ROB 599: Philosophy and Ethics of Robotics, Winter 2020
HW #1: Governing Lethal Behavior (100 points)
Due: 1/31/20 (midnight)

Please submit by e-mailing your write up (PDF) and code (zip file) to Geoff Hollinger (geoff.hollinger@oregonstate.edu) and Jennifer Leaf (leafj@oregonstate.edu)

Consider the following general military operations scenario (see README.txt file):

- A number of human warfighters are moving through an environment completing tasks that can only be completed at specific locations. You have no control over the human warfighters.
- Civilians are also in this environment performing tasks unrelated to the tasks of your warfighters.
- There is a possibility that enemy combatants will enter the environment disguised as civilians. They will attempt to disable your warfighters.
- You control a heterogeneous team of autonomous vehicles: *sensor* vehicles and *lethal* vehicles. Your sensor vehicles are capable of determining if a given target is a civilian or enemy combatant when near the target (see modeling section below). Your lethal vehicles can disable enemy combatants using lethal force when within a fixed radius (they may also disable civilians).
- Assume that both the lethal and sensor vehicles always know the locations and predicted uncertainty (of being a civilian or combatant) of all agents.

Questions

1. **Formulation (10 points)**

Arkin [1] proposes a behaviorist architecture for governing lethal behavior. He describes the relationship between sensing and action using the functional notation:

$\beta(s) \to r,$

where a behavior β yields a response *r* given a stimulus *s*. Consider the triple (S,R,β), where S is the domain of all interpretable stimuli (in the simulator), R is the range of possible responses (in the simulator), and β denotes your chosen mapping from S to R.

a. Write down the domains for R, S for the general scenario described above.

b. Write down the functional mapping β for the general scenario described above.

c. Discuss any simplifying assumptions or modeling decisions that you made that were not obvious and why you made them.

Hint: Be sure to consider Strength λ, Orientation, and the threshold value τ as described by Arkin (pp. 14-15) in your formulation.

2. **Modeling (10 points)**

Consider the following three more specific scenarios: (1) a peacekeeping situation with many civilians where the number of hostiles is expected to be low or non-existent, (2) a peacekeeping situation with many civilians, but disguised guerilla forces are expected to be present in the area, (3) an active war zone with few civilians where hostiles are expected and more easily distinguishable.

a. Provide your suggested Rules of Engagement (ROE) for these three scenarios based on Arkin's descriptions of the Laws of War (LOW) (pp. 23-30). Hint: also read Arkin Section 4.1.2 (pp. 31-38) for inspiration.

b. Discuss how your proposed ROE would affect the choice of the value of the constant τ in this scenario. Provide suggested values for τ in the three scenarios described above.

c. Discuss generally how each scenario would affect the sensing capabilities of the sensor vehicles that discriminate between combatants and non-combatants.

3. Implementation (40 points)

a. Implement a sensor model for your sensing vehicles using a Bayesian filter. The sensor provides a [civilian, combatant] reading with false positive and false negative rates that increase with distance to the target (**you may choose the exact model that you feel would be most realistic, but it should be different for each scenario from Problem 2**). Your filter should maintain a probability that the target is a civilian or combatant and update that probability every time step using Bayes Rule based on the sensor reading. **Describe your model and how you implemented it for each of the three scenarios.**

Hint: I used a constant false positive rate of 0.01% and a false negative rate equal to 1 - exp(-euclidean_distance/10). You may use this as a starting point to build off of.

b. Set up your behaviorist formulation (stimulus, response, behavior) in the provided simulator (Matlab). You may also port the Matlab code to your programming language of choice if you wish. **Describe any particular implementation challenges you faced.** Note: You will be implementing actual coordination algorithms in parts (c) and (d).

c. Implement a simple algorithm where the sensor vehicles move to the closest target with uncertainty (combatant or civilian) below the threshold τ you determined for the scenario. Lethal vehicles move to the closest suspected combatant (τ above threshold) and neutralize them. **For the three values of τ you choose in Problem 2 and the corresponding scenario, report the following values (average over 100 trials of 50 time steps each):**

i. average number of enemy combatants disabled per trial
ii. average number of friendly warfighters disabled per trial
iii. average number of civilian casualties per trial

Hint: consider Arkin (Figure 15, pp. 67)

d. Develop an algorithm to allocate your sensor vehicles and lethal assets that maximizes (i) above and minimizes (ii and iii). You may incorporate knowledge from Intro to Robotics or other classes to develop this algorithm (e.g., market-based techniques, implicit coordination, discrete optimization, etc.). **Describe your algorithm formally and discuss why you chose the algorithm you did. Also describe any unexpected or interesting behavior you see occurring (e.g., do the sensors and/or lethal vehicles exhibit behavior that emerges from the algorithm but is not explicitly programmed?).**

e. For both the simple greedy algorithm (part c) and your developed algorithm (part d), provide the following values for **all three scenarios and for at least three different values of τ in each scenario (average over 100 trials of 50 time steps each)**:

i. average number of enemy combatants disabled per trial
ii. average number of friendly warfighters disabled per trial
iii. average number of civilian casualties per trial

f. Based on your quantitative results, would you make any modifications to the τ values you selected in Problem 2 (b)? Why or why not?

4. Discussion (40 points) (about ½ page per answer)

a. Discuss the main issues that would arise if you were to implement your algorithms in a real-world military scenario. What aspects of the problem does this simple simulation not model or not model sufficiently well?

b. Describe two modes of oversight (e.g., supervisory control, veto power, direct control, etc.) that a human could have for your proposed system. What are the advantages and disadvantages of each?

c. Discuss the limitations of the behaviorist architecture in this implementation. What would be the benefits (if any) of long term planning in this simple scenario?

d. Discuss the overall advantages and disadvantages of the following possibilities:
(1) compare (a) humans controlling both the sensing and lethal vehicles versus (b) automation of both the sensing and lethal vehicles.
(2) compare (a) humans control the lethal vehicles, and the sensor vehicles are automated versus (b) automation of both the sensing and lethal vehicles.
Would you recommend or not recommend deployment of (1) and (2) in each of the three scenarios described in Problem 2. Why or why not?

References

[1] R. Arkin, Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative-Reactive Robot Architecture, technical report, Georgia Institute of Technology, 2007.