Universiteit
Leiden
The Netherlands

Understanding deep meta-learning

Huisman, M.

Citation

Huisman, M. (2024, January 17). *Understanding deep meta-learning*. *SIKS Dissertation Series*. Retrieved from https://hdl.handle.net/1887/3704815

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3704815

Note: To cite this publication please use the final published version (if applicable).

Understanding Deep Meta-Learning

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op woensdag 17 januari 2024

klokke 10:00 uur

door

Mike Huisman

geboren te Delft

in 1997

Promotor:

Prof.dr. A. Plaat

Co-promotor:

Dr. J. N. van Rijn

Promotiecommissie:

Prof.dr. H. C. M. Kleijn

Prof.dr. K. J. Batenburg

Prof.dr. E. O. Postma (Universiteit van Tilburg)

Dr. T. M. Moerland

Dr. S. Magliacane (Universiteit van Amsterdam)

This work was fully funded by Leiden University. SIKS Dissertation Series No. 2024-04. The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Copyright © 2023 Mike Huisman.

ISBN 978-94-93330-43-6

Contents

1 Introduction **7**
- 1.1 Background . . . 7
- 1.2 Motivation . . . 9
- 1.3 Research questions . . . 9
  - 1.3.1 Stateless neural meta-learning . . . 10
  - 1.3.2 Pre-training vs gradient-based meta-learning . . . 10
  - 1.3.3 LSTMs for few-shot learning . . . 11
  - 1.3.4 Subspace adaptation prior . . . 11
- 1.4 Outline of the dissertation . . . 11

2 Overview of Deep Meta-Learning **15**
- 2.1 Introduction . . . 15
- 2.2 Foundation . . . 17
  - 2.2.1 The meta abstraction . . . 17
  - 2.2.2 The meta-setup . . . 21
  - 2.2.3 The meta-learning field . . . 23
  - 2.2.4 Overview of the rest of this chapter . . . 25
- 2.3 Metric-based meta-learning . . . 25
  - 2.3.1 Example . . . 27
  - 2.3.2 Siamese neural networks . . . 28
  - 2.3.3 Matching networks . . . 29
  - 2.3.4 Prototypical networks . . . 30
  - 2.3.5 Relation networks . . . 32
  - 2.3.6 Graph neural networks . . . 33
  - 2.3.7 Attentive recurrent comparators . . . 33
  - 2.3.8 Metric-based techniques, in conclusion . . . 34
- 2.4 Model-based meta-learning . . . 35
  - 2.4.1 Example . . . 36
  - 2.4.2 Recurrent meta-learners . . . 36
  - 2.4.3 Memory-augmented neural networks (MANNs) . . . 37
  - 2.4.4 Meta networks . . . 39
  - 2.4.5 Simple neural attentive meta-learner (SNAIL) . . . 41
  - 2.4.6 Conditional neural processes (CNPs) . . . 42
  - 2.4.7 Neural statistician . . . 43
  - 2.4.8 Model-based techniques, in conclusion . . . 45

2.5 Optimization-based meta-learning . . . 45
2.5.1 Example . . . 46
2.5.2 LSTM optimizer . . . 46
2.5.3 LSTM meta-learner . . . 47
2.5.4 Reinforcement learning optimizer . . . 49
2.5.5 MAML . . . 50
2.5.6 iMAML . . . 52
2.5.7 Meta-SGD . . . 53
2.5.8 Reptile . . . 54
2.5.9 LEO . . . 56
2.5.10 Online MAML (FTML) . . . 57
2.5.11 LLAMA . . . 58
2.5.12 PLATIPUS . . . 60
2.5.13 Bayesian MAML (BMAML) . . . 61
2.5.14 Simple differentiable solvers . . . 62
2.5.15 Optimization-based techniques, in conclusion . . . 64
2.6 Concluding Remarks . . . 66
2.6.1 Overview . . . 66
2.6.2 Open challenges and future work . . . 67

3 Stateless Neural Meta-Learning using Second-Order Gradients 71
3.1 Introduction . . . 71
3.2 Related work . . . 72
3.3 Preliminaries . . . 73
3.3.1 MAML . . . 73
3.3.2 Meta-learner LSTM . . . 74
3.4 Towards stateless neural meta-learning . . . 75
3.4.1 Theoretical relationship . . . 75
3.4.2 Potential problems of the meta-learner LSTM . . . 76
3.4.3 TURTLE . . . 77
3.5 Experiments . . . 77
3.5.1 Hyperparameter analysis on sine-wave regression . . . 77
3.5.2 Few-shot sine wave regression . . . 78
3.5.3 Few-shot image classification . . . 78
3.5.4 Cross-domain few-shot learning . . . 80
3.5.5 Running time comparison . . . 80
3.6 Discussion and future work . . . 81

4 Understanding Gradient-Based Meta-Learning and Transfer Learning 85
4.1 Introduction . . . 86
4.2 Related work . . . 87
4.3 Background . . . 87
4.3.1 Supervised learning . . . 88
4.3.2 Few-shot learning . . . 88
4.3.3 Finetuning . . . 89
4.3.4 Reptile . . . 89
4.3.5 MAML . . . 90
4.4 A common framework and interpretation . . . 90

4.5 Experiments . . . 92
4.5.1 Toy problem . . . 94
4.5.2 Few-shot image classification . . . 95
4.6 Conclusion . . . 101

5 Are LSTMs Good Few-Shot Learners? 105
5.1 Introduction . . . 105
5.2 Related work . . . 107
5.3 Meta-learning with LSTM . . . 109
5.3.1 LSTM architecture . . . 109
5.3.2 Meta-learning with LSTM . . . 110
5.3.3 Problems with the classical LSTM architecture . . . 110
5.3.4 Towards an improved architecture . . . 111
5.4 Outer product LSTM (OP-LSTM) . . . 113
5.4.1 The architecture . . . 113
5.4.2 The algorithm . . . 115
5.5 Experiments . . . 116
5.5.1 Permutation invariance for the plain LSTM . . . 118
5.5.2 Performance comparison on few-shot sine wave regression . . . 119
5.5.3 Performance comparison on few-shot image classification . . . 120
5.5.4 Analysis of the learned weight updates . . . 121
5.6 Relation to other methods . . . 122
5.7 Conclusions . . . 124

6 Subspace Adaptation Prior for Few-Shot Learning 127
6.1 Introduction . . . 127
6.2 Related work . . . 128
6.3 Background on DARTS . . . 131
6.4 Subspace Adaptation Prior . . . 131
6.4.1 Intuition and operations . . . 132
6.4.2 The algorithm . . . 133
6.4.3 Analysis . . . 137
6.5 Experiments . . . 138
6.5.1 Sine wave regression . . . 138
6.5.2 The learned subspaces for sine regression . . . 139
6.5.3 Matching the problem structure . . . 139
6.5.4 Few-shot image classification . . . 140
6.5.5 Cross-domain few-shot image classification . . . 142
6.5.6 Effect of hard pruning . . . 143
6.5.7 The effect of the gradient order . . . 143
6.5.8 The learned subspaces for image classification . . . 144
6.5.9 Number of parameters and running time . . . 144
6.6 Conclusions . . . 145

7 Conclusions 149
7.1 Answers to research questions . . . 149
7.2 Future work . . . 152

Bibliography 164

A Hyperparameter details for TURTLE **165**
A.1 Used hyperparameters 165

B Additional experimental results for OP-LSTM **167**
B.1 Sine wave regression: additional results 167
B.2 Hyperparameter tuning 168
B.2.1 Permutation invariance experiments 168
B.2.2 Omniglot 168
B.2.3 miniImageNet and CUB 169
B.3 Robustness to random seeds 169
B.4 Within-domain 170
B.5 Cross-domain 171

C Additional experimental results for SAP **173**
C.1 Validation of re-implementation 173
C.2 Cross-domain few-shot image classification 174
C.3 The effect of hard pruning 174
C.4 The learned subspaces for image classification 174

Summary **177**

Dutch summary **181**

Acknowledgements **185**

Curriculum Vitae **187**

List of publications **189**

SIKS dissertation series **191**