

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none"><li>Zilu Tang</li><li>Muhammed Yusuf Kocyigit</li><li>Derry Wijaya</li></ul>	authors	<ul style="list-style-type: none"><li>Zilu Tang</li><li>Muhammed Yusuf Kocyigit</li><li>D. Wijaya</li></ul>	DUPLICATES	120
	title	AugCSE: Contrastive Sentence Embedding with Diverse Augmentations	title	AugCSE: Contrastive Sentence Embedding with Diverse Augmentations		
	publication_date	2022-10-20 03:52:07+00:00	publication_date	2022-10-20 00:00:00		
	source	SupportedSources.ARXIV	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal	None	journal	ArXiv		
	volume		volume	abs/2210.13749		
	doi		doi	10.48550/arXiv.2210.13749		
	urls	<ul style="list-style-type: none"><li>http://arxiv.org/pdf/2210.13749v1</li><li>http://arxiv.org/abs/2210.13749v1</li><li>http://arxiv.org/pdf/2210.13749v1</li></ul>	urls	<ul style="list-style-type: none"><li>https://www.semanticscholar.org/paper/5fd93eda50213f52d7f19f14efeb3ca5c58dc7bb</li></ul>		
	id	id6748581958260531620	id	id7830601838935962389		
	abstract	Data augmentation techniques have been proven useful in many applications in NLP fields. Most augmentations are task-specific, and cannot be used as a general-purpose tool. In our work, we present AugCSE, a unified framework to utilize diverse sets of data augmentations to achieve a better, general purpose, sentence embedding model. Building upon the latest sentence embedding models, our approach uses a simple antagonistic discriminator that differentiates the augmentation types. With the finetuning objective borrowed from domain adaptation, we show that diverse augmentations, which often lead to conflicting contrastive signals, can be tamed to produce a better and more robust sentence representation. Our methods achieve state-of-the-art results on downstream transfer tasks and perform competitively on semantic textual similarity tasks, using only unsupervised data.	abstract	Data augmentation techniques have been proven useful in many applications in NLP fields. Most augmentations are task-specific, and cannot be used as a general-purpose tool. In our work, we present AugCSE, a unified framework to utilize diverse sets of data augmentations to achieve a better, general-purpose, sentence embedding model. Building upon the latest sentence embedding models, our approach uses a simple antagonistic discriminator that differentiates the augmentation types. With the finetuning objective borrowed from domain adaptation, we show that diverse augmentations, which often lead to conflicting contrastive signals, can be tamed to produce a better and more robust sentence representation. Our methods achieve state-of-the-art results on downstream transfer tasks and perform competitively on semantic textual similarity tasks, using only unsupervised data.		
	versions		versions			