Master Thesis Defense Maryam Matin

Date: Monday, 23.03.2020
Time: 11:30 A.M.
Room: Online via Zoom.

Title: A Comprehensive Evaluation of Uncertainty Quantification Methods in Deep Learning

Abstract:
It's undefiable that deep learning models have revolutionized our modern data-pervasive societies. However, deep learning models come with few shortcomings. First, the weights of a neural network are fixed values and not distributions, therefore they can only make point predictions that are usually overconfident, and secondly, they are mostly uncalibrated. Therefore, it's important to have good uncertainty quantification methods which can determine if the model's predictions are reliable or not, especially in safety-critical tasks such as medical imaging for cancer detection or autonomous driving.

Most of the existing evaluations on the uncertainty quantification approaches ignore the effect of different models and different datasets for each model. Moreover, they usually confine their evaluations to less challenging datasets.

In this work, we conduct a comprehensive qualitative and quantitative analysis of three promising uncertainty quantification methods meaning Monte Carlo-Dropout, Deep Ensembles and the more recent and less studied Monte Carlo-DropConnect method. We evaluate the characteristics of the predictive uncertainties achieved by these methods in the context of image classification tasks in three common ConvNet models, meaning AlexNet, VGG16, and DenseNet on five datasets, namely CIFAR-10, CIFAR-100, Fashion-MNIST, SVHN, and FERPlus.

We investigate quantitative metrics such as classification accuracy, Negative Log-Likelihood (NLL) and Expected Calibration Error(ECE). As for the qualitative analysis, we look into the calibration curves. We also detect the five most uncertain images of each model-method combination per dataset, except CIFAR-100, and evaluated their barplots of predictive probability distributions and predictive entropy values as a function of method's number of samples/ensembles. Moreover, we track the changes in barplots of predictive distributions and entropy values of uncertain images for each model-dataset combination, except CIFAR-100, as the uncertainty method is varied. Finally, we evaluate the Out-of-Distribution(OOD) detection capabilities of the methods on three sets of experiments by plotting the ROC curves of In-Distribution vs Out-of-Distribution samples as the entropy threshold is varied.

The results demonstrate that Deep Ensembles clearly outperforms Monte Carlo methods in most cases with respect to all metrics and criteria, and Monte Carlo-DropConnect seems to be a better alternative to Monte Carlo-Dropout. We also observe that for datasets with high data uncertainty such as FERPlus, calibration of the models always worsens as the method's number of samples/ensembles increase, even as the classification accuracy improves.