Tech Review

Text classification methods typically require labeled documents as training data. Sourcing pre-labeled documents can be costly and difficult to obtain. The paper "Text Classification Using Label Names Only: A Language Model Self-Training Approach" proposes a weakly-supervised text classification model that only uses the label name of each class to train the model on unlabeled documents. The Label-Name-Only Text Classification model, or LOTClass for short, is built in three steps that achieve high accuracy against four different benchmark datasets using at most three words per class as a label name. LOTClass outperforms similar weakly-supervised models and approaches the performance of supervised models.

The start of constructing the LOTClass model begins with pre-trained neural models. Using a pre-trained neural model bootstrap the model with a knowledge base as well as generic features. While BERT was used in this instance, LOTClass can be adapted to use any other pre-trained neural language models such as ELMo, GPT, XLNet, or any BERT variant.

LOTClass is categorized as a weakly-supervised text classification model which aims to categorize text documents based on word-level descriptions. These models do not require pre-labeled documents and instead learn from general knowledge. By assigning pseudo labels this method is able to learn and detect category-indicative words.

Utilizing BERT as the backbone of the model, LOTClass is built using three core techniques. The first involves constructing a category vocabulary for each class that contains similar words with the label name. Next, the model collects category-indicative words to train itself to gain category information. The final step generalizes the model by self-training on the unlabeled data.

Category understanding using label name replacement uses a pre-trained BERT masked language model (MLM) to predict words that are similar to the label name. The context of the word is preserved based on each document. The top 50 predicted words are used to establish the category vocabulary. This method is effective at producing replacement words that have a similar context to the label name.

In order to find category-indicative words, LOTClass uses masked category prediction (MCP). This step uses the pre-trained language model to create contextualized word-level category supervision to train itself to predict the implied category of a word with the word masked. By masking out the category-indicative word during the training forces the model to

infer categories based on the context of the words.

The final step is to self-train the model on the unlabeled corpus. This refines the model for better generalization and allows the model to predict words without the mask. The paper also introduces the concept of soft-labeling during the self-training step. Soft labeling promotes high-confidence prediction and demotes low confidence ones. This step was shown to give better and more stable predictions compared to hard-labeling.

LOTClass is compared against other language models with four benchmark datasets: news topics from AG News, Wikipedia topics from DBPedia, movie review sentiment from IMDB, and product review sentiment from Amazon. These datasets range in a number of classes, training size, and test size. The results of the classification accuracy show that LOTClass outperforms other weakly supervised models including LOTClass without self-training. With self-training, LOTClass reaches a 90% accuracy against the benchmark datasets. This accuracy approaches semi-supervised and supervised models such as UDA, BERT, and char-CNN.

The LOTClass model introduced in this paper is capable of performing text classification using the label name of each class. By associating semantically related words to the label using label replacement, finding category-indicative words using masked predictions, and self-training, the model is able to outperform similar weakly-supervised models and approach more robust semi-supervised and supervised models. While LOTClass may struggle with label names that are generic and difficult to categorize, it is capable of classifying text documents where pre-labeled documents are not available. Building upon this method may prove useful for many applications and can be used where supervised methods are not feasible.

References

Meng, Yu, et al. “Text Classification Using Label Names Only: A Language Model Self-Training Approach.” Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, 2020, pp. 9006–17. DOI.org (Crossref), https://doi.org/10.18653/v1/2020.emnlp-main.724.