Guided Neural Architecture Search

Project Proposal For Bachelor Thesis

Motivation

Designing neural networks is a task that requires vast experience and time until a successful architecture is found. There are many tasks that rely on minor modifications to the architecture of the network for it to achieve an acceptable test loss on a unseen dataset.

Current work exists in which reinforcemnet learning [**?**], genetic algorithms [**?**] and other hybrid improved search methods exists [**?**]. All of these methods achieve very high accuracy rates, but rely on using exhaustive computational resources with up to 800 GPUs to explore their hypothesis space. A more practical method which can be deployed with realistic resources is yet to be found, although some initial work in this direction exists [**?**].

As of now, it is considered an art, rather than a methodological procedure to find a suitable neural network architecture that solves a task effectively. However, there are various applications in medicine and other domains that could strongly benefit from a 'one-click-training' on a dataset to generate and train a universal function approximator such as a neural network.

Bayesian optimization is a candidate to be considered as an efficient optimization technique that designs and trains a neural network on a given dataset. It automatically finds an architecture which maximizes the cost function for a given dataset. Bayesian Optimization is chosen for it's past empirical success in black-box optimization, and some guarantees to find optimal values in functions that have high cost to evaluate. Neural networks are chosen as machine learning classifiers, as these provide 'smoother' modifications, when compared to discrete model-selection amongst traditional machine learning models.

Scope And Challenges Of The Project

The goal of this project is to formulate and build a bayesian optimization framework that can be used to automatically design and train neural network architectures. The plan is to revise such a BO-algorithm, that will train a neural network that will achieve high performance on the MNIST dataset.
Initially and for simplicity, we set the depth of the network to a set size, and view each tunable parameter (as enumerated above) as a different dimension over which we can evaluate the function $\hat{f}$ on. The bayesian optimization can use a Gaussian Process, which will then get as an input the value of $\hat{f}(x)$, and over time explore the neural-network state-space until a suitable architecture is found.

MNIST is a simple dataset, and thus this project should be seen as a proof of concept, paving the way for future research in automated machine learning. Some challenges that will arise is the curse of dimensionality in the process of Gaussian Processes. However, for the scope of the MNIST dataset, we can keep the dimensions of the network simple at the beginning of this project, and also work on this issue if time allows for it.

References

[1] B. Zoph, Q. V. Le, and G. Brain, "NEURAL ARCHITECTURE SEARCH WITH REINFORCEMENT LEARNING." [Online]. Available: https://arxiv.org/pdf/1611.01578.pdf

[2] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, "Large-Scale Evolution of Image Classifiers." [Online]. Available: https://arxiv.org/pdf/1703.01041.pdf

[3] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *ArXiv*, p. 49, 2010. [Online]. Available: https://arxiv.org/pdf/1012.2599.pdf

[4] M. Wistuba, "Finding Competitive Network Architectures Within a Day Using UCT," 2017. [Online]. Available: http://arxiv.org/abs/1712.07420