



HOKKAIDO UNIVERSITY

Title	A Study on Machine Learning-based Approaches for Personality Identification and Translation [an abstract of dissertation and a summary of dissertation review]
Author(s)	RADISAVLJEVIC, Dusan
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第15665号
Issue Date	2023-09-25
Doc URL	https://hdl.handle.net/2115/90830
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Dusan_Radisavljevic_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 Dusan RADISAVLJEVIC

学 位 論 文 題 名

A Study on Machine Learning-based Approaches for Personality Identification and Translation

（人格特性の識別と翻訳のための機械学習アプローチに関する研究）

ソーシャルメディアサービスの普及により、私たちの交流はますますオンライン空間に移行している。この現象と様々なチャットアプリケーションの急速な発展により、私たちのコミュニケーションの大半はテキストメッセージで行われるようになってきている。そのため、テキストを通じて対話者が伝えようとする意図や感情を正しく解釈することがますます重要となっている。

人々のコミュニケーションを適切に解釈するためには、文脈や親しみやすさが大きく影響しているが、人々が示す特定のパターンには一貫性がある。パーソナリティ心理学ではこのようなパターンや人同士の個人レベルでの違いを、パーソナリティと呼ばれる概念に帰着させている。パーソナリティとは、行動、感情、認知のパターンに存在する個人差の総体であり、時間や文脈にかかわらず比較的一貫したままである。このように考えると、個性と一貫性の両方を表すパーソナリティは、コミュニケーションにおいて不可欠な役割を担っていると推察される。

個人のパーソナリティを正しく理解し、解釈することの意義は、多くの研究者が注目してきたところである。ここ数十年、技術の進歩や新しい計算アルゴリズムの活用が進み、パーソナリティコンピューティングの研究分野が形成されるに至った。パーソナリティコンピューティングはまだ比較的新しい研究分野であるが、その関心はますます高まってきている。しかし、この分野の新規性から、評価基準に研究標準がないため、作品の比較は困難である。さらに、性格に関連する情報をラベル付けたデータが容易に入手できないことも問題になっている。その主な要因は、異なる性格測定の好みとデータのプライバシーに対する懸念であることが多い。

この2つの問題は、パーソナリティコンピューティング研究に関する多くの研究において、主な障害として最も頻繁に挙げられており、私が異なる性格評価方法の接続の可能性を研究する最大の動機となっている。もし、異なる性格診断法をうまく接続することができれば、関係するすべての性格診断法の利用可能なデータを効果的に増やすことができる。さらに、パーソナリティの理解とコミュニケーションへの反映を中心としたアプローチをとることで、標準化された評価の枠組みを開発することに貢献し、その中で、異なる研究作品間でのパフォーマンスの違いをうまく再現し解釈することが可能となる。

この目的を念頭に置いて、本論文で説明される研究は、話者識別タスクから始まり、対話者のコミュニケーションのテキスト転写物のみを使用して対話者を識別する可能性を確立することを目的としている。この研究で提案された新しいトランスフォーマーベースのアプローチは、F1 指標で70% 以上のスコアを獲得し、ベースラインアプローチを凌ぐ確実性で、発話が誰のものかを予測できることが証明された。さらに、この実験では、商用ビデオゲームのダイアログのテキストトランスクリプトに基づく大規模なデータセットが得られ、70,000 を超える発話があることがわかった。私の知る限り、ファンタジーテキストを対話関連タスクに利用する見込みは、これまでの取り組みで検討されてきたが、市販のビデオゲームのデータをこのようなタスクに利用するのは、今回が初めての

ことである。

本研究では、テキストコミュニケーションに個人差があるか否かの答えが得られたことを踏まえ、さらに、テキストベースの特徴と2つの異なる性格評価モデルとの関係を調べることで、その違いを生み出す正確な理由を検証している。ビッグファイブとマイヤーズ・ブリッグス・タイプ・インディケーターという2つの性格診断モデルは、ソーシャルメディアプラットフォーム Reddit のテキストを分析したところ、いずれも特定の言語的特徴との相関が示された。この発見は、ビッグファイブの語彙的背景からビッグファイブの言語特性を確認するものであるが、同様の特性が、あまり研究されていない Myers-Briggs Type Indicator の性格モデルにも反映されている可能性があるという新しい知見を提供するものである。

これらの知見は、より入手しやすい Myers-Briggs Type Indicator と、もう一つの性格診断モデルであるエニアグラムのデータを、ビッグファイブの性格診断に変換するために、研究の後半で使用されている。実験のこの部分で取られた詳細なアプローチは、研究の再現性と比較可能性を保証するものである。その結果、シンプルなアプローチでありながら、ピアソン r 相関係数の評価指標において、測定ごとの相関強度が最大 13.2% 増加した。

また、評価基準を領域やデータの特性に合わせるため、Myers-Briggs Type Indicator と Enneagram を Big 5 評価に変換する際に最も優れたアプローチを、スピアマンの順位相関係数と平均 2 乗誤差の評価指標を用いて再評価した。これらの再評価により、当初の知見が確認され、このタスクに最も効果的と思われる機能とアルゴリズムの選択に関する主張がさらに実証された。