To learn intelligent, **general-purpose** behavior, learning each new skill from scratch isn’t going to cut it.

meta-learning  ←  learning priors, structure  
such that learning new tasks is fast
Can we learn a representation under which SL is fast and efficient?

Given 1 example of 5 classes:

Classify new examples

training data

test datapoints

meta-training

\( \mathcal{T}_1 \)

\( \mathcal{T}_2 \)

training classes

\( \vdots \)

\( \vdots \)

diagram adapted from Ravi & Larochelle ‘17
Can we learn a representation under which RL is fast and efficient?

Can we learn a representation under which imitation is fast and efficient?

Can we learn a representation under which **human imitation** is fast and efficient?

input human demo

resulting policy

Yu*, Finn*, Dasari, Zhang, Abbeel, Levine. One-Shot Imitation from Observing Humans via Domain Adaptive Meta-Learning. RSS ’18
Where do the tasks come from?
Meta-learning: manual algorithm design —> manual task distribution design

Are we simply kicking the can down the road?
Where do the tasks come from?

**self-driven**: propose your own tasks

**environment-driven**: dynamic, real-world environment
Meta-learn with only **unlabeled** images?

Construct tasks without labeled data?

Unsupervised learning (to get an embedding space) → Propose tasks → Run meta-learning

\[
\mathcal{T}_1 \quad \mathcal{T}_2
\]

\[
\{\mathbf{x}_i\}
\]

\[
\text{each image: point in } \mathbb{R}^n
\]

Result: representation suitable for learning downstream tasks

Hsu, Levine, Finn. Unsupervised Learning via Meta-Learning. ICLR’19
Can we meta-learn with only unlabeled images?

Unsupervised learning (to get an embedding space) → Propose tasks → Run meta-learning

A few options:
- BiGAN — Donahue et al. ’17
- DeepCluster — Caron et al. ’18
- Clustering to Automatically Construct Tasks for Unsupervised Meta-Learning (CACTUs)

MAML — Finn et al. ’17
ProtoNets — Snell et al. ’17

<table>
<thead>
<tr>
<th>method</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAML with labels</td>
<td>62.13%</td>
</tr>
<tr>
<td>BiGAN kNN</td>
<td>31.10%</td>
</tr>
<tr>
<td>BiGAN logistic</td>
<td>33.91%</td>
</tr>
<tr>
<td>BiGAN MLP + dropout</td>
<td>29.06%</td>
</tr>
<tr>
<td>BiGAN cluster matching</td>
<td>29.49%</td>
</tr>
<tr>
<td>BiGAN CACTUs MAML</td>
<td>51.28%</td>
</tr>
<tr>
<td>DeepCluster CACTUs MAML</td>
<td><strong>53.97%</strong></td>
</tr>
</tbody>
</table>

Same story for:
- 4 different embedding methods
- 4 datasets (Omniglot, CelebA, miniImageNet, MNIST)
- 2 meta-learning methods (*)
- Test tasks with larger datasets

*ProtoNets underperforms in some cases.

Hsu, Levine, Finn. Unsupervised Learning via Meta-Learning. ICLR ’19
What about unsupervised meta-RL?

Environment → Propose tasks → Run meta-RL

Result: Environment-specific RL algorithm
What about unsupervised meta-RL?

Environment $\rightarrow$ Propose tasks $\rightarrow$ Run meta-RL

Result: Environment-specific RL algorithm

- Discrete set of random tasks

\[ R(s, z) = \log p_D(z|s) \]
What about unsupervised meta-RL?

Environment $\rightarrow$ **Propose tasks** $\rightarrow$ Run meta-RL

**Result:** Environment-specific RL algorithm

- Discrete set of random tasks
- Learned diverse set of tasks by pushing apart discriminator classes

Environment $\rightarrow$ **Propose tasks** $\rightarrow$ Run meta-RL

- **Policy** $\rightarrow$ visit states that are discriminable
- **Discriminator** $\rightarrow$ predict skill from state

$$R(s, z) = \log p_D(z | s)$$

Eysenbach, Gupta, Ibarz, Levine. Diversity is All You Need. ICLR ‘19
Does it work?

- compare **UML-Random**, **UML-DIAYN**, **RL from scratch**
- measure learning performance on **test tasks with rewards**

2D Navigation  
Cheetah  
Ant

**Takeaway**: Relatively **simple** mechanisms for proposing tasks work surprisingly well.

Where do the tasks come from?

**self-driven**: propose your own tasks

**environment-driven**: dynamic, real-world environment
Deriving tasks from dynamic, real-world environments

- Motor malfunction
- Gradual terrain change

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Environments through Meta-RL. ICLR’19
Deriving tasks from dynamic, real-world environments

online adaptation = few-shot learning  

motor malfunction

gradual terrain change

$s_{t-k:t}, a_{t-k:t}$

tasks are temporal slices of experience

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Environments through Meta-RL. ICLR ‘19
one step of gradient descent
VelociRoACH Robot

Meta-train on variable terrains

Meta-test with slope, missing leg, payload, calibration errors
VelociRoACH Robot

Meta-train on variable terrains  Meta-test with slope, missing leg, payload, calibration errors

model-based RL  with MAML (ours)
(no adaptation)
VelociRoACH Robot

Meta-train on variable terrains  Meta-test with slope, **missing leg**, payload, calibration errors

model-based RL (no adaptation)  with MAML (ours)

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Environments through Meta-RL. ICLR’19
Where do the tasks come from?

**self-driven**: propose your own tasks

**environment-driven**: dynamic, real-world environment

Future work: combine both.
Self-driven + environment-driven tasks

- Gradual terrain change
- Motor malfunction

- Crawl towards parent with food
- Eat something
- Play with toy car

- Drive to work
- Help a co-worker
- Answer emails
- Obtain lunch

**sufficiently complex environment** + **basic drives** → **realistic & complex task distribution**

*e.g., the real world*

*survival, belonging, social interaction, curiosity, fulfillment*
Open Problems in Meta-Learning

**Data**

Where do the tasks come from?
(Defining one reward function is hard enough!)

Do we need all tasks to be available at once?
(i.e. moving away from iid sampling from task distribution)

In practice, data is available **incrementally**.
Can we learn priors when *only a few tasks are present*?

**Preliminary work:** *Online Meta-Learning*, Finn*, Rajeswaran*, Kakade, Levine. ICML ’19

**Algorithms**
Open Problems in Meta-Learning

**Data**

Where do the tasks come from?
(Defining one reward function is hard enough!)

Do we need all tasks to be available at once?
(i.e. moving away from iid sampling from task distribution)

**Algorithms**

How to reason over uncertainty w.r.t. prior?
(Bayesian DL is hard)
Ambiguity in Few-Shot Learning

Can we learn to generate hypotheses about the underlying function?

Existing meta-learning algorithms return the MAP, $\arg\max_{\phi_i} p(\phi_i | x_i^{\text{train}}, y_i^{\text{train}})$

Can we sample $\phi_i \sim p(\phi_i | x_i^{\text{train}}, y_i^{\text{train}})$?

- learning to actively learn
- learning to explore in meta-RL
- trading off weight of prior & posterior in variable-shot learning

Important for:

Gordon*, Bronskill*, et al. Meta-Learning Probabilistic Inference for Prediction. ICLR’19

Open Problems in Meta-Learning

**Data**

Where do the tasks come from?
(Defining one reward function is hard enough!)
Do we need all tasks to be available at once?
(i.e. moving away from iid sampling from task distribution)

**Algorithms**

How to reason over uncertainty w.r.t. prior?
(Bayesian DL is hard)
How to learn from off-policy data in meta-RL?
(fundamental challenges to making it off-policy)
Meta-training:
Learn from off-policy data

Meta-test time:
Iteratively explore, collect data, update policy

We have a mismatch…

Proposed Solution: *pseudo on-policy data* for adaptation, *off-policy data* for meta-optimization

*Thompson sampling* from learned latent task variable to explore
Open Problems in Meta-Learning

Data

Where do the tasks come from?
(defining one reward function is hard enough!)

Do we need all tasks to be available at once?
(i.e. moving away from iid sampling from task distribution)

How to reason over uncertainty w.r.t. prior?
(Bayesian DL is hard)

Algorithms

How to learn from off-policy data in meta-RL?
(fundamental challenges to making it off-policy)

realistic & complex task distributions + powerful meta-learning algs
recover effective priors
over intelligent behavior
Collaborators

Sergey Levine  Pieter Abbeel  Tianhe Yu  Anusha Nagabandi  Ignasi Clavera  Kate Rakelly  Aurick Zhou

Annie Xie  Sudeep Dasari  Simin Liu  Kyle Hsu  Abhishek Gupta  Ben Eysenbach  Kelvin Xu

Papers, data and code linked at: people.eecs.berkeley.edu/~cbfinn

Questions?