Data Analysis for Oil Production Prediction

Christine W. Chan
University of Regina, Canada

Hanh H. Nguygen
University of Regina, Canada

Xiongmin Li
University of Regina, Canada

INTRODUCTION

An economic evaluation of a new oil well is often required, and this evaluation depends heavily on how accurately production of the well can be estimated. Unfortunately, this kind of prediction is extremely difficult because of complex subsurface conditions of reservoirs. The industrial standard approach is to use either curve-fitting methods or complex and time-consuming reservoir simulations. In this study, we attempted to improve upon the standard techniques by using a variety of neural network and data mining approaches. The approaches differ in terms of prediction model, data division strategy, method, tool used for implementation, and the interpretability of the models. The objective is to make use of the large amount of data readily available from private companies and public sources to enhance understanding of the petroleum production prediction task. Additional objectives include optimizing timing for initiation of advanced recovery processes and identifying candidate wells for production or injection.

BACKGROUND

The production of an oil well is influenced by a variety of factors, many of which are unknown and unpredictable. Core logs, drill steam test (DST), and seismic data can provide geological information about the surrounding area; however, this information alone cannot explain all the characteristics about the entire reservoir. While core analysts and reservoir engineers can analyze and interpret geoscience data from several wells with the help of numerical reservoir simulations, the process is technically difficult, time consuming and expensive in terms of both labor and computational resources.

For a quick estimation of petroleum production only, the decline curve analysis remains the most popular methods among engineers (Baker et al., 1998; Li, Horne, 2003). However, a weakness with the decline curve analysis technique is that it is difficult to foresee which equation can adequately describe production of a reservoir. Moreover, a single curve is often inadequate for describing production data generated during the entire life of the reservoir. Therefore, fitting production data to a decline curve is a difficult process and can result in unreliable predictions (El-Banbi, Wattenbarger, 1996).

To overcome the weakness of conventional decline curve analysis techniques, we adopted data mining techniques for the task of modeling non-linear production data. We compared the use of neural networks versus curve estimation technique in the prediction of oil production (Nguyen et al, 2003) and found that with sufficient training data, the neural network approach can perform better on unknown data. In our exploration on modeling production data using artificial intelligence techniques, we introduced variations along six dimensions, which will be discussed in the next section.

MAIN FOCUS: VARIATIONS IN NEURAL NETWORK MODELING APPROACHES

Prediction Model

The first step of our research was to identify the variables involved in the petroleum production prediction task. The first modeling effort included both production time series and geoscience parameters as input variables for the model. Eight factors that influence production were identified; however, since only data for the three parameters of permeability, porosity, and first shut-in pressure were available, the three parameters were included in the model. The production rates of the three months prior to the target prediction month were also included as input variables. The number of hidden units was determined by trial and error. After training the neural network, a sensitivity test was conducted to measure the impact of each input variable on the output. The results showed that all the geoscience variables had limited (less than 5%) influence on the production prediction.

Therefore, the second modeling effort relied on a model that consists of time series data only. The training and testing error was only slightly different from those of the first model. Hence, we concluded that it is reasonable to omit the geoscience variables from our model. More details on the modeling efforts can be found in (Nguyen et al., 2004).

Data Manipulation Strategy

Since different ways of data preprocessing can influence model accuracy and usage, we investigated the following three approaches for processing the monthly oil production data (Nguyen and Chan, 2004 b): (1) **sequential**, when data from individual wells were arranged sequentially, (2) **averaging**, when data from individual wells were averaged over their lifetimes, and (3) **individual**, when data from each individual well were treated independently. Two types of models were used: one with past productions as input and another with time indices as input.

The results showed that with production volumes as input, the average and the individual approaches suffered from forecast inaccuracy when training data was insufficient: the resulting neural networks only performed well when the training data covered around 20 years of production. This posed a problem since many wells in reality cannot last that long, and it is likely that intervention measures such as water flooding would have been introduced to the well before that length of time has elapsed. Hence, these two approaches were not suitable for prediction at the initial stage of the well. On the other hand, the sequential approach used a large number of training samples from both early and late stages, therefore the resulting models worked better when provided with unseen data.

The neural networks with production months as input usually required less data for training. However, it was also observed that long-life wells generated smoother decline curves. We believe it is recommendable to use all three data preprocessing approaches, i.e. sequential, averaging and individual, when the time index is used. The individual approach can provide reference points for new infill wells while the other two approaches give some estimation of what would happen to a typical well in the study area.

Multiple Neural Networks

Several researchers have attempted to use multiple neural networks to improve model accuracy. Hashem et al. (1994) proposed using optimal linear combinations of a number of trained neural networks in order to integrate the knowledge required by the component networks. Cho and Kim (1995) presented a method using fuzzy integral to combine multiple neural networks for classification problems. Lee (1996) introduced a multiple neural network approach in which each network handles a different subset of the input data. In addition to a main processing neural network, Kadaba et al. (1989) used data-compressing neural networks to decrease the input and output cardinalities.

The multiple-neural-network (MNN) approach proposed in (Nguyen and Chan, 2004 a) aims to improve upon the classical recursive one-step-ahead neural network approach. The objective of the new approach is to reduce the number of recursions needed to reach the lead time. A MNN model is a group of neural networks working together to perform a task. Each neural network was developed to predict a different time period ahead, and the prediction terms increased at a binary exponential rate. A neural network that predict 2^n step ahead is called an n-ordered neural network.

The choice of binary exponential was made due to two reasons. First, big gaps between two consecutive neural networks are not desirable. Forecasting is a process of reducing the lead time to zero and smaller. The smaller the gaps are, the fewer steps the model needs to take in order to make a forecast. Secondly, binary exponential does not introduce bias on the roles of networks while a higher exponential puts a burden onto the lower-ordered neural networks.

To make a prediction, the neural network with the highest possible order is used first. For example, to predict 7 units ahead (x_{t+7}), a 2-ordered neural network is used first. It then calculates temporary variables backward using lower-ordered neural networks.

Related Content