cases	doc_1		doc_2		decision	id
	authors	• Tong Guo • Huilin Gao	**authors**	• Tong Guo • Huilin Gao	DUPLICATES	53
	title	Revisiting Semantic Representation and Tree Search for Similar Question Retrieval	**title**	Revisiting Semantic Representation and Tree Search for Similar Question Retrieval		
	publication_date	2019-08-22 11:44:12+00:00	**publication_date**	2019-09-06 00:00:00		
	source	SupportedSources.ARXIV	**source**	SupportedSources.INTERNET_ARCHIVE		
	journal	None	**journal**			
	volume		**volume**			
	doi		**doi**			
	urls	• http://arxiv.org/pdf/1908.08326v8 • http://arxiv.org/abs/1908.08326v8 • http://arxiv.org/pdf/1908.08326v8	**urls**	• https://web.archive.org/web/20200913110202/https://arxiv.org/pdf/1908.08326v8.pdf		
	id	id-8813666269953058087	**id**	id-3520330625225908533		
	abstract	This paper studies the performances of BERT combined with tree structure in short sentence ranking task. In retrieval-based question answering system, we retrieve the most similar question of the query question by ranking all the questions in datasets. If we want to rank all the sentences by neural rankers, we need to score all the sentence pairs. However it consumes large amount of time. So we design a specific tree for searching and combine deep model to solve this problem. We fine-tune BERT on the training data to get semantic vector or sentence embeddings on the test data. We use all the sentence embeddings of test data to build our tree based on k-means and do beam search at predicting time when given a sentence as query. We do the experiments on the semantic textual similarity dataset, Quora Question Pairs, and process the dataset for sentence ranking. Experimental results show that our methods outperform the strong baseline. Our tree accelerate the predicting speed by 500%-1000% without losing too much ranking accuracy.	**abstract**	This paper studies the performances of BERT combined with tree structure in short sentence ranking task. In retrieval-based question answering system, we retrieve the most similar question of the query question by ranking all the questions in datasets. If we want to rank all the sentences by neural rankers, we need to score all the sentence pairs. However it consumes large amount of time. So we design a specific tree for searching and combine deep model to solve this problem. We fine-tune BERT on the training data to get semantic vector or sentence embeddings on the test data. We use all the sentence embeddings of test data to build our tree based on k-means and do beam search at predicting time when given a sentence as query. We do the experiments on the semantic textual similarity dataset, Quora Question Pairs, and process the dataset for sentence ranking. Experimental results show that our methods outperform the strong baseline. Our tree accelerate the predicting speed by 500%-1000% without losing too much ranking accuracy.		
	versions		**versions**			