



多智能体强化学习在智慧城市中的应用

演讲人：蒋炆峻



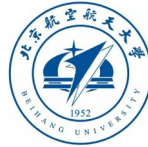
目录

CONTENTS

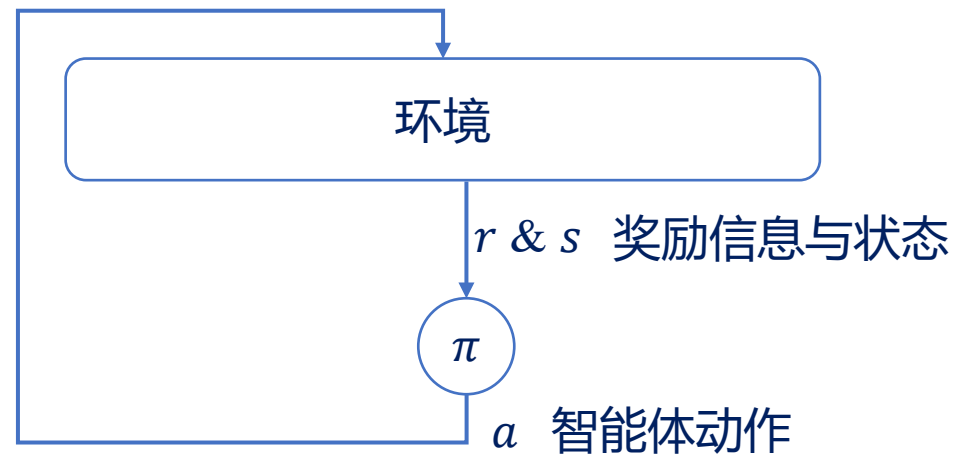
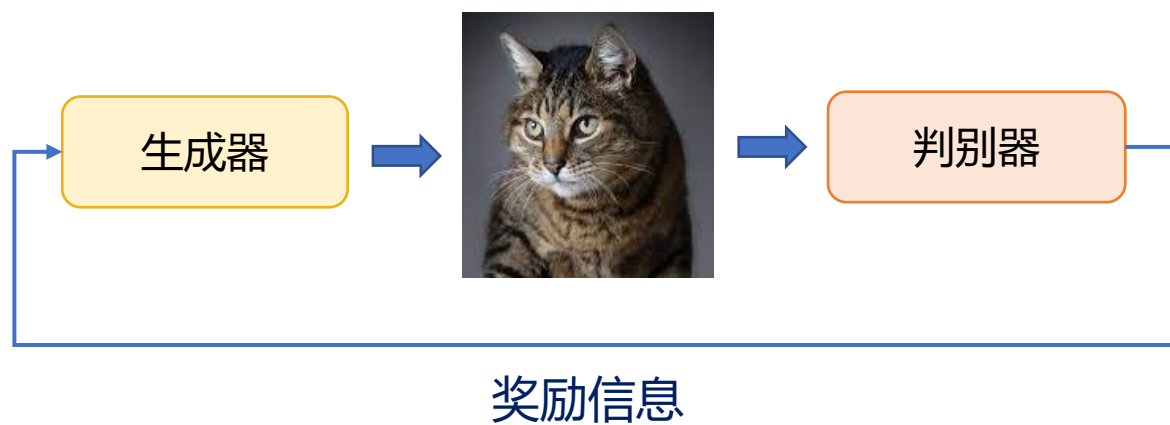


- 01** ▶ 多智能体强化学习介绍
- 02** ▶ 应用一：智能信号灯控制
- 03** ▶ 应用二：多智能体路径规划

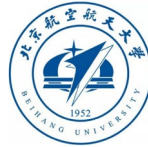
多智能体强化学习介绍



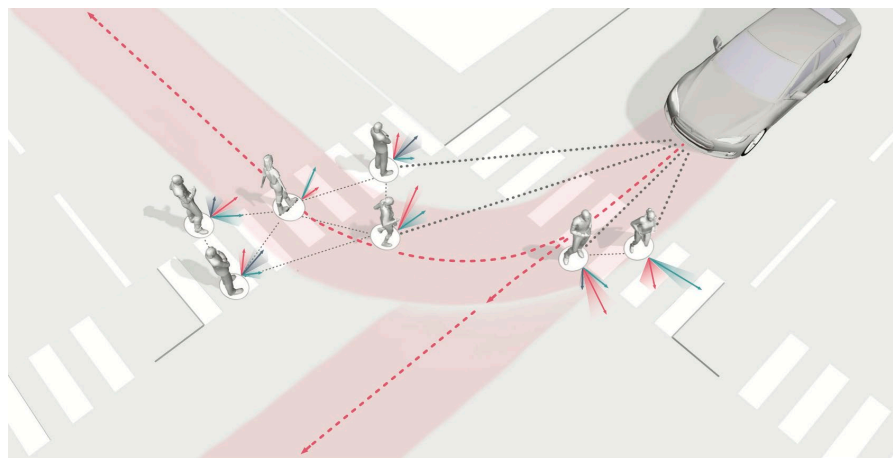
- 多智能体强化学习：环境中存在多个实体，同时与环境进行交互。
 - 多智能体 vs 单智能体



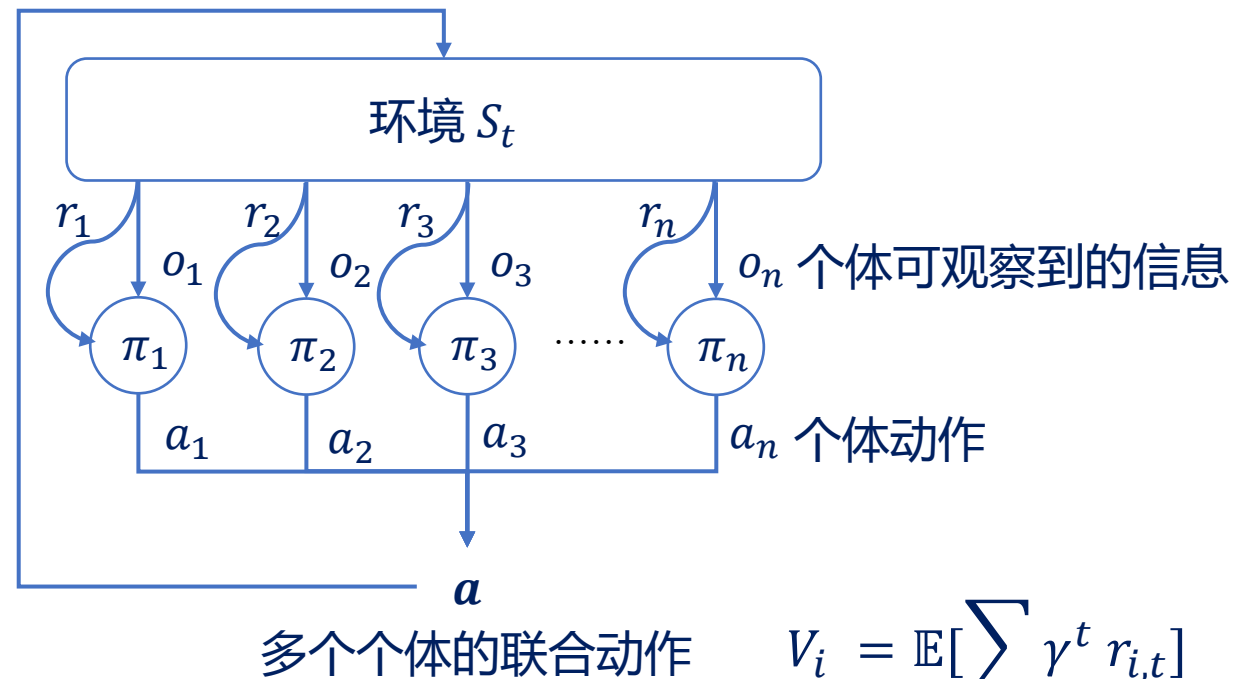
多智能体强化学习介绍



- 多智能体强化学习：环境中存在多个实体，同时与环境进行交互。



城市交通是多智能体系统

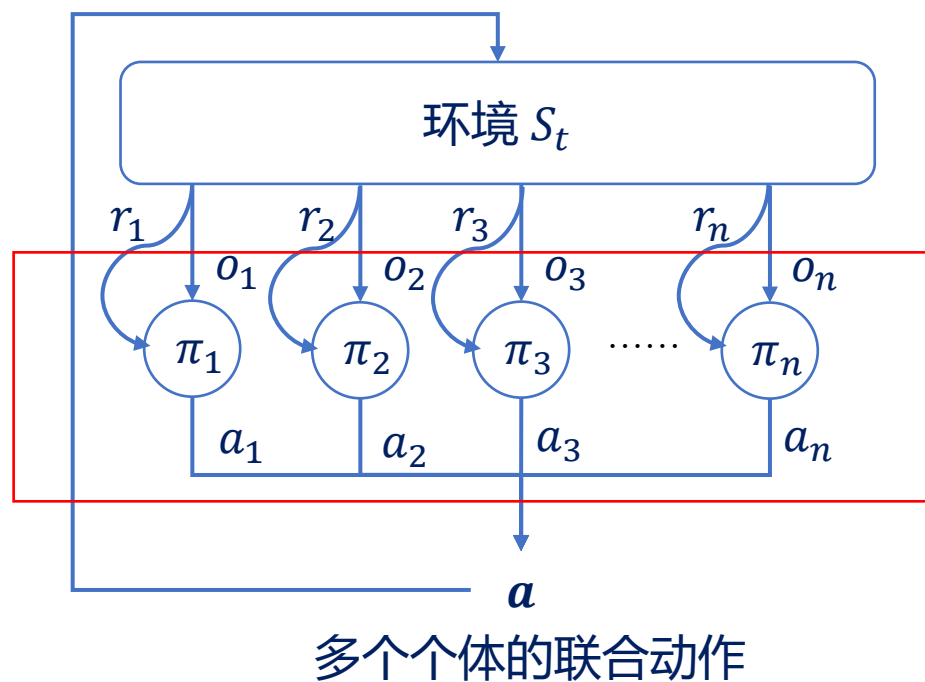


$$V_i = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r_{i,t} \right]$$

• 新挑战



- 新挑战：环境的非平稳性
 - 对于某一代理，其之前记录到的 <动作, 状态, 奖励> 信息是不可靠的。



解决方案 1: 中心化 (整体化)

缺陷: 不易部署与扩展

• 新挑战



目录

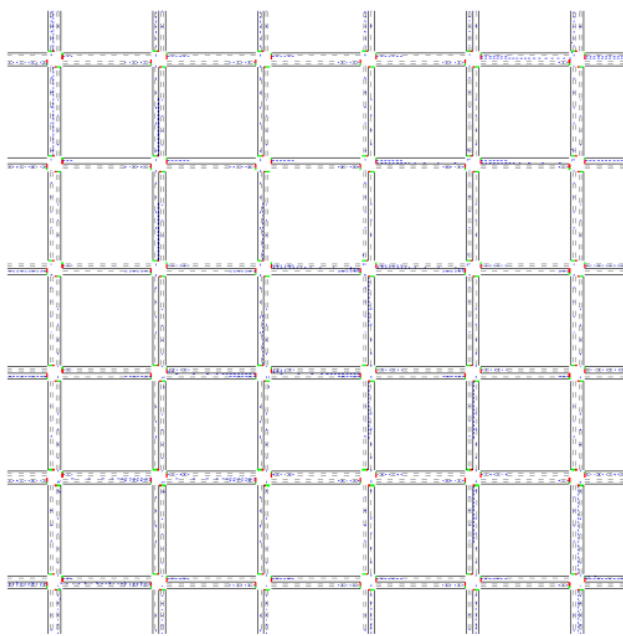
CONTENTS



- 01 ▶ 多智能体强化学习介绍
- 02 ▶ 应用一：智能信号灯控制
- 03 ▶ 应用二：多智能体路径规划

应用一：智能信号灯控制

- 智能实时调控路口交通信号灯，以减轻交通拥堵情况

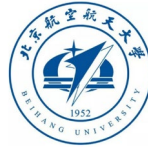


交通路口网络

多智能体协作解决



城市拥堵

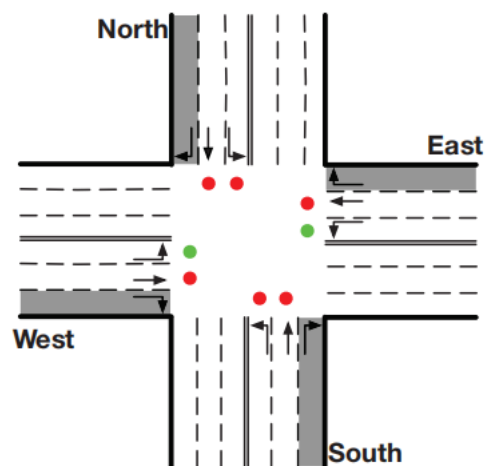


- 问题定义

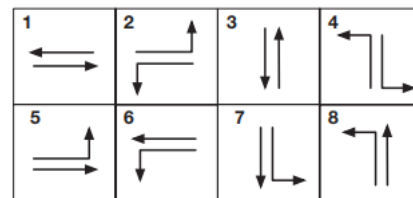
- 路口 Agent 如何根据观察到信息选择下一时刻的控制信号?

• 问题定义

- 路口 Agent 如何根据 **Observation** 观察到信息选择下一时刻的 **Action** 控制信号?
- 控制信号：允许哪些运动方向上的车辆可以通行。



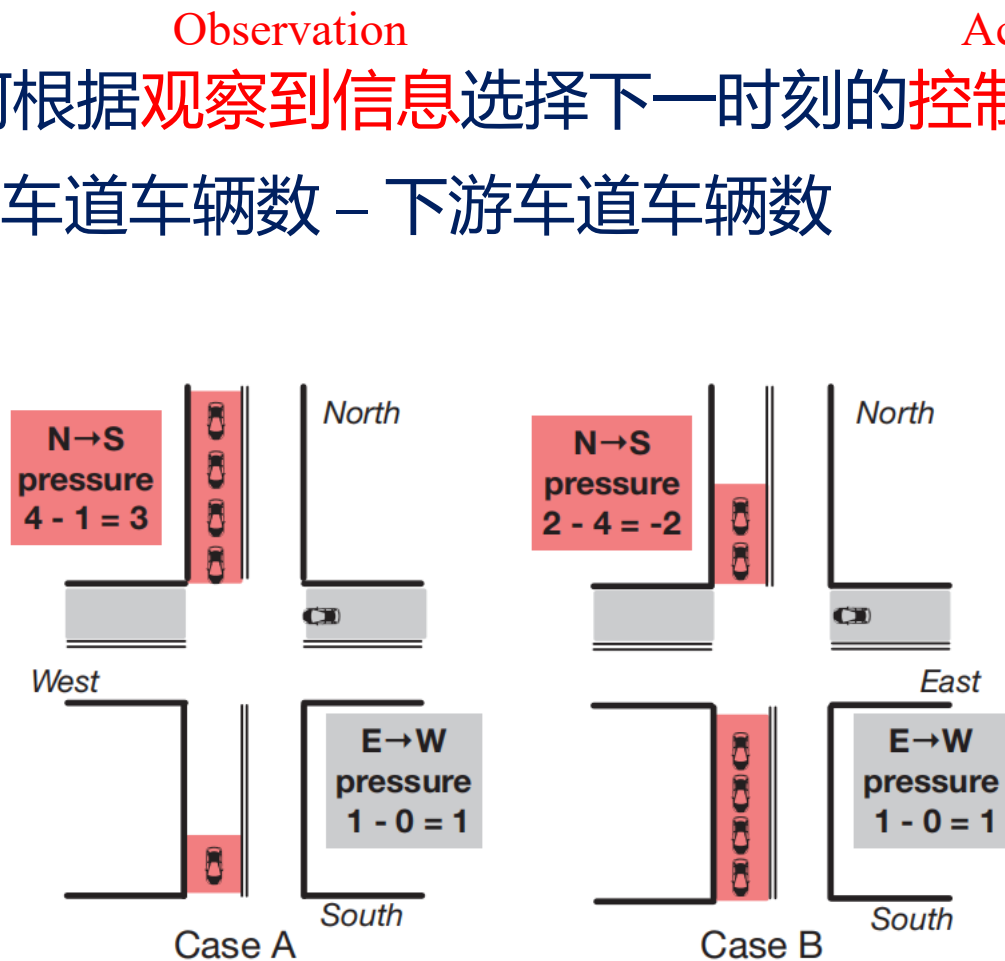
12 种运动方向



8 种控制信号

• 问题定义

- 路口 Agent 如何根据 **观察到信息** 选择下一时刻的 **控制信号**?
- 通行压力：上游车道车辆数 - 下游车道车辆数



- 问题定义

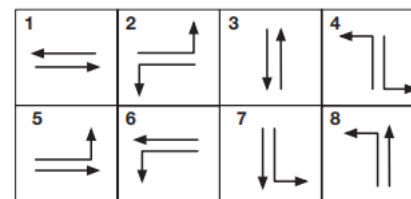
- 路口 Agent 如何根据**观察到信息**选择下一时刻的**控制信号**?

Observation

当前时刻的控制信号

12 个运动方向上的通行压力值

Action



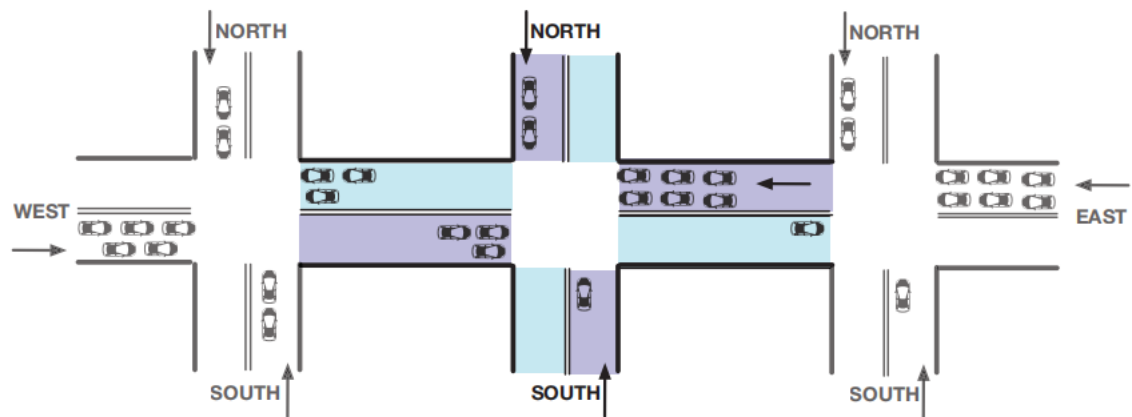
8 种控制信号中的一种

- 问题定义

- 优化的目标？如何定义奖励函数？

$$r_t^i = -P_t^i$$

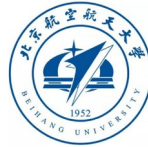
$$V^i = \max \sum_{t=0} \gamma^t r_t^i$$



$$\begin{aligned} \text{Pressure} &= |\# \text{queueing cars on entering lanes} - \# \text{queueing cars on exiting lanes}| \\ &= |3 + 2 + 6 + 1 - 3 - 0 - 1 - 0| \\ &= 8 \end{aligned}$$

路口压力 P_t^i

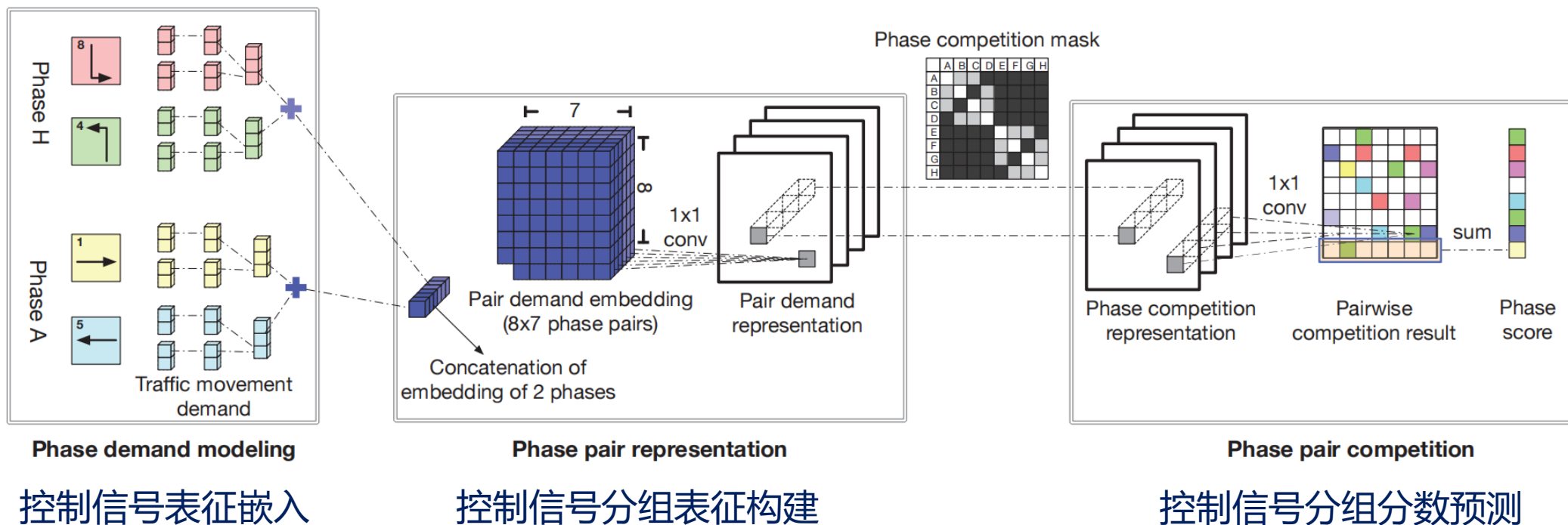
应用一：智能信号灯控制



• 策略网络

• 基于比较思想的控制信号打分网络

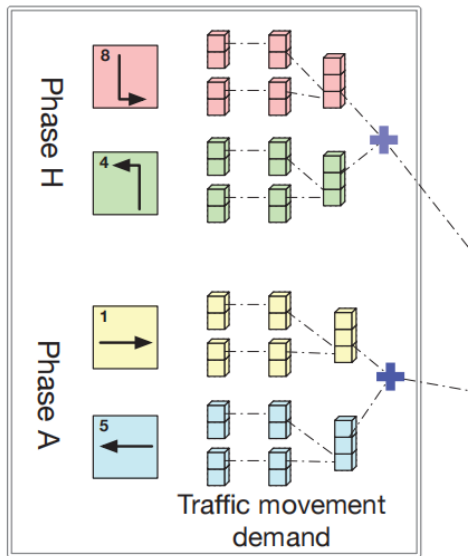
(控制信号 p , 控制信号 q) \rightarrow $score$



Learning Phase Competition for Traffic Signal Control. CIKM 2019.

Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control. AAAI 2020.

- 策略网络



Phase demand modeling

控制信号表征嵌入

Observation

当前时刻的控制信号

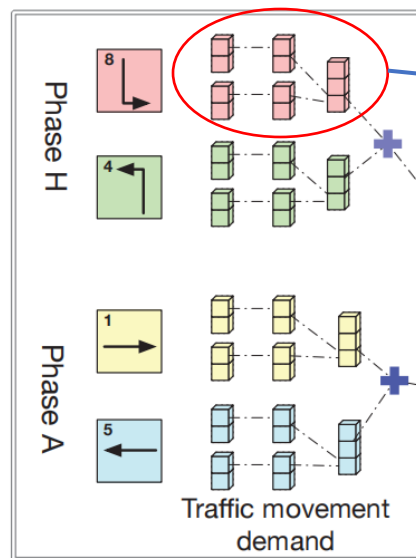
12 个运动方向上的通行压力值



嵌入候选控制信号的表征

- 策略网络

控制信号由两个运动方向组成



Phase demand modeling

控制信号表征嵌入

使用 MLP 嵌入

通行压力

当前时刻控制信号

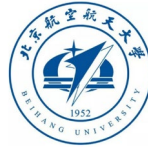
$$\mathbf{h}_i^v = \text{ReLU}(\mathbf{W}^v \mathbf{f}_i^v + \mathbf{b}^v), \quad \mathbf{h}_i^s = \text{ReLU}(\mathbf{W}^s \mathbf{f}_i^s + \mathbf{b}^s).$$

$$\mathbf{d}_i = \text{ReLU}(\mathbf{W}^h [\mathbf{h}_i^v, \mathbf{h}_i^s] + \mathbf{b}^h).$$

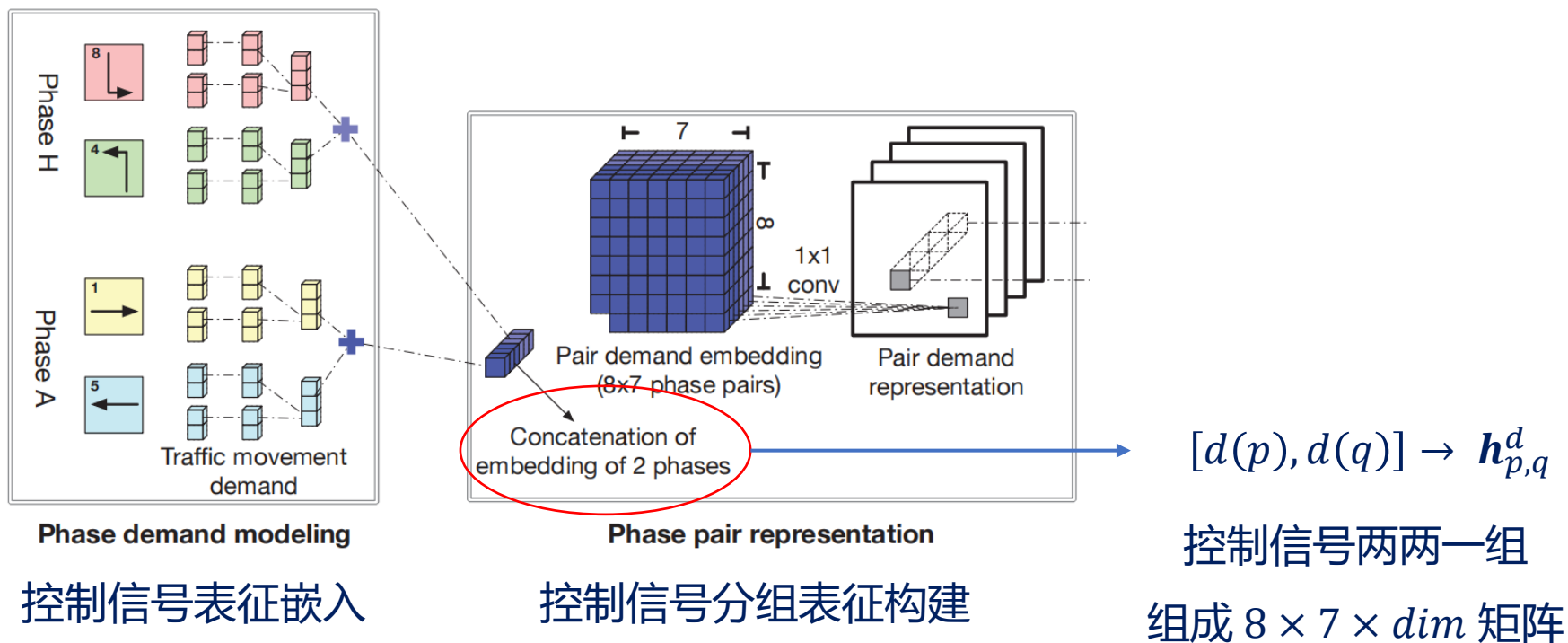
$$\mathbf{d}(\mathbf{p}) = \mathbf{d}_i + \mathbf{d}_j.$$

控制信号嵌入表征

应用一：智能信号灯控制



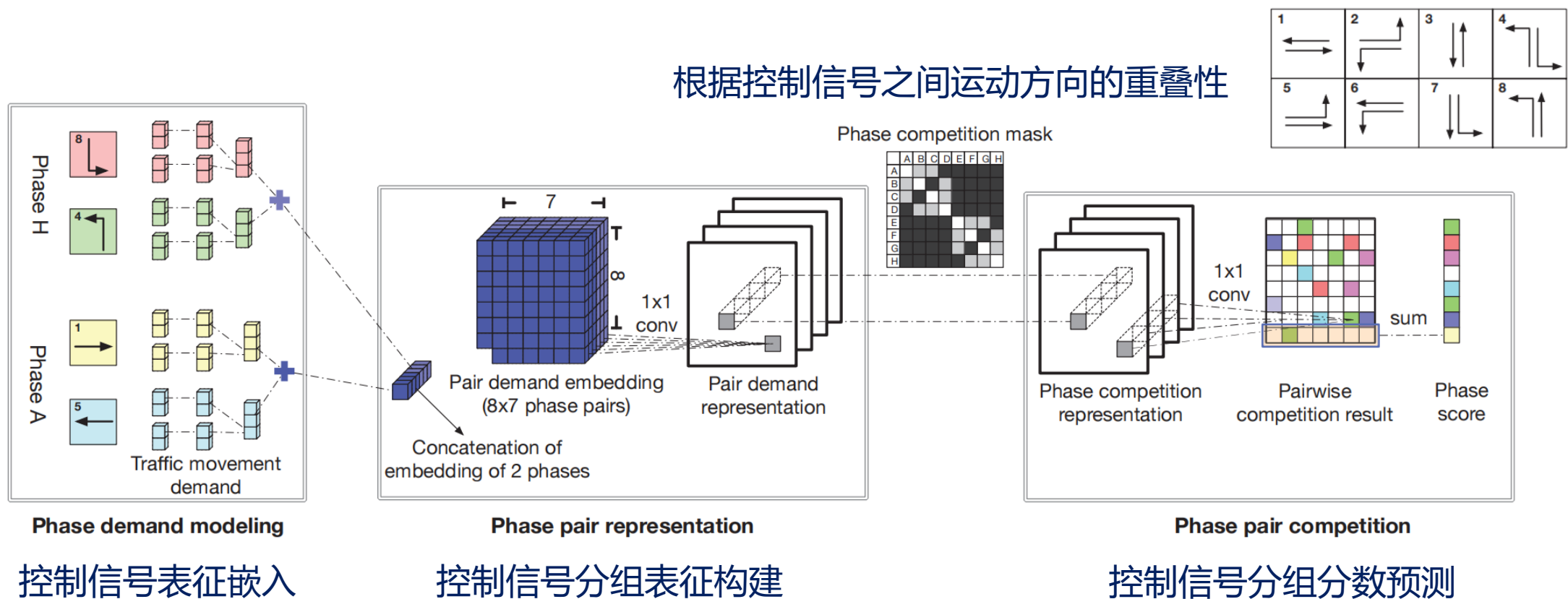
• 策略网络



Learning Phase Competition for Traffic Signal Control. CIKM 2019.

Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control. AAAI 2020.

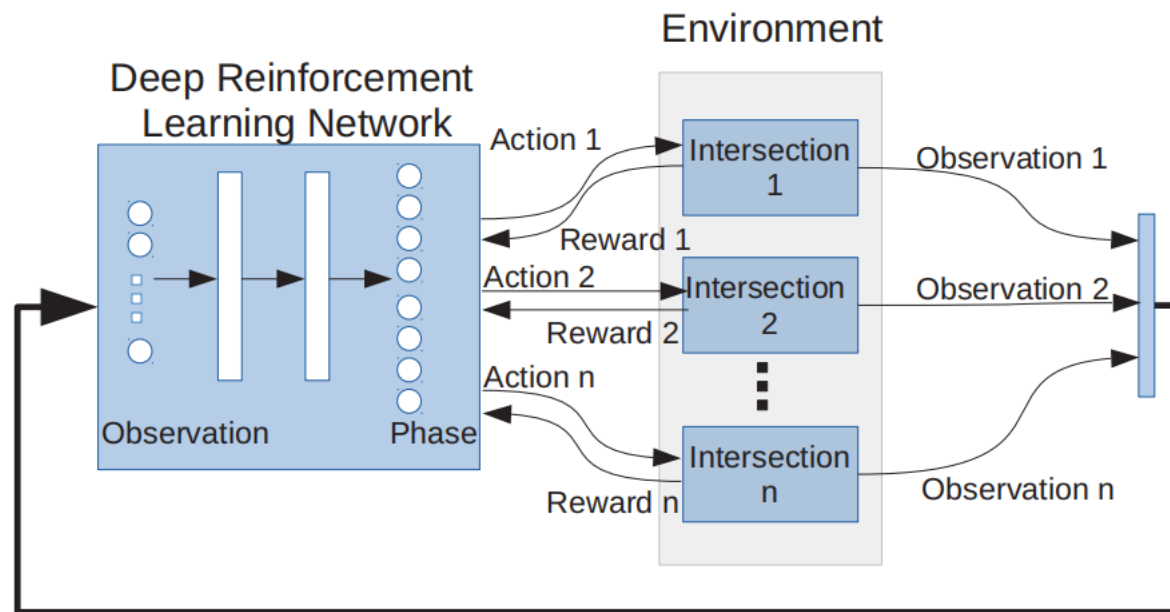
策略网络



Learning Phase Competition for Traffic Signal Control. CIKM 2019.

Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control. AAAI 2020.

- 整体框架



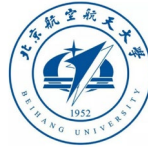
目录

CONTENTS



- 01** ▶ 多智能体强化学习介绍
- 02** ▶ 应用一：智能信号灯控制
- 03** ▶ 应用二：多智能体路径规划

应用二：多智能体路径规划



- 为多个智能体同时规划路径，以达成某种合作目标



无人机/车

工厂运输货物

智能体



无人驾驶车/服从导航的车

多车辆的行驶路线规划

- 多车辆自主建图问题

- 给定一组车辆，为其规划一组通行代价最小的路径，使得城市街区中每条道路都被访问过一定次数，从而更新地图街景信息。



地图采集车

采集更新



城市街区

- 多车辆自主建图问题

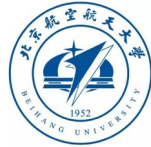
- 给定路网有向图 $G(V, E)$, 为 L 个 agent 各生成一条路径 $p^{(i)}$, 最终 V 中每个顶点被所有 agent 共同覆盖 M_v 次。

Observation

Agent 所观察到的局部信息 (位置)
从其他 Agent 收到的通信信息

Action

下一时间步前往的目标路段



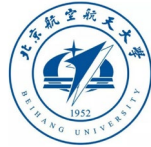
- 多车辆自主建图问题

- 给定路网有向图 $G(V, E)$, 为 L 个 agent 各生成一条路径 $p^{(i)}$, 最终 V 中每个顶点被所有 agent 共同覆盖 M_v 次。

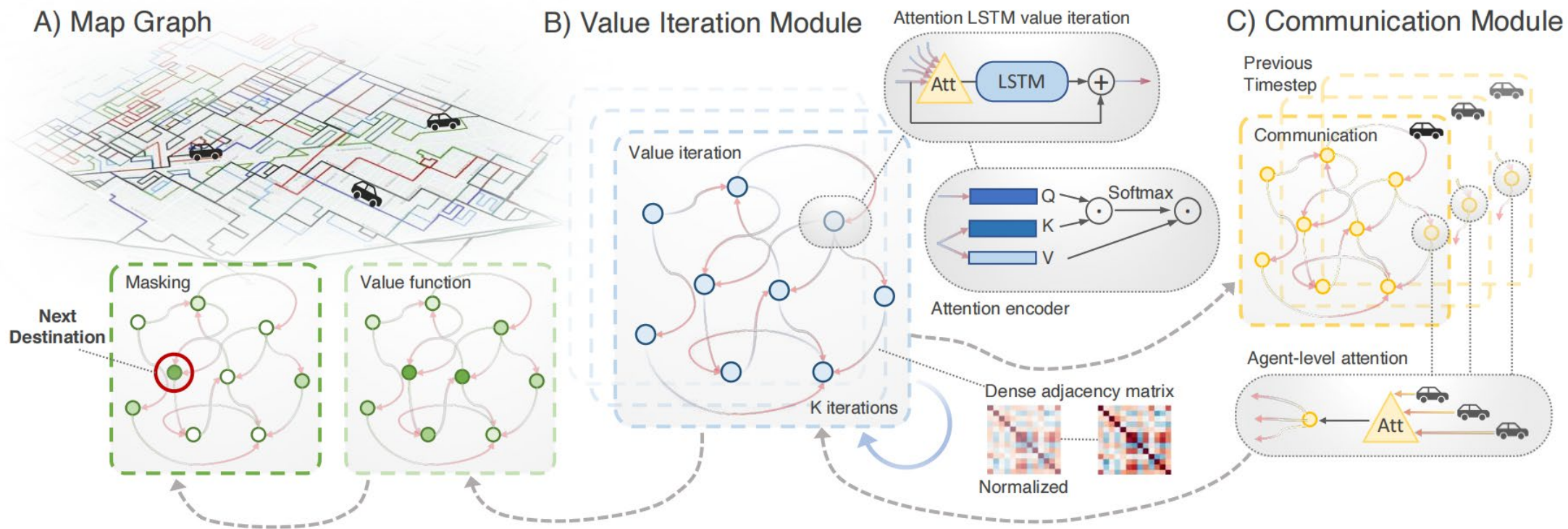
Reward

$$R = - \sum_i Cost(p^{(i)})$$

应用二：多智能体路径规划



• 策略网络



值迭代网络预测下一跳概率

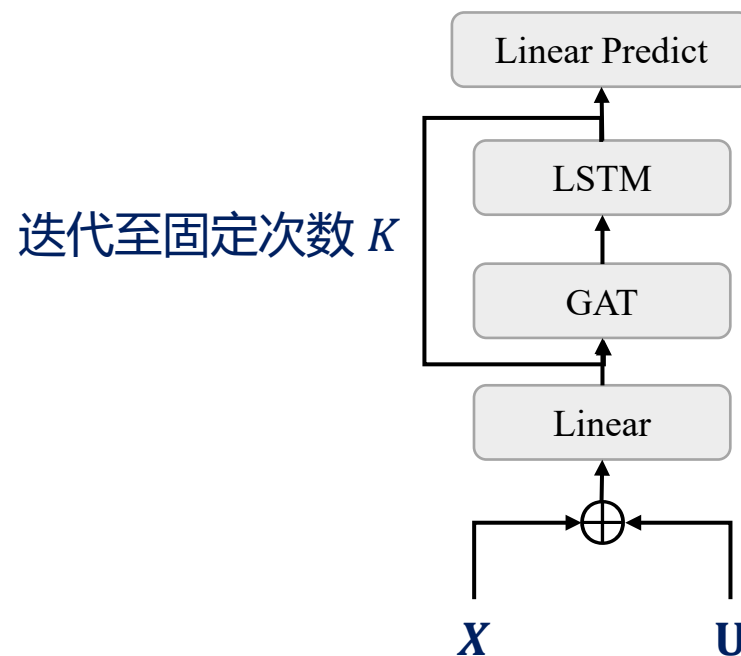
通信信息交互完成协作

- 策略网络

- 值迭代网络预测下一跳

路段初始特征 $\mathbf{X} = \{x_1, x_2, \dots, x_n\} \in R^{n \times d}$

通信初始信息 $\mathbf{U} = \{u_1, u_2, \dots, u_n\} \in R^{n \times d}$



$$\pi(a_t^i; s_t^i) = \text{softmax}(\mathbf{X}^{(K)} \mathbf{W}_{dec} + \mathbf{b}_{dec})$$

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \text{LSTM}(\text{GAT}(\mathbf{X}^{(k)}, A), \mathbf{H}^{(k)})$$

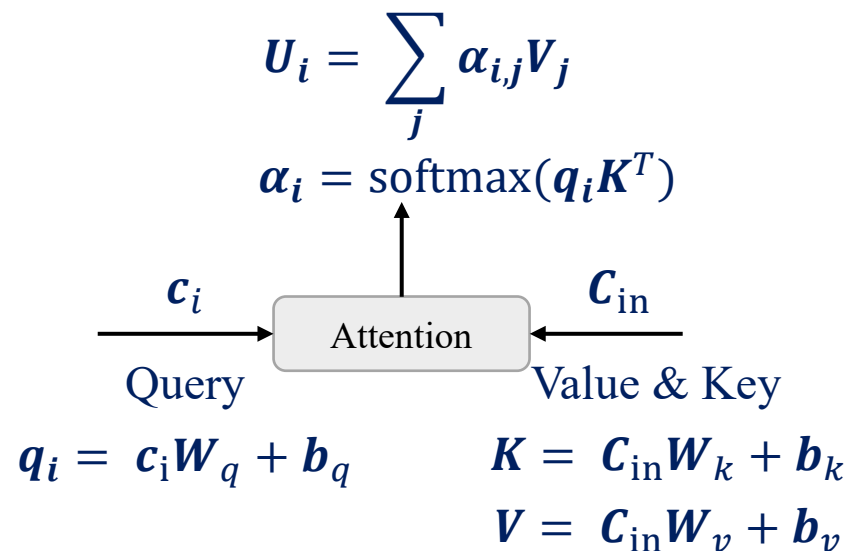
$$\mathbf{X}^{(0)} = (\mathbf{X} || \mathbf{U}) \mathbf{W}_{enc} + \mathbf{b}_{enc}$$

- 策略网络

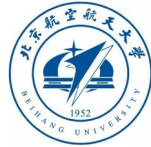
- 通信模块

Agent_{*i*} 输出通信向量 $\mathbf{c}_i = \mathbf{X}_i^{(K)} \mathbf{W}_{com} + \mathbf{b}_{com} \in R^{n \times d}$

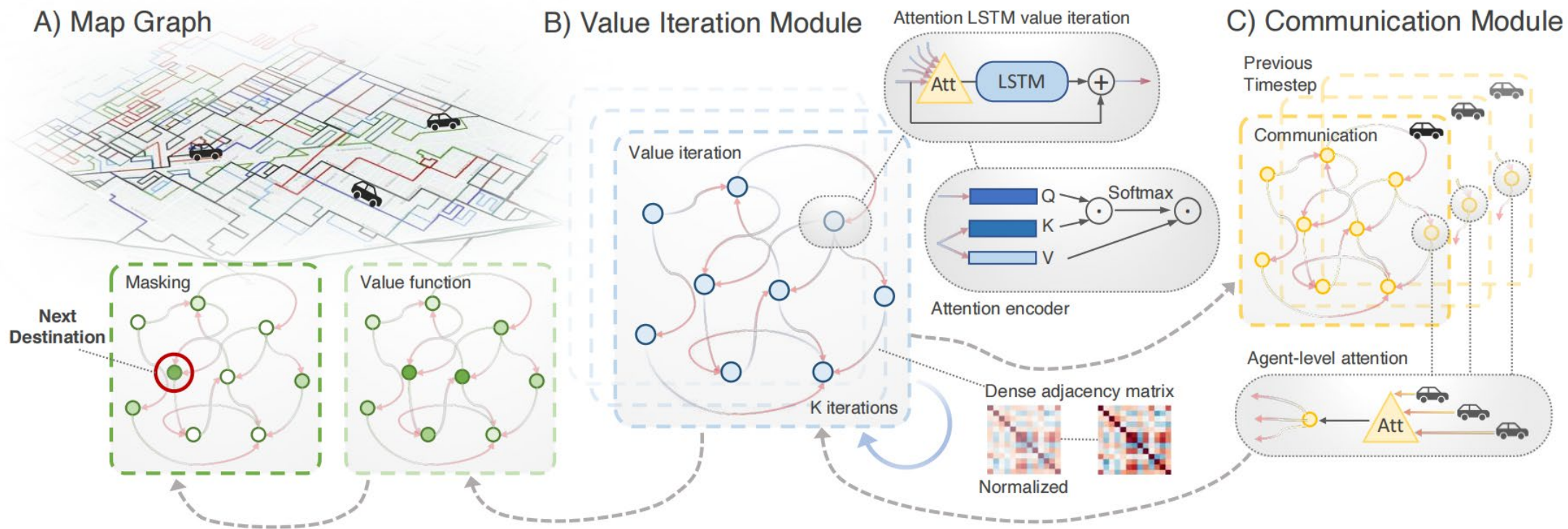
Agent_{*i*} 接收通信向量集合 $\mathbf{C}_{in} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L\}$



应用二：多智能体路径规划



• 策略网络



值迭代网络预测下一跳概率

通信信息交互完成协作



感谢聆听!