

日本語における筆者の性格推定の取組み

那須川 哲哉 上條 浩一

日本アイ・ビー・エム株式会社 東京基礎研究所

1. はじめに

近年、性格に関する研究が大きく進展している。性格を測る指標としてBig Five Model [1, 2] がデファクトスタンダードとなり、共通の枠組みで研究が進められるようになった上、長年にわたって蓄積されてきたデータの分析を通じて統計的に有意な知見が色々と得られるようになってきている [3]。この流れは、社会心理学の分野に限らない。文章の特徴に筆者の性格との関連性が見出せるという研究報告 [4] を背景に、自然言語処理の分野でも性格に関する研究への関心が高まっている。

筆者らは、日本語文章において性格と関連性の高い言語的特徴を調査 [5] すると共に、その特徴をベースにした日本語筆者の性格推定システムの実現に取り組んできた [6]。本稿では、その取組みを通じて得られた知見を示す。

2. 関連研究

性格の基本的な次元が下記の5つであるというBig Five Model [1, 2] が支持され、共通の枠組みに基づいて、科学的な研究が進められるようになっている。

- Openness to experience (知的好奇心)
独創的・好奇心が強い vs. 着実・警戒心が強い
- Conscientiousness (誠実性)
手際が良い・まめな人 vs. 楽天的・不注意
- Extraversion (外向性)
社交的・エネルギーが豊富 vs. 孤独志向・控えめ
- Agreeableness (協調性)
人当たりが良い・温情あり vs. 冷たい・不親切
- Neuroticism (感情起伏)
繊細・神経質 vs. 情緒安定・自信家

自然言語処理の分野では、テキスト中の表現とその筆者の性格との関連性を分析し、与えられたテキストから筆者の性格を推定する取組みが増えてきており、2013年と2014年にはShared Task [7, 8] も開催され、より効果的な手法の検討が進められている。

英語テキストを対象とした2013年のShared Taskの結果 [7] では、多様性の高い単語レベルの素性を使うボトムアップのアプローチよりも、Linguistic Inquiry and Word Count (LIWC) [9] のようにカテゴリ化された素性を用いるトップダウンのアプローチの方が精度が良いという知見が得られている。但し、このShared Taskで用いられたラベル付きゴールド

スタンダードのデータは、約2千4百件のエッセイ (平均語数: 8百語未満) 及び250ユーザのFacebookのデータであり、機械学習のためのデータとしては特に大きいサイズではないことから、学習用データの量が増えれば、傾向が変わる可能性が考えられる。

自然言語処理の分野における性格分析の取組みには英語のテキストを対象としたものが多く、日本語を対象とした取組みは、出始めている [10, 11, 12] もの、まだ数が少ない。テキストと性格の関連性には言語や文化による違いが現れてくる可能性があることを考えると、日本語における研究の余地はまだ大きいと感じられる。我々は、性格診断アンケートの結果として性格ラベルを付与した312ユーザのTwitterデータ (以下「ツイート」) を対象とし、LIWCをベースに日本語のどのようなカテゴリの表現がどのような性格と関連を示すかを調査した [5, 13]。各ユーザの平均ツイート数は855件であった。調査の結果、LIWCをベースにした日本語表現カテゴリのうち、7割以上という、多くのカテゴリが何らかの性格との関連性を示しているという知見が得られた。また、関連性が認められたカテゴリの中には、日本語向けに追加したカテゴリも含まれていた。この結果を踏まえ、我々は、日本語向けのカテゴリを拡張しつつ、より大規模なデータでの分析を行なった上で、日本語筆者の性格推定システムを実現した [6]。

3. 日本語において性格と関連性の高い表現

3.1. 第一調査

日本語においてどのような表現が性格を反映しているかを調査する上で、筆者らはまずLIWC [9] に着目した。LIWC2001の各カテゴリに対応する表現を認識する仕組みを IBM Watson Explorer Advanced Edition Analytical Components V11.0 [14] の言語処理機能を利用して実装した [13]。LIWC2001で辞書化されている68カテゴリの日本語化において、日本語に対応しないArticles、及びNon fluenciesと区別困難なFillersを除き、Prepositionsに関しては助詞を対応させた。さらに、心理学的知見から、日本人の性格と関連する可能性があると考えられる13のカテゴリを追加した。その中には、Relaxカテゴリ (「温泉」「癒し」他) などに加え、LIWC2001に含まれる一人称代名詞のサブカテゴリとしてのWatashiカテゴリ (「私」「あたし」他) やBokuカテゴリ (「僕」「ぼく」他) などが含まれる。

調査データとして、150回以上ツイートしている

Twitterユーザに、性格診断アンケート¹を受けてもらい、そのアンケートの結果から算出される性格プロフィールを各ユーザの正しい性格特性値（正解）として、各ユーザのツイートと結びつけた。性格プロフィールとしては、Big Fiveに加え、Kevin FordのUniversal Needs Map に沿ったNeeds（欲求）分析[15]とSchwartzの価値概説（Schwartz Value Survey）に沿ったValues（価値観）分析[16]の結果も対象とした（図1）。

性格プロフィール	
大分類	小分類
Big5 (個性)	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism (5)
Values (価値観)	Excitement, Harmony, Curiosity, Ideal, Closeness, Self-expression, Liberty, Love, Practicality, Stability, Challenge, Structure (12)
Needs (欲求)	Self-transcendence, Conservation, Hedonism, Self-enhancement, Open to change (5)

図1:性格プロフィールの分類

性格プロフィールと表現の関連性を分析するにあたり、まず、筆者 j の全ツイートデータ中に含まれる各カテゴリ $cat_i (i=1, \dots, 79)$ に該当する単語数を算出後、1文字数あたりの出現数 $a_{i,j}$ に換算し、 n 人分の出現ベクトル $\mathbf{a}_j = (a_{1j}, \dots, a_{79j})$ を求めた。さらに、 j のアンケート時の性格プロフィール $c_{kj} (k=1, \dots, 22)$ に対するスコアを $s_{kj} (0 \leq s_{kj} \leq 1)$ とし、各 k に対する n 人分のスコアベクトル $\mathbf{s}_k = (s_{k1}, \dots, s_{kn})$ を求めた。そして、各 \mathbf{a}_i と \mathbf{s}_k 間のピアソンの積率相関係数、および p 値を計算した。

その結果、2015年12月の時点で分析対象となった312ユーザ分のデータから得られた性格プロフィールとカテゴリとの相関関係の抜粋を表1に示す。

表1:外向性及び協調性と一部カテゴリとの相関関係

カテゴリ	外向性		協調性	
	相関係数	p 値	相関係数	p 値
Negations	-0.167	0.0027	-0.0311	0.584
Assents	0.155	0.0055	0.0539	0.342
助詞	-0.200	0.00027	-0.0691	0.222
Watashi	0.0741	0.191	0.127	0.0243
二人称	-0.0174	0.759	-0.152	0.0066

外向性に対し、Negationsカテゴリの表現（「～しない」「～せず」などの否定形）が負の相関を示し、Assentsカテゴリの表現（「OK」「認める」「了解」など）が正の相関を、助詞カテゴリが負の相関を示している。協調性に対しては、Watashiカテゴリの表現が正の相関を示し、二人称代名詞カテゴリの表現（「あなた」「お前」など）が負の相関を示すといった興味深い傾向が見られた。

但し、この時点では、データの偏りに関する懸念

1

<https://personality-insights-surveys.mybluemix.net/?s=gjdgxk>

が存在した。性格診断アンケートのサイトはインターネット上で公開していたものの、当初は、アンケート参加をもつばら日本IBMの社員に呼びかけていたことから、多くのデータ提供者がIBM社員及びその関係者であると推測された。アンケート参加者の性格プロフィールの分布自体は、図2の外向性の分布例のようになっており、極端な分布は見られないことが確認されている。しかし、年齢に関しては、表2の通り、参加者の半数強が35歳から54歳であった。

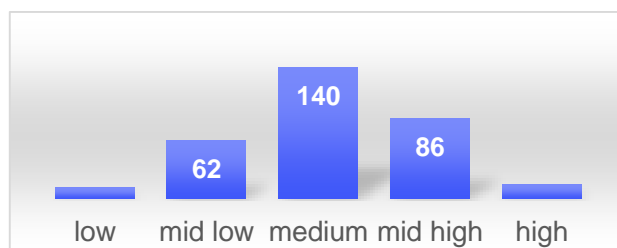


図2:2015年12月時点の分析対象者の外向性の分布

表2:第一調査アンケート参加者の年齢分布

年齢	17 以下	18-24	25-34	35-54	55 以上
人数	1	51	77	172	12

3.2. 第二調査

第一調査の後、様々なチャネルを通じ、性格診断アンケートへの参加を幅広い対象に募ると共に、Big Fiveの各プロフィールを更に6つのサブ・プロフィールに分けて分析する第二弾の性格診断アンケート²を開始した。その結果、2016年9月の時点で、1630人（平均年齢28歳、男女比61:39）分のデータ（平均ツイート数：1315件）を集めることができた。

表3:第二調査アンケート参加者の年齢分布³

年齢	17 以下	18-24	25-34	35-54	55 以上
人数	67	534	367	235	17

調査対象の表現として、第一調査で p 値が小さく相関係数の比較的大きかった助詞を細分化し、

- 格助詞カテゴリ
- 係助詞カテゴリ
- 終助詞カテゴリ
- 副助詞カテゴリ

を設定すると共に、日本語特有のカテゴリとして字種に着目した

- 漢字カテゴリ
- ひらがなカテゴリ
- カタカナカテゴリ
- アルファベットカテゴリ

及び、

- 遊びカテゴリ（「遊び」もしくは「ゲーム」）

2

<https://personality-insights-surveys.mybluemix.net/?s=gjdgxj>

³年齢未回答者を除く

● 旅行カテゴリ
を追加した。

結果的に、第二調査では各性格プロファイルに対して、各カテゴリのp値と相関係数の絶対値が下がる傾向が見られ、表4の通り、p値が0.05未満のカテゴリが増えた反面、相関係数の絶対値が0.1を超えるカテゴリの数が減少した。

表4: 性格プロファイルに対しp値の低いカテゴリ数 (相関係数の絶対値が0.1以上のカテゴリ数)

	第一調査 79 カテゴリ中		第二調査 90 カテゴリ中	
	p<0.05	p<0.01	p<0.05	p<0.01
知的好奇心	20 (19)	5 (5)	51 (11)	39 (11)
誠実性	23 (23)	11 (11)	36 (4)	25 (4)
外向性	14 (14)	7 (7)	44 (7)	25 (7)
協調性	15 (15)	5 (5)	49 (11)	34 (11)
感情起伏	25 (25)	17 (17)	51 (11)	39 (11)

但し、p値が0.01未満で相関係数の絶対値が0.1を超えるカテゴリの数は減っていない。

第二調査で追加したカテゴリのうち、Big Fiveの性格プロファイルのいずれかとの相関係数の絶対値が0.1を超え、その際のp値が0.01未満になったカテゴリは、

- 格助詞カテゴリ
- 係助詞カテゴリ
- ひらがなカテゴリ
- 遊びカテゴリ

の4カテゴリであった。

結果的に、第一調査で見られた表1の関係においては、表5の値となり、傾向が変化した。

表5: 外向性及び協調性と一部カテゴリとの相関関係

カテゴリ	外向性		協調性	
	相関係数	p 値	相関係数	p 値
Negations	-0.068	0.0061	-0.066	0.0072
Assents	0.078	0.0016	0.035	0.158
助詞	-0.030	0.226	-0.133	5.32×10^{-7}
Watashi	0.032	0.190	0.044	0.078
二人称	0.0129	0.604	-0.019	0.452

外向性に対する、Negationsカテゴリの表現の負の相関や、Assentsカテゴリの表現の正の相関が弱まった上、助詞カテゴリの表現との相関に有意性が見られなくなった。また、協調性に対する、Watashiカテゴリの表現との正の相関及び二人称代名詞カテゴリの表現との負の相関も有意性が認められなくなった。すなわち、3百件程度のデータからこういった傾向を結論付けるのは難しいと考えられる。

その反面、助詞と協調性に負の相関が見られるようになり、千6百件規模のデータにおいても、助詞や字種の使用頻度が性格を反映するという結果が得られた。例えば、格助詞カテゴリの表現に関しては、

Big Fiveの性格プロファイルのうち、協調性と負 (相関係数: -0.128 , p値: 1.59×10^{-7}) の、感情起伏と正 (相関係数: 0.129 , p値: 1.51×10^{-7}) の、知的好奇心と正 (相関係数: 0.257 , p値: 1.31×10^{-10}) の相関が認められた。また、ひらがなカテゴリの表現は、誠実性と負 (相関係数: 0.127 , p値: 2.11×10^{-7}) の相関が認められた。

4. 日本語筆者の性格推定システム

前述の結果から、日本語の性格分析に有効なカテゴリを集めたカテゴリ体系Japanese Category for Personality Identification (JCPI) を用いて性格を推定する仕組みを実装した。実装にあたっては、前述の調査で用いた性格診断アンケート回答者のツイートを利用し、アンケートから得られた性格プロファイルと各カテゴリの表現の出現頻度との回帰分析を行った。また、GloVe[17]の学習済みWord Vector (200次元)のデータの日本語表現部分を利用し、カテゴリ体系に依存しない文字Nグラム ($2 \leq N \leq 10$)の表現の情報を併用できるようにした。

1630人のデータを用いて、0から1の値をとるBig Fiveの各性格プロファイルに対して、10-分割交差検証により、実測値と推定値とのMean Absolute Error (MAE)を求めた結果を表6に示す。カテゴリ体系を用いるよりも、Word Vectorを用いた方が精度が高いが、カテゴリ体系とWord Vectorを併用した場合に、最も高い精度が得られた。

表6: 性格推定システムの性能 (MAE) 評価結果

	JCPI	Word Vec.	JCPI+WV
協調性	0.1057	0.1027	0.1001
誠実性	0.0962	0.0941	0.0939
外向性	0.1211	0.1158	0.1145
感情起伏	0.1186	0.1147	0.1122
知的好奇心	0.1099	0.1064	0.1063
平均	0.1103	0.1067	0.1054

実装したシステムは、IBM Watson Personality Insights日本語版として提供され、デモサイト⁴も公開されている。Big Fiveの各プロファイルと、各々の6つのサブ・プロファイルに加え、図1のValues (価値観)とNeeds (欲求)の推定結果が出力される。2017年1月現在、このデモサイト上では、日本語のツイートを10万人分処理した結果をベースに、百分位数のスコアによって、Twitterユーザの集団の中での性格特性の位置を示すようになっている [18]。

本システムを用いて、2016年4月14日以降に発生した熊本地震の被災者のツイートを分析した結果の一部を表7に示す。ユーザプロファイルやツイート発信地などの情報から熊本の住民と推測されるTwitter

⁴ <https://personality-insights-livedemo.mybluemix.net/>

ユーザのうち、ツイート数の多い3千人を対象として、ユーザ毎に、震災前の44日間（3月1日～4月13日）、震災中の9日間（4月14日～4月23日）、震災後の44日間（4月24日～6月7日）のツイートを収集し、各期間の各ユーザのツイートを（総語数3千語以上の場合に）入力して得られた結果の平均値である。

表7:熊本地震被災者の震災前後の性格推定値⁵

Big Five プロファイル サブプロファイル	震災前	震災中	震災後
感情起伏	0.521	0.538	0.512
低ストレス耐性	0.568	0.695	0.583
知的好奇心	0.506	0.487	0.495
大胆性	0.489	0.363	0.467
思考力	0.495	0.39	0.481
誠実性	0.539	0.518	0.535
注意深さ	0.513	0.702	0.532

Big Fiveプロファイルの値は概ね安定しているが、一部のサブプロファイルの値は、被災中に変化し、時間が経つと被災前の値に近づいている様子が見られる。例えば、低ストレス耐性や注意深さの値が被災中に大きく上がっており、ストレスが高まり、過敏になっている様子が反映されていると考えられる。

5. おわりに

日本語において性格と関連性の高い表現を調査し、そのデータを利用した性格推定の取組みについて示した。カテゴリ化された素性を用いるトップダウンのアプローチを試した結果、内容語だけでなく、助詞やひらがなの使用状況が性格プロファイルと統計的な有意性をもって相関しており、何を表現するかだけでなく、どう表現するかに性格が反映されているという知見が得られた。また、1630人分のツイートを対象とした調査では、Word Vectorを用いるボトムアップのアプローチが、カテゴリ化された素性を用いるトップダウンのアプローチよりも性格推定精度を高めることができ、両者を併用した場合にさらに精度が高まるという結果が得られた。

従来、テキストマイニングをはじめとするテキストデータの活用技術は、テキスト中に何が書かれているかを分析対象にしてきたが、どう書かれているかも分析することで、表現されている内容のより深い解釈につながる可能が考えられる。

謝辞

性格診断データの収集にあたって多くの方々のご協力をいただきました。また、熊本地震被災者の分析では、東京工業大学の青木竜哉氏のご協力をいただきました。ここに記して深謝いたします。

⁵ 2016年10月時点の調査結果

IBM © Watson Explorer Advanced Edition Analytical Components V11.0、IBM © Watson Personality Insights は International Business Machines Corporation の米国およびその他の国における商標。

参考文献

- [1] Goldberg, Lewis R. "An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology* 59.6: 1216, 1990.
- [2] McCrae, R. R. and John, O.P. "An introduction to the five-factor model and its applications." *Journal of Personality*, 60(2), 175-215, 1992.
- [3] 村上宣寛: 性格のパワー, 日経 BP 社, 2011.
- [4] Mairesse, F., Walker, M.A., Mehl, M.R., and Moore, R.K., Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial intelligence Research*, 30: 457-500, 2007.
- [5] 那須川哲哉, 上條浩一, 山本真大, 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回年次大会予稿集, pp.1181-1184, 2016.
- [6] Kamijo, Koichi, Tetsuya Nasukawa, and Hideya Kitamura. "Personality Estimation from Japanese Text." In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pp. 101-109. 2016.
- [7] Celli, Fabio, Fabio Pianesi, David Stillwell, and Michal Kosinski, "Workshop on Computational Personality Recognition: Shared Task", AAI Technical Report WS-13-01, 2013.
- [8] Celli, Fabio, Bruno Lepri, Joan-Isaac Biel, Daniel Gatica-Perez, Giuseppe Riccardi, and Fabio Pianesi. "The workshop on computational personality recognition 2014." In *22nd ACM international conference on Multimedia*, pp. 1245-1246., 2014.
- [9] Pennebaker, James W., Martha E. Francis, and Roger J. Booth. "Linguistic inquiry and word count: LIWC 2001." *Mahway: Lawrence Erlbaum Associates* 71 (2001): 2001.
- [10] 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回年次大会予稿集, pp.900-903, 2013.
- [11] 岡本拓馬, 松本和幸, 吉田稔, 北研二. ナイーブベイズ法を用いた Twitter による性格推定, 言語処理学会第20回年次大会予稿集, pp.1123-1125, 2014.
- [12] 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, 第29回人工知能学会全国大会, 4J1-5, 2015.
- [13] 山本真大, 那須川哲哉, 上條浩一, 北村英哉, LIWC2001 手作業翻訳の方針と半自動翻訳手法の提案, 言語処理学会第22回年次大会予稿集, pp.1185-1188, 2016.
- [14] Wei-Dong Zhu, et al. IBM Watson Content Analytics: Discovering Actionable Insight from Your Content. An IBM Redbooks publication. ISBN-10:0738439428. 2014.
- [15] Yang, H., and Li, Y. "Identifying user needs from social media." IBM Research Report, RJ10513 (ALM1309-013), 2013.
- [16] Ryan L. Boyd1, Steven R. Wilson, James W. Pennebaker1, Michal Kosinski, David J. Stillwell, and Rada Mihalcea, "Values in Words: Using Language to Evaluate and Understand Personal Values." In *9th International AAAI Conference on Web and Social Media*, p.31-40, 2015.
- [17] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global Vectors for Word Representation." In *EMNLP*, vol. 14, pp. 1532-43. 2014.
- [18] Personality Insights
<https://www.ibm.com/watson/developercloud/personality-insights.html>