		doc_1		doc_2	decision	id
cases	**authors**	• Danqi Chen • Xingcheng Yao • Tianyu Gao	**authors**	• Tianyu Gao • Xingcheng Yao • Danqi Chen	DUPLICATES	241
	title	SimCSE: Simple Contrastive Learning of Sentence Embeddings	**title**	SimCSE: Simple Contrastive Learning of Sentence Embeddings		
	publication_date	2021-04-18 00:00:00	**publication_date**	2021-04-18 00:00:00		
	source	SupportedSources.PAPERS_WITH_CODE	**source**	SupportedSources.SEMANTIC_SCHOLAR		
	journal		**journal**	ArXiv		
	volume		**volume**	abs/2104.08821		
	doi		**doi**	10.18653/v1/2021.emnlp-main.552		
	urls	• https://arxiv.org/pdf/2104.08821v4.pdf • https://github.com/PaddlePaddle/PaddleNLP/tree/develop/examples/text_matching/simcse • https://aclanthology.org/2021.emnlp-main.552.pdf	**urls**	• https://www.semanticscholar.org/paper/c26759e6c701201af2f62f7ee4eb68742b5bf085		
			id	id-4280444153600027295		
	id	id7918049846954081274				
	abstract	This paper presents SimCSE, a simple contrastive learning framework that greatly advances state-of-the-art sentence embeddings. We first describe an unsupervised approach, which takes an input sentence and predicts itself in a contrastive objective, with only standard dropout used as noise. This simple method works surprisingly well, performing on par with previous supervised counterparts. We find that dropout acts as minimal data augmentation, and removing it leads to a representation collapse. Then, we propose a supervised approach, which incorporates annotated pairs from natural language inference datasets into our contrastive learning framework by using "entailment" pairs as positives and "contradiction" pairs as hard negatives. We evaluate SimCSE on standard semantic textual similarity (STS) tasks, and our unsupervised and supervised models using BERT base achieve an average of 76.3% and 81.6% Spearman's correlation respectively, a 4.2% and 2.2% improvement compared to the previous best results. We also show -- both theoretically and empirically -- that the contrastive learning objective regularizes pre-trained embeddings' anisotropic space to be more uniform, and it better aligns positive pairs when supervised signals are available.	**abstract**	This paper presents SimCSE, a simple contrastive learning framework that greatly advances the state-of-the-art sentence embeddings. We first describe an unsupervised approach, which takes an input sentence and predicts itself in a contrastive objective, with only standard dropout used as noise. This simple method works surprisingly well, performing on par with previous supervised counterparts. We find that dropout acts as minimal data augmentation and removing it leads to a representation collapse. Then, we propose a supervised approach, which incorporates annotated pairs from natural language inference datasets into our contrastive learning framework, by using â€œentailmentâ€ pairs as positives and â€œcontradictionâ€ pairs as hard negatives. We evaluate SimCSE on standard semantic textual similarity (STS) tasks, and our unsupervised and supervised models using BERT base achieve an average of 76.3% and 81.6% Spearmanâ€™s correlation respectively, a 4.2% and 2.2% improvement compared to previous best results. We also showâ€”both theoretically and empiricallyâ€”that contrastive learning objective regularizes pre-trained embeddingsâ€™ anisotropic space to be more uniform, and it better aligns positive pairs when supervised signals are available.		
	versions		**versions**			