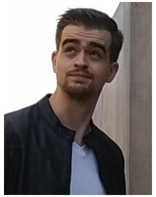
Improving genetic algorithms applied to hyper-parameter optimization using weight merging

The Team



Angel L. González

MSc Data Science Student

Varrun Varatharajan

MSc Autonomous Systems Student

Research Objectives & Goals

Question

Weight merge improves performance of genetic algorithms applied to HPO?

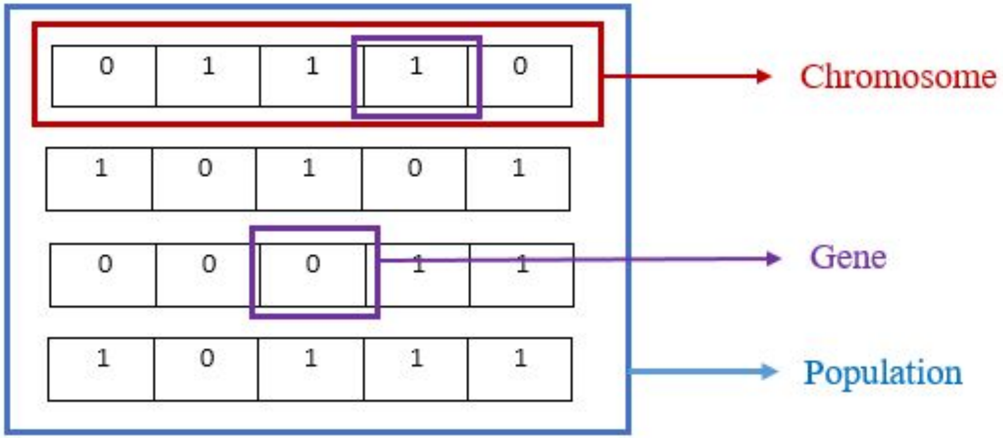
Hypothesis

Weight merge can improve fitness function performance and accelerate optimum search

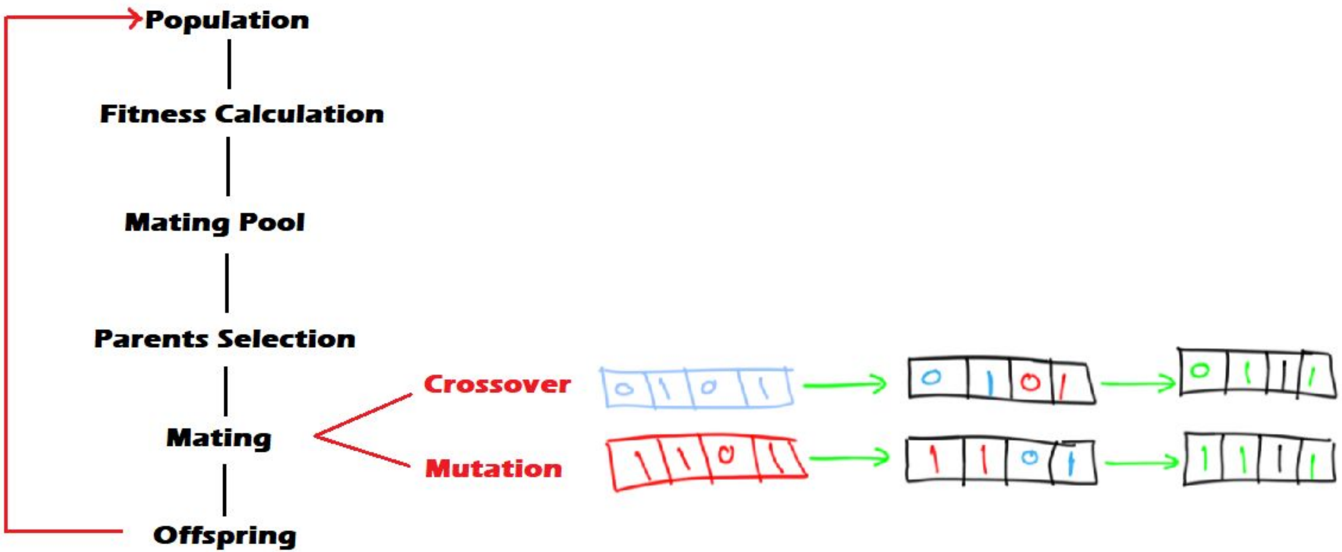
Objective

Design, build and test a tool that improves this algorithm's performance

Genetic algorithms



Genetic algorithms



Background

Comparing backpropagation with a GA for NN training

First use of genetic algorithms to train neural networks

Optimizing NN hyper-parameters through an GA

GA successfully applied in HO

NN weight selection using GA

First test of combination of ANN weights (but not merge)

Large-scale evolution of image classifiers

They tried to merge models without success

> such choices to others. In a separate experiment, we attempted recombining the trained weights from two parents in the hope that each parent may have learned different concepts from the training data. In a third experiment, we recombined structures so that the child fused the architectures of both parents side-by-side, generating wide models fast. While none of these approaches improved our recombination-free results, further study seems warranted.

Automatically designing NN using GA for image classification

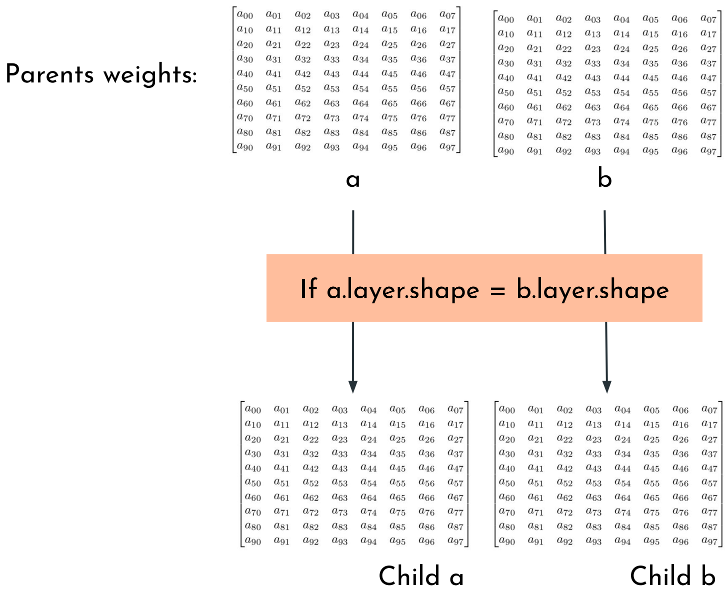
Efforts should be made to improve fitness functions

> several algorithms based on evolutionary computation techniques have been developed. In future, we will place efforts on developing effective evolutionary computation methods to significantly speed up the fitness evaluation of CNNs.

The weight merging technique

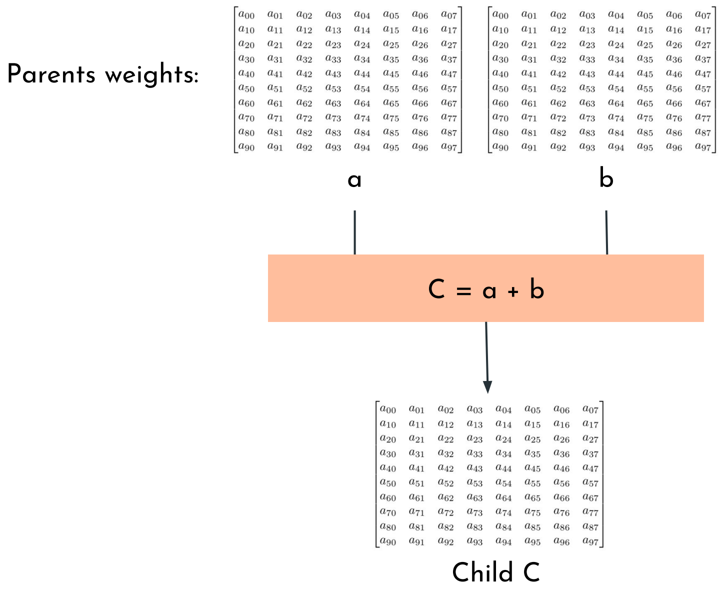
In the paper:

Large-scale evolution of image classifiers



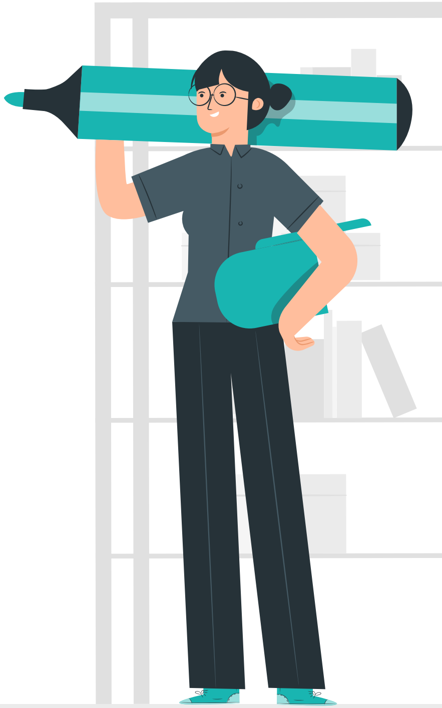
In our project

Weigh merge technique



Research Methodology

- 2 algorithms per experiment
  - Basic genetic algorithm
  - Genetic algorithm + weight merging
- Collect the **duration** of each execution of a fitness function
- Compare for each experiment
  - The distribution of the fitness function duration for each algorithm
  - Duration of the entire algorithm
  - Performance of the final model (error/accuracy)



Datasets

MNIST

70,000 handwritten digits

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10

Boston Housing Price

Boston area housing census figures

506 cases, 14 attributes

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides
0	7.4	0.70	0.00	1.9	0.076
1	7.8	0.88	0.00	2.6	0.098
2	7.8	0.76	0.04	2.3	0.092
3	11.2	0.28	0.56	1.9	0.075
4	7.4	0.70	0.00	1.9	0.076

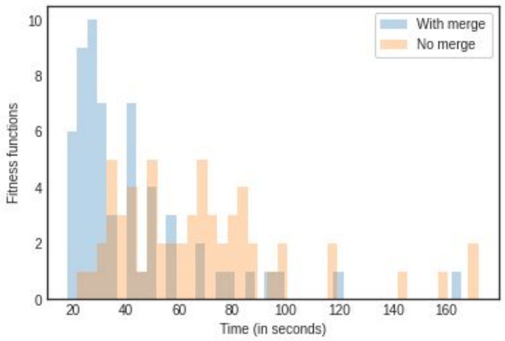
Wine Quality

12 attributes of different wines

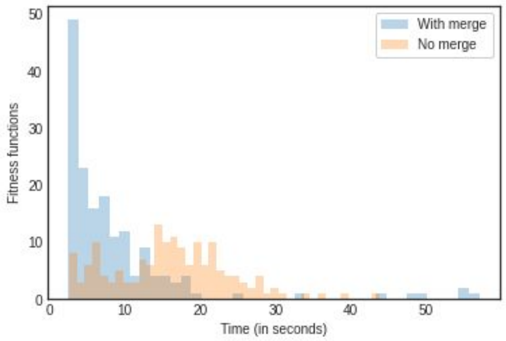
* Y. LeCun, C. Cortes, and C. J. Bruges, “The mnist database of handwritten digits.”
** Harrison, D. and Rubinfeld, D.L. 'Hedonic prices and the demand for clean air', J. Environ. Economics & Management, vol.5, 81-102, 1978.
*** P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

Results

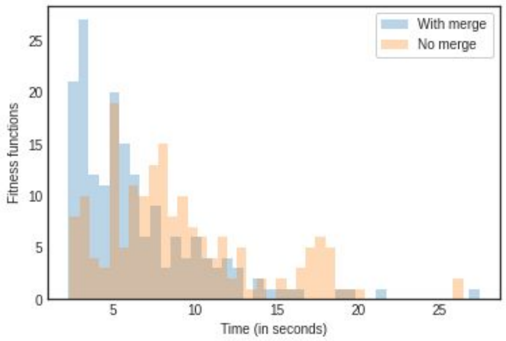
Dataset	Average time (basic)	Average time (merge)	Time skewness (basic)	Time skewness (merge)	Result (basic)	Result (merge)
Minst	70.7	47.3	1.709575	4.503455	accuracy 0.9641	accuracy 0.9745
The Boston house-price	15.8	7.3	0.231586	1.732404	MAE 3.0085	MAE 3.0086
Wine Quality	9.26	6.65	1.471797	2.675971	MAE 0.4493	MAE 0.4804



MNIST



Boston housing price



Wine Quality

Analysis

Results from weight merging method indicate:

- ✔ No negative impacts on performance (accuracy and mean errors)
- ✔ Reduces the average training time by at least 25%*

* With the tested datasets

Conclusion and future work

Weight merge effective in preserving training data in GA

Layer shape is not an obstacle to the implemented technique

Distributed learning. Merge weights trained with different chunks of data

Conv. layers or RNN with weights matrices of different dimensionality

Questions?