Approximating Model Uncertainty in Deep Learning

Raghav Mehta

McGill University, Montreal QC, Canada
raghav@cim.mcgill.ca

Abstract. Deep Learning has gained a high amount of attention in applied Machine Learning fields like Computer Vision, Natural Language Processing, Information Retrieval etc., in the last decade. These Deep Learning based methods achieve state-of-the-art results in many problems of the above mentioned fields. However, these methods do not capture model uncertainty associated with the problem. Classical Bayesian Learning methods allow us to capture and represent the model uncertainty, but they are computationally costly. In this literature survey, we will talk about the latest development in the field of Deep Learning, popularly known as Bayesian Deep Learning, which associate the classical Bayesian learning methods with Deep Learning methods.

1 Introduction

Deep Learning has become omni-present in almost all of the applied machined learning fields due to recent advances. For example, in the field of Computer Vision, Deep Learning tools achieve state-of-the-art results in the varied range of problems like classification, segmentation, object detection, depth regression, etc. Though, these results are promising and gives better performance, they don't provide us the associated uncertainty with the output. This is due to the fact that Deep Learning gives point-wise deterministic prediction for a single input instead of the distribution of prediction. In many crucial applications like Medical Image Processing, self-driving cars, etc., it is of paramount importance to get associated uncertainty with the output. This is due to the fact that a wrong decision in the above mentioned field can be a difference between life-and-death. Take for example a case reported in May 2016 [10], where a fatality was reported as a self-driving car of Telsa was not able to successfully differentiate between sky and a white side of trailer. Similarly, in case of automatic medical diagnosis, we don't want a patient to be identified as a healthy patient when the automatic Diagnosis system is not confident in its prediction. Instead of that it is preferable if such a case is referred to a senior doctor for better diagnosis.

2 Literature Review

Bayesian Methods provide the inherent model uncertainty associated with the system, but they are computationally costly. While, with recent advances in

Graphical Processing Units (GPU), Deep Learning has become computationally cheap, but they don't provide model uncertainty. In a seminal paper by Gal and Ghahramani [1], they close the gap between Bayesian methods and Deep Learning based methods. They mathematically prove that the use of dropout, a widely popular regularization technique, after every weight layer a in deep network can be interpreted as a Bayesian approximation of deep Gaussian Process (GP) model. Specifically, they shows that the dropout minimizes the KL divergence between an approximate distribution and the posterior of a deep Gaussian Process. At test time to get an associated uncertainty with the output, same input is passed through the network with dropout, N times which results in N Monte-Carlo (MC) samples of the output. Now, uncertainty is calculated as a variance of this MC samples. Authors refer to this technique as MC-Dropout. They the MC-Dropout based uncertainty estimate with classical variational inference (VI) and probabilistic back-propagation (PBP) based methods and show that MC-Dropout outperforms these methods in approximating uncertainty estimation. Experiments were done for two tasks, classification (MNIST digit dataset) and regression.

Above mentioned paper lead to significant interest of applied machine learning field in Bayesian Neural Network, which in-turn resulted in many papers [4] [3] [11] [12] [13] which use [1] in various applications to approximate uncertainty in the model output. Bayesian SegNet [4] was one of the first paper to use MC-Dropout based uncertainty estimate in a classical computed vision task of semantic segmentation. In paper, authors extended the paper on SegNet, which was the state-of-the-art deep network at that time, to included MC-Dropout based uncertainty estimate. It was shown that as we increase the number of MC samples overall accuracy of the model prediction increases. And the performance of the system is higher than standard weight averaging method. They also experimented with various variant of Bayesian SegNet where they placed the dropout layers at difference depth and side (Encoded and Decoder). They concluded that when dropout is applied only for at the layers with the lowest resolution, we get maximum performance. The experiments were done on standard datasets for semantic segmentation, i.e. Pascal VOC12, CamVid, and SUN RGB-D. In the last experiment, it was reported that augmenting any semantic segmentation architectures like FCN, SegNet, or DenseNet can help in improving the performance of the networks by 2-3%.

In [3], authors used the approximated uncertainty in the paradigm of Active Learning (AL). Authors showed that when the acquisition function in AL, function to choose image to be marked by expert, is dependent on approximated uncertainty using MC-Dropout, it gives better performance than other state-of-the-art AL methods on MNIST dataset. They explored various information theoretical functions like Entropy, Mutual Information (MI), Variation Ration (VR), and Mean Standard Deviation (MSD) etc. as an uncertainty estimate. They concluded that MI, VR, and Entropy outperforms MSD, while VR achieves the best performance among all others. Usefulness of uncertainty in a clinically relevant task of Cancer Diagnosis form image data of skin segments was also reported.

In [11], authors explored the above mentioned uncertainty measurements in task of Multiple Sclerosis segmentation and detection. Through experimentation on private multi-cite Mutlimodality MRI dataset, it was shown how these uncertainty measurements can be useful in choosing better operating points. In [12] authors evaluate MC-dropout uncertainty measures in diagnosing diabetic retinopathy (DR) from fundus images and show that it captures uncertainty better than straightforward alternatives. The uncertainty informed decision referral was shown to improve diagnostic performance. In [13], authors proposed and validated that propagating MC-Dropout based uncertainty through the two stage system results in improvement of the overall performance of system in terms of both accuracy and confidence for the task of Pulmonary Nodule Detection.

Although, MC-Dropout based uncertainty estimate allows us to model uncertainty it doesn't say anything about the data uncertainty. In the following paper [2], authors divide the uncertainty in mainly two parts: (i) Epistemic Uncertainty which captures the uncertainty associated with the model parameters. This uncertainty arises as we our model parameters are not able to capture the true distribution of the model which generated our data. This uncertainty can become zero if we have infinite amount of data as it will allow our model parameter to learn the true distribution of data generation model. MC-Dropout based uncertainty estimate captures Epistemic uncertainty. This uncertainty is useful for capturing out-of-data examples. For example, if we train our model on MRI scans from Phillip and Siemens scanners but during test time we use MRI scans from GE scanners then Epistemic uncertainty for MRI scan from GE scanners will be higher than Phillips or Siemens scanners. (ii) Aleatoric Uncertainty, which captures the uncertainty associated with the input data. This is further further divided it into two parts: Homoscedastic and Heteroscedastic uncertainty. Homoscedastic uncertainty remains constant with the input but instead it is dependent on the task at hand, while Heteroscedastic uncertainty changes with change in input. Take for example the task of MRI segmentation. In this case if we use MRI scan of same patient with different noise than the Heteroscedastic uncertainty will change, but Homoscedastic uncertainty will remain constant. But if we change the task from segmentation to regression than for the same MRI scan of patient with same noise will results in changes of Homoscedastic uncertainty, but Heteroscedastic uncertainty will remain constant.

In [2], authors showed how the network can learn to predict Heteroscedastic uncertainty as an additional output for the task of classification and segmentation. They interpret this learned uncertainty as a learned loss attenuation which makes the loss more robust to noisy data. In the end, they combined both Epistemic and Heteroscedastic (Aleatoric) uncertainty. Through the experiment on CamVid and NYUv2 indoor scene dataset for the task of semantic segmentation, authors showed why it is necessary to combine both the uncertainty estimate. It was reported that the model performance increases when both these uncertainties are modeled in comparison to only one. Experimentation was also done on Make3D dataset and NYUv2 Depth dataset for the task of pixelwise depth regression. Through these experiments, they showed how Epistemic uncertainty decreases when model is trained using more data but it doesn't result in decrease of Aleatoric uncertainty.

As mentioned above Homoscedastic uncertainty can be thought as a task dependent uncertainty which changes with change in task. In [5], authors used this Homoscedastic uncertainty to weigh different losses in a multi-task setting. They show how a single parameter associated with each task can help in learn the relative weights of the tasks in a multi-task setting and how this weight is inversely proposnal to the Homoscedastic uncertainty of the task. They experimented on CityScape dataset for the task of Semantic Segmentation, Instance Segmentation, and Depth Estimation. Their results indicate that multi-task learning indeed helps to improve performance and when the weights of each task are learned, using Homoscedastic uncertainty, it gives the better performance in comparison to when weights are manually tuned.

Recently, in [7] authors extended the work of [2] by using Ensemble of Network and Adversarial Training. It was reported that when an ensemble of same network with different initialization and randomized batch input is used with Heteroscedastic uncertainty estimation technique of [2] is used, the network is able to represent the uncertainty estimate in a better way. They also showed that when examples of training dataset is augmented with their corresponding adversarial examples, uncertainty estimation improves. They experimented on sythesised as well as real dataset for regression task, and for classification task they used MNIST, SVHN, and ImageNet datasets, to show advantage of both Ensemble Learning and Adversarial data augmentation. By experimenting on NotMNIST dataset, they showed that uncertainty estimation for an unknown test class is better represented when using Ensemble and adversarial augmentation in compare to [2].

Inspired by the results of [7] where adversarial augmentation of training dataset improved the uncertainty estimation. The authors proposed and validated that just by using augmentation at the test time, network is able to produce better uncertainty estimation. They experimented on Kaggle dataset of Diabetic Ratinopathy (DR) for 5-class classification task. They used following augmentation technique for test-time augmentation: random crop and resize, random brightness, hue, saturation, and contrast adjustments, random horizontal and vertical flips, and random rotation.

One of the disadvantage of model presented in [2] is that it introduces and additional output of the network to learn Heteroscedastic Uncertainty. This increases the burden on the network. In addition to this, learned Uncertainty (variance) is also not directly dependent on the model output (mean). Although, this assumption of independence can be true in some scenarios, it may not always be true. In [6] paper, authors derive of an approximation of the uncertainty such there is no need for extra parameters to learn Heteroscedastic Uncertainty. They show that this approximation of Heteroscedastic Uncertainty represents associated uncertainty with output in a better way that the approximation derived in

[2], through experimentation on Ischemic Stroke Lesion Segmentation (ISLES) 2015 challenge dataset (an MRI scan dataset).

All the techniques to estimate uncertainty have been related to Dropout at Test time, or MC-Dropout. In a recent paper [8], Batch Normalization (BN) was casted as a Bayesian approximation method, similar to Dropout. They call this technique Monte Carlo Batch Normalization (MCBN). They take advantage of the fact that the parameters of BN varies with different batches. At the test time, we can get uncertainty estimate by passing the same image through network with different Batches, where in each batch one image is the test image, while the rest of them are sampled from training images. Through, experimentation on various regression dataset for regression task and on Pascal VOC dataset for segmentation task it was reported that MCBN gives better uncertainty estimation than MC-Dropout based technique on metrics like Predictive Log Likelihood (PLL) and Continuous Ranked Probability Score (CRPS). One of the disadvantage of MCBN technique is that it requires the availability of training data during test time, which may not be feasible in real scenario where many deep networks are deployed.

3 Conclusion

In this brief literature survey of recent advances in Bayesian Deep Learning methods, we saw how simple regularization technique like Dropout and Normalization technique like Batch Normalization can be viewed a Bayesian Approximation, which allows us to get uncertainty estimation of the model. Various extensions of MC-Dropout were shown to improve uncertainty estimation and model performance. In [2] and [5], MC-Dropout (Epistemic) uncertainty was augmented with either Heteroscedastic or Homoscedastic Uncertainty estimate, which gave improved performance and better uncertainty estimation. It is of interest to see if using all three (Epistemic, Heteroscedastic, and Homoscedastic) uncertainty can further improve the performance of the system.

References

1. Gal, Y. and Ghahramani, Z., 2016, June. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (ICLR) (pp. 1050-1059).
2. Kendall, A. and Gal, Y., 2017. What uncertainties do we need in bayesian deep learning for computer vision?. In Advances in neural information processing systems (NIPS) (pp. 5574-5584).
3. Gal, Y., Islam, R. and Ghahramani, Z., 2017. Deep bayesian active learning with image data. arXiv preprint arXiv:1703.02910.
4. Kendall, A., Badrinarayanan, V., and Cipolla, R., 2017. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. In Proceedings of the British Machine Vision Conference (BMVC).
5. Kendall, A., Gal, Y. and Cipolla, R., 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 7482-7491).

6. Kwon, Y., Won, J.H., Kim, B.J., and Paik, M.C., 2018. Uncertainty quantification using Bayesian neural networks in classification: Application to ischemic stroke lesion segmentation. In: Medical Imaging with Deep Learning (MIDL).
7. Lakshminarayanan, B., Pritzel, A. and Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in Neural Information Processing Systems (NIPS) (pp. 6402-6413).
8. Azizpour, H., Teye, M. and Smith, K., 2018. Bayesian uncertainty estimation for batch normalized deep networks. In International Conference on Machine Learning (ICML).
9. Ayhan, M.S. and Berens, P., 2018. Test-time Data Augmentation for Estimation of Heteroscedastic Aleatoric Uncertainty in Deep Neural Networks. In: Medical Imaging with Deep Learning (MIDL).
10. NHTSA. PE 16-007. 2017. Technical report, U.S. Department of Transportation, National Highway Traffic Safety Administration. Tesla Crash Preliminary Evaluation Report.
11. Nair, T., Precup, D., Arnold, D.L. and Arbel, T., 2018, September. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (pp. 655-663). Springer, Cham.
12. Leibig, C., Allken, V., Ayhan, M.S., Berens, P. and Wahl, S., 2017. Leveraging uncertainty information from deep neural networks for disease detection. Scientific reports, 7(1), p.17816.
13. Ozdemir, O., Woodward, B. and Berlin, A.A., 2017. Propagating Uncertainty in Multi-Stage Bayesian Convolutional Neural Networks with Application to Pulmonary Nodule Detection. Bayesian Deep Learning Workshop in association with Advances in Neural Information Processing Systems (NIPS).