

cases	doc_1		doc_2				decision	id
							NOT DUPLICATES	417
	authors	<ul style="list-style-type: none">Yan ZhangRuidan HeZuozhu LiuKwan Hui LimLidong Bing	authors	<ul style="list-style-type: none">Yan ZhangRuidan HeZuozhu LiuKwan Hui LimLidong Bing				
	title	An Unsupervised Sentence Embedding Method byMutual Information Maximization	title	An Unsupervised Sentence Embedding Method by Mutual Information Maximization				
	publication_date	2020-09-25 00:00:00	publication_date	2020-09-25 00:00:00				
	source	SupportedSources.OPENALEX	source	SupportedSources.SEMANTIC_SCHOLAR				
	journal	arXiv (Cornell University)	journal					
	volume		volume					
	doi	None	doi	10.18653/v1/2020.emnlp-main.124				
	urls	<ul style="list-style-type: none">https://openalex.org/W3089244638	urls	<ul style="list-style-type: none">https://www.semanticscholar.org/paper/574c7c7260a795d2906f03b5229ace3f54ee0ab3				
	id	id-4688814547794997643	id	id-1858939200859417787				
	abstract		abstract	BERT is inefficient for sentence-pair tasks such as clustering or semantic search as it needs to evaluate combinatorially many sentence pairs which is very time-consuming. Sentence BERT (SBERT) attempted to solve this challenge by learning semantically meaningful representations of single sentences, such that similarity comparison can be easily accessed. However, SBERT is trained on corpus with high-quality labeled sentence pairs, which limits its application to tasks where labeled data is extremely scarce. In this paper, we propose a lightweight extension on top of BERT and a novel self-supervised learning objective based on mutual information maximization strategies to derive meaningful sentence embeddings in an unsupervised manner. Unlike SBERT, our method is not restricted by the availability of labeled data, such that it can be applied on different domain-specific corpus. Experimental results show that the proposed method significantly outperforms other unsupervised sentence embedding baselines on common semantic textual similarity (STS) tasks and downstream supervised tasks. It also outperforms SBERT in a setting where in-domain labeled data is not available, and achieves performance competitive with supervised methods on various tasks.				
	versions		versions					