

# 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 [3] [5] 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) [4] BERT and RoBERTa. As the article about GPT-2 [4] shows, the architecture was pretrained in a self-supervised way to understand the language first, and later finetuned for specific tasks, which is the reason why we used it in this project, leveraging the potential of it to model the English language, and providing it with inputs and labels for our specific classification task. We are aware of the fact that GPT-2 is not the current state of the art model in OpenAI's series of language models, however due to computational costs, we did not want to use a too large model with several billion parameters like GPT-3 [2] or GPT-NEOX [1]. GPT-2 small was used with its 124 million parameters.

## 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

## 3 Results

**Table 1.** Evaluation results

Model	Accuracy	F1 score	Precision	Recall	Execution speed
GPT-2@cat@100	0.549	0.543	0.55	0.549	0.0467
GPT-2@bin@100	0.524	0.516	0.540	0.524	<b>0.0324</b>
GPT-2@cat@250	0.70	0.70	0.70	0.70	0.0479
GPT-2@bin@250	0.69	0.6889	0.6945	0.69	0.0577
<b>GPT-2@cat@500</b>	<b>0.837</b>	<b>0.838</b>	<b>0.84</b>	<b>0.837</b>	0.067
GPT-2@bin@500	0.82	0.819	0.821	0.82	0.0831

In Table 1 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) represents the maximum sequence length the model was trained with. The execution speed column indicates how quickly a batch is processed by the model in seconds. It was measured on an NVIDIA P100 GPU during inference.

In paper [5] the models had 4 binary classifiers as output, thus showing the metrics separately for each character in the MBTI type, however since in this project we used both categorical and binary outputs, a fairer comparison is done by converting the binary outputs to categorical with some postprocessing steps and calculating the error using that format.

As the table shows, the sequence length of 500 was the best performing by far, while the categorical outputs outperformed the binary outputs by a small margin. Compared to the difference in performance, the difference between the models in terms of execution speed was negligible.

## 4 Conclusion and Future work

### 4.1 Conclusion

TODO - write something about the difficulties during the project

## 4.2 Future work

TODO

## References

1. Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., He, H., Leahy, C., McDonell, K., Phang, J., Pieler, M., Prashanth, U.S., Purohit, S., Reynolds, L., Tow, J., Wang, B., Weinbach, S.: Gpt-neox-20b: An open-source autoregressive language model (2022). <https://doi.org/10.48550/ARXIV.2204.06745>, <https://arxiv.org/abs/2204.06745>
2. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners (2020). <https://doi.org/10.48550/ARXIV.2005.14165>, <https://arxiv.org/abs/2005.14165>
3. 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>
4. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners (2019)
5. 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>