cases	doc_1	doc_2	decision	id
	see doc_1 below	see doc_2 below	DUPLICATES	353

doc_1

authors	• Asa Cooper Stickland • Iain Murray
title	BERT and PALs: Projected Attention Layers for Efficient Adaptation in Multi-Task Learning
publication_date	2019-02-07 15:05:46+00:00
source	SupportedSources.ARXIV
journal	None
volume	
doi	
urls	• http://arxiv.org/pdf/1902.02671v2 • http://arxiv.org/abs/1902.02671v2 • http://arxiv.org/pdf/1902.02671v2
id	id7430136925069115060
abstract	Multi-task learning shares information between related tasks, sometimes reducing the number of parameters required. State-of-the-art results across multiple natural language understanding tasks in the GLUE benchmark have previously used transfer from a single large task: unsupervised pre-training with BERT, where a separate BERT model was fine-tuned for each task. We explore multi-task approaches that share a single BERT model with a small number of additional task-specific parameters. Using new adaptation modules, PALs or `projected attention layers', we match the performance of separately fine-tuned models on the GLUE benchmark with roughly 7 times fewer parameters, and obtain state-of-the-art results on the Recognizing Textual Entailment dataset.
versions	

doc_2

authors	• Cooper Stickland, Asa • Murray, Iain
title	BERT and PALs: Projected Attention Layers for Efficient Adaptation in Multi-Task Learning
publication_date	2019-05-15 00:00:00
source	SupportedSources.CORE
journal	
volume	
doi	None
urls	• https://core.ac.uk/download/322483062.pdf
id	id-4417393995638584962
abstract	Multi-task learning shares information between related tasks, sometimes reducing the number of parameters required. State-of-the-art results across multiple natural language understanding tasks in the GLUE benchmark have previously used transfer from a single large task: unsupervised pre-training with BERT, where a separate BERT model was fine-tuned for each task. We explore multi-task approaches that share a single BERT model with a small number of additional task-specific parameters. Using new adaptation modules, PALs or `projected attention layers', we match the performance of separately fine-tuned models on the GLUE benchmark with roughly 7 times fewer parameters, and obtain state-of-the-art results on the Recognizing Textual Entailment dataset.Comment: Accepted for publication at ICML 201
versions	