Robust and Reliable Uncertainty Estimation for Deep Classifiers with Dirichlet Networks

Location: Laboratoire CEDRIC, Conservatoire National des Arts et Métiers (Paris 3^e)
Duration: 6 months, starting Mar. or Apr. 2021
Project Supervisors: Prof. Nicolas Thome and Charles Corbière

Overview

Reliable uncertainty estimation with deep neural networks is a challenging yet fundamental requirement for deploying classifiers in open-world conditions. Failing to detect possible errors and abnormal samples may carry serious repercussions in critical visual recognition applications such as autonomous driving [McAllister et al., 2017] and medical diagnosis [Heckerman et al., 1992]. Despite the tremendous predictive performance achieved the past few years, modern neural networks (NNs) have been shown to yield over-confident predictions [Guo et al., 2017], to be unreliable when exposed to out-of-distribution examples [Hein et al., 2019] and to have difficulties with detecting their own misclassifications [Hendrycks and Gimpel, 2017].

Recently, a new promising line of research, Dirichlet Networks [Malinin and Gales, 2018], address these issues by proposing to learn a Dirichlet distribution over class probabilities. Predictions of a neural networks are treated as subjectives opinions and training consists of learning a function that collect the evidence leading to this opinions. These approaches allow to better represent misclassifications and out-of-distribution samples on the simplex (Fig.1).

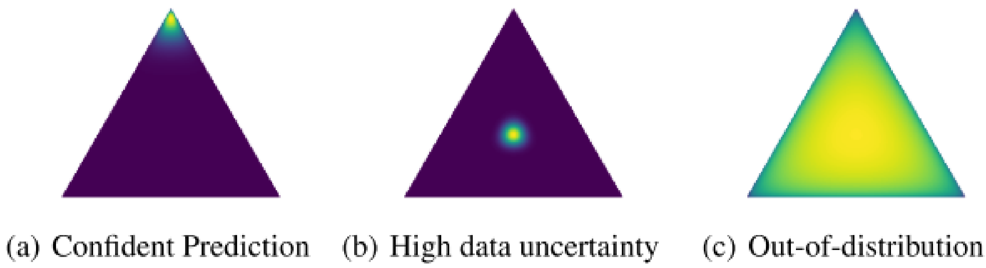

Figure 1: Desired behaviors of a Dirichlet distribution over class probabilities

However, to obtain such desired behaviors, current methods rely on using an explicit out-of-distribution dataset in training, which is an unrealistic assumption. Likewise, selection or creation of such auxiliary data draws some questions about the guarantees of characterizing such potentially infinite large distribution, especially for high dimensional data such as images

Objectives

The goal of this project is to **propose a generative approach for Dirichlet Networks to reliably estimate uncertainty for in- and out-distribution samples without requiring access to an existing outliers dataset at training time**. The approach will be evaluated in computer vision applications. Proposed objectives are:

- Review literature on Dirichlet Networks and out-of-distribution (OOD) detection;
- Re-implement baselines to generate OOD samples proposed in [Charpentier et al., 2020] and [Sensoy et al., 2020] ;
- Develop an approach based on generative modelling or OOD samples synthetization;
- Conduct experimental validation on standard computer vision datasets.

Qualifications

- Last year student enrolled in a Master's degree (Engineering School or University) in Applied Mathematics, Computer Vision, Computer Science or Artificial Intelligence.
- Experience in one or more areas of computer science and/or applied mathematics, such as computer vision, machine learning, deep learning, statistics or similar.
- Experience with Python and ideally with one or more purpose libraries (pytorch, TensorFlow, opencv, numpy, ...)
- Preference will be given to candidates who wish to pursue a PhD at the end of the internship (starting Sep./Oct. 2020)

Organisation

The position is a 6-month internship with a flexible starting date in March or April 2021. Successful candidate will be part of the VERTIGO team of the Center for Studies and Research in Computer Science and Communication (CEDRIC). The CEDRIC is an academic research lab bringing more than 80 researchers in 7 teams who pursue teaching and research activities in digital sciences at the National Conservatory of Arts and Crafts (Cnam). The internship is located in Paris, 3^e arrondissement, France. The project will be supervised by Prof. Nicolas Thome and Charles Corbière.

Application

Please send your application (CV + a short motivation letter) by mail to `nicolas.thome@cnam.fr` and `charles.corbiere@valeo.com`

References

[Charpentier et al., 2020] Charpentier, B., Zügner, D., and Günnemann, S. (2020). Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. In *Advances in Neural Information Processing Systems*.

[Guo et al., 2017] Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the International Conference on Machine Learning (ICML)*.

[Heckerman et al., 1992] Heckerman, D., Horvitz, E., and Nathwani, B. (1992). Toward normative expert systems: Part i the pathfinder project.

[Hein et al., 2019] Hein, M., Andriushchenko, M., and Bitterwolf, J. (2019). Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Hendrycks and Gimpel, 2017] Hendrycks, D. and Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks.

[Malinin and Gales, 2018] Malinin, A. and Gales, M. (2018). Predictive uncertainty estimation via prior networks. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[McAllister et al., 2017] McAllister, R., Gal, Y., Kendall, A., van der Wilk, M., Shah, A., Cipolla, R., and Weller, A. (2017). Concrete problems for autonomous vehicle safety: Advantages of Bayesian deep learning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

[Sensoy et al., 2020] Sensoy, M., Kaplan, L., Cerutti, F., and Saleki, M. (2020). Uncertainty-aware deep classifiers using generative models. In *Proceedings of the AAAI Conference on Artificial Intelligence*.