Efficient Hyperparameter Optimization of Deep Learning Algorithms Using Deterministic RBF Surrogates

ILIJA ILIEVSKI
PHD CANDIDATE
NATIONAL UNIVERSITY OF SINGAPORE
Deep neural networks are great but...

• They have **many** hyperparameters

• They are **very sensitive** to hyperparameter values

• Very **hard to guess** good hyperparameter values

• Solution: use **hyperparameter optimization algorithms**

• number of layers and neurons
• learning rate and momentum
• dropout rate
• weight initialization
• and many others...
Hyperparameter optimization is **not** easy

\[ E_{val} = F(x) \]

- One hyperparameter evaluation requires DNN training to convergence that can **take several hours**
- The hyperparameter space has **large number of local minima**
- Difficult **non-convex optimization in high dimensions**
Our approach: Use surrogate model...

- Approximate the expensive hyperparameter evaluation with a surrogate model:

\[ F(x) \approx S_n(x) = \sum_{i=1}^{n} \lambda(\|x - x_i\|)^3 + b^\top x + a \]

- Experiments show that Cubic Radial Basis Function surrogate with polynomial tail fits the highly non-convex and spiky hyperparameter space surprisingly well
and Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:

1. Rank the candidate points according to weighted average of the surrogate value and distance to the current best solution
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:

1. Rank the candidate points according to weighted average of the surrogate value and distance to the current best solution.
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:

1. **Rank** the candidate points according to weighted average of the surrogate value and distance to the current best solution

2. Perform **expensive evaluation of the highest ranked** hyperparameter values
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:

1. Rank the candidate points according to weighted average of the surrogate value and distance to the current best solution
2. Perform expensive evaluation of the highest ranked hyperparameter values
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:

1. Rank the candidate points according to weighted average of the surrogate value and distance to the current best solution

2. Perform expensive evaluation of the highest ranked hyperparameter values
Dynamic Coordinate Search

Explore the surrogate space for optimal hyperparameter values by evaluating candidate points around the current best found solution:

1. **Rank** the candidate points according to weighted average of the surrogate value and distance to the current best solution

2. Perform **expensive evaluation of the highest ranked** hyperparameter values
Results: Optimizing 19 CNN hyperparameters
Results: Optimizing 19 CNN hyperparameters
Thank you. Questions?


Ilija Ilievski, Taimoor Akhtar, Jiashi Feng, and Christine Annette Shoemaker

Paper available at:
https://ilija139.github.io

Code:
https://github.com/ilija139/HORD