Comparative Analysis of Curriculum Strategies in training Meta-Learning

Maria Mihai
author

Matthijs Spaan
responsible professor

Joery de Vries
supervisor

1. Background

Meta-Learning = distilling insights from previous learning experiences so that new tasks are learned more efficiently [1]
Curriculum-Learning = generating orders in which to present learning examples [2]
Self-paced Learning = creating curricula solely based on learning progress [2]
Focus of our research:

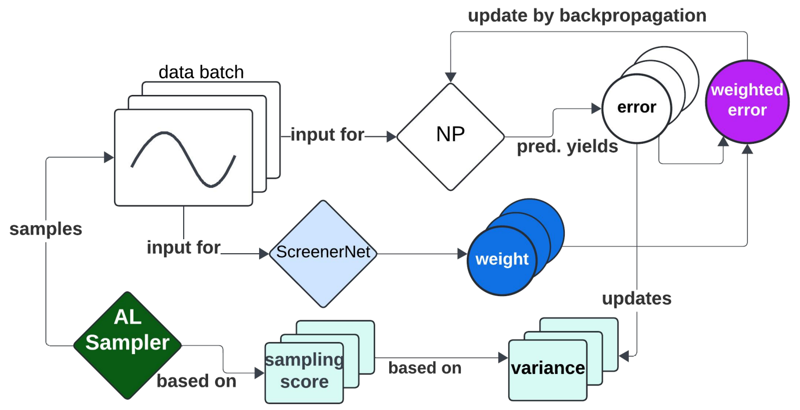
ScreenetNet [3]
- model-based approach to SPL
- targets the loss function
- predicts and emphasizes difficult tasks
- Expected effects:
  - increasing accuracy for difficult tasks
  - accelerating convergence

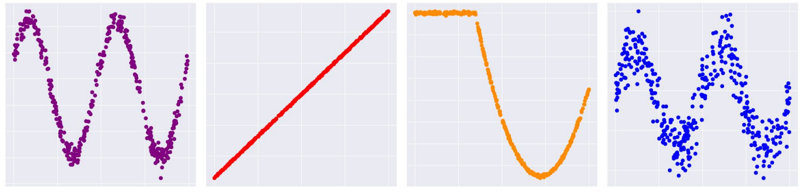
Active Bias [4]
- statistics-based approach to SPL
- targets the sampling of tasks
- estimates and emphasizes uncertain tasks
- Expected effects:
  - increasing noise robustness
  - increasing generalization performance

2. Research Questions

What are the separate and joint effects of ScreenerNet and Active Bias on the performance of Neural Processes being trained for solving 1-D function regression tasks?
Approached from 3 perspectives:
- Noise Robustness (accuracy after being trained with noisy data)
- Generalization (accuracy on task types different from those seen during training)
- Training Efficiency (convergence, training loss evolution).

3. Methodology



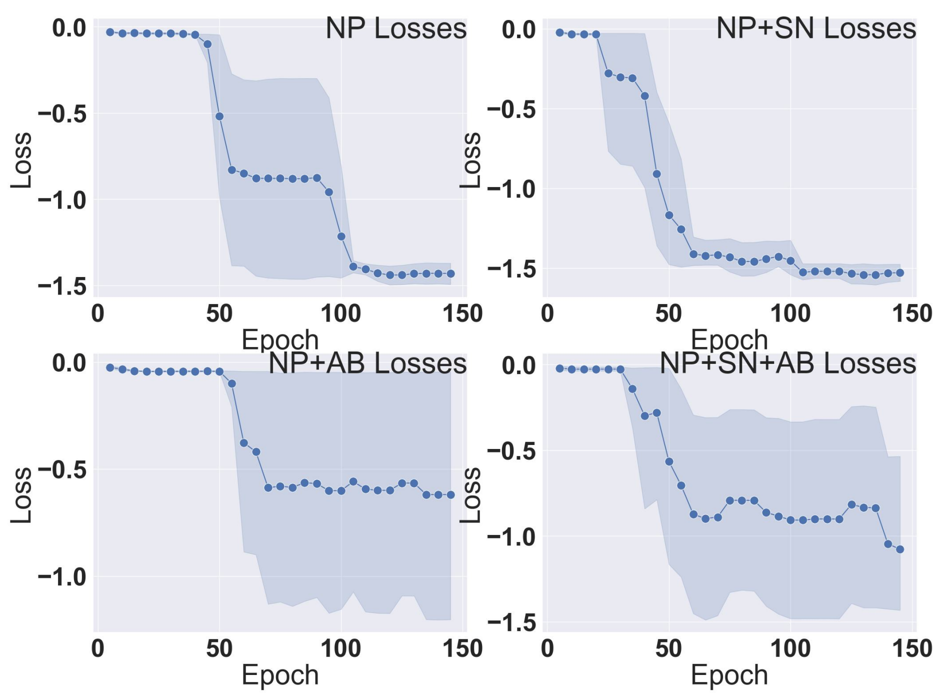


Models under comparison: NP, NP+SN, NP+AB, NP+SN+AB.
Training procedure:
- all models are trained and evaluated on the same data
- each model is trained 5 times
- after every 5 epochs, evaluate NLL loss on test set of same distribution
- two training data-sets: different noise levels (50%: 0.0 noise, 25%: 0.1 noise, 25%: 0.1 noise), and different function families (50%: sinus, 25%: quadratic, 25%: slope).

Testing procedure:
- testing each trained model on a single function type at a time, and aggregating the results for the 5 runs.
- experiments are separated into in-distribution (ID) - same task types as seen during training and out-of-distribution (OOD) - different task types; an informal measure of task similarity was used to generate 2 types of OOD experiments.

4. Results and Discussion

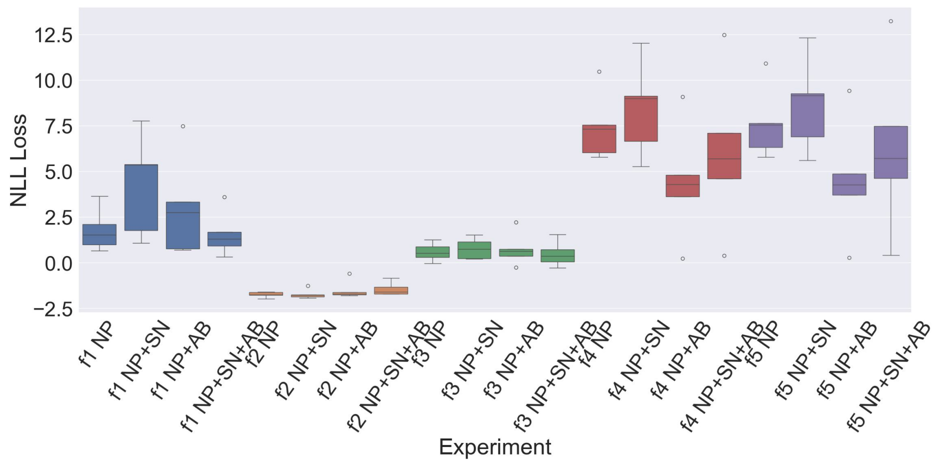
Noise Robustness



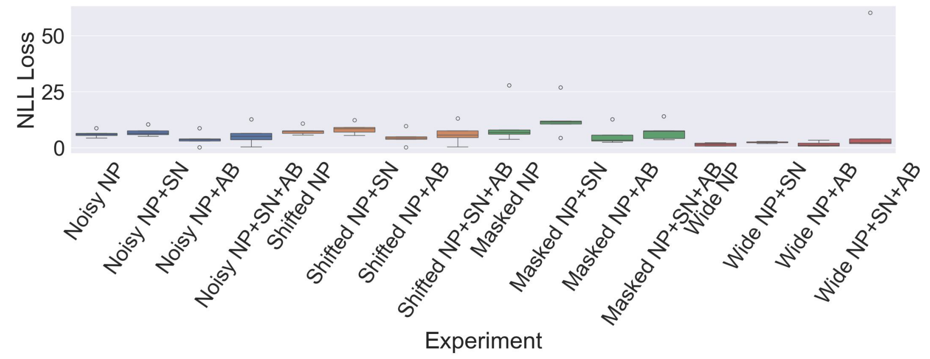
a. training on noisy dataset

- Noisy set-up impacted the training efficiency of NP+AB (for 3 out of the 5 runs, no evolution past the initial state).
- NP+SN showed a more stable evolution (lower loss variance over 5 runs).
- In NP+SN+AB, SN appears to partially mitigate AB's delayed convergence (happened in only 1 run).

Generalization



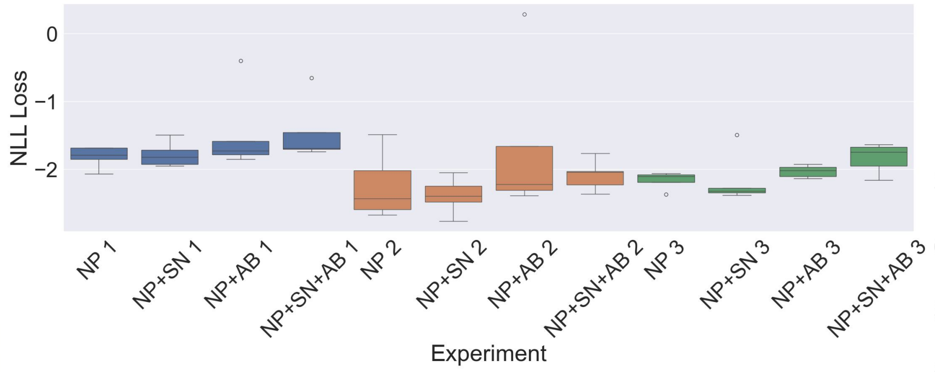
b. final accuracy for OOD but similar tasks



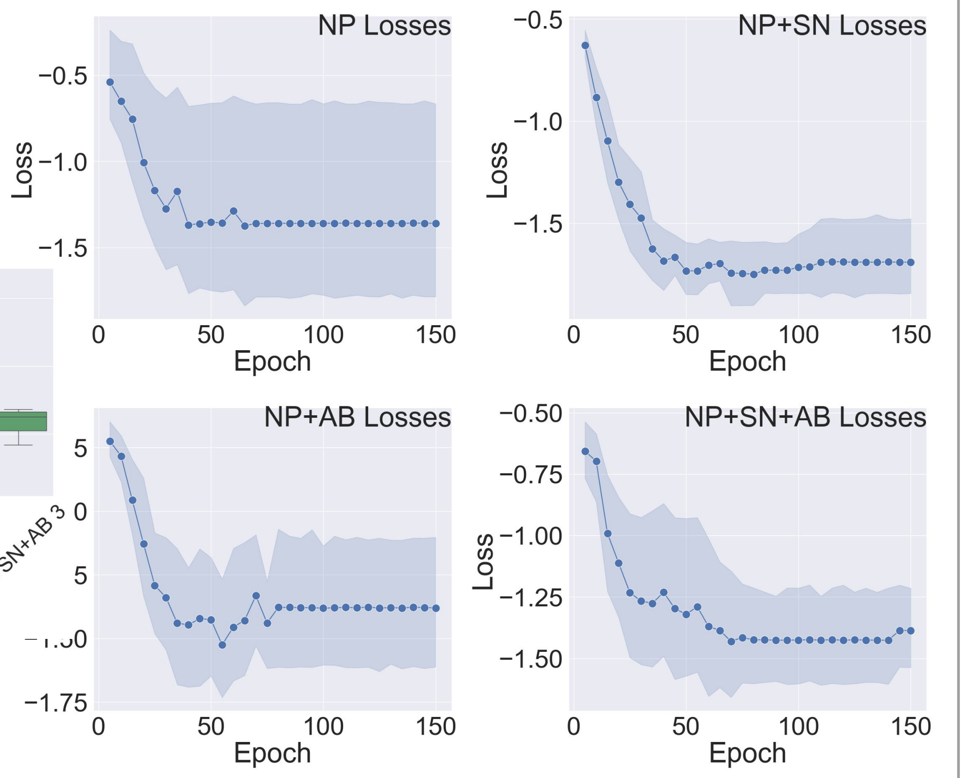
c. final accuracy for OOD but more different tasks

- NP+SN has lower accuracy than the baseline for OOD tasks;
- NP+AB achieves better accuracy for some unseen task types (but no identified pattern).
- NP+SN+AB has similar generalization performance to NP+AB, but large outliers.

Training Efficiency



d. final accuracy for ID tasks



e. training on clean dataset with different function types.

- NP+SN achieves better performance for the task types seen during training, and stabilizes the loss evolution (lower loss variance than the other models);
- NP+AB is out-performed by NP for ID task types and converges later than other models;
- NP+SN+AB has similar results to NP+AB for ID tasks but more stable loss evolution.

5. Conclusions

- ScreenerNet achieves more stable loss evolution and slightly better accuracy for seen task types, but lower accuracy for different task types than those seen during training.
- Active Bias outperforms the baseline in some generalization tasks, but the results are mixed. When trained on noisy data, training efficiency is affected.
- Combining ScreenerNet and Active Bias slightly increases ScreenerNet's generalization performance and stabilizes Active Bias' training evolution under noisy conditions.

6. Future work and limitations

- Investigating more complex architectures for ScreenerNet and the trade-off between prediction accuracy and training efficiency;
- Analyzing Active Bias' delayed convergence under noisy datasets;
- Investigating different ways of combining ScreenerNet and Active Bias

References

[1] Hospedales, Timothy, et al. "Meta-learning in neural networks: A survey." IEEE transactions on pattern analysis and machine intelligence 44.9 (2021): 5149-5169.
[2] Soviany, Petru, et al. "Curriculum learning: A survey." International Journal of Computer Vision 130.6 (2022): 1526-1565.
[3] Kim, Tae-Hoon, and Jonghyun Choi. "Screenernet: Learning self-paced curriculum for deep neural networks." arXiv preprint arXiv:1801.00904 (2018).
[4] Chang, Haw-Shiuan, Erik Learned-Miller, and Andrew McCallum. "Active bias: Training more accurate neural networks by emphasizing high variance samples." Advances in Neural Information Processing Systems 30 (2017).
[5] Garnelo, Marta, et al. "Neural processes." arXiv preprint arXiv:1807.01622 (2018).

M.T.Mihai-1@student.tudelft.nl

Research Project - CSE 3000

TU Delft Delft University of Technology