Benchmark results

The results of our benchmark for several Language Models using data from MMS.

Our preliminary results has been presented in (Rajda et al. 2022) and finally presented **in (Augustyniak et al. 2023) review at NeurIPS'23**.

Benchmark results - F1 Macro scores

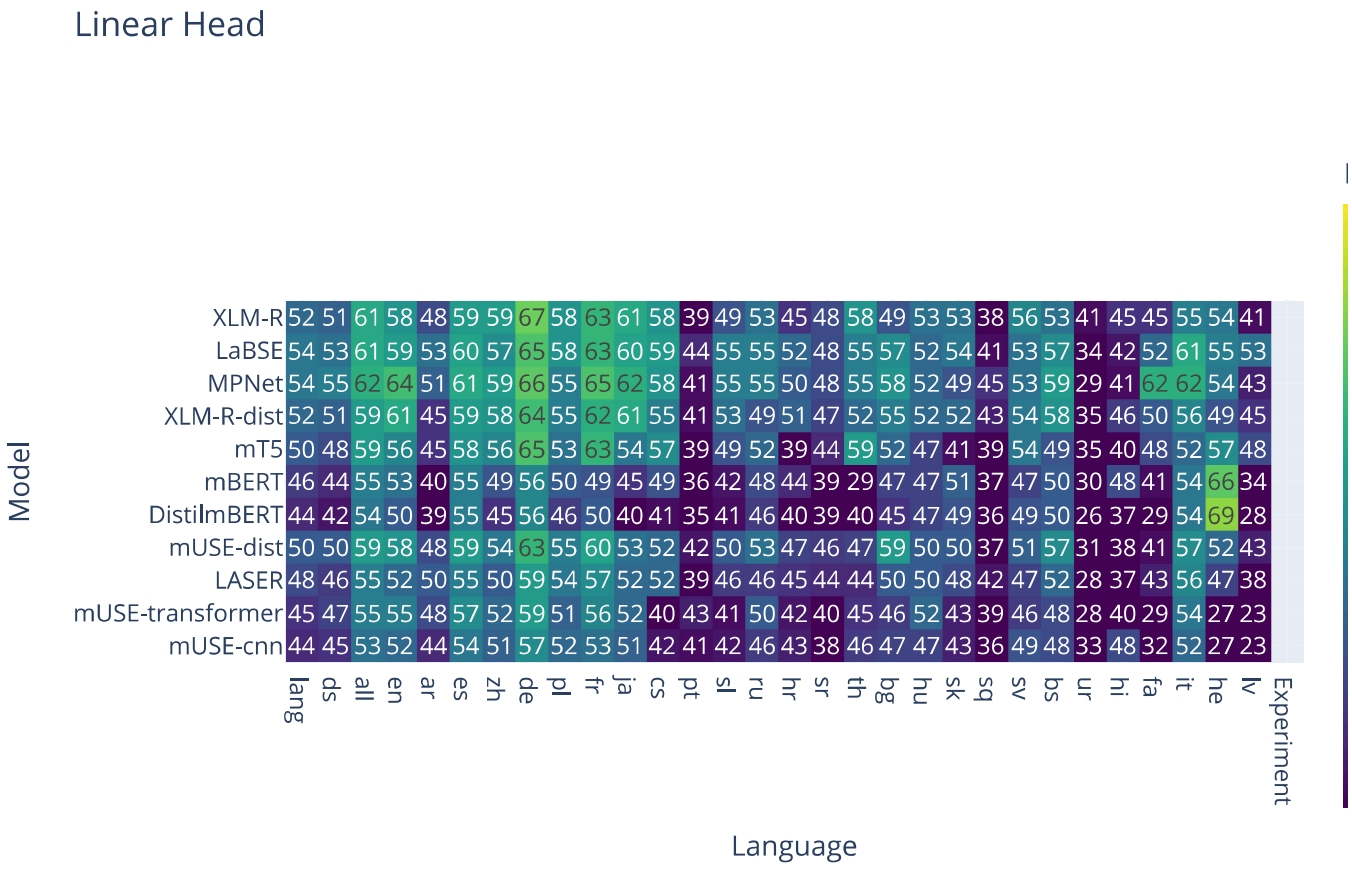
Models

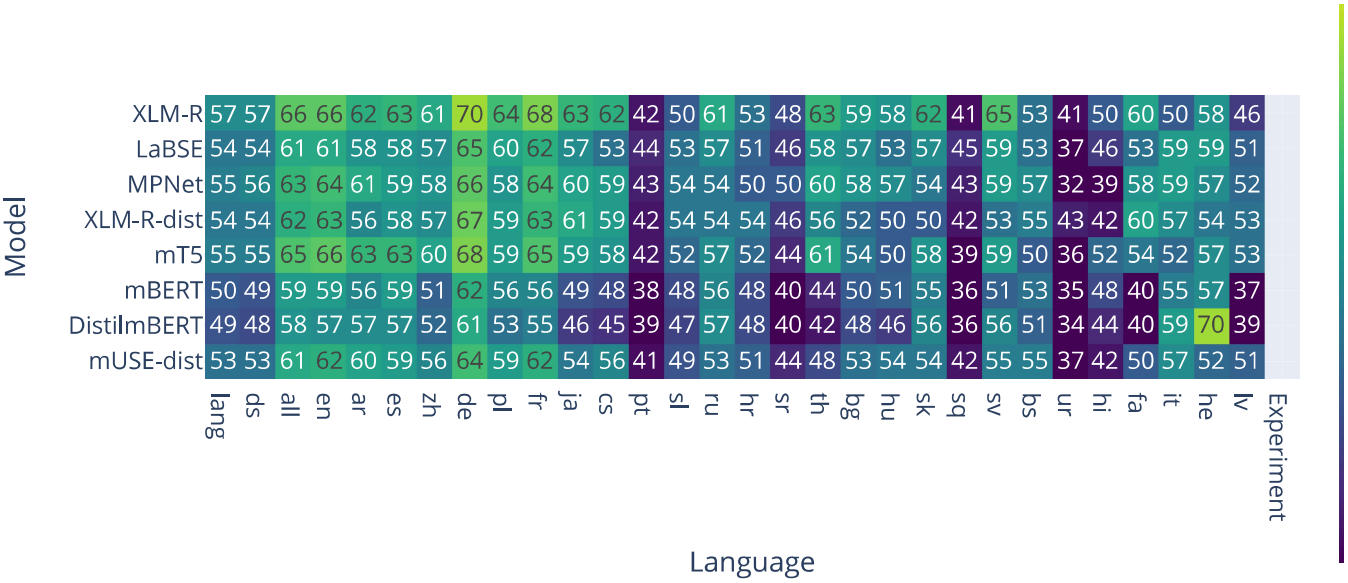
Model	Inf. time [s]	#params	#langs	base	data	reference
mT5	1.69	277M	101	T5	CC^b	(Xue et al. 2021)
LASER	1.64	52M	93	BiLSTM	$OPUS^c$	(Artetxe and Schwenk 2019)
mBERT	1.49	177M	104	BERT	Wiki	(Devlin et al. 2019)
MPNet**	1.38	278M	53	XLM-R	$OPUS^c$, $MUSE^d$, $Wikititles^e$	(Reimers and Gurevych 2020)
XLM-R-dist**	1.37	278M	53	XLM-R	$OPUS^c$, $MUSE^d$, $Wikititles^e$	(Reimers and Gurevych 2020)
XLM-R	1.37	278M	100	XLM-R	CC	(Conneau et al. 2020)
LaBSE	1.36	470M	109	BERT	CC, Wiki + mined bitexts	(Feng et al. 2020)
DistilmBERT	0.79	134M	104	BERT	Wiki	(Sanh et al. 2020)
mUSE-dist**	0.79	134M	53	DistilmBERT	$OPUS^c$, $MUSE^d$, $Wikititles^e$	(Reimers and Gurevych 2020)
mUSE-transformer*	0.65	85M	16	transformer	mined QA + bitexts, SNLI	(Yang et al. 2020)

Model	Inf. time [s]	#params	#langs	base	data	reference
mUSE-cnn*	0.12	68M	16	CNN	mined QA + bitexts, SNLI	(Yang et al. 2020)

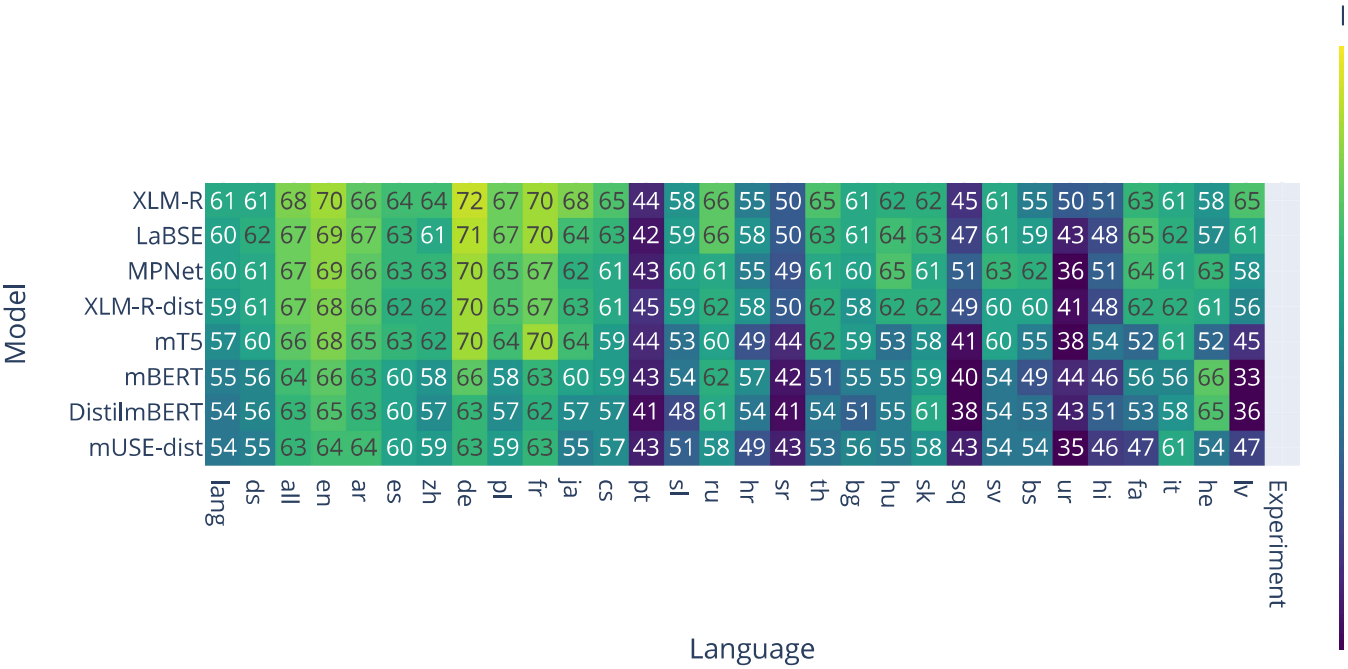
- * mUSE models were used in TensorFlow implementation in contrast to others in torch
- a Base model is either monolingual version on which it was based or another multilingual model which was used and adopted
- b Colossal Clean Crawled Corpus in multilingual version (mC4)
- c multiple datasets from OPUS website (https://opus.nlpl.eu)
- d bilingual dictionaries from MUSE (https://github.com/facebookresearch/MUSE)
- e just titles from wiki articles in multiple languages

Results





Fine-tuning



References

Artetxe, Mikel, and Holger Schwenk. 2019. “Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond.” *Transactions of the Association for Computational Linguistics* 7 (September): 597–610. https://doi.org/10.1162/tacl_a_00288.

Augustyniak, Łukasz, Szymon Woźniak, Marcin Gruza, Piotr Gramacki, Krzysztof Rajda, Mikołaj Morzy, and Tomasz Kajdanowicz. 2023. “Massively Multilingual Corpus of Sentiment Datasets and Multi-Faceted Sentiment Classification Benchmark.” https://arxiv.org/abs/2306.07902.

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. “Unsupervised Cross-Lingual Representation Learning at Scale.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–51. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.747.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–86. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423.

Feng, Fangxiaoyu, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2020. “Language-agnostic BERT Sentence Embedding.” *Computing Research Repository* arXiv:2007.01852. https://arxiv.org/abs/2007.01852.

Rajda, Krzysztof, Lukasz Augustyniak, Piotr Gramacki, Marcin Gruza, Szymon Woźniak, and Tomasz Kajdanowicz. 2022. “Assessment of Massively Multilingual Sentiment Classifiers.” In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, 125–40. Dublin, Ireland: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.wassa-1.13.

Reimers, Nils, and Iryna Gurevych. 2020. “Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation.” In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4512–25. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.365.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.” *Computing Research Repository* arXiv:1910.01108. https://arxiv.org/abs/1910.01108.

Xue, Linting, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. “MT5: A Massively Multilingual Pre-Trained Text-to-Text Transformer.” In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 483–98. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.41.

Yang, Yinfei, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, et al. 2020. “Multilingual Universal Sentence Encoder for Semantic Retrieval.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 87–94. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-demos.12.