cases	doc_1		doc_2		decision	id
			authors	• Linhan Zhang • Qian Chen • Wen Wang • Chong Deng • Xin Cao • Kongzhang Hao • Yuxin Jiang • Wei Wang	DUPLICATES	99
	authors	• Zhang, Linhan • Chen, Qian • Wang, Wen • Deng, Chong • Cao, Xin • Hao, Kongzhang • Jiang, Yuxin • Wang, Wei	**title**	Weighted Sampling for Masked Language Modeling		
	title	Weighted Sampling for Masked Language Modeling	**publication_date**	2023-02-28 01:07:39+00:00		
	publication_date	2023-02-27 00:00:00	**source**	SupportedSources.ARXIV		
	source	SupportedSources.OPENALEX	**journal**	2023 IEEE International Conference on Acoustics, Speech and Signal Processing		
	journal	arXiv (Cornell University)	**volume**			
	volume		**doi**			
	doi	10.48550/arxiv.2302.14225	**urls**	• http://arxiv.org/pdf/2302.14225v1 • http://arxiv.org/abs/2302.14225v1 • http://arxiv.org/pdf/2302.14225v1		
	urls	• https://openalex.org/W4322759629 • https://doi.org/10.48550/arxiv.2302.14225	**id**	id6527591747762610780		
	id	id7715361091835219495	**abstract**	Masked Language Modeling (MLM) is widely used to pretrain language models. The standard random masking strategy in MLM causes the pre-trained language models (PLMs) to be biased toward high-frequency tokens. Representation learning of rare tokens is poor and PLMs have limited performance on downstream tasks. To alleviate this frequency bias issue, we propose two simple and effective Weighted Sampling strategies for masking tokens based on the token frequency and training loss. We apply these two strategies to BERT and obtain Weighted-Sampled BERT (WSBERT). Experiments on the Semantic Textual Similarity benchmark (STS) show that WSBERT significantly improves sentence embeddings over BERT. Combining WSBERT with calibration methods and prompt learning further improves sentence embeddings. We also investigate fine-tuning WSBERT on the GLUE benchmark and show that Weighted Sampling also improves the transfer learning capability of the backbone PLM. We further analyze and provide insights into how WSBERT improves token embeddings.		
	abstract		**versions**			
	versions					