

AI赋能的RIS信道信息获取与无线资源优化

东南大学信息科学与工程学院 李骏

2025.08

- ◆ **1. AI赋能RIS的研究背景**
- ◆ **2. 基于生成式AI的高效RIS信道获取**
- ◆ **3. 基于强化学习的RIS信道动态感知**
- ◆ **4. 基于预训练决策模型的RIS资源分配**

智能通信产业发展

二十届三中全会将**新一代信息技术与人工智能**作为我国战略发展方向



AI-RAN联盟 (2024年2月)

由英伟达、爱立信、微软、诺基亚等十余家公司牵头成立了**AI-RAN联盟**
旨在将人工智能深度融入无线通信，实现**内生智能的通信网络架构**

智能通信产业正处于稳步增长阶段

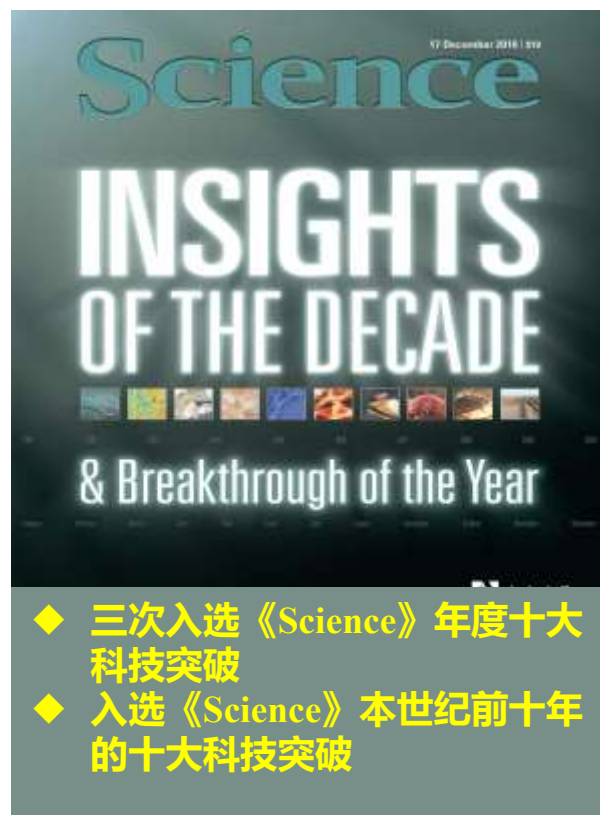
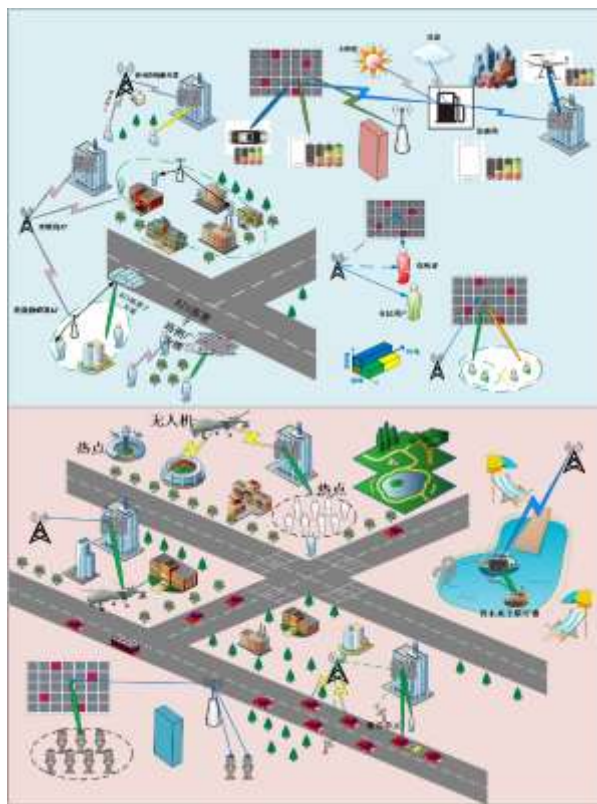


移动通信系统正向**内生智能架构**转变



RIS技术的研究重要性

智能超表面（RIS），能够通过数字化的电磁波改善无线传输环境，增强覆盖质量与范围，具备**易部署、低功耗**等特点，被认为是**6G变革性技术**



美国国防部六大颠覆性 基础技术

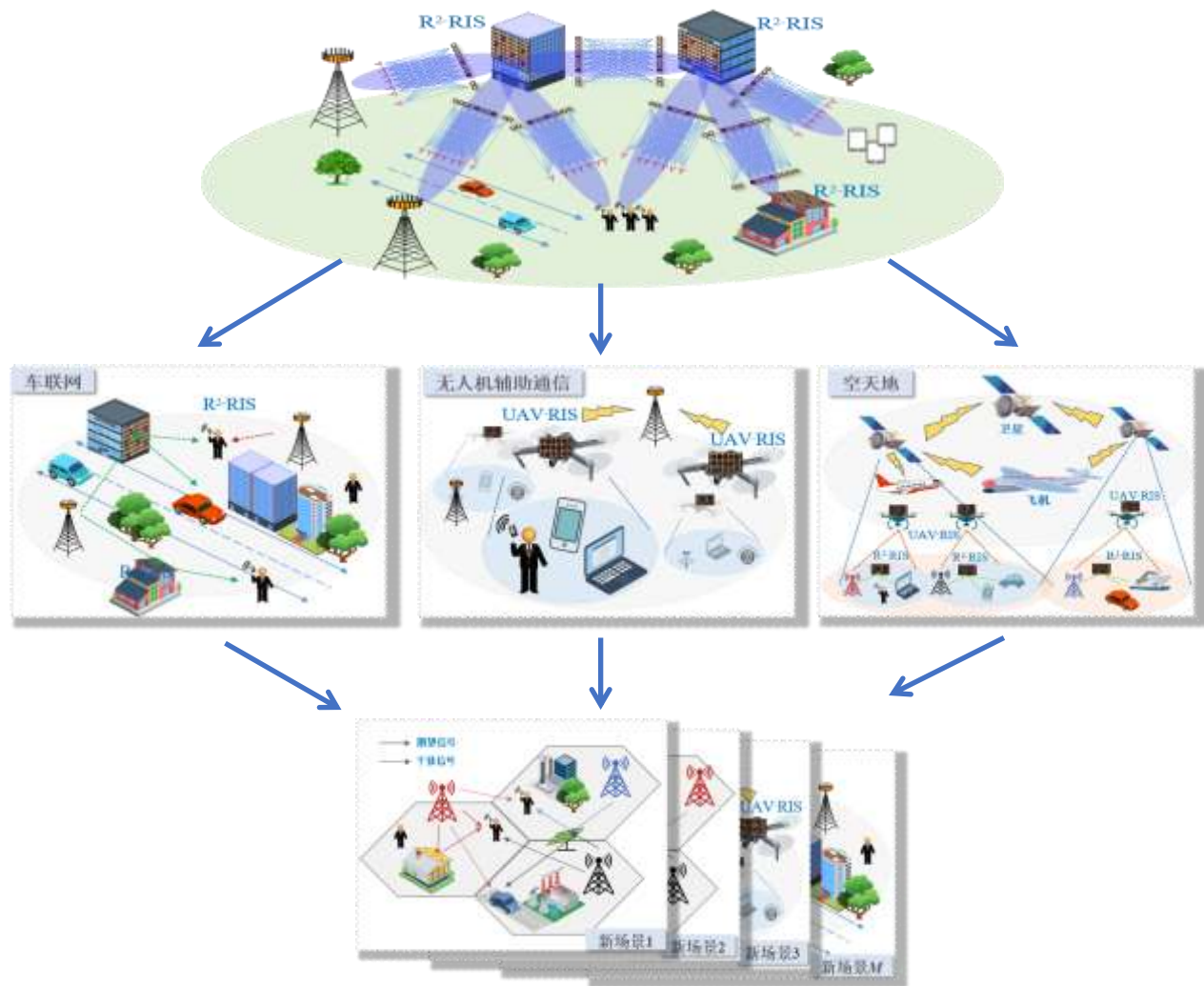
Basic Research Areas

- Six Disruptive Basic Research Areas
 - Engineered Materials (metamaterials and plasmonics)
 - Quantum Information and Control
 - Cognitive Neuroscience
 - Nanoscience and Nanoengineering
 - Synthetic Biology
 - Computational Modeling of Human and Social Behavior

超材料位列首位

为调控电磁波开拓了新的自由度

RIS技术面临的挑战



空间**超**高维：
信道获取开销大

环境**强**随机：
资源适配实时差

场景**快**迁移：
决策泛化收敛难

人工智能技术可高效解决上述问题，AI赋能的RIS覆盖增强技术研究势在必行

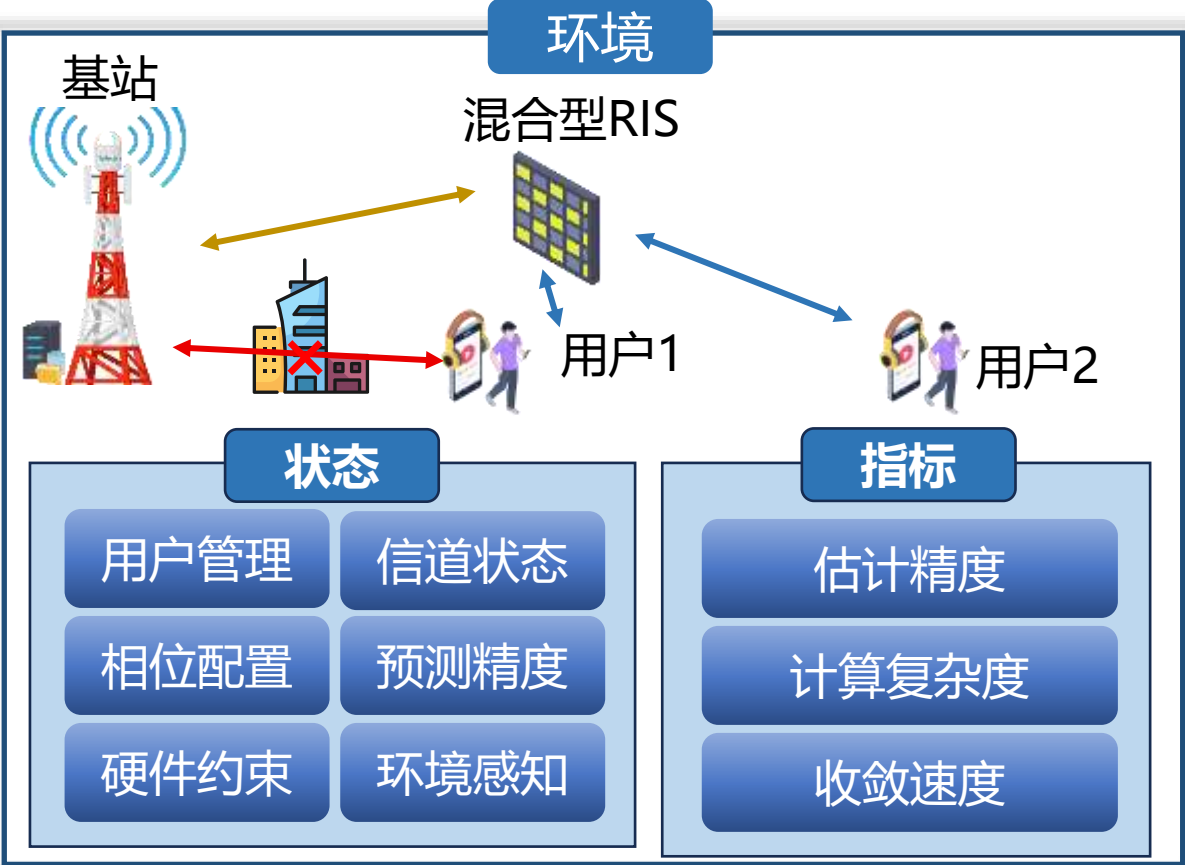
AI赋能RIS信道估计研究现状

传统方法挑战:

- 级联信道估计困难，不适合实时动态环境
- 导频开销巨大，影响频谱效率和系统容量

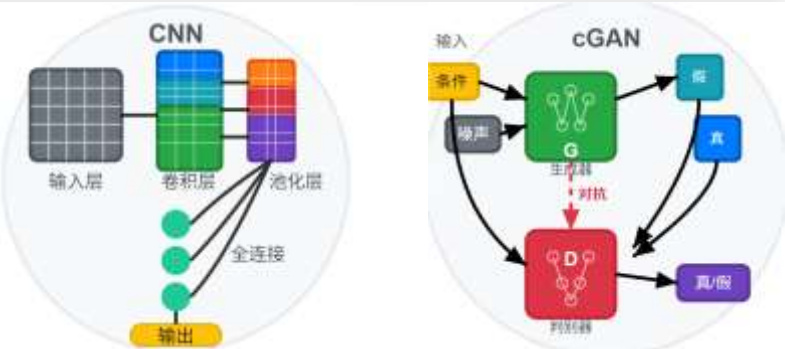
现有研究已用AI技术对RIS信道估计进行了探索

- 基于深度学习优化信道估计方法

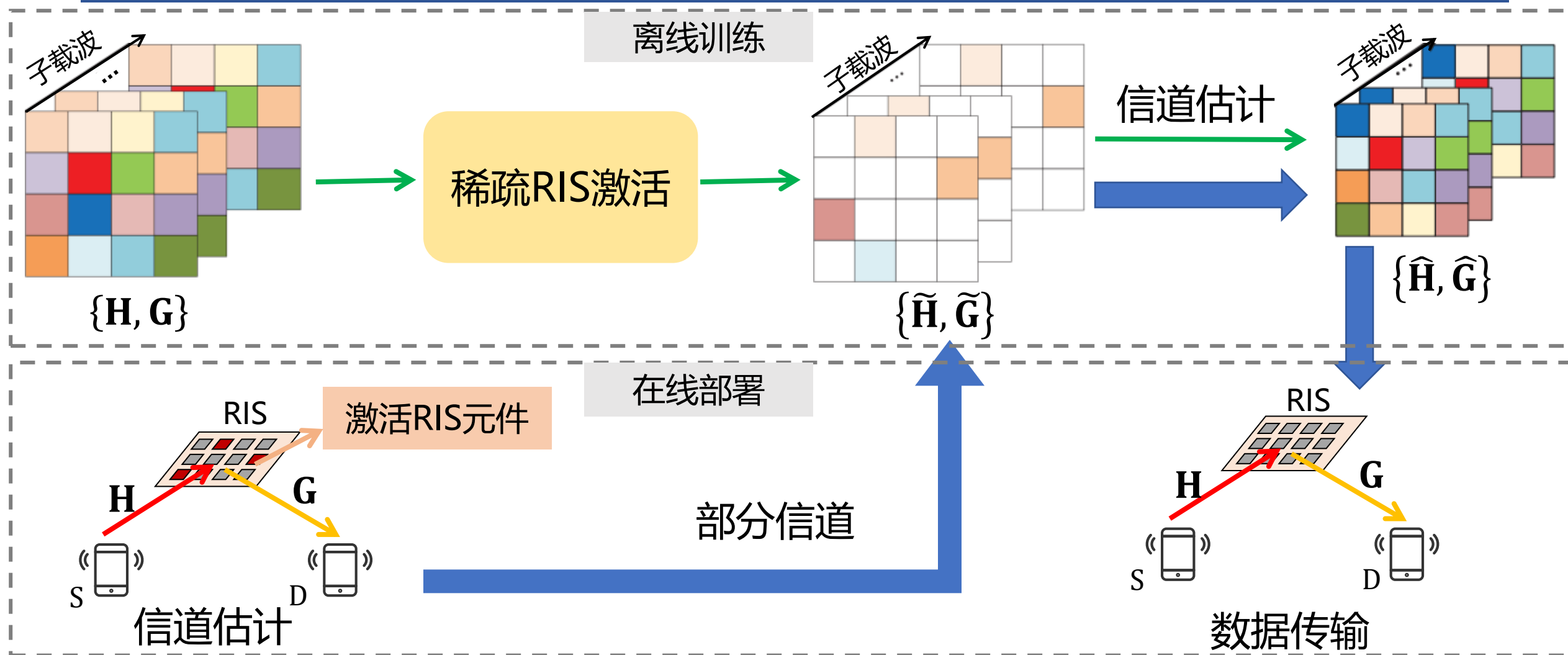


AI方法

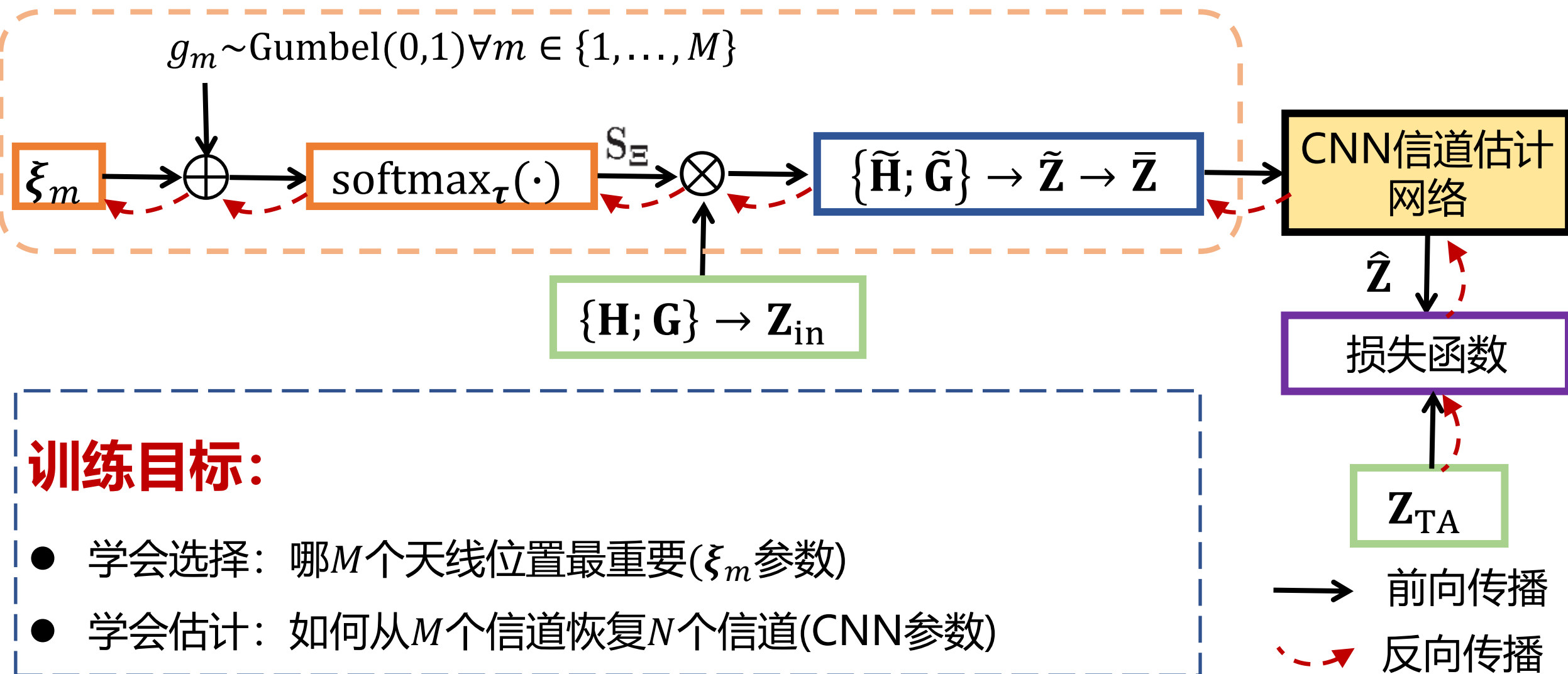
- 基于卷积神经网络(CNN)构建信道估计网络
- 基于条件生成对抗网络(cGAN)构建信道估计框架



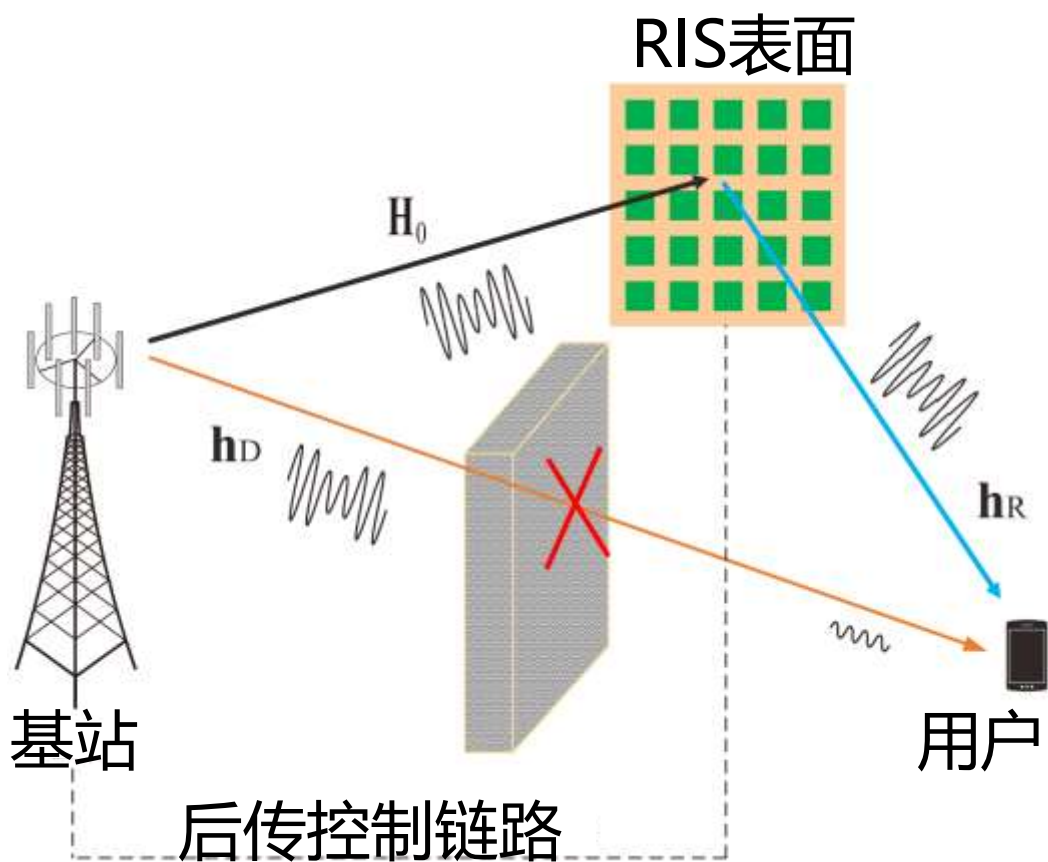
基于卷积神经网络的RIS信道生成



基于卷积神经网络的RIS信道生成



基于条件生成对抗网络的RIS信道生成



CNN信道估计的**核心缺陷**

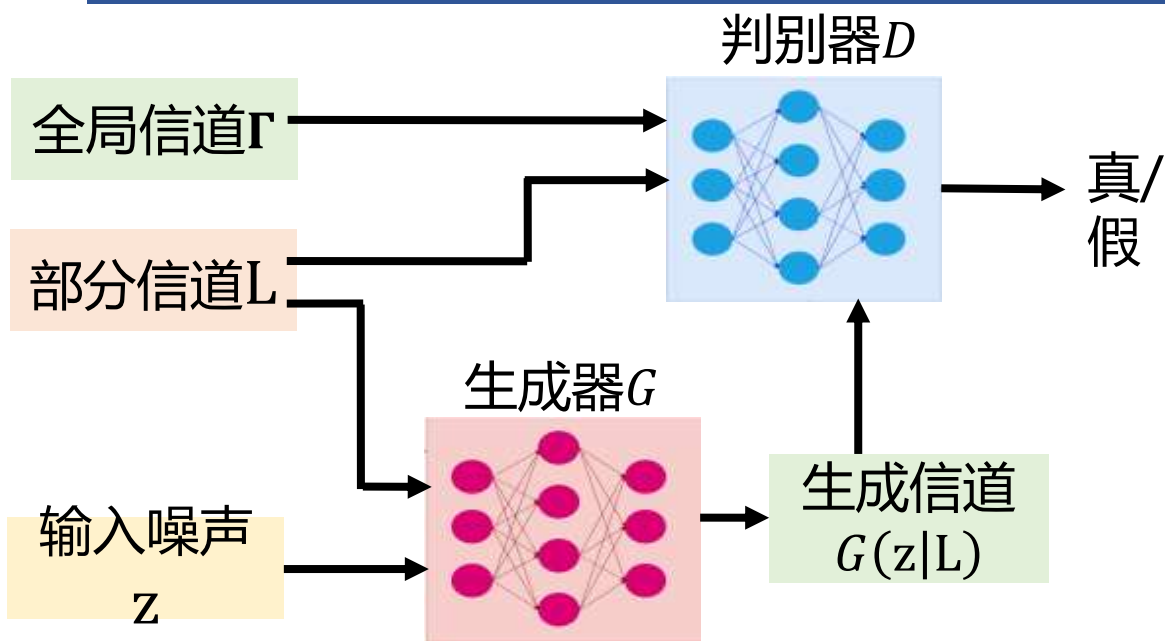
- **过度平滑问题**: CNN倾向于产生平滑结果丢失信道的特征
- **确定性映射**: 无法捕获信道固有的随机性和不确定性



解决方法

通过设计生成器 G 和判别器 D 相互对抗训练的cGAN架构, 以部分信道信息作为条件信息, 生成更真实的完整级联信道信息

基于条件生成对抗网络的RIS信道生成



- $G(z|L)$: 给定标签L, G 根据噪声 z 生成信道
- $D(x|L)$: 正确判别真实信道信息 x 的概率
- $D(G(z|L))$: 判别器认为生成信道是真的概率
- D 希望 $D(G(z|L))$ 很小, 即正确识别假信息
- G 希望 $D(G(z|L))$ 很大, 即成功骗过判别器

$$\mathcal{L}_2 = \mathbb{E}\{\|G(z|L) - \Gamma\|^2\}$$

总损失函数

$$\min_G \max_D V(G, D) = \underbrace{\mathbb{E}_{x \sim p_{\Gamma}(x)} \{\log(D(x|L))\}}_{D \text{ 正确识别真实信道信息的能力}} + \underbrace{\mathbb{E}_{z \sim p_Z(z)} \{\log(1 - D(G(z|L)))\}}_{G \text{ 骗过 } D \text{ 的能力}} + \underbrace{t\mathcal{L}_2}_{G \text{ 生成准确信息的能力}}$$

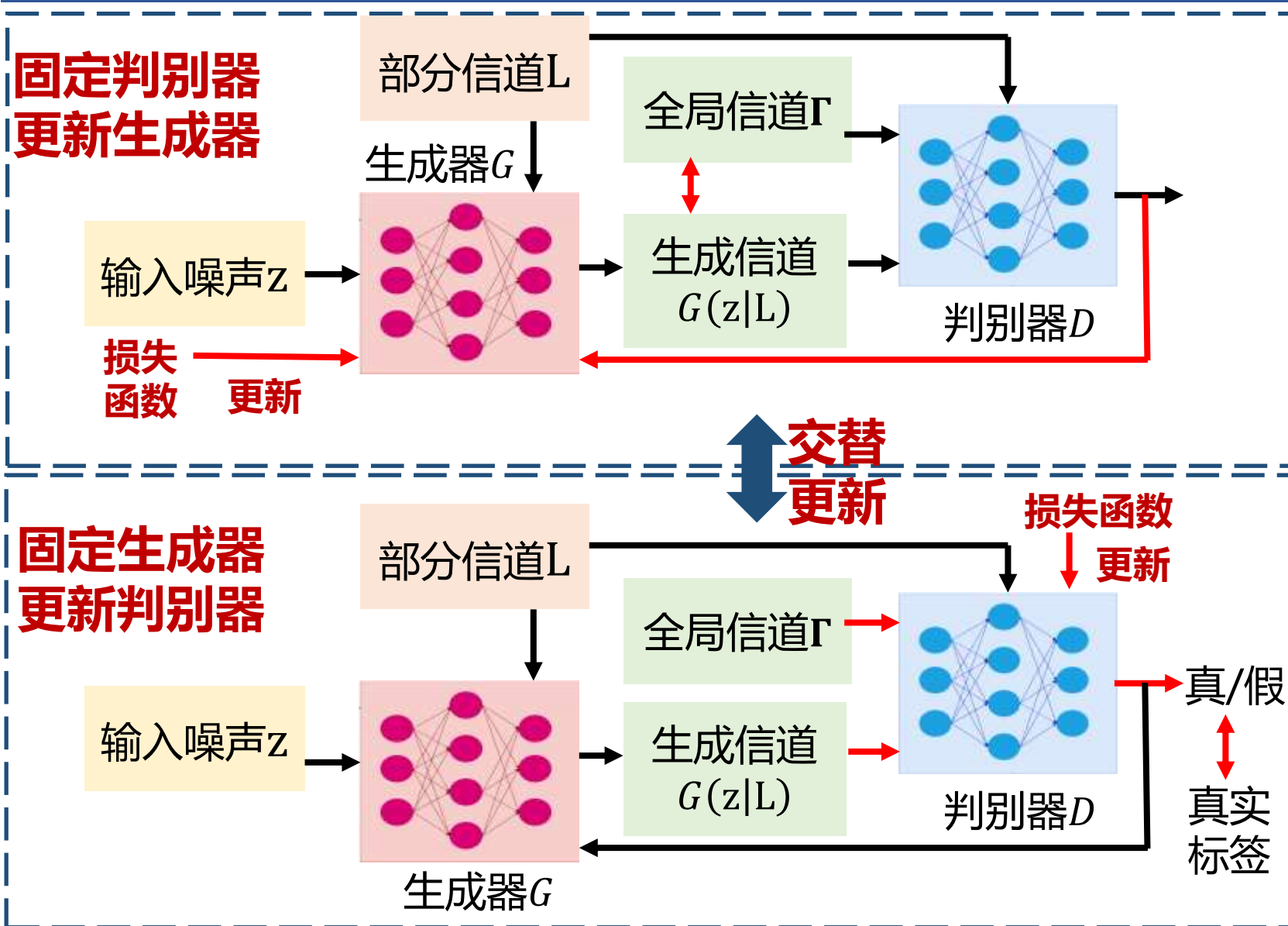
判别器损失函数:

$$\mathbb{E}_{x \sim p_{\Gamma}(x)} \{\log(D(x|L))\} + \mathbb{E}_{z \sim p_Z(z)} \{\log(1 - D(G(z|L)))\}$$

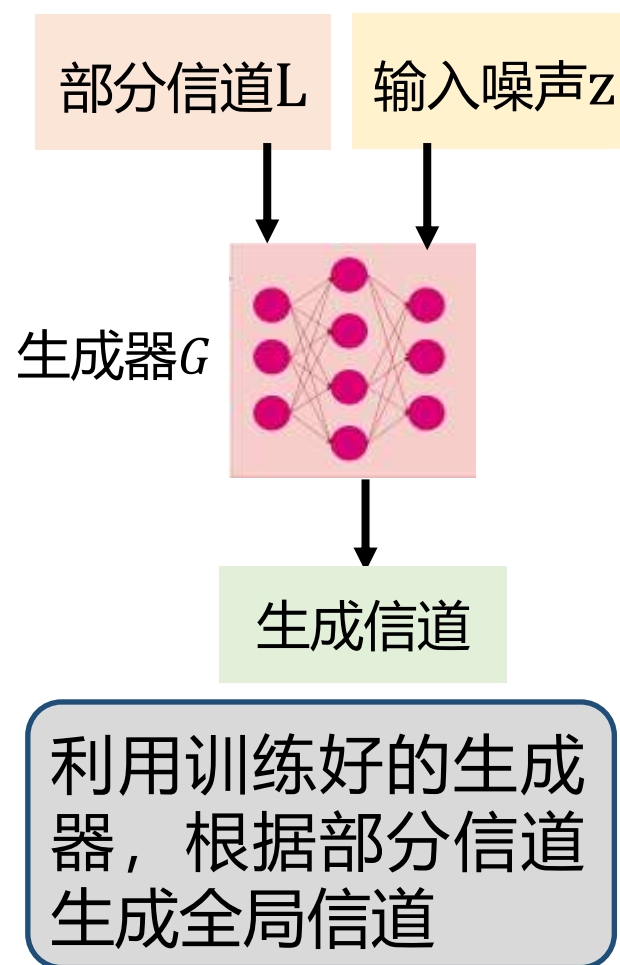
生成器损失函数:

$$\mathbb{E}_{z \sim p_Z(z)} \{\log(1 - D(G(z|L)))\} + t\mathcal{L}_2$$

基于条件生成对抗网络的RIS信道生成



推理流程



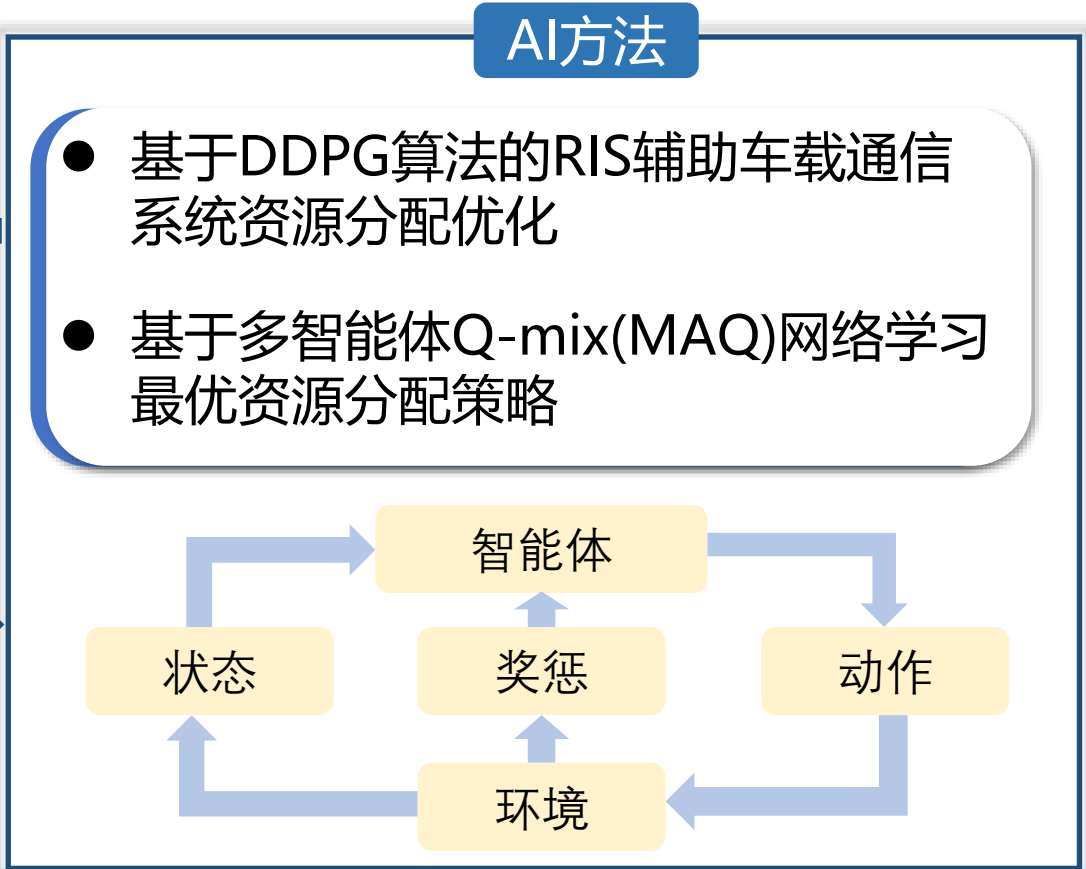
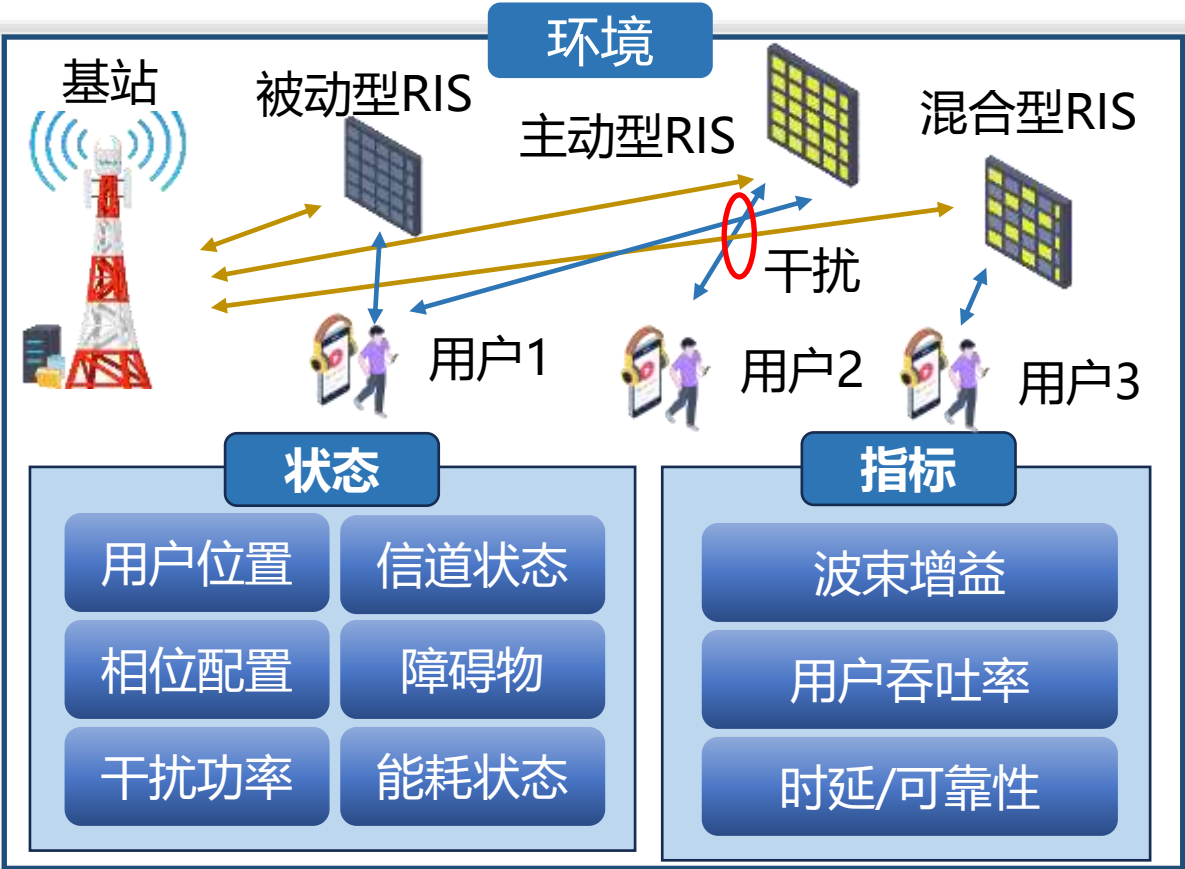
AI赋能RIS资源调度研究现状

传统方法挑战:

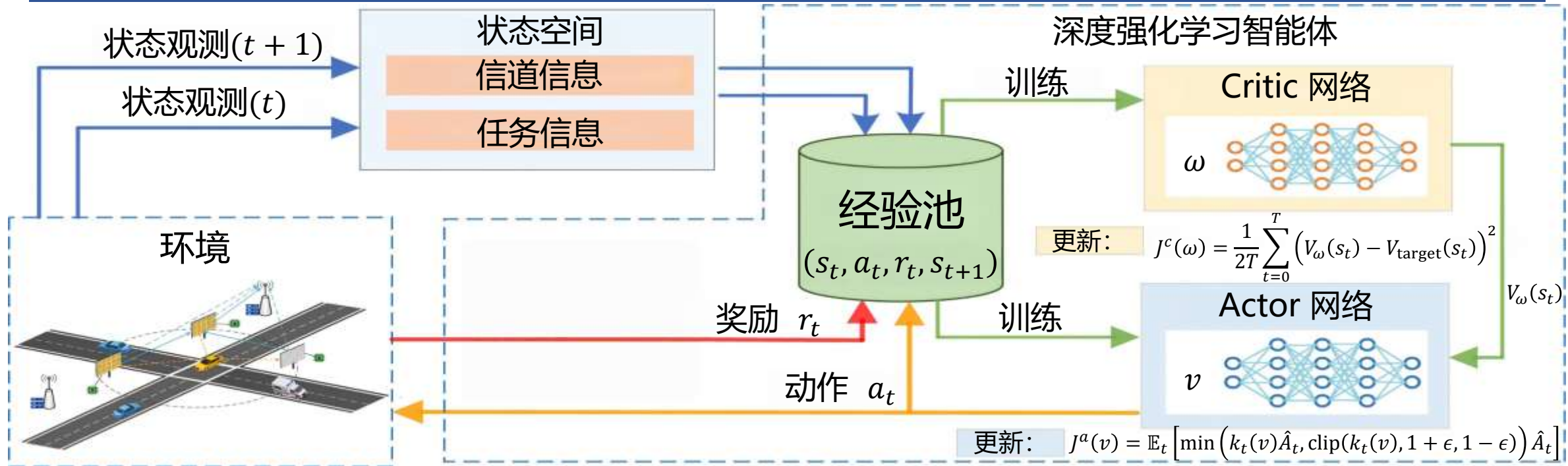
- 凸优化方法收敛慢，多维资源耦合优化困难
- 静态分配策略，无法适应动态负载

现有的研究已使用AI技术对RIS资源调度进行了探索

- 基于深度强化学习的动态场景资源调度
- 基于多智能体强化学习的协同优化调度



基于深度强化学习的RIS资源调度



状态空间定义

- 车辆-RIS信道增益: 车辆到RIS的信道质量
- RSU-RIS信道增益: RSU到RIS的信道质量
- RIS间信道增益: RIS之间的信道质量
- 任务数据量: 各车辆的任务比特数
- 时延约束: 各车辆任务的执行截止时间

动作空间定义

- 卸载目标 α : 离散变量, 决定车辆 m 是否将任务卸载到RSU k
- RIS相位偏移 θ : 连续变量, 控制各RIS反射元件的相位偏移 $[0, 2\pi]$
- 混合空间处理: 将离散变量 α 转换为连续形式统一优化, 便于PPO算法优化

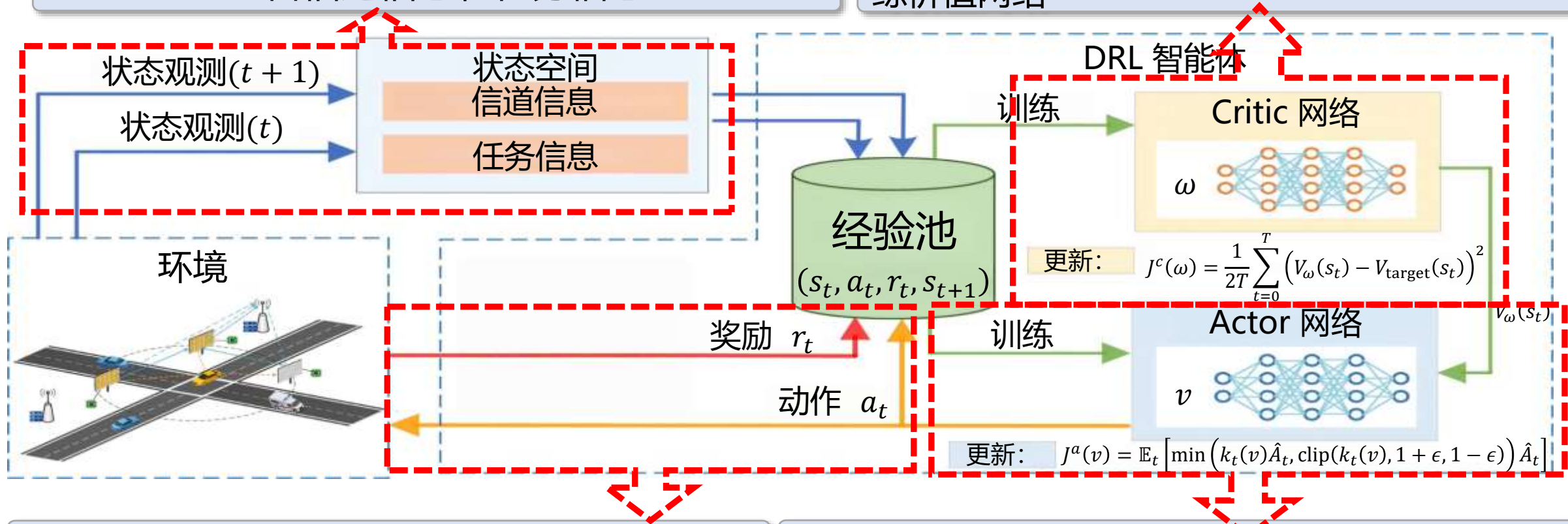
奖励函数定义

- 瞬时总卸载效率(SOE) $r = \frac{L_{\text{total}}}{E_{\text{total}}}$
- 分子 L_{total} : 车辆成功卸载任务比特总数
- 分母 E_{total} : 网络总能量消耗
- 优化目标: 最大化卸载效率, 即用最少能耗完成最多卸载任务
- 约束考虑: 时延、SINR等系统约束条件

基于深度强化学习的RIS资源调度

智能体从VEC环境中获取当前状态 s_t
包含信道信息和任务信息

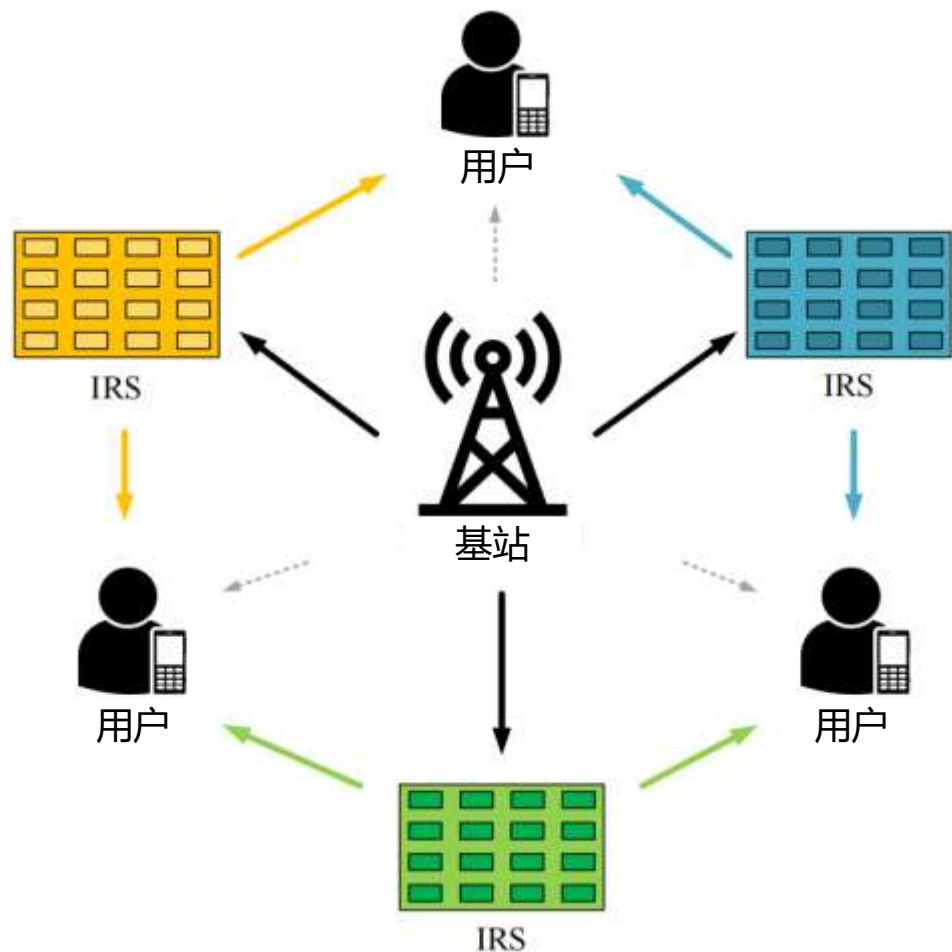
Critic网络更新：使用MSE损失函数 $J^c(\omega)$ 训练价值网络



环境交互：将动作 a_t 应用到双RIS辅助VEC网络
环境中执行，环境根据SOE性能计算即时奖励 r_t

Actor网络根据状态 s_t 输出动作 a_t (卸载决策 α 和相位偏移 θ)
Actor网络更新：使用PPO目标函数 $J^a(v)$ 更新策略网络

基于多智能体强化学习的RIS资源协同调度



核心优化目标

在多个RIS辅助的下行链路多用户MISO通信系统场景下，联合优化

- 基站发射波束成形
- RIS离散相移波束成形
- RIS相移分辨率和反射元件开/关状态

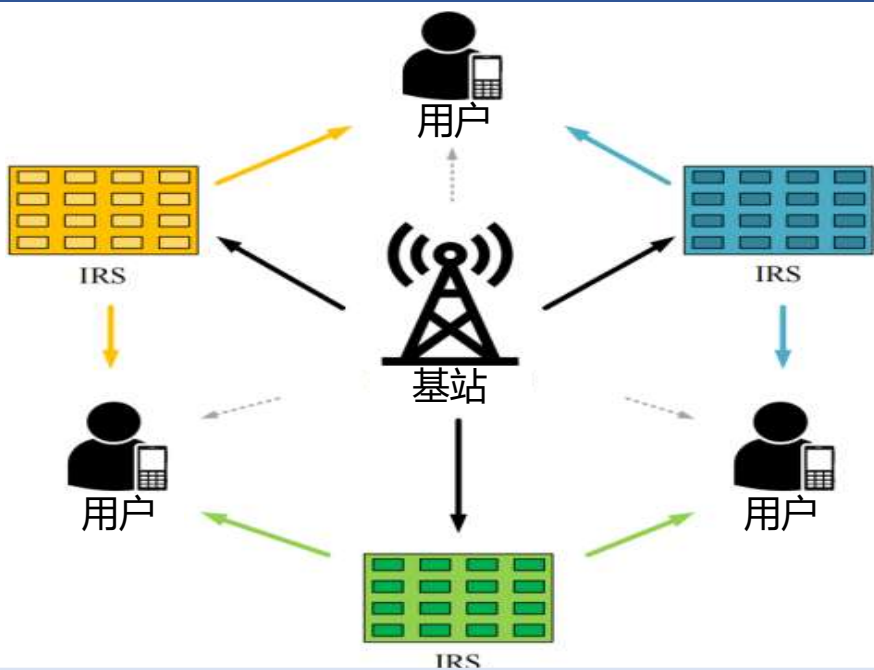
以最大化长期平均系统数据速率

解决方法

采用多智能体强化学习(MARL)框架求解

- 以信道状态信息和RIS能量状态为状态空间
- 以波束成形、相移配置和开/关状态为动作空间
- 使用多智能体Q-mix(MAQ)网络学习最优策略

基于多智能体强化学习的RIS资源协同调度



奖励函数定义

总奖励 $r_t \triangleq \sum_{k=1}^K R_{k,t} + \xi_1 \sum_{k=1}^K \min\{R_{k,t} - R_{k,t}^{\text{req}}, 0\} + \xi_2 \min\{P^{\text{max}} - \sum_{k=1}^K P_{k,t}, 0\} + \xi_3 \sum_{l=1}^L \min\{E_l(t) - E_l^{\text{min}}, 0\}$

- 第一项：系统总的数据传输速率
- 第二项：用户需求满足情况
- 第三项：功率是否超限
- 第四项：RIS能量是否不足

状态空间定义

RIS局部状态 $s_{l,t} \triangleq \left[\left\{ \exp\left(-\frac{d_{l,k}}{d_{l,0}}\right) \cdot f_k(R_{k,t-1}) \right\}_{k \in \mathcal{K}}, E_{l,t} \right]$

- 用户重要性 $\exp\left(-\frac{d_{l,k}}{d_{l,0}}\right)$ 表示用户 k 离 RIS l 有多近，越近越重要
- 服务质量 $f_k(R_{k,t-1})$ 表示用户 k 的需求是否得到满足
- $E_{l,t}$ 表示剩余能量

基站局部状态 $s_{B,t} \triangleq [\{P_{k,t}\}_{k \in \mathcal{K}}]$

- 表示基站给每个用户分配的发射功率

全局状态 $s_t \triangleq [\{s_{l,t}\}_{l \in \mathcal{L}}, s_{B,t}]$

动作空间定义

RIS局部动作 $a_{l,t} \triangleq [b_l, \theta_{l,t}, \rho_{l,t}]$

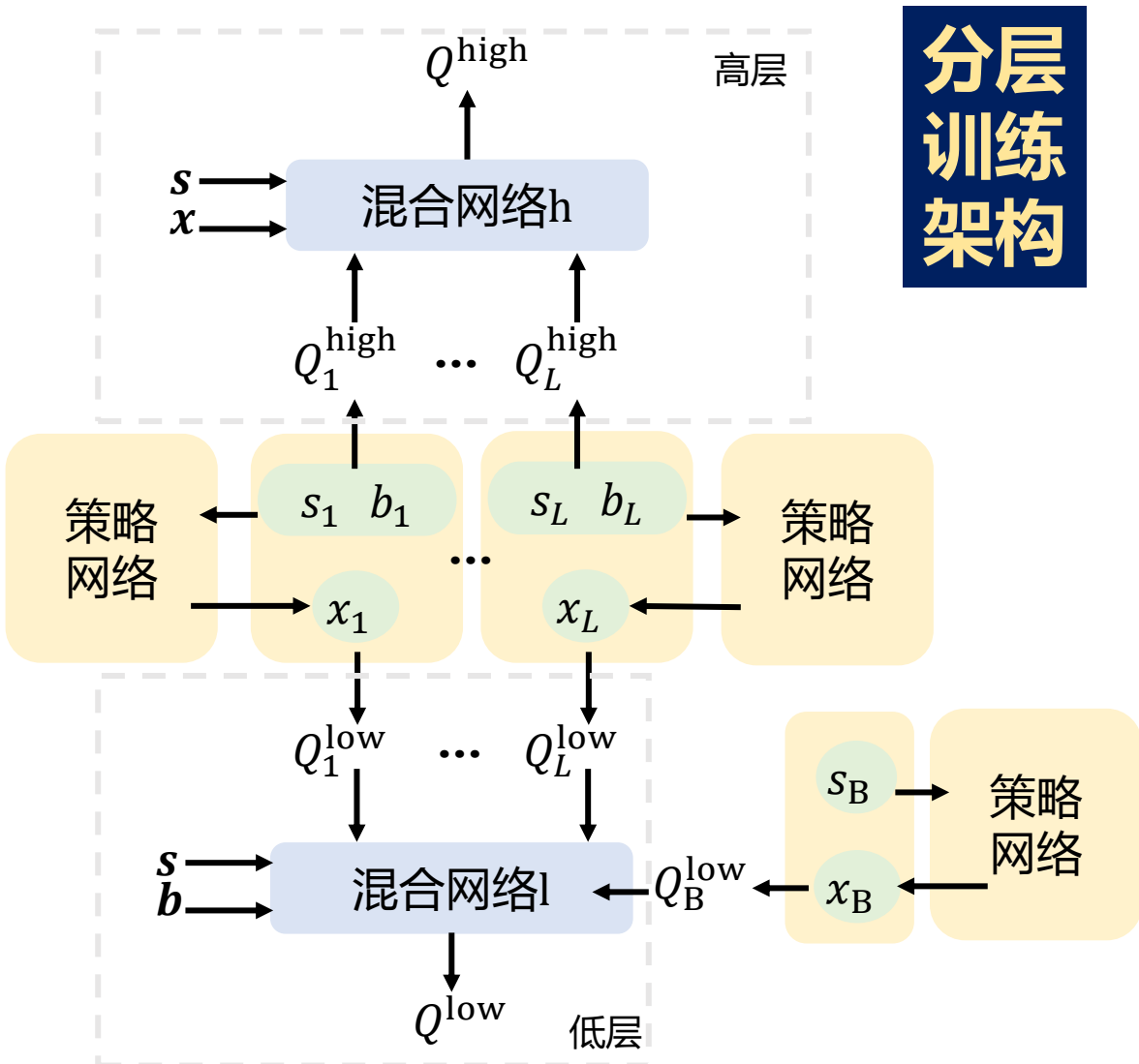
- 分辨率选择 b_l ，相移向量 $\theta_{l,t}$ ，开关状态 $\rho_{l,t}$

基站局部动作 $a_{B,t} \triangleq [\{v_{k,t}\}_{k \in \mathcal{K}}]$

- 表示基站给每个用户调制的波束赋形

全局动作 $a_t \triangleq [\{a_{l,t}\}_{l \in \mathcal{L}}, a_{B,t}]$

基于多智能体强化学习的RIS资源协同调度



高层决策

- 各RIS智能体根据局部状态 s_i 通过高层Q网络计算 Q_i^{high} 来选择相移分辨率 b_i
- 高层混合网络将所有个体Q值融合为全局 Q^{high} 评估整体分辨率配置的价值

执行

低层执行

- 策略网络根据状态 s 和分辨率 b 生成具体动作 x_i 和 x_B
- 各智能体通过低层Q网络计算 Q_i^{low} 和 Q_B^{low} 评估动作效果
- 低层混合网络融合产生全局 Q^{low} 评估整体执行方案

样本采集

参数更新

学习更新

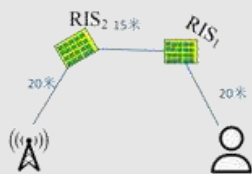
- 系统收集经验并分别更新高层网络(优化 Q^{high} 指导分辨率选择)
- 低层网络优化 Q^{low} 指导动作执行和策略网络参数
- 通过不断迭代使多智能体学会协调配合以最大化系统性能

- ◆ 1. AI赋能RIS的研究背景
- ◆ 2. 基于生成式AI的高效RIS信道获取
- ◆ 3. 基于强化学习的RIS信道动态感知
- ◆ 4. 基于预训练决策模型的RIS资源分配

级联RIS信道场景

应用场景

长距离通信



复杂遮挡环境



应急通信



多跳RIS的技术挑战:

- 信道维度爆炸性增长
- 级联信道建模困难
- 毫秒级信道状态更新需求
- 计算复杂度高

解决方案

利用生成式模型在**高维度、高质量生成**的先天优势来对多跳RIS进行信道生成

级联RIS系统框图

RIS₁阵元为 M_1

RIS₂阵元为 M_2

基站天线为 N

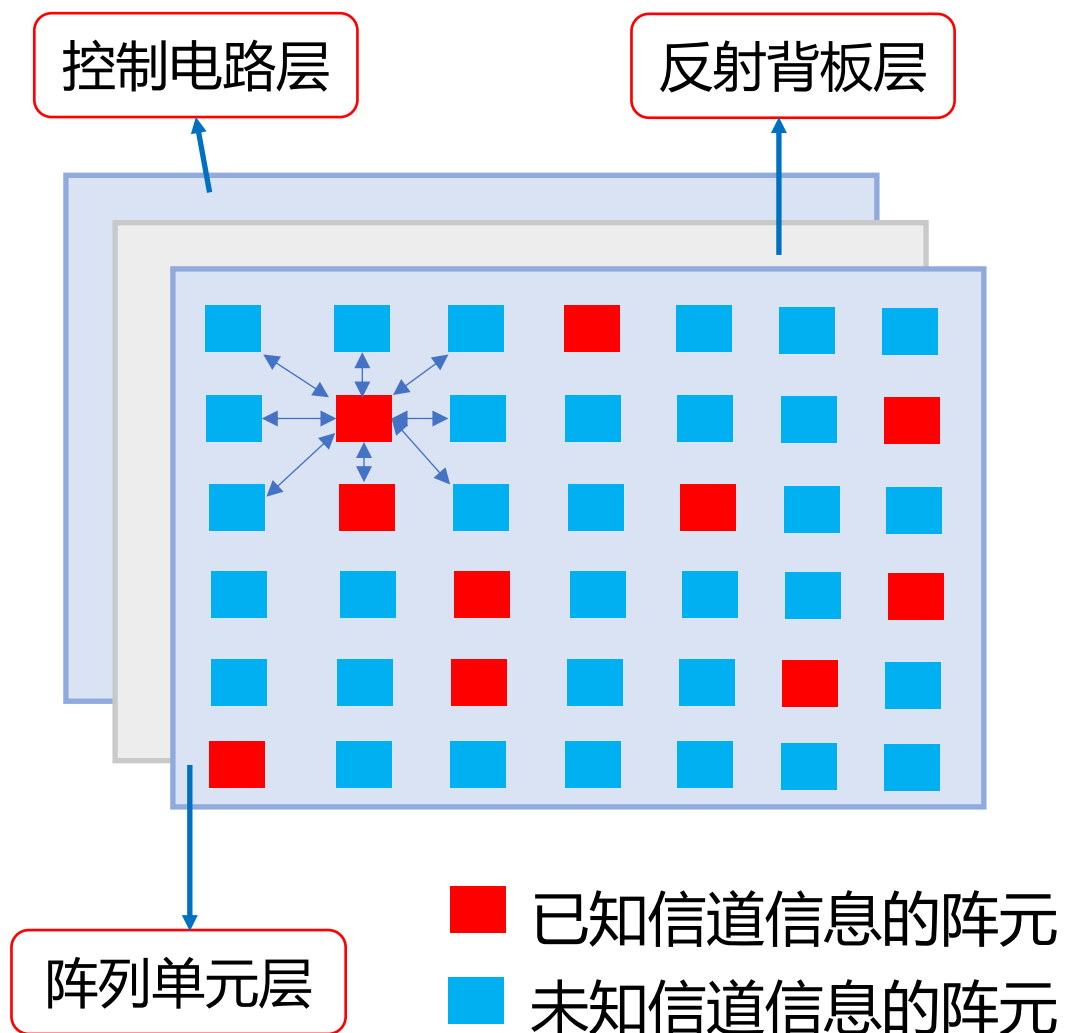
用户为单天线



基本元素:

- 用户信号通过RIS₁与RIS₂到达基站
- 用户到RIS₁信道: $\mathbf{u} \in \mathbb{C}^{M_1 \times 1}$
- RIS₁到RIS₂信道: $\mathbf{D} \in \mathbb{C}^{M_2 \times M_1}$
- RIS₂到基站信道: $\mathbf{G} \in \mathbb{C}^{N \times M_2}$

RIS生成式信道获取建模



高成本挑战：对RIS每个阵元都进行信道估计会导致高昂的导频开销，减少数据传输的可用时间

精度限制：仅利用部分阵元对应链路的信道信息无法进行准确的波束赋形设计

考虑以上特性，利用**RIS阵元与阵元之间的空间相关性**预测所有阵元对应的 CSI

假设空间相关系数矩阵（不唯一）：

$$\Omega_{i,j} = \begin{cases} 1, & k = 0 \\ \frac{\sin(k)}{k}, & \text{otherwise} \end{cases} \quad k = \frac{2\pi|i-j|d}{R}$$

i, j 为阵元, d 为阵元之间的距离, R 为波长

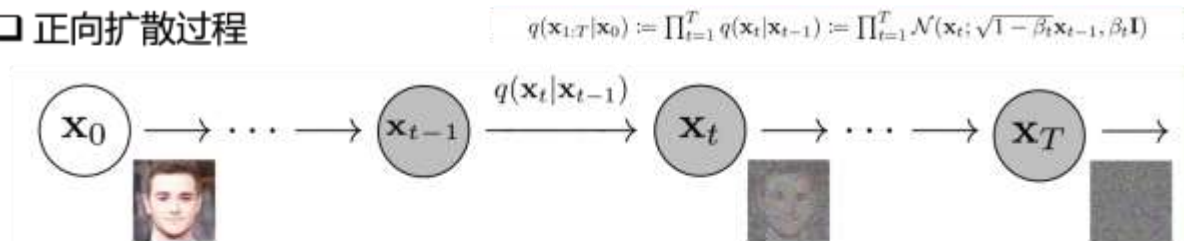
扩散模型 (DDPM) 介绍

扩散模型包括两个过程：

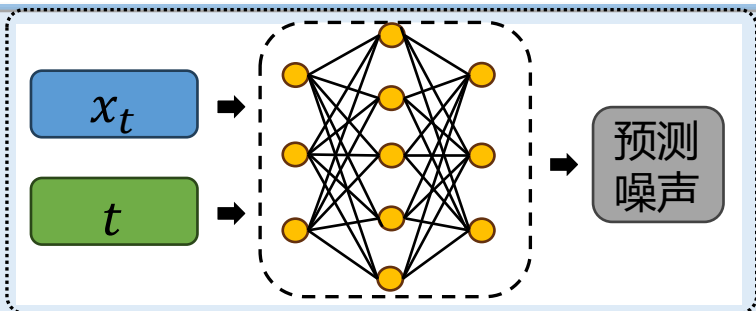
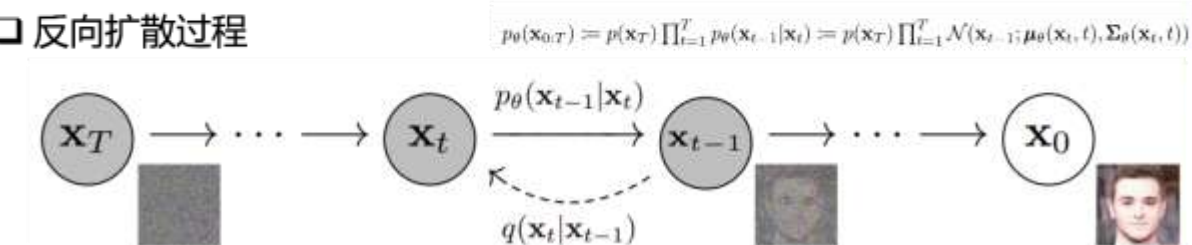
前向过程 (forward process) 和 **反向过程** (reverse process)，其中前向过程又称为扩散过程 (diffusion process)

无论是前向过程还是反向过程都是一个**参数化的马尔可夫链 (Markov chain)**，其中反向过程可用于生成数据样本（作用类似GAN中的生成器，只不过GAN生成器会有维度变化而DDPM的反向过程没有维度变化）

□ 正向扩散过程



□ 反向扩散过程



DDPM的训练目标是**最小化重建噪声的均方误差(MSE)**

$$\mathcal{L}_{\text{DDPM}} = E_{t, x_0, \epsilon} (||\epsilon - \epsilon_\theta(x_t, t)||)^2$$

即，用神经网络预测加噪前的噪声，并最小化预测误差

扩散模型优势：



训练稳定性

扩散模型采用确定性的噪声添加过程，无需复杂的对抗训练机制，确保了训练过程的**稳定性**和可重复性



生成质量

通过渐进式去噪过程，扩散模型能够生成具有最先进质量的数据，为复杂数据分布建模提供了强有力支撑



计算效率

扩散模型具有高度并行化的计算特性，能够利用现代GPU架构实现高效处理，适合**大规模实际应用部署**

基于条件扩散模型的信道生成算法

1 数据准备工作



对于单条链路而言，信号从用户出发，经过 将全部阵元信道信息作为**训练样本**，作实值化处理

- RIS₁的第 m_1 个阵元
- 到达RIS₂的第 m_2 个阵元
- 最终到达基站的第 n 个天线

数据准备



$$\mathcal{H} \in \mathbb{C}^{N \times M_2 \times M_1} \rightarrow \mathbf{x}_0 \in \mathbb{R}^{2NM_2M_1 \times 1}$$

将部分阵元信道信息作为**训练条件**，作实值化处理



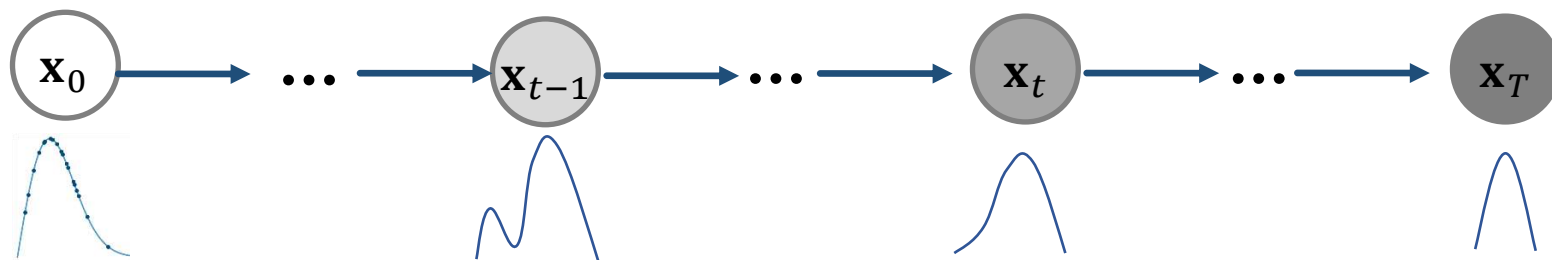
$$\mathcal{H}^P \in \mathbb{C}^{N \times M_2^P \times M_1^P} \rightarrow \mathbf{x}^P \in \mathbb{R}^{2NM_1^P M_2^P \times 1}$$

这条单链路总信道 $h_{m_1, m_2, n} = g_{n, m_2} d_{m_2, m_1} u_{m_1}$

总信道链路集合为 $\mathcal{H} = \{h_{m_1, m_2, n}, \dots, h_{M_1, M_2, N}\}$

基于条件扩散模型的信道生成算法

2 正向扩散阶段 核心是通过逐步添加高斯噪声，使 $\mathbf{x}_0$ 的分布逐渐“扩散”为近似标准高斯分布



- 输入 $\mathbf{x}_0 \in \mathbb{R}^{2NM_1M_2 \times 1}$
- 在正向阶段中，向 $\mathbf{x}_0$ 的每一行元素添加**服从高斯分布**的人工噪声 $\epsilon_t \in \mathbb{R}^{2NM_1M_2 \times 1}$
- 经过 T 步加噪，最终到 $\mathbf{x}_T$ ，使得 $\mathbf{x}_T$ 近似服从标准高斯分布

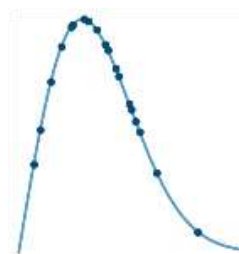
加噪公式

- $\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \epsilon_t$
- 递推可得第 t 步噪声 $\mathbf{x}_t$ 公式
$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \bar{\epsilon}_t$$
，其中 $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$

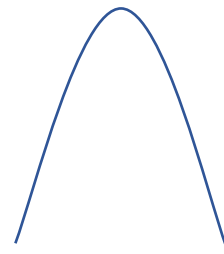
因此我们可以从 $\mathbf{x}_0$ 直接一步加噪得到 $\mathbf{x}_t$

为何要不断扩散加噪？

原始数据服从复杂的分布
 $\mathbf{x}_0 \sim p(\mathbf{x}_0)$



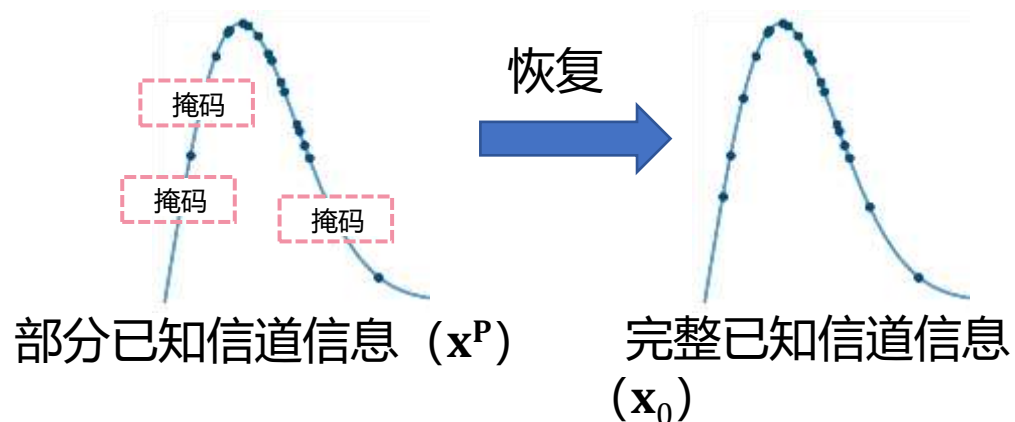
逐步添加高斯噪声
慢慢变成纯高斯噪声



只需要学习从简单分布逐步还原到复杂分布，而不是一次性生成数据

基于条件扩散模型的信道生成算法

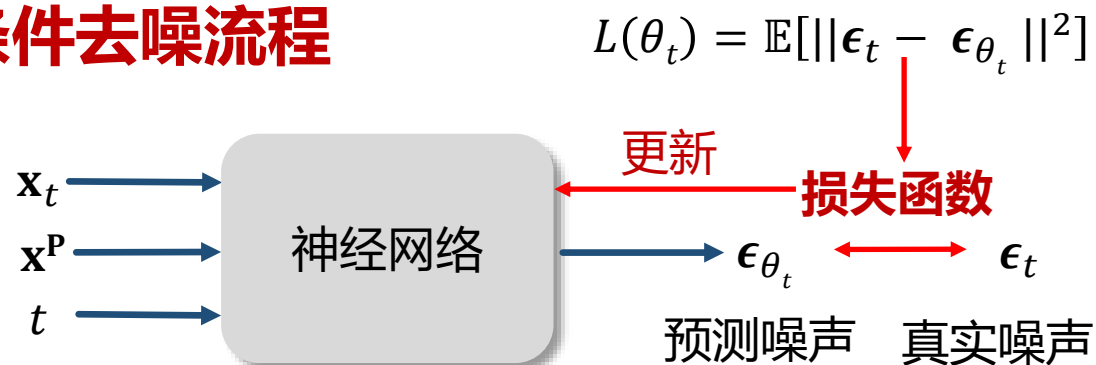
3 条件反向去噪训练阶段



条件信道生成原理

- 信道就像一张拼图，我们只知道其中一部分图。扩散模型就是要学会怎么还原其他部分的拼图，一步一步的减小噪声，还原整张拼图
- 我们加入的已知信道作为条件就是要告诉模型一边修复信道一边参考已有的“线索”

条件去噪流程



- 输入：
 - ① 正向扩散阶段获得的 $\mathbf{x}_t \in \mathbb{R}^{2NM_1M_2 \times 1}$
 - ② 部分信道作为条件 $\mathbf{x}^P \in \mathbb{R}^{2NM_1^pM_2^p \times 1}$
 - ③ 时间步 t
- 通过每次去噪过程中，由神经网络预测输出的噪声 ϵ_{θ_t} 与 ϵ_t 计算损失函数 $L(\theta_t) = \mathbb{E}[\|\epsilon_t - \epsilon_{\theta_t}\|^2]$
- 模型学习到了正向过程中该步添加的噪声 ϵ_t ，为后续推理生成阶段提供帮助

基于条件扩散模型的信道生成算法

3 条件反向去噪训练阶段

1. 随机选择一个时间步并进行编码

从数据集采样一批次的全部信道

将信道掩码之后做为条件

$t \xrightarrow{\text{编码}} \text{时间步嵌入}$

$\mathcal{H} \rightarrow \mathbf{x}_0 \in \mathbb{R}^{2NM_2M_1 \times 1}$

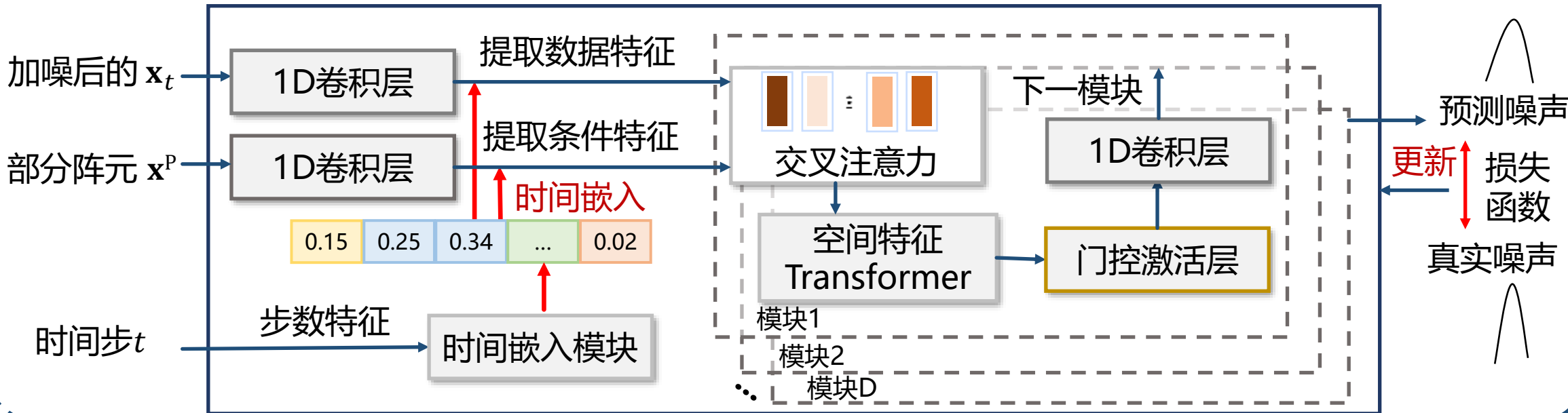
$\mathcal{H}^P \rightarrow \mathbf{x}^P \in \mathbb{R}^{2NM_1^pM_2^p \times 1}$

2. 对初始 $\mathbf{x}_0$ 进行加噪

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_t$$

得到第 t 步噪声

3. 训练神经网络



基于条件扩散模型的信道生成算法

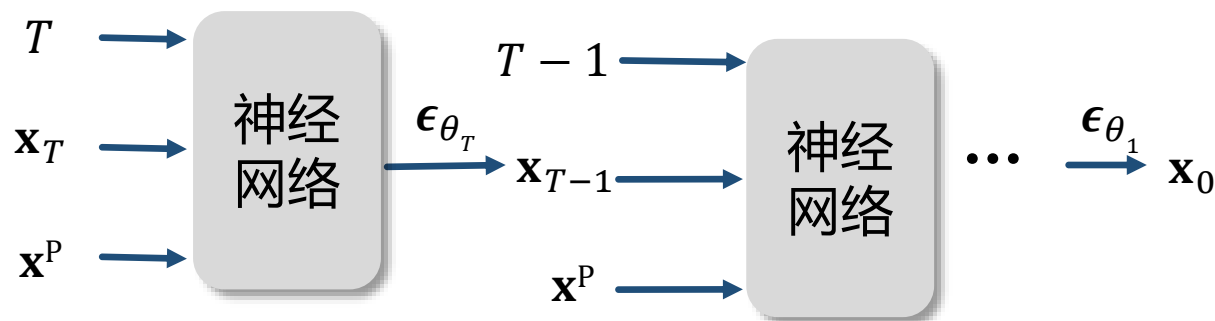
4 条件反向去噪生成阶段

- **初始化**: 从标准高斯分布中采样初始噪声向量作为初始带噪数据 $\mathbf{x}_T \in \mathbb{R}^{2NM_1M_2 \times 1}$
- **去噪**: 输入 $\mathbf{x}_T$ 、 t 、 $\mathbf{x}^P$ 到已训练好的神经网络模型 (模型已经训练好噪声 ϵ_θ)

$$\begin{aligned} & \mathbf{x}_{t-1} \\ &= \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta_t} \right) \\ &+ \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} (1 - \alpha_t) \mathbf{z}_t, t = T, T - 1, \dots, 1 \end{aligned}$$

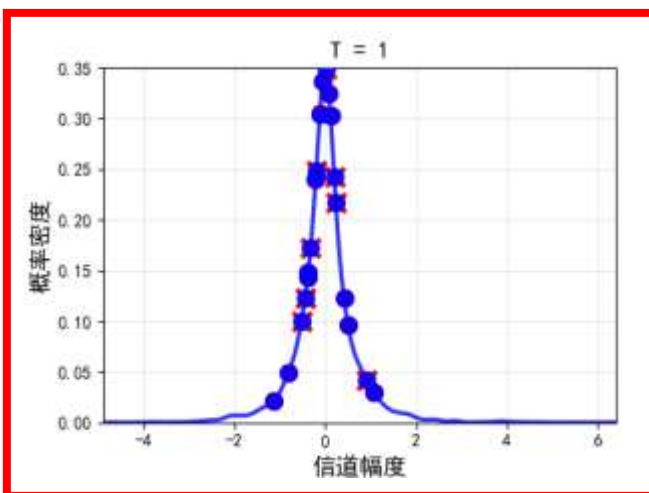
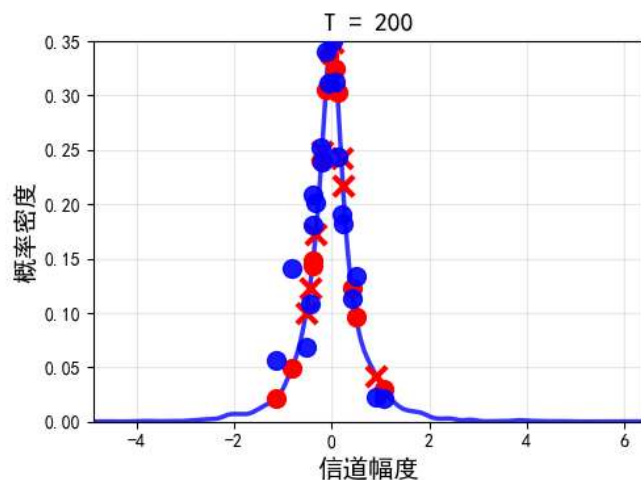
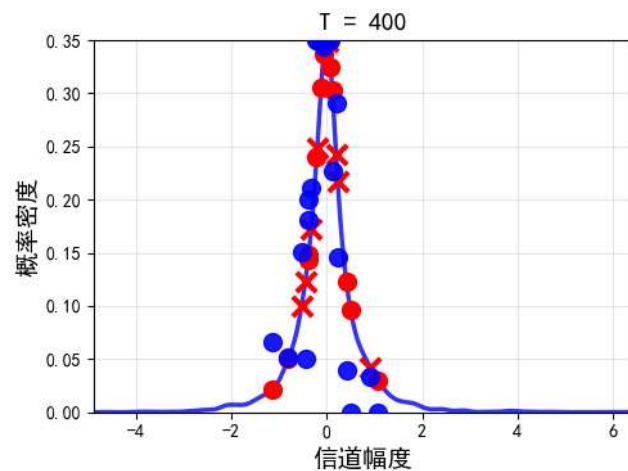
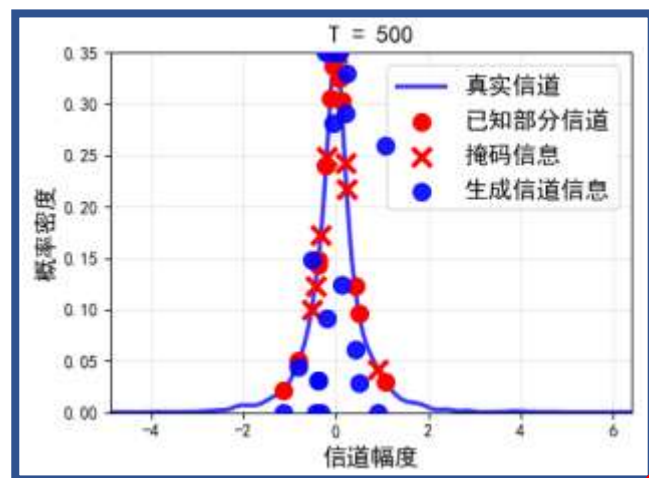
- **迭代去噪 (从 $t = T$ 到 $t = 1$)**: 从训练阶段获得的每一步噪声进行去噪从而得到 $\mathbf{x}_0 \in \mathbb{R}^{2NM_1M_2 \times 1}$
- **恢复信道**: $\hat{\mathcal{H}}$

条件去噪生成流程图



- 神经网络接受当前时刻的带噪样本、部分条件信息和当前的时间步
- 根据训练过程学习到的噪声去指导带噪样本进行去噪
- 模型最终输出生成后的全部信道样本

条件扩散模型信道生成过程



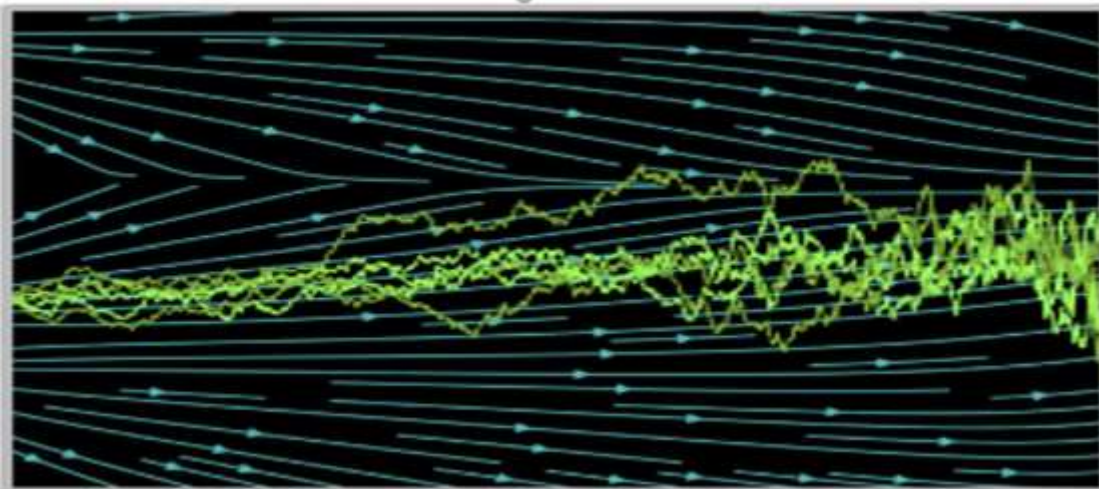
左图展示了条件扩散模型的信道生成过程，对比了模型生成的信道与真实信道的差异，我们能观察到：

- 在去噪初期阶段，条件扩散模型**生成的样本**与**真实样本**之间存在明显差异
- 随着条件扩散模型逐渐去噪，生成信道逐**渐逼近真实信道**，最终得到的信道与真实信道样本高度吻合
- 条件扩散模型学习的信道分布足够接近真实的信道

校正流模型介绍

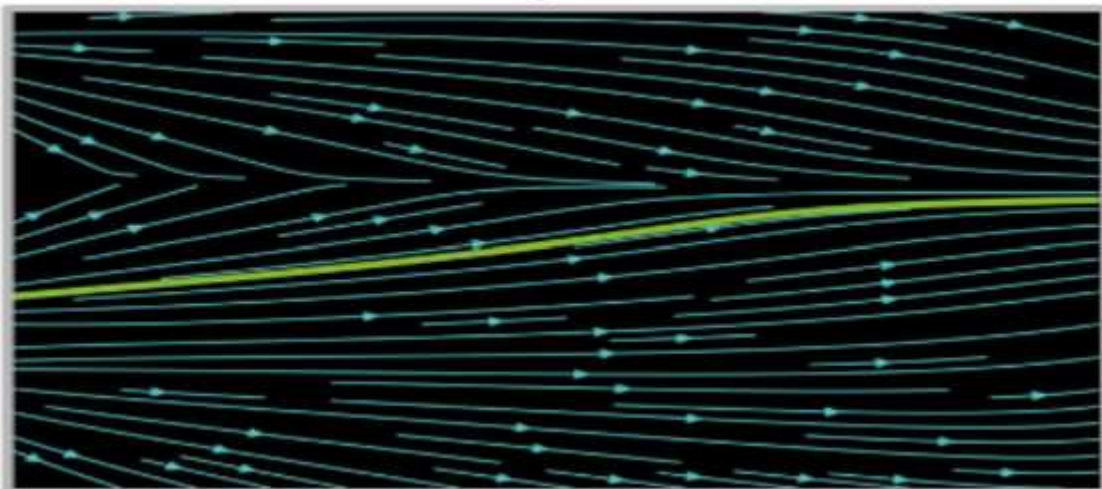
DDPM的数据生成过程依赖**随机微分方程 (Stochastic Differential Equation, SDE)** 具有强随机性，每一步涉及噪声采样

- **数学形式**: 通过SDE建模数据演化路径
- **训练阶段**: 建模数据向噪声的扩散过程并训练神经网络预测去噪方向
- **推理阶段**: 从噪声出发，迭代执行反向去噪过程

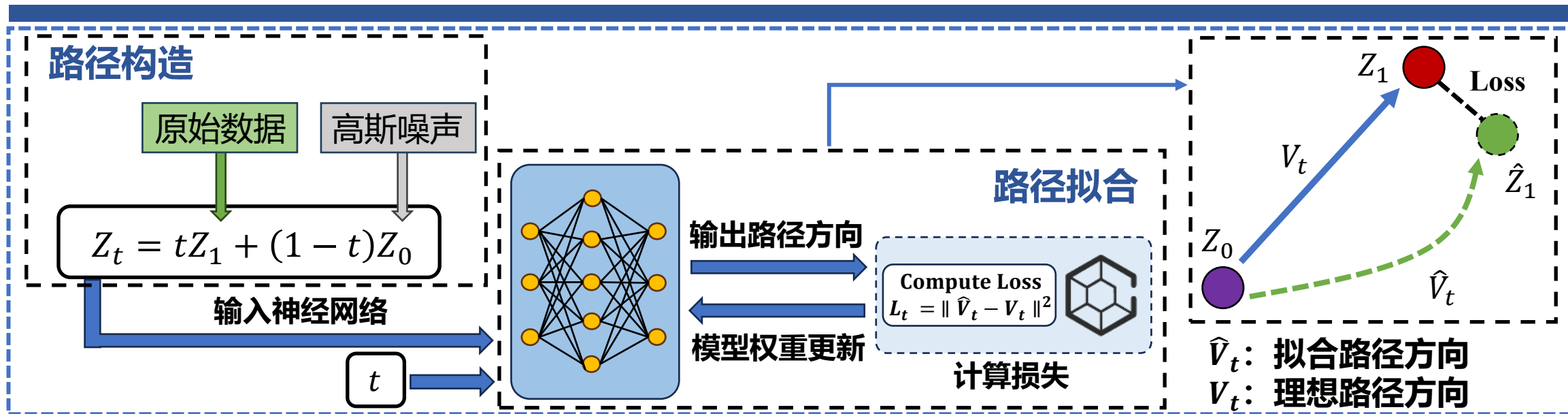


校正流 (Rectified Flow) 直接学习从初始噪声分布到目标数据分布之间的**确定性演化路径**，该路径由**常微分方程 (Ordinary Differential Equation, ODE)** 所驱动

- **数学形式**: 通过ODE建模数据演化路径
- **训练阶段**: 拟合真实数据与初始噪声之间插值路径上的路径引导
- **推理阶段**: 通过求解ODE从噪声生成样本效率高，误差可控



校正流模型训练过程



训练过程主要分为两个部分：**路径构造与路径拟合**

路径构造： 使用线性插值构造路径，对于每个数据对 (Z_0, Z_1) 构造**统一的线性插值路径**

$Z_t = tZ_1 + (1 - t)Z_0, t \in [0, 1],$
其中 $Z_0 \sim \mathcal{N}(0, I)$ 是高斯噪声， $Z_1 \sim p_{\text{data}}$ 是目标数据

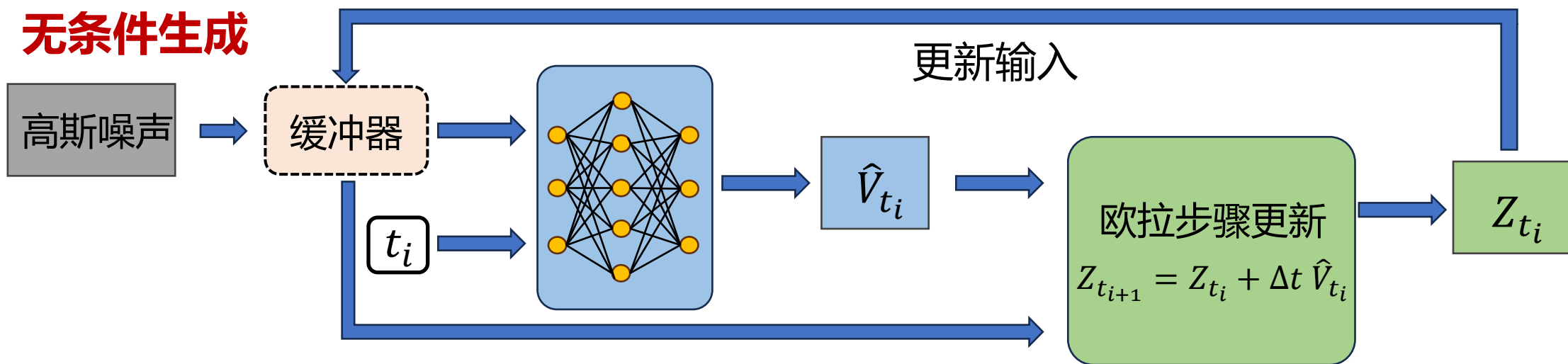
路径拟合： 以**差分形式**拟合路径直接计算真实路径方向

$$V_t = \frac{dZ_t}{dt} = Z_1 - Z_0,$$

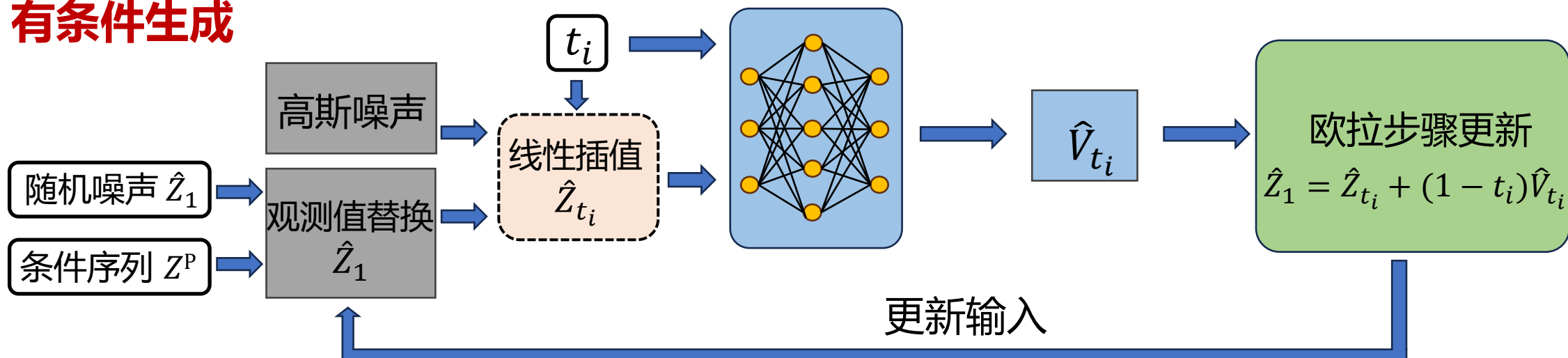
输入路径点 Z_t 和时间步 t ，输出在时间步 t 上的路径方向，实现**最小化拟合误差**

校正流模型推理过程

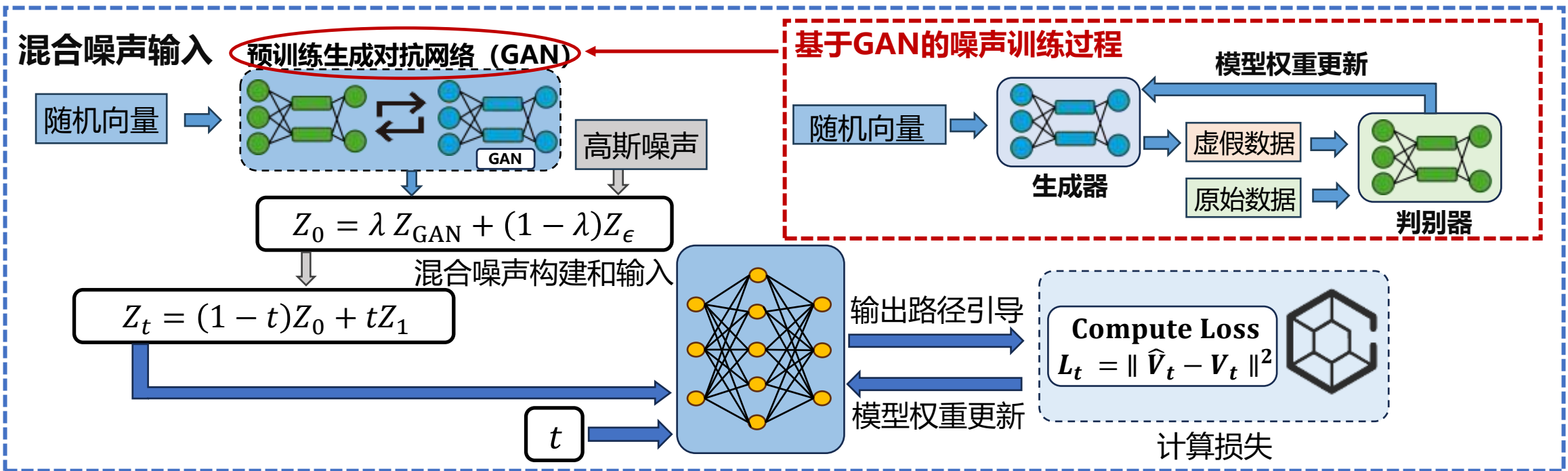
无条件生成



有条件生成



校正流模型的噪声输入改进



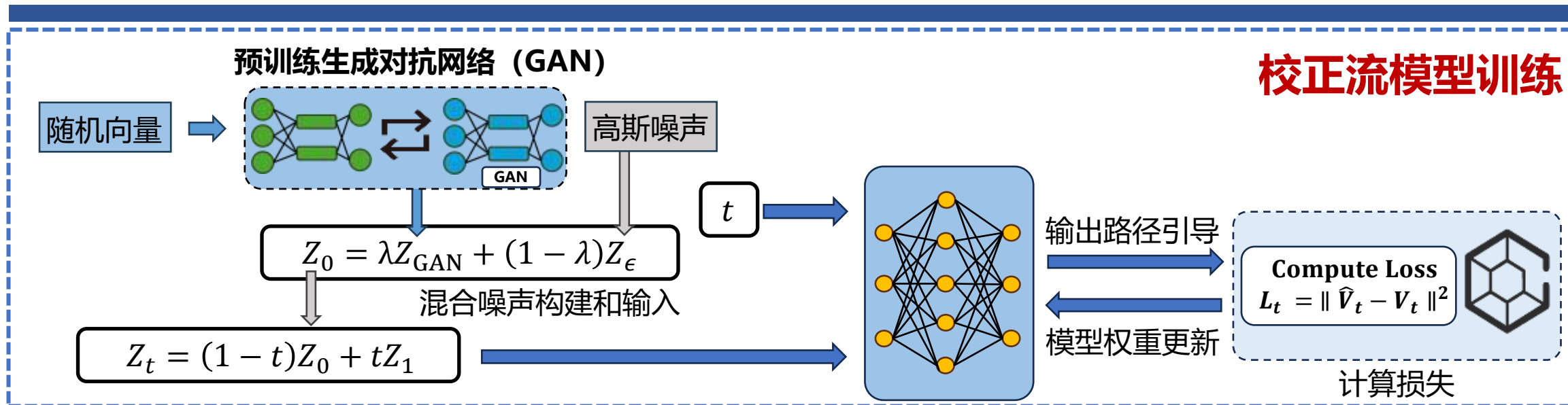
噪声输入存在问题:

- 高斯噪声与数据分布不对齐, 结构错配严重
- 高斯噪声出发点太散, 路径长, 模型很难拟合
- 高斯噪声对复杂分布建模能力弱

混合噪声引入动机:

- 增强语义先验, 混合噪声引入真实先验, 提升输入语义一致性
- 改善路径几何, 生成的初始状态更靠近目标数据流形
- 提升表达能力, 兼顾分布覆盖与特征保真度, 支持复杂多模态分布建模

校正流模型辅助RIS信道生成



单次训练迭代的完整流程:

1. 数据采样:

- 从训练集中采样真实样本 $\mathbf{H} \in \mathbb{C}^{N \times M_2 \times M_1} \rightarrow Z_1 \in \mathbb{R}^{2NM_2M_1 \times 1}$
- 生成混合噪声 $Z_0 = \lambda Z_{\text{GAN}} + (1 - \lambda) Z_{\epsilon}$
- 从Logit-Normal分布采样时间步 t , $t \in [0, 1]$

2. 构造中间状态

- 计算插值状态: $Z_t = tZ_1 + (1 - t) Z_0$
- 计算目标方向: $V_t = Z_1 - Z_0$

3. 模型前向传播

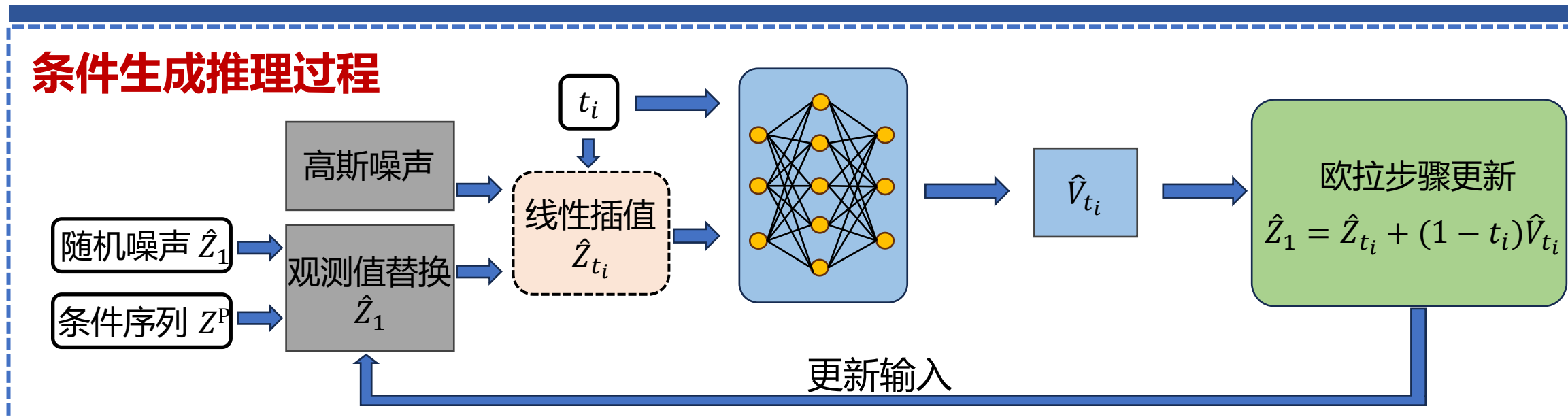
- 将 Z_t 和 t 输入到流匹配网络 G_{θ}
- 获得预测的速度场: $\hat{V}_t = G_{\theta}(Z_t, t)$

4. 损失计算

- 计算预测误差: $L_t = \|\hat{V}_t - V_t\|^2$
- 损失衡量预测方向与理想方向的差距, 使模型逐步学会正确的传输方向

校正流模型辅助RIS信道生成

条件生成推理过程



部分已知阵元的级联信道: $\mathbf{H}^P \in \mathbb{C}^{N \times M_2^P \times M_1^P} \rightarrow \mathbf{Z}^P \in \mathbb{R}^{2NM_1^P M_2^P \times 1}$

流匹配模型采样过程:

1. 随机初始化: 用标准正态分布的随机噪声初始化目标序列

列 $\hat{\mathbf{Z}}_1 \sim \mathcal{N}(0, \mathbf{I})$

2. 计算自适应时间步:

$$t_i = (i/N)^k, \text{ for } i = 0 \text{ to } N - 1$$

3. 采样高斯噪声: $\mathbf{Z}_0 \sim \mathcal{N}(0, \mathbf{I})$

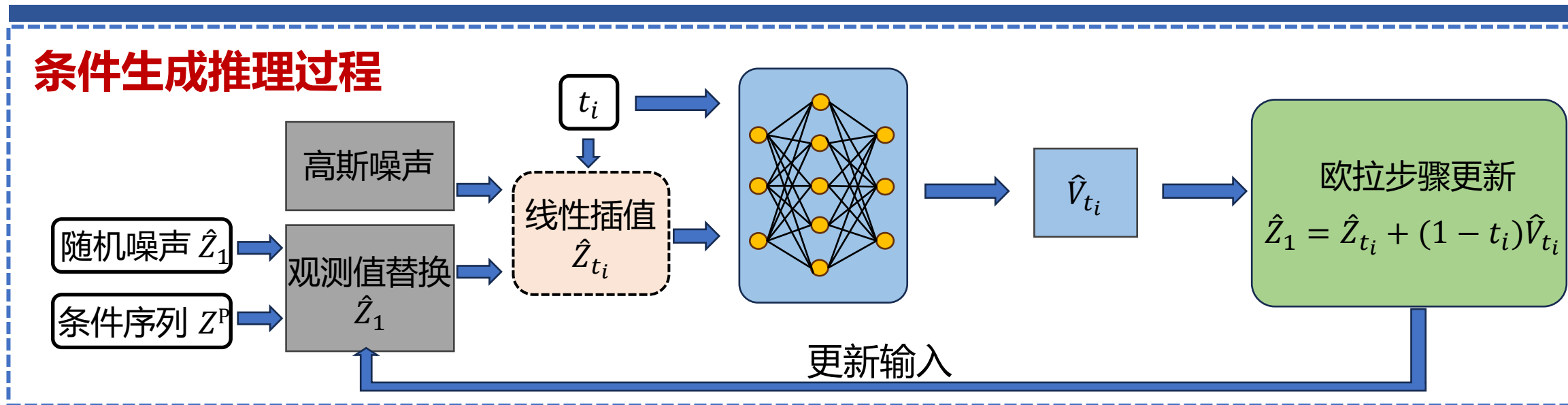
4. 观测值替换:

$$\hat{\mathbf{Z}}_1 = \hat{\mathbf{Z}}_1 \odot (1 - \mathbf{M}) + \mathbf{Z}^P \odot \mathbf{M}$$

通二进制掩码 $\mathbf{M}$ 将目标序列 $\hat{\mathbf{Z}}_1$ 中观测位置的值替换为给定的条件值, 实现条件控制

校正流模型辅助RIS信道生成

条件生成推理过程



5. 线性插值 $\hat{Z}_{t_i} = t_i \hat{Z}_1 + (1 - t_i) Z_0$

- 在当前状态 $\hat{Z}_1$ 和噪声 Z_0 之间进行加权平均
- t_i : 权重系数, 控制插值比例
- $\hat{Z}_{t_i}$: 插值后的中间状态

6. 流匹配预测 $\hat{V}_{t_i} = G_\theta(\hat{Z}_{t_i}, t_i)$

- G_θ 训练好的流匹配神经网络
- 速度场 $\hat{V}_{t_i}$, 指示如何从当前状态向目标分布移动

7. 欧拉步骤更新 $\hat{Z}_1 = \hat{Z}_{t_i} + (1 - t_i) \hat{V}_{t_i}$

- 欧拉方法: 数值积分的一阶方法
- 更新后的 $\hat{Z}_1$, 更接近目标分布

8. 完成 N 次迭代后, $\hat{Z}_1$ 从初始噪声转换为符合目标分布的样本, $\hat{Z}_1$ 在观测位置匹配给定条件, 在未观测位置是合理的生成值

最终生成全部级联信道:

$$\hat{Z}_1 \in \mathbb{R}^{2NM_1^P M_2^P \times 1} \rightarrow \hat{\mathcal{H}} \in \mathbb{C}^{N \times M_2 \times M_1}$$

仿真参数设置



环境仿真参数设置:

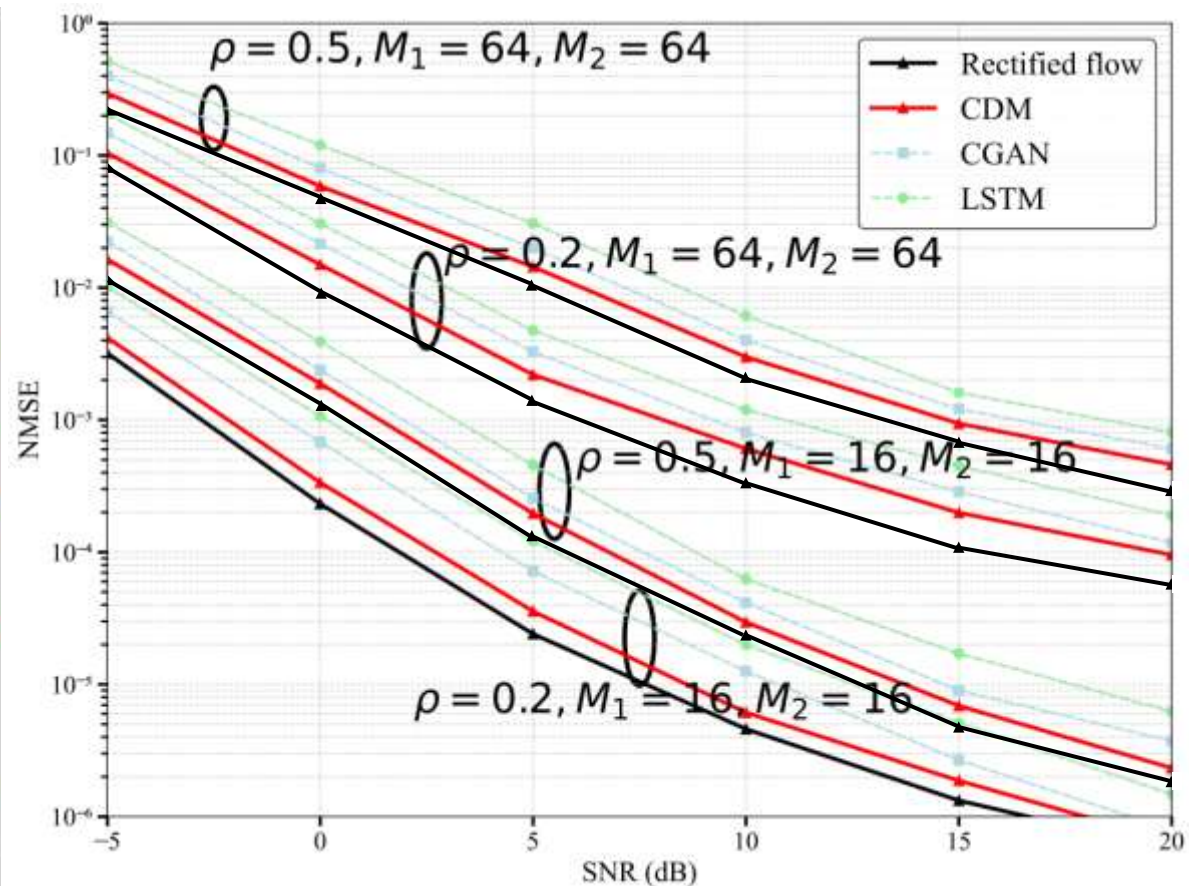
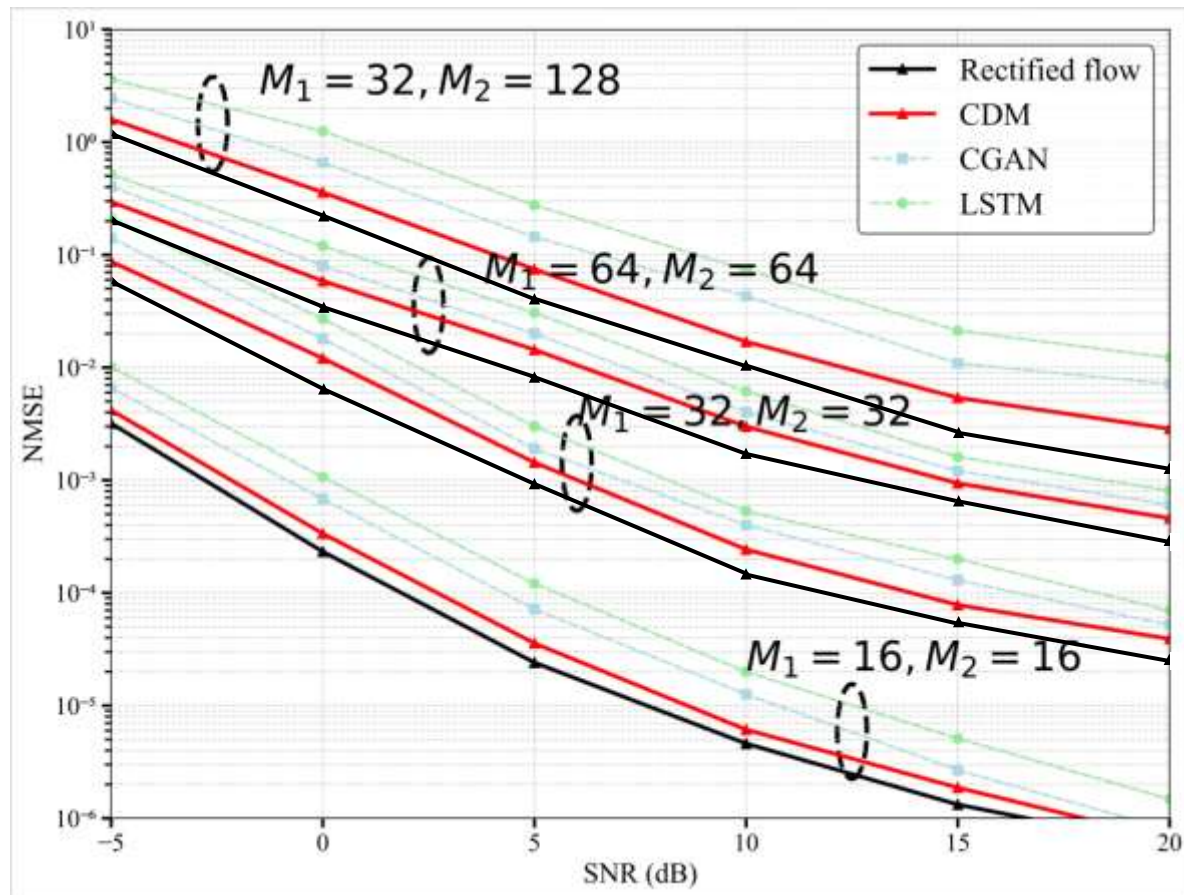
- RIS₁的阵元数 M_1 分别设置为16、32、64
- RIS₂的阵元数 M_2 分别设置为16、32、64、128
- 基站天线为 $a = 4$
- 空间相关性矩阵参数定义: 阵元间距 $d = 0.5$ 波长
- RIS之间的距离: 10m

对比算法:

- (1) Rectified flow: 流匹配模型+混合噪声
生成步数 $T = 200$
混合噪声 λ 设置为0.8
 - (2) CDM: 条件扩散模型
生成步数 $T = 200$
 α_t 范围是[0.98, 0.999]
 - (3) CGAN: 条件对抗生成式网络
 - (4) LSTM: 长短期记忆网络
- ρ : 掩码比例设置为0.2、0.5
 - 训练样本数为50000
 - 模型训练轮数300轮
 - 学习率设置为0.0001
 - 权重衰减0.00001
 - 批次大小128

评估指标:
$$\text{NMSE} = \mathbf{E} \left[\frac{\|\widehat{\mathbf{H}} - \mathbf{H}\|_2}{\|\mathbf{H}\|_2} \right]$$

仿真结果



相较于传统深度学习算法，如GAN和DDPM，流匹配模型以更低的复杂度达到最优性能

- ◆ 1. AI赋能RIS的研究背景
- ◆ 2. 基于生成式AI的高效RIS信道获取
- ◆ 3. 基于强化学习的RIS信道动态感知
- ◆ 4. 基于预训练决策模型的RIS资源分配

RIS信道估计频率与系统性能之间的权衡

核心问题：如何决定何时进行信道估计？

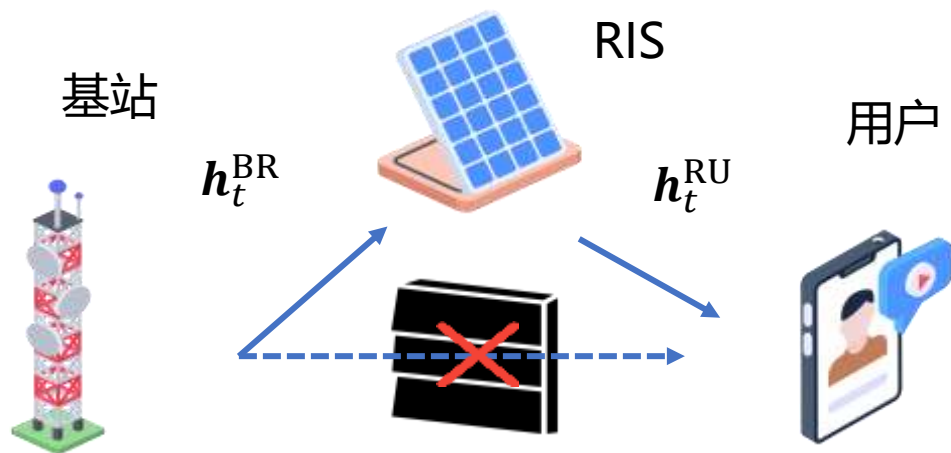
- 想要在RIS系统中实现良好波束成形效果的前提是获取准确的信道状态信息，这需要进行频繁的信道估计，而**信道估计具有时间成本**，会压缩有效通信时间并影响通信性能



- 如果并非每个时刻都进行信道估计则无法获得实时信道状态信息，**基于过时信息将影响RIS波束赋形设计及系统性能**

因此，我们需要设计一套**自适应信道估计间隔决策机制**，根据实时状态智能决定是否估计信道，从而在估计开销与通信性能之间实现最优权衡

基于 RIS 辅助 SISO 通信系统



- 系统由基站、智能反射面和用户设备组成
- 假设**BS与UE之间直达链路被阻挡**（最坏情况分析），所有通信信号通过IRS反射路径
- BS和UE均为单天线；RIS由 N 个反射单元组成，每个单元有幅度系数和相位系数，幅度固定为1，相位可调节

基本元素：

- 基站到RIS信道： $\mathbf{h}_t^{\text{BR}}$
- RIS到用户信道： $\mathbf{h}_t^{\text{RU}}$
- RIS波束赋形矩阵： $\mathbf{\Phi} = \text{diag}(\boldsymbol{\phi})$
- 等效信道： $\mathbf{h}_t = \text{diag}[(\mathbf{h}_t^{\text{RU}})^{\text{H}}] \mathbf{h}_t^{\text{BR}}$
- 高斯马尔可夫信道： $\mathbf{u}_t = \rho \mathbf{u}_{t-1} + \sqrt{1 - \rho^2} \mathbf{n}_t$
- 用户可达速率： $R_t = \log_2(1 + P|\boldsymbol{\phi}_t^T \mathbf{h}_t|^2 / \sigma_z^2)$

优化目标：在有限的时间资源下，通过智能决定何时进行信道估计并设计RIS波束，使系统在估计开销与通信性能之间取得平衡，从而**最大化长期平均吞吐量**

优化问题定义

定义信道估计指示变量 $\alpha_t \in \{0,1\}$

- $\alpha_t = 1$: 在时隙 t 执行信道估计, 可获得最新CSI, 但占用 $\delta\tau$ 的传输时间
- $\alpha_t = 0$: 在时隙 t 不执行信道估计, 无法获取最新CSI, 但占用 τ 的传输时间

$$\begin{aligned} \mathbf{P}: \quad & \max_{\{\alpha_t, \phi_t, \forall t\}} \mathbb{E} \left[\sum_{t=1}^T [\alpha_t(1-\delta)R_t + (1-\alpha_t)R_t] \tau \right], \\ \text{s.t.} \quad & 0 \leq \phi_{t,n} < 2\pi, \forall n \in \mathcal{N}, \\ & \alpha_t \in \{0,1\}. \end{aligned}$$

估计频率与系统性能的权衡难题

优化问题具有以下特点:

- **信道估计与RIS波束赋形决策高度耦合:** 两者相互影响、相互依赖
- **问题呈顺序优化特性:** 不同阶段的决策变量互为输入与约束
- **决策依赖不完美且滞后的CSI:** 基于历史或误差CSI进行策略选择



考虑以上特性, 我们将问题建模成
双层部分可观测马尔可夫决策过程

双层部分可观测马尔可夫决策过程

在该框架下，我们将每个时隙拆分为两个连续决策阶段：

- **第一阶段：**基于上一时隙的通信速率和CSI可用性，判断当前时隙是否需要进行**信道估计**
- **第二阶段：**根据第一阶段决策结果获得的CSI以及RIS系统参数，决策当前时隙的最优**波束赋形**

奖励被定义为当前时隙的用户可达速率，反映了信道估计开销与波束赋形性能之间的权衡

- **状态：**真实的基站-RIS-用户设备的级联信道信息

- **第一阶段观测：**前一时隙的可达数据速率

$$\mathbf{o}_{t,1} \triangleq [a_{t-1}(1 - \delta)R_{t-1} + (1 - a_{t-1})R_{t-1}]\tau$$

- **第一阶段动作：**当前时隙是否进行信道估计

- **第二阶段观测：**第一阶段动作执行完后更新的信息

$$\mathbf{o}_{t,2} \triangleq \begin{cases} \{\tilde{\mathbf{h}}_t, 0\} & \text{if } a_t = 1 \\ \{\tilde{\mathbf{h}}_{t-k}, k\} & \text{if } a_t = 0 \end{cases}$$

- **第二阶段动作：**当前时隙的 RIS 波束赋形向量

- **奖励：**当前时隙的用户可达速率

$$r_t \triangleq [a_t(1 - \delta)R_t + (1 - a_t)R_t]\tau$$

面临的挑战

- 决策过程中，仅依赖当前时隙的观测是不够的，因为它是对真实状态的非完美认知，因此我们需要整合历史的观测序列

● 通信质量反馈: $\mathbf{o}_{\leq t,1} \triangleq \{\mathbf{o}_{1,1}, \dots, \mathbf{o}_{t,1}\}$

● 信道状态反馈: $\mathbf{o}_{\leq t,2} \triangleq \{\mathbf{o}_{1,2}, \dots, \mathbf{o}_{t,2}\}$

- 双层部分可观测马尔可夫决策过程的目标是**找到最优信道估计和波束赋形策略**

● 信道估计策略: $\pi(a_{t,1} | \mathbf{o}_{\leq t,1}, \mathbf{o}_{\leq t-1,2})$

● 波束赋形策略: $\pi(a_{t,2} | \mathbf{o}_{\leq t,1}, \mathbf{o}_{\leq t,2})$

● 状态价值函数: $V^\pi(\mathbf{s}_t) = \mathbb{E}_{a_{t,1}, a_{t,2}, \dots, a_{T,1}, a_{T,2} \sim \pi} \left[\sum_{k=t}^T \gamma^{k-t} r_k | \mathbf{s}_t \right] \longrightarrow \pi^* = \arg \max_{\pi} V^\pi(\mathbf{o}_{\leq t,1}, \mathbf{o}_{\leq t,2})$



- **挑战一**: 缺乏关于信道的先验知识，通过**基于模型的方法**来求解价值函数相对困难
- **挑战二**: 经典强化学习算法处理较长的观测历史序列面临**维度灾难问题**

基于循环神经网络的信道估计与波束赋形算法

方案一：基于**循环神经网络**双层深度强化学习

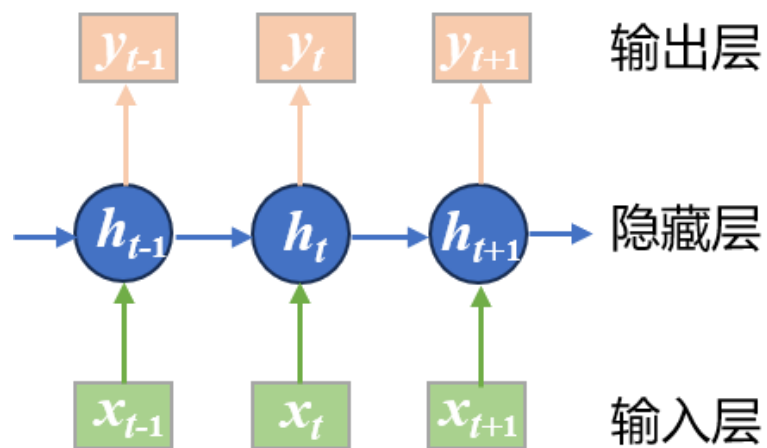
循环神经网络 (RNN)

- 通过递归连接**共享参数**
- 当前时间步利用前一时间步的**隐状态**来捕捉时间序列中的**依赖关系**

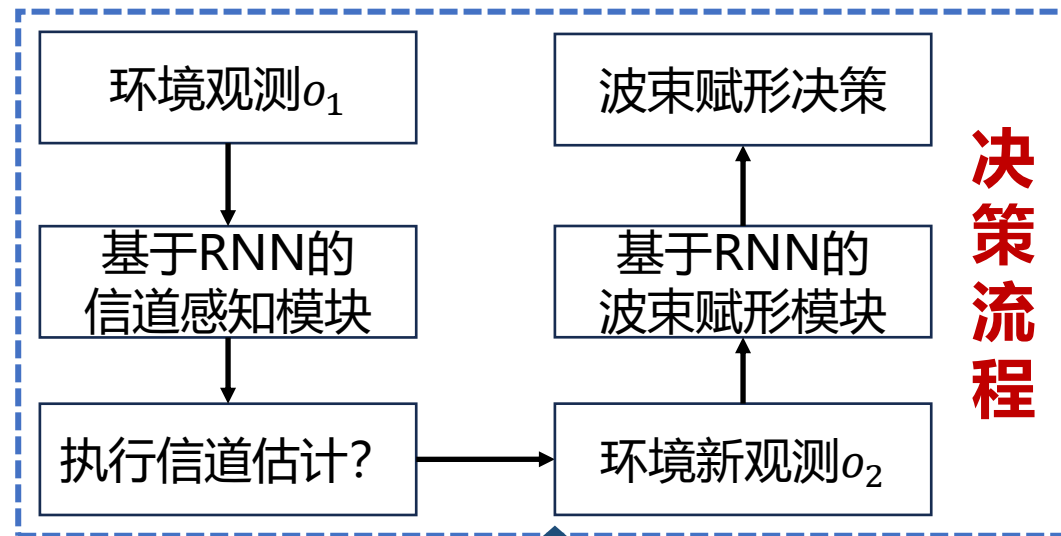
提供策略网络做决策

RNN 提取到的系统隐含状态

观测到的历史数据、CSI可用性等



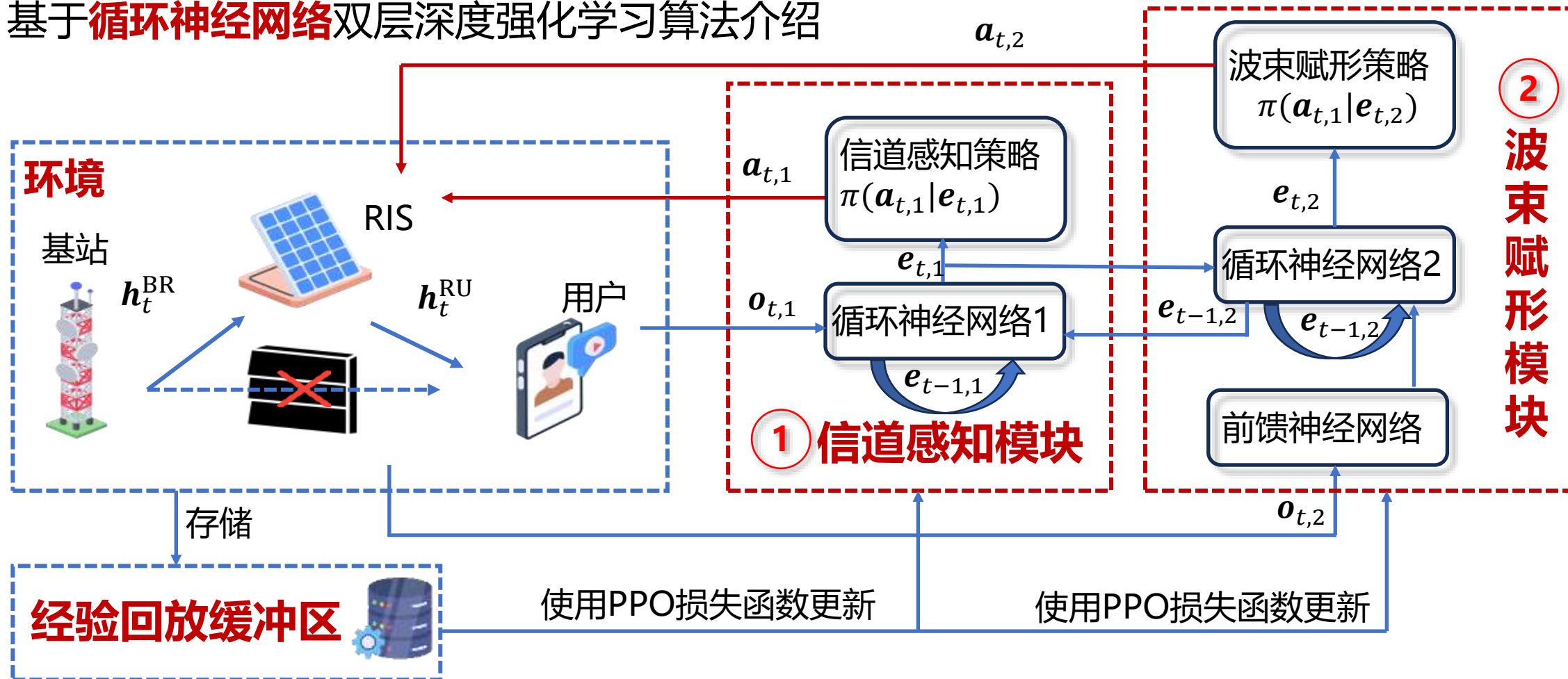
循环神经网络 (RNN)



RNN用于编码部分可观测的历史信息
(如历史速率、过往信道估计决策)
生成隐藏状态表示, 辅助第一阶段
(信道估计模块) 与第二阶段 (波束
赋形模块) 的策略选择

算法流程

基于循环神经网络双层深度强化学习算法介绍



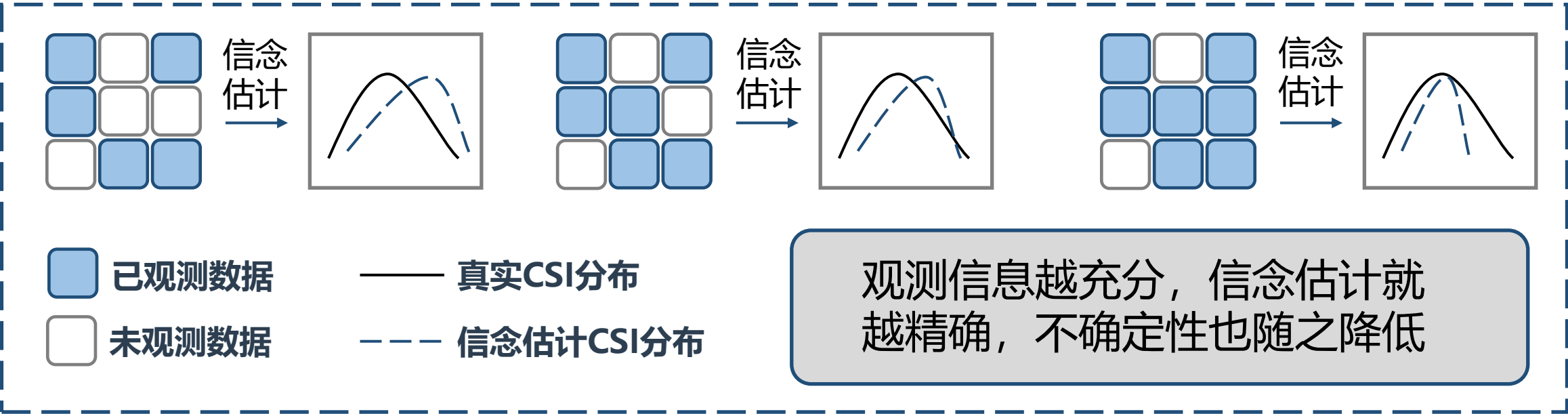
缺点: RNN为黑箱方法, 对隐藏状态的可解释有限

信念概念引入

基于循环神经网络的算法**未进行显示建模**，并且**缺乏可解释性**



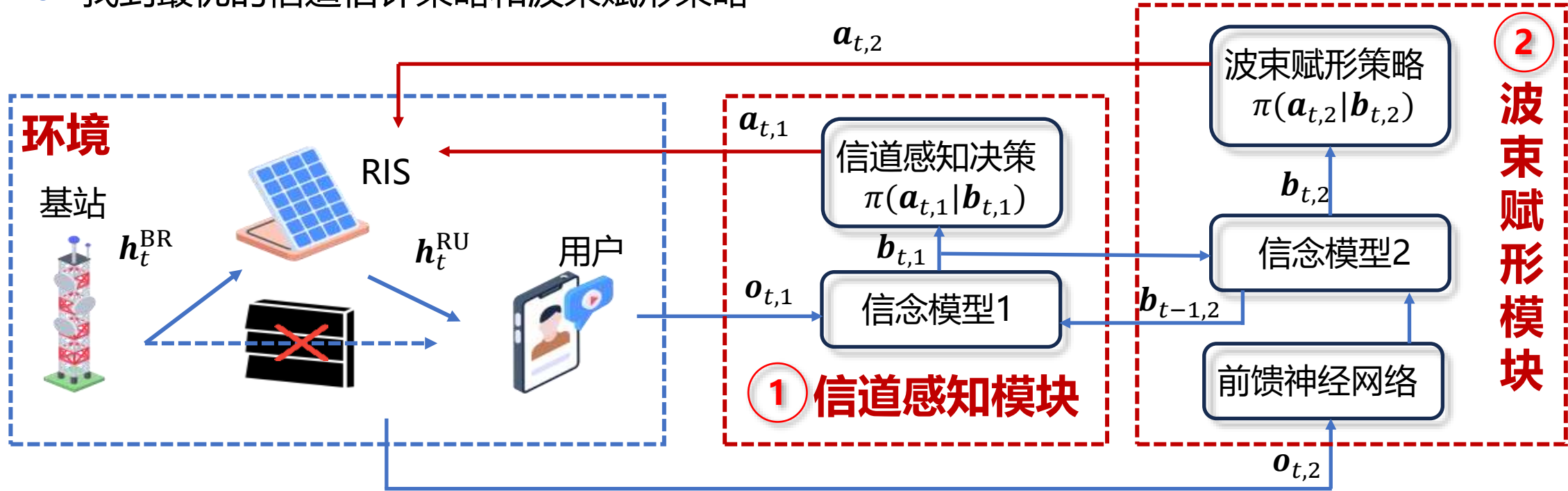
因我们用一个具有明确物理意义的、**可解释**的变量来**显式地建模**和推断不确定的信道状态，而不是依赖一个抽象的隐藏状态。这个变量就是“**信念**”，它被定义为在给定历史观测的前提下，关于 CSI 的概率分布：



基于信念推断的信道估计与波束赋形算法

方案二：基于信念推断的双层深度强化学习

- 引入信念：在给定历史观测的前提条件下，关于 CSI 的概率分布： $b_{t,1}(s_t) \triangleq \Pr(s_t|o_{\leq t,1}, \mathbf{o}_{< t,2})$
- 借助信念状态对 CSI 进行**显式建模**，并基于信念进行决策
- 找到最优的信道估计策略和波束赋形策略



变分下界

问题：直接求真实CSI后验分布很难，由贝叶斯公式：

$$P(\mathbf{s}_t | o_{\leq t,1}, \mathbf{o}_{\leq t,2}) = \frac{P(\mathbf{o}_{t,2} | \mathbf{s}_t) P(\mathbf{s}_t | o_{\leq t,1}, \mathbf{o}_{< t,2})}{\int P(\mathbf{o}_{t,2} | \mathbf{s}_t) P(\mathbf{s}_t | o_{\leq t,1}, \mathbf{o}_{< t,2}) d\mathbf{s}_t} \rightarrow \text{分母是一个高维积分，无法计算}$$

因此，我们借助**变分推断 (Variational Inference)** 的方法，通过学习一个更简单的、可以用神经网络参数化的分布来逼近真实的分布，在变分推断的框架下，我们可以得到观测的对数似然 $\log P(o_{\leq t,1}, \mathbf{o}_{< t,2})$ 的证据下界 (ELBO)：

$$\begin{aligned} \log P(o_{\leq t,1}, \mathbf{o}_{< t,2}) \\ \geq \mathbb{E}_{\mathbf{b}_{\leq t,2} \sim q(\cdot)} [\log g(o_{\leq t,1}, \mathbf{o}_{< t,2} | \mathbf{b}_{\leq t,2})] &\longrightarrow \text{重建项} \\ - D_{KL}[q(\mathbf{b}_{\leq t,2} | o_{\leq t,1}, \mathbf{o}_{< t,2}) || p(\mathbf{b}_{\leq t,2})] &\longrightarrow \text{正则项} \end{aligned}$$

启发

生成模型 g_ϕ
信念推断模型 q_ξ
信念转移模型 p_ψ

信念更新目标函数

在变分推断的框架下，我们推导出

- **第一阶段**信念更新的目标函数

$$\mathbb{E}_{b_{1:t,2} \sim q_\xi} \left\{ \sum_{k=1}^t \log g_\varphi(o_{k,1}, \mathbf{o}_{k,2} | b_{k,2}) \right\} - \sum_{k=1}^t D_{\text{KL}}[q_\xi(b_{k,2} | b_{k,1}, \mathbf{o}_{k,2}) \parallel p_\psi(b_{k,2} | b_{k,1})]$$

- **第二阶段**信念更新的目标函数

$$\mathbb{E}_{b_{1:t,1} \sim q_\xi} \left\{ \sum_{k=1}^t \log g_\varphi(\mathbf{o}_{k-1,2}, o_{k,1} | b_{k,1}) \right\} - \sum_{k=1}^t D_{\text{KL}}[q_\xi(b_{k,1} | b_{k-1,2}, o_{k,1}) \parallel p_\psi(b_{k,1} | b_{k-1,2})]$$

生成模型 g_φ

根据信念 b 重建观测 o

信念推断模型 q_ξ

根据观测 o 推测信念 b

信念转移模型 p_ψ

预测下一信念

策略/价值网络目标函数

算法策略网络和价值网络的更新在PPO算法的框架下：

- **第一阶段** 信道估计策略更新的目标函数

$$J(\omega^c) = -\mathbb{E}_{\mathbf{b}_{t,1}, \mathbf{b}_{t,2}, \mathbf{a}_{t,1}} \{ \min[\varrho_t(\omega^c) \tilde{A}_t(\mathbf{b}_{t,2}), \text{clip}(\varrho_t(\omega^c), 1 - \epsilon, 1 + \epsilon) \tilde{A}_t(\mathbf{b}_{t,2})] \}$$

- **第二阶段** 波束赋形策略的目标函数

其中 ϱ_t 表示新老策略比例： $\varrho_t(\omega^b) = \frac{\pi(\mathbf{a}_{t,2}|\mathbf{a}_{t,2}, \omega^b)}{\pi(\mathbf{a}_{t,2}|\mathbf{a}_{t,2}, \hat{\omega}^b)}$

$$J(\omega^b) = -\mathbb{E}_{\mathbf{b}_{t,2}, \mathbf{a}_{t,2}} \{ \min[\varrho_t(\omega^b) \tilde{A}_t(\mathbf{b}_{t,2}), \text{clip}(\varrho_t(\omega^b), 1 - \epsilon, 1 + \epsilon) \tilde{A}_t(\mathbf{b}_{t,2})] \}$$

其中 $\tilde{A}_t(\mathbf{b}_{t,2})$ 为优势函数：

$$\tilde{A}_t(\mathbf{b}_{t,2}) = \sum_{k=1}^K \gamma^{k-1} r_{t+k} + \gamma^{K-1} V(\mathbf{b}_{t+K} | \omega^v) - V(\mathbf{b}_{t,2} | \omega^v)$$

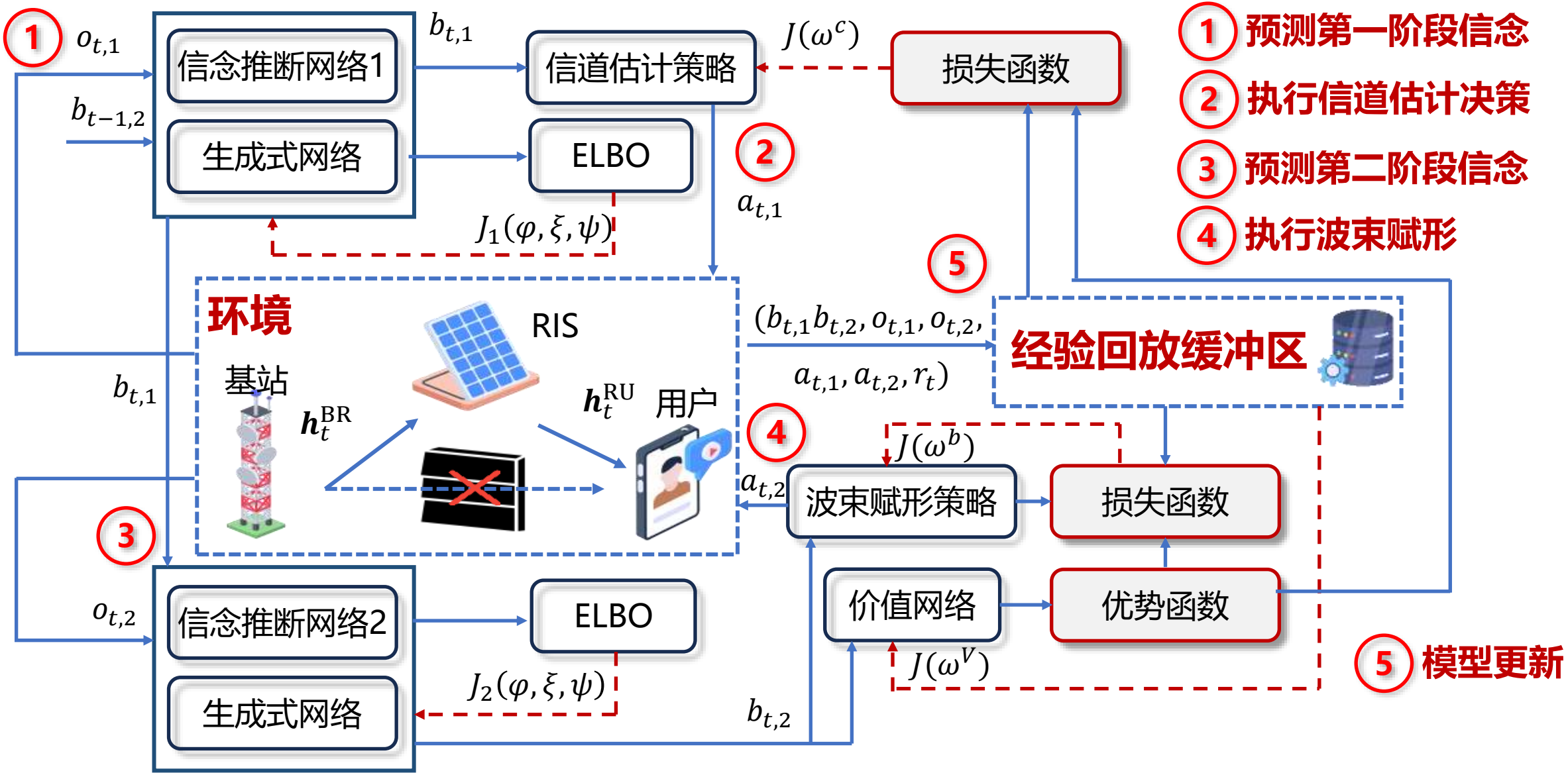
- **价值网络** 更新的目标函数：

$$J(\omega^v) = -\mathbb{E}_{\mathbf{b}_{t,2}} \left[\left(\sum_{k=1}^K \gamma^{k-1} r_{t+k} - V(\mathbf{b}_{t,2} | \omega^v) \right)^2 \right]$$

- 总目标函数：

$$J(\Theta) \triangleq J(\omega^a) + \lambda^c J(\omega^c) + \lambda^v J(\omega^v) + \lambda^e [J_1(\varphi, \xi, \psi) + J_2(\varphi, \xi, \psi)]$$

基于信念推断的信道估计与波束赋形算法算法流程



仿真结果

- 周期性近端策略优化算法 (Periodic-PPO)
- 基于 RNN 的 PPO 算法 (RNN-based PPO)

- 周期性梯度下降算法 (Periodic-GD)

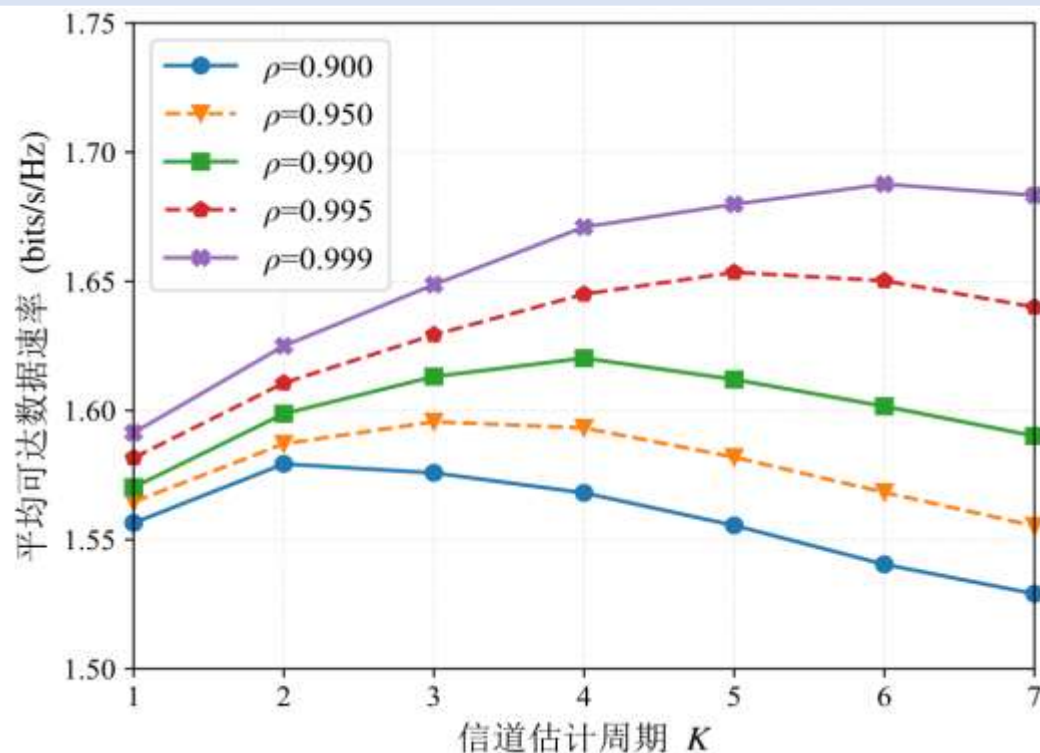


图 1 不同信道相关系数下平均可达数据速率与周期性信道估计间隔的关系图

周期性信道估计的方法存在一个**最优**信道估计间隔

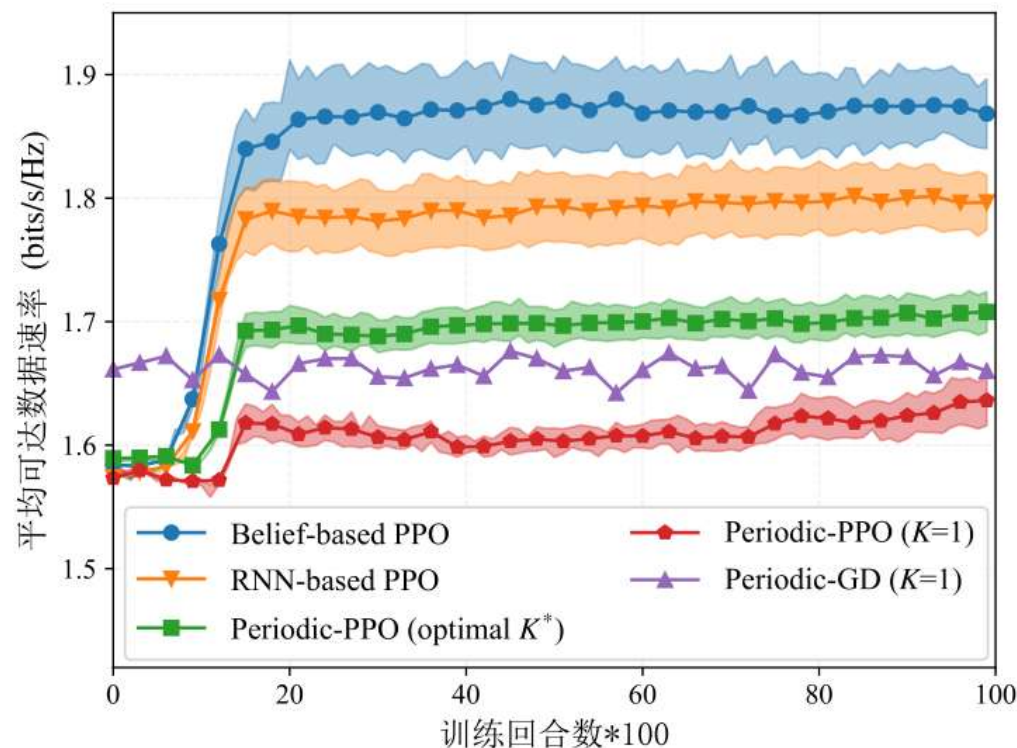


图 2 不同算法的平均可达数据速率比较图

所提算法的性能超过周期式信道估计算法

仿真结果

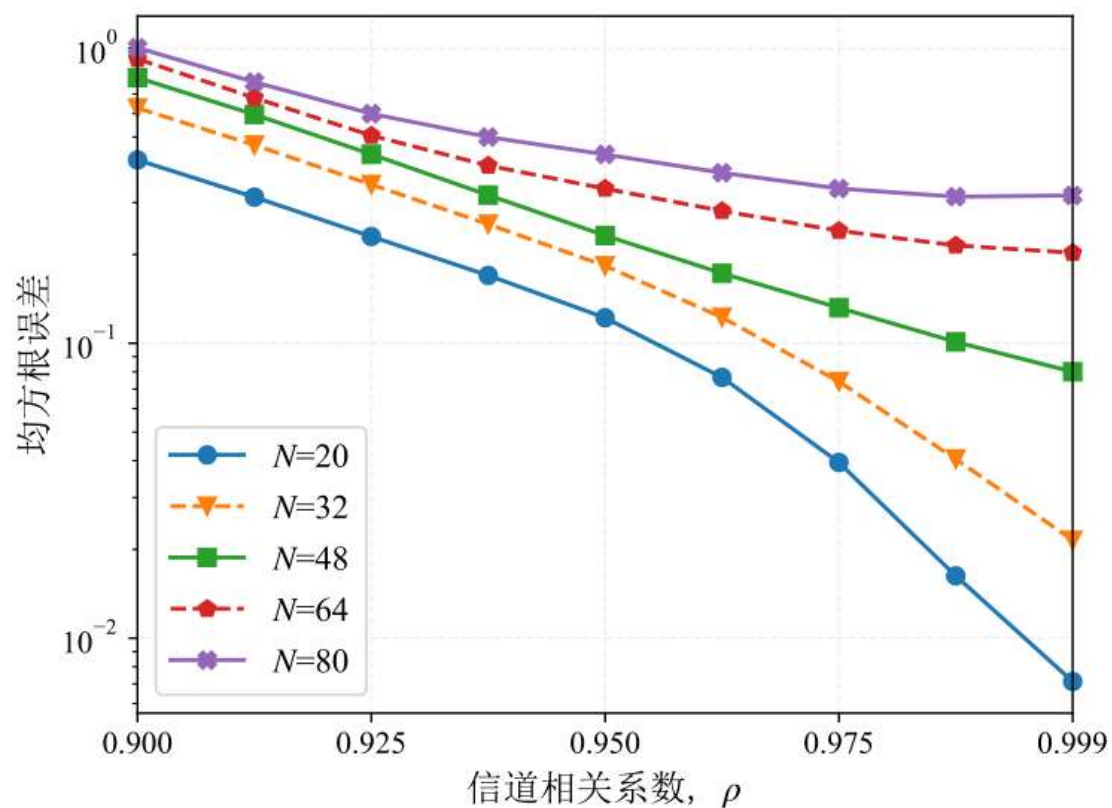


图 3 所提算法 NMSE 性能与信道相关系数关系图

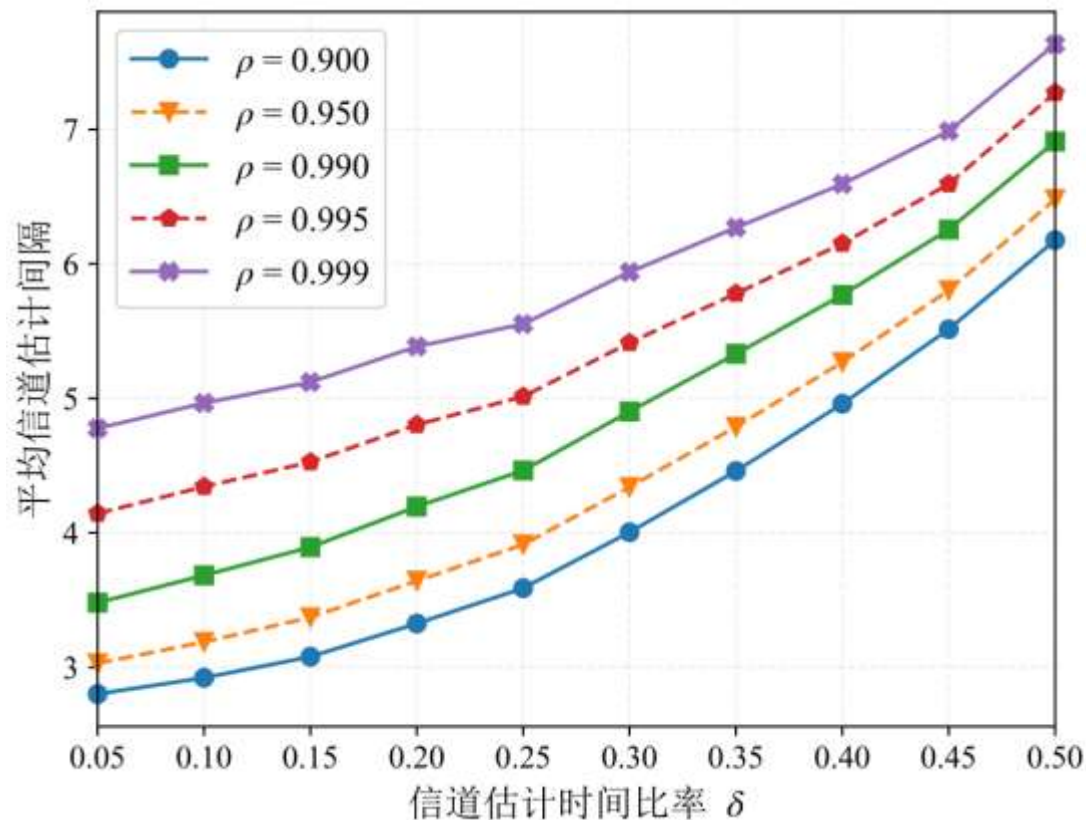


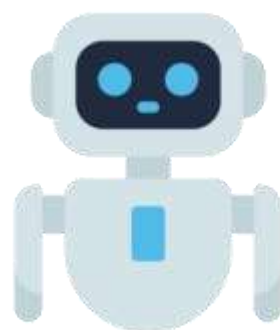
图 4 平均信道估计间隔与信道估计时间比率的关系图

- 所提自适应信道估计算法能较好地对CSI进行建模
- 并非每个时隙都需要进行信道估计，**合理选择信道估计的间隔**对于提升系统性能至关重要

- ◆ 1. AI赋能RIS的研究背景
- ◆ 2. 基于生成式AI的高效RIS信道获取
- ◆ 3. 基于强化学习的RIS信道动态感知
- ◆ 4. 基于预训练决策模型的RIS资源分配

强化学习缺点

智能体与环境交互，通过观察**环境状态**、采取**动作**并接收**奖励**，以**试错**的方式优化策略，即在给定状态下应采取的最优行动序列



智能体



状态 奖励



环境



长期依赖性

在处理需要**长期规划**或**稀疏奖励**的任务时，传统强化学习模型难以捕捉利用长期的历史信息



训练稳定性

传统强化学习算法依赖于精心设计的网络超参数，容易受到**初始条件**和**环境变化**的影响



策略泛化性

传统强化学习针对特定环境进行优化，对于未见过的**新环境**，泛化能力较弱，已有策略难以迁移需要**重新训练**

Transformer原理

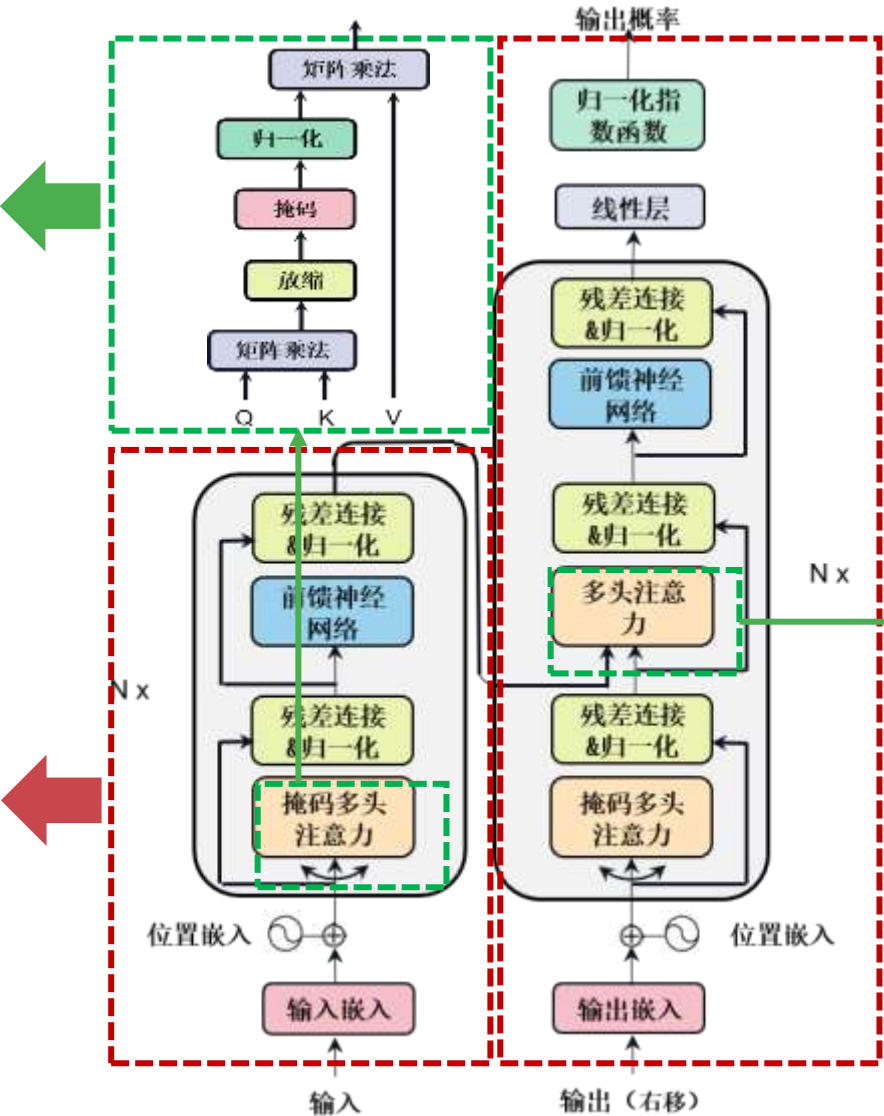
注意力机制

- 评估序列中每个元素对当前任务的贡献大小从而捕捉**全局依赖**
- 对于输入数据进行嵌入特征提取，计算矩阵 **Q(查询)**，**K(键值)**，**V(值)** 计算注意力权重

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

编码器

利用**位置编码**、**注意力机制**和**前馈神经网络**确保全局信息的有效传递以增强模型的表达能力



多头注意力机制

每个头部关注输入序列的不同部分，**并行处理**多种关系，提升表示能力

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

解码器

通过**注意力机制**结合编码器输出的信息，确保生成的输出前后相关，提升输出质量

Transformer优点

长序列处理能力

- **捕捉长程依赖**

直接在序列中的任意位置建立全局依赖，避免了长距离依赖信息衰减的问题

- **可扩展性**

利用注意力机制可以处理较长的任务序列，而不依赖固定的输入长度

- **并行化计算**

并行化所有输入的计算，不受序列位置的限制大大加快了训练速度

预训练能力

- **迁移泛化能力**

在大规模数据集获取的预训练模型可以通过**微调**快速适应新的下游任务

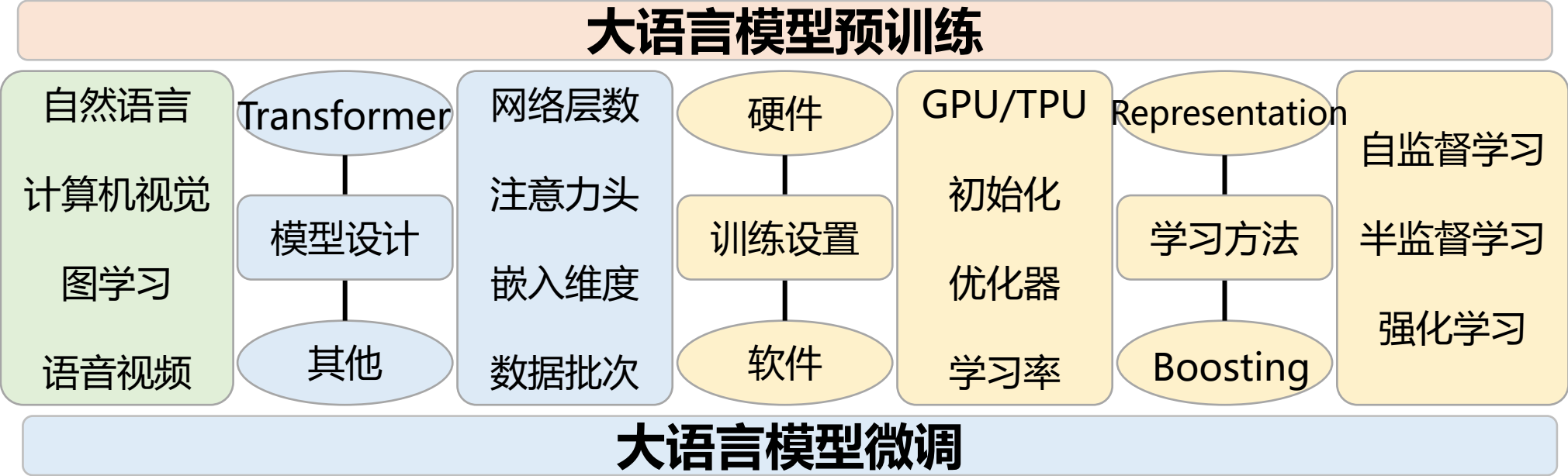
- **支持多任务学习**

利用一个模型处理**多个任务**（如图像识别，文本分类，序列决策等）

- **提高训练效率**

利用预训练的知识，能够更好地初始化**新任务模型**，避免随机参数训练

基于Transformer的大语言模型



ChatGPT

OpenAI开发，具备强大的文本生成和理解能力，支持复杂分析任务



Baichuan

百度开发，具备广泛的中文数据集训练背景，支持多语言处理



ChatGLM

清华开发，面向中文场景的对话模型，具备对中文上下文的深度理解能力



Qwen

阿里巴巴开发，结合多任务学习，具备生成和理解复杂语言任务的能力



LLaMA

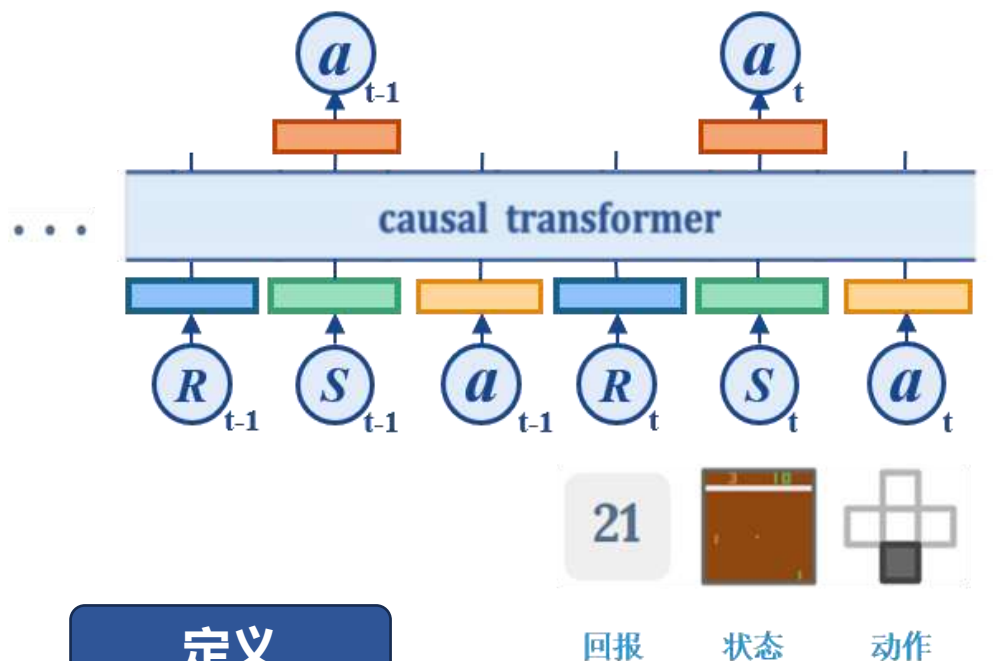
Meta开发，轻量化设计，能够在较小的计算资源下提供较高的性能



Moss

复旦开发，开源的中文对话模型，具备对话、问答、多任务处理能力

Decision Transformer预训练决策模型



定义

Decision Transformer是将Transformer架构应用于强化学习中的一种模型，将强化学习问题转化为**序列预测**问题，利用**历史轨迹**生成最优策略

工作原理

- **序列表示**: 构建状态、动作和未来累计奖励的序列
- **因果掩码**: 预测每一步动作时只使用历史的序列信息
- **监督学习**: 预测的动作尽可能接近真实动作
- **最佳预测**: 给定历史轨迹和目标奖励生成最佳动作序列

核心优势



稳定能力优



优化效率高



泛化能力强

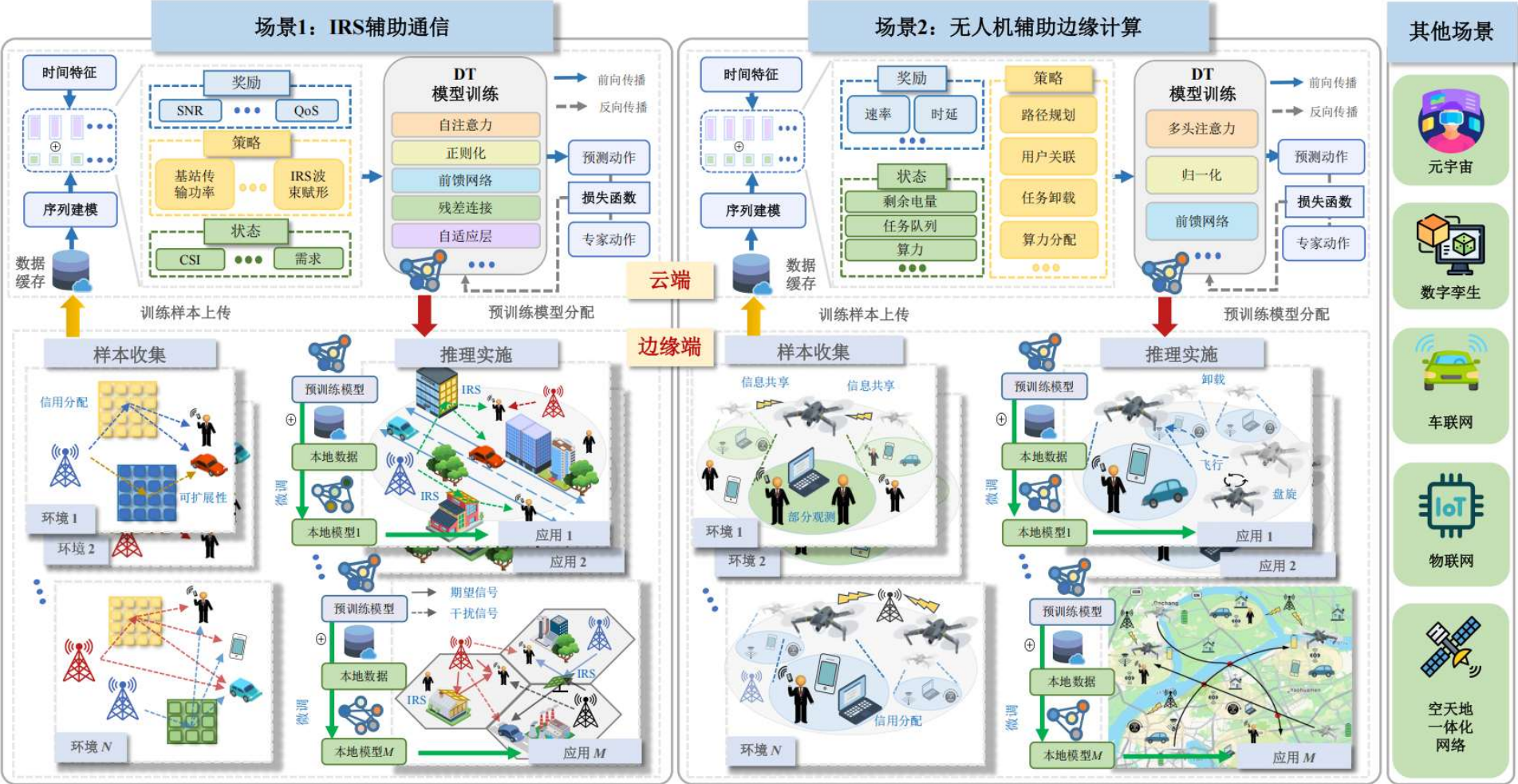
Decision Transformer总体应用框架

场景挑战

动态网络拓扑
可变服务需求
不可预测干扰
.....

方法限制

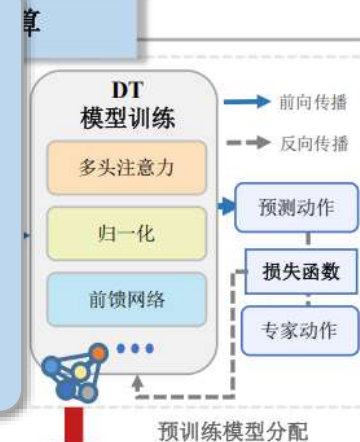
训练效率低
泛化能力弱
资源消耗大
.....



Decision Transformer总体应用框架

数据收集

- 不同的通信任务，存在不同的资源管理决策问题
- 对于特定的通信任务，收集不同场景下的状态、动作和奖励的数据样本
- 将不同通信任务所收集到的训练样本上传到云端



其他场景



元宇宙



数字孪生



车联网



物联网



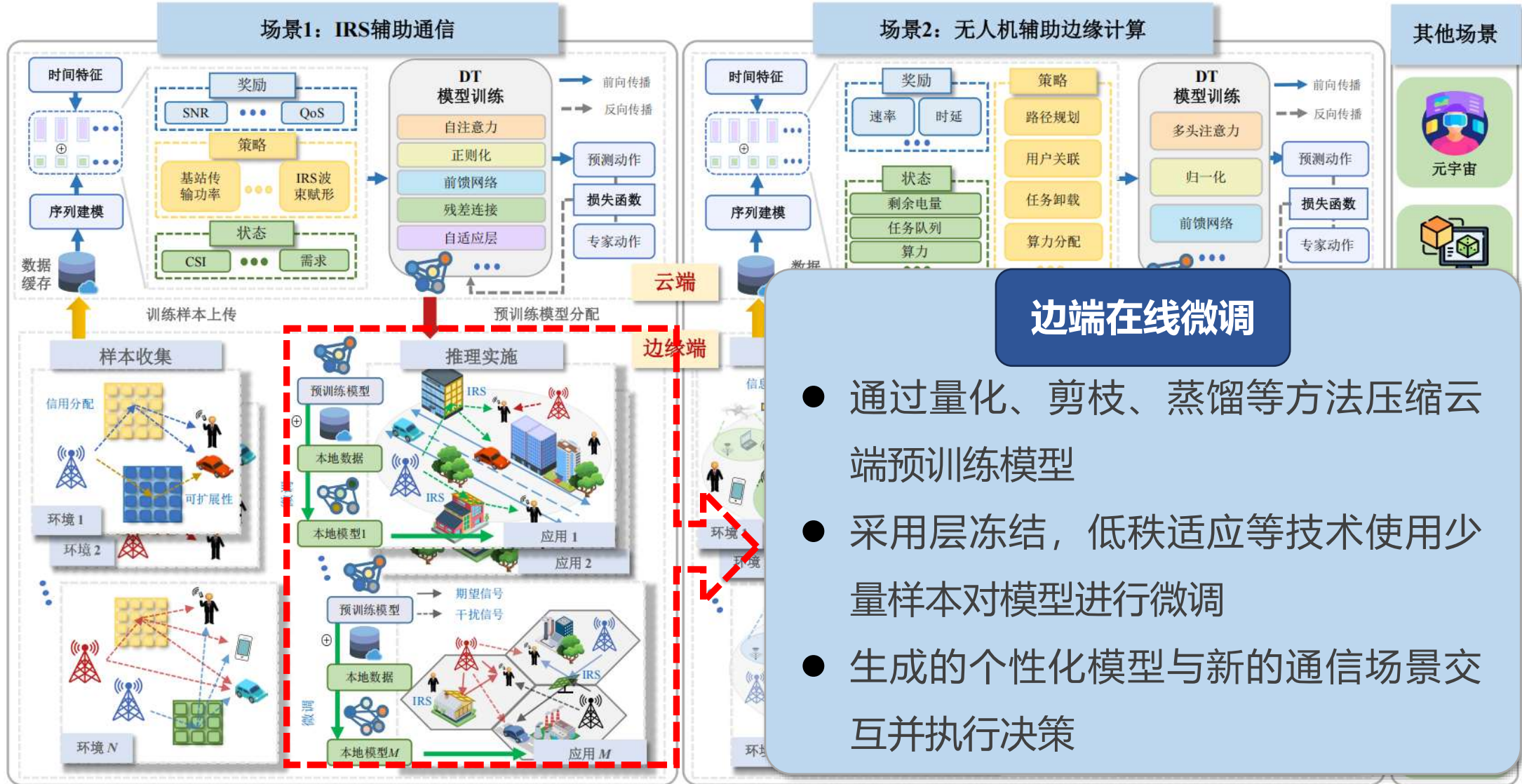
空天地
一体化
网络



Decision Transformer总体应用框架



Decision Transformer总体应用框架

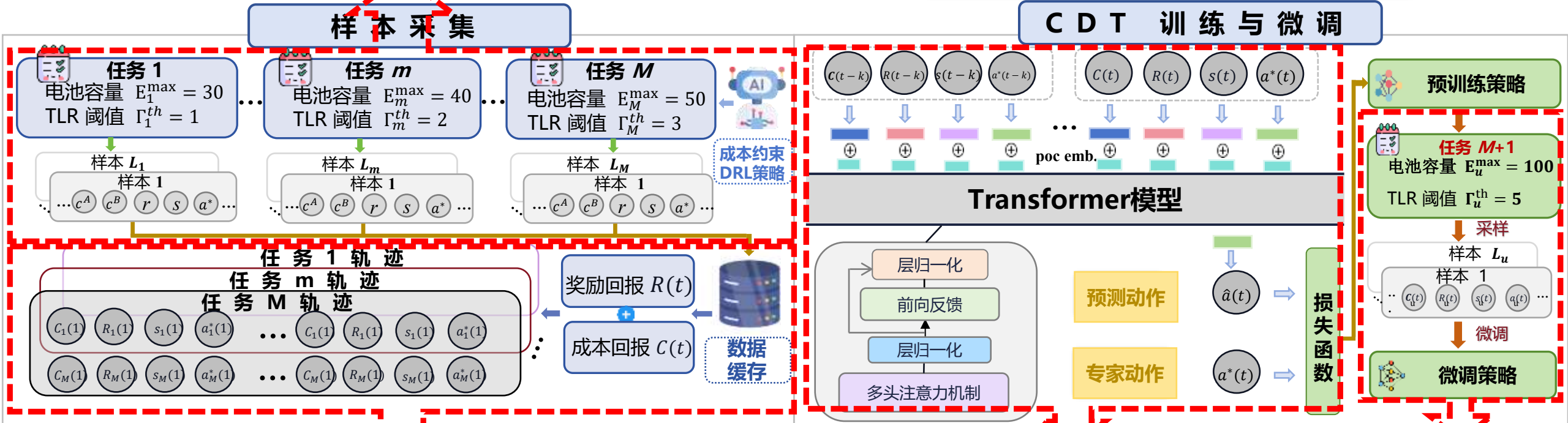


基于条件约束的Decision Transformer框架

多任务数据收集：用PPO-Lag在M个不同约束阈值下收集专家轨迹

核心改进与优势

- 约束感知能力
- 跨约束泛化



数据库构建：所有样本存入数据库
轨迹重构处理：将样本重构为包含奖励回报和成本回报的序列

CDT策略训练：训练Transformer网络，最小化动作预测误差
奖励估计网络训练：训练网络 ϕ 学习基于约束预测奖励回报

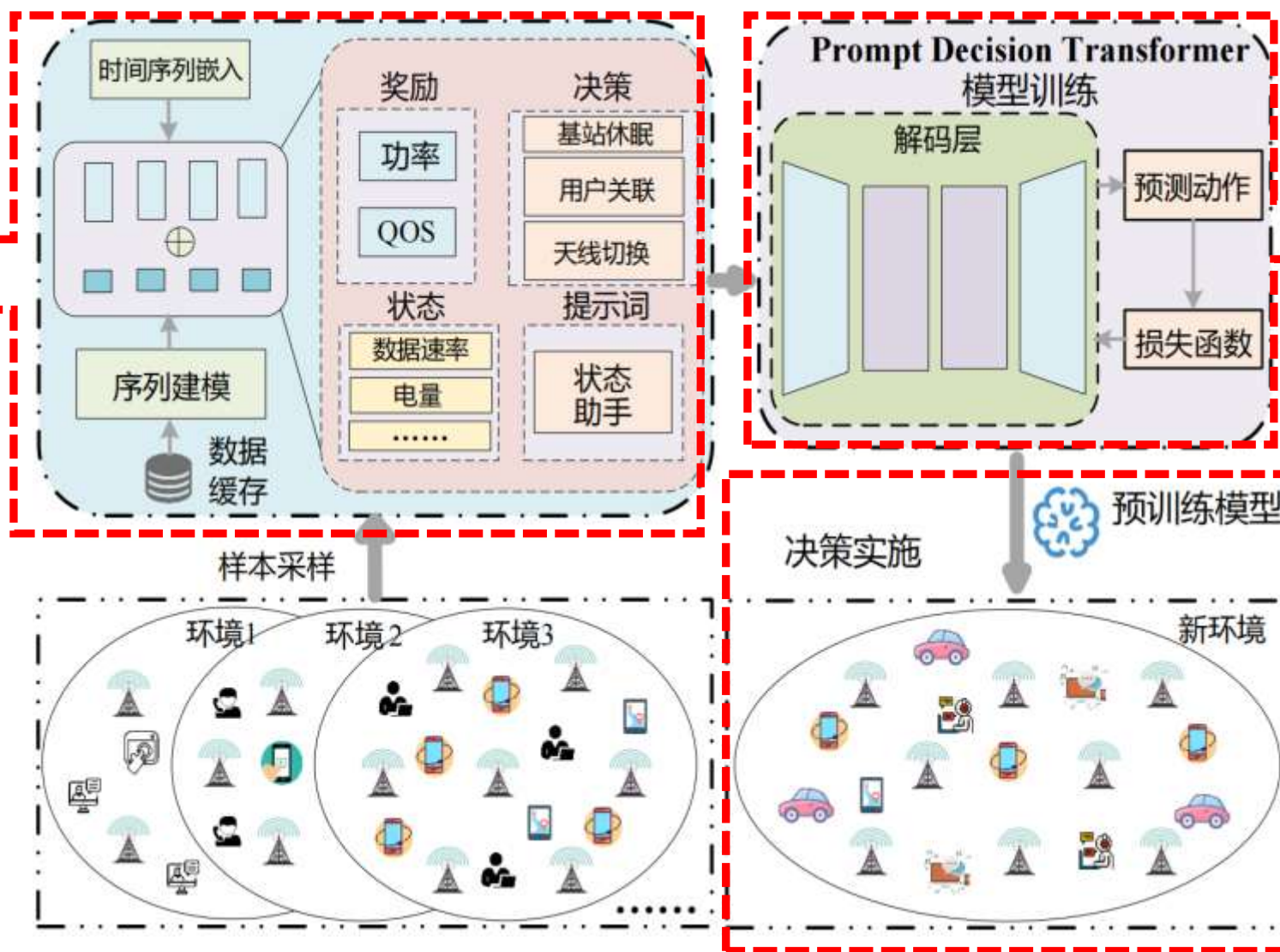
微调流程

- 约束初始化
- 策略交互执行
- 少样本微调

基于提示增强的Decision Transformer框架

数据收集：从多个通信环境中收集基站运行的历史经验数据

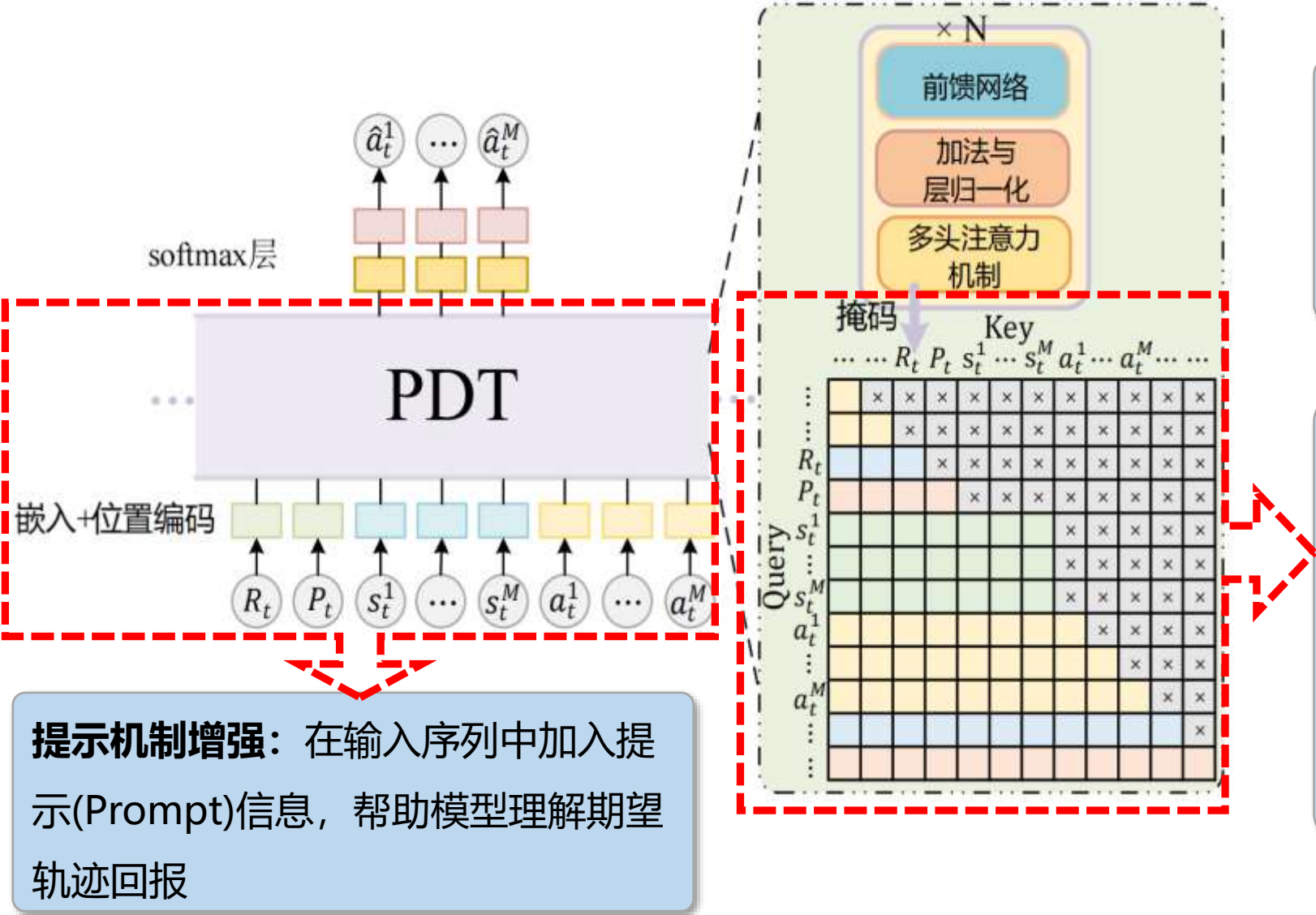
数据预处理：将收集的数据存储到缓冲区并通过顺序建模组织成轨迹序列格式



模型训练：将预处理后的数据集输入PDT进行预训练，通过解码器层学习最优的基站节能策略

模型部署：训练完成的模型可直接部署到新环境中进行决策，无需重新训练

基于提示增强的Decision Transformer框架



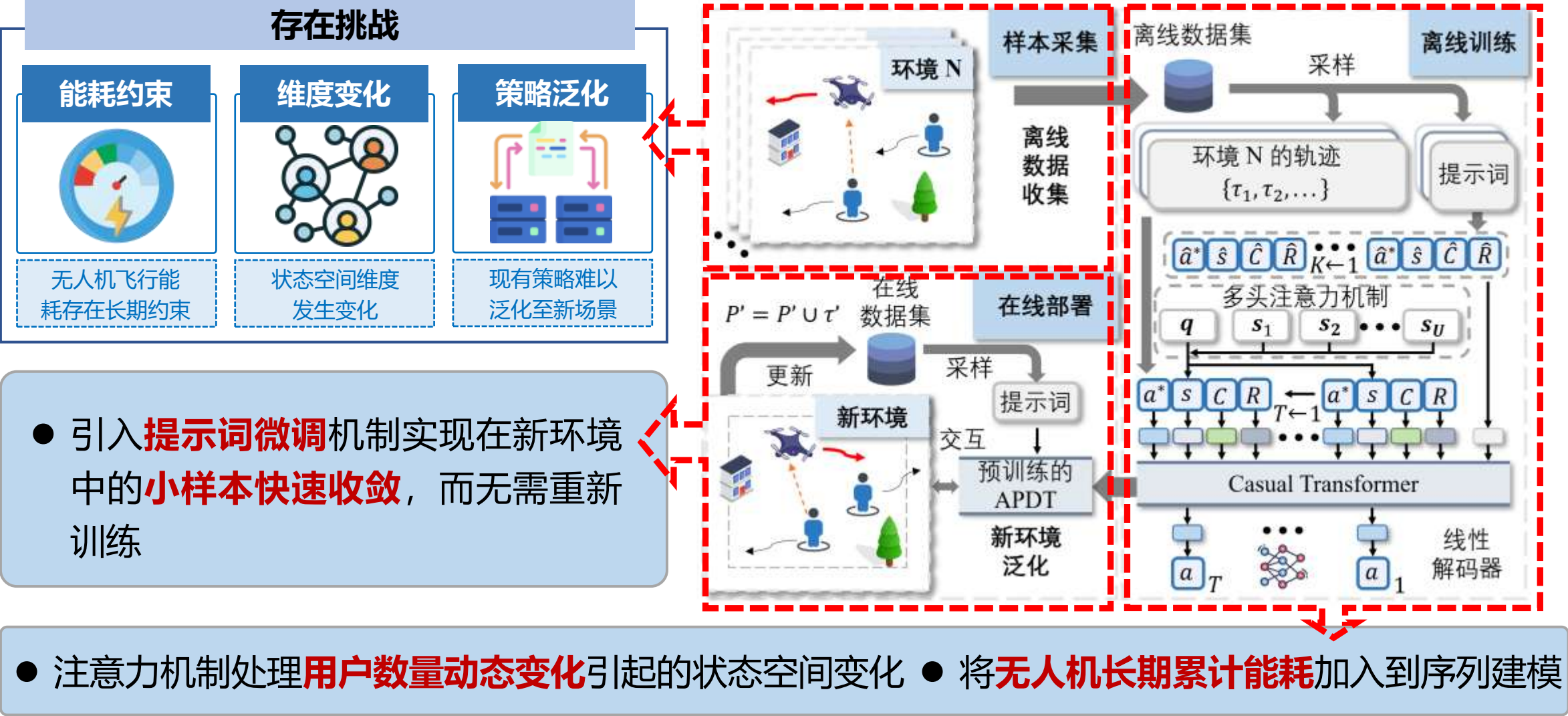
核心优势：

克服DT对数据集质量的过度依赖，实现多基站场景的灵活适应

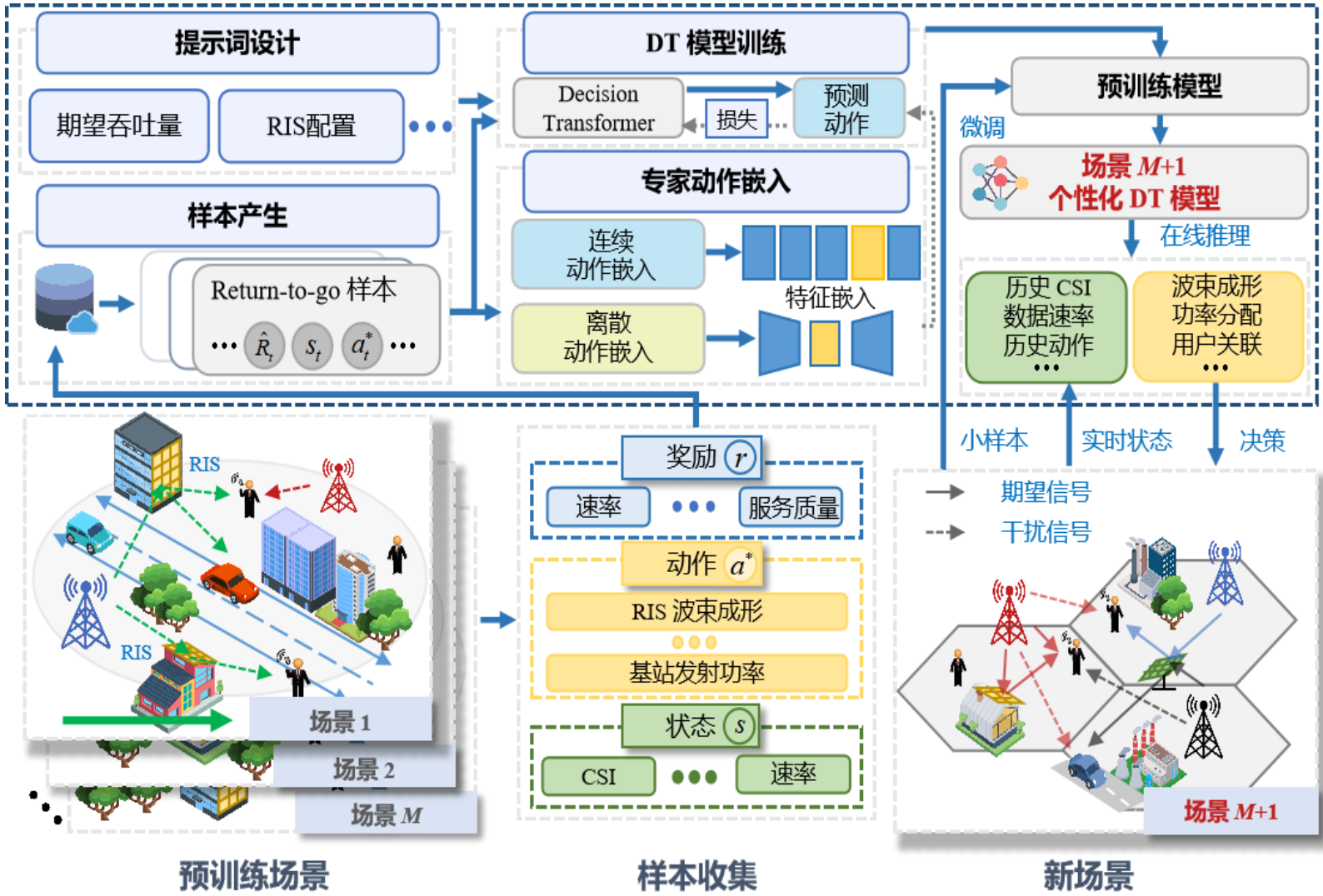
状态动作解耦： 每个基站的状态和动作独立处理，适应不同数量基站的任务

注意力掩码优化： 允许各基站观察同时间步其他基站状态，确保决策公平性

注意力增强的提示Decision Transformer框架



Decision Transformer在RIS辅助通信系统中的应用



场景特点

- 不同场景**IRS数目**不同, **阵元数目**不同
- 不同场景**用户数目**不同, **用户需求**不同
- **动作空间**维度大, 呈指数级增长

场景建模

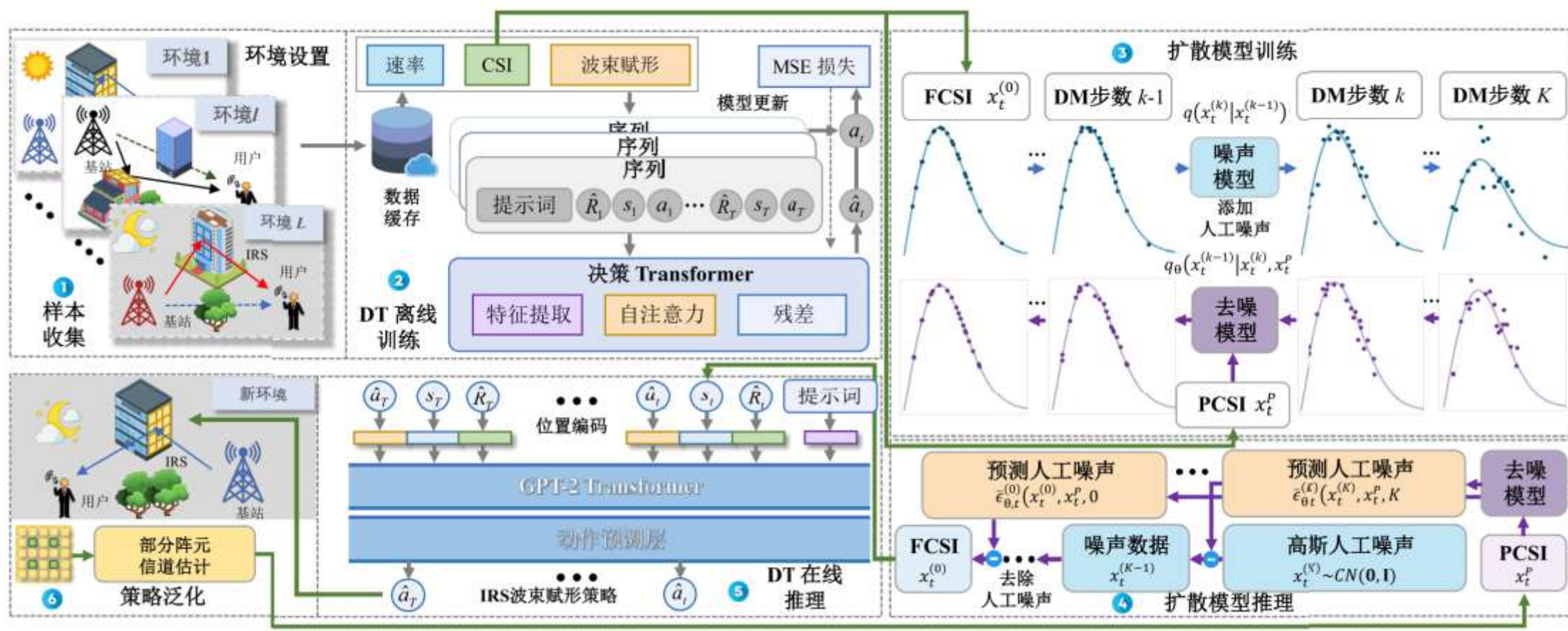
- **状态:** 信道状态信息, 历史接收速率等
- **动作:** IRS相移设计, 基站发射功率等
- **奖励:** 用户接收速率, 用户服务质量等

解决方法

- 根据不同场景设置不同的**提示词**, 指导模型识别场景信息
- **动作嵌入**技术, 降低复杂度

马尔可夫决策过程建模与样本重构

- **状态**: 由扩散模型生成的基站、RIS 和用户之间的级联信道信息
- **动作**: RIS 波束赋形向量
- **奖励**: 用户可达速率
- **样本序列**: 未来累积奖励, 状态, 动作序列 $\tau_t = (\hat{R}_1, s_1, a_1, \dots, \hat{R}_t, s_t, a_t)$



框架介绍

问题与动机

模型依赖

传统优化依赖
精确信道模型



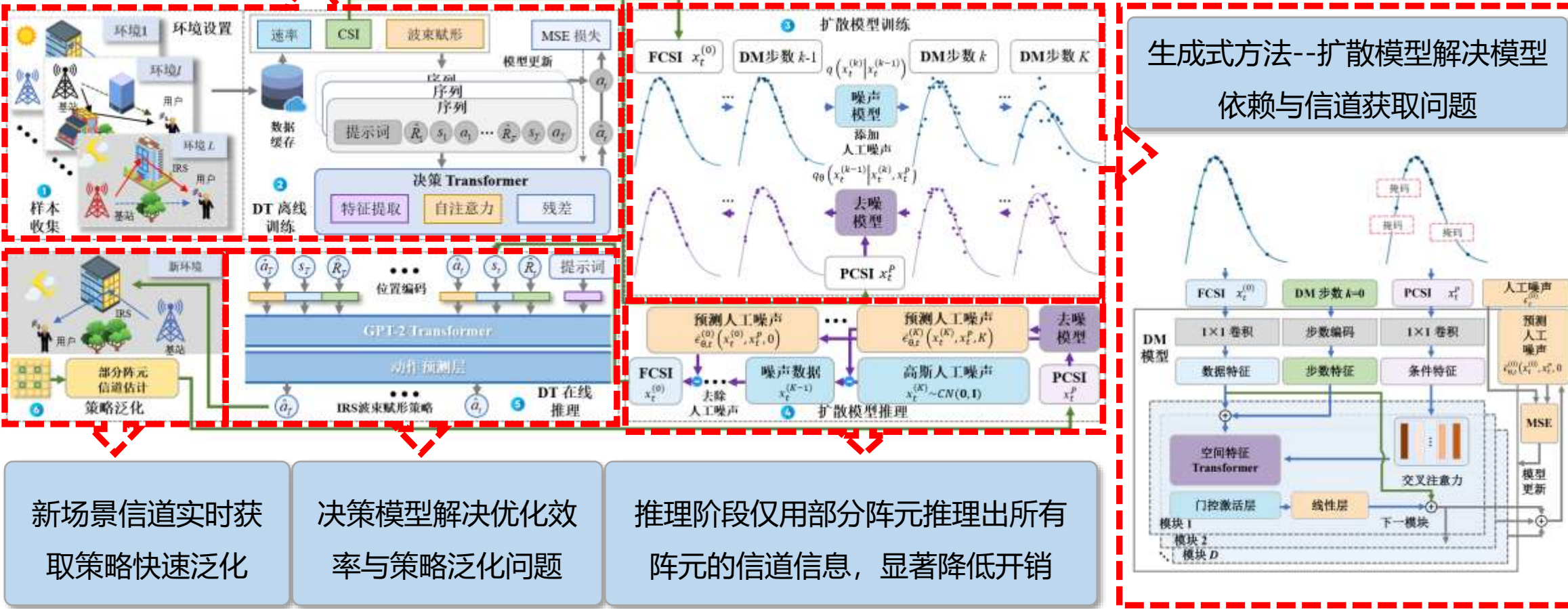
优化效率

值函数近似
存在差错传播



策略泛化

现有策略难以
泛化至新场景



新场景信道实时获取策略快速泛化

决策模型解决优化效率与策略泛化问题

推理阶段仅用部分阵元推理出所有阵元的信道信息，显著降低开销

训练与推理过程

● 决策模型预训练过程：

- **初始化DT网络和环境参数**

for 环境 $n = 1, 2, \dots, L$ do

- **样本收集：**训练强化学习算法，从各种环境中收集离线序列数据存入缓存器 B

- **提示设计：**

① 为每个环境设计独特提示信息和上下文

for 训练回合 $i = 1, 2, \dots, I_s$ do

② 从缓存器中采样批次数据

③ 预测动作： $\hat{a}_t = f_{\omega}(\tau_t)$

- **监督学习：**计算损失 $L = \mathbb{E}[(f_{\omega}(\tau_t) - a_t)^2]$ 并更新参数 $\omega \leftarrow \omega - \lambda \nabla_{\omega} L$

● 新环境小样本微调与推理过程：

- **初始化网络和环境参数，准备少量样本**

- **模型微调：**使用少量样本微调预训练的DEDT模型适应新环境

- **获取信道：**

① 获取信道状态信息 H_0^p 并转换为向量形式，通过扩散去噪生成准确信道估计

for $k = 1, 2, \dots, K$ do

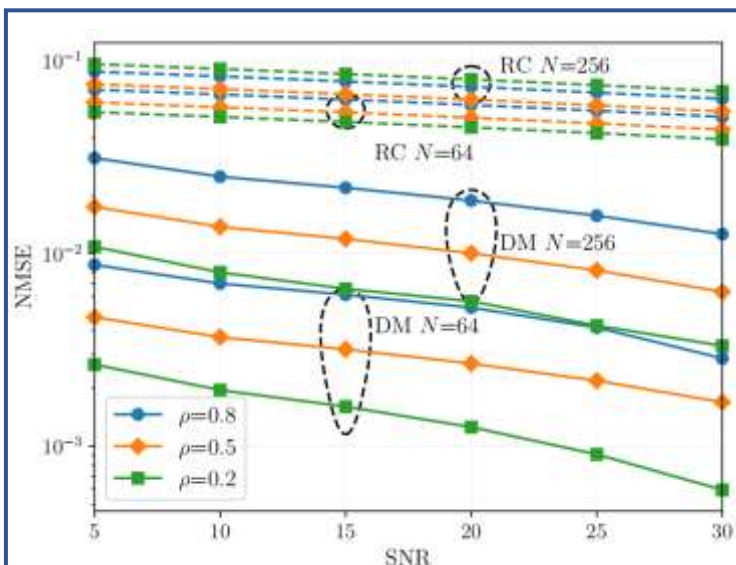
② 预测噪声 $\hat{\epsilon}_{\theta}(k) = f(x_t(k), k)$ 并执行数据去噪

③ 转换为信道形式 s_t 并更新序列 τ_t

- **波束赋形：**基于更新序列输出预测动作 $\hat{a}_t = f_{\omega}(\tau_t)$ 与环境交互获取奖励并更新目标回报和序列

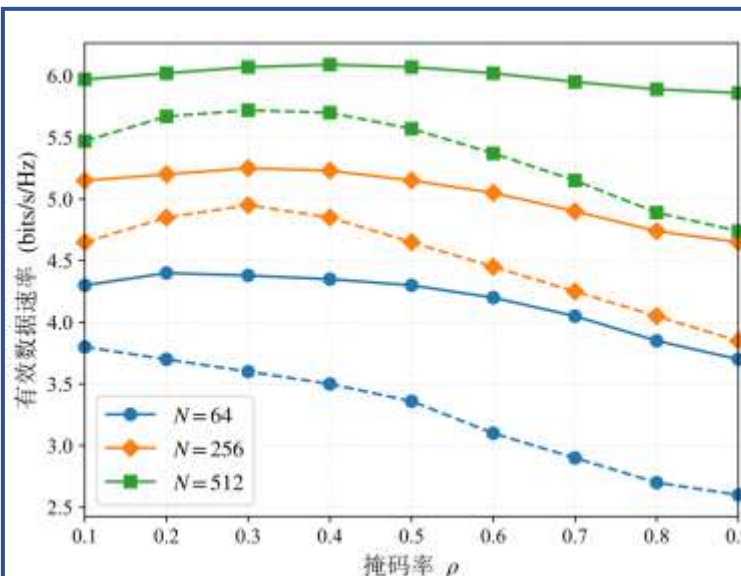
仿真结果

不同信道掩码比率和IRS阵元数量下NMSE与SNR的关系图



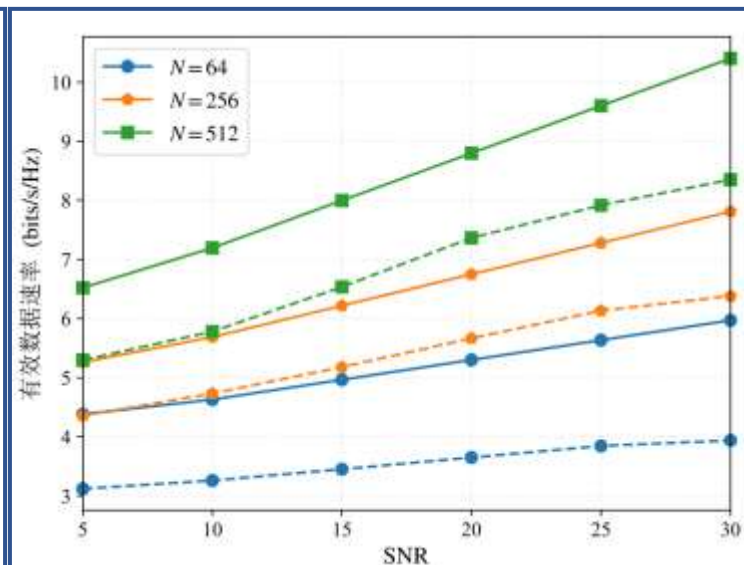
DEDT算法生成信道在不同掩码比率和IRS阵列规模下均显著**优于随机信道方法**

有效掩码速率与信道掩码率的关系图



DEDT算法有效掩码速率在各掩码率下**均优于RCDT**，体现其优越的性能与权衡能力

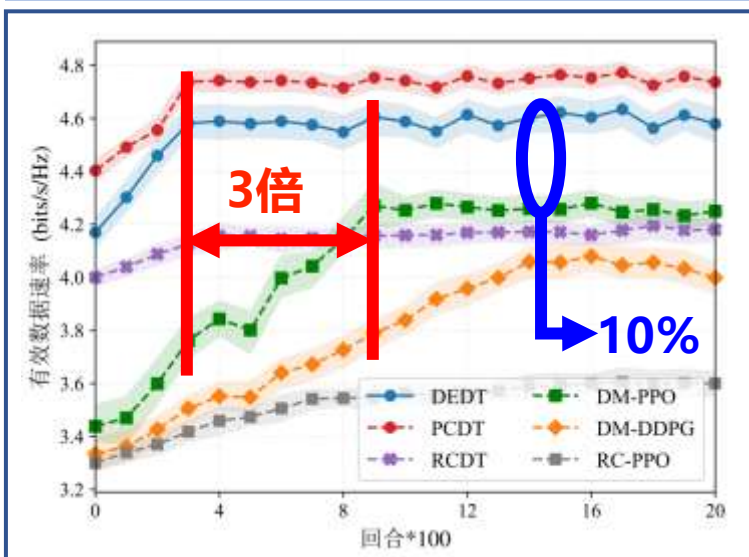
有效数据速率与SNR的关系图



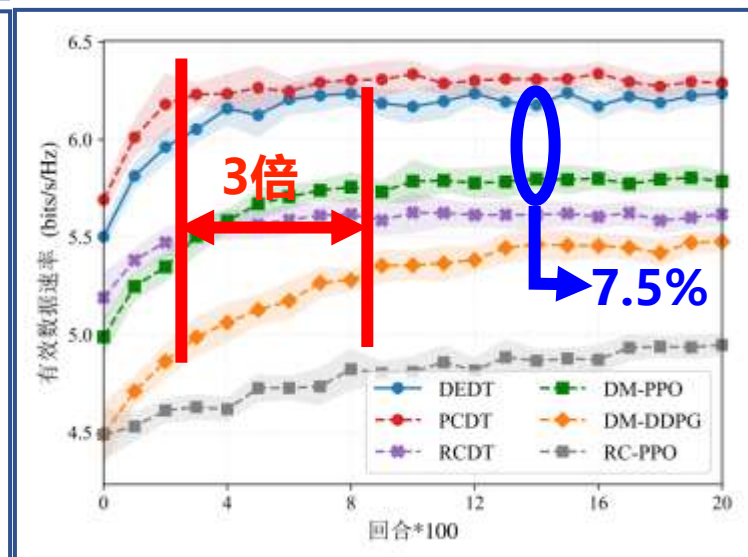
DEDT在不同SNR和IRS条件下始终优于RCDT，**推断更准，速率更高**

仿真结果

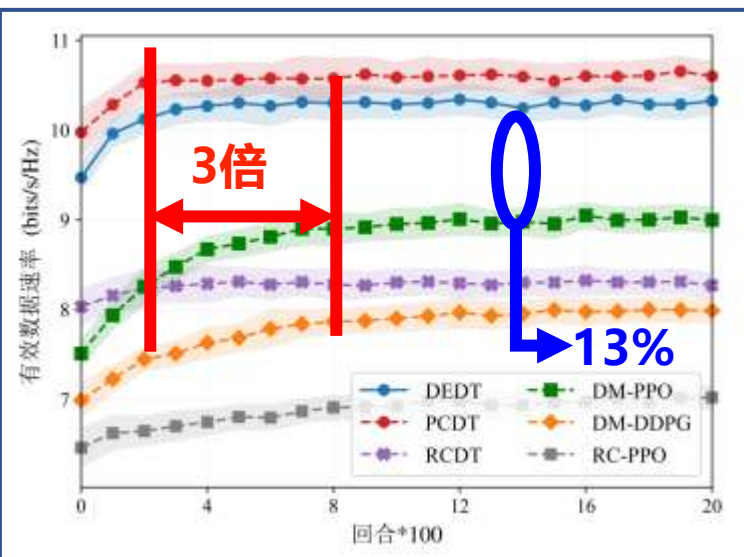
DEDT与基线算法的性能及收敛速度比较图 ($N=64$, SNR=5dB)



DEDT与基线算法的性能及收敛速度比较图 ($N=256$, SNR=15dB)



DEDT与基线算法的性能及收敛速度比较图 ($N=512$, SNR=30dB)



基准算法



- 随机信道近端策略优化算法 (RC-PPO)
- 随机信道决策 Transformer (RCDT)
- 基于扩散模型的近端策略优化算法 (DM-PPO)
- 基于扩散模型的深度确定性策略梯度算法 (DM-DDPG)

相比于传统的从零开始学习的强化学习算法，所提方法在性能上提升了**7.5%**以上，且收敛速度快了**3倍**



東南大學
SOUTHEAST UNIVERSITY

敬请批评指正！