機械翻訳を利用した多言語文書の筆者の性格推定

那須川 哲哉　上條 浩一　金山 博　村岡 雅康
日本アイ・ビー・エム株式会社 東京基礎研究所

1. はじめに

心理学の発展に伴い、Big Five Model[1,2]によって、人の性格特性が数値化できるようになっている。その結果、仲の良い人、特にパートナー関係にある人は性格が近い傾向にあるといった報告[3]が出てきており、例えば、性格特性を推定することで、性格が近そうな人を特定し、グループ形成に活かすといった応用が考えられるようになってきている。

自然言語処理の分野では、文章の特徴に筆者の性格との関連性が見出せるという報告[4]に基づき、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いる。メールやSNSなど日常的にオンライン上でテキストを作成する機会が多い昨今、そのテキストから性格を推定することができれば、従来は多様な質問[5,6,7]に回答する必要のあった性格特性の調査が飛躍的に容易になり、個々人の性格に適応する多様な応用につながると考えられる。そのため、2013年と2014年にはShared Task[8,9]も開催され、より効果的な手法の検討が進められている。しかし、その取り組みの多くが英語のテキストを対象としたものであり、テキストから性格推定を行えるシステムが開発されている言語は限られているのが現状である。

性格が近く相性の良い可能性が高そうな人を探すという応用を考えた場合、母国語が同じでコミュニケーションが取り易い相手よりも、母国語が異なる相手においてこそ、共感を得やすく、努力してでもコミュニケーションを取りたいと思える候補者を見出すことに価値があると考えられる。例えば、未知の環境に入って様々な不安を抱えている留学生にとっては、早い段階での親しい人との巡り合いが、留学生活の充実度を高めてくれる可能性がある。長期的には多様な個性との出会いが重要である可能性もあるものの、短期的には、例えば、海外旅行時などに、ガイドやツアー仲間などで、性格の近い人間を選ぶことができれば、より快適な経験につながる可能性が高そうである。そのためには多言語のテキストにおいて筆者の性格推定を可能にする必要がある。

我々は、日本語を対象とした筆者の性格特性推定システムを開発した[10]が、そのためにまず必要となったのが、筆者の性格特性と紐付いた学習用の日本語テキストデータであった。性格特性を調査するには、調査協力者に様々な質問に回答してもらう必要がある。性格特性を調べるための質問には[5,6,7]をはじめ複数の種類が存在するが、英語で開発されたものが多く、この質問のための適切な表現を日本語で用意した上で、調査協力者を募るという作業が必要になる。この表現を自動的に獲得する試みも行われている[11]が、現時点では、人手による翻訳が不可欠である。従って、テキストからの性格推定システムを新しい言語において構築するためには、性格特性を調査するための質問をその言語で用意し、さらにその言語での調査協力者を募り、質問に回答してもらうと共に、協力者が記述したテキストデータを収集する必要があり、大きなコストがかかる。

言語毎の性格推定システムを構築することなく、多様な言語のテキストから筆者の性格推定を実現するためには、機械翻訳を利用することが考えられる。近年では多様な言語の機械翻訳システムが構築されているため、性格推定システムが存在しない言語に関しては、性格推定システムが存在する言語に翻訳した上で性格推定を行えば良いという考えである。しかし、性格推定に寄与する表現には、言語特有で、翻訳結果にうまく反映されない可能性のある表現が含まれる。例えば、我々が日本語の性格分析に有効な表現のカテゴリを集めたカテゴリ体系のJapanese Category for Personality Identification (JCPI)[10]には、例えば英語などの、他言語において直接的に対応する表現のない助詞や文字種が含まれている。そのため、翻訳時に継承されない表現や内容がノイズとなって性格推定の精度を落とす可能性が考えられる。

機械翻訳を利用した多言語文書の筆者の性格推定を精度良く実現するため、我々は、逆翻訳を利用し、機械翻訳によって原文における分布が変化してしまう表現のカテゴリを特定した上で、分布の変化を考慮して性格推定を行う仕組みを考案した。

本稿では、次節において、性格推定システムの概要と、単純に機械翻訳をした上で性格推定を行なう場合の推定精度に関する実験結果を示す。次に、第3節において、逆翻訳によって表現の分布が変化するカテゴリと変化内容を示し、その結果を考慮した上で性格推定を行う仕組みと、その効果を第4節で示す。

2. 性格推定システムと機械翻訳の利用

我々が日本語版の性格推定を実現したシステムは、現在、IBM Watson Personality Insights (以降PI)という名称で公開されている[12]。任意のテキスト

を入力すると、その筆者の性格特性として、性格の基本的な次元を下記の5つとするBig Five Modelなどの値を出力す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･エネルギッシュ vs. 孤独志向･控えめ
- Agreeableness（協調性）
  人当たり良い･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

本システムは、現時点では、英語、スペイン語、日本語、アラビア語、韓国語に対応している。各言語のデータにおいて、調査協力者の質問回答結果から得られた、0から1の値を取る性格特性値をテキストから再現する際のMean Absolute Error（MAE)は0.09から0.12程度である[13]。

機械翻訳による翻訳結果を用いた場合に、得られる性格特性がどの程度変化するかを調査した結果を表１に示す。Big Five Modelのうち、Neuroticismに関しては､アラビア語向けのPIで出力しないため、対象外とした。ここでは、テキストとしてツイッターのデータを利用し､日本語のツイッターユーザ100名のテキストをベースにした。データ収集など処理の便宜上の理由に加え、同等の内容で比較できるという点から、この100名のテキストを、機械翻訳により、英語、スペイン語、アラビア語に翻訳した結果を、各言語のテキストのデータとした。機械翻訳システムとしては、Microsoft Translator[1]を用いた。各言語のPIで得られた性格特性を正解とした場合に、機械翻訳で日本語に翻訳してから日本語版のPIで得られた性格特性との乖離がどの程度発生するかをMAEの値で示している。

表1:機械翻訳テキストによる性格推定のMAE

Big Five	アラビア語	スペイン語	英語
Openness	0.0332	0.0478	0.0929
Conscientiousness	0.0525	0.0284	0.0671
Extraversion	0.0653	0.0519	0.0622
Agreeableness	0.0279	0.0343	0.106
平均	0.0447	0.0406	0.082

言語や性格特性の軸によって異なるが､MAEの値で､0.03から0.11程度の乖離が発生するという結果が得られた。

3. 翻訳による性格特性関連表現分布の変化

翻訳したテキストを用いることで性格推定結果が変化する理由として、性格特性の推定に寄与する表現の分布が翻訳によって変化することが考えられる。

一般的に、性格推定システムを構築する上では、トップダウンに定義したカテゴリもしくはボトムアップに集約したカテゴリに属する表現が、ある著者によって記述されたテキスト中にどの程度の割合で含まれているかを算出し、その分布と質問への回答によって得られた性格特性の値とを回帰分析することで、性格推定モデルを構築し、テキストのみからの著者の性格推定を実現する。そのため、各カテゴリに属する表現の分布が変化すると、推定される性格特性が変化することになる。

我々がPI用に定義した日本語の性格分析用のカテゴリ体系JCPIの各カテゴリにおける出現分布の変化状況の抜粋を表2に示す。

日本語のテキストを一旦、英語、スペイン語、アラビア語に機械翻訳した上で、逆翻訳して日本語に戻すと、助詞に関しては、格助詞の出現頻度が概ね半分に減っているのに対し、副助詞は若干増えており､スペイン語やアラビア語を経由した場合が2割弱の増加なのに対し、英語を経由した場合は、4割以上増加しているという結果が得られた。数字に関しては、どの言語でも2割程度に減っており、内容を調べてみると、例えば「一番良い」や「一番悪い」といった表現が「最高」や「最低」になっているなど、数字を使わない表現に変換されているケースが見受けられた。代名詞に関しては、代名詞全体の出現頻度は1割から2割程度の増加にすぎないが、一人称代名詞の割合が3割程度に減少しており､二人称代名詞の割合が2倍以上に増加するといった結果が得られた。これは、翻訳時に、対象言語に応じて代名詞が補完されており、逆翻訳においては、二人称代名詞が訳文に残るのに対し、一人称代名詞は省略される傾向が強いという現象が発生しているためであった。

4. 多言語向けの性格特性推定手法

前節の結果を踏まえ、性格特性推定システム未対応言語に対しては、機械翻訳システムを用い、性格特性推定システム対応言語に翻訳して性格推定を行うと共に、性格特性関連表現の分布の翻訳による変化を捉え、性格特性推定値を調整する仕組みを開発した。試行錯誤の結果たどり着いた性格特性調整パラメータ生成の仕組みを図１に示す。

基本的な処理の流れは下記のようになる。

1. 機械翻訳システムを用い､日本語のテキストJを性格特性システム未対応言語Z及び､性格推定システム対応言語Y_1, .., Y_n に翻訳する
2. Zに翻訳されたテキストZJ、及び、Y_1, .., Y_n に翻訳されたテキストY_1J, .., Y_nJを機械翻訳システムで日本語に逆翻訳し、JZJ及びJY_1J, .., JY_nJを生成する

[1] https://www.microsoft.com/en-us/translator/translatorapi.aspx

表 2:他言語に翻訳してから逆翻訳した日本語テキストにおける性格分析用カテゴリ表現の出現頻度分布の変化

カテゴリ	格助詞	副助詞	数字	全代名詞	一人称単数代名詞	二人称代名詞	カタカナ語
(a) 元の日本語データにおける出現頻度（千語中の語数）	110.809	8.740	34.911	14.826	5.840	0.369	75.026
(b) 英語に翻訳してから逆翻訳した日本語データにおける出現頻度	55.165	12.592	6.582	18.014	2.603	0.855	0.252
(c) スペイン語に翻訳してから逆翻訳した日本語データにおける出現頻度	58.512	10.300	6.714	16.895	2.352	0.780	0.375
(d)アラビア語に翻訳してから逆翻訳した日本語データにおける出現頻度	55.695	10.335	7.554	18.575	2.428	0.834	0.328
英語経由の逆翻訳による出現頻度の変化率（b/a）	0.498	1.441	0.189	1.215	0.446	2.317	0.003
スペイン語経由の逆翻訳による出現頻度の変化率（c/a）	0.528	1.178	0.192	1.140	0.403	2.113	0.005
アラビア語経由の逆翻訳による出現頻度の変化率（d/a）	0.503	1.183	0.216	1.253	0.416	2.258	0.004

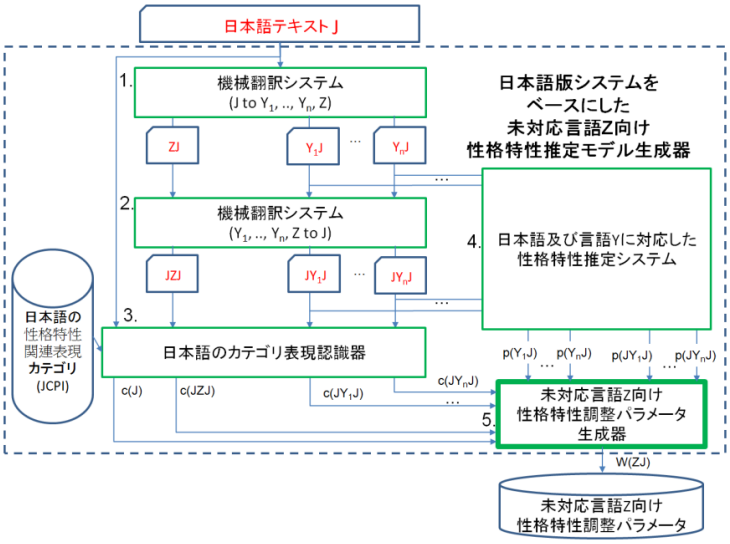

図1:性格特性推定システム未対応言語向けの性格特性調整パラメータの生成

3. J及びJZJとJY$_1$J，..，JY$_n$Jにおける性格特性関連表現カテゴリの分布を解析する
4. Y$_1$J，..，Y$_n$Jに対しては、各言語の性格特性推定システムを、JY$_1$J，..，JY$_n$Jに対しては、日本語の性格特性推定システムを用いて、各々の性格特性推定値を出力する
5. 3で得られた各言語のテキストにおける性格特性表現カテゴリの分布情報と、4で得られた、性格推定システム対応言語における翻訳前後での性格特性推定値の値から、未対応言語Z向け性格特性調整パラメータを生成する

この流れで生成されたパラメータを用い、性格特性推定システム未対応言語Zのテキストから性格特性を推定する関数 $\boldsymbol{pm}(Z)$ を下記のように定義した。

$$\boldsymbol{pm}(Z)=\boldsymbol{W}(ZJ)\boldsymbol{c}(JZ)+\boldsymbol{p}(JZ) \quad (1)$$

ここで、$\boldsymbol{p}(JZ)$はZを日本語に翻訳したテキストを日本語版のPIにかけて得られる性格特性推定値であり、$\boldsymbol{c}(JZ)=(c_1(JZ),\ldots,c_s(ZJ))^{\mathrm{T}}$は、Zを日本語に翻訳したテキストにおける性格特性関連表現カテゴリ（JCPI）の分布の値であって、sはカテゴリ数を示している。$\boldsymbol{W}(ZJ)$は日本語から翻訳された未対応言語Zのテキストに対する重み付き係数行列であり、s次元の重みベクトル$\boldsymbol{w}$で構成される。

$$\boldsymbol{W}(ZJ)=(\boldsymbol{w}_1(ZJ),\ldots,\boldsymbol{w}_p(ZJ))^{\mathrm{T}}$$
$$\boldsymbol{w}_k(ZJ)=(w_{k1},\ldots,w_{ks})^{\mathrm{T}}$$

（pはBig Fiveなどの構成要素から成る性格特性プロフィールの数。） ここで、言語に依存しないグローバル・パラメータ$\boldsymbol{G}=(\boldsymbol{g}_1,\ldots,\boldsymbol{g}_p)^{\mathrm{T}}$、$\boldsymbol{g}_k=(g_{k1},\ldots,g_{ks})^{\mathrm{T}}$、k =1,…,p を定義し、$\boldsymbol{w}_k(ZJ)$を下記のように求めることにする。

$$\boldsymbol{w}_k(ZJ)=\boldsymbol{g}_k\cdot\boldsymbol{v}(ZJ) \quad (2)$$

ここで、$\boldsymbol{v}(ZJ)=(v_1(ZJ),\ldots,v_s(ZJ))^{\mathrm{T}}$が未対応言語Zと日本語における性格特性関連表現カテゴリの分布の類似性を示す値のベクトル表現である。この関係を成立させるため、$\bar{\boldsymbol{c}}(J)$と$\bar{\boldsymbol{c}}(JZJ)$を引数とする関数$\boldsymbol{H}$を定義する。

$$\boldsymbol{v}(ZJ)=\boldsymbol{H}(\bar{\boldsymbol{c}}(J),\ \bar{\boldsymbol{c}}(JZJ)) \quad (3)$$

ここで、$\bar{\boldsymbol{c}}(A)=(\bar{c}_1(A),\ldots,\bar{c}_s(A))^{\mathrm{T}}$は言語Aのテキストにおける性格特性関連表現カテゴリの平均出現分布であり、下記のように求める。

$$\overline{c_l}(A)=\frac{1}{N}\sum_{i=1}^{N}c_l(T_i^A) \quad (4)$$

T_i^Aは筆者iが言語Aで記述したテキストを示しており、Nは筆者の数である。筆者i による言語Aのテキストにおけるカテゴリlの表現の数を単語あたりに正規化したものが$\boldsymbol{c_l}(T_i^A)$となる。言語Aと言語Bの特徴が似ていて、カテゴリ表現の分布の変化が小さいカテゴリでは、$\bar{c}_l(ABA)$を$\bar{c}_l(A)$で割った値が1に近い値になるものと考えられる。そのため、この値が1に近いカテゴリの重みを大きくするように

$\boldsymbol{H}(\boldsymbol{x},\boldsymbol{y})=(H(x_1,y_1),\ldots,H(x_s,y_s))^{\mathrm{T}}$を定義する。

$$H(x,y) = \begin{cases} \min(1/|1-r|, a_0) & r_0 \le r \le r_1, \\ 0 & \text{otherwise}, \end{cases}$$
$$r = \begin{cases} \frac{y}{x} & x \neq 0, \\ r_2 & x = 0, \end{cases} \quad (5)$$

ここで、a_0、r_0、r_1、r_2、は負でない定数であり、かつ、$r_1 < r_2$ である。

次に、性格特性推定結果を最適化するように性格特性推定システムに対応している言語Yのテキストデータを用いて$\boldsymbol{G}$の値を求める。具体的には、上記の式(1)から(5)に対して、下記の$\Delta serr_k$ の値が各性格特性 k (k=1, …, p)について最小化されるように$\boldsymbol{g}_k$を求めた。

$$\Delta serr_k = \sum_{i,j} (pm_k(T_i^{Y_j}) - p_k(T_i^{Y_j}))^2 \quad (6)$$

こうして得られた性格特性調整パラメータを用いて最終的に得られた各言語のテキストの性格特性のMAEを表3に示す[2]。表1の値との差分を表1の値で割った結果を改善率とした。表1の値と比べ、全てのケースにおいて精度が改善されるという結果が得られた。

表3:提案手法で調整した性格推定のMAEと(改善率)

Big Five	アラビア語	スペイン語	英語
Openness	0.0325 (2.1%)	0.0477 (0.2%)	0.0718 (22.7%)
Conscientiousness	0.0175 (66.7%)	0.0186 (34.5%)	0.0285 (57.5%)
Extraversion	0.0486 (25.6%)	0.0505 (2.7%)	0.0481 (22.7%)
Agreeableness	0.0252 (9.7%)	0.0269 (21.6%)	0.0615 (42.0%)
平均	0.031 (30.6%)	0.0359 (11.6%)	0.0525 (36.0%)

5. おわりに

言語毎の性格推定システムを構築することなく、機械翻訳を利用して、多様な言語のテキストからの筆者の性格推定を精度良く実現するための取り組みを示した。

性格特性推定エンジンが存在する言語のテキストを機械翻訳してから別言語のエンジンで性格特性を推定すると、元の言語で推定される値と比較して、MAEの値で、0.03から0.11程度乖離するという結果が得られた。この原因として、性格推定に寄与する表現カテゴリに属する表現の出現分布が機械翻訳によって変化することが考えられたため、実際にカテゴリの分布の変化を調査したところ、日本語を、英語・スペイン語・アラビア語のいずれかに翻訳してから日本語に逆翻訳した場合、どのケースでも一人称代名詞の割合が30%程度に減少し、二人称代名詞の割合が2倍以上に増加するといった結果が得られた。そのため、逆翻訳における分布の変化を考慮して、性格推定における各表現カテゴリの重みを調整するようにした。さらに、性格特性推定結果を最適化するように性格特性推定システムに対応している複数言語のテキストデータを用いてパラメータの調整をおこなうようにした[3]ところ、性格特性の軸によって数値が異なるものの、概ね高いレベルで、全言語の全ての性格特性において、MAEを改善することができた。

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] Youyou, Wu, David Stillwell, H. Andrew Schwartz, and Michal Kosinski. "Birds of a Feather Do Flock Together: Behavior-Based Personality-Assessment Method Reveals Personality Similarity Among Couples and Friends." Psychological science 28, no. 3: 276-284. 2017.

[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] P.T. Costa and R.R. McCrae. Revised NEO Personality Inventory (NEO PI-R) and NEO FiveFactor Inventory (NEO-FFI). Psychological Assessment Resources, 1992.

[6] 下仲順子，中里克治，権藤恭之，高山緑．NEO-PI-R，NEO-FFI 共通マニュアル．東京心理, 1999.

[7] S. D. Gosling, P. J. Rentfrow, and Jr. Swann, W. B. A very brief measure of the big five personality domains. Journal of Research in Personality, Vol. 37, pp. 504-528, 2003.

[8] Celli, Fabio, Fabio Pianesi, David Stillwell, and Michal Kosinski, "Workshop on Computational Personality Recognition: Shared Task", AAAI Technical Report WS-13-01, 2013.

[9] 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.

[10] 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 回年次大会予稿集, pp.807-810, 2017.

[11] 植田晋平，河原大輔，黒橋禎夫，岩井律子，井関龍太，熊田孝恒，パーソナリティ表現の自動翻訳の試み，言語処理学会第 22 回年次大会予稿集, pp.282-285, 2016.

[12] IBM Watson Developer Cloud: Personality Insights, https://personality-insights-demo.ng.bluemix.net/

[13] IBM Cloud 資料: Personality Insights, "The science behind the service", https://console.bluemix.net/docs/services/personality-insights/

[2] 定数は a_0=2、r_0=0.8、r_1=1.2、r_2=2.0 に設定した。

[3] 英語の実験ではスペイン語とアラビア語のシステムに限定するなど、実験対象言語の性格特性推定システムは外して調整した。