Dual-stage Attention-based recurrent Neural Network for Time Series Prediction.

In this work we address the Nonlinear autoregression exogenous (NARX) problem, where we try to predict the current value of time series based upon its previous value as well as the current and the past values of multiple driving series. The approach for this predictive modeling exercise is based on Dual stage attention-based recurrent neural network (DA-RNN) technique. The technique is developed by Qin, et al. 2017 at the university of California San Diego. (see the attached paper).

DA-RNN:

The Dual stage attention-based recurrent neural network (DA-RNN) is an NARX (nonlinear autoregressive exogenous) model designed to:

- Capture the long-term temporal dependencies;
- Select the relevant driving series to make prediction.

The technique consists of two major stages:

Stage 1: it introduces an input attention mechanism to adaptively extract relevant driving series at each time step by referring to the previous encoder hidden state.

Stage 2: it uses temporal attention mechanism to select relevant encoder hidden states across all time steps.

Problem: Brief description:

Given the previous values of the target series $(y_1, y_2, y_3, \dots, y_{t-1})$ with $y_{t-1} \in \mathrm{IR}$

As well as the current and the past value of n driving (exogenous) series $(X_1, X_2, X_3, \dots, X_t)$ with $X_t \in \mathrm{IR}^n$

The NARX model aims to learn a non-linear mapping to the current value of the target series

$$\hat{y}_t = F(y_1, y_2, y_3, \dots, y_{t-1}, X_1, X_2, X_3, \dots, X_t)$$

Where F is the mapping nonlinear function to learn.

LSTM/GRU for solving the problem:

Recurrent Neural Network is a type of deep neural network specially designed for sequence modeling; however, it suffers from vanishing/exploding gradient and consequently it does not capture the long term dependencies. LSTM / GRU have overcome these limitations.

LSTM/GRU units Encoder-Decoder neural networks are popular for sequence modeling and have shown some success in machine translation. The idea is based on encoding the source sentence as fixed -length vector and use the decoder to generate a translation. However, the performance of these encoder decoder networks deteriorates rapidly as the length of the sequence increases. These has provided the bases to build attention-based encoder-decoder network that employs an attention mechanism to select part of the hidden states across all the time steps.

For time series prediction, we use two-stages attention based Recurrent Neural Network DA-RNN. In this workflow, the first attention mechanism aims at extracting the relevant driving series at each time step by referring to the previous encoder hidden state. The second attention mechanism is a temporal attention that aims at selecting relevant encoder hidden states across all time steps. These attention mechanisms can select the most relevant input features as well as capture the long term temporal dependencies of time series appropriately.

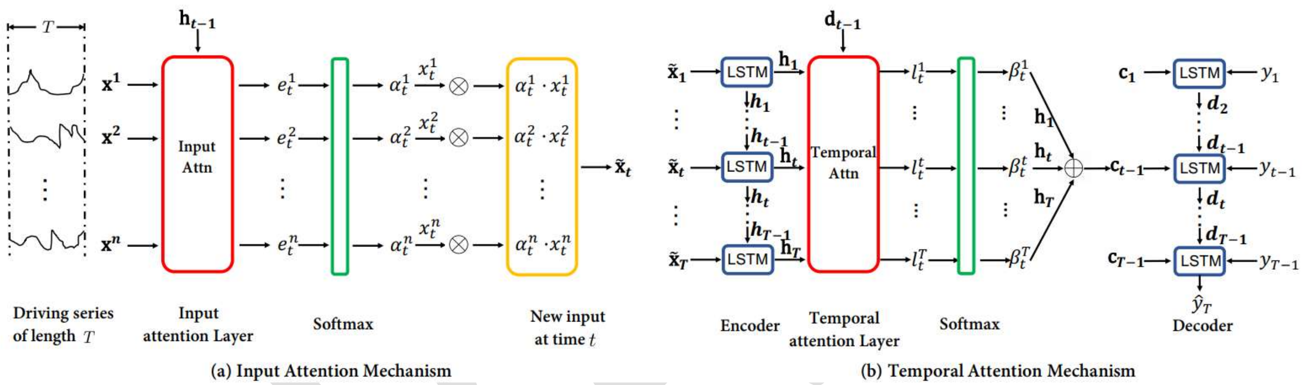

Figure 1:Illustration of dual-stage attention-based recurrent neural network. (a) The input attention mechanism computes the attention weights α_t^k for multiple driving series $\{x^1, x^2, x^3, \ldots, x^n\}$ conditioned on the previous hidden state h_{t-1} in the encoder and then feeds the newly computed $\hat{x}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \ldots, \alpha_t^n x_t^n)^T$ into the encoder LSTM unit. (b) the temporal attention system computes the attention weights β_t^t based on the previous decoder hidden state d_{t-1} and represents the input information as a weighted sum of the encoder hidden states across all the time steps. The generated context vector c_t is then used as an input to the decoder LSTM unit. The output $\hat{y}_T$ of the last decoder LSTM unit is the predicted result.