Training Neural Networks with PSO and Other Population-Based Algorithms

Anna Rakitianskaia
annar@cs.up.ac.za

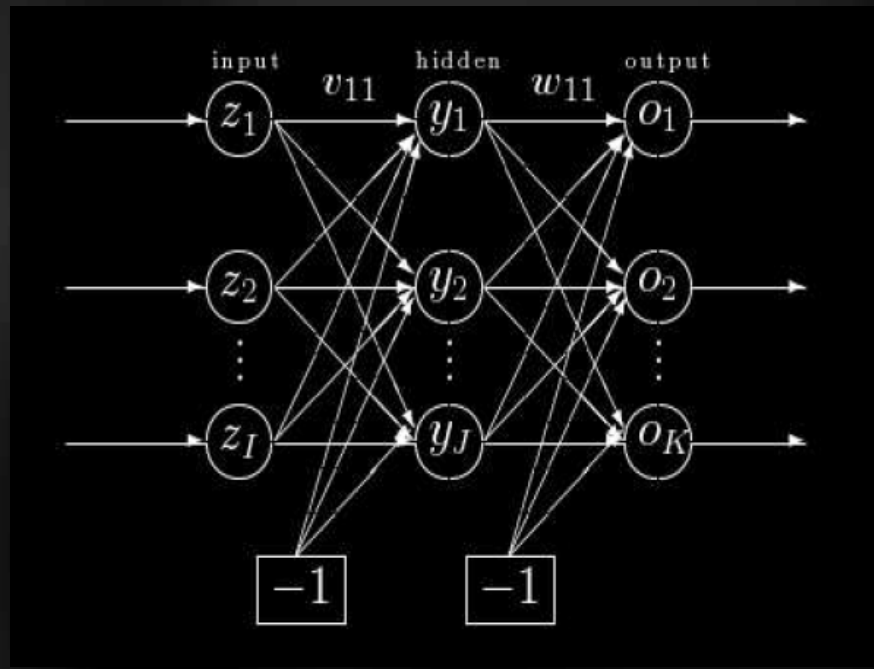
UNIVERSITY OF PRETORIA, SOUTH AFRICA

Outline

- Neural Networks
- Evolutionary Computation and PSO
- Classic vs. Alternative training
- Optimal Solution
- Example Applications
- Conclusion

Neural Network

- A set of interconnected neurons
- Universal Optimizer:
  - NN can represent any non-linear function provided there are enough units in the hidden layer
- NN training involves finding the optimal set of connection weights



Back Propagation

- Back Propagation of Error
  - Gradient descent based method that updates NN connection weights in order to minimize the resulting error
- It is essentially a hill-climbing algorithm
- Activation functions have to be differentiable

Alternatives?

- Population-based algorithms:
  - Evolutionary Algorithms (EA)
  - Particle Swarm Optimization (PSO)

Evolutionary Algorithms

- An evolving population of chromosomes
- Based on Darwinian principles
  - “Survival of the fittest”
  - Adoption of pseudo-biological operators such as selection, crossover and mutation
- Each chromosome represents a complete NN (weight vector)

Particle Swarm Optimization

- Models the social behavior of a bird flock

(1) $$\vec{v}_i(t) = \omega\vec{v}_i(t-1) + c_1\vec{r}_1(\vec{x}_{pbest}(t) - \vec{x}_i(t)) + c_2\vec{r}_2(\vec{x}_{gbest}(t) - \vec{x}_i(t))$$

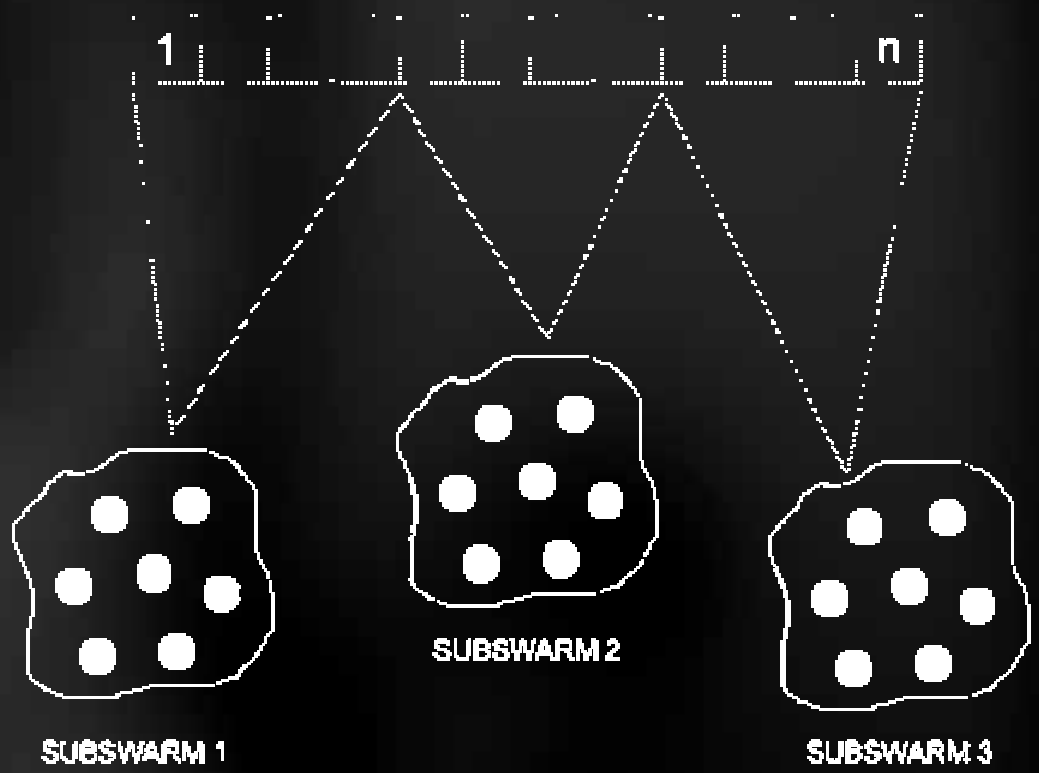
(2) $$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t)$$

Algorithm Outline

1. Initialize the swarm of N particles
2. For each particle *i* in [1, *N*], set *pbest* and *gbest*, and update particle velocity and position according to equations (1) and (2)
3. Repeat step 2 until a stopping condition is met

Cooperative PSO

- Split the search space dimension-wise into mutually exclusive subsets
- Assign a separate subswarm to each subset
  - Fitness function is undefined for the subspaces
  - How can the subswarm solutions be combined into one?



Cooperative PSO

Context Vector

- Maintains the complete solution vector used by the subswarms for particle fitness evaluation
- It is a combination of the solutions found by the subswarms

Algorithm Outline

1. Initialize K subswarms
2. For each subswarm:
   A. For each particle:
      i. Evaluate fitness using the context vector
      ii. Adjust velocity and position
      iii. Determine personal best
   B. Determine global best
3. Repeat step 2 until a stopping condition is met

Cooperative PSO

- Advantages
  - Reduced problem complexity
  - Has been shown to improve the performance of PSO on large-scale high-dimensional problems
  - Refined search
  - Easy to parallelize

Training NN with PSO

- Each particle represents a complete NN (weight vector)
- PSO and EA can also be used to optimize NN topology

Classic vs. Alternative

- Back Propagation:
  - Prone to premature convergence on local optima due to its hill-climbing nature
  - Can only work with differentiable activation functions
  - Sensitive to the starting point (initial weight values) and algorithm coefficients such as learning rate

Classic vs. Alternative

- Population based algorithms:
  - Stronger global search properties
  - Does not restrict activation function selection
  - Independent from the initial values, since multiple starting points are used

Classic vs. Alternative

- Drawbacks of the population based algorithms :
  - Slower convergence
  - Poor exploitation of good areas

EA vs. PSO

- EA
  - Originally designed for binary problem representations
  - Slow convergence

- PSO
  - Designed for continuous problem representations
  - Faster convergence

Empirical results have shown that PSO tends to outperform both EA and standard back propagation *(E. Massio Grimaldi, F. Grimaccia, M. Mussetta, R. E. Zich, "PSO as an Effective Learning Algorithm For Neural Network Applications," pp. 557-560, 2004 3rd International Conference on Computational Electromagnetics and Its Applications Proceedings)*

Optimal solution

- Start searching with a population based algorithm *(exploration)*
- Once a good area has been located, switch to back propagation *(exploitation)*
  - Restriction: activation functions have to be differentiable for back propagation to work

Example Applications

- PSO outperformed BP on a selection of problems
  - Diameter error prediction in boring machines
  - Prediction of pollutant levels in downtown area of Hong Kong
  - Training of product unit NNs
  - QSAR studies of inhibitors of platelet-derived growth factor receptor phosphorylation
  - Differential protection scheme for power transformers
  - Etc...

Conclusion

- Population-based algorithms are a viable alternative to back propagation
- PSO showed to be particularly valuable in cases where a high number of local minima is known to exist
- An optimal solution may be to start the search with a population-based algorithm and then refine the found solution with back propagation

Questions/Comments?

Thank you!