

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

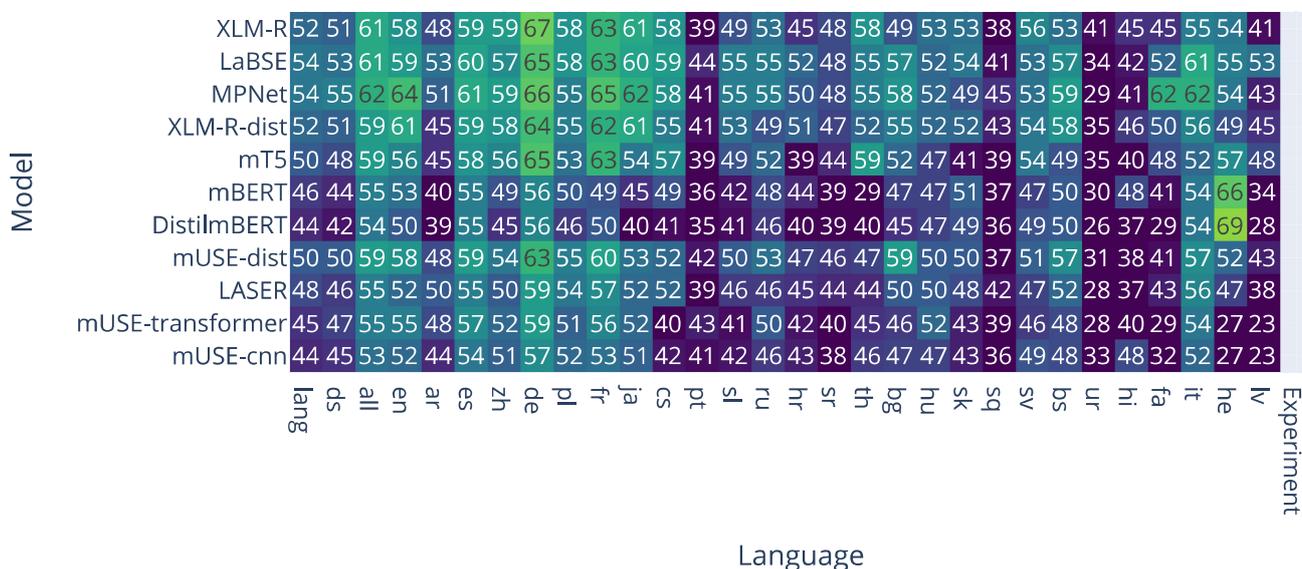
Model	Inf. time [s]	#params	#langs	base	data	reference
mT5	1.69	277M	101	T5	<i>CC</i> ^b	(Xue et al. 2021)
LASER	1.64	52M	93	BiLSTM	<i>OPUS</i> ^c	(Artetxe and Schwenk 2019)
mBERT	1.49	177M	104	BERT	Wiki	(Devlin et al. 2019)
MPNet**	1.38	278M	53	XLNet	<i>OPUS</i> ^c , <i>MUSE</i> ^d , <i>Wiktitles</i> ^e	(Reimers and Gurevych 2020)
XLNet-dist**	1.37	278M	53	XLNet	<i>OPUS</i> ^c , <i>MUSE</i> ^d , <i>Wiktitles</i> ^e	(Reimers and Gurevych 2020)
XLNet	1.37	278M	100	XLNet	CC	(Conneau et al. 2020)
LaBSE	1.36	470M	109	BERT	CC, Wiki + mined bitexts	(Feng et al. 2020)
DistilBERT	0.79	134M	104	BERT	Wiki	(Sanh et al. 2020)
mUSE-dist**	0.79	134M	53	DistilBERT	<i>OPUS</i> ^c , <i>MUSE</i> ^d , <i>Wiktitles</i> ^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

Linear Head



BiLSTM Head

Model	Language																				Experiment									
	lang	ds	all	en	ar	es	zh	de	pl	fr	ja	cs	pt	sl	ru	hr	sr	th	bg	hu		sk	sq	sv	bs	ur	hi	fa	it	he
XLM-R	57	57	66	66	62	63	61	70	64	68	63	62	42	50	61	53	48	63	59	58	62	41	65	53	41	50	60	50	58	46
LaBSE	54	54	61	61	58	58	57	65	60	62	57	53	44	53	57	51	46	58	57	53	57	45	59	53	37	46	53	59	59	51
MPNet	55	56	63	64	61	59	58	66	58	64	60	59	43	54	54	50	50	60	58	57	54	43	59	57	32	39	58	59	57	52
XLM-R-dist	54	54	62	63	56	58	57	67	59	63	61	59	42	54	54	54	46	56	52	50	50	42	53	55	43	42	60	57	54	53
mT5	55	55	65	66	63	63	60	68	59	65	59	58	42	52	57	52	44	61	54	50	58	39	59	50	36	52	54	52	57	53
mBERT	50	49	59	59	56	59	51	62	56	56	49	48	38	48	56	48	40	44	50	51	55	36	51	53	35	48	40	55	57	37
DistilmBERT	49	48	58	57	57	57	52	61	53	55	46	45	39	47	57	48	40	42	48	46	56	36	56	51	34	44	40	59	70	39
mUSE-dist	53	53	61	62	60	59	56	64	59	62	54	56	41	49	53	51	44	48	53	54	54	42	55	55	37	42	50	57	52	51

Fine-tuning

Model	Language																				Experiment									
	lang	ds	all	en	ar	es	zh	de	pl	fr	ja	cs	pt	sl	ru	hr	sr	th	bg	hu		sk	sq	sv	bs	ur	hi	fa	it	he
XLM-R	61	61	68	70	66	64	64	72	67	70	68	65	44	58	66	55	50	65	61	62	62	45	61	55	50	51	63	61	58	65
LaBSE	60	62	67	69	67	63	61	71	67	70	64	63	42	59	66	58	50	63	61	64	63	47	61	59	43	48	65	62	57	61
MPNet	60	61	67	69	66	63	63	70	65	67	62	61	43	60	61	55	49	61	60	65	61	51	63	62	36	51	64	61	63	58
XLM-R-dist	59	61	67	68	66	62	62	70	65	67	63	61	45	59	62	58	50	62	58	62	62	49	60	60	41	48	62	62	61	56
mT5	57	60	66	68	65	63	62	70	64	70	64	59	44	53	60	49	44	62	59	53	58	41	60	55	38	54	52	61	52	45
mBERT	55	56	64	66	63	60	58	66	58	63	60	59	43	54	62	57	42	51	55	55	59	40	54	49	44	46	56	56	66	33
DistilmBERT	54	56	63	65	63	60	57	63	57	62	57	57	41	48	61	54	41	54	51	55	61	38	54	53	43	51	53	58	65	36
mUSE-dist	54	55	63	64	64	60	59	63	59	63	55	57	43	51	58	49	43	53	56	55	58	43	54	54	35	46	47	61	54	47

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, Łukasz 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>.