NEAP-F: Network Epoch Accuracy Prediction Framework (Student Abstract)

Arushi Chauhan[1], Mayank Vatsa[2], Richa Singh[2]
[1] IIIT Delhi, India,
[2] IIT Jodhpur, India
arushi16019@iiitd.ac.in,{mvatsa, richa}@iitj.ac.in

Abstract

Recent work in neural architecture search has spawned interest in algorithms that can predict the performance neural networks using minimum time and computation resources. We propose a new framework, Network Epoch Accuracy Prediction Framework (NEAP-F) which can predict the testing accuracy achieved by a convolutional neural network in one or more epochs. We introduce a novel approach to generate vector representations for networks, and encode "ease" of classifying image datasets into a vector. For vector representations of networks, we focus on the layer parameters and connections between the network layers. A network achieves different accuracy on different image datasets; therefore, we use the image dataset characteristics to create a vector signifying the "ease" of classifying the image dataset. After generating these vectors, the prediction models are trained with architectures having skip connections seen in current state-of-the-art architectures. The framework predicts accuracies in order of milliseconds, demonstrating its computational efficiency. It can be easily applied to neural architecture search methods to predict the performance of candidate networks and can work on unseen datasets as well.

Introduction and Objective

In this paper, we propose Network Epoch Accuracy Prediction Framework (NEAP-F) which, based on network architecture and image dataset characteristics, predicts the performance of the network on an image dataset without training it. We have also outlined the formal representations of the network architectures and image dataset characteristics that influence classification accuracy, along with comparing the performance of known regression methods in predicting the accuracy of networks. NEAP-F can be used to predict the testing accuracy for an epoch, or to predict the training curve over several epochs, which can be incorporated in neural architecture search techniques to evaluate how well a candidate network performs on an image dataset, and use this evaluation to propose networks. Through this work, we reduce the computational time and resources involved in examining network performance on an image dataset by substituting training with prediction of its training performance. The image dataset vector generation approach enables the network to predict a sample network's performance on unseen image datasets. We provide a vector representation for network architectures with skip connections by modelling them as graphs. Finally, we demonstrate the results of sample networks trained on CIFAR-10 (Krizhevsky, Nair, and Hinton 2009), MNIST (LeCun, Cortes, and Burges 2010) and SVHN (Netzer et al. 2011) datasets with their testing accuracies over initial epochs of training.

Proposed Framework

The inputs to NEAP-F are sample network architecture, epoch and image dataset representations, and the intended output is the accuracy of the network. The steps involved in the proposed framework are summarized below.

Architecture Data Points Generation: The sample networks have been trained on 3 different datasets, CIFAR-10 (Krizhevsky, Nair, and Hinton 2009), SVHN (Netzer et al. 2011) and MNIST (LeCun, Cortes, and Burges 2010). To accommodate sample networks with skip connections in the framework, we trained networks on CIFAR-10 and MNIST datasets with skip connections. We augment the dataset with more sample architectures using the MetaQNN architecture samples given in the dataset in (Baker et al. 2017). We have 2748 sample architectures and their training curves, a total of 50670 pairs - each having a $< sample\ network, image\ dataset, epoch >$ triplet and its corresponding accuracy.

Vectorization of Network Architecture Samples and "Ease" of Classifying Dataset: To convert sample networks to vectors, we consider an analogy between networks and graphs. Each edge (connection) starts from a source layer (node) i and ends at a destination layer j. This helps to incorporate skip connections in network architecture vector. To each edge, we have assigned a weight which can be seen in equation 1 below. The value $weight(i, j)$ signifies the proportional number of channels that the source layer contributes to the destination layer using this edge.

$$weight(i, j) = \frac{number\ of\ output\ channels\ of\ layer\ i}{number\ of\ input\ channels\ of\ layer\ j} \tag{1}$$

We have encoded the edges using equation 2 below

$$edge(i, j) =< type(i), parameters(i), i, j, weight(i, j) > \tag{2}$$

Model	MSE	MAE	Relative
Ordinary Least Squares	0.0113	0.0794	0.1534
Bayesian Ridge Regressor	0.0113	0.0794	0.1542
SVM Regression	0.0076	0.0627	0.1211
XGBoost Regression	0.0078	0.0558	0.1242
Decision Tree Regressor	0.0071	0.0322	0.0727
Random Forest Regressor	**0.0038**	**0.0287**	**0.0687**

Table 1: Testing Error of Different Regression Models

where, $type(i)$ returns the value mapping of type of layer i and $parameters(i)$ returns the parameters of layer i. Stacking each edge vector into a single vector and padding it gives the vector representation of the network. The intrinsic features of the image dataset contribute to the performance of a network on it, apart from network architecture and its hyperparameters. We have identified number of classes, class imbalance, minimum and maximum inter-class similarity, and minimum and maximum intra-class similarity in an image dataset as factors pivotal to predicting the accuracy of the network. (Abramovich and Pensky 2019) experimentally showed that more the number of classes, higher is the classification accuracy. Intuitively, a dataset with clearly separated clusters for each class will be easier to classify, hence, the clusters should have lower intra-class distances, and higher inter-class distances.

Accuracy Prediction NEAP-F combines the vector representations of sample architectures, image datasets along with epoch to output the accuracy of the sample network on a particular epoch and image dataset. Hence, this is a regression task and we have compared the performance of ordinary least squares (OLS), Bayesian ridge (BvR), SVM, XGBoost, decision tree, and random forest regression models. The performance is evaluated using mean squared error, mean absolute error, and relative error.

Experiments and Results

The experiments have been performed on a single Tesla K80 GPU and are completed in order of miliseconds, demonstrating that the framework is computationally efficient. The dataset is split into 80% training and 20% testing for evaluating the performance of the models. We train the regressor models to predict the accuracy values, which are in the range [0,1]. Table 1 shows the error of regression models on the testing set. The models are trained on ten-fold cross validation and the hyperparameters are chosen through grid search. We observe that random forest regressor gives the lowest error, followed closely by the decision tree regressor, while OLS and BvR yield the highest errors, indicating poor performance.

The regression models are trained on subsets of size smaller than the original dataset. The samples for the subsets are chosen randomly and the mean squared error and mean absolute error of these models on the testing set are noted. As we can see in Figure 1, the error declines as the subset size increases for all models. Decision tree regressor shows the sharpest decline as subset size increases while

Model	CIFAR10	MNIST	SVHN
SVM Regression	0.0063	0.0402	0.0064
XGBoost Regression	0.0060	0.0265	0.0044
Decision Tree	0.0040	0.0547	0.0052
Random Forest	**0.0023**	**0.0230**	**0.0032**

Table 2: Prediction MSE Errors on Different Datasets

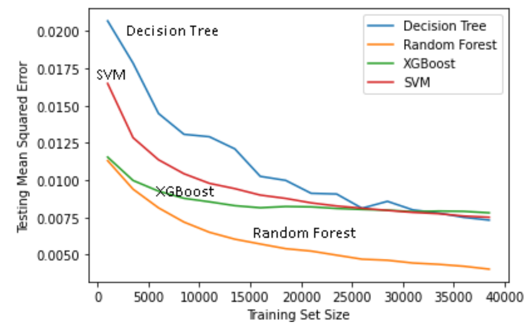

Figure 1: MSE of regression models on smaller subsets.

random forest regressor gives the lowest error for all subset sizes. This experiment demonstrates that the models can give competent performance vis-a-vis lower mean squared error even when less data is available for training. We further examined the performance of regression models in predicting the accuracy achieved by sample networks trained on specific image datasets. A new mini-dataset of network samples and epoch accuracies is prepared for each image dataset and 80:20 train-test split in each such mini-dataset is made for this experiment. Table 2 shows the mean squared error of the models on the networks trained on CIFAR-10, MNIST and SVHN separately. From the experiments, we infer that random forest regressors perform the best out of all the regression models.

References

Abramovich, F.; and Pensky, M. 2019. Classification with many classes: Challenges and pluses. *Journal of Multivariate Analysis* 174(C).

Baker, B.; Gupta, O.; Rasker, R.; and Naik, N. 2017. Accelerating Neural Architecture Search using Performance Prediction. In *International Conference on Learning Representations Workshop*.

Krizhevsky, A.; Nair, V.; and Hinton, G. 2009. CIFAR-10 (Canadian Institute for Advanced Research). http://www.cs.toronto.edu/ kriz/cifar.html. Visited on 2020-03-16.

LeCun, Y.; Cortes, C.; and Burges, C. 2010. MNIST handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist* 2. Visited on 2020-03-17.

Netzer, Y.; Wang, T.; Coates, A.; Bissacco, A.; Wu, B.; and Ng, A. 2011. Reading Digits in Natural Images with Unsupervised Feature Learning. In *Advances in Neural Information Processing Systems*.