DOI: 10.19783/j.cnki.pspc.230917

基于深度强化学习的 Π 型阻抗匹配网络多参数最优求解方法

胡正伟，夏思懿，王文彬，曹旺斌，谢志远

（华北电力大学电子与通信工程系，河北 保定 071003）

摘要：针对电力线信道阻抗变化复杂、负载阻抗不匹配造成通信质量差等问题，提出一种基于深度强化学习的 Π 型阻抗匹配网络多参数最优求解方法，并验证分析了深度强化学习对于寻找最优匹配参数的可行性。首先，建立 Π 型网络结构，推导窄带匹配和宽带匹配场景下的最优匹配目标函数。其次，采用深度强化学习，利用智能体的移动模拟实际匹配网络的元件参数变化，设置含有理论值与最优匹配值参数的公式作为奖励，构建寻优匹配模型。然后，分别仿真验证了窄带匹配和宽带匹配两种应用场景并优化模型的网络参数。最后，仿真结果证明，经过训练后的最优模型运行时间较短且准确度较高，能够较好地自动匹配电力线载波通信负载阻抗变化，改善和提高电力线载波通信质量。

关键词：深度强化学习；电力线通信；窄带匹配；宽带匹配

Multi-parameter optimal solution method for Π-type impedance matching networks based on deep reinforcement learning

HU Zhengwei, XIA Siyi, WANG Wenbin, CAO Wangbin, XIE Zhiyuan

(Department of Electrical & Electronic Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: There are problems of complex power line channel impedance variation and poor load impedance mismatch. Thus a multi-parameter optimal solution method for a Π-type impedance matching network based on deep reinforcement learning is proposed, and the feasibility of deep reinforcement learning for finding the optimal matching parameters is verified and analyzed. First, the Π-type network structure is established to derive the objective function for the optimal matching in the narrowband matching and broadband matching scenarios. Secondly, deep reinforcement learning is used to use the movement of the agent to simulate the component parameters of the actual matching network, and set the formula containing the theoretical value and the optimal matching value of the parameters as a reward to build the optimal matching model. Then, this paper separately verifies the network parameters of narrowband matching and broadband matching application scenarios and optimizes the network parameters of the model. Finally, the simulation results prove that the trained optimal model has short running time and high accuracy. It can better automatically match the load impedance change of power line carrier communication, and improve the quality of power line carrier communication.

This work is supported by the General Program of National Natural Science Foundation of China (No. 52177083).

Key words: deep reinforcement learning; power line communication; narrowband matching; broadband matching

0 引言

随着科技的进步，电力线通信技术飞速发展，对电力线载波通信质量也提出了更高的要求[1-3]。电力线环境比较恶劣，负载阻抗变化复杂，当信号源阻抗与负载阻抗不等时，传输线上会存在入射波和反射波，传送的功率不能完全被负载吸收，从而降低接收端负载的接收功率，导致信号不能有效地传输到接收端，降低通信质量。为了实现从信号源到负载的最大功率转移，阻抗匹配是必需环节。

电力线通信系统按传输的频带宽度可以分为窄带电力线系统和宽带电力线系统。窄带电力线系统可以提供各种自动化和控制应用等服务[4]，数据的传输速率较低；宽带电力线载波通信[5-6]，指载波信号工作频率在 2 MHz 以上的电力线通信技术，可以

基金项目：国家自然基金面上项目资助(52177083)；国家自然科学基金青年科学基金项目资助(62001166)

提供多种电信服务，数据的传输速率可达 2 Mbits/s 以上。窄带电力线系统在频率上对网络的阻抗做了简化处理，适用于网络阻抗随频率变化不大的情况[7]。宽带电力线通信系统网络拓扑结构比较复杂，线路阻抗随着频率的变化较大，对通信来说是非常不利的。因此有必要研究匹配网络如何分别在窄带、宽带电力系统中进行阻抗匹配。

目前对于实现电力线阻抗匹配功能的方法，主要分为两类。一类是对网络结构进行分析[8-9]，文献[10-11]通过改变匹配电路中的器件匝数和电容、电感值的大小进行阻抗匹配。文献[12]提出了一种基于非线性优化方法实现宽带阻抗匹配电路的方法，消除耦合电容器和耦合变量器漏感引起的工作衰减。另一类是对匹配网络的器件参数进行优化，使用遗传算法、粒子群优化算法等智能算法寻找最优匹配参数值。一般多采用 L 型匹配网络进行阻抗匹配[13-16]，这种方法电路结构简单，无法应用于宽频匹配领域。多级匹配网路结构复杂，只能应用于宽带领域，无法适用于窄带匹配领域。相比之下，Π 型匹配网络具有谐波抑制能力强、电路结构相对简单、适应场景广泛、不存在匹配禁区等优势，可以更好地与负载阻抗、电路工作频率进行匹配。

由于Π型阻抗匹配网络的问题中包含若干等式约束，可以将其转化为多目标约束优化问题[17]，采用智能算法解决此类问题。文献[18]提出了乌鸦搜索算法(crow search algorithm, CSA)，通过基于乌鸦的智能行为来解决目标约束优化问题，此算法需要较多的目标函数和约束条件。文献[19]提出在模型预测控制的框架下，将多目标问题转化为约束二次规划问题，但是在后续的算法使用过程中涉及到的参数数量较大，实验较为复杂。本文采用深度强化学习的算法，深度强化学习参数少、准确度较高，且较于其他智能算法，增加经验池、目标网络的特点可以有效避免智能算法陷入局部最优的情况。

深度强化学习方面，文献[20]最先使用一个多层感知器来近似表示 Q 值函数，并提出了神经拟合 Q 迭代(neural fitted Q iteration, NFQ)算法。文献[21]将卷积神经网络与 Q 学习算法相结合，提出了深度 Q 网络(deep Q-network, DQN)模型。DQN 算法在电网系统中得到广泛应用[22-24]。文献[25]将电压管控问题转化为马尔科夫决策问题，采用多智能体深度确定性策略梯度算法对智能体进行离线训练。

为了解决 Π 型匹配网络的多参数最优求解问题，本文提出了基于深度强化学习的阻抗匹配网络参数估计方法。结合特定的Π型阻抗匹配网路，设置电容电感的取值范围空间，使用深度强化学习方法在搜索空间中找到阻抗匹配公式的最优解，以此实现匹配功能。本文利用深度强化学习具有识别环境动态变化并通过学习优化自身行为的特性，动态调整匹配网路的电容、电感值，以此在整个通信带宽下实现最大传输功率。并在窄带匹配环境和宽带匹配环境下分别对电路进行了仿真对比分析，验证了深度强化学习算法的精准性、有效性。

1　理论依据

阻抗匹配是改善电力线载波通信的一个重要研究技术，在电路端口加入无源网络，改善电路端口的等效输入阻抗。图 1 为阻抗匹配系统的工作原理图，网络结构选择为可调Π型阻抗匹配网络，因此需要根据阻抗测量结果调整电感 L、电容 C1 和电容 C2 的取值，以此实现阻抗匹配。

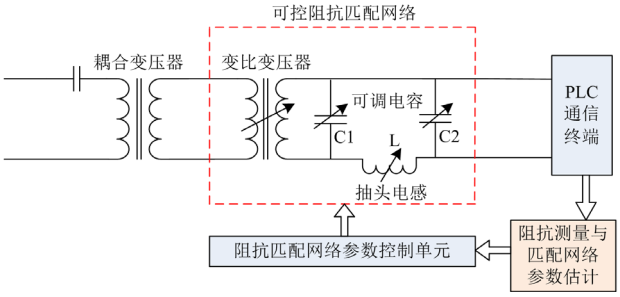

图 1　阻抗匹配系统的工作原理图

Fig. 1 Working principle of impedance matching system

Π型阻抗匹配网路的等效阻抗原理图如图2所示。

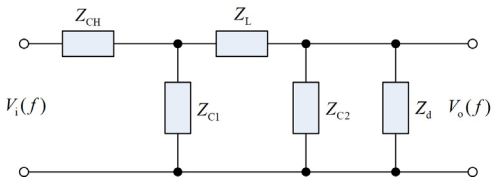

图 2　Π 型网络等效阻抗原理图

Fig. 2 Equivalent impedance schematic of Π-type network

式(1)、式(2)给出了匹配网络的输出电压 $V_o(f)$ 及传递函数 $H(f)$ 与阻抗之间的计算表达式。

$$V_o(f)=\frac{(Z_L+Z_{C2}//Z_d)//Z_{C1}}{Z_{CH}+(Z_L+Z_{C2}//Z_d)//Z_{C1}}\cdot\frac{Z_{C2}//Z_d}{Z_L+Z_{C2}//Z_d}\cdot V_i(f) \tag{1}$$

$$H(f)=\frac{V_o(f)}{V_i(f)}=\frac{Z_{CH}+(Z_L+Z_{C2}//Z_d)//Z_{C1}}{(Z_L+Z_{C2}//Z_d)//Z_{C1}}\cdot\frac{Z_L+Z_{C2}//Z_d}{Z_{C2}//Z_d} \tag{2}$$

式中：$V_o(f)$为输出电压；$V_i(f)$为输入电压；Z_{CH}为信道的阻抗值；Z_{C1}、Z_{C2}分别为电容 C1、电容 C2 的阻抗值；Z_L为电感 L 的阻抗值；Z_d为通信终端负载。

匹配网络中的阻抗参数 $Z_L=\mathrm{j}2\pi fL$、$Z_{C1}=\mathrm{j}2\pi fC_1$、$Z_{C2}=\mathrm{j}2\pi fC_2$，可通过调整电容 C1、电容 C2 和电感 L 的阻抗值使传递函数 $H(f)$ 达到最优匹配要求。因此，图 2 中的信道阻抗状态函数$|H_k(f)|$及其对应的最优匹配状态函数$|H_k(f)|_{\mathrm{opt}}$由式(3)表示。

$$\begin{cases}|H_k(f)|=\left|\dfrac{Z_{CHk}+(Z_{Li}+Z_{C2i}//Z_d)//Z_{C1i}}{(Z_{Li}+Z_{C2i}//Z_d)//Z_{C1i}}\cdot\dfrac{Z_{Li}+Z_{C2i}//Z_d}{Z_{C2i}//Z_d}\right| \\ |H_k(f)|_{\mathrm{opt}}=\left|\dfrac{Z_{CHk}+(Z_{Lk}+Z_{C2k}//Z_d)//Z_{C1k}}{(Z_{Lk}+Z_{C2k}//Z_d)//Z_{C1k}}\cdot\dfrac{Z_{Lk}+Z_{C2k}//Z_d}{Z_{C2k}//Z_d}\right|\end{cases} \tag{3}$$

式中：Z_{CHk}为信道 k 时刻的阻抗；Z_{C1i}、Z_{C2i}、Z_{Li}分别为匹配网络电容 C1、C2 和电感 L 的初始状态参数取值；Z_{C1k}、Z_{C2k}、Z_{Lk}分别为匹配网络电容 C1、C2 和电感 L 在信道 k 时刻的最佳匹配参数取值。

因此匹配问题可看作对式(3)的多参数求解问题，即给出(Z_{C1k},Z_{C2k},Z_{Lk})的组合值满足$|H_k(f)|_{\mathrm{opt}}$。为了降低计算量，本文将多参数求解问题转换为从(Z_{C1i},Z_{C2i},Z_{Li})到(Z_{C1k},Z_{C2k},Z_{Lk}) 3 个参数取值组合的寻优问题，即寻找一个最佳组合(Z_{C1k},Z_{C2k},Z_{Lk})满足式(3)中的$|H_k(f)|_{\mathrm{opt}}$。

图 3 为阻抗网络参数最优选择的几何空间示意图，从空间几何角度看，将阻抗匹配问题转换为从坐标(Z_{C1i},Z_{C2i},Z_{Li})到寻找满足式(3)的最佳坐标(Z_{C1k},Z_{C2k},Z_{Lk})的寻优过程。但该点的坐标未知，需要不断在边界为 $x=Z_{C1\_max}$、$y=Z_{C2\_max}$、$z=Z_{L_max}$的立方体内寻找。

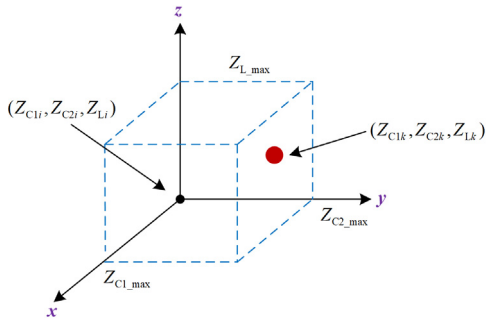

图 3 阻抗网络参数最优选择的几何空间示意图

Fig. 3 Geometric space diagram for optimal selection of impedance network parameters

电力线匹配问题可分为窄带匹配和宽带匹配。窄带匹配主要解决当频点固定时，负载阻值变化，匹配网络各参数随之变化的问题。宽带匹配主要解决当负载阻值固定时，一段频率范围内取不同频点数，各频率点的匹配效果随匹配网络参数变化而变化的问题。

1.1 窄带匹配

为了保证接收端负载的接收功率最大，需在信号源与负载间加入匹配网络，调节匹配网络参数，使信号源阻抗、负载与匹配网络所构成的等效阻抗达到共轭匹配条件。

窄带情况下的阻抗匹配，假设电路只在特定的频率点工作，在此基础上进行相应的匹配计算。如图 2 Π 型网络等效阻抗图，电源被等效为信号源$V_i(f)$与信道阻抗 Z_{CH} 串联，通信终端负载 $Z_d=R+\mathrm{j}X$；而电感 L、电容 C1、电容 C2 则构成了连接电源与负载之间的 Π 型匹配网络。

根据图 2 可计算出各段电阻值为

$$Z_1=\frac{R+\mathrm{j}X}{1-wXZ_{C2}+\mathrm{j}wZ_{C2}R} \tag{4}$$

$$Z_2=\mathrm{j}wZ_L+Z_1 \tag{5}$$

$$Z_{in}=\frac{Z_2}{1+\mathrm{j}wZ_{C1}Z_2} \tag{6}$$

式中：R 为电力线负载的阻抗实部值；X 为电力线负载阻抗虚部值；w 为电路所处的工作频率；Z_1为负载 Z_d 与电容 C2 并联后的等效电阻；Z_2 为等效电阻 Z_1 与电感 L 串联后的等效总电阻；Z_{in} 为负载与 Π 型匹配网络所构成的输入阻抗。

当输入阻抗 Z_{in} 与负载 Z_d 达到共轭匹配时，实现阻抗匹配效果。

1.2 宽带匹配

宽频阻抗匹配问题要考虑整个频带范围内的频点，而不像窄带匹配，只针对某一频率。因此引入转换功率增益(transducer power gain, TPG)这一概念。对于任意的负载阻抗和信号源的情况下，定义负载吸收的实际功率和最大功率 P_L 的比值为转化功率增益 T_{PG}，如式(7)和式(8)所示。

$$T_{PG}(W_k)=\frac{P_L}{P_A}=\frac{4R_{CH}(W_k)R_{in}(W_k)}{(R_{CH}(W_k)+R_{in}(W_k))^2+(X_{CH}(W_k)+X_{in}(W_k))^2} \tag{7}$$

$$f_{itness}=\frac{1}{m}\sum_{i=1}^{m}T_{PG}(W_k) \tag{8}$$

式中：$T_{PG}(W_k)$为频点 W_k 处的转换功率增益；P_A 为

转移功率增益最大时负载的接收功率；$R_{\mathrm{CH}}(W_k)$、$X_{\mathrm{CH}}(W_k)$分别为信道阻抗的实部和虚部；$R_{\mathrm{in}}(W_k)$、$X_{\mathrm{in}}(W_k)$分别为负载与匹配网路等效输入阻抗的实部和虚部；f_{itness}为衡量宽带匹配标准的目标函数；m为整个频带范围内的频点数。

当到达宽频最优匹配$R_{\mathrm{in}}=R_{\mathrm{CH}}$，$X_{\mathrm{in}}=-X_{\mathrm{CH}}$时，$f_{\mathrm{itness}}$函数值可达到最优值 1。

2　基于 DQN 的匹配网络参数求解建模

2.1 DQN 算法原理

通过以上对窄带、宽带匹配网络进行分析，本文采取 DQN 的方法[26-27]，令智能体在搜索空间内反复进行探索，寻找最优坐标点，并输出到坐标点的最优路径，从而得到最优匹配结果。DQN 算法训练如图 4 所示。

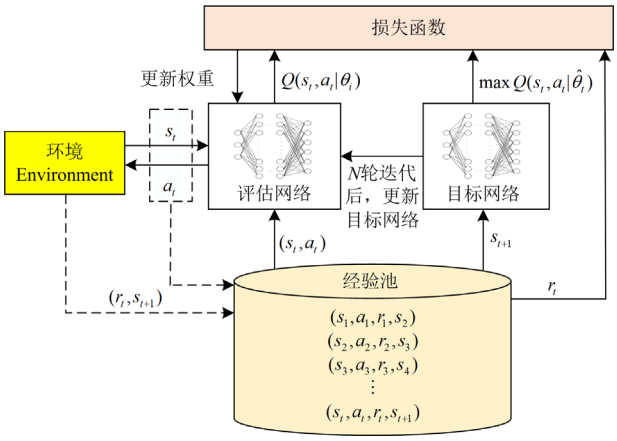

图 4 DQN 算法训练图

Fig. 4 Training diagram of the DQN algorithm

传统的强化学习算法通过建立 Q-table 存储索引后的价值函数，表格存储量较小，运行较慢，效果不明显。深度强化学习算法[28-29]，用神经网络替换传统强化学习算法的 Q-table，以估计每个动作的价值，完成从离散状态空间到连续状态空间的转变，将智能体在搜索空间的状态作为 Q Network 的输入，输出每个动作对应的 Q 值，得到将要执行的动作，效果优于传统的强化学习。

深度强化学习的任务就是找到一个满足匹配式(6)、式(8)的最优策略。由图 4 可知，首先初始化两个评估网络 Q 和目标网络 $\hat{Q}$，并初始化神经网络的权重参数$\theta_i=\hat{\theta}_i$。在每一个回合中，采用智能体 Agent 与环境 Environment 交互的情况，在每一次交互的过程中，都会得到一个状态 s_t，智能体通过遍历状态 s_t 下所有动作的价值，选择价值最大的动作进行输出，因此定义策略 π 下的动作价值函数 $Q^{\pi}(s_t,a_t)$ 为

$$Q^{\pi}(s_t,a_t)=E[r_{t+1}+\gamma r_{t+2}+\gamma^2 r_{t+3}+\cdots+|s_t,a_t]=E_{s_{t+1}}[r_{t+1}+\gamma Q^{\pi}(s_{t+1},a_{t+1})|s_t,a_t] \tag{9}$$

式中：E 为均方误差函数；$E_{s_{t+1}}$ 为在状态 s_{t+1} 时的均方误差函数；r_{t+1} 为 $t+1$ 时刻的奖励值；γ 为折扣率，表示后续状态对当前状态的回报影响；s_t 为智能体 t 时刻的状态；a_t 为智能体 t 时刻所执行的动作。

智能体依据 Q 值最大的一个动作作为最优动作，得到最优动作价值函数表达式，如式(10)所示。

$$Q^{\pi^*}(s_t,a_t)=\max_{\pi}Q^{\pi}(s_t,a_t)=E_{s_{t+1}}[r_{t+1}+\gamma\max_{a_{t+1}}Q^{\pi^*}(s_{t+1},a_{t+1})|s_t,a_t] \tag{10}$$

式中：$Q^{\pi^*}(s_t,a_t)$ 为智能体执行最优策略 π^* 时，在状态 s_t 下执行最优动作 a_t 后获得的最大累积回报的期望值。

在智能体获得奖励后，将执行最优动作，并将所获得的经验样本 (s_t,a_t,r_t,s_{t+1}) 放到回放经验池中。智能体从经验池中随机抽取小批量经验样本数据。

接下来是计算每一个状态的目标函数，目标函数 Y_i 计算如式(11)所示。

$$Y_i=r+\gamma\max_{a_{t+1}}Q(s_{t+1},a_{t+1}|\theta_i) \tag{11}$$

式中：θ_i 为神经网络权重参数；$Q(s_{t+1},a_{t+1}|\theta_i)$ 为评估网络的值函数。

计算每一个状态的目标值，通过目标网络 $\hat{Q}$ 执行动作后的奖励更新 $\hat{Q}$ 值。损失函数 $L(\theta_i)$ 为

$$L(\theta_i)=E_{(s_t,a,r,s_{t+1})}[(Y_i-Q(s_t,a_t|\theta_i))^2] \tag{12}$$

式中：$Q(s_t,a_t|\theta_i)$ 为目标网络的值函数；$E_{(s_t,a,r,s_{t+1})}$ 为经验样本 (s_t,a,r,s_{t+1}) 时的均值函数。通过对损失函数中的参数 θ_i 求偏导，得到梯度公式如式(13)所示。

$$\nabla_{\theta_i}L(\theta_i)=E_{(s_t,a,r,s_{t+1})}[(Y_i-Q(s_t,a_t|\theta_i))\nabla_{\theta_i}Q(s_t,a_t|\theta_i)] \tag{13}$$

式中，∇_{θ_i} 为对损失函数中的参数 θ_i 求偏导。

使用梯度下降法更新评估神经网络的参数 θ_i。经过 N 次迭代后，将当前目标网络权重参数传递给评估网络，对权重参数进行更新。DQN 算法的最终目标是随着迭代次数增加，寻找当 $\hat{Q}=Q$ 时的最优位置。

2.2 深度强化学习的步骤

深度强化学习的算法流程如图 5 所示，代码具体实现步骤如下。

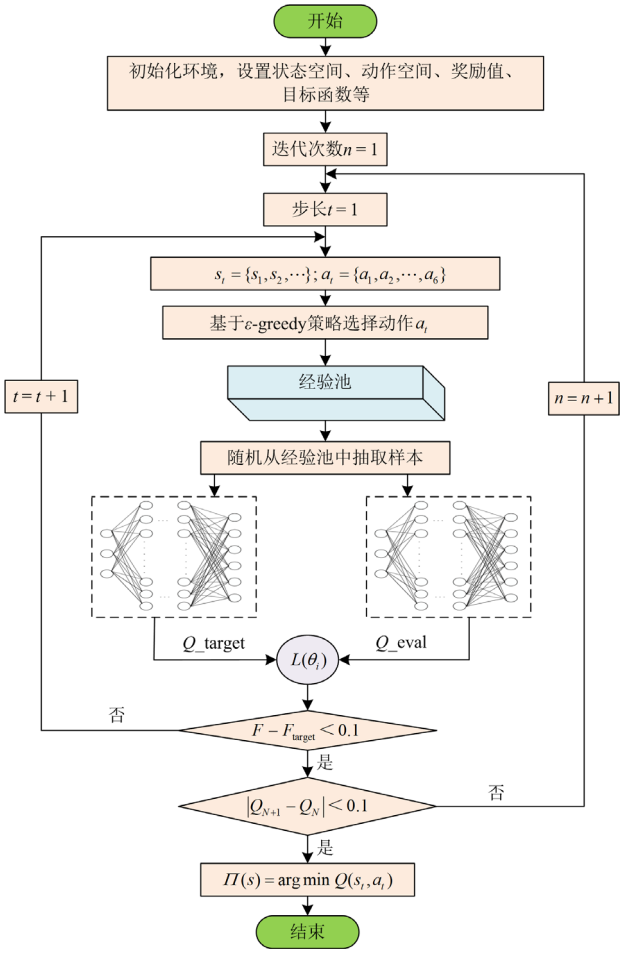

图 5 DQN 算法流程图

Fig. 5 Flow chart of the DQN algorithm

1) 搭建实验环境，初始化环境。设置状态空间、动作空间、奖励值、目标函数和深度强化学习的学习率、隐含层个数、经验池容量、衰减率等。

2) 探索阶段，智能体与环境进行交互，得到状态、动作信息，采用 ε-greedy 贪心策略搜索，即在网络产生决策的同时又能以一定概率探索其他可能的最优行为。

3) 存储和抽取经验样本。将所获取的经验样本存入至经验池中，不断更新经验池中的经验样本，并随机抽取小批量的经验样本作为评估网络、目标网络的输入。

4) 计算损失函数。计算两个网络的均方误差作为损失函数，以梯度下降算法更新评估网络的权重参数。

5) 判断算法所得函数值是否达到最优匹配时的目标函数值 F_{target}，若是，继续下一步；反之，重复流程。

6) 判断迭代结束。判断相邻迭代次数的累计评估 Q 值之间差值的绝对值 $|Q_{N+1}-Q_N|$ 是否小于一定的阈值，若是则迭代结束，否则继续迭代训练。

7) 确定并保存最优策略，结束程序。

2.3 基于 DQN 的阻抗匹配参数求解算法

智能体、环境、动作、状态、奖励函数是 DQN 算法的关键要素，在本文的应用场景下，代表的物理意义如下。

1) 智能体 Agent：将 Π 型匹配网络中的器件电容 C1、电容 C2、电感 L 对应的点坐标设置为智能体。在算法中，Agent 在空间内移动探索，对应实际应用中电容、电感器件数值的调整。

2) 动作 action：action 对应可调电容、电感数值的变化，在算法中将每个可调器件的 action 定义为两个，一共 3 个器件，共 6 个动作，分别为(1, 0, 0)、(−1, 0, 0)、(0, 1, 0)、(0, −1, 0)、(0, 0, 1)、(0, 0, −1)，即 Agent 在搜索空间下进行搜索时，Agent 可以向上或向下移动，表示在实际情况下，可调电容 C1、电容 C2、电感 L 的数值加一、减一。

3) 环境 Environment：即 DQN 算法中的搜索空间大小，器件电容 C1、电容 C2、电感 L 的可调范围分别对应搜索空间 x、y、z 轴的范围大小。

① 窄带场景

电力线载波通信的带宽一般为 3~500 kHz，负载变化范围在 10~500 Ω时，可调电容的调节范围一般在 150 nF 以内，可调电感的调节范围一般在 70 μH 以内。因此，搜索空间 x、y、z 轴分别设置为(1,150)、(1,150)、(1,70)。

② 宽带场景

频率范围为 2~50 MHz 时可调电容的调节范围在算法中一般在 1nF 以内，可调电感的调节范围一般在 1 μH 以内。因此，搜索空间 x、y、z 轴分别设置为(1,100)、(1,100)、(1,100)。

4) 状态 state：状态为 Agent 在搜索空间内每执行一个动作后，所处的位置和该位置下目标函数值。

① 窄带场景

在窄带匹配场景下，状态为坐标点位置和该坐标对应的输入等效阻抗值。

② 宽带场景

在宽带匹配场景下，状态为坐标点位置和对应该坐标点的 f_{itness}。Agent 在搜索空间内，依据动作随机走动，计算每一坐标点的 f_{itness}。

5) 奖励 reward：因为寻找满足最优匹配条件的坐标点是未知的，无法明确给定每一步的奖励值，因此将奖励设置为式(14)和式(15)，其中变量为坐标点的目标函数值与最优目标函数值的差值，以此来明确每一步的奖励。

① 窄带场景

在窄带匹配下，奖励函数 F_{reward} 表达式为

$$F_{\text{reward}}=\mathrm{e}^{-(t_{\text{arget}}-50)} \tag{14}$$

式中，t_{arget} 为深度强化学习算法计算出的等效输入阻抗值。

算法在搜索空间内进行搜索，计算每一个位置的等效输入阻抗值，将等效输入阻抗与信号源输入阻抗之间的差值作为奖励函数的变量。采用此奖励公式，当搜索结果越接近最优匹配结果，两者差值越小，奖励值越小，反之越大，以此达到搜索最优匹配效果。

② 宽带场景

奖励函数如式(15)所示。

$$F_{\text{reward}}=\mathrm{e}^{-|f_{\text{itness}}-1|}+100f_{\text{itness}} \tag{15}$$

算法通过计算每一个坐标点的目标函数 f_{itness}，将目标函数 f_{itness} 与理想状态 f_{itness} 为 1 时的差值作奖励函数的变量，采用此奖励函数公式。通过训练，可让智能体在训练过程中，得到最优匹配解，达到解决阻抗问题的最优效果。

3　仿真实验与分析

3.1　仿真参数设置

DQN 算法所用到的全连接神经网络结构如图 6 所示，神经网络输出 6 个动作 a_1 — a_6。

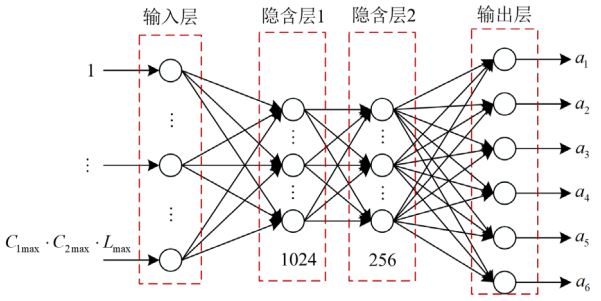

图 6　全连接神经网络结构图

Fig. 6 Convolutional neural network structure diagram

本文采用 DQN 算法进行训练，采用 stable_baseline3 的框架，DQN 算法的参数设置如表 1 所示。

3.2　深度强化学习的参数设置

3.2.1 边界设置

在 DQN 算法中，确定搜索空间大小是十分重要的。搜索空间过大，智能体的搜索时间长，导致算法运行时间较长；搜索空间过小，会导致搜索的最优结果不符合最优要求。

由于匹配问题属于多参数优化问题，存在多个最优解情况。图 7 右侧色彩条不同颜色代表 49.8~50.3 Ω的等效阻抗值。图 7 为窄带匹配最优点分布图，在电源阻抗为 50 Ω、负载阻抗为 150+50j 的情况下，在搜索空间内，符合最优匹配条件的坐标点。由图 7 可以看出，$C_1\in(0,100)$、$C_2\in(0,100)$、$L\in(0,50)$，在此范围构成的三维空间中可搜索到多个最优匹配值，因此将搜索空间设置在该区域内。

表 1　DQN 算法参数设置

Table 1 DQN algorithm parameter settings

DQN 算法参数	数值
窄带匹配学习率	0.003
宽带匹配学习率	0.0005
神经网络隐含层个数	2
神经网络隐含层点数	1024, 256
窄带搜索最大步数	200
宽带搜索最大步数	400
贪婪率	0.05
经验池存储样本数	2000
目标网络更新次数间隔	100

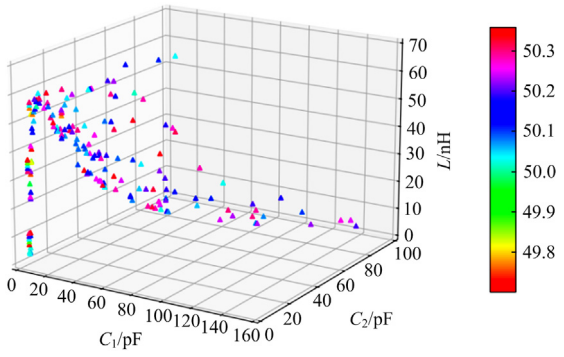

图 7　窄带匹配最优点分布图

Fig. 7 Narrowband matching optimal points distribution

在宽带匹配情况下，f_{itness} 函数作为衡量宽带匹配效果的衡量标准，其值越接近 1，负载接收功率越大，各频点的匹配效果越好。但在实际情况中，各频点下负载接收功率增益不能达到 100%，f_{itness} 值应是无限接近于 1 的小数。图 8 右侧色彩条表示不同 f_{itness} 的数值。图 8 为负载为 $150+50\text{j}$ 时，遍历整个搜索空间，每个坐标点 f_{itness} 情况分布。

由图 8 可知，当 $C_1\in(0,15)$、$C_2\in(0,10)$、$L\in(0,60)$ 时，f_{itness} 较大，约为 0.847，匹配效果较好。随着边界值增大，f_{itness} 逐渐减小，在 $C_1=50$ pF，$C_2=50$ pF，$L=1$ μH 时，f_{itness} 达到 0.071，没有参考意义。适当减少无参考意义的搜索空间，可以缩短搜索时间，提高搜索效率，所以将 x、y、z 轴边界设置为(0, 50)、(0, 50)、(0, 60)。

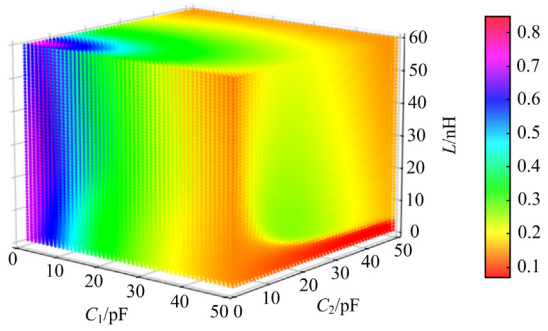

图 8 宽频匹配结果

Fig. 8 Broadband matching results

3.2.2 训练次数

图 9 为在负载150＋50j时，窄带匹配场景中不同频率下，随着训练迭代次数的增加，DQN 寻优所得等效输入阻抗最优值变化图。由图 9 可得，随着迭代次数的增加，训练次数在 10 000 次后，迭代效果趋于稳定，最优匹配所得等效阻抗约为 50.088 Ω，接近信号源阻抗 50 Ω，所以将窄带匹配的训练次数设置为 10 000 次。

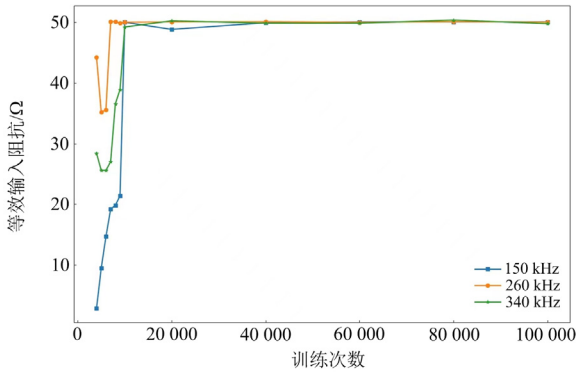

图 9 窄带匹配不同频率训练次数不同 DQN 所得最优结果图

Fig. 9 Optimal results obtained from DQN with different frequency training times for narrowband matching

图 10 为宽频匹配下，DQN 算法得出的最优 f_{itness} 随训练次数变化而变化的过程。训练次数达到 10 000 次，负载为100＋20j时，优化结果趋于 0.814；负载为300＋10j时，优化结果趋于 0.734；负载为450＋20j时，优化结果趋于 0.633。并随着训练次数的增加，f_{itness} 的结果值趋于稳定，因此宽带匹配的训练次数可设置为 10 000 次。

3.3 基于 DQN 的窄带实验结果与分析

利用 python 搭建窄带匹配网络仿真模型，窄带匹配网络采用一级 Π 型匹配网路，电路所处的工作频率分别选取 100 kHz、260 kHz 和 480 kHz 进行仿真分析。

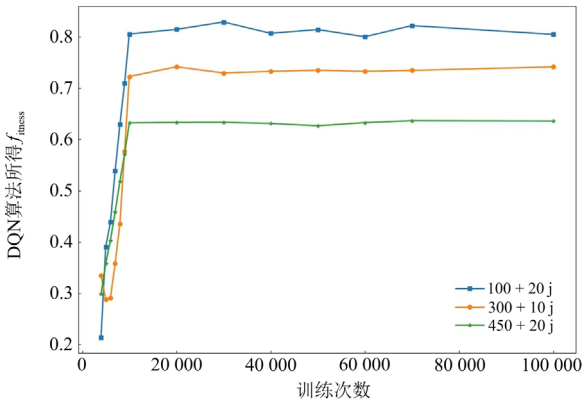

图 10 宽带匹配不同训练次数下 DQN 所得最优结果图

Fig. 10 Optimal results obtained from DQN with different training times under broadband matching

1) 等效输入阻抗值

等效输入阻抗值是负载与 Π 型电路的等效阻抗。当等效输入阻抗与信号源阻抗达成共轭匹配，电路可达到最优匹配状态，等效输入阻抗可以作为衡量 DQN 算法准确度的一个重要标准。在仿真建模中，在 100~500 kHz 中选取 150 kHz、260 kHz、340 kHz 3 个频率点，负载从(50, 500) Ω 范围内进行仿真。图 11 为在不同频率点下，负载为 10~500 Ω 时，DQN 算法计算所得的等效输入阻抗匹配结果与信号源阻抗的相对误差分布图。由图 11 可知，不同频率下，负载阻值在(50, 500) Ω 范围内，DQN 的匹配效果较好。

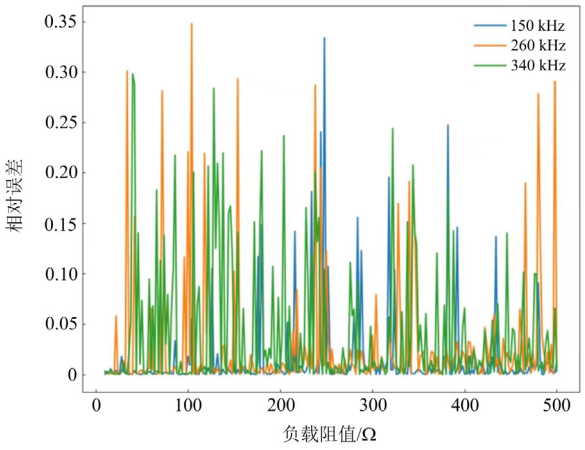

图 11 不同频率下匹配结果相对误差分布图

Fig. 11 Relative error distribution of matching results at different frequencies

2) 统计数据

单次实验数据具有偶然性，重复 DQN 窄带匹配代码 50 次，记录最优结果，如图 12 所示。当负

载为150＋50j时，频率分别为 150 kHz、260 kHz、340 kHz 时，经过 50 次的反复实验，DQN 算法所得最优坐标相对应的等效输入阻抗值均接近 50 Ω，50 次实验匹配效果达到 100%，证明 DQN 算法运行具有稳定性。

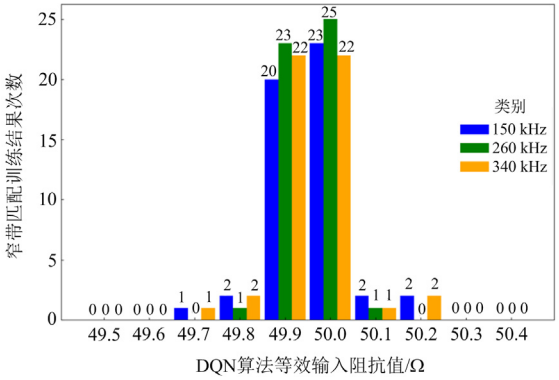

图 12　窄带匹配统计数据

Fig. 12 Narrowband matching statistics

3) 方法对比

将 DQN、遍历法与粒子群算法进行对比。遍历法首先通过对搜索空间进行量化处理，将搜索空间 x、y、z 轴以单位 1 进行进行量化，将搜索空间切割100×100×50 的坐标点，计算每一个坐标点的等效输入阻抗值，对比所有结果选取最优位置解。粒子群算法在窄带匹配下参数设置：加速度常数分别为 0.9 和 0.4，惯性权重为 0.6，粒子群数为 20，迭代次数为 100 次。图 13 描述了 3 种算法搜索到的最优值对比。

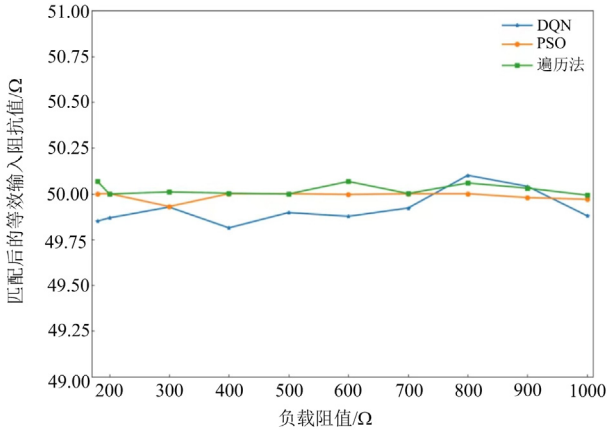

图 13　算法寻优值对比

Fig. 13 Comparison of algorithm search values

对比 3 种方法得到的最优匹配结果。遍历法找到最优点需要迭代 600 000 次，DQN 算法训练过程寻优点需要迭代 10 000 次，粒子群算法迭代需要 100 次。DQN 需要迭代次数小于遍历法，大于粒子群算法，最优匹配结果与理论值相差较小，效果较好。

运行时间是衡量算法时间复杂度的评价指标，运行时间越短表明算法的时间复杂度越小、适应性越好。使用 python 中的 datetime 模块对 3 种方法的运行程序进行计时，对比 3 种方法的运行时间如表 2 所示。DQN 算法经过训练后，保存了大量的经验样本和最优路径模型，其运行时间是指 DQN 经过训练后的最优模型运行时间。由表 2 可以看出 DQN 的运行时间小于遍历法和粒子群算法，运行时间较短。

表 2　DQN 窄带匹配方法对比

Table 2 Comparison of DQN narrowband matching methods

算法	遍历法	粒子群算法	DQN
运行时间/s	23.3457	0.5292	0.4489

通过方法对比，DQN 算法的特点如下。

① DQN 是一种基于神经网络的强化算法，可以通过大量的训练数据和奖励值来学习最优策略，在离散空间下获得最优解，匹配结果准确度较高，且运行时间较小。

② DQN 算法可以动态适应环境变化。DQN 算法的训练过程所需要的迭代次数较多，但是智能体经过训练后，可得到较好的训练效果并保存了最优模型，存储了大量的策略知识。当负载阻抗值变化后，不需再次训练，最优模型仍可以得到较好的匹配效果。

3.4　基于 DQN 的宽带匹配训练结果

1) 转换功率增益

转换功率增益 T_{PG}，用来衡量加入匹配网络后在各频率点负载的接收功率大小，各频点下 T_{PG} 越接近 1，证明负载接收功率越大，匹配效果越好。在仿真实验中频率范围设置在 2~50 MHz 内，设置频率点为 512 个。图 14—图 16 为在同一负载时，加入匹配网络并使用 DQN 算法寻优，各频率点的 T_{PG} 和无匹配网络时各频率点 T_{PG} 的变化趋势。通过运行 DQN 算法程序，当负载为 25＋30j时，通过 DQN 寻优算法得出结果为 $C_1=110$ pF，$C_2=170$ pF，$L=210$ nH，对应 f_{itness} 为 0.905；当负载为 200＋50j时，通过 DQN 寻优算法得出结果为 $C_1=20$ pF，$C_2=40$ pF，$L=340$ nH，f_{itness} 为 0.793；当负载为 400＋50j时，通过 DQN 寻优算法得出结果为 $C_1=10$ pF，$C_2=30$ pF，$L=520$ nH，f_{itness} 为 0.662。

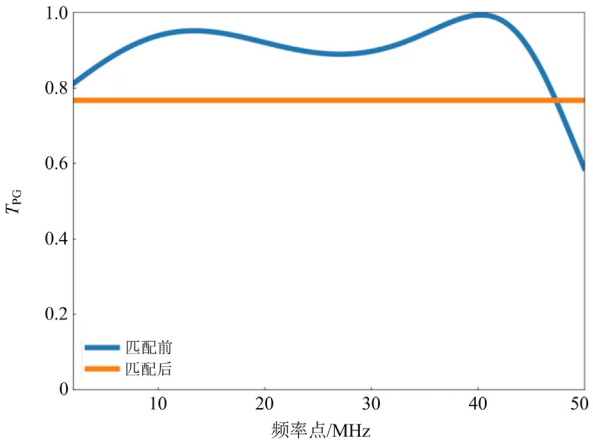

图 14　25 + 30j 宽频各频率点下 T_{PG} 值

Fig. 14 T_{PG} values at each frequency point of 25 + 30j broadband

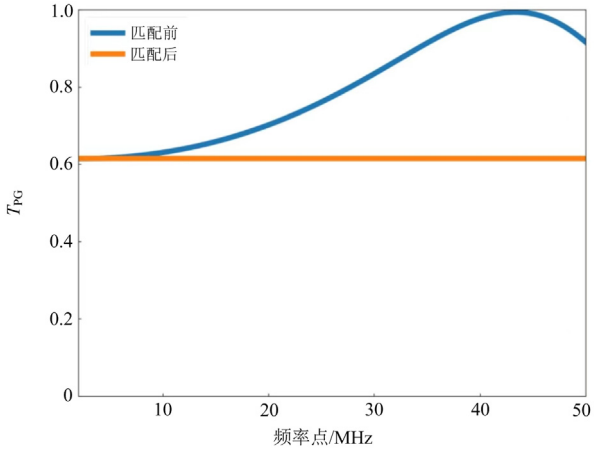

图 15　200 + 50j 宽频各频率点下 T_{PG} 值

Fig. 15 T_{PG} values at each frequency point of 200 + 50j broadband

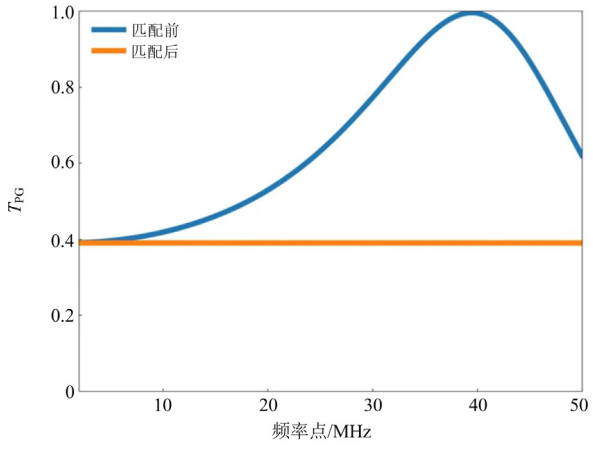

图 16　400 + 50j 宽频各频率点下 T_{PG} 值

Fig. 16 T_{PG} values at each frequency point of 400 + 50j broadband

由图 14—图 16 可知，不同负载在最优匹配点下，各频率点 T_{PG} 的变化趋势。通过分别比较加入匹配网络和无匹配网络的 T_{PG} 曲线与横轴之间的面积，估算 DQN 算法的匹配效果，如表 3 所示。

表 3　DQN 匹配前后效果对比

Table 3 Comparison of DQN before and after matching effects

负载值	无匹配网络/s	加入匹配网络/s	效率提高/%
25 + 30j	36.781 60	43.444 17	13.880
200 + 50j	29.538 46	38.007 84	17.645
400 + 50j	18.731 70	31.771 30	27.166

通过对比可得，DQN 计算最优结果的效率比匹配前提高约 13%以上。

2) 统计数据

重复 DQN 宽频匹配代码 50 次，记录最优值结果，由图 17 可知，负载为$150+50\mathrm{j}$时，通过遍历法计算所得加入匹配网络后 f_{itness} 最大值为 0.849，因为搜索空间内不存在坐标点，可使 f_{itness} 值达到 0.9 及以上，所以在实验结果中(0.9, 1.0)之间，数据样本数为 0。DQN 算法计算最优值在(0.8, 0.85)之间的数据样本为 45；当负载为$300+10\mathrm{j}$时，f_{itness} 最大值在(0.7, 0.75)之间，数据样本数为 35；当负载为$450+20\mathrm{j}$时，f_{itness} 最大值在(0.6, 0.7)之间，数据样本数为 50。

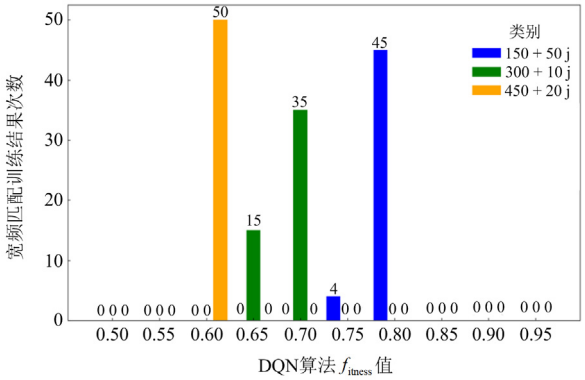

图 17　宽频统计数据概率分布图

Fig. 17 Probability distribution of broadband statistics

3) 方法对比

将 DQN、遍历法与粒子群算法进行对比。遍历法，首先通过对搜索空间进行量化处理，将搜索空间 x、y、z 轴以单位 1 进行进行量化，将搜索空间切割 50×50×60 的坐标点，通过对空间每个点都进行匹配公式计算结果，对比所有结果选取最优位置解。粒子群算法在宽带匹配下设置：加速度常数分别为 0.9 和 0.4，惯性权重为 0.6，粒子群数为 20，迭代次数为 100。

图 18 记录了加入匹配网络前、加入匹配网络后

利用遍历法、粒子群算法、DQN 算法后，f_{itness} 值的变化趋势。由图 18 可知，DQN 寻优所得到的最优值与遍历法、粒子群算法相比，准确度较高，误差较小且迭代次数较少。DQN 利用奖励函数寻最优值，可以减少搜索步数和搜索时间。

对比 3 种方法的运行时间，DQN 算法运行时间指 DQN 经过训练后的最优模型运行时间。由表 4 可看出，DQN 的运行时间小于遍历法和粒子群算法，运行时间最短。

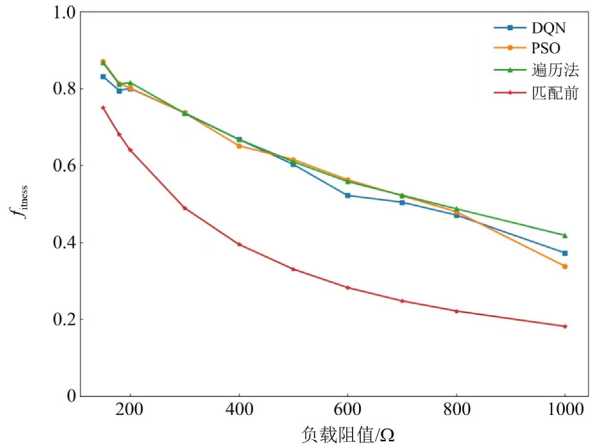

图 18 算法寻优值对比

Fig. 18 Comparison of algorithm search values

表 4 DQN 宽带匹配方法运行时间对比

Table 4 Comparison of DQN broadband matching methods

算法	遍历法	粒子群算法	DQN
程序运行时间/min	46.9833	2.2167	0.8

4 结论

本文提出一种基于深度强化学习的Π型阻抗匹配网络多参数最优求解方法，并分别建模仿真了窄带匹配和宽带匹配两种应用场景，优化了模型参数，有效缩短了算法的搜索时间。经实验测试证明，深度强化学习与遍历法和粒子群方法相比，本文所提方法准确度较高且运行时间较短，算法的应用场景更多，能够较好地自动匹配电力线载波通信负载阻抗变化，改善和提高电力线载波通信质量。电力载波通信的时变性对阻抗的实时匹配提出了更高的要求，因此进一步降低匹配运算的时间，保证网络能获得较快的匹配速度，将是下一步研究的方向。

参考文献

[1] 高鸿坚, 谢宏伟, 陆旭, 等. 长距离电力线载波通信数字前端技术[J]. 中国电力, 2023, 56(3): 128-136.
GAO Hongjian, XIE Hongwei, LU Xu, et al. Digital front end technology for long-distance power line communication[J]. Electric Power, 2023, 56(3): 128-136.

[2] 王贤辉, 徐鲲鹏, 李铮, 等. 低压电力线载波通信零线耦合方法研究[J]. 电力科学与技术学报, 2023, 38(4): 222-229.
WANG Xianhui, XU Kunpeng, LI Zheng, et al. Research of neutral line coupling method in low voltage power line carrier communication[J]. Journal of Electric Power Science and Technology, 2023, 38(4): 222-229.

[3] 张培玲, 赵可可. 基于单频通信的低压电力线通信系统设计与实现[J]. 中国电力, 2023, 56(3): 118-127, 136.
ZHANG Peiling, ZHAO Keke. Design and implementation of low voltage power communication system based on single frequency communication[J]. Electric Power, 2023, 56(3): 118-127, 136.

[4] 王炳庭. 直流电力线通信阻抗匹配方法及阻抗匹配耦合器研究[D]. 南京: 南京邮电大学, 2020.
WANG Bingting. Research on DC power-line communication impedance matching method and impedance matching coupler[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2020.

[5] LE Jian, WANG Cao, ZHOU Wu, et al. A novel PLC channel modeling method and channel characteristic analysis of a smart distribution grid[J]. Protection and Control of Modern Power Systems, 2017, 2(2): 146-158.

[6] 张慧. 低压宽带电力线信道和噪声建模研究[D]. 北京: 华北电力大学, 2019.
ZHANG Hui. Research on modeling low voltage broadband power line channels and noise[D]. Beijing: North China Electric Power University, 2019.

[7] 魏绍亮, 张帅, 程奉玉, 等. 低压电力线通信带宽自适应分配策略研究[J]. 电力系统保护与控制, 2021, 49(21): 123-131.
WEI Shaoliang, ZHANG Shuai, CHENG Fengyu, et al. Adaptive bandwidth allocation strategy for low-voltage power line communication[J]. Power System Protection and Control, 2021, 49(21): 123-131.

[8] 翟明岳, 徐志强, 王九金. 宽带电力线通信系统中的资源分配综述[J]. 电网技术, 2010, 34(5): 173-179.
ZHAI Mingyue, XU Zhiqiang, WANG Jiujin. A survey on resource allocation in broadband power line communication system[J]. Power System Technology,

2010, 34(5): 173-179.

[9] 陈浩. 配电网电力线载波通信网络匹配优化与组网方法[D]. 北京: 华北电力大学, 2021.
CHEN Hao. Matching optimization and networking method for power line carrier communication network in distribution network[D]. Beijing: North China Electric Power University, 2021.

[10] GAYATHRI, SMITHA D, RANI, et al. Adaptive impedance matching system for broadband power line communication using RC-filters[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 7: 11823-11832.

[11] CHOI W H, PARK C Y. A simple line coupler with adaptive impedance matching for Power line Communication[C] // 2007 IEEE International Symposium on Power Line Communications and its Applications (ISPLC 2007), March 26-28, 2007, Pisa, Italy: 186-190.

[12] 郭以贺, 杜思思. 一种中压电力线通信阻抗匹配电路设计[J]. 电力系统保护与控制, 2017, 45(11): 102-107.
GUO Yihe, DU Sisi. Design of impedance matching circuit for MV power line communication[J]. Power System Protection and Control, 2017, 45(11): 102-107.

[13] 贾男. 电力线载波通信自适应阻抗匹配方法研究[D]. 北京: 华北电力大学, 2018.
JIA Nan. Research on adaptive impedance matching method for power line carrier communication[D]. Beijing: North China Electric Power University, 2018.

[14] 吴迪. 基于 ST7538 的窄带电力线通信系统研究[D]. 哈尔滨: 哈尔滨工业大学, 2008.
WU Di. Research on narrowband power line communication system based on ST7538[D]. Harbin: Harbin Institute of Technology, 2008.

[15] 范函, 张浩. 一种电力线载波通信自适应阻抗匹配方案[J]. 电力系统保护与控制, 2009, 37(8): 79-82.
FAN Han, ZHANG Hao. An adaptive impedance matching scheme for power line carrier communication[J]. Power System Protection and Control, 2009, 37(8): 79-82.

[16] 胡文婧. 中压电力线载波通信自适应阻抗匹配算法研究[D]. 北京: 华北电力大学, 2019.
HU Wenjing. Research on adaptive impedance matching algorithm for medium voltage power line carrier communication[D]. Beijing: North China Electric Power University, 2019.

[17] 马永杰, 陈敏, 龚影, 等. 动态多目标优化进化算法研究进展[J]. 自动化学报, 2020, 46(11): 2302-2318.
MA Yongjie, CHEN Min, GONG Ying, et al. Research progress of dynamic multi-objective optimization evolutionary algorithm[J]. Journal of Automation, 2020, 46(11): 2302-2318.

[18] ASKARZADEH A. A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm[J]. Computers and Structures, 2016, 169: 1-12.

[19] LIU Zhenze, YUAN Qing, NIE Guangming, et al. A multi-objective model predictive control for vehicle adaptive cruise control system based on a new safe distance model[J]. International Journal of Automotive Technology, 2021, 22(2): 475-487.

[20] RIEDMILLER M. Neural fitted Q iteration-first experiences with a data efficient neural reinforcement learning method[C] // 16th European Conference on Machine Learning (ECML 2005), October 3-7, 2005, Porto, Portugal: 317-328.

[21] ASSEMAN, ALEXIS, ANTOINE, et al. Accelerating deep neuroevolution on distributed FPGAs for reinforcement learning problems[J]. ACM Journal on Emerging Technologies in Computing Systems, 2021, 17(2): 1-17.

[22] 王子晗, 高红均, 高艺文, 等. 基于深度强化学习的城市配电网多级动态重构优化运行方法[J]. 电力系统保护与控制, 2022, 50(24): 60-70.
WANG Zihan, GAO Hongjun, GAO Yiwen, et al. Multi-level dynamic reconfiguration and operation optimization method for an urban distribution network based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(24): 60-70.

[23] 孙立钧, 顾雪平, 刘彤, 等. 一种基于深度强化学习算法的电网有功安全校正方法[J]. 电力系统保护与控制, 2022, 50(10): 114-122.
SUN Lijun, GU Xueping, LIU Tong, et al. A deep reinforcement learning algorithm-based active safety correction method for power grids[J]. Power System Protection and Control, 2022, 50(10): 114-122.

[24] YIN Xiuxing, LEI Meizhen. Jointly improving energy efficiency and smoothing power oscillations of integrated offshore wind and photovoltaic power: a deep reinforcement

learning approach[J]. Protection and Control of Modern Power Systems, 2023, 8(2): 420-430.

[25] 徐博涵, 向月, 潘力, 等. 基于深度强化学习的含高比例可再生能源配电网就地分散式电压管控方法[J]. 电力系统保护与控制, 2022, 50(22): 100-110.
XU Bohan, XIANG Yue, PAN Li, et al. Local decentralized voltage management of a distribution network with a high proportion of renewable energy based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(22): 100-110.

[26] 刘亚辉, 申兴旺, 顾星海, 等. 面向柔性作业车间动态调度的双系统强化学习方法[J]. 上海交通大学学报, 2022, 56(9): 1262-1275.
LIU Yahui, SHEN Xingwang, GU Xinghai, et al. A dual-system reinforcement learning method for flexible job shop dynamic scheduling[J]. Journal of Shanghai Jiaotong University, 2022, 56(9): 1262-1275.

[27] 伊娜, 徐建军, 陈月, 等. 基于深度强化学习的多阶段信息物理协同拓扑攻击方法[J]. 电力工程技术, 2023, 42(4): 149-158.
YI Na, XU Jianjun, CHEN Yue, et al. A multi-stage coordinated cyber-physical topology attack method based on deep reinforcement learning[J]. Electric Power Engineering Technology, 2023, 42(4): 149-158.

[28] 宋伟业, 刘灵玥, 阎洁, 等. 基于深度强化学习的海上风电集群自进化功率平滑控制方法[J]. 中国电力, 2023, 56(3): 36-46.
SONG Weiye, LIU Lingyue, YAN Jie, et al. Self- evolving power smooth control method for offshore wind power cluster based on deep reinforcement learning[J]. Electric Power, 2023, 56(3): 36-46.

[29] 韩保军, 高强, 代飞, 等. 基于协同奖励函数多目标强化学习的智能频率控制策略研究[J]. 电力科学与技术学报, 2023, 38(2): 18-29.
HAN Baojun, GAO Qiang, DAI Fei, et al. Intelligent frequency control strategy based on multi-objective reinforcement learning of cooperative reward function[J]. Journal of Electric Power Science and Technology, 2023, 38(2): 18-29.

收稿日期：2023-07-17；　修回日期：2023-10-25

作者简介：

胡正伟(1978—)，男，博士，副教授，研究方向为电力线通信技术、FPGA 边缘计算及神经网络加速；

夏思懿(1998—)，女，硕士研究生，研究方向为电力线通信技术；E-mail: 13358842962@163.com

王文彬(2000—)，男，硕士研究生，研究方向为电力线通信技术。E-mail: 1464766492@qq.com

（编辑　张　颖）