Optimizing Interaction Enjoyment with Reinforcement Learning on Autonomous Robotic Behavior

Sarah Darrow and Cindy L. Bethel, PhD (Mentor)

Therabot
A SUPPORT COMPANION

Introduction

- TherabotTM is a therapeutic assistive robotic dog that is intended to act as a support companion for people diagnosed with Post-Traumatic Stress and trauma-related disorders.
- The focus of this research effort is the development and evaluation of life-like and functional behaviors
- A network of sensors including capacitive touch sensors, gyroscopes, and accelerometers has been developed and used to collect data on when, where, and how users interact with the robot
- Q-learning with Boltzman-Distributed exploration is used to analyze this data and determine how to alter the robot's behavior during different interactions
- This allows the robot to act autonomously with realistic movements that resemble a real dog naturally responding to interactions with a human.

Goals

1. Optimal distribution of autonomous movements for interaction enjoyment
2. Responsive to touch
3. Personalizable to each patient

Approach

Autonomous Movement

Each action has a set of weights measuring the predicted enjoyment of an action in a state

- These weights are updated as the actions are performed

$$\pi(a) = \pi(a|a_{t-1}) * \sum_{i=1}^{n} \pi(a|s_i)$$

An action a is randomly selected from a weighted distribution based on:

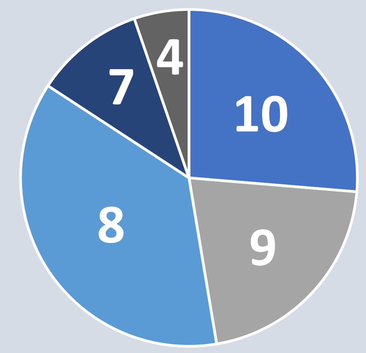
- The last action a_{t-1}
- Each of the sensors being touched $s_{1..n}$

The speed of the action and time between actions is selected at random

- The minimum speed is lower when more sensors are touched

Estimating Enjoyment

- Touch sensor data was recorded during interactions with 20 participants
- Each participant was asked to rate their interaction
- The data and scores were used with a support vector machine regression to estimate user enjoyment



Recorded Ratings

Features Used for Training:

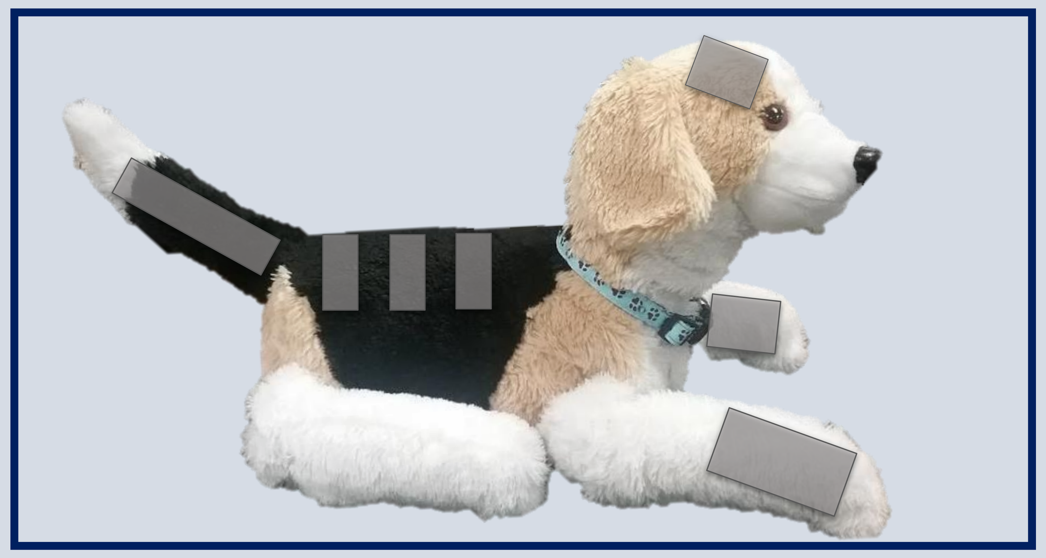
- Average magnitude of sensors being touched
- Number of sensors touched
- Number of new touches/ releases
- Percentage of each sensor magnitude in sum of all magnitudes

Updating Weights

$$\pi(a|a_{t-1}) = \pi\,(a|a_{t-1}) + \alpha(S - S_a)$$
$$\pi(a|s_i) = \pi\,(a|s_i) + \alpha(S - S_a)$$

After each sensor evaluation, the weights for the last 3 actions taken are updated, along with the weights for the sensors touched when that action was chosen

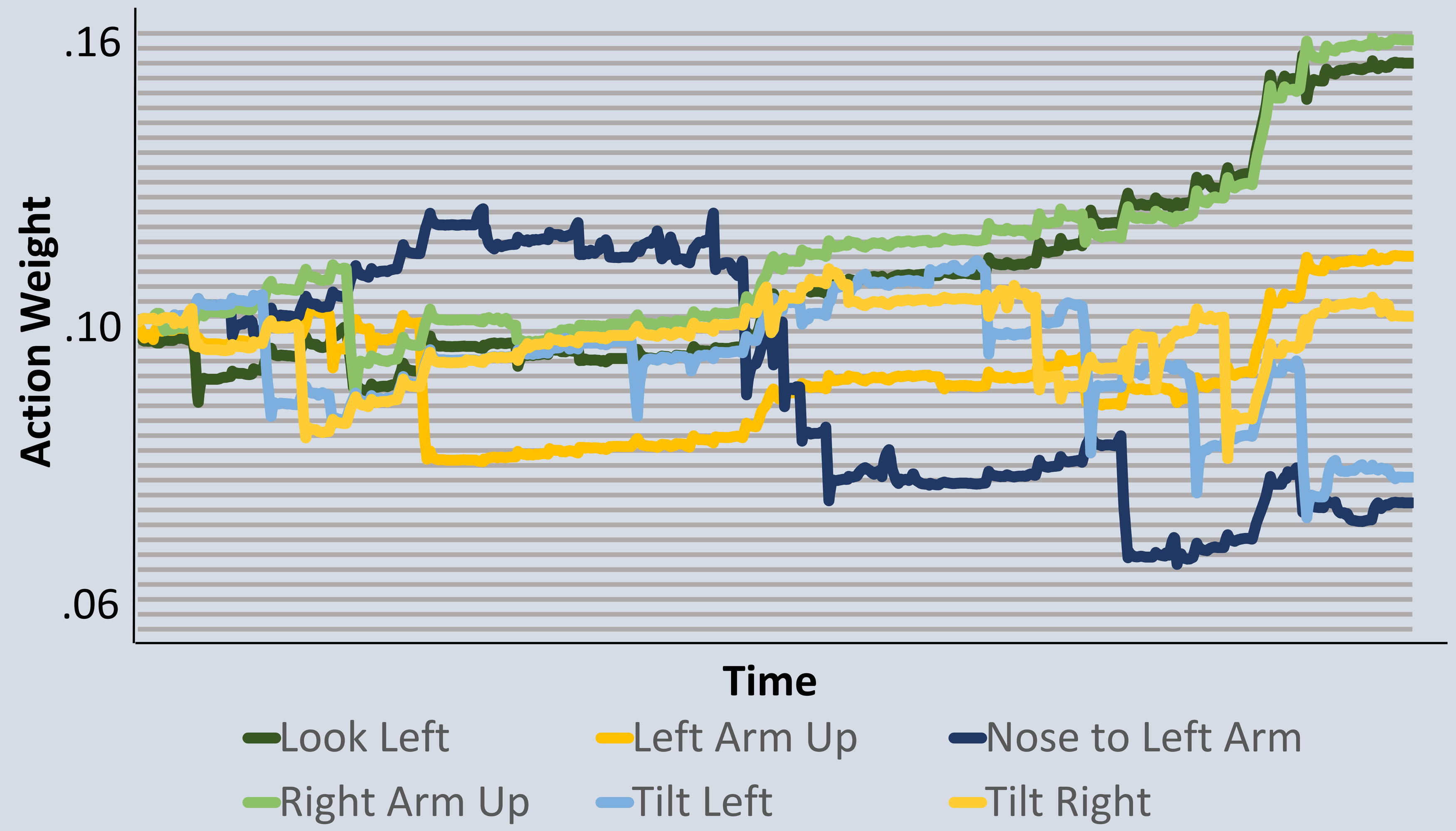
- Based on the enjoyment score S relative to the past scores S_a



7 Capacitive Touch Sensors

Results

An example of action weights changing over a period of 13 minutes



Conclusions

- The robot is capable of naturally moving autonomously and responding to tactile interactions
- The strategy for choosing movements is updated over time to adjust for each patients' preferences
- More data (especially negative feedback) should be collected to ensure scoring enjoyment is accurate

Acknowledgements

- The Therabot™ team: Zachary Henkel, Kenna Baugus, Nicholas Dinep-Schneider, John Atkins, and Nikhil Lokhande
- This project was funded by the James W Bagley College of Engineering Undergraduate Research Stipend

References

- Stiehl, Walter Dan and Cynthia Breazeal. “Affective Touch for Robotic Companions.” *ACII* (2005)
- David Silvera-Tawil, David Rye, and Mari Velonaki. 2015. Artificial skin and tactile sensing for socially interactive robots. Robot. Auton. Syst. 63, P3 (January 2015), 230-243.
- Even-Dar, Eyal, and Yishay Mansour. "Learning rates for Q-learning." *Journal of Machine Learning Research* 5.Dec (2003): 1-25.

MISSISSIPPI STATE UNIVERSITY

STARS SOCIAL, THERAPEUTIC & ROBOTIC SYSTEMS LAB

MISSISSIPPI STATE UNIVERSITY JAMES WORTH BAGLEY COLLEGE OF ENGINEERING