Course project proposal

Estimating Certainty of Deep Learning – implementation and performance analysis

Understanding what a model "knows" and what it "does not know" is a critical part of machine learning systems. Unfortunately, today's deep learning algorithms are usually unable to reliably estimate their own (un)certainty. [1]

- *In May 2016 we tragically experienced the first fatality from an assisted driving system. According to the manufacturer's blog, "Neither Autopilot nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied."*
- *In July 2015, an image classification system erroneously identified two African American humans as gorillas, raising concerns of racial discrimination. See the news report here.*

If both these algorithms were able to assign a high level of uncertainty to their erroneous predictions, then each system may have been able to alert an operator or choose a safer option.

Background

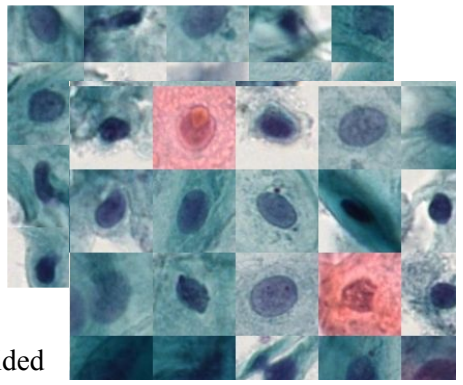
Deep neural networks (NNs) are powerful predictors that have recently achieved impressive performance on a wide spectrum of tasks. However most modern deep learning models do not provide certainty of their own prediction. Often users interpret low activations as "uncertain"/"unconfident" and high activations as "certain"/"confident" predictions, relying on the softmax output (values within (0,1)) as a probability or certainty measure of the model. However, such predicted probabilities, produced by the output of a softmax layer, can not reliably be used as true prediction probabilities (as a confidence for each label). In practice, they tend to be too high - ***neural networks are 'overconfident' in their predictions***.

Accurately quantifying the predictive certainty in NNs is a challenging and yet unsolved problem which has received great interest from the community over the past few years. (Un)certainty computation in deep learning is essential for the design of robust and reliable systems and vital for usage in critical applications such as medical diagnostics or self driving vehicles.

A multitude of approaches have been proposed. Bayesian NNs, which learn a distribution over weights, are among state of the art for estimating predictive uncertainty. These approaches are theoretically appealing and popular in accademia; there is e.g. a dedicated workshop at NeurIPS 2019 on Bayesian Deep Learning, http://bayesiandeeplearning.org/. However, these techniques generally require significant modifications to the training procedure and are computationally expensive. Different approximations are therefore proposed; recent promising approaches include a weight-perturbation approach implemented within the Adam optimizer [2] and Stochastic Weight Averaging [3]. Other approaches aim to produce well-calibrated uncertainty estimates using ensemble methods [4] or by approximating a Bayesian model using dropout [5,6]. More straightforward calibration approaches include temperature scaling [7] and regression based ones [8].

Project description

The task is to evaluate several existing approaches to reach well calibrated estimates of certainty for deep learning based image classification. Intuitively, a calibrated model means that for an input, whenever it predicts an output with a certainty 0.7, then that output should occur 70% of the time. The selection of methods to be evaluated is a part of the project task. A number of promising methods to be considered are proposed in [2-8].



Medical application

The most effective way of decreasing cancer mortality is early detection, which makes screening for cancer highly desired. Computer assisted analysis of cytology slides is a requirement for cost effective medical care. For safe diagnostics and functional inclusion in our healthcare system, reliable certainty estimates are essential. Implemented methods will be evaluated on the task of separating cell nuclei into the classes Cancer and Healthy in a provided annotated data set.

The project work should include

- Preparation of the project plan and distribution of the tasks within the team.
- A survey of the relevant literature and selection of 3-5 state of the art methods for the task.
- Implementation of the selected methods in a common environment (e.g., Matlab or Python).
- Quantitative evaluation of the selected methods on the provided data.
- Writing of the project report.

References

[1] https://alexgkendall.com/computer_vision/bayesian_deep_learning_for_safe_ai/

[2] Khan, M.E., Nielsen, D., Tangkaratt, V., Lin, W., Gal, Y. and Srivastava, A. "Fast and Scalable Bayesian Deep Learning by Weight-Perturbation in Adam." In International Conference on Machine Learning, pp. 2616-2625. 2018.

[3] Maddox, W., Garipov, T., Izmailov, P., Vetrov, D. and Wilson, A.G. "A simple baseline for bayesian uncertainty in deep learning." arXiv preprint arXiv:1902.02476 (2019).

[4] Lakshminarayanan, B., Pritzel, A. and Blundell, C. "Simple and scalable predictive uncertainty estimation using deep ensembles." In Advances in Neural Information Processing Systems, pp. 6402-6413. 2017.

[5] Gal Y, Ghahramani Z. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning." In International Conference on Machine Learning, pp. 1050-1059. 2016.

[6] Gal, Y. "Uncertainty in deep learning." PhD diss., PhD thesis, University of Cambridge, 2016.

[7] Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q. "On Calibration of Modern Neural Networks." In International Conference on Machine Learning, pp. 1321-1330. 2017.

[8] Kuleshov, V., Fenner, N. and Ermon, S. "Accurate Uncertainties for Deep Learning Using Calibrated Regression." In International Conference on Machine Learning, pp. 2801-2809. 2018.

Contact

Joakim Lindblad
Centre for Image Analysis, Department of Information Technology, Uppsala University
joakim@cb.uu.se

Nataša Sladoje
Centre for Image Analysis, Department of Information Technology, Uppsala University
natasa.sladoje@it.uu.se