工作进度报告

使用深度强化学习算法优化边缘计算任务卸载问题

范也 MG21320002

已完成工作

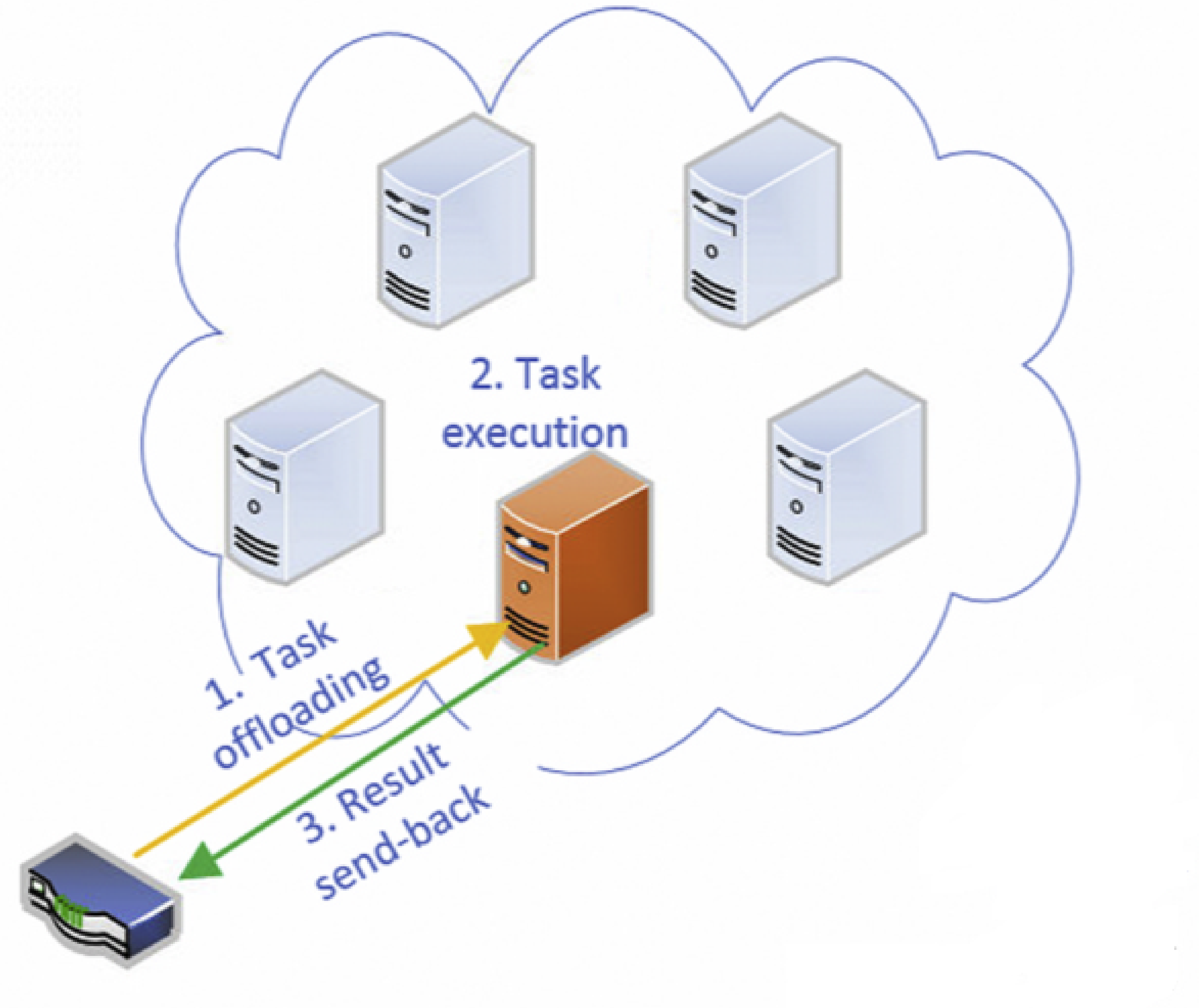
1. 调查20几篇DRL+任务卸载的论文
2. 编写完成基本的实验框架，包含
    1. 边缘计算任务卸载的模拟环境
    2. 使用较先进的DRL算法：T3C、SAC、PPO进行调度

后续工作

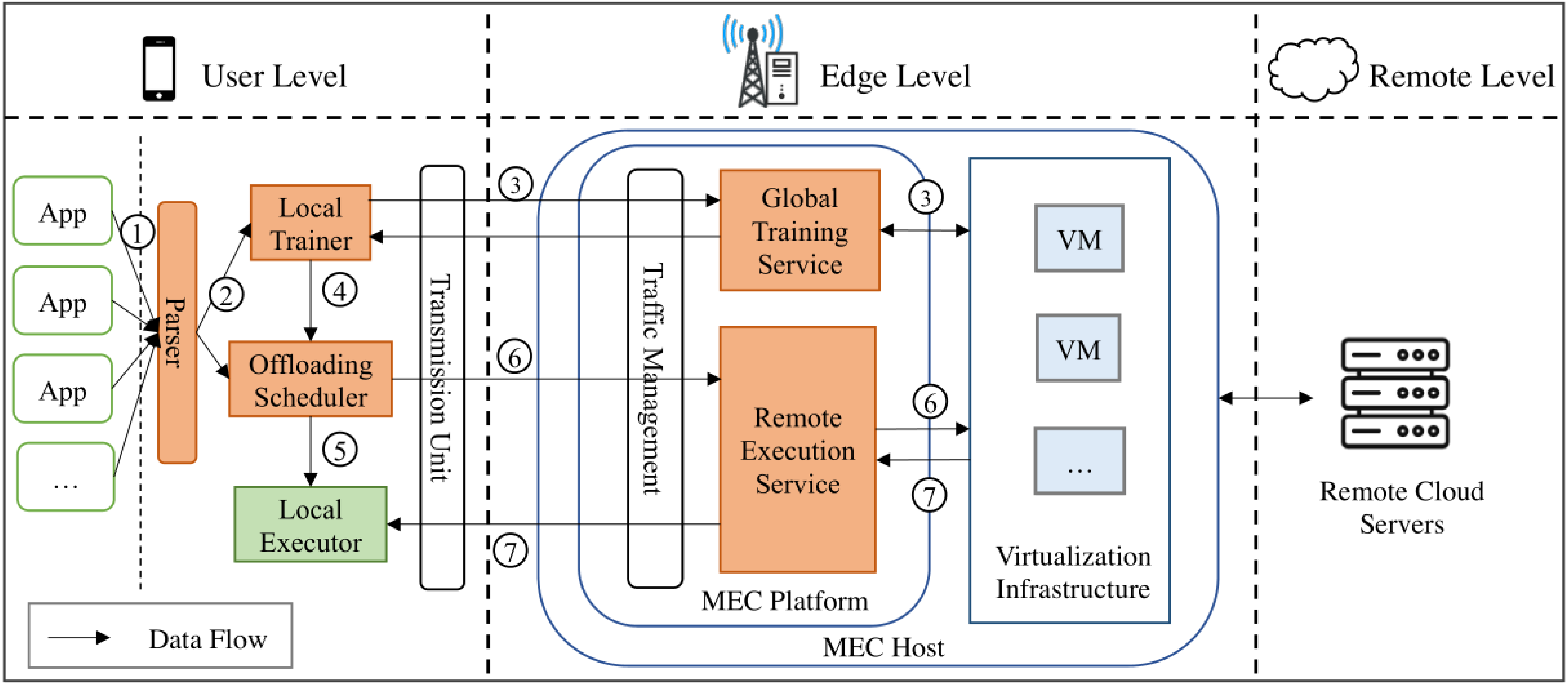
1. 对现有的论文进行较详细的排查，找到比较好的优化目标 2 week
    1. 优化现有的环境和算法
    2. 调参
2. 整合discrete action的DRL算法
3. 实现baseline：贪心、穷举、无调度、线性松弛、DQN

MEC offloading 简介

1. 云计算的不足
2. 工厂、AR游戏...
3. 任务调度：最优化问题
    1. 传统方式（凸优化、非凸优化）、遗传算法
    2. game theory
    3. 深度强化学习算法：速度快、泛化性强
4.
$$\begin{array}{ll} \min_{(\mathbf{a})} & Time + Battery, \\ \text{s.t.} & a_i \in \{0,1\}, \forall i \in \mathcal{M}, \end{array}$$



(a)

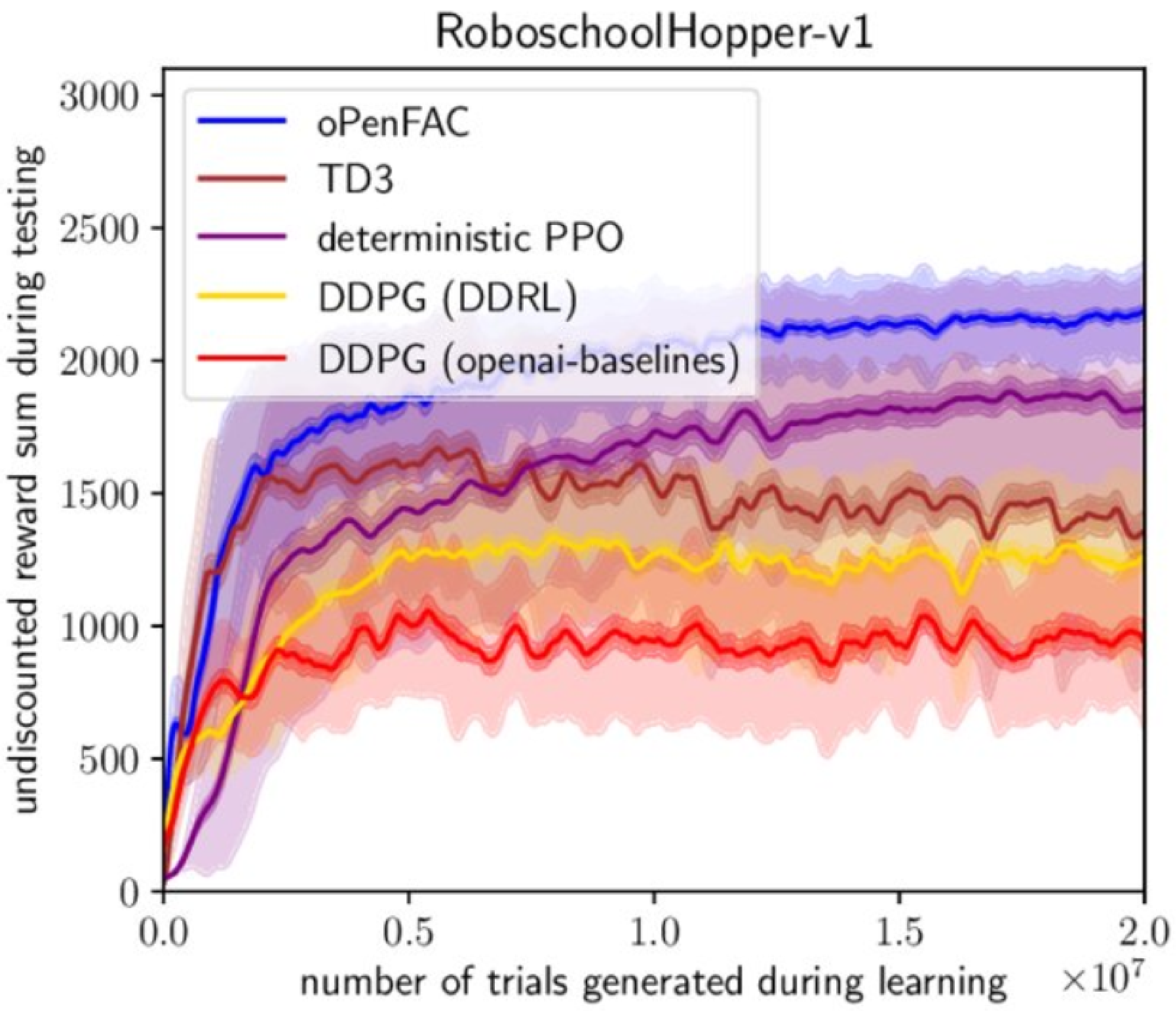

User Level
Edge Level
Remote Level
App
App
App
...
1
Parser
2
Local Trainer
4
Offloading Scheduler
5
Local Executor
Transmission Unit
3
6
7
Traffic Management
Global Training Service
Remote Execution Service
3
6
7
MEC Platform
VM
VM
...
Virtualization Infrastructure
MEC Host
Remote Cloud Servers
Data Flow

已读论文

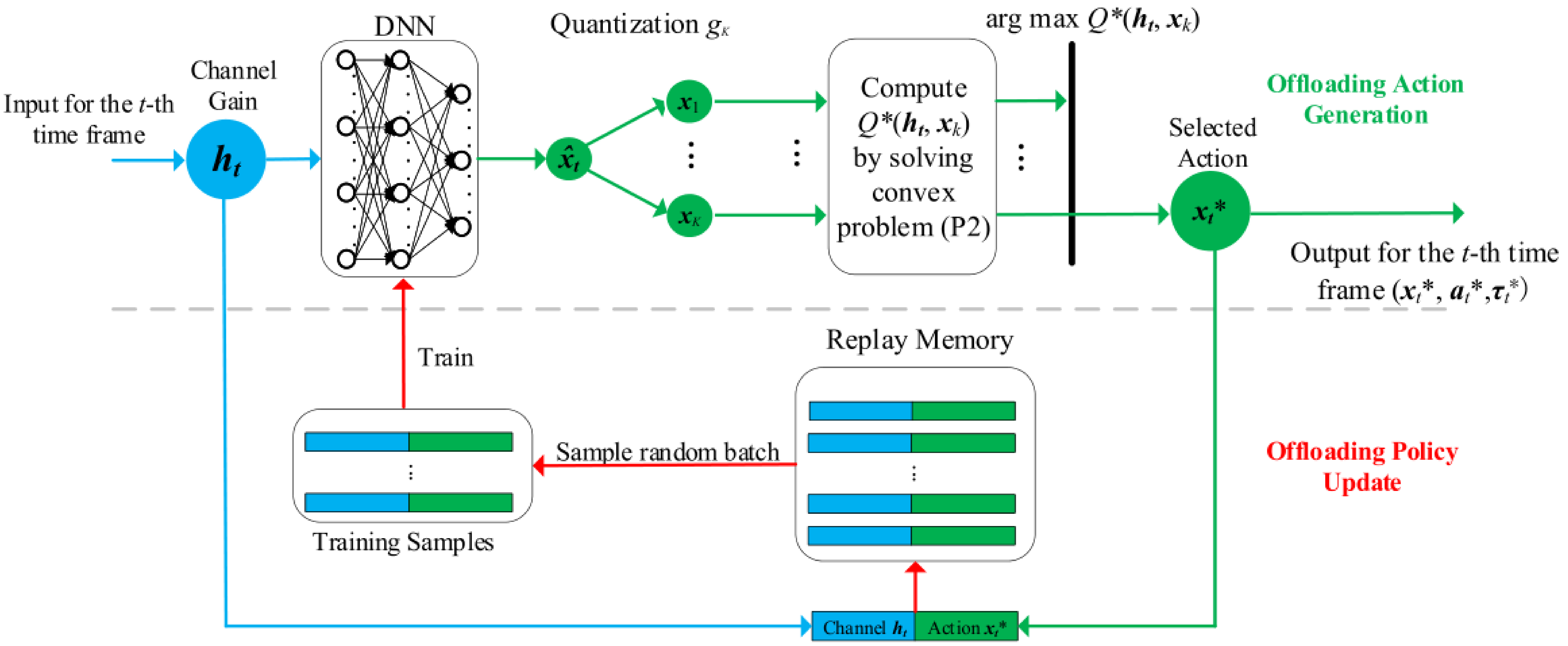
DRL + MEC offloading + Lyapunov + ...

1. 现有论文中使用的DRL算法比较落后：
   REINFORCE、DDPG、A3C、DQN
2. 算法框架基本一样，都是简单的模拟环境+DRL算法
3. 优化目标（公平性、效率、安全性、电量）、考虑的环境（多agent、连续、离散、时间分配方式）各有不同
4. 通过算法中的微调，lyapunuv 或 np-hard证明增加工作量

目前先实现基本的框架，根据具体目标进行微调



Modeling method	Centralized/ distributed	Scalability	Static/ dynamic	Achieved objectives
(Non)convex optimization	Centralized	Not scalable	Static	Global offloading optimization
MDP	Centralized/ distributed	Not scalable	Dynamic	Local offloading optimization
Game theory	Distributed	Scalable	Dynamic	Nash equilibrium
Lyapunov optimization	Centralized	Relatively scalable	Static	Offloading optimization Lyapunov drift
Machine learning	Centralized	Scalable	Dynamic	Intelligent decision



DNN
Quantization g_K
arg max $Q^*(\boldsymbol{h}_t, \boldsymbol{x}_k)$
Channel Gain
Input for the t-th time frame
$\boldsymbol{h}_t$
$\hat{x}_t$
$\boldsymbol{x}_1$
$\boldsymbol{x}_K$
Compute $Q^*(\boldsymbol{h}_t, \boldsymbol{x}_k)$ by solving convex problem (P2)
Offloading Action Generation
Selected Action
$\boldsymbol{x}_t^*$
Output for the t-th time frame $(\boldsymbol{x}_t^*, \boldsymbol{a}_t^*, \boldsymbol{\tau}_t^*)$
Train
Replay Memory
Sample random batch
Training Samples
Offloading Policy Update
Channel $\boldsymbol{h}_t$
Action $\boldsymbol{x}_t^*$

模拟实现

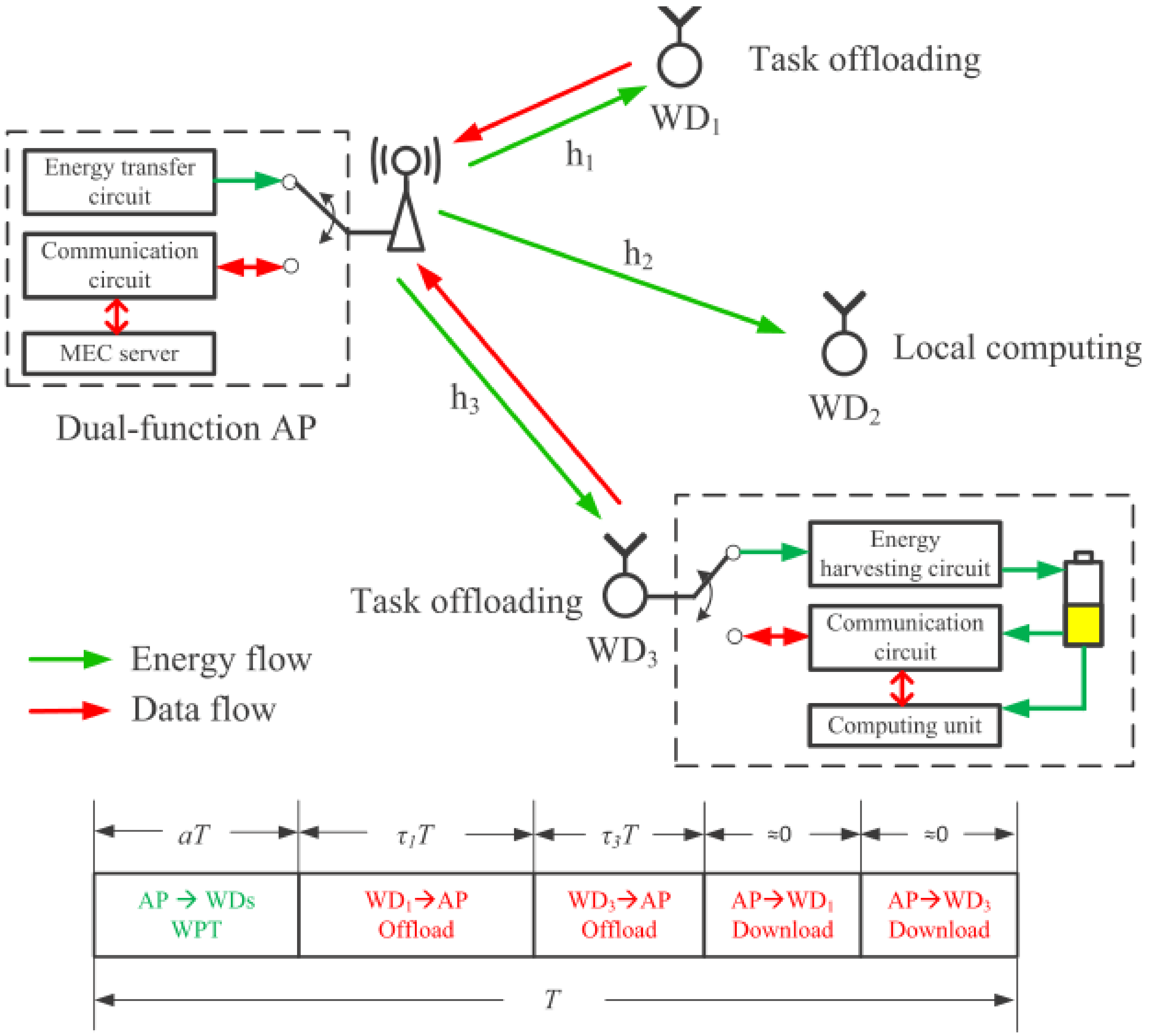
1063 line

已经包含功能特性

1. 电量消耗
2. 处理速度
3. 任务缓冲
4. 数据传输
5. 信道占用

拟添加：

1. 虚拟机配置
2. 位置移动
3. 多server



DRL算法

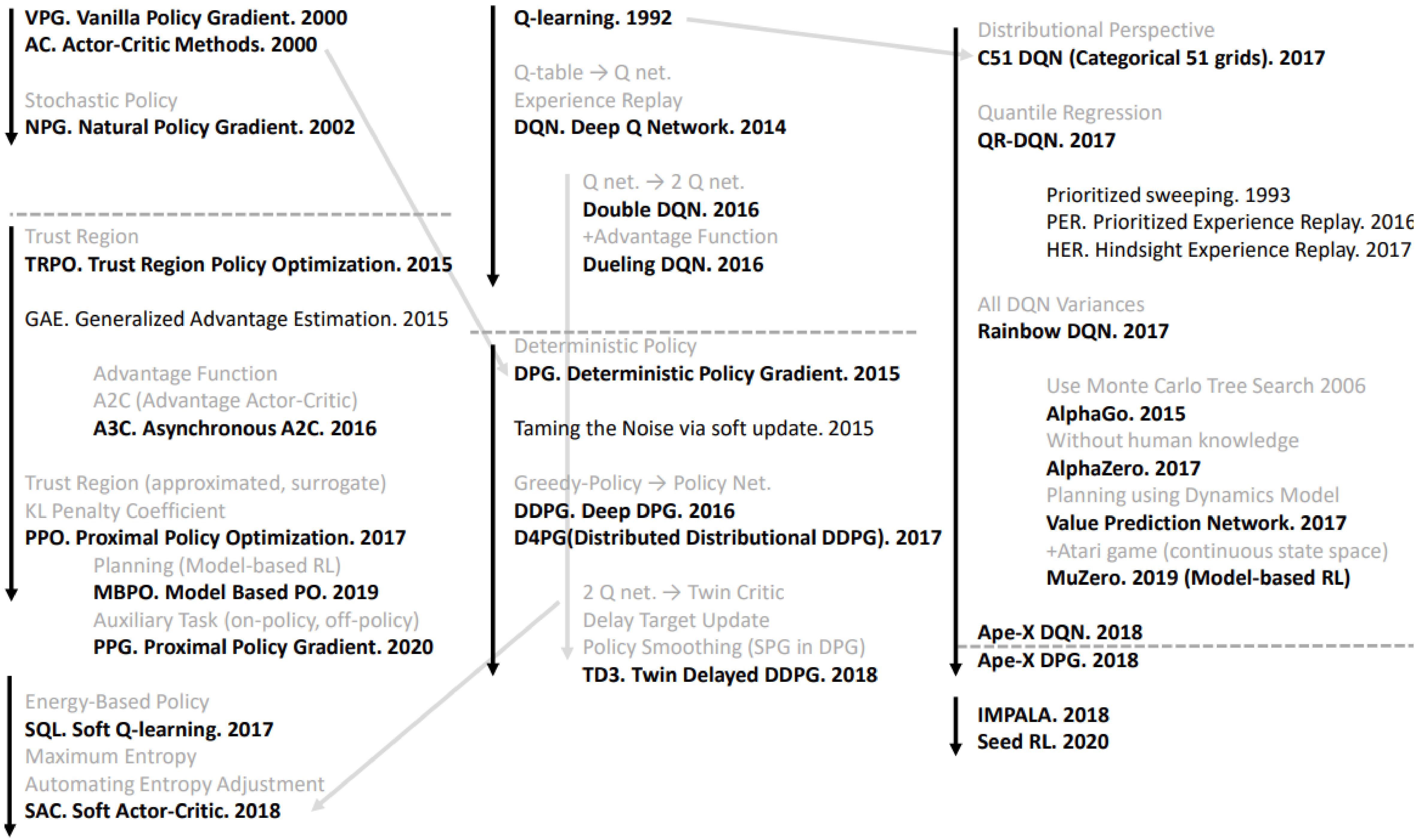
~1000 line

已包含大部分连续动作空间算法：

- TD3、PPO、SAC、DDPG

拟实现其他算法：

- D3QN（离散）、D4PG(连续)



后续工作计划

较新颖的优化目标

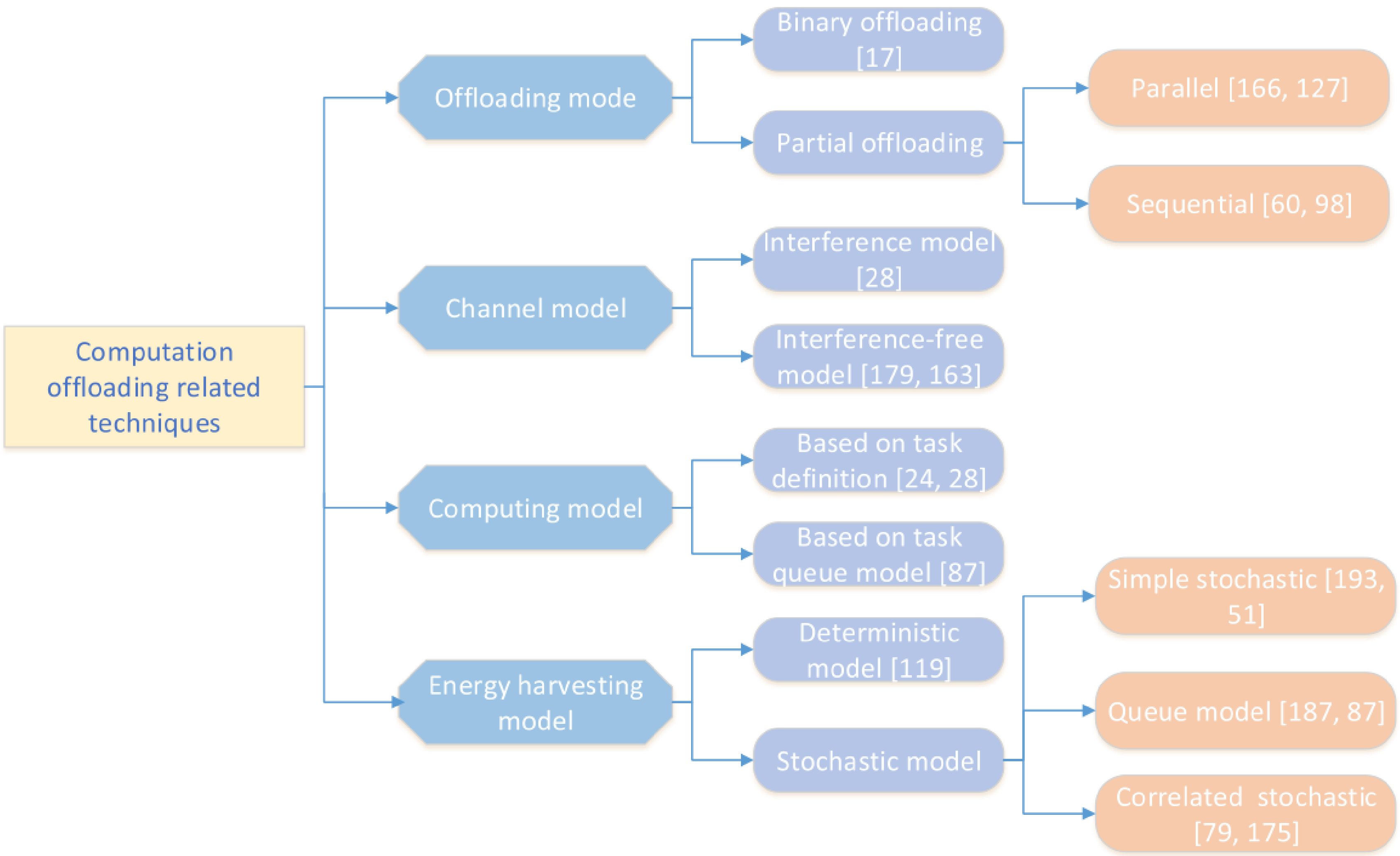
将现有论文中的

- 优化目标
- 特性（安全性、泛化性、算法性能）
- 卸载模型
- 数学理论
- 环境配置

进行总结，找出比较好的优化目标，对现有算法进行调整

实验

- baseline： 贪心、穷举、无调度、线性松弛、DQN
- discrete： D3QN、Rainbow、…



总结

- 基本实现实验框架，包含模拟环境和调度算法

后续工作：

1. 选择更新颖的优化目标和环境设置
2. 调参、基线实验

问题

论文评价