

cases	doc_1		doc_2		decision	id
					DUPLICATES	322
	authors	<ul style="list-style-type: none">Yuki AraseJun'ichi Tsujii	authors	<ul style="list-style-type: none">Yuki AraseJunichi Tsujii		
	title	Transfer Fine-Tuning: A BERT Case Study	title	Transfer Fine-Tuning: A BERT Case Study		
	publication_date	2019-09-01 00:00:00	publication_date	2019-09-03 00:00:00		
	source	SupportedSources.OPENALEX	source	SupportedSources.INTERNET_ARCHIVE		
	journal	arXiv (Cornell University)	journal			
	volume		volume			
	doi	10.18653/v1/d19-1542	doi			
	urls	<ul style="list-style-type: none">https://openalex.org/W2988707515https://doi.org/10.18653/v1/d19-1542https://www.aclweb.org/anthology/D19-1542.pdf	urls	<ul style="list-style-type: none">https://web.archive.org/web/20200924234559/https://arxiv.org/pdf/1909.00931v1.pdf		
	id	id257162581231611739	id	id166437410020634166		
	abstract		abstract	A semantic equivalence assessment is defined as a task that assesses semantic equivalence in a sentence pair by binary judgment (i.e., paraphrase identification) or grading (i.e., semantic textual similarity measurement). It constitutes a set of tasks crucial for research on natural language understanding. Recently, BERT realized a breakthrough in sentence representation learning (Devlin et al., 2019), which is broadly transferable to various NLP tasks. While BERT's performance improves by increasing its model size, the required computational power is an obstacle preventing practical applications from adopting the technology. Herein, we propose to inject phrasal paraphrase relations into BERT in order to generate suitable representations for semantic equivalence assessment instead of increasing the model size. Experiments on standard natural language understanding tasks confirm that our method effectively improves a smaller BERT model while maintaining the model size. The generated model exhibits superior performance compared to a larger BERT model on semantic equivalence assessment tasks. Furthermore, it achieves larger performance gains on tasks with limited training datasets for fine-tuning, which is a property desirable for transfer learning.		
	versions		versions			