带光滑约束的大地电磁深度强化学习反演

曾晨瑞，熊杰*，曹振，张倩玮，袁梦姣

长江大学电子信息与电气工程学院，湖北 荆州 434023

第一作者，E-mail：2022710599@yangtzeu.edu.cn

* 通信作者，E-mail：Xiongjie@yangtzeu.edu.cn

2024-06-24 收稿；2024-11-16 修回；2024-11-21 接受

基金项目：国家自然科学基金项目(62273060)

中图分类号：P631.325；TP18　　文章编号：2096-8523(2026)01-0302-12　　doi: 10.19509/j.cnki.dzkq.tb20240349

摘　要：反演是处理大地电磁测深数据的关键步骤之一，得到了学者的广泛研究。其中基于数据驱动的反演方法主要包括带监督反演和半监督反演，对无监督反演的研究较少。DQN(deep Q-network)是一种经典的深度强化学习算法，作为无监督反演方法最近被用于解决一维大地电磁反演问题，该方法具有不需要训练数据集、对初始模型依赖较小、多次反演能够得到反演结果的概率分布等优点，但存在反演结果不集中的问题。提出了带光滑约束的大地电磁强化学习反演方法(smooth DQN，简称 SDQN)，该方法基于强化学习框架，将反演问题看成马尔可夫决策问题，并分别定义环境、奖励、智能体等术语；然后将正则化反演的模型约束项引入到奖励中来，从而引导智能体不断调整预测模型的电阻率参数以得到更符合模型约束的结果。理论模型反演结果表明，相较于 DQN 反演和 Occam 反演方法，在相同反演次数情况下 SDQN 方法反演不同噪声水平的观测数据时结果更稳定。西藏扎西康矿集区的大地电磁实测数据反演结果与 Occam 反演结果基本吻合并与已有的地质解释资料一致。SDQN 方法具有反演结果更集中、对观测数据的抗噪能力更强的优点，是解决大地电磁反演问题的新工具。

关键词：深度强化学习；大地电磁反演；光滑约束；DQN 方法；SDQN 方法

大地电磁测深法(magnetotelluric sounding，简称 MT)通过天然电磁场来研究岩层的电学性质及其分布特征，在油气勘探、地热等领域应用广泛[1-2]。反演问题是其研究的重点问题之一[3]。经典的大地电磁反演如阻尼最小二乘法、Occam 法、非线性共轭梯度法等[4-13]，利用目标函数梯度信息迭代求解，具有反演速度快的优点，但同时存在过于依赖初始模型的问题。随着地球物理反演方法的发展，模拟退火、粒子群优化、差分进化、蚁群优化等全局寻优的反演算法因容易跳出局部最优，克服了初始模型依赖问题，得到了广泛研究和应用[14-19]，但存在计算反演时间较长的问题。

随着人工智能的发展，基于数据驱动的深度学习反演方法近年来被应用于大地电磁反演领域[20-25]。深度学习反演可分为带监督反演、半监督反演、无监督反演三大类。带监督反演是先采用大量标签数据集 (往往是合成数据集) 训练深度神经网络，然后用训练完毕的网络实现反演功能[25-30]；有时也将实测数据作为标签以得到更符合实际的反演解。在实测数据匮乏的情况下，带监督反演方法存在局限，半监督反演可解决此类反演问题[31-32]；无监督反演则是研究在完全没有标签数据的情况下如何进行反演[33-34]，其中 WANG 等[33] 将深度强化学习中的 DQN 算法与一维大地电磁反演相结合，首次提出了无监督学习的地球物理反演方法。该方法将反演问题转变成马尔可夫决策问题，通过对初始模型的每一层参数进行扰动来不断修正模型，在满足终止条件后得到一次反演结果，通过多次反演得到反演解的概率分

著录信息：曾晨瑞，熊杰，曹振，等. 带光滑约束的大地电磁深度强化学习反演 [J]. 地质科技通报，2026，45(1)：302-313.

ZENG Chenrui, XIONG Jie, CAO Zhen, et al. Magnetotelluric inversion using deep reinforcement learning with a smooth constraint[J]. Bulletin of Geological Science and Technology, 2026, 45(1): 302-313.

© Editorial Office of Bulletin of Geological Science and Technology. This is an open access article under the CC BY-NC-ND license.　https://dzkjtb.cug.edu.cn

布，进而依据概率分布得到最终的反演结果。该方法的提出具有开创性意义，然而，反演时间较长，反演结果不够稳定。为此，将光滑约束引入到强化学习反演，以稳定反演求解过程。通过实验表明，带光滑约束的 DQN 强化学习反演算法 (SDQN) 反演过程更加稳定、反演结果更加集中。最后通过实测资料反演进一步检验了本方法的实用性。

1 方法原理

1.1 强化学习

强化学习是一种在有限离散环境中使用马尔可夫决策过程构造的无监督学习算法。比较著名的有 Q-learning 算法[35-36]。其优化思想为：智能体接受当前环境下的状态信息，根据策略选择动作，并得到来自环境的奖励；智能体能够根据奖励来优化自己的策略，进而选取更优的动作。其中，动作的选取依赖于动作价值函数，也叫做 Q 函数，它表示不同状态下做出不同动作时对应的价值，Q 函数的定义为：

$$Q(s_t,a_t)=E[R_t+\gamma Q(s_{t+1},a_{t+1})], \tag{1}$$

式中：$Q(s_t,a_t)$ 为在当前状态 s_t 下动作 a_t 对应的 Q 值；E 为期望；R_t 为执行动作 a_t 时获得的奖励（可视为短期回报）；$Q(s_{t+1},a_{t+1})$ 为下一个状态 s_{t+1} 下动作 a_{t+1} 的 Q 值（可视为长期回报）；γ 为折扣因子，取值范围在 0 ~ 1 之间，控制智能体在长期回报和短期回报之间取得平衡。

因此，Q 函数可理解为在状态 s_t 下执行动作 a_t 后能够得到的所有奖励的期望值。Q 值越大，意味着执行该动作越有可能获得更高的长期回报，智能体就越有可能选择该动作。如果 Q 函数的估计存在偏差或不准确，智能体就不能够做出最优动作，达不到理想的学习结果。Q 函数的更新依赖于环境的奖励，更新公式为：

$$Q'(s_t,a_t)=Q(s_t,a_t)+\eta[R_t+\gamma\max\{Q(s_{t+1},a_{t+1})\}-Q(s_t,a_t)], \tag{2}$$

式中：η 为学习率，控制智能体依赖先验信息的程度；$\max\{Q(s_{t+1},a_{t+1})\}$ 为下一个状态 s_{t+1} 中的最大 Q 值；$Q'(s_t,a_t)$ 为更新后在状态 s_t 下动作 a_t 对应的 Q 值。

智能体通过 Q 函数的不断更新来逐步逼近最优策略 $\pi^*(s_t)$，使得智能体能够在状态 s_t 下选取最优的动作 a，最优策略 $\pi^*(s_t)$ 定义为：

$$\pi^*(s_t)=\arg\max_a\{Q(s_t,a_t)\}, \tag{3}$$

式中：$\arg\max_a\{Q(s_t,a_t)\}$ 为选取 Q 函数中 Q 值最大的动作 a。

当 Q 函数的更新趋于收敛时，智能体在状态 s_t 下选取的 Q 值最大动作 a_t 即为最优动作。

1.2 深度强化学习

强化学习中的 Q 函数可以用 Q 表来表示，但 Q 表难以解决大规模问题，因此 DQN 算法的深度强化学习应运而生[37]。DQN 算法使用深度神经网络替代了 Q-learning 算法中的 Q 表。Q 值近似过程表示为：

$$Q(s,a,\theta)\approx Q(s,a), \tag{4}$$

式中：θ 为深度神经网络的网络参数。

DQN 算法相较于 Q-learning 算法有以下几个改进方面：①DQN 算法拥有一个用于存储经历 $e_t=(s_t,a_t,r_t,s_{t+1},\text{done})$ 的经验池 $D=\{e_1,\ldots,e_n\}$，智能体每次学习时都会从经验池中随机选取一定数量的经历进行学习。②DQN 算法中包含 2 个结构相同的深度神经网络。第一个网络称为预测网络，用于预测当前状态 s_t 下各动作的 Q 值，其网络参数随着 Q 值的更新而实时更新；第二个网络称为约束网络，它用于得到下一个状态 s_{t+1} 下对应的最大 Q 值 $\max\{Q(s_{t+1},a_{t+1},\theta_t')\}$，以进行 Q 值的更新，约束网络的参数在 Q 值每隔固定更新次数后更新一次，更新时直接复制预测网络的网络参数。约束网络的引入是为了防止预测网络因实时更新而不断振荡导致不能收敛。

结合式 (2)，可以定义预测网络的损失函数 (MSE) 为：

$$L_t(\theta_t)=E[(R_t+\gamma\max_{a_{t+1}}\{Q(s_{t+1},a_{t+1},\theta_t')\}-Q(s_t,a_t,\theta_t))^2], \tag{5}$$

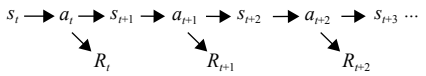
式中：θ_t 为预测网络的网络参数；θ_t' 为约束网络的网络参数。

智能体学习的本质是对预测网络的参数进行更新，从而用网络来近似大规模的 Q 表。

2 带光滑约束的深度强化学习反演

2.1 关键术语的定义

智能体接受环境提供的状态信息 s_t 来做出动作 a_t，并得到来自环境的奖励 R_t。智能体不断与环境交互的过程可以用图 1 所示的马尔可夫决策过程来表示。对于一维大地电磁反演问题，首先要将该问题转化成马尔可夫决策问题。存在一个预测模型，智能体能够不断调整预测模型的模型参数，使预测模型正演响应尽可能逼近真实的观测数据。基于该



s_t, s_{t+1}, s_{t+2}, s_{t+3} 分别为 t, t+1, t+2, t+3 时的状态; a_t, a_{t+1}, a_{t+2}, a_{t+3} 分别为状态 s_t, s_{t+1}, s_{t+2}, s_{t+3} 下的动作; R_t, R_{t+1}, R_{t+2} 分别为执行动作 a_t, a_{t+1}, a_{t+2},时获得的奖励

图 1　马尔可夫决策过程

Fig. 1　Markov decision process

方法, 分别设计状态、动作、奖励如图 1 所示。

2.1.1　状态设计

定义的状态如式(6)所示, 为了让智能体能够找到一个合适的预测模型, 使得预测数据和观测数据尽可能拟合, 定义的状态包括真实观测数据和预测模型正演得到的预测数据, 具体来说包括真实视电阻率(ps^{obs}), 真实视相位(φ^{obs}), 预测模型正演得到的视电阻率(ps^{pre})和视相位(φ^{pre})4 个数据。智能体能够逐层对预测模型的电阻率参数进行扰动, 因此, 状态里面还包括当前正在反演第几层($layer$)的信息。如: 智能体对当前预测模型的第 i 层做出动作后, 状态随之更新, 智能体能够根据更新后的状态继续对第(i+1)层进行扰动。因此, 定义的总的状态($state$)为:

$$state=\left\{ps^{\mathrm{obs}},\varphi^{\mathrm{obs}},ps^{\mathrm{pre}},\varphi^{\mathrm{pre}},layer\right\}。\tag{6}$$

2.1.2　动作设计

定义的动作包括 3 个, 分别是: ①对当前层不施加扰动; ②对当前层的电阻率在对数域上+0.05; ③对当前层的电阻率在对数域上−0.05。设置对数域是为了让智能体能够更快地找到适合的预测模型来满足目标函数的终止条件, 同时让高阻和低阻的分布更为均衡。

2.1.3　奖励设计

奖励可以根据反演问题的目标函数来设计。传统反演方法中的目标函数 ($\varphi(m)$) 可以写为:

$$\varphi(m)=\varphi_{\mathrm{d}}(m)+\lambda\varphi_{\mathrm{m}}(m),\tag{7}$$

式中: $\varphi_{\mathrm{d}}(m)$ 为数据约束项; λ 为正则化因子; $\varphi_{\mathrm{m}}(m)$ 为模型约束项; m 为预测模型电阻率。

设计的数据约束项和模型约束项如式(8)和式(9)所示:

$$\varphi_{\mathrm{d}}(m)=\sum_{i=1}^{N}\frac{\left|\frac{d_i^{\mathrm{obs}}-d_i^{\mathrm{pre}}}{d_i^{\mathrm{obs}}\alpha}\right|}{N},\tag{8}$$

$$\varphi_{\mathrm{m}}(m)=\left\|\nabla m_{\log}\right\|_1,\tag{9}$$

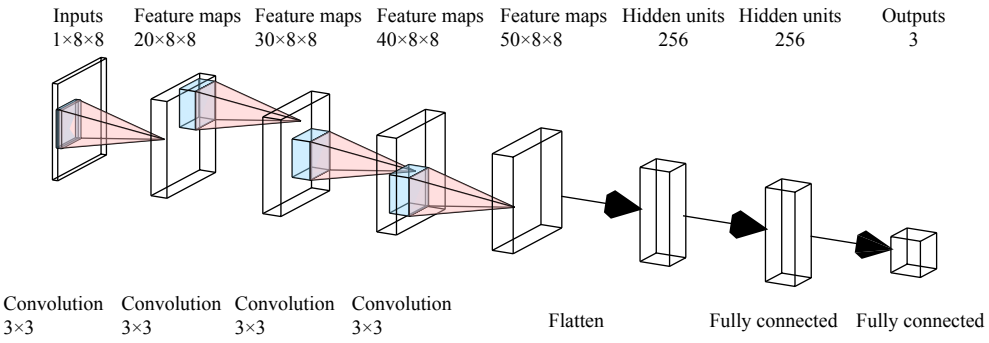
式中: N 为频点数; d_i^{obs} 为观测数据; d_i^{pre} 为预测模型的正演响应; α 为观测数据中噪声的百分比, 设为随机高斯数的标准差; $\|\ \|_1$ 为 L1 范数; ∇ 为求梯度; $m_{\log}$ 为预测模型中对数域上的电阻率。

式 (8) 确保了数据误差权重在所有频点上处于相同水平, 式 (9) 确保了模型约束权重在所有层上处于相同的水平。

在施加光滑约束的情况下使观测数据尽可能拟合, 因此反演过程中的奖励由式 (7) 控制, 如果式 (7) 比上一次的值要更小, 则奖励为正 10, 否则为负 10; 而反演的终止条件由式 (8) 控制, 即 $\varphi_{\mathrm{d}}(m)\leqslant 1$ 时终止, 完成一次反演, 终止时奖励为正 20。

2.2　智能体设计

智能体通过 DQN 来选择动作, SDQN 方法中的 DQN 结构如图 2 所示。其中网络的输入来自环境的状态信息。对于观测数据和响应数据频点数的不同, 可以灵活地设置反演第几层的参数个数。如: 在频点数为 15 的情况下, 观测数据和响应数据合并后共有 4×15=60 个数; 可以用冗余且相同的 4 个数表示当前反演第几层。因此, 整个状态的维数为 60+4=64, 可以转换成 8×8 的二维张量, 这样做的目



Inputs. 输入, 数字为输入形状; Feature maps. 特征图, 数字为特征图的形状; Hidden units. 隐藏单元, 数字为隐藏单元个数; Convolution. 卷积操作, 数字为卷积核; Faltten. 展平操作; Fully connected. 全连接操作; Outputs. 输出

图 2　DQN 结构

Fig. 2　Structure of DQN

的是希望智能体能够像看图片一样处理状态信息。然后这些状态信息分别经过了通道数为 20, 30, 40, 50 的卷积层来提取特征，卷积核大小为 3×3，采用同尺寸填充(same)的方式填充，由于卷积层的参数较少，所以没有加入池化层和随机失活(dropout)。最后将卷积层的输出展平并通过 2 个全连接层，每一层有 256 个神经元。它们进一步提取特征，最后输出 3 个动作对应的 Q 值。整个网络的激活函数为 relu 函数，网络优化器为 Adam。

智能体可以选择最大的 Q 值所对应的动作来作为当前的最优动作。同时，为了让智能体能够在解空间上进行充分探索，智能体还将以一定概率(称为探索因子)随机选择动作，本方法中探索因子设为固定值。智能体选择动作后作用于环境并接受奖励、下一个状态和当前反演是否完成等信息，将这些信息放入经验池中，当预测网络达到更新条件时，便从经验池中随机取一个批量的样本进行学习。当约束网络达到更新条件时，便将预测网络的网络参数复制到约束网络。智能体与环境不断交互来积累经验，在前期主要通过试错来调整预测模型，这使反演结果不稳定。智能体依据经验不断学习如何调整来获得更大的回报(更多的正奖励)，随着多次的反演，这一调整过程趋于确定，带来了更稳定的反演结果。该学习过程没有标签数据来指导，因此为无监督学习方法。算法流程如图 3 所示。同文献 [33] 的 DQN 方法一样，总共反演 300 次来获得一个统计结果，用来比较相同反演次数下的反演情况，相关的超参数设置见表 1。

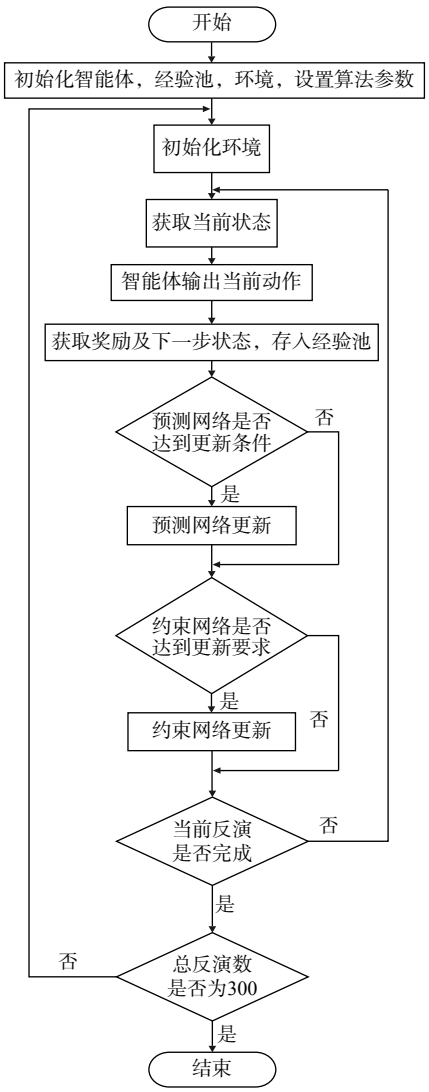


图 3　算法流程图

Fig. 3　Algorithm flowchart

表 1　SDQN 算法超参数设置

Table 1　Hyperparameter settings of SDQN

超参数	数值
小批量大小	32
经验池大小	2 500 000
约束网络更新频率	500
折扣因子	0.95
预测网络更新频率	5
学习率	0.001 3
探索因子	0.1

3　模型实验

3.1　五层模型实验

构造了 5 层理论模型作为实验，同时设计了 10 层的预测模型。理论模型和预测模型对应关系示意如图 4 所示，其中的 5 层理论模型的电阻率和厚度见表 2。对理论模型的正演响应加入 2% 的高斯噪声作为真实观测数据，它为 0.001 ~ 1 000 Hz 内对数域上等间隔采样的 15 个点。为了讨论 SDQN 算法在不同初始模型下对反演结果的影响，10 层预测模型有 3 种不同的初始值，10 层的预测模型的初始参数见表 3。

经过测试发现 SDQN 方法中正则化因子在 0.5 的情况下效果最好。图 5 为 DQN 和 SDQN 2 种方法的训练曲线对比图，从图 5a ~ c 可以看出，2 种方法的网络模型更新次数都随着反演的进行振荡下降，SDQN 方法的网络模型更新次数在反演阶段的中后期振荡幅度更小，表明 SDQN 方法的网络参数更新更平稳。从图 5d ~ f 可以看出，在不同的初始模型下 2 种方法选取动作对应的 Q 均值都随着反演的进行振荡上升并趋于收敛，且 SDQN 方法在整个反演

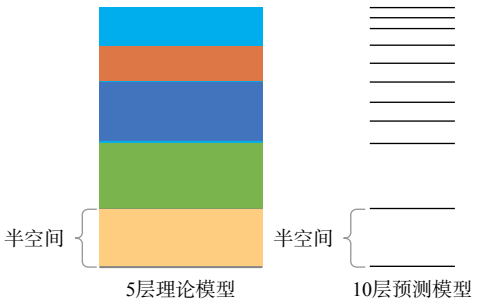


图 4　模型构造对应关系示意图

Fig. 4　Schematic diagram of correspondence in model construction

阶段的 Q 均值略低于 DQN 方法，这是由于在整个反演阶段加入了光滑约束，所以智能体更难获得奖励。从图 5g ~ i 可以看出，在历代反演的结果和理论模型的均方差损失中 SDQN 方法每次反演的均方差损失震荡幅度要更小，表明 SDQN 方法每次得到的反演解更集中，提高了反演结果的稳定性。

2 种方法得到的最终反演结果如图 6 所示。阴影部分是最终解可能的概率分布，它是多次反演解合成起来的结果。将每一层中电阻率出现概率最大的结果作为这一层最终的预测结果。在 DQN 方法

表 2　5 层理论模型电阻率和厚度

Table 2　Resistivity and thickness of five-layer synthetic model

层号	电阻率/(Ω·m)	厚度/m
1	200	1 820
2	20	1 900
3	100	4 535
4	10	2 150
5	100	半空间

表 3　10 层预测模型 3 种初始模型参数设置

Table 3　Three initial model parameter settings of ten-layer predictive model

层号	1	2	3	4	5	6	7	8	9	10
电阻率/(Ω·m)	10	10	10	10	10	10	10	10	10	10
	100	100	100	100	100	100	100	100	100	100
	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000
厚度/m	500	600	720	865	1 035	1 245	1 495	1 795	2 150	半空间

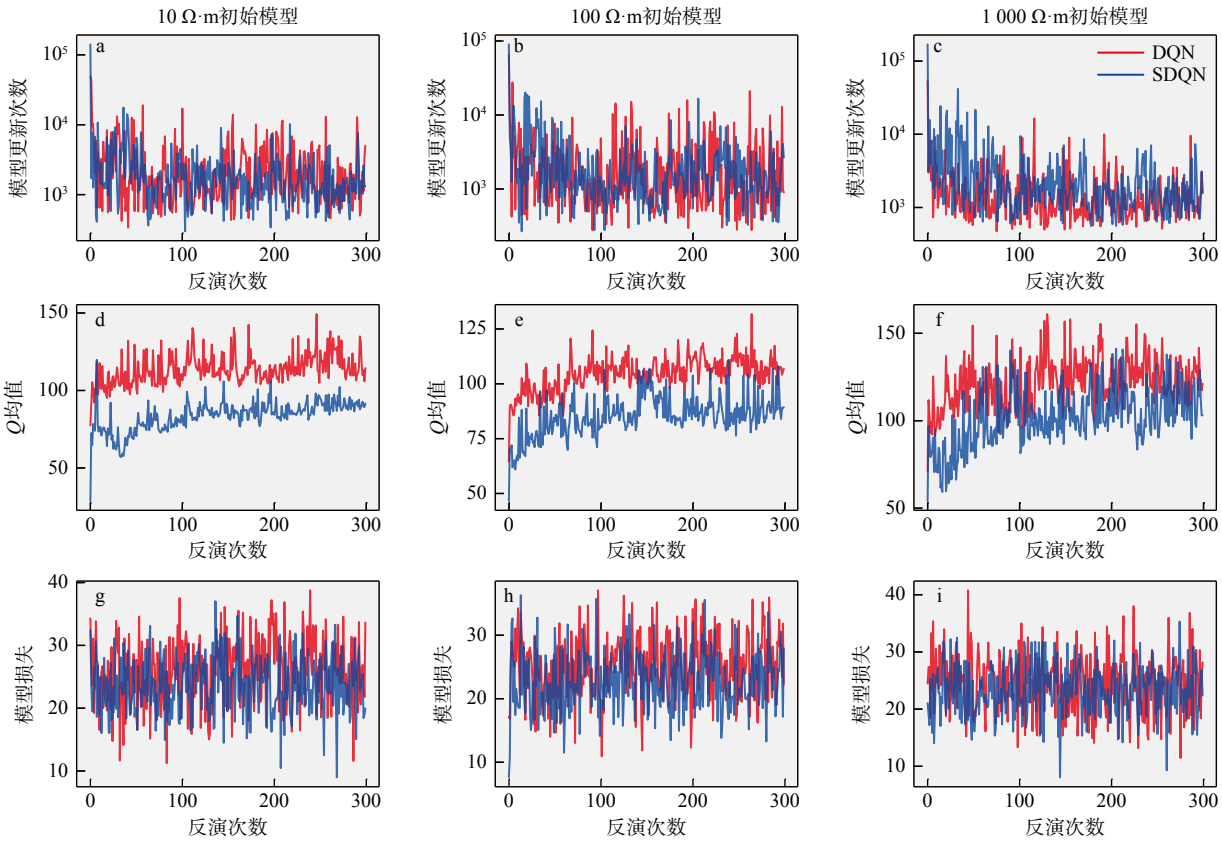


图 5　不同初始模型下 DQN 和 SDQN 方法的迭代曲线

Fig. 5　Training curves of DQN and SDQN methods under different initial models

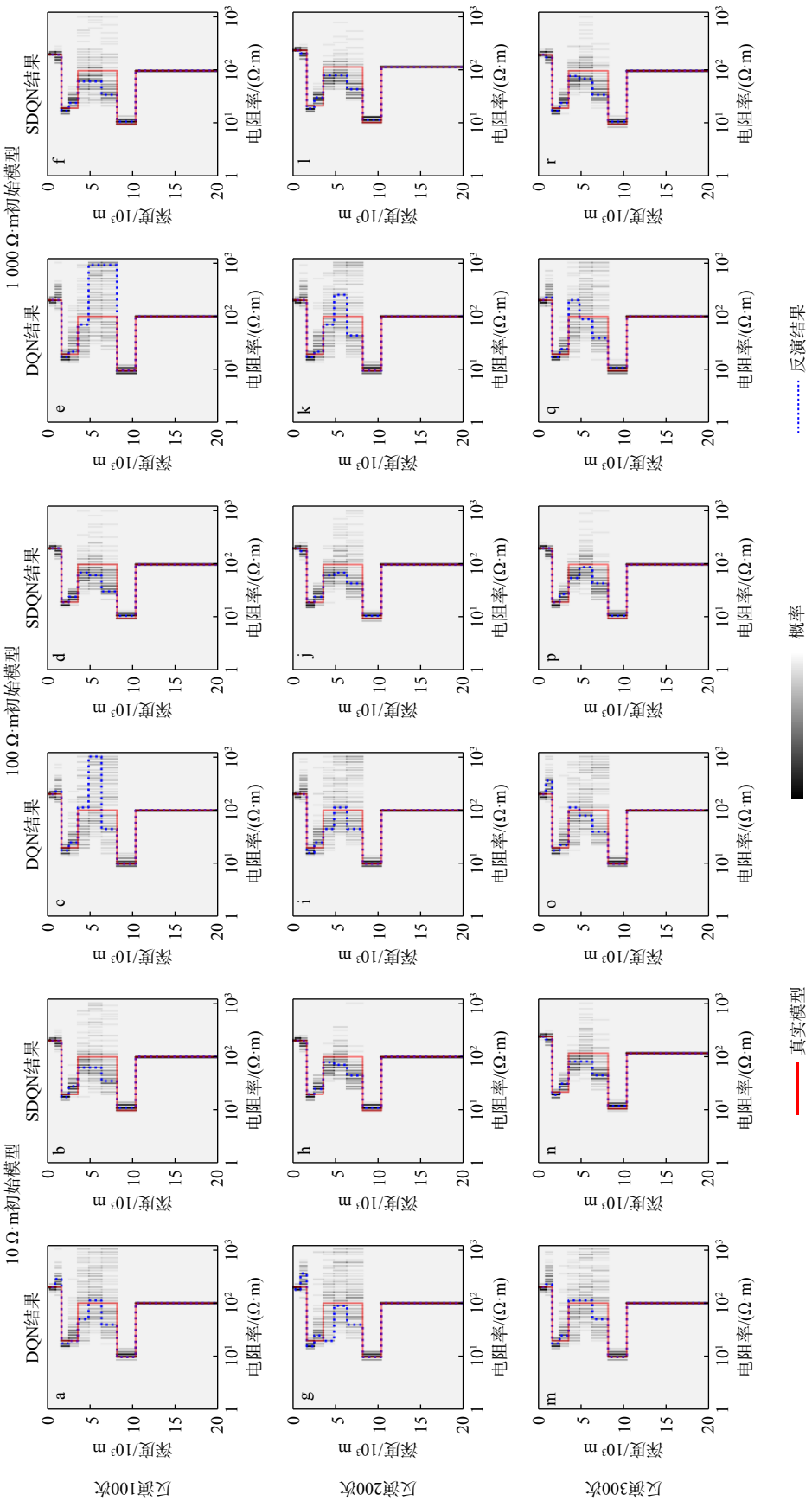


图 6　不同初始模型下 DQN 和 SDQN 方法的反演结果

Fig. 6　Inversion results of DQN and SDQN methods under different initial models

中, 每张图的阴影部分分布较大, 表明反演的结果不确定信息更大。而在 SDQN 方法中, 每张图的阴影分布的范围要更集中, 不确定信息更小。

图 6c, d, i, j, o, p 的反演结果计算 2 种方法在初始模型为 100 Ω·m 下不同反演次数的最终反演结果的均方误差和所有反演结果的均方误差的方差, 如表 4 所示。从表 4 可以看出, DQN 方法在不同反演次数下最终结果的均方误差差异较大, 且反演 300 次后的最终结果没有 SDQN 方法好; SDQN 方法的反演结果均方误差的方差均明显优于 DQN 方法, 表明 SDQN 方法的反演结果概率上更接近理论模型。

3.2　噪声测试

为了讨论 SDQN 方法对噪声的敏感程度, 另设计了 8 层理论模型, 模型的参数如表 5 所示, 观测数据同样取 0.001 ~ 1 000 Hz 内对数域上等间隔采样的 15 个点。构造的 20 层预测模型的初始参数如表 6 所示, 每一层的初始模型的电阻率都为 100 Ω·m, 且层厚按 1.12 的比率增长。对观测数据分别加入 2%, 5%, 10% 的高斯噪声进行测试, 反演中式 (8) 的噪声百分比 (α) 对应设置为 2%, 5%, 10%。同样对每次测试反演 300 次, 并与经典的 Occam 方法[38] 进行对比。Occam 方法的初始模型参数也按表 5 设置, 反演中已知厚度, 仅反演电阻率。2 种方法的终止条件都设为数据拟合误差小于 1, 正则化因子都设为 0.01, 测试得到的反演结果如图 7 所示。

表 4　初始模型为 100 Ω·m 时反演结果的误差和方差

Table 4　Error and variance of inversion results for initial model at 100 Ω·m

反演次数	最终结果的均方误差/($\Omega^2\cdot m^2$)		所有结果的均方误差的方差/($\Omega^4\cdot m^4$)	
	DQN	SDQN	DQN	SDQN
100	99.97	14.23	2 532.45	296.60
200	13.355	15.094	3 019.32	427.82
300	25.332	12.084 5	2 134.51	201.47

图 7 中的 3 种方法的拟合误差均小于 1, 且都能较好地反映出真实模型曲线的基本形态, 但能够看出 Occam 方法对噪声较为敏感, 在噪声为 5% 的情况下中间层对模型曲线的吻合情况没有 SDQN 和 DQN 方法好, 所以 DQN 和 SDQN 方法对噪声的鲁棒性较好。为了进一步对比 SDQN 和 DQN 方法的结果, 绘制了 SDQN 和 DQN 方法在 300 次反演中每一层结果的箱型图如图 8 所示, 每层结果的数据

表 5　8 层理论模型电阻率和厚度

Table 5　Resistivity and thickness of eight-layer synthetic model

层号	1	2	3	4	5	6	7	8
电阻率/(Ω·m)	2 500	1 000	100	10	100	25	10	2.5
厚度/m	600	1 400	2 200	3 400	7 000	9 000	11 000	半空间

表 6　20 层预测模型初始模型参数设置

Table 6　Initial model parameter settings for twenty-layer the predictive model

层号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
电阻率/(Ω·m)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
厚度/m	600	672	752	842	944	1 057	1 184	1 326	1 485	1 663	1 863	2 087	2 337	2 618	2 932	3 284	3 678	4 119	4 613	半空间

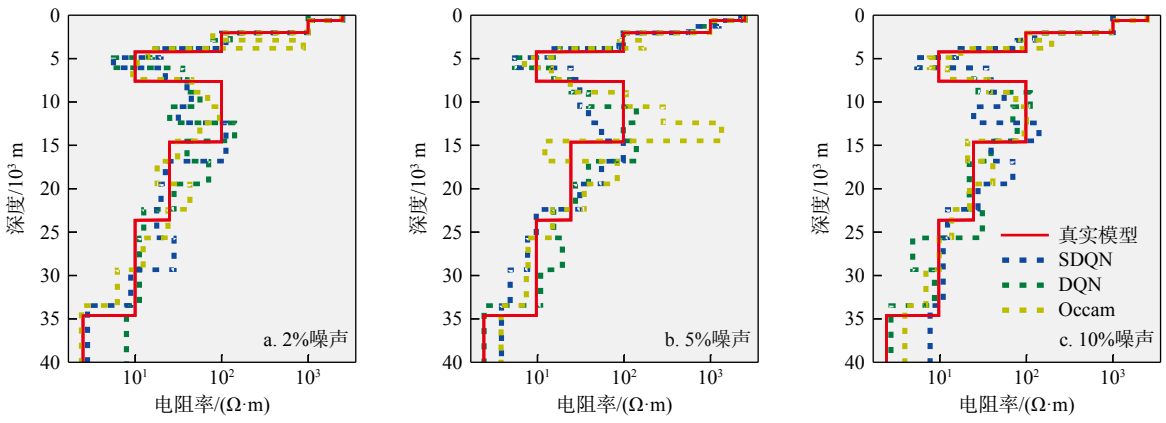


图 7　SDQN、DQN、Occam 在噪声为 2%(a), 5%(b), 10%(c) 下的反演结果

Fig. 7　Inversion results of SDQN, DQN, and Occam at noise levels of 2%(a), 5%(b), and 10%(c)

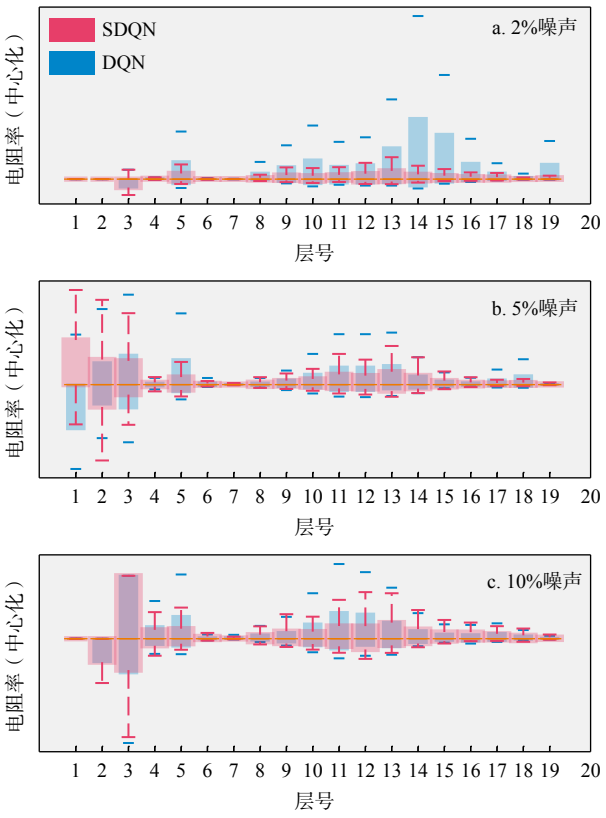


图 8　不同噪声水平下 SDQN、DQN 方法的反演结果箱型图

Fig. 8　Box plots of inversion results of SDQN and DQN methods at different noise levels

做了去中心化处理以方便对比。可以看见在大部分层下 SDQN 方法的箱型图上限和下限小于 DQN 方法，且上下四分位数小于 DQN 方法，表明 SDQN 方法的反演结果更集中。

从数据上看，图 9 为在噪声水平为 5% 的情况下 2 种方法反演结果的正演响应曲线，*MAPE* 为数据的平均绝对百分比误差，2 种方法的响应曲线都能大致拟合观测数据，误差都在可接受的范围内。

从模型上看，图 10 为噪声水平为 5% 时不同反演次数下 DQN 和 SDQN 方法的反演结果，可以看到 SDQN 方法在不同反演次数下的反演结果都较为稳定且都能较好地贴近理论模型，而 DQN 方法的前几次反演并不稳定。对于 SDQN 方法，认为少量的反演次数是可行的，这将极大缩短反演时间，例如只需要反演 50 次，耗时仅为 3 500 s 就能得到比较好的反演结果，DQN 需要反演 300 次，耗时 6 000 s，才能达到相近的反演结果。且在光滑程度上，SDQN 方法最终结果的光滑度为 1 322.82，DQN 方法最终结果的光滑度为 2 225.17，说明 SDQN 方法学到了更为光滑的结果。

相应的反演时间如表 7 所示，SDQN 方法在反演前期由于更难获得奖励，要花费更长的时间才能达到反演目标，但是在反演中后期，SDQN 方法一旦收敛，反演时间会比 DQN 方法的反演时间更短。

表 7　5% 噪声水平下 DQN 和 SDQN 方法的反演时间

Table 7　Inversion times of DQN and SDQN methods at 5% noise level

反演次数	反演时间/s	
	DQN	SDQN
50	3 287.29	3 562.45
100	4 164.09	4 520.86
150	4 918.86	5 144.10
200	5 628.15	5 745.47
250	6 304.54	6 299.31
300	6 944.33	6 841.33

4　实测资料处理

为了进一步验证 SDQN 方法的实用性，对西藏

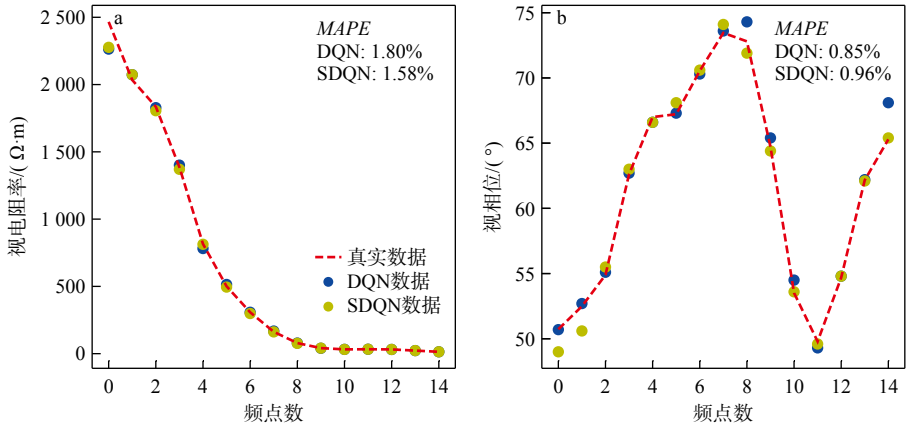


图 9　5% 噪声水平下 SDQN、DQN 方法的数据拟合图（*MAPE* 为平均绝对百分比误差）

Fig. 9　Data fitting curves of SDQN and DQN methods at 5% noise level

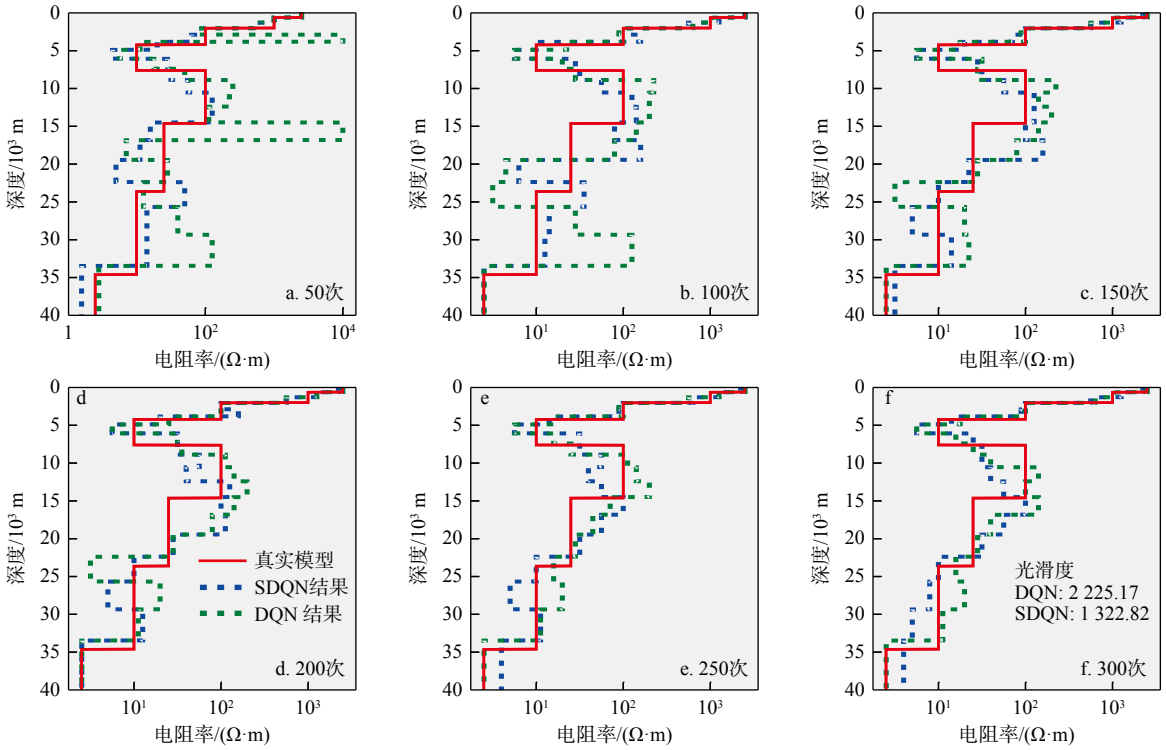


图 10　5% 噪声水平时不同反演次数下 DQN 和 SDQN 方法的反演结果

Fig. 10　Inversion results of DQN and SDQN methods under different inversion times at 5% noise level

扎西康矿集区的部分大地电磁的实测结果数据进行反演对比[39], 该数据集的采样点的频率范围为 0.001 ~ 320 Hz。测点位置如图 11 所示, 测点区出露主要包括长英质火山岩、石英云母片岩、第四系、片麻岩等地层单元[40], 对位于第四系连续的 3 个测点进行反演, 测点数据经过了平滑处理, 同时设置了噪声百分比 α 为 0.05, Occam 法和 SDQN 方法的正则化因子为 0.01, 按照表 6 的方式设置了 20 层的预测模型, 预测模型的初始电阻率都为 100 Ω·m, 第一层的厚度为 800 m, 后面的每层厚度按照 1.13 的比率增加。结合文献 [40], 反演结果及其解释如图 12 所示。

反演结果和 Occam 反演结果在大部分层上基本相似, 3 个测点下的地质模型总体呈现出 3 层结构, 呈 EW 向伸展, 与实际资料能够很好的吻合, 说明了 SDQN 方法的有效性。

5　结论及讨论

(1)不同的初始模型反演结果表明, SDQN 方法相比于 DQN 方法, 同样不依赖于初始模型, 且能够更快地收敛, 反演的结果更稳定。

(2)不同的抗噪实验表明, SDQN 方法相比于 DQN 方法, 鲁棒性同样强, 且反演的时间更快, 结果更集中。

(3)实测资料反演结果表明 SDQN 方法能够较好的反演出地下介质的电性关系, 证实了本方法的实用性。

(4)相对于传统的反演方法, SDQN 方法仍然比较耗时, 更快的强化学习反演方法和二维大地电磁强化学习反演将是下一步的研究方向。

所有作者声明不存在利益冲突。

The authors declare that no competing interests exist.

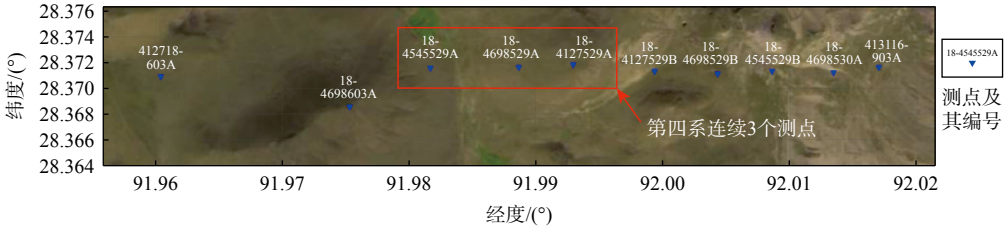


图 11　测点分布图

Fig. 11　Site distribution map

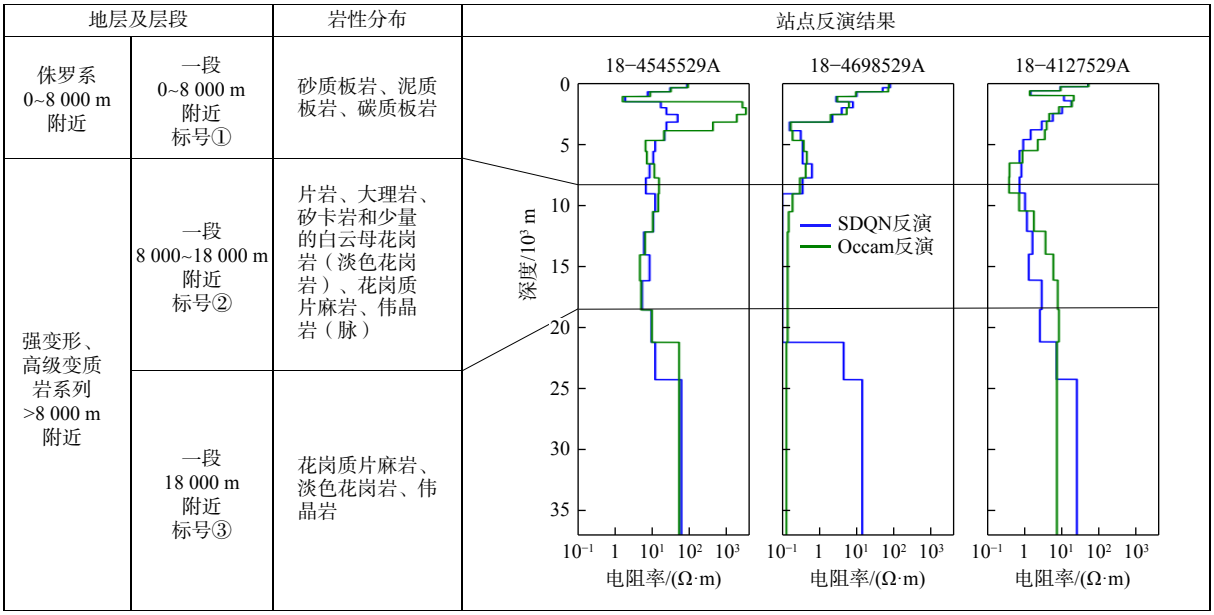


图 12　实测反演结果

Fig. 12　Inversion results of measured data

参考文献：

[1] 秦策, 刘幸飞, 王绪本, 等. 基于自适应有限元正演的大地电磁法三维反演算法研究 [J]. 地球物理学报, 2022, 65(6): 2311-2325.
QIN C, LIU X F, WANG X B, et al. Three-dimensional inversion of magnetotelluric based on adaptive finite element method[J]. Chinese Journal of Geophysics, 2022, 65(6): 2311-2325. (in Chinese with English abstract)

[2] 董珮瑶, 杜利, 赵磊, 等. 机器学习模型在地热开发水温预测中的应用 [J]. 地质科技通报, 2025, 44(3): 388-398.
DONG P Y, DU L, ZHAO L, et al. Application of machine learning models for groundwater temperature prediction in geothermal development[J]. Bulletin of Geological Science and Technology, 2025, 44(3): 388-398. (in Chinese with English abstract)

[3] 喻国, 肖骑彬, 李满. 大地电磁二维各向异性反演及其在青藏高原北部的应用 [J]. 地球物理学报, 2021, 64(6): 2108-2126.
YU G, XIAO Q B, LI M. Two-dimensional magnetotelluric inversion with anisotropy and its application in the northern Tibetan Plateau[J]. Chinese Journal of Geophysics, 2021, 64(6): 2108-2126. (in Chinese with English abstract)

[4] 唐荣江, 王绪本, 甘露. 一种利用特征值性质的 MT 阻尼最小二乘反演 [J]. 石油物探, 2017, 56(6): 898-904.
TANG R J, WANG X B, GAN L. A damped least square inversion for MT utilizing eigenvalue property[J]. Geophysical Prospecting for Petroleum, 2017, 56(6): 898-904. (in Chinese with English abstract)

[5] 吴小平, 徐果明. 大地电磁数据的 Occam 反演改进 [J]. 地球物理学报, 1998, 41(4): 547-554.
WU X P, XU G M. Improvement of Occam's inversion for MT data[J]. Acta Geophysica Sinica, 1998, 41(4): 547-554. (in Chinese with English abstract)

[6] 姜奋勇, 叶益信, 陈海文, 等. 基于非结构网格的带地形 MT 二维 Occam 反演及应用 [J]. 物探与化探, 2022, 46(2): 482-489.
JIANG F Y, YE Y X, CHEN H W, et al. Application of 2D inversion of magnetotelluric data bearing terrain information based on an unstructured mesh[J]. Geophysical and Geochemical Exploration, 2022, 46(2): 482-489. (in Chinese with English abstract)

[7] 徐凤姣, 谢兴兵, 郭全仕. 不同极化模式的二维大地电磁非线性共轭梯度反演及应用研究 [J]. 石油物探, 2022, 61(1): 174-182.
XU F J, XIE X B, GUO Q S. Study and application on the nonlinear conjugate gradient inversion method for 2D magnetotelluric with different polarization modes[J]. Geophysical Prospecting for Petroleum, 2022, 61(1): 174-182. (in Chinese with English abstract)

[8] 管贻亮, 李予国, 胡祖志, 等. 大地电磁非线性共轭梯度一维反演 [J]. 石油物探, 2014, 53(6): 752-759.
GUAN Y L, LI Y G, HU Z Z, et al. Nonlinear conjugate gradients algorithm for 1D magnetotelluric inversion[J]. Geophysical Prospecting for Petroleum, 2014, 53(6): 752-759. (in Chinese with English abstract)

[9] 彭勃, 强思远, 施小清. 基于贝叶斯实验设计优化跨孔高密度电阻率法监测四维水文地质过程 [J]. 地质科技通报, 2025, 44(5): 293-301.
PENG B, QIANG S Y, SHI X Q. Optimizing 4D hydrogeological process monitoring using cross-hole electrical resistivity tomography(CHERT) *via* Bayesian experimental design[J]. Bulletin of Geological Science and Technology, 2025, 44(5): 293-301. (in Chinese with English abstract)

[10] SEILLÉ H, VISSER G. Bayesian inversion of magnetotelluric data considering dimensionality discrepancies[J]. Geophysical Journal International, 2020, 223(3): 1565-1583.

[11] LIU X, ZHENG F W. Axis anisotropic Occam's 3D inversion of tensor CSAMT in data space[J]. Applied Geophysics, 2025, 22(2): 252-263.

[12] NÁDASI E, GRIBENKO A V, ZHDANOV M S. Large-scale inversion of magnetotelluric data using regularized Gauss-Newton method in the data space[J]. Pure and Applied Geophysics, 2022, 179(10): 3785-3806.

[13] 梁生贤, 吾守艾力·肉孜, 廖国忠, 等. 大地电磁线性反演算法比较 [J]. 地球物理学进展, 2014, 29(6): 2702-2707.
LIANG S X, WUSHOUAILI R Z, LIAO G Z, et al. Comparison and aanalysis of two-dimensional linear algorithm inversion for magnetotelluric[J]. Progress in Geophysics, 2014, 29(6): 2702-2707. (in Chinese with English abstract)

[14] 周俊杰, 胡英才, 刘祜, 等. 考虑先验信息的大地电磁模拟退火反演 [J]. 地质论评, 2023, 69(增刊 1): 465-467.
ZHOU J J, HU Y C, LIU H, et al. Inversion of magnetotelluric simulated annealing considering prior information[J]. Geological Review, 2023, 69(S1): 465-467. (in Chinese with English abstract)

[15] 潘新朋, 刘志顺, 高大维, 等. 岩石物理驱动的储层物性参数非线性地震反演方法 [J]. 地球物理学报, 2024, 67(3): 1237-1254.
PAN X P, LIU Z S, GAO D W, et al. Rock-physics-driven nonlinear seismic inversion for petrophysical parameters of reservoir[J]. Chinese Journal of Geophysics, 2024, 67(3): 1237-1254. (in Chinese with English abstract)

[16] 熊杰, 孟小红, 刘彩云, 等. 基于差分进化的大地电磁反演 [J]. 物探与化探, 2012, 36(3): 448-451.
XIONG J, MENG X H, LIU C Y, et al. Magnetotelluric inversion based on differential evolution[J]. Geophysical and Geochemical Exploration, 2012, 36(3): 448-451. (in Chinese with English abstract)

[17] 丁鑫, 郭乙霏, 方宏远, 等. 基于蚁群算法的层状结构介电特性反演方法 [J]. 人民黄河, 2019, 41(6): 107-110.
DING X, GUO Y F, FANG H Y, et al. The inversion method of dielectric properties in layered structure based on ant colony algorithm[J]. Yellow River, 2019, 41(6): 107-110. (in Chinese with English abstract)

[18] 李丽丽, 李长伟, 程勃, 等. 基于改进粒子群算法的大地电磁反演 [J]. 科学技术与工程, 2023, 23(26): 11098-11107.
LI L L, LI C W, CHENG B, et al. Magnetotelluric inversion based on improved particle swarm optimization algorithm[J]. Science Technology and Engineering, 2023, 23(26): 11098-11107. (in Chinese with English abstract)

[19] LI R H, GAO L, YU N, et al. Memetic strategy of particle swarm optimization for one-dimensional magnetotelluric inversions[J]. Mathematics, 2021, 9(5): 519.

[20] 刘高村, 王绪本, 袁崇鑫, 等. Transformer 网络在大地电磁反演成像中的应用 [J]. 物探化探计算技术, 2023, 45(4): 484-496.
LIU G C, WANG X B, YUAN C X, et al. Application of transformer network in magnetotelluric inversion imaging[J]. Computing Techniques for Geophysical and Geochemical Exploration, 2023, 45(4): 484-496. (in Chinese with English abstract)

[21] XU K J, LIANG S Y, LU Y, et al. Magnetotelluric data inversion based on deep learning with the self-attention mechanism[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5917910.

[22] GAO C, LI Y B, WANG X Q. AUTL: An attention U-Net transfer learning inversion framework for magnetotelluric data[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 3004505.

[23] JEVINANI M R, DEHKORDI B H, FERGUSON I J, et al. Correction to: Deep learning-based 1-D magnetotelluric inversion: Performance comparison of architectures[J]. Earth Science Informatics, 2024, 17(2): 1679-1680.

[24] LIAO X L, ZHANG Z H, YAN Q X, et al. Inversion of 1-D magnetotelluric data using CNN-LSTM hybrid network[J]. Arabian Journal of Geosciences, 2022, 15(17): 1430.

[25] CONWAY D, HEINSON G, REES N, et al. Time-lapse inversion of one-dimensional magnetotelluric data[J]. Earth, Planets and Space, 2018, 70(1): 27.

[26] 李思平, 刘彩云, 熊杰, 等. 基于改进残差网络的大地电磁反演研究 [J]. 物探与化探, 2023, 47(6): 1508-1518.
LI S P, LIU C Y, XIONG J, et al. Magnetotelluric inversion based on an improved residual network[J]. Geophysical and Geochemical Exploration, 2023, 47(6): 1508-1518. (in Chinese with English abstract)

[27] 王方, 熊杰, 田慧潇, 等. 基于深度学习的大地电磁二维反演方法 [J]. 地质科技通报, 2024, 43(2): 344-354.
WANG F, XIONG J, TIAN H X, et al. Two-dimensional magnetotelluric inversion method based on deep learning[J]. Bulletin of Geological Science and Technology, 2024, 43(2): 344-354. (in Chinese with English abstract)

[28] 廖晓龙, 张志厚, 姚禹, 等. 基于卷积神经网络的大地电磁反演 [J]. 中南大学学报(自然科学版), 2020, 51(9): 2546-2557.
LIAO X L, ZHANG Z H, YAO Y, et al. Magnetotelluric inversion based on convolutional neural network[J]. Journal of Central South University (Science and Technology), 2020, 51(9): 2546-2557. (in Chinese with English abstract)

[29] 韩盈, 安志国, 底青云, 等. 基于循环神经网络的大地电磁信号噪声压制研究 [J]. 地球物理学报, 2023, 66(10): 4317-4331.
HAN Y, AN Z G, DI Q Y, et al. Research on noise suppression of magnetotelluric signal based on recurrent neural network[J]. Chinese Journal of Geophysics, 2023, 66(10): 4317-4331. (in Chinese with English abstract)

[30] LI G, GU X J, CHEN C J, et al. Low-frequency magnetotelluric data denoising using improved denoising convolutional neural network and gated recurrent unit[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5909216.

[31] 王永昌, 刘彩云, 熊杰, 等. 基于生成对抗网络的半监督地震波阻抗反演 [J]. 现代地质, 2024, 38(6): 1585-1593.
WANG Y C, LIU C Y, XIONG J, et al. Semi-supervised seismic wave impedance inversion based on generative adversarial networks[J]. Geoscience, 2024, 38(6): 1585-1593. (in Chinese with English abstract)

[32] YANG B, XU K J, LIU Z. Fuzzy constrained inversion of magnetotelluric data using guided fuzzy C-means clustering[J]. Surveys in Geophysics, 2021, 42(2): 399-425.

[33] WANG H, LIU Y H, YIN C C, et al. Stochastic inversion of magnetotelluric data using deep reinforcement learning[J]. Geophysics, 2022, 87(1): 49-61.

[34] ZHU X Y, DONG H F. Shear wave velocity estimation based on deep-Q network[J]. Applied Sciences, 2022, 12(17): 8919.

[35] DELL'AVERSANA P. Reinforcement learning in optimization problems. Applications to geophysical data inversion[J]. AIMS Geosciences, 2022, 8(3): 488-502.

[36] JANG B, KIM M, HARERIMANA G, et al. Q-learning algorithms: A comprehensive classification and applications[J]. IEEE Access, 2019, 7: 133653-133667.

[37] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement

learning with double Q-learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1): 2094-2100.

[38] KAMINSKY A. ZondMT1D[CP/OL]. (2024-01-26) [2025-11-20]. http://zond-geo.com/english/zond-software/electromagnetic-sounding/zondmt1d/.

[39] 梁生贤. 扎西康矿集区大地电磁测深数据(2018-2022)[DB/OL]. (2022-12-31)[2025-11-20]. https://data.tpdc.ac.cn/zh-hans/data/b55ce907-6829-46de-9548-c58f539e69c4/.
LIANG S X. Magnetotelluric data of Zhaxikang ore district (2018-2022)[DB/OL].(2022-12-31)[2025-11-20].https://data.tpdc.ac.cn/zh-hans/data/b55ce907-6829-46de-9548-c58f539e69c4/. (in Chinese with English abstract)

[40] 焦彦杰, 黄旭日, 李光明, 等. 藏南扎西康矿集区深部结构与成矿: 来自地球物理的证据 [J]. 地球科学, 2019, 44(6): 2117-2128.
JIAO Y J, HUANG X R, LI G M, et al. Deep structure and mineralization of Zhaxikang ore-concentration area, South Tibet: Evidence from geophysics[J]. Earth Science, 2019, 44(6): 2117-2128. (in Chinese with English abstract)

Summary for “带光滑约束的大地电磁深度强化学习反演”

Magnetotelluric inversion using deep reinforcement learning with a smooth constraint

ZENG Chenrui，XIONG Jie*，CAO Zhen，ZHANG Qianwei，YUAN Mengjiao

School of Electronic Information and Electrical Engineering, Yangtze University, Jingzhou Hubei 434023, China

First author，E-mail：2022710599@yangtzeu.edu.cn

* Corresponding author， E-mail：Xiongjie@yangtzeu.edu.cn

Abstract：Inversion is one of the key steps in processing magnetotelluric sounding data and has been widely studied by scholars. The data-driven inversion approaches mainly include supervised inversion and semi-supervised inversion, but there is limited research on unsupervised inversion. [**Objective**] Deep Q-network (DQN), a classical deep reinforcement learning algorithm, has recently been applied to one-dimensional magnetotelluric inversion problems as an unsupervised inversion approach. It has the advantages of not requiring a training dataset, being less dependent on the initial model, and being able to obtain the probability distribution of inversion results through multiple inversions. However, it suffers from the problem that the inversion results are not concentrated. [**Method**] This paper proposed a magnetotelluric inversion method based on deep reinforcement learning with a smooth constraint (SDQN). This method was based on the framework of reinforcement learning, treated the inversion problem as a Markov decision problem, and defined terms such as environment, reward, agent. Then, the model constraint term of regularized inversion was introduced into the reward function, guiding the agent to continuously adjust the resistivity parameters of the prediction model to obtain results that were more consistent with the model constraints. [**Results**] The inversion results of the synthetic model showed that, compared with the DQN inversion and Occam inversion methods, the SDQN method produced more stable results when inverting observed data at different noise levels under the same number of iterations. The inversion results of the magnetotelluric measured data in the Tashikang mine area of Xizang were largely consistent with the Occam inversion results and aligned with the existing geological interpretations. [**Conclusion**] The SDQN method has the advantages of more concentrated inversion results and stronger noise resistance to observed data, making it a new tool for solving the problem of magnetotelluric inversion.

Key words: deep reinforcement learning；magnetotelluric inversion；smooth constraint；DQN method；SDQN method

Received：2024-06-24；Revised：2024-11-16；Accepted：2024-11-21

doi: 10.19509/j.cnki.dzkq.tb20240349