自动驾驶战术决策中规划与深度强化学习的结合

汇报人：李泽漩

目 录

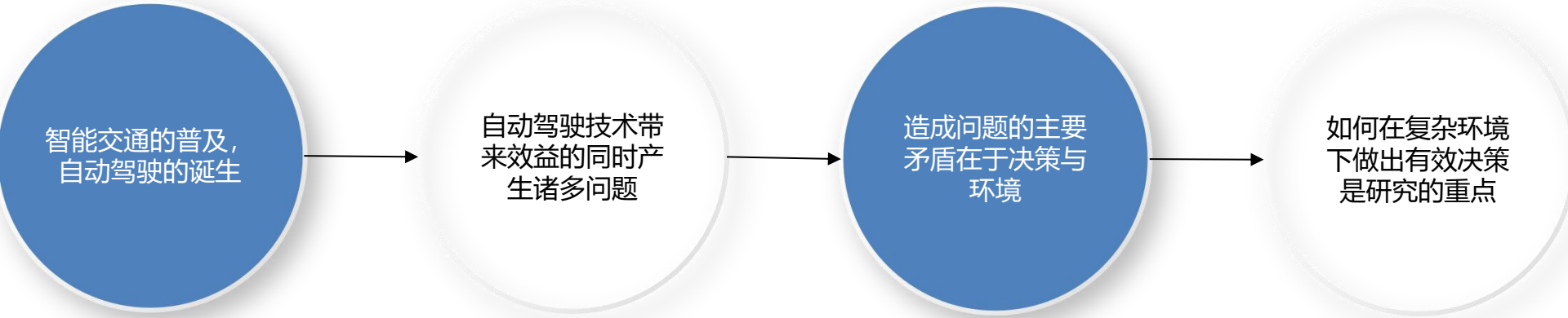
CATALOG

PART 01

第一部分
研究背景

课题背景及内容

CLICK TO INPUT YOUR TITLE



PART 02

第二部分
研究现状

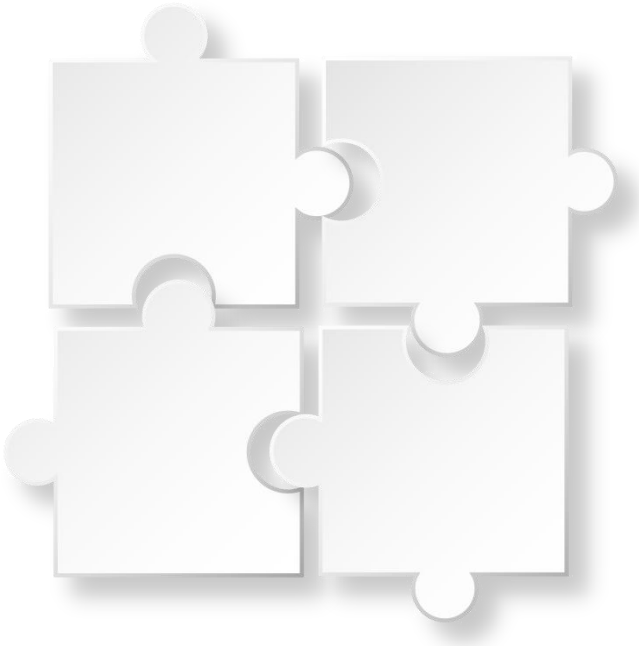
研究现状

01.基于规则的方法

从数据中发现结构，并基于发现的结构进行聚类，将聚类结果用于后续的数据分析。

02.运动规划问题的方法

将决策行为看做运动规划问题，在给定的位置A 与位置 B 之间找到一条符合约束条件的路径。



03.蒙特卡洛树搜索

通过某种“试验”的方法，得到这种事件出现的频率，或者这个随机变数的平均值，并用它们作为问题的解。

04.强化学习

用于描述和解决智能体在与环境的交互过程中通过学习策略以达成回报最大化或实现特定目标的问题

研究现状

基于规则的方法缺乏推广到未知情况的能力，这使得很难将它们扩展到现实世界驾驶的复杂性。

蒙特卡洛树搜索方法由于现实中存在大量可能的场景，且有限的计算资源降低了解决方案的质量，无法预先计算出通常有效的策略。

运动规划问题的方法，顺序设计首先需要预测周围车辆的轨迹，再相应的规划自我车辆的轨迹，导致在轨迹规划过程中无法考虑相互作用的反应行为。

强化学习方法的缺点是需要大量的训练样本才能达到收敛，同时受到信用分配问题的影响。

PART 03

第三部分
研究方法

研究方法

结合规划和强化学习的特性，通过训练神经网络来引导MCTS到达搜索树的相关区域，同时利用MCTS改进神经网络的训练过程

研究方法

01

在线使用时，计划可以被任何一个合理的决定打断，即使只有一次迭代也会返回所学到的动作，更多的计算时间改善了结果。

02

方法通用，可以适用于不同的驾驶场景

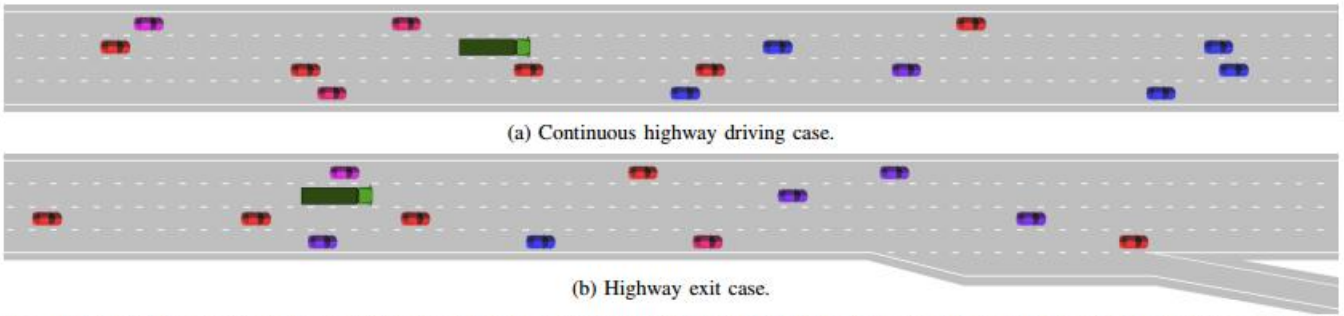
03

在预测时考虑了其他车辆的行驶意图，算法是在连续状态空间运行的。

04

AlphaGo Zero算法扩展到了一个具有连续状态空间的领域，不能使用自对弈。

研究方法



(a) Continuous highway driving case.

(b) Highway exit case.

Fig. 1. Examples of the two test cases. (a) shows an initial state for the continuous highway driving case, whereas (b) shows the exit case, when the ego vehicle is approaching the exit on the right side of the road. The ego vehicle is the green truck, whereas the color of the surrounding vehicles represent the aggressiveness level of their corresponding driver models, see Sect. IV-E. Red is an aggressive driver, blue is a timid driver, and the different shades of purple represent levels in between.

a图为公路连续驾驶案例，b图为高速公路出口案例

研究方法

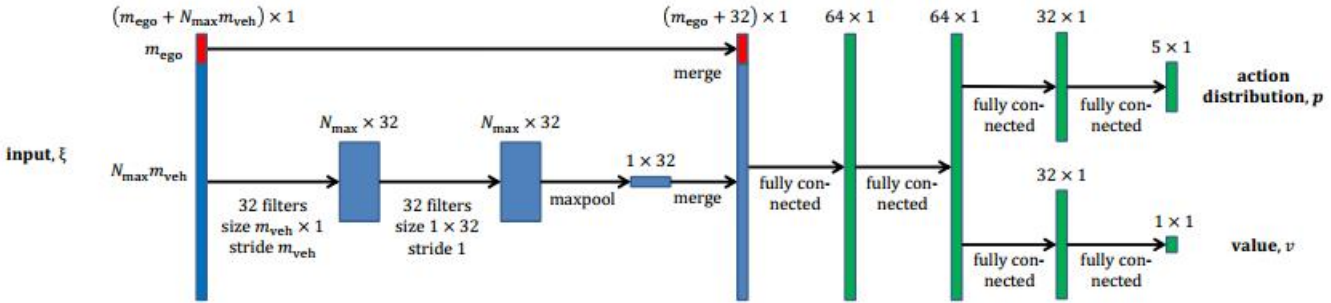
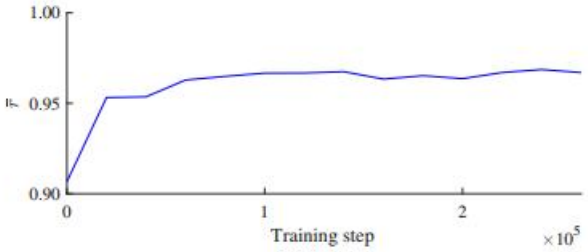


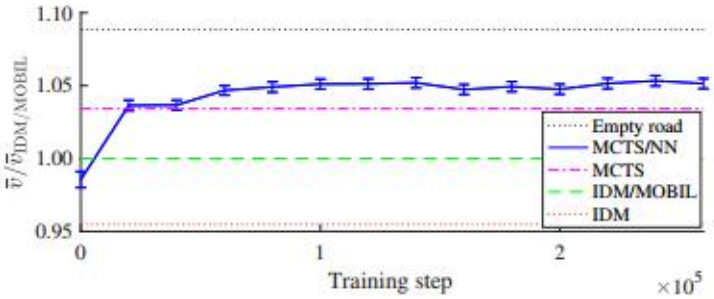
Fig. 2. The figure illustrates the neural network architecture that was used in this study. The convolutional and max pooling layers create a translational invariance between the input from different surrounding vehicles, which makes the ordering and the number of vehicles irrelevant.

神经网络架构

研究方法

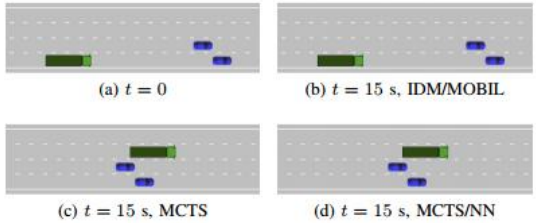


该图显示了在训练过程中每一步所获得的训练值。其中，每一步的最大可能奖励为1，当智能体偏离预期速度或改变车道时，回报会减少。在案例中，经过了20000步的训练，智能体的表现明显提高了。随着训练的增加，获得的平均奖励会略有增加，在大约10万步左右，表现会稳定下来。

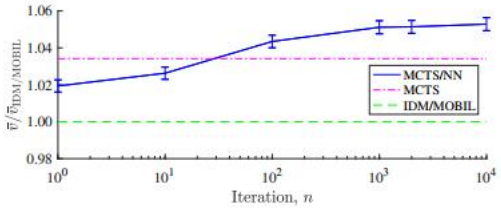


该图显示了应用IDM时的平均速度，整个过程IDM始终保持在其原始车道上，因此可以认为是最低性能。此外还有道路空无一人时的理想平均速度。从图中可以看出，标准MCTS模型的表现优于IDM/MOBIL模型，而MCTS/NN模型很快与MCTS模型的表现相匹配，并在大约6万步的训练后超过了MCTS模型。

研究方法

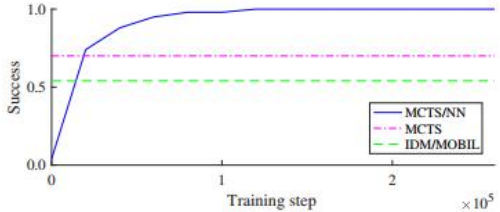


该图显示了模型中需要规划的情况。自我车辆被放置在相邻车道上的两辆慢速车辆后面，驾驶员被设定为谨慎行驶。自我车辆的最佳行为是向左变道两次，以便超车，标准MCTS模型和经过训练的MCTS/NN模型都解决了这种情况，而IDM/MOBIL模型则被卡在原始车道上。

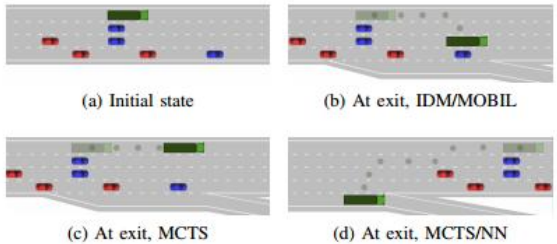


该图为MCTS迭代次数n对训练性能的影响，仅执行一次迭代其性能即优于IDM/MOBIL。在大约30次迭代时，MCTS/NN超过了使用2000次迭代的标准MCTS，并且在n = 1000时性能稳定。

研究方法



高速公路出口情况涉及到通过出口与不通过的情况，所以其目标为到达出口，其次是通过出口。在大多数情况下，模型都能学会如何到达出口。标准MCTS基线方法成功率为70%，改良IDM/MOBIL基线方法成功率为54%。

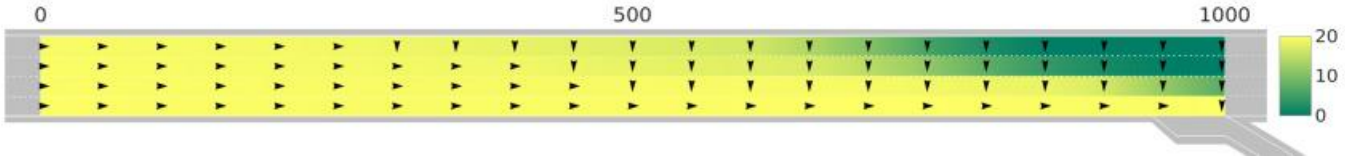


该图显示了规划的情况。自我车辆从离出口300米的最左边车道出发，6辆车位于其他车道，其中3辆车的驾驶员设定为胆小型，其余3辆车的驾驶员设定为进攻型，所有车辆以21米/秒的初始速度启动。在这种情况下，自我车辆到达出口的唯一方法是首先减速，然后向右变道，这只有经过训练的MCTS/NN模型才能实现，标准的MCTS模型在树状搜索中未找到出口，一直停留在原车道上。IDM/MOBIL代理加速到25米/秒，并在没有到达出口的情况下尽可能地向右改变车道。

PART 04

第四部分
实验结果分析

实验结果分析



该图显示了在没有其他车辆存在的情况下，接近出口时不同状态下的学习值和所采取的行动。如预期所示，对于靠近出口的状态，除了最右边的车道之外，所有车道的学习值都在减小。在远离出口的地方，模型总是选择保持在当前车道并保持当前速度，而在靠近出口的地方，模型则向右变道，使其进入最右边的车道。

实验结果分析

本文的研究结果表明，该框架将规划和学习相结合，可用于创建自动驾驶战术决策代理。对于两个概念上不同的高速公路驾驶案例，生成的智能体比单独使用MCTS形式的规划或单独以训练神经网络形式的学习表现更好。该模型的性能也优于基于IDM和MOBIL模型的基线方法。所提出的框架是灵活的，可以很容易地适应其他驾驶环境。

PART 05

第五部分
思考与总结

思考与总结

本文中将规划与深度学习相结合并利用卷积神经网络CNN进行训练的研究思想值得学习借鉴，因为采用规则的方法与单独采用蒙特卡洛树搜索的方法一般情况下不能达到预期效果，采用深度学习又需要大量的训练样本，通过训练神经网络将所需训练的范围缩小到某一区域从而减少了训练量。
顺着该研究思想，可以将其推广至类似的项目中，例如自动驾驶与人车交互相当于将本文中的研究模型里的车车交互改为人车交互，或者是自动驾驶车辆eHMI影响下行人群体过街认知-决策-行为模型研究中用于车辆自动驾驶的决策生成。

请老师批评指导