# 机器学习模型在地热开发水温预测中的应用

董珮瑶[1,2]，杜利[1,2]，赵磊[1,2]，包一凡[1,2]，尹茂生[3*]

1. 深层地热富集机理与高效开发全国重点实验室，北京 100083；
2. 中石化新星（北京）新能源研究院有限公司，北京 100083；
3. 南方科技大学环境科学与工程学院，广东 深圳 518055

第一作者，E-mail：dongpeiyao.xxsy@sinopec.com

* 通信作者，E-mail：myin@eitech.edu.cn

2024-02-23 收稿；2024-06-24 修回；2024-12-17 接受

基金项目：中国石化集团科研项目(JP23087; JP23178)

中图分类号：TP181；P314；TK529　　文章编号：2096-8523(2025)03-0388-11　　doi: 10.19509/j.cnki.dzkq.tb20240063

**摘　要：**地热作为一种清洁能源具有广阔的应用前景，可持续地开发和利用地热资源中地热水的温度评估是重要的研究课题。人工智能技术已成为矿产和油气资源勘探开发研究的热点和前沿方向，然而在地热资源开发方面相关研究和应用较少。剖析了油气资源开发中大数据与人工智能应用的重要价值，对当前地热资源开发中人工智能技术的应用与探索进行了介绍。以陕西咸阳地热田为例，采用长短期记忆(long short-term memory，简称 LSTM)神经网络构建了以灌定采模式下单井水温的时间序列模型；采用随机森林和 XGBoost 算法，建立了多个井地热水温度的预测模型。研究结果表明，建立的机器学习模型在地热水温度预测方面表现优秀，模型准确度均在 95% 以上，且速度快。该地区地热水温度的首要影响因素是取水段顶深，模型验证了渭北断裂带对热储的重要作用。实例应用验证了机器学习模型在解决地热资源开发复杂难题中的优越性，人工智能技术的合理应用能够为地热资源的高效开发和科学降本提质增效提供更多有效的决策依据。

**关键词：**地热开发；机器学习；模拟；地下水；水温预测

地热资源作为一种可再生能源，具有显著的可持续性开发潜力[1]。在地热资源开发中，为了合理地布置地热钻井和预测回灌时热突破时间(冷水从回灌井出发到达生产井所需的时间)，需要准确了解地热水的温度分布和地热井运行时的水温变化过程。中国在地热资源的勘探与利用方面的重视程度相对较低，投入不足[2]，导致相关资料数据相对贫乏[3]。此外，考虑到中国地热资源开发起步于 20 世纪 70 年代初[4]，发展历史较短，受到地理及地质条件的限制，当前的技术方法和基础数据存在局限性[5-6]。油井伴生地热资源丰富，地热开发策略可以借鉴传统的油气开发模式[7]。值得注意的是，随着大数据和人工智能技术的迅速进步，相关方法在解决复杂非线性问题方面已展现出广阔的应用前景。目前，在油气资源开发和产量预测等领域已开展了机器学习技术的应用探索[8]，并且预示着机器学习技术在地热资源开发领域可能取得创新性的进展。

油气资源开发中的关键决策如层系组合、井位选择和注采量调整等，通常依赖于数值模拟和生产优化方法，但这些方法存在求解难度大、计算耗时长且收敛速度慢的问题[9]。此外，对单井初始产量、递减率和累计产量等开发指标的预测也是至关重要的[10]。传统预测方法依赖数值模拟和地质工程技术，但存在周期长和准确度低的问题[9]。为克服这些局限性，人工智能技术的应用日益增多，特别是在挖掘实时生产数据中蕴含的生产信息方面，通过建立

**著录信息：**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 Peiyao, DU Li, ZHAO Lei, 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.

© 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

油藏或单井的代理模型，代替传统的数值模拟计算或产能分析方法，实现对剩余油分布和生产指标的快速准确预测[11]。例如，采用长短期记忆（LSTM）神经网络方法构建产能时序模型，使用历史生产数据来预测油井未来的产油量[12]。基于人工神经网络的井底压力预测模型，通过输入井底温度、溶解气油比和油气水产量等监测数据进行训练[13]。这些创新技术为油气资源开发提供了新的分析和预测工具，有助于提高决策的准确性和效率，可以借鉴应用于地热资源开发。

在地热资源开发方面，可以利用基于流动传热偏微分控制方程的数值模拟软件（如 TOUGH、COMSOL Multiphysics 等）进行地热水温计算[14-15]，然而数值模拟方法所需的水文地质资料丰富，野外调查、模型求解计算和优化耗时较长[14]。此外，由于一些地区碳酸盐岩储层的非均质性强，难以精细刻画复杂地质条件下的地下水热运移[16-17]。因此，有必要开发准确度高、速度快、不受地质条件限制的具有普适性的地热水温度模拟预测方法。已有研究[18]将机器学习方法应用于深部地热水的开发评价和可持续发展分析。例如，张娟娟等[19]基于 BP 神经网络构建了天津地热资源开发能力评价指标体系。董华松等[20]利用小波神经网络实现了地热影响因素定量分析，并进行了模型参数优化。赵磊等[21]利用 LSTM 解决了地热可持续利用技术中预测回灌井日压力的难题。最新的研究重点在于利用机器学习开发新的反演方法，以便从示踪剂试验数据中研究裂隙孔径分布或传热表面积[22]。例如，WU 等[23]和 HAWKINS 等[24]使用机器学习方法确定了裂隙孔径分布，并预测了热量下降。以上研究进展表明，人工智能和机器学习技术正成为提高地热资源开发效率和精确度的关键工具。

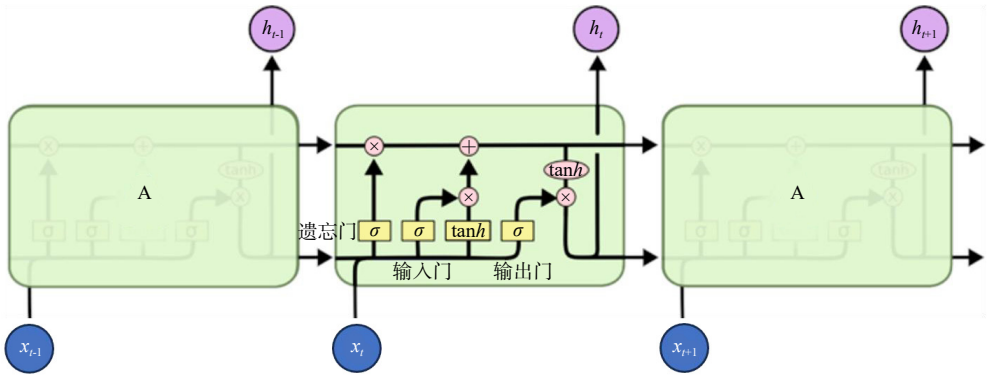
陕西咸阳地热田是关中盆地最重要的地热田之一，关中盆地地热资源储量大、开发历史悠久。咸阳是地幔隆起区，具有良好的地热背景和较高的热流背景[25]。截至 2016 年底，咸阳市共有热水井 105 口，回灌井 8 口，市区地热水开采量 $709.55\times10^4$ $m^3/a$[26]。本研究围绕陕西咸阳地区地热井的多年监测数据和抽水试验数据展开，旨在通过先进的数据分析技术，提升地热资源开发的效率和效果。为此，本研究采用 LSTM 神经网络算法构建了以灌定采模式下单井水温的时间序列模型；采用随机森林和 XGBoost（eXtreme gradient boosting）算法，分别构建了多个井地热水温度的分析模型；模拟了在特定采灌模式下以及多个井开发条件下的温度变化，进而探讨了影响该地区地热水温度的各种因素。实际应用表明，本研究构建的机器学习模型能够有效利用有限的资料预测地热水温度变化，从而证实了大数据和人工智能技术在解决地热资源开发中的复杂难题上具有准确度高、速度快的优势。不仅展示了先进的人工智能技术在地热资源开发领域的实际应用价值，也为地热资源的高效开发提供了科学的降本提质增效策略，为地热行业技术人员提供了更多有效的决策依据。研究成果对于推动地热能源的持续、科学和高效开发具有重要意义。

## 1 理论与方法

### 1.1 集成学习与深度学习算法

集成学习算法在机器学习中通过组合多个学习器来解决分类和回归问题，常优于单一学习器[27]。集成学习算法大致可以分为 3 种类型：Bagging、Boosting 和 Stacking。本研究主要采用 Bagging 和 Boosting 方法进行地热井水温预测。Bagging 方法通过整合多个基模型的预测结果来提高整体性能，各个基学习器相互独立，允许并行处理。Bagging 方法致力于在过拟合与欠拟合之间找到一个平衡点，其最典型的代表是随机森林。Boosting 方法起始于弱学习器，并通过加权训练逐步加强其性能。与 Bagging 相比，Boosting 方法通常能够获得更好的效果，但它无法并行训练，并且由于对错误样本给予更多关注，也可能导致过拟合[28]。Boosting 算法中最为人熟知的包括自适应增强（adaptive boosting，简称 AdaBoost）和梯度提升机（gradient boosting machine，简称 GBM），GBM 通过梯度下降方法优化模型，XGBoost 是其一种变体[29]。

深度学习算法作为机器学习的一个重要分支，基于神经网络算法的发展，已经逐渐成为人工智能和数据分析领域的主流方法[30]。特别是在时间序列分析方面，深度学习提供了多种有效的处理方法，包括循环神经网络（recurrent neural networks，简称 RNN）中的 LSTM 模型、卷积神经网络和基于注意力机制的模型。RNN 是一种包含循环的神经网络结构，允许信息在网络中持久化。RNN 中的每个单元（图 1）可视为神经网络的一个缩影，它接收特定时刻的输入 $x_t$，并输出相应的结果 $h_t$，$\sigma$ 和 tan$h$ 是激活函数。RNN 结构使得信息能够从一个时间步骤传递到下一个时间步骤。LSTM 神经网络是 RNN 的一个特




A 为一个单元; $\sigma$ 和 tan$h$ 均为激活函数; $x_{t-1}$, $x_t$ 和 $x_{t+1}$ 分别为 $t$-1, $t$ 和 $t$+1 时刻的输入; $h_{t-1}$, $h_t$ 和 $h_{t+1}$ 分别为 $t$−1, $t$ 和 $t$+1 时刻的输出

图 1　LSTM 神经网络结构原理

Fig. 1　LSTM neural network structure schematic

殊变体[31], LSTM 单元可通过遗忘门、输入门和输出门进行有选择地添加或遗忘信息(图 1)。时间序列数据的长度、维数和数量的增加会使得 RNN 的学习能力降低, LSTM 神经网络成功地解决了传统 RNN 的缺陷。

### 1.2　模型评价指标

为了全面评估模型的泛化能力, 必须采用适当的评估标准。评价指标的选择不仅对模型性能的解释产生影响, 而且对模型的优化方向具有指导意义。地热水温度预测属于回归类问题, 在回归模型评估中, 常用的指标包括决定系数($R^2$)、解释方差($EV$)、平均绝对误差($MAE$)、均方根误差($MSE$)以及计算时间等。$R^2$ 和 $EV$ 是评价模型拟合效果的重要指标; $R^2$ 和 $EV$ 的值越接近 1, 表明模型对数据的解释能力越强, 预测效果越佳; $MAE$ 和 $MSE$ 则直接衡量预测值与实际值之间的偏差, 二者的值越小, 表示模型的预测准确度越高, 误差越小。除此之外, 计算时间也是一个重要的考量因素, 尤其是在需要快速响应的实际应用中。

### 1.3　特征重要性分析

特性重要性是指标识出与模拟结果相关性最高的特征因素变量, 有助于深入理解所处理问题的本质, 还能通过特征选择改进模型性能。计算特征重要性有 2 种常用的与模型无关的方法: 基于排列的计算方法和 SHAP(shapley additive exPlanations)计算方法。它们适用于多种算法, 如 XGBoost 和神经网络等。

LUNDBERG 等[32] 提出的 SHAP 方法是一种可以解释机器学习模型的特征归因方法(公式(1)), 在揭示潜在现象方面表现出卓越的性能[33]。该方法通过评估特征对模型输出的边际贡献, 从全局和局部层面解释机器学习模型。它构建了一个加性模型, 将预测值分解为特征的贡献, 计算每个特征的 $SHAP$ 值, 即其在所有特征组合中的边际贡献均值。

$$g(\boldsymbol{z}') = \varphi_0 + \sum_{j=1}^{M} \varphi_j \boldsymbol{z}'_j, \tag{1}$$

式中: $g$ 为解释模型; $\boldsymbol{z}' \epsilon \{0,1\}^M$ 为联盟向量; $\varphi_0$ 为平均贡献值; $M$ 为最大联盟的数目; $\varphi_j$ 为特征 $j$ 的 Shapley 值; $\boldsymbol{z}'_j$ 为特征 $j$ 是否存在。

## 2　实例应用

### 2.1　研究区概况

咸阳地热田位于关中盆地中部腹地, 如图 2。东起泾河与渭河交汇处, 西至武功−扶风县界, 南临渭河, 北接宝鸡峡引渭工程北干渠, 东西长近 100 km, 南北平均宽约 8 km。关中盆地为一新生界断陷盆地, 其北部为斜坡带, 南部为坳陷区, 由 2 个斜列的次级凹陷组成, 即西安凹陷和固市凹陷。

关中盆地受近 EW 向、NE 向及 NW 向正倾滑断层系统的控制。渭北断裂是关中盆地的控盆断裂, 也是盆地基底分界, 两侧地层变化大。渭北断裂整体呈近 EW 走向, 大体沿渭河呈向南凸出的弧形展布, 延伸长度超过 300 km, 为一南倾正断层。渭北断裂穿咸阳而过, 在咸阳地区走向 NEE, 倾角 45° ~ 80°, 断层切割深度在 20 km 以上, 上新统被错断 700 ~ 1 000 m。由于渭北断裂规模大、切割深、继承性活动强烈, 与深部热源沟通紧密, 因而是控制咸阳地热田的主要储热储水和导热导水断裂。盆地内燕山期花岗岩的放射性铀的衰变可能形成 $9.4\times10^{22}$ MJ 的总热量[26], 放射性铀衰变和地幔隆起是咸阳地热

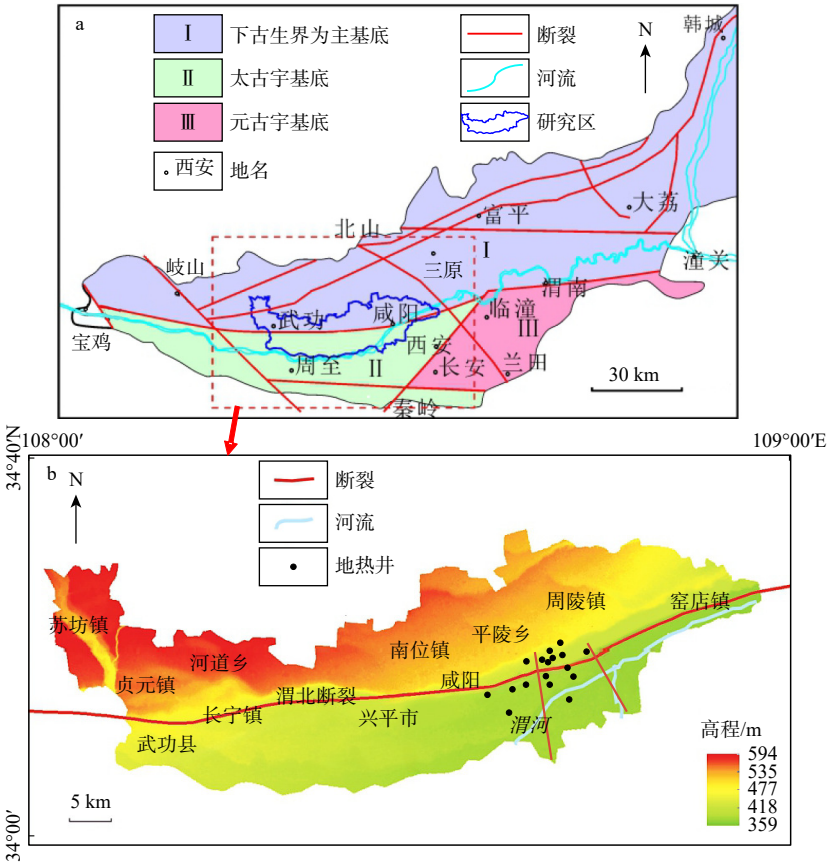



图 2　陕西咸阳地热田地质构造 (a) 和地热井分布 (b) 图

Fig. 2　Geological structures (a) and producing well locations (b) of the Xianyang geothermal field, Shaanxi Province

田形成的主要原因。

目前咸阳地热田开发的主力热储层是新近系的张家坡组，蓝田灞河组和高陵群，岩性为砂岩，深度为 800 ~ 4 000 m。它们是咸阳地热田的主要开采层，顶面埋深分别为 [800, 1 500)，[1 500, 2 700)，[2 700, 3 200] m，砂岩的孔隙度分别为 10% ~ 20%，15% ~ 35% 和 5% ~ 20%。

### 2.2　抽水试验与数据

抽水试验是确定含水层水文地质参数最常用的技术方法，此外，根据抽水试验可以评估含水层的富水程度与地热井的产水能力，初步获得水温[14]。咸阳 54 口完钻地热井均按《供水水文地质勘察规范》(GB50027-2001)[34] 进行了大、中、小三落程稳定流抽水试验，记录水位和出水流量，试验结束后观测恢复水位。采用温度计测量出水温，水温 2 h 观测一次，每次测量温度计的浸水时间不少于 10 min。地热井运行时，井口位置安装温度计和传感器，连续监测每日采灌流量、水温。结合钻井资料和抽水试验建立地热井试验资料数据集，包含除中落程试验井水温外的 16 个特征(表 1)，部分井的抽水试验实测数据(多次测量平均值)如表 2 所示。

### 2.3　以灌定采时间序列模型

采用陕西咸阳的一组地热采灌井(YGSY1

**表 1　咸阳 54 口地热井试验资料数据**

Table 1　Content of test data of 54 geothermal wells in Xianyang

| 序号 | 特征名称 | 单位 |
|---|---|---|
| 1 | 井类型 | / |
| 2 | 经度 | ° |
| 3 | 纬度 | ° |
| 4 | 井口海拔 | m |
| 5 | 完钻井深 | m |
| 6 | 井斜深 | m |
| 7 | 取水段顶深 | m |
| 8 | 取水段底深 | m |
| 9 | 取水段中深 | m |
| 10 | 静水位 | m |
| 11 | 中落程降深 | m |
| 12 | 中落程流量 | $m^3/h$ |
| 13 | 大落程降深 | m |
| 14 | 大落程流量 | $m^3/h$ |
| 15 | 小落程降深 | m |
| 16 | 小落程流量 | $m^3/h$ |
| 17 | 中落程水温 | ℃ |

**表 2　咸阳部分地热井的抽水试验实测数据**

Table 2　Pumping test data of some geothermal wells in Xianyang

| 井名 | 取水段顶深/m | 取水段底深/m | 取水段中深/m | 静水位/m | 中落程降深/m | 中落程流量/($m^3 \cdot h^{-1}$) | 中落程温度/℃ |
|---|---|---|---|---|---|---|---|
| WR3 | 2060.90 | 2724.90 | 2392.90 | 12 | 19 | 120.00 | 82 |
| WR5 | 1581.50 | 2961.60 | 2271.55 | −52.5 | 28 | 121.03 | 75 |
| WR6 | 2003.31 | 3033.78 | 2518.55 | 6 | 18 | 130.66 | 83 |
| WR9 | 1055.20 | 1599.13 | 1327.17 | −8 | 14 | 102.00 | 61 |
| WT2 | 909.70 | 1699.93 | 1304.82 | −9 | 31 | 120.00 | 50 |

和 YGSY2)的数据建立单井温度时间序列模型。YGSY2 为 YGSY1 生产井的回灌井，选取 YGSY1 井的抽水流量、YGSY2 井的注水水温和流量的日变化数据作为模型多个输入特征，以 YGSY1 井的水温为目标。采用顺序划分方法对数据进行划分，将 2020 年每日观测数据用于模型训练。由于非供暖季地热井不运行，抽水流量为 0 $m^3/h$，模型中为保证时间序列的连续性，假设非供暖季地热井取水温度为 0℃，实际并未进行地热水的抽取[35]。

深度学习中 RNN 的特殊变体 LSTM 神经网络算法适用于该时间序列分析预测问题，因此利用 LSTM 神经网络算法建立多变量时间序列模型，模拟 YGSY1 井出水温度随时间的变化过程。综合 1.2 节中的各评价指标，采用网格搜索和交叉验证的方法优化调整神经网络模型的参数，即通过网格搜索遍历所有的参数组合，每组参数都用交叉验证进行评估，从而选出最优的参数组合来建立模型。主要优化调整的参数包括时间步长、网络层数、隐藏层神经元、学习率和迭代次数，模型中的各参数优选结果如表 3 所示。

**表 3　LSTM 模型参数优选结果**

Table 3　LSTM model parameter optimization results

| 参数 | 参数取值 | 优选值 |
|---|---|---|
| 时间步长 | 5，10，20，40 | 5 |
| 网络层数 | 1，2 | 1 |
| 隐藏层神经元 | 128，256 | 128 |
| 学习率 | 0.01，0.005，0.001 | 0.001 |
| 迭代次数 | 50，100，200 | 200 |

利用未参与训练过程的 2021—2022 年的结果对模型的效果进行测试，模型模拟计算结果和实际观测结果对比如图 3，YGSY1 生产井的供暖季水温稳定在 83 ~ 85℃。尽管短期内未模拟到温度显著下降，即回灌导致的热突破的发生，但采用 LSTM 模型方法可以预测该采灌模式下未来数十年的每日井水温度。LSTM 神经网络对前序数据变化非常敏感，随着时间推移实时监测的采灌井数据量增加，模型训练和测试的数据集也可以不断增加，一旦发生影响生产井水温的回灌流量和温度的信息，模型可以通过计算当日及此后的水温，快速响应出是否有温度降低的趋势，有利于可持续开发中热突破的预报和预警。同时，可利用该模型计算当前采灌井距下，热突破发生的时间，从而评估当前采灌量的合理性[36]。

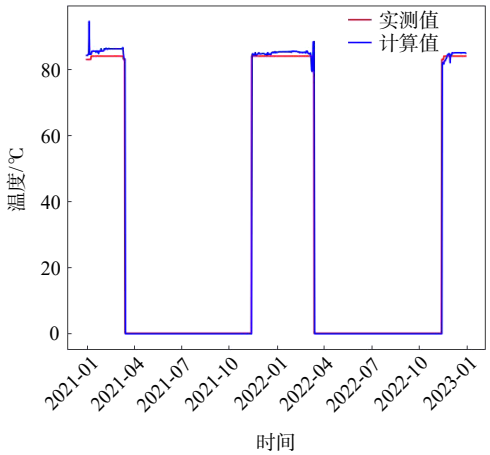



图 3　YGSY1 井水温模拟计算结果和实际观测结果对比

Fig. 3　Comparison of groundwater temperature simulation results and observation results in Well YGSY1

### 2.4　区域多井开发预测模型

采用咸阳 54 口地热井的抽水试验数据建立多井开发温度预测模型。根据现有的大、中、小三落程抽水试验资料数据，获取的 16 个特征如表 1 所示，包括井类型、井位(经度、纬度、井口海拔)、完钻井深、取水段顶深、取水段底深、取水段中深、静水位、中落程降深、中落程流量等，以中落程水温为计算目标。为消除不同特征间量纲的影响，将数据归一化，即将数据线性地转化到 [0, 1] 范围。采用随机取样法对样本数据集进行划分，按训练集和测试集比例为 8∶2，即随机抽取 43 口井的试验数据作为训练集，11 口井的试验数据作为测试集。

由于采用区域内多个井的试验数据属于模拟多个井在空间上的温度分布问题，不考虑温度随时间的变化，利用随机森林和 XGBoost 算法分别建立模型。综合 1.2 节中的各评价指标，采用网格搜索和交

叉验证的方法优化调整随机森林和 XGBoost 模型的参数。随机森林模型主要优化调整的参数包括决策树的个数、决策树最大特征数、决策树最大深度、叶子节点含有的最少样本和节点可分的最小样本数；XGBoost 模型主要优化调整的参数包括学习率、树的最大深度、训练中树的个数、最小叶子节点样本权重和以及模型参数 *gamma*，模型中的各参数优选结果如表 4 所示。

利用优选后的参数建立相应的地热井水温预测模型，并利用未参与训练过程的随机抽取的 11 口井的试验数据对模型的效果进行模拟测试，随机森林模型和 XGBoost 模型模拟计算结果和实际观测结果的对比如图 4，5。随机森林模型模拟结果中，仅 WT2 井和 WT1 井的计算值与实测值差异较大，但差异均不超过实测值的 10%，而 XGBoost 模型对 11 口井的水温计算结果均较好。因此，应优先选择 XGBoost 模型进行地热井水温预测分析。

**表 4　随机森林和 XGBoost 模型参数优选结果**

Table 4　Optimization results of random forest and XGBoost model parameters

| 模型 | 参数 | 参数取值 | 优选值 |
|---|---|---|---|
| 随机森林 | 决策树的个数 | 40，80，100，200，1 000 | 200 |
| | 决策树最大特征数 | ‘auto’，‘sqrt’，‘log2’ | ‘auto’ |
| | 决策树最大深度 | None，10，20，30 | 10 |
| | 叶子节点含有的最少样本 | 2，3，5 | 3 |
| | 节点可分的最小样本数 | 9，10，11 | 9 |
| XGBoost | 学习率 | 0.01，0.05，0.1 | 0.1 |
| | 树的最大深度 | 2，3，4，5 | 3 |
| | 训练中树的个数 | 50，100，200 | 100 |
| | 最小叶子节点样本权重和 | 1，3，5，7 | 1 |
| | 模型参数*gamma* | 0，0.1，0.3，0.5 | 0 |

## 3　分析与讨论

### 3.1　模型评价

在采用 LSTM 模型预测 YGSY1 生产井的出水

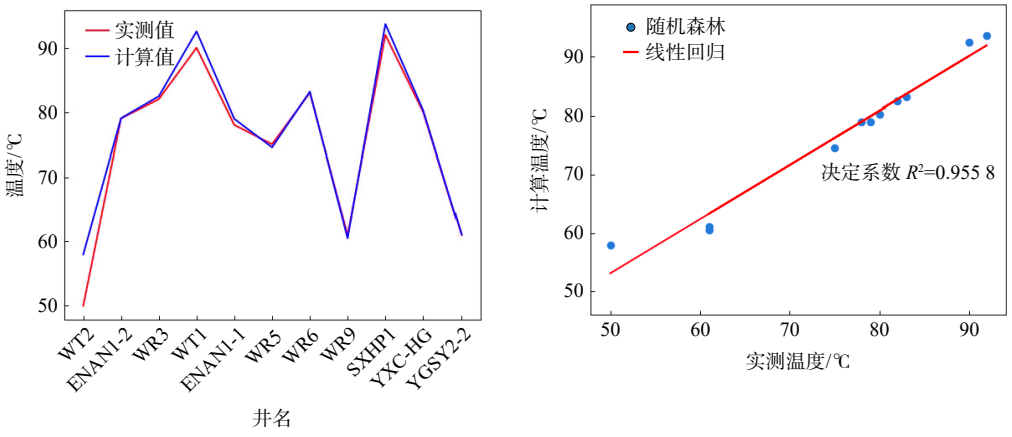



图 4　地热井水温模拟计算结果和实际观测结果对比（随机森林模型）

Fig. 4　Comparison between the simulation results of geothermal well groundwater temperature and the observation results (random forest model)

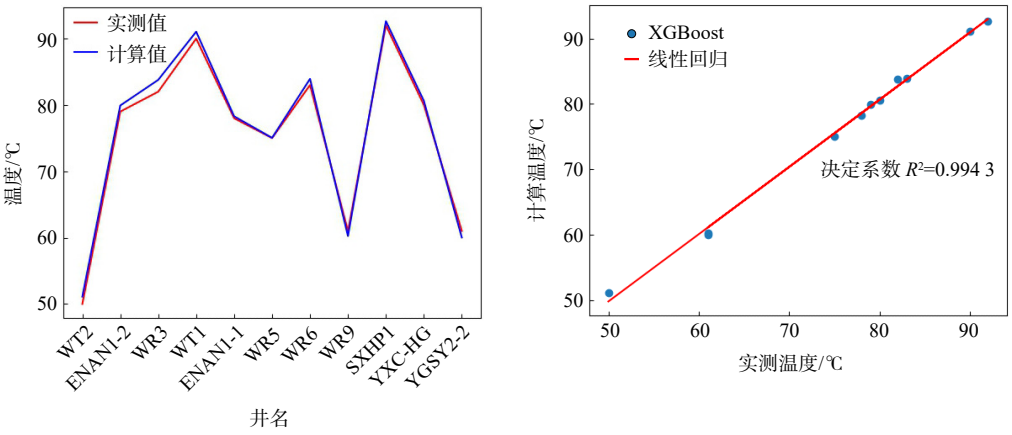



图 5　地热井水温模拟计算结果和实际观测结果对比（XGBoost 模型）

Fig. 5　Comparison between the simulation results of geothermal well groundwater temperature and the observation results (XGBoost model)

温度随时间的变化中，如图 6，通过增加循环迭代次数对模型进行改进评估[37]。将迭代次数增加到 50 次时，能使训练集和测试集的平均绝对误差 *MAE* 从 0.4 下降到 0.04，*MAE* 在 150 次迭代后趋于稳定。如表 5 所示，LSTM 模型模拟结果的决定系数 $R^2$，即预测准确度在 95% 以上，适用于对多变量时间序列的准确预测。采用 LSTM 模型预测 2021−2022 年时间精度为每日的地热水温度时，计算时间约为 17 s，其速度和精度优于现有的对该研究区的地热开发数值模拟方法[38]。

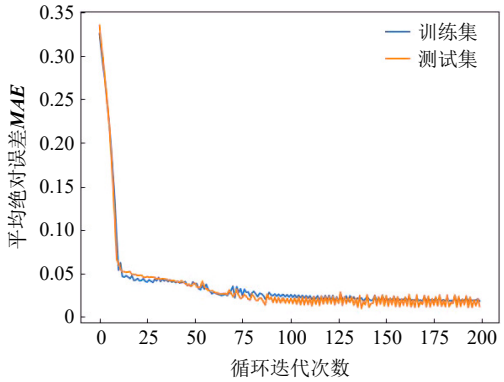



图 6　LSTM 模型平均绝对误差 *MAE* 值随循环迭代次数的变化

Fig. 6　Variation of *MAE* value of LSTM model with the epoch size

**表 5　模型评价指标对比**

Table 5　Comparison of model evaluation indexes

| 评价指标 | LSTM | 随机森林 | XGBoost |
|---|---|---|---|
| 决定系数$R^2$ | 0.974 1 | 0.955 8 | 0.994 3 |
| 解释方差*EV* | 1.000 0 | 0.965 2 | 0.996 0 |
| 平均绝对误差*MAE* | 0.019 8 | 0.021 6 | 0.013 0 |
| 均方根误差*MSE* | 0.005 8 | 0.001 7 | 0.000 2 |
| 计算时间/s | 16.562 5 | 0.131 8 | 0.092 8 |

对比图 4，5，随机森林模型和 XGBoost 模型的决定系数 $R^2$ 均大于 0.95，表明这 2 种方法预测准确度均较高。如表 5 所示，在 11 口地热井的水温预测中，随机森林模型的平均绝对误差为 0.021 6，XGBoost 模型的平均绝对误差为 0.013 0，2 种方法中 XGBoost 模型的准确度更高，计算时间更短，表现效果要优于随机森林模型。

以咸阳地热田为例地热水温度预测的 2 种场景中，采用的 LSTM 神经网络、随机森林模型和 XGBoost 模型 3 种机器学习方法均表现出准确度高、模型计算速度为秒级的优势。此外，这些方法无需大量的地质勘测资料，无需确定渗透系数等影响地热水温度的水文地质参数[23-24]，根据举出的 2 种应用场景，仅需要单井每日的流量观测数据或部分地热井的抽水试验数据，因此具有一定的迁移性和普适性。

### 3.2　模拟结果解释

对随机森林和 XGBoost 模型中的特征变量进行了基于 *SHAP* 值的重要性分析，结果如图 7 所示。分析结果表明，抽水试验资料的 16 个关键特征中 9 个具有较高的重要性，其中影响地热水温度的首要因素是取水段顶深。从水文地质学角度分析，完钻井深和取水段中深是第二、第四重要的影响特征，因为井水温度直接与钻井深度和取水段位置有关[39]。此外，重要性排序第三的特征纬度也被证明是咸阳地区地热水温的驱动因素，因为它关系着地热井到呈 EW 走向的渭北断裂带[26] 的距离，进而影响地热井水温度。而井口海拔的平均 *SHAP* 值只有 0.007 9，表明它几乎不影响井水温度。

进一步分析纬度与中落程流量特征的交互效应对井水温度的影响作用，如图 8，中落程流量低的井样本(图中蓝色点集中区)，相应的纬度较低(归一化

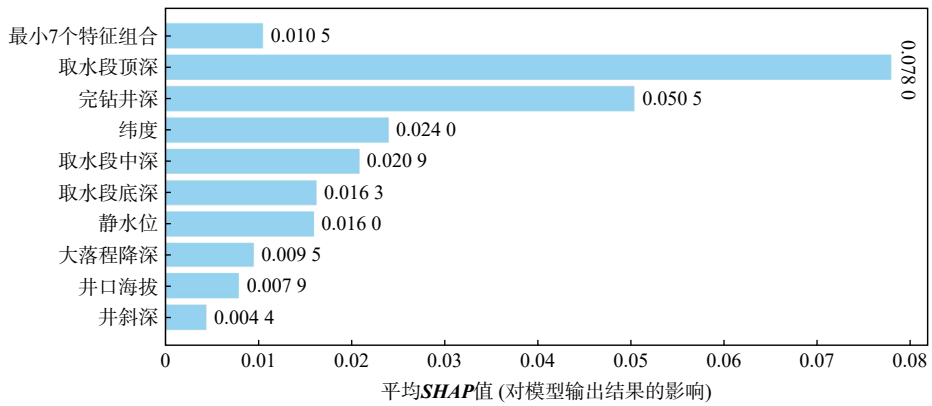
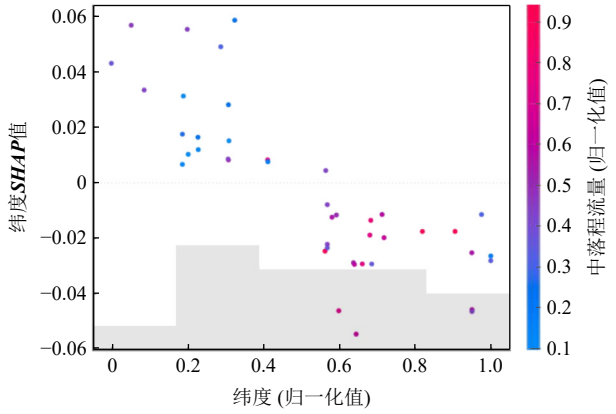



图 7　根据特征重要性对水温影响因素排序

Fig. 7　Rank influencing factors of groundwater temperature according to feature importance




灰色区域表示落在该纬度范围内的井样本频数统计，即灰色柱状高的部分表示该纬度范围内的井样本多

图 8　纬度与中落程流量特征交互 *SHAP* 依赖图

Fig. 8　Feature dependence *SHAP* plot of the interaction between latitude and medium drawdown flow

值小于 0.5)，其 *SHAP* 值一般大于零，表明较低的中落程流量和较低的纬度，对水温有正面的影响(正相关)。此外，图 8 中的点还呈现出随纬度升高 *SHAP* 值下降的分布规律，纬度越低的井其对应的 *SHAP* 值越大，说明纬度对地热水温度的影响越大。

根据机器学习模型进行特征重要性分析，利用 SHAP 方法进行模型解释，可以通过数据分析将机器学习模型预测结果与相关水文地质和地热成因机制联系起来[40]，明确了地热开发取水段位置的选择和生产井位置对地热水温的重要意义，这也是该模型方法独特的优势[41]。

## 4　结　论

(1)LSTM 神经网络模型能较为准确地模拟多变量特征下地热生产井水温的时间序列，训练集和测试集准确度均在 95% 以上，基于该方法，可以预测特定采灌状况下未来的地热井水温度变化，有利于热突破分析计算。

(2)选择经度、纬度、井口海拔、完钻井深、中落程流量、静水位等 16 个特征模拟计算抽水试验的中落程水温时，随机森林模型和 XGBoost 模型的预测准确度均在 95% 以上，模型计算速度快，XGBoost 模型的表现效果更好。

(3)特征重要性分析表明，取水段顶深是影响地热水温的首要影响因素，此外，完钻井深和纬度对咸阳地区地热水温起着重要作用。

(4)尽管受样本数据量限制，选取合适的机器学习模型能够满足地热井取水段温度的预测准确度和影响因素分析，为地热资源开发和钻井提供数据参考。

所有作者声明不存在利益冲突。

The authors declare that no competing interests exist.

**参考文献：**

[1] 汪集旸，胡圣标，庞忠和，等. 中国大陆干热岩地热资源潜力评估 [J]. 科技导报，2012，30(32)：25-31.
WANG J Y, HU S B, PANG Z H, et al. Estimate of geothermal resources potential for hot dry rock in the continental area of China[J]. Science & Technology Review, 2012, 30(32): 25-31. (in Chinese with English abstract)

[2] 武选民，柏琴，苑惠明，等. 冰岛地热资源开发利用现状 [J]. 水文地质工程地质，2007，34(5)：129-130.
WU X M, BAI Q, YUAN H M, et al. Development and utilization of geothermal resources in Iceland[J]. Hydrogeology & Engineering Geology, 2007, 34(5): 129-130. (in Chinese with English abstract)

[3] 姜光政，高堋，饶松，等. 中国大陆地区大地热流数据汇编(第四版)[J]. 地球物理学报，2016，59(8)：2892-2910.
JIANG G Z, GAO P, RAO S, et al. Compilation of heat flow data in the continental area of China(4th edition)[J]. Chinese Journal of Geophysics, 2016, 59(8): 2892-2910. (in Chinese with English abstract)

[4] 陈墨香，汪集旸. 中国地热研究的回顾和展望 [J]. 地球物理学报，1994，37(增刊 1)：320-338.
CHEN M X, WANG J Y. Review and prospect on geothermal studies in China[J]. Chinese Journal of Geophysics, 1994, 37(S1): 320-338. (in Chinese with English abstract)

[5] 文冬光，张二勇，王贵玲，等. 干热岩勘查开发进展及展望 [J]. 水文地质工程地质，2023，50(4)：1-13.
WEN D G, ZHANG E Y, WANG G L, et al. Progress and prospect of hot dry rock exploration and development[J]. Hydrogeology & Engineering Geology, 2023, 50(4): 1-13. (in Chinese with English abstract)

[6] GUO Q H, WANG Y X, LIU W. Hydrogeochemistry and environmental impact of geothermal waters from Yangyi of Tibet, China[J]. Journal of Volcanology and Geothermal Research, 2009, 180(1): 9-20.

[7] GONG B, LIANG H B, XIN S L, et al. Numerical studies on power generation from co-produced geothermal resources in oil fields and change in reservoir temperature[J]. Renewable Energy, 2013, 50: 722-731.

[8] 邹才能，杨智，朱如凯，等. 中国非常规油气勘探开发与理论技术进展 [J]. 地质学报，2015，89(6)：979-1007.
ZOU C N, YANG Z, ZHU R K, et al. Progress in China's unconventional oil & gas exploration and development and theoretical technologies[J]. Acta Geologica Sinica, 2015, 89(6): 979-1007. (in Chinese with English abstract)

[9] 张凯，赵兴刚，张黎明，等. 智能油田开发中的大数据及智能优化理论和方法研究现状及展望 [J]. 中国石油大学学报(自然科学

版), 2020, 44(4): 28-38.
ZHANG K, ZHAO X G, ZHANG L M, et al. Current status and prospect for the research and application of big data and intelligent optimization methods in oilfield development[J]. Journal of China University of Petroleum (Edition of Natural Science), 2020, 44(4): 28-38. (in Chinese with English abstract)

[10] 钟仪华, 王淑宁, 罗兰, 等. 用深度学习挖掘油田开发指标预测模型的知识 [J]. 西南石油大学学报(自然科学版), 2020, 42(6): 63-74.
ZHONG Y H, WANG S N, LUO L, et al. Knowledge mining for oilfield development index prediction model using deep learning[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2020, 42(6): 63-74. (in Chinese with English abstract)

[11] 王洪亮, 穆龙新, 时付更, 等. 基于循环神经网络的油田特高含水期产量预测方法 [J]. 石油勘探与开发, 2020, 47(5): 1009-1015.
WANG H L, MU L X, SHI F G, et al. Production prediction at ultra-high water cut stage via recurrent neural network[J]. Petroleum Exploration and Development, 2020, 47(5): 1009-1015. (in Chinese with English abstract)

[12] 谷建伟, 周梅, 李志涛, 等. 基于数据挖掘的长短期记忆网络模型油井产量预测方法 [J]. 特种油气藏, 2019, 26(2): 77-81.
GU J W, ZHOU M, LI Z T, et al. Oil well production forecast with long-short term memory network model based on data mining[J]. Special Oil & Gas Reservoirs, 2019, 26(2): 77-81. (in Chinese with English abstract)

[13] 刘想平, 汪崎生, 汤军. 用神经网络建立自喷井井底流压预测模型 [J]. 石油勘探与开发, 1997, 24(5): 92-94.
LIU X P, WANG Q S, TANG J. An application of neural network in developing a model for predicting flowing bottomhole pressure of flowing wells[J]. Petroleum Exploration and Development, 1997, 24(5): 92-94. (in Chinese with English abstract)

[14] 张进平, 王煜曦, 刘桂宏, 等. 通州区地热资源优化开采模式动态研究 [J]. 工程力学, 2022, 39(6): 247-256.
ZHANG J P, WANG Y X, LIU G H, et al. Dynamic study on optimal exploitation mode of geothermal resources in Beijing Tongzhou District[J]. Engineering Mechanics, 2022, 39(6): 247-256. (in Chinese with English abstract)

[15] 段晓飞, 康凤新, 吴晓华, 等. 基于采灌均衡模拟的砂岩热储合理采灌井距计算方法 [J]. 地质科技通报, 2024, 43(5): 170-180.
DUAN X F, KANG F X, WU X H, et al. A methodology for determining the optimal well spacing in sandstone geo-thermal reservoirs through production-reinjection equilibrium simulation[J]. Bulletin of Geological Science and Technology, 2024, 43(5): 170-180. (in Chinese with English abstract)

[16] 马峰, 高俊, 王贵玲, 等. 雄安新区容城地热田碳酸盐岩热储采灌数值模拟 [J]. 吉林大学学报(地球科学版), 2023, 53(5): 1534-1548.
MA F, GAO J, WANG G L, et al. Numerical simulation of exploitation and reinjection of carbonate geothermal reservoir in Rongcheng geothermal field, Xiong'an New Area[J]. Journal of Jilin University (Earth Science Edition), 2023, 53(5): 1534-1548. (in Chinese with English abstract)

[17] 赵敬波, 刘健, 周志超, 等. 基于离散裂隙网络模型的地下水流并行模拟方法 [J]. 地质科技通报, 2023, 42(4): 55-64.
ZHAO J B, LIU J, ZHOU Z C, et al. Parallel groundwater flow simulation method based on a discrete fracture network model[J]. Bulletin of Geological Science and Technology, 2023, 42(4): 55-64. (in Chinese with English abstract)

[18] XU T F, YUAN Y L, JIA X F, et al. Prospects of power generation from an enhanced geothermal system by water circulation through two horizontal wells: A case study in the Gonghe Basin, Qinghai Province, China[J]. Energy, 2018, 148: 196-207.

[19] 张娟娟, 林建旺, 林黎, 等. 基于 AHP 和 BP 神经网络的深部地热水可持续开发能力评价 [J]. 地下水, 2008, 30(6): 46-48.
ZHANG J J, LIN J W, LIN L, et al. Evaluation of deep geothermal water sustainable development ability based on AHP and BP neural network[J]. Ground Water, 2008, 30(6): 46-48. (in Chinese with English abstract)

[20] 董华松, 黄文辉. 利用遗传算法优化的小波神经网络实现地热资源预测 [J]. 资源与产业, 2014, 16(3): 101-106.
DONG H S, HUANG W H. Prediction of geothermal resources by means of wavelet neural network optimized by genetic algorithm[J]. Resources & Industries, 2014, 16(3): 101-106. (in Chinese with English abstract)

[21] 赵磊, 张海雄, 贺婷婷. 基于 LSTM 神经网络的地热井回灌压力预测方法研究 [J]. 当代化工研究, 2023(4): 119-123.
ZHAO L, ZHANG H X, HE T T. Reinjection pressure prediction of geothermal well with LSTM neural network model[J]. Modern Chemical Research, 2023(4): 119-123. (in Chinese with English abstract)

[22] REN Y Q, KONG Y L, PANG Z H, et al. A comprehensive review of tracer tests in enhanced geothermal systems[J]. Renewable and Sustainable Energy Reviews, 2023, 182: 113393.

[23] WU H, FU P C, MORRIS J P, et al. Characterization of flow and transport in a fracture network at the EGS Collab field experiment through stochastic modeling of tracer recovery[J]. Journal of Hydrology, 2021, 593: 125888.

[24] HAWKINS A J, FOX D B, KOCH D L, et al. Predictive inverse model for advective heat transfer in a short-circuited fracture: Dimensional analysis, machine learning, and field demonstration[J]. Water Resources Research, 2020, 56(11): e2020WR027065.

[25] JIANG G Z, HU S B, SHI Y Z, et al. Terrestrial heat flow of continental China: Updated dataset and tectonic implications[J]. Tectonophysics, 2019, 753: 36-48.

[26] 罗璐, 朱霞, 何春艳, 等. 陕西咸阳地热田地热流体成因研究 [J]. 地质论评, 2019, 65(6): 1422-1430.
LUO L, ZHU X, HE C Y, et al. Study on the genesis of geothermal fluid in Xianyang geothermal field[J]. Geological Review, 2019, 65(6): 1422-1430. (in Chinese with English abstract)

[27] 杨剑锋, 乔佩蕊, 李永梅, 等. 机器学习分类问题及算法研究综述 [J]. 统计与决策, 2019, 35(6): 36-40.
YANG J F, QIAO P R, LI Y M, et al. A review of machine-learning classification and algorithms[J]. Statistics & Decision, 2019, 35(6): 36-40. (in Chinese with English abstract)

[28] BI Y X, GUAN J W, BELL D. The combination of multiple classi-

fiers using an evidential reasoning approach[J]. Artificial Intelligence, 2008, 172(15): 1731-1751.

[29] CHEN T Q, LI H, YANG Q, et al. General functional matrix factorization using gradient boosting[C]//Anon. Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia, USA: [s. n. ], 2013: 436-444.

[30] 尹宝才, 王文通, 王立春. 深度学习研究综述 [J]. 北京工业大学学报, 2015, 41(1): 48-59.
YIN B C, WANG W T, WANG L C. Review of deep learning[J]. Journal of Beijing University of Technology, 2015, 41(1): 48-59. (in Chinese with English abstract)

[31] GRAVES A. Supervised sequence labelling with recurrent neural networks[M]. Berlin: Springer, 2012.

[32] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[J]. Neural Information Processing Systems, 2017: 21889700.

[33] GUIDOTTI R, MONREALE A, RUGGIERI S, et al. A survey of methods for explaining black box models[J]. ACM Computing Surveys, 2019, 51(5): 1-42.

[34] 中华人民共和国国家质量监督检验检疫总局, 中华人民共和国建设部. 供水水文地质勘察规范: GB50027-2001[S]. 北京: 中国标准出版社, 2001.
General Administration of Quality Supervision Inspection and Quarantine of the People's Republic of China, Ministry of Construction of the People's Republic of China. Standard for hydrogeological investigation of water-supply: GB50027-2001[S]. Beijing: China Standards Press, 2001. (in Chinese)

[35] 赵志宏, 刘桂宏, 王佳铖, 等. 城市深层地热能可持续开采多场耦合效应数值模拟研究进展 [J]. 煤炭学报, 2023, 48(3): 1126-1138.
ZHAO Z H, LIU G H, WANG J C, et al. Coupled multi-field effect on sustainable development of deep geothermal energy in cities[J]. Journal of China Coal Society, 2023, 48(3): 1126-1138. (in Chinese with English abstract)

[36] HUO A D, WANG X F, LYU Y, et al. Simulation research on the reinjection temperature fields of deep geothermal wells based on real-scale experiment[J]. Energy Exploration & Exploitation, 2019, 37(2): 646-662.

[37] 陶晔, 杜景林. 基于随机森林的长短期记忆网络气温预测 [J]. 计算机工程与设计, 2019, 40(3): 737-743.
TAO Y, DU J L. Temperature prediction using long short term memory network based on random forest[J]. Computer Engineering and Design, 2019, 40(3): 737-743. (in Chinese with English abstract)

[38] 柯婷婷, 黄少鹏, 许威, 等. 关中盆地沣西地区地热对井采灌开发模式的数值模拟 [J]. 第四纪研究, 2019, 39(5): 1252-1263.
KE T T, HUANG S P, XU W, et al. Numerical modeling of doublet well system for extracting heat from sandstone geothermal reservoir: A case study of Fengxi area, the Guanzhong Basin, NW China[J]. Quaternary Sciences, 2019, 39(5): 1252-1263. (in Chinese with English abstract)

[39] 高俊义. 地热温度和生产井深度对岩体水流-传热温度影响机理研究 [J]. 地球物理学进展, 2020, 35(5): 1659-1664.
GAO J Y. Study on mechanism of the influence of geothermal temperature and production well depth on water flow and heat transfer temperature in rock mass[J]. Progress in Geophysics, 2020, 35(5): 1659-1664. (in Chinese with English abstract)

[40] 杨国华, 李婉露, 孟博. 基于机器学习方法的地下水氨氮时空分布规律 [J]. 吉林大学学报(地球科学版), 2022, 52(6): 1982-1995.
YANG G H, LI W L, MENG B. Spatiotemporal distribution of groundwater ammonia nitrogen based on machine learning methods[J]. Journal of Jilin University (Earth Science Edition), 2022, 52(6): 1982-1995. (in Chinese with English abstract)

[41] 肖立志. 机器学习数据驱动与机理模型融合及可解释性问题 [J]. 石油物探, 2022, 61(2): 205-212.
XIAO L Z. The fusion of data-driven machine learning with mechanism models and interpretability issues[J]. Geophysical Prospecting for Petroleum, 2022, 61(2): 205-212. (in Chinese with English abstract)

Summary for “机器学习模型在地热开发水温预测中的应用”

# Application of machine learning models for groundwater temperature prediction in geothermal development

DONG Peiyao[1,2], DU Li[1,2], ZHAO Lei[1,2], BAO Yifan[1,2], YIN Maosheng[3*]

1. State Key Laboratory of Deep Geothermal Resources, Beijing 100083, China;
2. SINOPEC Star (Beijing) New Energy Research Institute Co., Ltd., Beijing 100083, China;
3. School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen Guangdong 518055, China

First author, E-mail: dongpeiyao.xxsy@sinopec.com

* Corresponding author, E-mail: myin@eitech.edu.cn

**Abstract**: Geothermal energy, as a clean energy source, has broad application prospects. The temperature assessment of geothermal water is a key topic in the sustainable development and utilization of geothermal resources. [**Objective**] Artificial intelligence (AI) technology has become a hotspot and frontier direction in the exploration and development of mineral, oil, and gas fields. However, studies on its application in geothermal field development are limited. This paper first discusses the significant value of big data and AI in oil and gas field development, and then reviews the current applications of AI in geothermal field development. [**Methods**] Taking the Xianyang geothermal field in Shaanxi Province as a case study, a time series model for the temperature of geothermal water in a single well was constructed using a long short-term memory (LSTM) neural network under the predetermined production mode. Additionally, the random forest and XGBoost algorithms were used to predict the groundwater temperature of multiple geothermal wells. [**Results**] The accuracy of the three models was above 95%, and their running speed was fast. The depth of the water intake section's top was found to be the primary influencing factor on geothermal water temperature in the area. The model also confirmed that WeiBei fault zones play an important role in heat storage. [**Conclusion**] The application of these models demonstrates the superiority of machine learning in addressing complex problems in geothermal field development. The reasonable application of AI can provide a more effective decision-making basis for the efficient development of geothermal fields, as well as for scientific cost reduction, quality improvement, and efficiency enhancement.

**Key words:** geothermal development; machine learning; modeling; groundwater; groundwater temperature prediction

Received: 2024-02-23; Revised: 2024-06-24; Accepted: 2024-12-17

doi: 10.19509/j.cnki.dzkq.tb20240063