cases	doc_1	doc_2	decision	id
	<table><tr><td>authors</td><td>• Bohan Li • Hao Zhou • Junxian He • Mingxuan Wang • Yiming Yang • Lei Li</td></tr><tr><td>title</td><td>On the Sentence Embeddings from Pre-trained Language Models</td></tr><tr><td>publication_date</td><td>2020-11-02 00:00:00</td></tr><tr><td>source</td><td>SupportedSources.OPENALEX</td></tr><tr><td>journal</td><td>arXiv (Cornell University)</td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td>10.48550/arxiv.2011.05864</td></tr><tr><td>urls</td><td>• https://openalex.org/W3098400973 • https://doi.org/10.48550/arxiv.2011.05864 • http://arxiv.org/pdf/2011.05864</td></tr><tr><td>id</td><td>id7566594427820948189</td></tr><tr><td>abstract</td><td></td></tr><tr><td>versions</td><td></td></tr></table>	<table><tr><td>authors</td><td>• Lei LI • Yiming Yang • Mingxuan Wang • Junxian He • Hao Zhou • Bohan Li</td></tr><tr><td>title</td><td>On the Sentence Embeddings from Pre-trained Language Models</td></tr><tr><td>publication_date</td><td>2020-11-02 00:00:00</td></tr><tr><td>source</td><td>SupportedSources.PAPERS_WITH_CODE</td></tr><tr><td>journal</td><td></td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td></td></tr><tr><td>urls</td><td>• https://arxiv.org/pdf/2011.05864v1.pdf • https://github.com/InsaneLife/dssm • https://aclanthology.org/2020.emnlp-main.733.pdf</td></tr><tr><td>id</td><td>id6959606668887333883</td></tr><tr><td>abstract</td><td>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.</td></tr><tr><td>versions</td><td></td></tr></table>	DUPLICATES	256