CS230

First Application of Oil Production Forecasting Using Transfer Learning of Pre-Trained Deep N-BEATS Model

Zainab Hussain Alali

Department of Energy Resources Engineering
zainab93@stanford.edu

Abstract

Fluid rate measurement and forecasting in the oil and gas industry is crucial for the field development. The goal of this project is to develop a virtual-based flow meter to forecast the oil rate of two wells from the Norwegian Norne field with limited history (500 time steps), using deep learning models. This paper introduces a novel idea showing the first time implementation of transfer learning and pre-trained deep Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (N-BEATS) model in oil production forecasting. Prior work was limited on applying feature-based linear regression algorithms and traditional sequence deep learning models, mainly RNNs, LSTMs, and Long Short Time Series Network (LSTnet), to predict the pressure response of a single well with a single fluid. In this paper we extended the application of deep learning research by introducing two new methods, an attention-based model using the Temporal Fusion Transformer (TFT) and a transfer learning approach using pre-trained N-BEATS on M4 series. Both TFT and the pre-trained N-BEATS models outperformed traditional LSTM model, increasing the test score (MAE) by 0.04 and 0.08, respectively, demonstrating excellent matching results by N-BEATS. We concluded that using transfer learning and pre-trained N-BEATS model eliminates the previous disadvantages of LSTM models requiring multivariate features and a large training history. This research shows auspicious results for using transfer learning and N-BEATs model in the petroleum industry, especially for new or green fields with limited historical data.

1 Introduction

In the oil and gas industry, fluid rates prediction is critical to ascertain future well planning and field development requirements. Nowadays, many operators are digitizing their oil fields by installing downhole and surface sensors on individual wells, providing continuous and real-time measurement of rates and pressure. These sensors are typically expensive costing around 290,000$ including maintenance cost which can be avoided using virtual-based flow meters via machine learning. Existing research on this subject shows implementation of feature-based linear regression algorithms and deep learning architecture, mainly RNNs, LSTMs and LSTnet, to predict the pressure response of a single well with a single fluid [1], [2], [3], [4]. In this project, we introduce two new deep models for time series forecasting, an attention-based TFT model and a transfer learning approach using a pre-trained N-BEATS model on M4 series to predict oil rate profiles in a multi-well reservoir system.

2 Dataset

The dataset compiled for this project is adopted from an offshore field in the Norwegian sea, called the Norne field. Production data of the Norne field and its associated simulation model are published for public research. The dataset mainly includes multivariate time series data of wells' rates and pressure. The historical field data of two wells, containing 500 timesteps, were extracted from the simulation model. In this research, we aimed to develop a model to predict the wells' oil rates. Other time series data including, pressure and water rate were used as past features. A sliding window size of 31 timesteps was used, consisting of 30 past timesteps as input features and one future timestep to predict the target oil rate. The dataset was split into 85% for training, 432 historical time steps, and 15% for testing, 76 future time steps. The dataset was scaled prior to training to equalize features' weights and improve convergence of the deep learning models

3 Methods

In this work, we mainly relied on applicable deep learning models designed to handle the sequence dependency of time series data. In our first application, we trained the dataset using the LSTM model to predict the wells' oil rates [3]. One of the main challenges of this project is the limited history of the field having a small dataset size and limited features. The target oil rate data was also complex with no clear trend or seasonality. Hence, it was rather challenging to fit the LSTM model requiring enormous iterations and complex hyperparameters' tuning. Even with that being done, the model might not generalize well for the test dataset. The second model we used was the Temporal Fusion Transformer (TFT). The TFT model integrates the mechanisms used in LSTM layers, the attention heads in transformers and the Gated Residual Network (GRN) to learn the relationship along the time axis. The TFT model applied on the dataset outperformed the LSTM model, however, the prediction has not reached the desired level yet. To overcome these challenges, we adopted a new transfer or meta learning approach for time series prediction. We utilized a recently developed deep learning model called N-BEATS, Neural Basis Expansion Analysis for Interpretable Time Series Forecasting, to pre-train it on the large M4 time series dataset and use it to predict the target oil rate in our testing set.

3.1 LSTM Model

LSTM models are capable to learn long-term dependencies and perform multivariate time series prediction. BlockRNN architecture was built utilizing the Block Recurrent Neural Network model (BlockRNN) provided by the Darts and Pytorch library. After numerous iterations, our best BlockRNN model comprised of 2 LSTM layers with 1000 units followed by three fully connected layers of sizes 512,512 and 1024, respectively. A batch size of 96 was used, refer to Table 1 summarizing the model's parameters. Adam optimizer was used with a learning rate of $\mathbf{10^{-4}}$ to minimize the Mean Squared Error (MSE) loss function. The model was set for a target iteration of 500 epochs while adopting an early stopping technique to monitor the losses with a patience of 50 epochs. During hyperparameter tuning, it was found that larger model architectures with small learning rates and larger batch size are required to minimize the loss function.

Table 1:Key parameters of BlockRNN model

Parameters	**Value**
Number of LSTM layers	2
Number of units in LSTM	1000
Learning rate	0.00001
Number of Fully connected layers (FC)	3
Number of neurons in FC layers	512, 512, 1024
Epochs	85
Batch size	96
Optimizer	Adam
Loss function	MSE

3.2 Temporal Fusion Transformer (TFT)

Temporal fusion transformer is an attention-based deep learning model for time series forecasting. The building block for TFT consists of a Gated Residual Network (GRN) comprises of two dense layers and two activation functions, Exponential Linear Unit (ELU) and gated linear unit (GLU), allowing for both skip connections and gating for efficient information flow. It also contains a Variable Selection Network (VSN) for selecting the most relevant features at each time step. Time-dependent processing is based on, an LSTM encoder-decoder for local processing and a self-attention layer for learning long range dependencies across different time steps [5]. The architecture used in this project was built utilizing the Darts and Pytorch library, as showing in Figure 1. The model comprised of a single LSTM layer , 64 hidden neurons of each dense layer of the GRN and a single self- attention layer, refer to Table 2. The TFT is trained by minimizing the quantile loss function summed across a certain quantile, (q) ∈ [0.1, 0.5, 0.9] , refer to Equation 1. A batch size of 96 and Adam optimizer were used with a learning rate of $\mathbf{10^{-4}}$ to minimize the quantile loss function centered around 0.5. We followed the same training procedure used in the previous LSTM model for monitoring the loss applying early stopping technique. Using grid search, it was found that less attention head layers minimizes the loss function.

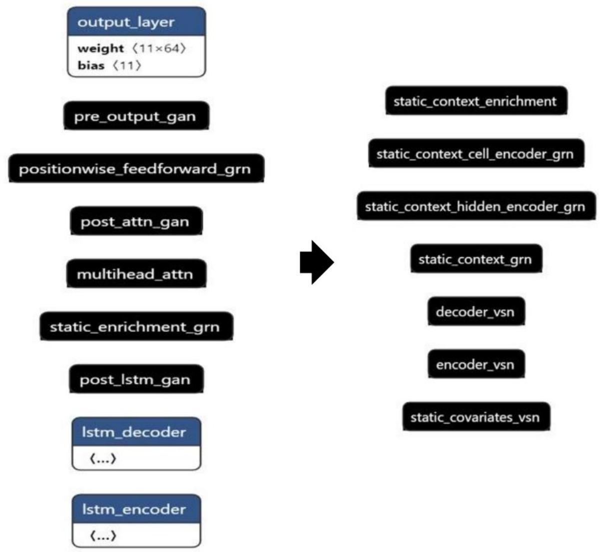

Figure 1: TFT model architecture

Table 2:Key parameters of TFT model

Parameters	Value
Number of LSTM layers	1
Number of neurons in the hidden layer	64
Learning rate	0.0001
Number of attention heads	1
Epochs	135
Batch size	96
Optimizer	Adam
Loss function	Quantile (0.5)

$$QL(y, \hat{y}, q) = \max\left[q\left((y - \hat{y}), (1 - q)(y - \hat{y})\right]\right.$$ Equation 1

3.3 Transfer Learning Using Pre-Trained N-BEATS on M4 Dataset

N-BEATS is a new state of the art deep neural model developed in 2019 to handle time series prediction tasks. The basic building block of N-BEATS consists of a multi-layer fully connected network with ReLU activation function. The block predicts two expansion coefficients, forward (forecast) and backward (backcast). The blocks are organized into stacks using doubly residual stacking principle. Partial forecasts predicted by each block are hierarchically aggregated first at the stack level and then at the overall network level [6]. The N-BEATS algorithm outperformed all existing deep learning models when applied on the M4 dataset improving forecasting accuracy by 11%. The M4 is a large and highly heterogeneous dataset containing a collection of 100,000 time series data from business, financial and economic forecasting problems. The N-BEATS model was pre-trained on the M-4 series using the hyperparameters listed in with mainly 2 layers and 20 stacks utilizing a gpu machine, as adapted from [7] and [8]. The Symmetric Mean Absolute Percentage Error (SMAPE) was used as a loss function during training,Table 3. A transfer or meta learning approach was used to predict the oil rate series of our testing dataset via sharing the N-BEATS model parameters and weights learned from the M4 series. Figure 2 shows our adopted transfer learning modeling approach.

Table 3: Key parameters of N-BEATS model

Parameters	**Value**
Number of stacks	20
Number of blocks	1
Number of layers	2
Layer width	136
Expansion coefficient	11
Learning rate	0.0001
Optimizer	Adam
Loss function	SMAPE

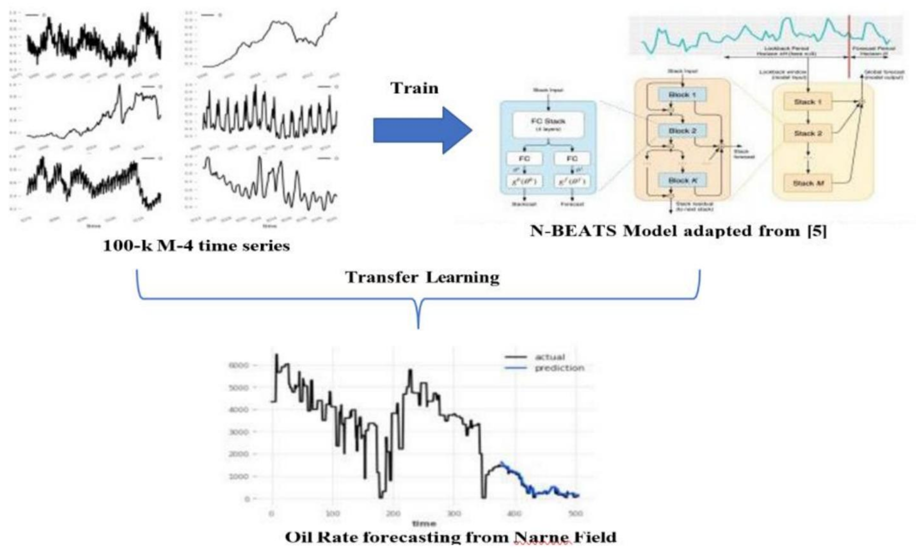

Figure 2:N-BEATS model architecture

$$SMAPE = \frac{100\%}{n} \sum \frac{2 \times |y - \hat{y}|}{|y| + |\hat{y}|} \qquad \text{Equation 2}$$

4 Result and Discussion

With the tuned hyperparameters currently achieved for the BlockRNN ,TFT and N-BEATS models, the N-BEATS model outperformed both the BlockRNN and TFT models. Figure 3 shows the prediction results of well 1 and well 2 for all models for the testing dataset. Table 3 also benchmarks the achieved MSE and RMSE for all models. N-BEATS model was able to achieve excellent match of the forecasted oil rates from the test data for well-1 and well-2. This transfer learning approach from the abundant M4 time series data was able to overcome the limitation of small dataset size yielding successful prediction without supplying the model with any covariate features.

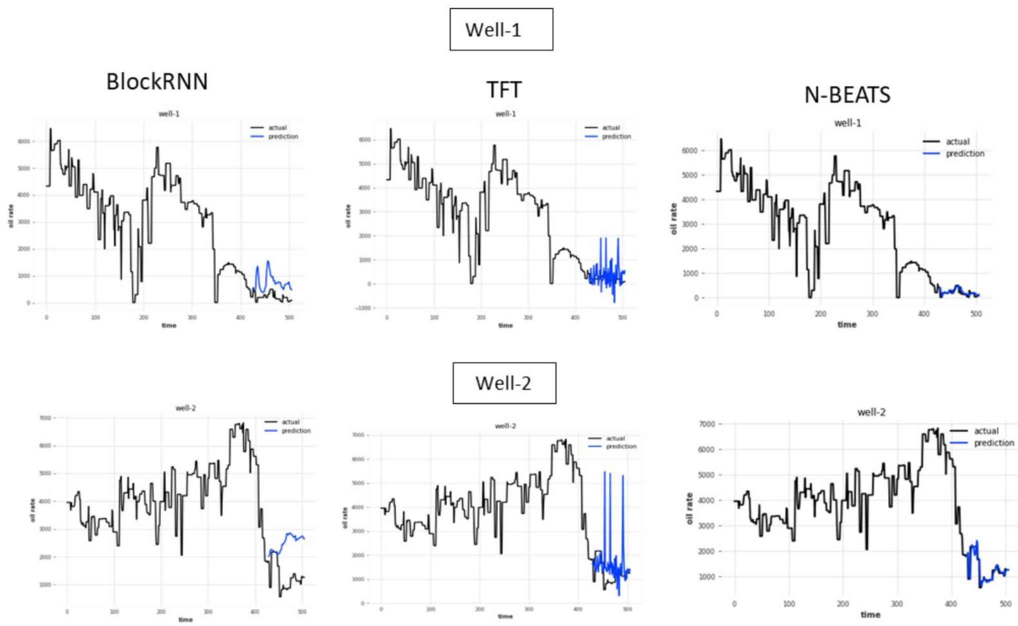

Figure 3:Oil rate prediction using three models BlockRNN, TFT and N-BEATS

Table 4: Testing scores comparison of all models for well-1

	well-1		
	BlockRNN	TFT	N-BEATS
MAE	0.09	0.06	0.01
RMSE	0.1	0.09	0.02

Table 5: Testing scores comparison of all models for well-2

	well-2		
	BlockRNN	TFT	N-BEATS
MAE	0.22	0.08	0.03
RMSE	0.24	0.1	0.04

3 Future Remaining work

In this paper, we used a single-time step prediction during forecasting. For future work, we will use the N-BEATS model for multi-time steps prediction using two techniques, a single-shot and autoregression. Forecasting multi-time steps in the future will be a game changer in the oil and gas industry because it will allow for predicting multi-years performance of new fields with limited data and provides a reliable replacement for costly downhole sensors in digital fields.

4 Contribution

This study was performed by Zainab Al-Ali .Many thanks to the project mentor Vincent Li for the support and guidance, and for Professor Andrew Ng and Kian Katanforoosh for teaching the deep learning course.

7 References

[1] Tian, Chuan , and Roland N. Horne. "Applying Machine-Learning Techniques To Interpret Flow- Rate, Pressure, and Temperature Data From Permanent Downhole Gauges." SPE Res Eval & Eng 22 (2019): 386–401. doi: https://doi.org/10.2118/174034-PA

[2] Tian, Chuan , and Roland N. Horne. "Recurrent Neural Networks for Permanent Downhole Gauge Data Analysis." Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, October 2017. doi: https://doi.org/10.2118/187181- MS

[3] Li, Yuanjun , Sun, Ruixiao , and Roland Horne. "Deep Learning for Well Data History Analysis." Paper presented at the SPE Annual Technical Conference and Exhibition, Calgary, Alberta, Canada, September 2019. doi: https://doi.org/10.2118/196011-MS.

[4] Orta Alemán, Dante , and Roland Horne. "Improved Robustness In Long-term Pressure Data Analysis Using Wavelets and Deep Learning." Paper presented at the SPE Annual Technical Conference and Exhibition, Virtual, October 2020. doi: https://doi.org/10.2118/201597-M

[5] Bryan et al., "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting" ,Sep,2020, 1912.09363.pdf (arxiv.org)

[6] Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019, September). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting.

[7] Oreshkin et al., "Meta-learning framework with applications to zero-shot time-series forecasting", 2020, https://arxiv.org/abs/2002.02887

[8] Herzen, J et al., " Transfer Learning For Time Series Forecasting", May 2022, https://unit8.com/resources/transfer-learning-for-time-series-for