Reinforcement Learning:

Autonomous Vehicles and Collision Ethics

Andre Dugas

University of Colorado, Boulder

Abstract

In October of 2018, Edward Awad et al. published a study regarding socially acceptable machine ethics. In the study, millions of participants weighed the value of different types of people and animals in a scenario similar to the famous philosophical trolley problem. The study, entitled, “The Moral Machine Experiment”, inspired me to create a reinforcement learning “vehicle” in a grid world state that could test these scenarios and make smart decisions that fall in line with the ethical standard discovered by Edward et al.

The goal of this study is to use a Q-learning reinforcement algorithm to model a scenario in which the car must navigate towards an end goal while avoiding two character obstacles which are valued differently according the study. The grid world will be designed in a way that forces the vehicle to take a route risks hitting one of the two character obstacles. As the vehicle agent learns which route is optimal, I will to observe which obstacle the vehicle risks hitting.

1. A Background on Autonomous Vehicles and Collision Ethics

As the prospect of widespread autonomous driving draws nearer, a myriad of underlying problems have begun to surface. Among them is that of encoding ‘ethical decisions’ in the driving software that guides passengers. Specifically, in the scenario where a car must choose between two dangerous collisions: Either hitting a pedestrian or hitting an object and risking the passengers. In this case, a vehicle should ideally be able to make a logical, hard-coded decision that saves one party, and puts the other at risk of injury or death.

As noted in recent publications on this topic, such as “The Social Dilemma of Autonomous Vehicles” (Bonnefon et al), decisions about this type of scenario must be made before self-driving cars become a global commodity (p. 1573). Yet, due to the subject’s subjective nature, this means that one side of the ethical debate must win out and implement their standards into autonomous vehicles (AV’s). To do this, three different considerations must be made. First, there is the utilitarian point of view to save the maximum number of lives in a collision between varying numbers of people. This idea is widely accepted and, as Anderson et

al states, “machines can follow the theory of utilitarianism at least as well as human beings” (p. 18). Thus, a popular theory that is easy to encode should provide a simple solution to the issue at hand. But the second scenario further complicates the issue. The second consideration is of a scenario where a car must choose between two people of different social status. For example, this could be a pregnant woman crossing a street and an influential CEO in the car. Utilitarianism has a much harder time choosing a side in this scenario. The final consideration stems from the economics behind AV’s. A customer of an AV would most likely prefer that their car makes decisions to keep them safe at all times. This would include the trolley-like dilemma of picking who to put at risk. Naturally, the AV companies would like to best suit their customers. This poses the question of encoding a bias to save the passenger more often that not. Being such a multifaceted issue, it becomes clearer that the trolley problem for AV’s is not one with a simple solution.

https://www.scientificamerican.com/article/driverless-cars-will-face-moral-dilemmas/
Credit: Iyad Rahwan

After considering all of these ethical points of view, I decided to focus my study on the value of life issue, in which a car must make a moral decision between two different members of society. This project does not test utilitarian ethical dilemmas, or driver/passenger biases.

2. Previous Approaches to Machine Ethics in AV’s

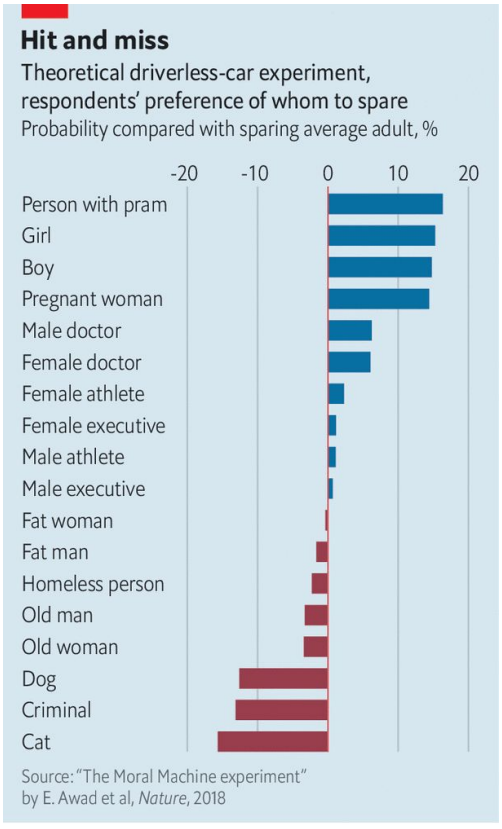
Despite the rapidly advancing design and creation autonomous vehicles, according to publicly available information, the ethical dilemmas posed above have yet to be implemented in any AV software. Perhaps in this innovative intersection of artificial intelligence and transportation, ethics will be one of the last considerations to be made.

Regardless, hypothetical solutions have been suggested for solving the issue. These range from complete randomness of choice in ethical dilemmas to always saving the driver, to using historical data from human drivers who had to make similar decisions.

I believe all of these ideas have strengths and weaknesses, and perhaps the best solution is one that weighs each of these decision types as input and then reacts. However, as noted in section 1, this study only tests one of the proposed solutions (public preference data) to expose the outcomes of one type of policy.

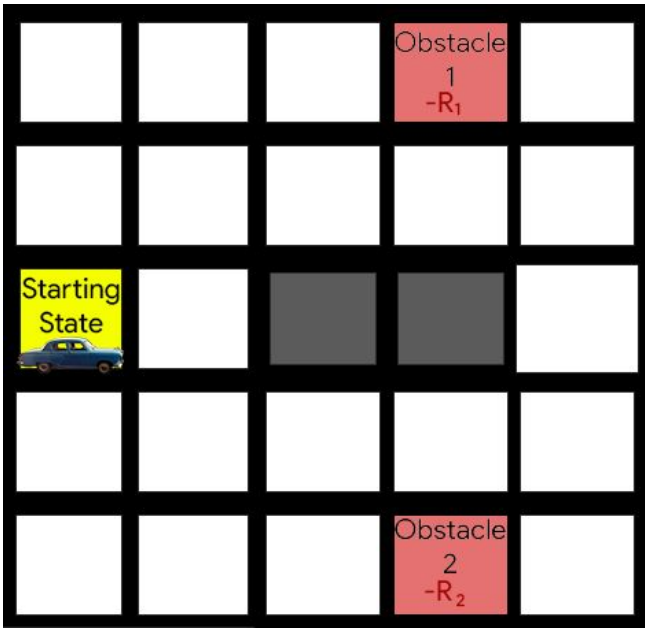
3. Overview of Implemented Artificial Intelligence Methods 1pg

In this study I am testing public preferences of life values to guide an AV's decision-making process. To do this, I have created a grid world environment that will be explored by an agent (representing the AV). This environment contains a goal state for the AV to reach as efficiently as possible. There also exists two pedestrians along the way: these will be the independent variables which alter the car's policy. The AV receives a penalty if it enters into the same state as a pedestrian. My goal is to test a variety of scenarios using different obstacles as independent variables.The correlating penalties will be assigned based on the statistics below, which have been gathered by Awad et al through a public poll of preferences. The obstacles are ranked from top to bottom by highest preference to be saved.



To explore the environment and decide which path to take to the goal, the car will follow a policy that it creates and updates with each decision that it makes. This method is an applied example Q-value reinforcement learning using an Markov Decision Process (MDP).

I have implemented this algorithm in python, using classes to portray the various interacting pieces of my model. These classes include the agent (car), the environment, the states, and the MDP. Each class has methods that will be called upon to gather data and ultimately form an optimal policy for maneuvering through the grid world. It is important to note that unlike the common implementation of a reinforcement learning grid world, this version does not have an end goal that gives the car a desirable reward. This is because the environment is designed to replicate the a trolley problem between two of the members from the above chart. This means that the only way for the car to complete the 'game' is to choose to hit one of the obstacles. Because of this, the obstacle states are called 'end-states', as the game comes to an end when the agent reaches one of these states. To avoid an infinite game in which the agent avoids both obstacles, I award a negative penalty for staying alive, which is given after each move that the agent makes. To create a scenario in which a car would not have an option to escape a dangerous outcome, I will be using a grid world that contains a perimeter of bad outcomes. I further incentivize the car to approach one of these dangerous states by inversely discounting the penalties of hitting the obstacles making them more severe over time.

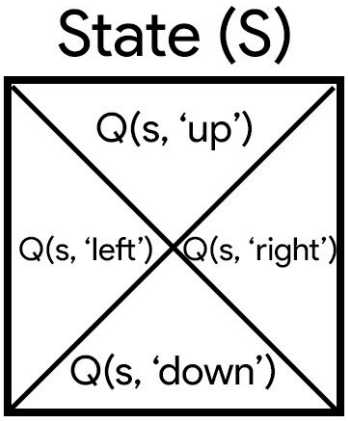


*Penalties are denoted by -R1, -R2 and correlate with the values listed in the above chart

The decisions made by the agent are based on calculations in each state that estimate the potential penalty of taking each available action. The agent is able to move up, down, left, or right at any time, yet if it hits a perimeter wall, it will bounce back into the same state and receive the per-move living penalty. Additionally, each decision made by the agent to move will be

accompanied by a randomness (noise) factor. This represents the uncertainty of the real world and puts the agent at a higher risk of hitting an obstacle if it is in a neighboring state.

The equation used to calculate values of taking actions in each state is called the Bellman Equation, and in because this study is using Q-learning, it has been further expanded into its Q value form. Q-values are used to describe the potential penalty for each available action in a state, and a state's value is then summarized by its maximum Q-values. The equation can be found below along with the parameters and a brief explanation of their effect on the Q-value outcome.



Value of state, V(s)=max Q(s,a)

$$Q(s,a) = R(s,a) + \gamma \sum_{s'} \left(P(s,a,s')V(s')\right)$$

Bellman Equation for Q(s, a)

- Penalty function (R(s,a)) = living penalty (-10) for all states and obstacle penalties for obstacle states
    - The living penalty incentivizes a shorter path solution, as it penalizes each move.
    - The obstacle penalty correlates with the polled value of life from Awad et al study.
- Discount penalty factor (ϒ) = 1.10 (above 1 because rewards are negative)
    - This num ensures that the agent will approach an end state as quickly as possible, as to not unnecessarily raise its final penalty sum.
- Probability function (P(s, a, s')

    - This function represents the noise of the environment. That means that when an action is attempted by the agent, there is only a 90% percent chance that the intended action is made. 10% of the time, a different action will be made at random.
- Value of next State (V(s'))
    - The value of the next state is its max Q-value. This represents the best (least severe) penalty value that the next state has to offer.
    - Recall: The value of the current state is updated by the max of its 4 Q-values after they are calculated.

For a variety of obstacle combinations (chosen at random), I have run 1000 iterations in the grid world and recorded the optimal policies that result.

4. Empirical Results

Unfortunately, my program produced a fatal error that resulted in constant Q-values during each iteration. Despite being unable to debug this error as of December 13, 2018, I can gather a good idea of what the results would show.

When running properly, my model will return a policy that guides the car in the path towards the lowest socially valued obstacle. Changing parameters such as increasing the noise in the environment would further incentivize the agent to create a policy that favors driving near the lower risk obstacle. This is the goal of the reinforcement learning agent because it minimizes risk of hitting the highly valued obstacle. A future update will contain the bug-free version of the project, which will provide empirical data to support this claim.

5. Conclusion and Further Discussion

This project is meant to serve as a small example of the possibility for solving complex ethical dilemmas using artificial intelligence methods such as Q-value reinforcement learning. Artificial Intelligence is a tool to be used in harmony with heuristics that humans decide upon. In this case, the heuristics fall under the topic of ethical decision making and morality. This AV model shows that even for widely debated topic, a reasonable solution can be reached by using artificial intelligence.

Moving forward, I would hope to enhance the clarity of this project, as well as expand its application beyond the trolley problem and into a wider bucket of autonomous vehicle problems. The next step would be creating a new heuristic that factors in the other ethical ideas, such as passenger bias and utilitarianism.

Works Cited

1. Awad, Edmond et al. "'Moral Machine Experiment': Large-Scale Study Reveals Regional Differences In Ethical Preferences For Self-Driving Cars." *Science Trends*, 2018, doi:10.31988/scitrends.41760.
2. Greene, Joshua D. "Solving the Trolley Problem." *A Companion to Experimental Philosophy*, 2016, pp. 173–189., doi:10.1002/9781118661666.ch11.
3. A., Stafylopatis et al. "Reinforcement Learning Based Autonomous Vehicle Navigation in a Dynamically Changing Environment." *European Journal of Operational Research*, 15 May 1996, doi:10.5353/th_b3970738.
4. Anderson, Michael et al. "Machine Ethics: Creating an Ethical Intelligent Agent." *AI Magazine*, vol. 28, no. 4, 2007.
5. Bonnefon, Jean-Francois et al. "The Social Dilemma of Autonomous Vehicles." *ScienceMag*, vol. 352, no. 6293, 24 June 2016.
6. "Whom Should Self-Driving Cars Protect in an Accident?" *The Economist*, The Economist Newspaper, 27 Oct. 2018, www.economist.com/science-and-technology/2018/10/27/whom-should-self-driving-cars-protect-in-an-accident.