“Humies” Competition
GECCO 2018

Emergent solutions to high dimensional multi-task reinforcement learning

Stephen Kelly & Malcolm Heywood
Why does the result qualify as human competitive?

Game title:
- Atari
- Doom

Visual State $s(t)$

End-of-Evaluation

Game score

Atomic Action $a(t)$
Visual RL dominated by Deep learning

- DQN (2015)
  - Visual RL on Atari Learning Environment (49 titles)
  - Q-learning with Deep learning
  - Cropped visual image (84 × 84)
  - Frame stacking (removes the interleaving of sprites & stochastic properties)
  - “able to surpass the performance of all previous algorithms and achieve a level comparable to that of a professional human games tester across a set of 49 games” [Nature (2015) Vol. 518]
- One policy per game title
- Learning parameters and DNN topology identified a priori
Visual RL Compared to ‘human’
100 (algorithm – rnd)/(Human – rnd)

Log ( % human )

Algorithm Better than Human

Human level

Algorithm Worse than Human

Statistically equivalent

July 2018
Humies
Visual RL and Multi-task learning

- Multiple game titles played by single agent
- Single title DQN provides the baseline
- Best DNN result needs prior knowledge regarding parameters and topology
- Constitutes an example of a task pertaining to ‘Artificial General Intelligence’
Multi-title TPG versus Single-title DQN

![Chart showing the comparison between Multi-title and Single-title DQN scores for various games like Alien, Battle Zone, Asteroids, and others. The chart compares the performance across different groups (Group 1, Group 2, Group 3).]
Why [is our entry] ‘best’ in comparison to other entries?

• Single title task
  – TPG provides solutions competitive with human and DQN
  – Agents have to be competitive over multiple game titles

• Multi-title task
  – TPG multi-task solution is competitive with DQN trained under single title setting
  – DNN state-of-the-art in single task does not address Multi-title task

• TPG for Single title task a special case of TPG for Multi-title task
The ‘icing on the cake’

• TPG addresses multiple issues simultaneously:
  – **Complexity of topology** is emergent and:
    • Highly modular
    • Unique to the task
    • Explicitly reflects a decomposition of the task
  – **No image specific instructions** just:
    • Four 2 Argument operators \{+ , − , × , ÷\}
    • Three 1 Argument operators \{log, exp, cosine\}
    • One conditional operator
  – TPG **highly efficient computationally**
  – Some examples...
Teams (nodes) per graph emerge...

[ditto pixels used]
Emergent discovery of Multi-title solutions
Run time complexity

**DQN**
- $\approx 1.6$ million weights in MLP
- $\approx 3.2$ million convolution operations in DNN
- 3.2 GHz Intel i7-4700s
  - 5 decisions per second
- GPU acceleration
  - 330 decisions per second

**TPG**
- Single title
  - 71 – 2346 Instructions (avg)
- Multi title
  - 413 – 869 Instructions (avg)
- 2.2 GHz Intel E5-2650
  - Single title:
    - 758-2853 decisions per sec.
  - Multi-title
    - 1832-2922 decisions per sec.
Questions?