Introduction

To evaluate the potential of underground reservoirs, hydrocarbon productivity is one essential property. Only if the productivity is higher enough to cover the exploitation cost, a reservoir is economically exploitative. Thus, productivity evaluation is essential in reservoir characterization and development, which effectively guides the deployment of well sites and the exploitation engineering.

Many researches have been done on productivity prediction and evaluation. Considering the influence of grain size and porosity on permeability, Hogg et al. (1996) predicted well productivity from grain analysis and logging while drilling. Using commonly available open-hole log data, Cheng et al. (1999) developed a simple and cost effective method to predict initial productivity. Linking wireline logs with formation testing and/or core data, Liu et al. (2000) predicted productivity of multi-layer clastic reservoirs. Productivity depends not only on the storage and permeability of the reservoir formations, but also on the engineering factors such as hydraulic fracturing. Huang et al. (2015) derived a steady productivity formula of fracturing directional wells in anisotropic reservoirs based on point-source potential function. Considering the compressibility of both fluids and solids, Chen et al. (2019) developed a three-dimensional unified pipe-network method to understand oil flow behavior in fractured unconventional reservoirs, and further evaluated the influence of fractures on oil production.

Physics-based productivity prediction methods often apply to certain type of reservoirs, which require in-depth geological understanding to achieve better predictions. Benefit from the development of big data and machine learning, some data-driven methods have been developed to perform productivity evaluation. Pan et al. (2015) and Hu et al. (2018) both introduced neural networks into productivity prediction by establishing relationship between productivity and well measurements. However, log curves were averaged to set as inputs of neural network models, discarding structural characteristics of reservoir formations.

In this paper, we develop a hybrid deep neural network (HDNN) model to perform productivity prediction. The hybrid model is a concatenate network consist of multilayer perceptron (MLP) and convolutional neural network (CNN). This architecture takes advantage of both discrete numerical inputs and structural log curves.

Theory

There are multiple sources of data that may have influences on reservoir productivity, such as core, well logging, well tests, and engineering operations. The data type of various data includes numerical type, category, image, text and so on. In addition, sampling presentation of different data is various. For example, some test or engineering data, such as facies description or hydraulic fracturing, are discrete. While some geophysical properties are presented as structured log curves, whose sampling index is continuous. Discrete data indicates integral effect of reservoir formations. However, structural data reflects more details within the formations. Considering that production is a criterion of reservoir formation, more comprehensive data should be taken into consideration to perform accurate productivity prediction, especially the structural data.

![Figure 1 Architecture of HDNN for productivity prediction, where FC indicates fully-connected layer.](image)
To better use various attributes of various types and formats and to fully learn the relationship between measurements and target production, a HDNN (shown as Figure 1) is proposed. The proposed network is a combination of MLP and CNN, taking numerical, categorical and structural data as mixed inputs. MLP is adopted to process numerical and categorical inputs. CNN is applicable to extract high-hierarchy features from structural data. After feature leaning for different formats of data separately, outputs of MLP and CNN are concatenated to achieve a final evaluation of production.

The procedure of productivity prediction based on HDNN comprises following four steps:

- **Data preparation and preprocessing**

  The inputs to HDNN, including numerical, categorical and structural data, should be firstly gathered from various wells. Then some preprocessing works are taken to prepare well-sampled dataset. For categorical data, one-hot encoding is utilized to perform quantitative transformation. Log curves are extracted according to the depth range of each reservoir formation of each well. Since the thicknesses of various formations are generally different, resampling is required to obtain uniform-sampled data. Furthermore, numerical and structural data are normalized to avoid the effect of the inconsistent scale. Besides features, hydrocarbon production of each formation is set as target label. Finally, the prepared dataset is separated into training and test data for model training and performance evaluation.

- **Deep neural network construction**

  After preprocessing, the inputs for machine learning are numerical and structural data. To better use the mixed inputs, MLP and CNN are adopted to separately cope with numerical and structural data. MLP consists of multiple fully connected (FC) layers. CNN is composed of several convolutional network units and can effectively excavate inherent features from structural data. For structural logging data, a one-dimensional convolution layer is utilized. Following separate feature leaning, outputs of MLP and CNN are concatenated to predict productivity.

  The training dataset is represented as \( \{ \mathbf{X}^N, \mathbf{X}^S, Y_i \} \), where \( \mathbf{X}^N \) and \( \mathbf{X}^S \) are numerical and structural data, \( Y \) is observed production, \( i \) indicates instance index, varying from 1 to \( M \). Setting the nonlinear mapping function from input features to target production as \( F \), the HDNN model is expressed as

  \[
  \hat{Y}_i = F(\mathbf{X}^N_i, \mathbf{X}^S_i),
  \]

  where \( \hat{Y}_i \) is the predicted production. For the proposed HDNN, \( F \) is composed of multi-layer FC layers, convolution layers, pooling layers and activation functions.

- **Hyperparameter tuning and model training**

  To perform optimization, mean squared error (MSE) loss function is adopted, expressed as

  \[
  MSE = \frac{1}{M} \sum_{i=1}^{M} \left\| F(\mathbf{X}^N_i, \mathbf{X}^S_i) - Y_i \right\|^2.
  \]

  The performance of deep learning depends not only on the dataset, but also on many hyperparameters, such as the number of network layers, the number of nodes, the number of convolution kernels, the size of convolution kernels, the learning rate, etc. For this reason, hyperparameter tuning is essential to achieve better performance. In addition, the well-trained model is applied to test data to evaluate its generalization performance. Some quality control analyses are carried out to validate the consistency of the predicted and observed productions.

- **Productivity prediction and evaluation**

  With above steps, a HDNN model for productivity prediction is well trained. The well-trained model can represent the complex nonlinear relationship between various geological and engineering measurements and the target hydrocarbon production. Then possible production of some unknown wells can be predicted, further guiding well location deployment and development engineering.
Examples

The HDNN model is applied to a development oil block, with 180 development wells. Seven types of log curves, namely caliper, acoustic time, gamma ray, spontaneous potential, shale content, deep and shallow lateral resistivity, are available for all wells. Initial oil production is set as target label. Besides log curves, some formation or engineering attributes with relation to initial production are taken into consideration. These attributes include formation thickness, formation median depth, perforation thickness and perforation number. Table 1 shows available data features and their corresponding data types. Log curves of a certain formation are shown in Figure 2. Due to the restriction of available measurements, only numerical and structural data are provided here for initial oil production prediction.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Numerical Data</th>
<th>Structural Data (Log Curves)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formation Thickness</td>
<td></td>
<td>CAL</td>
</tr>
<tr>
<td>Formation Depth</td>
<td></td>
<td>AC</td>
</tr>
<tr>
<td>Perforation Thickness</td>
<td></td>
<td>GR</td>
</tr>
<tr>
<td>Perforation Number</td>
<td></td>
<td>SP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VSH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LLD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LLS</td>
</tr>
</tbody>
</table>

Figure 2 Demonstration of log curves of a certain reservoir formation.

In addition to the HDNN model, a typical MLP and CNN model are also adopted to make comparison. The MLP and CNN model share same architectures with the assembled MLP and CNN modules in HDNN. For MLP or CNN model training, single type of data is taken as input, and the corresponding network module is directly connected to the last FC layers to achieve output, without concatenate operation. For the MLP model, 11 numerical features are prepared as inputs, where structural log curves are averaged as additional seven features. During model training, same optimizer and loss function are adopted.

Figure 3 Training performance of MLP model, CNN model and HDNN model in sequence.

Figure 3 shows the training performance of MLP, CNN and HDNN model for productivity prediction. The performance of MLP model indicates that it runs into overfitting, which may be caused by low correlation of features to target label and relative deeper network layers. However, CNN model and HDNN model both show better convergence, with training error and validation error declining. Comparatively, HDNN model exhibits smoother fluctuation, better validation error descent and lower mean absolute error. That is to say, considering comprehensive mixed inputs, HDNN model performs well than typical CNN model. Following model training, test data are utilized to perform quality control. As showed in Figure 4, measured and predicted oil production from three deep neural network models are demonstrated in crossplots. Meanwhile, squared correlation coefficient ($r^2$) is also displayed. In accordance with the analyses from training performance, the HDNN model is superior to typical CNN and MLP model, demonstrating best correlation with highest $r^2$. 

82nd EAGE Conference & Exhibition 2020
8-11 June 2020, Amsterdam, The Netherlands
Figure 4 Crossplots of predicted production versus real production for MLP, CNN and HDNN model.

Conclusions

Taking mixed data as inputs, a hybrid deep neural network combined of MLP and CNN is developed to predict reservoir productivity. As a data-driven method, the proposed HDNN model overcomes the weaknesses of some physics-based methods, providing wide applicability. It neither bases on some theoretical hypotheses nor limits to specific type of reservoir. Instead of taking statistic average of log curves as input, the HDNN model takes a variety of types of data from various sources into account. The mixed data provides more aspects of features. In particular, structural log curves can be used to exploit cumulative effect within reservoir formations and are especially suitable for productivity evaluation. CNN model takes advantage of structural log curves and demonstrates high performance over MLP model. Furthermore, the proposed HDNN model combines MLP and CNN, and establishes more accurate nonlinear relationship between input data and target production. Examples highlights the superiority of HDNN model over MLP or CNN model. In addition, the proposed HDNN model realizes end-to-end learning, avoiding tedious works such as feature extraction and selection. In conclusion, HDNN model helps to perform incisive data mining and improve accuracy and generalization of productivity prediction. Furthermore, the architecture of HDNN model can be further generalized to perform other prediction tasks which providing mixed inputs.

References


