Bayesian Learning of Generative Adversarial Networks

Yuan Liu [*1] **Bowei Liu** [*1] **Mengdi Yang** [*1]

1. Overview

In recent years, supervised learning on images using convolutional neural networks (CNNs)(Krizhevsky et al., 2012) has been studied thoroughly and successfully, contrast to the trend in unsupervised learning. However real-world problems often involve an unlimited number of unlabeled images and videos that only unsupervised learning can deal with. Especially, unsupervised generative learning of images can function as a good reference and guidance for notoriously tough tasks such as image generation. To utilize Probabilistic Graphical Model theory in vanilla deep learning models, we use probabilistic backpropagation (PBP) (Hernández-Lobato & Adams, 2015) as an gradients computation method to introduce uncertainty in the weights learned throughout all our models. The goal is to achieve faster convergence rate and competitive accuracy compared to the original result from Hernández-Lobato's paper. Also, other Bayesian techniques will be implemented and compared with PBP.

To further verify the superior performance in terms of converge epochs and over-fitting robustness of probabilistic backpropagation when compared with traditional backpropagation methods use Stochastic Gradient Decent (LeCun et al., 2015), we intend to implement this technique to more complex models like Generative Adversial Networks (GAN) (Radford et al., 2015). Typical GAN models utilize batch normalization (BN) (Ioffe & Szegedy, 2015) and dropout (Gal & Ghahramani, 2016) as Baysian ways to introduce uncertainty to the model while PBP is never tried in GAN. We divide our work into two major parts, one is the theory part, which we focus on establish a Bayesian interpretation of typical deep learning methods like BN and dropout. Then prove these techniques are superior not only limited to simple network structre but also as sophiscated as dual networks, for example, GAN. In the experiment section, we first try to show that PBP is possible to achieve better performance in terms of convergence speed in typical neural networks, the extend this to more complicated network structures.

2. Literature review

We focus on four papers relevant to Bayesian Learning of deep neural networks to give us a sense of how Bayesian Learning is implemented. Two of them are about the probabilistic backpropagation (Blundell et al., 2015) (Hernández-Lobato & Adams, 2015), the other two illustrate dropout (Gal & Ghahramani, 2016) (Kingma et al., 2015).

2.1. Weight uncertainty in neural network

This paper introduces a new algorithm for learning neural networks with uncertainty on the weights called Bayes by Backprop. Traditional weights in neural networks are point estimates which tends to overfit the training data and cannot assess the uncertainty in the training data. All weights in this new approach are represented as distributions over possible weight values. The proposed algorithm optimizes a well-defined objective function to learn a distribution on the weights of a neural network.

On the MNIST dataset, this approach achieves similar result as dropout technique. Bayes by Backprop approach is able to learn the trade of between exploration and exploitation for the problem of contextual bandits. The paper also demonstrated on a simple non-linear regression problem that the uncertainty introduced in regions with little or no data helps the neural network to better predict unseen data.

2.2. Probabilistic neural network

In traditional neural network, weights are deterministic. On the contrary, in a probabilistic neural network model, each individual weight is Gaussian distributed. In Hernández-Lobato's paper, a claimed scalable method for learning Bayesian neural networks is proposed in order to tackle existing Bayesian techniques' lack of scalability to large dataset and network sizes. This method is called probabilistic backpropagation (PBP). In the experiments section of his paper, the performance (running time) of a single layer neural network between variantional inference (VI), standard stochastic gradient descent via backpropagation (BP) and PBP are compared. This is an unfair comparison since they included the running time of searching hyperparameters. Though PBP can automatically adjust its hyperparameters. In terms of validation accuracy, PBP is similar to BP. A comparison of these three methods in a multi-layer neural network is also illustrated in his paper. The result shows that PBP and PB has comparable performance, while both methods surpass VI in terms of prediction error. But nowadays VI is seldom used in most circumstances. Then only useful conclusion from the experiments is that PBP could be as good as BP as a backpropagation method. But it has no decisive advantage over BP. From his paper, there is no direct evidence that PBP is scalable in large scale dataset and neural network models. All experiments are done in toy experiments, which cannot support the paper's claim of the scalability of PBP.

2.3. Dropout as a Bayesian approximation

Gal's paper shows that dropout is indeed a Bayesian approximation. Then it further extends the result to show that a nereal model uncertainty can be obtained from dropout NN models. In the experiments section of Gal's paper, a predictive performance experiment is conducted. Droupout significantly outperforms all other models both in terms of RMSE as well as test log-likelihood on most of the datasets. In another experiment, which introduces model uncertainty in reinforcement learning, it obtains much more average reward in terms of the same batches executed. Overall, Gal's paper demonstrates the outstanding performance of dropout methods. As a matter of fact, dropout, batch normalization and momentum are the most prevalent techniques used in today's neural networks.

In Diederik's paper, they propose a local reparameterizaton technique, which translates uncertainty about global parameters into local noise that is independent across data points in the minibatch, in order to significantly reduce the variance of stochastic gradients for variational Bayesian inference (SGVB) of a posterior over model parameters, while retaining parallelizability. This local reparameterization method has lower computational complexity and shows the lowest variance among all variational dropout estimators for all conditions in the experiment. Additionally, they demonstrate a connection between dropout and SGVB model with local reparameterization, and suggest variational dropout, an extension of Gaussian dropout where optimal dropout rates are inferred from the data, rather than fixed in advance. This adaptive variational versions of regular dropout perform equal or better than their non-adaptive counterparts and standard dropout under all tested conditions in the experiment, especially for the smaller networks.

3. Dataset

To reduce the demand of GPU resources in the training process, we choose image-oriented models. Hence, our datasets are also image-based. We start with toy datasets to verify that our implementation works as expected. Then we use more sophisticated datasets like MNIST (LeCun et al., 2010) and CIFAR-10 (Krizhevsky et al., 2014). If possible, we show our work scalable by using selected subset of ImageNet (Deng et al., 2009) or other similar datasets.

4. Plan and expected result

We plan to set up a Github repository for version control and code coordination. Weekly meeting for in person coding and discussion will also be scheduled to keep every member informed and to synchronize progress. By midway report we aim to finish implementing probabilistic backpropagation for a baseline CNN so that we can measure its performance. This midway goal is to ensure we could test this approach further on more complex models like GANs. However it is not guaranteed that this midway goal is met due to unexpected complexities encountered. If that happens we will focus our effort solely on improving the baseline CNN instead of moving on to GANs. To reduce the amount of computation force required to train a network, we use image-based models, especially, DCGAN(Radford et al., 2015) or LAP-GAN(Denton et al., 2015).

As for division of work, Yuan Liu will lead design the infrastructure and implementation will be equally divided among all members.

References

Blundell, Charles, Cornebise, Julien, Kavukcuoglu, Koray, and Wierstra, Daan. Weight uncertainty in neural networks. *arXiv preprint arXiv:1505.05424*, 2015.

Deng, Jia, Dong, Wei, Socher, Richard, Li, Li-Jia, Li, Kai, and Fei-Fei, Li. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255. IEEE, 2009.

Denton, Emily L, Chintala, Soumith, Fergus, Rob, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pp. 1486–1494, 2015.

Gal, Yarin and Ghahramani, Zoubin. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059, 2016.

Hernández-Lobato, José Miguel and Adams, Ryan. Probabilistic backpropagation for scalable learning of bayesian neural networks. In *International Conference on Machine Learning*, pp. 1861–1869, 2015.

Ioffe, Sergey and Szegedy, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456, 2015.

Kingma, Diederik P, Salimans, Tim, and Welling, Max. Variational dropout and the local reparameterization trick. In *Advances in Neural Information Processing Systems*, pp. 2575–2583, 2015.

Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Krizhevsky, Alex, Nair, Vinod, and Hinton, Geoffrey. The cifar-10 dataset. *online: http://www. cs. toronto. edu/kriz/cifar. html*, 2014.

LeCun, Yann, Cortes, Corinna, and Burges, CJ. Mnist handwritten digit database. *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, 2, 2010.

LeCun, Yann, Bengio, Yoshua, and Hinton, Geoffrey. Deep learning. *nature*, 521(7553):436, 2015.

Radford, Alec, Metz, Luke, and Chintala, Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.