Multi-DDPG Humanoid Control

XXXX XXXXX

August 2016

Contents

Statement of Intent

We propose to use deep neural networks in conjunction with reinforcement learning for the control of a human-like simulated-robots' joints to better achieve human-level control and life-like movement. We will use proven techniques for attacking problems of continuous control such as DQN and the DDPG algorithm [6, 7, 8]. We propose a unique application by creating separate groups of networks, each running the DDPG algorithm, to control each joint of a simulated robot individually. We intend to create a novel algorithm to allow communication between networks and ability to revise policy and value decisions in control.

Background and Significance

One of the greatest difficulties in creating an android type robot is movement control. While people would love to use such robots for assistance in a variety of environments it is currently in-feasible. Current algorithms cannot even get a simulated robot to express human like movement and adaptability. For this reason a majority of robots we see in industry are not human-like, are stationary, or maneuver throughout their environment through the use of wheels.

One example of the best achieved efforts toward android type robots is the Atlas robot [2]. While the abilities of Atlas are phenomenal they clearly lack the grace and efficiency with which a human moves as well as the speed. Others have found ways to express grace and stride [3] but lack speed. While incremental, these efforts do prove the desire and need for humanoid robots is strong.

Recent advances in computer capabilities, success of neural nets and deep learning [1, 9], and discoveries in reinforcement learning [8, 11, 12] and continuous control [6] have allowed for new discoveries and more human-level control. These discoveries have made bench-mark results in playing Atari2600 games [6], controlling stationary and moving 2D and 3D simulated agents of many kinds [5, 6], and playing games such as Go [10] at a human level.

Methodology and Procedures

We will be using some frameworks for neural nets, deep learning, reinforcement learning, and physics simulation that have already been proven. Google's Tensorflow python libraries will be used to set and implement our neural networks as well as provide the algorithms for deep learning. We will implement the DDPG [6] algorithm and architecture with Tensorflow's libraries ourselves. We will use OpenAI Gym's framework for our reinforcement learning environment to help us set up states, observations and rewards needed for learning. Finally we plan to use MuJoCo's physics simulator to imitate our android in realistic situations. These will be the bases upon which we will frame our research and results.

First we will produce a robust network running the basic DDPG algorithm on a series of single joint continuous problems. Solving multiple environments will help show the robustness of the network and also the optimizer. Next we will apply this same network to solving continuous multiple joint problems, from there we will attempt to control the same problems using the same network but with a copy for every joint. Here we will seek to find novel ways to allow these networks to communicate and revise policy and value decisions made. Once successful we will apply the same steps to the humanoid robot control problem, first using the standard DDPG algorithm then applying our novel approach of one network per joint. Reiterating the ability of our one network to reproduce results previously achieved will again create assurances of the robustness of our networks.

We expect that our algorithm will produce better results than previously achieved and allow for more human level control. Some studies have intimated that connecting the joints neurologically similar to the way a human body is can achieve better and more natural results [4].

Human and Animal Subjects

There will be no human or animal subject involved in our research

Preliminary Outline

A. Title Page and Author

B. Abstract

C. Acknowledgments

D. Dedication

E. Table of Contents

F. Introduction

  i. Motivation and Objectives

  ii. Contributions

  iii. Statement of Originality

  iv. Publications

Preliminary Research

To Prepare for this thesis and proposal, I have been working on solving classic control problems with deep neural nets following the work of Sutton and Barto [11] in reinforcement learning. I have also been working on learning the details of neural nets, deep Learning and optimization [1, 9]. I have read about the difficulties of controlling human like robots in John Baichtal's Robot Builder. Beyond this I have been collecting documents on current research in the area, and have a list of future reading and future research needed before working on our novel algorithm. I am continuing to build this list. I Hope to build on the work of DeepMind and improve upon their algorithms. I have researched algorithms invented by them for reinforcement learning and deep learning and implemented them to achieve similar results. Their work is among the latest and groundbreaking research in the field of machine learning achieving benchmark results [6, 7, 8]

Qualifications of Investigator

Joshua Powers is a senior in the computer science program at Brigham Young University. He has been working in the Perception Control and Cognition Research Lab run by David Wingate, working on deep learning and reinforcement learning problems. Joshua has created several deep neural networks that solve various classical control problems using the DQN algorithm [7]. Joshua will also be finishing two courses on machine learning and specifically neural networks by the end of this year. Joshua is in the honors program and has spent significant time drawing connections between disciplines and attending special learning events. Joshua has also been a teacher assistant for many of the computer science classes at Brigham Young University. He has a 3.77 GPA has been on the Dean's List for The College of Physical and Mathematical Sciences and currently has a full-ride academic scholarship.

Qualifications of Adviser

David Wingate received a B.S. and M.S. in computer science from Brigham Young University in 2002 and 2004, and a Ph.D. in computer science from University of Michigan in 2008. He was a postdoc and research scientist at MIT with a joint appointment in the Laboratory for Information Decision Systems and the Computational Cognitive Science group. Before joining BYU, he was the director of Lyric Labs, and advanced R&D group inside Analog Devices, Inc., where he directed research at the intersection of machine learning and hardware.

Dr. Wingate's research interests lie at the intersection of probabilistic programming, hardware accelerated probabilistic inference and machine learning. His work spans diverse topics in audio processing, Bayesian nonparametrics, reinforcement learning, massively parallel processing,visual perception, dynamical systems modeling, and signal processing. He is currently pursuing applications in robotics and low-cost medical imaging.

Schedule

September 1st: Begin research on thesis, implement the Actor Critic algorithm on a single joint and achieve simple continuous control, begin search for current research on communicating neural nets.

September 15th: Finish single joint environment and begin work and double joint environment.

October 1st: Finish double joint environment and recreate results for the current Mujoco simulator using the actor critic method, begin attempts on novel algorithm for double joint control with two neural nets rather than one.

October 31st: Review results from current tests, if favorable continue multi-neural net control and apply to Mujoco robot. Possible reconsider algorithm.

December 1st: Begin writing preliminary results for thesis, start work on human level joint robot for Mujoco simulator. Continue development of multi-neural net agent. Ask is it good, can it be better?

January 1st: Start working with human level joint robot in Mujoco. Apply current algorithm making changes as needed. Work on rough draft for thesis

February 1st: Prepare rough draft and results from research for review by advisor

February 12th: Submit draft to department advisor and honors department, start working on thesis defense

March 1st: Complete thesis defense draft, prepare for submission

March 11th: Completed thesis defense papers

April 1st: Review thesis and defense, prepare for the defense

April 20th: Thesis presentation and defense

Expenses and Budget

Our proposed research will have a budget of 550 dollars. This will be enough to cover our only expense which will be subscribing for access to MuJoCo's physics simulator.

References

[1] Ian Goodfellow Yoshua Bengio and Aaron Courville. Deep learning. Book in preparation for MIT Press, 2016.

[2] Boston-Dynamics. Boston Dynamics: Dedicated to the Science and Art of How Things Move., 2016.

[3] Steve Collins, Andy Ruina, Russ Tedrake, and Martijn Wisse. Efficient bipedal robots based on passive-dynamic walkers. *Science (New York, N.Y.)*, 307(5712):1082–5, feb 2005.

[4] Tao Geng, Bernd Porr, and Florentin W o rg o tter. Fast Biped Walking with a Sensor-driven Neuronal Controller and Real-time Online Learning. *The International Journal of Robotics Research*, 25(3):243–259, 2006.

[5] Online Submission Id. Terrain-Adaptive Locomotion Skills using Deep Reinforcement Learning. *To appear in ACM TOG*, 35(4):1–10, 2014.

[6] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, pages 1–14, sep 2015.

[7] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing Atari with Deep Reinforcement Learning. *arXiv preprint arXiv: ...*, pages 1–9, dec 2013.

[8] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[9] A. Michaael Nielsen. *Neural Networks and Deep Learning.* Determination Press, 2015.

[10] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[11] R.S. Sutton and A.G. Barto. Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, 9(5):1054–1054, 1998.

[12] Csaba Szepesvári. Algorithms for Reinforcement Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 4(1):1–103, 2010.