本人複数”カテゴリに属する確率を予測するモデルが最も良い性能を示した（F 値：0.789）。本研究は、Twitter ユーザのプロフィール画像とユーザ行動の関係の網羅的な理解と、ユーザ行動に基づいて同じカテゴリの画像を利用するユーザの発見に貢献すると考えられる。

本稿の構成は以下のとおりである。最初に関連研究について述べ、データの収集について言及する。次いで、分析手法とその結果、および考察を示す。最後に、本研究の制約について議論した後、まとめを述べる。

2 関連研究

2.1 ソーシャルメディアにおけるプロフィール画像

これまでに、プロフィール画像によるユーザの自己表現について調査した研究が存在する。Zhao と Jiang は、アメリカと中国の大学生を対象に、どのようなプロファ

イル画像が利用されているかを調べた [15]。この研究では、アメリカの大学生は中国の大学生より友人と写った写真を利用する傾向にあると結論付けている。Kapidzic は、Facebook 上の 288 名の大学生に対して、プロフィール画像を選択する動機とナルシズムの関係を分析した [6]。ここでは、ナルシズムが強いほど自身の魅力や個性が伝わる写真を選択しやすいと述べられている。

性別や年齢とプロフィール画像の関係に関する研究もいくつか存在する。Siibak は、プロフィール画像を通じた若い男性の自己表現について分析している [13]。15 歳から 28 歳の男性が性的で恋愛に適した写真を用いて、自身の魅力を表現することに努めているとこの研究は述べている。Siibak と Hernwell が、スウェーデンとエストニアの 10 代の若者のプロフィール画像を調べたところ、12 歳から 16 歳の女性は、プロフィール画像を加工し自身をより魅力的に表現していることが分かった [14]。

2.2 本研究の立ち位置

過去の研究とは異なり、本研究では Twitter 上のユーザのプロフィール画像とユーザ行動の関係を明らかにする。具体的に、我々は (1) リツイート、リプライ、URL、ハッシュタグの添付を含む様々なユーザ行動について調査し、(2) 統計的解析によりプロフィール画像のカテゴリ間に存在するユーザ行動の傾向の差異を定量的に明らかにし、(3) ユーザ行動を基に、対象ユーザがあるカテゴリに属するか否かを予測するモデルを構築した。

3 プロファイル画像のカテゴリ

3.1 定義

我々は、プロフィール画像に写っている対象に基づき、以下の 13 種類のカテゴリに分類した：本人一人、自画像、顔隠し、本人複数、他人、文字、ロゴ、オタク、キャラクタ、動物、オブジェ、景色、たまご。表 1 に各カテゴリに関する説明を示す。また、図 1 には、それぞれのカテゴリにおけるプロフィール画像の例を示す。

表 1 に示されるカテゴリ群は著者の経験に基づいて構築されたものであるため、以下の疑問に答えなければならない：Q1. “複数人によるプロフィール画像の分類はどの程度一致するのか”，Q2. “これらのカテゴリは Twitter 上に存在するプロフィール画像の種類をどの程度網羅できているか”。本稿では、Q1 を一致性の課題、Q2 を網羅性の課題と呼ぶ。次節では、これらの課題の検証について述べる。

3.2 一致性の検証

ここでは、一致性の課題の検証を行う。我々は、以下の手順に従ってこの検証を実施した。

まず、我々は 4 人の実験協力者を集めた。彼らは本検証に当たって報酬を受け取る。次に、Twitter Sample

表 1 プロファイル画像のカテゴリ

カテゴリ	説明
本人一人	本人の写真
自画像	本人のイラストや似顔絵
顔隠し	部分的に顔を隠している人の写真
本人複数	複数人で写っている写真
他人	ユーザ以外の人が映っている写真（有名人や子供）
文字	文字から構成される画像
ロゴ	ロゴの画像
オタク	アニメや漫画に登場する美少女キャラクターの画像
キャラクタ	上記のオタク以外の有名なキャラクターの画像
動物	鳥や猫、犬などの動物の写真
オブジェ	ボール、バイク、コップなどの所有物の写真
景色	自然風景の写真
たまご	Twitter のデフォルト画像

API を用いて、無作為に日本人の Twitter ユーザを収集した。収集期間は 2013 年 9 月 18 日から 10 月 17 日とした。その結果、20,833,001 のツイートと 4,394,542 のユニークユーザが得られた。このユニークユーザの中から、我々は無作為に 300 ユーザを抽出した。抽出されたユーザを対象に、300 のプロフィール画像を取得し、本検証に用いた。最後に、実験協力者に 300 のプロフィール画像を 13 種類のカテゴリに分類するように依頼した。

分類された結果に対して、我々は Fleiss のカッパ係数を用いて一致率を検証した。得られた係数の値は 0.704 となり、4 人の実験協力者の分類結果は実質的に一致しているという結果となった。

3.3 網羅性の検証

次に、網羅性の課題の検証を行う。

我々はまず、4 人の実験協力者を集めた。彼らは本検証に当たって報酬を受け取っている。続いて、4 人の実験協力者に、それぞれ異なる 300 のプロフィール画像を分類するように依頼した。これらのプロフィール画像は、3.2 節で述べた 4M のユーザから無作為に抽出されたユーザのものである。分類を依頼する際、我々は「その他」というカテゴリを用意した。最終的に、カテゴリのラベルを付与された 1200 のプロフィール画像が得られた。

本検証の対象ユーザの中に、アカウントを凍結された、また、サービスを退会したユーザが存在した。これらのユーザの情報は取得できないため、我々は彼らを「凍結・退会」ユーザとして分類し、検証の対象から除いた。1200 の対象ユーザのうち、113 ユーザが「凍結・



図1 各カテゴリにおけるプロフィール画像の例

退会」として認められた。従って、残存する 1087 ユーザを本検証の対象とする。

1087 ユーザの中で、「その他」に分類されたユーザ数は 93 であった。ここで、13 種類のカテゴリがどの程度 Twitter 上のプロフィール画像を網羅できているかを網羅性能力と呼ぶ。網羅性能力は、全体の対象プロフィール画像のうち、13 種類のいずれかに分類されたプロフィール画像の割合と定義する。本検証における 13 種類のカテゴリの網羅性能力は、0.914 $((1087 - 93)/1087)$ となった。我々が構築した 13 種類のカテゴリは、Twitter 上のプロフィール画像を十分網羅できると言える。

4 ユーザ行動

本研究で対象とするユーザ行動について述べる。我々は、ユーザのフォロー、ツイート、リツイート、リプライ、URL、ハッシュタグに関する行動に注目し、以下の 7 つの特徴量を定義する。

フォロー。 ユーザのフォロー数を *followees*、フォロワー数を *followers* とし、フォロー行動に関する特徴量 *FF* を以下のように定義する。

$$FF = \frac{followers}{followees} \quad (1)$$

この値が高いユーザは、情報を大きく拡散できる能力を有すると言える。

ツイート。 ユーザが Twitter を利用してきた期間を *UDays*、ユーザの全ツイート数を *Tweets_{All}* とし、ツイート行動に関する特徴量 *R_{tw}* を以下のように定義する。

$$R_{tw} = \frac{Tweets_{All}}{UDays} \quad (2)$$

これは、1 日に投稿する平均ツイート数となる。

リツイートある対象期間におけるユーザのリツイート数を *Retweets*、ツイート数を *Tweets* とする。リツイ

ート行動に関する特徴量 *R_{rt}* を以下のように定義する。

$$R_{rt} = \frac{Retweets}{Tweets} \quad (3)$$

これは、ユーザの情報拡散に対する積極性に関わる指標である。また、ユーザのリツイートされたツイートの数を *Retweeted* とし、リツイート行動に関する特徴量 *R_{rted}* を以下のように定義する。

$$R_{rted} = \frac{Retweeted}{Tweets} \quad (4)$$

この特徴量から、情報拡散されるような投稿をどれだけ行っているかを推定できる。

リプライ。 ユーザのリプライ数を *Replies* とし、リプライ行動に関する特徴量 *R_{reply}* を以下に定義する。

$$R_{reply} = \frac{Replies}{Tweets} \quad (5)$$

ユーザがどの程度コミュニケーションを好むかを表現する指標である。

ハッシュタグ。 ユーザのハッシュタグを含むツイート数を *Hashtags* とし、ハッシュタグ行動に関する特徴量 *R_{hash}* を以下に定義する。

$$R_{hash} = \frac{Hashtags}{Tweets} \quad (6)$$

ハッシュタグを利用することで、自身のツイートが検索されやすくなる。ユーザが、自身のツイートを通じて、フォロワー以外のユーザとコミュニケーションをとる機会を増加させようとする意図が反映される指標である。

URL。 ユーザの URL を含むツイート数を *URLs* とし、URL 行動に関する特徴量 *R_{url}* を以下に定義する。

$$R_{url} = \frac{URLs}{Tweets} \quad (7)$$

この値が高いユーザは、より詳細な情報を積極的にフォロワーに届けようとしていると解釈できる。

上述されたユーザ行動に関する 7 つの特徴量を行動特性と呼び、本研究の分析対象とする。

表2 各カテゴリにおける行動特性の中央値と標準偏差

カテゴリ	FF		R_{tw}		R_{rt}		R_{rted}		R_{reply}		R_{hash}		R_{url}	
	中央値	標準偏差	中央値	標準偏差	中央値	標準偏差	中央値	標準偏差	中央値	標準偏差	中央値	標準偏差	中央値	標準偏差
本人一人	1.137	1187.609	3.483	15.043	0.066	0.110	0.049	0.248	0.432	0.265	0.000	0.141	0.099	0.280
自画像	1.026	77.705	4.550	21.747	0.060	0.151	0.102	0.268	0.398	0.260	0.015	0.159	0.102	0.268
顔隠し	0.990	11881.358	9.067	29.614	0.082	0.148	0.021	0.153	0.469	0.283	0.018	0.150	0.025	0.198
本人複数	0.994	3.926	7.867	8.414	0.074	0.153	0.042	0.605	0.673	0.217	0.000	0.024	0.000	0.077
他人	0.964	1720.484	5.650	22.367	0.041	0.191	0.030	0.219	0.500	0.299	0.002	0.073	0.019	0.156
文字	1.189	8748.564	2.783	21.475	0.010	0.212	0.210	0.346	0.058	0.278	0.000	0.224	0.191	0.378
絵	2.000	1497.370	2.200	13.991	0.002	0.176	0.316	0.345	0.073	0.293	0.004	0.263	0.377	0.386
オタク	1.060	371.456	26.033	36.187	0.062	0.160	0.026	7.366	0.353	0.247	0.032	5.245	0.026	0.142
キャラクター	0.963	8.965	5.200	20.734	0.039	0.172	0.019	1.338	0.468	0.281	0.005	0.154	0.022	0.210
動物	0.901	378.935	2.317	22.860	0.039	0.255	0.000	0.195	0.367	0.294	0.000	0.199	0.032	0.283
オブジェ	0.917	88.759	6.167	28.211	0.025	0.177	0.007	0.141	0.427	0.283	0.000	0.089	0.013	0.210
景色	0.888	180.888	2.583	19.062	0.027	0.218	0.001	0.227	0.302	0.313	0.000	0.151	0.013	0.276
たまご	0.657	395.017	0.300	22.865	0.000	0.138	0.000	0.208	0.000	0.282	0.000	0.203	0.000	0.326

5 手法

5.1 データの収集

7つの行動特性を算出するために、我々は以下のデータを要する：*followers*, *followees*, *Tweets_{All}*, *UDays*, *Tweets*, *Retweets*, *Retweeted*, *Replies*, *Hashtags*, *URLs*. これらのデータを取得するため、まず、3.2, 3.3節とは別に、各カテゴリに対し100人のユーザを人手で収集した（合計1300ユーザ）. 次に、Twitter REST APIを用いて、対象ユーザから上述のデータを、2014年2月23日から3月25日の期間に取得した.

対象ユーザの全ての行動特性を算出したところ、表2に示す結果が得られた. 表2は、プロフィール画像の各カテゴリごとに、対象ユーザの行動特性の中央値と標準偏差がまとめられている. 全体的に、外れ値の影響により、標準偏差が中央値に対して比較的高い値を示している. 中央値から、各カテゴリに属するユーザの行動特性を大まかに把握できる. 一方で、定量的な解釈ができないため、表2のみから結論を導くことは困難である.

5.2 分析手法

定量的および統計的に行動特性とプロフィール画像の関係性を明らかにするために、我々はロジスティック回帰分析を行った. これにより、複数の説明変数が1つの独立変数に与える影響を分析する. 本研究の場合、あるカテゴリに属するか否かを二値の独立変数とし、7つの行動特性を説明変数とする. これより、あるカテゴリに属するために有意に機能する行動特性の効果を発見する.

回帰分析は以下の手順に従う. まず、対象ユーザがあるカテゴリに属しているか否かによって2群に分割した. そのカテゴリに属しているユーザには正(1)、属していないユーザには負(0)が割り当てられる. 次に、全ての行動特性を、平均0.0、分散1.0となるように正規化を行った. 最後に、各カテゴリに対してロジスティック回帰分析を行った.

6 結果と考察

6.1 行動特性の効果

表3に回帰分析の結果を示す. ここで、 β は偏回帰係数を示す. 偏回帰係数は、他の説明変数の値が一定だと仮定した時、対象の説明変数が独立変数に与える影響力の大きさと極性を表す.

本人一人カテゴリにおいて、 R_{tw} は負の影響を示している ($\beta = -0.51, p < 0.01$). Liuらの研究で議論されているが、プロフィール画像に1人の顔が映っているユーザは慎重であり、社会的に望まれるような行為を好む[9]. 従って、本人一人ユーザは、Twitter上で投稿する内容を熟慮して選択している可能性がある. また、本人一人ユーザは、リプライ行動 ($\beta = 0.32, p < 0.01$) やURLに関する行動 ($\beta = 0.35, p < 0.01$) をとる傾向にある. 上述したように、本人一人ユーザは社会的に期待されるように振る舞う[9]ため、他のユーザとコミュニケーションを取ったり、より詳細な情報をフォロワーに届けたりすることで、他者と社会的関係を構築・維持していると思われる.

本人一人ユーザとは対照的に、顔隠しユーザとオタクユーザは投稿の頻度が高い結果となった ($\beta_s = 0.31, 0.71, ps < 0.001$). Joinsonは、Computer-Mediated Communicationにおいて視覚的匿名性が担保されると、相手を知らない場合でも、自身を積極的に表現するようになること述べている[5]. この知見は、顔隠しユーザの投稿に対する積極性を説明するものであると言える. 表2に示される R_{tw} の中央値から、オタクユーザは1時間に少なくとも1回の投稿を行っていることが分かる. アニメや漫画がテレビで放送されている時、それらに関する内容で投稿されたツイートは“実況ツイート”と呼ばれる. 人手で確認したところ、オタクユーザのツイートは他カテゴリに属するユーザより短い傾向にあることが分かった. つまり、オタクユーザは、短い文章で

表3 ロジスティック回帰分析の結果 (β :偏回帰係数, $S.E.$:標準誤差, AIC :赤池情報量基準 [2]. $^*...p < 0.01$, $^{**}...p < 0.001$)

行動特性	本人一人		自画像		顔隠し		本人複数		他人		文字		ロゴ	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.
FF	-0.17	0.31	-1.75	2.11	0.18	0.09	-123.09	118.65	0.00	0.12	0.15	0.07	-0.12	0.18
R_{rw}	-0.51 *	0.18	-0.13	0.12	0.31 **	0.08	-0.43	0.18	-0.04	0.11	-0.07	0.12	-0.43	0.17
R_{rt}	-0.26	0.13	-0.06	0.12	0.03	0.11	0.09	0.13	0.04	0.11	0.56 **	0.13	0.15	0.14
R_{rted}	-0.32	0.60	-0.02	0.15	-3.03	1.55	0.02	0.12	-0.30	0.66	-0.02	0.12	0.01	0.10
R_{reply}	0.32 *	0.12	0.12	0.12	0.16	0.12	1.00 **	0.13	0.27	0.11	-0.91 **	0.18	-0.43 *	0.16
R_{url}	0.35 *	0.11	0.19	0.11	-0.22	0.17	-1.81 *	0.56	-0.45	0.19	0.36 **	0.09	0.63 **	0.09
R_{hash}	-1.97	1.19	-0.25	0.65	-0.06	0.28	-14.52	6.69	-1.67	1.57	-0.29	0.54	-0.02	0.14
AIC	682.13		712.53		670.41		555.84		695.03		644.68		627.18	

行動特性	オタク		キャラクタ		動物		オブジェ		景色		たまご	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.
FF	-0.16	0.53	-44.06	37.55	-0.24	0.68	-0.92	2.73	-0.94	1.56	-0.11	0.33
R_{rw}	0.71 **	0.08	-0.08	0.11	-0.22	0.13	0.18	0.09	-0.24	0.13	-0.03	0.11
R_{rt}	0.04	0.14	-0.14	0.13	0.26	0.10	0.00	0.12	0.26	0.11	-0.13	0.21
R_{rted}	0.58	0.27	0.05	0.06	-3.60 *	1.36	-4.21	1.72	-1.84	1.14	-4.71 *	1.45
R_{reply}	-0.03	0.14	0.25	0.11	-0.15	0.13	-0.02	0.12	-0.30	0.13	-0.94 **	0.16
R_{url}	-0.87 **	0.22	-0.13	0.14	0.15	0.12	-0.05	0.15	0.06	0.12	0.02	0.11
R_{hash}	3.50 **	0.76	-0.21	0.66	-0.10	0.31	-1.78	1.44	-1.12	1.04	-0.67	0.84
AIC	570.08		696.27		697.23		695.60		699.50		631.03	

数多くの実況ツイートを投稿していると推察される。オタクユーザが少ない単語で投稿することを考慮すると、URL リンクによる詳細な情報の共有に対して消極的である ($\beta = -0.87, p < 0.001$) ことは直感的に理解される。また、彼らはハッシュタグを多く利用する傾向にあり ($\beta = 3.50, p < 0.001$)、同じ興味を持つユーザに対して積極的に情報を公開していると思われる。

本人複数カテゴリに対しては、 R_{reply} が有意な正の効果を示した ($\beta = 1.00, p < 0.001$)。この結果は、彼らが他のユーザとのコミュニケーションを積極的に行う動機を有することを意味する。本人複数ユーザは友人や家族と共に写った写真を用いるため、コミュニケーションに対する動機が彼らのプロフィール画像と利用傾向の双方に影響を与えたと考えられる。

文字カテゴリとロゴカテゴリにおいて、 R_{url} は有意な正の効果を示した ($\beta_s = 0.36, 0.63, p_s < 0.001$)。また、どちらのカテゴリに対しても、 R_{reply} は負の効果を示した (文字: $\beta = -0.91, p < 0.001$; ロゴ: $\beta = 0.43, p < 0.01$)。これらの2つのカテゴリ内には、企業や組織のアカウントが数多く存在する。従って、URL リンクを積極的に活用することで詳細な情報をユーザに発信し、効果的な宣伝を行おうとしていると考えられる。また、これらのアカウントが、ある特定の個人との Twitter 上での会話を避ける傾向にあることは直感的である。

たまごユーザは新たな経験に対する積極性が低いと指摘されている [9]。加えて、彼らは Twitter の新規ユーザである可能性が高い。特に、 R_{rted} と R_{reply} に関しては有意な負の効果を示している ($\beta_s = -4.71, -0.94, p_s < 0.01, 0.001$)。彼らは繋がっているユーザ数が少ないため (フォロワー数, フォロワー数の中央値: 39.5, 21.0)、リツイートされる頻度や返信する頻度が低くなる。

現段階で、動物カテゴリに R_{rted} が有意に負の効果を持つ ($\beta = -3.60, p < 0.01$) 理由を解釈できない。今後、これらのユーザが動物の写真を選択する動機を詳細に調査し、行動特性との相関を分析する必要がある。

6.2 プロファイル画像カテゴリの予測

あるユーザの行動特性を基に、そのユーザがあるカテゴリに属するか否かを予測するロジスティック回帰モデルを構築した。この予測モデルはカテゴリごとに構築されるため、我々は13種類のモデルを構築した。

あるカテゴリに対する予測モデルの構築の手順を示す。まず、正解データセットと不正解データセットを作成する。正解データセットには、そのカテゴリに属する全てのユーザが含まれる (100 ユーザ)。その他のカテゴリに属する 1200 ユーザの中から無作為に 100 ユーザを選択し、それらを不正解データセットとした。次に、これらのデータセットに対して 10 分割交差検定により、モデルの評価を行った。モデルの性能の評価には F 値を用いた。F 値とは、精度と再現率の調和平均である。高ければ高いほど、モデルの性能は良いと解釈される。

図 2 は、各カテゴリに対する予測モデルの F 値の平均を示す。図 2 において、エラーバーは F 値に対する誤差範囲を示している。ここで、カテゴリを X とした時、そのカテゴリの予測モデルの F 値を F_X 、誤差範囲を SE_X とする。本人複数カテゴリに対する予測モデルが最も良い性能を示した ($F_{本人複数} = 0.789, SE_{本人複数} = 0.058$)。6.1 節で述べられたように、本人複数ユーザはリプライを頻繁に行うが、URL の利用に関しては消極的である。この利用傾向が他のどのカテゴリのユーザとも異なるため、予測モデルは本人複数ユーザを比較的容易に識別できたと言える。

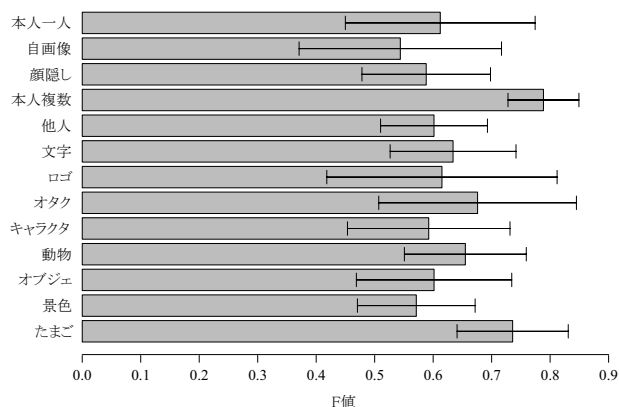


図2 各カテゴリに対する予測モデルのF値

本人複数カテゴリに次いで、たまごカテゴリとオタクカテゴリに対するモデルの性能が良い結果を示した ($F_{たまご} = 0.734$, $SE_{たまご} = 0.090$; $F_{オタク} = 0.676$, $SE_{オタク} = 0.160$)。たまごカテゴリには初心者が多く、活発に Twitter を利用していないという点で、彼らの利用傾向は独特である。特に、彼らのツイートはリツイートされない傾向にあり、返信機能を利用する頻度も低いことが分かっている (6.1 節)。オタクユーザは、他のカテゴリのユーザと比較して、投稿する頻度とハッシュタグを利用する頻度が高い (6.1 節)。

7 本研究の制約

本研究にはいくつかの制約が存在する。1つ目は、対象ユーザが日本人ユーザに限定されている点である。ソーシャルメディアの利用目的や利用傾向は文化的差異の影響を受けると指摘されている [7]。プロフィール画像の選択やその動機は文化によって異なるかもしれない。2つ目は、対象ユーザを収集する際、Twitter Sample API を利用した点である。この API は、リアルタイムで投稿されたツイートを取得するため、我々のデータセットには、積極的に投稿するユーザが偏って収集されている可能性がある。3つ目に、ロジスティック回帰分析は、相関関係を発見するものであり、因果関係について議論できない。行動特性とプロフィール画像の因果関係を理解するためには、今後さらなる統計的分析が必要である。

8 結論

本稿で、我々は最初に、13 種類の Twitter 上のプロフィール画像のカテゴリを定義した。回帰分析を用いて、カテゴリ間で 7 種類の行動特性の違いについて検証した。その結果、いくつか有意な相関を発見した。

さらに我々は、ユーザの行動特性を基にそのユーザがあるカテゴリに属するか否かを予測するロジスティック回帰モデルを構築した。その結果、本人複数カテゴリに対する予測モデルの性能が最良であった。

我々は、本稿が Twitter ユーザのプロフィール画像と行動特性の関係の理解に貢献することを期待する。今後は、文化的背景によって利用されるプロフィール画像がどの程度異なるかを検証する予定である。

参考文献

- [1] Abel, F., Gao, Q., Houben, G. J. and Tao, K.: Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web, *The Semantic Web: Research and Applications*, pp. 375–389. 2011.
- [2] Akaike, H. Information theory and an extension of the maximum likelihood principle, in *Selected Papers of Hirotugu Akaike*, pp. 199–213, Springer, 1998.
- [3] Golbeck, J., Robles, C. and Turner, K.: Predicting Personality with Social Media, in *Proc. of ACM CHI EA '11*, pp. 253–262, 2011.
- [4] Hum, N. J., Chamberlin, P. E., Hambricht, B. L., Portwood, A. C., Schat, A. C. and Bevan, J. L.: A picture is worth a thousand words: A content analysis of Facebook profile photographs, *Computers in Human Behavior* **27**(5), pp. 1828–1833. 2011.
- [5] Joinson, A. N. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity, *European Journal of Social Psychology* **31**(2), pp. 177–192. 2001.
- [6] Kapidzic, S. Narcissism as a predictor of motivations behind Facebook profile picture selection., *Cyberpsychology, Behavior and Social Networking* **16**(1), pp. 14–19. 2013.
- [7] Kim, Y., Sohn, D. and Choi, S. M.: Cultural difference in motivations for using social network sites: A comparative study of American and Korean college students, *Computers in Human Behavior* **27**(1), pp. 365–372. 2011.
- [8] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a Social Network or a News Media?, in *Proc. of ACM WWW'10*, pp. 591–600, 2010.
- [9] Liu, L., Preotiuc-Pietro, D., Samani, Z. R., Moghadam, M. E. and Ungar, L.: Analyzing Personality through Social Media Profile Picture Choice, in *Proc. of ICWSM'16*, 2016.
- [10] Pan, Y., Cong, F., Chen, K. and Yu, Y.: Diffusion-aware personalized social update recommendation, in *Proc. of ACM RecSys'13*, pp. 69–76, 2013.
- [11] Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M. and Crowcroft, J.: The personality of popular facebook users, in *Proc. of ACM CSCW'12*, pp. 955–964, 2012.
- [12] Ramage, D., Dumais, S. and Liebling, D.: Characterizing Microblogs with Topic Models, in *Proc. of ICWSM'10*, pp. 130–137, 2010.
- [13] Siibak, A. Constructing masculinity on a social networking site, *Young* **18**(4), pp. 403–425. 2010.
- [14] Siibak, A. and Hernwall, P.: 'Looking like my favourite Barbie' – Online Gender Construction of Tween Girls in Estonia and in Sweden, *Studies of Transition States and Societies* **3**(2), pp. 57–68. 2011.
- [15] Zhao, C. and Jiang, G.: Cultural differences on visual self-presentation through social networking site profile images, in *Proc. of ACM CHI'11*, pp. 1129–1132, 2011.