



(12)发明专利申请

(10)申请公布号 CN 109379752 A

(43)申请公布日 2019.02.22

(21)申请号 201811049938.7

(22)申请日 2018.09.10

(71)申请人 中国移动通信集团江苏有限公司
地址 210029 江苏省南京市鼓楼区虎踞路
59号

申请人 中国移动通信集团有限公司

(72)发明人 乔勇

(74)专利代理机构 北京东方亿思知识产权代理
有限责任公司 11258

代理人 彭琼

(51)Int.Cl.

H04W 24/02(2009.01)

H04W 24/10(2009.01)

H04B 7/0413(2017.01)

G06N 3/04(2006.01)

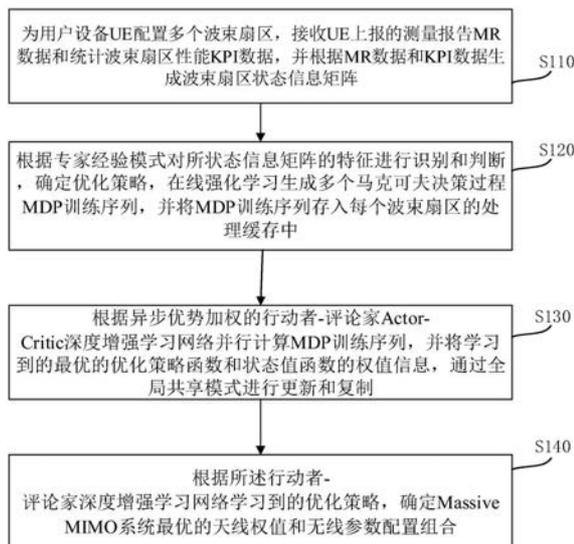
权利要求书4页 说明书19页 附图5页

(54)发明名称

Massive MIMO的优化方法、装置、设备及介
质

(57)摘要

本发明实施例提供的Massive MIMO的优化方法、装置、设备及介质,根据专家经验模式对状态信息矩阵的特征进行识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将MDP训练序列存入每个波束扇区的处理缓存中;根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制;根据行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合,实现优化调整与现网性能评估的在线实时交互,实时采用最优的天线权值和无线参数配置组合,实现网络性能最佳。



1. 一种大规模多入多出Massive MIMO系统优化方法,其特征在于,所述方法包括:

为用户设备UE配置多个波束扇区,接收UE上报的测量报告MR数据和统计波束扇区性能KPI数据,并根据所述MR数据和所述KPI数据生成波束扇区状态信息矩阵;

根据专家经验模式对所述状态信息矩阵的特征进行识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将所述MDP训练序列存入每个波束扇区的处理缓存中;

根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制;

根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合。

2. 根据权利要求1所述的方法,其特征在于,所述为用户设备UE配置多个波束扇区,包括:

采用波束赋形的方式形成多个所述波束扇区覆盖不同空间区域的UE;

根据所述波束扇区的参考信号接收功率强度,确定所述UE接入的波束扇区。

3. 根据权利要求1所述的方法,其特征在于,

所述波束扇区状态信息矩阵的方程为:

$$X_t = \{X_{\text{throughput}}, X_{\text{rsrp}}, X_{\text{Ni}}, X_{\text{TA}}, X_{\text{AOA}}, \dots\}^T,$$

其中, $X_{\text{throughput}}$ 为波束扇区的吞吐量效率, X_{rsrp} 为波束扇区的覆盖水平, X_{Ni} 为波束扇区的干扰水平, X_{TA} 为在该波束扇区下UE用于调整其主小区物理上行链路控制信道PUCCH/物理上行链路共享信道PUSCH/探测参考信号SRS上行发送的时间, X_{AOA} 为确定UE的波达角度, t 为时刻, $t=1,2,3\cdots T$ 。

4. 根据权利要求1所述的方法,其特征在于,所述根据专家经验模式对所述状态信息矩阵的特征进行识别和判断,确定优化策略,包括:

根据专家经验模式对所述波束扇区不同状态信息进行识别和判断;

从天线权值和无线参数的集合中选取相应的天线权值和无线参数;

对选取的所述天线权值和无线参数进行优化;

根据贪婪算法进行所述优化,确定优化策略。

5. 根据权利要求4所述的方法,其特征在于,所述天线权值和无线参数的集合的方程为:

$$A = \{\mathbf{a}_{\text{SS-RSRP/CSI-RSRP_power}}, \mathbf{a}_{\text{azimuth_angle}}, \mathbf{a}_{\text{elevation_angle}}, \mathbf{a}_{\text{horizontal_HPBW}}, \mathbf{a}_{\text{vertical_HPBW}}, \mathbf{a}_{\text{handover_parameter}}, \mathbf{a}_{\text{reselection_parameter}}, \dots\},$$

其中, $\mathbf{a}_{\text{SS-RS/CSI_RS_power}}$ 为同步信号参考信号接收功率SS-RSRP/信道信息状态参考信号接收功率CSI_RSRP功率参数优化调整, $\mathbf{a}_{\text{azimuth_angle}}$ 为波束方位角优化调整, $\mathbf{a}_{\text{elevation_angle}}$ 为波束俯仰角优化调整, $\mathbf{a}_{\text{horizontal_HPBW}}$ 为波束水平波瓣宽度优化调整, $\mathbf{a}_{\text{vertical_HPBW}}$ 为波束垂直波瓣宽度优化调整, $\mathbf{a}_{\text{handover_parameter}}$ 为Massive MIMO小区切换参数优化调整, $\mathbf{a}_{\text{reselection_parameter}}$ 为Massive MIMO小区重选参数的优化调整;

所述优化策略的表达式为：

$$\pi(\mathbf{a}_t | S_t),$$

其中， π 为优化策略函数， \mathbf{a}_t 为天线权值和无线参数优化调整， S_t 为波束扇区的网络状态函数。

6. 根据权利要求1-5任一项所述的方法，其特征在于，所述MDP训练序列的方程为：

$$(S_1, \mathbf{a}_1, S_2, r_1, \dots, S_t, \mathbf{a}_t, S_{t+1}, r_t, \dots, S_{T-1}, \mathbf{a}_{T-1}, S_T, r_T),$$

其中， r_t 为性能回报， \mathbf{a}_t 为选取的所述天线权值和无线参数优化调整， S_t 为波束扇区的网络状态函数。

7. 根据权利要求6所述的方法，其特征在于，所述性能回报的方程为：

$$r_t = \eta \times r_{\text{CASE}_t} + (1 - \eta) \times r_{\text{CESE}_t},$$

其中， r_{CASE_t} 为小区平均吞吐效率， r_{CESE_t} 为小区边缘吞吐效率， η 为小区平均吞吐效率和小区边缘吞吐量频谱效率的期望平衡的权重系数，且 $0 \leq \eta \leq 1$ 。

8. 根据权利要求1所述的方法，其特征在于，所述根据专家经验模式对所述状态信息矩阵的特征进行识别和判断，确定优化策略，包括：

根据专家经验模式对有助于优化网络性能的特征进行识别和判断，确定优化策略；

其中，所述有助于优化网络性能的特征包括所述波束扇区的弱覆盖特征、波束扇区间干扰特征、波束扇区方向与业务UE分布特征。

9. 根据权利要求1所述的方法，其特征在于，所述异步优势加权的行动者-评论家深度增强学习网络并行计算所述MDP训练序列，并将学习到的最优的优化策略函数和状态值函数的权值信息，通过全局共享模式进行更新和复制，包括：

根据深度神经网络，构造基于行动者-评论家Actor-Critic结构的深度增强学习网络，其中，所述深度增强学习网络包括Actor网络模块和Critic网络模块，其中，Actor网络模块用于优化策略函数和调整天线权值和无线参数，所述Critic网络模块用于估计状态值函数；

判断各个所述波束扇区的网络性能是否最佳；

若各个所述波束扇区的网络性能是最佳，则保持当前的天线权值和无线参数配置；

若各个所述波束扇区的网络性能不是最佳，则进入深度增强学习模式。

10. 根据权利要求9所述的方法，其特征在于，所述则进入深度增强学习模式，包括：

根据所述MDP训练序列，并行计算基于优势加权的k步k-step时序差分误差，通过评估所述状态值函数逼近实际回报的程度来评估所述优化策略函数下的优化程度；

所述Critic网络模块根据所述状态值函数，确定所述Critic网络模块的梯度信息，并将所述Critic网络模块的梯度信息反馈至所述Massive MIMO系统共享的Critic网络模块；

根据所述Critic网络模块的评估标准，累计更新所述Actor网络模块的优化策略梯度，并将所述Actor网络模块的梯度信息反馈至所述Massive MIMO系统共享的Actor网络模块；

所述Massive MIMO系统共享的Actor-Critic网络汇总所有所述梯度信息，并异步生成对应的权值参数；

将所述权值参数复制到所述波束扇区的Actor-Critic结构的深度增强学习网络中，直至所有所述波束扇区能够学习到所述Massive MIMO系统的优化策略函数和评估能力。

11. 根据权利要求9或10所述的方法，其特征在于，所述优化策略函数的表达式为：

$\Pi(a_t|S_t;u)$,

所述状态值函数的表达式为:

$V(S_t;v)$,

其中, Π 为优化策略函数, u 和 v 分别为优化策略函数和状态值函数的深度增强学习网络权值参数。

12.根据权利要求10所述的方法,其特征在于,所述根据所述MDP训练序列,并行计算基于优势加权的 k 步 k -step时序差分误差,通过评估所述状态值函数逼近实际回报的程度来评估所述优化策略函数下的优化程度,还包括:

采用基于优势加权的 k 步 k -step时序差分误差作为各个所述波束扇区的优势函数,所述优势函数的方程为:

$$A_t = \sum_{i=0}^k W^i \sum_{j=0}^k (\gamma^j r_{t+i} + \gamma^k V(S_{t+k};v)) - V(S_{t+k};v)$$

其中, W^i 为加权函数, γ 为性能回报的折扣因子,且 $\gamma \in [0,1]$ 。

13.根据权利要求10所述的方法,其特征在于,所述Actor网络模块的优化策略的目标函数为:

$$F_{\Pi}(u) = \log \Pi(a_t|S_t; u)(G_t^w - V(S_t;v)),$$

其中, $\Pi(a_t|S_t;u)$ 为优化策略函数, $V(S_t;v)$ 为状态值函数, G_t^w 为基于优势加权的 k -step时序差分误差的目标函数。

14.根据权利要求1所述的方法,其特征在于,所述根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合之后,还包括:

判断整个Massive MIMO系统的性能回报是否达到最大;

若是,则保持最优的天线权值和无线参数配置组合;

若否,则进入深度增强学习模式,直到整个Massive MIMO系统的性能回报达到最大。

15.一种Massive MIMO系统优化装置,其特征在于,所述装置包括:

矩阵生成模块,用于为用户设备UE配置多个波束扇区,接收UE上报的测量报告MR数据和统计波束扇区性能KPI数据,并根据所述MR数据和所述KPI数据生成波束扇区状态信息矩阵;

训练序列生成模块,用于根据专家经验模式对所述状态信息矩阵的特征进行识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将所述MDP训练序列存入每个波束扇区的处理缓存中;

深度增强学习模块,用于根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制;

优化模块,用于根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合。

16.一种Massive MIMO系统优化设备,其特征在于,包括:至少一个处理器、至少一个存储器以及存储在所述存储器中的计算机程序指令,当所述计算机程序指令被所述处理器执行时实现如权利要求1-14中任一项所述的方法。

17. 一种计算机可读存储介质,其上存储有计算机程序指令,其特征在于,当所述计算机程序指令被处理器执行时实现如权利要求1-14中任一项所述的方法。

Massive MIMO的优化方法、装置、设备及介质

技术领域

[0001] 本发明涉及通信技术领域,尤其涉及一种大规模多输入多输出(Massive Multiple Input Multiple Output,Massive MIMO,MM)的优化方法、装置、设备及介质。

背景技术

[0002] Massive MIMO技术是第五代移动通信技术5G中提高网络覆盖和系统容量的关键技术。在大规模MIMO系统中,基站配置有大量的天线,数目通常有几十、几百甚至几千根,是现有MIMO系统天线数目的数倍、数十倍以上。Massive MIMO充分利用大规模天线阵列和空间自由度,改进了长期演进(Long Term Evolution,LTE)的广播机制,通过波束赋形的机制发射多个波束扇区,提高发射电磁波的功率密度,增强了覆盖范围。但是,Massive MIMO系统的权值参数,随着波束扇区数目和动态扫描维度的增加,波束扇区天线权值和无线参数的组合数由200多个增加至10000多个,再加上小区重选参数、小区切换参数,使波束扇区天线权值和无线参数的组合数量更多,导致Massive MIMO系统的优化调整变得异常复杂。

[0003] 已有的针对Massive MIMO网络扇区的优化方案大致有三种,以下分别做简要介绍:

[0004] 第一种,按照网络规划中的仿真模型进行Massive MIMO网络扇区优化调整。

[0005] 按照网络规划中使用的地图匹配,利用热点场景模型进行业务预测、覆盖仿真和干扰仿真等处理,模型可输出规划仿真后的波束扇区天线权值。规划仿真的目的是尽量使Massive MIMO基站的各个扇区的主瓣方向指向用户需求集中的方向,使用户获得尽可能强的信号,同时还可规划仿真同频、邻频扇区的无线参数,使用户在使用业务过程中获得尽可能大的增益。

[0006] 第二种,通过人工调整的方式进行Massive MIMO网络扇区优化调整。

[0007] 通过Massive MIMO网络统计得到不同方向的业务分布数据、覆盖数据和干扰数据等,大致判断用户业务的聚集程度,再使用测试仪表,进行实地测试,发现覆盖、干扰较差的区域,然后通过人工调整的方式修改Massive MIMO网络的天线权值和无线参数,实现对Massive MIMO网络扇区的权值和无线参数的优化调整。

[0008] 第三种,基于轮询迭代方式进行Massive MIMO网络扇区优化调整。

[0009] 使Massive MIMO网络的每个波束扇区对所有天线权值、无线参数进行轮询遍历,对采集到的海量网络数据进行存储、评估和测试,并以Massive MIMO网络的性能增益为目标,对采集到的样本数据与天线权值以及无线参数的不同组合进行寻优迭代,最终得到Massive MIMO网络扇区最佳的参数组合,从而可实现Massive MIMO网络扇区优化调整。

[0010] 虽然上述这些方案能够在一定程度上实施Massive MIMO网络扇区优化调整,但是,对于快速发展的5G通信技术而言,移动通信基站的建设更新速度很快,已有的方案已远远不能满足实际应用的需求。

[0011] 具体而言,对于第一种,按照网络规划中的仿真模型进行Massive MIMO网络扇区优化调整的方案,仿真模型中每一批新增入网的Massive MIMO基站的网络规划都是严格按

照规划仿真计算出的天线方位角、下倾角、半功率角、参考信号功率等参数设置。但是,当不断有新增基站入网时,已经入网的相邻Massive MIMO基站就需要相应地优化调整;且由于新的基站入网后可使无线环境发生变化,这导致按照规划仿真的参数配置缺乏可持续性,后期网络性能下降。

[0012] 对于第二种,通过人工调整的方式进行Massive MIMO网络扇区优化调整的方案,由于Massive MIMO网络波束扇区天线权值和无线参数的组合数超过10000,基于人工调整的方案不但严重依赖操作人员的经验,而且需要后台数据分析和实地测试,分问题类型、分时间段、分批次地调整网络参数,每次优化评估均要耗费大量人力成本。特别是在Massive MIMO网络规模入网后,需要优化调整的参数数量巨大,人工调整方式显然已无法满足应用需求,效率极低。

[0013] 对于第三种,基于轮询迭代方式进行Massive MIMO网络扇区优化调整,处理时需要对所有天线权值、无线参数进行逐个轮询遍历,采集得到的海量样本数据需要存储和计算,对存储器和计算能力均有很高要求;通过采集到的历史样本数据,再迭代得到最优的权值参数组合是基于旧的优化策略得到的结果,一旦网络环境发生变化,需要采集新的数据才能迭代更新优化策略;并且,由于每个Massive MIMO网络的扇区的无线环境不同,往往需要逐个轮询迭代进行优化配置,即使应用基于有效样本得出的预测模型进行参数组合配置,仍存在误差,仍需要返回现网继续测试和评估,后期还需要进行多次参数修正,整个过程耗时耗力。

发明内容

[0014] 本发明实施例提供了一种Massive MIMO系统优化方案,使Massive MIMO系统不仅在优化效率和网络评估能力可以在线同时提升,而且实现优化调整与现网性能评估的在线实时交互,实时采用最优的天线权值和无线参数配置组合,实现网络性能最佳。

[0015] 第一方面,本发明实施例提供了一种Massive MIMO系统优化方法,方法包括:

[0016] 为用户设备UE配置多个波束扇区,接收UE上报的测量报告MR数据和统计波束扇区性能KPI数据,并根据所述MR数据和所述KPI数据生成波束扇区状态信息矩阵;

[0017] 根据专家经验模式对所述状态信息矩阵的特征进行识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将所述MDP训练序列存入每个波束扇区的处理缓存中;

[0018] 根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制;

[0019] 根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合。

[0020] 在第一方面的一些实施例中,采用波束赋形的方式形成多个所述波束扇区覆盖不同空间区域的UE;

[0021] 根据所述波束扇区的参考信号接收功率强度,确定所述UE接入的波束扇区。

[0022] 在第一方面的一些实施例中,所述波束扇区状态信息矩阵的方程为:

[0023] $X_t = \{X_{\text{throughput}}, X_{\text{rsrp}}, X_{\text{Ni}}, X_{\text{TA}}, X_{\text{AOA}}, \dots\}^T$,

[0024] 其中, $X_{\text{throughput}}$ 为波束扇区的吞吐量效率, X_{rsrp} 为波束扇区的覆盖水平, X_{Ni} 为波束扇区的干扰水平, X_{TA} 为在该波束扇区下 UE 用于调整其主小区物理上行链路控制信道 PUCCH/物理上行链路共享信道 PUSCH/探测参考信号 SRS 上行发送的时间, X_{AOA} 为确定 UE 的波达角度, t 为时刻, $t=1, 2, 3 \dots T$ 。

[0025] 在第一方面的一些实施例中, 所述根据专家经验模式对所述状态信息矩阵的特征进行识别和判断, 确定优化策略, 包括:

[0026] 根据专家经验模式对所述波束扇区不同状态信息进行识别和判断;

[0027] 从天线权值和无线参数的集合中选取相应的天线权值和无线参数;

[0028] 对选取的所述天线权值和无线参数进行优化;

[0029] 根据贪婪算法进行所述优化, 确定优化策略。

[0030] 在第一方面的一些实施例中, 所述天线权值和无线参数的集合的方程为:

[0031] $A = \{\mathbf{a}_{\text{SS-RSRP/CSI-RSRP_power}}, \mathbf{a}_{\text{azimuth_angle}}, \mathbf{a}_{\text{elevation_angle}}, \mathbf{a}_{\text{horizontal_HPBW}}, \mathbf{a}_{\text{vertical_HPBW}}, \mathbf{a}_{\text{handover_parameter}}, \mathbf{a}_{\text{reselection_parameter}}, \dots\}$,

[0032] $\mathbf{a}_{\text{SS-RS/CSI_RS_power}}$ 为同步信号参考信号接收功率 SS-RSRP/信道信息状态参考信号接收功率 CSI-RSRP 功率参数优化调整, $\mathbf{a}_{\text{azimuth_angle}}$ 为波束方位角优化调整, $\mathbf{a}_{\text{elevation_angle}}$ 为波束俯仰角优化调整, $\mathbf{a}_{\text{horizontal_HPBW}}$ 为波束水平波瓣宽度优化调整, $\mathbf{a}_{\text{vertical_HPBW}}$ 为波束垂直波瓣宽度优化调整, $\mathbf{a}_{\text{handover_parameter}}$ 为 Massive MIMO 小区切换参数优化调整, $\mathbf{a}_{\text{reselection_parameter}}$ 为 Massive MIMO 小区重选参数的优化调整;

[0033] 其中, $\mathbf{a}_{\text{SS-RS/CSI_RS_power}}$ 为同步信号参考信号接收功率 SS-RSRP/信道信息状态参考信号接收功率 CSI-RSRP 功率参数优化调整, $\mathbf{a}_{\text{azimuth_angle}}$ 为波束方位角优化调整, $\mathbf{a}_{\text{elevation_angle}}$ 为波束俯仰角优化调整, $\mathbf{a}_{\text{horizontal_HPBW}}$ 为波束水平波瓣宽度优化调整, $\mathbf{a}_{\text{vertical_HPBW}}$ 为波束垂直波瓣宽度优化调整, $\mathbf{a}_{\text{handover_parameter}}$ 为 Massive MIMO 小区切换参数优化调整, $\mathbf{a}_{\text{reselection_parameter}}$ 为 Massive MIMO 小区重选参数的优化调整;

[0034] 其中, $\mathbf{a}_{\text{SS-RS/CSI_RS_power}}$ 为同步信号参考信号接收功率 SS-RSRP/信道信息状态参考信号接收功率 CSI-RSRP 功率参数优化调整, $\mathbf{a}_{\text{azimuth_angle}}$ 为波束方位角优化调整, $\mathbf{a}_{\text{elevation_angle}}$ 为波束俯仰角优化调整, $\mathbf{a}_{\text{horizontal_HPBW}}$ 为波束水平波瓣宽度优化调整, $\mathbf{a}_{\text{vertical_HPBW}}$ 为波束垂直波瓣宽度优化调整, $\mathbf{a}_{\text{handover_parameter}}$ 为 Massive MIMO 小区切换参数优化调整, $\mathbf{a}_{\text{reselection_parameter}}$ 为 Massive MIMO 小区重选参数的优化调整;

[0035] 所述优化策略的表达式为:

[0036] $\pi(\mathbf{a}_t | S_t)$,

[0037] 其中, π 为优化策略函数, \mathbf{a}_t 为天线权值和无线参数优化调整, S_t 为波束扇区的网络状态函数。

[0038] 在第一方面的一些实施例中, 所述 MDP 训练序列的方程为: $(S_1, \mathbf{a}_1, S_2, r_1, \dots, S_t, \mathbf{a}_t, S_{t+1}, r_t, \dots, S_{T-1}, \mathbf{a}_{T-1}, S_T, r_T)$,

[0039] 其中, r_t 为性能回报, \mathbf{a}_t 为选取的所述天线权值和无线参数优化调整, S_t 为波束扇区的网络状态函数。

[0040] 在第一方面的一些实施例中, 所述性能回报的方程为:

[0041] $r_t = \eta \times r_{\text{CASE}_t} + (1-\eta) \times r_{\text{CESE}_t}$,

[0042] 其中, r_{CASE_t} 为小区平均吞吐效率, r_{CESE_t} 为小区边缘吞吐效率, η 为小区平均吞吐效率和小区边缘吞吐量频谱效率的期望平衡的权重系数, 且 $0 \leq \eta \leq 1$ 。

[0043] 在第一方面的一些实施例中, 所述根据专家经验模式对所述状态信息矩阵的特征进行识别和判断, 确定优化策略, 包括:

[0044] 根据专家经验模式对有助于优化网络性能的特征进行识别和判断, 确定优化策略;

[0045] 其中, 所述有助于优化网络性能的特征包括所述波束扇区的弱覆盖特征、波束扇区间干扰特征、波束扇区方向与业务 UE 分布特征。

[0046] 在第一方面的一些实施例中, 所述异步优势加权的行动者-评论家深度增强学习

网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制,包括:

[0047] 根据深度神经网络,构造基于行动者-评论家Actor-Critic结构的深度增强学习网络,其中,所述深度增强学习网络包括Actor网络模块和Critic网络模块,其中,Actor网络模块用于优化策略函数和调整天线权值和无线参数,所述Critic网络模块用于估计状态值函数;

[0048] 判断各个所述波束扇区的网络性能是否最佳;

[0049] 若各个所述波束扇区的网络性能是最佳,则保持当前的天线权值和无线参数配置;

[0050] 若各个所述波束扇区的网络性能不是最佳,则进入深度增强学习模式。

[0051] 在第一方面的一些实施例中,所述则进入深度增强学习模式,包括:

[0052] 根据所述MDP训练序列,并行计算基于优势加权的k步k-step时序差分误差,通过评估所述状态值函数逼近实际回报的程度来评估所述优化策略函数下的优化程度;

[0053] 所述Critic网络模块根据所述状态值函数,确定所述Critic网络模块的梯度信息,并将所述Critic网络模块的梯度信息反馈至所述Massive MIMO系统共享的Critic网络模块;

[0054] 根据所述Critic网络模块的评估标准,累计更新所述Actor网络模块的优化策略梯度,并将所述Actor网络模块的梯度信息反馈至所述Massive MIMO系统共享的Actor网络模块;

[0055] 所述Massive MIMO系统共享的Actor-Critic网络汇总所有所述梯度信息,并异步生成对应的权值参数;

[0056] 将所述权值参数复制到所述波束扇区的Actor-Critic结构的深度增强学习网络中,直至所有所述波束扇区能够学习到所述Massive MIMO系统的优化策略函数和评估能力。

[0057] 在第一方面的一些实施例中,所述优化策略函数的表达式为:

[0058] $\Pi(at|St;u)$,

[0059] 所述状态值函数的表达式为:

[0060] $V(St;v)$,

[0061] 其中, Π 为优化策略函数, u 和 v 分别为优化策略函数和状态值函数的深度增强学习网络权值参数。

[0062] 在第一方面的一些实施例中,所述根据所述MDP训练序列,并行计算基于优势加权的k步k-step时序差分误差,通过评估所述状态值函数逼近实际回报的程度来评估所述优化策略函数下的优化程度,还包括:

[0063] 采用基于优势加权的k步k-step时序差分误差作为各个所述波束扇区的优势函数,所述优势函数的方程为:

[0064] $A_t = \sum_{i=0}^k W^i \sum_{i=0}^k (\gamma^i r_{t+i} + \gamma^k V(S_{t+k}; v)) - V(S_{t+k}; v)$

[0065] 其中, W^i 为加权函数, γ 为性能回报的折扣因子,且 $\gamma \in [0, 1]$ 。

[0066] 在第一方面的一些实施例中,所述Actor网络模块的优化策略的目标函数为:

[0067] $F_{\Pi}(\mathbf{u}) = \log \Pi(\mathbf{a}_t | \mathbf{S}_t; \mathbf{u})(G_t^W - V(\mathbf{S}_t; \mathbf{v}))$,

[0068] 其中, $\Pi(\mathbf{a}_t | \mathbf{S}_t; \mathbf{u})$ 为优化策略函数, $V(\mathbf{S}_t; \mathbf{v})$ 为状态值函数, G_t^W 为基于优势加权的k-step时序差分误差的目标函数。

[0069] 在第一方面的一些实施例中,所述根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合之后,还包括:

[0070] 判断整个Massive MIMO系统的性能回报是否达到最大;

[0071] 若是,则保持最优的天线权值和无线参数配置组合;

[0072] 若否,则进入深度增强学习模式,直到整个Massive MIMO系统的性能回报达到最大。

[0073] 第二方面,本发明实施例提供了一种Massive MIMO系统优化装置,装置包括:

[0074] 矩阵生成模块,用于为用户设备UE配置多个波束扇区,接收UE上报的测量报告MR数据和统计波束扇区性能KPI数据,并根据所述MR数据和所述KPI数据生成波束扇区状态信息矩阵;

[0075] 训练序列生成模块,用于根据专家经验模式对所述状态信息矩阵的特征进行识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将所述MDP训练序列存入每个波束扇区的处理缓存中;

[0076] 深度增强学习模块,用于根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制;

[0077] 优化模块,用于根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合。

[0078] 本发明实施例提供了一种Massive MIMO系统优化设备,包括:至少一个处理器、至少一个存储器以及存储在存储器中的计算机程序指令,当计算机程序指令被处理器执行时实现如上述实施方式中第一方面的方法。

[0079] 第四方面,本发明实施例提供了一种计算机可读存储介质,其上存储有计算机程序指令,当计算机程序指令被处理器执行时实现如上述实施方式中第一方面的方法。

[0080] 本发明实施例提供的Massive MIMO系统优化方法、装置、设备及介质,为用户设备UE配置多个波束扇区,各个波束扇区接收UE上报的测量报告MR数据和统计波束扇区性能KPI数据,并根据所述MR数据和所述KPI数据生成波束扇区状态信息矩阵;根据专家经验模式对所述状态信息矩阵的特征进行识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将所述MDP训练序列存入每个波束扇区的处理缓存中;根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制;根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合,使Massive MIMO系统不仅在优化效率和网络评估能力可以在线同时提升,而且实现优化调整与现网性能评估的在线实时交互,实时采用最优的天线权值和无线参数配置组合,实现网络性能最佳。

附图说明

[0081] 为了更清楚地说明本发明实施例的技术方案,下面将对本发明实施例中所需要使用的附图作简单地介绍,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0082] 图1示出了本发明实施例提供的Massive MIMO的优化方法的示意图;

[0083] 图2示出了本发明一个示例中Massive MIMO系统中天线模块示意图;

[0084] 图3示出了本发明一个示例中Massive MIMO系统波束扇区示意图;

[0085] 图4示出了本发明实施例提供的行动者-评论家深度增强学习网络的示意图;

[0086] 图5示出了本发明一个示例中Massive MIMO系统波束扇区优化示意图;

[0087] 图6示出了本发明一个示例中Massive MIMO系统中多个波束扇区优化示意图;

[0088] 图7示出了本发明实施例提供的Massive MIMO的优化装置的结构示意图;

[0089] 图8示出了本发明实施例提供的Massive MIMO的优化设备的硬件结构示意图。

具体实施方式

[0090] 下面将详细描述本发明的各个方面的特征和示例性实施例,为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细描述。应理解,此处所描述的具体实施例仅被配置为解释本发明,并不被配置为限定本发明。对于本领域技术人员来说,本发明可以在不需要这些具体细节中的一些细节的情况下实施。下面对实施例的描述仅仅是为了通过示出本发明的示例来提供对本发明更好的理解。

[0091] 需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0092] 下面结合附图,详细描述根据本发明实施例的Massive MIMO系统优化方法、装置、设备及介质。应注意,这些实施例并不是用来限制本发明公开的范围。

[0093] 图1是示出根据本发明实施例的Massive MIMO系统优化方法的示意图。如图1所示,本发明实施例中的Massive MIMO系统优化方法包括以下步骤:

[0094] S110,为用户设备(User Equipment,UE)配置多个波束扇区,接收UE上报的测量报告MR数据和统计波束扇区性能KPI数据,并根据所述MR数据和所述KPI数据生成波束扇区状态信息矩阵。

[0095] S120,根据专家经验模式对所述状态信息矩阵的特征进行识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将所述MDP训练序列存入每个波束扇区的处理缓存中。

[0096] S130,根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制。

[0097] S140,根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合。

[0098] 其中,用户设备UE也可以被称为终端,用户设备可以是手机、平板电脑、笔记本电脑等。

[0099] 在S110中,作为一个示例,采用波束赋形的方式为UE配置N个波束扇区(Sector_1、Sector_2、...Sector_i、...Sector_n),其中,N为正整数,N个波束扇区覆盖不同空间、不同区域的UE,每个波束扇区都拥有不同的波束索引,覆盖一个UE的多个波束扇区的参考信号接收功率(Reference Signal Receiving Power,RSRP)的强度是不同的。UE从多个波束扇区中选择一个波束扇区的RSRP强度最强的接入,其中,RSRP可以是信道信息状态参考信号接收功率(Channel Status Information_Reference Signal Receiving Power,CSI_RSRP),也可以是同步信号参考信号接收功率(Synchronous Signal_Reference Signal Receiving Power,SS_RSRP)。

[0100] 图2示出了本发明一个示例中Massive MIMO系统中天线模块示意图,如图2所示,Massive MIMO系统中天线由 $M_g \times N_g$ 个子天线模块组成,各子天线模块的中心的水平方向间距为 $d_{g,h}$,各子天线模块的中心的垂直方向间距为 $d_{g,v}$,每个子天线模块包括 $M \times N \times Q$ 个天线阵子,该 $M \times N \times Q$ 个天线阵子组成一个收发单元(Transceiver Unit, TXRU),每个TXRU对应一个波束扇区(Sector),其中,M是每个列中具有相同极化的天线元件的数目,N是列数,Q表示极化数量,其中Q取1表示单极化,Q取2表示双极化。

[0101] 图3示出了本发明一个示例中Massive MIMO系统波束扇区示意图,如图3所示,Massive MIMO天线结构实现空间分集、空间复用和波束赋形功能都是以子天线块为单位,每个子天线模块相当于多天线中的每个子天线;图3中以子天线块为单位实现波束赋形功能,因为每个子天线块实际上是一个阵元数为 $M \times N \times Q$ 个的阵列模块,由 $N \times M$ 个子天线块组成的多天线是一个二维系统,由 $M \times N \times Q$ 个阵元组成的阵列是一个三维系统,所以,Massive MIMO天线中的总阵元数为 $M_g \times N_g \times M \times N \times Q$ 个。

[0102] 在S110中,作为一个示例,MM系统中的每个波束扇区Sector_i,可根据波束扇区接收UE上报的测量报告(Measurement Report,MR)数据和波束扇区性能(Key Performance Indicators,KPI)数据,生成对应时刻的波束扇区的状态信息矩阵 X_t 。例如,在时刻t,根据Massive MIMO系统中各个波束扇区Sector的吞吐量效率、覆盖水平、干扰水平、UE分布等状态信息,可以用表达式(1)来表示单个波束扇区的状态信息矩阵,如下:

$$[0103] \quad X_t = \{X_{\text{throughput}}, X_{\text{rsrp}}, X_{\text{Ni}}, X_{\text{TA}}, X_{\text{AOA}}, \dots\}^T \quad (1)$$

[0104] 其中, X_t 表示时刻t($t=1,2,3,\dots,T$)当前波束扇区的状态信息矩阵;

[0105] $X_{\text{throughput}}$ 表示波束扇区的吞吐量效率,这里,波束扇区的吞吐量效率包括小区平均吞吐量效率 $X_{\text{throughput_averager}}$ 和小区边缘吞吐量效率;

[0106] X_{rsrp} 表示波束扇区的覆盖水平,其中, $X_{\text{rsrp_serving_sector}}$ 表示在该波束扇区下UE上报测量的服务扇区的参考信号接收功率情况, $X_{\text{rsrp_neighbor_sector}_1}$ 、 $X_{\text{rsrp_neighbor_sector}_2}$ 、 $X_{\text{rsrp_neighbor_sector}_3}$ 、...、 $X_{\text{rsrp_neighbor_sector}_n}$ 表示在该波束扇区下UE上报测量的所有相邻波束扇区的参考信号接收功率情况,其中,n表示相邻波束扇区的个数,且相邻波束扇区的参考信号接收功率情况按照信号强度由强到弱降序排列;

[0107] X_{Ni} 表示波束扇区的干扰水平,其中, $X_{Ni_UL_sinnr}$ 表示所有UE上报测量的波束扇区的上行信号链路质量, $X_{Ni_DL_sinnr}$ 表示所有UE上报测量波束扇区的下行信号链路质量; X_{TA} 表示在该波束扇区下UE用于调整其主小区物理上行链路控制信道(Physical Uplink Control Channel,PUCCH)/物理上行链路共享信道(Physical Uplink Shared Channel,PUSCH)/探测参考信号(Sounding Reference Signal,SRS)上行发送的时间,进而确定UE距离Massive MIMO视频系统的远近;

[0108] X_{AOA} 表示确定UE的波达角度,其中,在Massive MIMO系统的立体空间中包括水平波达角度 $X_{Horizontal_AOA}$ 和垂直波达角度 $X_{Vertical_AOA}$ 。

[0109] 在本发明的实施例中,每个波束扇区都有对应的状态信息矩阵,其他波束扇区的状态信息矩阵都可以根据上述表达式(1)来确定。

[0110] 在本发明的实施例中,波束扇区的状态信息矩阵不仅仅包括波束扇区的吞吐量效率、覆盖水平、干扰水平、UE分布等状态信息,还可以包括其他对优化网络性能有用的特定的状态信息,其他特定的状态信息可根据实际需求进行设置。

[0111] 在本发明的某些实施例中,可将MR数据和KPI数据进行归一化处理,即将MR数据和KPI数据统一映射到 $[0,1]$ 的区间上,形成标准化数值,便于进行相关数据处理。

[0112] 举例来说,根据测量报告MR数据和波束扇区性能KPI数据,生成波束扇区的状态信息矩阵,并将涉及到的数据进行归一化处理,每个波束扇区都有对应的状态信息矩阵,便于特征的识别和判断,便于对波束扇区的并行处理,且数据进行归一化处理,能够防止参数异常,计算速度快。

[0113] 在S120中,作为一个示例,根据专家经验模式对Massive MIMO系统下,各自不同的波束扇区的状态信息矩阵 X_t 的不同网络状态 S_t 的特征进行识别和判断,并针对性地从Massive MIMO系统优化调整所涉及的天线权值和无线参数的集合中,选取相应的天线权值和无线参数,并对选取的天线权值和无线参数进行优化动作 α ,为了获取该优化策略下的最大性能回报 r ,采取贪婪算法进行优化调整,进而形成专家经验模式下的特定优化策略 $\pi(\alpha_t | S_t)$;根据特定优化策略,在线强化学习生成马克可夫决策过程(Markov Decision Processes,MDP)训练序列,MDP训练序列为 $(S_1, \alpha_1, S_2, r_1, \dots, S_t, \alpha_t, S_{t+1}, r_t, \dots, S_{T-1}, \alpha_{T-1}, S_T, r_T)$ 。

[0114] 在本发明实施例中,特定优化策略 $\pi(\alpha_t | S_t)$ 是每个波束扇区(Sector_i)不同的网络状态通过专家经验模式对状态信息矩阵 X_t 中的特征进行自动识别和判断,选取相应的天线权值和无线参数组合进行优化动作 α ,形成的专家经验模式下的特定优化策略 π 。其中, S 表示MM系统下波束扇区的网络环境或网络性能的状态函数, X_t 为时刻 t 波束扇区的状态信息矩阵 X_t ,对应的 $S_{t+1} = (S_t, \alpha_t, X_{t+1})$,对应的 $S_t = [X_t]$,即, S_1 是波束扇区的状态信息矩阵 X_1 的状态函数, S_t 是波束扇区的状态信息矩阵 X_t 的状态函数。

[0115] A 为Massive MIMO小区优化调整所涉及的天线权值和无线参数的设置的集合,优化动作 $\alpha \in A$,集合 A 中天线权值和无线参数的所有优化动作 α 均被专家经验模式下的特定优化策略 π 所映射。

[0116] 作为一个示例,可以用表达式(2)来表示集合 A 。

[0117] $A = \{\alpha_{SS-RS/CSI-RS_power}, \alpha_{azimuth_angle}, \alpha_{elevation_angle}, \alpha_{horizontal_HPBW}, \alpha_{vertical_HPBW}, \alpha_{handover_parameter}, \alpha_{reselection_}$

parameter, ...} (2)

[0118] 其中, $\alpha_{SS-RS/CSI_RS_power}$ 为SS-RS/CSI_RS功率参数优化调整, $\alpha_{azimuth_angle}$ 为波束方位角优化调整, $\alpha_{elevation_angle}$ 为波束俯仰角优化调整, $\alpha_{horizontal_HPBW}$ 为波束水平波瓣宽度优化调整, $\alpha_{vertical_HPBW}$ 为波束垂直波瓣宽度优化调整, $\alpha_{handover_parameter}$ 为Massive MIMO小区切换参数优化调整, $\alpha_{reselection_parameter}$ 为Massive MIMO小区重选参数优化调整, 等。

[0119] 在本发明的实施例中, 集合A不仅仅包括覆盖功率参数、波束方位角、波束俯仰角、波束垂直波瓣宽度、小区切换参数、小区重选参数等天线权值和无线参数的优化动作, 还包括其他优化调整所涉及的天线权值和无线参数的优化动作, 其他优化调整所涉及的天线权值和无线参数可根据实际需求进行设置。

[0120] 作为一个示例, 为了在执行优化动作后获得波束扇区的最大性能回报奖励, 采用贪婪算法调整优化动作, r 表示针对 S_t 下执行优化动作 a_t 后的性能回报, 即 $r(S_t, a_t, S_{t+1})$ 为标量函数, 即Massive MIMO系统中波束扇区根据当前的网络状态执行优化动作 a 之后, 波束扇区得到的性能回报, 可以通过表达式 (3) 来计算波束扇区的性能回报:

$$[0121] \quad r_t = \eta \times r_CASE_{t+1} + (1-\eta) \times r_CESE_t \quad (3)$$

[0122] 其中, r_CASE_t 为小区平均吞吐量效率的性能回报, r_CESE_t 为小区边缘吞吐量效率的性能回报, η 为小区平均吞吐量效率和小区边缘吞吐量效率的平衡的权重系数, 且, $0 \leq \eta \leq 1$ 。例如, UE数量相对较少推荐取值95%, UE数量相对较多推荐取值60%。 η 的取值可以根据UE数量进行设置。

$$[0123] \quad \text{其中, } r_CASE_t = \begin{cases} 1, & r_CASE_{t+1} \geq r_CASE_t \\ -1, & r_CASE_{t+1} < r_CASE_t \end{cases},$$

$$[0124] \quad r_CESE_t = \begin{cases} 1, & r_CESE_{t+1} \geq r_CESE_t \\ -1, & r_CESE_{t+1} < r_CESE_t \end{cases}$$

[0125] 其中 $r_CASE_t = \text{traffic.throughput.BitRate.}$ 为波束扇区业务在统计周期 t 内统计到业务吞吐率;

[0126] $r_CESE_t = \frac{1}{n} \times \sum_{i=1}^n \text{traffic.throughput.BitRate.LowCQI.sample } i$ 为波束扇区业务在统计周期进行采样, 获取到低信道质量指示 (Channel Quality Indicator, CQI) 时的业务吞吐率的平均速率, 其中, $\text{LowCQI} \leq 7$, 根据实际支持CQI表格决定, 取对应的正交相移键控 (Quadrature Phase Shift Keying, QPSK) 调制的CQI指示。

[0127] 在本发明的实施例中, 通过UE反馈的CQI指标反馈无线信道质量, 把信道质量量化成0~15的序列 (4bit数来承载), 并定义为CQI, 演进型基站 (Evolved Node B, eNodeB) 根据上报的CQI来决定编码方式。

[0128] 作为一个示例, 表1是4-bit CQI Table数据信息表, CQI index等于15, 对于CQI指示对应的QPSK的格式, 可等效为小区边缘环境。

[0129] 表1

CQI index	modulation	code rate x 1024	efficiency
0	out of range		
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
[0130] 4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
[0131] 12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

[0132] 作为一个示例,根据专家经验模式对波束扇区的弱覆盖特征进行识别和判断,确定特定优化策略,并生成MDP训练序列过程包括:

[0133] 根据专家经验模式在波束扇区的状态信息矩阵 X_t 中,识别和判断出覆盖特征,在该波束扇区下所有UE上报测量的服务扇区参考信号(SS_RS/CSI_RS)的平均接收功率 $X_{\text{rsrp_serving_sector}}$ 小于一定的门限 $X_{\text{min_rsrp_threshold}}$,即服务扇区参考信号平均接收功率:

$$[0134] \quad X_{\text{rsrp_serving_sector}} = \frac{1}{n} \times \sum_{i=1}^n (X_{\text{rsrp_serving_sector}_1} + X_{\text{rsrp_serving_sector}_2} + \dots + X_{\text{rsrp_serving_sector}_n}) < X_{\text{min_rsrp_threshold}}$$

[0135] 对选取的覆盖功率参数SS-RS/CSI-RS_{power}进行优化动作 α ,即 $\alpha_{\text{SS-RS/CSI-RS_power}}$,以0.5dB为调整步进,为了获取特定优化策略下的最大性能回报 r ,采用贪婪算法调整覆盖功率,覆盖功率调整范围为 $[0, P_{\text{max}} \text{ dB}]$,进而形成专家经验模式下的特定优化策略,通过在线强化学习将专家模式下的特定优化策略生成MDP训练序列,MDP训练序列为 $(S_1, \alpha_1, S_2, r_1, \dots, S_t, \alpha_t, S_{t+1}, r_t, \dots, S_{T-1}, \alpha_{T-1}, S_T, r_T)$ 。

[0136] 在本发明实施例中,门限 $X_{\text{min_rsrp_threshold}}$ 可以根据经验进行设置。

[0137] 作为另一个示例,根据专家经验模式下对波束扇区间的干扰特征进行识别和判

断,确定特定优化策略,并生成MDP训练序列过程包括:

[0138] 根据专家经验模式在波束扇区的状态信息矩阵 X_t 中,识别和判断出干扰特征,在该波束扇区下UE上报测量的服务扇区参考信号(SS_RS/CSI_RS)平均接收功率 $X_{rsrp_serving_sector}$ 与该波束扇区下所有UE上报测量的第N个最强相邻波束参考信号平均接收功率的差值小于一定的门限 D_{rsrp} ,且所有UE上报测量波束扇区的下行信号链路质量 $X_{Ni_Dl_sinr}$ 小于一定的门限 $X_{min_Dl_sinr}$,即 $(X_{rsrp_serving_sector}-X_{rsrp_neighbor_sector_N}) < D_{rsrp}$ 且 $X_{Ni_Dl_sinr} < X_{min_Dl_sinr}$ 。

[0139] 对选取的扇区水平方向半功率波瓣宽度 $horizontal_HPBW$,或,扇区垂直方向半功率波瓣宽度 $vertical_HPBW$ 进行优化动作 α ,即 $\alpha_{horizontal_HPBW}$ 或 $\alpha_{vertical_HPBW}$,以1-3度为调整步进,为了获取特定优化策略下的最大性能回报 r ,采用线性递减的贪婪算法调整波瓣宽度,调整范围为[3,65]度,进而形成专家经验模式下的特定优化策略,通过在线强化学习将专家经验模式下的特定优化策略生成MDP训练序列,MDP训练序列为 $(S_1, \alpha_1, S_2, r_1, \dots, S_t, \alpha_t, S_{t+1}, r_t, \dots, S_{T-1}, \alpha_{T-1}, S_T, r_T)$ 。

[0140] 在本发明实施例中,门限 D_{rsrp} 和门限 $X_{min_Dl_sinr}$ 可以根据经验进行设置,如, D_{rsrp} 为3dB, $X_{min_Dl_sinr}$ 为-3dB。

[0141] 作为另一个示例,根据专家经验模式对波束扇区的方向与业务UE分布偏差特征识别和判断,确定特定优化策略,并生成MDP训练序列过程包括:

[0142] 在波束扇区的状态信息矩阵 X_t 中,根据UE的业务服务质量(Quality of Service, QOS)等级和业务量需求标记方式的上行物理信号入射方向、UE上行传输时间提前量(Timing Advanced,TA)分布与Massive MIMO基站天面法线水平(垂直)面夹角 φ_{AOA} (θ_{ZOA})取均值计算,即 $\varphi_{AOA} = \frac{1}{n} \times \sum_{i=1}^n (\varphi_{AOA_1} + \varphi_{AOA_2} + \dots + \varphi_{AOA_n})$,

[0143] 确定标记有业务量UE范围,其中,QOS等级包括,11为高优先级有业务量需求的UE,10为低优先级有业务量需求的UE,01为高优先级零业务量需求的UE,00为低优先级零业务量需求的UE。对波束扇区的方位角进行优化动作,即 $\alpha_{azimuth_angle}$,为了获取特定优化策略下的最大性能回报 r ,采用贪婪算法调整波束扇区的方位角,调整范围为 $[-2 \times \varphi_{AOA}, 2 \times \varphi_{AOA}]$,对波束扇区的俯仰角进行优化动作,即 $\alpha_{elevation_angle}$,为了获取特定优化策略下的最大性能回报 r ,采用贪婪算法调整波束扇区的俯仰角,调整范围为 $[-2 \times \theta_{ZOA}, 2 \times \theta_{ZOA}]$,进而形成专家经验模式下的特定优化策略,通过在线强化学习将专家模式下的特定优化策略生成MDP训练序列,MDP训练序列为 $(S_1, \alpha_1, S_2, r_1, \dots, S_t, \alpha_t, S_{t+1}, r_t, \dots, S_{T-1}, \alpha_{T-1}, S_T, r_T)$ 。

[0144] 作为一个示例,根据每个波束扇区的特定优化策略,在线强化学习生成马克可夫决策过程MDP训练序列,并将MDP训练序列存到各自的波束扇区的强化学习网络架构的处理缓存中。

[0145] 在本发明的实施例中,专家经验模式对所述波束扇区的状态信息特征不仅仅包括弱覆盖特征、波束扇区间干扰特征、波束扇区方向与业务UE分布特征,还可以包括其他对有助于优化网络性能的状态信息特征,其他特征同样可以根据实际需求设置特征条件和优化动作,通过专家经验模式进行识别和判断,形成优化策略;在线强化学习生成多个马克可夫决策过程MDP训练序列,并将MDP训练序列存到各自的波束扇区的强化学习网络架构的处理

缓存中。

[0146] 在本发明实施例中,根据专家经验模式对Massive MIMO系统下各自不同的波束扇区的状态信息矩阵 X_t 的不同网络状态的特征进行识别和判断,并针对性地从Massive MIMO系统优化调整所涉及的天线权值和无线参数的集合中,选取相应的天线权值和无线参数,并对选取的天线权值和无线参数进行优化动作 α ,为了获取波束扇区的最大性能回报 r ,采取贪婪算法进行优化调整,进而形成专家经验模式下的特定优化策略 $\pi(\alpha_t | S_t)$;根据特定优化策略,在线强化学习生成马克可夫决策过程MDP训练序列,并将MDP训练序列存入各个波束扇区的处理缓存中,既避免了需要大量存储空间来存储历史样本数据,节约了存储空间,可以加速MM系统优化策略的探索和收敛,使Massive MIMO系统的网络性能快速进入局部最优的状态,调整效率高,减少了人工成本,并能够保证Massive MIMO系统网络稳定性。

[0147] 图4示出了本发明实施例提供的行动者-评论家深度增强学习网络的示意图。如图4所示,行动者-评论家深度增强学习网络,包括如下步骤:

[0148] S131,根据深度神经网络,构造基于行动者-评论家Actor-Critic结构的深度增强学习网络,其中,所述深度增强学习网络包括Actor网络模块和Critic网络模块,其中,Actor网络模块用于优化策略函数和调整天线权值和无线参数,所述Critic网络模块用于估计状态值函数。

[0149] S132,判断各个所述波束扇区的网络性能是否最佳。

[0150] S133,根据所述MDP训练序列,并行计算基于优势加权的 k 步 k -step时序差分误差,通过评估所述状态值函数逼近实际回报的程度来评估所述优化策略函数下的优化程度。

[0151] S134,所述Critic网络模块根据所述状态值函数,确定所述Critic网络模块的梯度信息,并将所述Critic网络模块的梯度信息反馈至所述Massive MIMO系统共享的Critic网络模块。

[0152] S135,根据所述Critic网络模块的评估标准,累计更新所述Actor网络模块的优化策略梯度,并将所述Actor网络模块的梯度信息反馈至所述Massive MIMO系统共享的Actor网络模块。

[0153] S136,所述Massive MIMO系统共享的Actor-Critic网络汇总所有所述梯度信息,并异步生成对应的权值参数。

[0154] S137,将所述权值参数复制到所述波束扇区的Actor-Critic结构的深度增强学习网络中,直至所有所述波束扇区能够学习到所述Massive MIMO系统的优化策略函数和评估能力。

[0155] 在S131中,作为一个示例,针对Massive MIMO系统下每个波束扇区,将深度神经网络作为函数逼近器,构造基于行动者-评论家(Actor-Critic)结构的深度增强学习网络,其中,深度神经网络由 n 个卷积层和 m 个全连接层组成, n 和 m 均为正整数。

[0156] 图5示出了Massive MIMO系统波束扇区优化示意图,如图5所示,行动者-评论家(Actor-Critic)结构的深度增强学习网络由Actor网络模块和Critic网络模块组成,其中,Actor网络模块用于优化策略函数 $\Pi(\alpha_t | S_t; u)$ 和相对应的优化动作(调整天线权值和无线参数),Critic网络模块用于估计状态值函数 $V(S_t; v)$,其中 u 和 v 分别为优化策略函数和状态值函数的深度增强学习网络权值参数。

[0157] 每个波束扇区都独自拥有结构相同的Actor网络模块,即拥有深度增强学习优化

策略函数 $\Pi(\mathbf{a}_t | \mathbf{S}_t; \mathbf{u})$, 和Critic网络模块, 即拥有状态值函数 $V(\mathbf{S}_t; \mathbf{v})$, 且每个波束扇区都独自拥有基于行动者-评论家 (Actor-Critic) 结构的深度增强学习网络权值参数 \mathbf{u}, \mathbf{v} 。

[0158] 每个波束扇区根据深度增强学习优化策略函数 $\Pi(\mathbf{a}_t | \mathbf{S}_t; \mathbf{u})$, 执行优化动作 \mathbf{a} 来调整天线权值和无线参数, 为了得到最大的性能回报 r , 状态值函数 $V(\mathbf{S}_t; \mathbf{v})$ 根据MM网络给出的真实回报与自己的评估差异来调整自动的评估标准, 使得评估更加精确; Actor网络模块的优化策略函数 $\Pi(\mathbf{a}_t | \mathbf{S}_t; \mathbf{u})$ 根据状态价值函数 $V(\mathbf{S}_t; \mathbf{v})$ 的评估, 调整MM网路优化策略, 选择更优的天线权值和无线参数组合, 执行优化动作 \mathbf{a} , 获得更多的MM网路性能回报。

[0159] 在本发明实施例中, 借助于专家经验模式下在线强化学习获得个马尔可夫决策过程 (MDP) 训练序列, 每个波束扇区 (Sector_i) 进行深度增强学习时, 避免随机优化动作导致 Massive MIMO网络性能的下降, 同时减少不必要的天线权值和参数组合, 加快了优化策略和网络评估的训练, 更快地获得最佳网络性能。

[0160] 在S132中, 作为一个示例, 对于MDP训练序列的每一个时刻 $(\mathbf{S}_t, \mathbf{a}_t)$, 其目标都是获得波束扇区最大化的性能回报 r , 即使波束扇区的网络性能达到最佳。判断在时刻 t , 波束扇区的网络性能是否最佳, 若是, 则后续的优化动作的累计回报 $G_{t:T} = 0$, 则保持最优的天线权值和无线参数配置。如果在时刻 t , 波束扇区的网络性能仍没有达到最佳状态, 则进入深度增强学习模式。若后续的优化动作的累计性能回报约等于状态值函数的评估值时, 即 $G_{t:T} \approx V(\mathbf{S}_t; \mathbf{v})$, 表示从状态 \mathbf{S}_t 开始, 波束扇区在优化策略函数下执行优化动作 \mathbf{a} , 直到网络性能最佳时的得到的累计性能回报。

[0161] 在S133中, 作为一个示例, 深度增强学习模式包括, 采用基于优势加权的 k -step 时序差分误差, 作为评估每个波束扇区 (Sector_i) 的天线权值和无线参数所进行优化动作 \mathbf{a} 的优势函数 A_t 。具体地, 通过状态值函数 $V(\mathbf{S}_t; \mathbf{v})$ 去逼近实际性能回报程度来评估优化动作的好坏, 根据贝尔曼方程获得优化动作的累计性能回报, 可以根据下面的表达式 (4) 来计算:

$$[0162] \quad G_t^i = \sum_{i=0}^k \gamma^i r_{t+i} + \gamma^k V(\mathbf{S}_{t+k}; \mathbf{v}) \quad (4)$$

[0163] 其中, γ 为性能回报的折扣因子, 且, $\gamma \in [0, 1]$, 当 γ 接近于0时着重瞬时回报, 当 γ 接近于1时着重长期回报。后续 k -step 的上界为 t_{\max} , $k = T(t_{\max}) - t$;

[0164] r 是波束扇区的对应的即时奖励, $r_t = \eta \times r_CASE_t + (1-\eta) \times r_CESE_t$, 其中 r_CASE_t 为小区平均吞吐效率, r_CESE_t 为小区边缘吞吐效率, η 为为小区平均吞吐效率和小区边缘吞吐量频谱效率的期望平衡的权重系数, 且 $0 \leq \eta \leq 1$ 。

$$[0165] \quad \text{其中, } r_CASE_t = \begin{cases} 1, & r_CASE_{t+1} \geq r_CASE_t \\ -1, & r_CASE_{t+1} < r_CASE_t \end{cases},$$

$$[0166] \quad r_CESE_t = \begin{cases} 1, & r_CESE_{t+1} \geq r_CESE_t \\ -1, & r_CESE_{t+1} < r_CESE_t \end{cases},$$

[0167] 其中 $r_CASE_t = \text{traffic.throughput.BitRate}$. 为波束扇区业务在统计周期 t 内统计到业务吞吐率;

[0168] $r_CESE_t = \frac{1}{n} \times \sum_{i=1}^n \text{traffic.throughput.BitRate.LowCQI.sample } i$ 为波束扇区业务在统计周期进行采样, 获取到低CQI时的业务吞吐率的平均速率, 其中, 低CQI ≤ 7 , 根据实际支持CQI表格决定, 取对应的QPSK调制的CQI指示。

[0169] 为了加速深度增强学习效率,对k-step性能回报的状态值评估进行优势加权,其加权函数为表达式(5):

$$[0170] \quad W^i = (V(S_{t+i}) / \sum_{i=0}^k V(S_{t+i})) \quad (5)$$

[0171] 基于优势加权的k-step时序差分误差的目标函数TD-target为表达式(6):

$$[0172] \quad G_t^w = \sum_{i=0}^k W^i G_t^i \quad (6)$$

[0173] 因此,基于优势加权的k-step时序差分误差作为波束扇区(Sector_i)评估调整天线权值和无线参数所进行优化动作a的优势函数A_t为:

$$[0174] \quad A_t = G_t^w - V(S_t; v) = \sum_{i=0}^k W^i \sum_{i=0}^k (\gamma^i r_{t+i} + \gamma^k V(S_{t+k}; v)) - V(S_{t+k}; v) \quad (7)$$

[0175] 对应的S_{t+k}是MDP训练序列中(S_t, A_t)的评估,对k步中的每一个时刻进行迭代以及对应的状态值,遇到最后小于长度为k的步长,进行降级处理。

[0176] 在本发明实施例中,k-step时序差分误差评估方法没有立即对下一个时刻的状态S_{t+1}进行误差评估,一方面,充分考虑Massive MIMO系统的无线网络的连续性和稳定性;另一方面,利用已经执行k-step的MDP训练序列,计算多步的回报在最优天线权值和无线参数附近快速迭代寻优,更逼近完成所有优化调整步骤的状态值评估,迭代计算量小,使对优化策略调整和状态价值评估的学习效率更高。

[0177] 在S134中,作为一个示例,对于每个波束扇区(Sector_i)通过并行计算出的基于优势加权的k-step时序差分误差A_t,迭代k-step时序差分误差评估,对应的S_{t+k}时刻是MDP中(S_t, a_t)的评估,对k步中的每个时刻进行迭代,并代入Critic网络模块中的状态值函数V的损失函数,损失函数为:

$$[0178] \quad F_V(v) = (G_t^w - V(S_t; v))^2 \quad (8)$$

[0179] 其中, $i \in (t, t+k)$,通过梯度下降算法计算,使得Critic网络模块的权值参数的梯度值 $\nabla_v F_V(v) = \nabla_v (G_t^w - V(S_t; v))^2$ 最小化,并通过累计更新的方式将每个波束扇区(Sector_i)中Critic网络模块累计梯度去更新存储的所有MM系统共享的Critic网络梯度信息dv_s,即 $dv_s \leftarrow dv_s + \nabla_v (G_t^w - V(S_t; v))^2$ 。

[0180] 在S135中,作为一个示例,将基于优势加权的k-step时序差分误差对每个波束扇区(Sector_i)评估优化动作a的优势函数A_t作为基函数,来累计更新每个波束扇区(Sector_i)的Actor网络模块中优化策略函数梯度。

[0181] 在每个波束扇区(Sector_i)的Actor网络模块中优化策略函数梯度的寻优过程中,选取优化调整的参数集合A中天线权值和无线参数的最优设置或最优组合。

[0182] 优化策略的目标函数为:

$$[0183] \quad F_\Pi(u) = \log \Pi(a_t | S_t; u) A_t = \log \Pi(a_t | S_t; u) (G_t^w - V(S_t; v)) \quad (9)$$

[0184] 通过 $i \in (t, t+k)$ 对梯度上升方式计算使Actor网络模块的权值梯度值 $\nabla_u F_\Pi(u) = \nabla_u \log \Pi(a_i | S_i; u) (G_t^w - V(S_i; v))$ 最大化,并通过累计梯度更新,将各个波束扇区(Sector_i)中Actor网络模块的累计梯度去更新所有Massive MIMO系统共享的Actor网络梯度信息du_s,即:

[0185] $du_s \leftarrow du_s + \nabla_u \log \Pi(a_i | S_i; u) (G_i^W - V(S_i; v))$ 。

[0186] 在S136中,作为一个示例,图6示出了Massive MIMO系统中多个波束扇区优化示意图,如图6所示,所有Massive MIMO系统共享的Actor-Critic网络汇总到来自不同波束扇区(Sector_i)的Actor-Critic网络权值梯度 dv_s 和 du_s ,通过深度Actor-Critic网络异步生成所有Massive MIMO系统共享的深度Actor-Critic网络的权值参数 u_s 和 v_s 。

[0187] 在S137中,作为一个示例,将所有Massive MIMO系统共享的权值参数 u_s 和 v_s 信息,复制到各个波束扇区(Sector_i)的Actor-Critic网络中,重复S133至S136,使得各个波束扇区(Sector_i)能够学习到所有的所有Massive MIMO系统共享的权值参数 u_s 和 v_s 信息系统的优化策略函数和评估能力。

[0188] 在本发明实施例中,利用基于优势加权的k-step时序差分误差估计快速迭代寻优,通过深度网络优化函数的权值信息共享模式,每个波束扇区优化策略函数和评估能力可以不用本地样本也能学习,入网的Massive MIMO系统越多,寻找最优天线权值和无线参数配置速度越快。基于优势加权的k-step优化方案与基于轮询迭代方式的优化方案的效果对比如表2所示:

[0189] 表2

[0190]

	轮询迭代	K-step
Massive MIMO天线权值和参数	10000+	5000+
n个波束扇区	n*10000+	5000+/n
m个Massive MIMO系统	m*n*10000+	5000+/(m*n)

[0191] 例如,对于一个200个波束扇区的Massive MIMO系统规模性部署区域来说,每个参数(或组合)样本100个,完成所有参数迭代至少需要16亿次的迭代,而在本发明实施例中,由于对每个波束扇区进行分布式处理,专家经验模式下参数组合迭代可减少50%,采用k-step回报寻优学习3-4次完成所有参数寻优,天线权值和无线参数配置寻优速度明显提升。

[0192] 在S140中,作为一个示例,Massive MIMO系统中的每个波束扇区(Sector_i)通过深度增强学习到的优化策略函数和评估能力,针对各自不同的网络状态,相应地采取最优的天线权值参数和网络参数配置组合,如果整个Massive MIMO系统达到最佳状态则保持参数设置,否则继续进入深度增强学习模式,继续寻求最优的天线权值参数和网络参数配置组合,使网络性能回报达到最大。

[0193] 在本发明实施中,Massive MIMO系统基于优势加权的k-step时序差分误差估计和深度网络优化策略和状态值函数权值共享和复制的方式,使Massive MIMO系统不仅在优化效率和网络评估能力可以在线同时提升,而且实现优化调整与现网性能评估的在线实时交互,实时采用最优的天线权值和无线参数配置组合,实现网络性能最佳。

[0194] 基于同样的发明构思,本发明实施例还提供了一种Massive MIMO的优化装置。如图7所示,本发明实施例提供的Massive MIMO的优化装置,包括:

[0195] 矩阵生成模块701,用于矩阵生成模块,用于为用户设备UE配置多个波束扇区,接收UE上报的测量报告MR数据和统计波束扇区性能KPI数据,并根据所述MR数据和所述KPI数据生成波束扇区状态信息矩阵;

[0196] 训练序列生成模块702,用于根据专家经验模式对所述状态信息矩阵的特征进行

识别和判断,确定优化策略,在线强化学习生成多个马克可夫决策过程MDP训练序列,并将所述MDP训练序列存入每个波束扇区的处理缓存中;

[0197] 深度增强学习模块703,用于根据异步优势加权的行动者-评论家Actor-Critic深度增强学习网络并行计算所述MDP训练序列,并将学习到的最优的优化策略函数和状态值函数的权值信息,通过全局共享模式进行更新和复制;

[0198] 优化模块704,用于根据所述行动者-评论家深度增强学习网络学习到的优化策略,确定Massive MIMO系统最优的天线权值和无线参数配置组合。

[0199] 在一个实施方式中,矩阵生成模块701具体用于:采用波束赋形的方式形成多个所述波束扇区覆盖不同空间区域的UE;

[0200] 根据所述波束扇区的参考信号接收功率强度,确定所述UE接入的波束扇区。

[0201] 在一个实施方式中,矩阵生成模块701具体用于:所述波束扇区状态信息矩阵的方程为:

[0202] $X_t = \{X_{\text{throughput}}, X_{\text{rsrp}}, X_{\text{Ni}}, X_{\text{TA}}, X_{\text{AOA}}, \dots\}^T$,

[0203] 其中, $X_{\text{throughput}}$ 为波束扇区的吞吐量效率, X_{rsrp} 为波束扇区的覆盖水平, X_{Ni} 为波束扇区的干扰水平, X_{TA} 为在该波束扇区下UE用于调整其主小区物理上行链路控制信道PUCCH/物理上行链路共享信道PUSCH/探测参考信号SRS上行发送的时间, X_{AOA} 为确定UE的波达角度, t 为时刻, $t=1,2,3\cdots T$ 。

[0204] 在一个实施方式中,训练序列生成模块702具体用于:根据专家经验模式对所述波束扇区不同状态信息进行识别和判断;

[0205] 从天线权值和无线参数的集合中选取相应的天线权值和无线参数;

[0206] 对选取的所述天线权值和无线参数进行优化;

[0207] 根据贪婪算法进行所述优化,确定优化策略。

[0208] 在一个实施方式中,训练序列生成模块702具体用于:所述天线权值和无线参数的集合的方程为:

[0209] $A = \{a_{\text{SS-RSRP/CSI-RSRP_power}}, a_{\text{azimuth_angle}}, a_{\text{elevation_angle}}, a_{\text{horizontal_HPBW}}, a_{\text{vertical_HPBW}}, a_{\text{handover_parameter}}, a_{\text{reselection_parameter}}, \dots\}$,

[0210] $a_{\text{SS-RS/CSI_RS_power}}$ 为同步信号参考信号接收功率SS-RSRP/信道信息状态参考信号接收功率CSI_RSRP功率参数优化调整, $a_{\text{azimuth_angle}}$ 为波束方位角优化调整, $a_{\text{elevation_angle}}$ 为波束俯仰角优化调整, $a_{\text{horizontal_HPBW}}$ 为波束水平波瓣宽度优化调整, $a_{\text{vertical_HPBW}}$ 为波束垂直波瓣宽度优化调整, $a_{\text{handover_parameter}}$ 为Massive MIMO小区切换参数优化调整, $a_{\text{reselection_parameter}}$ 为Massive MIMO小区重选参数的优化调整;

[0211] 其中, $a_{\text{SS-RS/CSI_RS_power}}$ 为同步信号参考信号接收功率SS-RSRP/信道信息状态参考信号接收功率CSI_RSRP功率参数优化调整, $a_{\text{azimuth_angle}}$ 为波束方位角优化调整, $a_{\text{elevation_angle}}$ 为波束俯仰角优化调整, $a_{\text{horizontal_HPBW}}$ 为波束水平波瓣宽度优化调整, $a_{\text{vertical_HPBW}}$ 为波束垂直波瓣宽度优化调整, $a_{\text{handover_parameter}}$ 为Massive MIMO小区切换参数优化调整, $a_{\text{reselection_parameter}}$ 为Massive MIMO小区重选参数的优化调整;

[0212] 所述优化策略的表达式为:

[0213] $\pi(a_t | S_t)$,

[0214] 其中, π 为优化策略函数, a_t 为天线权值和无线参数优化调整, S_t 为波束扇区的网络状态函数。

[0215] 在一个实施方式中,训练序列生成模块702具体用于:MDP训练序列的方程为:

[0216] $(S_1, a_1, S_2, r_1, \dots, S_t, a_t, S_{t+1}, r_t, \dots, S_{T-1}, a_{T-1}, S_T, r_T)$,

[0217] $(S_1, a_1, S_2, r_1, \dots, S_t, a_t, S_{t+1}, r_t, \dots, S_{T-1}, a_{T-1}, S_T, r_T)$,

[0218] 其中, r_t 为性能回报, a_t 为选取的所述天线权值和无线参数优化调整, S_t 为波束扇区的网络状态函数。

[0219] 在一个实施方式中, 训练序列生成模块702具体用于: 性能回报的方程为:

$$[0220] \quad r_t = \eta \times r_CASE_t + (1 - \eta) \times r_CESE_t,$$

[0221] 其中, r_CASE_t 为小区平均吞吐效率, r_CESE_t 为小区边缘吞吐效率, η 为小区平均吞吐效率和小区边缘吞吐量频谱效率的期望平衡的权重系数, 且 $0 \leq \eta \leq 1$ 。

[0222] 在一个实施方式中, 训练序列生成模块702具体用于: 根据专家经验模式对有助于优化网络性能的特征进行识别和判断, 确定优化策略;

[0223] 其中, 所述有助于优化网络性能的特征包括所述波束扇区的弱覆盖特征、波束扇区间干扰特征、波束扇区方向与业务UE分布特征。

[0224] 在一个实施方式中, 深度增强学习模块703具体用于: 根据深度神经网络, 构造基于行动者-评论家Actor-Critic结构的深度增强学习网络, 其中, 所述深度增强学习网络包括Actor网络模块和Critic网络模块, 其中, Actor网络模块用于优化策略函数和调整天线权值和无线参数, 所述Critic网络模块用于估计状态值函数;

[0225] 判断各个所述波束扇区的网络性能是否最佳;

[0226] 若各个所述波束扇区的网络性能是最佳, 则保持当前的天线权值和无线参数配置;

[0227] 若各个所述波束扇区的网络性能不是最佳, 则进入深度增强学习模式。

[0228] 在一个实施方式中, 深度增强学习模块703具体用于:

[0229] 根据所述MDP训练序列, 并行计算基于优势加权的k步k-step时序差分误差, 通过评估所述状态值函数逼近实际回报的程度来评估所述优化策略函数下的优化程度;

[0230] 所述Critic网络模块根据所述状态值函数, 确定所述Critic网络模块的梯度信息, 并将所述Critic网络模块的梯度信息反馈至所述Massive MIMO系统共享的Critic网络模块;

[0231] 根据所述Critic网络模块的评估标准, 累计更新所述Actor网络模块的优化策略梯度, 并将所述Actor网络模块的梯度信息反馈至所述Massive MIMO系统共享的Actor网络模块;

[0232] 所述Massive MIMO系统共享的Actor-Critic网络汇总所有所述梯度信息, 并异步生成对应的权值参数;

[0233] 将所述权值参数复制到所述波束扇区的Actor-Critic结构的深度增强学习网络中, 直至所有所述波束扇区能够学习到所述Massive MIMO系统的优化策略函数和评估能力。

[0234] 在一个实施方式中, 深度增强学习模块703具体用于: 优化策略函数的表达式为:

$$[0235] \quad \Pi(a_t | S_t; u),$$

[0236] 所述状态值函数的表达式为:

$$[0237] \quad V(S_t; v),$$

[0238] 其中, Π 为优化策略函数, u 和 v 分别为优化策略函数和状态值函数的深度增强学习网络权值参数。

[0239] 在一个实施方式中, 深度增强学习模块703具体用于: 采用基于优势加权的k步k-

step时序差分误差作为各个所述波束扇区的优势函数,所述优势函数的方程为:

$$[0240] \quad A_t = \sum_{i=0}^k W^i \sum_{i=0}^k (\gamma^i r_{t+i} + \gamma^k V(S_{t+k}; v)) - V(S_{t+k}; v)$$

[0241] 其中, W^i 为加权函数, γ 为性能回报的折扣因子,且 $\gamma \in [0, 1]$ 。

[0242] 在一个实施方式中,深度增强学习模块703具体用于:Actor网络模块的优化策略的目标函数为:

$$[0243] \quad F_{\Pi}(u) = \log \Pi(\alpha_t | S_t; u)(G_t^w - V(S_t; v)),$$

[0244] 其中, $\Pi(\alpha_t | S_t; u)$ 为优化策略函数, $V(S_t; v)$ 为状态值函数, G_t^w 为基于优势加权的k-step时序差分误差的目标函数。

[0245] 在一个实施方式中,优化模块704具体用于:判断整个Massive MIMO系统的性能回报是否达到最大;

[0246] 若是,则保持最优的天线权值和无线参数配置组合;

[0247] 若否,则进入深度增强学习模式,直到整个Massive MIMO系统的性能回报达到最大。

[0248] 另外,结合图1至图7述的本发明实施例的Massive MIMO系统优化方法和Massive MIMO系统优化装置可以由Massive MIMO系统设备来实现。图8示出了本发明实施例提供的Massive MIMO系统设备的硬件结构示意图。

[0249] Massive MIMO系统设备可以包括处理器801以及存储有计算机程序指令的存储器802。

[0250] 图8是示出能够实现根据本发明实施例的通信方法和网络服务器的计算设备的示范性硬件架构的结构图。如图8所示,计算设备800包括输入设备801、输入接口802、处理器803、存储器804、输出接口805、以及输出设备806。

[0251] 其中,输入接口802、处理器803、存储器804、以及输出接口805通过总线810相互连接,输入设备801和输出设备806分别通过输入接口802和输出接口805与总线810连接,进而与计算设备800的其他组件连接。

[0252] 具体地,输入设备801接收来自外部的输入信息,并通过输入接口802将输入信息传送到处理器803;处理器803基于存储器804中存储的计算机可执行指令对输入信息进行处理以生成输出信息,将输出信息临时或者永久地存储在存储器804中,然后通过输出接口805将输出信息传送到输出设备806;输出设备806将输出信息输出到计算设备800的外部供用户使用。

[0253] 计算设备800可以执行本申请上述的通信方法中的各步骤。

[0254] 处理器803可以是一个或多个中央处理器(英文:Central Processing Unit, CPU)。在处理器801或处理器701是一个CPU的情况下,该CPU可以是单核CPU,也可以是多核CPU。

[0255] 存储器804可以是但不限于随机存储存储器(RAM)、只读存储器(ROM),可擦除可编程只读存储器(EPROM)、光盘只读存储器(CD-ROM)、硬盘等中的一种或多种。存储器804用于存储程序代码。

[0256] 可以理解的是,在本申请实施例中,图7提供的任一模块或全部模块的功能可以用图8所示的中央处理器803实现。

[0257] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用全部或部分地以计算机程序产品的形式实现,所述计算机程序产品包括一个或多个计算机指令。在计算机上加载或执行所述计算机程序指令时,全部或部分地产生按照本发明实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL)或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输)。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质,(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质(例如固态硬盘Solid State Disk(SSD))等。

[0258] 本说明书的各个部分均采用递进的方式进行描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点介绍的都是与其他实施例不同之处。尤其,对于装置和系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例部分的说明即可。

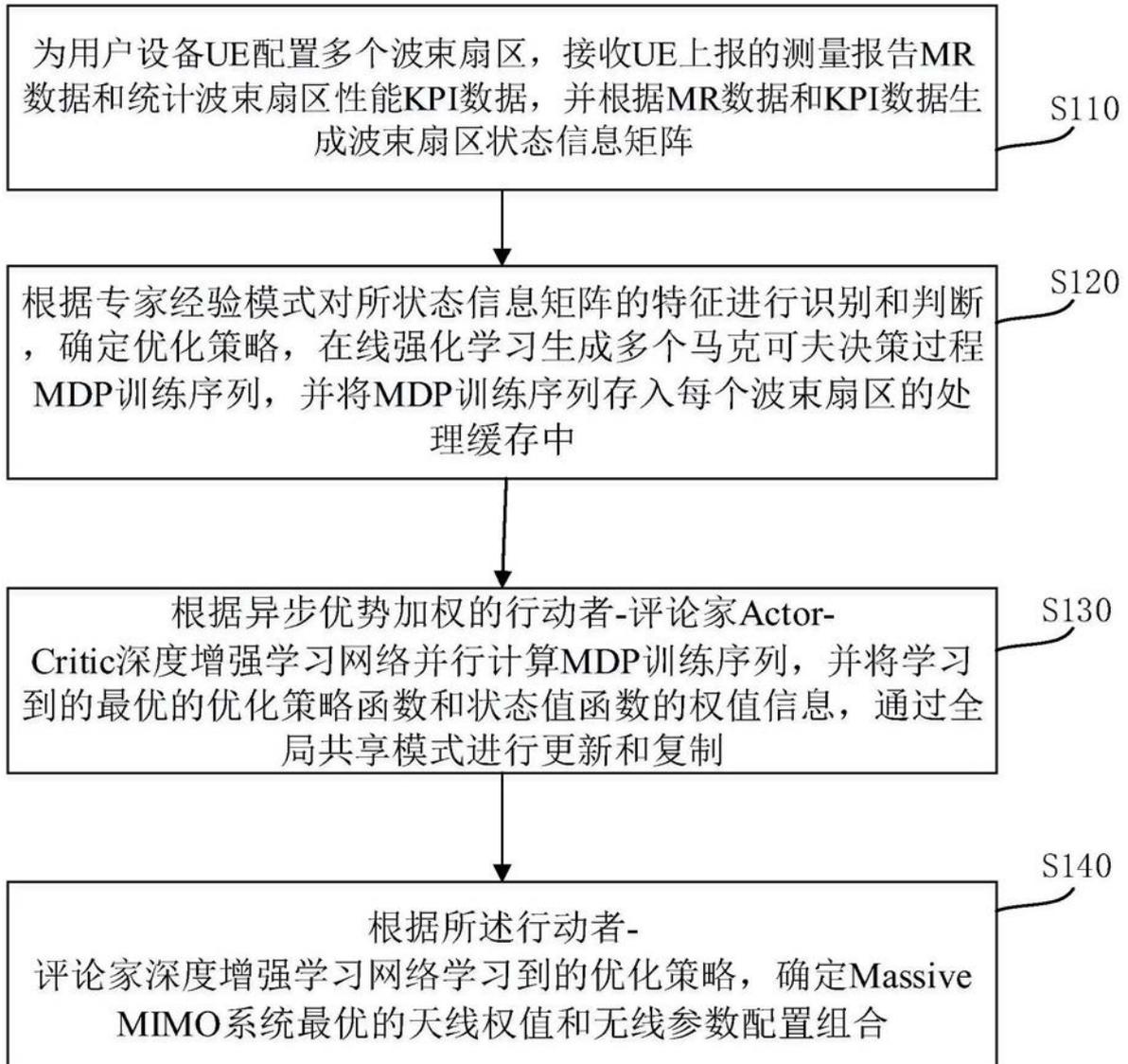


图1

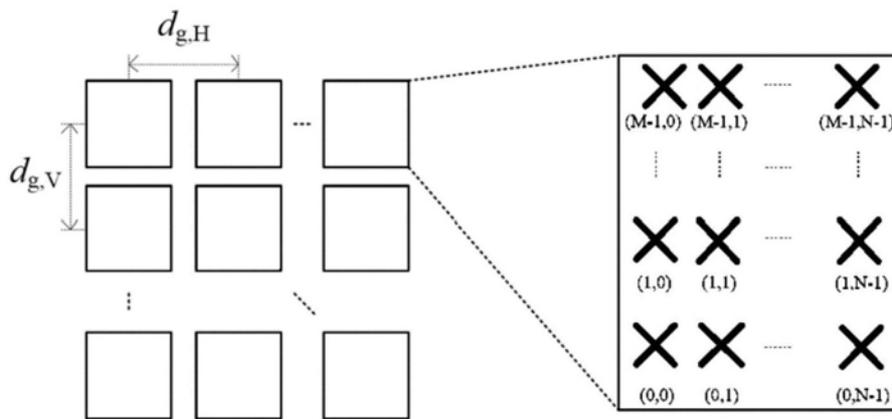


图2

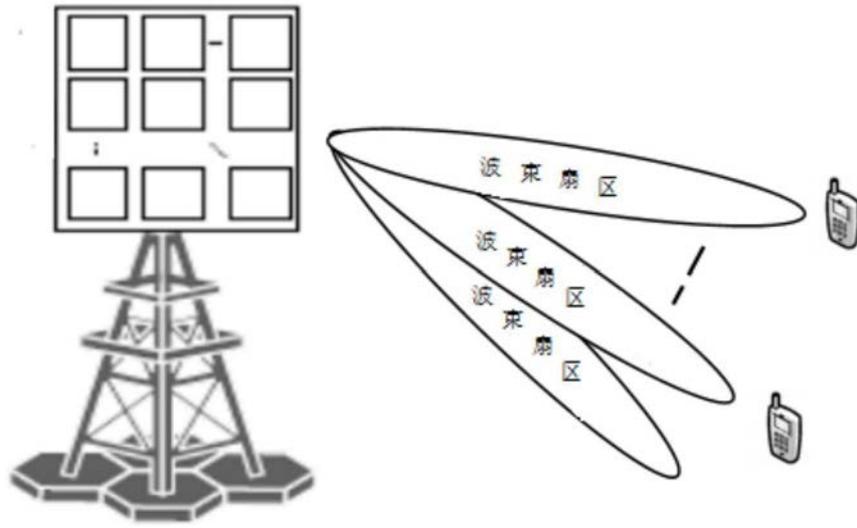


图3

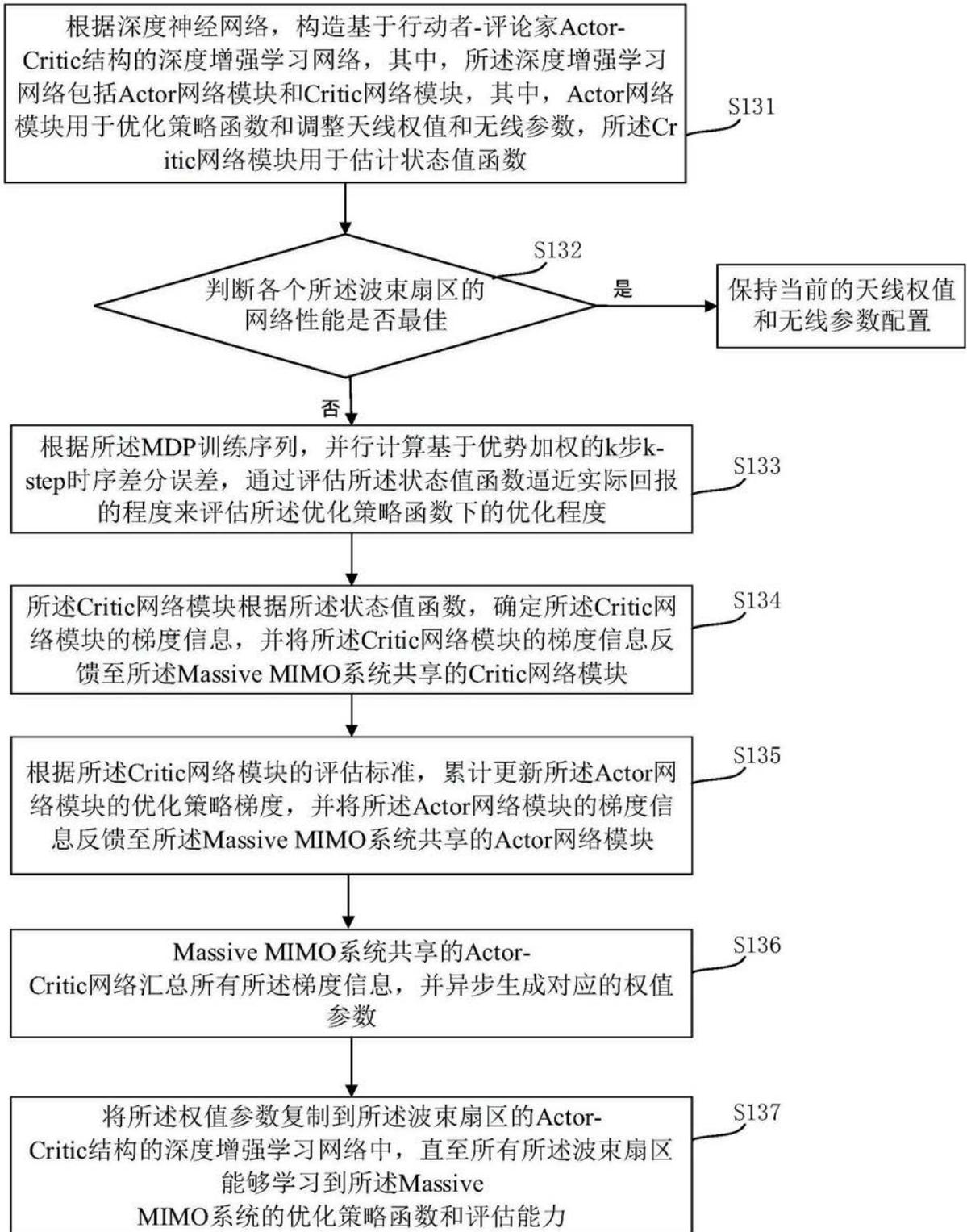


图4

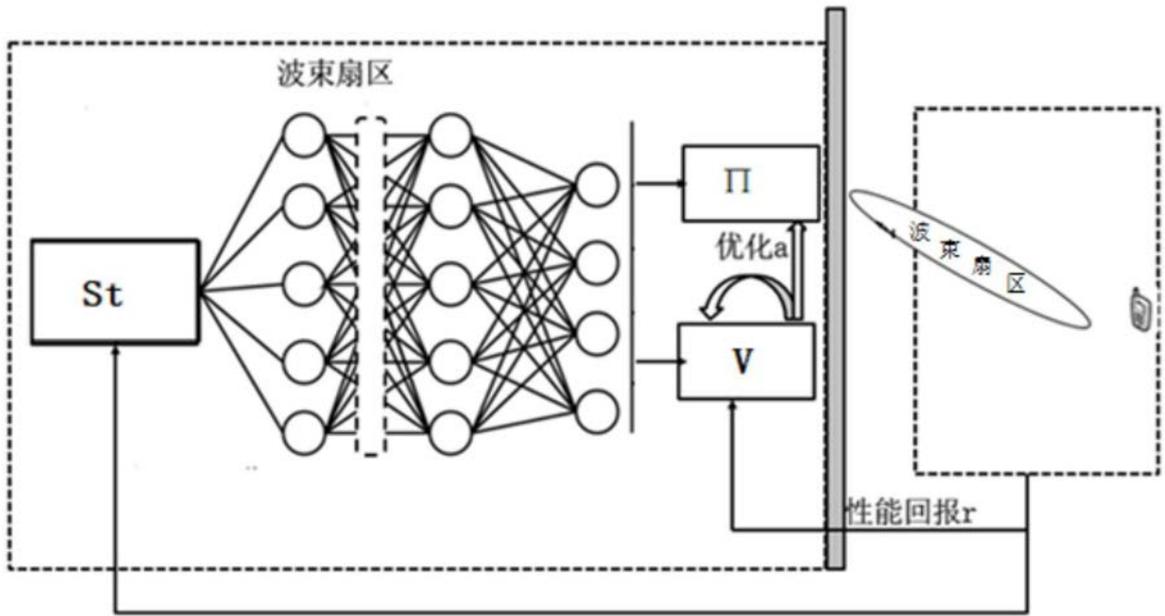


图5

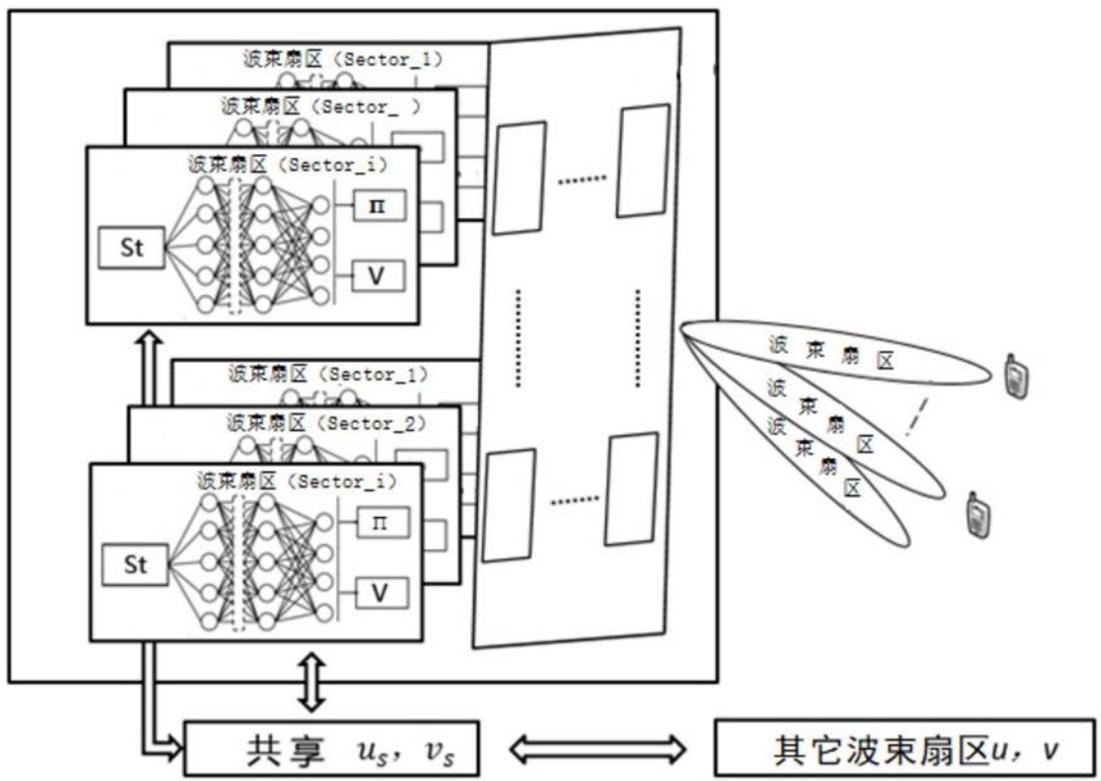


图6

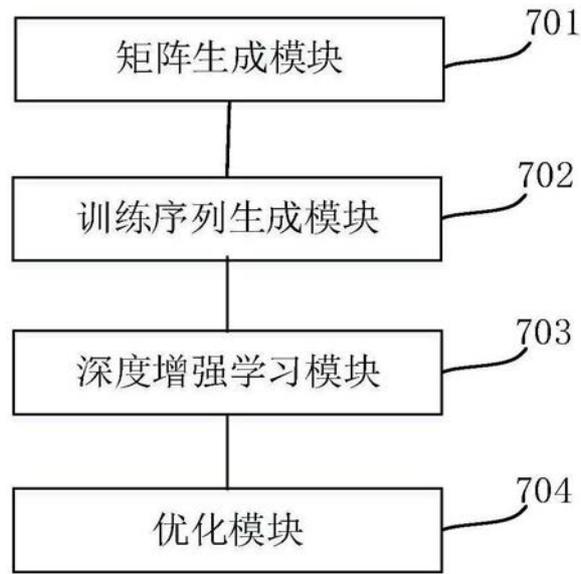


图7

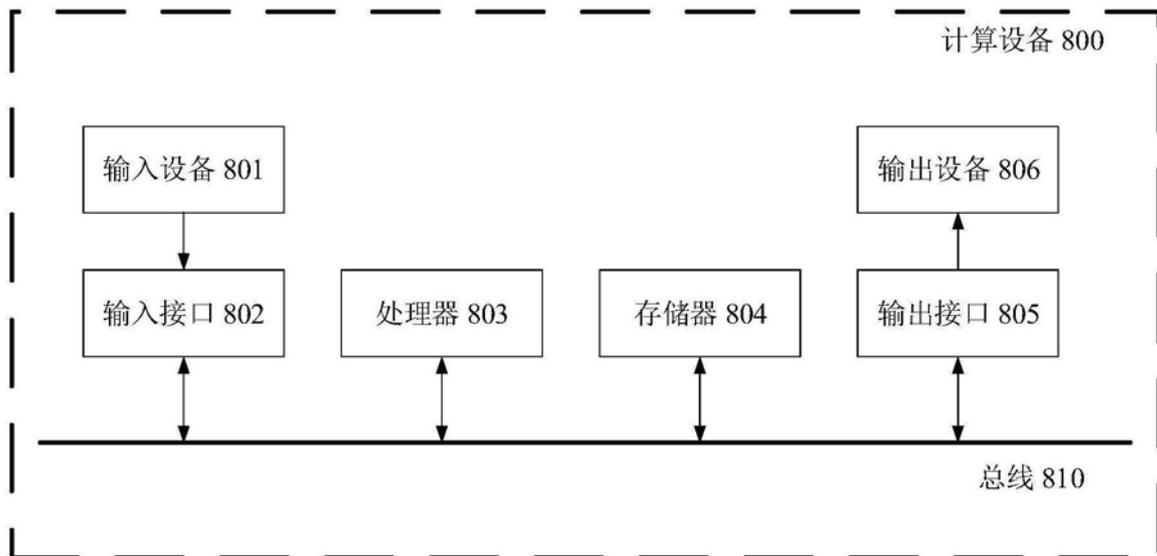


图8