Motivational Social Companion Robot

Lakhan Kamireddy & Shruthi Sukumar

March 11, 2019

Abstract

Humans are known to often display sub-optimal behaviors like procrastination or impulsivity when having to complete a task, or be able to delay gratification. Here we consider the specific scenario in which people do not stick to their work-out goals due to laziness or lack of motivation, despite setting goals to do so. We propose building a motivational agent that builds a descriptive model of the human's behaviour and uses this model to decide on appropriate interventions to motivate people to workout. We use a Markov Decision Process (MDP) based model for the human which the robot uses to predict behaviour. The robot itself takes decisions as a reinforcement learning agent that solves a 2-armed bandit problem. To validate that our system works, we propose running a control study of humans sticking to their workout goal. We then run a group our subjects with our agent and observe their persistence with respect to their goal. We expect to see an improved persistence among subjects in our test study compared to the control.

1 Main Deliverables

In this document we propose an end-to-end agent that is able to interact with the human to probe their current state-of-mind, provide interventions to improve motivation, test effectiveness of the motivation and update its estimate of the human model. To satisfy this, we theoretically design a 2-agent learning system, one agent being the human, and the other being our robot. Henceforth we will denote our robot as *MSC* that stands for motivational social companion. A pictorial representation or *system diagram* of our end-to-end system is presented in Figure 1.

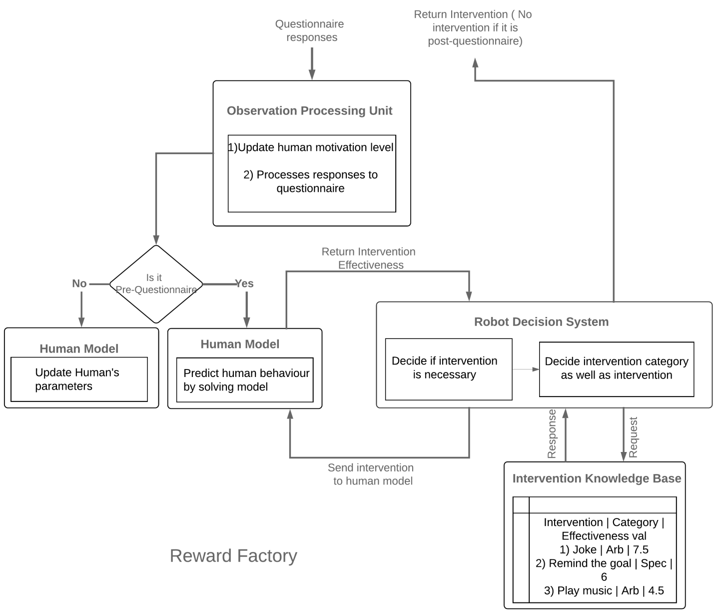

Figure 1: System Diagram our end-to-end system

Following is a brief description of the main components in the system diagram seen in Figure 1.

1.1 Observation Processing Unit

Lead : Shruthi & Lakhan

This block will involve parsing the inputs from the human subject. This will be basically be responses to a questionnaire from the robot. The purpose of this is to determine how motivated the human is to decide what intervention to give. Also determines how much of the goal the human has completed up until now. A post-workout questionnaire helps robots determine how accurate their model of the human is and update it if necessary.

1.2 Human Model And Solver

Lead: Shruthi
This block contains the model of the human based on which the robot operates. To predict behaviour due to an intervention, the model for the human needs to be solved online. Details about this will be discussed under Technical Components.

1.3 Human Model Parameter Updates

Lead: Shruthi
Based on the subject's responses to a post-workout questionnaire, the robot needs to improve its estimate of the human-specific parameters. This enables the robot to build a different model for every subject.

1.4 Robot Decision-Making System

Lead: Lakhan
This is the main decision block of the robot with respect to interventions to be dispensed. The robot needs to communicate with the human model block to identify how the simulated human will respond to different interventions. There are two main intervention categories: 1) Arbitrary intervention ,e.g., playing a song and 2) Goal-specific intervention which improves the salience of the reward of achieving the goal for the human. The specific intervention we consider for this case is reminding the human of a task. The robot also needs to communicate with the intervention database to determine what reward to test on the human model. It also updates the database based on the response of the model. After determining the best intervention, it then decides whether to exploit it or explore a new intervention to the human.

2 Technical Components

Here we list the technical components (mostly theoretical) that we will be over the course of this project. These subsystems will be developed largely in ROS and integrated with a webpage for this project. The development on ROS allows for using better robotic systems in the future to which our system can be adapted.

2.1 Observation Processing Unit

Lead: Shruthi & Lakhan Pre-workout and post-workout questionnaires will be displayed (question-by-question) to the human using a web page. Subjects will respond to this brief questionnaire that is meant to assess their mood. This will involve interfacing our back-end logic which we will build on the robot operating system (ROS) with `html` and other web-development languages.

2.2 Robot's model of human agent—MDP framework

Lead: Shruthi The human model based on which the robot will make decisions will be a MDP model with descriptive variables in the state space and reward function. A basic form of this MDP model is described below:

State Space:

$$S = \{m, g, c, q\}$$

where m is the motivational level of the human. m is a categorical variable with 5 possible values designating a discrete motivational state $m = 0, 0.25, 0.5, 0.75, 1$. g is the goal exercise, e.g.,want to exercise for 10 hours and c is the cumulative time of exercise done over the time horizon until now. q is the binary variable that determines if it is pre-workout or post-workout surveys.

Action Space:

$$a = x_i$$

$$x_i \in [0, 1]$$

where x_i is the fraction of g that the human chooses to exercise

Reward Function:

$$r(s_i, a_i, s_{i+1}) = \begin{cases} \beta * m_i - e * x_i + \alpha, & \text{if } c_i + x_i * g < g \\ \beta * m_i * 10, & \text{if } c_i + x_i * g \geq g \end{cases}$$

where i specifies the discretized time in the MDP that is considered. β is the individual dependent parameter of the subjective value of a reward, based on the internal state of mind of a human. b_i is the subjective value of the robot's arbitrary intervention if any are specified

State Transition Function: Goal is assumed to be constant over the horizon of the MDP

$$g_{i+1} = g_i = g$$
$$c_{i+1} = c_i + x_i$$
$$m_{i+1} = \max\{m_i + k * 0.25, 1\}$$

where T_i reported exercise time by the human in the $i + 1^{\text{th}}$. k is an integer that represents the gain of the arbitrary intervention provided by the robot that helps increase human's motivation level. k is a categorical random variable.

Bellman equation based solution

$$V^*(s_i) = \max_{a_i^*} \mathbb{E}\{r(s_i, a_i) + \gamma V^*(s_{i+1})\}$$

Here, γ is the discount factor that represents the human's lack of foresight towards their long term goal, i.e., their inability to delay gratification. The goal-specific intervention is assumed to modulate γ through some function, e.g.,

$$\gamma = \gamma + 0.05$$

.

Building and solving this MDP model will be done on ROS. This will be the back-end to our webpage interface that the human sees.

2.3 Human Model Update

Lead: Shruthi

Based on human's responses to post-workout questionnaire, the robot needs to update its assumption about human parameters. In the basic MDP example listed above, these parameters are β, e and *alpha*. These need to be updated based on some error-driven learning algorithm, the error here being the discrepancy between predicted human behaviour and empirical observation. This algorithm will also be integrated with ROS.

2.4 Robot Decision-Making System

Lead: Lakhan

The robot's decision-making system accesses the intervention knowledge base to determine which is the most rewarding intervention and starts with testing that on the newly updated human model. The robot will use a metric that determines if the intervention is good enough to improve motivation and then stop testing the human models with any further interventions. Essentially the robot cycles through the knowledge base from most to least rewarding and stops when the improvement by the human model satisfies some criteria (T.B.D) of motivation improvement. Once the robot is satisfied it has an effective intervention to give, it solves a multi-armed bandit problem to explore or exploit based on the knowledge base, that results in the best possible outcome. You could imagine that at the beginning, the robot explores more owing to it's low confidence of it's model of the human, but starts exploiting as it learns the human model better.

3 Evaluation Criteria

Following is a rudimentary of the steps we will take to evaluate our system:

1. Divide the group of recruited subjects into control and test groups
2. For each group see fraction of the total number of people that completed their workout goal
3. Across all subjects in each group, compute fraction of workout goal completed and average

In test group subjects are surveyed with the web-application + ros backend In control group, the subjects are surveyed with the web-application only that looks similar to the front-end seen by the test group, but it has no back-end that helps it provide an intervention.

Survey format:
Pre-survey to be conducted with robot interaction system
TBD: Post survey conducted through email/face-to-face

1. Is this pre-workout or post-workout interaction ?
2. What is your name/participant id for identification ?
3. Did you workout today ?
4. How much more workout have you done since our last interaction ?
5. On a scale of 0 - 4 how motivated are you now ?

4 Project Schedule

The following is the schedule, listed by week, that we expect to follow over the course of the semester to complete our project.

- Mar 11 - Shruthi - Modeling the human, ROS (basic tutorial - sample implementation)
  Lakhan - Start researching multi-armed bandit problem, ROS (basic tutorial - sample implementation)
- Mar 18 - Shruthi - inverse reinforcement learning, ROS continued
  Lakhan - Model of the robot decision system based on multi-armed bandit, ROS continued
- Mar 25 - Shruthi & Lakhan - Start coding up ros nodes and web application April 1 - Shruthi & Lakhan - Continue coding ros nodes and web application, work on presentation
- April 8 - Shruthi & Lakhan - Start testing our application - using us as experiment subjects, if we are using some solvers (open-source)/ package use that
  April 9 (Mid-sem presentation)
- April 15 - Shruthi & Lakhan - Fix any bugs/issues in code, Start recruiting human subjects, and finalize questionnaire
- April 22 - Shruthi & Lakhan - Evaluate on humans, start writing research paper
- April 29 - Shruthi & Lakhan - Evaluate on humans, continue writing research paper

5 Risks & Challenges

1. *Completion rate:* Human subjects are not paid - incentive to take the questionnaire regularly (risks of some subjects skipping questionnaires at times)
2. Risk of humans not behaving according to the built human model. Human behavior is very noisy and unpredictable
3. Risk of running into problems due to uncontrollable environmental factors w.r.t workout
4. *Attrition:* May run into risks due to the length of our human evaluation protocol/study (span of a week or more)
5. Concern - Not all survey participants go to the gym to workout at the same time - robot interaction with the participants can get trickier due to schedule conflicts
6. Challenge - System planned is abstract - so no guarantee work-flow will go as per the current plan
7. Challenge - Hard-coded interventions may turn out to be cliche to participants

6 Written Example Of Demonstration

- Person comes in to take the pre-workout questionnaire.
- Based on responses, robot provides an intervention.
- Person continues to workout.
- Post-questionnaire to assess if the intervention worked (for test group).
- Post-questionnaire to record the progress towards goal status (for both control and test group).
- Continue the evaluation study over a predetermined span of a few days.