Deep Probabilistic Forecasting with Monte-Carlo Dropout in Neural Networks

Novin Shahroudi
Institute of Computer Science, University of Tartu

Forecasting is an act of predicting the future that comes with a degree of uncertainty. That is why a probabilistic expression of the future is more robust and uncertain nature better than other forms. Deep Learning is taking a prevalent role in today’s industries. It is being employed in more mission-critical tasks such as self-driving cars as well as tasks that may not require a safety measure, yet require some measure of reliability such as in energy power forecasting. Taking uncertainty into account in these models is a key enabler to account for safety and reliability in these models. This poster demonstrates some of the experiments performed for my master thesis to perform probabilistic forecasting using a type of Stochastic/Neural Approximation for neural networks. Reliable forecasting models play a crucial role in operations of electricity grid. Wind power is expected to provide at least 35% of the electricity production in Estonia and Sweden in the upcoming decade. Therefore, this work was motivated to be applied on Wind Power Forecasting.

Background

Time-series Forecasting

In the Time series forecasting, a history of data up to time is given and a prediction should be made for the future for a given number of steps. Data points are dependent on each other and may not be identified equally. Each step in the forecast referred to as a lead time, and the whole forecast steps referred to as horizon. Equation (1) denotes the observed values of a time series from the past to the present and the future.

\[ y_0, \ldots, y_t, y_{t+1}, \ldots, y_{t+n} \in \mathbb{N} \]  

(1)

Additional features could be incorporated in univariate time series forecasting that it is referred to as exogenous variables.

Bayesian Deep Learning

A Bayesian neural network (BNN) [5] considers a distribution over its parameters as well as its outputs as opposed to a generic Artificial Neural Network (ANN) which may be referred to as non-Bayesian Neural Networks. Calculating the posterior of a Bayesian model is intractable for neural networks due to hummingous parameters space even with relatively small networks. In [7], authors introduce Monte-Carlo Dropout (MCD) network as an approximation of BNN that gives reasonably good results to approximate a Bayesian Neural Network.

Probabilistic forecasting

Probabilistic or Dense Forecasting is the most informative form of forecast that estimates a probability distribution rather than a point forecast for each time step. It is common to demonstrate a probabilistic forecast using fan-charts, each fan representing a quantile of the probability distribution as depicted in Fig. 5.

Evaluation metrics

CRPS [7] evaluates quality of a probabilistic forecast represented by a predictive distribution function F(y|t) against the true value or observed probability passed through a heaviside function H(y-g) as shown in Eq. (2). An important property of CRPS is that it measures the accuracy, as well as the sharpness of the predictions which measures calibratedness of the forecasts as is also reflected in the score. When CRPS normalized, it is denoted as NCRPS.

\[ \text{CRPS}(y) = -\frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left( F(y|t) - H(y-g) \right) ^2 \, dy \]  

(2)

\[ \text{Mean Squared Error (MSE)} = \frac{1}{n} \sum_{i=1}^{n} (y_{t+i} - y_{t+i}) ^2 \]  

(3)

Data

For the experiments with Synthetic data, the Timesynth was used to generate and sample from a sinusoidal signal with frequency. For the experiments with real data, wind power datasets from Global Energy Forecasting Competition 2014 [5] was used. Setup of the data performed according to [4]. Two sets of experiments conducted on this dataset, a univariate forecasting, and univariate forecasting with 4 wind speed features as other predictors.

Baseline Models

Naive or linear forecast is a well-known time-series forecasting baseline. It uses the last observation as the forecast.

Methods

Probabilistic Baseline

Naive method extended to provide probabilistic baseline where a given quantile of the training data is being used as the forecast.

Models

Methods explained in this section were mainly inspired by [6]. A Multi-layer Perceptron (MLP), and a Recurrent Neural Network (RNN) with CRPS and NCRPS were used. Both models provided two outputs, one estimating the mean and the other the variance of the forecast. 100 units were used with a dropout rate of 0.5 for both models. For the RNN, recurrent dropout was used. The batch size for all experiments set to 32. All models were run for 10 epochs with Adam optimizer and Cyclical Learning Rate Schedul uler [8]. Activation function for the first layer set to linear for the MLP and tanh for the RNN. Activation of the first dense output that estimate the mean set to hard sigmoid, and the second dense output that estimates the logarithm of the variance set to linear.

\[ L(y) = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( y_i - \hat{y}_i \right) ^2 + \log(\sigma(\hat{y}_i)) \right] ^2 \]  

(4)

Uncertainty in the model also referred to as Epistemic uncertainty. It can be caused by model misspecification (a wrong parametric family), error in the estimation of the model’s parameters when modeling, exposing the model to novel examples and any other systematic error in the model. This can be estimated by Dropout at test time. It produces different possible outputs that can be thought to be produced by different sub-networks which would imply the variance of the model. A neural network can also be thought of as an ensemble method for reducing the variance of prediction by using a number of unbiased models.

Scenario Forecasting

The roll-forward forecasting produces biased results as a result of the model’s ignorance caused by treating forecast values as new observations. To tackle this issue scenario forecasting was used which also enabled the model to produce different forecast trajectories as another informative form of forecasting. Since the forecasts were probabilistic, one could sample from the forecast distribution and the sampled values could be used instead as the next value in the input sequence forecasting for all possible alternative futures.

Results

All results obtained by evaluation on 20 different splits of the validation set. Input size for the Synthetic data was 21 and 4 for the GEPCom’14 with forecast horizon of size 48. Results of 5 variation of each model per each architecture are being reported here. 1 of which corresponded to the Bayesian models and to a non-Bayesian model (ReLU). The Bayesian models are denoted as ALL for Aleatoric, the EP for Epistemic, the ALEP for Aleatoric+Epistemic, and SFELP for the Scenarios+Epistemic. Finally, the probabilistic baseline denoted by QLaiue.

Conclusion

Successful application of the method proposed in [6] for probabilistic forecasting. A naive baseline proposed by us which helps with diagnosis and it is a good baseline for evaluation of models calibration. The proposed model, SFELP, improved the results both from the perspective of performance, and calibration. RNN was more effective especially in the case of the real dataset with multiple variables involved in the forecasting.

References


Acknowledgements

I would like to wholeheartedly acknowledge my master thesis supervisors Prof. Dr. Miroslav Halilović of the Institute of Computer Science at University of Zira, Erik Ylip of the Swedish Institute of Computer Science at Research Institutes of Sweden (SICS), and Sjurhild Hard Gift of Greenworks A.B. I am also grateful for the encouraging advice and support from Sofiya Demchuk.