



(12) 发明专利申请

(10) 申请公布号 CN 114449550 A

(43) 申请公布日 2022. 05. 06

(21) 申请号 202011224030.2

(22) 申请日 2020.11.05

(71) 申请人 中国电信股份有限公司
地址 100033 北京市西城区金融大街31号

(72) 发明人 刘心唯 李力卡 张慧嫦 王昆
何忠江 刘翼

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038
专利代理师 王玉玺

(51) Int.Cl.
H04W 24/02 (2009.01)
H04W 52/02 (2009.01)

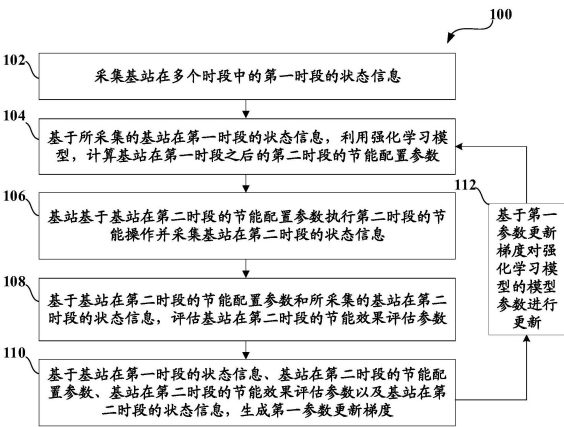
权利要求书2页 说明书10页 附图3页

(54) 发明名称

基站节能系统的参数优化方法和装置

(57) 摘要

本公开涉及一种用于优化基站的节能配置参数的方法,包括:基于采集的基站在多个时段中的第一时段的状态信息,利用第一学习模型,计算基站在第一时段之后的第二时段的节能配置参数;基于基站在第二时段的节能配置参数执行第二时段的节能操作并采集基站在第二时段的状态信息;基于基站在第二时段的节能配置参数和所采集的基站在第二时段的状态信息,评估基站在第二时段的节能效果评估参数;基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息,生成第一参数更新梯度;以及基于第一参数更新梯度对第一学习模型的模型参数进行更新。



1. 一种用于优化基站的节能配置参数的方法,包括:
采集基站在多个时段中的第一时段的状态信息;
基于所采集的基站在第一时段的状态信息,利用第一学习模型,计算基站在第一时段之后的第二时段的节能配置参数;
基于基站在第二时段的节能配置参数执行第二时段的节能操作并采集基站在第二时段的状态信息;
基于基站在第二时段的节能配置参数和所采集的基站在第二时段的状态信息,评估基站在第二时段的节能效果评估参数;
基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息,生成第一参数更新梯度;以及
基于第一参数更新梯度对第一学习模型的模型参数进行更新。
2. 根据权利要求1所述的方法,其特征在于,还包括:
基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息,对第二学习模型进行训练。
3. 根据权利要求2所述的方法,其特征在于,还包括:
利用经训练的第二学习模型,预测基站的状态信息;
基于所预测的基站的状态信息,利用第一学习模型,计算基站的节能配置参数;以及
基于经所预测的基站的状态信息得到的基站的节能配置参数,执行基站的节能操作并评估基站的节能效果评估参数。
4. 根据权利要求3所述的方法,其特征在于,还包括:
基于所预测的基站的状态信息、经所预测的基站的状态信息得到的基站的节能配置参数以及由此评估得到的节能效果评估参数,生成第二参数更新梯度;以及
基于第二参数更新梯度对第一学习模型的模型参数进行更新。
5. 根据权利要求4所述的方法,其特征在于,
第一学习模型包括价值网络模型和策略网络模型,以及
以使价值网络模型的输出值最大为基准,生成第一参数更新梯度和第二参数更新梯度。
6. 根据权利要求4所述的方法,其特征在于,
在生成一次第一参数更新梯度期间,能够生成多次第二参数更新梯度。
7. 根据权利要求1所述的方法,其特征在于,
所述状态信息包括平均下行物理资源块prb利用率、平均流量以及累计能耗中的至少一项。
8. 根据权利要求1所述的方法,其特征在于,
所述节能配置参数为节能操作的判断阈值门限;并且
所述状态信息中的每个特征对应于两个判断阈值门限,其中所述两个判断阈值门限中的至少一个为下限阈值,且所述两个判断阈值门限中的至少一个为上限阈值。
9. 根据权利要求5所述的方法,其特征在于,
所述策略网络模型通过全连接网络计算基站在第一时段之后的第二时段的节能配置参数。

10. 一种用于优化基站的节能配置参数的装置,包括:
存储器,其上存储有指令;以及
处理器,被配置为执行存储在所述存储器上的指令,以执行以根据权利要求1至9中的任一项所述的方法。

11. 一种计算机可读存储介质,包括计算机可执行指令,所述计算机可执行指令在由一个或多个处理器执行时,使得所述一个或多个处理器执行根据权利要求1至9中的任意一项所述的方法。

基站节能系统的参数优化方法和装置

技术领域

[0001] 本公开总体上涉及通信领域与人工智能领域,更具体地涉及能够对基站节能系统的参数进行优化的方法和装置。

背景技术

[0002] 随着移动通信技术的发展,现今已发展到了第五代移动通信,基站的各项性能也随之得到了提升,从而能够适应高带宽、低时延和大连结的不同需求场景。但在人们享受质量越来越高的信息服务的同时,基站侧的压力也在不断增加,其所消耗的能源也在日益剧增。由此导致了运营商每年在基站能耗上的花费也相应地在上涨。鉴于此,根据不同基站在不同时间段上具有不同程度负荷的基本特点,可以考虑在基站的负荷较低的时候对其进行节能操作从而实现在对业务几乎无太大影响的前提下,尽可能地节约基站的能耗,以降低运营商的成本。构建这样的基站节能系统一方面需要从硬件层面上扩展和提升基站的节能能力;另一方面则需要从控制层面上,让基站能自动且智能地运用这些节能能力,即建立一套智能的基站节能策略。

[0003] 然而,当前的基站节能系统大多是基于传统的数学建模构建分析模型,进而通过解最优化问题得到节能参数,最后根据节能配置参数执行节能操作,而这期间并不存在闭环效果反馈。也就是说,目前大多数的基站节能策略是基于诸如基站负载、业务量等方面的kpi的指标进行阈值门限判断来决定是否对目标基站进行节能操作,因而无法精确描述实际场景,并且所得到的节能参数存在根本性的偏差、可信度低且无法根据节能效果做持续优化。

[0004] 因此,本发明旨在提出一种改进的基站节能系统的参数优化方法和装置。

发明内容

[0005] 在下文中给出了关于本公开的简要概述,以便提供关于本公开的一些方面的基本理解。但是,应当理解,这个概述并不是关于本公开的穷举性概述。它并不是意图用来确定本公开的关键性部分或重要部分,也不是意图用来限定本公开的范围。其目的仅仅是以简化的形式给出关于本公开的某些概念,以此作为稍后给出的更详细描述的前序。

[0006] 根据本公开的一个方面,提供一种用于优化基站的节能配置参数的方法,包括:采集基站在多个时段中的第一时段的状态信息;基于所采集的基站在第一时段的状态信息,利用第一学习模型,计算基站在第一时段之后的第二时段的节能配置参数;基于基站在第二时段的节能配置参数执行第二时段的节能操作并采集基站在第二时段的状态信息;基于基站在第二时段的节能配置参数和所采集的基站在第二时段的状态信息,评估基站在第二时段的节能效果评估参数;基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息,生成第一参数更新梯度;以及基于第一参数更新梯度对第一学习模型的模型参数进行更新。

[0007] 根据本公开的另一个方面,提供一种用于优化基站的节能配置参数的装置,包括:

存储器,其上存储有指令;以及处理器,被配置为执行存储在所述存储器上的指令,以执行以根据权利要求1至9中的任一项所述的方法。

[0008] 根据本公开的又一个方面,提供一种计算机可读存储介质,包括计算机可执行指令,所述计算机可执行指令在由一个或多个处理器执行时,使得所述一个或多个处理器执行根据权利要求1至9中的任意一项所述的方法。

[0009] 通过以下参照附图对本发明的示例性实施例的详细描述,本发明的其它特征及其优点将会变得更为清楚。

附图说明

[0010] 构成说明书的一部分的附图描述了本公开的实施例,并且连同说明书一起用于解释本公开的原理。

[0011] 参照附图,根据下面的详细描述,可以更清楚地理解本公开,其中:

[0012] 图1示出了根据本公开的实施例的基站节能系统的参数优化方法;

[0013] 图2示出了根据本公开的另一实施例的基站节能系统的参数优化方法;

[0014] 图3示出了根据本公开的另一实施例的基站节能系统的参数优化方法的示意图;以及

[0015] 图4示出了可以实现根据本公开的实施例的计算设备的示例性配置。

具体实施方式

[0016] 参考附图进行以下详细描述,并且提供以下详细描述以帮助全面理解本公开的各种示例实施例。以下描述包括各种细节以帮助理解,但是这些细节仅被认为是示例,而不是为了限制本公开,本公开是由随附权利要求及其等同内容限定的。在以下描述中使用的词语和短语仅用于能够清楚一致地理解本公开。另外,为了清楚和简洁起见,可能省略了对公知的结构、功能和配置的描述。本领域普通技术人员将认识到,在不脱离本公开的精神和范围的情况下,可以对本文描述的示例进行各种改变和修改。

[0017] 在现今大数据驱动的时代,基站节能参数的确定需要由实际数据驱动,并且基站历史的各项kpi信息以及执行节能策略后的效果记录的可获取性也为闭环的节能参数优化奠定了数据基础。针对现有节能参数的确定方法无法构成闭环优化且脱离大数据分析的问题,本专利提供了一种基于深度强化学习的节能参数的闭环优化方法。该方法分析了节能闭环控制过程的各环节的数据,为节能参数的优化寻找最优的调整方向,从而使得节能效果以梯度的效率获得提升。

[0018] 接下来,将参照图1至图4对根据本发明的基站节能系统的参数优化方法和装置进行详细描述。

[0019] 数学符号说明

[0020] $state_t$:描述基站在 $t-1$ 时刻至 t 时刻的时间段内的状态的数据元组,简写为 s_t 。

[0021] $(\theta_1, \theta_2, \dots, \theta_m)_t$:描述在 t 时刻至 $t+1$ 时刻的时间段内,用于判断节能操作的判断阈值门限,简写为 θ_t 。

[0022] $reward_t$:描述在 t 时刻至 $t+1$ 时刻的时间段内,根据人为设定的量化公式,基于 s_t 计算得到的基站节能效果的量化结果,简写为 r_t 。

[0023] $\text{value}(s_t, \theta_t; w)$ 或 $\text{value}(s_t; w)$: 强化学习模型中的价值网络模型, 在输入 (s_t, θ_t) 或 s_t 后, 其输出对应的状态-动作对的价值或状态的价值, 其中 w 为该神经网络中的参数。

[0024] **$\text{policy}(s_t; \varphi)$** : 强化学习模型中的策略网络模型, 在输入 s_t 后, 其输出节能门限 θ_t , 其中 φ 为该神经网络中的参数。

[0025] $\text{model}(s_t, \theta_t)$: 监督学习模型, 输入 s_t 或 (s_t, θ_t) , 预测 s_{t+1} 。

[0026] s'_t : 监督学习模型生成的基站的数据元组, 区别于从真实环境中采集到的 s_t 。

[0027] θ'_t : 策略网络根据 s'_t 得到用于判断节能操作的判断门限, 区别于 θ_t 。

[0028] r'_t : 根据 s'_t 计算得到基站节能效果的量化结果, 区别于 r_t 。

[0029] 图1示出了根据本公开的实施例的基站节能系统的参数优化方法100。

[0030] 如图1所示。在步骤S102处, 采集基站在多个时段中的第一时段的状态信息。

[0031] 例如, 采集当前所观测到的基站的状态信息, 诸如客户感知相关的特征、业务和负荷类指标等。在根据本公开的一些实施例中, 状态信息可以包括平均下行物理资源块利用率(平均prb利用率)、平均流量以及累计能耗等。

[0032] 在根据本公开的一些实施例中, 假设基站数据的采集间隔为1小时, 并以基站在这期间的平均下行物理资源块利用率、平均流量、以及能耗这三个维度特征来描述基站的状态, 则截止至8:00时刻的过去1小时的基站状态的统计信息可以用如下数据元组来表示:

[0033] $s_{8:00} = (\text{prb}_{8:00}, \text{traffic}_{8:00}, \text{energy}_{8:00})$

[0034] 其中, $\text{prb}_{8:00}$ 表示7:00至8:00期间的平均下行物理资源块利用率, $\text{traffic}_{8:00}$ 表示7:00至8:00期间的平均流量, $\text{energy}_{8:00}$ 表示7:00至8:00期间基站的累计能耗。

[0035] 要注意的是, 虽然上述步骤以1小时为间隔来采集基站状态的观察值, 但这并不意味着对本发明的限制。例如, 采集间隔也可以为30分钟、2小时、5小时、1天、1个月等。也就是说, 采集间隔的选取可以根据具体的场景而不同。

[0036] 在步骤S104处, 基于所采集的基站在第一时段的状态信息, 利用强化学习模型, 计算基站在第一时段之后的第二时段的节能配置参数。

[0037] 在根据本公开的一些实施例中, 强化学习模型包括价值网络模型和策略网络模型。

[0038] 在根据本公开的一些实施例中, 可以基于Actor-Critic算法家族中的算法来构建强化学习模型。

[0039] 关于价值网络模型, 其可以是输入 (s_t, θ_t) 的形式, 也可以是仅输入 s_t 的形式。具体是哪种形式取决于具体选用Actor-Critic算法家族中的哪一种, 例如, A2C、A3C、PG、DPG、DDPG等。其中, 价值网络模型的一种实现方式可以通过全连接网络实现。也就是说, 只要能实现输入 s_t 或 (s_t, θ_t) , 输出一个标量的形式即可。

[0040] 关于策略网络模型, 其可以通过全连接网络实现输入 s_t , 输出 θ_t 。策略网络模型根据基站在第一时段的采集的状态信息, 计算基站在第一时段之后的第二时段的节能配置参数 $(\theta_1, \theta_2, \dots, \theta_m)$, 并输出到基站侧, 其中节能配置参数 $(\theta_1, \theta_2, \dots, \theta_m)$ 是用于基站进行节能操作判断的阈值门限。

[0041] 例如, 仍以7:00时刻开始, 1小时为间隔为例, 将 $s_{8:00}$ 输入到策略网络 **$\text{policy}(s_t; \varphi)$** 后将得到作用于8:00至9:00时段的节能配置参数, 即基站在8:00至9:00

时段执行各项节能操作的判断阈值门限 $(\theta_1, \theta_2 \cdots \theta_m)_{8:00}$, 为方便表示, 下面用 $\theta_{8:00}$ 代替。

[0042] 在根据本公开的一些实施方式中, 关于如何从 s 得到 θ , 一种做法是让 s 通过全连接网络从而输出 θ 。其数学形式可以写成如下形式:

$$[0043] \quad \theta \stackrel{\text{def}}{=} \text{policy}(s; \varphi) = \text{dense}(s; \varphi)$$

[0044] 其中, φ 表示该神经网络的参数。

[0045] 要注意的是, 在步骤S102中采集的状态信息中的每个特征 (诸如平均下行物理资源块利用率、平均流量以及累积能耗) 对应于两个判断阈值门限, 其中所述两个判断阈值门限中的至少一个为下限阈值 θ_1 , 且所述两个判断阈值门限中的至少一个为上限阈值 θ_2 , 其中 $\theta_1 \leq \theta_2$ 。也就是说, 在以平均下行物理资源块利用率、平均流量以及累积能耗作为参数维度的示例中, θ_m 为 θ_6 , 且平均下行物理资源块利用率对应于判断阈值门限 θ_1 和 θ_2 , 平均流量对应于判断阈值门限 θ_3 和 θ_4 , 累积能耗对应于判断阈值门限 θ_5 和 θ_6 。

[0046] 在步骤S106处, 基站基于基站在第二时段的节能配置参数执行第二时段的节能操作并采集基站在第二时段的状态信息。

[0047] 例如, 仍以 7:00 时刻开始, 1 小时为间隔为例, 如上所述, 通过策略网络 $\text{policy}(s_{8:00}; \varphi)$ 可以得到 $\theta_{8:00} = (\theta_1, \theta_2 \cdots \theta_m)_{8:00}$ 。在这之后, 这组配置参数将作为 8:00 至 9:00 期间基站判断是否执行某些节能操作的判断阈值门限。

[0048] 在根据本公开的一些实施例中, 节能操作包括但不限于常见的基站节能操作, 诸如符号关断, 载波关断, 通道关断等。

[0049] 以下行物理资源块利用率为例, 判断规则可以是: 如果平均下行物理资源块利用率小于 θ_1 则执行载波关断操作; 如果平均下行物理资源块利用率大于 θ_2 则停止载波关断操作; 如果平均下行物理资源块利用率介于 θ_1 与 θ_2 之间, 则保持之前的节能决策指令, 即如果平均下行物理资源块利用率从小于 θ_1 的状态转移到了介于 θ_1 与 θ_2 之间, 节能决策指令沿用小于 θ_1 时的执行载波关断的决定, 如果 prb 利用率从大于 θ_2 的状态转移到了介于 θ_1 与 θ_2 之间的状态, 则节能决策指令沿用大于 θ_2 时的停止载波关断的决定。

[0050] 也就是说, 通过 $\text{policy}(s_t; \varphi)$ 输出的 θ_t , 从而间接地控制基站在时刻 t 至时刻 $t+1$ 期间的节能操作。

[0051] 此外, 基站同时也会采集对应于这期间的状态信息 s_{t+1} 。因此, 新的状态信息 s_{t+1} 又可以作为策略网络模型的输入, 从而计算得到新的节能配置参数 θ_{t+1} 。由此, 可以持续得到节能配置参数, 从而实现基站的持续节能操作。也就是说, 上述步骤S102-S106是可以循环执行的。

[0052] 下面仍以 7:00 时刻开始, 1 小时为间隔为例简述基站执行节能操作并采集状态数据的循环过程:

[0053] 8:00 时刻, 基站采集得到状态信息 $s_{8:00}$, 并通过策略网络得到 $\theta_{8:00} = \text{policy}(s_{8:00}; \varphi)$;

[0054] 8:00 至 9:00 期间, 基站基于 $\theta_{8:00}$ 判断得到节能操作指示;

[0055] 9:00 时刻, 基站采集得到状态信息 $s_{9:00}$, 并通过策略网络得到 $\theta_{9:00} = \text{policy}(s_{9:00}; \varphi) \cdots$

[0056] 在步骤S108处,基于基站在第二时段的节能配置参数和所采集的基站在第二时段的状态信息,评估基站在第二时段的节能效果评估参数。

[0057] 在根据本公开的一些实施例中,可以根据收集得到的状态信息与节能操作数据,通过预先设定的节能效果评价机制,对所做的节能策略进行评分,即量化节能效果。所述节能效果评价机制可以根据需要按照不同的场景来设定。

[0058] 例如,仍以7:00时刻开始,1小时为间隔为例,通过上述步骤S102-S106得到了 $s_{8:00}, \theta_{8:00}, s_{9:00}, \theta_{9:00}, \dots$ 的数据序列。从而可以对节能的效果进行量化评估。例如,若希望保证流量不受到影响的前提下尽量节能,则回报的计算公式可以写成平均流量traffic与能耗energy的函数,即 $\text{reward} = g(\text{energy}, \text{traffic})$,一种设计思路可以是该函数随energy的增加而递减,随traffic的增加而增加。

[0059] 因此,通过如下回报计算公式,可以量化每个时间段的节能效果,如8:00至9:00期间的节能效果可以表示成:

[0060] $\text{reward}_{9:00} = g(\text{energy}_{9:00}, \text{traffic}_{9:00})$

[0061] 其中, $\text{reward}_{9:00}$ 表示8:00至9:00期间的节能操作效果的评估参数; $\text{energy}_{9:00}$ 表示8:00至9:00期间基站的累计能耗; $\text{traffic}_{9:00}$ 表示8:00至9:00期间的平均流量。为简化表示,在下文中,用r表示reward,用 $r_{9:00}$ 表示 $\text{reward}_{9:00}$ 。

[0062] 要注意的是,以上的节能效果量化评价方法仅仅是示例,而并不构成对本发明的限制。只要能够对节能操作的效果进行评价,可以采用任何适当的评价方法,也可以根据实际的场景选取适当的评价方法。

[0063] 在步骤S110处,基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息,生成第一参数更新梯度。

[0064] 如上所示,通过上述步骤S102-S108,可以得到多组下列形式的数据元组:

[0065] $(s_{8:00}, \theta_{8:00}, r_{8:00}, s_{9:00})$

[0066] $(s_{9:00}, \theta_{9:00}, r_{9:00}, s_{10:00})$

[0067] \dots

[0068] $(s_t, \theta_t, r_t, s_{t+1})$

[0069] 基于上述数据元组中的一个或多个,以使价值网络模型的输出值最大为基准,可以生成第一参数更新梯度 $(\Delta w, \Delta \varphi)$ 。

[0070] 在步骤S112处,基于第一参数更新梯度对强化学习模型的模型参数进行更新。

[0071] 例如,基于在步骤S110中得到的第一参数更新梯度 $(\Delta w, \Delta \varphi)$ 对价值网络模型和策略网络模型的参数进行更新,诸如对价值网络模型中的参数w和策略网络模型中的参数 φ 进行更新。

[0072] 此外,虽然未示出,但可以在利用强化学习模型计算节能配置参数之前对强化学习模型进行训练,以满足预定的节能配置参数的计算精度要求。

[0073] 此外,虽然未示出,但上述步骤S102-S112是可以迭代执行的。

[0074] 如上所述,图1所示的基站节能系统的参数优化方法100描述了利用深度强化学习算法,基于真实环境中采集的基站的状态数据,结合节能配置参数数据和节能效果进行强

化学习算法模型的参数更新,更新后的强化学习模型根据基站状态信息生成新的节能配置参数,由此构成闭环优化方法。

[0075] 此外,还可以利用监督学习模型辅助加速强化学习模型的更新来缓解强化学习模型对与真实环境交互的强依赖。

[0076] 图2示出了根据本公开的另一实施例的基站节能系统的参数优化方法200。为了清晰地示出了监督学习模型对强化学习模型的模型参数优化的辅助,图2中同样示出了图1所示的方法100中的各项步骤,但再次省略对其的描述。

[0077] 如图2所示,在步骤S202处,基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息,对监督学习模型进行训练。

[0078] 关于监督学习模型,其任务是模拟无线传播环境的特征,在输入 s_t 或 (s_t, θ_t) 之后,其预测并输出 s_{t+1} 。因此可以利用 $(s_t, \theta_t, r_t, s_{t+1})$ 来对监督学习模型进行训练。也就是说,可以基于所采集的真实环境的状态信息 s_t 对监督学习模型进行训练,直至监督学习模型预测并输出的状态信息满足精度要求。

[0079] 关于监督学习模型的选取,没有硬性的要求,只要能实现输入 s_t 或 (s_t, θ_t) ,预测并输出 s_{t+1} 即可。在根据本公开的一些实施方式中,可以基于随机森林算法、LSTM网络等来构建监督学习模型。

[0080] 在步骤S204处,利用经训练的监督学习模型,预测基站的状态信息。

[0081] 如参照步骤S202所述,在完成对监督学习模型的训练之后,可以利用经训练的监督学习模型预测基站在下一时段的状态信息。也就是说,除了基于真实环境的基站的状态信息之外,经训练的监督学习模型还可以基于预测得到的状态信息作为输入来预测下一状态信息。

[0082] 在步骤S206处,基于所预测的基站的状态信息,利用强化学习模型,计算基站的节能配置参数。

[0083] 步骤S206的操作类似于步骤S104的操作,因此省略对其的详细描述。

[0084] 在步骤S208处,基于经所预测的基站的状态信息得到的基站的节能配置参数,执行基站的节能操作并评估基站的节能效果评估参数。

[0085] 步骤S208的操作类似于步骤S106和S108的操作,因此省略对其的详细描述。

[0086] 在步骤S210处,基于所预测的基站的状态信息、经所预测的基站的状态信息得到的基站的节能配置参数以及由此评估得到的节能效果评估参数,生成第二参数更新梯度 $(\Delta w', \Delta \phi')$ 。

[0087] 步骤S210的操作类似于步骤S110的操作,因此省略对其的详细描述。

[0088] 在步骤S212处,基于第二参数更新梯度对强化学习模型的模型参数进行更新。

[0089] 如参照图1所述,基于强化学习模型可以得到第一参数更新梯度 $(\Delta w, \Delta \phi)$,而在步骤S210中,基于监督学习模型和强化学习模型的组合操作可以得到第二参数更新梯度 $(\Delta w', \Delta \phi')$ 。在下文中,将详细描述分别基于第一参数更新梯度 $(\Delta w, \Delta \phi)$ 和第二参数更新梯度 $(\Delta w', \Delta \phi')$ 来对强化学习模型的参数进行更新的过程。

[0090] 如上所述,在基站从真实环境中采集各项状态数据的过程中,还可以通过对监督

学习模型进行训练,进而基于经训练的监督模型的预测结果来模拟真实环境的反馈,从而丰富基站的状态信息数据,以增强和优化对强化学习模型的模型参数的更新。此外,基于预测得到的基站的状态信息还可以加速强化学习模型的训练。

[0091] 此外,虽然未示出,但上述步骤S204-S212是可以迭代执行的。

[0092] 另外,要注意的是,如参照图1和图2所述,由于监督学习模型的预测是可以实时执行的,因此在以预定时间段执行图1所示的根据本公开的实施例的基站节能系统的参数优化方法100期间,可以循环多次执行基于监督学习模型的状态信息预测、节能配置参数计算、节能效果评估参数评估等相关过程,从而在生成一次第一参数更新梯度($\Delta w, \Delta \varphi$)期间,可以生成多次第二参数更新梯度($\Delta w', \Delta \varphi'$)。也就是说,在预定时间段持续期间,可以基于多次生成的第二参数更新梯度($\Delta w', \Delta \varphi'$)对强化学习模型的参数进行更新。基于此,能够实现在与真实环境进行交互的同时,融合模拟情景的辅助,以及在获取下一时段的真实环境数据之前,基于从模拟情景获得的数据对模型参数持续进行更新。

[0093] 图3示出了根据本公开的另一实施例的基站节能系统的参数优化方法200的示意图。图3所示的过程是对图2所示的方法200的进一步详细说明,即对强化学习与监督学习的融合方案的进一步说明。

[0094] 如块302所示,针对通过强化学习模型得到的数据和监督学习模型得到的数据,使用相同的价值网络参数 w 和策略网络参数 φ 。换句话说,基于相同的价值网络参数 w 和策略网络参数 φ 同时进行两组不同的数据采集流程:第一组用于基站与真实环境的情景;而第二组用于基站与监督模型的模拟交互情景。

[0095] 如块304所示,通过强化学习模型基于第一组数据采集流程与真实环境发生交互并持续积累数据($s_t, \theta_t, r_t, s_{t+1}$)。

[0096] 如块306所示,在基于第一组数据采集流程与真实环境发生交互期间,通过监督学习模型基于第二数据采集流程与模拟情景发生交互并持续积累模拟操作的数据($s'_t, \theta'_t, r'_t, s'_{t+1}$)。

[0097] 如块308所示,通过数据($s_t, \theta_t, r_t, s_{t+1}$)和数据($s'_t, \theta'_t, r'_t, s'_{t+1}$)可以分别计算得到关于价值网络和策略网络的第一参数更新梯度($\Delta w, \Delta \varphi$)和第二参数更新梯度($\Delta w', \Delta \varphi'$)。从而,分别基于第一参数更新梯度($\Delta w, \Delta \varphi$)和第二参数更新梯度($\Delta w', \Delta \varphi'$)来更新(w, φ)。

[0098] 图4示出了能够实现根据本公开的实施例的计算设备400的示例性配置。

[0099] 计算设备400是能够应用本公开的上述方面的硬件设备的实例。计算设备400可以是被配置为执行处理和/或计算的任何机器。计算设备400可以是但不限制于工作站、服务器、台式计算机、膝上型计算机、平板计算机、个人数据助手(PDA)、智能电话、车载计算机或以上组合。

[0100] 如图4所示,计算设备400可以包括可以经由一个或多个接口与总线402连接或通信的一个或多个元件。总线402可以包括但不限于,工业标准架构(Industry Standard Architecture, ISA)总线、微通道架构(Micro Channel Architecture, MCA)总线、增强ISA(EISA)总线、视频电子标准协会(VESA)局部总线、以及外设组件互连(PCI)总线等。计算设

备400可以包括例如一个或多个处理器404、一个或多个输入设备406以及一个或多个输出设备408。一个或多个处理器404可以是任何种类的处理器,并且可以包括但不限于一个或多个通用处理器或专用处理器(诸如专用处理芯片)。处理器402例如可以被配置为实现如上参照图1-图3所述的方法中的各项操作。输入设备406可以是能够向计算设备输入信息的任何类型的输入设备,并且可以包括但不限于鼠标、键盘、触摸屏、麦克风和/或远程控制器。输出设备408可以是能够呈现信息的任何类型的设备,并且可以包括但不限于显示器、扬声器、视频/音频输出终端、振动器和/或打印机。

[0101] 计算设备400还可以包括或被连接至非暂态存储设备414,该非暂态存储设备414可以是任何非暂态的并且可以实现数据存储的存储设备,并且可以包括但不限于盘驱动器、光存储设备、固态存储器、软盘、柔性盘、硬盘、磁带或任何其他磁性介质、压缩盘或任何其他光学介质、缓存存储器和/或任何其他存储芯片或模块、和/或计算机可以从其中读取数据、指令和/或代码的其他任何介质。计算设备400还可以包括随机存取存储器(RAM) 410和只读存储器(ROM) 412。ROM 412可以以非易失性方式存储待执行的程序、实用程序或进程。RAM 410可提供易失性数据存储,并存储与计算设备400的操作相关的指令。计算设备400还可包括耦接至数据链路418的网络/总线接口416。网络/总线接口416可以是能够启用与外部装置和/或网络通信的任何种类的设备或系统,并且可以包括但不限于调制解调器、网络卡、红外线通信设备、无线通信设备和/或芯片集(诸如蓝牙™设备、802.11设备、WiFi设备、WiMax设备、蜂窝通信设施等)。

[0102] 根据本发明的基站节能系统的参数优化方法和装置使用基于强化学习的算法框架进行基站节能系统的配置参数寻优,并且通过整合监督学习加速了强化学习模型的模型参数更新和同步。相对于现有技术而言,本发明提出的技术方案提出了从“节能参数确定”到“执行节能操作”再到“节能效果反馈”的节能参数闭环迭代优化流程,参数的优化方向由数据作为核心驱动,可信度高;结合监督学习和强化学习模型,利用监督学习辅助加速强化学习模型的更新,缓解强化学习对与真实环境交互的强依赖;从而解决了传统的数学建模无法精确描述实际场景、所得的节能参数存在根本性的偏差、可信度低且无法根据节能效果做持续优化的问题。

[0103] 本公开可以被实现为装置、系统、集成电路和非瞬时性计算机可读介质上的计算机程序的任何组合。可以将一个或多个处理器实现为执行本公开中描述的部分或全部功能的集成电路(IC)、专用集成电路(ASIC)或大规模集成电路(LSI)、系统LSI,超级LSI或超LSI组件。

[0104] 本公开包括软件、应用程序、计算机程序或算法的使用。可以将软件、应用程序、计算机程序或算法存储在非瞬时性计算机可读介质上,以使诸如一个或多个处理器的计算机执行上述步骤和附图中描述的步骤。例如,一个或多个存储器以可执行指令存储软件或算法,并且一个或多个处理器可以关联执行该软件或算法的一组指令,以根据本公开中描述的实施例提供各种功能。

[0105] 软件和计算机程序(也可以称为程序、软件应用程序、应用程序、组件或代码)包括用于可编程处理器的机器指令,并且可以以高级过程性语言、面向对象编程语言、功能性编程语言、逻辑编程语言或汇编语言或机器语言来实现。术语“计算机可读介质”是指用于向可编程数据处理提供机器指令或数据的任何计算机程序产品、装置或设备,例如磁盘、光

盘、固态存储设备、存储器和可编程逻辑设备 (PLD)，包括将机器指令作为计算机可读信号来接收的计算机可读介质。

[0106] 举例来说，计算机可读介质可以包括动态随机存取存储器 (DRAM)、随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦只读存储器 (EEPROM)、紧凑盘只读存储器 (CD-ROM) 或其他光盘存储设备、磁盘存储设备或其他磁性存储设备，或可以用于以指令或数据结构的形式携带或存储所需的计算机可读程序代码以及能够被通用或专用计算机或通用或专用处理器访问的任何其它介质。如本文中所使用的，磁盘或盘包括紧凑盘 (CD)、激光盘、光盘、数字多功能盘 (DVD)、软盘和蓝光盘，其中磁盘通常以磁性方式复制数据，而盘则通过激光以光学方式复制数据。上述的组合也包括在计算机可读介质的范围内。

[0107] 提供本公开的主题作为用于执行本公开中描述的特征的装置、系统、方法和程序的示例。但是，除了上述特征之外，还可以预期其他特征或变型。可以预期的是，可以用可能代替任何上述实现的技术的任何新出现的技术来完成本公开的部件和功能的实现。

[0108] 另外，以上描述提供了示例，而不限制权利要求中阐述的范围、适用性或配置。在不脱离本公开的精神和范围的情况下，可以对所讨论的元件的功能和布置进行改变。各种实施例可以适当地省略、替代或添加各种过程或部件。例如，关于某些实施例描述的特征可以在其他实施例中结合。

[0109] 另外，在本公开的描述中，术语“第一”、“第二”、“第三”等仅用于描述目的，而不能理解为指示或暗示相对重要性和顺序。

[0110] 类似地，虽然在附图中以特定次序描绘了操作，但是这不应该被理解为要求以所示的特定次序或者以顺序次序执行这样的操作，或者要求执行所有图示的操作以实现所希望的结果。在某些情况下，多任务处理和并行处理可以是有利的。

[0111] 另外，本公开的实施方式还可以包括以下示例：

[0112] (1) 一种用于优化基站的节能配置参数的方法，包括：

[0113] 采集基站在多个时段中的第一时段的状态信息；

[0114] 基于所采集的基站在第一时段的状态信息，利用第一学习模型，计算基站在第一时段之后的第二时段的节能配置参数；

[0115] 基于基站在第二时段的节能配置参数执行第二时段的节能操作并采集基站在第二时段的状态信息；

[0116] 基于基站在第二时段的节能配置参数和所采集的基站在第二时段的状态信息，评估基站在第二时段的节能效果评估参数；

[0117] 基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息，生成第一参数更新梯度；以及

[0118] 基于第一参数更新梯度对第一学习模型的模型参数进行更新。

[0119] (2) 根据 (1) 所述的方法，其特征在于，还包括：

[0120] 基于基站在第一时段的状态信息、基站在第二时段的节能配置参数、基站在第二时段的节能效果评估参数以及基站在第二时段的状态信息，对第二学习模型进行训练。

[0121] (3) 根据 (2) 所述的方法，其特征在于，还包括：

[0122] 利用经训练的第二学习模型，预测基站的状态信息；

[0123] 基于所预测的基站的状态信息，利用第一学习模型，计算基站的节能配置参数；以

及

[0124] 基于经所预测的基站的状态信息得到的基站的节能配置参数,执行基站的节能操作并评估基站的节能效果评估参数。

[0125] (4) 根据(3)所述的方法,其特征在于,还包括:

[0126] 基于所预测的基站的状态信息、经所预测的基站的状态信息得到的基站的节能配置参数以及由此评估得到的节能效果评估参数,生成第二参数更新梯度;以及

[0127] 基于第二参数更新梯度对第一学习模型的模型参数进行更新。

[0128] (5) 根据(4)所述的方法,其特征在于,

[0129] 第一学习模型包括价值网络模型和策略网络模型,以及

[0130] 以使价值网络模型的输出值最大为基准,生成第一参数更新梯度和第二参数更新梯度。

[0131] (6) 根据(4)所述的方法,其特征在于,

[0132] 在生成一次第一参数更新梯度期间,能够生成多次第二参数更新梯度。

[0133] (7) 根据(1)所述的方法,其特征在于,

[0134] 所述状态信息包括平均下行物理资源块prb利用率、平均流量以及累计能耗中的至少一项。

[0135] (8) 根据(1)所述的方法,其特征在于,

[0136] 所述节能配置参数为节能操作的判断阈值门限;并且

[0137] 所述状态信息中的每个特征对应于两个判断阈值门限,其中所述两个判断阈值门限中的至少一个为下限阈值,且所述两个判断阈值门限中的至少一个为上限阈值。

[0138] (9) 根据(5)所述的方法,其特征在于,

[0139] 所述策略网络模型通过全连接网络计算基站在第一时段之后的第二时段的节能配置参数。

[0140] (10) 一种用于优化基站的节能配置参数的装置,包括:

[0141] 存储器,其上存储有指令;以及

[0142] 处理器,被配置为执行存储在所述存储器上的指令,以执行以根据(1)至(9)中的任一项所述的方法。

[0143] (11) 一种计算机可读存储介质,包括计算机可执行指令,所述计算机可执行指令在由一个或多个处理器执行时,使得所述一个或多个处理器执行根据(1)至(9)中的任一项所述的方法。

[0144] 虽然已经通过示例对本公开的一些特定示例性实施例进行了详细说明,但是本领域的技术人员应该理解,以上示例仅是为了进行说明,而不是为了限制本公开的范围。在此公开的各示例性实施例可以任意组合,而不脱离本公开的精神和范围。本领域的技术人员还应理解,可以对示例性实施例进行多种修改而不脱离本公开的范围和精神。本公开的范围由所附权利要求来限定。

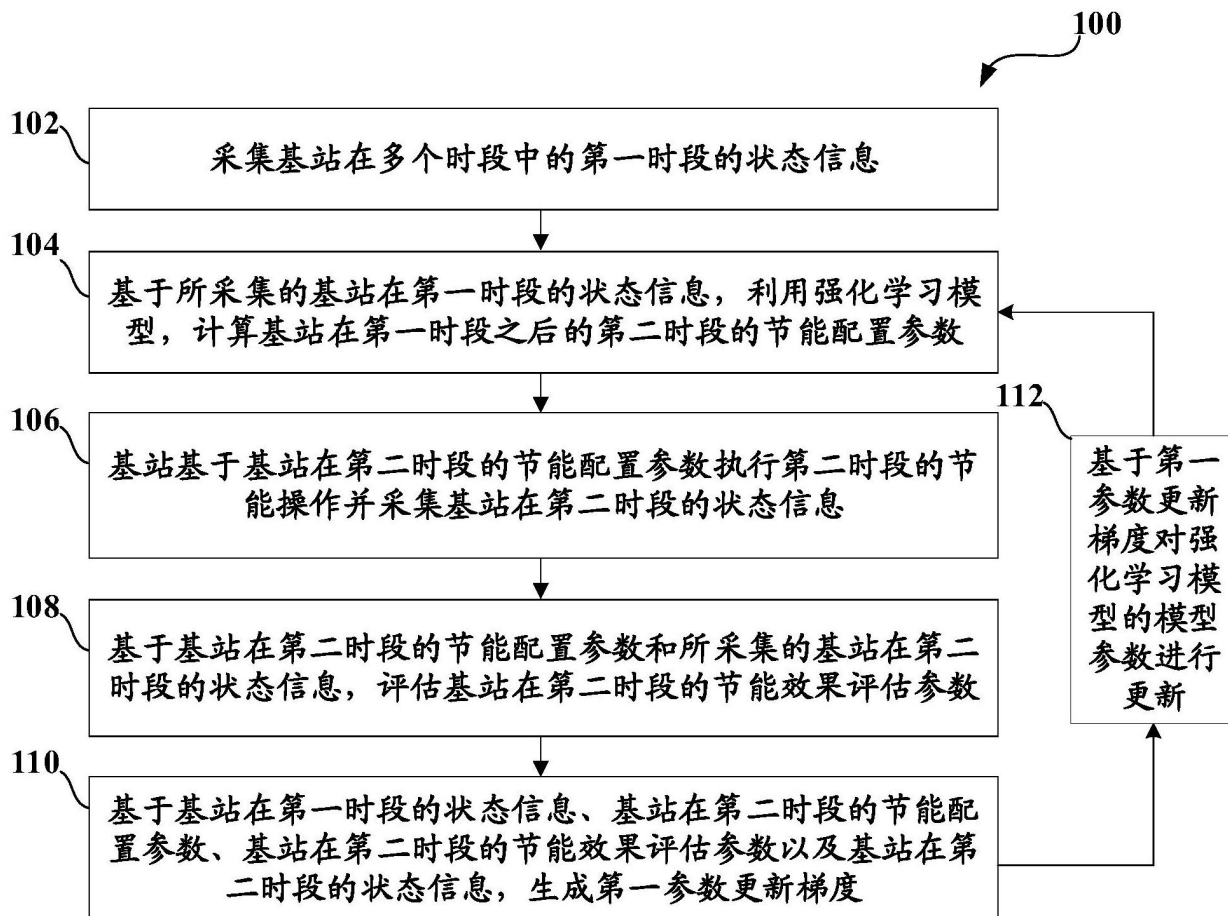


图1

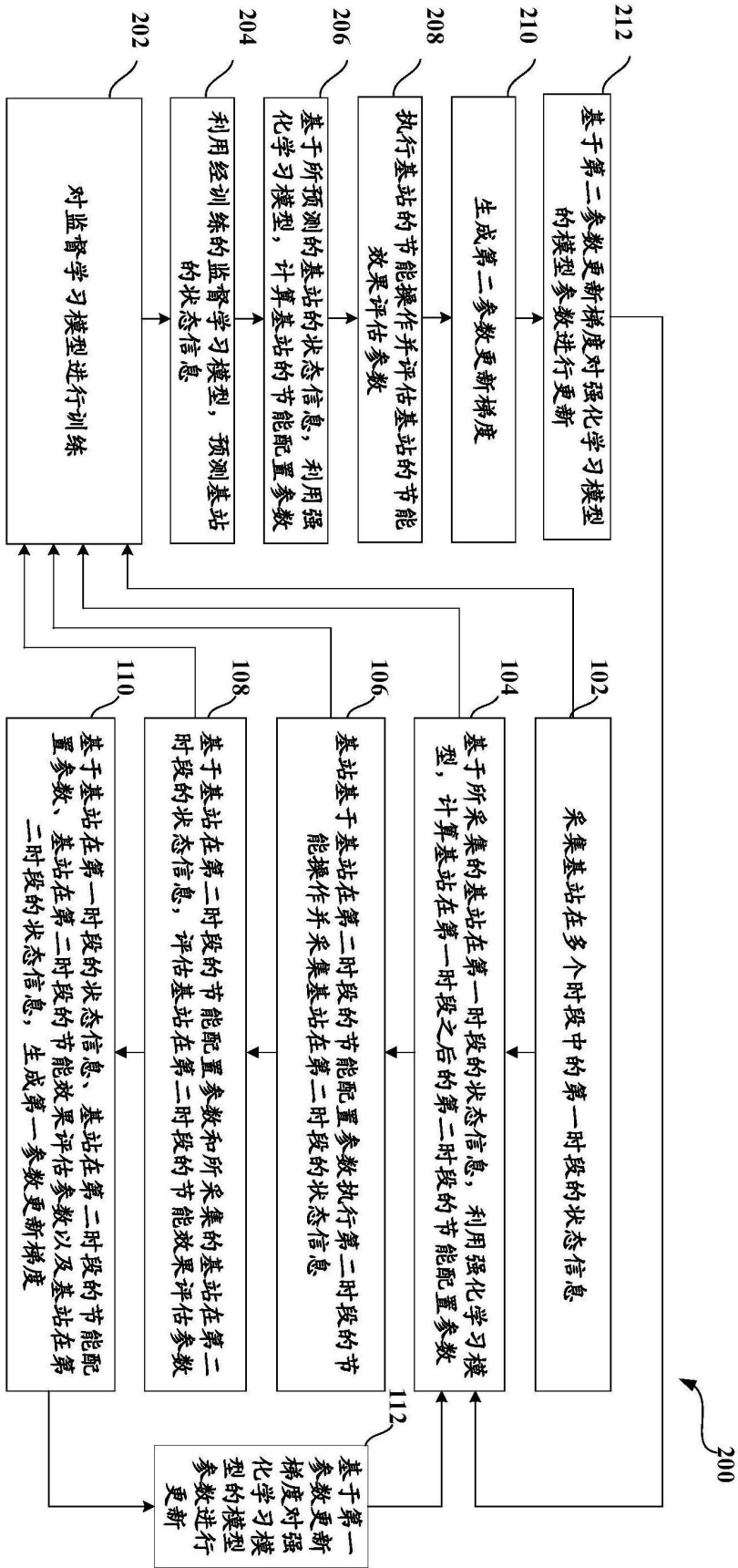


图2

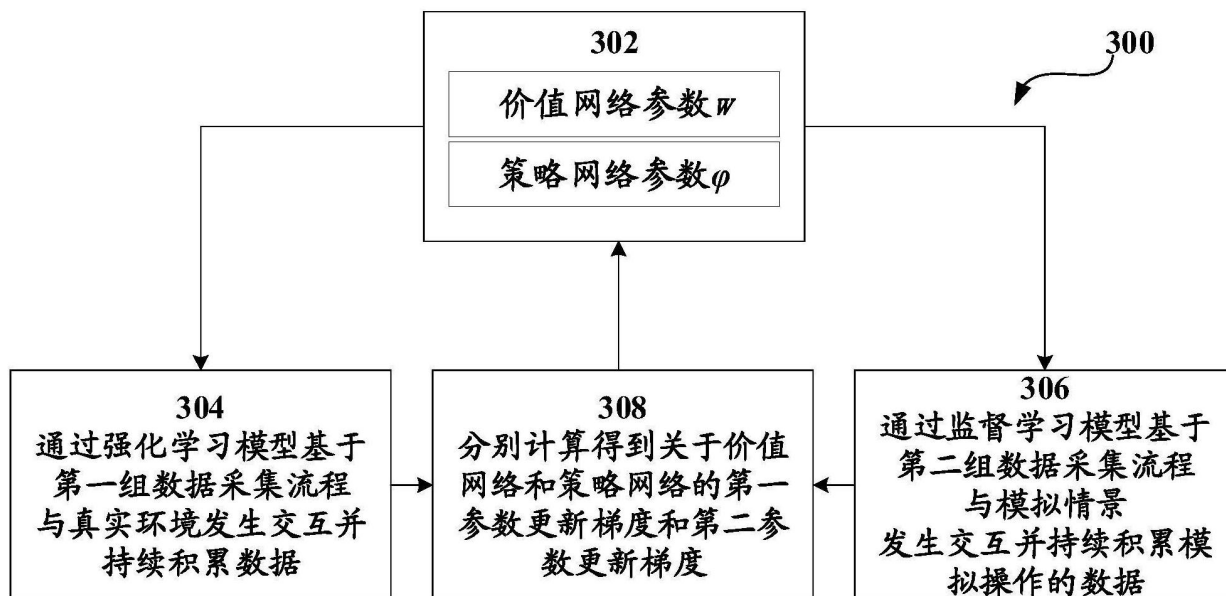


图3

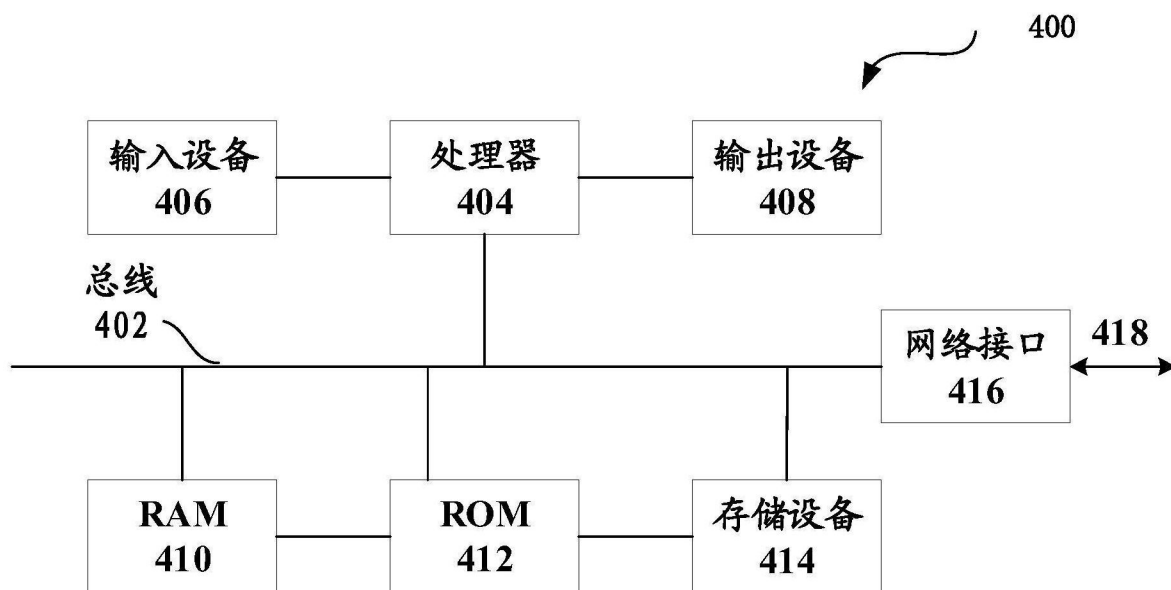


图4