

ニューラルネット言語モデルによる Twitter 上の発言からの 5 因子モデルに基づく性格分析

塚野 駿[†] 柴田 千尋[†] 政倉 祐子[‡] 田胡 和哉[†]
TSUKANO Shun SHIBATA Chihiro MASAKURA Yuko TAGO Kazuya

1. 背景と目的

日本の SNS 利用者数は 2014 年末の時点で 6,023 万人を占める。また、平成 27 年版情報通信白書によると Twitter の利用率は 31.0% と LINE, Facebook に次いで高い数字である。Twitter は最大 140 文字の短文を投稿できる SNS であり、ユーザの周囲の出来事やユーザと他ユーザの交流等が口語的な文章で多く投稿される。故に、文章中に含まれる単語や言い回し等からユーザの人物像が表れやすいと推定され、長浜ら[1]によって性別・出身エリアといったユーザの人物像を推定する試みが行われている。

本稿では Twitter 発言からユーザの性格を分析することを目的とし、ニューラル言語モデルによる学習を用いて Twitter 発言から 5 因子モデルと近似した性格分析を行う手法について述べる。

2. 利用データおよび提案する分析方法

本稿では Twitter 発言から性格分析を行う手法としてニューラルネット言語モデルの利用を行うことを提案する。

より具体的には、まず、多数のアクティブな Twitter ユーザの最新発言を複数収集する。次に、学習器にかけるため、形態素解析を行い、発言に含まれる単語を抽出する。その後、各単語について 100 次元のベクトルを算出し、これを元の単語の発言ユーザ、名詞、形容詞など、9 種類の品詞および記号に分類する。品詞毎にこのベクトルの平均を取り、各ユーザの各品詞のベクトルをニューラルネット言語モデルの学習データとして使用する。これと同時に各ユーザに性格分析の指標とするラベルとなる数値を割り振り、学習と評価に利用する。

今回性格分析の指標とする性格特性論は 5 因子モデルである。これはビッグ・ファイブとも

呼ばれ、人の性格を外向性・協調性・勤勉性・神経症傾向・開放性の 5 要素で表す事が出来るとする理論である。この理論は人の性格を先の 5 要素に基づいて数値的に表すことから機械的な分析に適しており、奥村ら[2]を始めとして多くの性格分析に利用されている。

ここではニューラルネット言語モデルを利用して 5 因子理論に基づいた性格分析を行い、Twitter 発言からより正確な性格分析を行う事を目指す。

3. 提案手法の詳細および評価実験

本稿では 5 因子モデルに基づいた性格分析を行う手法を提案しているが、5 因子モデルはアンケートの結果に基づく各個人の自己による性格分析であり、第三者視点の性格判断とは異なる可能性がある。第三者視点の性格判断とは、第三者がツイート発言を読み、各因子を外部より判断したものである。従って、本稿では、前者を「**第三者分析**」、後者を「**自己分析**」と呼ぶ。本稿では、1) 性格の第三者分析と自己分析の比較、2) 第三者分析をラベルとした発言からの性格の学習、の二つの実験を行った。

3. 1 第三者分析と自己分析の結果比較

ここでは Twitter 発言の分析結果と 5 因子モデルに基づいた性格分析を比較し、両者の相関性を分析する。

その為に、Twitter アカウントを保有するユーザの協力を得て、Twitter 発言と各ユーザの 5 因子モデルに基づく性格分析のアンケート結果を収集する。ここでは性格分析のアンケートとして小塩らの「日本語版 Ten Item Personality Inventory (TIPI-J)」[3]を使用した。

その後、各ユーザの最新 200 発言を発言に含まれる単語、話題数、他者を対象とした発言数等から分析し、5 因子モデルの性格指標である外向性・協調性・勤勉性・神経症傾向・開放性の各項目に対して 1 から 5 の 5 段階評価を割り振った。この結果と TIPI-J の結果を比較し、両者の相関を求める。本研究では、11 名の被験者に対してアンケートを行った。

Personality Analysis through Big Five Personality Model from Twitter Sentences using a Neural Network Language Model

[†] Tokyo University of Technology

{ c011232769, shibatachh, ktago }@edu.teu.ac.jp

[‡] Aichi Shukutoku University

ymasa@asu.aasa.ac.jp

3. 2 発言からの第三者分析結果の学習

ここではツイート発言とそれに基づく性格分析の結果を学習し、学習結果を利用した分析の正確性を求める。

学習データとして Twitter ユーザ約 700 名の最新 200 発言を取得し、各データについて節 3.1 と同様の分析を行い、5 因子モデルに基づいた各項目への 5 段階評価を割り振った。

その後、発言内容に対して MeCab を用いた形態素解析を行い、発言内に含まれる各単語を主要な 10 種類の品詞毎に分類する。なお、品詞が取れなかったり、それ以外の品詞であった場合は破棄する。

分類された各単語に対して word2vec を利用して 100 次元のベクトルを算出し、これをユーザ・品詞毎に平均したベクトル(1,000 次元)をデータとして使用する。学習には、3 層のニューラルネットワークを用いた。最終層は softmax 関数、他の層は全て ReL 関数を用いた。また、すべての層で dropout を 0.5 の確率で行った。各層のニューロン数は次のようにとった：入力層:1000, 1 層-3 層目:200, 60, 30, 出力層:3。

最終的にこのデータの内 500 人分を訓練データ、200 人分をテストデータとして学習を 200 回行い、評価を行った。

4. 実験結果と考察

節 3.1 の実験について、ツイートの分析による 5 段階の評価結果と TIPI-J の 14 段階の評価結果について、項目ごとに相関係数を算出した(表 1)。

表 1. 性格の第三者分析結果と自己分析(TIPI-J)結果の相関係数

項目名	相関係数
外向性	0.71
協調性	0.68
勤勉性	0.78
神経症傾向	0.16
開放性	0.51

無相関検定の結果、外向性、協調性、勤勉性については有意 ($p < 05$)、開放性については有意ではないものの高い相関を示した。一方、神経症傾向についてのみ、有意な相関が見られなかった。

次に、節 3.2 の学習結果について述べる。表 2 に、テストデータに対する F 値および精度を、5 因子モデルの項目毎に示す。比較のため、同データをナイーブベイズ法で学習した結果の評価も同時に示している。ナイーブベイズ法では、比較および結果の改善のため、節 3.2 の手法で用いた品

詞に属する単語のみを用い、それ以外の単語については同様に破棄した。また、スムージングのための係数は 0.1 とした。

表 2. 性格分析の学習結果の評価

	提案手法		ナイーブベイズ法	
	正解率	F 値	正解率	F 値
外向性	0.53	0.52	0.51	0.49
協調性	0.54	0.46	0.60	0.61
勤勉性	0.76	0.74	0.67	0.67
神経症傾向	0.36	0.35	0.32	0.33
開放性	0.50	0.51	0.47	0.47

表 2 から、今回の実験においてニューラル言語モデルによる学習がナイーブベイズ法に対して勤勉性においてやや精度が高く、外向性・神経症傾向においてわずかに精度が高いといえる。ただし、協調性において精度が劣っており、この点は調整を要する。

また、両学習手法の精度が勤勉性は高く、神経症傾向は低くなっている。これは勤勉性が話題の比率等、今回の手法で読み取り易い情報を評価基準としており、神経症傾向は文章の言い回しの統一性の低さ等の読み取り辛い情報を評価基準としていた為であると考えられる。

5. 結論

本稿ではニューラルネット言語モデルによる学習がナイーブベイズ法に対して、協調性以外においてやや高い精度を示し、特に第三者分析と自己分析の相関係数の高い勤勉性に高い傾向を示した。ただし、神経症傾向と開放性は第三者分析と自己分析の相関係数の時点で低い値を示しており、第三者分析基準の改定が必要と思われる。

今後の予定としては、学習データの単語の分類法等に着目し、学習精度の改善を図る予定である。平行して TIPI-J のサンプルの追加収集を行い、Twitter 発言の分析の改善ないし Twitter 発言と TIPI-J の分析結果自体を学習データとし、より TIPI-J に近い分析結果を目指す。

参考文献

- [1] 長浜裕貴, 遠藤聡志, 當間愛晃, 赤嶺有平, 山田考治: “ユーザツイート解析による人物像推定手法の提案と検討”, 情報処理学会第 76 回全国大会 (2014)
- [2] 奥村紀之, 金丸裕亮, 奥村学: “感情判断と Big Five 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 29 pp.1-4 (2015)
- [3] 小塩 真司, 阿部 晋吾, カトローニ ピノ: “日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み”, パーソナリティ研究 2012 第 21 巻第 1 号 pp.40-52 (2012)