UDC 622.244.6.05:681.3.06.

DRILLING RATE OF PENETRATION PREDICTION USING ARTIFICIAL NEURAL NETWORK: A CASE STUDY OF ONE OF IRANIAN SOUTHERN OIL FIELDS

Mehran Monazami
MSc Student, Ahwaz Faculty of Petroleum, PUT, Ahwaz, Iran
m. monazami 86@gmail.com

Abdonabi Hashemi
Assistant Professor, Ahwaz Faculty of Petroleum, PUT, Ahwaz, Iran

Mehdi Shahbazian
Assistant Professor, Ahwaz Faculty of Petroleum, PUT, Ahwaz, Iran

Abstract. Rate of Penetration (ROP) estimation is a key parameter in drilling optimization, due to its crucial role in minimizing drilling operation costs. However, a lot of unforeseen factors affect the ROP and make it a complex process and consequently difficult to predict. This paper presents an application of Artificial Neural Network (ANN) methods for estimation of ROP among drilling parameters obtained from one of Iranian southern oil fields, according to the fact that this method is useful when relationships of parameters are too complicated. The method is proposed as a more effective prognostic tool than are currently available procedures. The methodology enables drilling industry personnel to estimate the ROP not only during well planning procedure but also during drilling. A three layer feed-forward network has been selected which has the best correlation coefficient in testing the models. Simulation results show that the ANN approach is superior to the conventional methods in drilling rate prediction accuracy.

Keywords: Artificial Neural Network, Rate of Penetration, Drilling efficiency, ROP Prediction.

Introduction

In the recent years, drilling optimization techniques have been used to reduce drilling operation costs. This would be done by reducing the operation time, since time is always money in drilling operations (Rahimzadeh et al. 2010). The researchers in the drilling engineering fields are always looking for the prediction of unexpected events and optimizing the related parameters. Predicting the Rate of Penetration is of a great attention
for drilling engineers due to its effect on the optimization of various parameters that leads to reduction of the costs (Bataee et al. 2011).

There is no exact mathematical relation between drilling rate and different drilling variables because not only a large number of uncertain drilling variables influence drilling rate, but also their relationship is nonlinear and complex (Mendes et al. 2007). Penetration rate is affected by many parameters such as bit hydraulics, weight on bit, rotary speed, bit type, mud properties, formation characteristics, etc. (Akgun 2007). In drilling industry, to carry out save and environmental friendly drilling operations with cost effective well construction various methods are used from different disciplines. Most important disciplines are communication and computer technologies, as it contributes in drilling optimization (Jahanbakhshi 2012). Computer technologies and programs has propelled that they can solve complex problems in few minutes.

Several ROP models have been proposed. Among these models, the most well-known ones are Bourgoyne and Young, and Warren’s models. However, they did not provide satisfactory accuracy. In each of these models different parameters have been used to estimate the ROP. In this paper, a new model based on the Neural Networks (NN) is developed using neural network fitting tool of the MATLAB programming software. Implication of this model to our data showed the proficiency of the model in comparison to the other models.

**Artificial Neural Network and It’s Architecture Design**

NNs are massively parallel–distributed processing units known as neurons. These simple neurons have certain performance characteristics in common with biological neurons. ANNs provides a non-linear mapping between inputs and outputs by its intrinsic ability (Bataee et al. 2011). The success in obtaining a reliable and effective network depends largely on the correct data preprocessing, architecture selection and network training choice strongly (García-Pedrajas et al. 2003).

The most common NN architecture is the feed-forward neural network that is a network structure in which the information will propagates in one direction, from input to output. In developing the networks among two layered, three layered, and four layered networks, the three layered showed the lowest network error. Also, different structures in three layered have been tested. Finally, as it is shown in **Figure 1**, a three layer feed-forward network with 'tansig' activation function for hidden layer and 'purelin' for output layer and full connection topology between layers is used. Back–Propagation algorithm with Levenberg–Marquardt training function has been used for training. This algorithm can approximate any nonlinear continuous function to an arbitrary accuracy (Bontempi et al. 2001).
The network is trained by performing optimization of weights for each node interconnection and bias terms; until the output values at the output layer neurons are as close as possible to the actual outputs (Wang & Ding 2003). The Mean Squared Error (MSE) of the network is defined as equation 1, which is used to show the performance of the network training.

\[
MSE = \frac{1}{2} \sum_{k=1}^{G} \sum_{j=1}^{m} [Y_j(k) - T_j(k)]^2
\]

Where \( m \) is the number of output nodes, \( G \) is the number of training samples, \( Y_j(k) \) is the expected output, and \( T_j(k) \) is the actual output.

It should be noted that a potential difficulty with the use of powerful non-linear regression methods is the possibility of over-fitting data. In the above developed model, to avoid this difficulty, the field data are divided into three sets: training subset constitutes 80% of the total data and validation and testing subsets that include 10% of the database. The training set is used to calibrate the model. The validation set is used to ensure the generalization of the developed network during the learning phase, and the testing set is used to examine the final performance of the network (Shadizadeh 2010). In the training process, the desired output in the training set is used to help the network adjust the weights between its neurons or processing elements (Mohaghegh et al. 1995; Hassoun 1995). In addition, the proper selection of the number of neurons in the hidden layer can avoid the over-fitting of neural network effectively.

**Figure 1.** Architecture of three layer ANN.
Data Acquisition and Preprocessing

A total number of 336 cases were collected from the daily drilling reports in one of the southern Iranian oil fields. The parameters that were collected as the input data are as follows: mud properties, hole geometry information, drill string & bottom hole assembly size, pipes specification, operational parameters, formation characteristics. Mud properties are Mud Weight (MW), Plastic Viscosity (PV) and Yield Point (YP) (Shadizadeh 2010). Understanding the influence of the input parameters is considered the primary concern when developing ANN models. Introducing more input parameters than required will result in a large network size and consequently decrease learning speed and efficiency (Goda et al. 2005). In this study it is tried to eliminate both dependant parameters and those that result in higher training error to decrease the number of inputs. Finally selected parameters are according Table 1.

Table 1. Selected Parameters as NN inputs

<table>
<thead>
<tr>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drill Collar Outside Diameter (in)</td>
</tr>
<tr>
<td>Drill Collar Length (m)</td>
</tr>
<tr>
<td>Kick of Point (m)</td>
</tr>
<tr>
<td>Azimuth (degree)</td>
</tr>
<tr>
<td>Inclination Angle (degree)</td>
</tr>
<tr>
<td>Weight on Bit (lb)</td>
</tr>
<tr>
<td>Flow Rate (gallon per minute)</td>
</tr>
<tr>
<td>Bit Rotation Speed (rotary per minute)</td>
</tr>
<tr>
<td>Mud Weight (lb/ft³)</td>
</tr>
<tr>
<td>Solid Percent</td>
</tr>
<tr>
<td>Plastic Viscosity (cp)</td>
</tr>
<tr>
<td>Yield Point (lb/100ft²)</td>
</tr>
<tr>
<td>Measured Depth (m)</td>
</tr>
</tbody>
</table>

We can make the networks life a lot easier by giving it data scaled in such a way that all the weights can remain in small, predictable ranges (Goda et al. 2005).

To scale the data for a particular input X, find the maximum X (X_max) for that input, the minimum X (X_min) and find the scaled value of any input X (X_s) using the following equation:

\[ X_s = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \]
Training the Network

After the data in the database spreadsheet model is divided into training, cross validation, and testing subsets, a supervised training network is used to learn among the inputs data. The neural network learns to infer the relationship between parameters by iteratively adjusting the weighting factors in a two-stage propagate/adapt cycle. In the first stage of this cycle, the input values are propagated through each layer of the network until an output is generated. Doing the training process several times using different numbers of hidden neurons, finally a network with 13, 14 and 1 neuron in input, hidden, and output layers respectively, show higher prediction precision. The weighting factors are initially random so that the input values are transformed to the output values, but with no meaningful pattern (Shippen & Scott 2002). These outputs are compared to the desired output, and then an error signal is calculated.

Results

We were evaluated the simulation performance of the ANN model on the basis of mean MSE and efficiency coefficient ‘R’. Table 2 gives the MSE and R values for the three groups of data. Performance and Training state of ANN are shown in Figures 2 and 3, respectively. Regression plot of predicted ROP against Field data is shown in Figure 4. Figure 5, 6 and 7 show the extent of the match between the measured and predicted ROP values using ANN model in terms of a scatter diagram for training, validation and testing data sets respectively. As it is shown in Figure 7, the output of the model, simulated with testing data, shows a good agreement with the target.

<table>
<thead>
<tr>
<th>Number of Data</th>
<th>MSE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.0504</td>
<td>0.98921</td>
</tr>
<tr>
<td>Validation</td>
<td>0.0601</td>
<td>0.97219</td>
</tr>
<tr>
<td>Testing</td>
<td>0.0591</td>
<td>0.98913</td>
</tr>
</tbody>
</table>
Figure 2. Performance plot of ANN.

Figure 3. Training state plot of ANN.
Figure 4. Regression plot of ANN.
Figure 5. Training Data-Comparison between field and predicted ROP

Figure 6. Validation Data-Comparison between field and predicted ROP
Figure 7. Testing Data-Comparison between field and predicted ROP

Conclusions

- High accuracy for ROP prediction using selected parameters illustrates that they have essential role in drilling operation efficiency.
- The performance of the network model depends largely on the size and accuracy of the database and the variables selected for the analysis. The more the numbers of data the high will be the reliable and efficient result.
- The ROP performance predictions chart indicated that proper and accurate prediction of rate of penetration was accomplished.
- The application of different drilling parameters data set collection is suggested in field of study in order to find the most consistent model.
- The application of neural network for prediction of ROP showed more confident results since it takes into account the effect of more drilling parameters.
- To increase the extension of model, it is necessary to use data from different oil field.
List of Symbols

ANN Artificial Neural Network
ROP Rate of Penetration
NN Neural Network
PV Plastic Viscosity
YP Yield Point
MW Mud Weight

References


