Jane Wang (DeepMind)

Eda Momennejad (Columbia)

Jacqueline Gottlieb (Columbia)

NeurIPS 2019

– Day 5 –

WORKSHOPS I

David Abel (Duke)

Raia Hadsell (DeepMind)

Rich Sutton (U Alberta)

Jane Wang (DeepMind): From Brains to Agents & Back

NEUROSCIENCE ⇢ **AGENT-BASED** ⇠ **MACHINE LEARNING**

- Discovery
- Neuron-Centred
- Behaviourism

Better Models (CCNNs, Data Analysis)

Point at inaccuracies

Engineering

Symbolic / NNs

Control

TWO TYPES OF LEARNING

1. INNATE BEHAVIOURS → Slow
2. LEARNED BEHAVIOURS → Fast

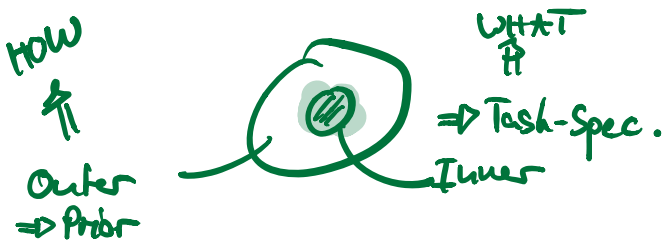
↓

BALDWINIAN EVOLUTION

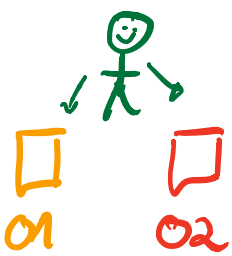
↳ Learning traits ⇒ Test on!

↓

META-LEARNING



HARLOW TASK (Wang et al. '18)



- Learn task structure
  - ↳ SLOW
- Fast adaptation

↳ ImpalaNet RL Experiments

⇒ Need many different images

⇒ GENERALIZATION BETTER W. MORE TRAINING TASKS

↳ PRIOR MISMATCH

COOKIE-CUTTER ANALOGY

- * How to learn?
- * What to learn?

→ Zador '19': Place Fields

Propensity innate ⇔ Content learned

META-LEARNING IN RL

- Memory-based agent (LSTM)
- Task distribution → Overlap

	BIOLOGICAL	ARTIFICIAL
INNATE	Ecological Evolved	Learned Struct. Priors
LEARNED	Lab Train	Env Train

UNDERSTANDING DYNAMIC

HIDDEN STATE DYNAMICS

Convergence to Task Repr. Clusters!

Jacqueline Gottlieb (Columbia): Info Demand - Simple & Complex

WHY WE CARE ⇒ Behaviorism / Cognitivism drastically simplify

- Signal Detection / RL
- Saliency Exp. Setup
  - assumes agent knows relevant signals, action space, rewards

INFO PROCESSING

ATTENTIONAL CONTROL = FRONTAL LOBE HOMUNCULUS
- classically modele w/o input!
- ⇓ CLOSE THE LOOP!

TOWARDS A NEW CYBERNETIC PARADIGM

INTRINSIC MOTIVATION

SELF-ORGANISATION: LEARNING W/O REWARDS

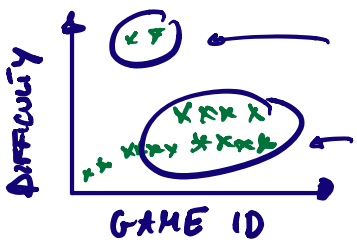
HUMAN INFO DEMAND → Complex setting ⇒ What to investigate?

→ Gottlieb & Oudeyer: LEARNING PROGRESS
↳ Reducible uncertainty

① Familiar items → Def Room
① Novel/Surprising → Trap

→ Optimal Student Paradigm: Many complications ⇒ Realisation of learning through uncertainty ⇓

→ Baranes et al. 14': 'Guitar Hero'



① Easy-2-Difficult Progression!
① Settling on Intermediate!

⇒ BEST LEARNERS CHOOSE INTERMEDIATE SELF-CHALLENGE LEVEL!

① RL model → Softmax Task Selection
↳ Selection sensitivity best model fit
⇒ Combines reward measure + progress!

→ Endogenous Task Selection: 2-TFC

Difficulty
① Variable Dims.
② Relevant

- Participants choose Task
  - Free to play
  - Max. Learning

TASK SELECTION DIFFERENCES

→ Learn Max. group challenges itself more, Free group more chill

↳ Individual Variation: Some participants in free group are highly motivated to explore task space

Ida Momennejad (Columbia): Predictive Cognitive Maps

HOW DO BRAINS BUILD COGNITIVE MAPS FOR LEARN. / PLANNING?

↳ Tolman: Cognitive Maps in Hippocampus → Behrens et al. 18'

MULTI-SCALE ⊕ MULTI-STEP

KIND OF REPRESENTATIONS → HOW TO LEARN?

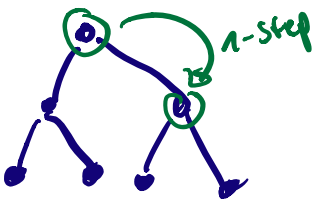
1-Step Transitions → MB Learner

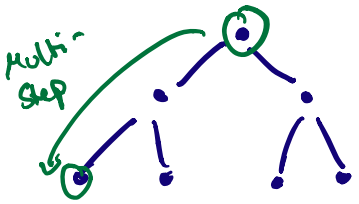
Successor Repr. (Dayan, 1993) → $V(s) = \sum_{s'} M(s,s') R(s')$

$$M_{\pi,\gamma}(s,s') = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t \mathbb{1}(s_t = s') \mid s_0 = s\right]$$

$$\hat{M}_{t+1}(s_t, s') \leftarrow \hat{M}_t(s_t, s') + \eta\left[\mathbb{1}(s_t = s') + \gamma \hat{M}_t(s_{t+1}, s') - \hat{M}_t(s_t, s')\right]$$

SUCCESSOR PRED. ERROR





⊙ Kim Stachenfeld Studies → Place & Grid Cells

⊙ Momennejad* et al. 17' ⇒ Reward vs. Transition Revaluation
⊕ Russek*

MB SR MB SR

↳ SUCCESSOR REPRESENTATION DYN → Best at explaining human ⊕ rodent behaviour!
↳ RSA analysis: Hippocampus, anterior PFC ↳ also for policy reveal.

↳ EPILEPSY SINGLE-CELL RECORDINGS → Place Cells ⇒ Room 2 Room Sw.

↳ LARGE-SCALE TORONTO MAP NAVIGATION → Levels of abstraction
↳ Familiar vs. GPS condition

POSTERIOR ← HIPPOCAMPUS → ANTERIOR

Posterior → Fine-Grained

Hippocampus → γ → 0.9 0.6 0.1 → "Hippocampal Gradients"

Anterior → Coarse

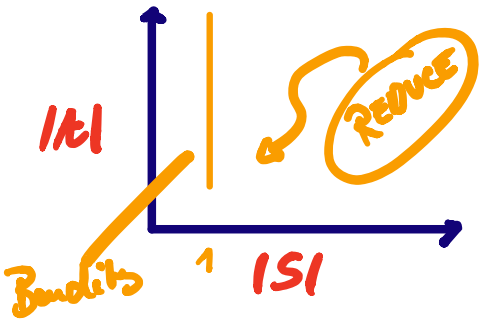
↳ DYN VALIDATION

Surprise ⇒ Replay + Replan
Momennejad et al. 18'
↳ prioritization ↑

⇒ PREDICTIVE HORIZON CONTROLLED BY DISCOUNT

David Abel (Duke): Abstraction & Meta RL

SPACE OF MDPS



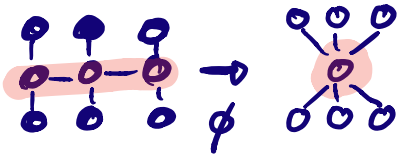
LEARN ABSTRACTIONS THAT MAKE RL EASIER

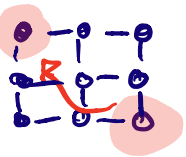
- STATE ABSTR.
- ACTION ABSTR.

$\phi : S \to S_\phi$

Option Framework $O \to \pi_o$ (Sutton et al., 1999)

STATE ABSTRACT ⇄ ACTION ABSTRACT





STATE ABSTRACTIONS: COMPRESSION VS. VALUE PRESERVATION

→ How small can $|S_\phi|$ be while still preserving 'good' solutions!

→ Info Bottleneck (Tishby et al., 99') ⇒ RL Abstraction Perspective:

SOURCE → ENCODER → PREDICTOR

$y \longrightarrow d(y, \hat{y})$

MDP → ϕ → AGENT π_ϕ

$\pi_E \longrightarrow d(\tilde{\pi}^S_E, \pi^S_\phi)$

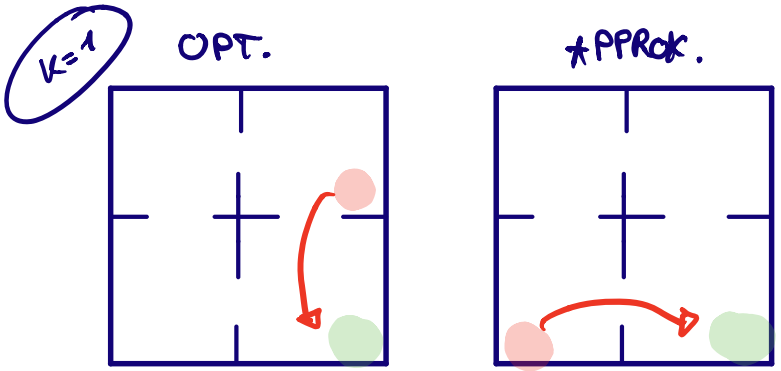
RL Formulation: $\min_\phi |S_\phi| + \beta \mathbb{E}[V^{\pi_E}(s) - V^{\pi^*_\phi}(s)] \leq$ DIB Objective

- SAMPLE EFFICIENCY
- $\beta = 0$: Only compression
- $\beta > 0$: Trade-off C vs. V

→ 4 Rooms Problem

→ Lunar Lander

ACTION ABSTRACTIONS: OPTIONS FOR VALUE PLANNING



⊙ Allow for better value propagation!

⊙ NP-Hard Problem

→ Not clear on qualitative difference

Raia Hadsell (DeepMind): Challenges for Continual & Meta RL

- Learning from i.i.d data
  - ↳ "1-page per topic in batch" analogy → INSTEAD: PROLONGED INSPECTION/CONSIDERATION ON ONE TOPIC
- How FAST/SLOW should learning algo be? → Adaptation/Evolution

META-LEARN

WHAT WE WANT

NOT TOO-FAST	NON-STATIONARY
NOT TOO-SLOW	NON-IID

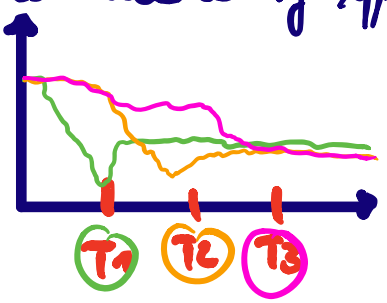
② MEMORY	OPTIMIZATION ①
COMPARE	TASK INFER.

---

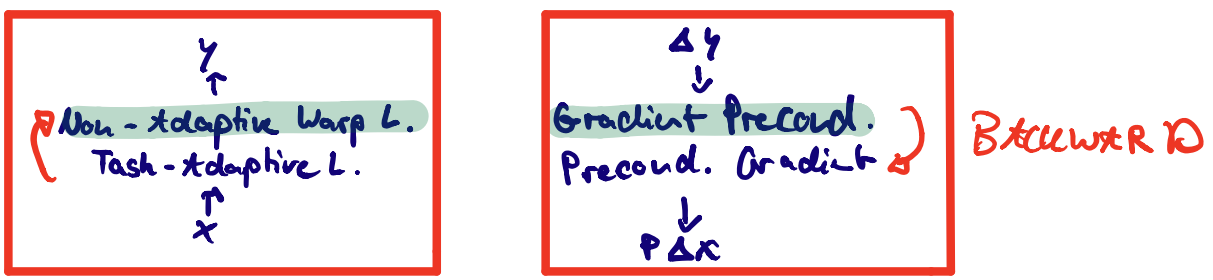
① WARPED GRADIENT DESCENT (WarpGrad) → S. FLENNERHAG

- Learn geometry does not depend on init but on search space
- Explicit precond. Δ over space
- Online w. MAML-type init!

→ Continual learning application (Sim Reg)



- Get faster at each new task!



---

② STABILIZING TRANSFORMERS FOR RL (GTrXL) → E. Parisotto

- Over time usage! Problem: Hard to optimize
- DM Lab application w. MERLIN (Wayne et al. 18)
  - Reactive Levels
  - Memory Levels

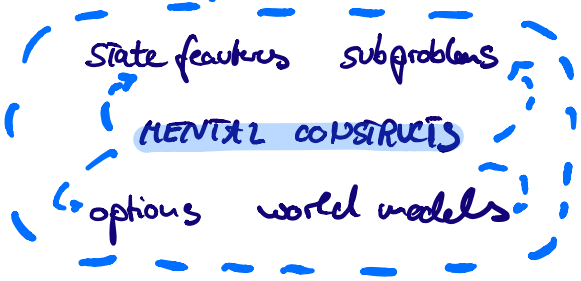
- Differences to TrXL
  - → Move layer norm inside ↳ better gradient flow
  - → GRU gating (works best!)
  - → Param pruning ≈ 22M.
  - → Unpack Sequences

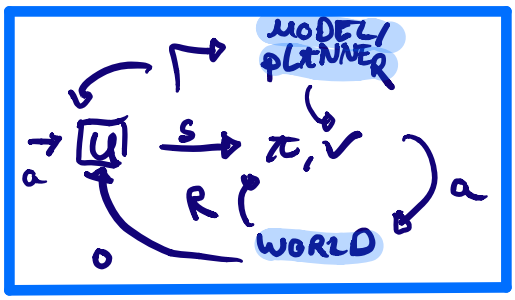
Rich Sutton (Uo tUberta): Towards a General-AI Architecture

WANT: GENERAL PRINCIPLES TO LEARN TO...
- PREDICT
- CONTROL

WE NEED
- Function approx.
- Partial Obs.
- Temp. Abstraction
- Non-Stationarity

"Super DYNA"





STATE FEATURE CONSTRUCTION
- Linear GVF
- Non-Linear Update
- SGD + Search
- Recurrent Process

DYNA (1990) PLANNING
- Learn model T, R
- DP-Style Planning
  - "Lookahead"
  - ⇒ Output 2 Feature

- Distribution Model
- Sampling Model
- Expectation Model
  Nothing lost if linear!

VALUE ITERATION WITH DISTRIBUTION/EXPECTATION MODEL

$$u_w(s) = w^T x(s) = \max_o \left[\hat{r}(s,o) + \sum_{s'} \hat{p}(s'|s,o)\, u_w(s')\right] = \max_o \left[\hat{r}(s,o) + w^T \hat{x}(s,o)\right]$$

- Linearity is key! ⇒ Nothing lost → Read out linear in learned features

SUBPROBLEMS ⇒ SHAPE REPRESENT. BEHAVIOR + HIGH-L. PLANNING

(A) What should problems be?
⇒ Each subproblem max. one state feature

(B) Where do subproblems come from?
⇒ Come from state features!

(C) How do subs help main?
⇒ Solution to sub is option max. feature
↳ Plan on top!

NON-STATIONARITY ⇒ MEMORY

Permanent	Transient
w_0	$\tilde{w}, w_0 + \tilde{w}$
α small	$\tilde{\alpha} >> \alpha$

Not Required if V exact

"SUPER-DYNA"