Prosub

Sujal Suri

sujal22514@iiitd.ac.in

April 14, 2025

Abstract

Open set semi-supervised learning is a realistic version of closed set semi supervised learning where unlabeled data can contain, Out of Bag classes. We in prosub leverage the unlabeled and potentially OOB data to jointly learn form samples with or without label

1 keywords

Prosub, Semi-Supervised learning, Open set learning

2 Datasets

2.1 Audio

Audio dataset we consider for the task is FSDKaggle2018, It contains 41 classes with around 10k samples. with variying number of samples for each class. 21 of which we consider OOB.

2.2 Text

For text dataset we consider DBOntology dataset, which contains 5,00,000 samples from 14 classes. 6 of which we consider as OOB

2.3 Image

For the image modality we consider Tiny image net dataset with 1,00,000 images with 64x64x3 dimensional images. We took 100 of the 200 classes as ID and other 100 as OOD, each class contains 500 samples.

3 Methods

3.1 Audio

For this modality we use a backbone of Resent18, with a trainable linear layer to get the feature dims.

3.2 Image

For this modality we use a backbone of Resent34, with a trainable linear layer to get the feature dims.

3.3 Text

For this modality we use a Distilled version of bert.

4 Hyperparameters

The hyperparameters used for all the methods are the same

- optimizer: SGD, with nestrov momentum
- criterion: Cross Entropy Loss
- batch ema : 0.2
- alpha id, beta id, alpha ood, beta ood :, 1, 1, 10, 10
- lr = 3e-3
- warm up: 5
- threshold: 0.9
- weight self supervised: 40

5 Training stratergy

Unfortunately the model did not learn much, due to some unidentified bug in the code. It gets stuck at the initial accuracy it has. Regardless all the model have been run for atleast 10 epochs, for the closed set classification we use fixmatch with a predefined threshold, there are 5 types of error for the model to learn from. Semisupervised, supervised, subspace, self supervised and regularization.