

# 実体験を考慮した性格特性推定システムの構築

安積佑和<sup>1</sup> 奥村紀之<sup>2</sup>  
<sup>1</sup> 大手前大学 <sup>2</sup> 大阪市立大学

## 概要

本研究では、ソーシャルメディアからの性格特性推定のために、被験者実験により収集した実体験と性格特性検査のための質問紙項目への回答（数値）を活用する手法について検討している。被験者実験においては、Big Five（主要5因子検査）への回答（数値）の他に、質問項目に該当する過去の実体験、または実体験がない場合は想定される状況について回答させた。収集した実体験例を基に Twitter のログから実体験例の拡張を行い、Word2Vec, Doc2Vec, Sentence BERT による性格特性の推定手法を構築している。

## 1 はじめに

2019 年末以降、世界的に新型コロナウイルス感染症が流行している。コロナ禍における自宅学習や在宅勤務などのライフスタイルの変化に伴い、うつや不安に関する報告も増加傾向にある [1, 2, 3]。これらの調査ではコロナ禍以前のうつや不安を訴える人の傾向と比較し、パンデミックにおけるその傾向はより強いものであると述べられている。そこで、ソーシャルメディアなどの発言履歴などから、在宅を余儀なくされている人々のうつや不安の兆候を抽出することができれば、メンタルヘルスの改善の一助となる可能性があると考えられる。

本研究では、ソーシャルメディアの記述を定量的に扱い、発言者のうつや不安の傾向を捉えるため、性格検査に基づく手法を検討している。本研究で対象とする性格特性の検査手法には、主要5因子検査 (Big Five) の一つである NEO-FFI を採用している。

藤倉ら [4] は音声対話やツイートの特徴が Big Five の各要素の分類に結びつくことを示している。那須川ら [5, 6] は内容語だけでなく、助詞やひらがなの使用状況も性格と関係し、何を表現するかだけでなく、どう表現するかにより性格が反映することを示している。我々は、より人間らしい性格推定を行う

ため、実際の被験者が質問紙項目へ回答する際の基準となった実体験に基づく手法を提案する。

## 2 関連研究

本研究では、性格特性の推定に Big Five を使用する。2.1 節では Big Five について述べ、2.2 節では本研究のベースラインとなる性格特性の抽出手法について解説する。

### 2.1 Big Five

性格特性を測る際の指標である Big Five は特性論 [7] の考え方であり、開放性 (Openness)、誠実性 (Conscientiousness)、外向性 (Extraversion)、調和性 (Agreeableness)、情緒不安定性 (Neuroticism) からなる5つの大きなパーソナリティ特性で人間を捉えようとする。

Big Five を測定するための尺度は Costa と McCrea による NEO-PI-R をはじめとして多数開発されている。日本語においては下仲ら [8] による NEO-PI-R 日本語版、村上ら [9] による主要5因子性格検査、和田 [10] による Big Five 尺度などがある。さらに、下仲ら [11] は240項目からなる NEO-PI-R の質問項目を60項目に短縮した日本版 NEO-FFI を開発している。これは各因子12項目合計60個の質問項目からなっている。各質問で最も当てはまらない0点から最も当てはまる4点の5段階で回答し、最終的には各因子最低0点から最高48点が算出される。本研究では、下仲らの日本版 NEO-FFI を使用して性格特性の検査、アンケート、および性格特性の推定を行う。

### 2.2 既存の性格推定手法

山下ら [12] の研究では、性格情報を反映した発話文生成を行っており、発話文生成に利用された性格推定手法をベースラインとして採用している。

この手法では本校で対象とする下仲らの日本版 NEO-FFI を用いており、各質問項目にキーワードを

設けている。性格推定をするテキスト中にそのキーワードが入っているか、またはキーワードが否定形に係っているかを見て得点の上下を行う。各因子が0点から48点で表されることから、中心の24点を基準とし、そこから点数の増減を行い、各因子の推定値を算出している。

### 3 実体験を考慮した性格特性推定手法

本研究では、統計的な情報による性格特性の抽出ではなく、人間の实体験に基づく推定手法を提案する。以下、性格推定の实体験の収集、実体験の拡張、性格特性推定の手法について述べる。

#### 3.1 実体験の収集アンケート

NEO-FFIの各質問項目に、人間であればどのように回答するかと考えた場合、被験者の実体験に基づいた判断が必要となる。ソーシャルメディアへの書き込みを投稿者の実体験とみなすことができれば、投稿者の判断基準に沿う性格特性の推定が可能となると考えられる。そこで、NEO-FFIの各質問項目に対し、0～4点の得点を判断するに至った具体的な実体験をアンケートにより収集する。

被験者17名に対し、日本版NEO-FFIの各質問項目への回答(数値)に加え、以下の3項目について回答させるアンケートを実施した。図1の例を示す。

- 各質問項目に対し0～4点の5段階の得点 (NEO-FFIで要求される各質問項目に対する0～4点の5段階の得点)
- 実体験の有無 (実体験があるかをはい、いいえの2択で回答)
- 実体験の記入 (自由記述で回答。複数の実体験がある場合は、体験毎に改行し文を区切る。実体験がない場合は、改行の数値を選んだ根拠を記入。根拠は可能な限り実生活に基づくものとする)

私は心配性ではない

最も当てはまらない 0 1 2 3 4 最も当てはまる

実体験の有無

はい ● いいえ ○

実体験の記入

忘れ物をしていないか心配になり、カバンの中を何度か確認する

図1 「私は心配性ではない」という質問項目に関する実体験の回答例。

アンケートを実施した結果、不適切な回答や、記入した点数と実体験の内容の不一致などの問題が確認された。以下に不適切な回答の例を示す。

- 自由記述で実体験とは言えない回答が含まれる。  
「特になし」や「あまり考えたことがなかった」など。

- 得点(回答の数値)と自由記述の実体験の内容の不一致がある。

例えば、「私は心配性ではない」という質問項目に対し、「面接やテストなど練習、復習していないと心配症になる」という回答に3点が付与されたり、「数は少ないが将来を見据えた決断時は心配面は比較的薄い。」という回答に1点が付与されるなど、実体験と数値の関係が反転しているものがある。(質問紙の逆転項目<sup>1)</sup>に多い。)

- 実体験の自由記述の一文の中に性格検査の得点(数値)と合う内容、合わない内容の両方が含まれる。

例えば「よく緊張したり、神経過敏になったりする。」という質問項目に対し、「部活動の試合前は緊張しないが、就職活動の面接前は緊張した。」ような回答がある。

- 自由記述で実体験らしい回答が少なくなる質問項目がある。

「私のことを自分勝手に、自分のことしか考えていない人間だと思っている人がいる」や「私を冷たく計算高いと見ている人がいる」といった回答者が人にどのように思われているかを問われる質問項目では実体験では判断しづらく、回答が少ない傾向がある。

これらの問題点に対し、2名での判定を行い、除外、修正の判断を行った。得点と実体験が一致していないものについては2名で協議を行い、点数をそのまま採用するか反転させるかを判断した。収集された実体験の文数、不適切として除外した数、点数を反転させた数、文章を修正した数を表1に示す。

全体数	除外	点数反転	文章修正
1186	211	17	5

表1 収集した実体験を補正した内訳

1) 他の質問項目とは測定の向きが逆になっている質問のこと。

### 3.2 ソーシャルメディアからの実体験の自動拡張

実体験は千差万別であり、種々のソーシャルメディアから文章を取得したとしても、必ずしも適切に実体験リストとの関連が計算できるとは限らない。そこで、NEO-FFI に回答したことがあるソーシャルメディアユーザの書き込みを利用し、NEO-FFI の各質問項目との類似性から実体験の自動拡張を考える。奥村ら [13] が収集した 483 名分の性格検査の結果、および Twitter から収集したツイートを、事例を自動で拡張する。ツイートから抽出した実体験には、そのユーザが回答した NEO-FFI の得点（数値）が付与されているため、NEO-FFI の質問項目に対する実体験として数値を取得することが可能となる。

実体験の自動拡張の前処理として、本稿で収集した実体験と拡張に利用するツイートは形態素解析により分割する必要がある。形態素解析器は MeCab<sup>2)</sup> [14] を使用し、辞書には ipadic-Neologd [15], [16], [17] を使用している。

実体験とツイートの類似度の算出には、Word2Vec [18], Doc2Vec [19], Sentence BERT [20] の 3 種類のモデルを使用し、モデルの適切さを評価する。Word2vec のモデルは、SNS データ、日本語 Wikipedia, 自動収集ウェブページから作成された「日本語大規模 SNS+Web コーパスによる単語分散表現モデル [21]」、Doc2Vec でのモデルは日本語 Wikipedia から作成された「日本語 Wikipedia で学習した Doc2Vec モデル<sup>3)</sup>」、Sentence BERT のモデルは日本語 Wikipedia から作成された「cl-tohoku/bert-base-japanese-whole-word-masking<sup>4)</sup>」を利用した。なお、Doc2Vec と Sentence BERT で利用したものは Wikipedia のモデルであるが、本稿の実験においては、ツイートによるファインチューニングは実施していない。

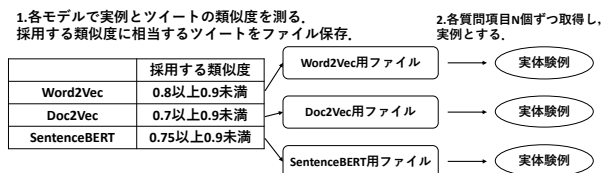


図2 実体験拡張の手順。

図2に示した手法により、実体験の拡張を行う。

- 2) <https://taku910.github.io/mecab/>
- 3) [https://yag-ays.github.io/project/pretrained\\_doc2vec\\_wikipedia/](https://yag-ays.github.io/project/pretrained_doc2vec_wikipedia/)
- 4) <https://github.com/cl-tohoku/bert-japanese/tree/v1.0>

次節以降、実体験の拡張数と、性格傾向の性能の関係を検証する。なお、図2における各モデルの類似度については、予備実験により最適と判定した数値を使用している。

## 4 性格特性の推定手法

前節までの検証実験において、Word2Vec のモデルでは類似度の値が高くなり文間の類似度の差が抽出しづらいという問題があり、Sentence BERT のモデルでは、筆者らの実験環境において計算時間がかかりすぎてしまうという問題があった。そのため、以降の性格特性の推定実験においては、Doc2Vec のモデルを採用している。検証には、Word2Vec, Doc2Vec, Sentence BERT の 3 種のモデルで拡張した実体験例を使用する。性格特性の推定が可能な類似度の閾値は、本稿の著者自身のテキストを利用し、最も著者の性格傾向に合う閾値として 0.5 を設定している。

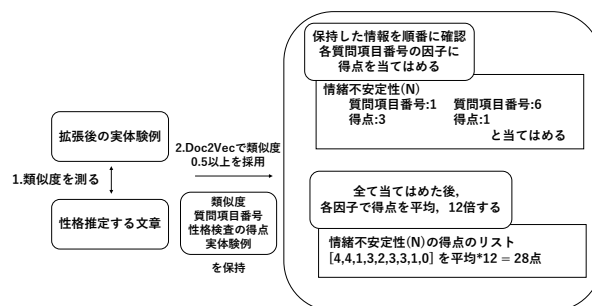


図3 性格推定の手順。

Doc2Vec を利用し、入力文と全ての実体験例との類似度を測定する。このとき、類似度が 0.5 未満となった入力文については性格特性の推定にしようしないものとしている。この得点の算出方法により、性格推定で得られる得点に変動を持たせることができる。

### 4.1 実体験の拡張回数と性格傾向推定の正確性

この節では実体験の拡張回数を変化させて性格推定を行い、性格推定に必要な事例の拡張回数を検討する。性格特性の推定モデルとして、使用する実体験例は少ない方が計算量の面で優位となるため、拡張に必要な最小限の文数のみを使用できるよう検証を行う。

実体験例を拡張する回数による変化を調査するため、NEO-FFI の各質問項目に対し、10 文 ~ 50 文まで 10 文刻みで調査した。この際、改行を 1 文として性

格特性の推定を実施している。実体験例の拡張に使用する文例はランダムに抽出している。そのため、拡張に使用した各モデルに対し、各文数で10回ずつ性格推定を行った。10回試行後、各因子での最高得点から最低得点の差は個数ごとに算出した平均値を表2である。なお、拡張に使用する文数の検証には計算時間の問題から Sentence BERT は使用していない。

	10文	20文	30文	40文	50文
Word2Vec	11.4	7.4	6.1	5.2	4.9
Doc2Vec	9.0	6.7	6.0	3.8	3.6

表2 最低得点と最低得点の差の平均値

さらに、上記の結果から拡張文数を100文~500文の100文刻みで再調査した。各手法、各個数で10回ずつ性格推定を行った。10回試行した結果、各因子での最高得点から最低得点の差を個数ごとに算出した平均値を表3に示す。

	100文	200文	300文	400文	500文
Word2Vec	3.6	3.2	2.6	3.6	4.8
Doc2Vec	3.4	2.8	2.4	3.0	3.2

表3 最低得点と最低得点の差の平均値

この結果から、100文から300文にかけて差が小さくなり、300文から500文にかけて差が大きくなることがわかる。そのため、1つの質問項目につき300文程度の実体験例の拡張が妥当だとわかった。

	N	E	O	A	C
実際の性格検査	34.5	22.1	21.5	30.4	23.8
2.2節の手法	22	26	24	20	30
Word2Vec	21	25	15	28	18
Doc2Vec	19	25	15	27	18
SentenceBERT	21	24	16	26	19

表4 全ての手法の比較

表4に筆者のテキストを用いた性格特性の推定結果と山下らの手法、および実体験に基づく推定手法の結果を示している。

## 4.2 5名分の性格特性の推定

実体験例を使用した性格特性の推定モデルの検証として、奥村らの使用していたNEO-FFIの解答結果と各回答者のTwitterのログを利用し、性格特性の抽出を行った。実際の性格検査の得点から推定した得点の差を集計し、絶対値に変換。各手法で平均をした値を表5に示す。

	A	B	C	D	E
Word2Vec	7.0	6.6	6.2	4.6	6
Doc2Vec	9.4	6.8	5.8	5.4	7
SentenceBERT	6.2	6.8	6.8	6.2	6.6

表5 5名分の性格特性推定結果の比較。AからEでそれぞれの被験者を表す。各数値は、Big FiveのOCEANの各因子の推定値と実際の回答結果との差の絶対値を算出し、5因子の平均として表示している。

この結果から、Word2Vecを利用して実体験例を拡張したモデルが良好であることが分かる。これは、Word2VecのモデルのみSNSを利用したモデルを利用しているため、Wikipediaで構築したDoc2VecおよびSentence BERTのモデルよりも類似度を算出しやすかった可能性がある。そのため、ファインチューニングを施したモデルを利用した実験を行う必要があると考えられる。

## 5 おわりに

本研究ではソーシャルメディアの記述を定量的に扱い、発言者のうつや不安の傾向を捉えるため、実体験に基づく性格特性の推定手法を検討した。ソーシャルメディアをベースとしたユーザ自身の実体験に基づく性格特性の抽出を行うため、従来手法に比べ良好な結果が得られている。一方で、計算資源による制約のため、類似度算出の際に十分な検証ができていない。今後の課題として、Doc2Vec、Sentence BERTのファインチューニングを実施したモデルによる性格特性の推定モデルの検証や、ソーシャルメディアを利用して構築されたモデルの検証があげられる。

## 謝辞

本研究の実施に際して、性格特性検査のためのアンケートにご協力いただいた皆様にこの場を借りて感謝の意を表します。

## 参考文献

- [1] Edmond Pui Hang Choi, Bryant Pui Hung Hui, and Eric Yuk Fai Wan. Depression and anxiety in hong kong during covid-19. **Environmental Research and Public Health**, 2020.
- [2] Selçuk Özdin and Şükriye Bayrak Özdin. Levels and predictors of anxiety, depression and health anxiety during covid-19 pandemic in turkish society: The importance of gender. **Social Psychiatry**, 2020.
- [3] Mark Shevlin, Orla McBride, Jamie Murphy, Jilly Gibson Miller, Todd K. Hartman, Liat Levita, Liam Mason, Anton P. Martinez, Ryan McKay, Thomas V. A. Stocks, Kate M. Bennett, Philip Hyland, Thanos Karatzias, and Richard P. Bentall. Anxiety, depression, traumatic stress and covid-19-related anxiety in the uk general population during the covid-19 pandemic. **BJPsych Open**, 2020.
- [4] 藤倉将平, 小川義人, 菊池英明. 音声対話およびtwitterにおけるユーザのパーソナリティ自動推定. 言語処理学会第19回年次大会発表論文集, 2013.
- [5] 那須川哲哉, 上條浩一, 山本眞大, 北村英哉. 日本語における筆者の性格推定のための言語的特徴の調査. 言語処理学会第22回年次大会発表論文集, 2016.
- [6] 那須川哲哉, 上條浩一. 日本語における筆者の性格推定の取組み. 言語処理学会第23回年次大会発表論文集, 2017.
- [7] 小塩真司. はじめて学ぶパーソナリティ心理学. ミネルヴァ書房, 2010.
- [8] 下仲順子, 中里克治, 権藤恭之, 高山緑. 日本版 neo-pi-r の作成とその因子的妥当性の検討. 性格心理学研究, 1998.
- [9] 村上宣寛, 村上千恵子. 主要5因子性格検査の世代別標準化. 性格心理学研究, 1999.
- [10] 和田さゆり. 性格特性を用いた bigfive 尺度の作成. 心理学研究, 1996.
- [11] 下仲順子, 中里克治, 権藤恭之, 高山緑. 日本版 neo-pi-r, neo-ffi 使用マニュアル. 東京心理株式会社, 1999.
- [12] 山下紗苗, 奥村紀之. 性格情報を反映した発話文生成. 人工知能学会全国大会論文集, 2021.
- [13] 奥村紀之, 金丸裕亮, 奥村学. 感情判断と bigfive を用いたブログ著者の性格推定に関する調査. 人工知能学会全国大会論文集, 2015.
- [14] 工藤拓, 山本薫, 松本裕治. Conditional random fields を用いた日本語形態素解析. 情報処理学会, 2004.
- [15] Sato Toshinori. Neologism dictionary based on the language resources on the web for mecab, 2015.
- [16] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き用辞書生成システム neologd の運用 — 文書分類を例にして —. 自然言語処理研究会研究報告, pp. NL-229-15. 情報処理学会, 2016.
- [17] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第23回年次大会 (NLP2017), pp. NLP2017-B6-1. 言語処理学会, 2017.
- [18] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. **In ICLR Workshop Papers**, 2013.
- [19] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. **In International Conference on Machine Learning**, 2014.
- [20] Nils Reimers and Iryna Gurevych. Sentence-bert:sentence embeddings using siamese bert-networks. **In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing**, 2019.
- [21] 松野省吾, 水木栄, 榊剛史. 日本語大規模 sns+web コーパスによる単語分散表現のモデル構築. 人工知能学会全国大会論文集, Vol. JSAI2019, pp. 4Rin113-4Rin113, 2019.