

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none">Xuandong ZhaoZhiguo YuMing WuLei Li	authors	<ul style="list-style-type: none">Xuandong ZhaoZhiguo YuMing-li WuLei Li	DUPLICATES	127
	title	Compressing Sentence Representation for Semantic Retrieval via Homomorphic Projective Distillation	title	Compressing Sentence Representation for Semantic Retrieval via Homomorphic Projective Distillation		
	publication_date	2022-03-15 07:05:43+00:00	publication_date	2022-03-15 00:00:00		
	source	SupportedSources.ARXIV	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal	None	journal	ArXiv		
	volume		volume	abs/2203.07687		
	doi		doi	10.48550/arXiv.2203.07687		
	urls	<ul style="list-style-type: none">http://arxiv.org/pdf/2203.07687v1http://arxiv.org/abs/2203.07687v1http://arxiv.org/pdf/2203.07687v1	urls	<ul style="list-style-type: none">https://www.semanticscholar.org/paper/b3a42f3ffb7d08fa9e57159c98543302c489b436		
	id	id1290883239565901286	id	id5606957588713600798		
	abstract	How to learn highly compact yet effective sentence representation? Pre-trained language models have been effective in many NLP tasks. However, these models are often huge and produce large sentence embeddings. Moreover, there is a big performance gap between large and small models. In this paper, we propose Homomorphic Projective Distillation (HPD) to learn compressed sentence embeddings. Our method augments a small Transformer encoder model with learnable projection layers to produce compact representations while mimicking a large pre-trained language model to retain the sentence representation quality. We evaluate our method with different model sizes on both semantic textual similarity (STS) and semantic retrieval (SR) tasks. Experiments show that our method achieves 2.7-4.5 points performance gain on STS tasks compared with previous best representations of the same size. In SR tasks, our method improves retrieval speed (8.2\times\$) and memory usage (8.0\times\$) compared with state-of-the-art large models.	abstract	How to learn highly compact yet effective sentence representation? Pre-trained language models have been effective in many NLP tasks. However, these models are often huge and produce large sentence embeddings. Moreover, there is a big performance gap between large and small models. In this paper, we propose Homomorphic Projective Distillation (HPD) to learn compressed sentence embeddings. Our method augments a small Transformer encoder model with learnable projection layers to produce compact representations while mimicking a large pre-trained language model to retain the sentence representation quality. We evaluate our method with different model sizes on both semantic textual similarity (STS) and semantic retrieval (SR) tasks. Experiments show that our method achieves 2.7-4.5 points performance gain on STS tasks compared with previous best representations of the same size. In SR tasks, our method improves retrieval speed (8.2\text{---}) and memory usage (8.0\text{---}) compared with state-of-the-art large models. Our implementation is available at https://github.com/XuandongZhao/HPD.		
	versions		versions			