cases	doc_1		doc_2		decision	id
	authors	• Vladimir Araujo • André Carvallo • Carolina Aspillaga • Denis Parra	authors	• Vladimir Araujo • Andrés Carvallo • C. Aspillaga • Denis Parra	DUPLICATES	261
	title	On Adversarial Examples for Biomedical NLP Tasks	title	On Adversarial Examples for Biomedical NLP Tasks		
	publication_date	2020-04-23 00:00:00	publication_date	2020-04-23 00:00:00		
	source	SupportedSources.OPENALEX	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal	arXiv (Cornell University)	journal	ArXiv		
	volume		volume	abs/2004.11157		
	doi	None	doi			
	urls	• https://openalex.org/W3017885892	urls	• https://www.semanticscholar.org/paper/6f509be31111ed828486198ad3ca41acb5c1c257		
	id	id-6493339733811671171	id	id4600450713927831273		
	abstract		abstract	The success of pre-trained word embeddings has motivated its use in tasks in the biomedical domain. The BERT language model has shown remarkable results on standard performance metrics in tasks such as Named Entity Recognition (NER) and Semantic Textual Similarity (STS), which has brought significant progress in the field of NLP. However, it is unclear whether these systems work seemingly well in critical domains, such as legal or medical. For that reason, in this work, we propose an adversarial evaluation scheme on two well-known datasets for medical NER and STS. We propose two types of attacks inspired by natural spelling errors and typos made by humans. We also propose another type of attack that uses synonyms of medical terms. Under these adversarial settings, the accuracy of the models drops significantly, and we quantify the extent of this performance loss. We also show that we can significantly improve the robustness of the models by training them with adversarial examples. We hope our work will motivate the use of adversarial examples to evaluate and develop models with increased robustness for medical tasks.		
	versions		versions			