Improving the performance of Recurrent Neural Networks for time series prediction by combining Long Short-Term Memory and Attention Long Short-Term Memory

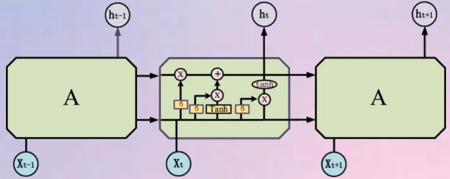
Aaron van Diepen <a.j.vandiepen@student.tudelft.nl>

Computer Science and Engineering, TU Delft

Does a combination of Long Short Term Memory with Attention Long Short Term Memory improve accuracy for time-series classification using Machine Learning as compared to using either separate from each other?

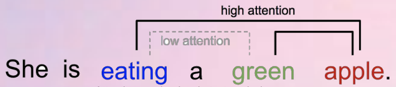
Long Short Term Memory

- Part of recurrent neural networks
- Connections between nodes form a directed graph
- Often used for sequence prediction problems



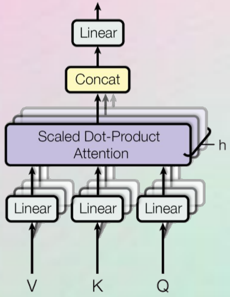
Multi-Head Attention

- Resembles the way in which humans focus more on certain things than others.



Attention mechanism in text analysis

- Computes attention value for every input
- Linearly transforms attentions into vector



Multi-head scaled dot-product attention

Attention Long Short Term Memory

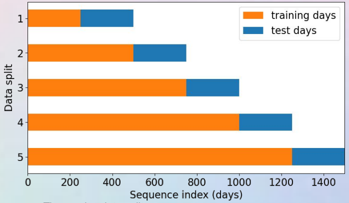
- Pytorch implementation created during project
- Multiplies multi-head attention with lower triangular matrix filled with 1's in order to preserve sequential data

Dataset

- Weather data intended for predicting dangerous levels of ozone
- Contains a lot of data gaps after 4 years
- First 1500 days will we be used
- 6 days of data will be used to predict peak temperature on the 7th day

Time Series Split

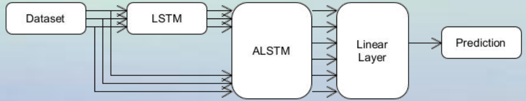
- In order to assure results are accurate
- Creates 5 different versions of the dataset
- Training followed by validation on the test days will be used for each model and each split
- Data used for testing will be withheld from normalization scaler
- Prevents future-looking by only training on data sequentially before testset



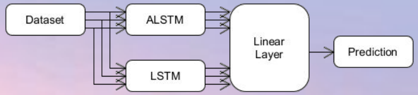
Time series data splits

Combining the LSTM and ALSTM

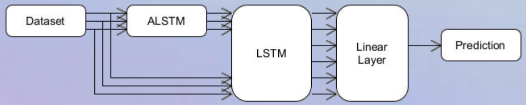
- Combining attention and memory states
- 3 types of combinations will be tested



Combination 2

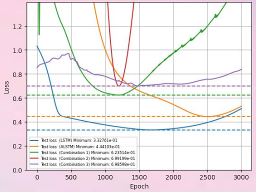


Combination 1

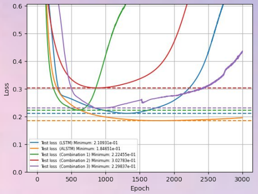


Combination 3

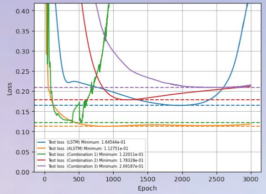
Final results



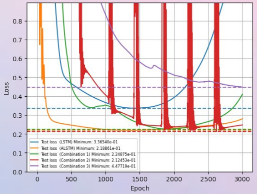
Testing loss graphs for all models on split 1 and their minimum loss achieved



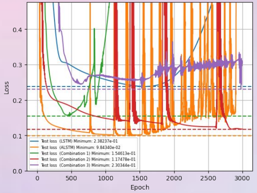
Testing loss graphs for all models on split 2 and their minimum loss achieved



Testing loss graphs for all models on split 3 and their minimum loss achieved



Testing loss graphs for all models on split 4 and their minimum loss achieved



Testing loss graphs for all models on split 5 and their minimum loss achieved

Model	AVG±STD (Split 1-5)	AVG±STD (Split 2-5)
LSTM	2.57E-1±7.60E-2	2.38E-1±7.27E-2
ALSTM	2.12E-1±1.39E-1	1.54E-1±5.75E-2
Combination 1	2.69E-1±2.03E-1	1.81E-1±5.11E-2
Combination 2	3.02E-1±2.32E-1	2.03E-1±7.74E-2
Combination 3	1.17E+0±1.86E+0	1.29E+0±2.13E+0

The average minimum loss and the standard deviation of all tested models for split 1-5 and split 2-5