

cases	doc_1		doc_2		decision	id
					DUPLICATES	175
	authors	<ul style="list-style-type: none">Hyun-Jin ChoiJudong KimSeongho JoeYoungjune Gwon	authors	<ul style="list-style-type: none">Hyunjin ChoiJudong KimSeongho JoeYoungjune Gwon		
	title	Evaluation of BERT and ALBERT Sentence Embedding Performance on Downstream NLP Tasks	title	Evaluation of BERT and ALBERT Sentence Embedding Performance on Downstream NLP Tasks		
	publication_date	2021-01-10 00:00:00	publication_date	2021-01-26 09:14:06+00:00		
	source	SupportedSources.OPENALEX	source	SupportedSources.ARXIV		
	journal	International Conference on Pattern Recognition	journal	None		
	volume		volume			
	doi	10.1109/icpr48806.2021.9412102	doi			
	urls	<ul style="list-style-type: none">https://openalex.org/W3162462834https://doi.org/10.1109/icpr48806.2021.9412102http://arxiv.org/pdf/2101.10642	urls	<ul style="list-style-type: none">http://arxiv.org/pdf/2101.10642v1http://arxiv.org/abs/2101.10642v1http://arxiv.org/pdf/2101.10642v1		
	id	id-9031418690049624146	id	id7220721565786038325		
	abstract		abstract	Contextualized representations from a pre-trained language model are central to achieve a high performance on downstream NLP task. The pre-trained BERT and A Lite BERT (ALBERT) models can be fine-tuned to give state-of-the-art results in sentence-pair regressions such as semantic textual similarity (STS) and natural language inference (NLI). Although BERT-based models yield the [CLS] token vector as a reasonable sentence embedding, the search for an optimal sentence embedding scheme remains an active research area in computational linguistics. This paper explores on sentence embedding models for BERT and ALBERT. In particular, we take a modified BERT network with siamese and triplet network structures called Sentence-BERT (SBERT) and replace BERT with ALBERT to create Sentence-ALBERT (SALBERT). We also experiment with an outer CNN sentence-embedding network for SBERT and SALBERT. We evaluate performances of all sentence-embedding models considered using the STS and NLI datasets. The empirical results indicate that our CNN architecture improves ALBERT models substantially more than BERT models for STS benchmark. Despite significantly fewer model parameters, ALBERT sentence embedding is highly competitive to BERT in downstream NLP evaluations.		
	versions		versions			