

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none">Luoqiu LiXiang ChenHongbin YeZhen BiShumin DengNingyu ZhangHuajun Chen	authors	<ul style="list-style-type: none">Bi, ZhenChen, HuajunChen, XiangDeng, ShuminLi, LuoqiuYe, HongbinZhang, Ningyu	DUPLICATES	231
	title	On Robustness and Bias Analysis of BERT-based Relation Extraction	title	On Robustness and Bias Analysis of BERT-based Relation Extraction		
	publication_date	2021-12-25 00:00:00	publication_date	2021-04-06 00:00:00		
	source	SupportedSources.INTERNET_ARCHIVE	source	SupportedSources.CORE		
	journal		journal			
	volume		volume			
	doi		doi	None		
	urls	<ul style="list-style-type: none">https://web.archive.org/web/20220103094036/https://arxiv.org/pdf/2009.06206v5.pdf	urls	<ul style="list-style-type: none">http://arxiv.org/abs/2009.06206		
	id	id8336599639660550199	id	id-165143594207494521		
	abstract	Fine-tuning pre-trained models have achieved impressive performance on standard natural language processing benchmarks. However, the resultant model generalizability remains poorly understood. We do not know, for example, how excellent performance can lead to the perfection of generalization models. In this study, we analyze a fine-tuned BERT model from different perspectives using relation extraction. We also characterize the differences in generalization techniques according to our proposed improvements. From empirical experimentation, we find that BERT suffers a bottleneck in terms of robustness by way of randomizations, adversarial and counterfactual tests, and biases (i.e., selection and semantic). These findings highlight opportunities for future improvements. Our open-sourced testbed DiagnoseRE is available in .	abstract	Fine-tuning pre-trained models have achieved impressive performance on standard natural language processing benchmarks. However, the resultant model generalizability remains poorly understood. We do not know, for example, how excellent performance can lead to the perfection of generalization models. In this study, we analyze a fine-tuned BERT model from different perspectives using relation extraction. We also characterize the differences in generalization techniques according to our proposed improvements. From empirical experimentation, we find that BERT suffers a bottleneck in terms of robustness by way of randomizations, adversarial and counterfactual tests, and biases (i.e., selection and semantic). These findings highlight opportunities for future improvements. Our open-sourced testbed DiagnoseRE is available in \url {https://github.com/zjunlp/DiagnoseRE} .Comment: work in progres		
	versions		versions			