Recurrent Neural Network Methodology for Prediction of SAGD performance utilizing petrophysical log data

*Namhwa Kim[1)], Kyungbook Lee[2)], Hyundon Shin[1)]**

1) Department of Energy Resources Engineering, Inha university, Republic of Korea

2) Korea Institute of Geoscience and Mineral Resources, Republic of Korea

Summary

The steam-assisted gravity drainage (SAGD) is the most implemented in-situ thermal recovery process to produce bitumen at Athabasca oilsands area in Canada since early 2000s. SAGD performance is very closely related to reservoir quality and its heterogeneity. However, it requires lots of static and dynamic data and their analysis to generate 3D reservoir model and its forward simulation model. In this study, we proposed machine learning (ML) model for prediction of SAGD production based on petrophysical log data. Long short-term memory (LSTM) algorithm, which is a promising recurrent neural network algorithm to train sequential data, was introduced to a data pair of log data and cumulative oil production, net present value at 7th operational year (CUMOIL7, NPV7) as input and output attributes, respectively.

Background Theory

Although there are several types of prediction method for SAGD performance, numerical simulation method has been widely used for oil sand reservoir (Gates et al., 2007; Su et al., 2013). Although this method deals with heterogeneous reservoir, it needs to construct three-dimensional reservoir model based on the petrophysical log data, core data and seismic data. The reservoir modeling requires cost and time, but still generates high uncertainty during the procedure (Pyrcz and Deutsch, 2014).

In recent years, various machine learning techniques got focus on their ability to approximate relationship between input and output attributes. Among the techniques, artificial neural network (ANN) is a non-linearly linked network between input and output attributes, it approximates the non-linear relationship by training data. However, for purpose-fit implementation of ANN, deciding what kind of layer to use for the neural network is very important. If the data set is formatted as sequential data, recurrent layer needs to be used. Especially, the recurrent layer is called as recurrent neural network (RNN). Long-short term memory (LSTM) is a promising algorithm to learn sequential data, because it was improved to reflect longer past data than existing algorithms.

According to Shin (2008), SAGD performance can be predicted using reservoir parameters such as permeability, oil saturation, porosity. These parameters are interpreted from petrophysical log data. Permeability is mainly related with lithofacies, and the main lithofacies in oil sand are sand and shale (Carrigy, 1966). Also, gamma ray log is used as qualitative indicator for existence and portion of shale in oil sands (Rider, 1996; McCormack, 2001). Saturation is related to resistivity, deep resistivity log is used to calculate water saturation via Archie's equation with assumptions (Asquith, 2004). Also, porosity is interpreted via several methods using several porosity logs.

This study tried to develop prediction model for SAGD performance at the early stage of field development when there is no 3D reservoir model. There exists lots of cored wells and production wells in Athabasca oil sands development area, that means oil sands project is a good candidate for develop a machine learning model. Thus, machine learning method was applied to capture underlying relationship between vertical heterogeneity of log data and SAGD performance. Especially, only petrophysical log data were used to reflect reservoir parameter.

However, there has been no previous attempt to utilize log data as input attribute, proper data pre-processing methods for machine learning has not been proposed. Therefore, in this study, it was tried to determine the optimal data pre-processing methods that can effectively learn the relationship between input data and output.

Methodology

In this research, the utilized log data and production data are from seven various SAGD fields located in Athabasca oil sands area, Canada. To reflect proper reservoir characteristics as possible, only log data which departs shorter than 50m from SAGD horizonal wells were chosen and SAGD wells which have more than seven years production history were selected.

As the SAGD performance indicator, cumulative oil production (CUMOIL7) and net present value at 7th year (NPV7) were calculated. For the NPV calculation, simple economic parameters were assumed: bitumen price ($32/bbl), steam cost ($8/bbl cold water equivalent) and discount rate (10%/year) (Nejadi et al, 2018). For the petrophysical well log data, gamma ray (GR) log, deep resistivity (RT) log and neutron porosity (NPHI) log were selected, and the bottom boundary of log data was set to the elevation of SAGD producer.

Total 65 pairs of data sets were obtained by above data selection. Then, the entire data sets were split to train, validation and test set; 53 train, 5 validation, 7 test sets. Also, the normalized GR, RT, NPHI logs were used as input attributes and normalized CUMOIL7, NPV7 were used as output attributes for machine learning.

For the machine learning procedure, Keras which is python-based machine learning library with TensorFlow backend was used. RNN structure was configured with one LSTM layer and output layer sequentially. The LSTM's number of cells was set to 10. To prevent over-fitting problem, L2 regularizer and drop-out (0.5) was set. Then, Adam was used as optimizer and loss function was set to mean-squared error (MSE). The batch size (53 case at one epoch), maximum epoch (5000 epoch) and early-stop function (stopping when over-fitting starts) was set.

For log data, 100cm log record interval was used to integrate different record interval such as 10cm, 12.5cm, 15.24cm. Then, RNN was trained using log data in full McMurray formation. Fig. 1 show the loss difference between train and validation at last epoch is big, it means RNN couldn't capture underlying relationship between input and output attributes enough. This result could happen if the pre-processed data had excessive amount of information to learn the relationship between input and output attributes. Thus, additional pre-processing works were needed to eliminate low impact data and to add meaningful data.

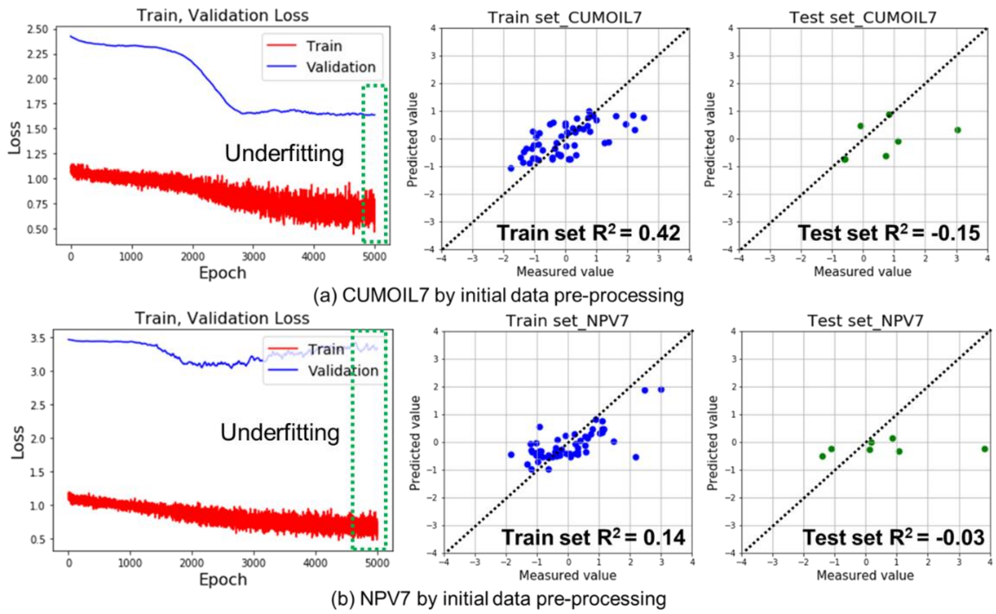


Fig. 1 Training and test results by initial data pre-processed data

The initial pre-processed log data covered full McMurray formation like Fig. 2 (a). However, pay zone, which have direct relationship with SAGD performance, is located around SAGD well elevation. It means that the log data can be reduced to focus on the pay zone by eliminating lower impact range like Fig. 2 (b), (c).

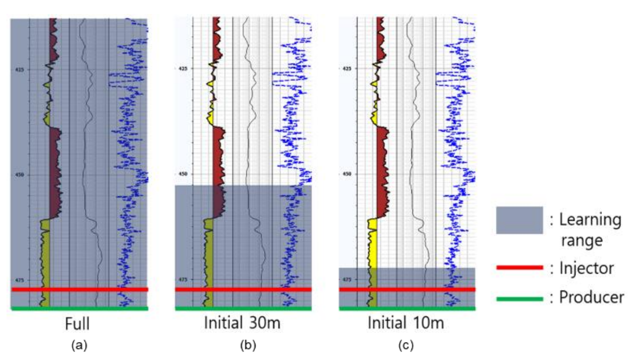
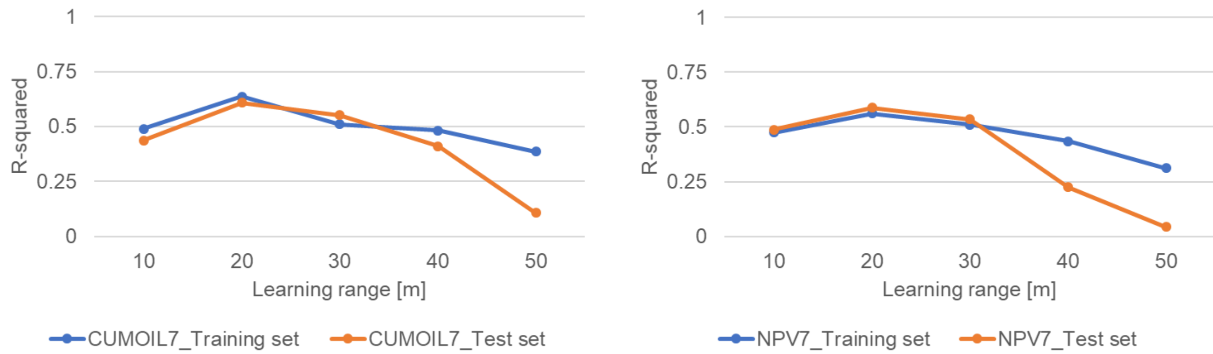
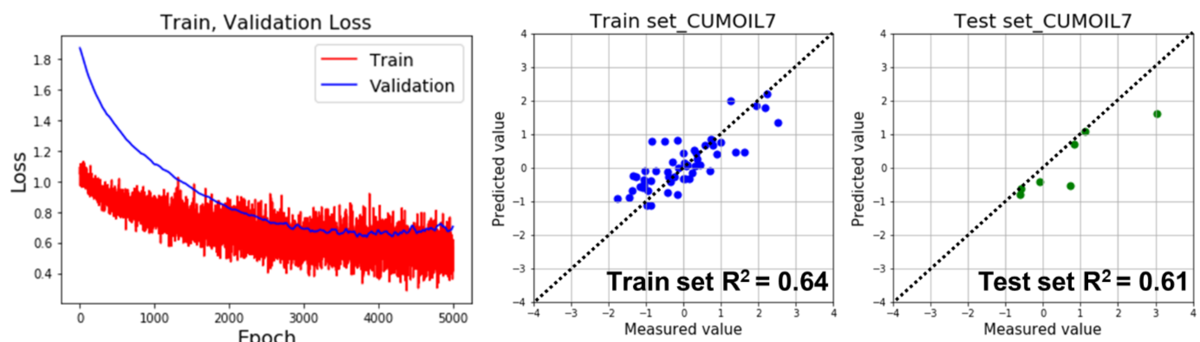


Fig. 2 Adjustment for learning range of log data. (a) Full range; (b) Initial 30m range; (c) Initial 10m range

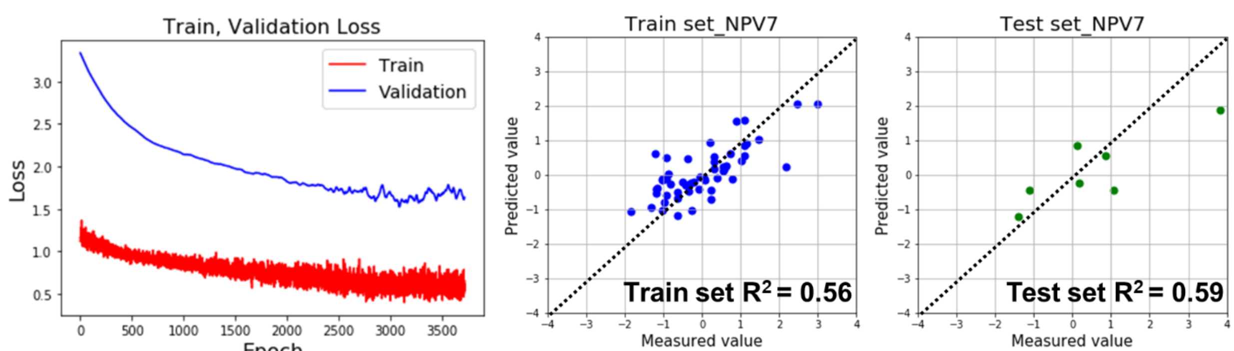
Fig. 3 shows that narrow learning ranges (10, 20, 30m) had better prediction performance than thicker ranges (40, 50m). This implies that data utilization only for pay zone improves the prediction performance, especially 20m learning range showed best performance. The result of 20m learning range gives the best prediction performance compared with full learning range.



(a) R-squared values of CUMOIL7, NPV7



(b) CUMOIL7 by 20-meter vertical range

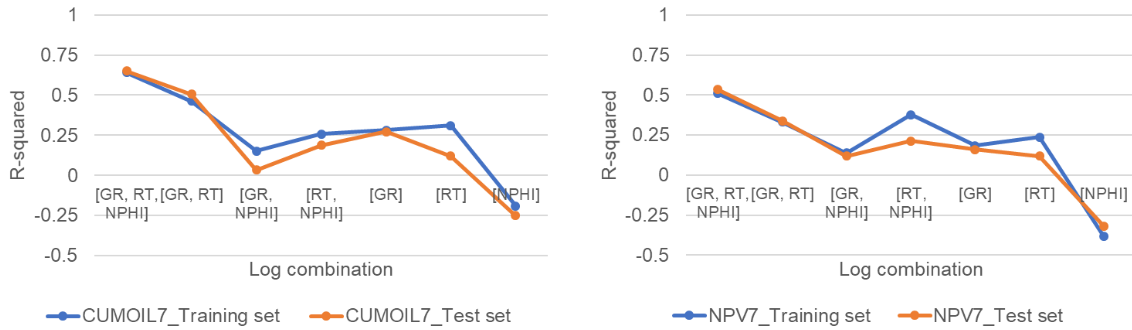


(c) NPV7 by 20-meter vertical range

Fig. 3 Prediction performance comparison for several learning ranges.

Each logging tools indicate different reservoir parameters. By sensitivity analysis on all possible logging tool combinations with GR, RT, and NPHI, it was compared how each logging tool affects the predictive performance (Fig. 5).

Furthermore, we did other sensitivity analysis how to learn log data such as input direction (from bottom to top or from top to bottom) and operational time.



(a) R-squared values of CUMOIL7, NPV7

Fig. 4 Prediction performance comparison for several log combinations.

Results and Discussion

In this study, SAGD performance prediction based on RNN was conducted by utilizing only petrophysical log data. Due to the limited number of data set (65), it was difficult to train enough for all log data in McMurray formation. Thus, it is needed to reduce the input attribute's complexity by adjusting learning range of log data. The result showed that 20m learning range was the best prediction performance and [GR, RT, NPHI] combinations is enough to fully reflect reservoir parameters for the SAGD process.

The study shows the possibility of applying deep learning tool for prediction of SAGD performance and its economic evaluation even though the limited number of subsurface data. In fact, the petrophysical log data exist one per legal subdivisions (LSD). The results of this study can be used as a decision-making tool during for the field development plan like new SAGD well drilling.

However, there are limitations to overcome in this study:1) No consideration about SAGD operational conditions, 2) lack of well log data set. Therefore, it is required further investigation to predict SAGD performance effectively.

Acknowledgements

This study was supported by the Demonstration-scale Offshore CO2 Storage Project in Pohang Basin, Korea (20162010201980). Also, this study was supported by the Korea Institute of Geoscience and Mineral Resources to obtain SAGD field data.

References

Asquith G B, Krygowski D, Gibson C R. Basic well log analysis. Tulsa: American Association of Petroleum Geologists, 2004.
Carrigy, M A. Lithology of the Athabasca oil sands. Canada.
ZHANG Dongxiao, CHEN Yuntian, MENG Jin. Synthetic well logs generation via Recurrent Neural Networks. PETROLEUM EXPLORATION AND DEVELOPMENT, 2018, 45(4): 598-607.
Gates, I. D., Kenny, J., Hernandez-Hdez, I. L., & Bunio, G. L. (2007, February 1). Steam Injection Strategy and Energetics of Steam-Assisted Gravity Drainage. Society of Petroleum Engineers. doi:10.2118/97742-PA
IHS Markit. 2018. AccuMapTM 28.03, Production Module in Alberta.
Lee, K., Lim, J., Yoon, D., & Jung, H. (2019, December 1). Prediction of Shale-Gas Production at Duvernay Formation Using Deep-Learning Algorithm. Society of Petroleum Engineers. doi:10.2118/195698-PA
McCormack, M. (2001, August 1). Mapping of the McMurray Formation for SAGD. Petroleum Society of Canada. doi:10.2118/01-08-01
Nejadi, S., Hubbard, S. M., Shor, R. J., Gates, I. D., & Wang, J. (2018, November 25). Optimization of placement of flow control devices under geological uncertainty in Steam Assisted Gravity Drainage. Society of Petroleum Engineers.
Michael J. Pyrcz, Clayton V. Deutsch. Geostatistical Reservoir Modeling. OUP USA. 2014. ISBN 0199731446.
Rider, M., "The Geological Interpretation of Well Logs", Gulf Publishing Company. Second edition. 1996.
Alaa Sagheer, Mostafa Kotb, Time series forecasting of petroleum production using deep LSTM recurrent networks, Neurocomputing, Volume 323, 2019, Pages 203-213, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2018.09.082.
Shin, H. (2008, January 1). A Reservoir Economic Indicator to Evaluate a SAGD Project. Petroleum Society of Canada. doi:10.2118/2008-072
Su, J.Y. Wang, I.D. Gates, SAGD well orientation in point bar oil sand deposit affects performance, Engineering Geology, Volume 157, 2013, Pages 79-92, ISSN 0013-7952, https://doi.org/10.1016/j.enggeo.2013.01.019.