UT Austin Villa: High Dimensional Parameter Optimization for Kicking

Patrick MacAlpine and Peter Stone

Department of Computer Science, The University of Texas at Austin

RoboCup 2017
Skill Description Language

Kicks represented as series of joint angle parameterized fixed poses

SKILL KICK_LEFT_LEG
KEYFRAME 1
setTarget JOINT1 $jointvalue1 JOINT2 $jointvalue2 ...
setTarget JOINT3 4.3 JOINT4 52.5
wait 0.08

KEYFRAME 2
increaseTarget JOINT1 -2 JOINT2 7 ...
setTarget JOINT3 $jointvalue3 JOINT4 (2 * $jointvalue3)
wait 0.08

Patrick MacAlpine (UT Austin)
Reinforcement Learning Direct Policy Search

- Learn a parameterized policy that determine an agent’s behavior (what actions an agent chooses based on state of environment)
- Optimization algorithm (CMA-ES) produces candidate parameters to evaluate on optimization task
- Optimization task evaluates parameters and returns fitness to optimization algorithm
More Parameters Produces Better Kicks

- 18 parameters produced 6 meter kick
- 24 parameters produces 11 meter kick
- 80 parameters produced 20 meter kick
More Parameters Produces Better Kicks

- 18 parameters produced 6 meter kick
- 24 parameters produces 11 meter kick
- 80 parameters produced 20 meter kick

How many and which parameters to choose to optimize?
More Parameters Produces Better Kicks

- 18 parameters produced 6 meter kick
- 24 parameters produces 11 meter kick
- 80 parameters produced 20 meter kick

How many and which parameters to choose to optimize?

Can we just optimize all parameters to perform optimization across largest set of potential possible kick motion policies?
Fast Walk Kick

Less than .25 seconds to execute, can kick the ball over 18 meters

- Optimize all joint angles except for head across 12 simulation cycles (≈ 260 parameters)
- > 1.5 average goal difference improvement against 2016 binaries
- Walk kick that does not require transition from stand position
Deep Learning of Kick

- Would like to learn parameters for >2000 output joint actions
- CMA-ES doesn’t scale well to 1000s of parameters
- Model kick as a neural network
- Input = time passed since kick started, output = joint angles
Deep Learning of Kick

- Use **supervised learning** (backprop) to obtain a **seed** for the neural network from currently optimized long kick

Seed produces 8 meter kick
Deep Learning of Kick

- Train neural network using Trust Region Policy Optimization (TRPO)
- Monotonic policy improvement during learning by constraining KL divergence of distribution of policy actions between iterations