

## 摘要

随着近年来自动驾驶技术的飞速发展，自动车（Automated Vehicles, AVs）或将成为未来人类的主要通行工具。而在全民自动驾驶时代真正到来之前，未来的 10~50 年仍将是人驾车（Human-driven Vehicles, HVs）与自动车共存的过渡时期。因此，研究自动车在混合交通流中的行驶策略及探究是否存在最佳自动车渗透率，使自动车以最快速度穿梭于道路的同时，更好地提升道路交通系统的通行能力是本文的研究重点。

首先，在 MDP（Markov Decision Process，马尔可夫决策过程）框架下，应用强化学习技术使自动车根据历史经验进行行驶策略的自我学习。自动车的训练过程基于表格式 Q 学习。具体地，自动车将采用 $\epsilon$ -贪婪搜索策略选取即时动作，令其朝着“利己”方向（即在保证安全、平稳的情况下使得自身速度尽可能大的方向）行驶。在该学习模式下进行学习的自动车不仅考虑即时奖励，还将未来收益计入权衡与预估范围内，使得自动车具有“长远”目光。训练完成的自动车将同时具有“预见”能力和“见缝插针”能力，以期更好地利用道路资源，提高其运行效率。

通过微观交通仿真，验证了自动车的“利己”策略在自由流、临界密度和低拥堵场景下的有效性，得到了不同场景下自动车的行驶策略规律。同时，针对自动车和人驾车对混合流的影响，得出了以下几点结论：1) 一定范围内自动车比例的增加可以提升混合流通行效率和稳定性，纯自动车交通流的通行能力相比纯人驾车流提升了 33.55%；2) 70% 在一定程度上可作为最佳自动车渗透率的参考比例；3) 在高自动车渗透率（90%~100%）下，仅少量人驾车就对混合流流量、平均行驶时间、拥堵程度有较大影响。在混合流中占比 0.33% 的人驾车对系统流量的影响约为其自身比例的 13 倍。

探究自动车利己策略在不同交通场景中的适用性以期削减不同场景中自动车的训练次数是本文的一个拓展研究内容。结果表明，中间密度的训练结果在低自动车渗透率下对其他密度完全适用，在中低自动车渗透率下对其他密度较适用。并且，中间自动车渗透率的训练结果仅对较低的自动车渗透率适用；当应用于高自动车渗透率时存在较大误差，特别地，当应用于纯自动车交通流时，误差最大。

本文一方面提出基于强化学习的自动车行驶策略，较好地符合自动车行驶的不确定性与智能性；另一方面，研究自动车对于混合交通流的影响为后续自动车上路后控制其数量、提升自动车速度以及系统效率打下基础。

**关键词：**自动驾驶车辆 利己策略 强化学习 人机混合交通流

## Abstract

With the rapid development of autonomous vehicle technology in recent years, automated vehicles (AVs) may become the main means of transportation in the future. Before the era of autonomous driving truly arrives, the next 10 to 50 years will still be a transitional period in which human-driven vehicle (HVs) and automated vehicles coexist. Therefore, this paper aims to research on the driving strategy of AVs in mixed traffic and explore the optimal penetration rate of AVs, so that AVs could shuttle through the road at its optimal speed, and better improve the capacity of road traffic system.

First, under the MDP (Markov Decision Process) framework, applied reinforcement learning technology to AVs, training them to self-learn driving strategies based on historical experience. The training process of AVs was carried out under the table-format-based Q-learning method. Specifically, AVs adopted the  $\epsilon$ -greedy search strategy to select the immediate action, enabling it move in the direction of "self-interest" (that is, to drive in the direction with the highest possible speed providing stable and safe condition). In this learning mode, AVs not only considered immediate rewards but also counted the future benefit into the scope of trade-offs and estimates, making the automated vehicle own a "long-term" vision. After the training, AVs had both "foreseeing" ability and "seeing stitching" ability in order to make better use of road resources and improve its operating efficiency.

Through microscopic traffic simulation, the effectiveness of the self-interest strategy of automated vehicles in free-flow, critical density, and low-congestion scenarios was verified, and the driving strategy of autonomous vehicles in different scenarios was obtained. At the same time, the following conclusions are drawn regarding the impact of AVs and HVs on mixed flow: 1) The increase in the proportion of automated vehicles within a certain range can improve the efficiency and stability of mixed traffic flow. The traffic capacity of pure automated vehicle traffic is increased by 33.55% compared to pure human-driven traffic; 2) 70% can be used as the reference ratio of the best automated vehicle penetration rate; 3) Under the high automated vehicle penetration rate (90% ~ 100%), only a small number of human-driven vehicles have a greater impact on the flow of the mixed traffic, the average driving time, and the degree of congestion . The impact of 0.33% human-driven vehicles on the overall flow of the system is about 13 times its own proportion.

To reduce the training times, we also explored the applicability of the self-interest strategy of AVs in different traffic scenarios. The results show that the training result of intermediate density scenario is fully applicable to other densities at low AV penetration rates; and are much

applicable to other densities at lower-middle AV penetration rates. Moreover, the training result of the intermediate AV penetration rate scenario is only applicable to the lower AV penetration rates; when applied to the high AV penetration rates, there is a large error, especially when applied to pure AV traffic flow, the error reaches maximum.

On the one hand, this paper proposed automated vehicle driving strategies based on reinforcement learning, which better meets the uncertainty and intelligence characteristics of autonomous driving; On the other hand, investigating the effect of automated vehicles on mixed traffic flow lays the foundation for controlling the number of automated vehicles on the road in the future, increasing the speed of AVs and improving the efficiency of traffic system.

**Keywords:** automated vehicles, self-interest strategy, reinforcement learning, man-machine mixed traffic

## 目录

摘 要.....	I
Abstract .....	II
1. 绪 论.....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 国内外研究现状.....	2
1.2.1 元胞自动机交通流模型.....	2
1.2.2 人机混合交通流.....	3
1.2.3 强化学习在自动车行为描述中的应用 .....	4
1.3 论文主要内容.....	5
1.3.1 研究目标和内容.....	5
1.3.2 技术路线.....	6
1.3.3 论文框架.....	7
2. 基于 Q 学习的自动车利己行为建模.....	8
2.1 人驾车建模与强化学习算法.....	8
2.1.1 人驾车建模.....	8
2.1.2 强化学习算法.....	9
2.2 自动车利己行为建模.....	14
2.2.1 混合流系统建模.....	14
2.2.2 状态空间 $S$ 的设计.....	15
2.2.3 动作空间 $A$ 的设计 .....	17
2.2.4 奖励函数 $R$ 的设计 .....	17
2.3 本章小结.....	18
3. 仿真与数值分析.....	19
3.1 仿真环境设置与训练.....	19
3.2 不同自动车渗透率下的混合交通流特性分析 .....	21
3.2.1 基本图 .....	21
3.2.2 时空图 .....	25
3.2.3 拥堵比例 .....	27
3.2.4 平均行驶时间 .....	29
3.3 自动车“利己”策略验证与分析 .....	30
3.3.1 自动车“利己”策略验证 .....	30
3.3.2 自动车“利己”策略分析 .....	31
3.4 场景适用性分析.....	36
3.4.1 同密度, 不同自动车渗透率场景 .....	36
3.4.2 不同密度, 同自动车渗透率场景 .....	36
3.5 本章小结.....	37
4. 结 论.....	39
4.1 工作总结和主要创新点 .....	39

# 北京工业大学毕业设计（论文）

---

4.1.1 工作内容.....	39
4.1.2 结论.....	39
4.1.3 主要创新点.....	40
4.2 未来展望.....	40
致谢.....	41
参考文献.....	42

## 1. 绪 论

### 1.1 研究背景及意义

#### 1.1.1 研究背景

##### （1）自动车的发展前景

自动驾驶汽车（Automated Vehicles, AVs，以下简称自动车），又称机器人汽车，它是一个包含环境感知、行为决策、行驶计划和智能控制的综合系统<sup>[1]</sup>。相较于手动驾驶汽车（Human-driven Vehicles, HVs，以下简称人驾车），自动车无需人为操作，且可根据实际道路情况自动调节行驶速度。随着过去二十年来传感和计算机技术的飞速发展，自动驾驶技术已日益成熟<sup>[2]</sup>。因此其有望逐步替代人驾车，使驾驶员在完全摆脱驾驶负担的同时，也大大减少由人为错误引起的交通事故<sup>[3]</sup>。

除能保障驾驶安全外，自动车还具有减少交通堵塞、降低出行成本、提高燃油效率并降低碳排放量、提升土地利用率等优势<sup>[4]</sup>。而提高道路通行能力在这其中受关注较高。通行能力是评估道路交叉口容量与通行效率的重要指标，如何提高交叉口通行能力是城市道路规划、建设和管理追求的目标<sup>[6]</sup>。由于自动车可通过减少反应时间、建立与周围车辆或环境的联系而有效地提升通行能力，因此研究者们希望通过构建基于自动车的网络来最大化道路通行能力，并减少车辆行驶时间<sup>[7]</sup>。

##### （2）未来人机混合驾驶的展望与研究的必要性

尽管近年来自动驾驶技术发展迅速，但其仍远远不能在短时间内完全取代人驾车。根据现有的自动驾驶技术的部署和采用，估计在未来的 10~50 年内，自动车将和人驾车共存<sup>[10]</sup>。因此，研究自动车与人驾车的交互对于今后（即自动车的时代）在交通控制和交通规划领域应用新兴技术以更好地服务出行者至关重要<sup>[12]</sup>。

##### （3）自动车研究技术的演变

交通流理论是研究交通科学与技术的基石。为研究和建立自动车的跟驰模型，近年来研究者们探索了较多的可能性，从传统的跟驰模型（如基于刺激的模型、基于安全距离的模型等）到基于数据驱动的跟驰模型，但这些模型在应用于自动跟车时仍具有一定局限性。比如，由于当前大多数传统的跟驰模型都经过简化，即它们仅包含较少的参数<sup>[13]</sup>，而使用较少的参数几乎无法对固有的复杂汽车跟驰过程进行建模。并且，通过经验数据校准的跟驰模型无法推广到未被校准过的交通场景中，导致泛化能力较差。因此，思考如何真正体现自动车的智慧行驶是解决上述模型局限性的关键。

近年来，强化学习（Reinforcement Learning, RL）技术兴起，并在图像识别、自然语言处理等多方面取得重大突破<sup>[22]</sup>。强化学习最初受动物行为训练的启发，训练员通过给予动物奖励或惩罚，以刺激动物形成行为与状态之间的某种反射机制<sup>[14]</sup>。强化学习解决的问题是：智能体（Agent）如何在无先验知识的情况下，仅通过设定给定目标来学习能

达到其目标的最优动作。因此，强化学习是一种免模型的学习方式，它可以很好地体现自动行驶的不确定性与智能性。将强化学习技术作为自动行驶的学习策略不失为一种较好的选择。

## 1.1.2 研究意义

- (1) 探索基于强化学习的自动行驶在人机混合流中的“利己”行驶策略，为后续自动行驶上路后控制其数量、提升自动行驶以及系统效率打下基础；
- (2) 研究人机混合驾驶下的交通流主要参数：速度、密度、流量之间的关系，及人机驾驶行为特征参数间的相互作用机理。为今后应对无人驾驶对交通环境造成的改变和影响具有重要的理论和现实意义；
- (3) 研究自动行驶利己策略在不同交通场景中的适用性，以削减不同场景中自动行驶的训练次数。

## 1.2 国内外研究现状

### 1.2.1 元胞自动机交通流模型

元胞自动机（Cellular Automata, CA）是一种微观数学模型：粒子在一个具有离散、有限状态的元胞组成的元胞空间上，按照一定的演化规则跳跃演化。而元胞（cell）则指被一定形式的规则网格分割而成的单元，其只能在有限的离散状态集中取值<sup>[15]</sup>。

由于交通系统本质上属于离散、非线性系统，因此可以将 CA 模型作为基础模型应用于描述或复现微观交通流的研究中。1986 年，Cremer 和 Ludwig 将 CA 的一个特例（lattice gas automata, LGA）运用到交通流的构建中，首次把 CA 理论知识与交通领域相结合<sup>[16]</sup>。1992 年，Nagel 和 Schreckenberg 提出了最经典的 NaSch TCA 模型，在确定性模型的基础上显示地引入了一个随机噪声，即“随机慢化”规则，以更好地模拟真实的道路情况<sup>[17]</sup>。同年，Biham 等创建了二维模型：BML-TCA，将其应用于城市道路交通路网的建模<sup>[18]</sup>。在此后，学者们多在前人的基础上对模型进行修改提升，以期更好地符合实际交通情况。1993 年，Takayasu 等在 CA-184 号规则的基础上提出了慢启动 ( $T^2$ -TCA) 模型<sup>[19]</sup>，为静止车辆设置了一个加速延迟，以解决前车与不动车辆的不同间距对静止车辆启动的影响问题。1995 年，Nagel 等提出了 STCA-CC 模型<sup>[20]</sup>，在 NaSch TCA 模型的基础上，禁止高速车辆的随机慢化，以解决在自由流中也会出现突发阻塞的现象。2000 年，Knospe 等引入了刹车灯对车辆产生的影响，建立了把驾驶员希望舒适和稳定驾驶需求考虑在内的舒适（comfortable driving, CD）模型<sup>[21]</sup>。2001 年，李晓白等考虑到车辆在  $t \rightarrow t + 1$  时间中前车的相对路程，首次建立了考虑前车速度效应的 VE 模型<sup>[22]</sup>。

车辆换道是车辆在多车道道路上的常见行为。1993 年，Nagatani 首次在 CA-184 模型的基础上，将车辆换道行为引入到双车道 TCA 的建模中<sup>[23]</sup>。Rickert 等在后续将此换

道规则引入 *NaSch TCA* 模型，以对单向双车道交通流进行建模<sup>[24]</sup>。Wagner 等评估 Rickert 等的模型，提出该模型并未很好的考虑交通流的相关因素（如密度反转现象），并因此引入了更加具体的安全限制因素，令车辆在换道时还需考虑目标车道后方的车辆以避免发生碰撞<sup>[25]</sup>。

上述改进使得元胞自动机交通流模型更加的丰富，也为本文对于人驾车的建模提供了理论基础。

## 1.2.2 人机混合交通流

随着自动驾驶技术的兴起，已有部分学者对自动驾驶与手动驾驶混合交通流进行了建模和交互分析。

Arnab Bose 等<sup>[26]</sup>通过分析手动驾驶与半自动车（假设半自动车的车头时距小于人驾车）混合流的  $q-k$  图，得出在相同交通密度下，混合交通的流量大于手动交通的流量。并且，半自动车辆的存在会增加冲击波（shock waves，即当前方车流密度大于后方车流密度时产生的不连续车流）的传播速度，但并不会影响车辆总的行驶时间。该文献还对停走状态下的混合交通流进行了分析：混合交通流车辆在停走状态下经历的平均延迟比纯手动交通流中的短，而停驶过程中处于停止状态的平均车辆数在手动和混合交通流中保持相同。

Jincai Chang 等<sup>[27]</sup>用宏观交通流微分方程模型描述自动车，并采用 CA 规则更新整个系统的车辆。通过改变混合流中的自动车渗透率，得出自动车的存在可以有效提升拥堵路段交通容量的结论。具体地，当自动车渗透率在 55% 时，道路通行能力增加得最为显著。

Danjue Chen 等<sup>[28]</sup>研究了影响人机混合流通行能力的若干因素，并给出了在交通流达到平衡状态时混合交通流运行能力的公式，即通行能力影响因素公式。公式考虑了自动车渗透率、自动车和人驾车的微观/介观特性以及不同的车道政策等因素。

Sina Bahrami 等<sup>[29]</sup>具体分析了自动车数量与混合流通行能力成正相关的原因。提出当只有少量自动车时，通行能力略微上升，原因在于自动车还未形成车队，因此不能较好地受益于 V2V 技术；当自动车渗透率继续上升时，通行能力急剧增加，因为这时自动车可以形成有较小跟驰间隙的车队。

Jiazu Zhou 等<sup>[30]</sup>为捕获混合交通流中车头时距的随机性和异质性，应用高斯混合模型（GMM）对四种类型的车头时距的分布进行建模。并基于车头时距分布模型建立纯交通流（HV 或 AV）和混合交通流的随机基本图（fundamental diagram, FD），探究不同因素对 FD 的影响。结果表明，较大的自动车渗透率有望提升交通流量，而更大的自动车队列强度则可能导致交通流量更多的随机性。但该研究也存在未考虑车道变更行为且仅考虑稳态流的缺陷。

Yangzexi Liu 等<sup>[31]</sup>探究了自动车及其渗透率对异构交通流动力学的影响。通过仿真实验得出相邻车道之间的车道变换频率随着交通密度沿着基本图曲线变化。并且，自

动车对集体交通流特征的影响主要与其变道和跟车时的智能操作有关，且跟车的影响更为明显。

Bokui Chen 等<sup>[1]</sup>对基于元胞自动机模型对混合交通流进行研究，并假设自动车能够知晓其前方多车的速度和位置，得出自动车在预见车辆数为 5 时效率最高。除自动车预见力意外，还研究了自动车渗透率、车辆密度及人驾车随机减速的可能性对道路通行能力的影响。

总结前人对于自动车和人驾车混合流的研究，有以下几点结论：1) 自动车的存在能够提升交通流量，尤其是道路通行能力和自由流速度；2) 可能存在最优自动车渗透率，当超过最优值时，自动车对混合流的影响可能减弱；3) 自动车可以提升混合流交通效率的原因可能在于其可以形成具有较小间隔的车队，但更大的队列强度或将导致车流运行中更大的随机性；4) 自动车和人驾车的不同车道政策对调控整体交通流性能存在一定的作用。

### 1.2.3 强化学习在自动车行为描述中的应用

上述前人的研究中，大多从缩短车辆反应时间、降低车头时距/车头间距等固定参数的设置来将自动车与人驾车区别开来。尽管这些传统的模型具有一定的可行性且为自动车跟车模型的建立提供了理论基础<sup>[32]</sup>，但其在应用于自动跟车时，具有准确性有限、泛化能力差、缺乏自适应更新等局限性<sup>[33]</sup>。因此，为了解决上述问题，学者们开始尝试通过建立数据驱动的自动车模型，来模拟类人的驾驶行为<sup>[34]</sup>。数据驱动的方法比传统模型更灵活，因为它们允许整合那些影响驾驶行为的其他参数，从而产生更丰富的模型。数据驱动的自动车跟驰建模研究可以分为两种主要类型：非参数回归和人工神经网络。然而，这些模型都没有真正从训练数据中学习决策机制，而是通过拟合数据来学习参数估计，因此仍未具有较好的泛化能力。

强化学习是一种以试错的机制设置智能体（Agent），与环境不断进行交互，通过最大化累积奖赏的方式来学习到最优策略的学习方法<sup>[35]</sup>。它可以有效地弥补传统数据驱动模型泛化能力不足的缺陷。近年来，有部分学者将强化学习技术作为自动车的行为学习策略，这符合自动车行为的智能性与不确定性。

Meixin Zhu 等<sup>[33]</sup>提出了基于深度强化学习的自动车跟车模型，它采用速度偏差作为奖励函数，并使用深度确定性策略梯度算法 DDPG 对模型进行优化。结果表明，该模型优于传统的和数据驱动的跟车模型，同时该模型也具有良好的泛化能力。

Jingqiu Guo 等<sup>[36]</sup>对人驾车和 CAV (connected & autonomous vehicles) 混合流在入匝、出匝两个场景中的运行特征进行了研究，并将强化学习技术作为 CAV 的学习方法。结果表明，道路的通行能力和车流平均速度都随着 CAV 渗透率的增加而有显著提升。并且，相邻车道的换道频率随密度的增长也有类似于基本图的曲线形状。这些发现表明，出/入匝对混合流的变道行为有重要影响。

Changxi You 等<sup>[37]</sup>着重于解决交通中自动车的规划问题。首先，设计相应马尔可夫

决策过程奖励函数，并使用强化学习技术确定自动车的最佳驾驶策略。其次，从专家驾驶员数据库中收集了许多示范，并使用逆强化学习技术基于数据学习最佳驾驶策略。仿真结果证明了使用强化学习和逆强化学习技术的自动车的理想行为。

上述学者的研究结果均表明强化学习的令自动车根据奖励函数从试错互动中学习的方法较好地使得自动车学习了符合其特性的行驶策略。然而，虽然现已有较多学者将强化学习或深度强化学习技术应用于自动车的行为学习中，但将自动车置于混合流中，研究其行驶策略等相关方向的文献仍有较大空白。因此，本文将结合元胞自动机交通流与强化学习技术，研究自动车在混合流中的“利己”行驶策略。同时，还将分析自动车“利己”策略对混合流交通特性的影响。

## 1.3 论文主要内容

### 1.3.1 研究目标和内容

#### 1.3.1.1 研究目标

不同于人驾车，自动车的加减速、换道、超车等行驶行为不受驾驶员的习惯、心理状态等主观因素所干扰。强化学习模式下的自动车通过不断与周围动态环境的交互而进行驾驶策略的自主学习，确保车辆可以安全、平稳、智能地行驶，避免手动驾驶反应时间长、适应性弱等缺陷。虽然目前自动车仍未大量普及，但在不久的将来，自动车与人驾车将在道路上共同运行，因此，本文的主要研究目标如下：

- (1) 基于强化学习技术，探究自动车在“利己”目标下的行驶策略，即在不同环境状态下的倾向动作分布；
- (2) 探究自动车渗透率的变化对人驾车和自动车混合交通流的影响，分析基本图、拥堵指数、行驶时间等相关指标；
- (3) 观察和验证自动车“利己”行驶策略的训练结果在不同交通场景下的适用性。

#### 1.3.1.2 研究内容

本文以元胞自动机交通流模型、强化学习技术为理论工具，对双车道自动车-人驾车混合流进行建模与仿真。通过对微观仿真结果的分析，得出自动车的“利己”行驶策略以及该策略对混合交通流特性的影响。具体的研究内容如下：

- (1) 建立基于 CA 与 RL 的双车道自动车-人驾车混合交通流模型

分别分析人驾车与自动车驾驶特性，建立基于 CA 的人驾车模型与基于 RL 的自动车模型。其中，自动车的行驶策略为“利己”，即在保证安全、平稳的前提下，通过采取加减速、换道、超车等动作尽量提升自身速度。

## （2）数值仿真，训练不同场景下自动车的“利己”策略

设置不同交通场景（不同密度或不同自动车渗透率），对自动车进行策略的训练。通过不断地探索-惩罚/奖励-学习，培养自动车的自我学习能力，包括“预见”能力与“见缝插针”能力。提取上述策略，观察其规律。

## （3）改变相关仿真参数，分析自动车“利己”策略对混合流的影响

通过设置不同自动车渗透率、人驾车换道频率、交通流密度，观察混合流流量、平均速度、行驶时间、拥堵程度等的变化。探究是否存在最优自动车渗透率以及人驾车或自动车的存在对混合流交通特性的影响。

## （4）分析并验证自动车“利己”策略适用性

为减少自动车训练次数，探究自动车行驶策略在不同场景下的适用性。具体包括：同密度，不同自动车渗透率的场景与不同密度，同自动车渗透率的场景。若存在通用性，则可节省训练时间，为今后自动车的复杂训练提供便利。

## 1.3.2 技术路线

本文的技术路线如下图所示：

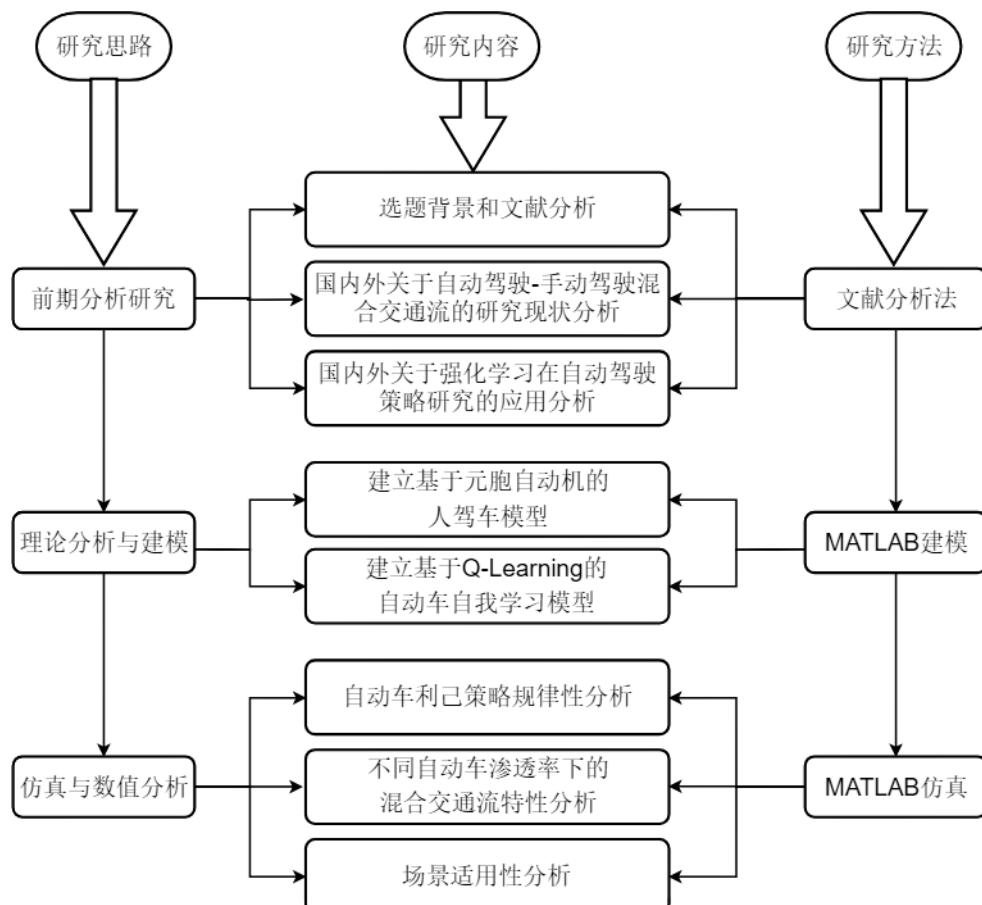


图 1 论文技术路线图

### 1.3.3 论文框架

本文的各章节安排如下：

第一章为“绪论”，介绍本文的研究背景及意义、国内外研究现状、本文的研究目标和内容、技术路线、论文框架、论文的理论意义及实际应用价值；

第二章为“基于 Q 学习的自动车利己行为建模”，首先介绍了基于元胞自动机的人驾车建模方法，其次论述了应用于自动车的 MDP 框架和强化学习算法（Q 学习），最后对自动车利己行为进行具体建模，包括设计描述 Q 表中的状态空间、动作空间和 Q 学习的奖励函数；

第三章为“仿真与数值分析”，对仿真结果进行三个维度的分析，即分析不同自动车渗透率下的混合交通流特性、验证并分析自动车利己策略的有效性和倾向性以及分析不同场景中自动车策略的适用性和通用性；

第四章为“结论”，对本文的工作进行总结，并对未来工作或研究方向进行展望。

## 2. 基于 Q 学习的自动车利己行为建模

本章将重点论述如何采用基于价值的强化学习方法（Q 学习）训练自动车，使其具有自我学习能力，从而按照“利己”策略穿梭于混合交通流中。在这之前，本章首先介绍了基于对称同向双车道元胞自动机模型的人驾车建模方法，并且对马尔可夫决策框架和强化学习算法进行了概述，以对后续自动车的建模打下理论基础。最后，在本文构建的双车道混合交通流模型中，人驾车与自动车的速度和位置更新均在元胞自动机框架下进行。

### 2.1 人驾车建模与强化学习算法

#### 2.1.1 人驾车建模

##### 2.1.1.1 元胞自动机交通流模型概述

元胞自动机交通流（Traffic Cellular Automata, TCA）模型是一类计算效率较高的微观交通流模型。它来自于统计力学，旨在通过对微观交互作用进行最少的描述和定义，来再现正确的宏观行为<sup>[39]</sup>。

TCA 在元胞自动机（Cellular Automata, CA）的基础上，对交通流进行建模，因此其本质上是一个离散的动态系统，即时间随着离散步长而前进，并且空间也是基于粗粒化处理（coarse-gained）的（比如，将道路离散成多个宽为 7.5m 的元胞，每个元胞为“空”或“占据”两种状态中的一种）。

TCA 模型多种多样，通常可首先根据车辆是否可以跨越多个元胞而分为单元胞或多元胞模型<sup>[39]</sup>。单元胞模型又可再细分为确定性模型（Deterministic Models），随机性模型（Stochastic Models）和慢启动模型（Slow-to-start Models）等。

由于本文的重心在于自动车的利己行为建模，因此，为同时兼顾人驾车行为建模的实际性与易操作性，本文参考随机性模型中的对称同向双车道元胞自动机模型（Symmetric Two-lane Cellular Automata, STCA）<sup>[38]</sup>，对人驾车进行行为建模。

##### 2.1.1.2 人驾车状态更新基本规则

参考 STCA 模型框架规则，按  $t$  到  $t+1$  的时间步长对所有驾车的状态进行更新。具体的速度与位置更新步骤如下：

- 1) 计算换道情况
- 2) 加速： $v_n \rightarrow \min(v_{max}, v_n + 1)$
- 3) 确定性减速： $v_n \rightarrow \min(v_n, d_n)$

4) 随机慢化:  $v_n \rightarrow \max(v_n - 1, 0)$

5) 位置更新:  $x_n \rightarrow x_n + v_n$

其中,  $v_n(t)$ ,  $v_{max}$  分别表示当前车辆  $n$  在  $t$  时刻的速度与车辆的最大速度;  $d_n$  表示当前车辆  $n$  距本车道前车的距离;  $x_n$  表示当前车辆  $n$  的位置。

在每个时间步内, 人驾车都将按照上述规则有序地进行状态更新, 并且始终首先计算换道情况 (具体的换道规则将在 2.1.1.3 节进行详细说明)。

加速规则体现了车辆欲保持高速行驶的意图, 但当前车速始终被最大车速所限制。

确定性减速规则保证了车辆行驶的安全性, 即除非车辆进行减速以适应与本车道前车的间距, 否则车辆将保持当前速度。

随机慢化规则考虑到人驾车行驶过程中可能存在的驾驶不稳定性<sup>[42]</sup>, 即在每个时间步内, 从均匀分布中得出一个随机数  $\xi(t) \in [0,1]$ 。然后将该数字与随机噪声参数  $P_{slowdown} \in [0,1]$  (称为减速概率<sup>[39]</sup>, 本文设置为 0.25) 进行比较。结果存在概率为  $P_{slowdown}$  的可能, 车辆将减速到  $v_n - 1$ 。根据 Nagel 和 Schreckenberg, 随机慢化规则捕获了由于人类行为或外部条件变化而引起的自然速度波动。

位置更新规则指出, 车辆根据其当前速度更新其位置。

### 2.1.1.3 人驾车换道规则

换道行为是车辆在多车道环境下的常见行为<sup>[42]</sup>, 驾驶员通过换道行为以期望提升或保持当前车速。基于改进 STCA 模型的换道规则<sup>[38]</sup>, 如果人驾车满足以下情况, 则其将以一定的概率  $p_{change}$  进行换道:

$$d_n < \min(v_n(t) + 1, v_{max}) \quad (1)$$

$$d_{n,other} > d_n \quad (2)$$

$$d_{n+1,other} > v_{max} \quad (3)$$

$$link_{n,other} \text{ is empty} \quad (4)$$

其中,  $v_n(t)$ ,  $v_{max}$  分别表示当前车辆  $n$  在  $t$  时刻的速度与车辆的最大速度;  $d_n$ ,  $d_{n,other}$ ,  $d_{n+1,other}$  分别表示当前车辆  $n$  与本车道前车、相邻车道前车和相邻车道后车的间距。

式 (1) 体现了车辆  $n$  的换道动机, 即当该车与前车的距离无法维持其继续加速或保持最大车速行驶时, 车辆  $n$  将考虑换道。式 (2) ~ (4) 保证了该车换道的可行性与安全性, 即车辆  $n$  在换道前还需考虑与相邻车道的前车是否有相比于本车道前车更大的间距、与相邻车道后车是否有安全间距以及相邻车道相同位置是否有车辆行驶。

### 2.1.2 强化学习算法

在动态交通中, 当前车辆可能会与其他车辆的作用而自然发生状态转换, 从而不受

驾驶员的控制。这一定程度上会与 MDP (Markov Decision Process, 马尔可夫决策过程) 假定的固定策略的原则产生矛盾。但是，即使存在诸多的不确定性，理性的驾驶员仍会计划自身的行驶轨迹，并根据其计划/策略采取行动<sup>[43]</sup>。

本文假设人类驾驶员从基于历史经验中学到的“标准”状态转换概率及其与交通密度的关系等的规划过程来做出这些决策。并且将这种决策模式应用于汽车运行方式中。因此，本文期望通过对历史仿真数据的学习，得出不同场景下的汽车最优行驶策略。而 MDP 则是用以对汽车与周围环境中的车辆之间的交互进行建模的最优选择。本节将首先介绍 MDP 的概念及其在建模中的应用，然后将引入强化学习技术以解决 MDP 问题，从而获得实现所需驾驶行为的最佳策略。

## 2.1.2.1 马尔可夫决策过程（MDP）概述

马尔可夫决策过程（MDP）是一个数学框架，它由 Bellman<sup>[44]</sup>率先提出，并以概率模型模拟主体与环境之间的相互作用。假定智能体（Agent）是与环境互动的学习者或决策者<sup>[45]</sup>。它在每个时间步上除了会更新对环境状态（State）的表示，还会收到环境对其在上一个时间步采取动作的反馈：奖励（Reward）。通过一定的策略对比后，智能体会对环境施加可能改变其未来状态的操作（Action）。智能体与环境之间的这种相互作用如图 2 所示。

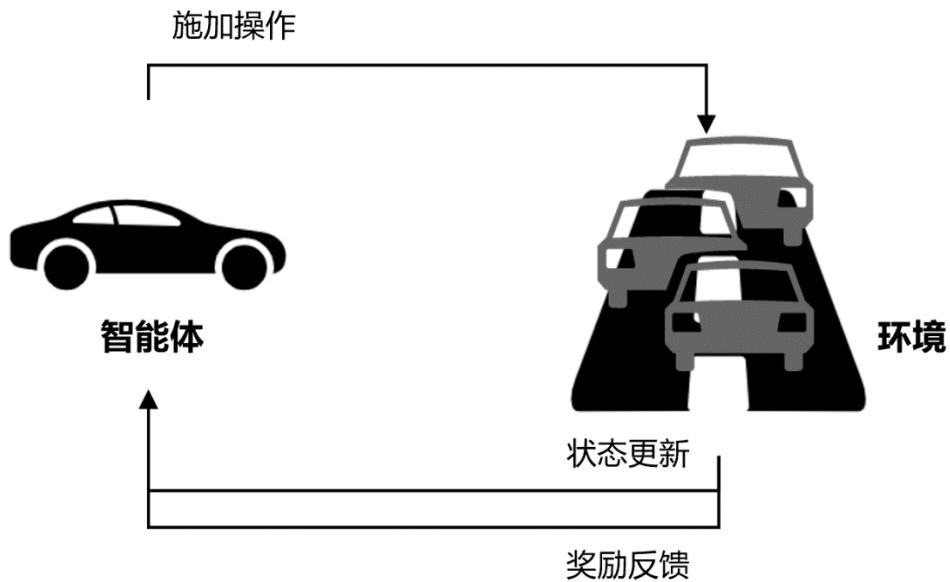


图 2 MDP 作用图

典型的 MDP 由 6 元组  $\{S, A, T, \gamma, D, R\}$  表示，其中  $S$  表示动态环境的（有限）可能状态集， $A$  表示智能体在特定状态下可以选择的（有限）动作集， $T(s, a, s')$  表示状态转移概率矩阵，提供系统在每对状态之间转换的概率， $\gamma \in (0, 1)$  代表保证总收益收敛的衰减率， $D$  是初始状态分布， $R(s)$  表示在当前状态  $s_t$  下，采取动作  $a_t$  后转移到下一状态  $s_{t+1}$  的瞬时奖励奖励函数<sup>[47]</sup>。在本文中，智能体即为汽车，它根据当前状态和直观确定的策略采

取动作。此动作的结果是随机的，并被参数化为转移概率函数。转移到新状态后，决策者会积累与该状态相关的一些奖励。决策过程将无限进行。

通常，术语“马尔可夫”表示给定当前状态，未来状态与过去状态之间是独立的，即马尔可夫性质<sup>[46]</sup>。具体地，MDP 假定在给定状态下执行操作的效果仅取决于当前状态-操作对，而不取决于先前的状态和操作，如式(5)所示。

$$p(s_{t+1}|s_t, a_t, s_{t-1}, a_{t-1}, \dots, s_0, a_0) = p(s_{t+1}|s_t, a_t) \quad (5)$$

MDP 的核心问题是找到智能体的策略 $\pi$ ，该策略 $\pi: S \rightarrow A$  指定在当前状态 $s_t$ 下要采取的动作 $a_t$ 。因此，MDP 的决策目标是找到在无限的范围内最大化累积折扣奖励的最优策略 $\pi^*$ ：

$$\pi^* = \arg \max_{\pi} E \left[ \sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(s_t)) \right] \quad (6)$$

其中， $\gamma$ 为衰减因子， $R(s_t, \pi(s_t))$ 表示智能体在当前状态 $s_t$ 下，根据策略 $\pi$ 执行动作而得到的奖励。

在已知策略 $\pi$ 的情况下，式(4)中的 MDP 将转换成带有状态转移概率 $p^\pi$ 的马尔可夫链（Markov Chain）：

$$p^\pi(s_{t+1}|s_t) = p(s_{t+1}|s_t, \pi(s_t)) \quad (7)$$

### 2.1.2.2 MDP 在建模中的应用

本文将交通流中每辆车（包括自动车和人驾车）的行驶过程定义为无限视野的 MDP。在 MDP 的框架下，在每个时间步  $t$  处，假设环境具有完全可观察性，则可通过观察周围的车辆和道路情况得到观察值 $o_t$ ，并将观察值离散化而获得每个车辆的状态 $s_t$ 。

每辆自动车可根据其策略 $\pi$ 和状态 $s_t$ 在每个时间步提取动作 $a_t$ ，即 $a_t = \pi(s_t)$ 。因此，通过更新交通流中所有车辆的动态即可达到新的状态 $s_{t+1}$ 。由于每辆自动车的当前状态仅受先前状态的影响，因此该过程满足马尔可夫性质。该决策过程如图 3 所示。

由于无法预见人驾车的行驶策略，因此该系统从自动车的角度来看是高度随机的。故采用强化学习方法来学习自动车的最优行驶策略。

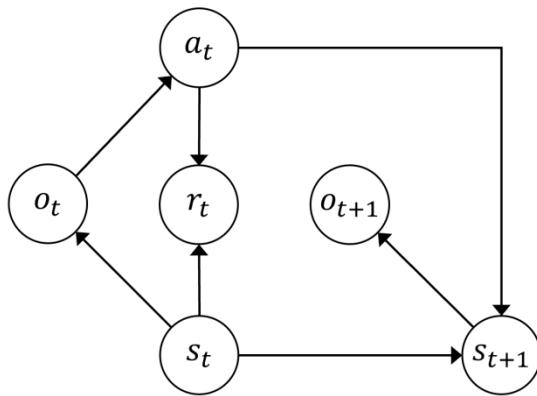


图 3 自动车决策过程图

### 2.1.2.3 强化学习求解方法

强化学习是受到生物能够有效适应环境的启发，以试错的机制与环境进行交互，通过最大化累积奖赏的方式来学习最优策略的一种学习方式<sup>[48]</sup>。强化学习中使用的主要求解方法可分为两类，即表格解法和函数逼近法<sup>[37]</sup>。

表格求解方法适用于解决状态和动作数量有限的 MDP 问题，具体包括蒙特卡洛、时间差分（Temporal Difference, TD）学习、自适应动态规划、策略梯度等<sup>[48]</sup>。而函数逼近法适用于解决大型或连续状态空间问题，通常可能使用一系列（非线性）函数来表示值、策略和奖励函数。从理论上讲，在监督学习领域中使用的所有方法都可以在强化学习中用作函数逼近器，例如人工神经网络<sup>[49]</sup>，朴素贝叶斯<sup>[50]</sup>，高斯过程<sup>[51]</sup>或支持向量机<sup>[52]</sup>。

由于 2.2 节中的 MDP 模型具有有限数量的状态和动作，并且假定智能体无法预测人驾车的行为，因此使用表格法来解决自动车的最优行驶策略问题。

具体的算法采用 TD 学习算法中的 Q 学习算法。TD 学习是动态规划和蒙特卡洛方法思想的结合。TD 学习无需完全了解环境，即无需模型即可实现，因此其优于动态规划算法。并且，相比较于蒙特卡洛，TD 学习是一种在线学习方法，即无需等到每个片段结束后才可得到回报。具体有关 Q 学习的原理将在下节介绍。

### 2.1.2.4 Q 学习

Q 学习是最早的在线学习算法，其同时也是强化学习最重要的算法之一<sup>[48]</sup>。Watkins 博士首先在其 1989 年发表的博士论文中引入了 Q 学习算法<sup>[53]</sup>。Q 学习的主要思路是通过构造 Q 函数，使得智能体选择当前状态对应的具有最高 Q 值的动作，并用执行该动作后得到的 Q 值更新该动作对应的 Q 值。具体的更新公式（Bellman 方程）如下：

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(R_t + \gamma \max_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \quad (8)$$

其中，学习率（步长） $\alpha \in [0,1]$ ，它决定了新获取的信息覆盖当前 Q 值的程度。若  $\alpha = 0$ ，则  $Q(s_t, a_t)$  不变；若  $\alpha = 1$ ，则旧价值将全部替换为新价值；若  $\alpha \in (0,1)$ ，则旧值将沿新值的方向微调旧价值。衰减率  $\gamma$  描述了智能体的未来奖励的重要性。具体地，若  $\gamma = 0$ ，则表示智能体仅在采取行动  $a_t$  后才考虑立即获得的奖励，因此智能体是“近视的”。随着  $\gamma$  逐渐接近 1，智能体将更多地考虑累积未来奖励，即变得越来越“有远见”。通常来说， $\gamma$  不等于 1，以避免智能体过于依赖将来的奖励。

Bellman 方程的运算过程如下：

- 1) 首先，智能体选择 Q 表中当前状态  $s_t$  对应的最大 Q 值所在动作  $a_t$ ，并得到奖励  $R_t$ 。
- 2) 由于价值不仅仅是动作的即时奖励 ( $R_t$ )；而是将来所能得到的最大预期回报。因此，状态/动作对的价值是指智能体刚得到的奖励 ( $R_t$ ) 与智能体将来预计得到的奖励 ( $\max_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1})$ ) 之和。
- 3) 引入  $\gamma$  折算未来奖励，避免智能体过于依赖未来奖励。
- 4) 现在得到的总和 ( $R_t + \gamma \max_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1})$ ) 是状态和动作对  $(s_t, a_t)$  的新价值，将新价值与之前的估计价值 ( $Q(s_t, a_t)$ ) 进行比较得出误差。
- 5) 用误差乘以学习率，从而将价值的旧估计转换为新值 ( $\alpha = 1$ )，或沿新值 ( $\alpha < 1$ ) 的方向微调旧价值。
- 6) 最后，得到的增量值将添加到旧估计值，从而更新 Q 表。

## 2.1.2.5 Q 表法在建模中的应用

本文将 Q 学习与表格法相结合，在 MDP 框架下采用 Q 表法作为自动车的学习方法。Q 表由状态和动作两个维度构成，Q 表中的每个条目都是一个 Q 函数（状态值函数）： $Q(s, a)$ 。Q 表具体的迭代更新过程为：

- 1) 初始化 Q 表中的各个 Q 值；
- 2) 智能体执行动作空间  $A$  中的随机动作  $a_t$  后，进入新状态  $s_t$  并从环境中收取奖励  $r_t$ ；
- 3) 智能体将奖励  $r_t$  作为新信息，根据式(8)贝尔曼方程更新前一状态  $s_t$  和刚采取的动作  $a_t$  对应的价值  $Q(s_t, a_t)$ ；
- 4) 重复 2) ~ 3) 至一个学习片段结束，进入下一个学习片段；
- 5) 重复上述迭代过程直至运行结束，Q 表中的 Q 值收敛到最佳 Q 函数，得到最优策略  $\pi^*$ 。

同时，为了平衡探索（exploration）与开发（exploitation），并充分体现 Q 学习的在线学习性<sup>[42]</sup>，这里采用一种启发式搜索策略，即  $\varepsilon$ -贪婪搜索策略选取即时动作。具体地，车辆  $n$  以  $1 - \varepsilon$  的概率执行 Q 表中状态  $s_t$  的最大 Q 值对应动作  $a_t$ ，以  $\varepsilon$  的概率随机执行

动作空间 $A$ 中的动作。算法 1 中列出了使用 $\epsilon$ -贪婪搜索算法的 Q 学习方法。

---

## 算法 1 Q-Learning using $\epsilon$ -greedy search

---

**Input:**  $S, A, \alpha, \gamma, \epsilon, R$

**Output:** the optimal state action function  $Q^*(s, a)$

```

1:  $Q \leftarrow Q_0$ 
2:  $Q(s_{final}, \cdot) \leftarrow 0$ 
3:  $Converge \leftarrow False$ 
4: while run time < time limit OR episode < episode limit do
5:    $t \leftarrow 0$ 
6:    $s_t \leftarrow s_0$ 
7:   while  $t \leq horizon limit$  do
8:     if  $rand() \leq \epsilon$  then
9:       Randomly choose  $a_t$  from action space;
10:    else
11:       $a_t = \arg \max_a Q(s_t, a)$ 
12:    end if
13:     $r_t \leftarrow$  reward function( $s_t$ )
14:     $s_{t+1} \leftarrow$  traffic model based on  $s_t$  and  $a_t$ 
15:    Bellman update on  $Q$ :  $Q(s_t, a_t) \leftarrow$ 
16:       $Q(s_t, a_t) + \alpha(r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t))$ ;
17:     $t \leftarrow t + 1$ ;
18:  end for
19: end for
20: return  $Q^*(s, a) \leftarrow Q(s, a);$ 

```

---

## 2.2 自动车利己行为建模

如前节所述，本文采用改进 STCA 模型对人驾车进行行为建模，采用基于表格的 $\epsilon$ -贪婪搜索 Q 学习方法作为自动车自我行驶策略学习的方法。本节将首先对人驾与自动车混合流进行总体交通建模，其次，对自动车利己行为进行具体建模，包括设计描述 Q 表中的状态、动作空间和 Q 学习的奖励函数。

### 2.2.1 混合流系统建模

首先，用于建模的 MDP 基于以下观察：考虑到如图 3 所示的双车道道路的典型交通场景，每辆车都以各自的速度行驶在道路中央。

其中，人驾车由于在驾驶员的操纵下行驶，因此其会根据本车与周围车辆的距离，不断地实时调整自身行驶策略（加速、减速或换道），并且考虑到行驶时的诸多不确定因素，人驾车会进行一定概率的突然减速。

对于自动车，本文假设其采取“利己”（即在保证安全的条件下，以自我利益最大化为目标，不考虑对周围车辆的影响）<sup>[42]</sup>策略，因此自动车将通过自学，参考历史经验（Q 表）和当前的环境，选取能使得其最大化“利己”奖励的动作。值得一提的是，本文中的“奖励”不仅为即时奖励，自动车还需考虑未来的得失来达到既定效率的目标，以期

培养一种“长远”目光的能力。而强化学习则是培养该学习能力的较好方法。本文首先令不同场景（不同自动车渗透率、不同车流密度）下的自动车从零学起，即无任何外界的预先知识（初始 Q 表为空），在每个场景下形成一套最优行驶策略。后续在验证场景间的策略适用性时，再令自动车从现有的策略（Q 表）学起。

本文假设所有车辆之间不会相互通信，但自动车间共享 Q 表，即同一场景下的所有自动车以同一张 Q 表为经验学习对象，并根据自身状态不断更新该 Q 表。下面将具体介绍 Q 表的状态和动作空间的设计以及 Q 学习奖励函数的表示。

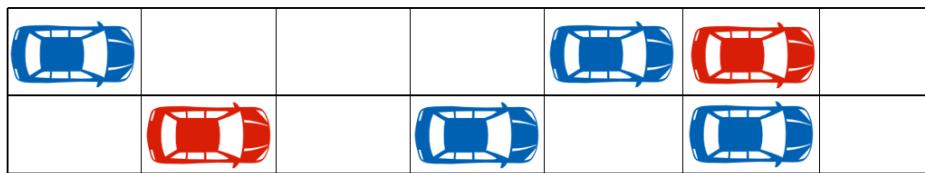


图 3 双车道道路交通场景，其中红车为 AV，蓝车为 HV

## 2.2.2 状态空间 $S$ 的设计

在定义状态空间之前，首先对自动车的观察空间进行描述以便更好地了解其周围环境。由于在本文中自动车间无法进行通信，因此假设自动车具有人类视野，即只能观察到其所在车道与相邻车道前后共四辆车的行为。并且，由于自动车采取“利己”策略，因此其与本车道后车的相对速度和相对距离不在其采取行动而考虑的范畴之内，故只将剩余三车的行为纳入其观察空间中。如图 4 所示，本文定义自动车的观察空间包含七个变量： $\{d_f, v_f, d_{f,other}, v_{f,other}, d_{r,other}, v_{r,other}, lane_id\}$ 。具体地，

- 1)  $d_f$  表示同一车道中，前车在 x 轴方向上与本车的相对位置；
- 2)  $v_f$  表示同一车道中，前车在 x 轴方向上与本车的相对速度；
- 3)  $d_{f,other}$  表示相邻车道中，前车在 x 轴方向上与本车的相对位置；
- 4)  $v_{f,other}$  表示相邻车道中，前车在 x 轴方向上与本车的相对速度；
- 5)  $d_{r,other}$  表示相邻车道中，本车在 x 轴方向上与后车的相对位置；
- 6)  $v_{r,other}$  表示相邻车道中，本车在 x 轴方向上与后车的相对速度；
- 7)  $lane_id$  表示车辆当前所在的车道 ID。

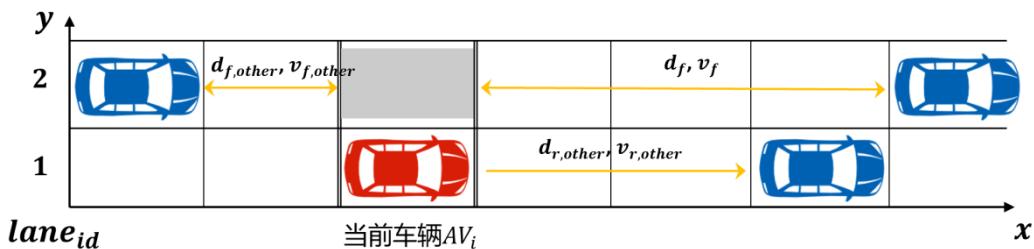


图 4 车辆观察空间示意图，其中红车为 AV，蓝车为 HV

完成对观察空间的定义后，需要将其转换为状态空间。为使状态空间包含有限个状态，且根据人类驾驶员对远近距离和车辆相对关系而非全信息的感知，本文将以上包含连续变量的观察值转换为以下离散值：

$$d_f = \begin{cases} 1, & d_f \leq \text{近距离} \\ 2, & \text{近距离} \leq d_f \leq \text{远距离} \\ 3, & d_f \geq \text{远距离} \end{cases} \quad (9)$$

$$v_f = \begin{cases} 1, & v_f < 0 \text{ (靠近)} \\ 2, & v_f = 0 \text{ (保持)} \\ 3, & v_f > 0 \text{ (远离)} \end{cases} \quad (10)$$

$$d_{f,other} = \begin{cases} 1, & d_{f,other} \leq \text{近距离} \\ 2, & \text{近距离} < d_{f,other} \leq \text{远距离} \\ 3, & d_{f,other} \geq \text{远距离} \end{cases} \quad (11)$$

$$v_{f,other} = \begin{cases} 1, & v_{f,other} < 0 \text{ (靠近)} \\ 2, & v_{f,other} = 0 \text{ (保持)} \\ 3, & v_{f,other} > 0 \text{ (远离)} \end{cases} \quad (12)$$

$$d_{r,other} = \begin{cases} 1, & d_{r,other} \leq \text{近距离} \\ 2, & \text{近距离} < d_{r,other} \leq \text{远距离} \\ 3, & d_{r,other} \geq \text{远距离} \end{cases} \quad (13)$$

$$v_{r,other} = \begin{cases} 1, & v_{r,other} < 0 \text{ (靠近)} \\ 2, & v_{r,other} = 0 \text{ (保持)} \\ 3, & v_{r,other} > 0 \text{ (远离)} \end{cases} \quad (14)$$

$$lane_{id} = \begin{cases} 1, & \text{右车道} \\ 2, & \text{左车道} \end{cases} \quad (15)$$

将相对距离按照近距离和远距离分割成三个区间段，则可将距离的连续值转换为离散值，其中，近距离设定为最大速度，远距离设定为2倍最大速度。同样的，将相对速度按照正在驶离、保持还是远离分割成三个区间段，则可将速度的连续值转换为离散值（如图5所示）。

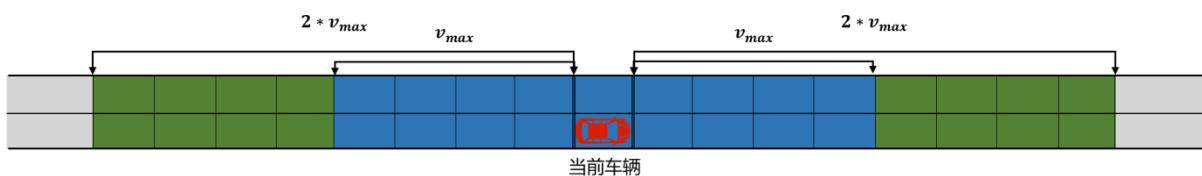


图5 道路分割示意图

由于本文设置道路为封闭边界，因此需要考虑临界状态。规定若前后车辆超出最大视线距离（即4倍最大速度），则认为该前后车处于“远距离”或“远离”状态。因此，自动车共有 $3^6 \times 2 = 1458$ 个不同的离散状态。将每个离散观测向量映射为1至1458之间的整数，则自动车的状态取值为1至1458之间的整数。

### 2.2.3 动作空间A的设计

一般而言，车辆行驶时存在以下八种不同动作：加速、强加速、保持、减速、强减速、换道保持、换道加速、换道减速。由于车辆采取换道行为通常是为了保持或提高车速，因此舍去换道减速这一动作。同时，考虑到在换道的一小段时间内，车辆的速度通常保持不变；且由于本文的时间步长为1s，符合一小段时间的界定，故也将换道加速这一动作舍去。因此，动作空间A包含六个基本动作：

- 1) 强加速,  $a_x = 3 \text{ cell/s}^2$
- 2) 加速,  $a_x = 1 \text{ cell/s}^2$
- 3) 保持,  $a_x = 0 \text{ cell/s}^2$
- 4) 减速,  $a_x = -1 \text{ cell/s}^2$
- 5) 强减速,  $a_x = -3 \text{ cell/s}^2$
- 6) 换道,  $a_x = 0 \text{ cell/s}^2$

为确保自动车在行驶过程中无追尾无碰撞等冲突的产生，还需对自动车添加一定的先验知识（比如当无换道空间时，自动车不可能采取换道操作），从而避免后续仿真过程有效性的缺失。

根据2.1.2.5节的论述，后续将采取 $\varepsilon$ -贪婪搜索策略在动作空间A中选取即时动作。

### 2.2.4 奖励函数R的设计

首先设置奖励函数为本车当前状态的速度与设定速度 $v_{nominal}$ （设为中间速度）之差，以使自动车尽量以较大速度行驶。但只考虑速度之差可能会导致自动车陷入不断地加减速循环之中。因此，为同时保证汽车行驶的效率与平稳性，在原有奖励函数的基础上，加入对于强加减速的惩罚因子 $a$ ，并对两者都配以一定的权重，即

$$R = \omega_1 v + \omega_2 a \quad (16)$$

其中，

- 1)  $v = 0.5(v_x^t - v_{nominal})$ ，表示自动车的当前速度与设定的预期速度之差，由于 $v_x^t - v_{nominal} \in [-2, 2]$ ，因此设定比例系数为0.5；
- 2)  $a \in \{-5, -1, 0\}$ ，表示发动机动作的惩罚因子，目标是期望自动车能够平稳运行，即避免不断的强加减速。因此，对于“强加速”和“强减速”的动作进行惩罚，即赋值“-5”；对于“保持”和“换道”的动作，由于二者的加速度为0，因此不进行惩罚；对于

其他所有动作，赋惩罚值为“-1”。

## 2.3 本章小结

本章首先引入 STCA 模型对人驾车进行行为建模，其次简要概述了 MDP 框架、强化学习原理以及 Q 学习算法，为后续的自动车建模进行理论铺垫。对于自动车的建模，本章采用表格式 Q 学习方法，令自动车朝着“利己”方向（即在保证安全、平稳的情况下使得自身速度尽可能大的方向）进行动作选择。具体地，自动车采用 $\epsilon$ -贪婪搜索策略选取即时动作，该策略较好地平衡了探索和开发，避免自动车陷入局部最优解中。在该学习模式下进行学习的自动车不仅考虑即时奖励，还将未来收益计入权衡与预估范围内，使得自动车具有“长远”目光。

本章在后半部分具体展示了 Q 表的设计过程。对于状态空间，考虑当前车辆的前后共三辆车，为当前车辆建立七维观察向量。为了获取有限的状态数量，本文考虑远近距离划分条件，将连续的观察值转化为离散的状态值，最终得到 1458 个状态值。对于动作空间，本文首先将所有驾驶员动作抽象划分为 8 个一般动作。其次，根据实际情况舍去换道加、减速动作后，剩余 6 个动作，构成自动车的动作空间。对于 Q 学习的奖励机制，本文在考虑速度差的基础上，增加了对于强加减速的惩罚项，避免自动车陷入不断地加减速循环之中。

总体上，本章详述了自动车“利己”策略的设计与训练步骤，为本文的重点章节。经过该训练方法训练后的自动车期望具备自我学习能力，包括“预见”能力和“见缝插针”能力，可以较好地利用道路资源，提高其运行效率。

### 3. 仿真与数值分析

本章将在第二章模型的基础上进行仿真，并对仿真结果进行三个维度的数值分析：首先，分析不同自动车渗透率下的混合交通流特性，以观察自动车和人驾车对混合流的影响；其次，验证并分析不同场景下的自动车利己策略，总结自动车采取策略的规律；最后，分析不同场景中自动车策略的适用性和通用性，以期削减自动车的训练次数。本章末尾验证了自动车速度的提升来源于对策略的自我学习，而不仅是“无随机慢化”。

#### 3.1 仿真环境设置与训练

仿真平台由 MATLAB 语言编写，首先令自动车在不同场景中从零学起，即 Q 表初始化为零。在后续的策略适应性测试中，再令自动车从已经获取的策略中学起，以验证不同场景下自动车策略的通用性。人驾车与自动车在仿真系统中的训练过程如图 6 所示。

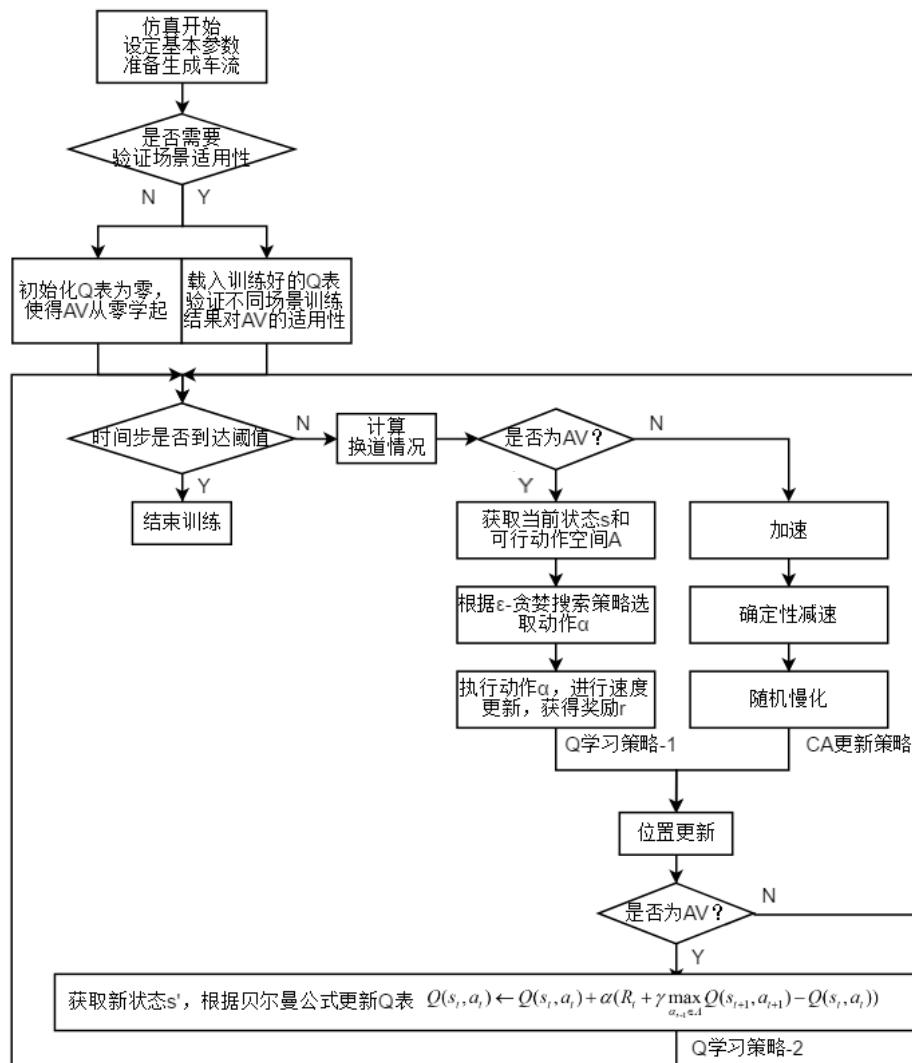


图 6 混合交通流仿真流程示意图

# 北京工业大学毕业设计（论文）

---

以长度为  $L = 10\text{km}$  ( $\sim 1334$  元胞) 的双车道道路作为仿真模拟环境。元胞的物理长度为  $7.5\text{m}$ , 即一辆汽车的长度。设置车辆最大速度  $v_{max} = 4 \text{ cell/s}$  ( $108\text{km/h}$ ), 交通期望速度  $v_{nominal} = 2 \text{ cell/s}$  ( $54\text{km/h}$ ), 人驾车的随机慢化概率  $P_{slowdown} = 0.25$ 。假设二分之一的驾驶员为激进型 (若存在换道空间则执行换道操作), 因此设置人驾车的换道概率  $P_{change} = 0.5$ , 不限制人驾车的换道次数。自动车的学习率  $\alpha = 0.2$ , 衰减率  $\gamma = 0.9$ , 探索概率  $\varepsilon = 0.1$  (仿真参数如表 1 所示)。

表 1 仿真参数及数值设定

参数	数值
道路长度 $L$	10 km
元胞长度 $cell_L$	7.5 m
车辆最大速度 $v_{max}$	4 cell/s ( $108\text{km/h}$ )
交通期望速度 $v_{nominal}$	2 cell/s ( $54\text{km/h}$ )
HV 随机慢化概率 $P_{slowdown}$	0.25
HV 换道概率 $P_{change}$	0.5
AV 学习率 $\alpha$	0.2
AV 衰减率 $\gamma$	0.9
探索概率 $\varepsilon$	0.1
仿真时长 $T$	10000 s
有效仿真时长 $T_{valid}$	5000 s
AV 渗透率 $\beta$	0 ~ 100 %
车辆总数 $N$	0 ~ 1600 辆
车流密度 $\rho_i$	0 ~ 140 veh/km
流量 $Q$	-
平均速度 $\bar{v}'(t)$	-

初始状态下道路车辆数设置为 0, 从  $t = 0\text{s}$  起每  $2\text{s}$  在每个车道的最初位置各生成一辆车 (若该位置有车占据, 则不生成车辆), 车辆的初始速度为最大速度。当每车道生成车辆总数到达设定值  $N$  时, 不再生成车辆。

以周期性边界进行仿真 (即道路首尾相连), 周期性边界可以控制车流的密度, 从而观察密度与速度和流量之间的联系。车流平均密度为每公里每车道平均的车辆数, 车流平均速度为单位时间内所有车辆速度总和的平均值, 流量为单位时间内通过道路某一横截面的车辆数<sup>[22]</sup>。

$$\rho_i = \frac{N_i}{L} \quad (17)$$

$$\bar{v}(t) = \frac{1}{T} \sum_{t=0}^T \frac{1}{N} \sum_{n=1}^N v_n(t) \quad (18)$$

$$Q = \rho_i \times \bar{v}(t) \quad (19)$$

仿真每一次运行演化时间为 10000 时间步，记录最后 5000 时间步内所有车辆的速度，求得每一时间步内所有车辆的平均速度，最后将得到的速度值再做时间平均，得到可用车辆平均速度  $\bar{v}'(t)$ ，即  $\bar{v}'(t) = \frac{1}{T_{valid}} \sum_{t=t_0}^T \frac{1}{N} \sum_{n=1}^N v_n(t)$ 。

## 3.2 不同自动车渗透率下的混合交通流特性分析

### 3.2.1 基本图

图 7 为不同自动车渗透率（0~100%）下的混合交通流基本图。将上方原始散点图进行拟合，得到下方曲线图。

如图 7-a 所示，随着自动车渗透率的提升，道路通行能力逐步上升，且在自动车渗透率为 100% 时达到最大值 2460veh/h，特别地，相比较于纯人驾车交通流，纯自动车交通流的通行能力增加了 33.55%。并且，混合流的平均车速也有一定的提升（图 7-b），因此可以证明在一定范围内，自动车的加入使得交通流的效率有所提升。

图 7-a 还显示了临界密度（即最佳车流密度）随自动车渗透率的变化。随着自动车渗透率的提升，临界密度逐渐后移，这表明在一定范围内，自动车的增加可以提升车流的稳定性。

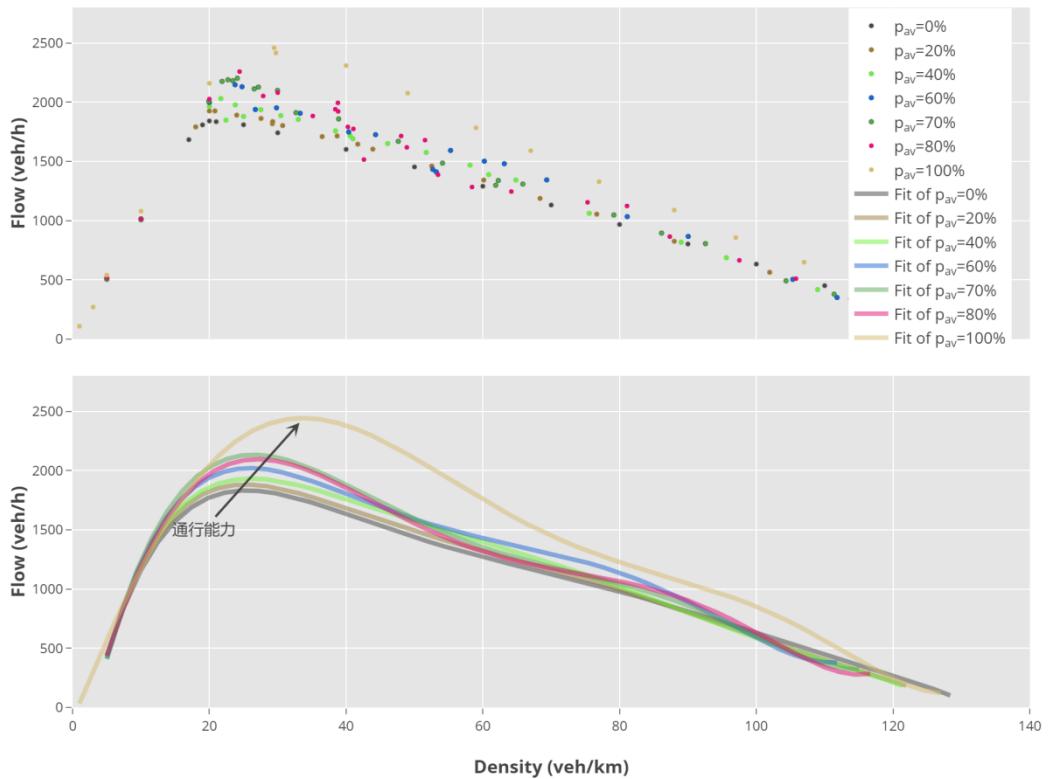


图 7-a 混合交通流密度-流量基本图（总图）

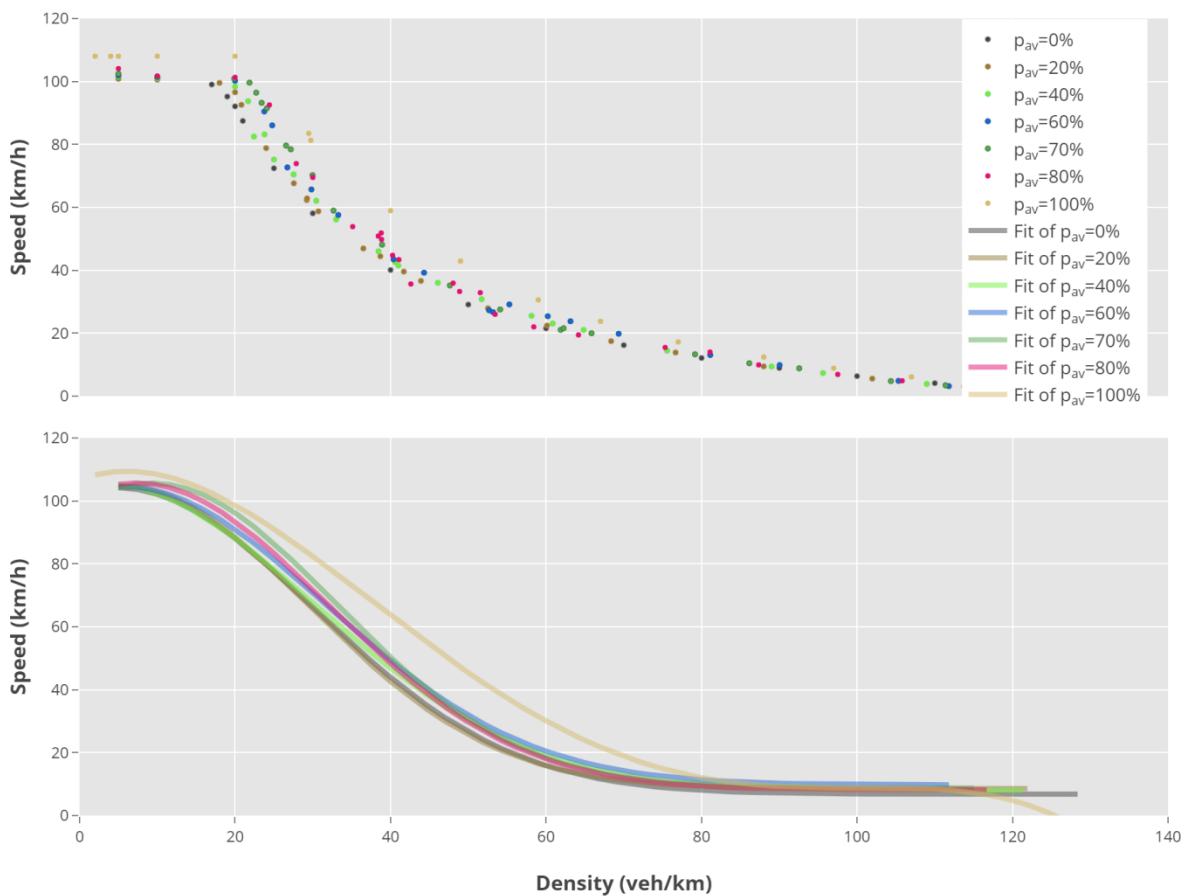


图 7-b 混合交通流密度-速度基本图

以上均为人驾车在不限制换道次数下的仿真结果，本文还对比了其与限制人驾车换道次数为1次的仿真结果(图7-c)。当自动车渗透率小于60%时，前者的仿真效果更好；当自动车渗透率等于60%时，两者较为一致；当自动车渗透率大于60%时，后者在临界密度之后的密度范围内仿真效果更好，且临界密度后移，车流更加稳定。上述结果表明人驾车的频繁换道在低自动车渗透率时可以较有效地利用道路资源，但在高自动车渗透率时，由于人驾车并未有“预见”能力，因此频繁换道或将增加道路拥堵的可能性。

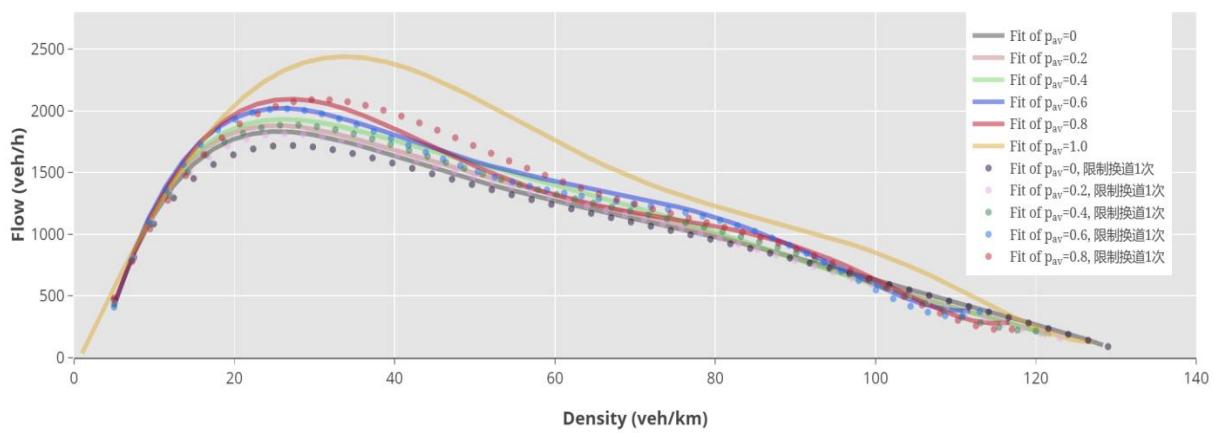


图 7-c 混合交通流密度-流量基本图，限制人驾车换道次数与不限制人驾车换道次数对比

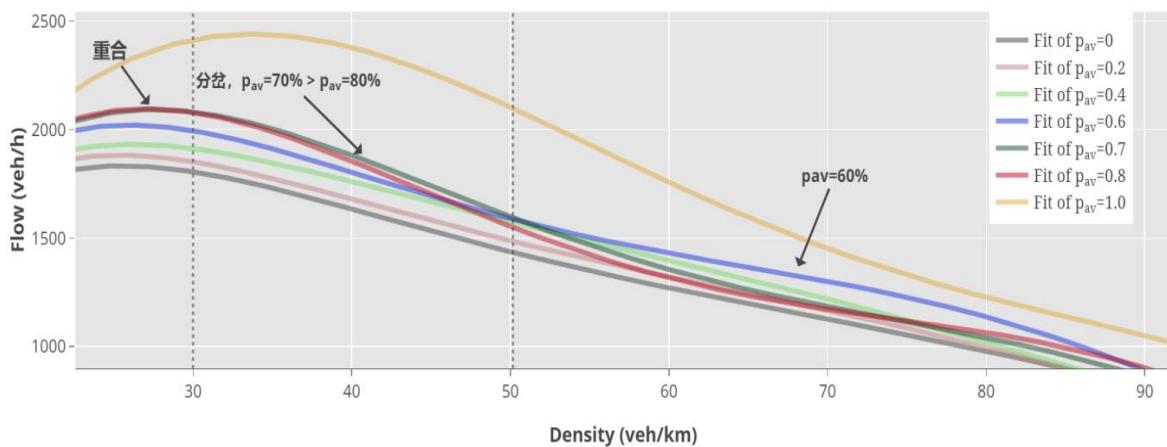


图 7-d 混合交通流密度-流量基本图（细节图）

由上文可知，纯汽车交通流的仿真效果最佳，但由于未来很长一段时间仍将面临汽车与人驾车共存的局面，因此探究混合交通流中是否存在最佳汽车渗透率（即对应的流量最大）是本文的重点之一。由图 7-a、7-d 可知，纯汽车交通流曲线与其他曲线在密度为 20veh/km~80veh/km 的区间内存在较大流量差。虽然汽车渗透率从 0 上升至 80%时，对应的流量有一定的上升趋势，但是每次增加的流量不明显；并且其最大值（80%对应的流量）与纯汽车曲线仍有约 250veh/h 流量的差距。同时，值得注意的是，汽车渗透率从 70%上升至 80%时，流量增加似乎有停滞趋势。因此猜测当汽车渗透率达到一定高度（比如 70%）时，汽车对于混合流的效率提升作用不再如之前一般显著，甚至有下降趋势，而此时人驾车对混合流的作用占主要部分。

为验证上述猜想，本文对比了密度为 30veh/h 左右，汽车渗透率为 100%、99.67%（每车道 1 辆人驾车）、99.33%（每车道 2 辆人驾车），98.33%（每车道 5 辆人驾车）、96.67%（每车道 10 辆人驾车）、93.33%（每车道 20 辆人驾车）共 5 种仿真场景的行驶轨迹，以探究在接近 100%汽车渗透率的场景下，人驾车的数量对于混合流流量的影响作用。

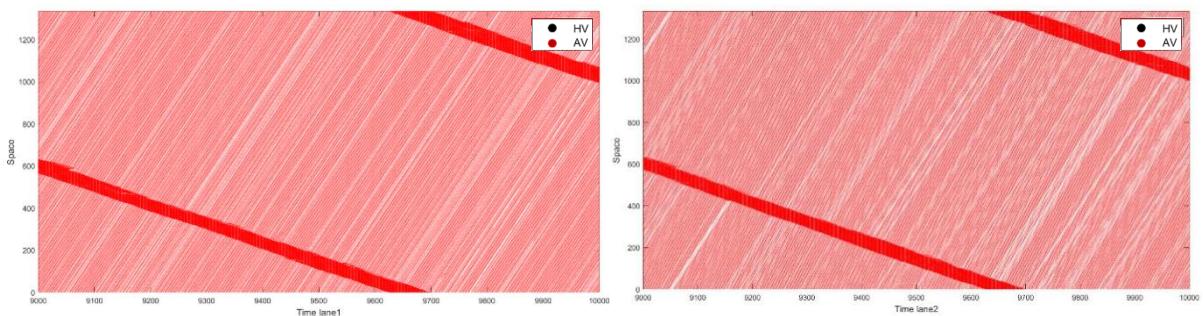


图 8 自动车渗透率为 100%的双车道时空图

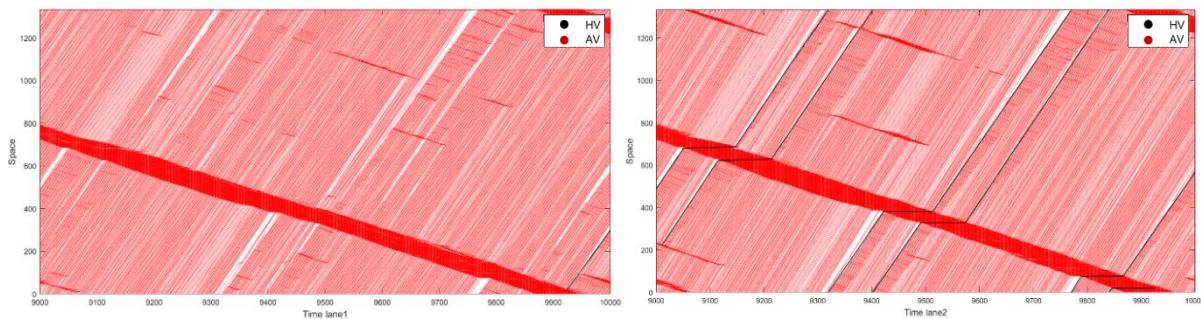


图 9 自动车渗透率为 99.67%（每车道 1 辆人驾车）的双车道时空图

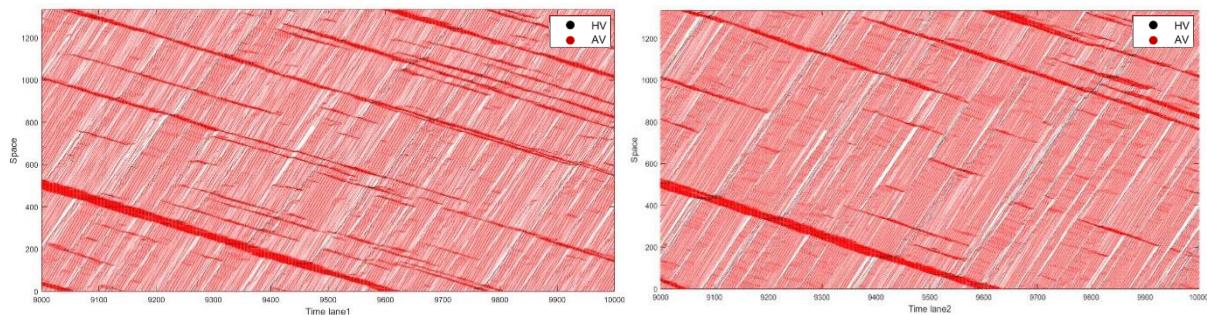


图 10 自动车渗透率为 96.67%（每车道 10 辆人驾车）的双车道时空图

对比图 8~图 10 可知，纯自动车交通流中自动车行驶轨迹整齐且车辆间无较大空隙，比较有效地利用了道路资源，对提升整体流量具有重要的作用（图 8）。而当每车道加入一辆人驾车后（图 9），时空图中出现了一定数量的空白部分。在不同车道中出现空白的原因不同，若为人驾车所在车道，则每处空白为一辆人驾车与其前方的自动车由于存在一定速度差而不断拉开距离所致；若为人驾车所在车道的相邻车道，则空白出现的原因为：人驾车所在车道由于人驾车的存在而出现道路空间的闲置，因此本车道车辆有较大倾向进行换道操作，这将导致本车道也出现闲置空间，因此在本车道同样位置也会出现空白部分。因此，虽然交通系统中仅有两辆人驾车（占比 0.33%），但其一定程度上造成了道路空间利用率的下降，最终导致系统整体流量相对于纯自动车交通流下降 100veh/km（占比 4.23%）。可见人驾车很大程度上影响了系统流量的变化。当每车道加入十辆人驾车后（图 10），人驾车驶过造成的道路空白出现得更为频繁，一定程度上造成了拥堵波个数的增加。

图 11 为对以上五种场景多次仿真后得到的平均流量与原基本图中其他自动车渗透率流量的对比图。虽然随机因素的存在导致了数据在一定范围内的波动，但该图总体上体现了人驾车在高自动车渗透率下对系统流量的关键影响作用。且当人驾车数量为 20（6.67%）时，即自动车渗透率为 93.33%时，流量与自动车渗透率为 30%时的流量接近。这也在一定程度上验证了上文“自动车超过一定比例（约为 70%）后，流量增加停滞、甚至有下降趋势”的论述。

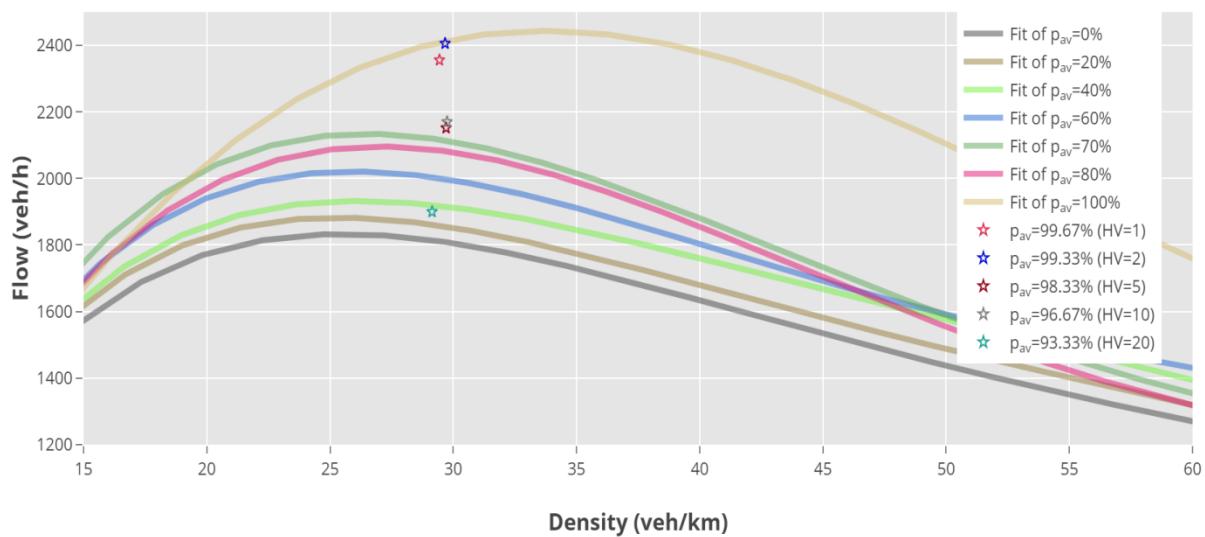


图 11 基本图对比图

### 3.2.2 时空图

由于当道路拥堵比例在 60%以下时，自动车的平均车速均大于人驾车平均车速（具体证明见 3.3.1 节），因此自动车的利己策略带来的速度提升效果在一定范围内得以从数值上证明。以下将从图中更直观地体现自动车相对于人驾车的速度提升。

对不同场景下的仿真结果作时间-空间图（以下简称时空图）以更细致地观察自动车的利己行为。图 12 为时空图的部分放大图，其中红线（点）代表自动车，蓝线（点）代表人驾车的运行轨迹。由于自动车采取利己策略以期望提升其自身车速，其在行驶过程中会与前车保持在一个较小的距离，因此若自动车后方为人驾车，则两者间车距会由于两者存在的速度差而不断自然拉大。但该距离差的拉大趋势在以下两种场景下有被打断的可能性：

- 1) 相邻车道上的车辆通过换道进入本车道，填补了距离，因此缩小了距离差（图 12-a）；
- 2) 前方的拥堵波传播至该时刻，导致自动车速度下降（图 12-b），从而缩小与后方人驾车的距离差。

而图 13 则较为清楚地展示了自动车的利己换道行为，其中红线代表自动车（2辆），黑线代表人驾车。当自动车遇到前方传至当前的拥堵波时，会在存在换道空间且换道利于后续行驶时进行换道以提升其自身速度。

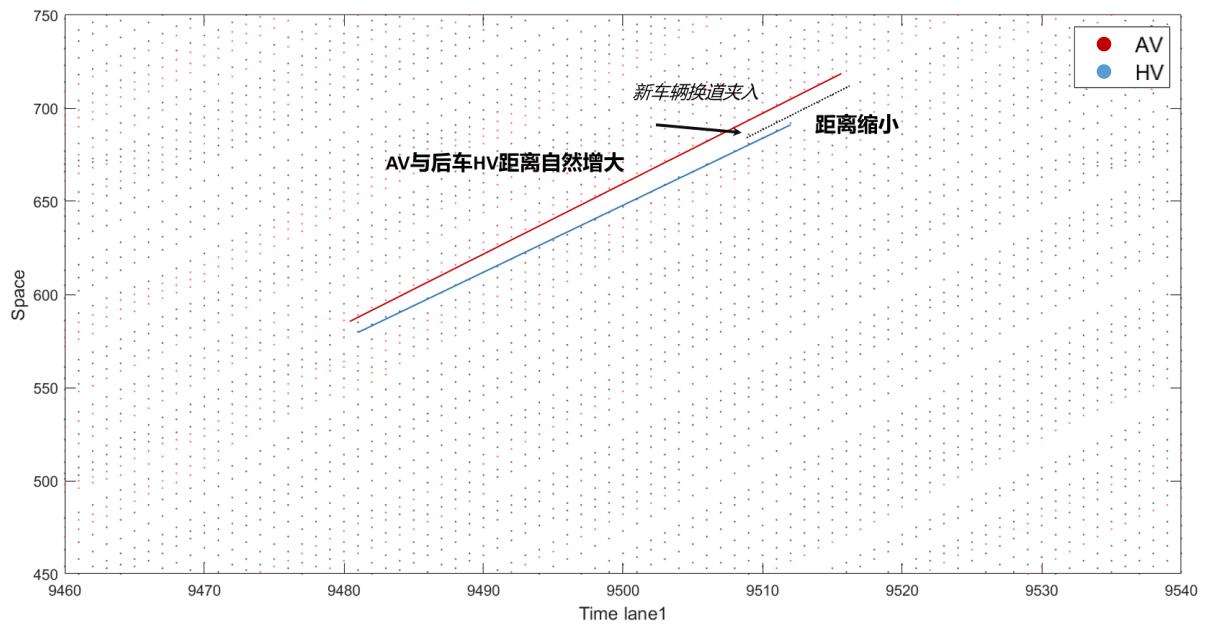


图 12-a 新车辆换道夹入导致的距离缩小

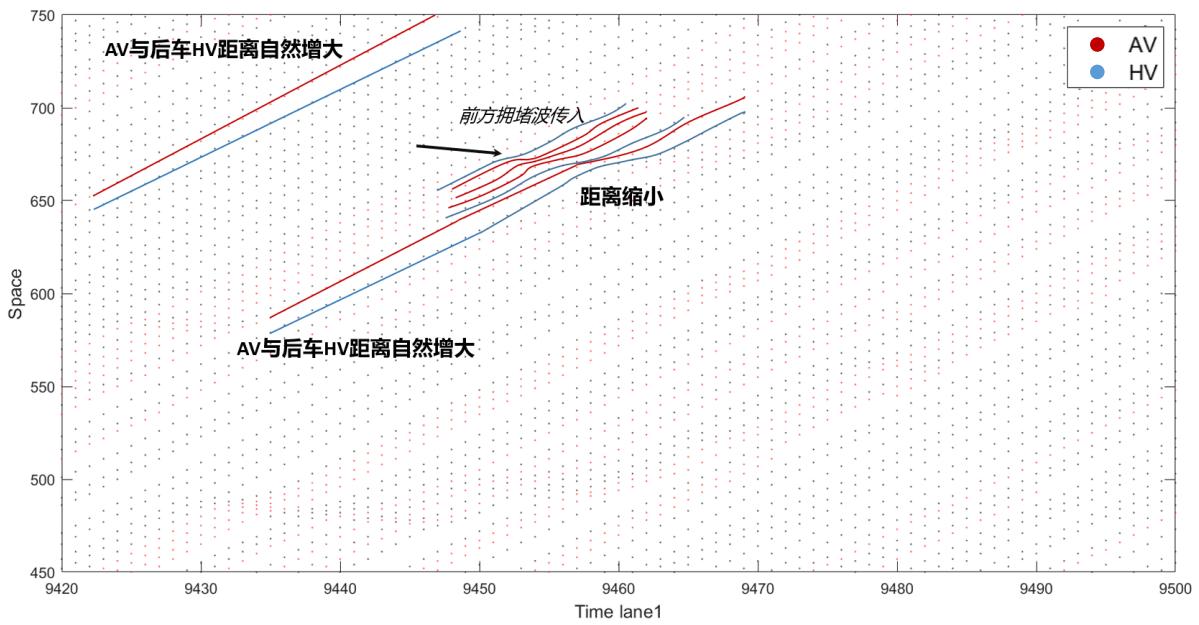


图 12-b 前方拥堵波传入导致的距离缩小

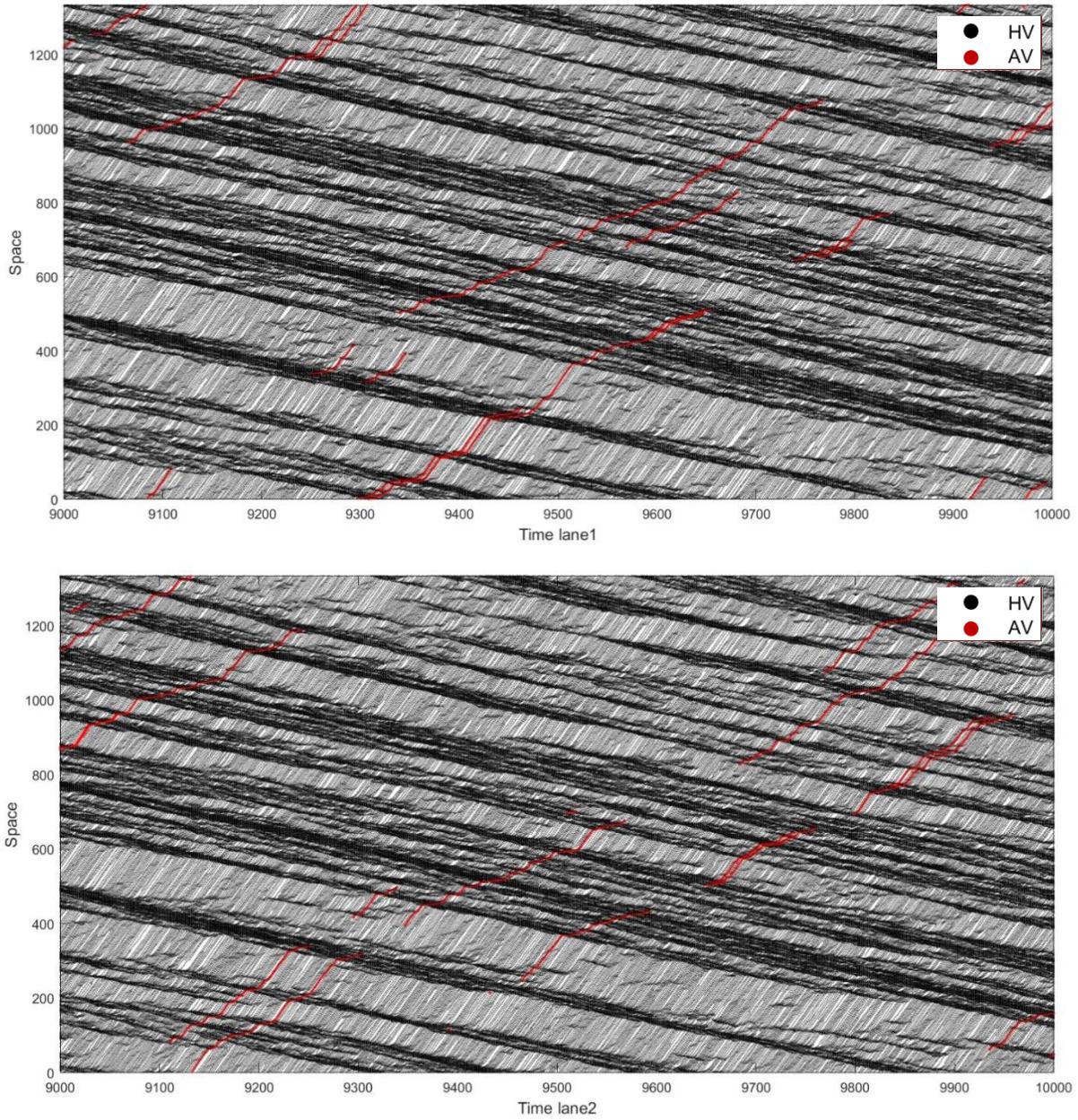


图 13 双车道时空图 (density=400veh/h, AV=2veh)

### 3.2.3 拥堵比例

由前文可知，自动车对混合交通流的流量与速度变化产生了较大的影响，后续将继续分析当自动车比例增加时交通拥堵情况的变化情况。

首先对拥堵车辆和拥堵比例进行定义。在某一时刻把速度小于等于 $1\text{cell/s}$ (即 $27\text{km/h}$ 时)的车辆定义为拥堵车辆，即任意时刻的拥堵情况可通过拥堵车辆的比例来表示交通流的拥堵程度，如下式描述：

$$CR = \frac{n}{\Delta TN}$$

其中， $CR$  代表拥堵比例， $n$  代表交通流中拥堵车辆的数量， $\Delta T$  代表仿真时间， $N$  代表总车数。

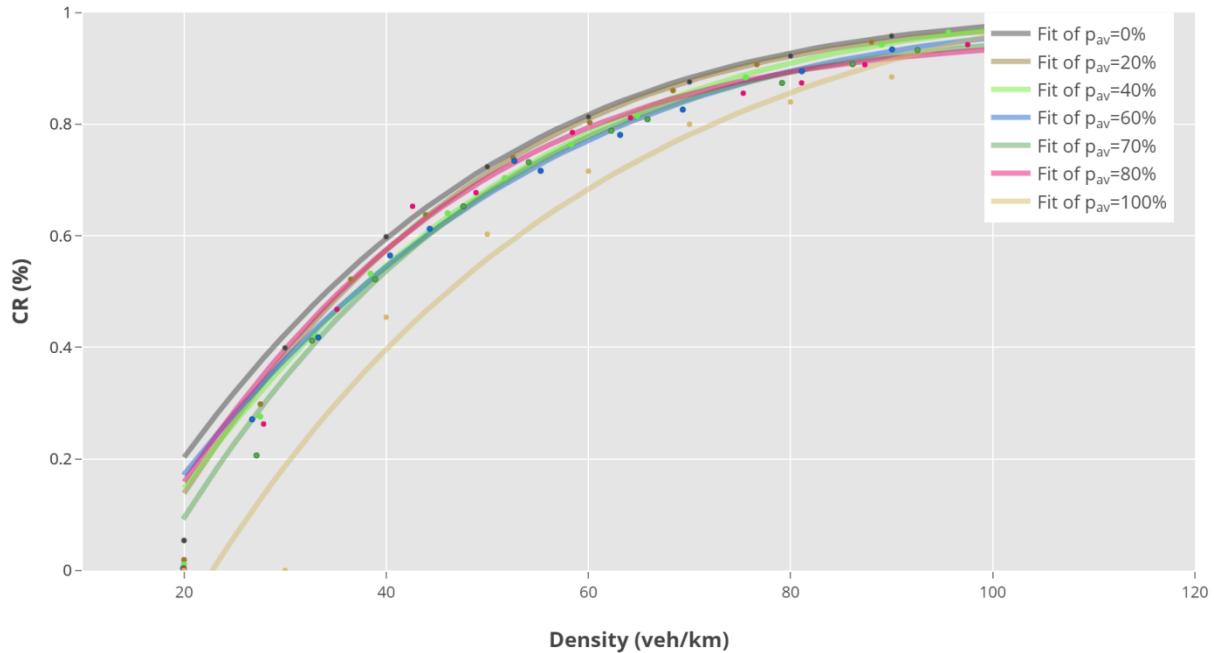


图 14-a 不同自动车渗透率下的拥堵比例曲线图（完整）

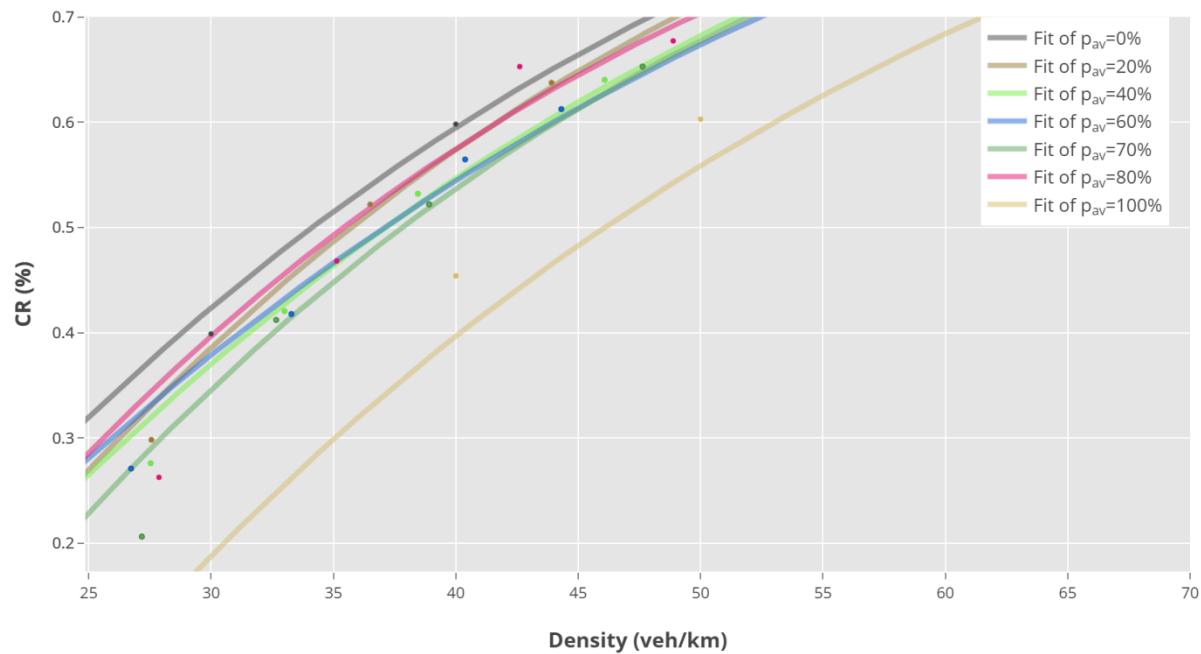


图 14-b 不同自动车渗透率下的拥堵比例曲线图（部分）

由拥堵比例曲线图 14-a 和 14-b 可知，自动车渗透率为 0~80% 的 CR 比例曲线接近（相差不超过 7%），而纯自动车交通流的 CR 比例可与纯人驾车交通流相差最高达到 20%。因此，本文更加细致地探究了纯自动车 CR 比例与其他自动车渗透率 CR 比例之

间存在较大差距的原因。

表 2 测试了 30veh/h 密度下的各自动车渗透率的 CR 比例。由于 3.2.1 节探讨了高自动车渗透率时人驾车对系统流量的关键作用，因此本节将继续验证此作用在 CR 比例中的体现。由表 2 可知，0~80% 自动车渗透率对应的 CR 比例相差约 10%，而 93.33%~100% 自动车渗透率对应的 CR 比例相差高达约 15%，这清楚地显示了少量人驾车对混合流拥堵程度的较大影响作用。

表 2 不同自动车渗透率下的 CR 比例 (density=30veh/h)

自动车渗透率 (%)	CR (%)
0	39.88
20	39.27
40	36.57
60	33.47
70	28.75
80	30.58
93.33 (20 辆 HV)	37.13
96.67 (10 辆 HV)	35.11
98.33 (5 辆 HV)	31.43
99.33 (2 辆 HV)	23.80
99.67 (1 辆 HV)	25.09
100	22.71

### 3.2.4 平均行驶时间

对相同距离下不同自动车渗透率的车流进行平均行驶时间的测定（表 3）。以中密度（40veh/km）车流为例，平均行驶时间随自动车渗透率的增加先下降（0~70%）后上升

表 3 不同自动车渗透率下的混合流平均行驶时间 (density=40veh/h, distance=2000m)

自动车渗透率 (%)	平均行驶时间 (s)
0	172
20	149
40	147
60	146
70	145
80	171
100	139

(70%~80%) 再下降 (100%)。并且, 为验证在高自动车渗透率下人驾车对平均行驶时间的影响, 同 3.2.3 节在系统中分别加入 1、2、5、10、20 辆人驾车 (下表中未列出), 观察平均行驶时间的变化趋势。结果表明, 该趋势与 3.2.3 节中拥堵比例趋势一致, 即人驾车在高自动车渗透率下存在关键性作用。同时, 70% 自动车渗透率也在一定程度上被多次证明为混合流中的最佳自动车渗透率。

### 3.3 自动车“利己”策略验证与分析

#### 3.3.1 自动车“利己”策略验证

由前文所述, 随机慢化是人类行为或外部条件变化而引起的自然速度波动, 且本文的仿真过程假设是在无外部条件干扰下进行的。因此前文在对自动车进行建模时未将随机慢化纳入速度更新规则中。然而, 为证明自动车速度的提升来源于对策略的自我学习, 而不仅是“无随机慢化”, 本节将对比同样引入随机慢化后自动车与人驾车的平均速度; 同时, 观察在不同自动车渗透率场景下自动车执行利己策略的结果。

图 15 体现了从自由流、最大流量到拥堵三个状态的演变过程中自动车与人驾车的速度对比。设置自动车数量为固定十辆, 以更好地观察极少辆自动车在混合流中的策略学习效果。由图 15-a 可知, 当混合流密度超过 40veh/km 后, 人驾车速度超过自动车。这是由于交通流出现大范围拥堵时 (密度为 40veh/km 时拥堵比例大于 60%), 自动车无足够的穿梭空间导致其速度无法得到有效地提升, 致使其速度小于人驾车。而当交通流尚未拥堵时, 道路中存在一定的空间可供自动车进行利己跟车和换道, 因此此时自动车的速度大于人驾车; 这也同样验证了在随机慢化规则下, 自动车仍可以有效地学习利己策略, 从而提升自身的速度。

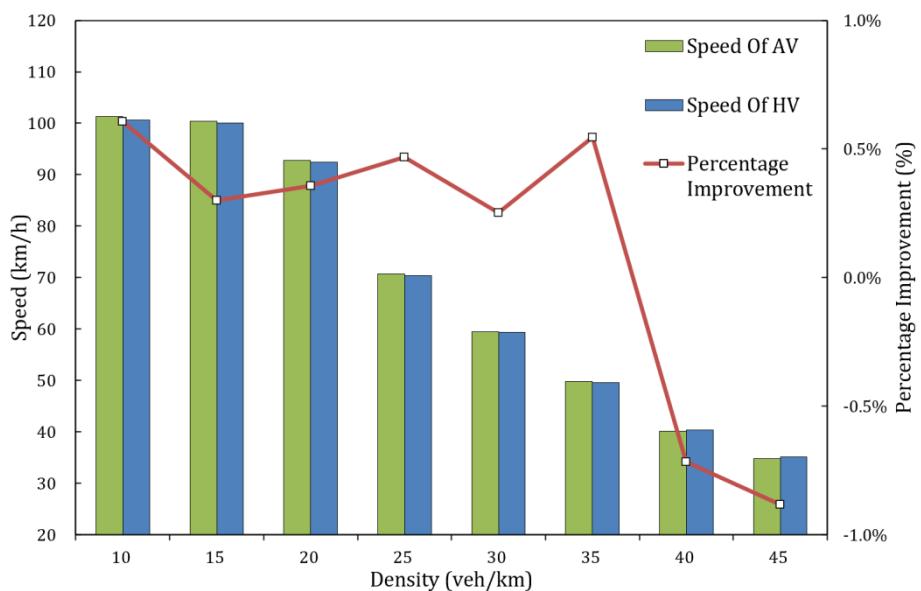


图 15-a 随机慢化规则下自动车与人驾车的速度对比 (自动车数量为 10 辆)

去除随机慢化规则后，在混有少辆自动驾驶车仿真实验的基础上，不断增加自动驾驶车数量，观察对应利己策略的有效密度范围。如图 15-b 所示，不同自动驾驶渗透率对应的自动驾驶利己策略的有效密度范围不同，但趋势同图 15-a（随机慢化规则下）较为一致，即在高密度（道路拥堵比例达到一定值）时无法应用该策略。

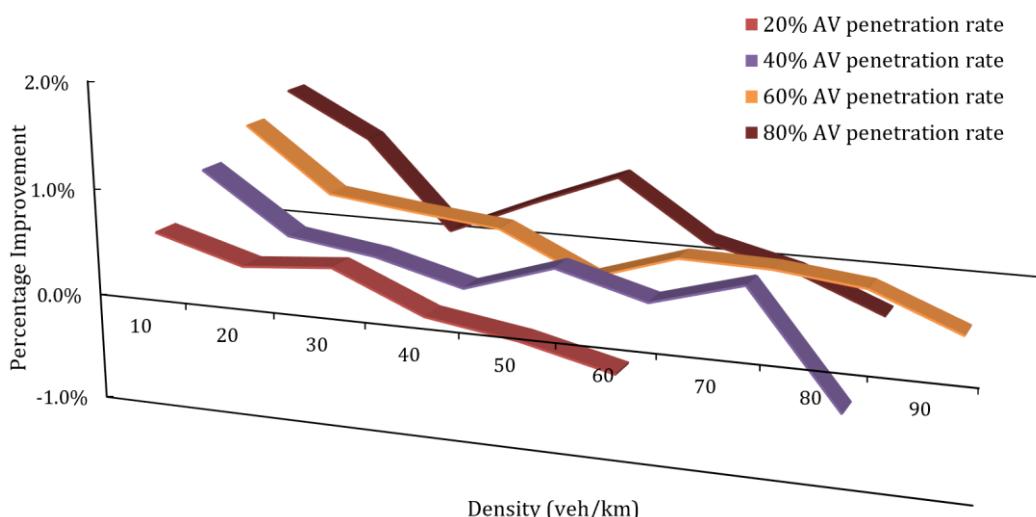


图 15-b 无随机慢化规则下自动驾驶与人驾车的速度差值百分比变化

### 3.3.2 自动车“利己”策略分析

提取不同场景（即不同密度、不同自动驾驶渗透率）下的自动驾驶训练结果（即 Q 表），对 Q 表按照状态分类展开分析。由于前文在定义状态空间时将距离与速度分成了三个区间段，因此，这里将对不同状态与不同动作的对应性规律进行作图分析，并将规律性总结与表 2-表 3 中。其中，下文中的六变量指 $\{d_f, v_f, d_{f,other}, v_{f,other}, d_{r,other}, v_{r,other}\} \in \{1, 2, 3\}$ （具体变量指代介绍见 2.2.2 节）。

总结表 4，在临界密度（即对应的流量最大）下，“强加速”动作基本发生在与周围车辆有较大相对距离和相对速度时，但在自动驾驶渗透率为 100% 时，自动驾驶表现得较为激进，即使在与前车相对距离较小且逐渐靠近时仍有一定概率选择采取强加速动作。

由于“加速”动作的限制条件相对于强加速较少，且自动驾驶在“利己”原则下倾向于加速，因此从表格数据看，在不同状态下都有可能采取加速动作。

“保持”动作的规律较为明显，当自动驾驶渗透率较低时，相对距离与相对速度倾向于集中在 1 或 2；当自动驾驶渗透率较高时，相对距离最高频率都在 2（即“近距离”至“远距离”范围内），而相对速度最高频率都为 1（即靠近）。

在临界密度下“减速”与“强减速”动作仅在自动驾驶渗透率为 100% 时较多发生。说明在临界密度且有人驾车存在时，自动驾驶运行较为平稳；当无人驾车存在时，虽然系统流量大于有人驾车存在的场景，但自动驾驶在运行时的减速、强减速操作使得其在追求速度和效率的同时也降低了平稳性和舒适性。

当存在人驾车时，“换道”动作较高频率发生在与周围车相对距离和相对速度都适中的情况下；而当不存在人驾车时，换道更为激进，即当相对速度为“靠近”时更易发生换道。

表 4-a 临界密度（20~30veh/km）下，不同自动车渗透率对应的最高频率策略规律

渗透率\动作	强加速	加速	保持
20%	$v_{f,other}, v_{r,other}$ 在 2 处，其余变量均在 3 处有最高发生频率。	$d_f = 3$ $v_f = 1$ $v_{f,other} = 2$ $v_{r,other} = 1$ $d_{f,other} = 3$ $d_{r,other} = 2$	
40%		$d_f \cdot v_f$ 未有明显动作倾向 $v_{f,other} = 1$ $v_{r,other} = 1$ $d_{f,other} = 1$ $d_{r,other} = 1$	六变量多集中在 1 或 2
60%		$d_f = 1$ $v_{r,other} = 1$ $v_f, v_{f,other}, d_{f,other}, d_{r,other}$ 未有明显动作倾向	
70%	六变量均在 3 处（即相对距离为“远距离”，相对速度为“远离”）有最高发生频率。	$d_f, v_f, v_{f,other}$ 未有明显动作倾向 $v_{r,other} = 1$ $d_{f,other} = 3$ $d_{r,other} = 3$	
80%		$d_f$ 未有明显动作倾向 $v_f = 2, 3$ $v_{f,other} = 1$ $v_{r,other} = 1$ $d_{f,other} = 3$ $d_{r,other} = 3$	$d_f, d_{f,other}, d_{r,other} = 2,$ $v_f, v_{f,other}, v_{r,other} = 1$
100%	$d_f = 1, 2$ $v_f = 1$ $v_{f,other} = 1$ $v_{r,other} = 2$ $d_{f,other} = 3$ $d_{r,other} = 3$	$d_f = 2$ $v_f = 1, 2$ $v_{r,other} = 1$ $d_{f,other} = 3$ $v_{f,other}, d_{r,other}$ 未有明显动作倾向	

表 4-b 临界密度 (20~30veh/km) 下, 不同自动车渗透率对应的最高频率策略规律

渗透率\动作	减速	强减速	换道
20%	出现频次较少, 可忽略不计	未采取该策略	除 $v_{f,other} = 1$ , 其余变 量均在 2 处有最高发生 频率
40%	未采取该策略		
60%			
70%			
80%	出现频次较少, 可忽略不计	仅出现一次, 可忽略不计	除 $v_{f,other}$ 未有明显动作 倾向, 其余变量均在 2 处有最高发生频率
100%	$d_f = 3$ $v_f = 1$ $v_{f,other} = 1$ $v_{r,other} = 2$ $d_{f,other} = 2$ $d_{r,other} = 1$	$d_f, d_{f,other}$ 未有明显动作倾向 $v_f = 1$ $v_{f,other} = 1$ $v_{r,other} = 1$ $d_{r,other} = 1$	$d_f = 3$ $v_f = 1$ $v_{f,other} = 1$ $v_{r,other} = 1$ $d_{f,other} = 2$ $d_{r,other} = 2$

总结表 5, 在中密度 (60veh/km) 下, “强加速” 动作发生规律与在临界密度下时较为一致, 即基本发生在与周围车辆有较大相对距离和相对速度时, 但在自动车渗透率为 100% 时, 自动车表现得较为激进。稍有不同的是, 在低自动车渗透率时, 中密度下的自动车在更多状态下会采取强加速动作。

相较于临界密度下的自动车, 中密度下的自动车采取“加速” 动作更有规律可循。该动作大多发生在与本车道前车为“近距离” 或“中距离” 且为“远离” 状态时, 且随着自动车渗透率的上升, 采取“加速” 动作的频率有所提高。

“保持” 动作在临界密度与中密度下表现得较为一致, 由于与相邻车道车辆的相对距离和速度集中在 1 或 2, 即换道条件并未达到最佳, 且与本车道前车的相对距离与相对速度也倾向于集中在 1 或 2, 即仍存在一定的行驶空间, 因此在此类状态下的动作倾向于“保持”。

相较于临界密度下的自动车, 中密度下的自动车由于无法始终维持较高速度, 因此有更高的频率采取“减速” 和“强减速” 动作。且随着自动车渗透率的增加 (20%~80%),

“强减速” 发生频率逐渐上升, 说明自动车的加入虽能提高系统运行效率, 但也降低了自动车运行的平稳性。另一方面, 当自动车渗透率达到 100% 时, “强减速” 频率又回到较低水平, 因此说明在中密度下, 人驾车的存在在一定程度上也是自动车运行不平稳的原因之一。并且, 从“减速” 至“强减速”的变化取决于自动车所处的状态的变化, 即与本车道前车从“保持” 转为“靠近”、与相邻车道前车从“保持” 转为“靠近” 等, 而在相对距离方面, 两者并未有明显差别。

中密度下的“换道” 动作比临界密度下的更为激进。其在与相邻车道车辆为“靠近” 时有最高的换道频率, 但其在不同相对距离上未有明显倾向。

# 北京工业大学毕业设计（论文）

表 5-a 中密度 (60veh/km) 下, 不同自动车渗透率对应的最高频率策略规律

渗透率\动作	强加速	加速	保持
20%	$d_f, v_f, v_{f,other}, d_{r,other}$ 未有明显动作倾向 $v_{r,other} = 2$ $d_{f,other} = 3$	$d_f = 1$ $v_f = 3$ $v_{f,other} = 3$ $v_{r,other} = 1$ $d_{f,other} = 3$ $d_{r,other} = 1$	$v_f$ 未有明显动作倾向 $d_f = 1, 2$ $v_{f,other} = 2$ $v_{r,other} = 1$ $d_{f,other} = 2$ $d_{r,other} = 1$
40%	$d_f, v_f, v_{f,other}, d_{r,other}, d_{f,other}$ 未有明显动作倾向 $v_{r,other} = 2, 3$	$d_f = 1, 2$ $v_f = 2, 3$ $v_{f,other} = 2$ $v_{r,other} = 3$ $d_{f,other} = 1$ $d_{r,other} = 1$	$v_f, v_{f,other}$ 未有明显动作倾向 $d_f = 2$ $v_{r,other} = 1$ $d_{f,other} = 2$ $d_{r,other} = 2$
60%		$d_f = 1$ $v_f = 3$ $v_{f,other} = 3$ $v_{r,other} = 3$ $d_{f,other} = 1$ $d_{r,other} = 3$	$v_f, v_{f,other}$ 未有明显动作倾向 $d_f = 1, 2$ $v_{r,other} = 1$ $d_{f,other} = 1, 2$ $d_{r,other} = 1, 2$
70%	六变量均在 3 处 (即相对距离为“远距离”, 相对速度为“远离”) 有最高发生频率。	$d_f = 1, 2$ $v_f = 3$ $v_{f,other} = 1, 3$ $v_{r,other} = 3$ $d_{f,other} = 1$ $d_{r,other} = 2, 3$	$d_f = 2$ $v_f = 1$ $v_{f,other} = 3$ $v_{r,other} = 1$ $d_{f,other} = 2$ $d_{r,other} = 2$
80%		$d_f = 1$ $v_f = 3$ $v_{f,other} = 3$ $v_{r,other} = 3$ $d_{f,other} = 1$ $d_{r,other} = 1$	$d_f = 1, 2$ $v_f = 1, 3$ $v_{f,other} = 3$ $v_{r,other} = 1, 3$ $d_{f,other} = 1, 2$ $d_{r,other} = 1, 2$
100%	$d_f, v_{f,other}$ 未有明显动作倾向 $v_f = 1, 2$ $v_{r,other} = 2$ $d_{f,other} = 3$ $d_{r,other} = 3$	$d_f = 3$ $v_f = 3$ $v_{f,other} = 3$ $v_{r,other} = 1$ $d_{f,other} = 1$ $d_{r,other} = 1$	$d_f = 2$ $v_f = 1$ $v_{f,other} = 1, 2$ $v_{r,other} = 1$ $d_{f,other} = 2$ $d_{r,other} = 2$

## 北京工业大学毕业设计（论文）

表 5-b 中密度 (60veh/km) 下, 不同自动车渗透率对应的最高频率策略规律

渗透率\动作	减速	强减速	换道
20%	$d_f = 3$ $v_f = 2$ $v_{f,other} = 2$ $v_{r,other} = 1$ $d_{f,other} = 3$ $d_{r,other} = 1$	出现频次较少, 可忽略不计	$d_f = 3$ $v_f = 1$ $v_{f,other} = 1$ $v_{r,other} = 1$ $d_{f,other} = 1, 2$ $d_{r,other} = 1, 2$
40%	$d_f = 2$ $v_f = 2$ $v_{f,other} = 2$ $v_{r,other} = 1$ $d_{f,other} = 3$ $d_{r,other} = 1$		$d_f, v_f, d_{r,other}, d_{f,other}$ 未有明显动作倾向 $v_{f,other} = 1$ $v_{r,other} = 1$
60%	$d_f$ 未有明显动作倾向 $v_f = 2$ $v_{f,other} = 2$ $v_{r,other} = 1$ $d_{f,other} = 1, 3$ $d_{r,other} = 1$	$d_f = 2, 3$ $v_f = 1$ $v_{f,other} = 1$ $v_{r,other} = 1$ $d_{f,other} = 2, 3$ $d_{r,other} = 1, 2$	
70%	$d_f, d_{f,other}$ 未有明显动作倾向 $v_f = 2$ $v_{f,other} = 2$ $v_{r,other} = 1$ $d_{r,other} = 1$	$v_{r,other} = 1$ $d_{f,other} = 2, 3$ $d_{r,other} = 2, 3$	$d_f, d_{r,other}$ 未有明显动作倾向 $v_f = 1, 2$
80%	$d_{r,other}, d_{f,other}$ 未有明显动作倾向 $d_f = 3$ $v_f = 2$ $v_{f,other} = 2$ $v_{r,other} = 1$		$v_{f,other} = 1$ $v_{r,other} = 1$ $d_{f,other} = 2, 3$
100%	$v_{f,other}, d_{r,other}, d_{f,other}$ 未有明显动作倾向 $d_f = 3$ $v_f = 1$ $v_{r,other} = 1$	$d_{r,other}$ 未有明显动作倾向 $d_f = 3$ $v_f = 1$ $v_{f,other} = 1$ $v_{r,other} = 1$ $d_{f,other} = 3$	$d_f, d_{r,other}, d_{f,other}$ 未有明显动作倾向 $v_f = 1, 2$ $v_{f,other} = 1$ $v_{r,other} = 1$

### 3.4 场景适用性分析

为验证特定场景下训练而成的自动车行驶策略在其他场景中的适用性，从而降低训练次数，本节将分别测试中间密度的策略在其他密度的适用性与中间自动车渗透率的策略在其他自动车渗透率的适用性。

#### 3.4.1 同密度，不同自动车渗透率场景

以自动车渗透率为 20% 和 40% 为例，将密度为 60veh/km 的训练结果分别应用至 0~125veh/km 的场景中，观察相同密度下两者流量的差别。

由图 16 可知，在自动车渗透率为 20% 时，红点与黑点基本吻合。而在自动车渗透率为 40% 时，除密度在 40~60veh/km 中新策略的效果劣于旧策略外，在其余密度下两者吻合得较好。因此可证明，中间密度的训练结果在低自动车渗透率下对其他密度完全适用，在中低自动车渗透率下对其他密度较适用。

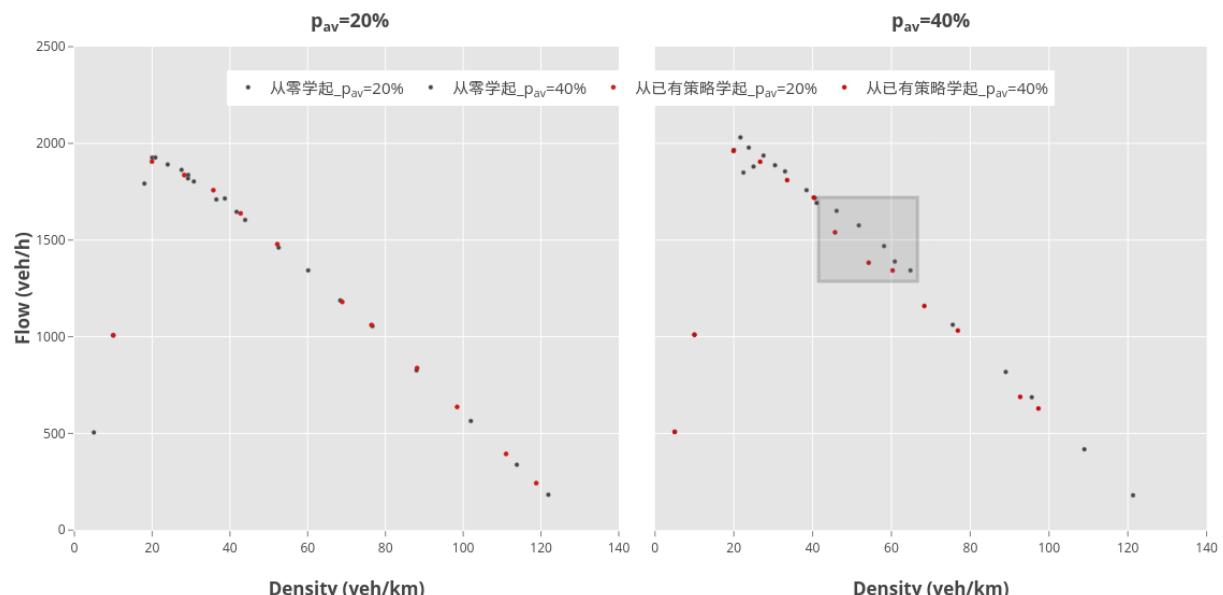


图 16 同密度不同自动车渗透率场景的适应性验证

#### 3.4.2 不同密度，同自动车渗透率场景

以密度为 30veh/km、40veh/km 和 50veh/km 为例，将自动车渗透率为 60% 的训练结果分别应用至自动车渗透率为 20%、40%、80%、100% 的场景中，观察相同密度下两者流量的差别。

由图 17 可知，在密度为 30veh/km 的场景下，新策略（四个深蓝色点）的训练效果均劣于旧策略（四条曲线）。具体地，在自动车渗透率为 20%、40%、80% 的场景中，新

策略的训练效果仅略劣于旧策略，流量差约为 30~40veh/h，而在自动车渗透率为 100% 的场景中，两者相差较大，流量差约为 200~300veh/h。

在密度为 40veh/km 的场景下，新策略（四个蓝绿色点）的训练效果也均劣于旧策略（四条曲线）。具体地，在低自动车渗透率（20%、40%）的场景下，旧策略与新策略对应的流量差距较小，约为 10~50veh/h；而在高自动车渗透率（80%、100%）的场景下，两者差距增大，流量差约为 100~400veh/h。

在密度为 50veh/km 的场景下，新策略（四个青绿色点）的训练效果在低自动车渗透率（20%、40%）的场景下适用性较好，误差在 10v~40veh/km 之内；在 80% 自动车渗透率场景下误差为 70veh/km 左右；而在 100% 自动车渗透率场景下，其误差达到了 300veh/h。

因此，以上分析可证明处于中间自动车渗透率的训练结果仅对较低的自动车渗透率适用（有一定误差，但误差较小）；当应用于高自动车渗透率时存在较大误差，特别地，当应用于纯自动车交通流时，误差最大。因此可以认为中自动车渗透率的训练结果不适用于纯自动车交通流。

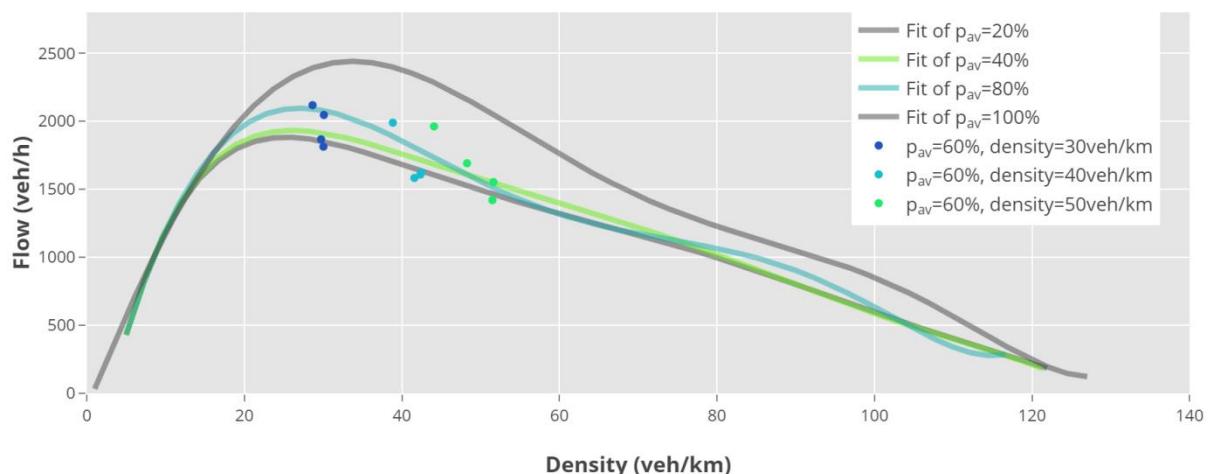


图 17 不同密度同自动车渗透率场景的适应性验证

## 3.5 本章小结

本章进行微观交通流仿真，并对仿真结果展开三个维度的分析：1) 分析不同自动车渗透率下的混合交通流特性，以观察自动车对混合流的影响；2) 验证并分析不同场景下的自动车利己策略，总结自动车采取策略的规律；3) 分析场景适用性，以削减自动车的训练次数。

为探究维度一，本章绘制了基本图、时空图、拥堵比例曲线图等以更加直观地观察自动车渗透率的变化对于混合流流量、速度、行驶时间、拥堵程度的影响。首先，相比于纯人驾车交通流，纯自动车交通流的通行能力提升了 33.55%，并且其平均车速也有大幅度提升，因此可以证明应用纯自动车流可增加机动车通行效率。同时，对比不同自

动车渗透率在不同密度上的流量，观察到临界密度随自动车渗透率的增加而后移，这表明自动车的加入提升了混合流的稳定性。其次，为探究是否存在最佳自动车渗透率（即在该渗透率下对应的流量最大、平均行驶时间最小、拥堵比例最小），本章对比了各自动车渗透率的仿真效果。虽然 0~93.33% 的自动车渗透率对应的各指标由于随机因素而存在一定波动性，但 70% 自动车渗透率在一定程度上可作为最佳自动车渗透率的参考比例。同时，本章也证明在高自动车渗透率（90%~100%）下，少量人驾车对混合流有较大影响。最后，时空图中车辆的轨迹验证了自动车通过更加智能的跟车（减小与前车间距）与换道行为达到利己的目的。

为探究维度二，本章首先验证了自动车利己策略在低拥堵道路条件下的有效性，同时得出“无随机慢化”与“有随机慢化”下的自动车利己策略的应用密度范围趋势较为一致。其次分析了自动车在不同状态下采取动作的倾向。具体地，为分析动作倾向规律，首先绘制了不同状态对应的动作频率图，而后又以表格的方式呈现动作倾向分布。最后，分别对临界密度（20~30veh/km）和中密度（60veh/km）下不同自动车渗透率的动作倾向分布展开分析。

为探究维度三，本章从两个角度，即同密度不同自动车渗透率和同自动车渗透率不同密度对不同场景的策略通用性展开验证。结果表明，中间密度的训练结果在低自动车渗透率下对其他密度完全适用，在中低自动车渗透率下对其他密度较适用。并且，中间自动车渗透率的训练结果仅对较低的自动车渗透率适用；当应用于高自动车渗透率时存在较大误差，特别地，当应用于纯自动车交通流时，误差最大。

## 4. 结 论

### 4.1 工作总结和主要创新点

#### 4.1.1 工作内容

本文以强化学习技术、元胞自动机交通流模型为理论工具，对双车道自动驾驶-人驾车混合交通流进行建模与仿真。通过对微观仿真结果的分析，得出自动驾驶的利己行驶策略以及该策略对混合交通流特性的影响，最后验证策略在不同场景中的适用性和通用性。具体的工作内容总结如下：

（1）建立 CA 框架下的基于 Q 学习的自动驾驶行驶策略学习模型

分别分析人驾车与自动驾驶驾驶特性，建立基于 STCA 的人驾车模型与基于 RL 的自动驾驶模型。其中，自动驾驶的行驶策略目标为“利己”，即在保证安全、稳定的前提下，通过采取加减速、换道、超车等动作尽量提升自身速度。自动驾驶具体的训练方法为表格式 Q 学习法，且其将采用 $\varepsilon$ -贪婪搜索策略选取即时动作。

（2）训练不同场景下的自动驾驶以获取其“利己”策略

设置不同交通场景(不同密度或不同自动驾驶渗透率)，对自动驾驶进行行驶策略的训练。通过不断地探索-惩罚/奖励-学习，培养自动驾驶的自我学习能力，包括“预见”能力与“见缝插针”能力。提取上述策略，观察其规律。

（3）改变相关仿真参数，分析自动驾驶“利己”策略对混合交通流的影响

通过设置不同自动驾驶渗透率、人驾车换道频率、交通流密度，观察混合流流量、平均速度、行驶时间、拥堵程度等的变化。探究是否存在最优自动驾驶渗透率以及人驾车或自动驾驶的存在对混合流交通特性的影响。

（4）分析并验证自动驾驶“利己”策略适用性

为减少自动驾驶训练次数，探究自动驾驶行驶策略在不同场景下的适用性。其中包括：同密度，不同自动驾驶渗透率的场景与不同密度，同自动驾驶渗透率的场景。若存在通用性，则可节省训练时间，为今后自动驾驶的复杂训练提供便利。

#### 4.1.2 结论

本节对以上四点工作所得出的结论总结如下：

- 1) 相对于传统的建模方法，基于强化学习的训练方法更符合自动驾驶行驶的不确定性与智能性。而自动驾驶“利己”策略在数值与图形上均得以验证：自动驾驶的平均速度在低拥堵场景中大于人驾车、自动驾驶通过智能跟车与换道获取长远利益；
- 2) 在自动驾驶-人驾车混合交通流中，道路通行能力、车流速度、车流稳定性均在一定范围内随着自动驾驶渗透率的增加而提升，并在自动驾驶渗透率为 100%时达到最大。同等

其他条件下，纯自动车交通流相对于纯人驾车交通流的通行能力提升 33.55%。值得注意的是，自动车渗透率增加到一定程度后，自动车对混合流的正向作用逐渐减弱，而 70%被证明是一个可参考的自动车渗透率临界点。大于 70%的自动车渗透率对混合流的效率提升不显著，甚至出现下降趋势；

- 3) 在高自动车渗透率（90%~100%）下，少量人驾车对混合流的流量、平均行驶时间、拥堵程度都有较大影响。在混合流中占比 0.33%的人驾车对系统整体流量的影响约为其自身比例的 13 倍；
- 4) 就自动车策略适用性而言，首先，中间密度的训练结果在低自动车渗透率下对其他密度完全适用，在中低自动车渗透率下对其他密度较适用。其次，中间自动车渗透率的训练结果仅对较低的自动车渗透率适用；当应用于高自动车渗透率时存在较大误差，特别地，当应用于纯自动车交通流时，误差最大。

### 4.1.3 主要创新点

- 1) 将强化学习技术作为自动车的策略学习方法，使得学习到的策略不受特定参数的限制，更加灵活，并且该“自我学习”能力符合自动车行驶的智能性与不确定性；
- 2) 提出“策略适应性”概念，并验证了部分场景下的自动车策略可以适用于其他场景，具有一定的泛化能力；
- 3) 结合强化学习技术与元胞自动机模型，从多个角度分析自动车对于混合交通流特性的影响，总结可参考的最佳自动车渗透率，并提出人驾车在高自动车渗透率下对混合流的关键性影响，为后续控制自动车与人驾车数量、提升交通系统效率打下基础。

## 4.2 未来展望

本文在强化学习的基础上训练了不同场景下的自动车利己策略，并且对未来自动车-人驾车混行的交通流进行了建模与仿真。但由于时间、精力的限制，本文的研究内容尚有许多地方需要填补、改进与升华。

首先，本文的双车道混合流模型是在元胞自动机框架下搭建的，元胞自动机虽是一种计算效率较高的微观交通流模型，但其过于离散化；并且，本文对人驾车进行行为建模而采用的 STCA 规则较为保守。因此，若要复现更真实的汽车运行状态，还需对元胞尺寸、更新规则做适当地修改和丰富。

其次，现阶段的双车道仿真道路难以覆盖更复杂的道路场景，而对称型车道也设置得略为简单。将来需要进一步修改道路模型，以求更贴近真实的道路环境。

并且，由于设备、时间和人力的限制，本文在仿真时未能对所有不同的场景进行测试，因此后续需要以更小的单位进行测试，填补之前的空缺，以确保更精确的仿真结果。

最后，通过对时空图的观察发现，自动车渗透率较高时自动车会自发形成具有较小间隔的车队，未来将把自动车列队研究作为后续的研究重点。

## 致谢

2016年秋，经历约莫七个半小时的高铁，来到了这个我将生活四年的地方——北工大，北京，陌生而憧憬、担忧又欢喜。

2020年夏，在距主校区1200公里外的家乡，想着大学时光竟只剩寥寥一月，有感伤，但更多的是对这四年大学生活的怀念、满足与感谢。

四年来，“平乐园100号”已成为“家”的代名词；从中蓝宿舍到校区骑车飞驰的日子似乎还在眼前；购物、上班、出游，搭乘十四号线不下数百次，“北工大西门”显得格外亲切；从旧图到新图，从中蓝到10号楼，在一教、三教、四教之间来回穿梭，在美食园风味、奥运、新食之中切换口味。

转眼间四年过去，从学子到校友不仅是身份的转变，更多是内心的沉淀。而内心得以沉淀，知识得以更加深厚、心智得以更加成熟，除自身的努力奋进外，离不开周围老师同学和亲人们的教导与关怀。

首先，感谢北京工业大学与城市交通学院给了我一个发挥自我能力的平台，使我在学校与学院的关怀与庇护下健康成长。

其次，感谢我的导师贺正冰教授在我毕设期间对我研究项目的指导与方向的纠正。虽然最后一学期我们并未实地碰面，但线上的不断督促、相关资料与讲座的分享、论文进展的定期检查都不断激励与敦促我将毕业论文这件事牢牢记在心中，并为之而努力。我最后能够顺利完成毕业设计并撰写完成毕业论文，离不开您的谆谆教导。

再者，感谢我的班主任熊杰老师、辅导员梁靖轩老师对我学业和生活方面的指导与关心。感谢我大学四年来的其他导师们，“新苗”计划导师荣建教授，创新导师于泉副教授，“杰出学子计划”导师翁剑成副教授以及王少帆导师等对我科研和学业的教导。同时，也感谢牟伦田副教授等任课老师对我课业的细心教授和课下的耐心解答。还要感谢我之前的上级王泽老师，对我工作的指导。

而后，感谢学长学姐们热心的为我解答学业和生活方面的问题，给予我的宝贵的学习资料帮助我提升了学习效率，叮嘱我的日常注意点避免我走了很多弯路；感谢我的各位同届小伙伴们时时激励我，忘不了和你们共同组队探讨问题，解决疑难的时光；感谢我的室友们，和你们从校园趣闻谈到天南海北的日子仍历历在目，感谢你们在我失意时的鼓励与陪伴。

最后，感谢我的父母和家人朋友们对二十多年来的支持与关怀，你们是我永远的后盾。也感谢参与评阅论文、出席答辩的各位老师、专家们提出的宝贵意见和建议。感谢不断拼搏的自己，愿我成为最好的自己。

## 参考文献

- [1] Bokui Chen, Duo Sun, Jun Zhou, Wengfai Wong, Zhongjun Ding. A future intelligent traffic system with mixed autonomous vehicles and human-driven vehicles[J]. Information Sciences, 2020.
- [2] S.D. Pendleton, H. Andersen, X. Du, X. Shen, M. Meghjani, Y.H. Eng, D. Rus, M.H. Ang. Perception, planning, control, and coordination for autonomous vehicles[J]. Machines, 2017, 5 (1): 6.
- [3] J Maddox. Improving driving safety through automation, congressional robotics caucus. National Highway Traffic Safety Administration, <https://docplayer.net/10832729-Improving-driving-safety-through-automation.html>, 2012/2020.
- [4] S. Bagloee, M. Tavana, M. Asadi, T. Oliver. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies[J]. Journal of Modern Transportation, 2012, 24 (4): 284-303.
- [5] Anderson, K. Nidhi, K. Stanley, P. Sorensen, C. Samaras, O. Oluwatola, Autonomous vehicle technology: A guide for policymakers[EB/OL]. Rand Corporation, [https://www.researchgate.net/publication/296697033\\_Autonomous\\_Vehicle\\_Technology\\_A\\_Guide\\_for\\_Policymakers/citation/download](https://www.researchgate.net/publication/296697033_Autonomous_Vehicle_Technology_A_Guide_for_Policymakers/citation/download), 2014/2020.
- [6] 苏镜荣,唐翀,程德勇.传统干路与单向二分路通行能力和延误对比[J].城市交通,2020,18(02):110-117+134.
- [7] Chen, Z., He, F., Yin, Y., Du, Y.. Optimal design of autonomous vehicle zones in transportation networks[J]. Transport. Res. Part B: Methodology, 2017, 99, 44–61.
- [8] Chen, Z., He, F., Zhang, L., Yin, Y.. Optimal deployment of autonomous vehicle lanes with endogenous market penetration[J]. Transport. Res. Part C: Emerg. Technol., 2016, 72, 143–156.
- [9] Levin, M.W., Boyles, S.D.. A cell transmission model for dynamic lane reversal with autonomous vehicles[J]. Transport. Res. Part C: Emerg. Technol., 2016, 68, 126–143.
- [10] Sina Bahrami,Matthew J. Roorda. Optimal traffic management policies for mixed human and automated traffic flows[J]. Transportation Research Part A,2020,135.
- [11] T. Litman, Autonomous vehicle implementation predictions[J], Victoria Transport Policy Institute, 2015, No.15-3326.
- [12] Danjue Chen,Soyoung Ahn,Madhav Chitturi,David A. Noyce. Towards vehicle automation: Roadway capacity formulation for traffic mixed with regular and automated vehicles[J]. Transportation Research Part B,2017,100.
- [13] Wang, X.R., Jiang, L., Li, Y., Lin, X., Zheng, Wang, F.Y.. Capturing car-following behaviors by deep learning[J]. IEEE Trans. Intell. Transport., 2017, Syst. 99, 1–11.
- [14] 刘赫. 动物行为训练的理论基础 [J]. 中国动物保健, 2014(2): 23-25.

- [15] 袁耀明. 交通流元胞自动机模型的解析和模拟研究[D].中国科学技术大学,2009.22.
- [16] M. Cremer, J. Ludwig, A fast simulation model for traffic flow on the basis of boolean operations[J]. *Math. Comput. Simul.*, 1986, 28 (4): 297– 303.
- [17] K. Nagel, M. Schreckenberg. A cellular automaton model for freeway traffic[J]. *Phys. I France*, 1992, 2: 2221–2229.
- [18] O. Biham, A.A. Middleton, D. Levine. Self-organization and a dynamical transition in traffic-flow models[J]. *Phys. Rev. A*, 1992, 46 (10): R6124–R6217.
- [19] M. Takayasu, H. Takayasu. 1/f noise in a traffic model[J]. *Fractals*, 1993, 1 (4): 860–866.
- [20] K. Nagel, M. Paczuski. Emergent traffic jams[J]. *Phys. Rev. E*, 1995, 51 (4): 2909–2918.
- [21] Knospe W, Schadschneider A, Schreckenberg M, et al. Towards a realistic microscopic description of highway traffic[J]. *Journal of Physics A General Physics*, 2000, 33(48): L477.
- [22] Li X, Wu Q, Jiang R. Cellular automaton model considering the velocity effect of a car on the successive car[J]. 2001, 64(6 Pt 2):066128.
- [23] T. Nagatani. Self-organization and phase transition in traffic-flow model of a two-lane roadway[J]. *Phys. A: Math. Gen.*, 1993, 26: 781.
- [24] M. Rickert, K. Nagel, M. Schreckenberg, A. Latour. Two lane traffic simulations using cellular automata[J]. *Phys. A*, 1996, 231: 534.
- [25] P. Wagner, K. Nagel, D.E. Wolf. Realistic multi-lane traffic rules for cellular automata[J]. *Phys. A*, 1997, 234: 687–698.
- [26] Arnab Bose, Petros Ioannou. Mixed manual/semi-automated traffic: a macroscopic analysis[J]. 2002, 11(6):439-462.
- [27] Jincai Chang, Zhuo Wang, Tong Xiao, et al. Modeling and simulations on automated vehicles to alleviate traffic congestion[J]. 2017, 3(2):112-125.
- [28] Danjue Chen, Soyoung Ahn, Madhav Chitturi, et al. Towards vehicle automation: Roadway capacity formulation for traffic mixed with regular and automated vehicles[J]. 2017, 100:196-221.
- [29] Sina Bahrami, Matthew J. Roorda. Optimal traffic management policies for mixed human and automated traffic flows[J]. 2020, 135:130-143.
- [30] Jiazu Zhou, Feng Zhu. Modeling the fundamental diagram of mixed human-driven and connected automated vehicles[J]. 2020, 115.
- [31] Yangzexi Liu, Jingqiu Guo, John Taplin, et al. Characteristic Analysis of Mixed Traffic Flow of Regular and Autonomous Vehicles Using Cellular Automata[J]. 2017.
- [32] Simonelli, F., Bifulco, G., De Martinis, V., Punzo, V.. Human-like adaptive cruise control systems through a learning machine approach[J]. *Appl. Soft Comput.*, 2009, 240–249.

## 北京工业大学毕业设计（论文）

---

- [33] Meixin Zhu, Xuesong Wang, Yinhai Wang. Human-like autonomous car-following model with deep reinforcement learning[J]. *Transportation Research Part C*, 2018, 97.
- [34] 山岩. 自动车拟人化换道决策和换道轨迹研究[D]. 长安大学, 2019.
- [35] LITTMAN M L. Reinforcement learning improves behaviour from evaluative feedback[J]. *Nature*, 2015, 521(7553): 445-451.
- [36] J. Guo, S. Cheng and Y. Liu. Merging and Diverging Impact on Mixed Traffic of Regular and Autonomous Vehicles[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [37] Changxi You, Jianbo Lu, Dimitar Filev, Panagiotis Tsotras. Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning[J]. *Robotics and Autonomous Systems*, 2019.
- [38] CHOWDHURY D, WOLF D E, SCHRECHENBERG M. Particle hopping models for two-lane traffic with two kinds of vehicles: effects of lane-changing rules[J]. *Phys. A: Statistical Mechanics and its Applications*, 1997, 235(3/4): 417-439.
- [39] Sven Maerivoet, Bart De Moor. Cellular automata models of road traffic[J]. *Physics Reports*, 2005, 419: 1-64.
- [40] K. Nagel, H.J. Herrmann, Deterministic models for traffic jams[J]. *Physics. A*, 1993, (199): 254.
- [41] Yangzexi L., Jingqiu G., John T., et al. Characteristic Analysis of Mixed Traffic Flow of Regular and Autonomous Vehicles Using Cellular Automata[J]. *Journal of Advanced Transportation*, 2017, 2017:1-10.
- [42] 郭静秋,方守恩,曲小波,王亦兵,刘洋泽西.基于强化协作博弈方法的双车道混合交通流特性[J].同济大学学报(自然科学版),2019,47(07):976-983.
- [43] Mohsen Kamrani, Aravinda Ramakrishnan Srinivasan, Subhadeep Chakraborty, Asad J. Khattak. Applying Markov decision process to understand driving decisions using basic safety messages data[J]. *Transportation Research Part C*, 2020, 115.
- [44] R. Bellman. A Markovian decision process[J]. *Math. Mech.*, 1957, 679–684.
- [45] R.S. Sutton, A.G. Barto. *Reinforcement Learning: An Introduction*[M], vol. 1, no. 1, MIT Press, Cambridge, 1998.
- [46] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*[J], Springer Science & Business Media, 2013, vol. 31.
- [47] 赵婷婷,孔乐,韩雅杰,任德华,陈亚瑞.模型化强化学习研究综述[J/OL].计算机科学与探索:1-11.<http://kns.cnki.net/kcms/detail/11.5602.tp.20200331.1819.002.html>, 2020-05-12.
- [48] 赵冬斌,邵坤,朱圆恒,李栋,陈亚冉,王海涛,刘德荣,周彤,王成红.深度强化学习综述:兼论计算机围棋的发展[J].控制理论与应用,2016,33(06):701-717.
- [49] Wang S.-C. *Artificial neural network*[M]. *Interdisciplinary Computing in Java Programming*, Springer,

## 北京工业大学毕业设计（论文）

---

2003, pp. 81-100.

- [50] Bernardo J.M., Smith A.F. Bayesian Theory[M]. John Wiley & Sons, Canada, 2001.
- [51] Shi J.Q., Choi T. Gaussian Process Regression Analysis for Functional Data[M]. CRC Press, 2011.
- [52] Cristianini N., Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods[M]. Cambridge University Press, 2000.
- [53] C.J.C.H. Watkins. Learning from Delayed Rewards (Ph.D. dissertation) [D]. King's College, Cambridge, 1989.