O. Vinyals
(DeepMind)
- AlphaStar -

S. Whiteson
(Uo Oxford)
- Bayes-Adaptive
RL -

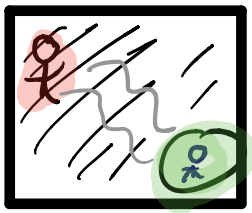
NeurIPS 2019
- Day 6 -
WORKSHOPS II

David Ha
(Google Brain)
- Embodiment -

Oriol Vinyals (DeepMind): AlphaStar – Grandmaster Level in StarCraft II using Multi-agent RL

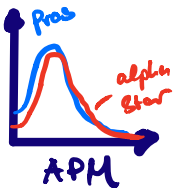
STARCRAFT II → A NEW CHALLENGE

① Collect Ressources
② Build base
③ Build units
④ Defeat agents



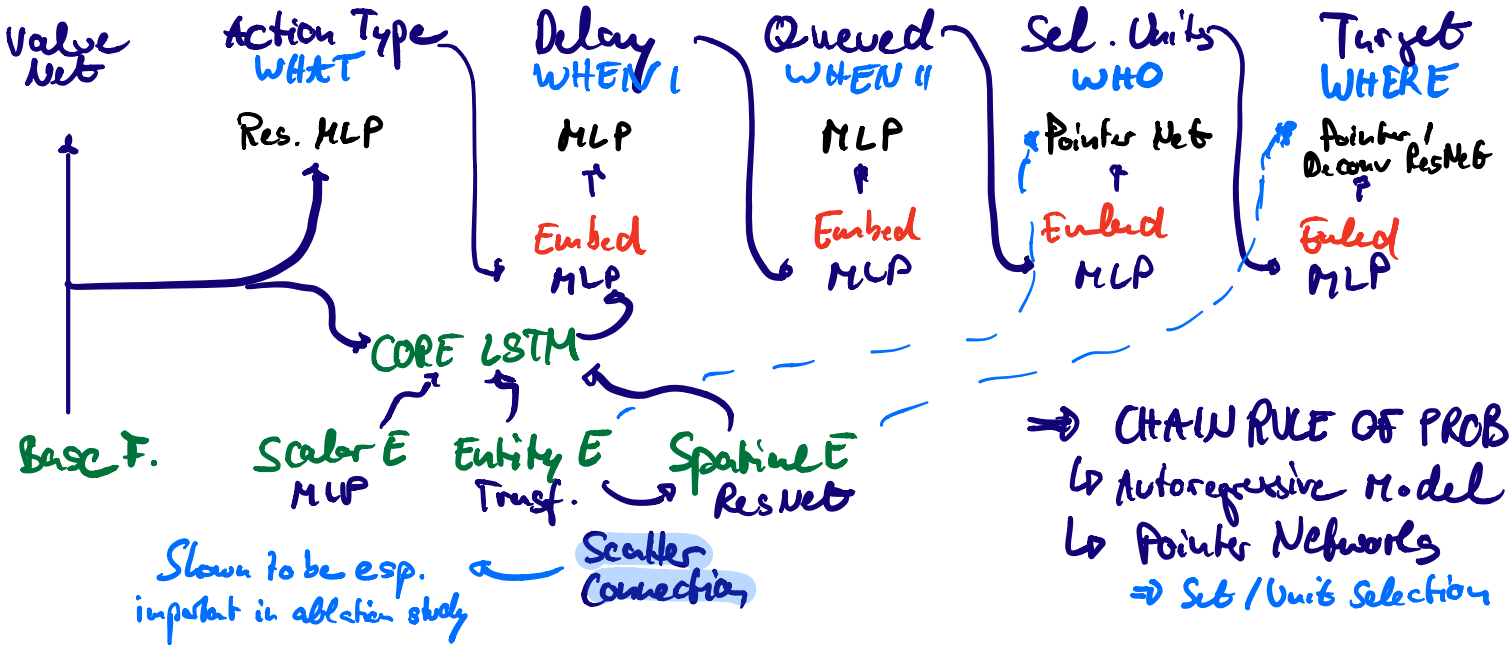
- Partially Obs.
- Multiagent
- 1000s of Moves
- Huge $\sim 10^{26}$

HUMAN-LIKE CONSTRAINTS



NO LONGER ZOOM OUT

ACTIONS & ARCHITECTURE → SEQ. MODELLING PARALLELS



→ CHAIN RULE OF PROB
↳ Autoregressive Model
↳ Pointer Networks
⇒ Sel / Unit Selection

EXPLORATION → BEHAVIOURAL CLONING + PRIOR REGULARIZ.

→ Long timescales very challenging
→ Naive worker rush ⇒ local Optimum
⇒ Instead: IMITATION LEARNING VIA BEHAVIOURAL CLONING

- ⊛ $\min KL(\pi^{exp} | \pi^{\alpha^*})$
- ⊛ Embedding Context z
  ↳ Only during training

→ RL Training ⊕ KL prior on expt behavior traces
→ TD(λ) version: UPGO // V-trace
↳ Off vs. On-Policy

ROBUSTNESS → LEAGUE

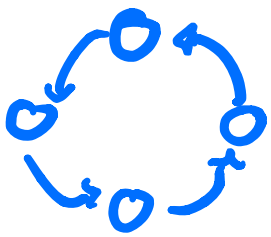
→ 'The League': Store Chkps.
→ Enforce diversity!
→ Hold-Out agent validation

Shimon Whiteson (UoOxford): Bayes-Adaptive DRL via Meta-L.

EXPLORATION → NAIVE ⇒ NO CONDITIONING OF BEHAVIOUR ON UNCERTAINTY EST.

Systematic Exploration: Rmax → Family PAC methods ⇒ Theoretical Guarantees

Poor init rewards
Over-Exploration?!
Tabular Method
↳ Pseudo-Count ⇒ Density Method



REAL vs. SIMULATED

- Rewards Real/Cost
- Balance Crucial
- Regret Min.

- Save env
- Sample-efficient
- Best arm ident.

⇒ CLAIM: There are no good methods for real exploration at these amenity!

↙ Principled Use of Prior Struct. Knowl.
↘ Reward Informed Density L. Most Cit

BAYESIAN RL

① Learning ⇒ Task Uncertainty

② Bayes-Adaptive ⇒ Hyper-State

$S^+ = S \times \mathcal{B}$ → Belief Space

↳ Explore only if it decreases task uncertainty + reward increase

CHALLENGES

① T, R Representation

② Intractable Planning ↘
③ Intractable Inference ↗ Instead: Posterior Sampling

Plan only in grad MDP ↙

⇒ No longer Bayes Optimal

variBAD ⇒ Meta-L. Family

$M_i \sim p(M)$ → Latent m_i

↳ Infer $p(m_i; h_t^-)$

⇒ LEARN TO APPROX. B-OPT. BEH.

↳ Work by Luisa Zintgraf

ENCODER q_ϕ → DECODER p_θ
$q_\phi \sim m$ → POLICY π_ψ

↳ Bayes-Adaptive Perspective on Meta-RL
↳ Gridworld Validation ⇒ Decodes Interpretation of Learn. Dynamics
↳ PEARL: Rakelly et al., 18' – Off-Policy
↳ Info Bottleneck Interpretation!
↳ HalfCheetah Vel.: Quickly deduce task!

David Ha (Google Brain): Innate Bodies, Brains & Minds

∝ ∝ DEAD FISH = BODIES AS EVOLVED PRIORS → EVOLUTIONARY PERSPECTIVE ON INDUCTIVE BIASES

PASSIVE WALKERS
McGeer 90'
Collins et al 01'

⇒ BUILDING BLOCKS

RL for Improv. Agent Design (Ha et al., 18')

DIRECTLY OPTIMIZE CONSTRAINTS OF AGENTS

BIPEDAL WALKER

COMPLEX ENV ↔ ROBUST POLICY?!
⇒ Path of least Resistance

Deep Image Prior
Ulyanov et al. 18'

Schmidhuber - LSTMs
↳ bias towards seq. l.

Anthony Zador: 'Critique of Pure Learning [...]' → Nvs. N

FINDING GOOD PRIORS FOR DRL → WANNs (Gaier & Ha 19')

- Search for architectures by de-emphasizing weights ⇒ RANDOMIZE
- NEAT-based evolutionary algo ⇒ Fast tuning of single shared weight
  - ↳ Bipedal Walker: „Learn" to ignore parts of inputs
  - ↳ MNIST: Ensemble-based classification

① INIT → ② EVALUATE → ③ RANK → ④ VARY

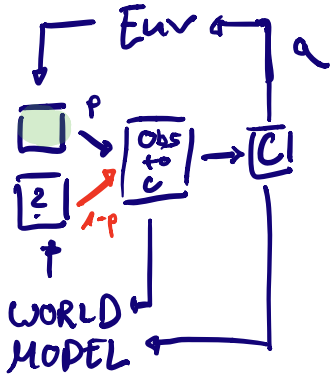
OPERATORS
INSERT NODE
ADD CONNECT.
CHANGE ACTIV.

⇒ AUTOMATED DISCOVERY – FOCUS ON NEW BUILDING BLOCKS

PREDICTING THE FUTURE PRIOR → OBSERVATIONAL DROPOUT

Ha & Schmidhuber 18'
Rec. World Models
↓
Scaling not goal by itself
↓
E ⇢ E*
TRANSFER
"WAKE ON, WAKE OFF"

- Freeman et al. 19':
  1-p% cases ⇒ „Blindfold" agent
  - ↳ Learned model only respects „important" aspects of env!
    ⇒ NOT PHYSICS EVOLVE



"A CREATURE DIDN'T THINK IN ORDER TO MOVE; IT JUST MOVED, AND BY MOVING IT DISCOVERED THE WORLD THAT THEN FORMED THE CONTENT OF ITS THOUGHT." – Andy ~~Wheat~~ Clark 18'