Research on Bayesian Deep Learning Networks

Sinan Talha KOŞAR
Computer Engineering
Middle East Technical University
Ankara, TURKEY
e2099190@ceng.metu.edu.tr

***Abstract*—Deep learning networks are powerful category in machine learning, but most of deep learning models cannot predict uncertainty and take advantage of probability theory. With recent achievements it merge with Bayesian approach. In recent years they gained great interest, which provides model for uncertainty with deep learning framework. As a result, we can say that intersection between deep learning and Bayesian probability is Bayesian Deep Learning which offers uncertainty estimation on the architecture.**

***Index Terms*—Machine Learning, Neural Network, Deep Learning, Bayesian Neural Network, Prediction Uncertainty**

I. Introduction

Basically, neural networks are inspired from human brain that aims to make computer learn the observations while deep learning implements a sequence of algorithms that aims to capture the relationship between observations without having prior knowledge about the task. On the other hand, Bayesian inference in retaining uncertainty in addition to more conventional inference which is missing in neural network architectures. Human brain is not good at analyzing the systems and discover relations among them with limited or inconsistent data but good at knowing these weaknesses and accomplish them with many applications. For example, consider a medical diagnosis systems, in which some patients have some, different for each other but not having all, of the symptoms of the same disease which have many combinations and relations. So, the power of deep learning modelling complex tasks leveraging the hierarchical depiction and power of Bayesian inference to make meaningful predictions with these limited/conflicting data with uncertainty management form more powerful model.

II. Need For Bayesian Deep Learning Networks

A. Current Situation

Deep learning networks can be constructed successfully in many domains even as sensitive domains which early detection, good prediction accuracy and confident results are really important to protect humans or computers like medical and security. Although there are lots of observed data in these domains, most of the data are imbalanced(nearly 5% of all tests result in positive cancer results, ¡1% email is spam), and this leads to the model being over-fitted to the over-sampled class. [1] - [3]

B. Bayesian Deep Learning Network

Bayesian deep learning networks, on the other hand, are more resistant to over-fitting, and can easily be modelled with small datasets. Bayesian inferences preserves uncertainty differently from more traditional statistical inferences. In parallel with training, by integrating out the parameters, the mean value is computed across many models to prevent over-fitting.

Design of Bayesian Deep Learning Network's effort increase as the system gets more complex, and there is no certain way to build Bayesian Deep Learning Network, so it is domain specific (even system specific) which is some kind of advantage in some point of view.

On the other hand, the Bayesian inferences have a more perceptive approach. Bayesian inferences gives ability to view a probability as a measure of the expectation, or confidence, that an occurrence occurs to deep learning networks. [1] - [3]

Philosophy of Bayesian inference depends on being rarely certain about the results but being very confident as beliefs. For example, medical patient having symptoms x and y but many diseases might causing them, different doctors might have different beliefs about disease, the philosophy is treating beliefs as probability and Bayesian inference are confident about the beliefs. [4]

III. How Bayesian Deep Learning Network Works?

In traditional learning networks, there are fixed weights and biases to decide how input will be transformed to output. In Bayesian inference, all weights and biases probability distribution applied to them. Therefore everything has probability distribution including model parameters. Think programming languages, we have variables and some values, and returns most updated value every-time; in comparison to that thinking there are similar entries in the Bayesian world called random variables that returns different values every-time. Obtaining new value is called sampling from random variables. Return value is depend on probability distribution associated with random variable. For example, in Bayesian Deep Learning Network, on multiple runs to identify an image, each time there are new set of sampled parameters(weights and biases), so instead of getting single set of output, multiple sets get. The multiple set of output represents the probability distributions, upon that confidence and uncertainty can be found. If an input image is something that is never seen before, then the uncertainty of classes will be high and said to be "I don't know about this input". [5]

IV. Bayesian Deep Learning Application Experiments

A. Bayesian Deep Learning Methods for Robust Computer Vision

Deep learning networks have became dominant place for computer vision problems. Current implementations provide a range of safety-critical functions, such as street-scene. Since false predictions might cause catastrophic consequences, these applications involve accurate measurement of predictive uncertainty. In Bayesian Deep Learning, predicting uncertainty can be broken into 2 parts that should be captured. Epistemic uncertainty and aleatoric uncertainty. Epistemic uncertainty accounts for instability in the deep learning network model parameters, while aleatoric uncertainty absorbs endogenous and irreducible data noise. Input-dependent aleatoric uncertainty about target y emerges from noise and ambiguities inherent in input x. In street-scene semantic segmentation, pixels at object boundaries are generally inconclusive, object positions are less certain in 3D detection due to noise, and the sensor resolution is limited. In computer vision, aleatoric uncertainty can be estimated by letting the parameters(weights and biases) to deep learning network outputs, modeling the target's $p(y|x)$ distribution. To classify, softmax output layer realize the predictive categorical distribution. For regression, Laplace and Gaussian models have been employed. For classification tasks, a predictive categorical distribution is commonly realized by a softmax output layer, although recent work has also explored Dirichlet models [16, 43, 35]. For regression, Laplace and Gaussian models have been employed. However, predicting the conditional distribution $p(y|x)$ directly with a deep learning network does not capture epistemic uncertainty, since information about the uncertainty in the parameters of the model is disregarded. This often results in highly confident predictions which are wrong, especially for inputs x which are not well represented by the training distribution. [6]

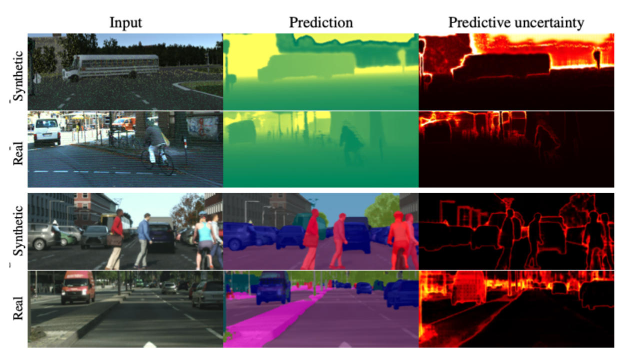

Fig. 1. DLN vs BDLN [6]

Predicting uncertainty are evaluated results are above. Input(on left), the prediction by deep learning network(at center) and estimated predictive uncertainty(on the right) i.e Bayesian Deep Learning Network. Black pixels lead to minimum uncertainty, white pixels to maximum uncertainty. Bayesian Deep Learning Networks produce rational estimates of uncertainty, even for simulated input data. In conclusion, comparison suggests that ensembling should be considered a go-to method for scalable Bayesian deep learning.

In this subsection, it is clearly shown the difference of deep learning network and Bayesian Deep Learning Network i.e implementation of capturing both aleatoric and epistemic uncertainty in deep learning models(making them Bayesian Deep Learning Networks for real-world problem.)

Conclusion

Bayesian Deep Learning Network proposed an evaluation for predictive uncertainty estimation it is theoretically attractive method for controlling over-fitting. Even with a small number of parameters. Bayesian deep learning methods' evaluation is challenging, seeking to determine the robustness and scalability of the methods. In last, Bayesian Deep Learning offers uncertainty estimation on the deep learning networks which is missing.

References

[1] Shridhar, K. (2020, February 17). Bayesian Neural Network Series Post 1: Need for Bayesian Networks. Retrieved June 26, 2020, from https://medium.com/neuralspace/bayesian-neural-network-series-post-1-need-for-bayesian-networks-e209e66b70b2 .

[2] Kendall, A. (2017, May 23). Deep Learning Is Not Good Enough, We Need Bayesian Deep Learning for Safe AI. Retrieved June 26, 2020, from https://alexgkendall.com/computer_vision/bayesian_deep_learning_for_safe_ai/

[3] Huynh, D. (2020, April 18). Bayesian deep learning with Fastai: How not to be uncertain about your uncertainty! Retrieved June 26, 2020, from https://towardsdatascience.com/bayesian-deep-learning-with-fastai-how-not-to-be-uncertain-about-your-uncertainty-6a99d1aa686e

[4] Davidson-Pilon, C. (2016). Introduction. In Bayesian methods for hackers: Probabilistic programming and bayesian inference. New York: Addison-Wesley.

[5] Chopra, P. (2019, March 06). Making Your Neural Network Say "I Don't Know" - Bayesian NNs using Pyro and PyTorch. Retrieved June 26, 2020, from https://towardsdatascience.com/making-your-neural-network-say-i-dont-know-bayesian-nns-using-pyro-and-pytorch-b1c24e6ab8cd

[6] Gustafsson, F. K., Danelljan, M., & Schön, T. B. (2019). Evaluating Scalable Bayesian Deep Learning Methods for Robust Computer Vision. Retrieved 2020, from http://bayesiandeeplearning.org/2019/papers/10.pdf