cases

	doc_1	doc_2	decision	id
authors	• Anne Lauscher • Ivan VuliÄ‡ • Edoardo Maria Ponti • Anna Korhonen • Goran GlavaÅ¡	• Anne Lauscher • Ivan Vulic • E. Ponti • A. Korhonen • Goran Glavavs	DUPLICATES	354
title	Specializing Unsupervised Pretraining Models for Word-Level Semantic Similarity	Specializing Unsupervised Pretraining Models for Word-Level Semantic Similarity		
publication_date	2019-09-05 11:49:40+00:00	2019-09-05 00:00:00		
source	SupportedSources.ARXIV	SupportedSources.SEMANTIC_SCHOLAR		
journal	None			
volume				
doi		10.18653/V1/2020.COLING-MAIN.118		
urls	• http://arxiv.org/pdf/1909.02339v2 • http://arxiv.org/abs/1909.02339v2 • http://arxiv.org/pdf/1909.02339v2	• https://www.semanticscholar.org/paper/3af5a41e43158a75bf7a8bb3f9517edc4163b3ca		
id	id-3847900239914390189	id-174404111900855418		
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 BERTâ€™s masked language modeling and next sentence prediction objectives with an auxiliary task of binary word relation classification. Our experiments suggest that our â€œLexically Informedâ€ BERT (LIBERT), specialized for the word-level semantic similarity, yields better performance than the lexically blind â€œvanillaâ€ 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, as well as large gains on lexical reasoning probes.		
versions				