Rapid Learning or Feature Reuse?
Towards Understanding the Effectiveness of MAML

Aniruddh Raghu*, Maithra Raghu*, Samy Bengio, Oriol Vinyals
Few Shot Learning

Many tasks, little data for each task

Task 1
Dog/Cat

Task 2
Chair/Lion

Task 3
Plane/Tree
Few Shot Learning

(Optimization-based) Meta Learning Algorithms

- Model Agnostic Meta Learning, (Finn et al), ICML 2017
Few Shot Learning

(Optimization-based) **Meta Learning Algorithms**

- Model Agnostic Meta Learning, *(Finn et al), ICML 2017*

**Outer Loop:** meta-initialization; **Inner Loop:** adaptation
Rapid Learning or Feature Reuse?

**Outer Loop**: meta-initialization; **Inner Loop**: adaptation

**Rapid Learning**

- θ₁
- θ₂
- θ₃

**Feature Reuse**

- θ₁
- θ₂
- θ₃

- Task 1
- Task 2
- Task 3

- **Outer loop**
- **Inner loop**
Rapid Learning or Feature Reuse?

How do hidden representations behave (during inner loop)?
Rapid Learning or Feature Reuse?

How do hidden representations behave (during inner loop)?

Measure Representation Similarity

SVCCA

https://github.com/google/svcca
Rapid Learning or Feature Reuse?

How do hidden representations behave (during inner loop)?

Measure Representation Similarity

Before inner loop

After inner loop
Rapid Learning or Feature Reuse?

SVCCA Before/After Inner Loop

SVCCA Similarity

Layer

Iteration 10000
Iteration 20000
Iteration 30000
Rapid Learning or Feature Reuse?

SVCCA Before/After Inner Loop

SVCCA Similarity

Layer

Very high similarity!
Rapid Learning or Feature Reuse?

SVCCA Before/After Inner Loop

Very high similarity!
ANIL: Almost No Inner Loop Algorithm
ANIL: Almost No Inner Loop Algorithm

- Removes inner loop for all but head of network
- Much more computationally efficient, same performance
- Insights into meta learning and few shot learning
## ANIL: Performance Results

Matches performance of MAML in few-shot classification and RL

<table>
<thead>
<tr>
<th>Method</th>
<th>Omniglot-20way-1shot</th>
<th>Omniglot-20way-5shot</th>
<th>MiniImageNet-5way-1shot</th>
<th>MiniImageNet-5way-5shot</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAML</td>
<td>93.7 ± 0.7</td>
<td>96.4 ± 0.1</td>
<td>46.9 ± 0.2</td>
<td>63.1 ± 0.4</td>
</tr>
<tr>
<td>ANIL</td>
<td>96.2 ± 0.5</td>
<td>98.0 ± 0.3</td>
<td>46.7 ± 0.4</td>
<td>61.5 ± 0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>HalfCheetah-Direction</th>
<th>HalfCheetah-Velocity</th>
<th>2D-Navigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAML</td>
<td>170.4 ± 21.0</td>
<td>-139.0 ± 18.9</td>
<td>-20.3 ± 3.2</td>
</tr>
<tr>
<td>ANIL</td>
<td>363.2 ± 14.8</td>
<td>-120.9 ± 6.3</td>
<td>-20.1 ± 2.3</td>
</tr>
</tbody>
</table>
ANIL and NIL (No Inner Loop)

NIL: No Inner Loop (at test time), performs equally well

<table>
<thead>
<tr>
<th>Method</th>
<th>Omniglot-20way-1shot</th>
<th>Omniglot-20way-5shot</th>
<th>MiniImageNet-5way-1shot</th>
<th>MiniImageNet-5way-5shot</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAML</td>
<td>93.7 ± 0.7</td>
<td>96.4 ± 0.1</td>
<td>46.9 ± 0.2</td>
<td>63.1 ± 0.4</td>
</tr>
<tr>
<td>ANIL</td>
<td>96.2 ± 0.5</td>
<td>98.0 ± 0.3</td>
<td>46.7 ± 0.4</td>
<td>61.5 ± 0.5</td>
</tr>
<tr>
<td>NIL</td>
<td>96.7 ± 0.3</td>
<td>98.0 ± 0.04</td>
<td>48.0 ± 0.7</td>
<td>62.2 ± 0.5</td>
</tr>
</tbody>
</table>
Thanks and Future Directions

http://learn2learn.net/tutorials/ani_tutorial/ANIL_tutorial/

- Exploring Medium Shot Learning?
- Meta-Learning as Pretraining?
- Learning Regimes for Interpolating between Rapid Learning and Feature Reuse?

Theoretical Analysis: Few-Shot Learning via Learning the Representation, Provably, (Du, Hu, Kakade, Lee, Lei)

Analyzing Feature Reuse: Rethinking Few-Shot Image Classification: a Good Embedding Is All You Need?, (Tian, Wang, Krishnan, Tenenbaum, Isola)