Minimum Miscibility Pressure Prediction Method Based On PSO-GBDT Model

Yawen He, **Weirong Li***, **Shihao Qian**, Xi'an Shiyou University, Xi'an, China

Abstract

With the development of EOR technology, CO_2 flooding was a very promising method to improve the recovery of conventional and unconventional oil reservoirs. MMP (minimum miscibility pressure) was one of the important parameters of the CO_2 flooding, and the use of an artificial intelligence algorithm can accurately predict the MMP, which was important to evaluate the effect of CO_2 flooding development in the reservoir.

This work presents methods to automatically find optimal parameter settings for machine learning model by using an evolutionary algorithm. In this paper, 195 sets of experimental data of MMP were collected and screened from a large amount of literature for model establish, and sensitivity analysis was performed with the Pearson's method for feature selection. Then, five machine learning algorithms were used for regression and comparison. Finally, a particle swarm optimization algorithm was used to optimize the parameters of the machine learning model with best performance. The accuracy of the training set obtained by the hybrid model was 99.9% and the accuracy of the test set was 97.6%. It indicated that the hybrid model are valid and accurate, and it can be used for MMP prediction in both laboratory and actual field.

Introduction

CO_2 flooding is considered to be one of the most effective EOR methods, particularly in developing light oil reservoirs. Depending on the injection pressure, there are three classification of CO_2 flooding, including miscible gas injection, partial miscible gas injection, and immiscible gas injection. The displacement efficiency of reservoir oil by CO_2 flooding is highly pressure dependent and miscible displacement is only achieved at pressures greater than a certain minimum pressure, termed the minimum miscibility pressure (MMP). The MMP is defined as the lowest pressure for which a given injected gas composition can develop miscibility through a multi-contact process with a specific reservoir at reservoir temperature. MMP is completely independent of reservoir heterogeneity, but it is a strong function of oil composition, composition of the injected gas, and reservoir temperature.

The methods used to predict the MMP under gas injection, miscible gas injection, or a dry gas cycling scheme, include lab tests, empirical correlations, equation of state (EOS) methods, and data-driven approaches. Specific lab test used to determine MMP include the swelling test, slim-tube test, rising bubble test, core flood, and other tests. Yellig and Metcalfe (1980) first proposed the experimental method to determine the CO_2-crude oil MMP through a thin tube. In the experiments, CO_2 was injected at a specific rate into a thin tube at different pressures to obtain a recovery-pressure relationship curve to determine the MMP. The rising bubble apparatus method (RBA) was proposed to determine MMP, in which an oil sample of a certain height is injected into a vertically placed visible high-pressure glass tube, followed by the injection of gas from the bottom of the glass tube at a constant rate, and MMP was determined from the shape of the bubbles and the distance they travel. Harmon and Grigg (1988) proposed a experiment to directly measure the relationship between the density and pressure of the injected rich gas phase, the MMP of the gas and crude oil miscible phase was determined by using the gas-oil dissolution characteristics.

DOI: 10.14800/IOGR.1219

Received July 08, 2022; revised August 12, 2022; accepted September 29, 2022.

*Corresponding author: weirong.li@xsyu.edu.cn

The empirical correlations are a simple and fast means to determine MMP. Various widely used correlations are outlined as follows. Orr and Silva (1987) proposed an empirical formula to determine the MMP for pure CO_2 and impure CO_2/crude oil system. The EVP correlation presented by Orr and Jensen (1986) can determine the MMP for low-temperature reservoirs (T<120°F). Yellig and Metcalfe (1980) proposed a correlation to forecast the CO_2 MMP with the temperature as the only correlating parameter from their experimental study. Alston et al. (1985) derived an empirical correlation to estimate the MMP for pure or impure CO_2/crude oil systems. Cronquist (1978) proposed an empirical formula to characterizes MMP as a function of reservoir temperature, molecular weight of the oil pentanes-plus fraction, and mole percentage of methane and nitrogen.

The EOS methods were established based on the theory of phase equilibrium of the system and was mainly used to map the relationship between the phase behavior of the CO_2 and crude oil system and the miscible function, and thus to obtain MMP for the system (Yellig and Metcalfe 1980; Holm 1987; Mungan 1981; Cronquist 1978). The existing EOS used to calculate MMP were the Peng-Robinson EOS (PR-EOS) (Silva and Orr 1987; Orr and Silva 1987; Alston et al. 1985), Nasrifar-Moshfeghian EOS (NM-EOS) (Nasrifar and Moshfeghian 2001), and the improved statistical fluid theory EOS (SSAFT-EOS) (Zhao et al. 2006). The EOS method mainly uses phase simulation technique to investigate the influence of injected gas on the properties of crude oil. The parameters of EOS were tuned by fitting the PVT experimental data to establish a phase model that conforms to the real fluid, and MMP of the oil and gas system is calculated by simulating the multi-stage contact experiment process (Al-Ajmim et al. 2011). The EOS method can establish a phase model that fits the flow state characteristics of oil and gas multiphase fluids, so it can accurately characterize the phase behavior of real reservoir fluids and estimate the physical parameters.

The lab measurements have a high accuracy, but it is very time consuming and expensive. The empirical correlations obtained under specific experimental conditions are often limited by failure to satisfy requirements. EOS methods usually require a large amount of preliminary experimental data to fit the PVT parameters.

Recently, artificial intelligence methods have become essential techniques for MMP prediction. They learn high-level features of PVT data using structures composed of several nonlinear transformations to classify large and complex data. The significant advantages of artificial intelligence methods are the high prediction accuracy and the ability to process large amounts of data in parallel. With the development of artificial intelligence techniques, researchers can build and apply the required models in different fields. **Table 1** shows the research related to machine learning algorithms for MMP prediction.

Table 1--Related works on machine learning based-MMP prediction model.

Ref.	Data	Model	Optimization Algorithm	Output	Input Parameters
Mousavi et al. 2008	44	LSSVM		Pure and impure CO_2 MMP	5
Chen et al. 2014	85	ANN	GA	Pure and impure CO_2 MMP	10
Choubineh et al. 2019	251	ANN		MMP of different injected gas and crude oil	
Dong et al. 2019		ANN	L2 Regularization、Dropout	MMP	
Saeedi and Soleimani 2020	144	ANN	TLBO、PSO	Pure and impure CO_2 MMP	8
Tian et al. 2020	152	BP	ABC、DA	Pure and impure CO_2 MMP	5
Zhang et al. 2020	170	BP、RF、SVM		Impure CO_2 MMP	6

(LSSVM: Least Squares Support Vector Machines; BP: Back Propagation; RF: Random Forest; SVM: support vector machines; TLBO: Teaching-learning-based optimization; ABC: Artificial Bee Colony; DA: Dragonfly algorithm)

The motivation of this work is to use an optimization algorithm to automatically learn an efficient architecture and best set of hyperparameters of machine learning algorithms without much human intervention. It is difficult for human experts to select the parameters of machine learning algorithms before applying it to solve any real-world problem. It usually demands human expertise and intensive

efforts to conduct experiments on many possible configurations of algorithms parameters to finalize one with relatively better performance. Researchers have developed hybrid versions of machine learning models for the optimization of parameter selection. Particle swarm optimization (PSO) is the most preferred selection of scholars to solve optimization problem as it has fewer hyperparameters, a simpler expression, and easy computation.

Based on previous studies, this paper introduces the machine learning algorithm to establish the MMP prediction model, and uses the PSO algorithm to optimize ML algorithm with the best performance to establish the MMP prediction model.

The rest of the paper is structured as follows: Section 2 explains the proposed PSO-based GBDT model architecture; Section 3 briefly presents data collected, analyzes the correlation between each parameter and MMP; Section 4 presents experimental results and discuss the performance in terms of accuracy; Chapter 5 concludes the research of this paper.

The Proposed Method

Gradient boosted decision trees (GBDT). GBDT was an ensemble of gradient boosting and decision trees. The algorithm classifies or regresses the data by using an additive model (a linear combination of basis functions) and by continuously reducing the residuals generated by the training process.

GBDT is calculated through multiple rounds of iterations, and each iteration generates a weak classifier. Each classifier is trained based on the gradient of the previous classifier (if the loss function is a squared loss function, the gradient is the residual value). The requirements for weak classifiers are generally simple enough and have low variances and high biases. In the training process reducing the deviation continuously improves the accuracy of the final classifier. The final total classifier is obtained by weighting and summing the weak classifiers obtained from each round of training.

The general steps of GBDT are shown in **Figure 1**. For each category, a regression tree is trained first, and the residuals are calculated for each category separately and repeated multiple calculations, and the final model is obtained. When predicting, the category with the highest probability is the corresponding category.

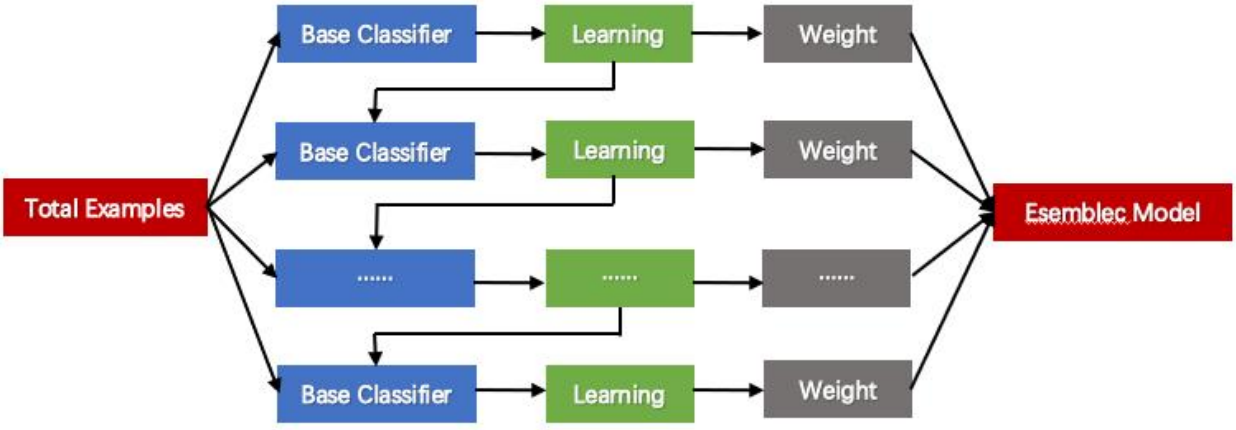

Figure 1—GBDT algorithm training process diagram.

Particle swarm optimization. PSO was an evolutionary computation technique from the study of bird predation behavior. It searches the optimal solution through collaboration and information sharing among individuals in the group.

PSO designed massless particle to simulate the birds in the birds' swarm. The particle has only two attributes: speed and position. Each particle individually searches for the optimal solution in the search space, which is recorded as the current individual extreme and shares the individual extreme with the other particles in the whole swarm. All the particles in the swarm adjust their velocity and position according to the current individual extreme they find and the current global optimal solution shared by the whole swarm. The process of PSO algorithm is illustrated in **Figure 2** as follows.

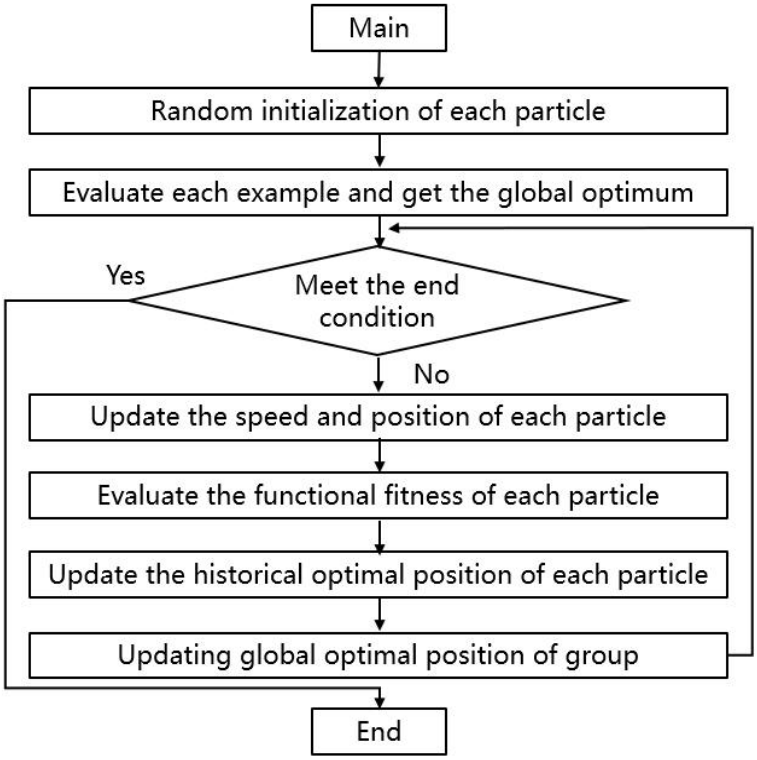

Figure 2—PSO algorithm flowchart.

Experiments

Data collection and pre-processing. To establish a highly reliable intelligent model, factors related to reservoir temperature (T) and relative molar fractions (mol%) of CO_2, N_2, CH_4, and C_2-Cn were collected as relevant parameters affecting the MMP.

This paper collects 395 data from a large number of literature (Lai et al. 2017; Cardenas et al. 1984; Eakin and Mitch 1988; Graue and Zana 1981; Kanatbayev et al. 2015; Spence and Watkins 1980; Harmon and Reid 1988; Thakur et al. 1984; Zhang et al. 2016; Al-Ajmi et al. 2009; Glasø 1985; Dicharry et al. 1973; Henry and Robert 1983; Khan et al. 1992) including laboratory measurement data and numerical simulation data. 195 sets of experimental data were obtained after sorting and screening for machine learning. As shown in **Figure 3**, the temperature, T , ranges mainly between 50 and 100°C; The molar fraction of CO_2 mainly concentrates between 0 and 25%; The molar fraction of N_2 is mainly from 0 to 10%; The component molar fraction of C_1 mainly concentrates between 0 and 60%; The component molar fraction of C_2~C_5 distributes between 0 and 10%; The component molar fraction of C_6 ranges from 0 to 5%; The molar fractions of C_{7+} ranges between 0 and 75%; MMP values mainly concentrates between 10 and 40 MPa.

Correlation analysis. Affected by the diversity of crude oil components, it is necessary to analyze the correlation between different components and MMP. The interaction between the various components also has a certain effect on MMP.

Temperature and molar percentages of each component were used as the influencing factors to predict MMP. Pearson's method was used to characterize the correlation between each component and MMP. As shown in **Figure 4**, the factors with the greatest influence on MMP was the reservoir temperature, followed by the molar fraction of the carbon component. The molar fractions of CO_2 and N_2 have little influence on MMP.

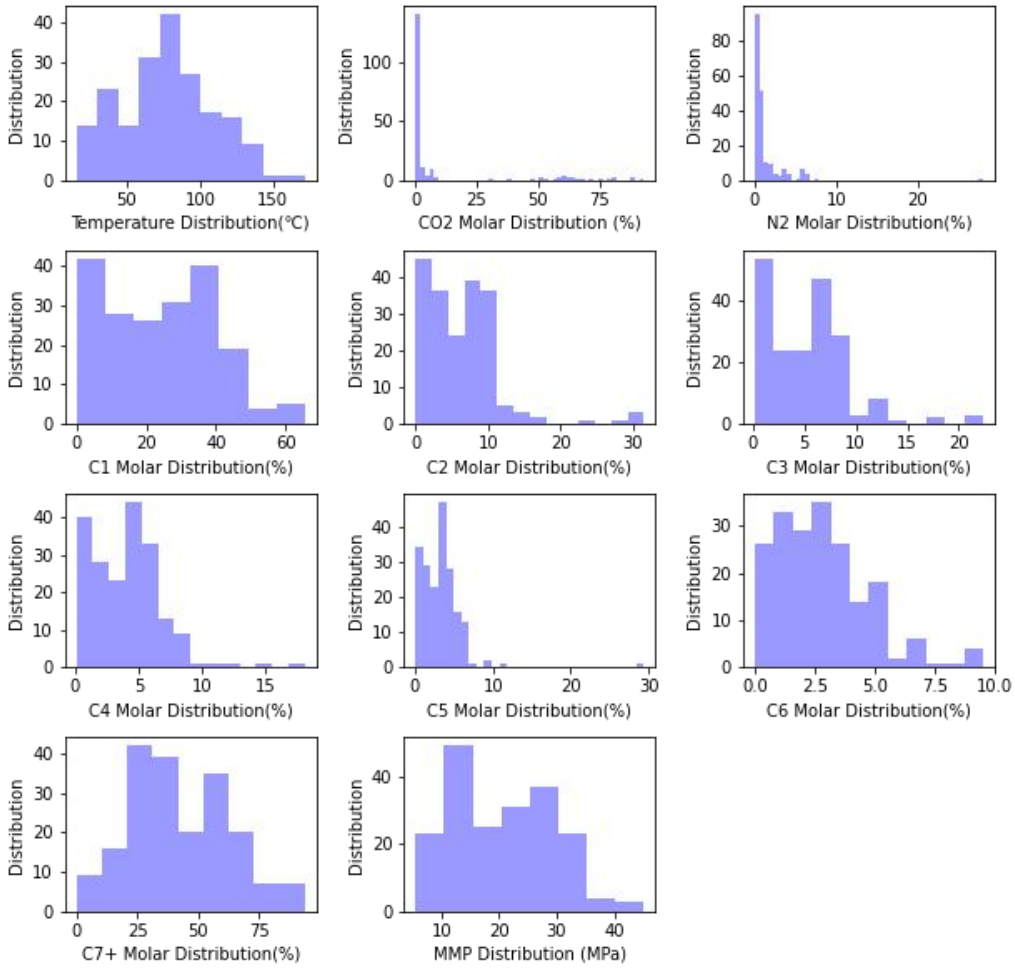

Figure 3—Input parameter distribution.

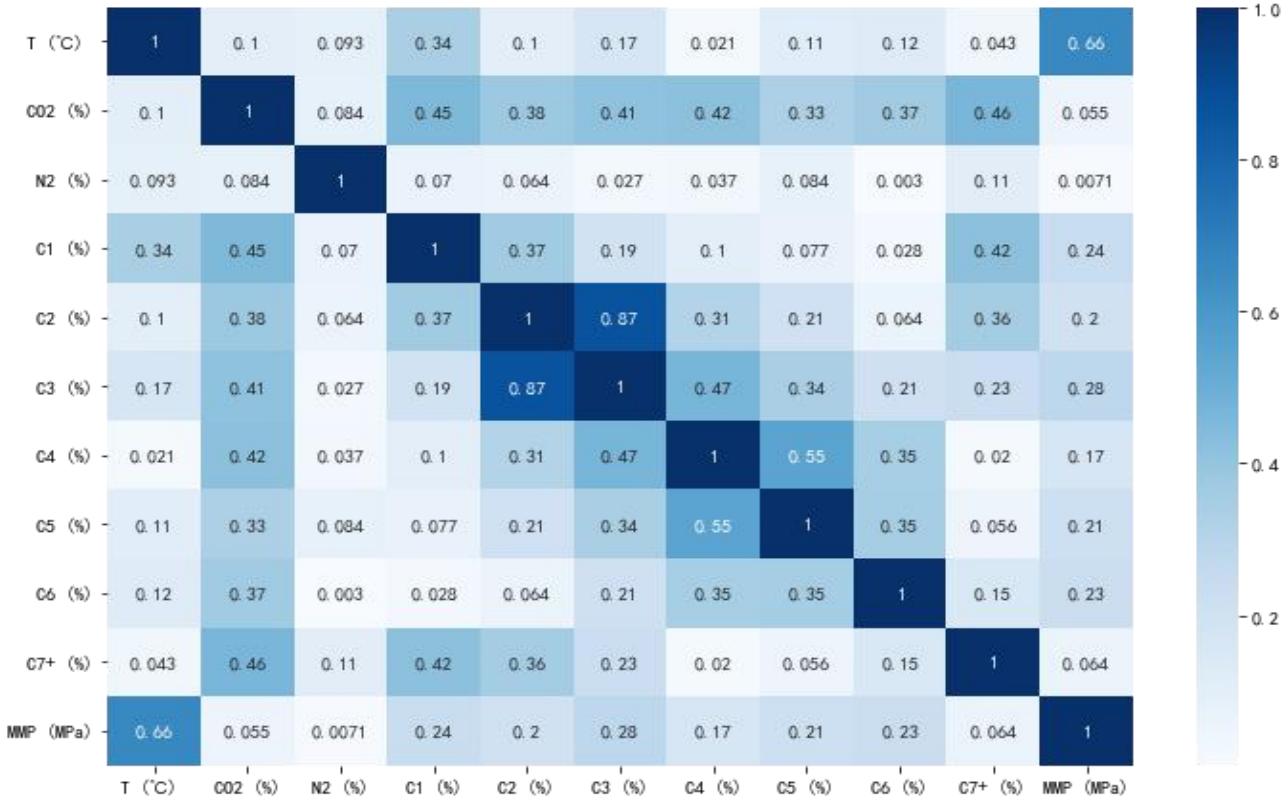

Figure 4—Schematic diagram of group correlation analysis.

Result and Discussion

The selection of proper hyperparameters of a GBDT model is very time-consuming process when it is to be decided on a trial-and-error basis. Thus, it is required to develop an automated approach which can reach to the best GBDT model with minimal human expertise and efforts. The proposed work uses the random but guided nature of PSO to find the best GBDT model. It optimizes its hyperparameters on a given dataset in predefined search-space.

After data pre-processing and correlation analysis, ten parameters, including *T*, the molar fraction of CO_2, N_2, C_1, C_2, C_3, C_4, C_5, C_6, and C_{7+} were used as inputs of the machine learning model, and the corresponding MMP was used as output for model training.

To ensure the coverage of the training and test sets as large as possible and without overlap, the entire database was randomly divided into two groups: the training set and the test set. The training set consists of 165 data points, and the test set with 30 data points was used for model validation. Five machine learning models are established for comparison, including LR, RR, RF, MLP, and GBDT.**Figure 5** shows the predicted results obtained by the five models. The horizontal coordinates in Figure 5 are the actual values and the vertical coordinates are the predicted MMP values. It shows that the GBDT model has the best performance, followed by random forest. The linear regression, ridge regression, and multilayer perceptron have poor performance.

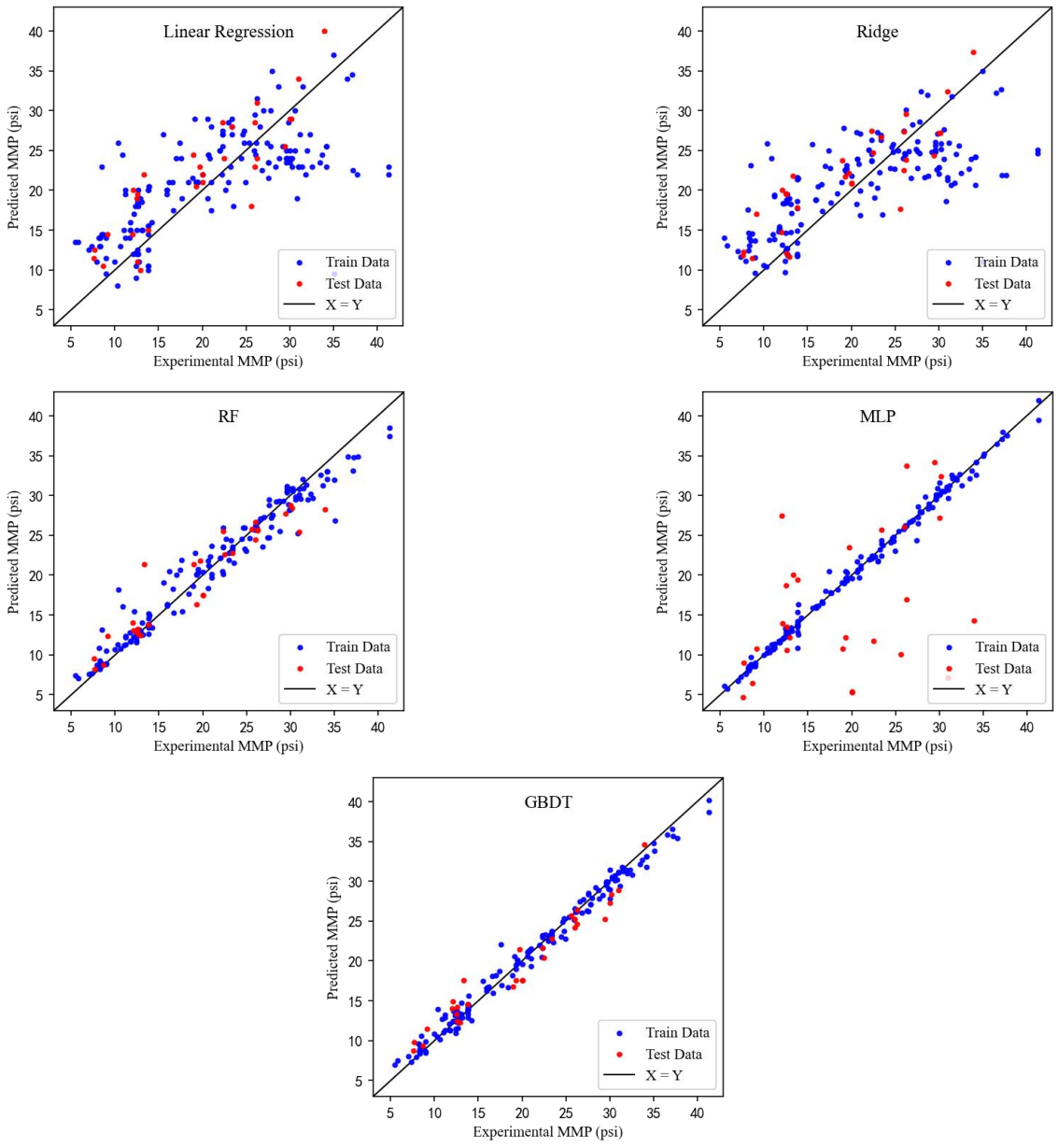

Figure 5—Performance comparison based on predicted and actual dataset.

Table 2 shows the accuracy of the five machine-learning models for MMP prediction. The result indicates that GBDT model has an accuracy of 98.5% and 93.7% in the training and test set, which outperforms the accuracy overwhelmingly to the other models.

Table 2—Current status of machine learning prediction MMP research.

Machine Learning Model	Accuracy	
	Train set	Test set
Linear regression	47.7%	67.0%
Ridge regression	47.8%	66.6%
Gradient boosting decision tree,	98.5%	93.7%
Random forest	95.1%	88.6%
MLP	99.3%	65.2%

Swarm optimization is performed by encoding hyperparameters of GBDT into particles. The fitness function represents the accuracy of the GBDT and has been passed on for generations. The fitness values are estimated over 100 generations to optimize MMP prediction. Four important parameters affecting the optimization of PSO were clarified. The PSO algorithm mainly optimizes four parameters: n-estimators, learning rate, max-depth, and alpha. The number of n-estimators is the number of decision trees, which is the amount of data evaluation. It has a monotonic effect on the accuracy of the model. The larger the n-estimators, the better the model. However, the accuracy of the model does not increase after the n-estimators reach a certain level. the optimal value of n-estimators is 413. And the value of the learning rate needs to be set within a certain range. High learning rate will lead to unstable learning. Too small learning rate increases the training time. The minimum miscibility pressure value of the learning rate is 0.83. The maximum parameter value for the maximum depth decision tree can be applied at high latitudes and low sample sizes. It is very effective to decide whether to increase the depth according to the result effect. Alpha is the weight of the L1 regularization term and can be used to speed up the computation in the case of high dimensionality. The Optimal value of alpha is 0.50. For the experiments now on, we split the dataset into 60% for training and 40% for evaluation. **Table 3** shows the range and optimal value of the four hyperparameters of PSO algorithm used in the experiment.

Table 3—The hyperparameters of PSO used in the experiment.

	N-estimators	**Learning Rate**	**Max-depth**	**Alpha**
Default value	100	1	10	0.9
Value ranges	10-1000	0.1-1	1-10	0.5-0.95
Optimal values	413	0.83	5	0.50

Figure 6 visualizes the performance of the PSO-GBDT model on training model. The curve is generated by plotting the prediction MMP against the actual MMP at various combination of input setting. Experiments show that the PSO optimizes the GBDT model well for MMP prediction. After the optimization, the PSO-GBDT model achieved 99.9% of accuracy in the training set and 97.6% of accuracy in the test set. The proposed hybrid model confirms to achieve a better performance. The accuracy of the hybrid model is improved by 1.4% and 3.9% for the training and test set, respectively.

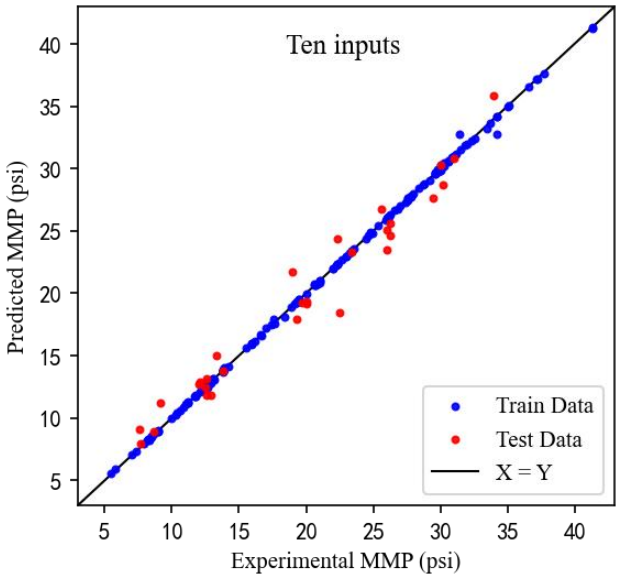

Figure 6—Performance of PSO-GBDT on training model .

Conclusions

Hyperparameter fine tuning has been an obstacle for obtaining a satisfying machine learning model due to the high cost of its trial-and-error process. To tackle this problem, we should speed up the searching efficiency as well as reduce the computation cost of fitness evaluation. We proposed a machine learning model optimized by PSO for efficient MMP prediction. To search for the optimal structure of machine learning model, we performed swarm optimization by encoding hyperparameters into particles. Main conclusions obtained by this work are as follows:

1. Temperature has the greatest influence on the MMP, followed by the molar fraction of carbon components. In contrast, the molar fractions of CO_2 and N_2 components have little effect on MMP.
2. After comparing the MMP prediction models established by the five algorithms, the comparison of accuracy shows that the GBDT model has the best performance, with 98.5% accuracy in the training set and 93.7% accuracy in the test set.
3. With combining PSO algorithm, the PSO-GBDT model was established. The accuracy of the training set is 99.9 %, and the accuracy of the test set is 97.6 %. The hybrid PSO-GBDT prediction model has improved the accuracy of both the training set and the test set, making the MMP prediction more accurate.

Conflicts of Interest

The author(s) declare that they have no conflicting interests.

References

Al-Ajmi, M., Alomair, O., and Adel E. 2009. Planning Miscibility Tests and Gas Injection Projects for Four Major Kuwaiti Reservoirs. Paper presented at the Kuwait International Petroleum Conference and Exhibition, Kuwait City, Kuwait, 12-14 December.

Al-Ajmim M., Tybjerg, P., Rasmussen, C., et al. 2011. EOS Modeling for Two Major Kuwaiti Oil Reservoirs. SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 25-28 September.

Alston, R., Kokolis, G., and James, C. 1985. CO_2 Minimum Miscibility Pressure: A Correlation for Impure CO_2 Streams and Live Oil Systems. *SPE Journal* **25**(2): 268-274.

Cardenas, R.L., Alston, R.B., Nute, A.J., et al. 1984. Laboratory Design of a Gravity-Stable Miscible CO_2 Process. *J Pet Technol* **36**(1): 111–118.

Chen, G., Fu, K., Liang, Z., et al. 2014.The Genetic Algorithm Based Back Propagation Neural Network for Mmp Prediction in CO_2-EOR Process. *Fuel* **126**: 202-212.

Cronquist, C. 1978. Carbon Dioxide Dynamic Miscibility with Light Reservoir Oils. Proc. Fourth Annual US DOE Symposium: 28-30.

Dicharry, R. M., Perryman, T.L., and Ronquille, J.D. 1973. Evaluation and Design of a CO_2 Miscible Flood Project-SACROC Unit, Kelly-Snyder Field. *J Pet Technol* **25**(3): 1309–1318.

Dong, P., Liao, X., Chen, Z., et al. 2019. An Improved Method for Predicting CO_2 Minimum Miscibility Pressure Based on Artificial Neural Network. A*dvances in Geo-Energy Research* **3**(4): 355-364.

Eakin, B.E., and Mitch, F.J. 1988. Measurement and Correlation of Miscibility Pressures of Reservoir Oils. Paper presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, 19-21 October.

Glasø, Ø.1985. Generalized Minimum Miscibility Pressure Correlation. *SPE J.* **25**(5): 927–934.

Graue, D. J. and Zana, E.T. 1981. Study of a Possible CO_2 Flood in Rangely Field. *J Pet Technol* **33**(1): 1312–1318.

Han, P., Jiang, Y., and Zhang, J. 1989. Prediction of Minimum Mixed-phase Pressure for Carbon Dioxide Oil Drive in Daqing oilfield. *Oilfield Chemistry* **6**(4): 309-316.

Harmon, R. A. and Reid, B. G. 1988. Vapor-Density Measurement for Estimating Minimum Miscibility Pressure. *SPE Res Eng* **3**(4): 1215–1220.

Henry, R. L. and Metcalfe, R. S. 1983. Multiple-Phase Generation During Carbon Dioxide Flooding. *SPE J.* **23**(3): 595–601.

Holm, W. L. 1987. Evolution of the Carbon Dioxide Flooding Processes. *Journal of Petroleum Technology* **39**(11): 1337-1342.

Lai, F., Li, Z., and Hu, X. 2017. Improved Minimum Miscibility Pressure Correlation for CO_2 Flooding Using Various Oil Components and Their Effects. *Journal of Geophysics and Engineering* **14**(2): 331–340.

Kanatbayev, M., Meisingset, K.K.K, and Uleberg, K. 2015. Comparison of MMP Estimation Methods with Proposed Workflow. Paper presented at the SPE Bergen One Day Seminar, Bergen, Norway, 19-22 April.

Khan, S.A., Pope, G.A., and Kamy, S. 1992. Fluid Characterization of Three-Phase CO_2/Oil Mixtures. Paper presented at the SPE/DOE Enhanced Oil Recovery Symposium, Tulsa, Oklahoma, 6-8 April.

Mousavi, D.S.A., Vafaie, S.M., Ameri, A., et al. 2008. Minimum Miscibility Pressure Prediction Based on a Hybrid Neural Genetic Algorithm. *Chem. Eng. Res. Des.* **86**(2): 173-185.

Mungan, N. 1981. Carbon Dioxide Flooding—Fundamentals. *Journal of Canadian Petroleum Technology* **20**(1): 87-92.

Nasrifar, K.H. and Moshfeghian, M. 2001. Vapor-liquid Equilibrium of LNG and Gas Condensate Mixtures by the Nasrifar-Moshfeghian Equation of State. *Fluid Phase Equil*. **200**(1):73-88

Orr, F.M and Jensen, C.M. 1986. Interpretation of Pressure-composition Phase Diagram for CO_2-crude Oil Systems. *SPE J* **24**(5):485-497.

Orr, F.M. and Silva, M.K. 1987. Effect of Oil Composition on Minimum Miscibility Pressure-Part 2: Correlation. *SPE Res Eng* **2**(4): 479-490. SPE-14150-PA.

Saeedi, D.A.H. and Soleimani, R. 2020. Prediction of CO_2-Oil Minimum Miscibility Pressure Using Soft Computing Methods. *Chemical Engineering & Technology* **43**(7): 1361-1371.

Silva, M.K. and Orr, F.M. 1987. Effect of Oil Composition on Minimum Miscibility Pressure-Part 1: Solubility of Hydrocarbons in Dense CO_2. *SPE Res Eng* **2**(4): 468-478. SPE-14149-PA

Spence, A.P. and Watkins, R.W. 1980. The Effect of Microscopic Core Heterogeneity On Miscible Flood Residual Oil Saturation. Paper presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas, 21-24 September.

Thakur, G.C., Lin, C.J., and Patel, Y.R. 1984. CO_2 Minitest, Little Knife Field, ND: A Case History. Paper presented at the SPE Enhanced Oil Recovery Symposium, Tulsa, Oklahoma, 19-21 April.

Tian, Y., Ju, B., Yang, Y., et al. 2020. Estimation of Minimum Miscibility Pressure During CO_2 Flooding in Hydrocarbon Reservoirs Using an Optimized Neural Network. *Energy Exploration & Exploitation* **38**(6): 2485-2506.

Yellig, W.F. and Metcalfe, R.S. 1980. Determination and Prediction of CO_2 Minimum Miscibility Pressures. *Journal of Petroleum Technology* **32**(1): 160-168.

Zhang, J., Zhang, X., and Dong, S. 2020. Estimation of Crude Oil Minimum Miscibility Pressure During CO_2 Flooding: A Comparative Study of Random Forest, Support Vector Machine, and Back Propagation Neural Network. 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC).

Zhang, K., Seetahal, S., Alexander, D., et al. 2016. Correlation for CO_2 Minimum Miscibility Pressure in Tight Oil Reservoirs. Paper presented at the SPE Trinidad and Tobago Section Energy Resources Conference, Port of Spain, Trinidad and Tobago, 4-6 June.

Zhao, G.B., Adidharma, H., Towler, B., et al. 2006. Minimum Miscibility Pressure Prediction Using Statistical Associating Fluid Theory: Two- and Three-Phase Systems. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, 24-26 September.

Yawen He, is a master candidate in Petroleum Engineering department at Xi'an Shiyou University. She has focused her research in areas involving reservoir simulation, well testing, and production analysis.

Weirong Li, is a Professor in the Petroleum Engineering Department at Xi'an Shiyou University. He research interests include unconventional resources/reserves estimates, reservoir simulation, well testing, and production analysis. Li holds a bachelor's degree in mathematics from Northeast Petroleum University, China; a master's degree in petroleum engineering from Research Institute of Petroleum Exploration and Development, China; and a PhD degree in petroleum engineering from Texas A&M University.

Shihao Qian, is a master candidate in Petroleum Engineering department at Xi'an Shiyou University. He has focused his research in areas involving reservoir simulation, well testing, and production analysis. He holds a BS degree in Petroleum Engineering from Xi'an Shiyou University.