A computational framework for interpretable text-based personality assessment from social media

Matej Gjurković, dipl. ing.

Mentor: Prof. Jan Šnajder, PhD
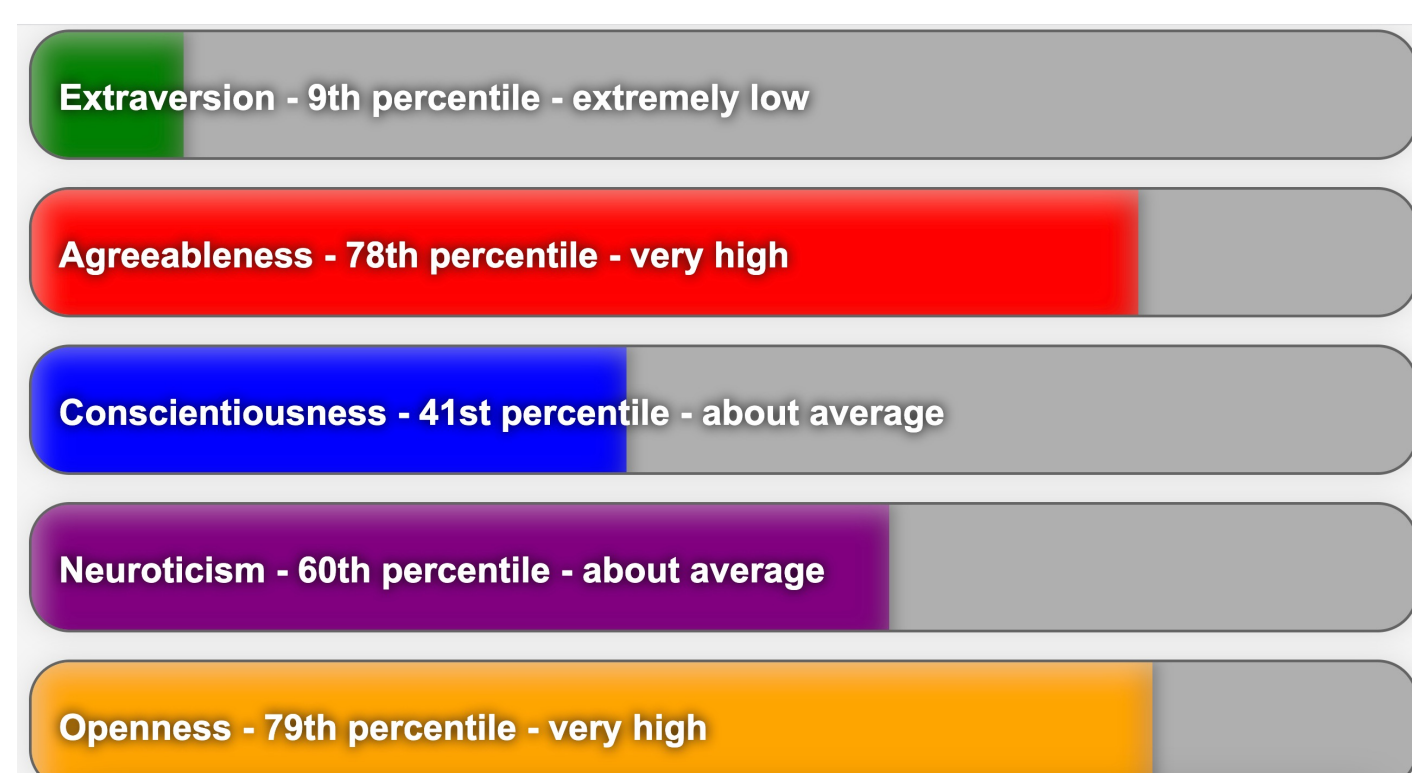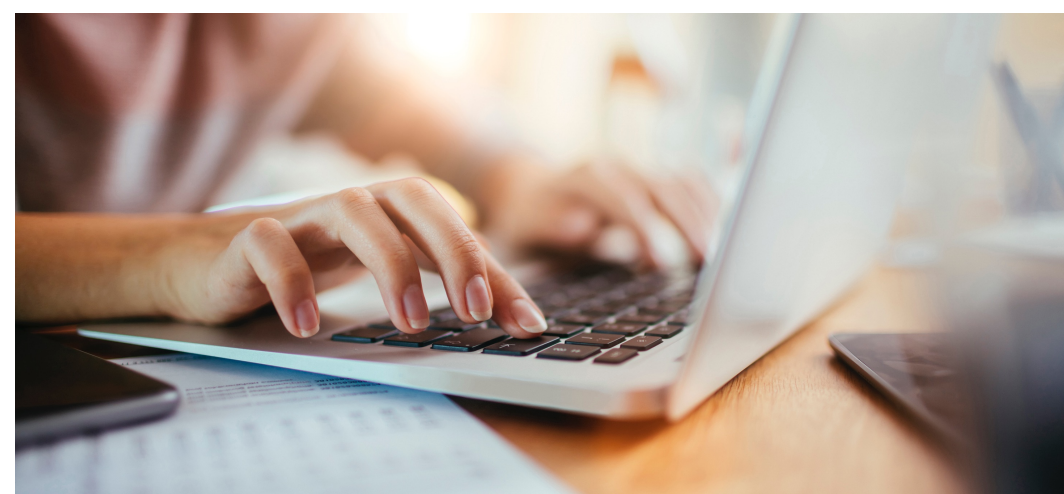University of Zagreb Faculty of Electrical Engineering and Computing

UNIVERSITY OF ZAGREB
Faculty of Electrical Engineering and Computing

1. Introduction

- The variability of language generation is influenced by various author characteristics, such as age, gender, culture, and **personality**
- To infer an author's personality, we can use **Natural Language Processing** to analyze both the **content** (*what* the text says) and the **style** (*how* it is expressed) of their writing

Extraversion - 9th percentile - extremely low
Agreeableness - 78th percentile - very high
Conscientiousness - 41st percentile - about average
Neuroticism - 60th percentile - about average
Openness - 79th percentile - very high

2. Problem Description

- **Lack** of **appropriate** personality-labeled datasets; limited in diversity of sources, authors, topics, and writing styles
- Mostly labeled with **Myers Briggs Type Indicator (MBTI)** but researchers use more valid **Big 5** personality model
- Lack of **interpretable** and **valid** NLP models that can support **evidence-based** judgment and mitigation of **confounders**

	Facebook (Big 5)	Twitter (MBTI)	Essays (Big 5)	Forums (MBTI)	Reddit (MBTI)
Anonymous	✗	+	✗	✓	✓
Diverse Topics	+	✗	✗	✗	✓
A large number of users & data	✓	✓	✗	✓	✓
Expressivity	+	✗	✓	✓	✓
Ease of labelling	✗	✓	✗	✓	❓

3. Objectives

Objective 1: Create a large-scale dataset from a new data source
- author-level labels: Big 5 scores with facets, age, gender, …

Objective 2: Develop framework for personality assessment from text which is:
- model agnostic (AI models are rapidly improving!)
- interpretable and explainable
- flexible: all levels of personality taxonomy (domain, facet)

4. Dataset creation

Step 1 – Identifyng the data source and labeling process [1]
- Reddit: a large number of users and texts per user, topical diversity due the anonymity
- Flairs – short subreddit/community descriptiors that are used to communicate age, gender – and on personality-related subreddits – personality type (MBTI)

Step 2 – semiautomatic labelling with Big 5 scores (and facets where available), gender, age, and location [2]

Result: the largest personality dataset of 17,640,062 comments written by 10,288 authors (MBTI, Big 5, gender, age, location)

5. Experiments

Exp. 1. Using abundant MBTI labels to predict Big 5

MBTI / Big 5	O	C	E	A	N
Introverted	−.062	−.062	**−.748**	−.055	**.157**
Intuitive	**.434**	−.027	−.042	.030	.065
Thinking	−.027	**.138**	−.043	**−.554**	**−.341**
Perceiving	**.132**	**−.575**	**.145**	.055	.031

Key finding: Big 5 score can be predicted without Big 5 labeled data (up to 0.265 Pearon's r)

Source set MBTI (n=8,691)

Adaptation set MBTI+Big 5 (n=382)

Target set Big 5 (n=1588)

Exp. 2. Personality-related biases in gender classification

|	Female			Male		
Variable	✓	✗	Δ	✓	✗	Δ
Age	26.78	25.83	**0.95***	25.46	26.90	−1.44
I/E	0.78	0.72	0.06	0.76	0.82	−0.06
N/S	0.86	0.91	−0.05	0.92	0.93	−0.01
T/F	0.47	0.64	**−0.17***	0.61	0.29	**0.32***
P/J	0.39	0.56	**−0.17***	0.53	0.39	**0.14***

Exp. 3. Benchmark prediction models

Key findings: 0.265 r (Openness) – 0.387 r(Extraversion), **more difficult to predict** than age (0.75 r) and gender (90.4 F1)

6. SIMPA: Statement-to-Item Matching Personality Assessment

Questionnaire items associated with positive and negative keys of facets and domains –> **Trait-relevant statements** (TRSes)
- **Idea**: detect **trait-indicative statements** (TISes) in comments based on **semantic similarity** to TRSes
- We base our framework on **Realistic Accuracy model** [2]
- Four stages: relevance, availability, detection, utilization
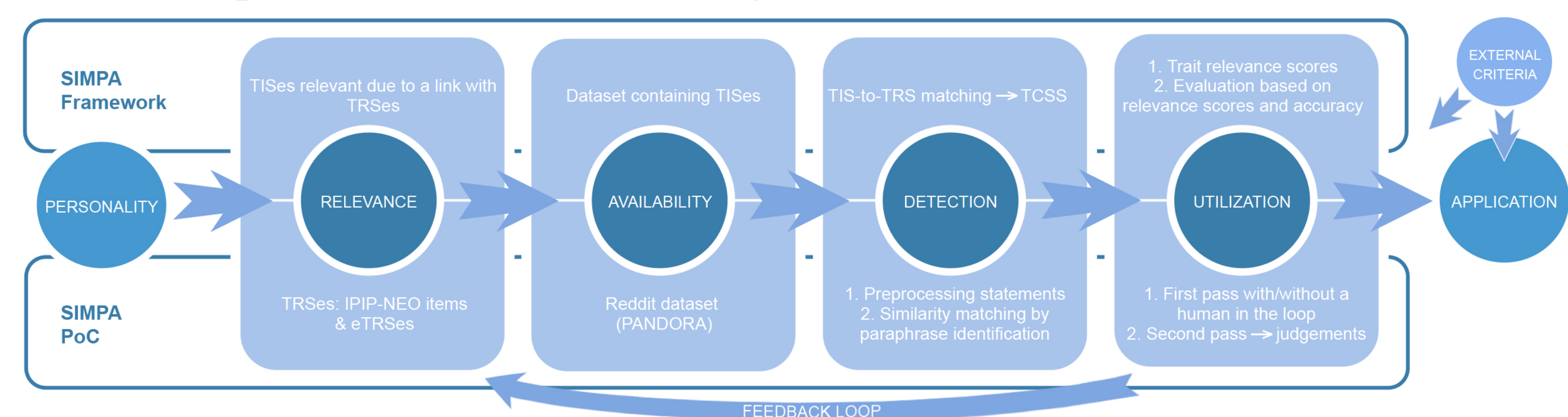
"Oh yeah I think this way too. It's massively overwhelming which is why I am lazy and I procrastinate.

TIS	TRS	Trait	Facet	Key
It's massively overwhelming	Often feel overwhelmed / Find it difficult to get down to work.	neuroticism	anxiety	plus
I am lazy and I procrastinate		conscientiousness	self-discipline	minus

Key findings: demonstrated **convergent** and **discriminant** **validity** in **supervised** and **unsupervised** setups while **increasing** predictive power (0.458 r for Extraversion)

7. Conclusion

- Created publicly available large-scale Reddit dataset labeled with Big 5 personality and demographics
- Through experiments demonstrated the usefulness of the dataset and the need for more valid and interpretable models
- Proposed, implemented and validated an interpretable approach for text-based personality assessment

References

[1] Matej Gjurković Jan Šnajder (2018). Reddit: A Gold Mine for Personality Prediction. Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES 2018), New Orleans, 87–97.

[2] D. C. Funder, Accurate personality judgment, Directions in Psychological Science 21 (2012) 177–182.

Contact
Matej Gjurković, dipl. ing.
matej.gjurkovic@fer.hr

PhD Day Faculty of Electrical Engineering and Computing, 1 June 2023