	doc_1		doc_2		decision	id
cases						

doc_1	
authors	• Bhattacharjee, K. • Ballesteros, M. • Anubhai, R. • Muresan, S. • Ma, J. • Ladhak, F. • Al-Onaizan, Y.
title	To BERT or Not to BERT: Comparing Task-specific and Task-agnostic Semi-Supervised Approaches for Sequence Tagging
publication_date	2020-01-01 00:00:00
source	SupportedSources.CROSSREF
journal	
volume	
doi	10.18653/v1/2020.emnlp-main.636
urls	• http://dx.doi.org/10.18653/v1/2020.emnlp-main.636
id	id-6747516891727132736
abstract	
versions	

doc_2	
authors	• Kasturi Bhattacharjee • Miguel Ballesteros • Rishita Anubhai • Smaranda Muresan • Jie Ma • Faisal Ladhak • Yaser Al-Onaizan
title	To BERT or Not to BERT: Comparing Task-specific and Task-agnostic Semi-Supervised Approaches for Sequence Tagging
publication_date	2020-10-27 04:03:47+00:00
source	SupportedSources.ARXIV
journal	None
volume	
doi	
urls	• http://arxiv.org/pdf/2010.14042v1 • http://arxiv.org/abs/2010.14042v1 • http://arxiv.org/pdf/2010.14042v1
id	id-6131347636948391843
abstract	Leveraging large amounts of unlabeled data using Transformer-like architectures, like BERT, has gained popularity in recent times owing to their effectiveness in learning general representations that can then be further fine-tuned for downstream tasks to much success. However, training these models can be costly both from an economic and environmental standpoint. In this work, we investigate how to effectively use unlabeled data: by exploring the task-specific semi-supervised approach, Cross-View Training (CVT) and comparing it with task-agnostic BERT in multiple settings that include domain and task relevant English data. CVT uses a much lighter model architecture and we show that it achieves similar performance to BERT on a set of sequence tagging tasks, with lesser financial and environmental impact.
versions	

decision	id
DUPLICATES	289