Shay Gury , Max Romnov , Noy Buganim

Professor Name : Gleb Ivashkevich

A Multi-Horizon Quantile Recurrent Forecaster

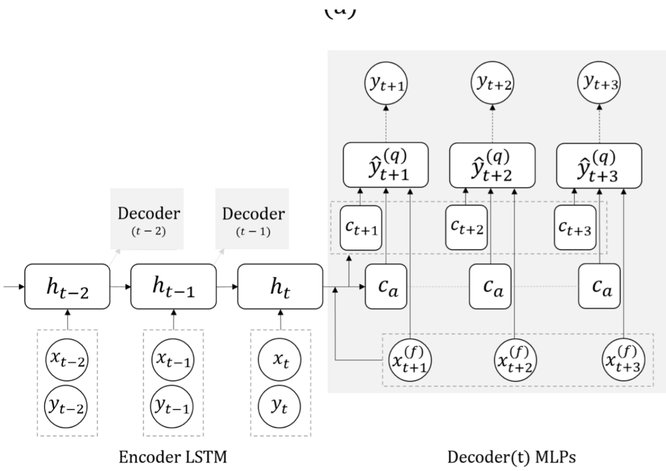
03 June 2022

Paper Review

The authors of this article propose a framework for general probabilistic multi-step time series regression - They propose MQ-R(C)NN: a Seq2Seq framework that generates Multi-horizon Quantile forecasts. The model is designed to solve the large scale time series regression problem, with an RNN encoder , and two steps linear decoder (global and local).

The model demonstrates an efficient training scheme for the combination of sequential neural nets and Multi-Horizon forecast. The approach improves stability and performance of encoder-decoder style recurrent nets by training on all time points where a forecast would be created, in a one pass over the data series, and a network that accommodates a previously little-attended issue: how to account for known future information, including the alignment of shifting seasonality and known events that cause large spikes and dips.

Usually recurrent models have the disadvantage that forecasted value is used as an input for the next generation, and this strategy allows prediction errors to accumulate and are faster to diverge, but the nature of this model combines 2 models at each step. Every point in history is encoded by an LSTM model, and decoded in two steps - one being a global look of the data (and adding the global context), and the other step is a local MLP that accounts for the next horizon. By minimizing the loss of the quantile of the predictive distribution (in different quantile percent of the probabilistic distribution) it gets the horizon context in regards to the full time series by using the combined context that produces a more accurate train to reduce the loss.



$$(c_{t+1}, \cdots, c_{t+K}, c_a) = m_G(h_t, x_{t:}^{(f)})$$

The first (*global*) MLP summarize the encoder output plus all future inputs into two contexts: a series of horizon-specific contexts ct+k for each of the K horizon points, and a horizon-agnostic context ca which captures common information:

$$(\hat{y}_{t+k}^{(q_1)}, \cdots, \hat{y}_{t+k}^{(q_Q)}) = m_L(c_{t+k}, c_a, x_{t+k}^{(f)})$$

The second (local) MLP applies to each horizon (k) and each quantile chosen (q) it combines future inputs (season or future events) . and the contexts from the global MLP

Both global and local decoder are nn.linear models with activation functions (local Relu)
Hyperparameters that can be chosen -

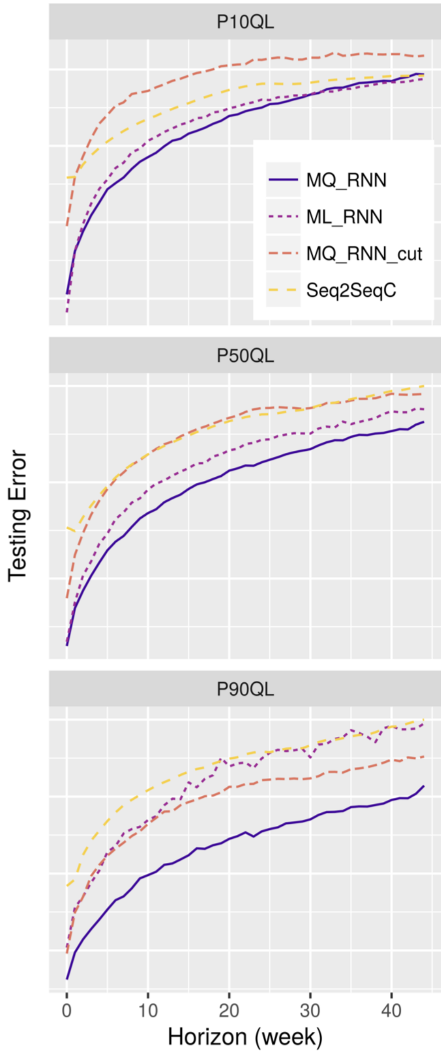
- Horizon size to forecast - days to the future
- Quantiles to forecast - probabilistic quantile data % (0.5 is MAPE)
- Context size - controls how far in the past the network can see
- Learning rate , batch size , num_epochs , hidden size , layer_size

<u>Setup :</u>In order to build the proposed model in the article , the authors consider the 3 next inputs:

1. temporal covariates available in history. $x^{(h)}$
2. knowledge about the future .$x^{(f)}$
3. static, time-invariant features. $x^{(s)}$)

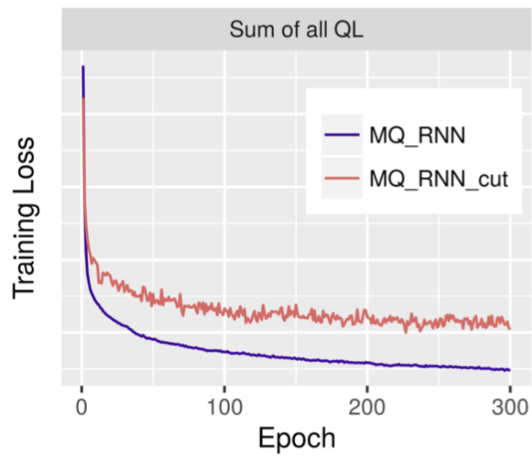
Each series is considered as one sample fed into a single RNN or CNN, even if they correspond to different items.

$p(y_{t+k,i}, \cdots, y_{t+1,i} | y_{:t,i}, x^{(h)}, x^{(f)}, x^{(s)})$



comparing result - in demanding forecasting problem Amazon

- Seq2SeqC Model - LSTM encoder/Decoder
- ML - RNN - using Log gaussian Likelihood (instead of quantile loss)
- MQ-RNN_cut - randomly cut on Forecast creation time (FCT)
- MQ-RNN - suggested article model based on quantile loss
  - P10 , P50 , P90 quantile (can be chosen as HyperParameter)



The cut MQ_RNN compared to MQ_RNN_cut training loss is better and faster since it’s forking to actual horizon the relationship of the decoders and not using random FCT effects better on the model .

The choice of encoder is not restricted to recurrent LSTM networks. Any neural net that has sequential or temporal structure and is compatible with forking-sequences, can serve as an encoder in the MQ-framework. Food example -

- NARX RNN (DiP- ietro et al, 2017), which computes hidden state h_t not only based on h_{t-1}, but also a specific set of other past states, e.g. $(h_{t-2}, \cdots, h_{t-D})$. This is also known as *skip-connections*. The presence of past states reduces the requirement on RNN cell’s ability to memorize long dependencies.(MQ-RNN_narx)
- WaveNet - all hidden states are from dilated causal convolutions.(MQ-CNN_wave)
- Lag-series - feed past series as lagged feature inputs, along with yt, into the recurrent layer at t. (MQ-RNN_lag)