

Appendix for the paper *Exploiting Transformer-based Multi-task Learning for the Detection of Media Bias in News Articles*

June 4, 2021

1 Datasets

In this section we provide more detailed information about the datasets discussed in.

Reddit data set (Reddit) [2]: The dataset contains 6861 comments extracted from the social network Reddit. The comments are posted in response to news articles and annotated for populist attitudes via Crowdsourcing. The authors operationalize populism as comments containing attitudes ranging from support to discrimination towards a social group. The labels are aggregated into a continuous scale ranging from 0 (*supportive*) to 1 (*discriminatory*).

Subjectivity data set (Subj) [5]: The dataset contains 10000 movie reviews for binary subjectivity classification. The reviews are labeled as *objective* or *opinionated*.

IMDb [4]: A dataset for binary sentiment classification comprising 50000 highly polar movie reviews with positive and negative sentiments.

Wikipedia data set (Wiki) [6]: The *Wiki Neutrality Corpus* consists of 180,000 sentence pairs extracted from English Wikipedia pages. The sentences are extracted from articles going against Wikipedia’s *NPOV*¹ rule. A sentence pair contains the original biased sentence and its revised neutral version.

Semantic Textual Similarity Benchmark (STS-B) data set [3]: In Semantic Textual Similarity tasks, models are trained to learn and measure semantic similarity of text sequences. STS-B contains 10943 multilingual and cross-lingual sentence pairs labeled in terms of similarity. The annotation scale ranges from 0 (*no overlap*) to five (*equivalence*), and the similarity labels are gathered via Crowdsourcing. The dataset is part of the GLUE benchmark [7].

Stanford Natural Language Inference (SNLI) corpus [1]: Natural Language Inference tasks are designed to train NLP models in distinguishing entailing from contradicting text passages. The SNLI dataset comprises 570000 sentence pairs labeled via Crowdsourcing. The crowdsourcers were asked to annotate the linguistic relation of two sentences with the labels *entailment*, *neutral*, or *contradiction*. The SNLI corpus is also part of the GLUE benchmark [7].

2 Hyperparameters

Platform	Google Colab
GPU	NVIDIA Tesla K80 (12 GB RAM)
Optimizer	AdamW ²
Batch size	32
Learning Rate	$5e^{-5}$

Table 1: Setup and MTL model parameters.

¹https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

References

- [1] Samuel R. Bowman et al. “A large annotated corpus for learning natural language inference”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sept. 2015, pp. 632–642. DOI: [10.18653/v1/D15-1075](https://doi.org/10.18653/v1/D15-1075). URL: <https://www.aclweb.org/anthology/D15-1075>.
- [2] Pere-Lluís Hugué Cabot et al. “Us vs. Them: A Dataset of Populist Attitudes, News Bias and Emotions”. In: *CoRR* abs/2101.11956 (2021). arXiv: [2101.11956](https://arxiv.org/abs/2101.11956). URL: <https://arxiv.org/abs/2101.11956>.
- [3] Daniel M. Cer et al. “SemEval-2017 Task 1: Semantic Textual Similarity - Multilingual and Cross-lingual Focused Evaluation”. In: *CoRR* abs/1708.00055 (2017). arXiv: [1708.00055](https://arxiv.org/abs/1708.00055). URL: <http://arxiv.org/abs/1708.00055>.
- [4] Andrew L. Maas et al. “Learning Word Vectors for Sentiment Analysis”. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics, June 2011, pp. 142–150. URL: <https://www.aclweb.org/anthology/P11-1015>.
- [5] Bo Pang and Lillian Lee. “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts”. In: *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. ACL ’04. Barcelona, Spain: Association for Computational Linguistics, 2004, 271–es. DOI: [10.3115/1218955.1218990](https://doi.org/10.3115/1218955.1218990). URL: <https://doi.org/10.3115/1218955.1218990>.
- [6] Reid Pryzant et al. “Automatically Neutralizing Subjective Bias in Text”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.01 (Apr. 2020), pp. 480–489. DOI: [10.1609/aaai.v34i01.5385](https://doi.org/10.1609/aaai.v34i01.5385). URL: <https://ojs.aaai.org/index.php/AAAI/article/view/5385>.
- [7] Alex Wang et al. “GLUE: A multi-task benchmark and analysis platform for natural language understanding”. In: *arXiv preprint arXiv:1804.07461* (2018).