

摘 要

本文提出了一个新的东南大学 \LaTeX 硕士研究生毕业论文模板，并说明了如何更优雅地写出一篇漂亮而无用的文章。

关键词： \TeX , \LaTeX , 学位论文

Abstract

This article proposes a new Southeast University master degree thesis \LaTeX template and explains how to elegantly write an article which is beautiful but full of shit.

Keywords: \TeX , \LaTeX , Thesis

目 录

摘 要	I
Abstract	III
第一章 绪论	1
1.1 研究工作的背景及意义	1
1.2 国内外研究	2
1.3 本文的贡献与创新	3
1.4 论文的结构安排	4
第二章 深度强化学习	5
2.1 强化学习	5
2.1.1 相关术语	5
2.1.2 基于价值的强化学习	5
2.1.3 基于策略的强化学习	6
2.1.4 价值与策略相结合的强化学习方法	6
2.2 深度学习	6
2.2.1 神经网络	6
2.2.2 激活函数	6
2.2.3 损失函数优化	6
2.3 深度强化学习	6
2.3.1 深度 Q 学习	6
2.3.2 近端策略优化	6
2.3.3 深度确定性策略梯度	6
第三章 仿真实验场景的设计与构建	7
3.1 城市交通仿真平台的可行性分析	7
3.1.1 Vissim 介绍	7
3.1.2 SUMO 介绍	8
3.1.3 Matsim 介绍	9
3.1.4 仿真平台的选择	10
3.2 基于 SUMO 的城市交通仿真平台	11
3.2.1 平台设计目标	11

3.2.2	功能模块简介	11
3.3	实验场景的选择与搭建	11
3.3.1	路网的编辑与生成	11
3.3.2	出行模式的设计	11
3.3.3	流量的生成	11
第四章	基于深度强化学习的模式与出发时间选择方法	13
4.1	马尔可夫决策过程框架	13
4.1.1	状态表示	13
4.1.2	动作空间	13
4.1.3	奖励函数	13
4.2	基于深度 Q 网络算法的模式与出发时间选择算法	13
4.2.1	神经网络结构设计	13
4.2.2	超参数的选择与标定	13
4.2.3	模型的优化	14
4.3	模型的训练与评估	14
4.3.1	数据收集与预处理	14
4.3.2	模型训练	14
4.3.3	模型的评估	15
第五章	针对多智能体的模式与出发时间选择方法	17
5.1	基于聚类的深度强化学习方法	17
5.1.1	DBSCAN 聚类方法	17
5.1.2	聚类参数的选择	17
5.1.3	深度强化学习模型的改进	17
5.2	模型的验证	17
5.2.1	与传统方法的对比	17
5.2.2	模型参数的灵敏性分析	17
第六章	总结与展望	19
6.1	总结	19
6.2	展望	19
致 谢		21
参考文献		23
作者简介		25

第一章 绪论

1.1 研究工作的背景及意义

随着人口、就业和社会活动的增加，出行需求的增长往往是城市发展不可逆转的结果。这将导致一些大型城市地区的交通状况恶化，在高峰期时，一些主干道的通行能力将会低于出行的出行需求，出现拥堵现象。为了获得准确的出行需求预测并实施有效的需求管理策略，研究人员和政府机构了解出行者在出行时如何进行决策将会至关重要。一旦决策者知道出行者在何时何地以及将采取什么模式出行，就可以提供有效的解决方案来缓解拥堵。因此，出行决策建模成为交通研究的关键。

研究人员通常将出行决策描述为不同维度的备选方案选择，例如出发时间、目的地、方式和路线。这些选择问题通常被描述为离散或者连续选择模型。早期的出行决策模型只考虑了一个维度，即从该选择维度的一组相互排斥的备选方案中选择一个备选方案。然而在实际生活中，需要结合不同行为维度的进行多维决策才足以支持日益增长的拥堵管理策略应用。

近年来，多维出行选择模型得到了更多的关注，因为与传统的一维选择模型不同，多维出行选择模型诠释了不同选择的相关性。在出行选择的问题中，模式选择和出行时间选择是两个非常重要的模块。从个人层面上看，这些都是出行者出行需求的重要性决策。在集计层面上，这决定了交通网络的荷载以及它的时空分布。不同交通方式的可行性与吸引力都取决于它的服务水平，如等待时间、出行时间、出行成本等，这可能会受到各种政策措施的影响，如高峰时段定价、拥堵定价、共乘或公交使用激励。对此类政策措施的评估需要一个出行模式和出发时间选择的综合框架。

模式选择与出行时间选择的研究大多基于随机效用最大化的离散选择模型。如嵌套 Logit 模型、交叉嵌套 Logit 模型和混合 Logit 模型可以应用于多维选择问题^[1]，因为它们具有建模不同选择维度之间相关性的能力。利用随机效用最大化的模型依靠着其强大的理论依据而被广泛地应用。基于随机效用的模型是可以解释出行选择的基本理论，而对于复杂的决策过程建模的适用性，尤其是在选择预测中，可能会受到随机效用函数中线性结构的限制。对于多维选择问题，不同维度之间的关联结构也需要预先确定。

效用的随机成分不仅可以解释出行者对与观察信息的局限性，而且可以考虑决策者的不完全信息和偏好的随机变化。然而，以下事实支持了对基于学习方法的出行选择模型的需求。首先，乘客在模式选择的决策过程，是由不同出行方式的服务水平信息告知和指导的。这些知识通常是通过各种方式获得的(包括出行经验)，并且会随着时间动态变化。第二，出行决策受到一些行为因素的影响，其中乘客更倾向于(更少)选择(改变)他们已经习惯的模式。第三，交通系统的随机性和时间依赖性最可能引起出行者的自适

应模式切换决策,在这种决策中,出行者可能会根据以往的经验更新他们对每种出行模式的预期效用。传统方法不能解决决策过程中涉及的时间维度。因此,与传统的选择建模方法相比,基于学习的出行决策模型更可取。

近年来,强化学习因其强大的探索能力和自主学习能力,已经与监督学习、无监督学习并称为三大机器学习技术 [2]。伴随着深度学习的蓬勃发展,功能强大的深度强化学习算法层出不穷,已经广泛应用于游戏对抗、机器人控制、城市交通和商业活动等领域,并取得了令人瞩目的成绩 [3]。后续大量的研究成果也表明,强化学习是实现通用人工智能的关键步骤。

出行选择的决策过程是一个复杂的过程,会受到环境的影响而不断地变化,通过建立传统的出行选择模型来解释出行行为的方法过于理想化。而此类场景很好地契合了强化学习“无模型、自学习、数据驱动”,使用强化学习的方法可以将此类复杂的模型使用深度神经网络进行描述,通过提取不同外界环境的特征数据如等待时间、出行成本等构建状态输入,再对出行者的出行行为进行优化,利用大数据训练网络增加其真实性和可靠性。相较于传统的离散选择模型,强化学习的方法对复杂的场景适应能力有极大的提升,并且适用的场景更加广泛。

1.2 国内外研究

在出行选择的模型中,通常使用基于随机效用的离散选择模型对不同维度的选择行为进行建模。从 McFadden[4] 在 1973 年提出了著名的 Multinomial Logit (MNL) 模型用于行为选择建模以来,Logit 系列模型就被广泛应用于出行决策问题。然而,MNL 存在一个被公认的问题:它假设了不相关的替代方案的独立性,也被称为 IIA (independence of irrelevant alternatives) 特性。这表示替代方案未观察到的特征彼此之间相互独立,然而在一些出行选择的问题中这个假设将会不成立。例如,在离散出发时间选择中,相邻出发时间区间的未观测特征往往表现出显著的相关性。为了解决这一问题,Ben-Akiva[5] 等在 1998 年提出了 Nested Logit (NL) 模型和有序广义极值模型 (Ordered Generalized Extreme Values, OGEV)。NL 模型能够识别嵌套组内不同替代方案之间的相关性。有序广义极值模型允许为每一对分组的备选方案提供一个相关参数。经过 Bhat[6-8] 在 1998 年的测试,得出的结论是,NL 和 OGEV 模型的性能都优于 MNL。在此之后,不同的研究人员针对问题的多样性提出了更先进的 NL 模型,如 Ben-Akiva 和 Bierlaire [9] 在 1999 年提出的 cross-nested Logit (CNL) 模型和 Lemp[10] 等在 2010 年提出的连续 CNL 模型。另一种改进的离散选择模型是 De Jong 等在 2003 年提出的 mixed Logit(MMNL) 模型,它通过改变 MNL 模型的参数随给定分布变化来考虑个体之间的异质性。然而,MMNL 的一个限制是,它需要对整个人口的参数分布进行特定的假设。这种限制可以通过潜在类 (LC) 模型来解决,该模型可以通过将总体划分为离散数量的类来捕获未观察到的偏好异质性。

另一种研究出行决策的主流方法是机器学习。与统计方法不同,在统计方法中,研

究人员试图确定模型结构和需要估计的参数,机器学习方法关注的是数据本身,并试图找到不同参数之间的关联。相较于随机效用的模型,机器学习模型的结构更加灵活,方便其探索不同特征之间的关联。针对出行决策的建模,主要有以下几种主流的机器学习方法:决策树模型 [11],神经网络模型 [12],以及支持向量机 [13]。与随机效用离散选择模型相比,这些机器学习方法可以处理大型数据库。然而,机器学习方法很少能捕捉到对出行行为研究较为重要的因素,包括时间价值 (VOT) 和弹性。此外,使用机器学习方法作为模型的主要框架还存在一个限制是机器学习模型对训练数据很敏感,在样本不足或有偏倚的情况下,会导致欠拟合或过拟合问题。

强化学习作为一种被广泛应用的学习机制,是利用环境的反馈评价作为学习的输入,学习主体拥有较强的环境适应能力的机器学习方法,因此适用于重复日变的交通决策场景中。强化学习被用来解决各种领域的顺序决策问题,如机器人控制、电子游戏和系统优化等。强化学习的理论为人类行为提供了可解释的心理学和神经科学视角,即人类如何在给定的环境中计划自己的行为。此外,强化学习框架提供了智能决策的数学形式化形式,在智能体控制中具有强大而广泛的适用性,可直接应用于控制理论中顺序决策问题的求解。在交通领域,强化学习方法也受到了广泛的应用,例如交通流管理、自动驾驶,以及路线规划 [14]。近期,一些研究已经采用强化学习方法来建模出行者日常活动计划以及出行决策。

现有的出行决策与出行需求预测的研究工作多使用基于价值的强化学习方法,Janssens[15]在 2007 年使用 Q-learning 的强化学习方法解决活动调度问题。Vanhulsel[16]等在 2009 年通过基于 Q-learning 的方法构建 MATSim 结构方程模型。Medhat[17]等在 2008 年开发了一个更全面的动态公共交通路径和出行活动选择模型,称为 MILITRAS 系统,其中的模型使用了预先设定的奖励(效用)函数。

近几年,深度强化学习在控制复杂智能体的决策行为上取得了巨大的成功,并将强化学习算法与许多神经相关因素的研究相结合,激发了大量使用人工神经网络作为通用函数逼近器的强化学习方法的研究。Hausknecht[18]等在 2016 年发表的著作研究了使用深度强化学习方法与多智能体合作行为。值得注意的是,它将多智能体研究中的矩阵博弈推广到更复杂的状态和行动空间。

1.3 本文的贡献与创新

在使用强化学习的仿真环境中,可以根据不同出行场景将出行者主要分成两种:有电子地图导航和无电子地图导航。在有电子地图导航的场景中,出行者信赖电子地图导航,会根据导航信息一般选择行程最短的模式和路径行驶。在此场景中,出行模式选择问题将转变为备选路径的行程时间预测和预估价问题。可以考虑使用预计到达时间(ETA)的计算方法解决 [22]。在无电子地图导航的场景中,出行者只能依靠自身过往经验,根据经验记忆选择效用最大的出行模式和路径。这种场景下,出行者的每次决策都会得到环境带来的不同反馈,与强化学习的思想相契合。因此这种场景可以使用强化学

习的模型解决。

相较于传统的离散选择模型，强化学习存在以下三点优势：

1. 强化学习模型直接与环境交互，减少了传统离散选择模型的假设限制。传统离散选择模型需要对环境的条件预先假设并检验，在复杂多变的环境下传统模型的弊端将会体现。
2. 强化学习的模型会减少采集数据的成本。一项新的交通政策在实施前需要大量的仿真验证，传统模型需要采集大量的现实数据来验证模型的有效性。强化学习可以基于智能体已知的场景，通过更改仿真环境中的基础设施或策略，使得智能体学习处理未知场景下的决策行为。
3. 考虑智能体记忆能力，贴近实际决策过程。在强化学习的模型中，智能体做出动作后会根据以往经验以及自身的探索不断优化不同决策行为的价值及策略，这与实际中人在进行决策时的惯性一致。

1.4 论文的结构安排

第二章 深度强化学习

2.1 强化学习

2.1.1 相关术语

智能体：任何有独立思想并且可以与所处环境进行交互的实体。在交通场景下，智能体可以是行人，车辆，信号灯等。

状态：当前时刻智能体对周围环境的感知。所有时刻的感知集合构成了状态空间。

动作：智能体在当前状态下采取的行动。在当前状态下能采取的所有动作构成了动作空间。

策略：智能体根据当前时刻的状态应采取哪个动作的控制准则。在数学上的含义为，使用概率密度函数表示智能体在每个状态下采取各个动作的概率。

奖励：在智能体采取动作后，环境对智能体的反馈效果。奖励 R 可以为正反馈或负反馈。

回报：智能体从当前时刻至行动结束所能获得的累积奖励之和。

状态转移：当智能体采取动作后，由当前状态转移到下一个状态的过程。状态转移过程大多数具有随机性，该随机性来源于环境。

2.1.2 基于价值的强化学习

在强化学习中，任意 t 时刻的状态 s_t 下执行策略 π 中的动作 a_t 都存在一个对应的奖励 R_t ，由于强化学习所研究的问题具有马尔可夫性，因此系统的整体回报 U^t 与当前时刻的奖励 R_t 和未来时刻的奖励 $R(t+n)$ 有关，所以存在等式：

式中，是折减因子（discount factor）。

2.1.3 基于策略的强化学习

2.1.4 价值与策略相结合的强化学习方法

2.2 深度学习

2.2.1 神经网络

2.2.2 激活函数

2.2.3 损失函数优化

2.3 深度强化学习

2.3.1 深度 Q 学习

2.3.2 近端策略优化

2.3.3 深度确定性策略梯度

第三章 仿真实验场景的设计与构建

3.1 城市交通仿真平台的可行性分析

3.1.1 Vissim 介绍

VISSIM (VISual SIMulation) 是由德国 PTV (Planung Transport Verkehr AG) 公司开发的一款功能强大的交通模拟软件。它被广泛应用于交通规划、设计、管理等领域。VISSIM 可以模拟各种复杂的交通场景,包括城市道路、高速公路、交叉口、公共交通,以及交通流、交通信号控制、公共交通路线规划、车辆路径规划等。本文将详细介绍 VISSIM 的功能和特点。

VISSIM 的基本功能包括:

1. 建模。VISSIM 的建模功能非常强大。用户可以使用 VISSIM 的图形用户界面轻松创建交通场景,如道路、交叉口和公共交通路线。用户还可以调整交通流率、车辆类型、行人流量和其他参数,以建立一个交通网络。在 VISSIM 的智能交通生成器的帮助下,可以快速、准确地创建交通场景。交通生成器使用真实的交通数据,为一天中的不同时段、工作日和周末创建真实的交通流模式。交通生成器还允许用户定制交通情景,并调整交通量、速度和其他参数。

2. 仿真模拟。VISSIM 提供了一个高保真的交通仿真引擎,可以高度准确地模拟各种交通场景。用户可以运行各种交通管理策略的模拟,如交通信号控制、公共交通路线规划和车辆路径规划。仿真引擎还可以生成实时交通数据,如车辆速度、行驶时间和延迟时间。有了这些数据,用户可以评估不同交通管理策略的性能,比较不同的方案,并做出明智的决定。

3. 分析功能。VISSIM 提供了一个强大的分析功能,使用户能够分析和可视化模拟结果。用户可以生成大量的图表和表格来分析交通流量、拥堵情况、车辆行驶时间、车辆速度和其他指标。分析功能还允许用户比较不同交通管理策略的性能,评估不同参数对交通流的影响。

VISSIM 还提供其他的高级功能:

1. 用户友好的界面。VISSIM 的界面非常友好,它允许用户轻松创建和编辑交通场景。拖放式界面和交通场景的图形表示使用户很容易建立复杂的交通网络。该界面还提供模拟结果的实时反馈,这有助于用户调整参数和评估不同交通管理策略的性能。

2. 定制化设计。VISSIM 提供了高度的自定义功能。用户可以调整交通流、车辆类型和行人流的参数,以创建真实的交通场景。用户还可以定制交通信号灯时间和公共交通时间表,以模拟不同的交通管理策略。此外,VISSIM 提供了一个脚本接口,允许用户通过编写自定义脚本来扩展软件的功能。

3. 多用户支持。**VISSIM** 支持多用户操作，这意味着多个用户可以同时在同一个模拟中工作。这一功能对于需要多个用户在模拟的不同部分工作的大型项目特别有用。

4. 插件支持。**VISSIM** 提供一个插件接口，允许用户扩展软件的功能。开发人员可以使用这个接口来创建插件，为软件增加新的功能，如自定义可视化和分析工具。

5. 多语言支持。**VISSIM** 支持多种语言，包括英语、德语和中文。这一特点使得来自不同国家和语言背景的用户可以使用该软件。

VISSIM 是一个全面而强大的交通模拟软件，为用户提供准确而真实的模拟结果。该软件支持各种类型的交通场景，从小型交叉口到大型城市网络。建模功能允许用户创建一个虚拟的交通网络，模拟引擎产生的实时交通数据可用于评估不同的交通管理策略和交通计划。该软件的高级功能，如定制、多模式模拟、行人模拟、排放和油耗以及驾驶辅助系统，为用户提供了一套全面的工具来评估交通网络的性能和影响。总之，**VISSIM** 是交通专业人员、研究人员和政策制定者设计、分析和优化交通系统的重要工具，以实现更安全、更高效和可持续的未来。

3.1.2 SUMO 介绍

SUMO (Simulation of Urban Mobility) 是由德国航空航天中心开发的一个综合性开源交通模拟软件。它是为城市和区域的交通流分析、优化和规划而设计的。在这篇文章中，我们将详细介绍 **SUMO** 的功能和特点。

SUMO 的基本功能包括：

1. 建模。**SUMO** 提供了强大的建模功能，允许用户创建一个虚拟交通网络。该软件支持各种道路网络，包括高速公路、公路、交叉口和环岛等。用户可以定义网络中的道路布局、交通流和车辆类型。此外，该软件还提供了从各种文件格式（如 **OpenStreetMap**）导入道路网络的工具。

2. 仿真技术。**SUMO** 使用微观交通模拟引擎，提供准确和真实的模拟结果。该软件可以模拟各种交通场景，如车辆路线、公共交通和行人运动。仿真引擎能够生成实时交通数据，如旅行时间、车辆速度和延迟时间。这些数据可以用来评估不同的交通管理策略和交通计划。

3. 可视化。**SUMO** 提供了一个用户友好的图形用户界面，允许用户对模拟结果进行可视化。该界面提供了一个交通网络的实时视图，并可以显示个别车辆、行人和公共交通的运动。此外，该软件还提供了生成各种图表和表格的工具，以总结模拟结果。

SUMO 的高级功能有：

1. 定制功能。**SUMO** 是高度可定制的，允许用户定义影响模拟的不同参数，如车辆类型、司机行为和交通信号。该软件还提供了一个脚本接口，允许用户通过编写自定义脚本来扩展软件的功能。用户还可以创建自定义的车辆模型，并将其导入到模拟中。

2. 多模式仿真。**SUMO** 支持多模式模拟，允许用户模拟多种交通方式，如汽车、公交车、火车和自行车。这个功能可以评估不同交通方式的性能，并对不同的交通计划进

行比较。

3. 优化功能。SUMO 提供优化功能，使用户能够改善交通网络的性能。该软件可以优化交通信号灯时间、车辆路线和公共交通时间表。优化功能允许用户找到最佳的交通管理策略和交通计划。

4. 并行化。SUMO 提供并行化功能，允许用户在多个处理器上运行模拟。这个功能加快了模拟时间，并能模拟更大的交通网络。

5. 集成。SUMO 可以与其他软件工具集成，如交通流模型、地理信息系统（GIS）和数据分析工具。集成功能使数据的交换和定制解决方案的开发成为可能。

总之，SUMO 是一个功能强大、用途广泛的模拟软件，为用户提供了一套全面的工具来模拟和分析各种交通情况。该软件支持多模式交通的模拟，包括汽车、公交车、自行车和行人，它还提供了一些高级功能，如交通需求生成、交通信号控制和车辆路由。SUMO 的灵活架构和开源代码库使其成为研究人员、交通专业人士和需要高度可定制模拟工具的决策者的热门选择。SUMO 有能力生成真实的交通数据，并提供对交通系统性能的洞察力，在设计、优化和评估交通系统方面发挥了关键作用，以实现更安全、更高效和可持续的未来。

3.1.3 Matsim 介绍

MATSim 是一个开源的、基于代理的、多模式的交通模拟软件，为城市或地区的个人行为 and 整个交通系统的建模和模拟提供了一个平台。该软件由瑞士苏黎世联邦理工学院（ETH Zurich）的一个研究团队历经数年开发，目前已被世界各地的研究人员、交通规划人员和政策制定者广泛使用。

MATSim 提供了一套全面的功能来模拟和仿真不同的交通场景。它将个人旅行者建模为代理人，根据他们的偏好和现有的交通选择做出决定。代理人可以选择不同的交通方式，如汽车、公共交通、步行或骑自行车，他们还可以选择最短、最快或最舒适的路线来到达目的地。该软件的先进模拟引擎可以生成大规模的模拟，可以在任何标准计算机系统上运行。

MATSim 软件基于模块化架构，允许用户根据自己的具体需求定制软件。这些模块可以被扩展或被其他模块取代，以模拟新的运输系统或包括新的功能。这一特点使该软件能够适应不同交通场景的具体要求，并使其成为研究人员和交通专业人士的热门选择。

MATSim 的主要特点之一是它能够模拟多模式的交通系统。它可以模拟广泛的交通选择，包括汽车、公共汽车、火车、自行车和步行，并且可以模拟不同交通方式之间的相互作用。这使用户能够评估不同交通系统的性能，并优化城市或区域内不同交通方式的使用。

MATSim 还包括一个全面的可视化工具，允许用户实时查看模拟的结果。该软件的可视化模块使用户能够看到模拟的代理人在交通网络中移动，做出决定，并到达他们的

目的地。可视化模块还可以显示模拟的交通流量、旅行时间和其他重要指标，可用于评估交通系统的性能。

MATSim 的另一个主要特点是其开源代码库。该软件是在 **GNU** 通用公共许可证下发布的，它允许用户修改软件并将其分发给其他人。这一特点使研究人员和交通专业人士能够合作和分享他们的工作，使 **MATSim** 成为交通规划和管理领域的研究和开发项目的热门选择。

此外，**MATSim** 软件已经由一个研究人员和交通专业人士团队进行了广泛的测试和验证。该软件已被用于对全球不同城市和地区的交通系统进行建模和模拟，包括苏黎世、柏林、悉尼、新加坡等。这有助于建立该软件的可信度和可靠性，并证明其对交通系统进行精确建模和模拟的能力。

总之，**MATSim** 是一款功能强大、用途广泛的交通仿真软件，为用户提供了一套全面的工具来模拟和仿真不同的交通场景。该软件的模块化架构、多模式模拟能力、先进的可视化功能和开源代码库使其成为研究人员和交通专业人士的热门选择。**MATSim** 能够准确地模拟交通系统并评估不同交通方案的性能，在设计、优化和评估交通系统以实现更安全、更高效和可持续的未来方面发挥了关键作用。

3.1.4 仿真平台的选择

在这三个交通模拟软件中，我会选择 **SUMO**，原因有很多。

首先，**SUMO** 是一个开源软件，这意味着它可以免费下载和使用。这对于预算紧张的人来说是一个很大的优势，因为使用 **SUMO** 没有任何许可费用。此外，**SUMO** 的开源性质意味着它是高度可定制的，这使得用户可以根据他们的具体需求来修改软件。

其次，**SUMO** 是一个高度通用的软件，可以模拟多模式的交通场景。它可以模拟汽车、公共汽车、自行车、行人，甚至火车。这意味着用户可以对各种交通场景进行建模和模拟，并评估不同交通方案的性能。

第三，**SUMO** 是一个高度精确的模拟软件。该软件以先进的交通流和车辆动力学模型为基础，提供高度真实和准确的结果。这使得 **SUMO** 成为交通专业人士、研究人员和政策制定者的重要工具，他们需要评估交通系统的性能，并优化它们，以实现更安全、更高效和可持续的未来。

第四，**SUMO** 有一个友好的界面，这使得初学者和高级用户都很容易使用。该软件的界面允许用户创建和修改交通网络，生成交通需求，并评估不同交通系统的性能。**SUMO** 还提供了一套全面的可视化工具，使用户可以实时查看模拟结果。

第五，**SUMO** 有一个庞大而活跃的用户社区，它提供了丰富的资源，包括教程、文档和用户论坛。这意味着，如果用户在使用该软件时遇到任何问题，可以很容易地找到支持和帮助。此外，用户社区还提供了一个合作和分享最佳实践和新发展的平台。

最后，**SUMO** 是交通行业中备受推崇和广泛使用的交通仿真软件。它已被用于模拟和仿真世界上许多城市和地区的交通系统，包括欧洲、亚洲和北美。这意味着 **SUMO** 已

经被交通专业人员广泛测试和验证，其结果也被政策制定者和决策者所信任。

总之，SUMO 是一个强大的、多功能的交通仿真软件，为用户提供了一套全面的工具来模拟和仿真不同的交通场景。它的开源性、多模式模拟能力、准确性、用户友好的界面、活跃的用户社区和受人尊敬的声誉使它成为交通专业人士、研究人员和决策者的热门选择。凭借其准确模拟交通系统和评估不同交通方案性能的能力，SUMO 在设计、优化和评估交通系统以实现更安全、更高效和可持续发展的未来方面发挥了关键作用。

3.2 基于 SUMO 的城市交通仿真平台

3.2.1 平台设计目标

3.2.2 功能模块简介

3.3 实验场景的选择与搭建

3.3.1 路网的编辑与生成

3.3.2 出行模式的设计

3.3.3 流量的生成

第四章 基于深度强化学习的模式与出发时间选择方法

4.1 马尔可夫决策过程框架

4.1.1 状态表示

4.1.2 动作空间

4.1.3 奖励函数

4.2 基于深度 Q 网络算法的模式与出发时间选择算法

4.2.1 神经网络结构设计

The first step in designing a DQN model is to choose an appropriate neural network architecture.

The architecture should be capable of taking the state as input and producing Q-values for each action in the action space as output.

Convolutional neural networks (CNNs) are commonly used for problems with image-based inputs, while fully connected neural networks can be used for problems with vector-based inputs.

The number and size of layers in the network can also vary depending on the complexity of the problem.

4.2.2 超参数的选择与标定

Hyperparameters play a crucial role in the performance of a DQN model.

The learning rate, batch size, discount factor, and other hyperparameters should be carefully selected to optimize the performance of the model.

The learning rate controls the rate at which the model updates its weights during training.

The batch size determines how many state-action pairs are processed in each training iteration.

The discount factor determines the balance between immediate and future rewards in the Q-value calculation.

Other hyperparameters, such as the size of the replay buffer or the frequency of target network updates, can also have a significant impact on model performance.

4.2.3 模型的优化

Various optimization techniques can be used to improve the stability and convergence of a DQN model.

Experience replay is a technique where past experiences are stored in a buffer and sampled randomly during training.

Target networks can be used to estimate the Q-values of the next state to improve stability during training.

Prioritized experience replay can be used to sample more important experiences more frequently.

Reward shaping can be used to modify the reward function to incentivize certain behaviors.

The selection and use of these optimization techniques can depend on the specific problem and the network architecture used.

4.3 模型的训练与评估

4.3.1 数据收集与预处理

Before training a DQN model, it is important to preprocess the input data to make it suitable for the neural network.

This can include steps such as normalization, scaling, and cropping of images.

The RL agent also needs to collect data by interacting with the environment.

During this process, the agent chooses actions based on the current state, receives a reward, and transitions to the next state.

The state, action, reward, and next state are stored in a replay buffer, which is then used for training.

4.3.2 模型训练

To train a DQN model, the neural network is updated using stochastic gradient descent with mini-batches of state-action pairs from the replay buffer.

During each training iteration, the model predicts Q-values for the current state, and the Q-value for the chosen action is compared to the target Q-value, which is calculated using the Bellman equation.

The difference between the predicted and target Q-values is used to calculate the loss, which is then backpropagated through the network to update the weights.

To improve stability and convergence, various techniques such as experience replay, target networks, and prioritized experience replay can be used.

4.3.3 模型的评估

Once the DQN model is trained, it is evaluated to assess its performance.

The agent interacts with the environment using the trained model, and the cumulative reward obtained over a fixed number of episodes is used as a metric of performance.

The agent may also be tested on a holdout set of data to assess its generalization capabilities.

If the performance of the model is not satisfactory, the hyperparameters or model architecture can be adjusted, and the training and evaluation process can be repeated.

第五章 针对多智能体的模式与出发时间选择方法

5.1 基于聚类的深度强化学习方法

5.1.1 DBSCAN 聚类方法

5.1.2 聚类参数的选择

5.1.3 深度强化学习模型的改进

5.2 模型的验证

5.2.1 与传统方法的对比

5.2.2 模型参数的灵敏性分析

第六章 总结与展望

6.1 总结

本模板基于宋睿同学发布在[SEU-master-thesis](#) 并在上述工作的基础上进行了微调, 解决了一些自己编写代码过程中 BUG。

6.2 展望

致 谢

感谢许元和樊智猛等前人的工作，没有他们的工作也就不会有这个模板的诞生。也感谢使用该模板的每一个人，因为你们的开放与进取心使得 \LaTeX 在东南大学的氛围越来越好。

参考文献

- [1] 李梦凡. 交通信息诱导下个体出行选择行为研究[D]. 合肥工业大学, 2018.

作者简介

知心哥哥 (1996.3.3 -), 男, 台湾南昌人, 现居台湾重庆, 为网易 CC 丢人主播, 主要研究方向有《黑旗》、《黑楼》和《黑暗剑》。

作者攻读硕士学位期间发表的论文

- [1]. **ZHi X**, WEI T, CHEN R, et al. Sea of Thieves: Fucking Animal[C]. 2017 810th International Conference on Disgraced (ICD). Chongqing, 2017. 1-6. (EI Indexed)
- [2]. **ZHi X**, WEI T, JI H, et al. Animal Crossing: Playing Together[J]. Nintendo Daily Journal, 2020, 13(3): 114-514. (SCI Indexed)
- [3]. 韦天, 知心哥哥. 你怪猎来我大圣: 3D 游戏之耻 [C]. 第 19 届口吐芬芳游戏评测国际会议 (SFGR). 台湾南昌, 2019. 1-14. (EI Indexed)

作者攻读硕士学位期间参与的研究课题

- [1]. **2018.5-2019.2**: 底特律便乘人暨强人工智能的实现
- [2]. **2020.1-2020.3**: 大老爹拿球的概率模型研究

