Gender Bias in Masked Language Models for Multiple Languages

1 Masked Language Models

We used Masked Language Models (**MLMs**) in eight languages for our experiments in the main paper: Japanese[1], German[2] (Chan, Schweter, and Möller 2020), Arabic[3] (Antoun, Baly, and Hajj 2020), Spanish[4] (Cañete et al. 2020), Portuguese[5] (Souza, Nogueira, and Lotufo 2020), Russian[6], Indonesian[7] and Chinese[8] (Cui et al. 2020).

As English MLMs, we use BERT[9], multilingual BERT[10] (Devlin et al. 2019), RoBERTa[11] (Liu et al. 2019), ALBERT[12] (Lan et al. 2019), DistilBERT[13], DistilRoBERTa[14] (Sanh et al. 2019), ConvBERT[15] (Jiang et al. 2020), XLM[16] (Conneau and Lample 2019), and Deberta[17] (He et al. 2020).

[1] https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking

[2] https://huggingface.co/deepset/gbert-base

[3] https://huggingface.co/aubmindlab/bert-base-arabertv01

[4] https://huggingface.co/dccuchile/bert-base-spanish-wwm-uncased

[5] https://huggingface.co/neuralmind/bert-base-portuguese-cased

[6] https://huggingface.co/blinoff/roberta-base-russian-v0

[7] https://huggingface.co/cahya/bert-base-indonesian-522M

[8] https://huggingface.co/hfl/chinese-bert-wwm-ext

[9] https://huggingface.co/bert-base-cased and https://huggingface.co/bert-large-uncased

[10] https://huggingface.co/bert-base-multilingual-uncased

[11] https://huggingface.co/roberta-base and https://huggingface.co/roberta-large

[12] https://huggingface.co/albert-base-v2

[13] https://huggingface.co/distilbert-base-cased

[14] https://huggingface.co/distilroberta-base

[15] https://huggingface.co/YituTech/conv-bert-medium-small

[16] https://huggingface.co/xlm-mlm-100-1280

[17] https://huggingface.co/microsoft/deberta-xlarge-v2

For Japanese MLMs, we evaluate four Japanese BERT models (**base-subword**[18], **large-subword**[19], **base-char**[20], **large-char**[21]), subword-based and character-based, with base and large sizes. For Russian, we use two MLMs – one trained on Wikipedia and news data (**wiki&news**)[22] and the other on OpenSubtitles (Lison and Tiedemann 2016) and SNS data (Shavrina and Shapovalova 2017).

References

Antoun, W.; Baly, F.; and Hajj, H. 2020. AraBERT: Transformer-based Model for Arabic Language Understanding. In *OSACT4*, 9–15. European Language Resource Association. ISBN 979-10-95546-51-1.

Cañete, J.; Chaperon, G.; Fuentes, R.; Ho, J.-H.; Kang, H.; and Pérez, J. 2020. Spanish Pre-Trained BERT Model and Evaluation Data. In *PML4DC*.

Chan, B.; Schweter, S.; and Möller, T. 2020. German's Next Language Model. In *COLING*, 6788–6796. International Committee on Computational Linguistics.

Conneau, A.; and Lample, G. 2019. Cross-lingual Language Model Pretraining. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *NeurIPS*, volume 32. Curran Associates, Inc.

Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; and Hu, G. 2020. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In *EMNLP: Findings*, 657–668. Association for Computational Linguistics.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*, 4171–4186. Association for Computational Linguistics.

[18] https://huggingface.co/cl-tohoku/bert-base-japanese-v2

[19] https://huggingface.co/cl-tohoku/bert-large-japanese

[20] https://huggingface.co/cl-tohoku/bert-base-japanese-char-v2

[21] https://huggingface.co/cl-tohoku/bert-large-japanese-char

[22] https://huggingface.co/DeepPavlov/rubert-base-cased

He, P.; Liu, X.; Gao, J.; and Chen, W. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Jiang, Z.-H.; Yu, W.; Zhou, D.; Chen, Y.; Feng, J.; and Yan, S. 2020. ConvBERT: Improving BERT with Span-based Dynamic Convolution. *Advances in Neural Information Processing Systems*, 33.

Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; and Soricut, R. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Lison, P.; and Tiedemann, J. 2016. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In *LREC*, 923–929. European Language Resources Association (ELRA).

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Sanh, V.; Debut, L.; Chaumond, J.; and Wolf, T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Shavrina, T.; and Shapovalova, O. 2017. To the methodology of corpus construction for machine learning:taiga syntax tree corpus and parser. *Corpora*, 78–84.

Souza, F.; Nogueira, R.; and Lotufo, R. 2020. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *BRACIS*.