

基于深度强化学习的电网机组组合算法

温裕鑫, 杨 军, 朱 旭

(武汉大学, 湖北 武汉 430072)

摘要:机组组合是电力系统经济调度中的关键问题。机组组合问题属于复杂混合整数非线性优化问题,大多数优化算法在求解速度、求解效率、计算复杂性等方面存在不足。随着人工智能技术的不断发展,智能算法在电力系统决策优化问题上逐渐得到应用,并取得较好效果。本文提出了一种基于深度强化学习的电网机组组合算法:首先构建以发电成本最小为目标函数的机组组合数学模型,并以功率平衡、机组出力限制和最小启停时间约束为约束条件,然后将机组组合问题在强化学习框架下完成描述,并对状态空间、动作空间、奖励函数进行定义,最后采用深度Q网络算法来训练求解机组组合问题模型。经算例分析及测试结果分析,验证了所提方法在机组组合问题求解上的有效性和优越性。

关键词:机组组合;深度强化学习;深度Q网络

中图分类号:TM73 **文献标志码:**A **文章编号:**1001-9898(2021)05-0006-05

Power Grid Unit Commitment Algorithm Based on Deep Reinforcement Learning

WEN Yuxin, YANG Jun, ZHU Xu

(Wuhan University, Wuhan 430072, China)

Abstract: Unit commitment is a key issue in the economic dispatch of power system. For solving this kind of complex mixed-integer nonlinear optimization problem, most optimization algorithms have shortcomings in solving speed, solving efficiency and computational complexity. With the continuous development of artificial intelligence technology, intelligent algorithms have gradually been applied in decision-making optimization problems in power system and achieved better results. This paper proposes a unit commitment algorithm based on deep reinforcement learning. Firstly, a unit commitment mathematical model is constructed with the objective function of minimizing generation cost and the constraints of power balance, generation limit and minimum uptime and downtime. Secondly under the framework of the reinforcement learning, a mathematical description of the unit commitment problem is established. And the corresponding state space, action space and reward function are defined. Finally, the DQN algorithms is applied to train the model to solve the unit commitment problem. The well-trained model verified the effectiveness and superiority of the method proposed in solving the unit commitment problem by analyzing the test results.

Key words: unit commitment; deep reinforcement learning; DQN

0 引言

随着电网规模扩大与用户量增大,对电力系统经济运行提出新的挑战。机组组合由于其显著的经济效益,一直以来都是电力系统研究中的热点和难点。机组组合问题优化求解指的是在一段时间周期内满足设备运行约束的前提下,通过制定周期内各机组的启停计划和出力计划,以达到整体运行成本最小化为目的。

从数学角度看,机组组合问题是一种高维、离

散、非线性的混合整数优化问题^[1]。许多研究者应用各种方法来求解这个数学优化问题,但对于此类问题是难以求得最优解的。在传统优化算法中,文献[2]提出一种优先顺序法求解机组组合问题,由于缺乏严格的理论依据,可行解的最优性无法求证。文献[3]提出采用动态规划法求解机组组合问题,通过排除不可行组合来降低计算量,但也同时可能将最优解排除。文献[4]采用拉格朗日松弛法对约束条件松弛,以形成拉格朗日问题。但由于该算法应用上存在对偶间隙,无法

收稿日期:2021-06-15

作者简介:温裕鑫(1999-),男,硕士研究生,主要研究方向为电力系统及其自动化。

得到最优解。在智能优化算法中,遗传算法、蚁群算法、粒子群算法等在机组组合问题上均有所应用^[5-10]。迄今为止,上述优化算法在计算速度、计算精度和求解效率上均有所不足。

随着人工智能技术在电力系统中的应用和发展,机器学习在机组组合问题的求解上具有更好的优势。深度学习和强化学习是机器学习的两大主流研究方向^[11]。深度学习能够提取海量数据中的特征信息,从而学习获得输入与输出的映射关系,具有强大的感知拟合能力。应用深度学习求解机组组合时海量数据中的决策数据无法保证都是理论最优解。强化学习能够在智能体与环境不断探索交互的过程中,学习状态与动作的策略模型,具有强大的探索决策能力。在应用强化学习求解机组组合时由于算法本身的局限性可能造成维数灾难问题。由此,引入深度强化学习方法,该方法能够在具有二者优点的同时避免存在的问题,更好地解决机组组合问题。

本文提出了一种基于深度强化学习的电网机组组合算法,构建了以发电成本最小为目标函数,功率平衡、机组出力限制和最小启停时间约束为约束条件的机组组合数学模型;在强化学习框架下定义了状态空间、动作空间、奖励函数,然后采用深度Q网络(Deep Q Network, DQN)算法结合LAMBDA迭代法来求解,并通过仿真计算验证了该算法的有效性。

1 机组组合数学模型

机组组合优化问题的目的就是获得显著的经济效益,即在满足设备运行约束条件下,求解能够使总发电成本最小的机组启停计划和出力计划。

1.1 目标函数

构建以总发电成本最小为目标函数的电网机组组合数学模型,总发电成本包括机组的燃料成本和机组的启停成本,具体表示如下:

$$\min \sum_{t=1}^T W_{G,t} \quad (1)$$

$$W_{G,t} = \sum_{i=1}^N (F_{G,i,t} + S_{on,i,t} + S_{off,i,t}) = \sum_{i=1}^N [(a_i P_{G,i,t}^2 + b_i P_{G,i,t} + c_i) X_{i,t} + S_i X_{i,t} (1 - X_{i,t})] \quad (2)$$

式中: $W_{G,t}$ 为全部机组在 t 时刻的发电成本; $F_{G,i,t}$

为机组 i 在 t 时刻的燃料成本; $S_{on,i,t}$ 为机组 i 在 t 时刻的启动成本; $S_{off,i,t}$ 为机组 i 在 t 时刻的停机成本; a_i 、 b_i 、 c_i 为机组 i 的燃料消耗系数; $P_{G,i,t}$ 为机组 i 在 t 时刻的出力; $X_{i,t}$ 为机组 i 的启停状态, 1 为开机状态, 0 为停机状态; S_i 为机组 i 的启动成本; N 为机组数量; T 为调度总时刻。

1.2 约束条件

机组组合问题的约束条件包括功率平衡约束、机组出力限制约束、最小启动时间约束、最小停机时间约束。

功率平衡

$$\sum_{i=1}^N X_{i,t} P_{G,i,t} = P_{L,t} \quad (3)$$

机组出力限制

$$X_{i,t} P_{G,i,\min} \leq P_{G,i,t} \leq X_{i,t} P_{G,i,\max} \quad (4)$$

最小启动时间

$$T_{G,i,t} \geq T_{on,i} \quad (5)$$

最小停机时间

$$-T_{G,i,t} \leq T_{off,i} \quad (6)$$

式中: $P_{L,t}$ 为 t 时刻的系统负荷需求; $P_{G,i,\min}$ 为机组 i 运行时的出力下限; $P_{G,i,\max}$ 为机组 i 运行时的出力上限; $T_{G,i,t}$ 为机组 i 在 t 时刻连续运行或停机的总时间数; $T_{on,i}$ 为机组 i 的最小连续运行时间数; $T_{off,i,t}$ 为机组 i 最小连续停机时间数。

2 强化学习框架下的机组组合描述

本文利用深度强化学习方法求解机组组合问题,首先需要在强化学习框架下完成机组组合描述转化。强化学习主要包含 3 种基本信号: 状态信号 S 、动作信号 A 和奖励信号 R 。在本文中, 机组系统是强化学习的环境, 智能体根据所含策略对从环境中获取包含机组状态的状态信号采取动作, 以获得成本最小的最优机组动作决策。

2.1 状态空间

对机组组合问题来说, 其状态空间应包括各机组启停状态、各机组连续运行/停运总时间数、以及负荷需求。因此, 对于机组组合的状态空间 S 可以表示为

$$S = \{P_{G,i,t} (i=1, \dots, N), T_{G,i,t} (i=1, \dots, N), P_{L,t}\} \quad (7)$$

2.2 动作空间

在机组组合问题中, 本文采用深度强化学习算法来求解机组启停计划, 并结合 LAMBDA 迭

代法在机组启停计划基础上求解机组出力计划。因此机组组合问题智能体的动作空间 A 应包含机组的启停动作,以含 5 台机组的系统为例,机组启停计划共有 $2N$ 种,则智能体相应地包含 $2N$ 个动作组合,可以表示为

$$A = \{A_G\} \quad (8)$$

式中: A_G 为 N 台机组的启停动作组合的集合。

2.3 提前终止机制

智能体在探索初期时,策略模型还未训练成熟,所采取的动作无法满足约束条件。因此对于这类约束条件需要设置提前终止机制进行处理。

(1)最小启停时间约束

智能体根据环境的状态采取机组启停动作,若机组启停动作不能满足最小启停时间约束,则该时刻以后的所有动作决策都不再具有意义。因此,需要设置提前终止机制以降低错误探索经验对模型训练的影响。

(2)功率平衡约束

本文求解机组组合问题时划分为机组启停计划和机组出力计划 2 个子问题进行求解。采用深度强化学习算法求解机组启停计划,采用 LAMBDA 迭代法求解机组出力计划,由于机组组合问题的拆分化使得求解过程存在分离性。功率平衡约束由 LAMBDA 迭代法进行考虑,若智能体给出的机组启停动作无法从根本上满足机组出力与负荷的平衡关系,则该动作决策是没有意义的。因此,需要通过设置提前终止机制来评判机组启停计划能否从根本上满足功率约束平衡。

2.4 奖励函数

在一般情况下,智能体的奖励应当由目标函数来表示,即:

$$r_t = -W_{G,t} \quad (9)$$

(1)当无法满足最小启停时间约束时,需要在式(9)叠加惩罚性质的奖励项,设置系数为 r_s ,同时需要乘以提前终止的时间数,如式(10)所示:

$$r_t = -W_{G,t} - r_s(T - T_{\text{end}}) \quad (10)$$

(2)当无法满足功率平衡约束时,需要在式(9)叠加惩罚性质的奖励项,设置系数为 r_d ,同时乘以功率差额,如式(11)所示:

$$r_t = -W_{G,t} - r_d \left| P_{L,t} - \sum_{i=1}^N P_{G,i,t} \right| \quad (11)$$

(3)当 2 种约束条件同时无法满足时,奖励函数如式(12)所示:

$$r_t = -W_{G,t} - r_s(T - T_{\text{end}}) - r_d \left| P_{L,t} - \sum_{i=1}^N P_{G,i,t} \right| \quad (12)$$

式中: r_t 为智能体在 t 时刻采取动作后获得的奖励; r_s 为系统无法满足最小启停时间约束时设置的惩罚系数; r_d 为系统无法满足功率平衡约束时设置的惩罚系数; T 为调度总时间; T_{end} 为终止时刻。

3 机组组合问题的深度强化学习求解方法

经典强化学习方法中由于其算法是以表格形式来存储状态动作对 (s, a) 的策略模型,当状态维数增大时,可能会造成维数灾难的问题。深度强化学习具有强化学习的决策能力以及深度学习的感知拟合能力,能够更好地解决机组组合问题。机组组合问题中机组启停动作是离散动作,而本文所采用的 DQN 算法能够很好地针对离散动作空间的问题进行求解。DQN 算法^[12]采用一个深度神经网络 θ 来拟合状态动作值函数 Q ,即获得策略模型,且这个网络具有一个目标网络 θ' 。

深度神经网络训练过程是通过最小化损失函数 $L(\theta)$ 来优化网络参数的:

$$L(\theta) = E(y_t - Q(s_t, a_t | \theta))^2 \quad (13)$$

式中: y_t 为目标 Q 值,如式(14)所示。

$$y_t = r_t + \gamma \max Q(s_{t+1}, a_{t+1} | \theta') \quad (14)$$

对网络参数的更新采用梯度下降算法,如式(15)所示。

$$\theta_{t+1} = \theta_t + \alpha [r_t + \gamma \max Q(s_{t+1}, a_{t+1} | \theta') - Q(s_t, a_t | \theta_t)] \nabla_{\theta_t} Q(s_t, a_t | \theta_t) \quad (15)$$

式中: α 为学习率; γ 为折扣因子。

每更新当前网络参数 C 次后,再根据式(16)将当前网络参数复制到目标网络中。

$$\theta' = \theta \quad (16)$$

本文建立基于 DQN 算法的电网机组组合训练框架,如图 1 所示。对于 DQN 算法,神经网络的输入为 $2N+1$ 维状态向量,输出是 N 维动作向量,即神经网络拟合的是状态动作值函数 $Q(s_t, a_t)$ 。

在 t 时刻,首先 DQN 智能体从机组环境中获取的状态 s_t ,包含各机组的启停状态、各机组连续运行或停机总时间数、以及负荷需求的状态值,再依据智能体的策略模型采取动作 a_t ,该决策动作即为机组组合中的机组启停计划,将机组启停计划输入到 LAMBDA 迭代法中求解机组出力计

划,再将机组启停计划和出力计划输入到机组系统环境中,则进入下一时刻状态 s_{t+1} ,并返回一个以发电成本、启停成本和提前终止惩罚的总和为成本的奖励值 r_t 。由此得到了一个数据序列 (s_t, a_t, r_t, s_{t+1}) ,将其保存在经验回放池中。DQN 算法设置经验回放池的目的是打破状态数据间的相关性,以更好地适用深度神经网络对输入数据是独立同分布的要求。DQN 智能体将从经验回放池中随机选取批量数据,并采用梯度下降法最小化损失函数来更新网络参数。DQN 算法求解机组组合流程框图如图 1 所示。

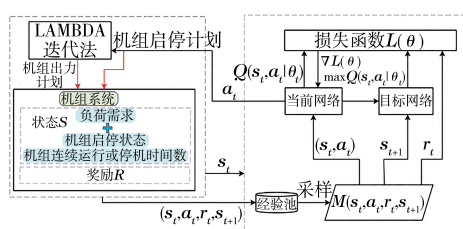


图1 DQN算法求解电网机组组合流程

4 算例分析

本文将针对含5台机组的算例系统进行仿真实验,以验证所提方法的有效性。5台机组的参数设置采用文献[13]中机组参数。DQN 算法超参数设置如表1所示。

表1 DQN算法超参数设置

项目	参数
学习率 α	0.000 01
折扣因子 γ	0.8
神经网络参数	隐藏层数为3层,神经元个数分别为400,200,100
训练总迭代次数	2×10^5
经验回放池容量	10^6

4.1 训练情况分析

记录当前迭代次数之前100次调度周期总奖励的平均值,并对曲线进行平滑处理,其结果如图2所示。从图2中可以看出,智能体在初始探索阶段中所获得的平均奖励值小。随着迭代次数的不断增加,智能体策略模型的参数不断优化,机组决策计划也会逐渐趋于成本最小化,平均调度周期奖励曲线则逐渐收敛。在训练过程中,由于提前终止机制的设置,当智能体给出的决策动作无法满足约束条件时,会获得较大的负值奖励。因此,当智能体触发提前终止机制时,就会在迭代次数 $4.5 \times 10^4 \sim 5 \times 10^4$ 区间中出现较大的下降波动。

当曲线逐渐趋于稳定时,仍然存在一定的波

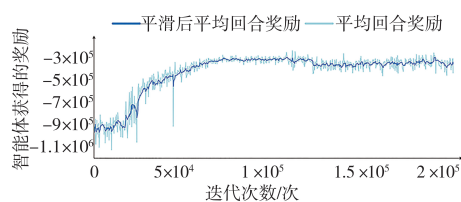


图2 平均调度周期奖励

动性,这是正常现象。在面对不同调度日时,负荷需求不同则对应的成本最优的动作决策不同,因此智能体获得奖励值也会不同,同时智能体在后期训练阶段仍然保持一定概率探索动作空间,所以仍然存在探索到较小奖励值的动作决策的可能性,由此两者共同影响平均总回报曲线的波动和振荡现象。

图3为损失函数曲线,随着回合次数的增加,损失函数逐渐减小并趋于稳定,由于本文采用的DQN算法每完成一个周期探索完成一次模型训练,因此记录次数与迭代次数相同。

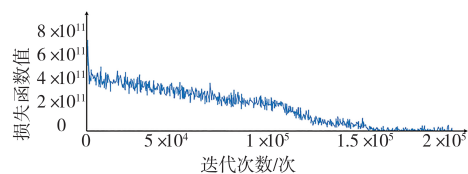


图3 损失函数曲线

图4为提前终止次数曲线,随着迭代次数的增加,提前终止次数在初始探索训练过程中,每1000次调度周期会大量触发提前终止机制。在经验回放池的样本数据不断积累过程中,智能体的模型训练朝着获得更大奖励的方向更新,使每1000次调度周期中发生的提前终止次数减少。当智能体逐渐趋于稳定时,提前终止次数仍然会稳定在一定范围内,这是由于训练后期阶段中仍然保持有一定概率探索动作空间。

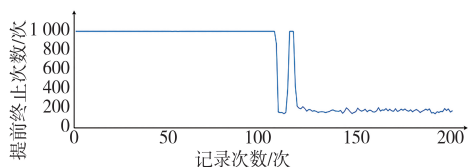


图4 提前终止次数曲线

4.2 调度结果分析

经过 2×10^5 次迭代训练(共耗时 52 min30 s)后,将训练成熟的模型对某一调度日进行测试以求解机组组合问题。采用 DQN 求解调度日的机组组合问题中得到机组出力与负荷需求曲线见图5。表2表示的是测试日机组总发电成本及智能体的决策时间。由表2可知,应用训练成熟的模

型求解调度日的机组组合问题时智能体给出决策动作(即机组启停计划和机组出力计划)的时间为0.07 s,在理论上与传统优化算法相比,DQN算法在求解机组组合问题上具有快速性。同时,当面对具有不同负荷需求的调度日时,DQN智能体仍然能够快速给出合理的决策动作,相比于传统优化算法而言,无需重复计算求解,因此,DQN算法在求解机组组合问题上具有泛化能力。由图5可以看出,对于调度日的机组出力计划能够较好地满足负荷平衡的需求。

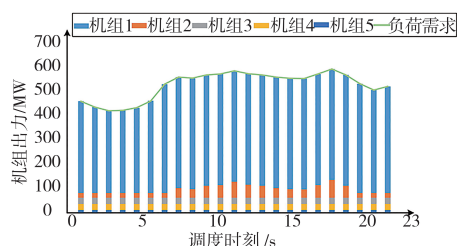


图5 调度日机组出力与负荷需求曲线

表2 调度日发电成本与决策时间

调度方法	燃料成本 /万元	启停成本 /万元	发电成本 /万元	决策时间 /s
DQN	198.77	1.22	199.98	0.07

由此可知,深度强化学习算法下的模型在训练成熟之后,针对本文所研究的决策问题上具有较高的决策效率。同时离线探索训练的时间并不影响模型的决策结果,当在相同机组参数设置下时,能够具有较好的泛化能力,不需要再花费大量的时间进行探索求解。

5 结论

本文提出了一种基于DQN算法的电网机组组合算法。将机组组合问题划分为机组启停计划和机组出力计划,通过DQN算法来求解机组启停计划并结合LAMBDA迭代法来求解机组出力计划。与传统算法相比,这种算法在模型训练完成后不需要对不同的调度日进行求解,而能够通过智能体的策略模型直接给出动作决策。根据模型训练及测试结果分析,验证了本文所提方法在机组组合问题求解上的有效性和优越性。

本文对于机组组合问题划分子问题的提出,使得求解问题的分离性,因此,在今后的工作中,为了更好地避免2个子问题的划分,需要对机组出力这类连续动作进行考虑及改进。本文在机组组合问题的求解思路及算法应用上,能够给

这类大规模混合整数规划问题的研究提供一种思路和解决方法。

参考文献:

- [1] MILAD T, WIM V A, ANTONIO F, et al. Large-scale unit commitment under uncertainty[J]. Quarterly Journal of the Belgian, French and Italian Operations Research Societies, 2015, 13(2): 115-171.
- [2] MORADI S, KHANMOHAMMADI S, HAGH M T, et al. A semi-analytical non-iterative primary approach based on priority list to solve unit commitment problem[J]. Energy, 2015, 88(8): 244-259.
- [3] SMRITI J, RAMESH K P, LATA G. Reliability Constrained Day Ahead Unit Commitment with Optimal Spinning Reserve Allocation for Solar Integrated Power System[C]. Institute of Electrical and Electronics Engineers, 5th IEEE International Conference on Recent Advances and Innovations in Engineering, 5th IEEE International Conference on Recent Advances and Innovations in Engineering, 2020: 1-6.
- [4] 何小宇, 张粒子, 谢国辉. 改进的拉格朗日松弛法求解机组组合问题[J]. 电力系统保护与控制, 2010, 38(17): 16-21.
- [5] 冀子臻, 顾圣平. 基于遗传萤火虫算法的含风电电力系统机组组合研究[J]. 水电能源科学, 2018, 36(12): 152-155.
- [6] NEMATI M, BRAUN M, TENBOHLEN S. Optimization of unit commitment and economic dispatch in microgrids based on genetic algorithm and mixed integer linear programming[J]. Applied Energy, 2018, 210(1): 944-963.
- [7] ZAND A, BIGDELI M, AZIZIAN D. A Modified ant colony algorithm for solving the unit commitment problem[J]. Advanced Energy: An International Journal(AEIJ), 2016(3): 15-27.
- [8] 李颖浩, 郭瑞鹏. 求解机组组合问题的多种群混沌蚁群算法[J]. 电力系统保护与控制, 2012, 40(9): 13-17.
- [9] SHUKLA A, LAL V N, SINGH S N. Profit-based unit commitment problem using PSO with modified dynamic programming[C]. IEEE, 2015 18th International Conference on Intelligent System Application to Power Systems (ISAP). 2015: 1-6.
- [10] 李 整. 基于粒子群优化算法的机组组合问题的研究[D]. 北京: 华北电力大学(北京), 2016.
- [11] 杨 挺, 赵黎媛, 王成山. 人工智能在电力系统及综合能源系统中的应用综述[J]. 电力系统自动化, 2019, 43(1): 2-14.
- [12] VOLODYMYR M, KORAY K, DAVID S, et al. Human-level control through deep reinforcement learning[J]. Nature, 2019, 518(7540): 29-33.
- [13] 熊 虎, 向铁元, 陈红坤, 等. 含大规模间歇式电源的模糊机会约束机组组合研究[J]. 中国电机工程学报, 2013, 33(13): 36-44.

本文责任编辑:王洪娟