Context-based Virtual Adversarial Training for Text Classification with Noisy Labels

Do-Myoung Lee, Yeachan Kim, Chang-gyun Seo
ShinhanCard, Deargen Inc., GC Company
LREC 2022

Introduction
Real World Datasets made by Human annotators or Users suffer from Label Noise. Furthermore, improving the quality of the datasets takes additional costs.

For that reason, we propose a simple but effective method Context-based Virtual Adversarial Training (ConVAT) which is robust to Noisy Label Datasets. Our approach is inspired by Virtual Adversarial Training (VAT).

The driving force behind ConVAT is to prevent the models from overfitting data points by enforcing consistency between the inputs and the perturbed inputs.

This strategy allows us to train a robust classifier against Label Noise without placing any burden to Computational Cost.

Dataset
Uniform label noise: In generating Noise from clean dataset, we flip the label from one category to another with the same probability across all the labels.

| 62.50% | 12.50% | 62.50% | 12.50% |
| 12.50% | 62.50% | 12.50% | 12.50% |
| 12.50% | 62.50% | 62.50% | 12.50% |
| 62.50% | 12.50% | 12.50% | 62.50% |

Random label noise: In generating Noise from clean dataset, labels are flipped from one labels to another based on a certain random distribution across all the labels.

| 58.33% | 18.75% | 10.41% | 12.50% |
| 12.50% | 58.92% | 14.28% | 14.28% |
| 21.05% | 21.05% | 50% | 7.89% |
| 15% | 17.50% | 17.50% | 50% |

C) Label Transition Matrix with 4 different types of Labels (Uniform)

D) Label Transition Matrix with 4 different types of Labels (Random)

Results
Denoising Effect of ConVAT
To analyze the denoising effects of ConVAT, we train our model and a baseline CNN model on the noisy dataset.

We visualize the context vectors of each sample from TREC datasets with 50% uniform noises through t-SNE (Van Der Maaten, 2014) and it is shown below in Figure E) and F).

In Figure E), the context vector are clustered with ambiguous decision boundaries. Moreover, data points labeled as ‘Y’ aren’t clustered. On the Contrary, context vectors in Figure F) are clearly clustered together with same labels. the scatter plot with overlayed clean labels shows our denoising effect of ConVAT on Label Noise.

E) Context Vectors from CNN

F) Context Vectors from ConVAT

Methodology
A) VAT

B) ConVAT

Training Step
1) Formulating Perturbation: Using context vector, we create a worst-case perturbation (adversarial example) for each data point to have a direction which maximize the classification loss. The adversarial example is obtained by computing the gradient of the context vector and is used to add an additional term to the given loss function.

2) Label Smoothing: Our additional term plays an important role in coping with Label Noise. Minimizing the distributional distance (KL-Divergence) between a normal sample and a perturbed sample enables the model to have an effect of Label Smoothing.

Conclusion
We have proposed context-based virtual adversarial training, coined ConVAT, a robust training method that prevents networks from overfitting to label noises. Furthermore, ConVAT is designed as a network-agnostic manner and has strong advantages in terms of the time complexity.

Comprehensive evaluation results have clearly shown that ConVAT is superior to previous works on Text Classification with Noisy Labels.