ISBN: 978-3-95450-241-7 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2022-TUP39

NEURAL NETWORK INVERSE MODELS FOR IMPLICIT OPTICS TUNING IN THE AGS TO RHIC TRANSFER LINE

J. P. Edelen, N. M. Cook, J. Einstein-Curtis, RadiaSoft LLC, Boulder, CO, USA
K. A. Brown, V. Schoefer, BNL, Upton, NY, USA

Abstract

One of the fundamental challenges of using machine-learning-based inverse models for optics tuning in accelerators, particularly transfer lines, is the degenerate nature of the magnet settings and beam envelope functions. Moreover, it is challenging, if not impossible, to train a neural network to compute correct quadrupole settings from a given set of measurements due to the limited number of diagnostics available in operational beamlines. However, models that relate BPM readings to corrector settings are more forgiving, and have seen significant success as a benchmark for machine learning inverse models. We recently demonstrated that when comparing predicted corrector settings to actual corrector settings from a BPM inverse model, the model error can be related to errors in quadrupole settings. In this paper, we expand on that effort by incorporating inverse model errors as an optimization tool to correct for optics errors in a beamline. We present a toy model using a FODO lattice and then demonstrate the use of this technique for optics corrections in the AGS to RHIC transfer line at BNL.

INTRODUCTION

Machine learning (ML) has seen a significant growth in its adoption for widespread applications. In particle accelerators, ML has been identified as having the potential for significant impact on modeling, operation, and controls [1, 2]. These techniques are attractive due to their ability to model nonlinear behavior, interpolate on complicated surfaces, and adapt to system changes over time. This has led to a number of dedicated efforts to apply ML, and early efforts have shown promise.

For example, neural networks (NNs) have been used as surrogates for traditional accelerator diagnostics to generate non-interceptive predictions of beam parameters [3, 4] or for a range of machine tuning problems utilizing inverse models [5]. When used in conjunction with optimization algorithms, neural networks have demonstrated improved switching times between operational configurations [6]. Neural network surrogate models have also been demonstrated to significantly speed up multi-objective optimization of accelerators [7]. Additionally, ML has been of interest for anomaly detection for root cause analysis [8] and for outlier detection, using large data-sets of known good operational states [9], using autoencoders.

In this work we seek to apply ML methods — for both tuning and anomaly detection — on the AGS to RHIC transfer line at Brookhaven National Laboratory. Specifically, we employ the use of inverse models for these applications. The application of inverse models for anomaly detection is a burgeoning area of research in many other fields that has not seen much attention in particle accelerators. Here we present our work towards implementing inverse models to detect errors in quadrupoles using only beam position monitors and corrector data. We will demonstrate the utility of this approach using a toy model, and then show how it scales to a larger system such as the AGS to RHIC transfer line. We will then show results of training inverse models using data from the machine and discuss future work for this effort.

FODO BENCHMARK

Before applying our technique to the ATR line, we first demonstrate the efficacy of our technique using a toy problem. Here we consider a linear system comprised of two quadrupoles, four beam position monitors and two correctors that operate in both the horizontal and vertical plane. The training data consisted of 5000 examples simulated by randomly changing corrector strength and the initial beam position. The model was trained to predict the corrector setting for a given set of BPM readings. The network architecture was optimized as a function of the number of layers as well as the nodes per layer to improve training loss without over-fitting. Figure 1 shows the model prediction compared to the ground truth for each of the correctors (kickers). The relationship is almost perfectly linear in all cases, indicating the model is well trained.

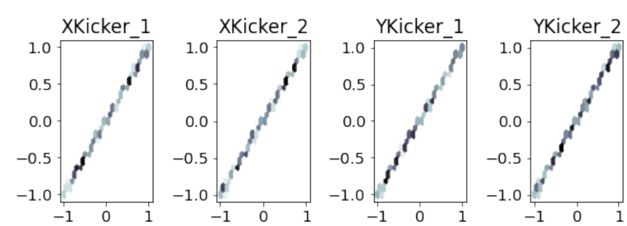

Figure 1: Real corrector setting on the x-axis with inverse model output result on the y-axis. The plots show a well-trained inverse model.

To evaluate the neural network against standard linear model benchmarks, we trained a linear model using the same data as the neural network. We also introduced different sextupole strengths to the lattice in order to understand the impact of a simple nonlinearity on the system. The linear model was compared to the neural network model for each case. Figure 2 shows the RMS prediction error on the test data for each of the correctors and the aggregate error which is computed as the sum of the squares of the individual corrector errors.

ISBN: 978-3-95450-241-7 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2022-TUP39

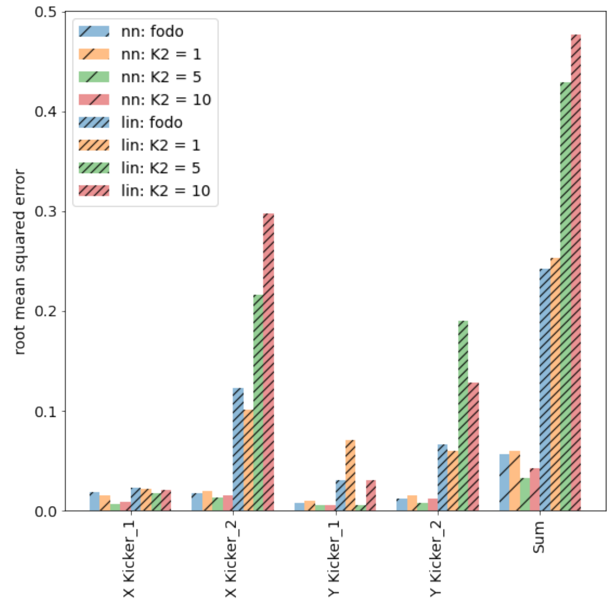

Figure 2: RMS prediction error for kickers in the FODO and nonlinear system based on model type and nonlinearity strength.

We see that the neural network does a better job at predicting the corrector settings even for the linear model case. There is also a weak improvement in the model prediction for the lattice with a sextupole component. This gives a good indication that the neural network will be a better solution in general even for the linear case.

Next we tested the model using data with systematically introduced errors in individual quadrupoles. We ran the MAD-X simulations with random corrector strengths and initial beam positions for a fixed quadrupole strength error. We then compared the predicted corrector settings to the actual corrector settings as we varied the quadrupole error. Figure 3 shows the model error as a function of both the quadrupole strengths being varied. Here we see a well-defined minima around the nominal lattice configuration, which is a quadrupole error of unity.

ATR STUDIES

Next we applied this concept to the UW line in the ATR. The problem is a bit more complex as there are now not just quadrupoles, correctors, and BPMs, but also combined function bends and vertical bends. Moreover there are comparatively few BPMs and correctors — only 26 and 14 respectively — making the model more subject to degeneracies. This combined with the 19 quadrupoles significantly increases the complexity of the problem space.

We trained the inverse model using 5000 samples, randomly varying the corrector strengths and beam initial positions. During our initial training of the inverse model four correctors (utv4, uth6, utv7, and wth1) were not well fit. In future work we will address this issue, however, for studies presented here we simply removed those four correctors

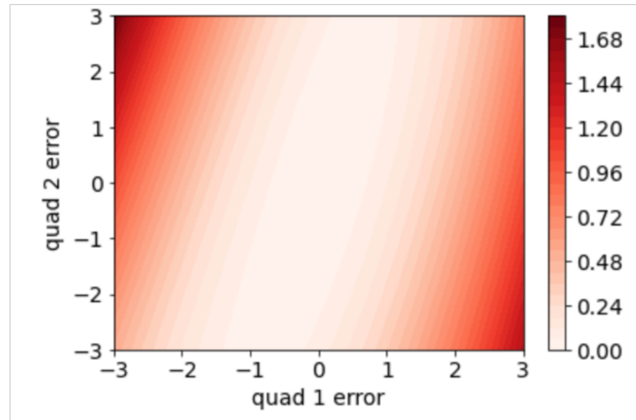

Figure 3: Model error as a function of quadrupole strength error as we vary both quads in the FODO cell.

from the prediction. For the correctors we did include we see very good agreement between the model prediction and the ground truth. We then included single quadrupole errors to evaluate the feasibility of detecting such an error with the inverse model. Figure 4 shows the predicted corrector settings vs the ground truth for three cases. Black is the beamline without any quadrupole errors. Blue is the beamline with a single quadrupole excitation error of negative 20%. Red is the same quadrupole error as blue but with a positive 20% excitation. Here we can clearly see the model can detect the error and even can pinpoint the location of the quadrupole error.

For tuning it is important to understand how the model error varies as each quadurpole magnet is changed. Figure 5 shows the model error as a function of quadrupole strength for six of the magnets in the beamline. Here we can clearly see that the model has very different sensitivities to different magnets. Moreover the model error is not always centered around the nominal excitation. To address this we consider the use of ensemble methods.

ENSEMBLE METHODS

Ensemble methods allow one to generate an aggregate prediction when individual models don't perform as well. This is especially helpful when the model is less sensitive to the particular quadrupole error. Figure 6 shows the model error as a function of quadrupole strength for an ensemble of 25 models. Each model is trained on the same data, however, random initializations are used for the weights.

When we aggregate the output of these models we see a clear improvement in the model error as a function of quadrupole strength. Figure 6 shows the median and the mean error of the ensemble predictions as a function of quadrupole strength. Note that the a quadrupole strength of zero is no error in the quadrupole setting. Compared with Fig. 5, the curves that were not centered around the correct quadrupole excitation are now behaving correctly. While the relationship between model error and quadrupole excitation is not the same strength along the beamline, this approach

ISBN: 978-3-95450-241-7
ISSN: 2673-5350
doi:10.18429/JACoW-IBIC2022-TUP39

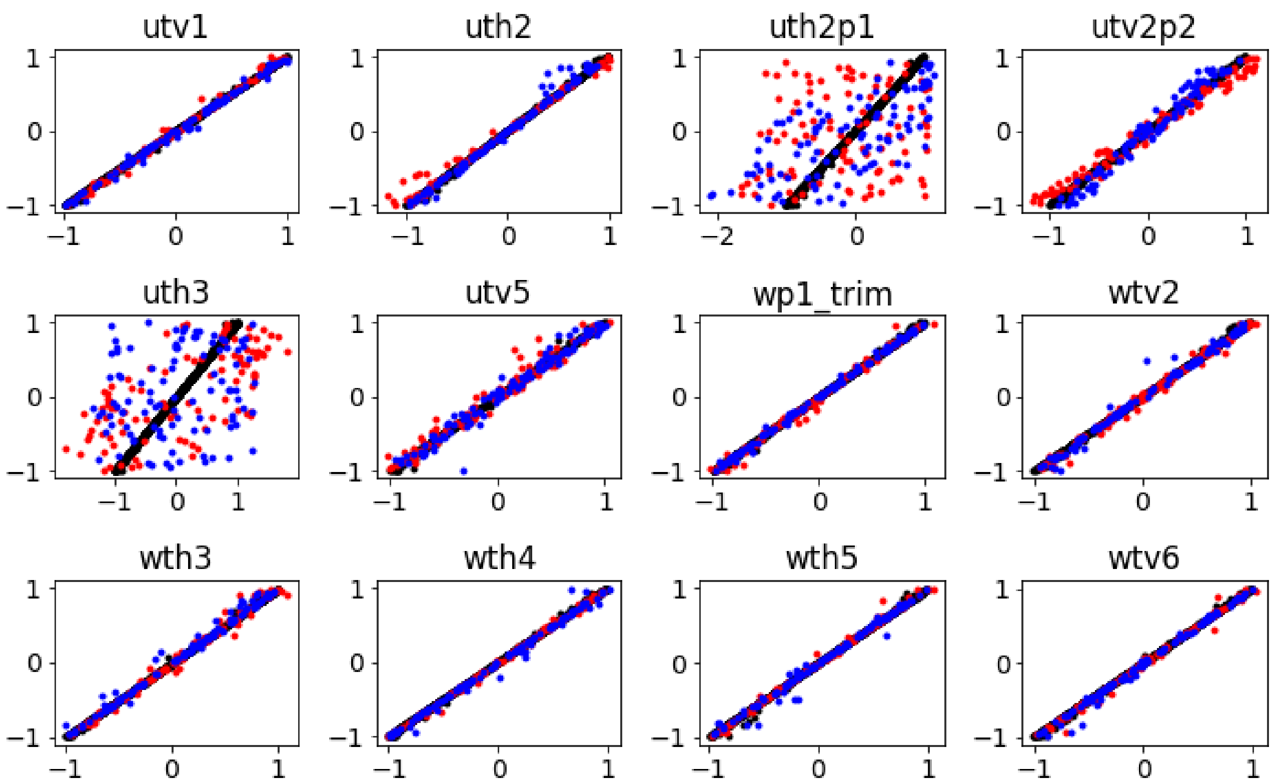

Figure 4: Black: model prediction vs ground truth with no quadrupole errors. Blue: model prediction vs ground truth for a negative 20% quadrupole error. Red: model prediction vs ground truth for a positive 20% quadrupole error.

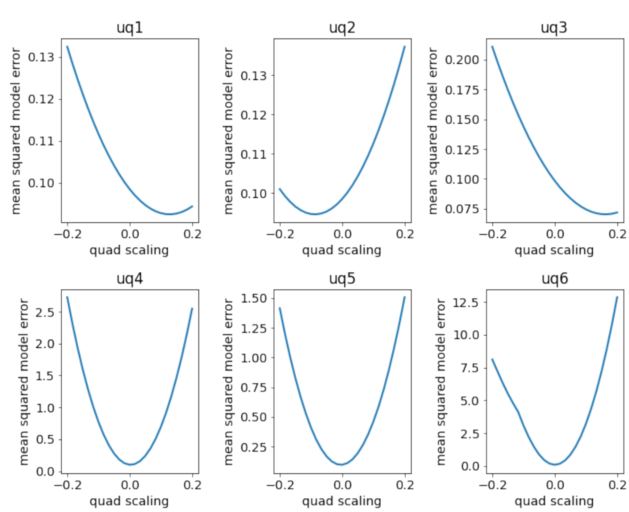

Figure 5: Model error as a function of quadrupole strength for six of the magnets.

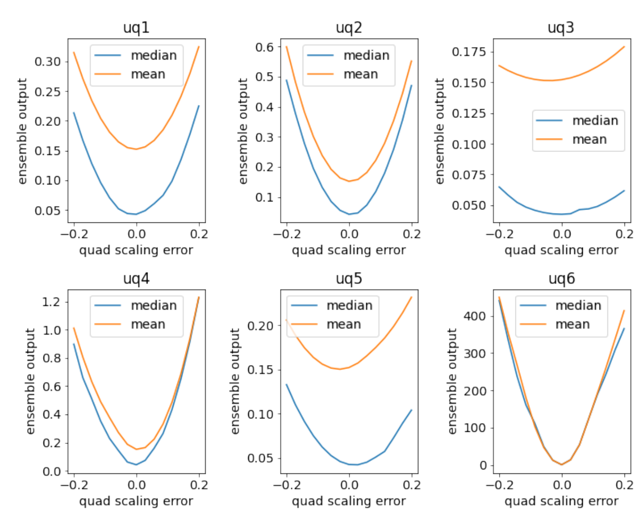

Figure 6: Model error as a function of the quadrupole excitation strength for six of the quadrupoles. Here we show the median and mean output of the ensemble.

demonstrates that an ensemble can improve the performance of the inverse model method.

Another advantage of the ensemble method is gaining uncertainty information. This is important when using the method as a diagnostic as the uncertainty in the diagnostic is an important consideration for operations. Figure 7 shows the ensemble output with the uncertainty (standard deviation in the ensemble) as the shaded region.

There is a clear correlation between the quadrupoles with high uncertainty and the quadrupoles that didn't have a good correlation between the model error and the quadrupole errors in Fig. 5.

ISBN: 978-3-95450-241-7 ISSN: 2673-5350 doi:10.18429/JACoW-IBIC2022-TUP39

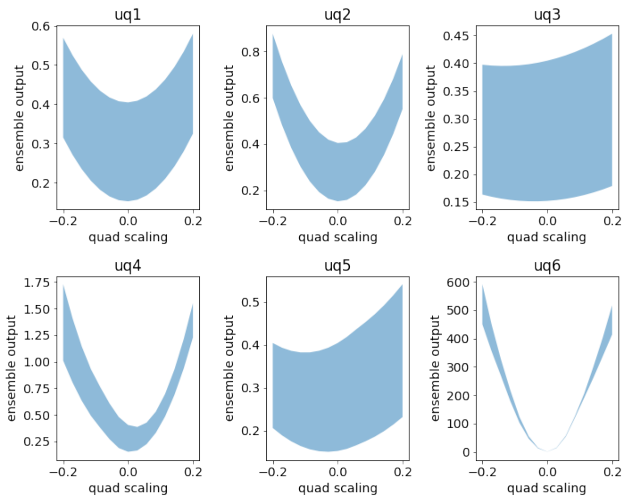

Figure 7: Model error as a function of the quadrupole strength for six of the quadrupoles. Here we show the ensemble mean and the variance.

CONCLUSIONS

In this paper we have refined the use of a neural network inverse model to detect quadrupole strength errors when trained on bpm and corrector data for a nominal setup. This method can be used as a diagnostic or for model-based tuning. We also show the use of ensemble methods to improve the quality of the model prediction; additionally, ensemble methods can provide a rudimentary uncertainty metric. We are continuing to refine our techniques in addition applying our technique to measured data collected during this period of performance.

ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Award Number DE-SC0019682

REFERENCES

[1] A. L. Edelen, S. Biedron, B. Chase, D. Edstrom, S. Milton, and P. Stabile, "Neural networks for modeling and control of particle accelerators," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 2, pp. 878–897, 2016. doi:10.1109/TNS.2016.2543203

[2] A. Edelen *et al.*, "Opportunities in machine learning for particle accelerators," 2018. doi:10.48550/arXiv.1811.03172

[3] C. Emma, A. Edelen, M. Hogan, B. O'Shea, G. White, and V. Yakimenko, "Machine learning-based longitudinal phase space prediction of particle accelerators," *Phys. Rev. Accel. Beams*, vol. 21, no. 11, p. 112 802, 2018. doi:10.1103/PhysRevAccelBeams.21.112802

[4] A. Edelen, S. Biedron, S. Milton, and J. Edelen, "First steps toward incorporating image based diagnostics into particle accelerator control systems using convolutional neural networks," 2016. doi:10.48550/arXiv.1612.05662

[5] A. L. Edelen, S. Biedron, S. V. Milton, and P. J. M. van der Slot, "Using A Neural Network Control Policy For Rapid Switching Between Beam Parameters in an FEL," in *Proc. FEL'17*, Santa Fe, NM, USA, Aug. 2017, pp. 488–491. doi:10.18429/JACoW-FEL2017-WEP031

[6] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and A. Lutman, "Demonstration of model-independent control of the longitudinal phase space of electron beams in the linac-coherent light source with femtosecond resolution," *Phys. Rev. Lett.*, vol. 121, no. 4, p. 044 801, 2018. doi:10.1103/PhysRevLett.121.044801

[7] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelmann, "Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems," *Phys. Rev. Accel. Beams*, vol. 23, no. 4, p. 044 601, 2020. doi:10.1103/PhysRevAccelBeams.23.044601

[8] J. P. Edelen and C. C. Hall, "Autoencoder based analysis of rf parameters in the fermilab low energy linac," *Information*, vol. 12, no. 6, p. 238, 2021. doi:10.3390/info12060238

[9] J. P. Edelen and N. M. Cook, "Anomaly detection in particle accelerators using autoencoders," 2021. doi:10.48550/arXiv.2112.07793