2017 Fall Robot Lab Undergraduate Final Progress Report

Robots Gain Social Intelligence Through Reinforcement Learning

B04901164 EE3 莫絲羽

1. Introduction

   This is the first semester to join the undergraduate program in Robot Lab. Zi-Yun has helped me out a lot getting familiar with all those programming environments together with the cooperation of this topic “Robots Gain Social Intelligence Through Reinforcement Learning”. Shih-Huan has also given me some pieces of advice with this project.

2. Project goal

   The goal is to make Pepper learn some basic social skills through deep reinforcement learning. However, due to the difficulties to teach Pepper about social intelligence, we have only picked four actions (Fig. 1.) including “do nothing”, “shake hands”, “say hello”, and “say goodbye” for Pepper to learn. The main point is that Pepper has to decide which action to take depending on his sight, which is a picture taken by his built-in camera.

Fig. 1. Four actions for Pepper to learn.

3. Progress

   A. Pepper APIs

      The first step to get familiar with Pepper was to figure out how to implement the APIs provided by NAOqi documentation. The followings are the APIs that I’ve learned this semester.

      a. Speech recognition: It is able to recognize words from vocabulary using the embedded speech recognition. In this project, it is used to recognize the response of human after each action is done.
      b. Motion control: The ALMotion module provides methods which facilitate making the robot move. The hands shaking is implemented by ALMotion.
      c. Face detection: ALFaceDetection is a vision module in which the robot tries to detect, and optionally recognize, faces in front of him. At first, we

want to utilize face detection to trigger the data collection episode of the scenario with human in front. Nevertheless, the performance is not quite good, so we try the following alternative.

d. Human detection: ALTracker, similar with face detection, but instead of tracking face, this module aims to track the whole human body. It is successfully implemented to detect whether there is a human passing by, and thus triggered to interaction with the person.
e. Laser detection: ALLaser Module retrieves data sent by the Laser head and stores it in an ALMemory key named: Device/Laser/Value. It is used to detect the distance between human and Pepper in order to let Pepper know how far he has to move closer to human.
f. Get images: It is used to take a picture with the built-in camera of Pepper and store it in our computer.

B. Navigation methods

In order to make Pepper return to its original point after each interaction, a navigation method has to be utilized. I've tried three of them and finally chose the last one. The following briefly shows the three methods

a. ROS SLAM

I've installed the ROS package and learned some basic commands of it. However, the area that Pepper will move is too small, so ROS is kind of too complicated this time. Thus, I tried to use landmark detection.

b. Landmark detection

I set up six landmarks on the floor outside MD205 along the path that Pepper would move. Unfortunately, due to the various facing direction of Pepper, it's quite impossible to make sure Pepper sees landmark every time he finishes an interaction.

c. Localization

ALLocalization is a module dedicated to the localization of the robot in an indoor environment. First, run the function "learnHome". Pepper would look around and learn the home node. For returning back to home, just call the function "goToHome", and Pepper would go back to the original point quite precisely. This is the method I used in the data collection finally.

C. Data collection

a. System structure

The data collection system structure is depicted in Fig. 2. There are three scenarios including "with human", "bye-bye", and "without human". Each scenario represents an episode. Only when a scenario is

terminated will the next one begin. In that way, we can make sure the numbers of data among these three scenario are balanced. For the scenario with human, human detection is utilized to make sure there is a human in front and would like to talk to Pepper. Laser detection would then calculate the distance between Pepper and human, and then approach the person.

One of the four actions will be randomly chosen by Pepper. According to the distance it is from the human and the response it received from the human, Pepper would get an immediate reward. How the reward is decided is shown in Table 1.

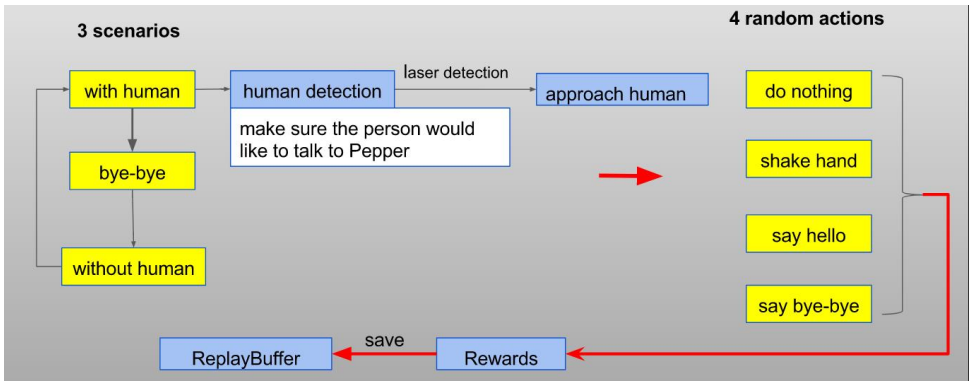

Fig. 2. The system structure of data collection

Table 1. Transition pair for reinforcement learning

If a person is detected and not too close to Pepper, get initial state					
Wish: Say "Good xx" and continue					
Action	Human's Respond	Reward	Next State	Terminate	Scenerio
Do nothing: 0	Do nothing	0.5	After 5 seconds	TRUE	Not face at Pepper
Do nothing: 0	Hello / Good xx	-1	After response is heard "Sorry for my ignorance"	FALSE Loop	
Good xx: 1	Do nothing	-0.5	After 5 seconds	TRUE	Not face at Pepper
Good xx: 1	Hello / Good xx	1	After response is heard Come closer to the person	FALSE Next process	Need depth camera on Pepper
Give its hand: 2	Do nothing	-1	After 5 seconds	TRUE	Not face at Pepper
Give its hand: 2	Hello / Good xx	-1	After response is heard "Sorry for my misunderstanding"	FALSE Loop	To far from Pepper
Give its hand: 2	Give his/her hand	-0.5	After touch sensor detects "I wish you a nice day."	TRUE	
Bye Bye: 3	Do nothing	-1	After 5 seconds "Sorry for my misunderstanding"	FALSE Loop	
Bye Bye: 3	Hello / Good xx	-1	After response is heard "Sorry for my misunderstanding"	FALSE Loop	

If a person is detected and close enough to Pepper, get initial state						
Wish: Give its hand and terminate						
Action	Human's Respond	Reward	Next State	Terminate	Scenerio	
Do nothing: 0	Do nothing	0.5	After 5 seconds	TRUE	Not face at Pepper	
Do nothing: 0	Hello / Good xx	-1	After response is heard "Sorry for my ignorance"	FALSE Loop		
Good xx	Do nothing	-1	After 5 seconds	TRUE	Not face at Pepper	
Good xx	Hello / Good xx	-0.5	After response is heard	FALSE Loop		
Give its hand	Do nothing	-0.5	After 5 seconds	TRUE	Not face at Pepper	
Give its hand	Hello / Good xx	-0.5	After response is heard "Sorry for my misunderstanding"	FALSE Loop	To far from Pepper	Should not appear
Give its hand	Give his/her hand	1	After touch sensor detects "I wish you a nice day."	FALSE Next process		
Bye Bye	Do nothing	-1	After 5 seconds "Sorry for my misunderstanding"	FALSE Loop		
Bye Bye	Hello / Good xx	-1	After response is heard "Sorry for my misunderstanding"	FALSE Loop		

If no one is found and no human in the image, get initial state			
Wish: Do nothing			
Action	Reward	Next State	Terminate
Do nothing	1	After the action is done	TRUE
Good xx	-1	After the action is done	TRUE
Give its hand	-1	After the action is done	TRUE
Bye Bye	-1	After the action is done	TRUE
If hearing human says "Bye Bye"			
Action	Reward	Next State	Terminate
Do nothing	-1	After the action is done	TRUE
Good xx	-1	After the action is done	TRUE
Give its hand	-1	After the action is done	TRUE
Bye Bye	1	After the action is done	TRUE

b. Transition pair

For each action, a transition pair would be stored in a replay buffer. The data format is a list consisting of [State, Action, Reward, Terminate, Next State]. “State” and “Next State” would be a picture taken before and after the action respectively. “Action” might be 0,1,2, or 3, which stands for “do nothing”, “say hello”, “shake hands”, and “say bye-bye” in order. “Reward” is the predetermined value that we gave for each condition. “Terminate” is a Boolean value indicating whether the episode ends after this action.

c. Demo video

I’ve recorded a video of data collection, including three episodes “with human”, “say bye-bye”, and “without human” sequentially.

Video link:

https://drive.google.com/file/d/1fbW3JjBu2lp7Qq9nE13HWN8Qui4iOGJ5/view?usp=sharing

4. Future work

A. Data collection

I hope to finish data collection in this winter vacation. Pepper would be placed outside MD205 to collect data automatically. In the meanwhile, Zi-Yun and I would give out a questionnaire asking for testers’ feedback and try to improve the structure.

B. Model training

After data collection, I would start to train the model using deep Q learning. If the result went out well, more aspects of social intelligence could be implemented.