

Twitter ユーザの人格特性の変化に関する調査と分析

富永 登夢^{†,a} 土方 嘉徳^{†,b}

† 大阪大学大学院 基礎工学研究科 †† 関西学院大学商学部

a) tominaga@hlab.sys.es.osaka-u.ac.jp b) contact@soc-research.org

概要 本研究では、Twitter ユーザを対象に、人格特性の変化と他ユーザとのインタラクションの関係を明らかにした。ここで対象としている人格特性は、Five Factor Model として知られる開放性、誠実性、外向性、調和性、神経症傾向の5次元で構成される。我々はまず、IBM の提供するモデルを用いて、Twitter ユーザのツイートから過去と現在の人格特性を推定した。次に、過去と現在の間に、Twitter ユーザの人格特性がどの程度変化しているかを確認した。さらに、上述の期間において、Twitter ユーザが受け取った返信、引用、そして登録に関するデータを取得し、それらと人格特性の変化量の関係を非線形重回帰分析により明らかにした。この分析から、他ユーザとのインタラクションが人格特性の変化と有意な相関関係を示すことが分かった。最後に、これらの結果を踏まえた考察と今後の展望について述べた。

キーワード Twitter, 人格特性, インタラクション, 非線形重回帰分析

1 背景

Twitter や Facebook, Instagram などのオンラインソーシャルネットワーク (以下, OSN) の出現により、大規模かつ明示的に供給されるパーソナルデータを活用した研究事例が数多く登場している [24, 16, 1]。例えば、OSN 上で獲得されるユーザに関する様々な情報を用いて、ユーザの人格特性の予測を試みる研究がこれまでに行われてきた [17, 5, 6, 18, 11]。近年では、IBM が、これまでの研究成果 [21, 15] を基盤とし、OSN 上のユーザの活動から彼らの人格特性を推定する技術を確立している。近年の人格特性に関する OSN 研究は、ユーザ個人の人格特性を加味したサービスの質の向上を目指して、OSN 上のユーザに関する情報を用いて彼らの心理的特徴の推定に注力している。

しかしながら、OSN 上で自身の発信した情報に対して向けられた他者からの多様な反応によって、ユーザの人格特性は変化する可能性がある。一般的には、人格構造は普遍的であると支持されてきた [14, 4] が、ある個人の人格特性の時間に伴う変化の存在も指摘されている [20, 19, 22]。

本研究の目的は、OSN ユーザの人格特性の変化と OSN 上での他者からの反応との関係を解明することである。これは、“人の性格が他者との交流や経験を通じて変化するのか?” という学術的示唆に富んだ疑問に対して1つの解を与えるものであり、ユーザの精神的状態とその要因を把握するという文脈において、心理学的にも重要な課題だと考えている。さらに、OSN 上のデータの利用により、心理学や行動学で伝統的に用いられる質問紙調査では達成できなかったユーザ規模での調査が可能と

なることも本研究の利点である。

我々は、Twitter を対象の OSN として取り上げ、約 1600 ユーザを対象に IBM の提供する推定技術を利用して彼らの人格特性を取得した。また、他のユーザからの反応に関するデータも取得し、人格特性との関係を分析した。その結果、人格特性の変化と Twitter における他者の反応は有意に相関することが分かった。

本稿は以下の構成に従う。まず、データの収集方法について言及したのちに、取得したデータの分析手法について説明する。得られた結果とそれに対する考察を述べ、最後に今後の展望を示す。

2 データ収集

2.1 対象ユーザの選別

我々は、設定言語を英語にしているユーザを、Twitter Sampling API を用いて 2016 年 9 月 3 日から 10 月 7 日の期間に無作為に取得した。Twitter ユーザの利用目的や利用形態は、ユーザの文化的背景の影響を大きく受けることが知られている [10, 23]。我々の研究は、人格特性や経験など個人のパーソナルデータを詳細に分析するものであり、ユーザの背後に存在する文化的背景は無視できない影響となりうる。そこで我々は、この影響を抑制するため、対象ユーザを英語話者ユーザに限定している。この手続きにより約 1100 万ユーザが取得された。

次に、約 1100 万ユーザの中から、投稿数が 2800 件以上 3200 件以下であるユーザを抽出した。ツイートを取得するための Twitter REST API は、あるユーザに対して最大 3200 件の投稿の取得を許可している。従って、対象ユーザの投稿数を 3200 件以下に限定すると、Twitter の利用を開始してから今日に至るまでの利用期間における利用履歴を観察できるという利点を持つ。こ

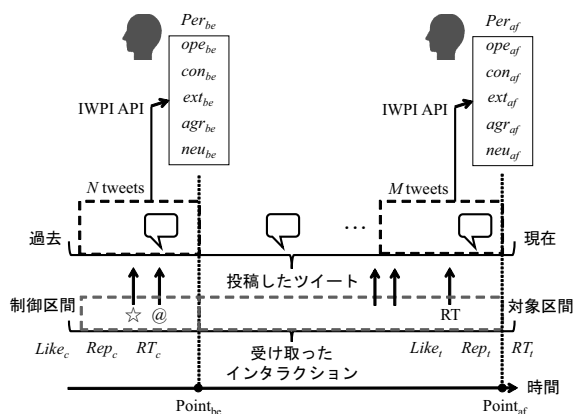


図 1: データの概観

の利点を考慮する理由は、人格特性の変化を観察する際に、この利用区間をユーザ間で統一しない場合、ユーザ毎に受け取る経験や交流の性質が大きく異なる可能性があるためである。例えば、利用し始めて間もない初心者と、長い期間使ってきた経験豊富なユーザとでは、周囲から受け取ってきたフィードバックの総量が異なるために、ある1つのフィードバックがユーザに対して有する重要性は大きく異なると思われる。この影響を最小限に抑えるため、利用し始めてから今日に至るまでの利用履歴が取得できるユーザに対象ユーザを限定する。投稿数が2800件以上であるという制約条件は、ユーザによって投稿数を大きく変動させないために設けた。

さらに、Twitterの利用期間が大きく異ならない範囲に属するユーザを抽出した。本研究は人格特性の変化を観察するため、人格特性の変化に要する期間をユーザ間で大きく異ならない範囲に設定する必要がある。そこで、上述したユーザ群における利用期間に対するユーザ数の分布を確認したところ、950日から1050日の利用期間においてなだらかな分布が確認された。そこで、この利用期間に含まれるユーザを抽出したところ、2510ユーザが確認された。我々は、データ取得の段階において、獲得されるデータに関する条件によってさらにユーザを選別している。その選別方法については次節にて述べる。

2.2 データセットの獲得

2.2.1 人格特性

人格特性については、我々はFive Factor Model [13]を採用する。これは、人格特性を以下の5次元の性質によって構成されるものとして定義されている：神経症傾向、外向性、開放性、調和性、誠実性。伝統的かつ一般的には、質問紙調査によって各次元の程度を取得する [4] が、近年ではSNS上の情報からユーザの人格特性が予測可能となっている。本研究では、その予測技術の1つであるIBM Watson Personality Insights API [8] (以下、IWPI API) を用いて、ユーザの人格特性を獲得す

る。この技術は、Twitterユーザを対象に、質問紙調査によって取得される人格特性 [4] を投稿の内容から推定するように学習したものである。IBMによれば、英語話者を対象にした調査では、この推定技術による推定結果と質問紙調査の人格特性は、弱いものの正の相関を示している ($r = 0.33$) [9]。この推定技術の構築プロセスと示された統計的結果から、これは本来的な人格特性を抽出する技術であると我々は解釈している。

我々は第一に、対象ユーザ群の全ツイートをTwitter REST APIによって取得した。次に、過去と現在のある時点における人格特性をIWPI APIにより推定する。IWPI APIは、人格特性の予測にユーザのツイート群を利用する。そこで、最初のN件 (図1左中) と最後のM件 (図1右中) をそれぞれ過去と現在の人格特性の予測に用いるツイート群とする。ここで、各ユーザに対してそれぞれのツイート群が1200単語有するようにNとMを設定している。URL及び画像のリンクはツイート群から除外している。また、1200単語有するように各ツイート群を構築した理由は、IWPI APIによって推奨されている必要単語数が1200と設定されているためである。それぞれのツイート群を T_{be} , T_{af} と呼ぶ。ここで、2510ユーザ (2.1節参照) の中から、それぞれのツイート群が投稿された期間が1週間以上1年未満であるユーザを抽出した。この制約によって最終的に選別された1636ユーザを分析の対象ユーザとする。最後に、各ツイート群からIWPI APIを通じてユーザの人格特性を獲得する。 T_{be} によって得られた人格特性を Per_{be} (図1左上)、 T_{af} による人格特性を Per_{af} (図1右上) と定義する。以降では、神経症傾向、外向性、開放性、誠実性、調和性の各性格因子はそれぞれ neu_{be} , ext_{be} , ope_{be} , con_{be} , agr_{be} 及び neu_{af} , ext_{af} , ope_{af} , con_{af} , agr_{af} と表現する。これらの値は、IWPI APIの仕様により、0から1の実数値を示す。

過去の研究 [15, 21] と同様に、IWPI APIはある時点の人格特性を推定するために、ユーザの過去の投稿に関する情報を用いる。従って、我々が取得した過去及び現在の人格特性である Per_{be} と Per_{af} はそれぞれ、 $Point_{be}$ (図1左下) と $Point_{af}$ (図1左下) における時点の人格を示す。

2.2.2 インタラクション

本研究では、Twitter上の引用、登録、及び返信をインタラクションと呼ぶ。具体的には、リツイート、Like、そしてリプライの機能を意味する。

あるユーザが受け取った引用と登録に関しては、取得した投稿に付随しているリツイート数、Like数を抽出した。ここで、そのユーザ本人のみに向けられた引用や登録を対象とするため、対象ユーザが引用リツイートし

た投稿はここでは除外している。次に、あるユーザの受け取った返信を取得するため、対象ユーザの全てのフォローの投稿の中から、その対象ユーザに向けられた返信となっているツイートを探し、その総数を取得した。

我々は、ユーザ間でこれまでに受け取ったインタラクションの頻度に対する認識が異なることを考慮し、 $Point_{be}$ までに受け取ったインタラクションの頻度と、 $Point_{be}$ から $Point_{af}$ までの期間に受け取ったインタラクションの頻度との関係を考慮する。これは、インタラクションの絶対量よりも、個人の経験における相対的な量の方が性格の変化に寄与すると考えられるからである。実際の計算手法は次章にて述べる。それぞれの期間を制御区間(図1左下赤枠)、対象区間(図1右下青枠)と定義する。制御区間 c と対象区間 t においてユーザが受け取った引用、登録、返信をそれぞれ、 RT_c , $Like_c$, Rep_c (図1左下)、及び RT_t , $Like_t$, Rep_t (図1右下)と表現する。

上記の手続きを1636の対象ユーザに対して、2016年12月21日から2017年1月12日の期間に実施した。

3 分析

3.1 人格特性の変化

対象ユーザの人格特性の変化を観察するため、全ての対象ユーザに対して $\Delta P = P_{af} - P_{be}$ ($P = neu, ext, ope, con, agr$)の計算を施す。各性格因子に対して、これらの値が正を示す場合は増加を、負を示す場合は減少を意味する。我々は、対象ユーザ群に対して、上記の値の頻度分布を確認する。

3.2 人格特性の変化とインタラクションの関係

人格特性の変化に与えるインタラクションの影響の大きさを解明するため、我々は各性格因子の変化量を目的変数とし、5つの重回帰分析を実施した。

説明変数は、対象区間において受け取ったインタラクションの相対的な頻度とする。ここで、あるユーザが対象区間において1日10件の引用を受け取ったと仮定する。このユーザが、制御区間において1日100件の引用を受け取っていた場合、対象区間における引用の頻度を少ないと感じる一方で、制御区間において1日1件の頻度でしか引用されていなければ、この対象区間における引用回数を多いと感じると思われる。このような個人の経験に基づくインタラクションに対する認識が人格特性の変化に関係すると考えられる。そこで、対象区間に受け取った登録、返信、引用の相対頻度 $\delta Like$, δRep , δRT を、以下の式によって定義する。

$$\delta F = \frac{F_t / Days_t}{F_t / Days_t + F_c / Days_c + 1} \quad (1)$$

$$F = Like, RT, Rep$$

ここで、 $Days_c$ と $Days_t$ は制御区間と対象区間の日数

表1: 人格特性の差分の分布に対する統計的データ

| 性格因子 | 平均 | 尖度 | 歪度 | 有意確率 |
|----------------|-------|-------|--------|--------|
| Δ_{neu} | 0.051 | 0.941 | 0.323 | <0.001 |
| Δ_{ext} | 0.002 | 0.707 | 0.119 | <0.001 |
| Δ_{ope} | 0.046 | 1.457 | -0.004 | <0.001 |
| Δ_{con} | 0.058 | 0.603 | 0.133 | <0.001 |
| Δ_{agr} | 0.073 | 0.194 | -0.002 | <0.05 |

である。分子は対象区間における受け取ったインタラクションの頻度、分母は対象区間と制御区間において受け取った頻度の和を示す。これらの値は0以上1未満の実数値を示し、対象区間に受け取ったインタラクションの頻度が高ければ1に、低ければ0に近づく。

本稿では、説明変数間の2次の交互作用を考慮した非線形重回帰分析を実施した。この際、交互作用による多重共線性の影響を考慮し、事前に全ての説明変数に対して平均0、分散1となるように正規化を行なった。また、AIC [2]に基づく変数増減法を用いた変数選択によるモデルの改良も実施した。

4 結果と考察

4.1 人格特性の変化

ここではまず、対象ユーザ全体の人格特性の変化の傾向を確認する。図2に、各性格因子の過去と現在の差分の分布を示す。ここでは、横軸に差分、縦軸にユーザ数(棒グラフ)とその密度曲線(曲線)が示されている。各性格因子の分布に対して統計的な特徴量を確認するため、我々はシャピロ・ウィルク検定を用いて正規性を検証すると共に、歪度と尖度を算出した。表1にその結果を示す。正規分布の場合、歪度、尖度は共に0を示す。正規分布に比べて、対象となる分布が大きい値に偏る場合に歪度は正の値となり、より尖った形状を示す場合に尖度は正の値となる。

シャピロ・ウィルク検定の結果から、全ての人格特性の分布は有意水準 $p = 0.05$ に対して正規性なしと判定される結果となった(表1)。これは、いずれの分布も尖度が高いことから(表1)、分布の中心にユーザが集中したためであると考えられる。各性格因子の平均値が0近辺であることを考慮すると、人格特性の変化の小さいユーザが多く見られたと結論づけられる。

一方で、図2では、各性格因子の変化量が0を中心に正負に均等に分布している様子が確認できる。これは、表1における歪度の低さからも言える。個々のユーザに注目すると、性格因子が増加したユーザも減少したユーザも一定数存在すると結論づけられる。図2aから、5種類の性格因子の中では、神経症傾向の歪度が最も大き

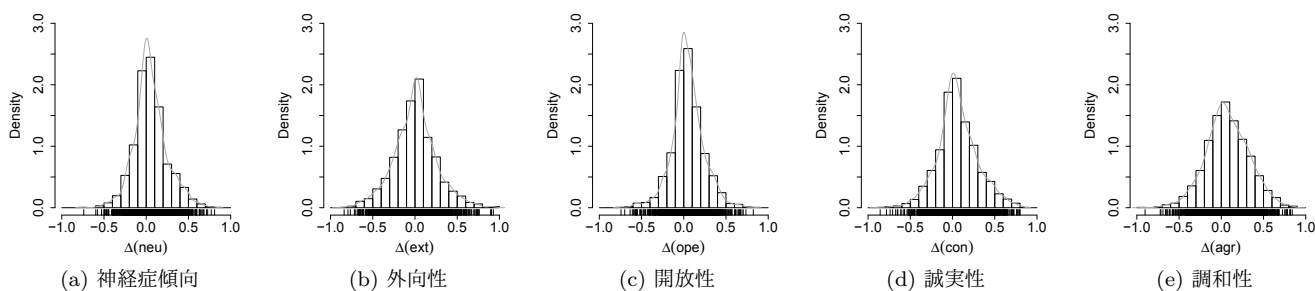


図 2: 各性格因子の過去と現在の差分に対するユーザ数の分布 (棒: 度数, 曲線: 密度曲線)

く正にわずかに偏っていることが分かる (Δneu : 0.323). Twitter ユーザは, わずかではあるが, サービスを利用し続けると神経症傾向が大きくなることを示す結果である.

4.2 人格特性の変化と交流及び経験の関係

本節では, 受け取ったインタラクションと人格特性の変化の相関関係を明らかにするために実施した重回帰分析の結果について述べる. それぞれの性格因子に対する重回帰分析の結果を, 表 2 に示す. この表では, 各性格因子を目的変数とした時の, インタラクションに関する説明変数 $\delta Like$, δRep , 及び δRT とそれらの交互作用の偏回帰係数と有意確率 ($*...p < 0.05$, $**...p < 0.01$) を示している. 偏回帰係数は, ある説明変数の目的変数に対する直接効果 (他の説明変数の値を一定にした時のその説明変数の効果) として解釈される. 以降では, 目的変数 y に対する説明変数 x の有する偏回帰係数を β_x^y と表現する. R^2 はモデルの適合率を示す決定係数であり, 0 から 1 の値を取る. 1 に近ければ近いほど良いモデルとして判断される. モデルの列に示される I と II は, それぞれの行において変数選択前, 変数選択後の結果を示す.

Δope は, $\delta Like$ から負の影響を受けることが分かった ($\beta_{\delta Like}^{\Delta ope} = -0.022*(II)$). つまり, 自身の投稿が他人に登録される経験が増加すると, 開放性が減少する傾向にあることを示している. 開放性とは, 多様性を好み, 知的好奇心を重視する性質を指す. 登録は, 周囲のユーザの関心や興味の範囲を知る重要な手掛かりとなるため, ユーザがフォローや閲覧者に合わせて投稿内容の話題を徐々に絞っていた可能性がある.

同様に, Δneu , Δcon と Δagr についても $\delta Like$ との負の関係が確認された ($\beta_{\delta Like}^{\Delta neu} = -0.055*(I), -0.053*(II)$; $\beta_{\delta Like}^{\Delta con} = -0.052*(I), -0.054*(II)$; $\beta_{\delta Like}^{\Delta agr} = -0.040*(I), -0.032*(II)$). 登録を受け取る頻度が増えることによって神経症傾向が減少する理由は, 登録の経験が増加することで自身が周囲のユーザに認められている [7] と認識するためであると考えられる. また, 誠実性や調和性の

減少もこの認識の影響を受けていると推察される. つまり, 登録の経験が増えることで深く思慮することなく投稿活動に従事したり, 登録を受け取る頻度が減ることにより周囲のユーザに対して協力的な姿勢を見せたりするようになると考えられる.

δRT は, Δneu に対して正の相関関係 ($\beta_{\delta RT}^{\Delta neu} = 0.023*(I), 0.024*(II)$), Δext に対して負の相関関係を示した ($\beta_{\delta RT}^{\Delta ext} = -0.028*(I), -0.021*(II)$). 引用の経験が増加すると, 自身の知らないより多くのユーザに自身の投稿が閲覧される機会が増える. 多くの場合, ユーザ自身で他者の引用を管理することはできない. 従って, 自身の意思に関係なく, 自身の投稿が多くのユーザに共有されてしまうことで, 周囲の反応に対して敏感になった (神経症傾向の増加) 可能性がある. また, 自身の投稿を閲覧するユーザを制御できないという認識により消極的になってしまった (外向性の減少) と考えられる. 一般的に, OSN 上で自身の投稿を見ている実際のユーザを把握することは困難であり [12], 引用の経験の増加によって自己提示戦略により困惑する [3] ようになったと言える.

Δext , Δcon , 及び Δagr は, いずれも δRep と正の相関関係を示した ($\beta_{\delta Rep}^{\Delta ext} = 0.025*(I), 0.026*(II)$; $\beta_{\delta Rep}^{\Delta con} = 0.017*(II)$; $\beta_{\delta Rep}^{\Delta agr} = 0.024*(I), 0.026*(II)$). ユーザは返信機能を使ってある特定の個人とコミュニケーションを図る. 引用や登録と比較すると, ユーザ間でより密なインタラクションが形成されていると言える. 従って, 返信を受け取る頻度と, 社交性 (外向性の要素), 几帳面さ (誠実性の要素), また良心的で協力的である性格 (調和性の要素) とが正に相関するという我々の結果は直感的に理解できる.

興味深いことに, Δcon のモデル I において, δRT と δRep はそれぞれ有意な相関関係を示していないが, それらの交互作用である $\delta RT * \delta Rep$ は, 正の相関係数を示した. これはモデル II においても確認される ($\beta_{\delta RT * \delta Rep}^{\Delta con} = 0.022*(I), 0.024*(II)$). これは, 引用と返信を受け取る経験がいずれも増加した場合, ユーザの誠実性が増加することを意味している. ユーザの投稿は, 通常自身の

表 2: 偏回帰係数と有意確率 (*... $p < 0.05$, **... $p < 0.01$)

| 性格因子 | モデル | $\delta Like$ | δRT | δRep | $\delta Like * \delta RT$ | $\delta RT * \delta Rep$ | $\delta Like * \delta Rep$ | R^2 |
|--------------|-----|---------------|-------------|--------------|---------------------------|--------------------------|----------------------------|-------|
| Δneu | I | -0.055** | 0.023* | 0.011 | -0.002 | 0.015 | -0.023** | 0.005 |
| | II | -0.053** | 0.024* | 0.011 | | 0.014 | -0.024** | 0.005 |
| Δext | I | 0.004 | -0.028* | 0.025** | -0.000 | -0.008 | 0.005 | 0.003 |
| | II | | -0.021* | 0.026** | | | | 0.005 |
| Δope | I | -0.023 | 0.002 | -0.001 | 0.001 | -0.001 | 0.009 | 0.001 |
| | II | -0.022* | | -0.000 | | | -0.008 | 0.002 |
| Δcon | I | -0.052** | 0.023 | 0.016 | 0.003 | 0.022* | -0.033** | 0.005 |
| | II | -0.054** | 0.023 | 0.017* | | 0.024** | -0.031** | 0.005 |
| Δagr | I | -0.040* | 0.011 | 0.024** | 0.003 | 0.005 | -0.021 | 0.004 |
| | II | -0.032* | | 0.026** | | | -0.013* | 0.006 |

フォローにのみ提示されるが、引用された場合は自身のフォロー以外のユーザに共有される。この際、その投稿に向けられた返信も紐づけられて共有されるため、自身に向けられた返信や会話がより多くのユーザに知られることになる。これにより、ユーザは軽率な行動を取りづらくなったと考えられる。また、知らない他人から返信を受け取ったために、ユーザがより慎重になった可能性もある。

一方で、 $\delta Like * \delta Rep$ は Δcon に対して負の相関関係を示した($\beta_{\delta Like * \delta Rep}^{\Delta con} = -0.033^{**}$ (I), -0.031^{**} (II))。引用とは異なり、基本的に登録や返信によってユーザの投稿が自身のフォロー以外に共有されることはない¹。つまり、これはフォローとの密なインタラクションの影響が反映されていると考えられる。特に、Twitter ユーザの Like は賞賛や理解を意味する [7] ため、それに伴う返信を受け取る機会の増加によって、ユーザにとってポジティブなコミュニケーションを形成するきっかけとなり、より気軽になれる(誠実性の減少)という結果は直感的に理解できる。このようなポジティブで密なコミュニケーションに関係する交互作用は、神経症傾向の減少や調和性の減少にもつながっていると考えられる($\beta_{\delta Like * \delta Rep}^{\Delta neu} = -0.023^{**}$ (I), -0.024^{**} (II); $\beta_{\delta Like * \delta Rep}^{\Delta agr} = -0.013^{*}$ (II))。

5 つの目的変数に対してそれぞれ重回帰分析を実施したが、いずれのモデルも非常に低い決定係数を示した(I: Δneu : 0.005, Δext : 0.003, Δope : 0.001, Δcon : 0.005, Δagr : 0.004)。また、変数選択によってモデルを改良したが、決定係数の十分な向上は確認されなかった(II: Δneu : 0.005, Δext : 0.005, Δope : 0.002, Δcon : 0.005, Δagr : 0.006)。我々の分析によって、説明変数と目的変数の間にいくつか有意な相関関係が発見されたが、目的変数を予測するモデルとしての性能は低いと言

¹近年では、Like や返信を行ったユーザの一部のフォローに対象ツイートが表示される機能が実装されている

える。本稿の分析では、登録、引用、返信の頻度にも注目している。予測モデルとしての性能を向上させるためには、インタラクションを通じて関わった相手との関係性やコミュニケーションの性質を詳細に分析し、説明変数として導入する必要がある。これらについては今後の検討課題とする。

5 本研究の制約と今後の課題

本研究の制約と限界について言及する。まず、対象ユーザの収集方法である。本研究では、投稿数や利用日数、使用言語によって対象ユーザを選別している。この対象ユーザ群が Twitter の代表的サンプルになっているか否かについては今後議論が必要である。ただし、我々は対象者を Twitter Sampling API で収集したユーザから抽出しており、代表性には最大限配慮している。また、これにより大規模な分析が可能となった点も強調する。

本研究は、Twitter ユーザを最も多く抱える英語圏のユーザに対象者を限定しているが、日本やブラジル、インドにも Twitter ユーザは多い。加えて、文化によって Twitter の利用形態は多様である [10, 23] ため、本研究の結果がいずれの文化圏でも同様の結果が得られるかは今後調査と分析が必要である。

我々の結果から、人格特性の変化とインタラクションの関係をより詳細に検証するためには、ユーザがインタラクションを通じて“誰と”・“どんな”やり取りを行なっているかを調査する必要性が示唆された。特に、返信機能を通じた会話の相手と性質について深く分析しなければならない。これらの調査と分析は、人格特性の変化の予測モデルの性能の向上に寄与すると期待される。

6 結論

本研究は、Twitter 上における人格特性の変化に対する受け取ったインタラクションの影響について分析し、

その関係の解明を試みるものである。我々は、既存の API を利用して Twitter 上の投稿から対象ユーザの人格特性を推定し、その変化を確認した。次に、対象ユーザがこれまでに受け取った登録や引用、返信と、推定された人格特性の変位の関係を重回帰分析によって明らかにした。我々の結果は、OSN における経験や交流が人格特性の変化に寄与することを示唆するものであり、学術的に重要な意義を含んでいる。今後は、残された課題の検討を進めると共に、サンプルの代表性やユーザの文化的背景に関する考察を深めていく。

参考文献

- [1] Abbar, S., Mejova, Y. and Weber, I.: You tweet what you eat: Studying food consumption through twitter, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3197–3206, ACM Press, 2015.
- [2] Akaike, H.: Information theory and an extension of the maximum likelihood principle, in *Proceedings of 2nd International Symposium on Information Theory*, pp. 267–281, Akademiai Kiado, 1973.
- [3] boyd, D. M.: Taken out of context: American teen sociality in networked publics, University of California, Berkeley, 2008.
- [4] Costa, P. T. and MacCrae, R. R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual, Psychological Assessment Resources, Incorporated, 1992.
- [5] Golbeck, J., Robles, C., Edmondson, M. and Turner, K.: Predicting personality from twitter, in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, pp. 149–156, IEEE, oct 2011.
- [6] Golbeck, J., Robles, C. and Turner, K.: Predicting personality with social media, in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, pp. 253–262, ACM, 2011.
- [7] Gorrell, G. and Bontcheva, K.: Classifying twitter favorites: Like, bookmark, or thanks?, *Journal of the Association for Information Science and Technology* **67**, pp. 17–25. jan 2016.
- [8] IBM, “IBM Watson Personality Insights.” <https://www.ibm.com/watson/services/personality-insights/>, 2017, visited.
- [9] IBM, “IBM Watson Personality Insights.” <https://console.bluemix.net/docs/services/personality-insights/science.html#science>, 2017, visited.
- [10] Jackson, L. A. and Wang, J. L.: Cultural differences in social networking site use: A comparative study of China and the United States, *Computers in Human Behavior* **29**(3), pp. 910–921. 2013.
- [11] Kosinski, M., Stillwell, D. and Graepel, T.: Private traits and attributes are predictable from digital records of human behavior, *Proceedings of the National Academy of Sciences* **110**(15), pp. 5802–5805. 2013.
- [12] Marwick, A. E. and boyd, D.: I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience, *New Media & Society* **13**(1), pp. 114–133. 2011.
- [13] McCrae, R. R. and Costa, P. T.: Validation of the Five-Factor Model of personality across instruments and observers, *Journal of Personality and Social Psychology* **52**(1), pp. 81–90. 1987.
- [14] Norman, W. T.: Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating, *The Journal of Abnormal and Social Psychology* **66**(6), pp. 574–583. 1963.
- [15] Plank, B. and Hovy, D.: Personality traits on twitter —or— How to get 1,500 personality tests in a week, in *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 92–98, 2015.
- [16] PreoĂciuc-Pietro, D., Volkova, S., Lampos, V., Bachrach, Y. and Aletras, N.: Studying user income through language, behaviour and affect in social media, *PloS one* **10**, p. e0138717. sep 2015.
- [17] Quercia, D., Kosinski, M., Stillwell, D. and Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter, in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, pp. 180–185, IEEE, oct 2011.
- [18] Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M. and Crowcroft, J.: The personality of popular facebook users, in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pp. 955–964, 2012.
- [19] Roberts, B. W., Caspi, A. and Moffitt, T. E.: The kids are alright: Growth and stability in personality development from adolescence to adulthood., *Journal of Personality and Social Psychology* **81**(4), pp. 670–683. 2001.
- [20] Robins, R. W., Fraley, R. C., Roberts, B. W. and Trzesniewski, K. H.: A longitudinal study of personality change in young adulthood, *Journal of Personality* **69**, pp. 617–640. aug 2001.
- [21] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dzurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P. and Ungar, L. H.: Personality, gender, and age in the language of social media: The open-vocabulary approach, *PLOS ONE* **8**(9), p. e73791. 2013.
- [22] Terracciano, A., McCrae, R. R. and Costa, P. T.: Intra-individual change in personality stability and age, *Journal of Research in Personality* **44**, pp. 31–37. feb 2010.
- [23] Trepte, S. and Masur, P. K.: Cultural differences in social media use, privacy, and self-disclosure, *Research report on a multicultural survey study*, tech. rep., 2016.
- [24] Volkova, S. and Bachrach, Y.: On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure, *Cyberpsychology, Behavior, and Social Networking* **18**(12), pp. 726–736. 2015.