

Classification of Myers–Briggs Type Indicator personality types using Natural Language Processing

Andor Kiss, Dóra Bányai, Milán Kriston, and Zoltán Kádár

Eötvös Loránd University

1 Literature search

As indicated in papers [1] [3] the state of the art models for text classification are transformer based architectures. Our idea was to use an LSTM-based baseline model and three different pretrained transformer architectures downloaded from huggingface, namely Generative Pre-trained Transformer 2 (GPT-2) [2] BERT and RoBERTa.

2 Individual contributions

2.1 Andor Kiss - TXC54G

- Team leader tasks - Git repo, weekly report to supervisor, Google Docs, LaTeX template
- Literature search
- Data exploration
- Data pipeline
- GPT-2 training and evaluation

2.2 Dóra Bányai - NEPTUN

- Literature search
- Data exploration
- RoBERTa

2.3 Milán Kriston - NEPTUN

- Literature search
- Data pipeline
- BERT

2.4 Zoltán Kádár - NEPTUN

- Literature search
- Data exploration
- LSTM-based baseline model

Table 1. Table captions should be placed above the tables.

Model	Accuracy	F1 score	Precision	Recall	Execution speed
GPT-2@cat@100	0.549	0.543	0.55	0.549	-
GPT-2@bin@100	0.524	0.516	0.540	0.524	-
GPT-2@cat@250	0.70	0.70	0.70	0.70	-
GPT-2@bin@250	0.69	0.6889	0.6945	0.69	-
GPT-2@cat@500	0.837	0.838	0.84	0.837	-
GPT-2@bin@500	0.82	0.819	0.821	0.82	-

3 Results

Where cat represents the model having 16 different output possibilities corresponding to the 16 personality types with softmax output activation, bin represents the model having 4 binary classifiers as the output layer predicting each character in the MBTI type, and the number (100, 250, 500) representing the maximum sequence length the model was trained with.

References

1. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., Gao, J.: Deep learning based text classification: A comprehensive review (2020). <https://doi.org/10.48550/ARXIV.2004.03705>, <https://arxiv.org/abs/2004.03705>
2. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners (2019)
3. dos Santos, V., Paraboni, I.: Myers-briggs personality classification from social media text using pre-trained language models. JUCS - Journal of Universal Computer Science **28**(4), 378–395 (apr 2022). <https://doi.org/10.3897/jucs.70941>, <https://doi.org/10.3897%2Fjucs.70941>