Proceedings World Geothermal Congress 2026, Calgary, Canada, June 8—11, 2026

Application of machine learning models in geothermal resource development - Taking groundwater temperature prediction as an example

Peiyao Dong
Sinopec Star Petroleum Co., Ltd, Beijing 100083, China
dongpeiyao.xxsy@sinopec.com

INTRODUCTION

Geothermal resources, as a form of renewable energy, have significant potential for sustainable development. Nevertheless, China's attention to the exploration and utilization of geothermal resources has been relatively low, with insufficient investment, resulting in a relative scarcity of basic data on geothermal resources and insufficient detail in existing information on resource distribution. Moreover, considering that the development of geothermal resources in China began in the early 1970s, the history of development is relatively short and is constrained by geographic and geological conditions, leading to limitations in current technological methods and basic data. Oil well-associated geothermal resources are abundant, and geothermal development strategies can draw on traditional oil and gas development models. It is worth noting that with the rapid advancement of big data and artificial intelligence technologies, these technologies have shown broad application prospects in addressing complex nonlinear problems. At present, machine learning technologies have already begun to be explored in fields such as oil and gas resource development and production forecasting, indicating the potential for innovative progress in the development of geothermal resources.

METHODS

In the current field of machine learning, ensemble learning algorithms have become a mainstream approach for solving problems such as classification and regression. The basic idea of this method is to complete a specific learning task by combining multiple learners, a strategy that generally achieves better performance than a single learner. Ensemble learning algorithms can be roughly divided into three types: Bagging, Boosting, and Stacking. Among them, this paper primarily applies the first two methods.

The LSTM neural network is a special variant of the RNN, proposed by Hochreiter & Schmidhuber in 1997 and recently improved and popularized by Alex Graves . LSTM units can selectively add or forget information through forget gates, input gates, and output gates, with each LSTM unit's input at a given moment including the previous moment's cell state, hidden state, and the current input. The LSTM was designed to address the long-term dependency problem and has successfully overcome the shortcomings of traditional RNNs, becoming the most popular form of RNN today.

RESULTS

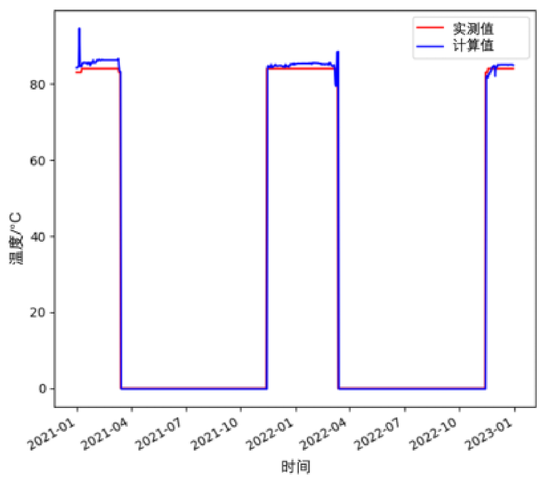


Figure 1. Comparison of groundwater temperature simulation results and observation results in well YGSY1

Table 1. Table 1 LSTM model parameter optimization results

Parameter	Value	Result
seq_length	5, 10, 20, 40	5
num_layers	1, 2	1
hidden-size	128, 256	128
learning rate	0.01, 0.005, 0.001	0.001
Epochs	50, 100, 200	200

A single-well temperature time series model was established using data from a group of geothermal injection wells in Xianyang, Shaanxi (YGSY1 and YGSY2). YGSY2 serves as the reinjection well for the production well YGSY1. Daily variation data of the pumping flow rate of YGSY1, and the injection water temperature and flow rate of YGSY2, were selected as multiple input features for the model, with the water temperature of YGSY1 as the target. The data were split using a sequential partitioning method, with observation data from 2020 used for model training. To ensure the continuity of the time series, it was assumed that the water temperature was 0 during the non-heating season.

A multivariate time series model was established using the LSTM neural network algorithm to simulate the temporal variation of the water temperature of YGSY1. Considering various evaluation indicators, the model parameters were optimized and adjusted using grid search and cross-validation; the main parameters optimized included time step, network layers, hidden layer neurons, learning rate, and number of iterations. The preferred results for each model parameter are shown in Table 1.

Since the test water data from multiple wells in the area was used, this simulates the problem of the spatial temperature distribution across multiple wells, without considering temperature changes over time. The XGBoost model was used to build the prediction model (Figure 2). Based on various evaluation metrics, grid search and cross-validation methods were adopted to optimize and tune the model parameters.

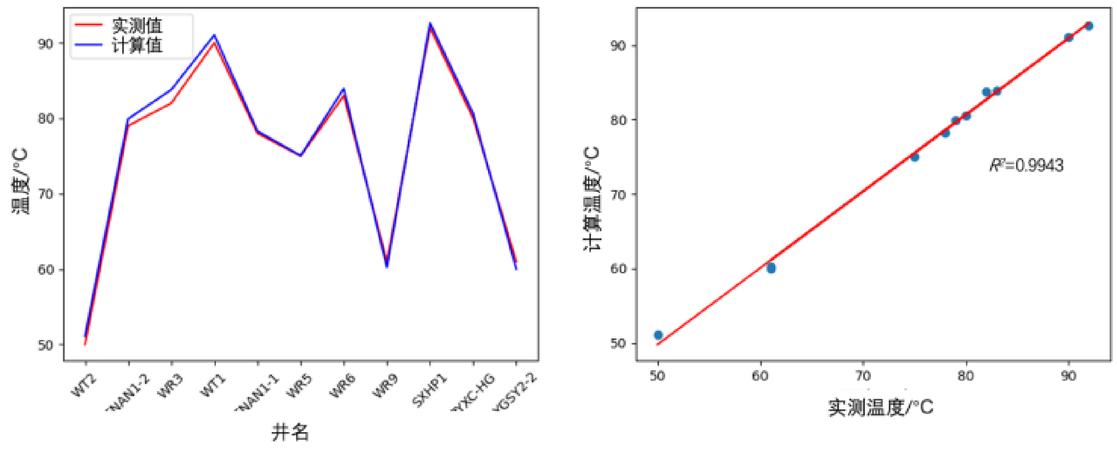


Figure 2 Comparison between the simulation results of geothermal well groundwater temperature and the observation results (XGBoost Model)

DISCUSSION

The analysis results indicate that 9 out of the 18 key features have high importance, with the primary factor being the temperature at the start of water production with a large drawdown. Drilling depth and the depth of the water intake section are the second and third most important influencing features. This finding aligns with the basic understanding of hydrogeology, as well water temperature is directly related to well depth and the location of water extraction. The average SHAP value of the static water level is only 0.0006, indicating that it has almost no effect on well water temperature. In addition, latitude has also been shown to be a driving factor of geothermal water temperature in the Xianyang area, as it relates to the distance from production wells to the east-west Weibei Fault Zone, thereby affecting the temperature of geothermal well water.

CONCLUSION

(1) The LSTM neural network model can accurately simulate the time series of geothermal production well water temperature under multivariate features, with both training and testing set accuracies exceeding 95%. Based on this, it is possible to predict future water temperature changes under this production and injection condition.

(2) When 18 features such as longitude, latitude, elevation, total drilling depth, flow rate, and static water level are selected to simulate and calculate the mid-depth water temperature during test water operations, both the random forest model and the XGBoost model achieve prediction accuracies above 95%, with the XGBoost model performing better.

(3) Feature importance analysis indicates that the test water mid-depth temperature is the primary factor affecting the mid-depth water output temperature. Additionally, total drilling depth and production interval depth play important roles in the water output temperature, while the output temperature is almost unrelated to the static water level. In summary, although limited by the sample data size, selecting an appropriate machine learning model can meet the prediction accuracy requirements for geothermal well production interval temperature, providing data reference for geothermal resource development and drilling.

REFERENCES

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.
GRAVES A. Supervised Sequence Labelling with Recurrent Neural Networks[M]. Heidelberg: Springer Berlin Press, 2012, 385.
LUNDBERG S M, LEE S I. A Unified Approach to Interpreting Model Predictions[C]. Advances in Neural Information Processing Systems Long Beach, CA: 31st Annual Conference on Neural Information Process Systems (NIPS), 2017.
GUIDOTTI R, MONREALE A, RUGGIERI S, et al. A Survey of Methods for Explaining Black Box Models[J]. ACM Computing Surveys (CSUR), 2018
ZHAO L, ZHANG H X, HE T T. Reinjection Pressure Prediction of Geothermal Well with LSTM Neural Network Model[J]. Modern Chemical Research,2023, 04: 119-123 (in Chinese with English abstract).
JIANG G, HU S, SHI Y, et al. Terrestrial heat flow of continental China: Updated dataset and tectonic implications[J]. Tectonophysics, 2019, 753: 36-48.
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.
WU H, FU P C, JOSEPH P M, 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.
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.