Active Learning of Deep Global Metamodels

Ross Alexander

Department of Aeronautics and Astronautics
Stanford University
Stanford, CA 94305
rbalexan@stanford.edu

I. PROJECT PROPOSAL

We propose to explore neural networks (NNs), Gaussian processes (GPs), and their deep variants (DNNs and DGPs) in global metamodeling tasks with active learning. Global metamodeling is the discipline concerned with constructing a surrogate model over the input space that is optimal with respect to the number of training examples. Traditionally, global metamodels are constructed following detailed sampling schemes. Using these sampling schemes, each of the training examples is evaluated (or labeled) and the global metamodel is constructed using all of the labeled examples. Since these global metamodels generally increase in accuracy with the number of training examples, it is relevant to identify samples that will lead to optimal metamodel accuracy increases utilizing active learning rather than traditional static sampling plans. Active learning in global metamodeling (also called adaptive sampling) can be done in a variety of ways, including mini-batch and sequential sampling.

We want to investigate neural networks and Gaussian processes and their deep variants due to the well-demonstrated expressive power of these models. There are some key differences between NNs and GPs that make them worth considering. DNNs and DGPs have grown in popularity due to their principled composability and their ability to learn strongly nonlinear mappings from inputs to latent variables to outputs. This expressivity is manifested in different ways in each model; in DNNs, the choices of activation functions, learning rates, and DNN architectures govern the learned representation, while in DGPs, the choice of mean function, covariance function (kernel), noise variance, and DGP architectures govern the learned representation [1]. While NNs and GPs both have methods for generating the expectation of the input space, GPs also provide methods for generating the variance of the input space. This offers much more flexibility for adaptive sampling objectives. On the other hand, the cost of training a GP using n samples is $\mathcal{O}(n^3)$, which leads to computational intractability when the dataset is large.

There are several relevant adaptive sampling techniques that have emerged from the surrogate optimization discipline [2]. Some interesting approaches for GPs are error-based, variance-based, and hybrid error-and-variance-based active learning [3]. For both NNs and GPs, other adaptive sampling techniques include incremental Latin hypercube sampling (iLHS), and some techniques based on Voronoi cell decomposition: local linear approximation (LOLA-Voronoi), cross-validation (CV-Voronoi), and k-fold cross-validation (KFCV-Voronoi) [4]–[6]. We also note mixture-of-experts methods with blending factors as an alternative to reduce GP training time in both low- and high-dimensional design spaces.

We will evaluate some of these models and adaptive sampling techniques on several general optimization benchmark functions of varying dimensionality and modality. Some useful accuracy metrics are root-mean-squared error (RMSE), integrated squared error (ISE), integrated squared variance (ISV), and integrated variance (IV). We can also report model generation time as a measure of time complexity. There are several machine learning packages for Julia (e.g. `Flux.jl`, `Scikit-Learn.jl`) and for Python (e.g. `Tensorflow`, `PyTorch`, `Scikit-Learn`). Additionally, there are several popular GP packages for Julia (e.g. `GaussianProcesses.jl`, `Stheno.jl`) and for Python (e.g. `GPy`, `GPFlow`, `GPyTorch`, `PyDeepGP`). We will likely have to build our own implementation of some adaptive sampling algorithms since these are not as widely proliferated.

Ultimately, improvements in global metamodeling using active learning on deep representations could lead to increased sample efficiency – saving time and money across many experimental domains including engineering, biology, computer science.

REFERENCES

[1] A. C. Damianou and N. D. Lawrence, "Deep Gaussian processes," in *Journal of Machine Learning Research*, 2013.

[2] H. Liu, Y. S. Ong, and J. Cai, "A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design," 2018.

[3] H. Liu, J. Cai, and Y. S. Ong, "An adaptive sampling approach for Kriging metamodeling by maximizing expected prediction error," *Computers and Chemical Engineering*, 2017.

[4] J. Eason and S. Cremaschi, "Adaptive sequential sampling for surrogate model generation with artificial neural networks," *Computers and Chemical Engineering*, 2014.

[5] K. Crombecq, L. De Tommasi, D. Gorissen, and T. Dhaene, "A novel sequential design strategy for global surrogate modeling," in *Proceedings - Winter Simulation Conference*, 2009.

[6] A. L. Kaminsky, Y. Wang, K. Pant, W. N. Hashii, and A. Atachbarian, "Adaptive sampling techniques for surrogate modeling to create high-dimension aerodynamic loading response surfaces," in *2018 Applied Aerodynamics Conference*, 2018.