Reliable Uncertainty Quantification of Deep Learning Models for a Free Electron Laser Scientific Facility

Lipi Gupta*
NERSC, Lawrence Berkeley National Laboratory
Berkeley, CA 94720
lipigupta@uchicago.edu

Aashwin Anand Mishra
SLAC National Accelerator Laboratory
Menlo Park, CA 94025
aashwin@slac.standord.edu

Auralee Edelen
SLAC National Accelerator Laboratory
Menlo Park, CA 94025
edelen@slac.stanford.edu

Abstract

Particle accelerators are essential instruments for scientific experiments. They provide different experiments with particle beams of different parameters (e.g. beam energies or durations). This is accomplished by changing a wide variety of controllable settings, in a process called tuning. This is a challenging task, as many particle accelerators are complex machines with thousands of components, each of which contribute sources of uncertainty. Fast, accurate models of these systems could aid rapid customization of beams, but in order to accomplish this reliably, quantified uncertainties are essential. We address the problem of obtaining reliable uncertainties from learned models of a noisy, high-dimensional, nonlinear accelerator system: the X-ray free electron laser at the Linac Coherent Light Source, which is a scientific user facility. We examine the efficacy of Bayesian Neural Networks (BNNs) to reliably quantify predictive uncertainty and compare these with Quantile Regression Neural Networks (QRNNs). The QRNN models provide mean absolute error on predictions that are consistent with the noise of the measured data. We find the BNN is sensitive to outliers and is substantially more computationally expensive, but it still captures the general trend of the target data.

1 Introduction and Motivation

At Free Electron Laser (FEL) facilities, such as the Linac Coherent Light Source (LCLS), electron beams are accelerated and converted into powerful X-rays to help advance research in different disciplines. The LCLS facilitates thousands of experiments each year that aid our understanding of phenomena such as photosynthesis [7] and molecular interactions for drug discovery [3]. Researchers have access to experimental stations for a limited window of time, and the required beam for each experiment needs to be set up on demand. This is challenging, as accelerators have high-dimensional, nonlinear parameter spaces. Beam time is in very high demand, making rapid and accurate methods for tuning critical for increasing scientific throughput. Machine learning based models of accelerator systems trained on historical measured data can aid the optimization process, but numerous sources of uncertainty (due to noisy signals, drift over time, intermittent anomalous behavior, etc) are present in accelerator systems and can lead to unrepresentative predictive models.

*

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Vancouver, Canada.

At the LCLS, having a model to predict the X-ray pulse energy (which is a critical beam parameter for FEL scientific users) based on accelerator settings would potentially be very useful for tuning. The RMS fluctuation in the measured data can be up to 0.3 mJ at LCLS and could be due to inherent noise in the measurement as well as upstream noise sources in the accelerator. Using historical measured data spanning several years, we investigate quantile regression neural networks (QRNN), and variational inference based Bayesian neural networks (BNNs) as approaches for producing uncertainty-aware predictions of the X-ray pulse energy. The dataset includes over 250,000 samples consisting of 76 tunable machine parameters and the resulting x-ray pulse energy, measured during periods of machine tuning. Though both QRNNs and BNNs are powerful methods, we found that a set of QRNNs can provide superior coverage on a test set of unseen data, despite learning on high-dimensional, noisy data. To assess model robustness, in this study, some of the models shown in section 2.1 and section 2.2 are trained without removing outliers (e.g. due to signal dropout from the detector). Models trained on data after removing erroneous or low-energy values are also shown, to provide a comparison of performance.

2 Results

2.1 Quantile Regression Neural Networks

Deterministic neural networks are an accurate and flexible approach to model a wide range of non-linear regression tasks. The training of such neural networks attempts to estimate the conditional mean of the target, via the minimization of a mean squared loss. However, in complex scientific applications, the data may have a significant noise component, and this noise process can be heteroscedastic. This is true for the dataset considered in this study. Additionally, the noise may correspond to an underlying distribution that is not approximated well by a Gaussian. For instance, the target data may be severely skewed, or strictly non-negative. One approach is to estimate the point predictions of different quantiles, using sets of quantile neural networks. For a random variable X with a cumulative distribution function, $F_X(x) = P(X \leq x)$, the t-quantile, q, is given by $\arg\min_q F_X(q) = t$. Quantile regression approaches invoke no assumptions about the parametric form of the final distribution.

We use QRNNs using the tilted loss on the measured FEL data set. The model consists of an input layer for all 76 scalar inputs (consisting of, for example, magnet settings and accelerating cavity settings), 8 fully connected hidden layers which decrease in the number of neurons from 80 to 10, by a factor of 10 neurons for each subsequent layer. The activation function for each layer was the hyperbolic tangent function. The final output layer of 1 scalar output provides the prediction of the photon energy. The median prediction (50% quantile), 97.5% quantile prediction, and 2.5% quantile prediction, were each fit using independent models. To create the median prediction, and the corresponding confidence intervals, the same training and validation data were used to train each independent model. A custom tilted loss function was written, and each model was optimized using Adam [6] for its given quantile. All QRNN models were trained with a batch size of 4096 samples for a maximum of 5000 training epochs. Early stopping was implemented, such that the loss on the validation set was monitored and training terminated when the loss showed no improvement for 500 epochs. The QRNN model was trained on 80% of the full dataset, with the remaining 20% split equally into a validation and test set. No further hyperparameter optimization on the model or training parameters described above was conducted. The results on the test set are shown in Fig. 1, but the initial architecture for the QRNN was based on a deterministic neural network that was previously tuned for good prediction of the mean X-ray pulse energy. The model performs very well, with a prediction coverage of 92.84% and a mean absolute error of 0.13 mJ. Fig. 2 shows QRNN predictions in time series form as opposed to being sorted by magnitude.

2.2 Bayesian Neural Networks

In BNNs the weights and biases are assumed to be random variables with corresponding probability distributions, leading to an output prediction that has an associated probability distribution as well. Given training data, Bayesian inference over these BNNs estimates the posterior probability distributions for these weights and biases. While querying from such a trained BNN, the mean of predictions is given via expectations over the posterior distributions: $P(y^*|x^*) = E_{P(W|D)}\left(P(y^*|x^*, W)\right)$.

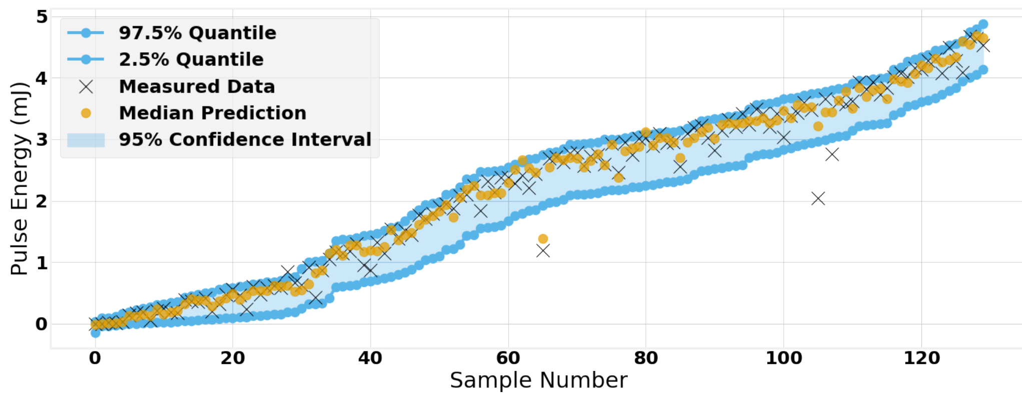

Figure 1: The QRNN model results for the test set, with samples sorted by mean prediction. The prediction coverage for these models is 92.84%. The test set mean absolute error (MAE) is 0.13 mJ.

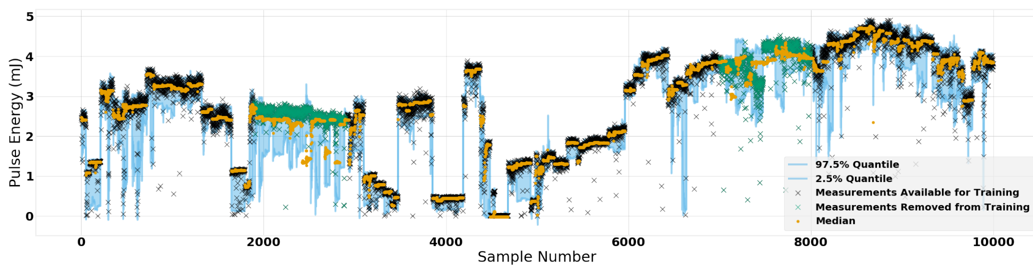

Figure 2: QRNN results shown in time-series form (increasing sample numbers are later in time).

Following best practice, we use a standard normal as the prior on the weights of the network. The joint prior over the entire set of weights and biases of the network is given by the product of the individual independent normal distributions over individual parameters, $P(W) = \prod_i \mathcal{N}(W_i|0,1)$.

For inference, approximate variational inference with the Bayes By Backprop algorithm [2] is used. The architecture is similar to the QRNN, with 7 fully connected hidden layers which decrease in the number of neurons from 80, 60, 50, 40, 30, 20, 10. To optimize the variational parameters, we utilize the Adam algorithm [6]. The model was trained using measured data, and samples below the detector sensitivity threshold (0.2mJ) were removed.

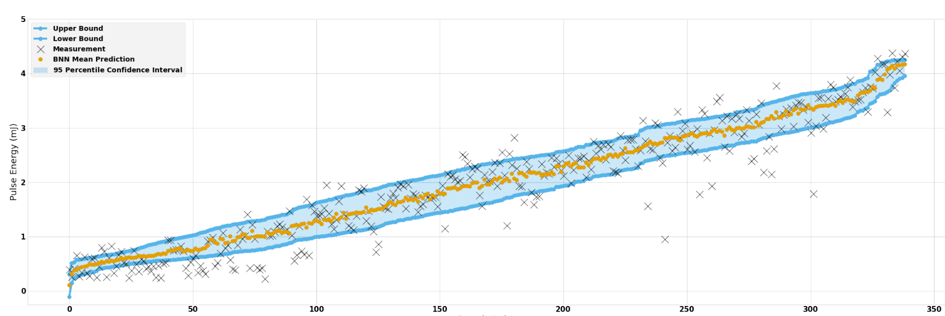

Figure 3: BNN results on the test set, with coverage at 67%. The MAE is 0.25 mJ. Samples are sorted by mean prediction.

The confidence interval estimated from the BNN predictions at each point do not provide 95% coverage, and are undercalibrated. While Fig. 3 does display that the BNN is beginning to capture trends in the data, compared to the QRNN result shown in Fig. 1, the BNN may require hyperparameter optimization to better predict the median measurement. The results reflect an inherent schism in BNN inference at present. While sampling based approaches for inference like Hamiltonian Monte Carlo provide reliable solutions, they do not scale to high dimensions. Approximate variational inference approaches have better scaling. However, they are extremely sensitive to hyperparameter choice and more importantly, may rely on substantial approximations to achieve scalability, resulting in limited or even unreliable uncertainty estimates [4, 1, 5].

3 Conclusions and Future Outlook

To integrate deep learning models into particle accelerator control and operation, the models must provide predictions with reliable uncertainty estimates consistent with the measured data. We examined the ability of quantile regression and Bayesian neural networks to provide predictions and uncertainty estimates for the X-ray pulse energy of the LCLS, given historical data spanning several years, a wide variety of operating modes, and 76 measured input variables. We find that QRNN models produce very reasonable median prediction and uncertainty estimates. The QRNN model shown in Fig. 1 achieves 92.84% coverage, while the optimized BNN provides only 67% coverage indicating undercalibration. While BNNs have shown excellent results on benchmark datasets, we find that in such complex problems approximate variation inference approaches may have limitations.

Because a goal for implementing model-based control of accelerators is to continuously retrain models as new data is collected, a model which is not sensitive to outliers would be a better candidate. Continuous retraining would address the irreducible uncertainty due to machine drift. An initial study examining this (by training on a sliding window of data and making predictions on future sections of the data) looks promising and will be continued in future work. Future work also includes applying feature selection on the inputs in this dataset, to determine if the dimensionality can be reduced. This will be coupled with hyperparameter optimization on the BNN model and training procedure to see if more reliable BNN models for this task can be obtained. The eventual goal is to bring a refined version of either the QRNN or BNN model into operational use at the LCLS.

4 Acknowledgments

We thank Joe Duris and Dylan Kennedy for putting together the routines to scrape the raw data from the Ocelot optimizer archives, which we then further processed. We also thank Joe Duris for his guidance on data cleaning for different FEL modes.

References

[1] J. Antorán, J. U. Allingham, and J. M. Hernández-Lobato. Depth uncertainty in neural networks. *arXiv preprint arXiv:2006.08437*, 2020.

[2] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. Weight uncertainty in neural networks. In *International Conference on Machine Learning*, pages 1613–1622, 2015.

[3] J.-P. Colletier, M. R. Sawaya, M. Gingery, J. A. Rodriguez, D. Cascio, A. S. Brewster, T. Michels-Clark, R. H. Hice, N. Coquelle, S. Boutet, G. J. Williams, M. Messerschmidt, D. P. DePonte, R. G. Sierra, H. Laksmono, J. E. Koglin, M. S. Hunter, H.-W. Park, M. Uervirojnangkoorn, D. K. Bideshi, A. T. Brunger, B. A. Federici, N. K. Sauter, and D. S. Eisenberg. De novo phasing with x-ray laser reveals mosquito larvicide binab structure, 2016. URL `https://doi.org/10.1038/nature19825`.

[4] A. Y. Foong, D. R. Burt, Y. Li, and R. E. Turner. Pathologies of factorised gaussian and mc dropout posteriors in bayesian neural networks. *CoRR, abs/1909.00719*, 2019.

[5] A. Y. Foong, Y. Li, J. M. Hernández-Lobato, and R. E. Turner. 'in-between'uncertainty in bayesian neural networks. *arXiv preprint arXiv:1906.11537*, 2019.

[6] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA,*

USA, May 7-9, 2015, Conference Track Proceedings, 2015. URL http://arxiv.org/abs/1412.6980.

[7] I. D. Young, M. Ibrahim, R. Chatterjee, S. Gul, F. D. Fuller, S. Koroidov, A. S. Brewster, R. Tran, R. Alonso-Mori, T. Kroll, T. Michels-Clark, H. Laksmono, R. G. Sierra, C. A. Stan, R. Hussein, M. Zhang, L. Douthit, M. Kubin, C. de Lichtenberg, L. Vo Pham, H. Nilsson, M. H. Cheah, D. Shevela, C. Saracini, M. A. Bean, I. Seuffert, D. Sokaras, T.-C. Weng, E. Pastor, C. Weninger, T. Fransson, L. Lassalle, P. Bräuer, P. Aller, P. T. Docker, B. Andi, A. M. Orville, J. M. Glownia, S. Nelson, M. Sikorski, D. Zhu, M. S. Hunter, T. J. Lane, A. Aquila, J. E. Koglin, J. Robinson, M. Liang, S. Boutet, A. Y. Lyubimov, M. Uervirojnangkoorn, N. W. Moriarty, D. Liebschner, P. V. Afonine, D. G. Waterman, G. Evans, P. Wernet, H. Dobbek, W. I. Weis, A. T. Brunger, P. H. Zwart, P. D. Adams, A. Zouni, J. Messinger, U. Bergmann, N. K. Sauter, J. Kern, V. K. Yachandra, and J. Yano. Structure of photosystem ii and substrate binding at room temperature, 2016. URL https://doi.org/10.1038/nature20161.