Building Versatile Agents through Unsupervised Interaction

Chelsea Finn

UC Berkeley  Google Brain  Stanford
Impressive Feats in AI

Why are these impressive?
They perform a complex task very well, sometimes even better than a human.

What is equally important: (but less impressive)
Generality & versatility.

How can we build generalists?
It turns out — the **simpler**, but **broader** capabilities are **really hard**. (Moravec’s Paradox)

This talk: can we do the **unimpressive** things?
**Versatility**: Can we leverage prior experience to quickly learn new concepts and adapt to new environments?

**Generality**: Can we learn from diverse experiences to generalize to many objects and goals?

Further, can we do this from raw sensory data with minimal supervision?

Grounded in robotics applications, but the underlying techniques are more general.
Generality: Can we learn from diverse experiences to generalize to many objects and goals?

Versatility: Can we leverage prior experience to quickly learn new concepts and adapt to new environments?

Further, can we do this from raw sensory data with minimal supervision?

Grounded in robotics applications, but the underlying techniques are more general.
Few-Shot Learning Example: Classification

Given 1 example of 5 classes:

- training data $D_{train}$
- test set $X_{test}$

Classify new examples

Diagram adapted from Ravi & Larochelle ’17
The Meta-Learning Problem

Meta-Supervised Learning:

Inputs: $D_{\text{train}}, x_{\text{test}}$

Outputs: $y_{\text{test}}$

Data: $\{D_i\}$

$D_i : \{(x, y)_j\}$

$y_{\text{test}} = f(D_{\text{train}}, x_{\text{test}}; \theta)$
Design of $f$?

- Hyp. space + Prior → Size principle → Examples

Recurrent network (LSTM, NTM, Conv)

$y_{test} = f(D_{train}, x_{test}; \theta)$

Santoro et al. ‘16, Duan et al. ‘17, Wang et al. ‘17, Munkhdalai & Yu ‘17, Mishra et al. ‘17, …

+ expressive, general
+ applicable to range of problems
- complex model for complex task of learning
- often large data requirements

and many many more approaches

Tenenbaum ‘99
Fei-Fei et al. ‘05
Lake et al. ‘11
Santoro et al. ‘16
Ravi & Larochelle ‘17
Andrychowicz et al. ‘16
Li & Malik ‘16
Vinyals et al. ‘16
Snell et al. ‘17
Hochreiter et al. ‘01
Snell et al. ‘17
Vinyals et al. ‘16
Hochreiter et al. ‘01
Andrychowicz et al. ‘16
Li & Malik ‘16
Santoro et al. ‘16
Ravi & Larochelle ‘17
and many many more approaches
Learning Few-Shot Adaptation

Fine-tuning

\[ \text{Fine-tuning} \quad \theta \leftarrow \theta - \alpha \nabla_\theta L_{\text{train}}(\theta) \]

Our method

\[
\min_\theta \sum_{\text{task } i} L_{\text{test}}^i (\theta - \alpha \nabla_\theta L_{\text{train}}^i(\theta))
\]

Key idea: Train over many tasks, to learn parameter vector \( \theta \) that transfers

Model-Agnostic Meta-Learning

Finn, Abbeel, Levine ICML '17
To give some intuition…

Finn, Abbeel, Levine ICML ’17
Can we learn a representation under which RL is fast and efficient?
MinImagenet Few-shot Classification

...and the results keep getting better

MinImagenet few-shot benchmark: 5-shot 5-way

Finn et al. ’17: 63.11%
Li et al. ’17: 64.03%
Kim et al. ’18 (AutoMeta): 76.29%

Program Synthesis

Learning to Learn Distributions

Federated Learning

Multi-Agent Competitions

Hierarchical Bayes Connection

Bayesian/Probabilistic MAML
**Versatility:** Can we leverage prior experience to quickly learn new concepts and adapt to new environments?

**Generality:** Can we learn from diverse experiences to generalize to many objects and goals?

Further, can we do this from raw sensory data with minimal supervision?
meta-training

\[ \mathcal{T}_1 \]

\[ \mathcal{T}_2 \]

Requires tasks constructed from labeled data

Requires demos for many previous tasks

Requires many tasks with corresponding reward functions
**Generality**: Can we learn from diverse experiences to generalize to many objects and goals?

**Versatility**: Can we leverage prior experience to quickly learn new concepts and **adapt to new environments**?

Further, can we do this from **raw sensory data** with **minimal supervision**?
Model-based RL

Collect data

Fit model

Plan using the model

Model-based RL achieves **zero-shot generalization** to new rewards.

But struggles when facing **new environments**.

Can we learn to adapt model quickly to new environments?
Can we learn to adapt model quickly to new environments?

gradual terrain change

motor malfunction

Anusha Nagabandi, Ignasi Clavera, Simin Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Real-World Environments through Meta-RL
Can we learn to adapt model quickly to new environments?

Online adaptation = few-shot learning

Tasks are temporal slices of experience

The task distribution is more-or-less free!

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Real-World Environments through Meta-RL
Meta-learning using either:
Recurrent Model (ReBAL)
MAML (GrBAL)

ONLINE ADAPTATION

Store recent history
\( \{s_{t-M:t}, a_{t-M:t}\} \)

Take action
\( a_t \) via
MPC

Model Adaptation
\( \theta'_* = u_{\psi}(s_{t-M:t}, a_{t-M:t}, \theta_*) \)

Meta-train a prior \( \theta_* \)
Dynamic Environments without Adaptation

Model-Based RL Only

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Real-World Environments through Meta-RL
Dynamic Environments with Online Adaptation with GrBAL

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Real-World Environments through Meta-RL
VelociRoACH Robot

Meta-train on variable terrains

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

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Real-World Environments through Meta-RL
Roach Robot

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

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Real-World Environments through Meta-RL
Roach Robot

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

GrBAL (ours)  Model-Based RL (no adaptation)

Nagabandi*, Clavera*, Liu, Fearing, Abbeel, Levine, Finn. Learning to Adapt in Dynamic Real-World Environments through Meta-RL
Roach Robot

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

GrBAL (ours)

Model-Based RL
(no adaptation)
gradual terrain change

motor malfunction

time
motor malfunction

gradual terrain change

iced terrain

\[ s_{t-k:t}, a_{t-k:t} \]

\( k \) time steps not sufficient to learn entirely new terrain

Continue to run gradient descent?

Nagabandi, Finn, Levine. Deep Online Learning via Meta-Learning
Online inference problem: infer latent “task” variable at each time step

Mixture of neural networks over task variable $T$, adapted continually: $\theta_t(T_i)$

Alternate between:

E-step: Estimate latent “task” variable at each time step $P(T_t)$ given data $x_t, y_t$

$$P(T_t = T_i|x_t, y_t) \propto p_{\theta(T_i)}(y_t|x_t, T_t = T_i)P(T_t = T_i)$$

likelihood of the data

prior

under task $T_i$.

M-step: Update mixture of network parameters

$$\theta_{t+1}(T_i) = \theta_t(T_i) - \beta P(T_t = T_i|x_t, y_t)\nabla_{\theta_t(T_i)} \log p_{\theta_t(T_i)}(y_t|x_t) \quad \forall T_i$$

gradient step on each mixture element, weighted by task probability

Note: Online learning with neural nets won’t work in the general case.
Does it work?

Crawler with crippled legs

- ours
- always take grad steps
- GrBAL (always reset to prior + 1 grad step)
- model-based, no adaptation
- model-based, grad steps

\{no meta-learning\}

Nagabandi, Finn, Levine. Deep Online Learning via Meta-Learning
Does it work?

Crawler with crippled legs

Next step:

Latent task distribution during online learning

Nagabandi, Finn, Levine. Deep Online Learning via Meta-Learning
Generality: Can we learn from diverse experiences to generalize to many objects and goals?

Versatility: Can we leverage prior experience to quickly learn new concepts and adapt to new environments?

Further, can we do this from raw sensory data with minimal supervision?
Learning Generalizable Models through Self-Supervision

**self-supervised robot learning**

Pinto & Gupta ’16

Levine, Pastor, Krizhevsky, Quillen ’16

**model-based control**

Nair*, Chen*, Agrawal*, Isola, Abbeel, Malik, Levine ’17

Arruda, Mathew, Kopicki, Mistry, Azad, Wyatt ’17

Petrovskaya, Park, Khatib ’07

Yu, Bauza, Fazeli, Rodriguez ’17

Our approach

**acquire a general-purpose model**

many objects, raw perceptual inputs, intuitive physics

**learn model from video data**
Collect **diverse** data in a **scalable** way

What can we learn from raw sensory data without notions of progress or success?
Learn to predict

$I_t, a_{t:t+H} \rightarrow I_{t:t+H}$

Contrast to:

Models capture **general purpose** knowledge about the world

Use **all** of the available supervision signal.

Also: No assumptions about task **representations**.
Are these models useful?

Plan to achieve human-specified goals
Planning with Visual Foresight

1. Consider potential action sequences
2. Predict the future for each action sequence
3. Pick best future & execute corresponding action
4. Repeat 1-3 to replan in real time

visual “model-predictive control” (MPC)

Overall System: Collect data, Train predictive model, Plan to achieve goals
How to predict video?

- deep recurrent network
- multi-frame prediction
- action-conditioned
- explicitly model motion

Finn, Goodfellow, Levine NIPS ’16
Finn & Levine ICRA’17
Which future is the best one?

Human specifies a goal by:

- Selecting where pixels should move.
- Providing an image of the goal.
- Providing a few examples of success.

Finn & Levine ICRA ’17
Ebert, Lee, Levine, Finn CoRL ’18
Xie, Singh, Levine, Finn CoRL ’18
How it works

Specify goal

Visual MPC execution

Visual MPC w.r.t. goal

How it works

Specify goal
(covering an object)

Visual MPC execution
How it works

Given 5 examples of success

Visual MPC with learned objective

infer goal classifier

visual MPC w.r.t. goal classifier

Xie, Singh, Levine, Finn. Few-Shot Goal Inference for Visuomotor Learning and Planning
Planning with a **single model** for many tasks
Demo at NIPS 2017: Long Beach, CA

planning with visual models

The students were unimpressed.
(but still had fun)

Demo at AI4ALL Outreach Camp
Takeaways

**Versatility:** Learn to adapt online to new drastically new conditions.

**Generality:** Learn from diverse experiences to generalize to many objects and goals.

With *minimal supervision* and *raw observations*.

What’s next?
long horizon reasoning

ability to convey and understand complex goals
“be gentle”, “cook dinner”, “try to console him”

Through composition of a general skill-set

And structured representations

Yu, Abbeel, Levine, Finn arXiv ‘18

Janner, Levine, Freeman, Tenenbaum, Finn, Wu ’18
long horizon reasoning
through composition of general skill-set
and structured representations

ability to **convey** and
**understand complex goals**
“be gentle”, “cook dinner”,
“try to console him”

continual learning

one step of adaptation
continual learning and adaptation

ultimately: **continual self-supervised learning** in the **real world**
Questions?

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