Game Theory Based Autonomous Driving Merge Project Final Presentation

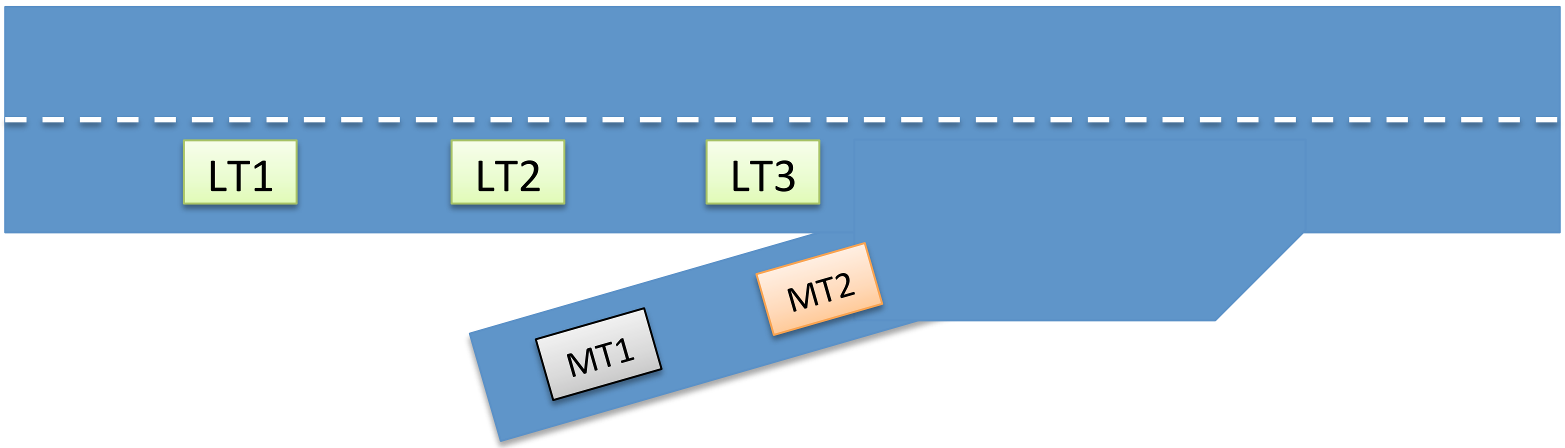
CIS 679, Computational Game Theory, Fall 2016

Larry Schester, 52734824

December 20, 2016

UNIVERSITY OF MICHIGAN-DEARBORN™

Problem Setup



- Project is focused on a vehicle merging into traffic in a controlled environment like a single-lane expressway on-ramp.
- Autonomous vehicle will be traveling on the on-ramp with the goal of merging with approaching in-lane traffic.
- Concept of multi-agent Q-learning that was presented in the lecture was applied to this problem.

Single Agent Q-Learning

- According to [2] (Watkins & Dayan, 1992) Q-learning principles are:
  - Simple agent-based learning method which determines optimal actions in a Markovian domain
  - An iterative approach which learns to improve by repetitive evaluation at particular states
  - Allows for reduced computing power
- A few examples of where Q-learning is used are: cellular phone channel allocation, voice dialog systems, robotic control, and computer vision [1]
- General formulas for single agent Q-learning are:

$$\boxed{Q : S \times A \to \mathbb{R}}$$

Where S represents the set of states, s, and A represents the set of actions, a; leading to the reward.

$$Q(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \left(\overbrace{\underbrace{r_{t+1}}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}}}^{\text{learned value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \right)$$

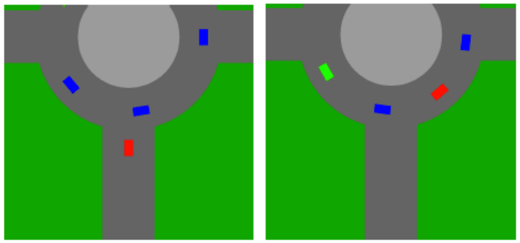
Formulas from: [3] Q-learning, Wikipedia.org, https://en.wikipedia.org/wiki/Q-learning, accessed 3 Dec 2016.

Multi-Agent Q-Learning

- Multi-agent Q-learning is an adaptation of single-agent Q-Learning
- Conceptually, the adaptation is rather straightforward, but there are key differences:
  - Now the environment is no longer considered static through the states, but it has additional agents that also adapt
  - This also means that the environmental states are no longer stochastically random – the other agents act in some *rational* way
- This increased level of more predictable interaction has been shown to outperform single-agent Q-learning (Littman, 1994, Claus and Boutilier, 1998, Hu and Wellman, 2000) [1]
- This increased performance because of the joint action of multiple players make it a game theory based concept; at least one study has been published by Boutilier (1999) [1] about this improved performance in coordination games

Development Steps

- A similar situation was proposed by Mobileye for merging into a roundabout. Their paper [6] discusses the use of game theory based reinforcement learning techniques to solve the problem, but doesn't seem to use them.

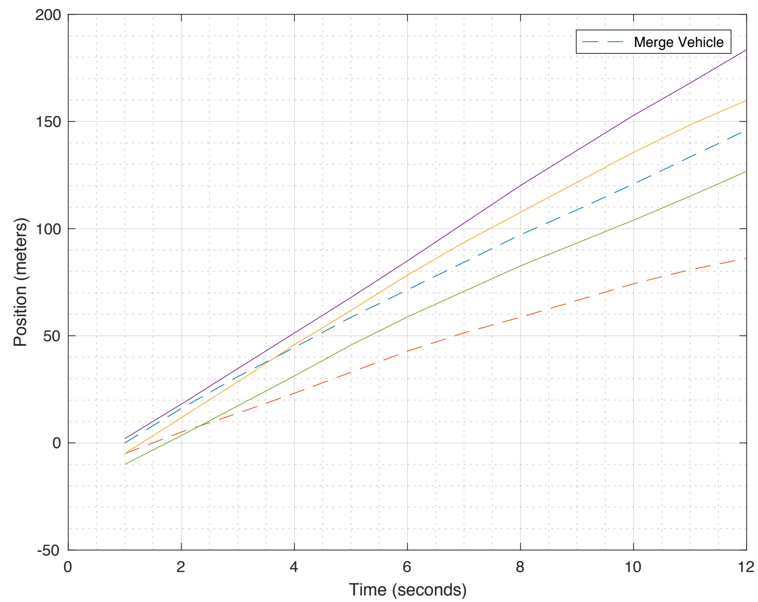


- The approach for the class project was to:
  - Develop a reward strategy where collisions are heavily penalized. Driver comfort is also considered as part of the reward.
  - Create a simulator for the in-lane/merge traffic situation described previously.
  - Utilize game theory techniques, specifically multi-agent Q-learning, to run simulations.
  - Analyze the results.

Program Function and Results

- The plot figure is a single-run simulation with three vehicles in-lane, two merging.
- The simulator runs in time steps, choosing randomly from a set of actions to take at each time-step state.
- This algorithm repeats this process for the number of episodes specified, feeding the information into the Q-learning formula to learn the best outcomes based on the current and next future state.
- The Q-learning formula uses calculations from the algorithm to map values taken from the time-step states into states for closing distance and closing speed, determining the best action for each state (see table for an example).

Simulator Output Plot After Single Run



		Merge Players			
	-5 m, 0 m/s	Decelerate	Maintain	Accelerate	Merge
Lane Players	Accelerate	0.0096	0.0096	**0.9579**	**0.9579**
	Decelerate	**0.9577**	0.0095	0.0096	0.0097
	Maintain	0.0096	**0.9576**	0.0095	0.0096

Example Learned Q-Values After 10,000 Episodes of Training

Next Steps

- The algorithm developed shows that it is feasible to apply game theory based techniques to automated driving.
- Additional game theory solutions, e.g. fictitious play, should be explored to determine which method is best.
- Simulations should be performed using real-world data from sensors used on vehicles today like camera, radar, and LIDAR.

Works Cited

[1] Hu, Junling, and Michael P. Wellman. "Nash Q-learning for general-sum stochastic games." *Journal of machine learning research* 4.Nov (2003): 1039-1069.

[2] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." *Machine learning* 8.3-4 (1992): 279-292.

[3] Q-learning, Wikipedia.org, https://en.wikipedia.org/wiki/Q-learning, accessed 3 Dec 2016.

[4] Hibino, Masato, et al. "Development Tool of Q-Nash Learning Agent for Intelligent System." *Network-Based Information Systems (NBiS), 2015 18th International Conference on*. IEEE, 2015.

[5] Nash, John. "Non-cooperative games." *Annals of mathematics* (1951): 286-295.

[6] "Long-term Planning by Short-term Prediction" by Shwartz, Ben-Zrihem, Cohen, and Shashua from Mobileye, accessed 9/26/2016, www.mobileye.com