cases	field	doc_1	doc_2	decision	id
cases	**authors**	• GlavaÅ¡, Goran • Korhonen, Anna • Lauscher, Anne • Ponti, Edoardo Maria • VuliÄ‡, Ivan	• Anne Lauscher • Ivan VuliÄ‡ • Edoardo Maria Ponti • Anna Korhonen • Goran GlavaÅ¡	DUPLICATES	47
	title	Specializing unsupervised pretraining models for word-level semantic similarity	Specializing Unsupervised Pretraining Models for Word-Level Semantic Similarity		
	publication_date	2020-01-01 00:00:00	2019-09-05 11:49:40+00:00		
	source	SupportedSources.CORE	SupportedSources.ARXIV		
	journal		None		
	volume				
	doi	10.18653/v1/2020.coling-main.118			
	urls	• https://core.ac.uk/download/425745327.pdf	• http://arxiv.org/pdf/1909.02339v2 • http://arxiv.org/abs/1909.02339v2 • http://arxiv.org/pdf/1909.02339v2		
	id	id7062737839762778110	id-3847900239914390189		
	abstract	Unsupervised pretraining models have been shown to facilitate a wide range of downstream NLP applications. These models, however, retain some of the limitations of traditional static word embeddings. In particular, they encode only the distributional knowledge available in raw text corpora, incorporated through language modeling objectives. In this work, we complement such distributional knowledge with external lexical knowledge, that is, we integrate the discrete knowledge on word-level semantic similarity into pretraining. To this end, we generalize the standard BERT model to a multi-task learning setting where we couple BERT's masked language modeling and next sentence prediction objectives with an auxiliary task of binary word relation classification. Our experiments suggest that our "Lexically Informed" BERT (LIBERT), specialized for the word-level semantic similarity, yields better performance than the lexically blind "vanilla" BERT on several language understanding tasks. Concretely, LIBERT outperforms BERT in 9 out of 10 tasks of the GLUE benchmark and is on a par with BERT in the remaining one. Moreover, we show consistent gains on 3 benchmarks for lexical simplification, a task where knowledge about word-level semantic similarity is paramount	Unsupervised pretraining models have been shown to facilitate a wide range of downstream NLP applications. These models, however, retain some of the limitations of traditional static word embeddings. In particular, they encode only the distributional knowledge available in raw text corpora, incorporated through language modeling objectives. In this work, we complement such distributional knowledge with external lexical knowledge, that is, we integrate the discrete knowledge on word-level semantic similarity into pretraining. To this end, we generalize the standard BERT model to a multi-task learning setting where we couple BERT's masked language modeling and next sentence prediction objectives with an auxiliary task of binary word relation classification. Our experiments suggest that our "Lexically Informed" BERT (LIBERT), specialized for the word-level semantic similarity, yields better performance than the lexically blind "vanilla" BERT on several language understanding tasks. Concretely, LIBERT outperforms BERT in 9 out of 10 tasks of the GLUE benchmark and is on a par with BERT in the remaining one. Moreover, we show consistent gains on 3 benchmarks for lexical simplification, a task where knowledge about word-level semantic similarity is paramount.		
	versions				