编号_____

南京航空航天大学

毕业论文

题　目　基于深度强化学习智能交通信号调度

学生姓名　陈建

学　　号　SX1916039

学　　院　计算机科学与技术学院

专　　业　计算机科学与技术

班　　级　1318001

指导教师　朱琨

二〇二一年十二月

南京航空航天大学
本科毕业论文诚信承诺书

本人郑重声明:所呈交的毕业论文(题目:基于深度强化学习智能交通信号调度)是本人在导师的指导下独立进行研究所取得的成果。尽本人所知，除了毕业论文中特别加以标注引用的内容外，本毕业论文不包含任何其他个人或集体已经发表或撰写的成果作品。

作者签名:　　年　月　日

(学号):

基于深度强化学习智能交通信号调度

摘 要

本文介绍如何使用NUAA Thesis 文档类撰写南京航空航天大学学位论文。

首先介绍如何获取并编译本文档，然后展示论文部件的实例，最后列举部分常用宏包的使用方法。

关键词：学位论文，模板，NUAA Thesis

NUAA2 Thesis Quick Start and Document Snippets

Abstract

This document introduces NUAA2 Thesis, the LaTeX document class for NUAA Thesis.

First, we show how to get the source code and compile this document. Then we provide snippets for figures, tables, equations, etc. Finally we enforce some usage patterns.

Key Words: NUAA thesis, document class, space is accepted here

目录

第一章　背景和研究意义

第二章 概述

2.1 交通信号概述

交通信号控制是一个重要而具有挑战性的现实问题，其目的是通过协调车辆在道路交叉口的运动来最小化所有车辆的通行时间。目前广泛使用的交通信号控制系统仍然严重依赖过于简化的信息和基于规则的方法。车联网技术的发展、硬件性能的提升以及人工智能技术的进步使得我们现在有更丰富的数据，更多的计算能力和先进的方法来驱动智能交通的发展。交通信号控制的目的为了方便车辆在交叉路口的安全和高效移动。安全是通过信号灯指定不同车道的车通行来分离相互冲突的运动实现的。为了能够有效地优化通行效率，已有的工作提出了不同的指标来量化通行效率，主要有以下三个：1. 通行时间：在交通信号控制中，车辆的行驶时间被定义为一辆汽车进入系统的时间与离开系统的时间的差值。最常见的优化目标之一是就是减少进过路口的所有车辆的平均通行时间。2. 队列长度：队列长度是指路口等待车辆的数量，越大的队列长度意味着越多的等待车辆，路口的通行效率越低，反之通行效率越高。3. 路口吞吐量：吞吐量是指在一定期间内进过路口完成通行的车辆数量。越大的吞吐量代表着越高的通行效率，所以很多工作将最大化吞吐量作为优化的目标。

2.2 基本术语

Approach: 指交叉路口的巷道。任何一个交叉路口都有两种 approach, 进入路口的 incoming approach 和离开路口的 outgoing approach。图 3.a 描述了一个典型的有 8 个 approach（四个入口，四个出口）的交叉路口。Lane: 一个 Approach 是由一组车道组成。与 Approach 的定义类似，车道也分为两种：转入车道（incoming lane）和转出车道 (outgoing lane)。Traffic movement：指的是车

流从一个 incoming approach 运动到另一个 outgoing approach，表示为，其中 ri 和 ro 分别表示 incoming lane 和 outgoing lane。通常，traffic movement 可以分为左转、直行以及右转三种，在少数特殊的路口也支持 U-turn 的 traffic movement。Movement signal: 根据 traffic movement 定义的运动信号，绿色代表可以通行，红色代表禁止通行。根据大多数国家的交通规则，右转的 traffic movement 是可以不受信号约束的。Phase: 信号灯的一个 phase(相位) 是指非冲突运动信号的组合，这意味着这些信号可以同时设置为绿色，而不会引起安全冲突。图 3.c 展示了最常用的四相位信号模式。Phase sequence: 相序，即一组相位的序列，它定义了一组相位及其变化顺序。Signal plan：信号计划，由一组相位序列及其相应的起始时间组成。通常表示为其中和分别代表相位及其开始时间。Cycle-based signal plan：周期性信号计划，与普通的信号计划不同的是其中的相位序列是按循环顺序工作的，可以表示为其中是重复出现的相位序列，是 j 周期中相位 pi 的起始时间。具体地，是第 j 周期的周期长度，是第 j 周期中的相位分裂比（phase split ratio），表示每个相位持续时间占总周期长度的比重。现有的交通信号控制方法通常在一天中重复类似的相位序列。

2.3 传统交通控制方法

2.3.1 Webster

对于单个交叉口，交通运输工程领域中的交通信号控制方法通常由三个部分组成：确定信号周期长度，确定信号相位序列以及相位分裂。Webster 是一种广泛使用的计算单个交叉路口的信号周期长度和相位分裂时间的方法。通过假设车流在一段时间内（例如，过去的五分钟或 10 分钟）是均匀到达的，可以计算出确切的最优周期和最佳相位分裂时间，从而最小化车量通行时间。

2.3.2 GreenWave

虽然使用 Webster 可以简单的控制单个交叉路口的交通信号，但是对于相邻的多个交叉路口，不能够简单地直接使用 Webster 来分别优化每一个路口，相邻路口信号灯的信号时间之间的偏移（即相邻路口信号周期起始时间的差

值）也需要进行优化，因为对于相距较近的路口来说，一个路口的控制策略可能会影响到其他路口。GreenWave 就是交通运输领域中最经典的协调相邻路口的信号控制方法，它通过优化相邻路口信号时间的偏移来减少车辆在某一方向行驶时的停留次数。这种方法可以形成沿指定交通方向的绿色信号波，在该方向行驶的车辆可以受益于渐进的绿色信号级联，而不会在任何交叉口停留，如下图所示：

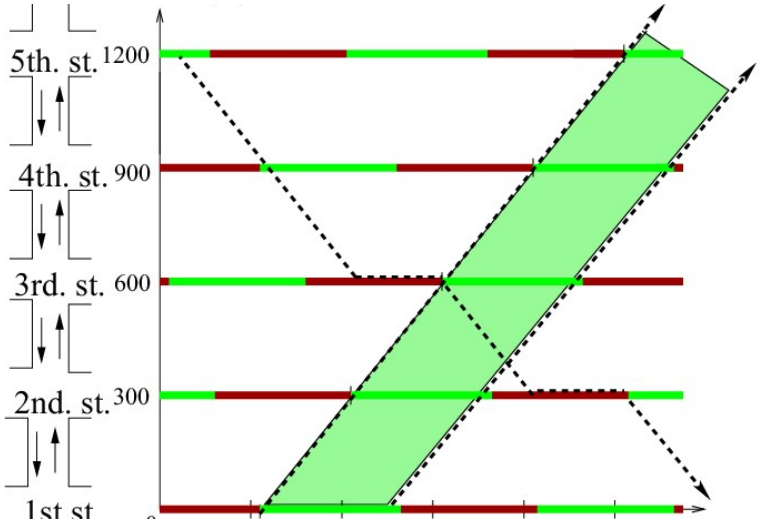

图 2.1　路口敏感性说明

2.3.3　Actuated Control

Actuated Control 根据当前相位和其他的竞争相位对绿色信号的请求来决定是否保持或者变化当前的相位。请求规则如下：1. 当目前相位的持续时间未达到最小时间周期时，或在当前相位对应入车道上有车辆进入，并且在接近信号的距离内时，就会产生延长绿色信号时间的请求，已让车辆可以直接通过路口。2. 当竞争相位的等待车辆数量大于一个阈值时，就会生成对绿色信号的请求。根据规则的差异，Actuated Control 主要可以分为 Fully-Actuated Control 和 Semi-Actuated Control 两种。

2.3.4 SOTL

Self-Organizing Traffic Light Control(SOTL) 是一种具有附加需求响应规则的 Fully-Actuated Control 方法。它与 Fully-Actuated Control 的主要区别在于当前相位的绿色信号请求定义（虽然它们都需要最小的绿色相位持续时间）：在 Fully-Actuated Control 中，当车辆接近信号灯时，就会产生延长绿色信号的请求，而在 SOTL 中，除非接近信号灯的车辆数量大于不一定是一个阈值，否则就不会产生请求。

2.3.5 Max-Pressure Control

Max-Pressure Control 的目的是通过最小化对应相位的压力（pressure）来平衡相邻路口之间的队列长度，从而降低过饱和的风险，其中压力的概念如下图所示：从形式来看，运动信号的压力可以定义为（交通运动的）传入车道上

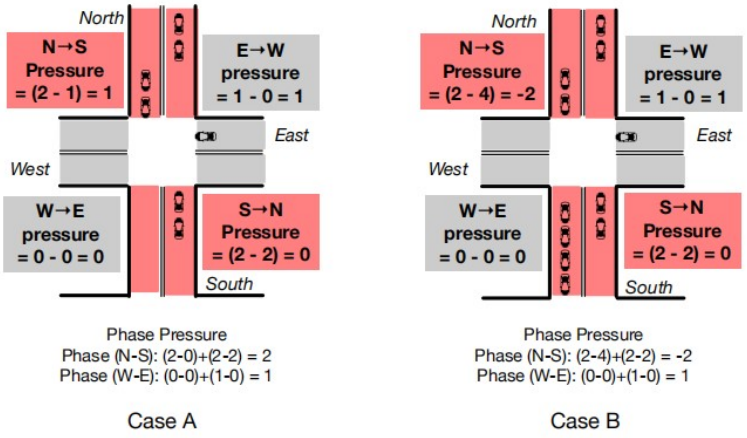

图 2.2 路口敏感性说明

的车辆数减去相应的传出车道上的车辆数；相位的压力定义为传入通道和传出通道上的总队列长度之间的差异。Varaiya 等人证明了当将优化目标设为最小化单个路口的相位压力时，Max-Pressure Control 可以最大限度地提高真个路网的吞吐量。

下表列出了每种方法的限制和要求：

表 2.1 常见的协作策略

方法	先验信息	输入	输出
Webster	相位序列	交通流量	基于周期的单个路口信号计划
GreenWave	信号计划	交通流量、速度限制、车道长度	基于周期的信号计划的偏移量
Actual Control，SOTL	相位序列	交通流量	是否变化到下一个相位
Max-Pressure Control	无	队列长度	所有交叉口的信号计划

2.4 基于强化学习的交通信号控制

最近，人们提出了不同的人工智能技术来控制交通信号，例如遗传算法、群体智能以及强化学习。其中在这些技术中，强化学习在近年来更具趋势。

2.4.1 强化学习概述

通常单智能体强化学习问题被建模成$\mathrm{MDP} < \mathcal{S}, \mathcal{A}, P, R, \gamma >$,其中$\mathcal{S}, \mathcal{A}, P, R, \gamma$分别表示状态集、动作集、概率状态转移函数、奖励函数和折扣因子。具体定义如下:

- $\mathcal{S}$：在时间步骤t，智能体得到一个观测状态$s^t \in \mathcal{S}$。
- $\mathcal{A}, P$：在时间步骤t，智能体采取一个动作$a^t \in \mathcal{A}$，然后环境根据状态转移函数转移到一个新的状态。

$$P\left(\mathrm{s}^{t+1} \mid \mathrm{s}^t, \mathrm{a}^t\right) : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S} \tag{2.1}$$

- R：在时间步骤t，智能体通过奖励函数获得一个奖励r^t。

$$R\left(s^t, a^t\right) : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R} \tag{2.2}$$

- γ：智能体的目标是找到一种使预期收益最大化的政策，即折扣奖励之和。折扣因子决定了即时奖励与未来奖励的重要性。

$$G^t := \sum_{i=0}^{\infty} \gamma^i r^{t+i} \tag{2.3}$$

2.4.2 基于强化学习交通信号控制框架

2.4.3 基本要素

第三章 工作

3.1 公平性研究

公平性问题是指，不同车辆通过同一个路口的通行时间可能有很大的差别，因为信号灯可能为了提高整体通行效率而牺牲一些车辆，让这些车辆多等待一些时间，即便这些车辆可能是先进入路口的，这对这些车来说是不公平的。一个好的控制策略应该在提高通行效率的同时能够保证每辆车所需的通行时间大致相同，也就是说，车辆通行时间的方差应该越小越好。但是已有的工作都是使用车辆的平均通行时间来衡量通行效率，很自然的忽略了公平性问题。

3.1.1 目标

本工作的目的是在提高通行效率（最小化平均通行时间）的同时，希望每条车道能够有尽可能相同的服务延迟（得到放行所需的时间）。这个目标可以用以下的 Jain Fairness Index(JFI) 指标来量化：

$$\mathcal{J} = \frac{(\sum_{i=1}^{M} \overline{D}_i)^2}{M \sum_{i=1}^{M} \overline{D}_i^2}, \tag{3.1}$$

其中 $\overline{D}_i$ 是第 i 条进近车道的平均延迟。当且仅当每一个 $\overline{D}_i$ 都相等时，这个指标达到最大值，即是 1。所以我们的目标也就是最大化这个指标。

3.1.2 智能体设计

状态表示

在 t 时刻的状态 $S(t)$ 由以下几个部分组成：

1. 交通流量：$\boldsymbol{V}(t) = \{V_1(t), V_2(t), \cdots, V_M(t)\}$。其中 $V_i(t)$ 表示第 i 条进近车道上车的数量。值得注意的是，由于右转不受限于信号灯的特殊性，这里我们不考虑右车道的交通流量。

2. 平均吞吐量：$\overline{\boldsymbol{L}}(t) = \{\overline{L}_1(t), \overline{L}_2(t), \cdots, \overline{L}_M(t)\}$。其中 $\overline{L}_i(t)$ 表示第 i 条进近车道的平均吞吐量。同上，不考虑右车道的平均吞吐量。
3. 信号相位：$\boldsymbol{P}(t)$ 是当前信号相位的数字化表示，1 表示绿色，可以通行；0 表示红色，禁止通行。

所以 $S(t) = \{\boldsymbol{V}(t)||\overline{\boldsymbol{L}}(t)||\boldsymbol{P}(t)\}$

动作选择

在本文中，动作选择机制是每次选择即将转换的信号相位。之后，交通信号灯将转换到这一新的相位并持续 Δt 的时间。为了安全起见，我们在两个不同的信号相位之间插入了 3 秒的黄色信号和 2 秒的红色信号。如果新选择的相位和当前相位相同，则不插入黄色和红色信号，以确保交通流畅。

奖励函数

受 PFS 分配原则的启发，我们设计了一个可以在效率和公平之间提供良好的平衡的奖励函数，如下所示：

$$r = -\sum_{i=1}^{M} \frac{Q_i(t)}{\overline{L}_i(t) + \delta}, \tag{3.2}$$

其中 $Q_i(t)$ 和 $\overline{L}_i(t)$ 分别是第 i 条进近车道的队列长度和平均吞吐量。在每一次调度后（这里，我们将一次动作选择视作一次调度），$\overline{L}_i(t)$ 按照以下方式进行更新：

$$\overline{L}_i(t) = (1 - \frac{1}{W})\overline{L}_i(t-1) + \frac{1}{W}L_i(t), \tag{3.3}$$

其中 $L_i(t)$ 是此次调度中车道 i 上得到放行的车的数量，W 是一个平衡通行效率和公平性的参数。另外，为了避免公式3.2的分母为 0，我们加上了一个可以忽略不计的正数 δ。

训练过程

训练过程伪代码如下：

3.2 通信问题研究

对于多路口的交通信号调度问题，协作（Coordination）可以有效地提升整体通行效率，以下列出几种常见的协作策略：

表 3.1 常见的协作策略

协作策略	目标	说明
Global single agent	$max_{\mathbf{a}}Q(s,\mathbf{a})$	s 是全局的环境状态，$\mathbf{a}$ 是所有路口的联合动作。
Independent RL without Communication	$max_{a_i}\sum iQ_i(o_i,a_i)$	o_i 是路口 i 的局部观测，a_i 是路口 i 的动作。
Independent RL with Communication	$max_{a_i}\sum iQ_i(\Omega(o_i,\mathcal{N}_i),a_i)$	$\mathcal{N}_i$ 是路口 i 的邻近路口的状态表示，$\Omega(o_i,\mathcal{N}_i)$ 是整合路口 i 及其邻近路口状态表示的函数。

3.2.1 图神经网络

3.2.1.1 分类及介绍

深度网络的研究推进了模式识别和数据挖掘领域的发展。借助于计算资源的高速发展（如 GPU），深度学习在欧几里得数据（如图像、文本和视频）中取得巨大的成功。但是在一些应用场景下，数据（图）是由非欧几里得域生成的，任然需要有效分析。例如，在电子商务领域，一个基于图的学习系统能够利用用户和商品之间的交互以实现精准的推荐。在化学领域，分子被建模为图，新药研发需要测定其生物活性。在论文引用网络中，论文之间通过引用关系互相连接，需要将它们分成不同的类别。

图数据的复杂性对现有机器学习算法提出了巨大的挑战，因为图数据是不规则的。每张图大小不同、节点无序，一张图中的每个节点都有不同数目的邻近节点，使得一些在图像中容易计算的重要运算（如卷积）不能再直接应用于

图。此外，现有机器学习算法的核心假设是实例彼此独立。然而，图数据中的每个实例都与周围的其它实例相关，含有一些复杂的连接信息，用于捕获数据之间的依赖关系，包括引用、朋友关系和相互作用。最近，越来越多的研究开始将深度学习方法应用到图数据领域。受到深度学习领域进展的驱动，研究人员在设计图神经网络的架构时借鉴了卷积网络、循环网络和深度自编码器的思想。

图神经网络的概念最早由 Gori[1] 等人提出，由 Scarselli[2] 等人进一步阐明。早期的初期是以迭代方式通过循环神经网络架构传播邻近信息来学习目标节点的表示，直至达到稳定的状态。

图神经网络可以分为：图卷积网络（Graph Convolution Network），图注意力网络（Graph Attention Network），图自编码器（Graph Auto-encoder），图生成网络（Graph Generative Network）和图时空网络（Graph Spatial-Temporal Network）。图卷积网络

图注意力网络是将注意力机制引入到基于空间域的图神经网络。图神经网络不需要使用拉普拉斯等矩阵进行复杂的计算，仅通过邻居节点的表征来更新目标节点的特征。由于能够放大数据中最重要部分的影响，注意力机制已经广泛应用到很多基于序列的任务中，图神经网络也受益于此，在聚合过程中使用注意力整合多个模型的输出。

图自编码器是一类图嵌入方法，其目的是利用神经网络将图的顶点表示为低维向量。典型的解决方案是利用多层感知机作为编码器来获取节点嵌入，其中解码器重建节点的邻域统计信息，如 positive pointwise mutual information（PPMI）或一阶和二阶近似值。

图生成网络的目标是在给定一组观察到的图的情况下生成新的图。图生成网络的许多方法都是特定于领域的。例如，在分子图生成中，一些工作模拟了称为 SMILES 的分子图的字符串表示。在自然语言处理中，生成语义图或知识图通常以给定的句子为条件。

图时空网络同时捕捉时空图的时空相关性。时空图具有全局图结构，每个节点的输入随时间变化。例如，在交通网络中，每个传感器作为一个节点连续记录某条道路的交通速度，其中交通网络的边由传感器对之间的距离决定。图形时空网络的目标可以是预测未来的节点值或标签，或者预测时空图标签。最近的研究仅仅探讨了 GCNs 的使用，GCNs 与 RNN 或 CNN 的结合，以及根据图结构定制的循环体系结构。

3.2.1.2 应用分类

以图结构和节点特征信息作为输入，根据输出的类别，可以将 GNN 的分析任务分为以下几类：节点级别：这一类聚焦于节点回归和节点分类任务边级别：图级别：

使用 GAT 来学习 communication，这样做有一下两个好处：1. 动态学习周边的路口的重要性：已有的工作直接将目标路口及其邻近路口的状态直接整合起来，这种做法实际上是默认每一个邻近路口对目标路口的影响力是相同的。实际上由于交通在时间和空间上的变化，同一个路口对于其目标路口的影响力也会发生变化。2. 免索性模型学习（Index-free modeling learning）：在多智能体场景下，通常需要使用参数共享（Parameter Sharing）来降低学习难度，从而加速学习。但是这一点在多路口信号控制场景下是不适用的，因为不同路口对其邻近路口的敏感性是不同的。例如，如下图所示，有 A，B 两个路口，A 对其 N-S 方向的交通更加敏感，B 对其 W-E 方向的交通更加敏感，如果直接将 A 的参数分享给 B，会导致 B 学习到的策略并不适用于自身的场景。

使用 IRL with communication 的方法确实可以有效的解决维度灾难的问题。已有的工作使用图神经网络来学习“交流”这一个过程。他们将每一个路口视作图中的一个节点，每条道路作为连接两个节点的边，很自然地可以将一张交通道路网建模成一个图。这种按路口建模的方式如下所示：

在这种建模方式下，每条车道的车辆以及当前的相位将作为该节点的特征。这种建模方式虽然可以很清晰的将多路口场景变成一张图。但是，因为是

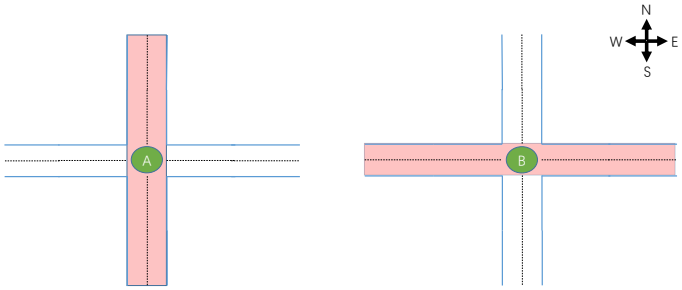

图 3.1　路口敏感性说明

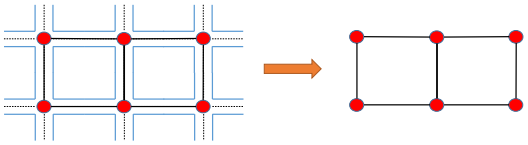

图 3.2　多路口建模成图（路口）

以一个路口为一个节点，所有车道的状态信息都整合到了一起，有些车道的的信息对目标节点是无用的，如下图所示：

路口 B 中只有 2 车道的交通流向与 A 车道有关，1、3 车道的车辆不会行驶到 A 路口。在信息传递的时候，如果将所有的信息都笼统地传递过去，将会增加 A 提取有效信息的难度，从而降低学习的效率。

此外

在本文中，我们采用 GAT 来学习 communication，不同的时，我们采用不同的图建模方式。我们不是按照路口来建图，而是按照道路来进行建模，即一

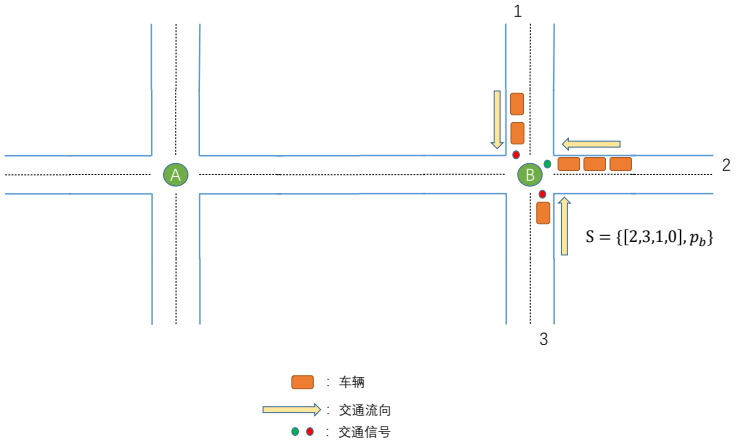

图 3.3　按路口建图模式下信息传递

条道路就是一个节点，如下图所示：

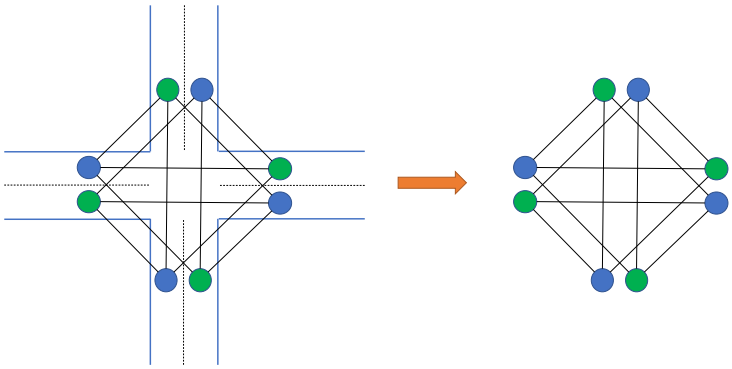

图 3.4　按道路建模成图

然后我们将相位信息加入到图的结构信息中，这里我们规定边有一个权重值 WE，如果在当前相位下，道路 i 到道路 j 之间是允许通行的，则连接这两个节点的边的权重 $WE_{ij}=1$，如果不允许通行，则 $WE_{ij}=0$。这个权重将用作后续剔除信息无关节点（该节点的信息对目标节点没有影响）。

这里我们使用注意力机制（Attention Mechanism）学习来学习邻近节点的表示对目标节点状态的影响，从而实现通信的目的。整个过程可以分解为以下几个步骤：

1. 观察交互（Observation Interaction）为了了解来自路口 j（源节点）的信息在确定路口 i（目标节点）的策略的重要性，我们首先嵌入（Embedding）来自前一层的这两个节点的表示并用来计算 e_{ij}（在确定路口 i 的策略时路口 j 的重要性），按照下列的操作：

$$e_{ij} = (h_i W_t)\dot{(}h_j W_s)^T, \tag{3.4}$$

其中 $W_s, W_t \in \mathbf{R}^{mxn}$ 分别是源节点和目标节点的嵌入参数。值得注意的是，这里 e_{ij} 不一定等于 e_{ji}。例如，

2. 无关信息剔除

$$e_{ij} = e_{ij} * WE_{ij}, \tag{3.5}$$

这里 WE_{ij} 是连接 i 和 j 的边的权重。

3. 计算邻域范围的注意力分布为了获取目标节点和源节点之间的注意力值，我们将目标节点及其邻近节点之间的互动分数（即，e_{ij}）进行归一化：

参考文献

[1] Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains[C]. Proceedings of Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., volume 2. IEEE, 2005. 729–734.

[2] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE transactions on neural networks, 2008, 20(1):61–80.

参考文献

[1] 本节演示如何手写参考文献目录
[2] 如果论文能用 biber 来管理参考文献的话，请使用 biber，不要手写
[3] 如果实在不方便用 biber 的话，可以使用这种方法来手写参考文献。格式完全手写会有点繁琐，而且不能在正文中引用。比如：
[4] KANAMORI H. Shaking without quaking[J]. Science, 1998, 279(5359): 2063.
[5] 吴云芳. 面向中文信息处理的现代汉语并列结构研究 [D]. 北京: 北京大学,2003[2013-10-14].

示例：[1] 这种写法不符合学校的要求，推荐使用这种写法[4]。

致　谢

在此感谢对本论文作成有所帮助的人。