cases	doc_1		doc_2		decision	id
	authors	• Yixuan Su and Fangyu Liu and Zaiqiao Meng and Tian Lan and Lei Shu and Ehsan Shareghi and Nigel Collier	**authors**	• Collier, Nigel • Lan, Tian • Liu, Fangyu • Meng, Zaiqiao • Shareghi, Ehsan • Shu, Lei • Su, Yixuan	DUPLICATES	141
	title	TaCL: Improving BERT Pre-training with Token-aware Contrastive Learning	**title**	TaCL: Improving BERT Pre-training with Token-aware Contrastive Learning		
	publication_date	2022-04-28 00:00:00	**publication_date**	2022-01-01 00:00:00		
	source	SupportedSources.INTERNET_ARCHIVE	**source**	SupportedSources.CORE		
	journal		**journal**			
	volume		**volume**			
	doi		**doi**	None		
	urls	• https://web.archive.org/web/20220504130909/https://arxiv.org/pdf/2111.04198v4.pdf	**urls**	• http://arxiv.org/abs/2111.04198		
	id	id-1118966649645273413	**id**	id4445726963631386986		
	abstract	Masked language models (MLMs) such as BERT and RoBERTa have revolutionized the field of Natural Language Understanding in the past few years. However, existing pre-trained MLMs often output an anisotropic distribution of token representations that occupies a narrow subset of the entire representation space. Such token representations are not ideal, especially for tasks that demand discriminative semantic meanings of distinct tokens. In this work, we propose TaCL (Token-aware Contrastive Learning), a novel continual pre-training approach that encourages BERT to learn an isotropic and discriminative distribution of token representations. TaCL is fully unsupervised and requires no additional data. We extensively test our approach on a wide range of English and Chinese benchmarks. The results show that TaCL brings consistent and notable improvements over the original BERT model. Furthermore, we conduct detailed analysis to reveal the merits and inner-workings of our approach.	**abstract**	Masked language models (MLMs) such as BERT and RoBERTa have revolutionized the field of Natural Language Understanding in the past few years. However, existing pre-trained MLMs often output an anisotropic distribution of token representations that occupies a narrow subset of the entire representation space. Such token representations are not ideal, especially for tasks that demand discriminative semantic meanings of distinct tokens. In this work, we propose TaCL (Token-aware Contrastive Learning), a novel continual pre-training approach that encourages BERT to learn an isotropic and discriminative distribution of token representations. TaCL is fully unsupervised and requires no additional data. We extensively test our approach on a wide range of English and Chinese benchmarks. The results show that TaCL brings consistent and notable improvements over the original BERT model. Furthermore, we conduct detailed analysis to reveal the merits and inner-workings of our approach.Comment: Camera-ready for NAACL 202		
	versions		**versions**			