(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113627533 B
(45)授权公告日 2023.11.10

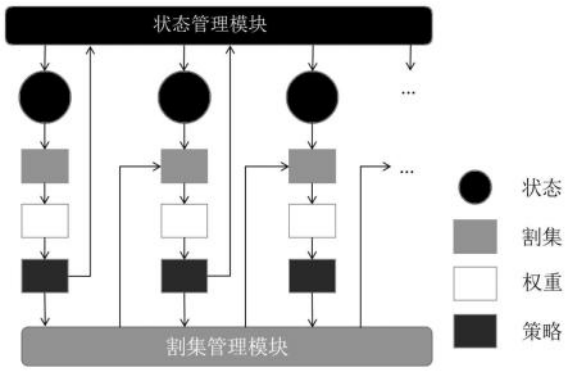
(21)申请号 202110920543.5
(22)申请日 2021.08.11
(65)同一申请的已公布的文献号
申请公布号 CN 113627533 A
(43)申请公布日 2021.11.09
(73)专利权人 北京邮电大学
地址 100876 北京市海淀区西土城路10号
专利权人 国网江苏省电力有限公司电力科学研究院
国网江苏省电力有限公司
国家电网有限公司
(72)发明人 李睿凡 王泽元 杜一帆 熊永平 刘子全
(74)专利代理机构 北京挺立专利事务所(普通合伙) 11265
专利代理师 高福勇
(51)Int.Cl.
G06Q 10/20(2023.01)
G06Q 10/0637(2023.01)
G06Q 50/06(2012.01)
G06N 3/092(2023.01)
G06N 7/01(2023.01)
(56)对比文件
CN 113036772 A,2021.06.25
WO 2021135554 A1,2021.07.08
朱德文.电梯群控动态配置的强化学习简化算法.中国电梯.2020,(第02期),全文.
审查员 冷凝
权利要求书3页 说明书8页 附图2页

(54)发明名称

一种基于强化学习的电力设备检修决策生成方法

(57)摘要

一种基于强化学习的电力设备检修决策生成方法涉及电力设备检修技术领域,解决了现有基于强化学习的建模策略的方式需要大量的数据且数据利用率不高的问题,包括:计算第一割集并据此计算电力设备引起电网停电损失的第一权重;将电力设备检修决策生成问题建模为一个马尔可夫决策过程,定义电力设备的运行状态;应用强化学习方法求解马尔可夫决策过程得到最优策略和最优策略的价值矩阵,第一权重加权到强化学习的电网的整体运行损失中,强化学习以最小化电网的整体运行损失为目标;计算第二割集并据此计算第二权重,加权到电网的整体运行损失中,改进最优策略。本发明够间接实现多个电力设备之间的通信,数据利用率高,在专业领域上的应用门槛较低。



CN 113627533 B

1.一种基于强化学习的电力设备检修决策生成方法,其特征在于,包括如下步骤:

步骤一、根据电网中所有电力设备的连接与导通情况,计算能够使当前电网停电的所有第一割集;根据第一割集计算电力设备引起电网停电损失的静态权重;

步骤二、将电力设备检修决策生成问题建模为一个马尔可夫决策过程,根据电网中所有电力设备的信息,定义电力设备有可能出现的若干个表示电力设备损坏程度的运行状态,若干个表示电力设备损坏程度的运行状态构成马尔可夫决策过程的状态集合;应用强化学习方法求解马尔可夫决策过程得到最优策略和最优策略的价值矩阵,静态权重加权到强化学习的电网的整体运行损失中,强化学习以最小化被加权静态权重的电网的整体运行损失为目标;

步骤三、根据电网待检修时其中所有电力设备的连接与导通情况,计算能够使当前电网停电的所有第二割集,根据第二割集计算电力设备引起电网停电损失的动态权重;静态权重加权到电网的整体运行损失中,以最小化被加权动态权重的电网的整体运行损失为目标,利用被加权动态权重的电网的整体运行损失和步骤二得到的价值矩阵改进最优策略,选取电网的整体运行损失最小的动作作为待检修电网的最终检修策略;

所述步骤二中应用强化学习方法求解马尔可夫决策过程得到最优策略和最优策略的价值矩阵包括如下步骤:

步骤2.1、初始化电力设备的价值矩阵V得到初始化的价值矩阵,初始化电力设备的策略π得到初始化的策略;

步骤2.2、利用静态权重对电网的整体运行损失进行加权,以最小化被静态权重加权的电网的整体运行损失为目标,根据被静态权重加权的电网的整体运行损失、初始化的策略和初始化的价值矩阵,利用贝尔曼方程更新价值矩阵;

步骤2.3、以最小化被静态权重加权的电网的整体运行损失为目标,根据被静态权重加权的电网的整体运行损失和最新的价值矩阵、利用贪心算法更新策略;以最小化被静态权重加权的电网的整体运行损失为目标,根据被静态权重加权的电网的整体运行损失和最新的策略,利用贝尔曼方程更新价值矩阵;

步骤2.4、判断是否实现被静态权重加权的电网的整体运行损失的最小化,若实现了电网的整体运行损失的最小化则步骤二完成,否则返回步骤2.3;

所述步骤三中改进最优策略包括如下步骤:

步骤3.1、利用动态权重对电网的整体运行损失进行加权,以最小化被动态权重加权的电网的整体运行损失为目标,根据被动态权重加权的电网的整体运行损失、最优策略和最优策略的价值矩阵,利用贝尔曼方程更新最优策略的价值矩阵;

步骤3.2、以最小化被动态权重加权的电网的整体运行损失为目标,根据被动态权重加权的电网的整体运行损失和最新的最优策略的价值矩阵、利用贪心算法更新最优策略;以最小化被动态权重加权的电网的整体运行损失为目标,根据被动态权重加权的电网的整体运行损失和最新的最优策略,利用贝尔曼方程更新最优策略的价值矩阵;

步骤3.3、判断是否实现被动态权重加权的电网的整体运行损失的最小化,若实现了被动态权重加权的电网的整体运行损失的最小化则选取电网的整体运行损失最小的动作作为待检修电网的最终检修策略,否则返回步骤3.2。

2.如权利要求1所述的一种基于强化学习的电力设备检修决策生成方法,其特征在于,

所述步骤一还包括获取电网中电力设备的检修损失R_M和损坏损失R_F的步骤，电网中电力设备的检修损失包括电力设备检修引起的电力设备个体的经济损失$R_{M,1}$和电力设备检修引起电网停电的经济损失$R_{M,2}$，电网中电力设备的损坏损失包括电力设备损坏引起的电力设备个体的经济损失$R_{F,1}$和电力设备损坏引起电网停电的经济损失$R_{F,2}$，步骤二所述的电网的整体运行损失为：

$$R_{sum}=R_{M,1}+R_{M,2}+\sum_{t=1}^{N_T}\left(R_{F,1}(t)+R_{F,2}(t)\right)$$

其中，t表示运行时刻，N_T表示电网运行总时长。

3.如权利要求1所述的一种基于强化学习的电力设备检修决策生成方法，其特征在于，所述第一割集和第二割集均为点割集，电力设备作为点割集的元素，点割集是指当点割集内的所有电力设备损坏时电网会出现停电，且不存在点割集的真子集内的所有电力设备损坏时电网会出现停电。

4.如权利要求1所述的一种基于强化学习的电力设备检修决策生成方法，其特征在于，所述电网的整体运行损失为：

$$R^{a}_{s_{device_i}}(cut)=-\left(R_M(s,a)+W_{device_i}(cut)\times R_F(s,a)\right)$$

其中，cut表示第一割集或第二割集，W_{device_i}表示通过第一割集或第二割集计算得到的设备i的权重，$R_M(s,a)$表示状态s动作a时电网中电力设备的检修损失，$R_F(s,a)$表示状态s动作a时电网中电力设备的损坏损失，$R^{a}_{s_{device_i}}(cut)$表示基于cut的在状态s动作a时的电网的整体运行损失。

5.如权利要求4所述的一种基于强化学习的电力设备检修决策生成方法，其特征在于，所述$W_{device_i}(cut)=\frac{1}{N}\sum_{m=1}^{N}\frac{1}{T_m}I(cut_m,device_i)$

其中，cut_m表示第m个割集，$device_i$表示电力设备i，i和m均为正整数，T_m表示当前cut中电力设备的总数量，N表示cut的个数，$I(cut_m,device_i)$的计算方式如下：

$$I(cut_m,device_i)=\begin{cases}0 & device_i \in cut_m\\ 1 & device_i \notin cut_m\end{cases}。$$

6.如权利要求4所述的一种基于强化学习的电力设备检修决策生成方法，其特征在于，步骤2.3中所述利用贝尔曼方程更新价值矩阵和步骤3.2中所述利用贝尔曼方程更新最优策略的价值矩阵的更新公式均为：

$$v_{new}(s)=\sum_{a\in A}\pi\left(a\middle|s_{device_i}\right)\left(R^{a}_{s_{device_i}}(cut)+\gamma\sum_{s'\in S}P^{a}_{ss'}v(s')\right)\quad(5)$$

再令$v()=v_{new}()$，用于公式(4)和下一次执行公式(5)；

步骤2.3中所述利用贪心算法更新策略和步骤3.2中所述利用贪心算法更新最优策略的更新公式均为：

$$\pi(\cdot|\cdot)=\pi\left(a\middle|s_{device_i}\right)=\arg\max_{a}\left(R^{a}_{s_{device_i}}(cut)+\gamma\sum_{s'\in S}P^{a}_{ss'}v(s')\right) \quad (4)$$

其中，A表示动作a的动作集合，γ表示折扣因子，S表示有限的状态集合，s和s'均属于S，s表示当前状态，状态s'表示转移后的状态，S_{device_i}是指设备i的当前状态s，$P^{a}_{ss'}$表示当前状态s到转移后的状态s'并采取动作a的概率转移矩阵，v()表示当前时间步下的价值矩阵，v_{new}()表示更新后的价值矩阵。

一种基于强化学习的电力设备检修决策生成方法

技术领域

[0001] 本发明涉及电力设备检修技术领域，具体涉及一种基于强化学习的电力设备检修决策生成方法。

背景技术

[0002] 电力设备检修是指在电路运行过程中，对设备进行检修以确保设备的良好运行状态，从而避免设备损坏对电网的运行产生较大的影响。现阶段电力设备检修的策略往往是通过人工决策，技术人员往往对电力设备管理有丰富的经验，并且有很多电力专业知识，其通过对设备状态进行评分来主观判断是否其需要检修。大部分工作主要是利用在线检测、离线检测和定期解体检测方法进行人工决策，然而，以上策略都是人工策略，往往要求技术人员需要丰富的经验，而且很多策略都是类似“制定检修策略”、“完善技能培训”等针对人的策略，不仅低效而且迁移能力很差。除了通过人工决策得到的检修的策略，目前电力设备检修方法还根据设备的运行状态进行评分，并根据检修分数对设备排序进行检修。基于人工的变电检修决策方法需要很多的专业知识并且对电路的运行状态有比较多的了解，基于设备评分的方法，往往也需要人工制定分数，并且迁移能力较差。

[0003] 电力设备检修可以理解为一个序列决策问题，利用强化学习建模检修策略和奖励的关系。强化学习受到动物学习中试错法的启发，将智能体(即电力设备)与环境交互得到的奖励值作为反馈信号对智能体进行训练，具有强大的探索能力和自主学习能力。多智能体强化指的是整个系统包含多于两个的智能体，并且智能体间存在一定关系，其应用于多个领域中，而且往往其可以得到超越人类经验的策略。多智能体强化学习已经在多个领域得到应用，比如，将深度Q网络应用在多智能体强化学习方法中，来指导服务云多工作流调度，达到了优化多工作流完成时间以及用户花费的效果；利用多智能体深度强化学习优化污水强化学习流程，将智能体与环境交互得到的奖励值作为反馈信号对智能体进行训练；利用协同多智能体自动构建对冲策略来减少投资组合中给定一组股票的损失风险。现今一些研究将电力设备检修与深度强化学习结合，设计了一种深度递归Q网络多智能体强化学习模型，其具有更好的优化与决策能力，而且维护成本很低，但目前流行的基于深度学习的强化学习通过神经网络表示状态并建模策略，需要大量的数据，且数据利用率不高，在专业领域上的应用门槛较高。

发明内容

[0004] 为了解决现有基于强化学习的建模策略的方式需要大量的数据且数据利用率不高的问题，本发明提供一种基于强化学习的电力设备检修决策生成方法。

[0005] 本发明为解决技术问题所采用的技术方案如下：

[0006] 一种基于强化学习的电力设备检修决策生成方法，包括如下步骤：

[0007] 步骤一、根据电网中所有电力设备的连接与导通情况，计算能够使当前电网停电的所有第一割集；根据第一割集计算电力设备引起电网停电损失的静态权重；

[0008] 步骤二、将电力设备检修决策生成问题建模为一个马尔可夫决策过程，根据电网中所有电力设备的信息，定义电力设备有可能出现的若干个表示电力设备损坏程度的运行状态，若干个表示电力设备损坏程度的运行状态构成马尔可夫决策过程的状态集合；应用强化学习方法求解马尔可夫决策过程得到最优策略和最优策略的价值矩阵，静态权重加权到强化学习的电网的整体运行损失中，强化学习以被静态权重加权的最小化电网的整体运行损失为目标；

[0009] 步骤三、根据电网待检修时其中所有电力设备的连接与导通情况，计算能够使当前电网停电的所有第二割集，根据第二割集计算电力设备引起电网停电损失的动态权重；动态权重加权到电网的整体运行损失中，以最小化被动态权重加权的电网的整体运行损失为目标，利用被加权动态权重的电网的整体运行损失和步骤二得到的价值矩阵改进最优策略，选取电网的整体运行损失最小的动作作为待检修电网的最终检修策略。

[0010] 本发明的有益效果是：

[0011] 本发明的一种基于强化学习的电力设备检修决策生成方法提出了用强化学习的动态规划方法求解，并且利用电网中的割集来计算设备的重要性并将其加权到动态规划求解过程中，并利用检修动作会引起割集变化的特点实现设备决策之间的联系。通过割集的改变并将融入决策过程中，能够间接实现多个设备之间的通信。本发明需要的数据较少，数据利用率高，反复利用电网中所有电力设备的连接与导通情况，反复利用电网的整体运行损失，在专业领域上的应用门槛较低。

附图说明

[0012] 图1为通过模拟得到的六种电网功率曲线图。

[0013] 图2为本发明的一种基于强化学习的电力设备检修决策生成方法的割集示意图。

[0014] 图3为本发明的一种基于强化学习的电力设备检修决策生成方法的状态和状态转移示意图。

[0015] 图4为本发明的一种基于强化学习的电力设备检修决策生成方法的策略优化流程简图。

[0016] 图5为本发明的一种基于强化学习的电力设备检修决策生成方法的流程图。

具体实施方式

[0017] 为了能够更清楚地理解本发明的上述目的、特征和优点，下面结合附图和具体实施方式对本发明进行进一步的详细描述。

[0018] 在下面的描述中阐述了很多具体细节以便于充分理解本发明，包括专业术语的解释说明，但是，本发明还可以采用其他不同于在此描述的其他方式来实施，因此，本发明的保护范围并不受下面公开的具体实施例的限制。

[0019] 一种基于强化学习的电力设备检修决策生成方法，包括如下步骤：

[0020] 步骤一、根据用于强化学习时的电网中所有电力设备的连接与导通情况，计算能够使当前电网停电的所有第一割集；根据第一割集计算电力设备引起电网停电损失的静态权重。

[0021] 步骤二、将电力设备检修决策生成问题建模为一个马尔可夫决策过程，根据电网

中所有电力设备的信息，定义电力设备有可能出现的若干个表示电力设备损坏程度的运行状态，得到的若干个表示电力设备损坏程度的运行状态构成马尔可夫决策过程的状态集合；应用强化学习方法求解马尔可夫决策过程得到最优策略和最优策略的价值矩阵，静态权重加权到强化学习的奖励中，即静态权重加权到电网的整体运行损失中，强化学习以最小化被加权静态权重的电网的整体运行损失为目标。

[0022] 步骤三、根据电网待检修时其中所有电力设备的连接与导通情况，计算能够使当前电网停电的所有第二割集，根据第二割集计算电力设备引起电网停电损失的动态权重；动态权重加权到电网的整体运行损失中，以最小化被加权动态权重的电网的整体运行损失为目标，利用被动态权重加权的电网的整体运行损失和步骤二得到的价值矩阵改进最优策略，选取被加权动态权重的电网的整体运行损失最小的动作作为待检修电网的最终检修策略。

[0023] 马尔可夫过程通常被用来建模强化学习，一个马尔可夫过程(Markov Decision Process，MDP)可用一个五元组<S，A，P，R，γ>表示，其中：S是一个有限的状态(state)集合；A是一个有限的动作(action)集合；P是一个概率转移矩阵，满足：$P_{ss'}^{a}=P[S_{t+1}=s'|S_t=s,A_t=a]$，s和s′均表示状态，状态s表示当前状态，状态s′表示转移后的状态，$P_{ss'}^{a}$表示当前状态s到转移后状态s′并采取动作a的概率转移矩阵，S_t表示t时刻的状态集合，t时刻即下文的运行时刻t，状态$s\in S_t$，S_{t+1}表示t+1时刻的状态集合，状态$s'\in S_{t+1}$，动作$a\in A_t$，A_t为t时刻的动作集合；R是一个奖励(reward)函数，满足：$R_s^a=E[R_{t+1}|S_t=s,A_t=a]$，$R_s^a$表示在当前状态s开始并采取动作a的奖励，即$R_s^a$表示没有割集加权下的奖励，E表示期望，$R_{t+1}$表示t+1时刻的奖励；γ是一个折扣因子，$\gamma\in[0,1]$；回报(return)$G_t$是对片段的评价，是指从时间t处开始所有的折扣奖励之和：

[0024] $$G_t=R_{t+1}+\gamma R_{t+2}+\cdots=\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}$$

[0025] 其中，R_{t+k+1}表示t+k+1时刻的奖励，k为大于等于0的整数；

[0026] 智能体的策略(policy)π是在给定状态下的动作分布：

[0027] $\pi(a|s)=P[A_t=a|S_t=s]$。

[0028] 状态价值函数$v_\pi(s)$描述的是智能体基于策略π在状态s开始所得到的期望回报值：$v_\pi(s)=E_\pi[G_t|S_t=s]$，$E_\pi[\,]$表示策略π对应的期望；

[0029] 动作值函数$q_\pi(s,a)$描述的是智能体基于策略π在状态s开始并采取动作a所得到的期望回报值：$q_\pi(s,a)=E_\pi[G_t|S_t=s,A_t=a]$。

[0030] 动态规划方法分为策略评估与策略迭代两部分。

[0031] 在策略评估中，每一轮迭代的近似可以使用状态价值函数v_π的贝尔曼方程进行更新，即对于任意$s\in S$有：

[0032] $$v_{new}(s)=\sum_{a\in A}\pi(a|s)\left(R_s^a+\gamma\sum_{s'\in S}P_{ss'}^{a}v(s')\right)\qquad(1)$$

[0033] 再令

[0034] $v()=v_{new}()$ (2)

[0035] 其中，$v()$表示当前时间步下的价值矩阵V，更新后的价值矩阵用于策略更新和价值矩阵更新，$v(s')$表示在状态s′和当前时间步的价值矩阵V，$v_{new}()$表示更新后的价值矩阵。显然，v_{π}是指在最优策略下的状态价值函数，步骤二中在保证v_{π}存在的条件下，序列$v()$将会收敛到v_{π}。

[0036] 在策略π改进中，我们使用贪心算法根据当前时间步的价值函数，即最新的$v()$，选择奖励最大的动作a来更新策略，其中 $P_{ss'}^{a}$ 表示在当前状态s下，执行动作a从状态s转移到状态s′的概率。对于每个状态s选择奖励最大的动作来实现策略迭代。，$\pi(a|s)$表示当前时间步的策略：

[0037] $$\pi(\cdot|\cdot)=\pi(a|s)=\arg\max_{a}\left(R_s^a+\gamma\sum_{s'\in S}P_{ss'}^{a}v(s')\right) \quad (3)$$

[0038] 在动态规划方法中，分别循环交替执行策略评估与策略迭代两部分，这两者互相影响，但他们整体会使得策略与价值函数都趋向最优解。

[0039] 本发明的思路为将电力设备看为强化学习中的智能体，将电力设备检修决策生成问题建模为一个马尔可夫决策过程，一种基于强化学习的电力设备检修决策生成方法的具体过程为：

[0040] 获取电网中各个电力设备的连接与导通情况、电网运行功率等信息。如图1模拟了6种不同的电网功率曲线的状态，功率图的横坐标均表示时间、单位为“天”，纵坐均标表示功率、单位为“MKW”。获取电网中不同电力设备的检修损失和损坏损失。电网的整体运行损失分为电网中电力设备的检修损失R_M和电网中电力设备的损坏损失R_F两种，电力设备的检修是指对尚未损坏的电力设备进行检查和维修，而损坏的电力设备则需要更换。电网中电力设备的检修损失包括电力设备检修引起的电力设备个体的经济损失$R_{M,1}$和电力设备检修引起电网停电的经济损失$R_{M,2}$，电网中电力设备的损坏损失包括电力设备损坏引起的电力设备个体的经济损失$R_{F,1}$和电力设备损坏引起电网停电的经济损失$R_{F,2}$。因此整个电网的整体运行损失(电网停电的经济损失)可以表示为：

[0041] $$R_{sum}=R_{M,1}+R_{M,2}+\sum_{t=1}^{N_T}\left(R_{F,1}(t)+R_{F,2}(t)\right)$$

[0042] 其中，$R_M=R_{M,1}+R_{M,2}$，$R_F=R_{F,1}+R_{F,2}$，t表示运行时刻，N_T表示电网运行总时长，电网停电的经济损失是会随着电网的运行功率不断改变，强化学习以最小化电网的整体运行损失为目标，即最小化电网的整体运行损失作为强化学习的奖励，权重加权到强化学习的奖励中即为权重加权到强化学习的最小化电网的整体运行损失中。

[0043] 根据电网的设备的连接与导通情况，计算能够使当前电网停电的所有割集。第一割集的第一和第二割集的第二，仅是为了区分两个割集可以不是相同的割集。割集是指当割集内的所有电力设备损坏时，该电网便会出现停电风险。割集为部分电力设备构成的集合，割集能够使当前电网停电，且不存在割集的真子集能够使当前电网停电。

[0044] 在多电力设备的情况下会存在两个问题:每个设备对电网整体运行的重要性不一致,设备的检修决策可能会影响其他设备的检修决策。为解决以上问题,本发明提出一种割集的方法来计算电网中各个设备的重要性程度,然后将其应用于动态规划求解。

[0045] 割集具体指的是点割集。点割集是原图中部分点的集合,如果去掉割集中的所有点,那么原图会分为两部分。具体定义如下:电网G为电力设备和电线构成的连通图,点集V为电网G中所有电力设备为元素构成的集合,即每个电力设备为一个点,设U是电网G的点集V的一个子集,如果在连通图G中删除U的所有点,则G-U不连通,并且不存在U的真子集使G-U不连通,就称点集U是图G的一个点割集。

[0046] 根据图2中电网中电力设备A、B、C、D和E的连接状态,人工得到当前电网的割集。如图2所示,原电路网格图的割集为{A,C,D}、{B,A}、{E,C,D}、{B,E}。

[0047] 另外,电力设备的检修也会引起割集的变化。当某一电力设备在进行检修时,一个新的割集将会根据其他电力设备的检修动作来生成,然后根据新的割集重新计算每个动作的奖励,从而实现不同电力设备的决策通信。

[0048] 通过电网的割集,能够计算每个电力设备引起电网整体运行损失的权重,通过将权重加权到强化学习的奖励中,便能使策略关注到不同设备对于电网的重要性。通过割集cut计算设备i($device_i$)的权重如下:

[0049] $$W_{device_i}(cut)=\frac{1}{N}\sum_{m=1}^{N}\frac{1}{T_m}I(cut_m,device_i)$$

[0050] 其中cut_m表示第m个割集,T_m表示当前割集中电力设备的总数量,N表示割集的个数,$I(cut_m,device_i)$的计算方式如下:

[0051] $$I(cut_m,device_i)=\begin{cases}0 & device_i \in cut_m\\ 1 & device_i \notin cut_m\end{cases}$$

[0052] 其中,i和m均为正整数。如图2所示,原电路网格图的割集为{A,C,D}、{B,A}、{E,C,D}、{B,E},那么电力设备A的权重计算方法为1/4*(1/3+1/2)≈0.2083,电力设备B、C等计算方法类似。

[0053] 基于设备权重进行强化学习动态规划求解。

[0054] 根据电力设备具体信息,定义电网中每个电力设备的表示电力设备损坏程度的运行状态,本实施方式中定义了四个状态,S1表示良好状态、S2表示轻度劣化状态、S3表示重度劣化状态、S4表示损坏状态,结合电力设备的一阶马尔可夫假设和电力设备的历史信息,计算电力设备的状态转移矩阵。历史信息目的是计算电力设备在不同运行状态间的转移概率,例如获取电力设备处于S1～S4的概率,然后通过贝叶斯方程计算转移概率,即电力设备的历史信息为电力设备处于各运行状态的概率。如图3所示,展示了一个电力设备的状态转移图,P_{bc}表示状态S_b到状态S_c的转移概率,λ_b表示S_b状态到损坏状态S4的概率,b取值1,2,3,c取值2,3,4。某个电力设备不检修时,状态转移的概率转移矩阵(又称状态转移矩阵)为:

[0055] $$\mathbf{P}=\begin{bmatrix}1-P_{12}-\lambda_1 & P_{12} & 0 & \lambda_1\\ 0 & 1-P_{23}-\lambda_2 & P_{23} & \lambda_2\\ 0 & 0 & 1-\lambda_3 & \lambda_3\\ 0 & 0 & 0 & 1\end{bmatrix}.$$

[0056] 状态转移的意思为t时刻的电力设备转移到t+1时刻的不同状态的概率，电力设备状态确定的时间单位为天。

[0057] 一个电力设备动作集合为[fix,nofix]，fix代表检修，nofix代表不检修。为了更加符合实际情况，在电力设备的S1状态只能选择“nofix”动作，在电力设备的S4状态只能选择“fix”动作，电力设备的其他状态的动作选择不进行限制。

[0058] 定义强化学习的整体优化目标，即以最小化电网的整体运行损失为目标。

[0059] 根据电力设备的权重，使用加权的动态规划算法对马尔可夫决策过程进行求解，在策略评估中加入静态权重或动态权重，如图4。

[0060] 在步骤二中应用强化学习方法求解马尔可夫决策过程得到最优策略和最优策略的价值矩阵包括如下步骤：

[0061] 步骤2.1、初始化电力设备的价值矩阵V得到初始化的价值矩阵，初始化电力设备的策略π得到初始化的策略；

[0062] 步骤2.2、利用静态权重对电网的整体运行损失进行加权，以最小化被静态权重加权的电网的整体运行损失为目标，根据被静态权重加权的电网的整体运行损失、初始化的策略和初始化的价值矩阵，利用贝尔曼方程更新价值矩阵；

[0063] 步骤2.3、以最小化被静态权重加权的电网的整体运行损失为目标，根据被静态权重加权的电网的整体运行损失和最新的价值矩阵、利用贪心算法更新策略；以最小化电网的整体运行损失为目标，根据被静态权重加权的电网的整体运行损失和最新的策略，利用贝尔曼方程更新价值矩阵；

[0064] 步骤2.4、判断是否实现被静态权重加权的电网的整体运行损失的最小化，若实现了被静态权重加权的电网的整体运行损失的最小化则步骤二完成，否则返回步骤2.3。

[0065] 在步骤三中改进最优策略包括如下步骤：

[0066] 步骤3.1、利用动态权重对电网的整体运行损失进行加权，以最小化电网的整体运行损失为目标，根据被动态权重加权的电网的整体运行损失、最优策略和最优策略的价值矩阵，利用贝尔曼方程更新最优策略的价值矩阵；

[0067] 步骤3.2、以最小化被动态权重加权的电网的整体运行损失为目标，根据被动态权重加权的电网的整体运行损失和最新的最优策略的价值矩阵、利用贪心算法更新最优策略；以最小化被动态权重加权的电网的整体运行损失为目标，根据被动态权重加权的电网的整体运行损失和最新的最优策略，利用贝尔曼方程更新最优策略的价值矩阵；

[0068] 步骤3.3、判断是否实现被动态权重加权的电网的整体运行损失的最小化，若实现了被动态权重加权的电网的整体运行损失的最小化则选取电网的整体运行损失最小的动作作为待检修电网的最终检修策略，否则返回步骤3.2。

[0069] 也就是在步骤二中，首先初始化电力设备的价值矩阵V和初始化电力设备的策略矩阵π；然后，利用贝尔曼方程结合优化目标进行策略评估和策略改进。步骤二中进行策略评估和策略改进多次，得到最终的价值矩阵V和策略π，得到最终的策略π称为最优策略，得到最终的价值矩阵V称为最优策略的价值矩阵V。

[0070] 根据最优策略和最优策略的价值矩阵，进行步骤三电网电力设备检修的动态策略生成，获取适合当前电网的电力设备检修策略，具体为：根据电网设备的运行状态，计算当

前电网的割集，并重新计算不同设备的权重$W_{device_i}(cut)$；利用得到的动态权重和最优策略的价值矩阵，重新进行策略改进步骤即改进所述最优策略，选取选择奖励最大的动作作为当前状态的最终检修策略。

[0071] 策略评估是指评估不同状态执行当前策略的价值R_s^a，即电网的整体运行损失，它包括两部分，即电力设备i在当前状态s时执行当前动作a的直接奖励$R^a_{s_{device_i}}$和在当前状态s时执行当前动作a引起状态转移的间接奖励，间接奖励是利用价值矩阵V表示，并利用权重(静态权重或动态权重)进行加权，并将其更新到价值矩阵V中。

[0072] 其中$R^a_{s_{device_i}}$计算方式如下：

[0073] $$R^a_{s_{device_i}}(cut) = -\left(R_M(s,a) + W_{device_i}(cut) \times R_F(s,a)\right)$$

[0074] 其中，cut表示第一割集或第二割集，W_{device_i}表示通过第一割集或第二割集计算得到的设备i的权重，$R_M(s,a)$表示状态s动作a时电网中电力设备的检修损失，$R_F(s,a)$表示状态s动作a时电网中电力设备的损坏损失，$R^a_{s_{device_i}}(cut)$表示基于cut的在状态s动作a时的电网的整体运行损失，即表示电力设备i在当前割集(第一割集或第二割集)加权下的奖励。也就是在计算电力设备引起的电网损失时，就可以利用割集计算得到的权重对电力设备引起的电网的整体运行损失进行加权，$W_{device_i}(\ )$表示静态权重或动态权重。然后再通过动态规划求解，从而获得电力设备对于电路连接的感知，并体现在策略中。

[0075] 加权的动态规划求解整体流程如图5所示，第一割集和第二割集均由割集管理模块提供，状态均由状态管理模块提供，电力设备的检修会引起电网割集的改变，在下一个状态中，需要根据当前割集计算各设备的权重。状态管理是指在当前时间段中，对处于S2和S3状态的设备依次进行检修决策。加权的动态规划得到对应当前策略的价值矩阵，利用当前的价值矩阵进行策略更新，选择奖励更多的动作作为当前状态的策略。根据公式(3)，将公式(3)的R_s^a换为$R^a_{s_{device_i}}(cut)$，那么步骤2.3中所述利用贪心算法更新策略和步骤3.2中所述利用贪心算法更新最优策略的更新公式均为：

[0076] $$\pi(\cdot|\cdot) = \pi\left(a\middle|s_{device_i}\right) = \arg\max_a\left(R^a_{s_{device_i}}(cut) + \gamma\sum_{s'\in S}P^a_{ss'}v(s')\right) \quad (4)$$

[0077] 其中，s_{device_i}是指设备i的当前状态s，$\pi\left(a\middle|s_{device_i}\right)$是指电力设备i在当前状态s下，选择最大奖励的动作作为相应策略，作为最新的策略π进行公式(5)。

[0078] 根据公式(2)，将公式(2)的R_s^a换为$R^a_{s_{device_q}}(cut)$，那么上述步骤2.3的利用贝尔曼方程更新价值矩阵和步骤3.2的利用贝尔曼方程更新策略的价值矩阵的更新公式均为：

[0079] $$v_{new}(s)=\sum_{a\in A}\pi\left(a\middle|s_{device_i}\right)\left(R^{a}_{s_{device_i}}(cut)+\gamma\sum_{s'\in S}P^{a}_{ss'}v(s')\right) \quad (5)$$

[0080] 再令v()＝v_{new}()，用于下一次执行公式(4)和下一次执行公式(5)。

[0081] 为构建多设备的电力检修策略，本发明提出了用强化学习的动态规划方法求解，并且利用电网中的割集来计算设备的重要性并将其加权到动态规划求解过程中，并利用检修动作会引起割集变化的特点实现设备决策之间的联系。通过割集的改变并将融入决策过程中，能够间接实现多个设备之间的通信。本发明需要的数据较少，数据利用率高，反复利用电网中所有电力设备的连接与导通情况，反复利用电网的整体运行损失，在专业领域上的应用门槛较低。通过仿真数据进行模拟，在多种电路下对策略进行评估，验证了本发明提出方法的有效性。

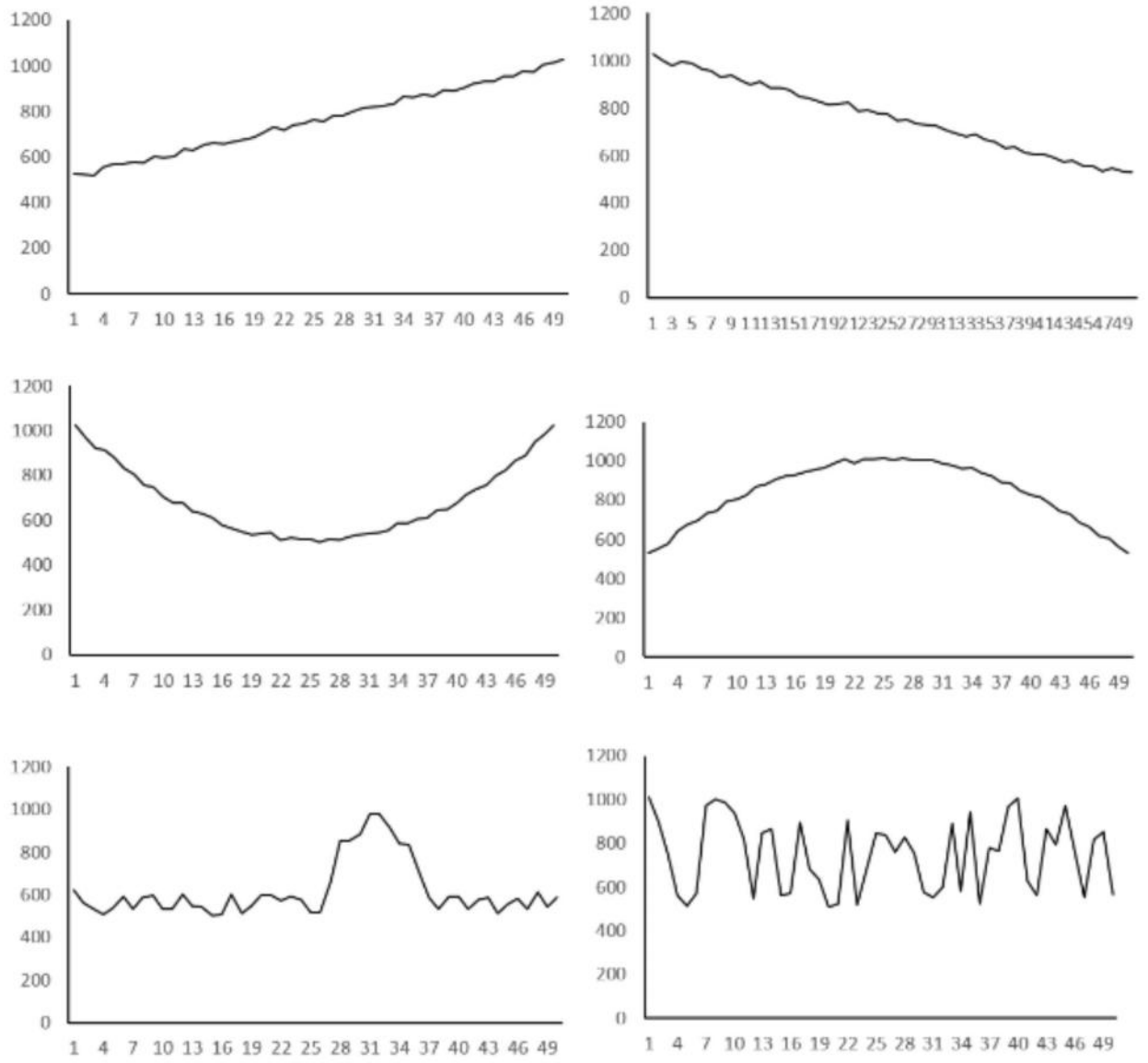

1200
1000
800
600
400
200
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49

图1

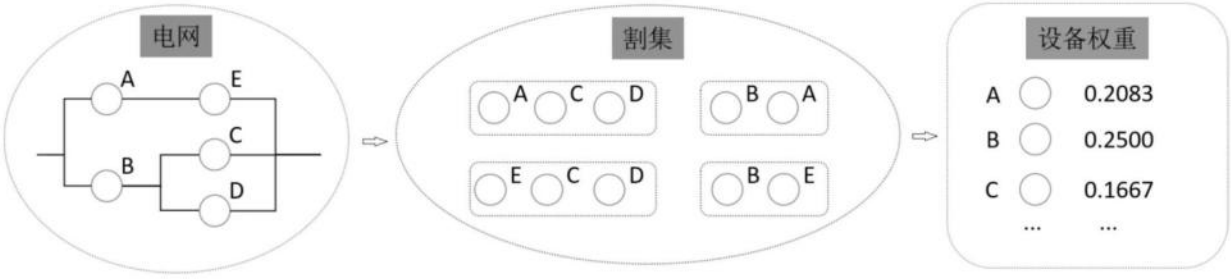

电网
A
E
C
B
D
割集
A C D
B A
E C D
B E
设备权重
A 0.2083
B 0.2500
C 0.1667
... ...

图2

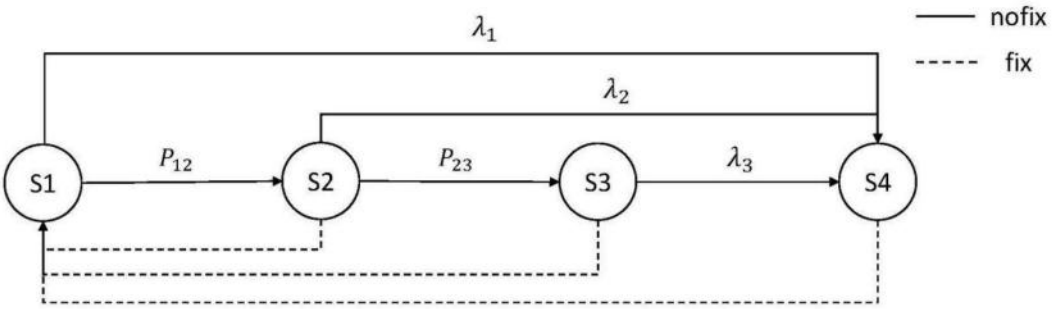

λ_1
nofix
fix
λ_2
S1
P_{12}
S2
P_{23}
S3
λ_3
S4

图3

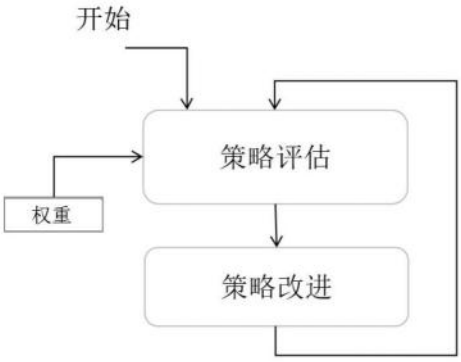

开始
策略评估
权重
策略改进

图4

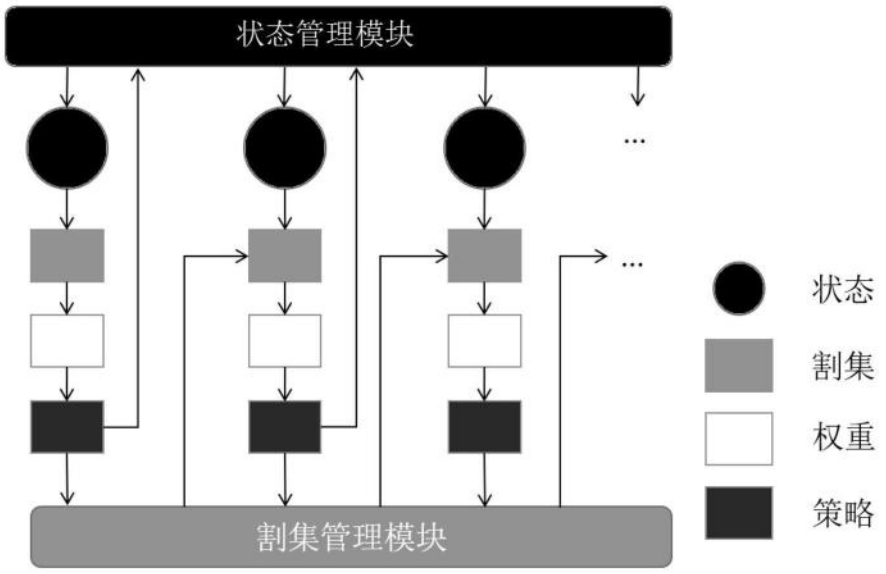

状态管理模块
...
...
状态
割集
权重
策略
割集管理模块

图5