Neural Architecture Search

Nikhil dwivedi
CSE Department
IIT Jodhpur
India
dwivedi.10@iitj.ac.in

Poonam Kashyap
CSE Department
IIT Jodhpur
India
kashyap.5@iitj.ac.in

Priyal Jain
CSE Department
IIT Jodhpur
India
jain.63@iitj.ac.in

***Abstract*—We have created number of CNN(convolutional neural network) models and then we have applied the search technique, here we have used hill climbing, to find the best model which will be classify the fashion-mnist dataset with high accuracy and with less parameters.**

I. Introduction

We have been given the fashion-mnist dataset which can be also downloaded from kaggle. The dataset consists of 70,000 28X28 grayscale images out of which 60,000 are for training set and 10,000 for testing set and we have to classify them into 10 classes using CNN model and for searching the best model we are using hill climbing searching technique.[1]

II. Load the fashion-mnist data

We have used fashion-mnist.loaddata() function to load the data and then we have stored the training data and testing data into corresponding variables.

III. CNN Architecture

We have made the base model using four layers.
a) Our first layer is normal(NC) layer which have parameters stride=1 ,padding=same ,kernal size is random from the list of 1 to 7 and activation function is also the list of 5 functions- relu,sigmoid,tanh,swish,gelu.
b) Our second layer is reduction(RC) layer which have parameters stride =2 ,padding=valid ,kernal size is random from the list of 1 to 7 and activation function is also the list of 5 functions- relu,sigmoid,tanh,swish,gelu.
c) Our first layer is normal(NC) layer which have parameters stride=1 ,padding=same ,kernal size is random from the list of 1 to 7 and activation function is also the list of 5 functions- relu,sigmoid,tanh,swish,gelu.
d) Our second layer is reduction(RC) layer which have parameters stride =2 ,padding=valid ,kernal size is random from the list of 1 to 7 and activation function is also the list of 5 functions- relu,sigmoid,tanh,swish,gelu.
e) Then we have the final layer(FL) which helps in classification.[2]

Therefore, in short our CNN architecture has layer in order of NC-RC-NC-RC-FL

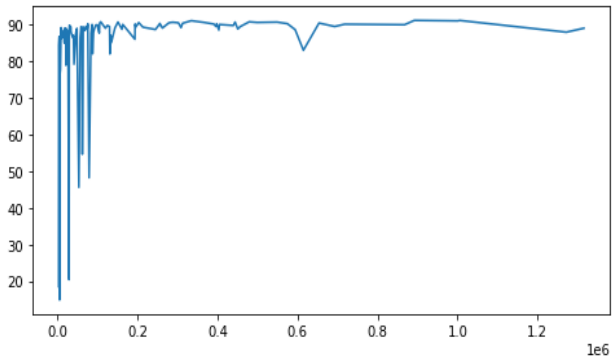

Fig. 1. Plot parameters versus testing accuracy

IV. Train the models and store the data

We have trained our model with different parameters in our CNN architecture.
We have trained around 120 models and then we have stored the data in csv file so that we don't have to train again and again before proceeding to the searching technique to find the best model.
The things that we have stored in csv file are Genome string ,Parameters ,Training Accuracy and Testing Accuracy in their respective order.The genome string is basically the string format of what parameters does the model has while training, so for different models genome string will be different.It contains the details of parameters used in our five layers with comma separated.[3]

V. Load the Dataset created after Training Models

After we got our data which we have created while training the different models we will load it and then we have plot the graph,as shown in the figure below Fig.[1],between parameters and testing accuracy as there is a trade-off between both the parameters so we have to find the model which balances both the parameters.

From the above graph we can easily see that when the parameters are too large the testing accuracy is generally high, but when the parameters are low the testing accuracy may or maynot be high ,so our intrest lies in the first half of the graph.[4]

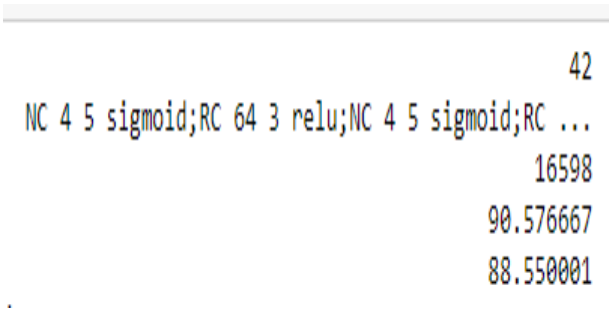

Fig. 2. Result of the best model

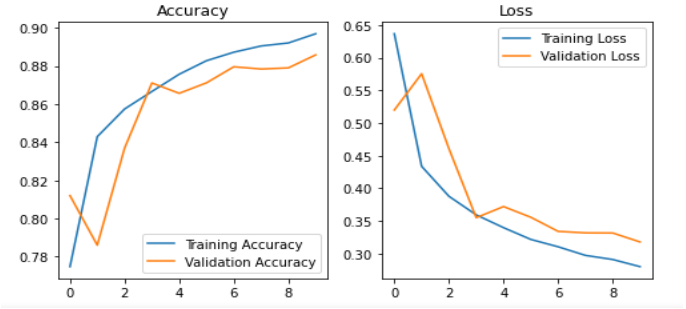

Fig. 3. Comparison between training data and validation data

VI. Searching Technique

Here we have used the Hill Climbing algorithm to find the best model in the collection of different models.

Basically, Hill Climbing algorithm is local search algorithm which continuously moves in the direction of maxima/minima and when it reaches the point where we can't move any further then it returns the optimal point. Here also we have used the similar approach.

At starting we have just chosen random point to start the local search and then we began to search the optimal point by moving left or right. So here basically two cases are there where we have to move left or right (in parameters list).

If suppose the left point is optimal than the current point then we move to the left i.e. testing accuracy is also increasing and parameters are decreasing , which we want as there will be no conflict between both the parameters but suppose the left point is not optimal then the current point then we have to check the right point, if the right point has higher testing accuracy then we will check the fitness function as it's not necessary that if the testing accuracy is increasing by slight margin and the parameters are increasing by the large margin then the right point will not be the optimal point for us as for increasing slight testing accuracy, it will cost us a lot of parameters.

There are some drawbacks of hill climbing so if the point is at maxima and it can't move further but there is still the optimal point then it is stuck there so to counter that we have created the count parameter and if it doesn't gives the optimal point than we will again start choosing the initial point and then we have to apply hill climbing again.[5]

VII. Training of best model

After applying the search technique we get our best model from our collection of different models.Now we got our corresponding genome string from which we can decode it and get the parameters like kernal size, filters and activation function and we can use that to train and fit our model.[7]

VIII. Results

After applying search technique we get the best model,the corresponding parameters are shown in Fig[2].

Although it is not guaranteed that everytime when we apply hill climbing we will get same model everytime as it will depends on hill climbing algorithm and trade-off between testing accuracy and parameters.

The parameters of model we are getting after hill climbing are

genome string -NC 4 5 sigmoid;RC 64 3 relu;NC 4 5 sigmoid;RC 64 3 relu;FL gelu;

total parameters - 16598

training accuracy - 90.576667

testing accuracy - 88.550001

Then we have plot the graph between the accuracy and the loss of both training data and validation data Fig[3]

References

[1] https://www.kaggle.com/zalando-research/fashionmnist.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, "Title of paper if known," unpublished.
[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.