Particle Swarm Optimization in Machine Learning

Michał Okulewicz, Julian Zubek

Institute of Computer Science
Polish Academy of Sciences

Statistical Machine Learning
09 January 2014
Presentation Plan

1. Introduction
2. Application to training of MLP
3. Application to training of SNN
4. Application to clustering
5. Application to full model selection
General machine learning task

Machine learning algorithm

ML algorithm = family of models + model selection

Family of considered models is called hypothesis space.

Choosing the best model is an optimization problem.

Note: Some methods do not describe hypothesis function explicitly (e.g. kNN).
• Decision Tree
  Hypothesis space: all possible partitions by trees.
  Model selection: multistep greedy search optimizing Gini coefficient at each split.
Optimization in machine learning

- Linear regression
  Hypothesis space:
  \[ y = \beta_0 + \beta_1 x \]
  Model selection: ordinary least squares estimator (closed-form).

- Logistic regression
  Hypothesis space:
  \[ \pi(x) = \frac{1}{1 + \exp(- (\beta_0 + \beta_1 x))} \]
  Model selection: Newton’s method (iterative root finding).
Optimization in machine learning

• Multi-layered perceptron
  Hypothesis space:

  \[ y(x) = \varphi \left( \beta_0^0 + \beta_1^0 \left[ \varphi \left( \beta_0^1 + \beta_1^1 \left[ \varphi \left( \beta_0^2 + \beta_1^2 x \right), \ldots \right]^T \right), \ldots \right]^T \right) \]

Model selection: Backpropagation (gradient descent optimization).
MultiLayer Perceptron

Interpretation:
- Artificial Neural Network modeling a neural system.
- A stack of logistic regression models.

Applications:
- Classification,
- Regression.

Problems:
- Standard Backpropagation algorithm might get lost in local minima (restarts needed).
- Needs tuning of a learning rate.
- Backpropagation is unsuitable for more than 2 hidden layers.
Spiking Neural Network

Interpretation:

- Artificial Neural Network taking into account timing of inputs.
- A set of differential equations for computing membrane potential of a neuron.

Applications:

- Sequence (time-series) analysis.
- Pattern recognition.

Problems:

- Tuning of the parameters is not easy (STDP and ReSuMe algorithms train only the weights, but not the recovery time, increase time and initial potential of the neurons).
Selected types of clustering

- Similarity based clustering with defined $K$.
- Capacitated clustering (possibly with maximum $K$).
- Cost based clustering (possibly with maximum $K$ and limited cluster capacity).
When global stochastic search is feasible?

- Search space is very large.
- Objective function has multiple minima.
- Function gradient is unknown.
- Non-standard evaluation criteria.
- It is easy to overfit*.

* Tom Dietterich, 1995

“[In machine learning] it appears to be better not to optimize!”
Particle Swarm Optimization

- Continuous iterative global optimization metaheuristic algorithm.
- Utilizes the idea of Swarm Intelligence.
- Optimization is performed by a set of simple beings called particles.
- Each particle has current location, velocity and memory of the best visited location.
- Particles communicate their best visited location to the set of their neighbours.
Particle Swarm Optimization

- Initialize swarm
- Evaluate particles
- [STOP conditions met]
- [STOP conditions not met]
- Update velocity
- Update position
- Repeat...

Michał Okulewicz, Julian Zubek

PSO in ML
tth iteration for the ith particle:

\[ v_i^{(t+1)} = c_1 u_1^{(1)} (x_{n[i]}^{(best)} - x_i^{(t)}) + c_2 u_2^{(2)} (x_i^{(best)} - x_i^{(t)}) + \omega v_i^{(t)} \] (1)

\[ x_i^{(t+1)} = x_i^{(t)} + v_i^{(t)} \] (2)
tth iteration for the i th particle:

\[ g_i^{(t)} = \frac{1}{3} \left[ 3x_i^{(t)} + \left( c_1 \left( \frac{\text{best}}{n[i]} - x_i^{(t)} \right) \right) + \left( c_2 \left( x_i^{(best)} - x_i^{(t)} \right) \right) \right] \]  

(3)

\[ x_i^{(t+1)} \sim \text{unif} \mathcal{B}_i(g_i^{(t)}, \|x_i^{(t)} - g_i^{(t)}\|) \]  

(4)

\[ v_i^{(t+1)} = \omega v_i^{(t)} + x_i^{(t)} - x_i^{(t)} \]  

(5)
Simulation on Rastrigin’s function for SPSO 2007
Simulation on Rastrigin’s function for SPSO 2011
**MLP: PSO + SGD BP**

Task: minimize MSE.

**Introduction**

- Application to training of MLP
- Application to training of SNN
- Application to clustering
- Application to full model selection

**Conclusions**
Task: minimize AUC of absolute value of membrane potential of Similarity Measure Neuron.
K-means clustering

Task: minimize distance from cluster centers to points belonging to clusters.
DVRP: example of cost based clustering

Task: minimize total routes length.
Full model selection

Standard machine learning
Model selection:
  • Tuning parameters of a function of a given class.

Meta-learning
Full model selection:
  • Choosing preprocessing algorithm and its parameters.
  • Choosing feature selection strategy and its parameters.
  • Choosing machine learning algorithm and its parameters.
PSO in full model selection

Particle Swarm Model Selection (H. J. Escalante, M. Montes, E. Sucar, 2009):

- Implemented on top of Challenge Learning Object Package (CLOP) for MATLAB.
- Used in Agnostic Learning vs Prior Knowledge 2007 challenge (75 competitors):
  - 8th place overall,
  - 5th place among agnostic methods,
  - 2th place among methods utilizing only standard CLOP algorithms.
- 2-6 hours needed for each dataset from the competition.
Conclusions

- PSO (or possibly other metaheuristics) could be applied to ML models fitting and selection.
- In standard approach it does not usually beat well known specific training algorithms.
- It could be used for a mathematically non-trivial models (SNN) or with non-standard fitness functions (we could optimize easier for a business criteria).
Maurice Clerc.

Ioan Cristian and Trelea.
The particle swarm optimization algorithm: convergence analysis and parameter selection.

X. Cui, T.E. Potok, and P. Palathingal.
Document clustering using particle swarm optimization.

Hugo Jair Escalante, Manuel Montes, Luis Enrique Sucar, Isabelle Guyon, and Amir Saffari.
Particle swarm model selection.

Yuan-wei Jing, Tao Ren, and Yu-cheng Zhou.
Neural network training using pso algorithm in atm traffic control.
Jan Karwowski, Michał Okulewicz, and Jarosław Legierski.
Application of particle swarm optimization algorithm to neural network training process in the localization of the mobile terminal.

J. Kennedy and R. Eberhart.
Particle swarm optimization.

Hongbo Liu, Bo Li, Xiukun Wang, Ye Ji, and Yiyuan Tang.
Survival density particle swarm optimization for neural network training.

Michael Meissner, Michael Schmuker, and Gisbert Schneider.
Optimized Particle Swarm Optimization (OPSO) and its application to artificial neural network training.

Ammar Mohemmed, Satoshi Matsuda, Stefan Schliebs, Kshitij Dhoble, and Nikola Kasabov.
Optimization of spiking neural networks with dynamic synapses for spike sequence generation using pso.
Bibliography III

Ben Niu and Li Li.
A hybrid particle swarm optimization for feed-forward neural network training.

Michał Okulewicz and Jacek Mańdziuk.
Application of particle swarm optimization algorithm to dynamic vehicle routing problem.

Xiaorong Pu, Zhongjie Fang, and Yongguo Liu.
Multilayer perceptron networks training using particle swarm optimization with minimum velocity constraints.

Y. Shi and R.C. Eberhart.
A modified particle swarm optimizer.

Y. Shi and R.C. Eberhart.
Parameter selection in particle swarm optimization.