	doc_1	doc_2	decision	id
cases	<table><tr><td>authors</td><td>• Yan Zhang • Ruidan He • Zuozhu Liu • Kwan Hui Lim • Lidong Bing</td></tr><tr><td>title</td><td>An Unsupervised Sentence Embedding Method byMutual Information Maximization</td></tr><tr><td>publication_date</td><td>2020-09-25 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>None</td></tr><tr><td>urls</td><td>• https://openalex.org/W3089244638</td></tr><tr><td>id</td><td>id-4688814547794997643</td></tr><tr><td>abstract</td><td></td></tr><tr><td>versions</td><td></td></tr></table>	<table><tr><td>authors</td><td>• Yan Zhang • Ruidan He • Zuozhu Liu • Kwan Hui Lim • Lidong Bing</td></tr><tr><td>title</td><td>An Unsupervised Sentence Embedding Method by Mutual Information Maximization</td></tr><tr><td>publication_date</td><td>2020-09-25 07:16:51+00:00</td></tr><tr><td>source</td><td>SupportedSources.ARXIV</td></tr><tr><td>journal</td><td>None</td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td></td></tr><tr><td>urls</td><td>• http://arxiv.org/pdf/2009.12061v2 • http://arxiv.org/abs/2009.12061v2 • http://arxiv.org/pdf/2009.12061v2</td></tr><tr><td>id</td><td>id3567519919786864619</td></tr><tr><td>abstract</td><td>BERT is inefficient for sentence-pair tasks such as clustering or semantic search as it needs to evaluate combinatorially many sentence pairs which is very time-consuming. Sentence BERT (SBERT) attempted to solve this challenge by learning semantically meaningful representations of single sentences, such that similarity comparison can be easily accessed. However, SBERT is trained on corpus with high-quality labeled sentence pairs, which limits its application to tasks where labeled data is extremely scarce. In this paper, we propose a lightweight extension on top of BERT and a novel self-supervised learning objective based on mutual information maximization strategies to derive meaningful sentence embeddings in an unsupervised manner. Unlike SBERT, our method is not restricted by the availability of labeled data, such that it can be applied on different domain-specific corpus. Experimental results show that the proposed method significantly outperforms other unsupervised sentence embedding baselines on common semantic textual similarity (STS) tasks and downstream supervised tasks. It also outperforms SBERT in a setting where in-domain labeled data is not available, and achieves performance competitive with supervised methods on various tasks.</td></tr><tr><td>versions</td><td></td></tr></table>	NOT DUPLICATES	418