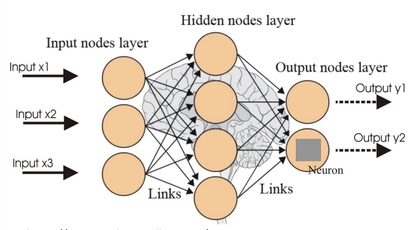
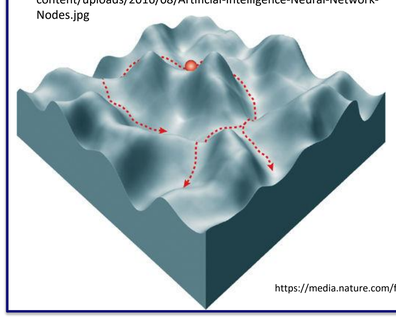
Efficient Learning in Deep Learning through Evolutionary Computations

Sai Krishna Kalluri, Technical University of Eindhoven, Eindhoven. Email: s.k.kalluri@student.tue.nl

Motivation



https://www.analyticsvidhya.com/wp-content/uploads/2016/08/Artificial-Intelligence-Neural-Network-Nodes.jpg



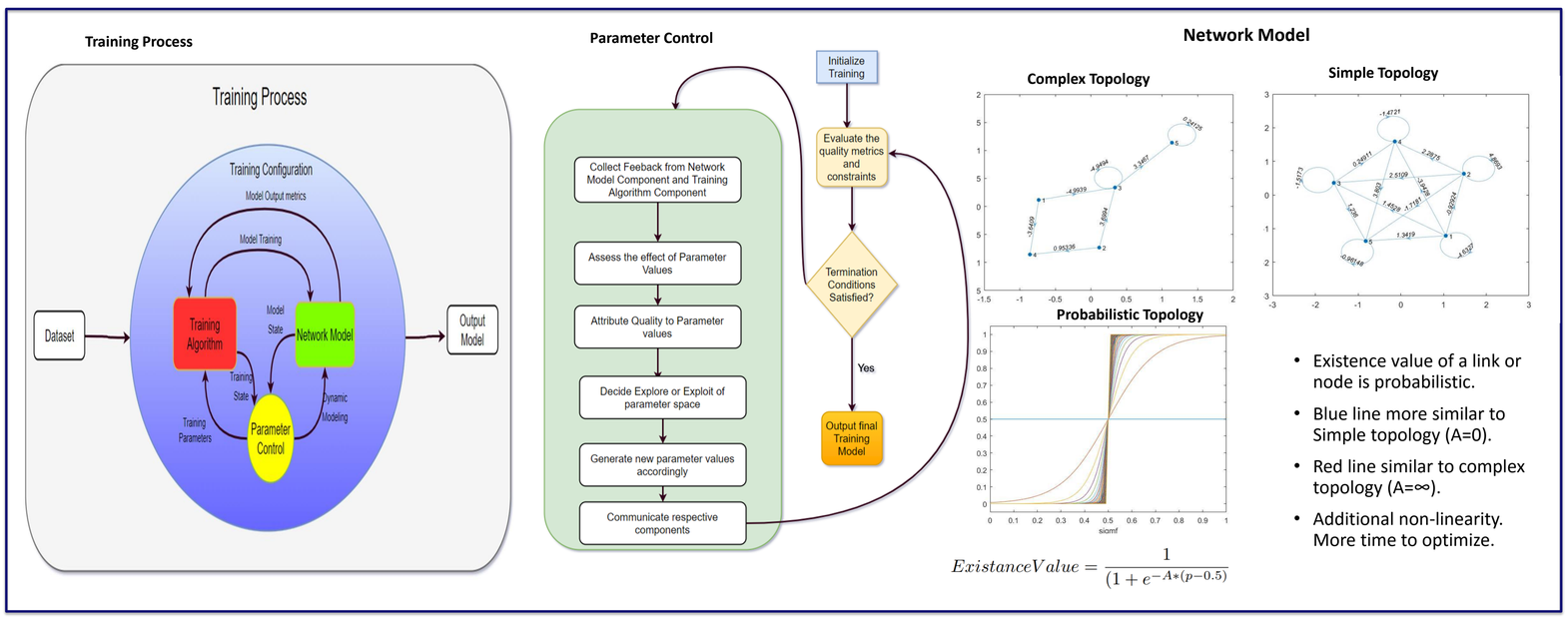
https://media.nature.com/full/nature-assets/nrm/journal/v10/n10/images/nrm2766-f2.jpg

Nature Reviews | Molecular Cell Biology

Training efficiently artificial neural networks without falling in local minima.

- Optimize topology in addition to weights.
- Reduce Sparse connections. Reduce memory consumption
- More generalization to new data.
- High quality with low training data
- Support Multi-Objective optimization and satisfy network constraints.

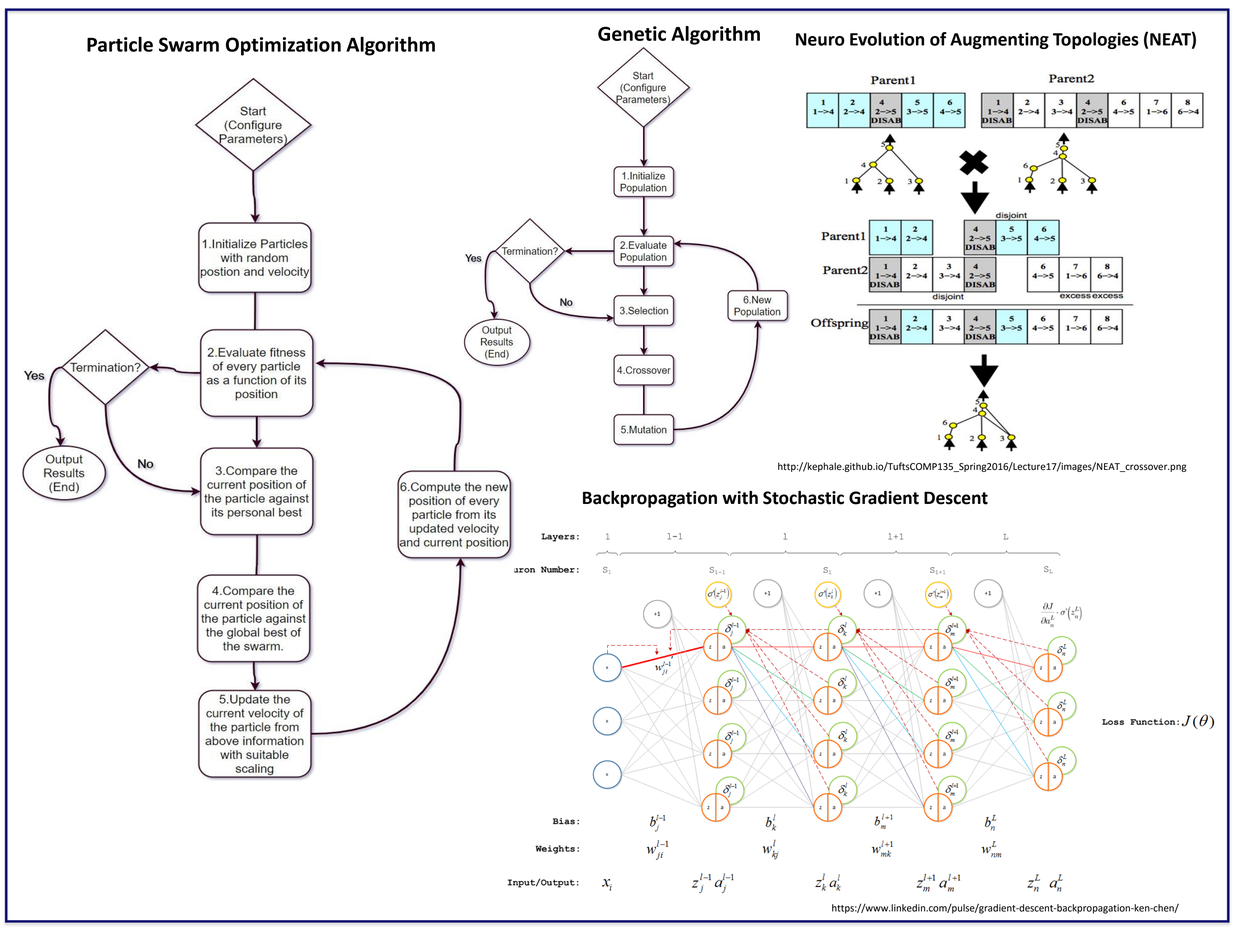
Methodology



$$ExistanceValue = \frac{1}{(1 + e^{-A*(p-0.5)}}$$

- Existence value of a link or node is probabilistic.
- Blue line more similar to Simple topology (A=0).
- Red line similar to complex topology (A=∞).
- Additional non-linearity. More time to optimize.

Training Algorithms



http://kephale.github.io/TuftsCOMP135_Spring2016/Lecture17/images/NEAT_crossover.png

https://www.linkedin.com/pulse/gradient-descent-backpropagation-ken-chen/

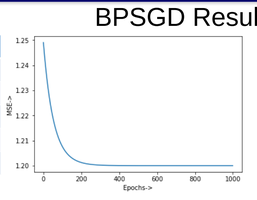
Experiments and Results

XOR Dataset Results

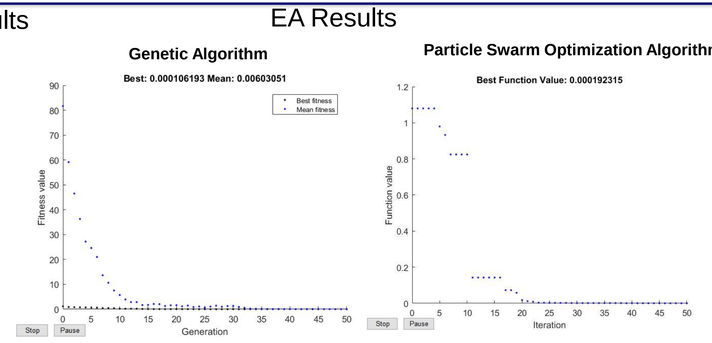
Dataset

X0	X1	Out
0	0	0
1	0	1
0	1	1
1	1	0
0.5	0.5	0

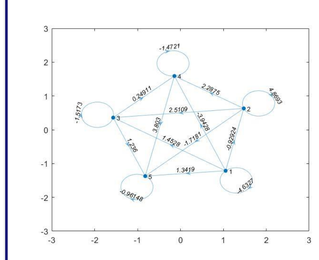
BPSGD Results



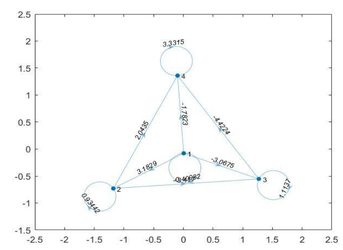
EA Results



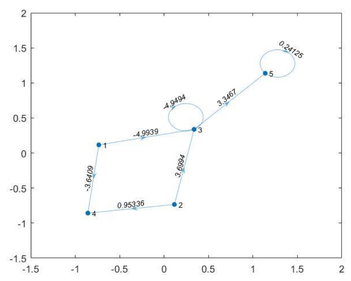
Genetic Algorithm+Simple (5 Nodes)

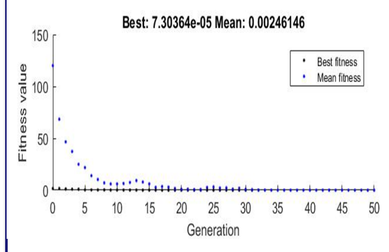


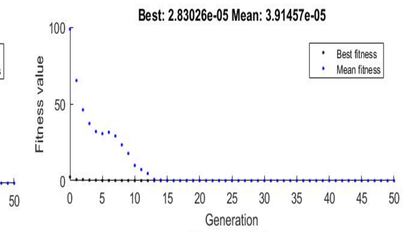
Genetic Algorithm+Simple (4 Nodes)

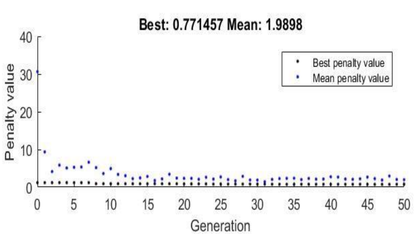


Genetic Algorithm+Complex (5 Nodes)

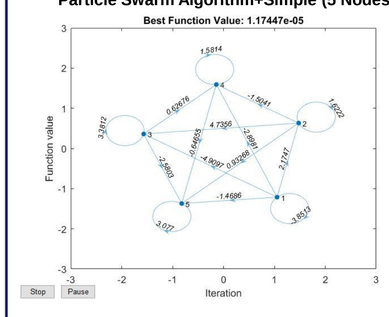




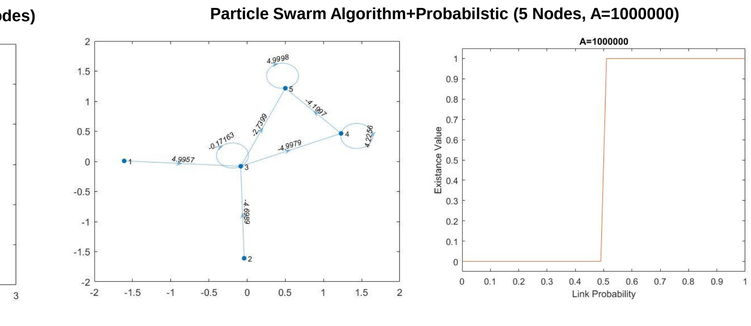


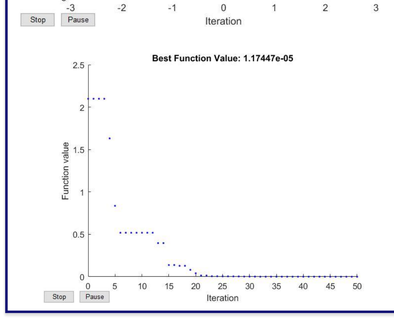


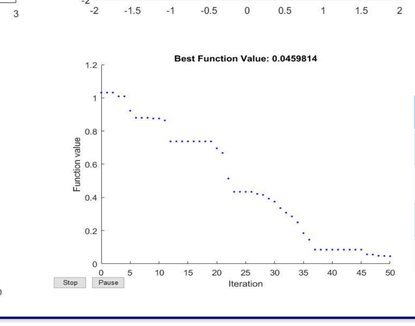
Particle Swarm Algorithm+Simple (5 Nodes)



Particle Swarm Algorithm+Probabilstic (5 Nodes, A=1000000)







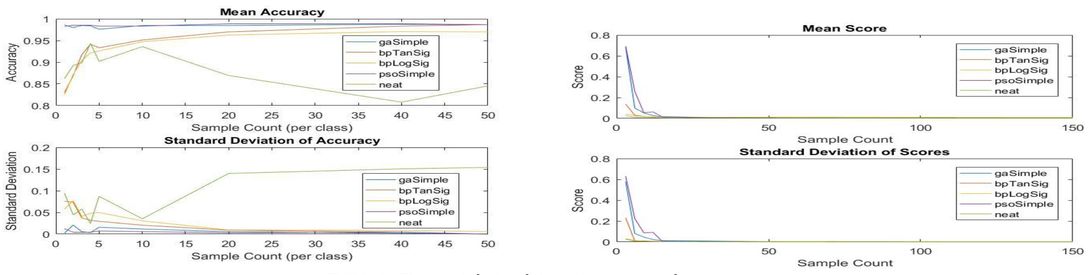
Pareto Front Fitness vs Memory

Configuration	Fitness Value	Memory (in links)
PSO+Simple (5 Nodes)	1.174477e-05	15
GA+Simple (4 Nodes)	2.8302e-05	10
PSO+Feedforward	0.000192315	9
PSO+Probabilistic (5 Nodes)	0.0459814	8

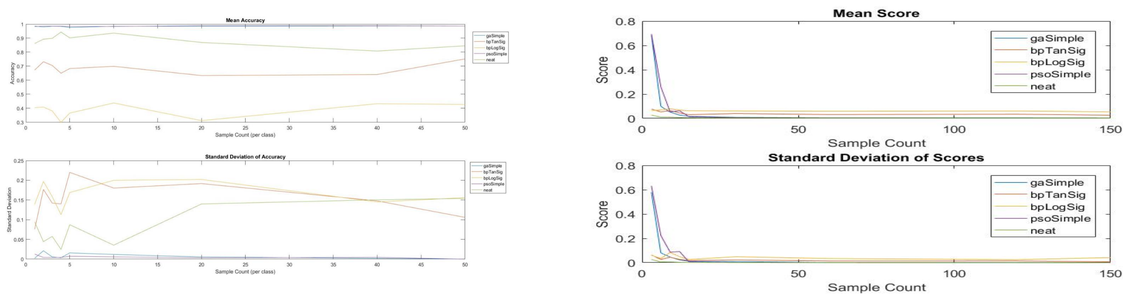
Classification/Regression Results

IRIS Dataset (Classification) — **Power Plant Dataset(Regression)**

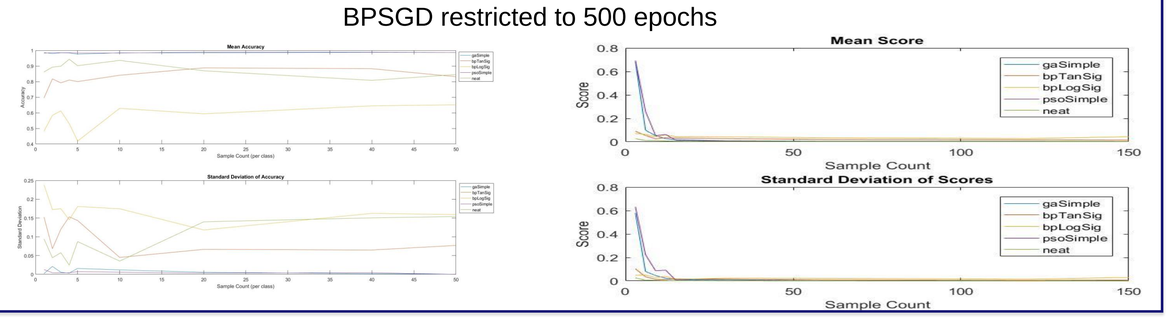
Complete Results (BPSGD allowed maximum of 50000 epochs)



BPSGD restricted to 200 epochs



BPSGD restricted to 500 epochs



V. Conclusions & Future work

Conclusions:

- Evolutionary algorithms are capable of finding better solutions compared to back propagation as shown in the above results.
- Results from XOR dataset show the limitations of Back Propagation in finding global minima.
- The results from IRIS and Power plant dataset show that EA's can perform better at low training sample size. PSO,GA in case of classification and NEAT in case of regression.
- Large amount of parallel nature of the algorithms make them an interesting prospect for future research.

Future Work:

- Benchmarking the performance of various configurations on high dimensional data (like MNIST) and other domains (like time series)
- Integrating parameter control into existing configurations.
- Porting the code to GPU's to take advantage of parallel processing capabilities of the algorithms.