

cases	doc_1		doc_2				decision	id
							DUPLICATES	183
	authors	<ul style="list-style-type: none">Seung Ki MoonSangwoo MoKimin LeeJae-Ho LeeJinwoo Shin	authors	<ul style="list-style-type: none">Jinwoo ShinJaeho LeeKimin LeeSangwoo MoSeung Jun Moon				
	title	MASKER: Masked Keyword Regularization for Reliable Text Classification	title	MASKER: Masked Keyword Regularization for Reliable Text Classification				
	publication_date	2021-05-18 00:00:00	publication_date	2020-12-17 00:00:00				
	source	SupportedSources.OPENALEX	source	SupportedSources.PAPERS_WITH_CODE				
	journal	arXiv (Cornell University)	journal					
	volume		volume					
	doi	10.48550/arxiv.2012.09392	doi					
	urls	<ul style="list-style-type: none">https://openalex.org/W3173275723https://doi.org/10.48550/arxiv.2012.09392http://arxiv.org/pdf/2012.09392	urls	<ul style="list-style-type: none">https://arxiv.org/pdf/2012.09392v1.pdfhttps://github.com/alinlab/MASKER				
	id	id8864293624899342553	id	id1670891010620103543				
	abstract		abstract	Pre-trained language models have achieved state-of-the-art accuracies on various text classification tasks, e.g., sentiment analysis, natural language inference, and semantic textual similarity. However, the reliability of the fine-tuned text classifiers is an often underlooked performance criterion. For instance, one may desire a model that can detect out-of-distribution (OOD) samples (drawn far from training distribution) or be robust against domain shifts. We claim that one central obstacle to the reliability is the over-reliance of the model on a limited number of keywords, instead of looking at the whole context. In particular, we find that (a) OOD samples often contain in-distribution keywords, while (b) cross-domain samples may not always contain keywords; over-relying on the keywords can be problematic for both cases. In light of this observation, we propose a simple yet effective fine-tuning method, coined masked keyword regularization (MASKER), that facilitates context-based prediction. MASKER regularizes the model to reconstruct the keywords from the rest of the words and make low-confidence predictions without enough context. When applied to various pre-trained language models (e.g., BERT, RoBERTa, and ALBERT), we demonstrate that MASKER improves OOD detection and cross-domain generalization without degrading classification accuracy. Code is available at https://github.com/alinlab/MASKER.				
	versions		versions					