

基于机器学习的地热采灌方案优化方法

王佳铖¹, 陈进帆¹, 赵志宏^{1*}, 谭现锋²

1. 清华大学土木工程系, 北京 100084

2. 山东省鲁南地质工程勘察院, 济宁 272100

摘要 为了解决基于数值模拟的优化方法, 通常需要大量模拟计算的问题, 在比较不同机器学习方法的预测性能后, 建立了基于多层感知机的代理模型, 以降低计算成本, 然后将其与遗传算法相结合, 提出了非均质地热储层中地热对井系统采灌方案优化方法。通过地热田对井系统的案例研究, 证明了所开发的采灌方案优化方法的合理性和有效性。结果表明, 基于代理模型的采灌方案优化方法能以更低的计算成本, 准确地找到给定开采井位置时最优的回灌井位置、采灌量和尾水温度。

关键词 地热对井系统; 机器学习; 采灌方案优化; 代理模型; 遗传算法

“碳达峰、碳中和”是推动中国经济高质量发展和生态环境高水平保护的内在要求, 而建设清洁低碳安全高效的能源体系是如期实现“双碳”目标的重要保障。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》指出要因地制宜开发利用地热能。中国中深层(200~3000 m)和深层(>3000 m)地热资源储量丰富^[1-3], 地热开采有单井、对井、群井3种模式, 其中以对井系统为主, 即通过开采井将热水采出, 经地面换热后再通过回灌井将尾水回注至热储层, 通过对流换热重新恢复高温。对井系统可有效维持热储压力,

但不合理的井间距、采灌量和尾水温度存在热突破的风险。为了最大限度地提升地热井产能, 通过抽水、回灌、示踪等现场试验确定热储的连通性及回灌井与开采井之间的水力联系, 在此基础上, 采用数学模型对不同开采和回灌情景下热储压力和温度的变化进行预测, 并优化设计采灌方案, 已成为地热资源高效开发利用的重要环节^[4-5]。

随着计算机技术的发展, 能够精细刻画热储层非均质性及其地应力、压力、温度动态演化规律的数值模拟方法已成为热储工程多场耦合效应研究与评价的主要方法^[6-7]。将参数敏感性分析与数值

收稿日期: 2022-08-31; 修回日期: 2022-09-30

基金项目: 国家重点研发计划项目(2019YFB1504103)

作者简介: 王佳铖, 博士研究生, 研究方向为地下工程, 电子信箱: wang-jc21@mails.tsinghua.edu.cn; 赵志宏(通信作者), 副教授, 研究方向为岩石力学与地下工程, 电子信箱: zhzhao@mail.tsinghua.edu.cn

引用格式: 王佳铖, 陈进帆, 赵志宏, 等. 基于机器学习的地热采灌方案优化方法[J]. 科技导报, 2022, 40(20): 93-100; doi: 10.3981/j.issn.1000-7857.2022.20.011

模拟相结合,可获得井距、采灌量、尾水温度等开采参数对地热井产能的影响规律,并识别出影响地热储层性能的关键参数,但无法精确确定开采参数的最优值^[8-10]。目前,地热开采参数优化设计方法主要有2类:一是基于数值模拟的优化方法,二是基于代理模型的优化方法。基于数值模拟的优化方法将模拟采灌条件下热储动态响应的数值方法与随机优化算法相结合,得到满足最大化/最小化物理或成本目标函数的最优解,该方法因需要遍历大量开采参数的取值以评估考虑每种可能的采灌方案对应的热储动态响应过程,其计算效率较低^[11-12]。为减少基于模拟的优化方法所需的计算负担,采用多元自适应回归样条、多元回归、叠加原理、快速推进等方法建立输入变量(如井位或采灌量)与输出响应(如开采井温度或压力)之间的代理模型,其具有足够的精度近似数值模拟的计算结果^[13-14]。再将代理模型与随机优化算法相结合,即可较为快捷地得到开采参数的最优解。但是,基于代理模型的优化方法的难点是如何构建合理准确的代理模型。

随着人工智能技术的快速发展,各类机器学习、深度学习已在国内外地热资源评价与开采优化中发挥作用^[15]。联合采用数值模拟方法与试验设计方法构建数据集,再通过决策树、随机森林、LightGBM和多层感知机等机器学习方法生成预测开采井温度、压力等参数的代理模型^[16-20]。可见,机器学习已在地热开发利用中发挥越来越重要的作用,但机器学习与地热开发利用的交叉融合仍处于起步阶段,还需进一步加强两者的交叉融合。因此,本研究建立三维热储数值模型,并通过不同情景下的采灌模拟构建训练集与测试集,提出基于机器学习代理模型的地热开采参数优化方法,最大限度地提高地热井产能。

1 基于机器学习代理模型的地热开采参数优化方法

基于代理模型对地热开采参数进行优化设计包括以下3步:(1)建立三维热储数值模型,并利用

监测数据对模型进行校准;(2)确定目标函数,构建数据集,建立开采参数和热储响应之间的机器学习代理模型;(3)基于代理模型与优化算法得到最优开采参数。针对地热对井系统中回灌井井位、采灌量、尾水温度3个开采参数进行优化设计,热储响应则包括开采井温度和水位。

1.1 三维热储数值模型

地热对井系统通常包括地质体(盖层、储层、断层等)和地热井(开采井、回灌井)2部分。采用多孔介质热弹性理论描述地质体中应力、渗流、传热三场耦合过程。为克服地质体与地热井尺寸相差悬殊的计算瓶颈,将地热井简化为一维线单元模型,考虑沿井筒轴向的渗流传热过程,井筒内流体与盖层的热交换通过等效换热系数来近似考虑^[21]。三维热储采灌模拟的具体算法请参考文献[20-22],不再赘述。

1.2 机器学习代理模型

1.2.1 目标函数

地热开采参数优化的目标是使 N 年内的经济收益最大化,定义一个目标函数(C),其包含以下2部分:第一部分是地热资源开采所能提供的收益(C_T),另一部分是由于开采井中的储层压力降低而产生的额外电力成本(C_P)。

$$C = \sum_{t=0}^N (C_T - C_P) \quad (1)$$

$$C_T = \rho_f \cdot C_{p,f} \cdot q \cdot \Delta T \cdot \eta \cdot p_h \quad (2)$$

$$C_P = \Delta P \cdot q \cdot p_e \quad (3)$$

式中, ρ_f 为流体密度; $C_{p,f}$ 为流体比热容; q 为采灌量; ΔT 为开采和尾水温度的差值; η 为热能利用效率; p_h 为热价; ΔP 为开采井内压力的降幅; p_e 为电价。

不合理的采灌井间距、回灌量和尾水温度可能会引起热突破,使得开采温度降低, C_T 减小;不合理的采灌井位置和开采量可能会引起开采井中的储层压力下降,使得将水抽到地面所需的电力成本增加, C_P 增大。因此,该目标函数综合考虑了热突破和储层压力下降的影响,并将不同量纲的物理值统一量化为经济收益。

1.2.2 数据集构建

构建开采参数和热储响应之间的机器学习代

理模型需要满足要求的数据集。假设开采井井位已确定,在热储数值模型中不同位置布置回灌井,进行不同情景的采灌模拟来获取该对井系统热储响应,重复该过程形成满足机器学习要求的数据集。

除开采参数外,预测变量还包括3个代表性区域的最大和最小渗透,如图1所示,其中, R 为采灌井间距; θ 为采灌井连线与水平线的夹角。代表性区域1是以回灌井为中心、 $0.5R$ 为半径的区域,该区域的渗透率极值可以表征回灌井周围的渗透率分布情况。代表性区域2、3是以采灌井连线的中间位置为中心, $0.75R$ 、 R 为半径的区域,该区域的渗透率极值可以表征优势渗流通道的位置。这3个代表性区域的最大和最小渗透率可以显著提高代理模型的预测精度,因此将作为代理模型的输入参数。值得注意的是,不同的输入参数和热储响应具有不同的维度和变化范围,因此使用线性函数归一化(min-max scaling)进行数据无量纲化处理。

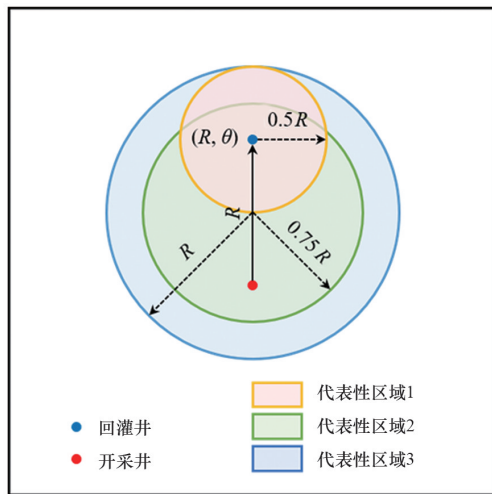


图1 3个代表性区域的预测变量示意

1.2.3 代理模型

采用4种基于树的机器学习模型,即决策回归树、随机森林、LightGBM和一种深度学习模型即多层感知机构建开采参数和热储响应之间的代理模型(图2)。

1) 决策回归树。

决策回归树(decision regression tree, DRT)是

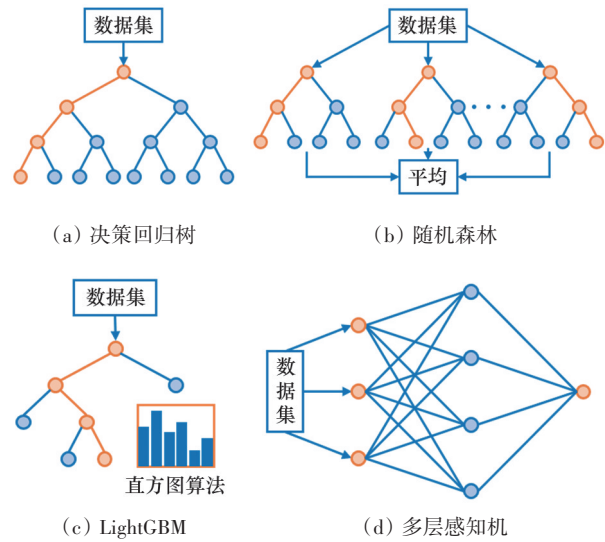


图2 机器学习方法流程

决策树(decision tree, DT)的回归分支。根据特征的重要性,对特征空间进行超平面划分,得到对应输出连续变量的叶节点,每个内部节点仅决策分支的方向,样本最终被划分至某个叶节点从而得到回归预测值。

2) 随机森林。

随机森林(random forest, RF)以决策树为基学习器,在随机采样基础上引入随机属性选择策略,内部节点分裂时特征集合为随机抽取的特征子集。RF通过样本扰动与属性扰动等方式,扩大决策树基学习器差异性,同时各种剪枝策略减小过拟合。

3) LightGBM。

LightGBM(LGBM)是梯度提升决策树(gradient boosting decision tree, GBDT)的衍生方法,通过直方图算法寻找最优切分点,减小内存消耗和训练时间。

4) 多层感知机。

多层感知机(multilayer perceptron, MLP)属于人工神经网络(artificial neural network, ANN),由输入层、隐藏层和输出层构成,激活函数能提高网络的非线性。网络采用误差反向传播算法(back propagation, BP),基于自动微分和梯度下降反向更新各层的连接权和偏置项,重复整个过程直至网络收敛。

为对比评估不同机器学习算法的预测精度,采用 R 方值(R^2)、均方根误差(RMSE)、平均绝对误差(MAE)和平均百分误差(MAPE)这4个指标对得到的代理模型进行综合评估

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n (\text{APE})_i = \frac{1}{n} \sum_{i=1}^n \left| \frac{y'_i - \hat{y}'_i}{y'_i} \right| \quad (7)$$

式中, i 为样本号; y 为真实值(标准化后); \hat{y} 为预测值(标准化后); \bar{y} 为真实值均值(标准化后); y' 为真实值(标准化前); \hat{y}' 为预测值(标准化前);APE为绝对百分误差(相对误差); n 为集合样本数。

采用机器学习算法库Scikit-learn来构建基于树的机器学习代理模型,基于TensorFlow平台搭建深度学习代理模型。通过网格搜索(grid search)和五折交叉验证算法(5-fold cross-validation)寻找验证集上表现最优的超参数,即将训练集划分为5个独立的折叠(fold),每轮搜索都依次将每个折叠作为验证集,其余折叠作为训练集,并取5次的评估指标平均值作为本轮搜索的模型误差值,直至搜索到使该误差值最小的最优超参数。该方法可有效避免陷入局部最优解。

1.3 开采参数优化设计

遗传算法(genetic algorithm, GA)是基于自然选择和遗传学的概念,以其高效率和强鲁棒性而得到广泛应用。在遗传算法中,每代种群通过选择、交叉、变异等运算生成下一代种群,种群中的个体适应度不断提高,并在达到预设的最大迭代次数或最小偏差时终止迭代。遗传算法采用群体搜索方式,并以一种概率方式迭代生成新的个体,从而增加了搜索过程的灵活性,避免搜索过程收敛于局部最优,具有较好的全局优化求解能力。将代理模型与遗传算法相结合,可以有效搜索目标函数的最大值,从而避免了大量的热储动态响应数值模拟。采

用高性能实用型进化算法工具箱Geatpy^[23]实现优化过程。

2 地热对井系统

2.1 工程概况

选取16 km²的面积作为数值模型区域,深度为1430~1130 m。热储的渗透率场由转向带法生成,被划分为边长为100 m的立方体区域(图3),相关长度为200 m,平均渗透率为4.83×10^{-11.8} m²,方差为0.5。假设开采井和回灌井的裸眼段长度均为100 m。对井系统参数取值如表1所示。

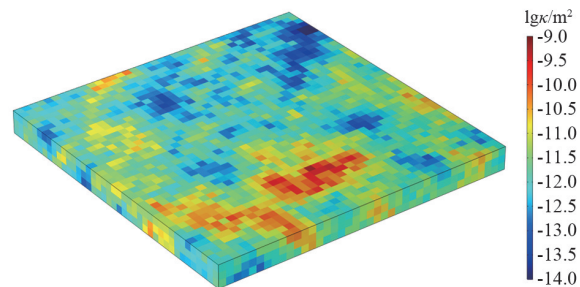


图3 热储渗透率场

表1 对井系统参数

	参数	取值
热储岩石	密度 $\rho_l / (\text{kg} \cdot \text{m}^{-3})$	2600
	渗透率 $\text{lg} \kappa / \text{m}^2$	-11.8
	比热容 $C_{p,s} / (\text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1})$	880
	导热系数 $k_{p,s} / (\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1})$	1.68
	孔隙率 φ	0.26
开采流体	密度 $\rho_l / (\text{kg} \cdot \text{m}^{-3})$	1000
	比热容 $C_{p,l} / (\text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1})$	4200
	导热系数 $k_{p,l} / (\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1})$	0.58
	电价 $p_e / (\text{元} \cdot (\text{kW} \cdot \text{h})^{-1})$	0.58
	热价 $p_h / (\text{元} \cdot \text{GJ}^{-1})$	88.9

储层中的初始水压场是在重力作用下产生的,因此侧边界上施加的水压与深度呈线性相关,考虑到数百米的储层厚度仅导致储层顶部和底部之间的微小温差,储层中的初始温度场假定为 $60 + 0.0025x$ ℃(x 为模型中的 x 向坐标值),并在深度方向为常数。对于渗流过程,侧边界为恒压边界,顶

部和底部边界为不透水边界。对于传热过程,侧边界为开边界,顶部和底部边界为恒温边界。

兼顾计算精度和效率,地热井及周围区域的网格细化,最大单元尺寸为 5 m,其余储层区域网格粗化,最大单元尺寸为 140 m。模型总计包含约 48000 个四面体单元,采灌井由 40 个一维线单元组成。

考虑了 50 a 的连续采灌时间,时间步长假定为 1 月,在此期间,连续监测开采井中温度和水位的变化,以计算目标函数 C 。

2.2 开采参数优化流程

首先构建数据集。假设开采井位于模型的中心,回灌井布置在距离开采井 200~1000 m 的范围内(图 4),采灌量在 80~120 m³/h 间,尾水温度在 25~35℃ 间,设置均匀分布随机数,在保证随机性的同时,使得数据集中的开采参数尽可能地覆盖整个优化空间。共生成 600 组数据,其中训练集占 85%,测试集占 15%。预测变量包括 4 个开采参数,即采灌井间距(R)、采灌井连线与水平线的夹角(θ)、采灌量(Q)、尾水温度(T_m),以及 6 个渗透率极值。目标函数由热储响应计算得到。然后基于不同的机器学习方法建立代理模型,通过评估指标选择预测精度最高的代理模型,在此基础上结合遗传算法优化地热开采参数。

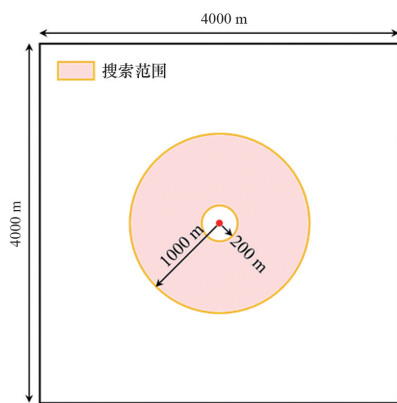


图 4 回灌井位置范围示意

3 结果与分析

3.1 代理模型误差分析

采用决策回归树、随机森林、LightGBM 和多层

感知机构建开采参数和热储响应之间的代理模型。不同机器学习方法训练得到的代理模型如表 2 所示, R^2 、RMSE、MAE、MAPE 这 4 个评价指标表明,基于多层感知机的代理模型是最优的,在训练集规模仅为 510 时, R^2 趋近于 1,说明代理模型能够以足够的精度预测目标函数(图 5)。因此,使用基于多层感知机的代理模型寻找最优地热开采参数。

表 2 代理模型评估指标

指标	DRT	RF	LGBM	MLP
R^2	0.965	0.992	0.990	0.999
RMSE	0.036	0.017	0.022	0.005
MAE	0.028	0.012	0.014	0.004
MAPE	0.020	0.009	0.010	0.003

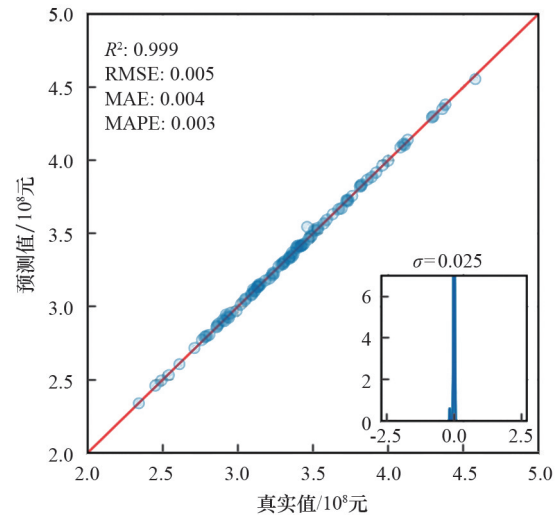


图 5 目标函数值散点图和误差直方图

3.2 开采参数优化结果

优化结果的准确性可以从 2 个方面确定:(1) 在使用基于代理模型的优化方法得到最优开采参数后,比较通过代理模型和直接模拟得到的目标函数值;(2) 比较基于代理模型和基于数值模拟的优化方法得到的最优开采参数和目标函数值。

遗传算法的种群规模和遗传代数根据对井系统的复杂性进行确定,同时优化了回灌井井位、采灌量和尾水温度多个开采参数,因此假定种群规模为 40,最大遗传代数为 100,交叉概率和变异概率设置为 0.95、0.05。

结合基于多层感知机训练得到的代理模型和

遗传算法,在已知范围内寻找最优开采参数。数值模拟每次运行时间约为 60 s,构建 600 组数据共需 10 h,优化过程约为 5 min。结果表明,通过基于代理模型的优化方法,回灌井的最优位置为 $x = 408.742$ m 和 $y = -214.040$ m,最优采灌量为 $Q = 120$ m^3/h ,最优尾水温度为 $T_m = 25^\circ\text{C}$,目标函数值为 4.632×10^8 元。基于这组采灌方案,用数值模拟可以得到 4.719×10^8 元,这与预测结果非常接近。同样,在相同种群规模和遗传代数下,可使用基于模拟的传统优化方法寻优。运行约 67 h 后,遗传代数达到最大预设值。最终优化结果为 $x = 410.080$

m 和 $y = -217.005$ m,最优采灌量为 $Q = 120$ m^3/h ,最优尾水温度为 $T_m = 25.007^\circ\text{C}$,目标函数值为 4.713×10^8 元。迭代对比结果如图 6 所示。2 种方法得到的开采参数非常接近,这表明基于代理模型的优化方法可以凭借更低的计算成本获得近似的优化结果。此外,针对选取的非均质渗透率场,采灌量和尾水温度分别取到了寻优范围内的极值,说明虽然较大的采灌量会引起开采井中的储层压力下降从而增加电力成本,且较低的尾水温度会引起热突破使得经济收益下滑,但两者的负面作用远小于其带来的经济效益增加幅度。

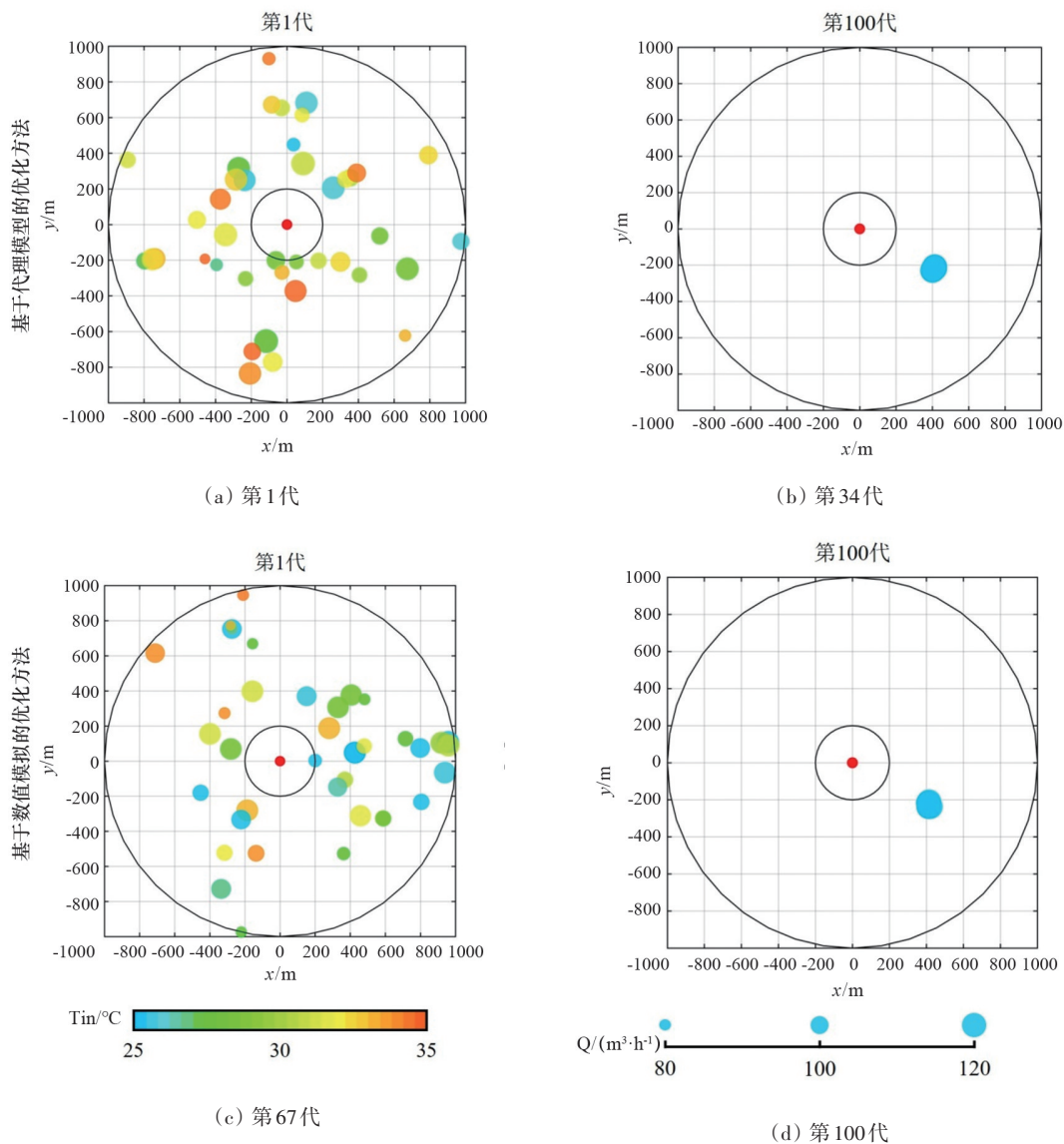


图6 不同优化方法的最优井位迭代结果对比

4 结论

开发了一种基于代理模型的优化方法来寻找热储中的最优井位、采灌量和尾水温度,与基于模拟的优化方法相比,该方法的计算成本更低。通过多层感知机机器学习方法和少量预测变量训练的代理模型能够准确预测热储的目标函数。将代理模型与遗传算法相结合,形成了基于代理模型的优化方法,证明了该方法的合理性和有效性。得到如下主要结论:相较于其他机器学习方法,多层感知机的预测性能最佳,在训练集规模较小时也能有较好的预测精度;基于代理模型的优化方法可以凭借更低的计算成本,准确地找到给定开采井位置时最优的回灌井位置、采灌量和尾水温度,且优化结果的误差在可接受范围内。

参考文献(References)

- [1] 汪集暘, 庞忠和, 孔彦龙, 等. 我国地热清洁取暖产业现状与展望[J]. 科技促进发展, 2020, 16(增刊1): 294-298.
- [2] 曹锐, 多吉, 李玉彬, 等. 我国中深层地热资源赋存特征、发展现状及展望[J]. 工程科学学报, 2022, (优先发表)
- [3] 王贵玲, 刘彦广, 朱喜, 等. 中国地热资源现状及发展趋势[J]. 地学前缘, 2020, 27(1): 1-9.
- [4] 孔彦龙, 庞忠和, 邵亥冰, 等. 面向成本的中深层地热储群井采灌优化布局研究[J]. 科技促进发展, 2020, 16(增刊1): 316-322.
- [6] O'Sullivan M J, Pruess K, Lippmann M J. State of the art of geothermal reservoir simulation[J]. *Geothermics*, 2001, 30(4): 395-429.
- [7] Pandey S N, Vishal V, Chaudhuri A. Geothermal reservoir modeling in a coupled thermo-hydro-mechanical-chemical approach: A review[J]. *Earth-Science Reviews*, 2018, 185: 1157-1169.
- [8] Liang X, Xu T F, Feng B, et al. Optimization of heat extraction strategies in fault-controlled hydro-geothermal reservoirs[J]. *Energy*, 2018, 164: 853-870.
- [9] Guo L L, Zhang Y B, Wang Z C, et al. Parameter sensitivity analysis and optimization strategy research of enhanced geothermal system: A case study in Guide Basin, Northwestern China[J]. *Renewable Energy*, 2020, 153: 813-831.
- [10] Mahmoodpour S, Singh M, Turan A, et al. Simulations and global sensitivity analysis of the thermo-hydraulic-mechanical processes in a fractured geothermal reservoir[J]. *Energy*, 2022, 247: 123511.
- [11] Samin M Y, Faramarzi A, Jefferson I, et al. A hybrid optimisation approach to improve long-term performance of enhanced geothermal system (EGS) reservoirs[J]. *Renewable Energy*, 2019, 134: 379-389.
- [12] Zhang S Y, Jiang Z J, Zhang S S, et al. Well placement optimization for large-scale geothermal energy exploitation considering nature hydro-thermal processes in the Gonghe Basin, China[J]. *Journal of Cleaner Production*, 2021, 317: 128391.
- [13] Asai P, Panja P, McLennan J, et al. Performance evaluation of enhanced geothermal system (EGS): Surrogate models, sensitivity study and ranking key parameters[J]. *Renewable Energy*, 2018, 122: 184-195.
- [14] Chen M J, Tompson A F B, Mellors R J, et al. An efficient optimization of well placement and control for a geothermal prospect under geological uncertainty[J]. *Applied Energy*, 2015, 137: 352-363.
- [15] Okoroafor E R, Smith C M, Ochie K I, et al. Machine learning in subsurface geothermal energy: Two decades in review[J]. *Geothermics*, 2022, 102: 102401.
- [16] Aydin H, Akin S, Senturk E. A proxy model for determining reservoir pressure and temperature for geothermal wells[J]. *Geothermics*, 2020, 88: 101916.
- [17] Zhou L, Zhang Y J, Hu Z J, et al. Analysis of influencing factors of the production performance of an enhanced geothermal system (EGS) with numerical simulation and artificial neural network (ANN)[J]. *Energy and Buildings*, 2019, 200: 31-46.
- [18] Jiang Z J, Zhang S Y, Turnadge C, et al. Combining autoencoder neural network and Bayesian inversion to estimate heterogeneous permeability distributions in enhanced geothermal reservoir: Model development and verification[J]. *Geothermics*, 2021, 97: 102262.
- [19] Gudala M, Govindarajan S K. Numerical investigations on a geothermal reservoir using fully coupled thermo-hydro-geomechanics with integrated RSM-machine learning and ARIMA models[J]. *Geothermics*, 2021, 96: 102174.
- [20] Wang J C, Zhao Z H, Liu G H, et al. A robust optimization approach of well placement for doublet in heterogeneous geothermal reservoirs using random forest tech-

- nique and genetic algorithm[J]. *Energy*, 2022, 254: 124427.
- [21] 赵志宏, 刘桂宏, 徐浩然. 深地能源工程热力多场耦合效应高效模拟方法[J]. *工程力学*, 2020, 37(6): 1-18.
- [22] Wang G L, Liu G H, Zhao Z H, et al. A robust numerical method for modeling multiple wells in city-scale geothermal field based on simplified one-dimensional well model[J]. *Renewable Energy*, 2019, 139: 873-894.
- [23] Jazzbin J. Geatpy: The genetic and evolutionary algorithm toolbox with high performance in python[EB/OL]. [2022-08-30]. <http://www.geatpy.com/>.

Optimizing development parameters of geothermal energy using machine learning technique

WANG Jiacheng¹, CHEN Jinfan¹, ZHAO Zhihong^{1*}, TAN Xianfeng²

1. Department of Civil Engineering, Tsinghua University, Beijing 100084, China

2. Shandong Lunan Geological Engineering Investigation Institute, Jining 272100, China

Abstract In order to solve the problem that the simulation-based optimization method usually requires a large number of simulations, in this paper, after comparing the prediction performance of different machine learning methods, a surrogate model based on MLP was developed to reduce the computational cost, which was then combined with genetic algorithm to develop an optimization method of development parameters for geothermal doublets in heterogeneous geothermal reservoirs. Through the case study of a doublet system, the reasonability and efficiency of the developed optimization method of development parameters were demonstrated. The results show that when given a certain position of production well, the surrogate model-based optimization method of development parameters can accurately find the optimal placement of injection well, the rate of production and injection, and the temperature of recharge water with lower computational cost.

Keywords geothermal doublet system; machine learning; optimization of development parameters; surrogate models; genetic algorithm ●



(责任编辑 刘志远)