张翔宇工作汇报（2022.11.16）

一、论文阅读

1.

题目：

Deep Reinforcement Learning with Enhanced Safety for Autonomous Highway Driving

摘要：

提出了一种用于自动驾驶的安全深度强化学习系统。提出的框架利用基于规则和基于学习的方法的优点来确保安全。我们的**安全系统由两个模块组成，即手工安全和动态学习安全。**手工制作的安全模块是基于常见驾驶实践的启发式安全规则，可确保与交通车辆的相对间隙最小。另一方面，动态学习的安全模块是数据驱动的安全规则，它从驾驶数据中学习安全模式。具体来说，如果将来的状态之一导致接近失误或碰撞，那么将为奖励函数分配负奖励，以避免碰撞并加速学习过程。我们在不同流量密度的模拟环境中演示了所提出框架的功能。结果表明，通过动态学习的安全模块增强了该策略的性能。

论文整体思路：

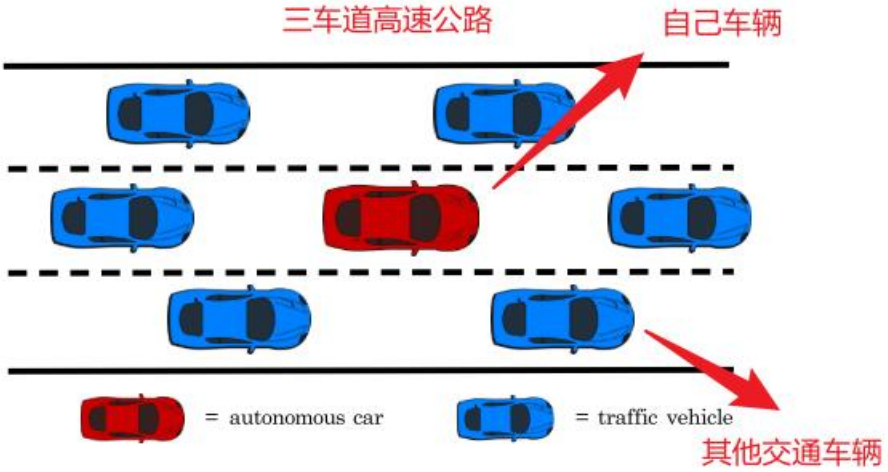
这篇论文主要使用了 $d_{TV} - T_{\min} \times v_{TV} > d_{TV\min}$ 这个启发式规则去判断当前状态是否安全。不满足这个条件就采取

$$a_{\text{safe}} = \begin{cases} \text{Hard brake} & \text{if } T_c \leq T_{hb}, \\ \text{Brake} & \text{if } T_{hb} < T_c \leq T_b, \\ \text{Maintain} & \text{if } T_b < T_c, \end{cases}$$

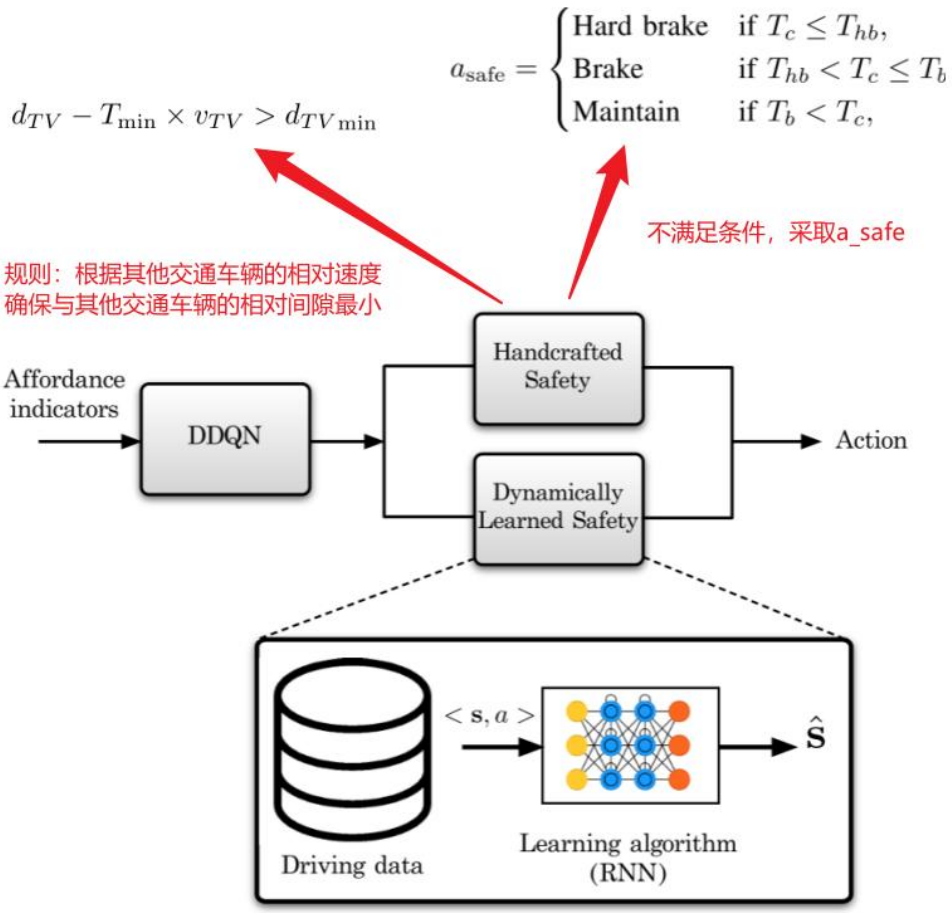
。但是当外部环境变化很快或遇到其他意外情况时，这个规则很难进行处理。为了解决这个问题，这篇文章引入了学习的机制，使用 RNN 预测有限范围内的未来状态，然后确定未来状态之一是否违反安全规则 $d_{TV} - T_{\min} \times v_{TV} > d_{TV\min}$。如果违反了这个规则，就设定负的 reward，以避免碰撞并加速学习过程。

通过与什么机制都不使用、和使用启发式规则，进行对比实验证明，启发式规则+学习机制，累计收益最高。相比于，只用启发式规则，加入学习机制，碰撞次数明显减少了。最终得出结论，加入学习机制，可以减少碰撞次数，提高安全性。

本文场景：


三车道高速公路
自己车辆
= autonomous car
= traffic vehicle
其他交通车辆

整体架构：


$a_{\text{safe}} = \begin{cases} \text{Hard brake} & \text{if } T_c \leq T_{hb}, \\ \text{Brake} & \text{if } T_{hb} < T_c \leq T_b \\ \text{Maintain} & \text{if } T_b < T_c, \end{cases}$
$d_{TV} - T_{\min} \times v_{TV} > d_{TV\min}$
不满足条件，采取a_safe
规则：根据其他交通车辆的相对速度
确保与其他交通车辆的相对间隙最小
Affordance indicators
DDQN
Handcrafted Safety
Dynamically Learned Safety
Action
$< s, a >$
$\hat{s}$
Driving data
Learning algorithm (RNN)

算法伪代码：

Algorithm 1 DDQN with enhanced safety module

Inputs: Trained RNN, prediction horizon k
Initialize: Safe buffer, collision buffer, Q-network, and target Q-network
while not done **do**
  Initialize cars and obtain affordance indicators s
  for length of an episode or collision **do**
    Perform ϵ-greedy and select action a_t
    if action is not safe **then**
      Store $(s_t, a_t, *, -R_{handcraft})$ in collision buffer and replace by safe action
      Apply action, observe the next state, and obtain reward
    if Static collision **then** ▷ Handcrafted safety module
      Reward $\leftarrow -R_{handcraft}$
      Store $(s_t, a_t, *, -R_{handcraft})$ in collision buffer
    else
      Store $(s_t, a_t, s_{t+1}, r_{t+1})$ in safe buffer
    Use RNN to predict $\hat{s}_{t+1}, \hat{s}_{t+2}, \ldots, \hat{s}_{t+k}$ ▷ Dynamically-learned safety module
    if Dynamic collision for *any* future (predicted) states **then**
      Reward $\leftarrow -R_{dynamic}$
      Store $(s_t, a_t, \hat{s}_{t+1}, -R_{dynamic})$ in collision buffer
    Sample random mini-batch $(s_\tau, a_\tau, s_{\tau+1}, r_{\tau+1})$, 50% from safe buffer and 50% from collision buffer
    Set $y_\tau = \begin{cases} r_{\tau+1} & \text{if sample is from collision buffer} \\ r_{\tau+1} + \gamma \hat{Q}\left(s_{\tau+1}, \text{argmax}_a\ Q(s_{\tau+1}, a, \theta_\tau), \hat{\theta}_\tau\right) & \text{if sample is from safe buffer} \end{cases}$
    Perform gradient descent on $\left(y_\tau - Q(s_\tau, a_\tau, \theta_\tau)\right)^2$ w.r.t θ

判断是否满足此规则

$d_{TV} - T_{\min} \times v_{TV} > d_{TV\min}$

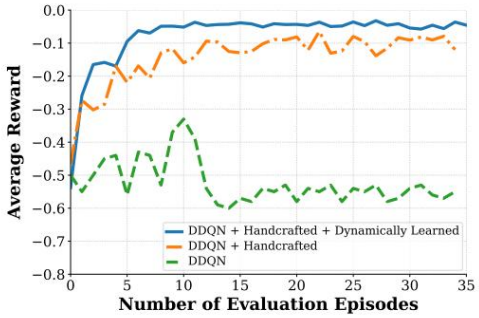
Handcrafted safety module：

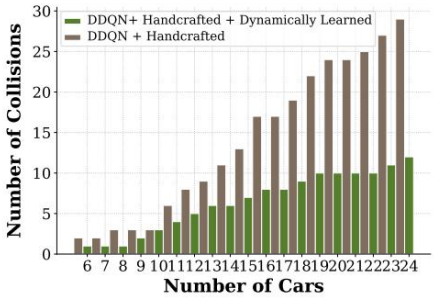
用规则 $d_{TV} - T_{\min} \times v_{TV} > d_{TV\min}$ 判定 ϵ-greedy 策略生成的 action 是否安全，不安全则进行负的 reward，存入 collision buffer；如果判定该行为是安全的，则存入 Safe buffer。

Dynamically-learned safety module：

输入状态和行为使用 RNN 进行下一时刻状态预测——>判断预测出来的状态是否安全（用什么进行的判定? $d_{TV} - T_{\min} \times v_{TV} > d_{TV\min}$ ）——>如果预测出来的状态会导致碰撞，将会对其进行施加负的 reward，存储到 collision buffer 中。

对比实验的结果：





结论：

左图展示出：使用动态学习的安全模块增强的策略优于仅使用手工制作的安全模块的策略，优于无任何安全模块的策略。

右图展示出：相比于只基于手工模块，通过动态学习的安全模块增强的策略导致的碰撞明显减少。整体趋势是随着车辆数量的增加，碰撞次数增加。

文章中的 heuristic safety rule：

判断是否符合此不等式：

$$d_{TV} - T_{\min} \times v_{TV} > d_{TV\,\min}$$

如果不满足此不等式，则采取

$$a_{\text{safe}} = \begin{cases} \text{Hard brake} & \text{if } T_c \leq T_{hb}, \\ \text{Brake} & \text{if } T_{hb} < T_c \leq T_b, \\ \text{Maintain} & \text{if } T_b < T_c, \end{cases} \tag{9}$$

where a_{safe} is the safe action and T_c is the time-to-collision. T_{hb} and T_b represent the thresholds above which the decision made by the RL is considered to be safe.

2.

题目：

Incorporating driver preferences into eco-driving assistance systems using optimal control

近年来，由于对气候变化的担忧，减少燃料消耗，从而减少运输造成的二氧化碳排放，已成为一个热门问题[1]。

经济驾驶行为，或 “生态驾驶”，已被建议作为一种可以通过鼓励驾驶员缓慢加速，预测信号和交通流量，以避免停车，保持平稳的速度，并避免空转的技术，减少道路车辆 10%的二氧化碳排放[2]。

针对这两个问题的潜在补救措施是向驾驶员提供有关其行为的主动反馈，例如通过在车辆内使用听觉，视觉或触觉人机界面 (HMIs) [8]，[9]。使用 V2X 通信 [10] 或机器学习 [11] 等方法改进这些接口有相当大的研究兴趣。

[1] M. P. Vandenbergh, J. Barkenbus, and J. Gilligan, “Individual carbon emissions: The low-hanging fruit,” UCLA L. Rev., vol. 55, p. 1701, 2007.

[2] J. N. Barkenbus, “Eco-driving: An overlooked climate change initiative,”Energy Policy, vol. 38, no. 2, pp. 762–769, 2010.

二、实践

1. 使用 K-Means 算法对之前暑假的数据进行聚类。

1.2　驾驶行为

在未实现高级自动驾驶之前，不同智能化、网联化等级的 CAV 将与人类驾驶汽车组成混合式交通，共享有限的道路资源，因此，驾驶人的驾驶行为与车辆燃油消耗密切相关。驾驶行为受多种因素影响，包括驾驶人性别、驾龄、职业、身体状况、性格、情绪、宗教信仰等生理与心理状况[23]。这些因素的差异决定了不同驾驶人在受到相同外部环境影响时做出不同的驾驶决策。Devlieger 等[24]定义了温和驾驶(速度为 1.00～1.45 $m \cdot s^{-1}$)、正常驾驶(1.45～1.90 $m \cdot s^{-1}$)和激进驾驶(1.90～2.45 $m \cdot s^{-1}$)3 种与燃油消耗相关的驾驶行为；Zorrofi 等[25]研究了驾驶行为对车辆燃油消耗的影响，指出激进驾驶行为在高速公路上增加了 33%的油耗，在城镇道路上增加了 5%的油耗；Jeffrey 等[26]通过轻型车的试验发现，城市道路中不同驾驶习惯和风格可引起约 30%的燃油消耗变化，高速公路会引起约 20%的燃油消耗变化，通过改变驾驶风格，激进驾驶行为可以减少 20%的燃油消耗，温和驾驶行为可以减少 5%～10%燃油消耗。

暂时设定K=3

使用 K-Means 算法之后，最终聚类结果：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		日期	发动机油温	机油压力	冷却剂温度	扭矩	瞬时油耗	整车负荷率	转速	GPS速度	加速度	坡度	趟次	驾驶风格
2	0	2021-08-03 07:24:54	27.1875	0	27	0.17	3.5	0.54	650.5	0	0	0		[2.0]
3	1	2021-08-03 07:25:00	27.1875	0	27	0.16	3.25	0.51	651	0	0	0		[2.0]
4	2	2021-08-03 07:25:06	27.1875	0	27	0.15	3.05	0.47	651	0	0	0		[2.0]
5	3	2021-08-03 07:25:12	27.1875	0	27	0.14	2.95	0.45	652.5	0	0	0		[2.0]
6	4	2021-08-03 07:25:18	27.1875	0	27	0.13	2.85	0.44	651	0	0	0		[2.0]
7	5	2021-08-03 07:25:24	27.1875	0	27	0.13	2.75	0.42	650	0	0	0		[2.0]
8	6	2021-08-03 07:25:30	27.1875	0	27	0.13	2.7	0.41	647.5	0	0	0		[2.0]
9	7	2021-08-03 07:25:36	27.1875	0	27	0.12	2.65	0.41	651.5	0	0	0		[2.0]
10	8	2021-08-03 07:25:42	27.1875	0	27	0.12	2.65	0.4	652.5	0	0	0		[2.0]
11	9	2021-08-03 07:25:48	28.09375	428	29	0.12	2.55	0.39	651	0	0	0		[2.0]
12	10	2021-08-03 07:25:54	28.09375	428	29	0.12	2.55	0.39	649	0	0	0		[2.0]
13	11	2021-08-03 07:26:00	28.09375	428	29	0.12	2.55	0.38	649	0	0	0		[2.0]
14	12	2021-08-03 07:26:06	28.09375	428	29	0.12	2.5	0.39	650	0	0	0		[2.0]
15	13	2021-08-03 07:26:12	28.09375	428	29	0.12	2.5	0.39	650.5	0	0	0		[2.0]
16	14	2021-08-03 07:26:18	28.09375	428	29	0.11	2.45	0.38	654	0	0	0		[2.0]
17	15	2021-08-03 07:26:24	28.09375	428	29	0.11	2.3	0.35	650.5	0	0	0		[2.0]
18	16	2021-08-03 07:26:30	28.09375	428	29	0.11	2.25	0.35	650	0	0	0		[2.0]
19	17	2021-08-03 07:26:36	28.09375	428	29	0.18	3.45	0.39	610	0	0	0		[2.0]
20	18	2021-08-03 07:26:42	28.09375	428	29	0.19	4.15	0.43	671.5	0	0	0		[2.0]
21	19	2021-08-03 07:26:48	28.09375	428	29	0.11	2.6	0.38	699	0	0	0		[2.0]
22	20	2021-08-03 07:26:54	28.1875	432	32	0.12	2.45	0.38	648	0	0	0		[2.0]
23	21	2021-08-03 07:27:00	28.1875	432	32	0.12	2.5	0.37	697	0	0	0		[2.0]
24	22	2021-08-03 07:27:06	28.1875	432	32	0.11	2.4	0.35	696.5	0	0	0		[2.0]
25	23	2021-08-03 07:27:12	28.1875	432	32	0.16	3.6	0.85	718	0	0	0		[2.0]
26	24	2021-08-03 07:27:18	28.1875	432	32	0.17	4.2	0.63	722.5	0	0	0		[2.0]
27	25	2021-08-03 07:27:24	28.1875	432	32	0.28	2.9	0.54	579	0	0	0		[2.0]
28	26	2021-08-03 07:27:30	28.1875	432	32	0.09	2.05	0.32	700	0	0	0		[2.0]
29	27	2021-08-03 07:27:36	28.1875	432	32	0.04	4.2	0.15	863	0	0	0		[2.0]
30	28	2021-08-03 07:27:42	28.1875	432	32	0.13	3.75	0.44	881.5	0	0	0		[2.0]
31	29	2021-08-03 07:27:48	28.1875	432	32	0.16	4.65	0.47	907	0	0	0		[2.0]
32	30	2021-08-03 07:27:54	28.375	500	36	0.22	4.75	0.59	789	0	0.53241	0		[2.0]
33	31	2021-08-03 07:28:00	28.375	500	36	0.28	12.75	0.67	1459.5	11.5	0.08796	2.83353		[1.0]
34	32	2021-08-03 07:28:06	28.375	500	36	0.1	2.35	0.32	692	13.4	0.30556	1.16414		[2.0]
35	33	2021-08-03 07:28:12	28.375	500	36	0.1	2.25	0.33	701.5	20	-0.2778	0		[2.0]
36	34	2021-08-03 07:28:18	28.375	500	36	0.28	1.65	0.93	830	14	-0.213	-1.2474		[2.0]
37	35	2021-08-03 07:28:24	28.375	500	36	0.2	7.9	0.61	1278.5	9.4	0.17593	-1.6788		[1.0]
38	36	2021-08-03 07:28:30	28.375	500	36	0.2	8.7	0.53	1460.5	13.2	0.2037	-1.2068		[1.0]
39	37	2021-08-03 07:28:36	28.375	500	36	0.08	1.55	0.5	667.5	17.6	-0.0093	0		[2.0]

2. 了解 DBSCAN 算法流程

DBSCAN算法

✓ 工作流程：

✎ 参数D：输入数据集

✎ 参数ϵ：指定半径

✎ MinPts：密度阈值

1. 标记所有对象为 unvisited;
2. Do
3. 随机选择一个 unvisited 对象 p;
4. 标记 p 为 visited;
5. If p 的 ε－领域至少有 MinPts 个对象
6. 创建一个新簇 C，并把 p 添加到 C;
7. 令 N 为 p 的 ε－领域 中的对象集合
8. For N 中每个点 p'
9. If p'是 unvisited;
10. 标记 p'为 visited;
11. If p'的 ε－领域至少有 MinPts 个对象，把这些对象添加到 N;
12. 如果 p'还不是任何簇的成员，把 p'添加到 C;
13. End for;
14. 输出 C;
15. Else 标记 p 为噪声;
16. Until 没有标记为 unvisited 的对象;

3. 和师兄讨论，订正了论文整体框架。

三、遇到的问题

1. 对驾驶行为识别方面，整体实现思路不是很明确。使用聚类进行驾驶行为识别，很难细致到具体行为。正常来说，使用聚类可以划分类别（激进、温和、正常），这样的话能称为驾驶行为识别吗？
2. 整体论文最后落点到生态驾驶上，是想通过对细致的、具体的行为进行油耗预测分析，从而通过建议等落点到生态驾驶上。如果使用聚类只划分了三种不同的驾驶风格的话，最后落点到生态驾驶上，分析风格太过于笼统？
3. 安全性驾驶，论文中的

A. Handcrafted safety module

The handcrafted safety module is based on well-known traffic rules that ensure a minimum relative gap to a traffic vehicle based on its relative velocity,

?

$$d_{TV} - T_{\min} \times v_{TV} > d_{TV\min}, \tag{8}$$

where d_{TV}, v_{TV} are the relative distance and relative velocity to a traffic vehicle, $T_{\min}$ is the minimum time to collision, $d_{TV\min}$ is the minimum gap which must be ensured before executing the action choice. If this condition is not satisfied, an alternate safe action will be provided. Specifically,

四、未来计划

1. 和老师、师兄师姐讨论，解决上述问题。
2. 对新的数据进行整理和处理。