Combating Label Noise in Selective Classifier using Abstention

Alchemy Club

Introduction

- Selective Classification allows a classifier to abstain from predicting some instances, thus trading off accuracy against coverage.
- Risk Coverage trade-off
    - $$\mathbf{Pr}_{S_m}\{R(f,g) > r^*\} < \delta,$$
- Label noise makes the classification problem difficult.
- Out-of-Distribution data exists in real work.

Related Work

- Selective classifier for image classification with label noise
  - different level of random label noise
  - different level of label noise for a specific class
- Utilize Deep Abstaining Classifier for improvement
- Add Out-of-Distribution Data during testing

Methodology

- Dataset: CIFAR10
  - 60000 32×32 color images in 10 classes
- Model: VGG-16
- Deep Abstaining Classifier as a Data cleanser
- Label Noise
  - Mislabel 10%, 20%,40% data uniformly
  - Mislabel 10%, 20%, 40% data in one class (cat)
- Out-of-distribution Dataset
  - SVHN Dataset
  - FakeData Dataset
  - Add 10%, 20%, 40% OOD data during testing

Results

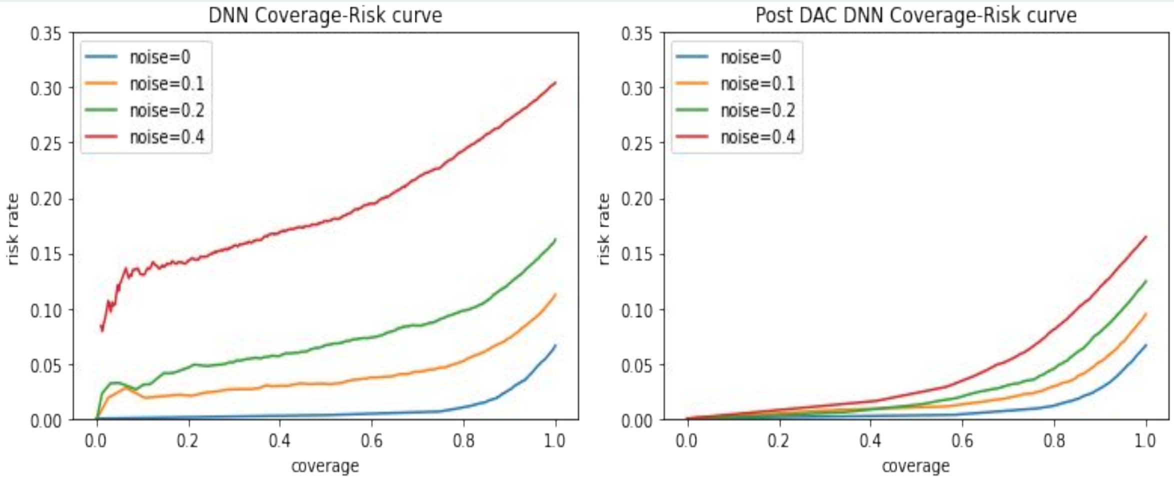

Figure 1. Risk-Coverage curve of **DNN** and **PostDAC DNN** model for data corrupted with **uniform label noise**

Noise Faction	***0***	***0.1***	***0.2***	***0.4***
DNN	**93.39 %**	88.78 %	83.79 %	69.59 %
Post DAC DNN	93.37 %	**90.54 %**	**87.54 %**	**83.53 %**

Table 1. Comparison of accuracy of **PostDAC DNN** against **DNN** model

Results

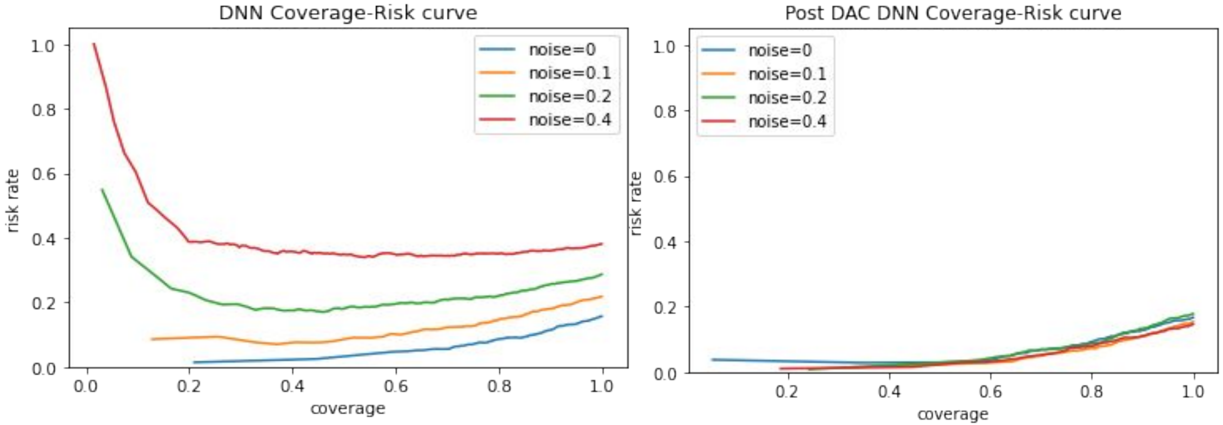

Figure 2. The "Cat" class Risk-Coverage curve of **DNN** and **PostDAC DNN** model for data corrupted with **single class label noise**

	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
0	93.7 / 94	97.7 / 97.7	91.7 / 92.5	**84.3 / 83.3**	94.5 / 94.5	88.4 / 88.3	96.3 / 95.2	95.2 / 94.9	97.5 / 96.9	94.6 / 93.7
0.1	93.5 / 93.2	97.5 / 96.7	92.2 / 91.2	78.2 / **84.6**	94.6 / 93.6	88.8 / 86.2	95.9 / 95.4	93.8 / 94.9	96.4 / 97.1	94.3 / 94
0.2	93.8 / 93.6	96.5 / 95.9	91.9 / 92	71.3 / **82.3**	94.1 / 94.1	91.1 / 87.5	95.3 / 96.1	94.8 / 95.4	96.2 / 97.5	94 / 94.8
0.4	92 / 94.4	96.7 / 92.4	91 / 91.2	61.9 / **85.4**	94.1 / 94.4	89.7 / 86.9	96.1 / 96.5	95 / 95.5	96.7 / 97.1	93.4 / 94.7

Table 2. Comparison of class-level accuracy of **PostDAC DNN** against **DNN** model

Results

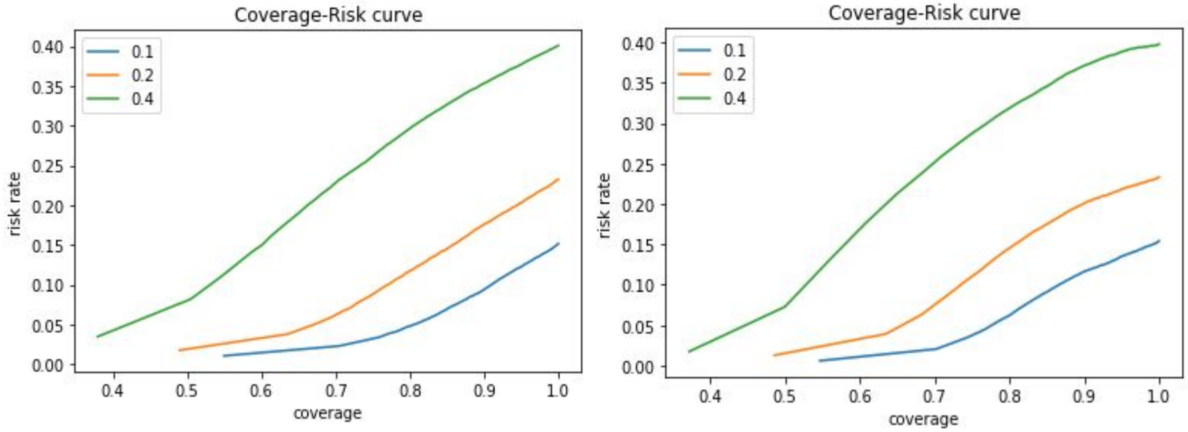

Figure 3. Risk-Coverage curve of **DNN selective classifier** on CIFAR 10 testset polluted with different fraction of **OOD data (FakeData/SVHN)**

Noise Faction	***0.1***	***0.2***	***0.4***
FakeData	84.59 %	76.67 %	60.23 %
SVHN	84.86 %	76.75 %	59.94 %

Table 3. Accuracy of **DNN selective classifier** on CIFAR 10 testset polluted with different fraction of **OOD data**

Takeaways

- **PostDAC DNN** selective classifier significantly **outperforms traditional DNN** selective classifier either when the training data is uniformly corrupted or has randomized single-class label.
- **PostDAC DNN** selective classifier can predict most samples with low risk (< 0.02) even when the training data is highly corrupted.
- There is a corresponding drop on accuracy for different level of OOD data and DNN selective classifier is not robust enough to address OOD data

Teamwork Partition

- Each team member has equal contribution to the project.
- Kai Wang and Ren Zhong are responsible for DAC implementation, corrupted dataset construction, and noisy data experiment.
- Chunchen Deng is responsible for the construction and analysis of the selective classifiers.
- Haobo Xu is responsible for building DNN model and performing OOD noise test.