

基于Q学习算法的综合能源系统韧性提升方法

吴熙¹,唐子逸¹,徐青山¹,周亦洲²

(1. 东南大学 电气工程学院,江苏 南京 210096;2. 河海大学 能源与电气学院,江苏 南京 210098)

摘要:将综合能源系统随机动态优化问题建模为马尔可夫决策过程,并引入Q学习算法实现该复杂问题的求解。针对Q学习算法的弊端,对传统的Q学习算法做了2个改进:改进了Q值表初始化方法,采用置信区间上界算法进行动作选择。仿真结果表明:Q学习算法在实现问题求解的同时保证了较好的收敛性,改进的初始化方法和采用的置信区间上界算法能显著提高计算效率,使结果收敛到更优解;与常规混合整数线性规划模型相比,Q学习算法具有更好的优化结果。

关键词:综合能源系统;孤岛运行;马尔可夫决策过程;Q学习算法;韧性

中图分类号:TM 73;TK 01

文献标志码:A

DOI:10.16081/j.epae.202002006

0 引言

台风、泥石流、地震等自然灾害和极端天气事件的发生会对电力系统造成严重破坏^[1-2]。电力系统承受这类高冲击低概率事件,使电力供应中断可能性最小的同时快速恢复到正常运行状态的能力被称为电力系统的韧性^[3-4]。韧性与表征可靠性的抗毁性存在较大差异,主要在于两者适用的运行场景不同,抗毁性针对的是系统正常运行场景,而韧性针对的是故障场景。随着人们对高冲击低概率事件危害性的认识日益加深,针对韧性的研究越来越受到重视^[5]。综合能源系统主要由电力系统和天然气系统构成。相较于电网的架空线路,天然气输气管道大多采用埋地式管道,对于台风、暴风雪等地表自然灾害具备一定的抵御能力^[6]。因此,电-气综合能源系统比电力系统具有更好的韧性。

根据响应过程时间的不同,韧性问题可以分为3个阶段:极端事件发生前的预防阶段、极端事件发生期间的生存阶段和极端事件结束之后的恢复阶段^[7]。文献[8-10]通过加强极端事件发生前的预防措施来提升综合能源系统韧性。在预防或恢复阶段提高系统韧性需要加固基础设施或者改变系统结构,而在生存阶段可以通过改进现有的保护控制机制或调度策略有效提高系统韧性^[11]。目前主要的研究仍集中于微网,针对如何提升生存阶段综合能源系统韧性的研究还十分有限。文献[11]提出了一种增强微网生存能力的方法,其目标是在系统恢复时间内最大限度地减少关键负荷切除,同时最大限度

地恢复非重要负荷。文献[12]提出了一种增强混合微网韧性的调度模型,该模型能提高孤岛微网关键负荷的生存能力。文献[13]提出了一种两阶段随机规划方法,充分利用微网内的资源,减轻电力中断的影响,有效地提高了微网的生存能力。相较于能量形式单一的微网,综合能源系统存在多种异质能源的耦合问题,具有随机性和强非线性,导致问题的求解更为困难。文献[6]通过蒙特卡洛模拟和马尔科夫状态转移对综合能源系统遭受不同类型自然灾害时的恢复力进行计算。文献[14]研究了生存阶段的韧性问题,在公共连接点供电中断情况下建立热电厂恢复策略模型对综合供能系统进行决策。在生存阶段提升综合能源系统韧性具有代价小、可行性高的优势,是本文的研究重点。

综合能源系统面临的不确定性主要分为2类:一类是运行参数不确定性^[15],包括分布式电源出力 and 负荷的不确定性,在资源有限的孤岛运行状态下,功率的意外波动可能会造成关键负荷的切除;另一类是与极端事件有关的不确定性,包括综合能源系统孤岛事件发生时间与持续时间。虽然许多极端事件可以被提前预知,但是由于故障发生的时间以及故障修复的时间未知,难以准确预测综合能源系统转入孤岛运行状态的时间和孤岛状态的持续时间。传统优化方法未考虑孤岛状态运行时间的不确定性,仅在预估故障恢复时间的情况下研究综合能源系统的优化运行问题。如果预估的持续时间过短,储能设备的能量在该策略中可能会过早耗尽,无法保证关键负荷持续运行;如果预估的持续时间过长,则可能导致孤岛过程中难以充分调用储能,从而造成过多的负荷切除。由于故障发生时间及恢复时间的不确定性,需要在调度范围内和调度范围之外考虑系统的生存能力。当考虑这2类不确定性时,综合能源系统的韧性问题可认为是随机动态系统的最优决策过程。马尔可夫决策过程MDP(Markov De-

收稿日期:2019-04-11;修回日期:2019-12-10

基金项目:国家电网公司科技项目(SGJSJX00YJJS1800721);
国家自然科学基金重点资助项目(51936003)

Project supported by the Science and Technology Project of SGCC(SGJSJX00YJJS1800721) and the Key Program of National Natural Science Foundation of China(51936003)

cision Process)理论是用于研究符合马尔可夫性的一类动态随机序贯决策问题的数学方法,可用来描述和处理大规模不确定性环境下的决策问题^[16]。在随机动态环境下,Q学习算法能以较小的计算代价获得较优解,对于解决复杂随机动态优化问题具有很好的效果。目前,Q学习算法已被应用于电力系统自动发电量控制^[17-18]、多能源系统安全运行^[19]、多能系统优化调度^[20-22]等领域。

本文将高冲击低概率事件发生后孤岛运行模式下提高综合能源系统生存能力(在不影响关键负荷供电的情况下,提高其他负荷供应的能力)问题建模为有限时段的MDP,引入Q学习算法实现该复杂问题的求解。在此基础上,对Q学习算法进行改进:改进了Q值表初始化过程和采用置信区间上界UCB(Upper Confidence Bound)算法进行动作选择,提高了计算效率,使结果收敛到更优解。

1 综合能源系统韧性问题的MDP建模

智能体每次做出的决策行为仅与当前时刻的调度周期、负荷需求、储能荷电状态(SOC)等状态信息有关,当智能体做出决策动作后,获得相应的“回报”并影响下一决策时刻的状态信息的大小;之后,在新的决策时刻,智能体又将根据新的状态信息采取新的决策动作。可见,未来状态仅与当前状态有关而与此前所经历的状态无关,即具有马尔可夫性。为快速、准确求解这种马尔可夫型动态随机序贯决策问题,本文将其建模为MDP。

1.1 韧性指标的构建

本文以用户热、电负荷综合损失作为韧性评价指标,包括用户电负荷中断损失和用户对于温度的不满意度,如式(1)所示。

$$R_{\text{res}} = \sum_t \sum_m \left(L_m P_{m,t} + \sigma_\theta \left| \theta_{\text{in},t+1} - \theta_{\text{opt}} \right|^2 \right) \quad (1)$$

其中, m 为电负荷等级分类; L_m 为 m 级电负荷的损失权重; $P_{m,t}$ 为 t 时刻 m 级电负荷的中断损失; σ_θ 为惩罚因子,根据用户对热舒适度要求确定; $\theta_{\text{in},t+1}$ 为 $t+1$ 时刻的室温; θ_{opt} 为最舒适的温度。

1.2 MDP模型组成元素

MDP离散控制过程包括状态空间 S 、动作空间 A 、转移概率 P 和奖励函数 R 这4个元素。

MDP的核心问题是找到最优的策略,当系统处于状态 s 时选择使得累计回报值最大的动作。

1.3 状态空间

将实时调度周期 T 、净电负荷功率 P_{nel} 、热负荷功率 H_1 、电储能系统存储能量值 w_{es} 、热储能系统存储能量值 w_{ts} 作为描述综合能源系统状态空间的状态量。其中,净电负荷功率 P_{nel} 为电负荷功率 P_{el} 和分布式发电设备出力 P_{dg} 的差值,为未知变量。

将一天等分成 $0-K$ 个决策周期,其中第 k 个决

策周期对应时段为 $[t_k, t_{k+1})$, t_k 为该周期决策时刻。将热负荷、净电负荷分别间隔离散为 $0-N_{\text{ht}}$ 和 $0-N_{\text{nel}}$ 个状态等级, N_{ht} 和 N_{nel} 分别为最大热负荷和净电负荷等级。电储能与热储能根据各自的最大容量,可以分别离散为 $0-N_{\text{es}}$ 和 $0-N_{\text{ts}}$ 个状态等级, N_{es} 和 N_{ts} 分别为最大电储能与热储能等级。

将综合能源系统在第 k 个决策周期的状态表示为 s_k ,如式(2)所示。

$$s_k = (t_k, p_{\text{nel},k}, h_{1,k}, w_{\text{es},k}, w_{\text{ts},k}) \in S \quad (2)$$

1.4 动作空间

本文中动作策略包括电储能系统、热储能系统充放电功率 P_{es} 和 H_{ts} 。动作空间 A 可以表示为:

$$A = \{P_{\text{es}}, H_{\text{ts}}\} \quad (3)$$

上述变量均为连续量,本文将其分别离散为一系列的离散值。电、热储能的充放电功率所取固定值的任一组合即对应一个动作策略。

将电储能系统充放电功率动作值离散为如下所示的 $2a+1$ 个固定值:

$$-P_{\text{es,max}}, -\frac{(a-1)P_{\text{es,max}}}{a}, \dots, -\frac{P_{\text{es,max}}}{a}, 0, \frac{P_{\text{es,max}}}{a}, \dots, \frac{(a-1)P_{\text{es,max}}}{a}, P_{\text{es,max}} \quad (4)$$

其中, $P_{\text{es,max}}$ 为电储能设备最大充放电功率。

热储能离散形式与电储能一致,不再赘述。

1.5 转移概率

假设当系统处于状态 s_k 时,选择了动作 a_k ,状态 s_k 到 s_{k+1} 的转移概率可以表示为:

$$P(s_{k+1} | s_k, a_k) = P(p_{\text{nel},k+1} | p_{\text{nel},k}) P(h_{1,k+1} | h_{1,k}) \times P(w_{\text{es},k+1} | w_{\text{es},k}) P(w_{\text{ts},k+1} | w_{\text{ts},k}) P(t_{k+1} | t_k) \quad (5)$$

其中,净电负荷功率的转移概率 $P(p_{\text{nel},k+1} | p_{\text{nel},k})$ 与净电负荷功率和分布式发电设备出力有关,可以表示为如式(6)所示。

$$P(p_{\text{nel},k+1} | p_{\text{nel},k}) = P(p_{\text{el},k+1} | p_{\text{el},k}) P(p_{\text{dg},k+1} | p_{\text{dg},k}) \quad (6)$$

假设净电负荷功率和分布式发电设备出力是未知的,在决策时刻 t_k 的转移概率 $P(p_{\text{nel},k+1} | p_{\text{nel},k})$ 也是未知的。然而,由于本文采用了Q学习算法求解问题,因此并不需要事先知道这些转移概率。本文采用的Q学习算法是一种无模型算法,该算法通过试错获得经验隐式学习转移概率。

1.6 奖励函数

奖励函数的设置会直接关系到算法的收敛速度和程度。当综合能源系统处于状态 s_k 时,选择动作 a_k 的奖励值 $r(s_k, a_k)$ 可以表示为:

$$r(s_k, a_k) = \alpha_k C_{\text{inc}}(w_{\text{es},k} - S_{\text{req},k}) - C_{\text{cpen},k} P_{\text{cels},k} - C_{\text{open},k} P_{\text{nels},k} - \lambda \left| \theta_{\text{in},k+1} - \theta_{\text{opt}} \right|^2 \quad (7)$$

$$\alpha_k = \begin{cases} 1 & w_{es,k} \geq S_{req,k} \\ 0 & \text{其他} \end{cases} \quad (8)$$

$$S_{req,k} = S_r - \omega k \quad (9)$$

$$C_{cpen,k} = C_{cp} - \beta_c k \quad (10)$$

$$C_{npen,k} = C_{np} - \beta_n k \quad (11)$$

$$C_{cp} > C_{np}, \beta_c \leq \beta_n \quad (12)$$

其中, α_k 为一个二进制变量, 如果第 k 个决策周期的电储能系统满足动态储能目标, 则其取值为 1, 否则取值为 0, 可通过式(8)得到; C_{inc} 为电储能系统激励因子; $S_{req,k}$ 为第 k 个决策周期的动态电储能目标, 可通过式(9)得到; $C_{cpen,k}$ 为第 k 个决策周期关键负荷的动态惩罚因子, 可通过式(10)得到; $C_{npen,k}$ 为第 k 个决策周期非关键负荷的动态惩罚因子, 可通过式(11)得到; $p_{cels,k}$ 为第 k 个决策周期切除的关键电负荷; $p_{nels,k}$ 为第 k 个决策周期切除的非关键电负荷; λ 为用户对于温度的不满意度惩罚系数; S_r 、 ω 分别为固定电储能目标、动态电储能目标系数; C_{cp} 、 β_c 分别为关键负荷的固定惩罚因子和动态惩罚系数; C_{np} 、 β_n 分别为非关键负荷的固定惩罚因子和动态惩罚系数。式(12)保证了关键负荷比非关键负荷的优先级更高。相较于固定惩罚因子, 设计动态惩罚因子可以更好地应对故障的清除时间比预期更早的情况。相比于固定惩罚因子的在各调度周期随机卸载, 采用动态惩罚因子将负荷减载转移到调度周期的相对后期, 避免了不必要的负荷减载, 从而提高了韧性。本文中的热负荷主要是空间热负荷, 在一定范围内, 用户不会感觉到温度变化的明显差异, 因此热负荷曲线可以转换为一个温度舒适度区间, 在满足用户舒适度的前提下具备一定的弹性。

为了考虑当前行为对未来奖励的影响, 第 k 个决策周期的总折扣奖励用 R_k 表示, 将其定义为时刻 t_k 的即时奖励与时刻 t_{k+1} 的折扣奖励之和, 如式(13)所示。

$$R_k = r(s_k, a_k) + \gamma \sum_{i=1}^{\infty} r(s_{k+i}, a_{k+i}) \quad (13)$$

其中, γ 为折现因子, 反映了未来回报的重要性。 γ 取值越小表示越重视当前回报。如 $\gamma=0$ 时表示只考虑当前奖励, 而 $\gamma=1$ 时表示系统对当前奖励和未来长期奖励的权重相等。

2 基于 Q 学习算法求解综合能源系统的韧性问题

2.1 Q 学习算法的基本原理

Q 学习算法是一种无模型强化学习算法, 可解决模型不完全可知情况下 MDP 的连续决策问题。 Q 学习不依赖于数学模型, 智能体只需获取外部环境的实时状态和实时反馈, 通过与外部环境的不断交互获得知识, 即可计算出智能体当前的最优动作策

略, 是一种基于值函数迭代的在线学习和动态最优技术^[16]。

Q 学习在第 k 个决策周期为每个状态-动作 $Q(s_k, a_k)$ 分配一个 Q 值, 并且在每次迭代时对其进行更新, $Q(s_k, a_k)$ 为在状态 s_k 采取行动 a_k 时动作值函数的期望。

$$Q(s_k, a_k) = E \left\{ r(s_k, a_k) + \gamma \sum_{i=1}^{\infty} r(s_{k+i}, a_{k+i}) \right\} \quad (14)$$

智能体通过不断的探索与试错获得知识经验, 反复更新动作值函数 $Q(s_k, a_k)$ 来逼近最优动作值函数 Q^* , 即动作值函数的最大期望。经过足够次数的迭代更新后, Q 值收敛到最大值。获得最大 Q 值的更新机制可用贝尔曼方程表示:

$$Q(s_k, a_k) = Q(s_k, a_k) + \alpha (r(s_k, a_k) + \gamma \max_{a' \in A} Q(s_{k+1}, a') - Q(s_k, a_k)) \quad (15)$$

其中, a' 为状态 s_{k+1} 下的可行动作; α 为学习速率, 它决定了新旧经验所占的比重。 α 越大, 表明保留之前训练的效果就越少, $\alpha=1$ 时只保留最新获得的经验。

2.2 改进的 Q 学习算法

针对 Q 学习算法的弊端, 本文对传统的 Q 学习算法做了 2 个改进: 改进了 Q 值表初始化过程, 采用 UCB 算法进行动作选择。前者可以提高收敛速度, 后者使得 Q 学习算法更容易跳出局部最优解, 做出更优的动作决策。

2.2.1 初始化方法

常规的 Q 学习初始化方法是在迭代开始之前将 Q 值表中所有元素置 0。在本文中, 状态-动作对的数量相当庞大, 因此需要花费大量的时间来探索和更新动作值函数 $Q(s, a)$, 收敛速度较慢。

为了提高计算效率, 本文提出了改进的初始化方法。在迭代开始之前, 将 Q 值表中的每个状态-动作对的值初始化为折现因子 $\gamma=0$ 时的即时奖励值 $r(s_k, a_k)$ 。计算即时奖励值 $r(s_k, a_k)$ 并不需要迭代计算, 只需 1 次计算即可得到, 利用该简单的初始化步骤可以显著缩短收敛时间。

2.2.2 UCB 算法

在 Q 学习算法的学习阶段, 成功的平衡探索与扩张对智能体的学习性能有着显著的影响。扩张是在目前已经给定的信息下做出最佳选择, 探索是通过做出其他选择收集更多的信息。常见的小量贪婪算法 (ϵ -greedy) 策略^[18,20] 对所有动作以等概率进行随机探索, 只关心每个动作的回报是多少, 而忽略每个动作的被选择次数等历史信息。该算法不再选中初始回报特别低的动作, 即使这个动作的回报只测试了很少的次数。而 UCB 算法不仅关注每个动作的奖励, 同样会关注每个动作被选择的次数, 并且动作选择依据的是数据计算的结果, 排除了初始阶段

单次随机坏结果的影响。因此,UCB算法更容易做出最优动作决策,收敛到更大的Q值。

在决策时刻 t_k ,智能体的目标是在状态 s_k 下选择具有最优动作值函数 Q^* 的动作 a 。但是,在时刻 t_k 真实的 $Q_k(s,a)$ 是未知的。在时刻 t_k 估计的Q值与真实的Q值之间存在误差项,如式(16)所示。

$$Q_k(s,a) = \langle Q(s,a,k) \rangle + (Q_k(s,a) - \langle Q(s,a,k) \rangle) \quad (16)$$

其中, $\langle Q(s,a,k) \rangle$ 为在决策时刻 t_k 估计得到的Q值; $(Q_k(s,a) - \langle Q(s,a,k) \rangle)$ 为对应的误差项。

霍夫丁不等式是限制这个误差的一种方法,表示如下:

$$Q_k(s,a) \leq \langle Q(s,a,k) \rangle + \sqrt{\frac{2 \ln N_t}{N(s,a)}} \quad (17)$$

其中, N_t 为到目前为止总共的迭代次数; $N(s,a)$ 为到目前每一个状态 s 下每一个动作 a 被执行的次数。

最优策略可表示为:

$$Q_k^*(s,a) = \langle Q(s,a,k) \rangle + \sqrt{\frac{2 \ln N_t}{N(s,a)}} \quad (18)$$

2.3 求解综合能源系统韧性的算法流程

本文提出的基于Q学习算法的综合能源系统韧性提升方法的步骤如下:步骤1,采用上文提到的初始化方法对所有状态-动作对 $Q(s,a)$ 进行初始化;步骤2,令学习次数 $N_{epi}=0$;步骤3,令决策周期 $k=0$,随机初始化系统状态 s_k ;步骤4,在状态 s_k 下根据UCB算法选择并执行动作 a_k ;步骤5,获得新的状态信息 s_{k+1} 和奖励值 $r(s_k,a_k)$;步骤6,根据式(15)更新Q值表;步骤7,令 $k=k+1$,若 $k \leq K$,返回步骤4,否则转至步骤8;步骤8,令 $N_{epi}=N_{epi}+1$,若 $N_{epi} \leq N_{epi,max}$,则返回步骤3,否则结束程序。

3 算例分析

为验证本文所提算法的有效性,构建了一个综合能源系统,该系统包括燃气轮机、余热锅炉、分布式发电机组、电储能设备、热储能设备,具体如图1所示。负荷包括电负荷和热负荷,其中电负荷又分

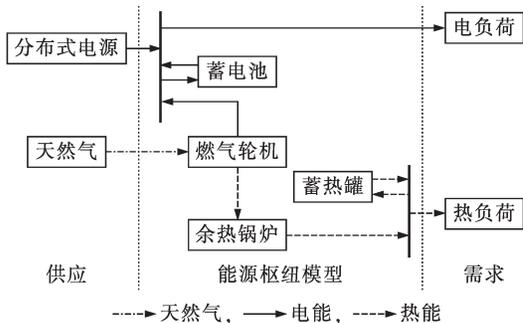


图1 综合能源系统结构

Fig.1 Structure of integrated energy system

为关键负荷和非关键负荷。极端事件发生后,综合能源系统与外部电网脱离,转入孤岛运行模式。

本文将0~1400 kW净电负荷功率以200 kW为间隔划分为7个状态;将0~1800 kW热负荷功率以300 kW为间隔划分为6个状态;电储能系统最大容量为1000 kW·h,容量上、下限分别为900 kW·h和200 kW·h。将存储容量以200 kW·h为间隔划分为5个状态;热储能容量为1000 kW·h,容量上、下限分别为900 kW·h和200 kW·h,以同样间隔划分为5个状态;以15 min为时间间隔,共24个决策时刻。总计 $7 \times 6 \times 5 \times 5 \times 24 = 25200$ 个状态组合。假定储能系统最大充放功率为每小时20%额定容量,即200 kW。将电储能系统充放电功率动作值以100 kW离散为以下5个固定值,即充电200 kW、充电100 kW、不充不放、放电200 kW、放电100 kW。热储能系统同样以100 kW离散为上述5个固定值。共计 $5 \times 5 = 25$ 个动作组合。总的状态-动作对为 $25200 \times 25 = 630000$ 。设Q学习算法的学习速率 $\alpha=0.1$ 、折扣因子 $\gamma=0.9$ 。设置 S_r, C_{ep}, C_{mp} 分别为0.75、20、10, ω, β_c, β_n 分别为0.1、0.2、0.2。并在MATLAB 2014b上进行仿真。

3.1 算法计算速度和收敛性比较

现实中极端事件发生的概率很小,获取可用于学习的完整、全面的真实数据比较困难。因此,本文采用蒙特卡洛法模拟法生成随机数列构成样本数据,获得了大量可用于学习的样本数据。

为了验证所提算法的优越性,在同样进行1000000次学习的情况下,对比了 ϵ -greedy算法和UCB算法在常规初始化方法和本文所提初始化方法下的学习速度,结果如表1所示。相比于常规初始化方法,本文所提初始化方法在采用 ϵ -greedy算法和UCB算法时的计算效率分别提高了11.9%和14.7%。在常规初始化方法和本文所提初始化方法下,UCB算法所花费时间比 ϵ -greedy算法分别减少了8.25%和11.17%。由此可见,本文采用的基于UCB算法的Q学习算法和提出的初始化方法都能够减少学习时间,提高收敛效率。

表1 学习速度比较

算法	学习花费时间/s	算法	学习花费时间/s
ϵ -greedy+常规初始化	1912.05	UCB+常规初始化	1754.23
ϵ -greedy+本文初始化	1684.31	UCB+本文初始化	1496.20

经过1000000次学习之后,基于UCB算法的Q学习算法已经基本收敛,而基于 ϵ -greedy算法的Q学习算法仍未收敛,如图2所示。并且基于UCB算法的Q学习算法比基于 ϵ -greedy算法的Q学习算法收敛到更大的Q值,这说明基于UCB算法的Q学习算法选择了更优的动作决策。

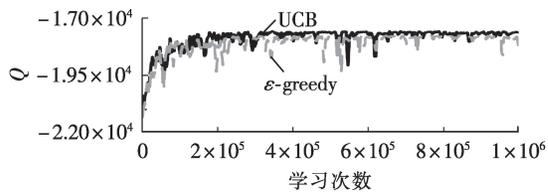


图2 算法收敛性比较

Fig.2 Comparison of algorithm convergence

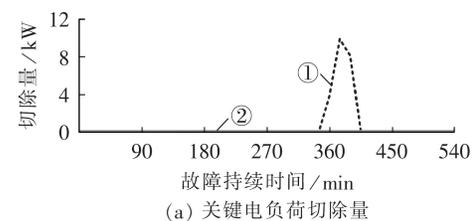
3.2 不确定性场景下的系统韧性对比

为验证本文提出的方法对上文提到的与极端事件有关的不确定性和运行参数不确定性有良好的应对能力,将本文提出的方法与常规的建模为混合整数规划模型的优化方法(下文简称为常规方法)在不确定场景下进行了对比。

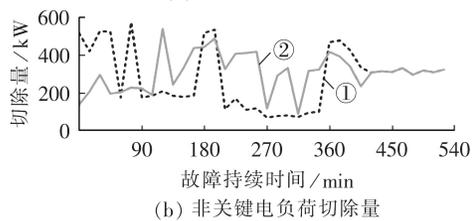
3.2.1 场景1

场景1假定不考虑预测误差,故障在发生9 h后被清除,故障清除时间晚于预期3 h。该场景下的负荷及分布式能源实时数据见附录图A1。

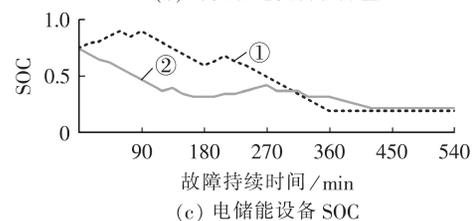
场景1下的仿真结果如图3所示。在该场景下,常规方法在符合故障预计清除时间360 min内没有切除关键负荷,但在故障持续时间超过预期之后的375~405 min时切除了关键负荷,而本文所提方法没有切除关键负荷,如图3(a)所示。这表明常规方法无法应对故障持续时间的不确定性。因为常规方法需要对故障持续时间进行事先预测,无法在实时状态下进行优化决策。从图3(b)中可以看出,因为采用了动态惩罚系数,本文所提算法将负荷减载转移到调度周期的相对后期,在故障发生的前2 h只切除



(a) 关键电负荷切除量



(b) 非关键电负荷切除量



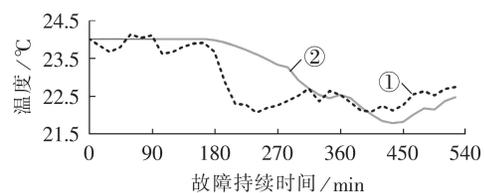
① 常规方法, ② 本文所提方法

图3 场景1下的仿真结果

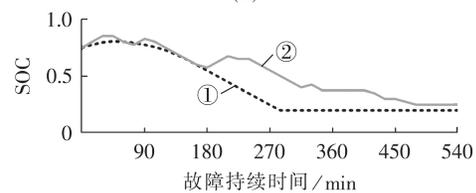
Fig.3 Simulative results under Scenario 1

了少量非关键负荷。如果故障在前2 h内被清除,常规方法将进行不必要的减载,而本文所提方法可以有效地避免不必要的负荷削减,很好地应对了事故清除时间早于预期的情况。在故障预计清除的360 min时,常规方法下电储能设备的能量已经完全被耗尽,在此后阶段电储能设备无法继续供电,造成了关键负荷的切除,无法保证综合能源系统在故障持续时间不确定场景下的生存能力,如图3(c)所示。而本文所提方法下电储能设备在故障预计清除的360 min时仍然留有12.5%的可用余量,在故障清除时间比预期长的情况下仍然保证了关键负荷的正常供电。

考虑到用户对温度存在舒适度范围以及建筑的蓄热特性,热负荷生存能力可以转化为对用户温度舒适度的评估。假设用户的舒适温度为24℃,在舒适温度的±2℃范围内波动属于用户可接受温度。在故障预计清除时间内,常规方法和本文所提方法都能保证室温处在用户温度舒适度范围内。但在故障持续时间超过预期之后,常规方法下室温持续下降,在故障持续了435~465 min时超出用户温度舒适度范围,而本文所提方法在所有时间内都能保证室温处在用户温度舒适度范围内进行调节,如图4(a)所示。在故障预计清除的360 min时,常规方法下热储能设备的能量已经达到下限,在此后阶段热储能设备无法继续供电,而本文所提方法下热储能设备在故障预计清除的360 min时仍然留有37.5%的余量,如图4(b)所示。因此,本文所提的方法能够有效应对故障持续时间的不确定性,保证了综合能源系统在故障持续时间不确定场景下的生存能力。



(a) 室温



(b) 热储能设备 SOC

① 常规方法, ② 本文所提方法

图4 场景1下的仿真结果

Fig.4 Simulative results under Scenario 1

3.2.2 场景2

场景2假定预测误差为±10%,故障在发生6 h后被清除,与预计时间相符。为了验证本文所提方法的可靠性,考虑了极端误差场景,即假设分布式发

电设备为负偏差,负荷需求为正偏差,实际的净负荷需求比预测值大。该场景下的负荷及分布式能源实时数据见附录图A2。在该场景下,因为存在预测误差,常规方法只能通过储能设备消除实时数据与预测之间的不平衡量,这会造成储能设备的能量过早释放,从而在故障后期无法为关键负荷提供能量。

在故障后期,常规方法切除了部分关键负荷,而本文所提方法没有切除关键负荷,如图5(a)所示。这意味着常规方法无法保证综合能源系统在孤岛状态下的生存能力。当故障持续了270 min时,常规方法的电储能SOC就已经达到下限值,而本文所提方法的电储能SOC在故障清除时仍留有一定余量,如图5(b)所示。本文所提方法在学习完成后能够响应不断变化的环境场景,根据实时信息进行决策,并不会受到预测误差的影响,可以有效应对分布式能源出力和负荷的不确定性。

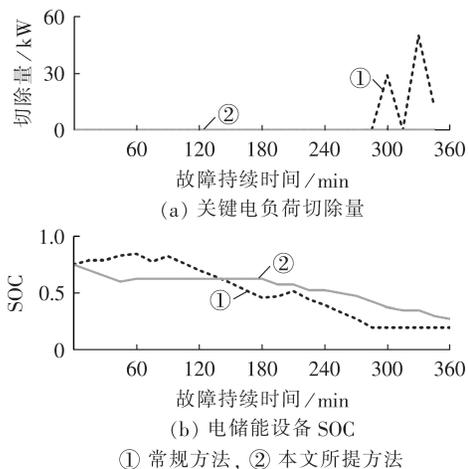


图5 场景2下的仿真结果

Fig.5 Simulative results under Scenario 2

在故障持续的时间内,常规方法无法保证室温完全处在用户温度舒适度范围内,而本文所提方法都能保证室温处在用户温度舒适度范围内进行调节,如附录图A3(a)所示。在存在预测误差的情况下,常规方法为了尽可能维持室温,过早地释放热储能设备的能量。当故障持续了215 min时,热储能设备就达到了容量下限值,比不存在预测误差提前了60 min,如附录图A3(b)所示。本文所提方法考虑了用户对于温度的不敏感性,合理地利用了热储能设备的能量,保证了室温在用户温度舒适度范围内,有效地应对了预测误差的不确定性。

4 结论

本文提出一种基于Q学习算法的综合能源系统韧性提升方法,通过优化本地微源供能及需求响应提升了综合能源系统在孤岛运行模式下的生存能力。将孤岛运行方式下综合能源系统随机动态优化

问题建模为MDP,引入Q学习算法实现该复杂问题的求解。Q学习算法可以在预测信息未知且没有先验知识的情况下通过经验学习来识别最优的动作决策行为。仿真结果验证了Q学习算法的有效性,其通过不断学习积累经验,实现了更好的收敛效果,在不确定性场景下能保证综合能源系统的生存能力。改进了的Q值表初始化方法及采用UCB算法进行动作选择可以显著提高计算效率,使结果收敛到更优解。并且,与常规混合整数线性规划模型相比,Q学习算法具有更好的优化结果。

目前,本文只解决了在生存阶段提升综合能源系统韧性的问题,下一步将研究故障前的预防策略以及资源受限情况下恢复策略的设计,形成一套针对综合能源系统的韧性提升全过程解决方案。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 吴勇军,薛禹胜,谢云云,等. 台风及暴雨对电网故障率的时空影响[J]. 电力系统自动化,2016,40(2):20-29.
WU Yongjun, XUE Yusheng, XIE Yunyun, et al. Space-time impact of typhoon and rainstorm on power grid fault probability[J]. Automation of Electric Power Systems, 2016, 40(2): 20-29.
- [2] 李更丰,黄玉雄,别朝红,等. 综合能源系统运行可靠性评估综述及展望[J]. 电力自动化设备,2019,39(8):12-21.
LI Gengfeng, HUANG Yuxiong, BIE Zhaozhong, et al. Review and prospect of operational reliability evaluation of integrated energy system[J]. Electric Power Automation Equipment, 2019, 39(8):12-21.
- [3] 高海翔,陈颖,黄少伟,等. 配电网韧性及其相关研究进展[J]. 电力系统自动化,2015,39(23):1-8.
GAO Haixiang, CHEN Ying, HUANG Shaowei, et al. Distribution systems resilience: an overview of research progress[J]. Automation of Electric Power Systems, 2015, 39(23):1-8.
- [4] GHOLAMI A, SHEKARI T, AMIRIOUN M H, et al. Toward a consensus on the definition and taxonomy of power system resilience[J]. IEEE Access, 2018, 6: 32035-32053.
- [5] 陈胜,卫志农,孙国强,等. 电-气互联综合能源系统安全分析与优化控制研究综述[J]. 电力自动化设备,2019,39(8):3-11.
CHEN Sheng, WEI Zhinong, SUN Guoqiang, et al. Review on security analysis and optimal control of electricity-gas integrated energy system[J]. Electric Power Automation Equipment, 2019, 39(8):3-11.
- [6] 齐世雄,王秀丽,邵成成,等. 极端事件下电-气混联综合能源系统的恢复力分析[J]. 电网技术,2019,43(1):41-49.
QI Shixiong, WANG Xiuli, SHAO Chengcheng, et al. Resilience analysis of integrated electricity and natural gas energy system under extreme events[J]. Power System Technology, 2019, 43(1):41-49.
- [7] HUSSAIN A, BUI V, KIM H. A proactive and survivability-constrained operation strategy for enhancing resilience of microgrids using energy storage system[J]. IEEE Access, 2018, 6: 75495-75507.
- [8] MANSHADI S D, KHODAYAR M E. Resilient operation of multiple energy carrier microgrids[J]. IEEE Transactions on Smart Grid, 2015, 6(5):2283-2292.
- [9] CONG H, HE Y, WANG X, et al. Robust optimization for

- improving resilience of integrated energy systems with electricity and natural gas infrastructures[J]. Journal of Modern Power Systems and Clean Energy, 2018, 6(5):1066-1078.
- [10] WU L, HE C, DAI C, et al. Robust network hardening strategy for enhancing resilience of integrated electricity and natural gas distribution systems against natural disasters[J]. IEEE Transactions on Power Systems, 2018, 33(5):5787-5798.
- [11] BALASUBRAMANIAM K, SARAF P, HADIDI R, et al. Energy management system for enhanced resiliency of microgrids during islanded operation[J]. Electric Power Systems Research, 2016, 137:133-141.
- [12] HUSSAIN A, BUI V, KIM H. Optimal operation of hybrid microgrids for enhancing resiliency considering feasible islanding and survivability[J]. IET Renewable Power Generation, 2017, 11(6):846-857.
- [13] GHOLAMI A, SHEKARI T, AMINIFAR F, et al. Microgrid scheduling with uncertainty: the quest for resilience[J]. IEEE Transactions on Smart Grid, 2016, 7(6):2849-2858.
- [14] 周玲, 吕林, 李婷, 等. 含分布式发电的热电联供系统负荷恢复策略[J]. 电力建设, 2017, 38(12):87-96.
ZHOU Ling, LÜ Lin, LI Ting, et al. Load recovery strategy for combined heat and power system with distributed generation [J]. Electric Power Construction, 2017, 38(12):87-96.
- [15] 雷金勇, 郭祚刚, 陈聪, 等. 考虑不确定性及电/热储能的综合能源系统两阶段规划-运行联合优化方法[J]. 电力自动化设备, 2019, 39(8):169-175.
LEI Jinyong, GUO Zuogang, CHEN Cong, et al. Two-stage planning-operation co-optimization of IES considering uncertainty and electrical/thermal energy storage[J]. Electric Power Automation Equipment, 2019, 39(8):169-175.
- [16] 葛少云, 李吉峰, 刘洪, 等. 考虑物理特征与行为因素的家庭用能特性建模[J]. 电力自动化设备, 2019, 39(3):36-44.
GE Shaoyun, LI Jifeng, LIU Hong, et al. Modelling of household energy consumption characteristics considering physical features and behavior factors[J]. Electric Power Automation Equipment, 2019, 39(3):36-44.
- [17] 殷林飞, 余涛. 基于深度 Q 学习的强鲁棒性智能发电控制器设计[J]. 电力自动化设备, 2018, 38(5):12-19.
YIN Linfei, YU Tao. Design of strong robust smart generation controller based on deep Q learning[J]. Electric Power Automation Equipment, 2018, 38(5):12-19.
- [18] 张孝顺, 李清, 余涛, 等. 基于协同一致性迁移 Q 学习算法的虚拟发电部落AGC功率动态分配[J]. 中国电机工程学报, 2017, 37(5):1455-1467.
ZHANG Xiaoshun, LI Qing, YU Tao, et al. Collaborative consensus transfer Q -learning based dynamic generation dispatch of automatic generation control with virtual generation tribe[J]. Proceedings of the CSEE, 2017, 37(5):1455-1467.
- [19] 曹茂森, 王蕾报, 胡博, 等. 考虑电-气耦合系统连锁故障的多阶段信息物理协同攻击策略[J]. 电力自动化设备, 2019, 39(8):128-136.
CAO Maosen, WANG Leibao, HU Bo, et al. Coordinated cyber-physical multi-stage attack strategy considering cascading failure of integrated electricity-natural gas system[J]. Electric Power Automation Equipment, 2019, 39(8):128-136.
- [20] LU R, HONG S H, ZHANG X. A dynamic pricing demand response algorithm for smart grid: reinforcement learning approach[J]. Applied Energy, 2018, 220:220-230.
- [21] 瞿凯平, 张孝顺, 余涛, 等. 基于知识迁移 Q 学习算法的多能源系统联合优化调度[J]. 电力系统自动化, 2017, 41(15):18-25.
QU Kaiping, ZHANG Xiaoshun, YU Tao, et al. Knowledge transfer based Q -learning algorithm for optimal dispatch of multi-energy system[J]. Automation of Electric Power Systems, 2017, 41(15):18-25.
- [22] 李怡瑾, 唐昊, 吕凯, 等. 源荷不确定冷热电联供微网能量调度的建模与学习优化[J]. 控制理论与应用, 2018, 35(1):56-64.
LI Yijin, TANG Hao, LÜ Kai, et al. Modeling and learning-based optimization of the energy dispatch for a combined cooling, heat and power microgrid system with uncertain sources and loads[J]. Control Theory & Applications, 2018, 35(1):56-64.

作者简介:



吴 熙

吴 熙(1987—),男,江苏靖江人,副教授,博士,主要从事电力系统振荡分析与控制、柔性交流输电系统、分布式控制等方面的研究(E-mail:wuxi112233@163.com);

唐子逸(1994—),男,浙江建德人,硕士研究生,主要研究方向为综合能源系统的运行与优化(E-mail:25804126@qq.com)。

(编辑 李玮)

Q -learning algorithm based method for enhancing resiliency of integrated energy system

WU Xi¹, TANG Ziyi¹, XU Qingshan¹, ZHOU Yizhou²

(1. School of Electrical Engineering, Southeast University, Nanjing 210096, China;

2. College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China)

Abstract: The stochastic dynamic optimization problem of integrated energy system is modeled as a Markov decision process, and Q -learning algorithm is introduced to solve this complex problem. In order to overcome the disadvantages of Q -learning algorithm, two improvements are made to the typical Q -learning: the Q table initialization method is improved and the upper bound convergence algorithm is adopted for the action selection. Simulative results show that Q -learning algorithm ensures better convergence while solving the problem, and the improved initialization method and the upper bound convergence algorithm can significantly improve the computational efficiency and make the results converge to a better solution. Moreover, compared with the conventional mixed integer linear programming model, Q -learning algorithm achieves better optimization results.

Key words: integrated energy system; islanded operation; Markov decision process; Q -learning algorithm; resiliency

附录

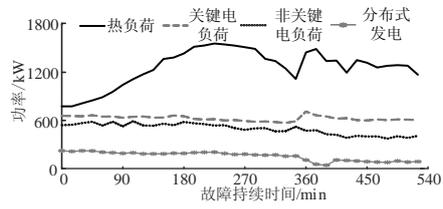


图 A1 场景 1 下负荷及分布式能源出力实时数据

Fig.A1 Real-time data of load and distributed generation under Scenario 1

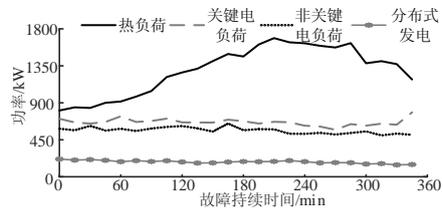


图 A2 场景 2 下负荷及分布式能源出力实时数据

Fig.A2 Real-time data of load and distributed generation under Scenario 2

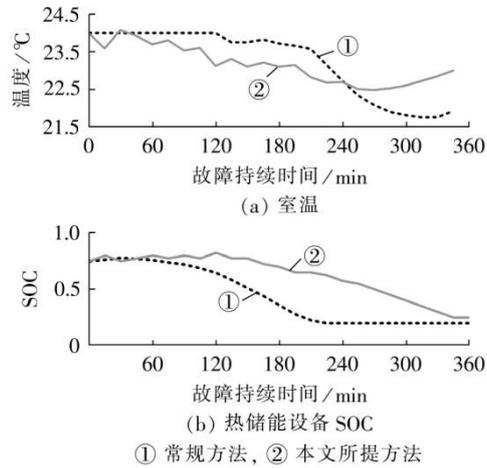


图 A3 场景 2 下的仿真结果

Fig.A3 Simulative results under Scenario 2