cases

doc_1		doc_2		decision	id
authors	• Bohan Li • Hao Zhou • Junxian He • Mingxuan Wang • Yiming Yang • Lei Li	**authors**	• Bohan Li and Hao Zhou and Junxian He and Mingxuan Wang and Yiming Yang and Lei Li	DUPLICATES	43
title	On the Sentence Embeddings from Pre-trained Language Models	**title**	On the Sentence Embeddings from Pre-trained Language Models		
publication_date	2020-11-02 13:14:57+00:00	**publication_date**	2020-11-02 00:00:00		
source	SupportedSources.ARXIV	**source**	SupportedSources.INTERNET_ARCHIVE		
journal	None	**journal**			
volume		**volume**			
doi		**doi**			
urls	• http://arxiv.org/pdf/2011.05864v1 • http://arxiv.org/abs/2011.05864v1 • http://arxiv.org/pdf/2011.05864v1	**urls**	• https://web.archive.org/web/20201113050429/https://arxiv.org/pdf/2011.05864v1.pdf		
id	id7619983922342773880	**id**	id3310671189580720983		
abstract	Pre-trained contextual representations like BERT have achieved great success in natural language processing. However, the sentence embeddings from the pre-trained language models without fine-tuning have been found to poorly capture semantic meaning of sentences. In this paper, we argue that the semantic information in the BERT embeddings is not fully exploited. We first reveal the theoretical connection between the masked language model pre-training objective and the semantic similarity task theoretically, and then analyze the BERT sentence embeddings empirically. We find that BERT always induces a non-smooth anisotropic semantic space of sentences, which harms its performance of semantic similarity. To address this issue, we propose to transform the anisotropic sentence embedding distribution to a smooth and isotropic Gaussian distribution through normalizing flows that are learned with an unsupervised objective. Experimental results show that our proposed BERT-flow method obtains significant performance gains over the state-of-the-art sentence embeddings on a variety of semantic textual similarity tasks. The code is available at https://github.com/bohanli/BERT-flow.	**abstract**	Pre-trained contextual representations like BERT have achieved great success in natural language processing. However, the sentence embeddings from the pre-trained language models without fine-tuning have been found to poorly capture semantic meaning of sentences. In this paper, we argue that the semantic information in the BERT embeddings is not fully exploited. We first reveal the theoretical connection between the masked language model pre-training objective and the semantic similarity task theoretically, and then analyze the BERT sentence embeddings empirically. We find that BERT always induces a non-smooth anisotropic semantic space of sentences, which harms its performance of semantic similarity. To address this issue, we propose to transform the anisotropic sentence embedding distribution to a smooth and isotropic Gaussian distribution through normalizing flows that are learned with an unsupervised objective. Experimental results show that our proposed BERT-flow method obtains significant performance gains over the state-of-the-art sentence embeddings on a variety of semantic textual similarity tasks. The code is available at https://github.com/bohanli/BERT-flow.		
versions		**versions**			