A Study of Generalization and Fitness Landscapes for Neuroevolution

Nuno M. Rodrigues[1], Sara Silva[1], Leonardo Vanneschi [2]

[1] LASIGE, Faculty of Sciences, University of Lisbon, Campo Grande, 1749-016 Lisbon, Portugal
[2] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal

Full Paper

data and systems intelligence

Motivation:

Fitness landscapes are a useful concept for studying the dynamics of meta-heuristics. So far, they have been successfully used for estimating the optimization capabilities of different flavors of evolutionary algorithms. However, they have never been used for studying the performance of machine learning algorithms on unseen data, and they have not been applied to studying neuroevolution landscapes.

Goals:

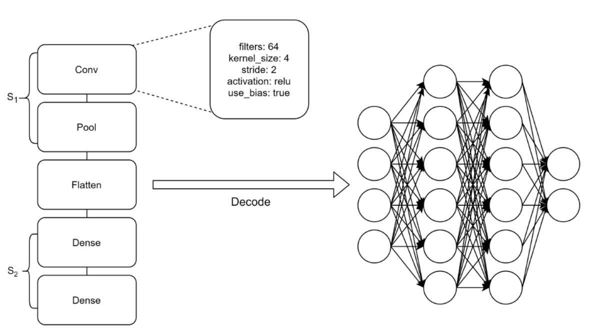
To apply fitness landscapes analysis to neuroevolution, using this concept to infer useful information about the learning and generalization ability of the machine learning method.

Methodology:

- Develop a grammar-based neuroevolution algorithm
- Define genetic operators
- Generate the landscapes
- Apply Autocorrelation, Entropic Measure of Ruggedness, Fitness Clouds, Density Clouds and Overfitting Measure

Grammar:

Conv ::	filters	|32,64,128,256|
	kernel_size	|2,3,4,5|
	stride	|1,2,3|
	activation	|relu, elu, sigmoid|
	use_bias	|true, false|
Pool ::	type	|Max, Avg|
	pool_size	|2,3,4,5|
	stride	|1,2,3|
Dense ::	units	|8,16,32,64,128,256,512|
	activation	|relu, elu, sigmoid|
	use_bias	|true, false|
Dropout ::	rate	[0.0 → 0.7]
Optimizer ::	learning_rate	|0.01, 0.001, 0.0001, 0.00001|
	decay	|0.01, 0.001, 0.0001, 0.00001|
	momentum	|0.99, 0.9, 0.5, 0.1|
	nesterov	|true, false|



Operators:

Conv ::	filters	|32,64,128,256|
	kernel_size	|2,3,4,5|
	stride	|1,2,3|
	activation	|relu, elu, sigmoid|
	use_bias	|true, false|
Pool ::	type	|Max, Avg|
	pool_size	|2,3,4,5|
	stride	|1,2,3|
Dense ::	units	|8,16,32,64,128,256,512|
	activation	|relu, elu, sigmoid|
	use_bias	|true, false|
Dropout ::	rate	[0.0 → 0.7]
Optimizer ::	learning_rate	|0.01, 0.001, 0.0001, 0.00001|
	decay	|0.01, 0.001, 0.0001, 0.00001|
	momentum	|0.99, 0.9, 0.5, 0.1|
	nesterov	|true, false|

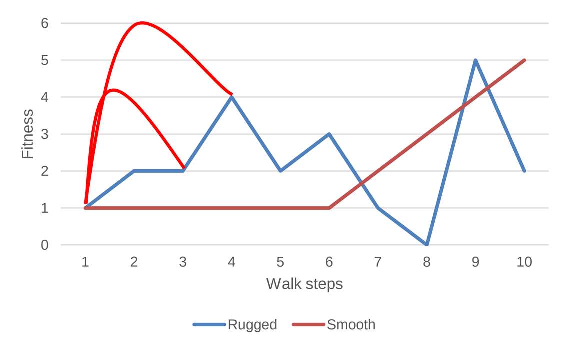
- Topological
- Parameters
- Learning

Problems:

We used four well know computer vision datasets: MNIST, Fashion-MNIST, CIFAR10 and SVHN.

SM **(Small & Mislabeled)** created to cause overfitting. Using the last 30% samples from MNIST and applying label corruption to the odd label values.

Measures:



Autocorrelation:
Used to measure the ruggedness of a landscape by comparing points at different time-steps.

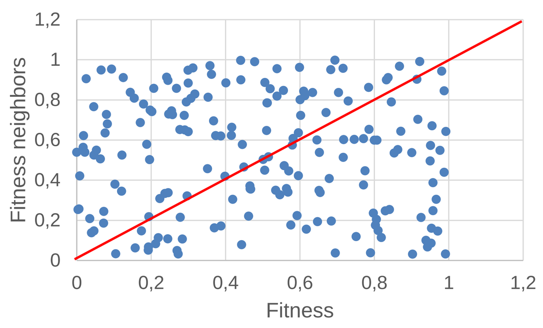
$$\hat{\rho}(k) = \frac{\sum_{t=1}^{n-k}(f(s_t)-\bar{f})(f(s_{t+k})-\bar{f})}{\sqrt{\sum_{t=1}^{n}(f(s_t)-\bar{f})^2}\sqrt{\sum_{t=1}^{n}(f(s_{t+k})-\bar{f})^2}},$$

where $\bar{f} = \frac{1}{n}\sum_{t=1}^{n} f(s_t)$

Encoding	00	01	$0\bar{1}$	10	11	$1\bar{1}$	$\bar{1}0$	$\bar{1}1$	$\bar{1}\bar{1}$
Classification	neutral	rugged	rugged	rugged	smooth	rugged	rugged	rugged	smooth
Shape									

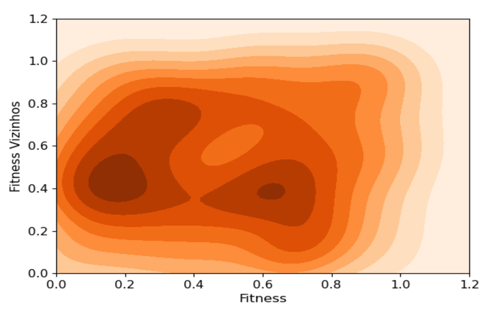
Entropic Measure of Ruggedness:
Indicator of the relationship between ruggedness and neutrality.
We can represent walk timeseries as a string $S(\varepsilon) = \{x_1.x_1 \dots .x_n\}$ where $x_1 \in \{\bar{1}, 0, 1\}$.

$$x_i = \Psi_{f_t}(i,\varepsilon) = \begin{cases} \bar{1}, if\ f_i - f_{i-1} < -\varepsilon \\ 0, if\ |f_i - f_{i-1}| \le -\varepsilon, \\ 1, if\ f_i - f_{i-1} > -\varepsilon \end{cases} \qquad \text{Entropy } H(\varepsilon) = -\sum_{p\neq q} P_{[pq]} \log_6 P_{[pq]},$$



Fitness Clouds:
Mapping between fitness of the individuals and fitness from their neighbors.
set of individuals $S = \{s_1, s_2, \dots, s_n\}$
set of neighbors $V(S_i) = \{v_1^i, v_2^i \dots, v_{m_i}^i\}$,

$$C = \{\left(f(s_i), f(v_k^i)\right), \forall i \in [1,n], \forall k \in [1,m_i]\}$$



Density Clouds:
Alternative measure to density of states.
Produces visual information regarding the density of the points of a fitness cloud.

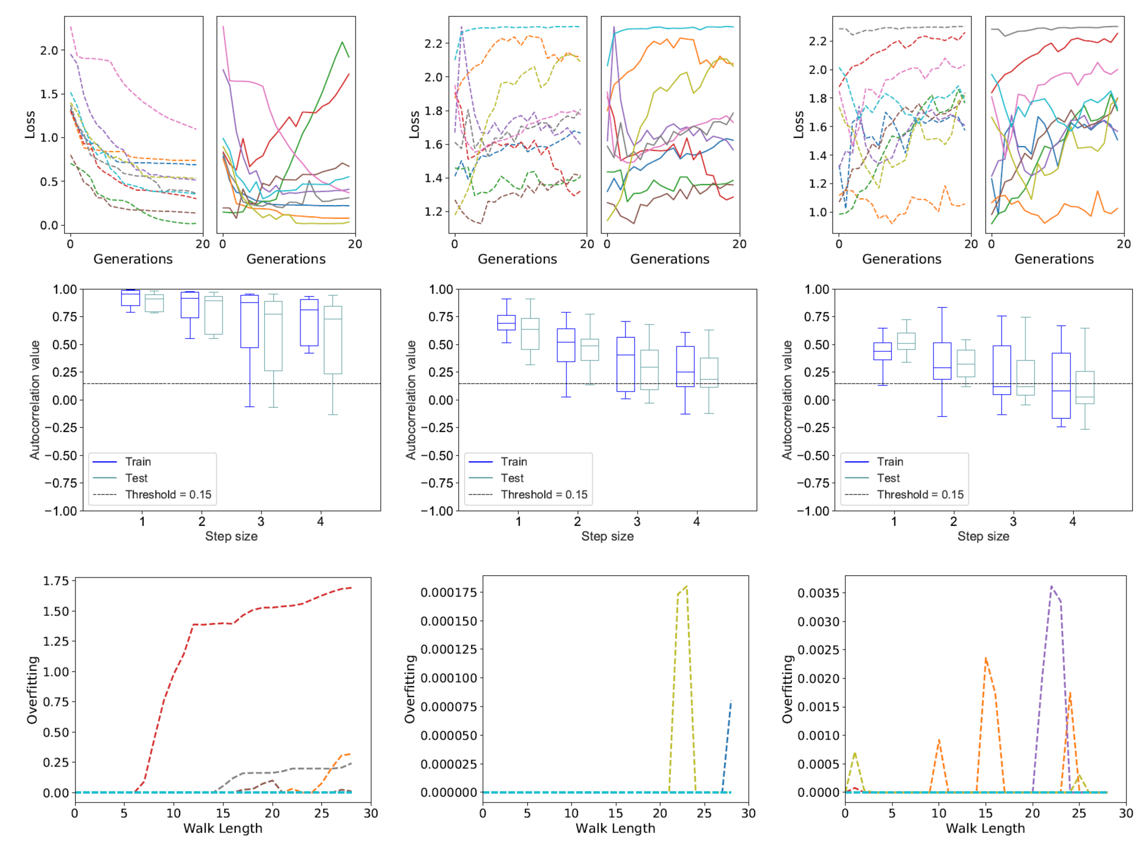
$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

Algorithm 2: Method used to measure overfitting in minimization problems.

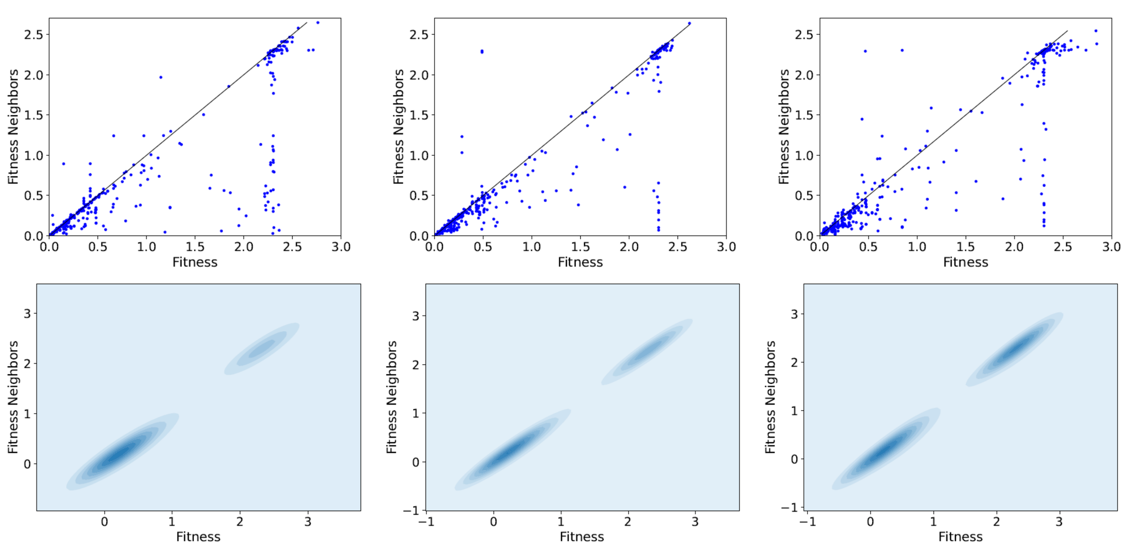
```
1  Length of the walk = n
2  Best test point, btp = 0
3  overfit(btp) = 0
4  for i = 1,2,...,n do
5      if training_fit(i) > test_fit(i) then
6          overfit(i) = 0
7      else
8          if test_fit(i) < test_fit(btp) then
9              btp = i
10             overfit(btp) = 0
11         else
12             diff_now = |training_fit(i)−test_fit(i)|
13             diff_btp = |training_fit(btp)−test_fit(btp)|
14             overfit(i) = max(0, diff_now−diff_btp)
```

Overfitting Measure:
Quantifies overfitting during the evolution of the algorithm.
Originally proposed for genetic programming.

Results:



CIFAR10 results for all 3 operators
Top row shows the evolutionary plots
Middle row shows the autocorrelation plots for 4 step sizes
Bottom row shows overfitting measure plots



		Learning	Parameter	Topology
MNIST	Train	76	81	81
	Test	75	79	76
FMNIST	Train	83	82	83
	Test	78	79	80
CIFAR10	Train	84	83	84
	Test	82	68	81
SVHN	Train	84	77	84
	Test	72	69	68
SM	Train	77	73	79
	Test	43	39	42

MNIST results for all 3 operators
Top row show the fitness clouds
Bottom row shows the density clouds

Table presents the percentage of points below or coincident with the identity line.

	Learning	Parameters	Topology
MNIST	0.29	0.45	0.43
FMNIST	0.41	0.47	0.43
CIFAR10	0.28	0.46	0.50
SVHN	0.41	0.43	0.44
SM	0.45	0.40	0.35

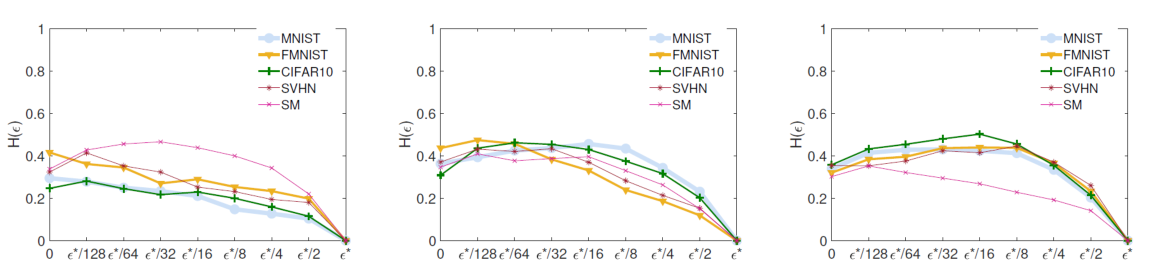


Table presents the R_f for all problems and operators.

$$R_f = max_{\forall \varepsilon \in [0,\varepsilon^*]} H(\varepsilon)$$

Plots show the different entropy values for a given set of sensitivity values.

Conclusions and Future work:

The neuroevolution algorithm works as intended and produces good results, all three mutation operators are viable and produce different types of landscapes, the measures are reasonable indicators regarding problem hardness and neutrality, and the overfitting measure was able to correctly capture the occurrence of overfitting.

We are currently extending this work, in collaboration with Prof. Gabriela Ochoa and Prof. Katherine Malan, to study both the Local Optima Networks and Fitness Distance Correlation of the landscapes produce by neuroevolution, as well as a full study of the architectural search space of CNNs for both learning and generalization landscapes.

ULISBOA · Ciências ULisboa · FCT Fundação para a Ciência e a Tecnologia · LASIGE driven by excellence · NOVA IMS Information Management School

Acknowledgements: This work was supported in part by FCT, Portugal, through funding of LASIGE Research Unit under Grant UIDB/00408/2020 and Grant UIDP/00408/2020, in part by BINDER under Grant PTDC/CCI-INF/29168/2017, in part by GADgET under Grant DSAIPA/DS/0022/2018, in part by AICE under Grant DSAIPA/DS/0113/2019), and in part by PREDICT under Grant PTDC/CCI-CIF/29877/2017.