多智能体强化学习在智慧城市中的应用

演讲人：蒋炆峻

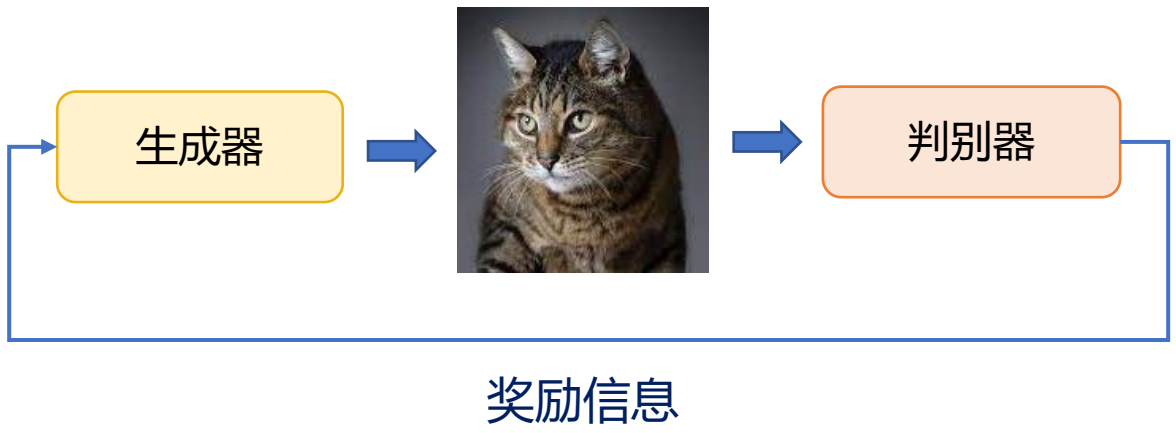
报告目录

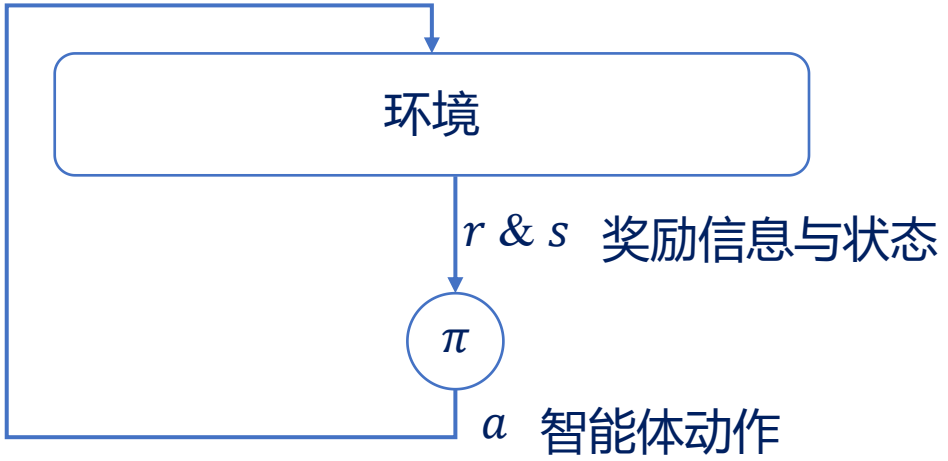
目录

CONTENTS

多智能体强化学习介绍

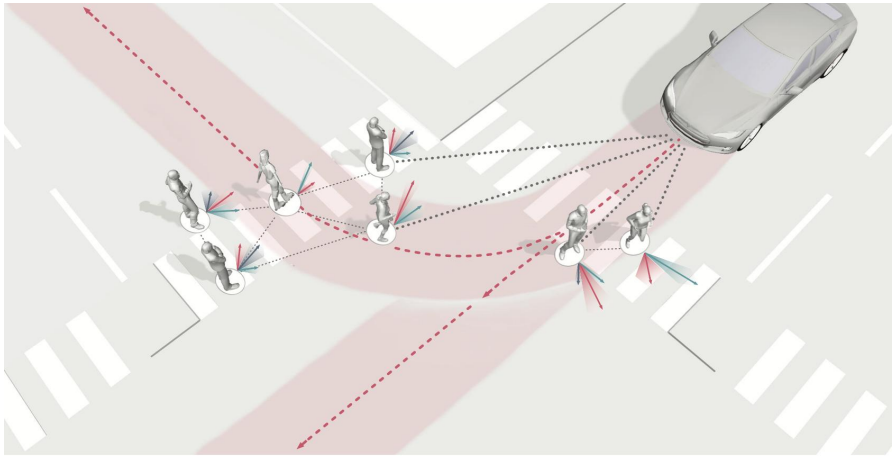
- 多智能体强化学习：环境中存在多个实体，同时与环境进行交互。
  - 多智能体 vs 单智能体



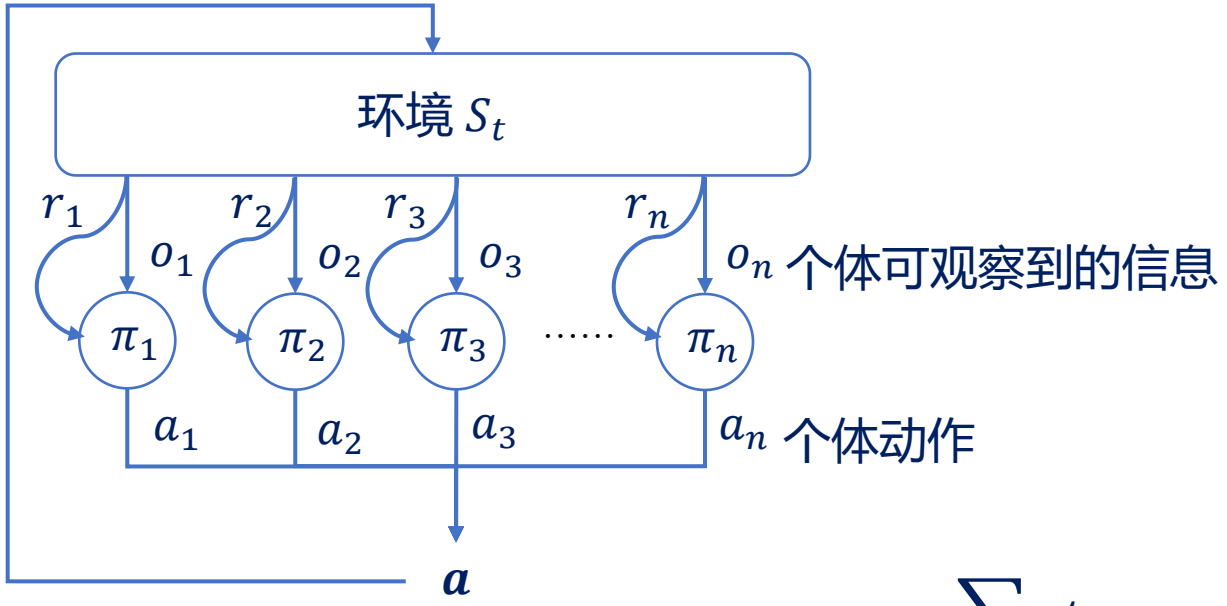


多智能体强化学习介绍

- 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



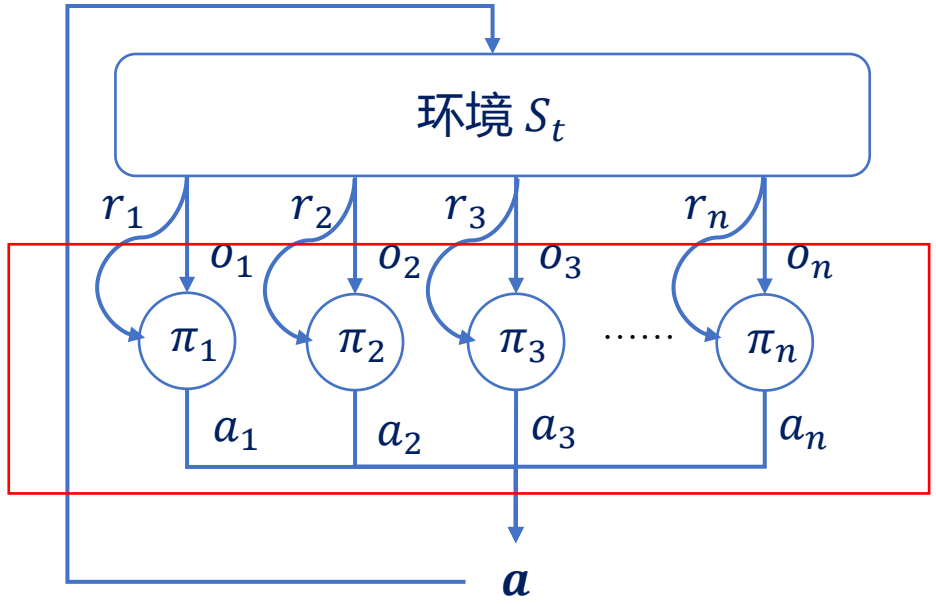
$$V_i = \mathbb{E}[\sum_{t=0} \gamma^t r_{i,t}]$$

多智能体强化学习介绍

- 新挑战

多智能体强化学习介绍

- 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。



多个个体的联合动作

解决方案 1：中心化（整体化）

缺陷：不易部署与扩展

多智能体强化学习介绍

- 新挑战

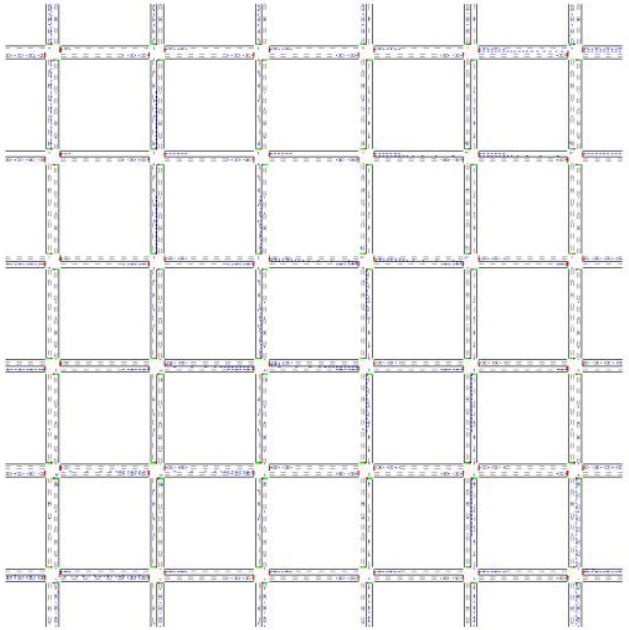
报告目录

目录

CONTENTS

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

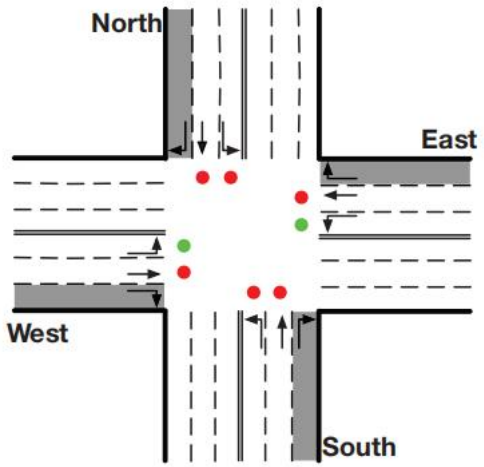
应用一：智能信号灯控制

- 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?

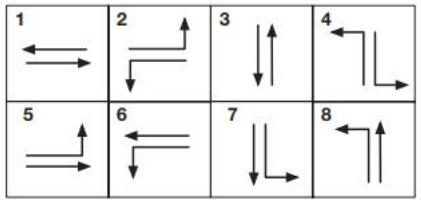
Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control. AAAI 2020.

应用一：智能信号灯控制

- 问题定义
  - 路口 Agent 如何根据观察到信息（Observation）选择下一时刻的控制信号（Action）？
  - 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

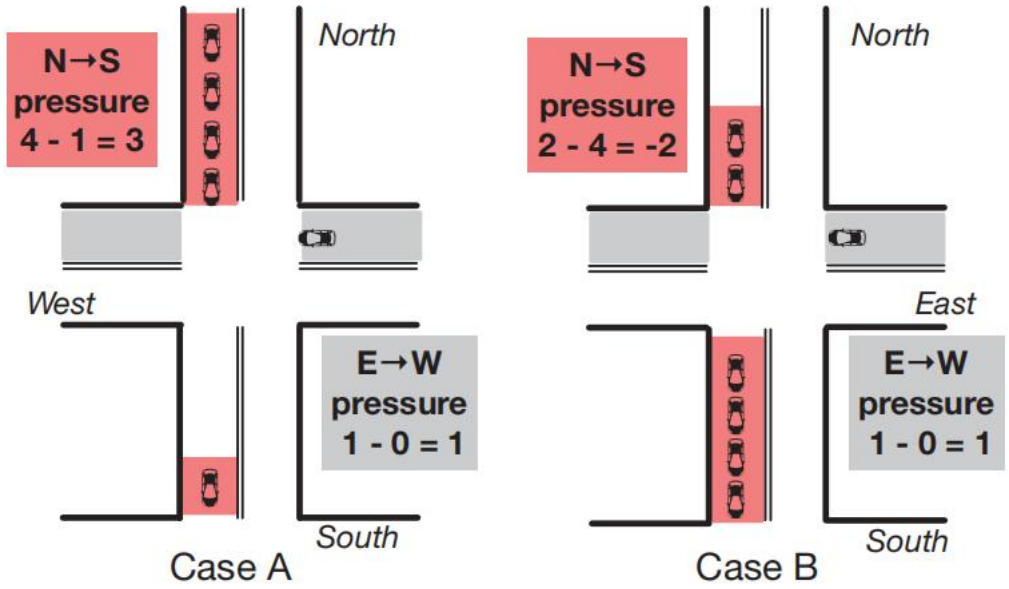


8 种控制信号

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

应用一：智能信号灯控制

- 问题定义
  - 路口 Agent 如何根据观察到信息（Observation）选择下一时刻的控制信号（Action）？
  - 通行压力：上游车道车辆数 – 下游车道车辆数



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

应用一：智能信号灯控制

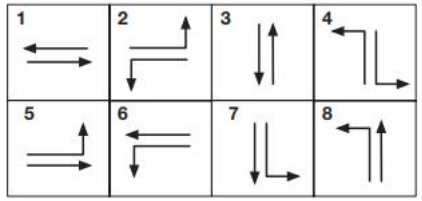
- 问题定义
  - 路口 Agent 如何根据观察到信息选择下一时刻的控制信号？

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

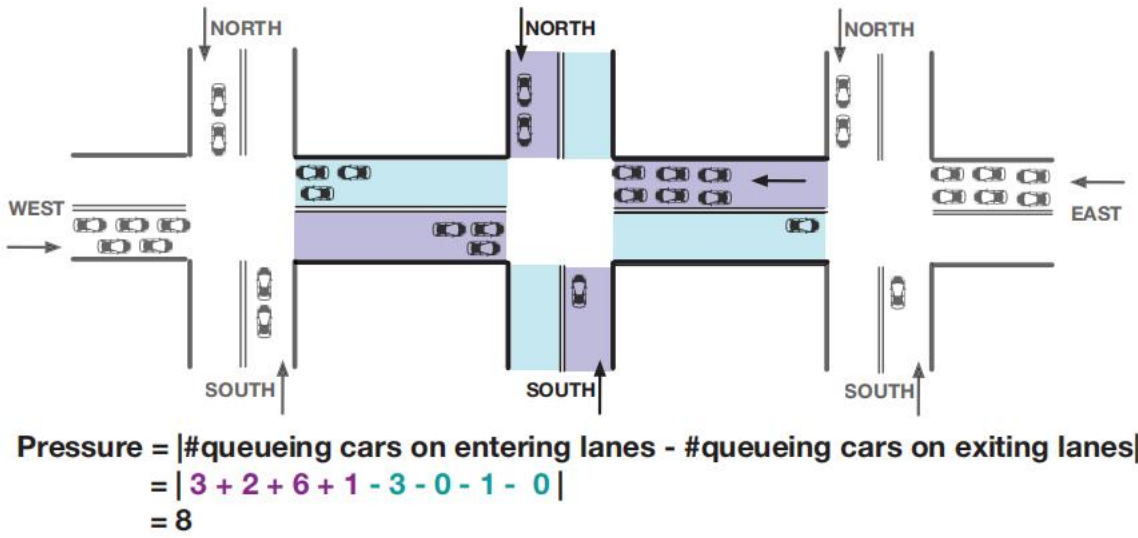
Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control. AAAI 2020.

应用一：智能信号灯控制

- 问题定义
  - 优化的目标？如何定义奖励函数？

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

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



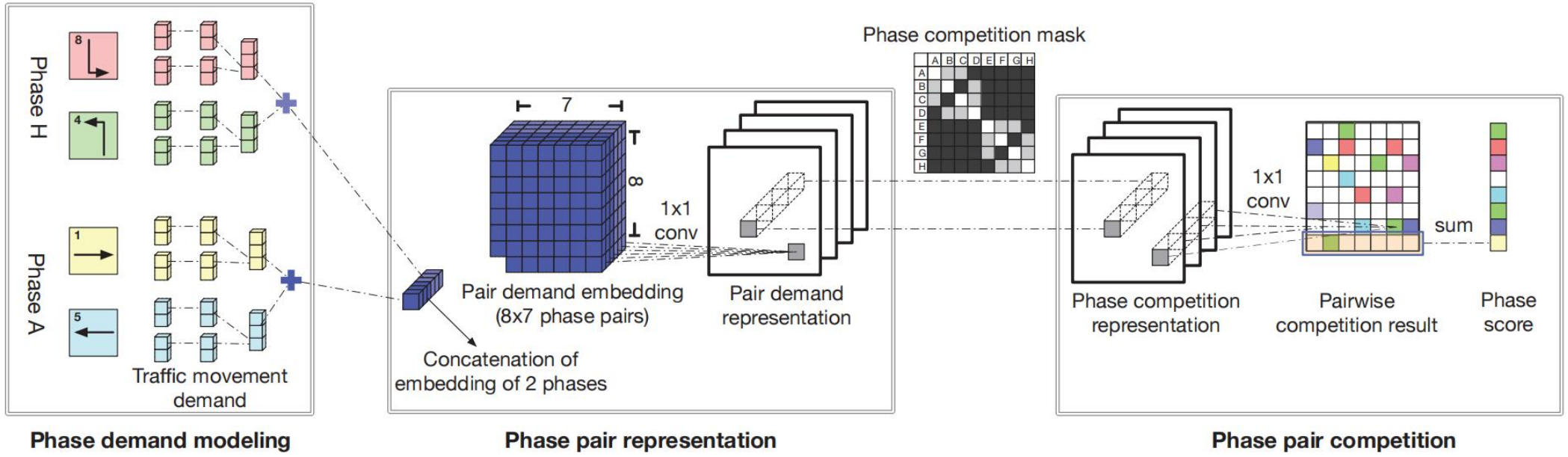
路口压力 P_t^i

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

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

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



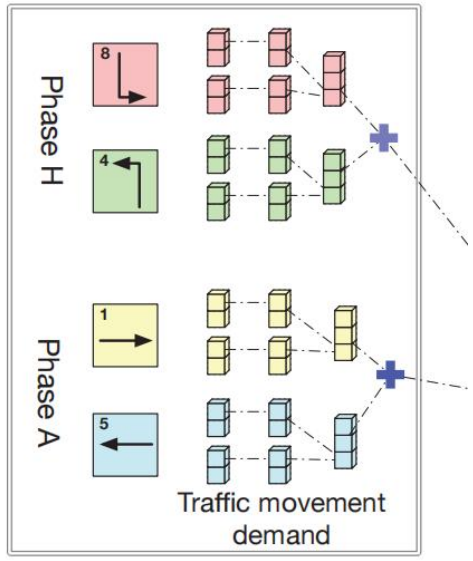
控制信号表征嵌入　　控制信号分组表征构建　　控制信号分组分数预测

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.

应用一：智能信号灯控制

- 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

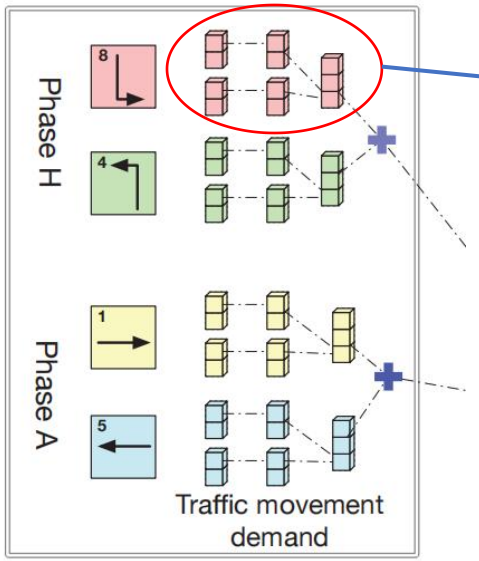
Learning Phase Competition for Traffic Signal Control. CIKM 2019.

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

应用一：智能信号灯控制

- 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

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

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

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

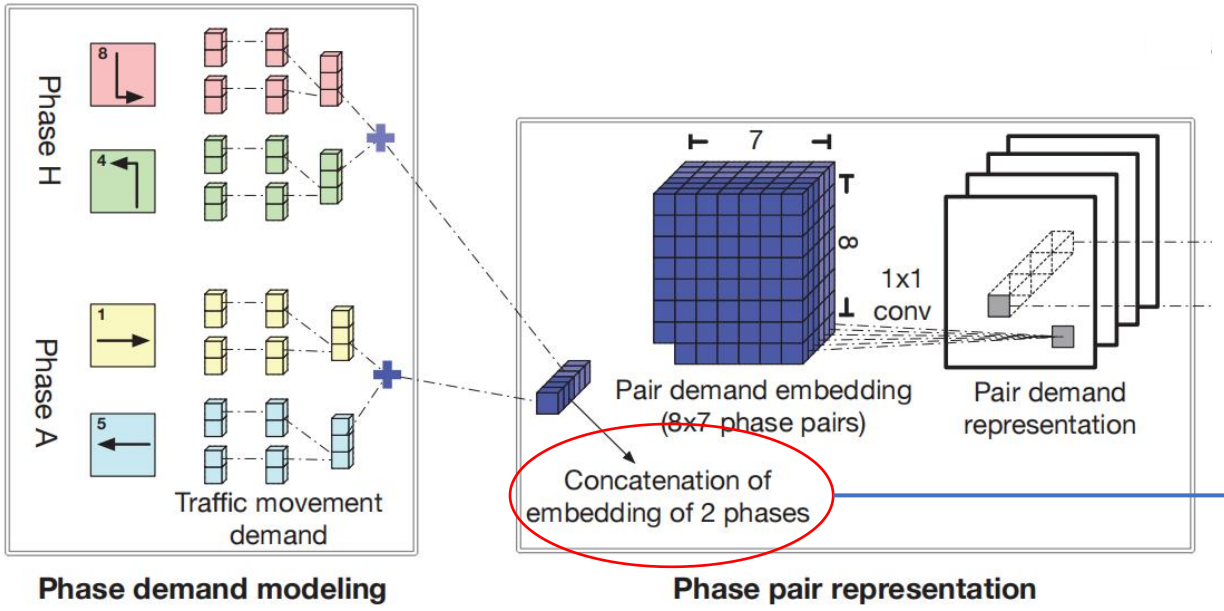
控制信号嵌入表征

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.

应用一：智能信号灯控制

- 策略网络



$[d(p),\ d(q)] \rightarrow \boldsymbol{h}_{p,q}^{d}$

控制信号表征嵌入

控制信号分组表征构建

控制信号两两一组

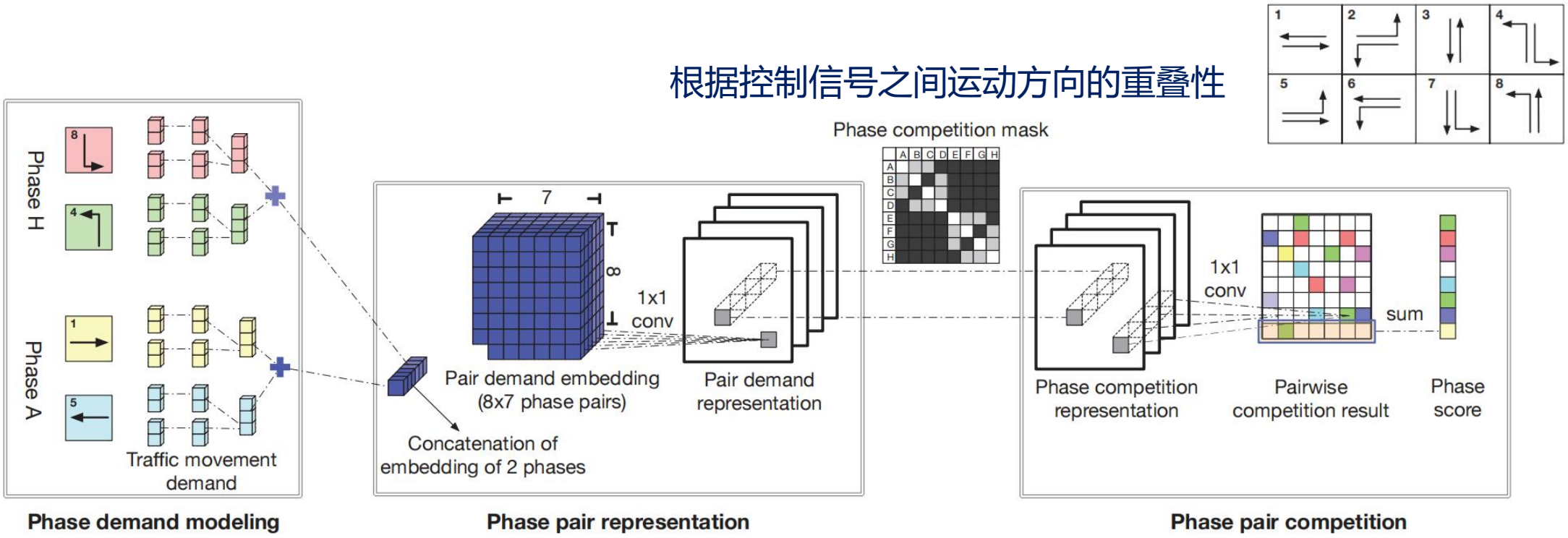
组成 $8 \times 7 \times dim$ 矩阵

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.

应用一：智能信号灯控制

- 策略网络



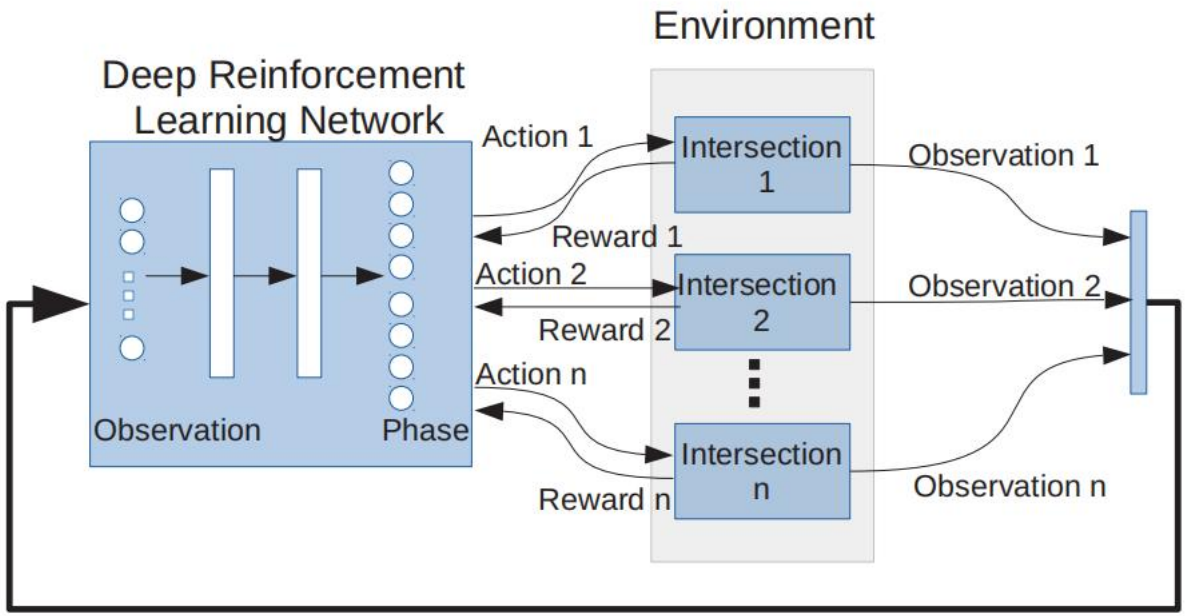
控制信号表征嵌入　　控制信号分组表征构建　　控制信号分组分数预测

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.

应用一：智能信号灯控制

- 整体框架



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

报告目录

目录

CONTENTS

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

应用二：多智能体路径规划

- 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

Multi-Agent Routing Value Iteration Network. ICML 2020.

应用二：多智能体路径规划

- 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	Action
Agent 所观察到的局部信息（位置） 从其他 Agent 收到的通信信息	下一时间步前往的目标路段

Multi-Agent Routing Value Iteration Network. ICML 2020.

应用二：多智能体路径规划

- 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。

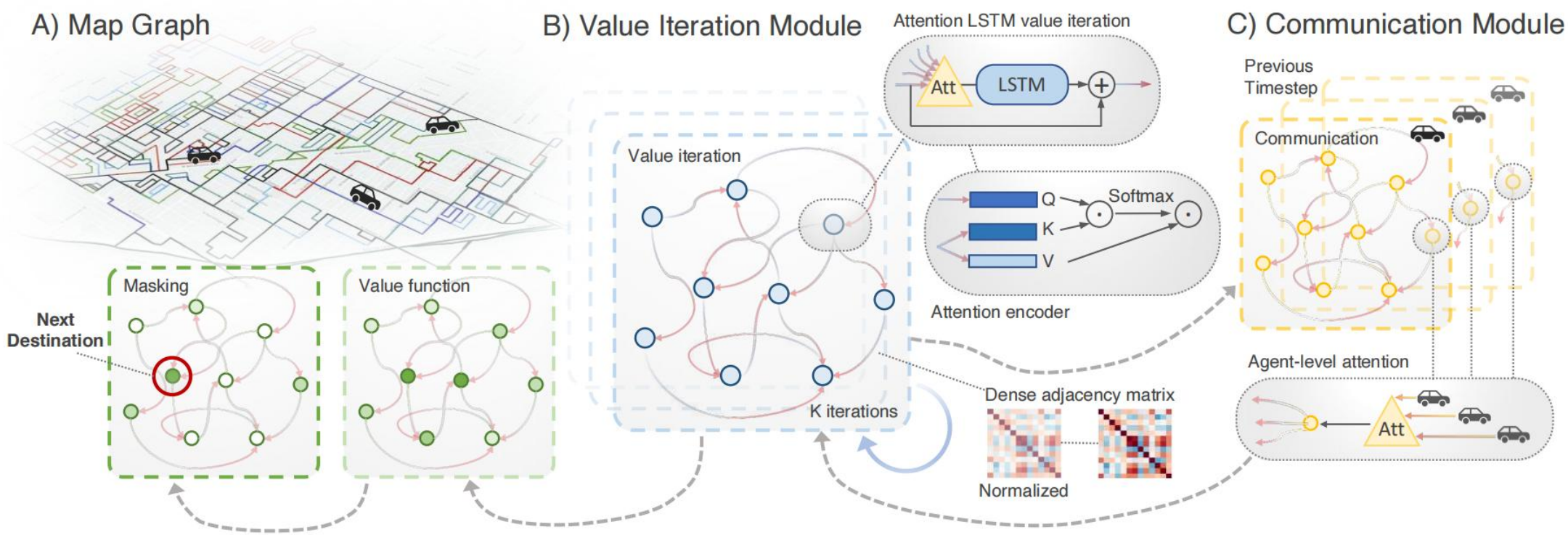
Reward

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

Multi-Agent Routing Value Iteration Network. ICML 2020.

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

Multi-Agent Routing Value Iteration Network. ICML 2020.

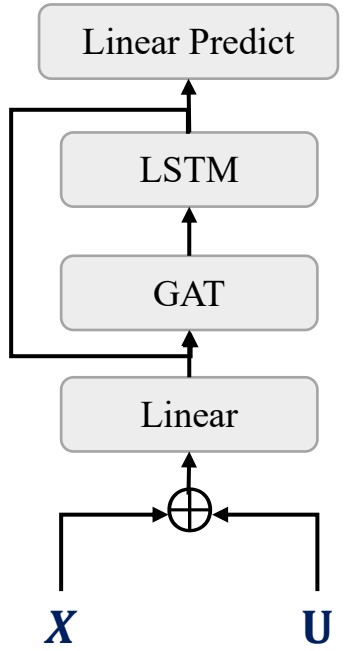
应用二：多智能体路径规划

- 策略网络
  - 值迭代网络预测下一跳

路段初始特征 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in R^{n \times d}$

通信初始信息 $\mathbf{U} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\} \in R^{n \times d}$

迭代至固定次数 K



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

$$\boldsymbol{X}^{(k+1)} = \boldsymbol{X}^{(k)} + \mathrm{LSTM}\left(\mathrm{GAT}\left(\boldsymbol{X}^{(k)}, A\right), \boldsymbol{H}^{(k)}\right)$$

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

Multi-Agent Routing Value Iteration Network. ICML 2020.

应用二：多智能体路径规划

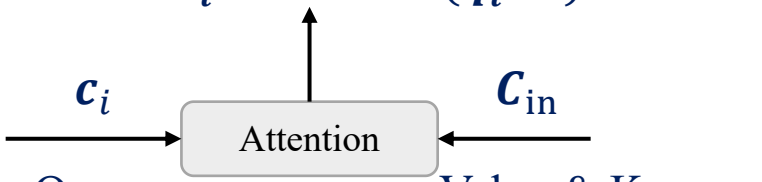
- 策略网络
  - 通信模块

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

Agent_i 接收通信向量集合 $\boldsymbol{C}_{\text{in}} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_L\}$

$$\boldsymbol{U}_i = \sum_j \boldsymbol{\alpha}_{i,j}\boldsymbol{V}_j$$

$$\boldsymbol{\alpha}_i = \text{softmax}(\boldsymbol{q}_i\boldsymbol{K}^T)$$



$$\boldsymbol{q}_i = \boldsymbol{c}_i\boldsymbol{W}_q + \boldsymbol{b}_q$$

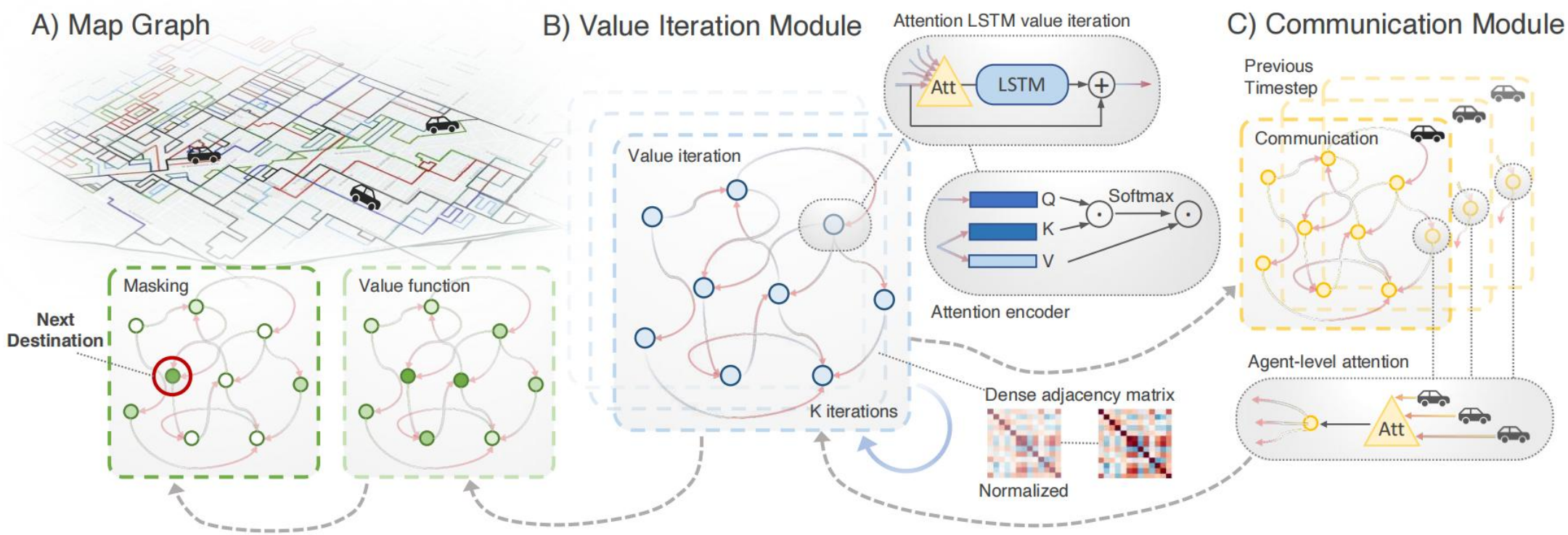
$$\boldsymbol{K} = \boldsymbol{C}_{\text{in}}\boldsymbol{W}_k + \boldsymbol{b}_k$$

$$\boldsymbol{V} = \boldsymbol{C}_{\text{in}}\boldsymbol{W}_v + \boldsymbol{b}_v$$

Multi-Agent Routing Value Iteration Network. ICML 2020.

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

Multi-Agent Routing Value Iteration Network. ICML 2020.

感谢聆听！