

STaMP: 個人の性格や政治的立場等の多面的特性と紐づく SNS データの構築及び文章スタイルによる個人特性予測

福島 汐音¹ 仲田 明良¹ 佐橋 優人¹ 増川 哲太¹ 三輪 洋文²
野中 尚人² 木下翔太郎³ 岸本 泰士郎³ 五十嵐 彰⁴ 岡久 太郎¹ 狩野 芳伸¹
¹ 静岡大学 ² 学習院大学 ³ 慶應義塾大学 ⁴ 大阪大学

{sfukuhata, anakada, ysahasi, tmasukawa, kano}@kanolab.net
okahisa-taro@inf.shizuoka.ac.jp {hirofumi.miwa, naoto.nonaka}@gakushuin.ac.jp
{shotaro.kinoshita, tkishimoto}@keio.jp igarashi.a.hus@osaka-u.ac.jp

概要

言語は人と人のコミュニケーションにおける重要なツールであると同時に、その表現には性別・年齢・学歴・政治的立場・性格特性などの個人特性が表出すると考えられる。本研究では、そうした個人を形成する特性とその個人が作成したテキストとの関係を明らかにすべく、日本語 SNS ユーザを対象に、1千人規模のクラウドソーシングによる個人特性に関連したアンケート調査を行い、STaMP データセットを構築した。また、独自に構築した文章スタイルのベクトル表現を獲得した言語モデルを用いて文章からの個人特性の予測を行い、個人特性と文章の関係に関する分析及び考察を行った。このような個人特性と SNS 上のテキストを紐づけたデータはこれまで存在せず、オンラインテキストに現れるユーザ行動を形づくる要素の解明に貢献すると期待される。

1 はじめに

ソーシャルネットワーキングサービス (Social Networking Service, 以下 SNS) の普及に伴い、インターネットでは人々による言語コミュニケーションが行われている。現実世界同様に、ユーザという人物を構成する様々な特性が、その文章形成に影響を及ぼしていると推測される。

異なる個人特性を持つユーザが、同じ内容について言及した文章を記した場合、命題的な内容が同じであったとしてもその文章の様相には細やかな違いが生じることがある。この差異の仕組みを解明出来れば、ユーザの個人特性とユーザの記す文章の関係を明らかにできる。

これまでも、SNS 上のテキストデータを元にユーザの特性を予測する研究は盛んにおこなわれてきた [1][2] が、それらは、オンラインの自発的テキストデータのみと閉じているものであった。

そこで本研究ではクラウドソーシングによるアンケートを実施して、オフラインの個人の特性に関するデータと、代表的な SNS の一つである Twitter(現 X) の投稿であるツイートデータ (以下、旧名の Twitter およびツイートと表記する) を収集し、それらを紐づけた STaMP (Social Texts and Multifaceted Personality) データセットを作成した¹⁾。データセットにおける個人特性の内容は、性別・年齢・最終学歴・収入・政治的立場・パーソナリティと非常に多岐にわたり、人々のオンラインでの振る舞いの表出としてのテキストと、現実の人間の属性の相関や推測につながると期待される。

作成したデータセットを用いて各個人特性予測のための学習データ及びテストデータを作成し、文章スタイルを元に事前学習を行ったモデルに対してファインチューニングを行ったうえで、個人特性の予測性能を評価した。また、既存の事前学習済みモデルを同条件でファインチューニングした場合とで予測精度を比較し、文章と個人特性の関係性について分析を行った。

2 関連研究

StyleSimCSE StyleSimCSE モデル [3] は、事前学習済み BERT[4] である bert-base-japanese-v3²⁾ をベースモデルとし、学習データをツイート、ラベルを発

1) 本データセットは一般公開予定である。なお、公開予定のデータセットはアンケート調査で収集したアカウントと紐づけられていない匿名化されたものである。
2) <https://huggingface.co/cl-tohoku/bert-base-japanese-v3>

言者のアカウントとして、同一アカウントによる発言か否かを分類する対照学習により、類似する SimCSE [5] と異なりスタイルなどテキストの非内容的ベクトル表現獲得を行うモデルである。

移民問題と SNS 分析 移民を対象とした社会科学の領域では、近年 Twitter を活用して人々の移民に態度を推定する研究が目立ってきている。例えばツイートから人々の移民に対する態度を推定し、アメリカの一部地域における移民政策導入直後に態度が悪化していると報告した研究がある [6]。同様に SentiStrength という SNS のテキストから感情を推定するソフトを使い、メディアで移民がより多く取り上げられている地域だと、移民に対する態度が否定的であったとしている研究がある [7]。

パーソナリティ研究と政治行動 近年のパーソナリティ研究では、外向性 (Extraversion)、協調性 (Agreeableness)、勤勉性 (Conscientiousness)、神経症傾向 (Neuroticism)、開放性 (Openness) の 5 つの要素から捉える Big Five などのモデルが主流となっている [8]。Big Five を測定する手法も多く開発されており、5 つの要素を 10 項目で測定可能な Ten Item Personality Inventory (TIPI)[9] などがよく用いられている。また、パーソナリティのネガティブな特性に焦点をあてたものとして、ナルシシズム (Narcissism)、マキャベリアニズム (Machiavellianism)、サイコパシー (Psychopathy) の 3 特性を総称した Dark Triad も注目されている [10]。これらのパーソナリティ、性格特性の違いは、個人の政治的態度や政治参加の違いとも関連付けた検討が行われてきている [11]。

3 STaMP データセットの構築

我々はアクティブな Twitter アカウントを対象にクラウドソーシング調査を複数回実施してアンケート結果と投稿データを収集し、発言者と文章・個人特性が紐づいたデータセットを構築した。

調査は以下の 3 つの基準すべてを満たす Twitter ユーザを対象に実施した。

1. 18 歳以上である。
2. 日本国籍を持っている。
3. 2021 年 12 月 31 日以前から Twitter を公開アカウントで利用しており、平均して 3 日に 1 回以上投稿している。

結果、1,594 アカウントから回答を得た。そのう

ち 605 件は Lancers クラウドソーシングによって、989 件は Yahoo クラウドソーシングによって収集した。実施時期により、収入・DTDD-J・TIPI-J に関連する 9 つの項目は Yahoo クラウドソーシングデータにのみ設問が存在する。

アンケートの設問で回答者の Twitter アカウントを尋ね、各アカウントのツイートを Twitter API v2 の Academic Research アクセスにより取得した。

3.1 パーソナリティに関する統計情報

本研究では性格特性の評価として Big Five と Dark Triad に相当するものとして、日本語版 Dark Triad Dirty Dozen (以下, DTDD-J) [12], 日本語版 TIPI (以下, TIPI-J) [13] を用いた。

収集したパーソナリティに対応するアンケートの設問一覧及び、ラベル決定のための計算式は、付録 A.1 と付録 A.2 に記載する。アンケートの各パーソナリティカテゴリのうち、DTDD-J と TIPI-J の回答結果の統計情報を以下の表 1 に示す。これらのスコアは付録 A.2 の計算式をもとに計算されており、スコアが高いほど各傾向が強いことを表す。

表 1 DTDD-J/TIPI-J に関する回答の統計情報

カテゴリ	Mean	Std	Min	Max	95 %CI / Lower	95 %CI / Upper
サイコパシー	10.74	3.03	4	20	10.54	10.93
ナルシシズム	10.49	3.79	4	20	10.24	10.73
マキャベリアニズム	9.36	3.81	4	20	9.11	9.61
外向性	7.13	2.79	2	14	6.95	7.31
協調性	9.31	2.47	2	14	9.15	9.47
勤勉性	6.6	2.69	2	14	6.42	6.77
神経症傾向	8.98	2.68	2	14	6.42	6.77
開放性	8.29	2.62	2	14	6.42	6.77

3.2 その他の設問

DTDD-J, TIPI-J の他に、政治的立場や個人特性に関する設問として、性別や年収、最終学歴、イデオロギーに関する設問を作成し調査を実施した。

4 個人特性の予測

アカウント単位での、性別・年齢・年収・最終学歴・イデオロギー・移民に対する態度・防衛費増額に対する態度・憲法改正に対する態度の 8 項目と DTDD-J と TIPI-J に含まれる 8 項目の計 16 項目についての分類タスクの予測精度の評価を行った。3 節のデータセットから各個人特性カテゴリに対応した学習用データセットを作成し、以下で述べる前処理を適用したうえで、事前学習済みモデルに対して作成したデータセットを用いてファインチューニングを行った。具体的には、我々が構築した StyleSimCSE の利用を提案手

法とし、京都大学が公開している日本語 BERT の deberta-v2-large-japanese³⁾を比較対象として用いた。これら事前学習済み BERT の最終層に、タスクに応じて多値分類用の全結合層を追加してファインチューニングを行った。ファインチューニング時の各種パラメータ設定は、付録 A.3 に記載する。

4.1 ツイートデータの前処理

投稿テキストと個人特性の関係を調査するため、引用リツイートやプロモツイート、ツイート内のメンション部分を除外した。文字数が 15 文字未満のツイートは内容に乏しい可能性があるため、100 文字以上のツイートは StyleSimSCE の事前学習設定に合わせ、除外した。総ツイート数が 80 件を超えるアカウントは直近 80 件のみを対象とし、80 件未満のアカウントは除外して、アカウントあたりのツイート数を 80 件にそろえた。結果、670 アカウント 53,600 ツイートとなった。

4.2 個人特性のラベリング

アンケートの各回答項目に関してラベリングを行った。その結果と各ラベルの人数の内訳を表 2 に示す。項目ごとの回答数の合計が一致しないのは、一部設問に無回答のアカウントがあること、前述のように一部設問が追加される前の調査があることに起因する。

表 2 個人特性のラベリングと各ラベルの人数内訳

カテゴリ	ラベル 1	人数	ラベル 2	人数	ラベル 3	人数
性別	男性	370	女性	272	-	-
イデオロギー	リベラル派	73	保守派	98	中道派	496
移民に対する態度	肯定的	273	否定的	155	ニュートラル	238
防衛費増額に対する態度	肯定的	39	肯定的	195	ニュートラル	432
憲法改正に対する態度	否定的	61	肯定的	145	ニュートラル	461
年収	200 万円未満	180	600 万円以上	106	200 万円以上 600 万円未満	74
最終学歴	中学校卒・高等学校卒	223	高等専門学校卒・大学卒・大学院卒・短期大学卒	330	その他	92
年齢	30 歳未満	155	50 歳以上	144	30 歳以上 50 歳未満	367
サイコパシー	Negative	93	Positive	106	Neutral	204
ナルシズム	Negative	94	Positive	125	Neutral	184
マキャベリアニズム	Negative	81	Positive	65	Neutral	257
外向性	Negative	122	Positive	76	Neutral	205
協調性	Negative	118	Positive	53	Neutral	232
勤勉性	Negative	54	Positive	133	Neutral	216
神経症傾向	Negative	159	Positive	93	Neutral	151
開放性	Negative	93	Positive	86	Neutral	224

なお、表 2 内のラベルについて、ラベル 3 は各項目における中立的なラベルである。また、政治的立場に関するカテゴリでは、ラベル 1 がリベラル的な立場に、ラベル 2 が保守的な立場に対応している。⁴⁾

4.3 訓練時のデータ分割と評価指標

表 2 のうちラベル 1、ラベル 2 の個人特性を対象として、個人特性ごとに 3 分割交差検証による訓練と

評価を行った。個人特性ごとにラベル比率が 1:1 になるようオーバーサンプリングを実施してからデータ分割し、その際訓練と評価でアカウントの重複が起こらないようにした。訓練データの 20 % は訓練時の検証データとして用いた。Accuracy およびラベルごとの F1 値を評価指標とした。

推論時には、対象アカウントの各ツイートに対する予測確率⁵⁾が 0.7 以上の結果に対して、2 値であるラベルの比率を計算し、設定した閾値により最終的な推測結果のラベルを決定した。

DTDD-J と TIPI-J については、学習モデル個別に、各項目それぞれで閾値を決定した。交差検証においてラベルごとの性能 (F1 値) の平均が最も高くなる閾値を探索的手法で求め、個人特性項目ごとに設定した (付録 A.4)。⁶⁾ DTDD-J と TIPI-J 以外の項目については閾値を 0.5 とした。

5 個人特性予測の結果

各個人特性項目における予測性能の評価結果を表 3 に示す。性別、年齢、最終学歴、移民に対する態度、勤勉性、開放性の 6 項目の予測においては高い性能であった。⁷⁾

表 3 各個人特性の予測精度の比較結果

カテゴリ	StyleSimCSE			deberta-V2-large-japanese		
	Accuracy	F-Score (ラベル 1)	F-Score (ラベル 2)	Accuracy	F-Score (ラベル 1)	F-Score (ラベル 2)
性別	0.7999	0.8076	0.7898	0.6849	0.7359	0.5745
年齢	0.8745	0.8713	0.8775	0.8290	0.8145	0.8411
年収	0.5194	0.5395	0.4961	0.4722	0.6200	0.1364
最終学歴	0.6379	0.6487	0.6242	0.5606	0.5365	0.5504
イデオロギー	0.5464	0.5714	0.5138	0.5773	0.7007	0.2802
移民に対する態度	0.6237	0.6906	0.5263	0.5330	0.6480	0.2881
防衛費増額に対する態度	0.5385	0.1579	0.6806	0.4974	0	0.6643
憲法改正に対する態度	0.5897	0.4226	0.6763	0.3631	0.1436	0.4610
サイコパシー	0.5285	0.5943	0.4190	0.5715	0.5	0.6250
ナルシズム	0.5487	0.4589	0.6084	0.4146	0.4927	0.2383
マキャベリアニズム	0.5925	0.6206	0.5599	0.5741	0.4889	0.6349
外向性	0.5614	0.6073	0.5027	0.5488	0.5395	0.5193
協調性	0.5127	0.5055	0.5135	0.5207	0.5773	0.3723
勤勉性	0.6386	0.6477	0.6278	0.5985	0.6151	0.5768
神経症傾向	0.5	0.4364	0.5398	0.5346	0.5156	0.5360
開放性	0.6075	0.6065	0.6032	0.4570	0.4896	0.3900

6 分析と考察

表 3 の個人特性の予測精度の結果から、文章スタイルが個人特性の予測において有用であると推察される。そこで、訓練後の StyleSimCSE モデルに対して、予測が当たっておりかつ予測確率が 0.7 以上のツイートを対象として分析を行った。まず、評価データに対する「移民に対する態度」「最終学歴」ごとの文章ベクトルを t-SNE [14] を用いて 2 次元に削減したうえでプロットした。次に、密度の高い位置

- 5) 各ラベルに対する予測確率をソフトマックス関数を使用し合計値が 1 になるように調整した後の値
- 6) StyleSimCSE, deberta-V2-large-japanese とともに、訓練データにおける予測精度 (F-Score) が最も高くなるような値のばらつきは少なかった
- 7) 表内で値が 0 になっている箇所は、予測が片方のラベルに傾いたためで、学習データの不足が原因と考えられる。

3) <https://huggingface.co/ku-nlp/deberta-v2-large-japanese>

4) 本研究では、個人特性の違いの予測を目的としているため、各カテゴリで中立的な個人特性を持つアカウントについては除外した

に存在するツイートを代表値として抽出して、「移民に対する態度」「最終学歴」とツイートの関係性を分析した。⁸⁾

6.1 移民に対する態度とツイートの関係性

文章ベクトルの分布(図 1)から,「移民に対する態度」の違いによって文章ベクトルに違いがあることが分かる. すべて異なるアカウントによるツイートを列挙した表 4・表 5 からは,「移民に対する態度」が肯定的なユーザは絵文字を多用し,否定的なユーザは使用しない傾向があることが分かる. 政治的な態度と文章との関連を示唆している.

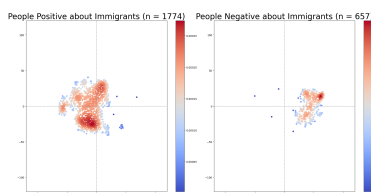


図1 ファインチューニング後の StyleSimCSE による「移民に対する態度」ごとの文章ベクトルを次元圧縮した図

表 4 「移民に対する態度」が肯定的なユーザのツイート例

ありがとうございます！ありがとうございます！
 魅力的な絵案！ありがとうございます！
 すばらしいです！お話を共有してくれてありがとう！
 いつもお楽しみ！お話を待たせて頂きます！ありがとうございます！
 一生懸命読んだ全巻の感想をありがとうございます！プロモーションビデオ編集中です！
 おお！感謝します！
 応援します！最近お忙しかれてませんが、引き続きよろしくおねがい！
 感謝します！唐は利用可能で、立てています！
 明日のセッションに向けて集まっていますので、これから毎週にオンライン生放送を行います！21:00の開始予定！参加のほどをぜひ、ご検討ください！
 感謝します！1月に登場します！(w)(w)！

表5 「移民に対する態度」が否定的なユーザのツイート例

ラスベガスモンスターゲームは、競馬場を模した、クリアルなゲームだ。

大手旅行会社の旅行がでたてしやうとこのコンゴに遊びが通りそう

ラムのラスベガスモンスターと戦って両面アライメントをたくさん使ったやうな量り感がすごいアクションゲームなんだね

ラッキーモンスターという出て来てもらってほしいし使えない

ファンタジーゲームの要素が出てくる。このファンタジーでRPG中々でたな

競馬RPGが素晴らしいやうにクリエーションの要素も含まれた。これだけあるか

チャットアプリもオンラインしながらプレイ中のRPGを一緒に遊べる方法はないかな

ゲームとかエンディング

競馬場を模したゲームがあるが、あのゲームなんかなってふくやう。ユーザーにはそういうメリットがあると思う

時空のバグも含まれるか。次の章の遊びを待つことに

6.2 最終学歴とツイートの関連性

同様の処理及び分析を「最終学歴」に対しても行った。文章ベクトルの分布(図2)から,「最終学歴」の違いによって文章ベクトルに違いがあることが分かる。ツイート例(表6・表7)では,「最終学歴」が中学校卒・高等学校卒であるユーザとそれ以外のユーザに分けて例示している。ここでも,絵文字の有無が違いとして見て取れ,高等教育を修了し

8) 表中の事例ツイートテキストは発信者が特定できない程度に加工したものである。

ていないユーザは絵文字が多く、発言の内容が日常生活に関するものが多いといった傾向がある。また高等教育を修了しているユーザは、発言の内容が政治に関するものが多い、スタイルが書き言葉的であるという傾向がある。

これから、最終学歴とその文章においても関連がある可能性が示唆される。なお、最終学歴と移民に対する態度の間には相関がみられなかった。

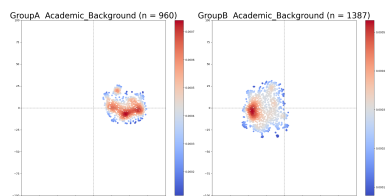


図2 ファインチューニング後の StyleSimCSE による「最終学歴」ごとの文章ベクトルを次元圧縮した図

表6 「最終学歴」の違いによるツイート例 (中学校卒, 高等学校卒)

[illegible]

表7 「最終学歴」の違いによるツイート例 (高等専門学校卒、大学卒、大学院卒)

[illegible]

7 おわりに

本研究では、クラウドソーシング調査による個人特性と紐づいた SNS 投稿データセットを構築し、実際に文章からの個人特性の予測を行うことで、文章と個人特性の関係性についての分析を行った。実験の結果、文章と個人特性には特性に関する明示的な内容以外の関係性がある可能性が高いことが分かった。今後の展望として、文章と個人特性の関係性のより詳細な分析や、その他の個人特性に関する予測精度の検証、分析結果を用いた言語モデルのアプリケーションへの応用、データ規模の拡大が挙げられる。

謝辞

本研究は JSPS 科研費 JP22H00804, JP21K18115, JP20K20509, JST AIP 加速課題 JPMJCR22U4, 及び セコム科学技術財団特定領域研究助成の支援を受けたものです。

参考文献

- [1] Ryohei Sasano Kosuke Yamada and Koichi Takeda. 「いいね」「シェア」をした投稿のテキストを利用した SNS ユーザの性格推定. 人工知能学会論文誌, 2020.
- [2] 石川晴貴, 荒澤孔明, 服部峻. 属性が性格に及ぼす作用の存在解明のための ツイッターユーザの属性抽出と性格ベクトル化. 電子情報通信学会技術研究報告 (IEICE Technical Report (Institute of Electronics, Information and Communication Engineers)), Vol. 119, pp. 7–12, 2020.
- [3] 仲田明良, 狩野芳伸. Style SimSCE: SNS ユーザ同一性に基づく対照学習によるスタイル類似性を捉えた文ベクトルの獲得. 言語処理学会年次大会発表論文集 (to appear), 2024.
- [4] Kenton Lee Jacob Devlin, Ming-Wei Chang and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. **arXiv preprint arXiv**, 2018.
- [5] Xingcheng Yao Tianyu Gao and Danqi Chen. SimCSE: Simple contrastive learning of sentence embeddings. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, p. 6894–6910, 2021.
- [6] René D. Flores. Do anti-immigrant laws shape public sentiment? a study of arizona’s sb 1070 using twitter data1. **American Journal of Sociology**, Vol. 123, No. 2, pp. 333–384, 2017.
- [7] Anastasia Menshikova and Frank van Tubergen. What drives anti-immigrant sentiments online? a novel approach using twitter. **European Sociological Review**, Vol. 38, pp. 694–706, 2022.
- [8] L. P. John, O. P. Naumann and C. J Soto. Paradigm shift to the integrative big five trait taxonomy. **Handbook of personality: Theory and research**, Vol. 3(2), pp. 114–158, 2008.
- [9] Rentfrow P. J. Gosling, S. D. and W. B. Swann Jr. A very brief measure of the big-five personality domains. **Journal of Research in personality**, Vol. 37(6), pp. 504–528, 2003.
- [10] D. L. Paulhus and K. M. Williams. The dark triad of personality: Narcissism, machiavellianism, and psychopathy. **Journal of Research in personality**, Vol. 36(6), pp. 556–563, 2002.
- [11] Pruyssers S. Chen, P. and J Blais. The dark side of politics: Participation and the dark triad. **Political Studies**, Vol. 69(3), pp. 577–601, 2021.
- [12] 田中圭介田村紋女, 増井啓太. 日本語版 dark triad dirty dozen (DTDD-J) 作成の試み. パーソナリティ研究, Vol. 24(1), pp. 26–37, 2015.
- [13] 小塩真司, 阿部晋吾. 日本語版 ten item personality inventory (TIPI-J) 作成の試み. パーソナリティ研究, Vol. 21(1), pp. 40–52, 2012.
- [14] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. **Journal of Machine Learning Research**, Vol. 9, pp. 2579–2605, 2008.

A 付録

A.1 各個人特性の判定に対応した設問の一覧

各個人特性の判定に対応した設問の一覧は以下の表 8 に示すとおりである。

設問 Q2-a については、「かなり増えたほうが良い」を 1、「かなり減ったほうがよい」を 5 として、5 段階での回答を求めている。設問 Q2-c ~ Q2-e, Q3, Q4 については、「そうは思わない」を 1, 「そう思う」を 5 として、5 段階での回答を求めている。設問 2-b は 0~100 の整数値での回答を求めている。ラベリングに際して、81 以上 100 以下を 1, 61 以上 80 未満を 2, 41 以上 60 未満を 3, 21 以上 40 未満を 4, 20 以下を 5 とした。設問 Q5 については、「200 万円未満」「200 万円以上 400 万円未満」「400 万円以上 600 万円未満」「600 万円以上 800 万円未満」「800 万円以上 1000 万円未満」「1000 万円以上 1500 万円未満」「1500 万円以上」の 7 段階での回答を求めている。また、設問 Q7-a~Q7-i の文頭にはそれぞれ以下の文言「以下の項目は、あなたにどの程度あてはまりますか。」「全くあてはまらない」を 1, 「非常にあてはまる」を 5 としたとき、最も近いと思うものを選んでください。」が含まれている。設問 Q8-a~Q8-j の文頭にはそれぞれ以下の文言「次のことばがあなた自身にどのくらい当てはまるかについて、」「全く違うと思う」から「強くそう思う」までのうちもっとも適切なものをそれぞれ選んでください。文章全体を総合的に見て、自分にどれだけ当てはまるかを評価してください。」が含まれている。設問 Q8-a Q8-j は 7 段階評価での回答を求めている。

表 8 各個人特性の判定に対応した設問の一覧

設問番号	設問	関連する個人特性
Q0	あなたの母国語なんですか	母国語
Q1-a	日本に定住しようと思っに来日する外国人は、もっと増えたいと思うですか、それとも減りたいと思うですか。	移民に対する態度
Q1-b	強い好意を 100% 強い反感を 0% とし、反感も反感も持たない場合は 50% としたとき、あなたの態度を数字で教えてください。 - 日本に定住している外国人	移民に対する態度
Q2-c	日本に定住しようと思っに来日する外国人が増えれば、犯罪発生率が低くなる	移民に対する態度
Q2-d	日本に定住しようと思っに来日する外国人は、日本から犯罪者を輩出する	移民に対する態度
Q2-e	一般的に言って、日本文化は、日本に定住しようと思っに来日する外国人によって徐々に変わってきている	移民に対する態度
Q3	「日本の防衛力はもっと強化すべきだ」という意見について、あなたのお友だちに近いものを選んでください	防衛費増額に対する態度
Q4	「憲法改正すべきだ」という意見について、あなたのお友だちに近いものを選んでください	憲法改正に対する態度
Q5	あなたの親戚全員の過去 1 年間の収入（税込み）は、次の中でどこに最も近いですか、 おおよそで結構ですのでお答えください	年収
Q6	あなたが就職に必要だった学校はどちらですか	最終学歴
Q7-a	私は他人から尊敬されても自分の思い通りにするところがある	マキヤベリアニズム
Q7-b	私は他人の心をだましたり嘘をついても自分の思い通りにするところがある	マキヤベリアニズム
Q7-c	私は他人の心にお金を使っても自分の思い通りにするところがある	マキヤベリアニズム
Q7-d	私は自分の利益の確保のために他人の心を manipulate するところがある	マキヤベリアニズム
Q7-e	私は、あまり自分のあやまちを認めることがない	サイコパス
Q7-f	私は、自分の利益の確保にはあまり関心がない	サイコパス
Q7-g	私は、どちらかというと周囲で人の評判を気にしない	サイコパス
Q7-h	私は、どちらかというと強い「聞くひねくれた人間である	サイコパス
Q7-i	私は、他人から尊敬されない人間に思われるのが嫌いだ	ナルシズム
Q7-j	私は、他人から尊敬してほしいと思いがちだ	ナルシズム
Q7-k	私は、高い身分や名声を手に入れたらいいと思う	ナルシズム
Q7-l	私は、他人からの尊敬は重要で価値がある	ナルシズム
Q8-a	私は自分自身のこと、近況で、外向的だと思う	外向性
Q8-b	私は自分自身のこと、ひかえめ、おとなしいと思う	外向性
Q8-c	私は自分自身のこと、他人に手紙を書き、もてなすを好むしやさいと思う	協調性
Q8-d	私は自分自身のこと、人に気をつかう、やさしい人間だと思う	協調性
Q8-e	私は自分自身のこと、しっかりしていて、自分に厳しいと思う	勤勉性
Q8-f	私は自分自身のこと、だらだらと、うそをついてると思う	勤勉性
Q8-g	私は自分自身のこと、心算で、うそを言わないと思う	神経症傾向
Q8-h	私は自分自身のこと、冷静で、気が配って定まっていると思う	神経症傾向
Q8-i	私は自分自身のこと、新しいことが好きで、変わった考えをもつと思う	開放性
Q8-j	私は自分自身のこと、想像力に富んだ、平凡な人間だと思う	開放性
Q9	あなたの年齢を教えてください	年齢

A.2 ラベル決定のための計算式と各ラベルの判定基準

ラベル決定のための計算式 [12][13] を以下に示す。

$$ImmigrantsScore = (Q2a + Q2b + Q2c + Q2d + Q2e) / 5$$

$$IdeologyScore = (ImmigrantsScore + Q3 + Q4) / 3$$

$$MachiavellianismScore = Q7a + Q7b + Q7c + Q7d$$

$$PsychopathyScore = Q7e + Q7f + Q7g + Q7h$$

$$NarcissismScore = Q7i + Q7j + Q7k + Q7l$$

$$ExtroversionScore = Q8a + (8 - Q8f)$$

$$CooperativenessScore = (8 - Q8b) + Q8g$$

$$DiligenceScore = Q8c + (8 - Q8h)$$

$$NeurologicalTendencyScore = Q8d + (8 - Q8i)$$

$$OpennessScore = Q8e + (8 - Q8j)$$

移民に対する態度は、ImmigrantsScore が 2.5 以下のものを否定的、3.5 以上のものを肯定的とした。防衛費増額に対する態度は、Q3 の回答が 1 のものを否定的、5 のものを肯定的とした。憲法改正に対する態度は、Q4 の回答が 1 のものを否定的、5 のものを肯定的とした。イデオロギーは、IdeologyScore が 2.5 以下のものをリベラル派、3.5 以上のものを保守派とした。

DTDD-J, TIPI-J に関するラベリングは、表 1 の結果と人数分布をもとにして閾値を設定し、閾値 1 より大きいものを Positive、閾値 2 より小さいものを Negative、それ以外を Neutral とした。各カテゴリと閾値の対応関係を以下の表 9 に示す。

表 9 DTDD-J/TIPI-J のラベル決定のための閾値

カテゴリ	閾値 1	閾値 2
Psychopathy	12	9
Narcissism	12	8
Machiavellianism	12	6
外向性	9	6
協調性	9	6
勤勉性	7	4
神経症傾向	11	9
開放性	10	7

A.3 ファインチューン時のパラメータ設定

各予測タスクにおけるファインチューニング時のパラメータ設定は以下の通りである。

lr = 2e-5, max epochs = 10 (early stopping により早期終了する可能性がある), batch size = 64

A.4 推論時の DTDD-J・TIPI-J の threshold

表 10 推論時の DTDD-J/TIPI-J の threshold

カテゴリ	StyleSimCSE	deberta-V2-large-japanese
サイコパス	0.43	0.44
ナルシズム	0.46	0.47
マキヤベリアニズム	0.48	0.49
外向性	0.42	0.46
協調性	0.45	0.46
勤勉性	0.45	0.43
神経症傾向	0.50	0.47
開放性	0.49	0.47