Intelligent coalbed methane production management and control technology based on reinforcement learning algorithm

Wang Hongli¹,a, Zhang Suian¹,b, Liu Bin¹,c

¹China University of Petroleum, Beijing 8, Fuxue Road, Changping District, Beijing 15210870454. China 102249

²China University of Petroleum, Beijing 8, Fuxue Road, Changping District, Beijing 13901129720. China 102249

³PetroChina Huabei Oilfield Company Renqiu, Hebei 18335512870. China 062552

¹5210870454@163.com

²cbmrc@126.com

³715500861@qq.com

ABSTRACT-The production control of coalbed methane wells has long been viewed as the most challenging step in its development process. For human engineers, they rely too much on previous experience. For artificial intelligence, there is no complete frame to use. Here we proposed a system with reinforcement learning algorithm to CBM production control optimization that used a proxy model to simulate the gas and water seepage in coal seam, and a ‘value networks’ to evaluate gas and water production capability and three control policy mode: bottomhole pressure (BHP) regression model, BHP reduction rate mode, BHP table to select moves. The system achieved a 20.99% and 38.14% increment in cumulative gas and water production, respectively.

1. INTRODUCTION

Production (or drainage) control is an important link in the development of coalbed methane (CBM). Unreasonable drainage and bottom-hole pressure (BHP) reduction strength will cause near-wellbore permeability to drop, restricting the spread of pressure drop funnels. It will cause single well productivity to be lower than expected finally[1-4].

Conventional CBM production policy divides the process into discharging stage, dropping water level stage, holding casing pressure stage, controlling BHP stage and keeping BHP stage. The principle of conventional CBM production policy is "Slow, long-term, continuous, and stable". Although this principle is in line with previous development experience, large-scale land surface CBM development of high-rank coal in China is still in the exploratory stage, and a set of production management system that adapt to different geological conditions has yet to be formed. Domestic researches always focus on the analysis of a certain controlling factor. High-rank reservoirs have the characteristics of complex pores and fractures, low permeability, and strong sensitivity. The gas in the coal seam is controlled by multiple factors, which cannot be solved by a single controlling factor. The production of CBM wells in China relies too much on experience, and it is difficult to effectively integrate the theory and practice formed by theory [4-7].
Reinforcement learning algorithms have achieved some success in a variety of domains [8-10], their applicability has previously been limited to domains in which useful features can be handcrafted, or to domains with fully observed, low-dimensional state spaces. Reinforcement learning algorithms are not yet popular in the oil and gas domain.

Talavera developed and implemented a predictive control model with reinforcement learning (MPCRL). The model was applied to control the oil production from a petroleum reservoir [22]. J. L. Guevara presented the application of reinforcement learning approach as an alternative to conventional algorithms for reservoir optimization. His objective is to find the steam injection rates at every time step that will maximize overall net present value at the end of the production [21].

For a given CBM well production control, no matter straight or horizontal well, the objective is to find the bottom hole pressure drop mode at each sampling instant to maximize a particular performance measure (e.g., daily production rate or cumulative production) while accounting for the complex reservoir dynamics. Here we use recent advances in training deep neural networks and reinforcement learning to develop a novel artificial agent, that can learn successful policies directly from automatic simulation.

2. METHODOLOGY

2.1. Reinforcement Learning Algorithm

Reinforcement learning is an unsupervised optimization method, inspired by behaviorist psychology, to find the best control strategies to achieve the desired objectives and also to maximize the defined benefits and rewards in a dynamic environment.

![Reinforcement learning framework]([19])

In reinforcement learning, a learner or agent, on its own learns step by step where to go and what to do by iteratively trying different possible actions in each situation [19]. In Reinforcement learning, our objective is to train an Agent to find an optimized Policy that will maximize the total Reward by continuous interactions with the Environment. A policy $\pi$ is defined as the behavior or action of the agent at each state of the environment or in a deterministic policy.

Reward represents a scalar feedback signal that indicates its performance at time step $t$. The total reward of a process is defined as:

$$G_t = r_{t+1} + \gamma r_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

where $\gamma$ is a discount factor.

The environment contains a Markov decision process (MDP) in which the current state captures all relevant information from the history. The change of state in MDP is defined using a probability function as:

$$\mathbb{P}[s_{t+1}|s_t] = \mathbb{P}[s_{t+1}|s_1, \ldots, s_t]$$

In this research value-based strategy is used, where during the MDP, an action-value function is approximated. This function is the expected overall reward starting from states, taking action $a$ and then following policy $\pi$. 

At every time step, a new action is chosen by following a policy improvement mechanism, such as $\epsilon$-greedy policy. In this policy, the agent will choose the optimal action (one that will offer the most return), also known as greedy action with a probability of $1 - \epsilon$, and a random action with a probability of $\epsilon$. This represents a trade-off between exploration and exploitation. Mathematically this can be expressed as:

$$q_\pi(s,a) = \mathbb{E}[G_t | s_t = s, a_t = a]$$  \hfill (3)

Where $m$ represents the number of possible actions.

\section*{2.2. Coalbed Methane Production Proxy Model}

The development of Environment for simulating gas and water transportation in coal seams is the premise for reinforcement learning. Though commercial simulation tools such as GEM and Eclipse have been widely used in optimization, the integration of any of these software packages in an auto-optimization program is usually problematic, and numerical simulation is too slow to combine with the continuous interactive processing of reinforcement learning. As a result, we used artificial neural networks to develop a CBM production proxy model for reservoir simulation. The simulator is essentially a dual-porosity, compositional model that is capable of simulating the sorption and diffusion phenomenon in matrix and fluid flow through the cleats. The proxy model is incorporated directly into the reinforcement learning algorithm as a state feedback simulator.

\section*{2.3. Coalbed Methane Production Control Optimization Pipeline}

For the specific optimization problem in this study, the objective is to find the optimal BHP drop mode of wells that maximize or minimize a given objective function.

The computational procedure (Figure 2) can be described as follows:

Step 1. Initialization. Detect the outliers in production data, and clean the data;

Step 2. Use the same block CBM production history data to develop a CBM production proxy model with artificial neural networks;

Step 3. Run history matching using the proxy model to modify the Environment;

Step 4. Build the Agent according to the special control problem;

Step 5. Run the optimization loop.

Fig.2 Diagram of the computational procedure for the optimization of CBM production control
We consider tasks in which the agent interacts with an environment through a sequence of observations, action sand rewards. The goal of the agent is to select actions in a fashion that maximizes cumulative future reward. We use a deep neural network to approximate the optimal action-value function from the cumulative water production.

Table 1  Agent components

| Reinforcement learning Agent elements | Elements setting                                      |
|---------------------------------------|------------------------------------------------------|
| States                               | Gas content                                          |
| Action                               | BHP reduction table                                  |
| Environment                          | Proxy model                                          |
| Reward                               | Normalize total time step cumulative water production |
| Greedy policy                        | \( \alpha : 0.9; \quad \gamma : 0.7 \)              |
| Reinforcement learning algorithm     | Q learning                                            |
| episodes                             | 32                                                   |

3. APPLICATION OF THE OPTIMIZATION PROCEDURE FOR CASE STUDIES

3.1. Geological Setting
At present, the high-rank coal-bed methane in the southern part of the Qinshui Basin in Shanxi has been developed commercially, and it is one of the important coal-bed methane industrial bases in China. In this section, we apply the CBM production control optimization pipeline in the Fanzhuang district.

The Fanzhuang district is located on the southeastern edge of the horseshoe-shaped slope belt in the Qinshui Basin. The strata in the area are broad and gentle, and generally inclined to the west. The dip angle of the strata is generally 2°-7°. It is mainly composed of small-distance normal faults and low gentle parallel folds. The direction of the folds is mainly north-north-east and nearly north-south. The coal seam is well preserved and the lateral distribution is stable. The thickness of the 3# coal seam is 5-8 m, the thickness of the 15# coal seam is 1.5-5 m, the buried depth of the 3# coal seam is 300-700 m, the thickness is 4.6-7 m, and the gas content is between 16-29 m³/t; The buried depth of the 15# coal seam is generally between 400-800 m, the thickness is 1.7-4.1 m, and the gas content is 15-27 m³/t. 3# and 15# coal seam have a high degree of thermal evolution and are anthracite. Formation tests show that coal reservoirs are under pressure-atmospheric pressure reservoirs with low pressure, low permeability, and strong heterogeneity.

3.2. CBM Production Data Preprocessing
Outliers are often present in large CBM well production datasets. Outlier refers to those data points with significant deviations caused by human factor, production equipment factor and computer factor.

We use exponential smoothing detection method, and set a 95% confidence interval to identify outliers, the formula is:

\[
\ell_t = \alpha Q_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \\
b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \\
Q_{t+1} = \ell_t + b_t
\]  

Where \( \ell_t \) is horizontal component, \( b_t \) is trend component, \( \alpha \) is moving average weight, \( \beta \) is exponential smoothing weight.
Taking Well A3612 as an example, the gas production rate, water production rate, BHP, casing pressure and dynamic water level before and after the data preprocessing are shown in Figure 3.

3.3. Proxy Model and History Matching

Impacted by uncertainties, which are derived from complexity, diversity of the coal formation condition, degree of accuracy in exploration and test method limitations, the parameters acquired during the test process may not reflect the true in situ conditions.

Parameters including the porosity, permeability, gas content, thickness, top depth and in situ pressure of the three seams were obtained from logging data of the drilled wells. Permeability data with anisotropy were unavailable and are assumed to identify along the butt cleats and face cleats in the history matching.

The relative permeability of CBM reservoirs to gas and water had a considerable effect on CBM production characteristics (Clarkson et al., 2011). Because the curve of Fanzhuang district was unavailable, the data was adjusted by trial and error (Fig. 5) until the simulated gas and water rates agreed with the field data within an acceptable range.

![Fig.3 Production data preprocessing result](image)

![Fig.4 Relative permeability to gas and water used in the history matching](image)
### Table 2  Property parameters of the field

| Parameters                        | Original value | History matching value |
|-----------------------------------|----------------|------------------------|
| Coal seam thickness (m)           | 4.5            | 4.5                    |
| Langmuir volume (m³/t)            | 23.5           | 23.5                   |
| Langmuir pressure (MPa)           | 2.2            | 2.2                    |
| Gas content in matrix (m³/t)      | 13.4           | 14.5                   |
| Cleat permeability (md)           | 0.8            | 1.75                   |
| Porosity (%)                      | 2              | 1.5                    |
| Initial in situ pressure (MPa)    | 6.2            | 6.2                    |
| Reservoir temperature (℃)         | 32.5           | 32.5                   |

History matching of gas and water rates was conducted for Well A3612. Well A3612 cumulative gas production and water cumulative production errors were determined to be 3.15% and 6.42%, respectively. History matching results using proxy model were illustrated in Figure 5.

![Fig.5  History matching result of proxy model](image)

3.4. **Production Control Policy Optimization**

We carry out the optimization of the CBM well production control policy by maximizing cumulative Reward in the previous proxy model with reinforcement learning Agent. Similarly, to achieve reliable conclusions, we conducted a total of ten optimization runs.

![Fig.6  Policy design](image)

We set out to create a single function that would be able to develop a wide range of Polices on a varied range of challenging tasks—a central goal of general artificial intelligence that has eluded previous efforts. To achieve this, we developed an BHP regression function (Figure 6(a)), a BHP reduction mode (Figure 6(b)), and a BHP table (Figure (c)), which is able to combine reinforcement learning with a class of polices. After we integrated the algorithm many times, we found that the second policy-BHP reduction mode works best.
Figure 7 shows average value and average accumulative value per trial for two strategy. As it can be seen, the optimal policy is gradually learned in less than 35 episode and the values of accumulative reward remains nearly constant in the last episodes. This fact implies that although the rewarding method is simple, it can successfully converge to an optimal policy. Furthermore, each step performance quality of the proposed method in the process is observed in (c) and (f).

![Fig.7 Training curve tracking the Agent's predict action value](image)

Fig. 8 contains the policy optimization results of production control in the BHP reduction rate and BHP, which maximizes cumulative gas production. No regular BHP reduction mode was observed. However, the following should be noted: 1) Although the new BHP reduction mode performs worse in terms of stability, shows stronger stability and 2) The action setting of the BHP mode alone cannot guarantee a high value for each action, as shown in figure 7(f).

![Fig.8 Original and optimal BHP reduction mode](image)

Gas and water productions and cumulative productions were compared by using the original and optimized Policy. Before conducting the optimization, all original BHP setting were removed from the model and the environment was assumed to be a virgin frontier. Due to well schedule management or well workover in the history of production, dynamic level and (BHP) vary with time, resulting in fluctuations in fluid rate and uncertainty in the prediction of cumulative production.

The total CGP and CWP for the wells with the original policy were determined to be $8.1 \times 10^4$ m$^3$ and 97 m$^3$, respectively; while the productions, optimized over 900 days, were $9.8 \times 10^4$ m$^3$ and 134 m$^3$, respectively.
respectively. An increment of 20.99% of gas production, and 38.14% of water production was observed with optimization.

The optimal policy was stable BHP and unstable BHP reduction combination. In other scenarios, which involved different BHP reduction rate, at least BHP do not fluctuate wildly. Such observations indicate that the BHP reduction rate control is favorable for the optimal control. As indicated in Figure 5(f), a region with relatively low value. An action in this region need control casing pressure and dynamic water level to cooperation.

4. CONCLUSION AND FUTURE WORK

Optimization of the production control of CBM vertical wells was conducted by maximizing the total value of each drainage stage and integrating a proxy model to the reinforcement learning optimization agent. The optimization results for single well production control in a modified environment indicated a satisfactory performance of agent. A comparison of cumulative gas production for the optimized and original control of wells highlights the necessity of using optimization runs in field applications. Infilling scenarios indicated that the optimal number of infill wells is one.

The success of the reinforcement learning algorithm in determining optimal production control policy in BHP control of CBM wells inspires the application of the algorithm into mixture control of BHP, casing pressure and dynamic water level, including optimization of fracturing design and operation schedules. Future research on the performance evaluations of different production control mode and reinforcement algorithms for these applications is also recommended.

ACKNOWLEDGMENTS

This research was conducted under the financial support of National Science and Technology Major Project of China (Grant No. 2017ZX05064004).

REFERENCES

[1] Clarkson C R, Qanbari F. 2016. A semi-analytical method for forecasting wells completed in low permeability, undersaturated CBM reservoirs. J. Journal of Natural Gas Science and Engineering, 2016,30:19-27. DOI= https://doi.org/10.1016/j.jngse.2016.01.040

[2] Seidle J P. 2002. Coal well decline behavior and drainage areas: Theory and practice. In SPE Gas Technology Symposium, (Calgary, 2002). DOI= https://doi.org/10.2118/75519-MS

[3] Sun Zheng, Shi Juntai, Wu Keliu, et al. 2018. A fully-coupled gas-water two phase productivity equations for low-permeability CBM wells. J. Journal of Petroleum Science and Engineering, 2018, 166: 611-620. DOI= https://doi.org/10.1016/j.ijheatmasstransfer.2018.05.078

[4] Li Guihong, Zhang Hong, Cui Yongjun, et al. 2005. A predictive model of gas content in coal reservoirs based on multiple stepwise regression analysis: a case study from Qinshui Basin. J. Coal Geology & Exploration, 2005, 33(3): 22-25.

[5] Hu Yafei, Zhang Suian, Wu Zhiying. 2013. The gas content modeling method based on geological statistical analysis: with seam No. 3 in southern Qinshui basin as an example. J. Coal Geology & Exploration, 2013,41(2):33-36.
[6] Chen Feng, Shen Dongri, Chen Yijun, et al. 2006. Application of wavelet fuzzy net in prediction of coal and gas outburst. Journal of Safety in Coal Mines, 2006(10):51-54.

[7] Abdulhai, B., Pringle, R., Karakoulas and G.J., 2003. Reinforcement Learning for True Adaptive Traffic Signal Control. J. Transp. Eng. 129 (3): 278–285. DOI= https://doi.org/10.1061/(ASCE)0733-947X(2003)129:3(278)

[8] Achiam, J., Held, D., Tamar A. and Abbeel, P., 2017. Constrained Policy Optimization. eprint arXiv:1705.10528 (posted May 2017) http://arxiv.org/abs/1705.10528

[9] Al-Gosayir, M., Leung, J. and Babadagli, T., 2012. Design of Solvent-Assisted SAGD Processes in Heterogeneous Reservoirs Using Hybrid Optimization Techniques. J. Can. Pet. Technol. 51 (6): 437–448. DOI= https://doi.org/10.2118/149010-pa

[10] Bao, X., Chen, Z.J., Wei, Y., Dong, C.C., Sun, J., Deng, H. and Yu, S., 2010. Numerical Simulation and Optimization of the SAGD Process in Surmont Oil Sands Lease. In Abu Dhabi International Petroleum Exhibition and Conference, (Abu Dhabi, UAE, 1-4 November). SPE-137579-MS. DOI= https://doi.org/10.2118/137579-ms

[11] Benndorf, J. and Jansen, J.D., 2017. Recent Developments in Closed-Loop Approaches for Real-Time Mining and Petroleum Extraction. J. Math. Geosci. 49 (3): 277–306. DOI= https://doi.org/10.1007/s11004-016-9664-8

[12] Chen, Y., Oliver, D.S. and Zhang, D., 2009. Efficient Ensemble-Based Closed-Loop Production Optimization. J. SPE J. 14 (4): 634–645. DOI= https://doi.org/10.2118/112873-PA

[13] Chen, Z., 2013. A Genetic Algorithm Optimizer with Applications to the SAGD Process. PhD Thesis, University of Calgary, Calgary, Canada (July 2013)

[14] Nikolaou, M.; Cullick, A. S.; Saputelli, L. 2006. Production Optimization: A moving Horizon Approach. Netherlands: SPE 99358. 2006. DOI= https://doi.org/10.2118/99358-MS

[15] Sarma, P. et al. 2006. Efficient real-time reservoir management using adjoint-based optimal control and model updating. J. Computational Geosciences. 10 (1). 3-36, March 2006. ISSN 1573-1499. DOI= https://doi.org/10.1007/s10596-005-9009-z

[16] Sarma, P. 2006. Efficient Closed-Loop Optimal Control of Petroleum Reservoirs under Uncertainty. Stanford University. Stanford. PhD Thesis.

[17] Chen, Y.; Oliver, D. S.; Zhang, D. 2009. Efficient Ensemble-Based Closed-Loop Production Optimization. J. Society of Petroleum Engineers Journal, p. 634-645, December 2009. DOI= https://doi.org/10.2118/112873-PA

[18] Wang, C.; Gaoming, L.; Reynolds, A. C. 2009. Production Optimization in Closed-Loop Reservoir Management. J. Society of Petroleum Engineers Journal. 506-523, September 2009.

[19] Sutton, R. & Barto, A. Reinforcement Learning: An Introduction (MIT Press, 1998).

[20] Thorndike, E. L. Animal Intelligence: Experimental studies (Macmillan, 1911).

[21] Guevara, J. L., Patel, R. G., & Trivedi, J. J.. 2018. Optimization of Steam Injection for Heavy Oil Reservoirs Using Reinforcement Learning. SPE International Heavy Oil Conference and Exhibition.

[22] Talavera, A. G., Tupac, Y. J., & Vellasco, M. M. B. R. 2010. Controlling Oil Production in Smart Wells by MPC Strategy with Reinforcement Learning. SPE Latin American and Caribbean Petroleum Engineering Conference.