

Attention LSTM

Forecasting flight passengers

Sina Sen 4821629

Problem

The introduction of Attention Long Short Term Memory (ALSTM) produces an alternative to Long Short Term Memory (LSTM) by aiming to optimize information passing via removing the complexity of the cells in LSTM. In this work, the results and comparison of the performance of LSTM algorithms versus ALSTM architectures is assessed through the forecast of airline passenger numbers over months. The results are analyzed through qualitative and quantitative data. The hypothesis made in this paper is that ALSTM will perform better than the LSTM, and the results shows that the hypothesis was correct under some circumstances, although by a small margin.

Background Information:

Recurrent Neural Network:

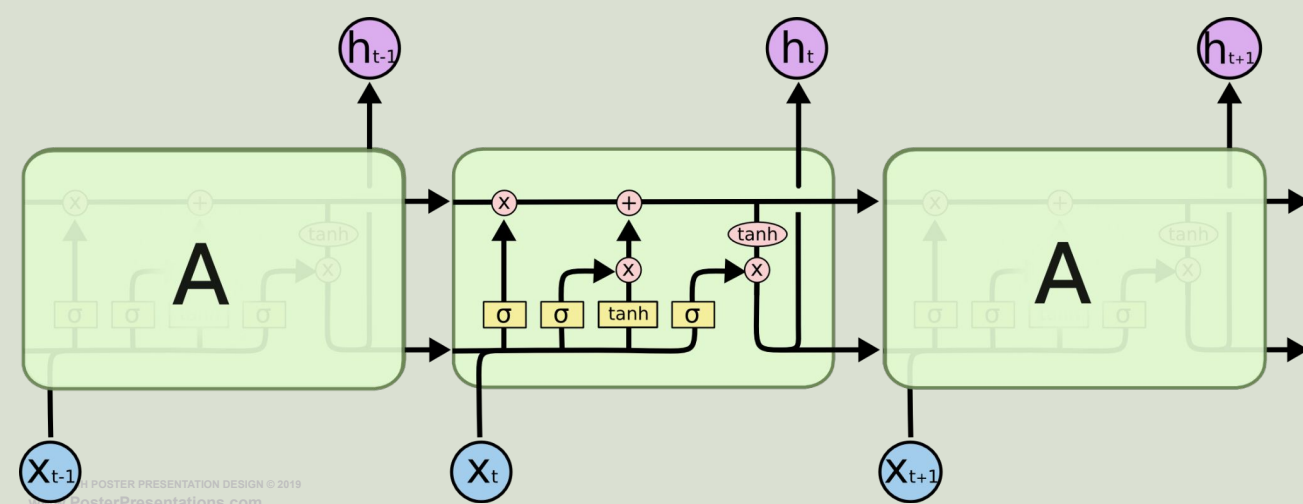
- Forecasting the future on the short term.
- Can neglect the context behind the input and the information passed through long sequences and long term dependencies.

Long Short Term Memory:

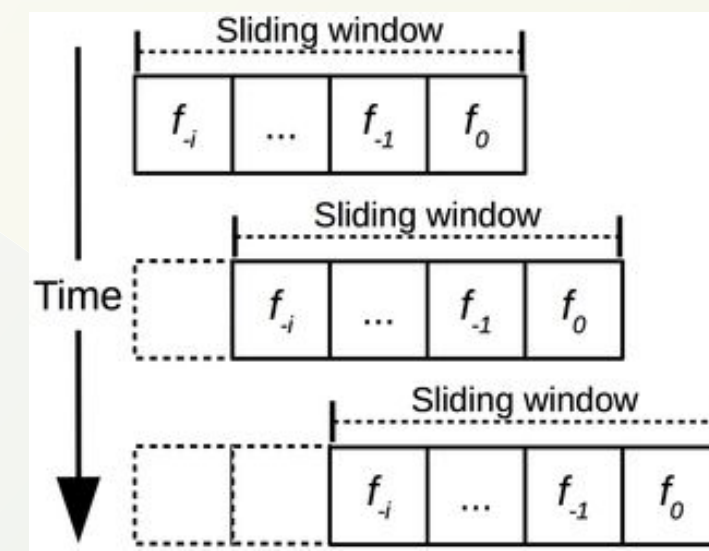
- Built to improve the performance and solve the problems of RNNs.
- Information passes through each step and information can be modified during the passes.
- Due to the necessity of information passing from one step to another and through lots of different intermediary steps, LSTMs might result in a loss of performance in terms of the ability to assess distant past inputs and to learn patterns.

Attention LSTM:

- Aims to improve the information transmission of LSTMs for long term dependencies
- Enables parallel processing of the input data,
- Allows the information transmission directly from each of the previous time steps



Flight Passenger Forecast



Task: To predict how many passengers will fly with the airline at the next time step. For this research, time step intervals used are months, as the data set used was convenient for this type of forecast.

Dataset: The dataset is a rather simple one, with month and passengers as its two columns. The dataset contains a span of 12 years, so 144 months.

Approach: the data from time steps 0 to t_i-1 is used to forecast the outcome of time step t_i , called **continuous approach**. **Sliding window approach** is also used to compare continuous approach with one that should favor LSTM more on paper.

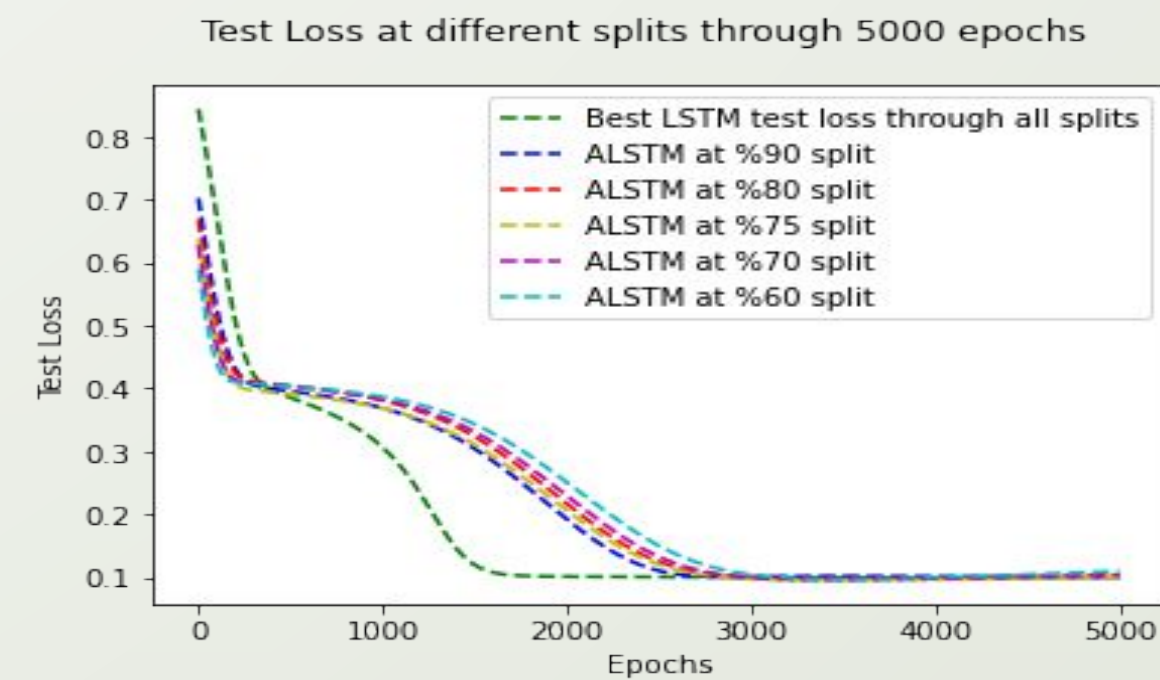
All of the following computations are averaged on 5 different runs to reduce variance. Furthermore, root mean squared error function is used to compute the loss. The graphs computed with loss values are all test losses.

Results

Continuous Approach

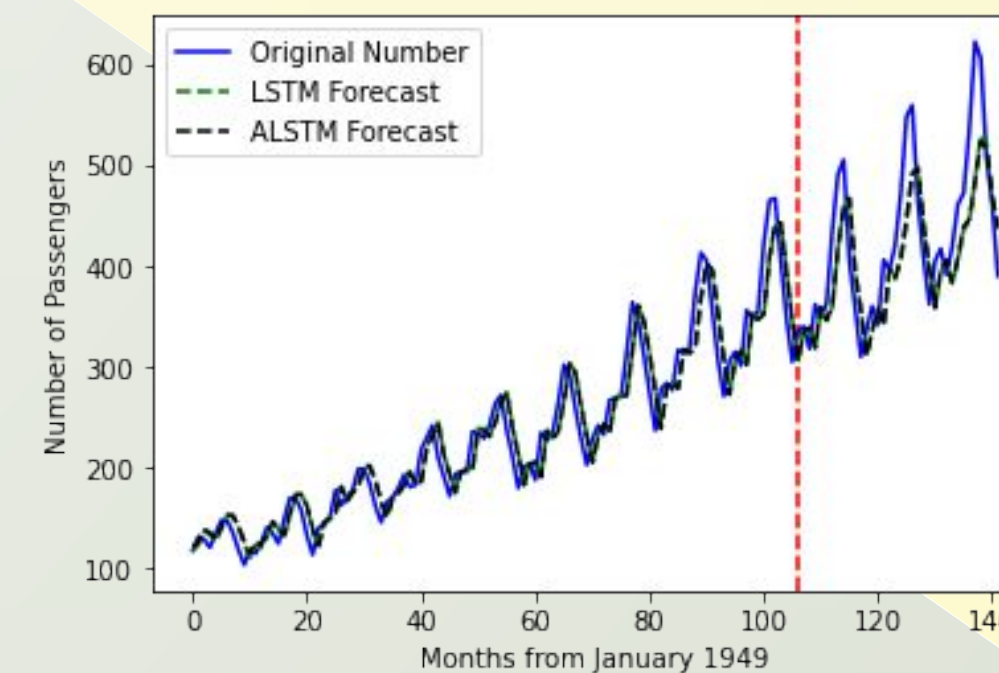
Train split	LSTM Test Loss Average	ALSTM Test Loss Average
0.90	0.099	0.100
0.80	0.105	0.098
0.75	0.106	0.095
0.70	0.108	0.096
0.60	0.114	0.097

Table 1: LSTM vs ALSTM Test Loss Through 5 different splits



Method	Mean	Standard Deviation
LSTM	0.106	0.004
ALSTM	0.097	0.002

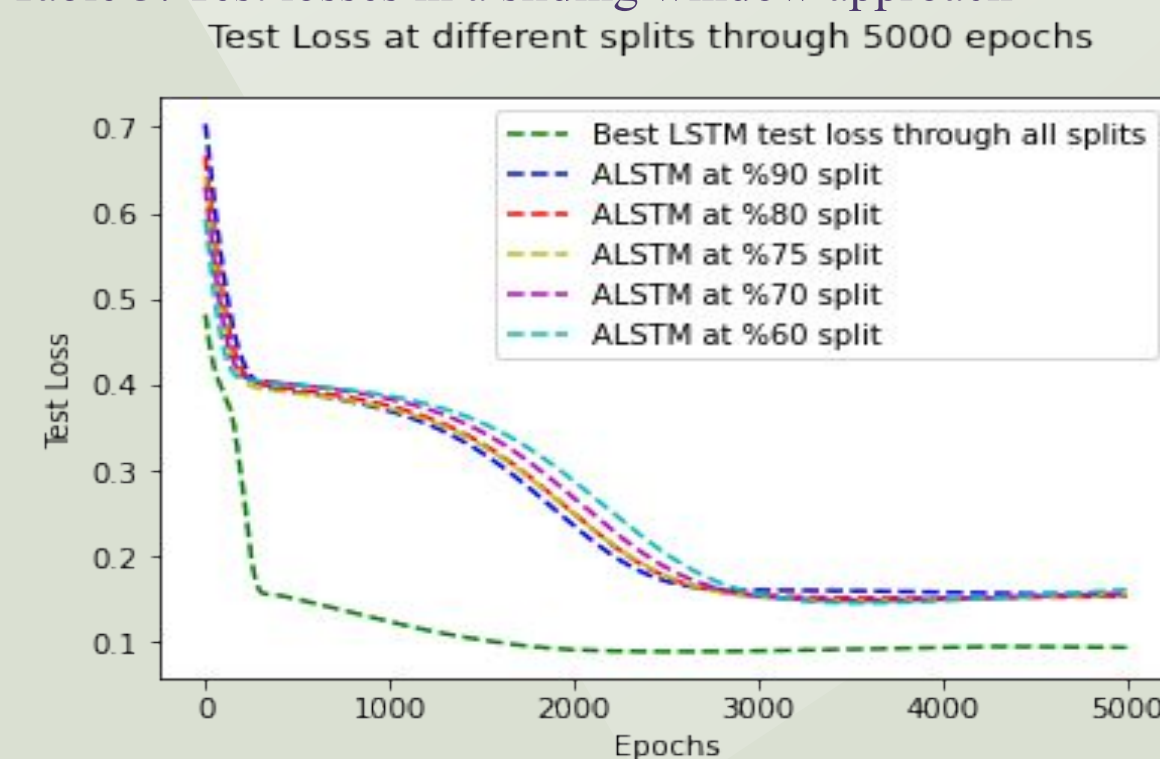
Table 2: Mean and Standard Deviation of the values on the left Number of Passengers Forecasts of LSTM and ALSTM



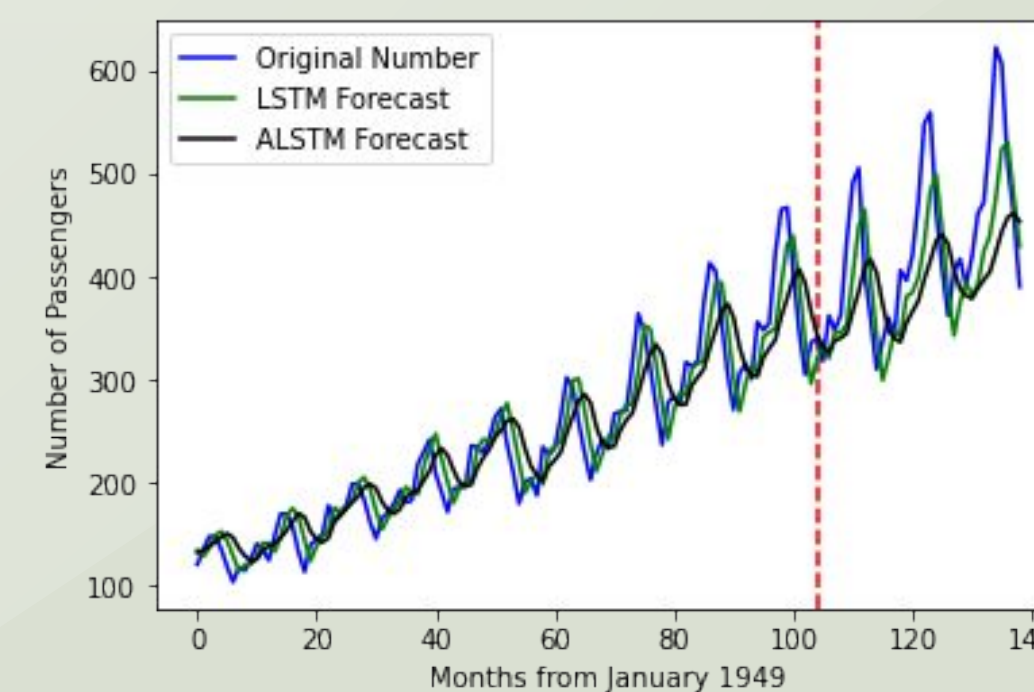
Sliding Window Approach

Train split	LSTM Test Loss Average	ALSTM Test Loss Average
0.90	0.088	0.154
0.80	0.091	0.151
0.75	0.093	0.148
0.70	0.101	0.148
0.60	0.109	0.145

Table 3: Test losses in a sliding window approach



Number of Passengers Forecasts of LSTM and ALSTM



Discussion and Conclusion

Continuous approach:

- ALSTM has an advantage over LSTMs. The average of ALSTM for different splits for was 2 standard deviations separated, or smaller, than the average of LSTM. This suggests that it is indeed a reasonable claim to conclude ALSTM performs better than LSTM in this approach.
- It is observed that this approach favors ALSTM, unlike sliding window

Sliding Window approach:

- LSTM has shown significant advantage over ALSTM in terms of performance, and LSTM's performance reduced by almost %100, while LSTM performance increased slightly.
- It is observed that because this approach uses limited data, it also limits ALSTM's ability to transmit information from all previous time steps, which is the main design goal behind ALSTM's architecture.

From this research, it can be concluded that within the scope of this project, ALSTM performed better than LSTM, but there is still room for improvement.

Future Work

- On the quantitative analysis side, the difference in the forecasted values of ALSTM and LSTM is so tiny that the two lines on forecast on the left are not distinguishable. This raises the question on if the advantage gained by ALSTM is significant enough.
- More testing