cases	doc_1	doc_2	decision	id

doc_1

authors	• Seung Ki Moon • Sangwoo Mo • Kimin Lee • Jae-Ho Lee • Jinwoo Shin
title	MASKER: Masked Keyword Regularization for Reliable Text Classification
publication_date	2020-12-17 00:00:00
source	SupportedSources.OPENALEX
journal	Proceedings of the ... AAAI Conference on Artificial Intelligence
volume	35
doi	10.1609/aaai.v35i15.17601
urls	• https://openalex.org/W3110664450 • https://doi.org/10.1609/aaai.v35i15.17601 • https://ojs.aaai.org/index.php/AAAI/article/download/17601/17408
id	id-3003058014721985410
abstract	
versions	

doc_2

authors	• Seung Jun Moon • Sangwoo Mo • Kimin Lee • Jaeho Lee • Jinwoo Shin
title	MASKER: Masked Keyword Regularization for Reliable Text Classification
publication_date	2020-12-17 04:54:16+00:00
source	SupportedSources.ARXIV
journal	None
volume	
doi	
urls	• http://arxiv.org/pdf/2012.09392v1 • http://arxiv.org/abs/2012.09392v1 • http://arxiv.org/pdf/2012.09392v1
id	id6862954679488916852
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	

decision	id
DUPLICATES	269