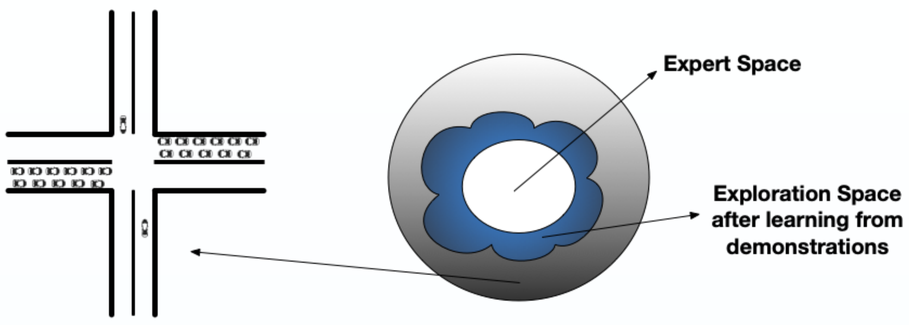
Learning Traffic Signal Control from Demonstrations

Yuanhao Xiong, Guanjie Zheng*, Kai Xu, Zhenhui (Jessie) Li*
xiongyh@zju.edu.cn, gjz5038@ist.psu.edu, kai.xu@tianrang-inc.com, jessieli@ist.psu.edu
Zhejiang University, *Pennsylvania State University, Shanghai Tianrang Intelligent Technology Co., Ltd

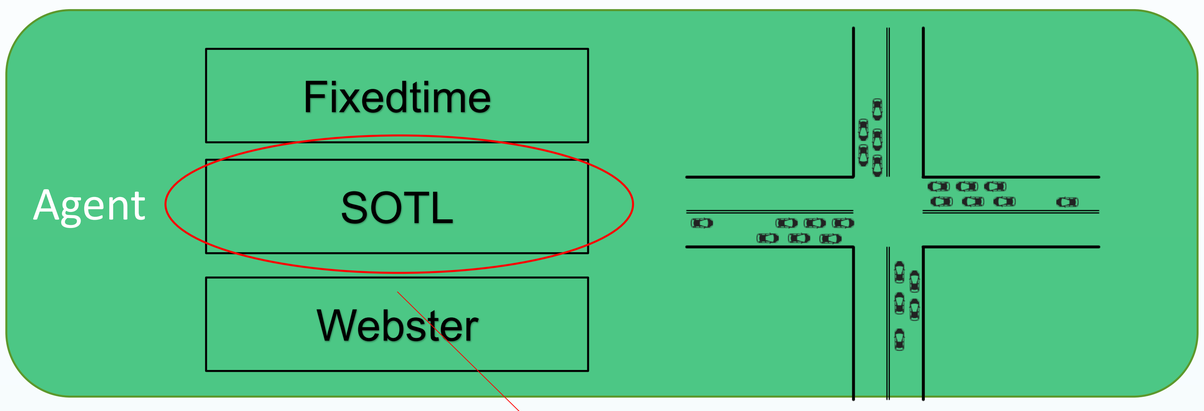
1. Introduction

Large exploration space



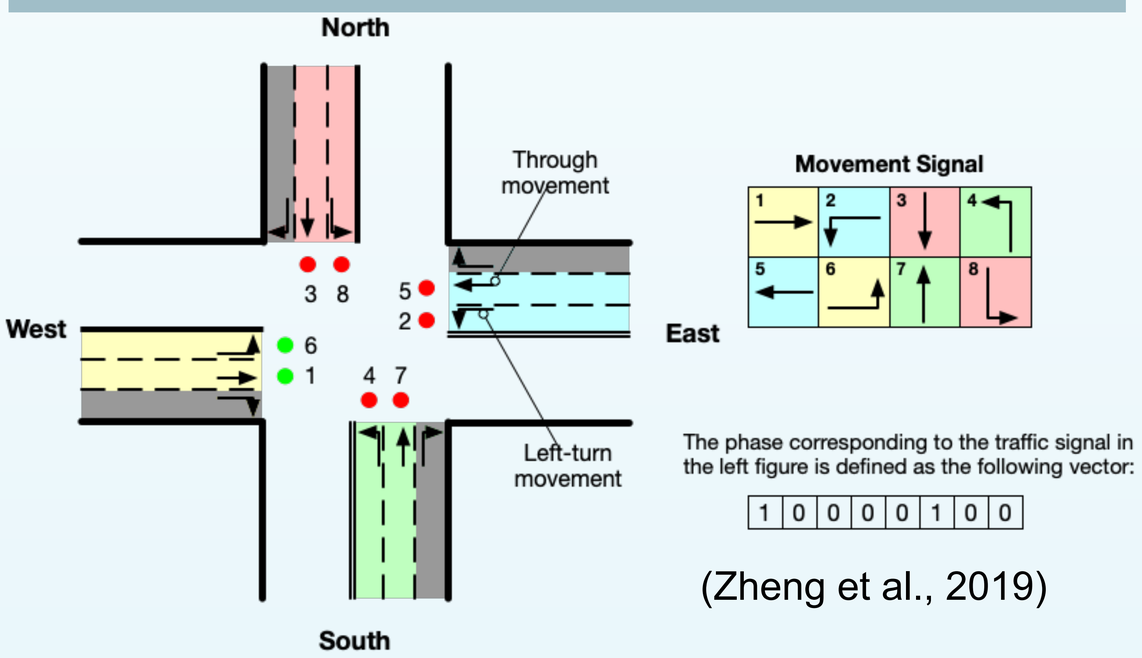
2. Why learning from demonstrations

Experts | **Imitation**



Self-Organizing Traffic Lights(SOTL) is used as the expert in our method. It controls the signal by calculating the number of waiting vehicles. Once the number exceeds a threshold, the signal becomes green for those vehicles to pass.

3. Setting



The phase corresponding to the traffic signal in the left figure is defined as the following vector:

1	0	0	0	0	1	0	0

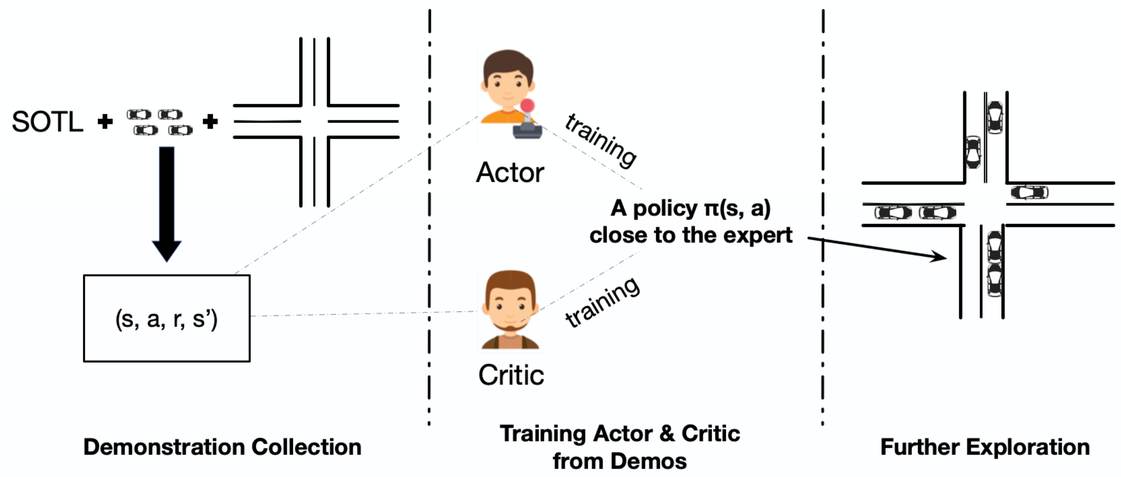
(Zheng et al., 2019)

- **State**: #vehicles on each traffic movement, current phase
- **Action**: the phase for the next time interval
- **Reward**: average queue length

We are implementing our model in Hangzhou China!

Try to find more related researches? Just scan QR code on the right or visit https://traffic-signal-control.github.io

4. Methodology



Figures of actor and critic are from https://www.freecodecamp.org/news/an-intro-to-advantage-actor-critic-methods-lets-play-sonic-the-hedgehog-86d6240171d/

- To make the policy network differentiable

$$a_{soft} = \text{softmax}((g_i + \pi)/\tau)$$

- With the action of demos a_D as the ground truth, the loss function can be derived:

$$L_{demo}(\theta_\pi) = \text{Cross-Entropy}(a_{soft}, a_D)$$

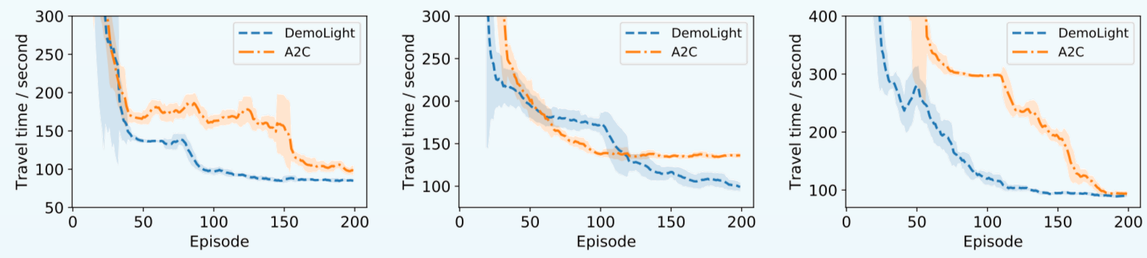
- Following Google's DQfD (Hester et al., 2018), the loss of the value network is computed as:

$$L_{demo}(\theta_Q) = L_{1-TD}(\theta_Q) + L_{n-TD}(\theta_Q) + L_{margin}(\theta_Q) + L_{reg}(\theta_Q)$$

5. Experiment

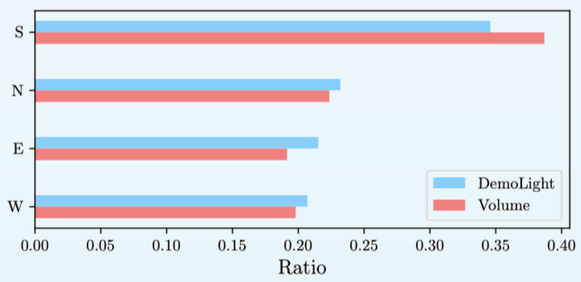
Table 1: Overall performance. Travel time is reported in the unit of second.

Model	City A			City B					City C		
	1	2	3	1	2	3	4	5	1	2	3
SOTL [2]	102.76	179.41	248.73	248.12	153.43	165.38	123.73	269.64	89.36	149.49	72.11
LIT [16]	103.31	122.88	154.77	346.30	146.88	139.83	104.84	551.65	93.95	130.11	48.43
A2C [1]	85.48	98.36	136.01	380.64	92.14	135.93	77.56	517.77	76.38	96.74	46.83
DemoLight	**76.64**	**85.16**	**85.93**	**116.10**	**92.02**	**97.88**	**76.39**	**183.70**	**72.44**	**91.26**	**43.04**
Improvement	10.34%	13.42%	36.82%	69.50%	0.13%	27.99%	64.52%	1.98%	5.30%	5.66%	8.09%

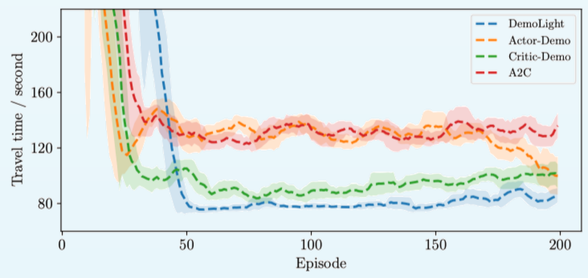


(a) City A #2 (c) City B # 3 (e) City C #2

Convergence Speed



Learnt Policy



Ablation Study

6. Acknowledgement

The work was supported in part by NSF awards #1652525, and #1618448. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing any funding agencies.

References

1. Wei et al., IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control
2. Wei et al., A Survey on Traffic Signal Control Methods
3. Zheng et al., Diagnosing Reinforcement Learning for Traffic Signal Control
4. Wei et al., PressLight: Learning Max Pressure Control to Coordinate Traffic Signals in Arterial Network
5. Zheng et al., Learning Phase Competition for Traffic Signal Control
6. Wei et al. , CoLight: Learning Network-level Cooperation for Traffic Signal Control
7. Xiong et al., Learning Traffic Signal Control from Demonstrations
8. Zhang et al., CityFlow: A Multi-Agent Reinforcement Learning Environment for Large Scale City Traffic Scenario
9. Hester et al., Deep Q-learning from demonstrations