

ITRG Project Proposal

Samuel Lerman

CSC 400: Problem Seminar

Prof. Randal Nelson

September 12, 2017

ABSTRACT

While researchers in the field of reinforcement learning have made great strides in defining mathematical models for human senses like sight and sound¹, other innate human qualities have often gone ignored. Empathy is seen in almost every animal, and is without a doubt one of the most valuable traits within a community of intelligent beings. We propose a simple context and methodology for beginning to apply the often human-like prowess of deep reinforcement learners to the problem of empathy. We introduce a variant to the game of Pong, which can be likewise abstracted to increasingly more complex contexts. Two paddles play the game of pong as per usual and one is defined to be the primary agent. A third paddle, representing our empathizer, is added and given complete free-range over its position and rotation on the board. The task of the empathizer is (1) to use a process of inverse reinforcement learning to estimate the objective function of the primary agent, and (2) to assist the primary agent in achieving its task. Ultimately, the empathizer should learn to keep the ball from crossing the primary agent's boundary and to bounce the ball past the boundary of the opponent. Our goal is to demonstrate that this is possible with the use of policy gradients, a technique of updating the gradients of deep neural networks to maximize an objective without supervision². A convolutional neural network is passed only the raw pixels of the game and a definition of the primary agent. The network then uses policy gradients to "empathize" with the primary agent and help achieve its goal. Such an algorithm can be abstracted with varying degrees of effectiveness to any video game in which an agent can be defined, and perhaps with time and continued research, a scenario in the real world to empathize with human agents.

PROJECT DESCRIPTION

Introduction

As artificial intelligence continues to advance, we as a species cannot ignore its potential impact on our civilization. Considerations that were once reserved for the realm of science fiction must now be taken seriously. There's generally a consensus that AI will have drastic effects on our society. World leaders and tech industry heads are openly discussing the need for precautions such as a universal basic income³, as automation threatens the foundation of our economy. Furthermore, once-fantastical ideas like the "singularity" and visions of apocalyptic or utopian fates are becoming increasingly more mainstream. With near-daily breakthrough achievements in the field, thanks to research organizations like DeepMind and OpenAI, these prophesies of monumental transformation seem increasingly more likely, and with industrial figures like Google Director of Engineering Ray Kurzweil⁴ and Tesla CEO Elon Musk⁵ championing such views, these predictions are increasingly more credible.

Regardless of where one's predictions fall—whether you envision a world post singularity in which a being with general intelligence is able to achieve any function with god-like ability, or a more grounded future in which simply a vast fraction of human jobs are displaced by robots—one thing is for certain: any individual or organization that monopolizes such a technology could have terrifying power over the whole of humanity. In today's day and age, control over production resources gives leverage to elites over workers and competitors, and creates unjust disparities⁶. This is made even worse if a single individual or private entity owns an algorithm or training dataset that functions to serve a critical role in our society. For example, if a proprietary Al automates the agricultural industry, the owners of such an algorithm would have a monopoly over a basic human need. More frighteningly, any military equipped with a proprietary Al would be unstoppable. Human soldiers would become obsolete, and whoever would control the algorithm would have dominion over the rest of the world. Simply said, as Al continues to advance, it becomes increasingly more important that we ensure it cannot be abused or exploited.

Another risk run by the advancement of artificial intelligence is the possibility of a poorly defined objective function. The Al itself may be benign, but it could still spell the end of humanity if the best way to optimize its own defined goal is something unanticipated. One might task it to feed everyone on the planet, and it may find that to do so most effectively would be to eradicate everyone from the planet, thus leaving no one to feed.

In any event, researchers need to begin defining—mathematically—certain precautionary measures. Here, we consider an ideal, romanticized solution. That is, we propose endowing such an intelligence with genuine compassion for humans, such that no matter who tasked it and no matter what it was tasked to do, failsafe empathy would prevent it from harming any individual or the larger community. But is empathy quantifiable? Can

one mathematically define an objective function that would represent human empathy? We propose (1) yes, it is possible, (2) a method for where to begin, and (3) what more needs to be done.

To argue (1), we simply point to the roughly 8 billion examples wandering the Earth today (and far more, if we count other species). Somehow, in the depths of the reader's consciousness, there is an inescapable empathy for her fellow man. One can see a smile on another's face and feel elated, or tears coming down someone's cheeks and deep pity. These instincts are ingrained in the human objective function and thus, assuming the brain consists of physical matter and operates on mathematical laws, empathy can be quantified.

Methodology

In lieu of a formal mathematical definition of empathy, we will formulate, for the purpose of an initial trial, one based on deep reinforcement learning.

Definition 1 An *empathizer* is a learning algorithm parametrized by contextual information (e.g. raw image pixels) and a definition of an agent (e.g. left paddle) or set of agents that outputs a policy for maximizing the average value of all agents' estimated objective functions.

Definition 2 An *agent* is any entity whose behavior is based on an objective function, defined with the hierarchy of needs and emotions explained below.

Definition 3 In the case of this experiment, the *primary agent* will be the singular agent—say, the left paddle—with which the empathizer empathizes.

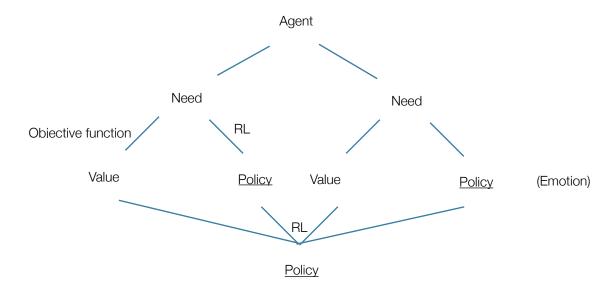
This project seeks to define the process of empathy for the empathizer with deep reinforcement learning techniques. Policy gradients are a proven technique for optimizing neural networks to play video games. Google's DeepMind used a single deep Q-learning algorithm to learn a wide array of Atari games⁷. They later showed that policy gradients, if configured properly, would be at least as effective⁸. Then in another groundbreaking paper, the team at DeepMind showed that policy gradients, in tandem with Monte Carlo tree search, could beat a champion at the game of Go⁹. In both of these cases, the architecture used was that of a convolutional neural network. CNNs are state of the art for image recognition. As of 2015, based on the benchmark of the ImageNet competition, the ability of CNNs to recognize images has surpassed that of humans¹⁰.

There is no doubt that an optimized CNN can learn to play Pong given raw image pixels and an objective function. However, the novelty of this project, and the challenge of it, is in having the rules, goal, and rewards of Pong be kept secret from the empathizer. Instead, the objective function of the empathizer must involve an estimation of the objective function of another observed agent, and a proportional reward. One question asked is: can we use CNNs and policy gradients to create an objective function based on the behavior of another agent?

The field of deep inverse reinforcement learning is not new, but sparsely has been tested in the context of video games. Deep inverse reinforcement learning algorithms have been experimented with in the last few years in the context of learning via demonstration¹¹. In this project, we are not interested in mimicking an agent, but rather assisting it.

We propose the following game: two agents play Pong regularly. One is defined to be the primary agent. Our empathizer is a third paddle, added to the board, and is given complete free-range, godlike power over the board. It must observe the game, in particular the primary agent, learn the primary agent's objective, and use its far greater capacity to help achieve that objective.

To connect the problem to humanity, we define a hierarchy of needs and emotions.



Definition 4 Each *need* (in the case of humans: food, water, warmth, love, empathy) is its own reinforcement learning model with an objective function *value* and a *policy* for every state.

Definition 5 An *emotion* is a state consisting of the unique union of value, policy pairs particular to each need of an agent.

In Pong, we define the needs and emotions of our paddles as such. A paddle has two needs: (1) to return the ball and (2) to make the opponent miss. These each have a policy derived from policy gradients and an objective function such that a success rewards +1 and a failure faults -1, where the total value for each is constrained to [-10, 10].

Another policy is learned to aggregate this vector of need values optimally (while this is redundant for Pong agents, it helps to abstract to more complex agents who have many needs, like human beings, and to whom the explicit aggregation cannot be known). A paddle's emotion is represented by its unique value, policy pairs. Therefore all of its emotions are the disjoint partition of of the set of all possible value, policy pairs belonging to an agent.

Every need an agent has consists of a value, policy pair. In this Pong context, an emotion is a 4-dimensional vector consisting of two values and two policies. Let:

 $F: Z^4 \longrightarrow Z^3$, where F is a scaled mapping between an emotion and an element of a discrete set of RGB colors (discrete simply to reduce complexity).

Given an emotion x, the agent's paddle displays the color F(x). These colors are analogous to body language and facial expressions in human beings. Therefore, the problem is formulated for more complex contexts, with the ultimate goal of application on human agents.

Potential Benefits & Future Steps

Pong is merely a starter context for quantifying human empathy. A similar abstraction could be tested on, for example, Mario—having an AI observe Mario's actions and learn to support him across the map. As the algorithm is improved and optimized i.e. as we better quantify empathy, we can increase the difficulty of the context.

- I/O interface that predicts your input
- A phone assistant that knows what you want before you ask
- Egalitarian/benevolent intelligence that understands humankind
- Many more

Resources and Timespan

To do this, one could use the Python deep learning framework TensorFlow to program a CNN with policy gradients. The main challenge is defining an inverse reinforcement learning algorithm, for which one would have to scour the literature. A preliminary product could take between a week or several months.

REFERENCES

- 1. http://ac.els-cdn.com/S0960982214010392/1-s2.0-S0960982214010392-main.pdf?
 http://ac.els-cdn.com/S0960982214010392/1-s2.0-S0960982214010392-main.pdf?
 http://ac.els-cdn.com/S0960982214010392/1-s2.0-S0960982214010392-main.pdf?
 http://ac.els-cdn.com/S0960982214010392/1-s2.0-S0960982214010392-main.pdf?
 http://ac.els-cdn.com/s0960982214010392/1-s2.0-S0960982214010392-main.pdf?
- 2. https://arxiv.org/pdf/1502.05477.pdf
- 3. http://www.businessinsider.com/president-obama-basic-income-debate-2016-10
- 4. https://www.theguardian.com/technology/2014/feb/22/robots-google-ray-kurzweil-terminator-singularity-artificial-intelligence
- 5. https://www.theguardian.com/technology/2017/feb/15/elon-musk-cyborgs-robots-artificial-intelligence-is-he-right
- 6. https://www.bbc.com/news/business-32824770
- 7. https://arxiv.org/pdf/1312.5602v1.pdf
- 8. https://arxiv.org/abs/1602.01783
- 9. https://gogameguru.com/i/2016/03/deepmind-mastering-go.pdf
- 10. https://arxiv.org/pdf/1502.01852.pdf
- 11. https://arxiv.org/pdf/1603.00448.pdf