Designing Convolutional Neural Network Architecture Using Genetic Algorithms

Ashray Bhandare and Devinder Kaur
Department of EECS, The University of Toledo, Toledo, Ohio, USA

Abstract – In this paper, *genetic algorithm* (GA) is *used to optimally determine the architecture of a convolutional neural network (CNN) that is used to classify handwritten numbers. The CNN is a class of deep feed-forward network, which have seen major success in the field of visual image analysis. During training, a good CNN architecture is capable of extracting complex features from the given training data; however, at present, there is no standard way to determine the architecture of a CNN. Domain knowledge and human expertise are required in order to design a CNN architecture. Typically architectures are created by experimenting and modifying a few existing networks. The* GA *determine the exact architecture of a CNN by evolving the various hyperparameters of the architecture for a given application. The proposed method was tested on the MNIST dataset. The results show that the genetic algorithm is capable of generating successful CNN architectures. The proposed method performs the entire process of architecture generation without any human intervention.*

Keywords: Convolutional Neural Network, Genetic Algorithm, MNIST data.

1 Introduction

The idea that programmable computers will become intelligent was conceived over a hundred years before one was built. AI has tackled and solved many problems that are intellectually difficult for human beings but relatively straightforward for computers. Such problems are defined by a set of mathematical rules. The challenge for AI is to transform tasks which are easy and intuitive for humans into formal procedures that a computer can understand. For example, it is easy for humans to recognize a face, a piece of music even when the data is corrupted or incomplete.

With the advancements in big data, Graphical Processing Unit (GPU) technology and algorithms there has been a lot of progress in the field of Deep Learning. Deep Learning is part of machine learning techniques and allows a machine to learn with experience and data. It makes use of artificial neural networks with more than one hidden layer. By implementing more layers and more neurons within a layer, it allows the network to understand complex ideas by building upon simpler ones. For example, a deep network can build the concept of an image of a car by combining simpler concepts, such as edges, corners, contour, and object parts [1].

Convolutional Neural Network (CNN) is one such type of deep networks. Yann LeCun carried out one of the first exercises on CNN. He taught a computer system how to recognize the differences between handwritten digits [2]. When the system chose incorrectly, he would correct it until the program figured out the reason it was wrong. These networks make use of the mathematical operation called convolution. Convolution is a specialized kind of linear operation. Unlike conventional neural networks, CNN's use this linear operation to obtain an intermediate output (feature) before using it as an input for the next layer. This is done in at least one of their layers. A typical CNN architecture consists of layers such as convolution layer, pooling layer, and Fully connected layer. Each of these layers consists of hyper-parameters that are chosen by researchers using new theoretical insights or intuition gained from experimentation. In this paper, we achieved the following objectives:

- Automate the process of CNN architecture selection.
- Achieve the architecture by evolving the hyperparameters of CNN using Genetic Algorithm (GA)
- Discover CNN architectures without any human intervention that perform well on a given machine-learning task.

GA is inspired by biological evolution, used to find globally optimal solutions and makes use of genetic operators such as selection, crossover, and mutation.

The goal of the proposed algorithms is to discover CNN architectures that perform well on a given machine-learning task with no human intervention. Over the course of many generations, Genetic Algorithm picks out the layers and hyperparameters of a CNN model. By discretizing and limiting the layer parameters to choose from, the algorithm is left with a finite but large space of model architectures to search from. It learns through random exploration and slowly begins to exploit its findings to select higher performing models. It receives the testing accuracy as a means of comparison between architectures and ultimately selects the best architecture. The entire process called an evolutionary experiment goes on for many generations until a fully trained suitable CNN model is generated.

This paper is organized as follows. In Section 2, the dataset used for this analysis is introduced. Section 3 presents the mathematical model of CNN. GA is explained in section 4. Section 5 talks about our proposed method to generate CNN architectures using GA. In Section 6, experiments and results are presented. Conclusions are presented in section 7.