

都市別の投稿を利用した市民の性格特性の分析

岩崎 大晟

企業のマーケティング戦略において、コミュニティの好みや行動の傾向を把握することは重要である。ソーシャルメディアの投稿には人々の好みや行動の傾向が反映されていると考えられ、実際に多くの企業でソーシャルメディアへの投資額は増加している。そこで本研究では、性格特性が個人の行動や反応の傾向に影響を与える重要な要素であることに着目し、個人の性格特性を主要な5つの因子で表現する Big Five Model を参照して、都市別の投稿を用いて市民の性格特性を分析することを目的とする。この際、都市間での差をより詳細に分析するために、Big Five Model の主要な因子ではなく、側面因子であるファセットの性格特性を対象にする。具体的には、市民という集団の性格特性を分析するにあたり、「利他主義」、「芸術的興味」、「冒険心」、「社交性」、「活発度」の5つの性格特性に焦点をあて分析を行う。

本研究では、大規模言語モデル RoBERTa をファインチューニングすることで、性格特性と経験性の予測モデルを作成する。作成したモデルを用いて各都市の投稿に対して自動でラベル付与を行い、市民の行動の傾向がどのように性格特性や経験性に反映されているかを検証する。

性格特性および経験性の予測モデルを学習するため、本研究では、福岡市、横浜市、札幌市の3都市を対象に、X (旧 Twitter) の投稿に対して人手で性格特性と経験性のアノテーションを行いデータセットを構築した。アノテーション作業では、全ての性格特性において、Fleiss の κ 係数を用いた判定者間一致度が 0.6 を超えたことから、判定者に依存しない妥当な基準でデータセットが作成されたことが示された。都市ごとのデータで学習したモデルについてはほとんどの性格特性において F 値は 0.8 を超え、都市を横断して学習したモデルについては全ての性格特性において F 値は 0.8 を超えた。経験性予測モデルにおいても F 値は 0.8 を超え、投稿における性格特性および経験性が高い精度で予測可能であることが示された。

学習したモデルを用いて未知の投稿に対してラベルの自動付与を行い、都市間で市民特性の比較を行った。その結果、「芸術的興味」、「社交性」、「活発度」の3つの性格特性については、都市間で母比率の Z 検定 (有意水準 5%) で有意差が見られる結果となった。具体的には、「芸術的興味」、「社交性」においては札幌市が有意に高い結果となった。「活発度」では福岡市と札幌市が横浜市に比べて有意に高い結果となった。また、本研究では Big Five Model の主要な5つの因子を測定するのではなく、ファセットに着目し分析を行ったことで、札幌市の開放性の高さに貢献する要因についても考察することができた。このように、本研究の提案手法により、一部の性格特性については都市間の差を検出できることが明らかになり、さらにファセットに着目することで都市の特徴を把握できることが示された。

(指導教員 関 洋平)