W-Transformer : A Wavelet-based Transformer Framework for Univariate Time Series Forecasting

Léna Sasal, Tanujit Chakraborty and Abdenour Hadid
Sorbonne Center for Artificial Intelligence, Sorbonne University, Abu Dhabi, United Arab Emirates
lena.sasal@sorbonne.ae

Motivation

Forecasting the future movement and value of time series is a key component of formulating effective strategies in most business.

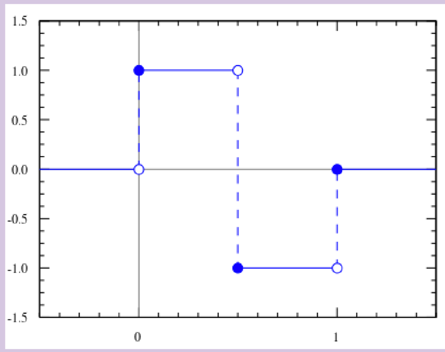
Contribution :

- We propose a novel model (namely W-Transformers) for time series forecasting combining wavelet decomposition and Transformers in the ensemble approach.
- We perform extensive experiments with multiple univariate time series datasets, demonstrating better performance compared with the state-of-the-art forecast methods for short and long-term forecasting.

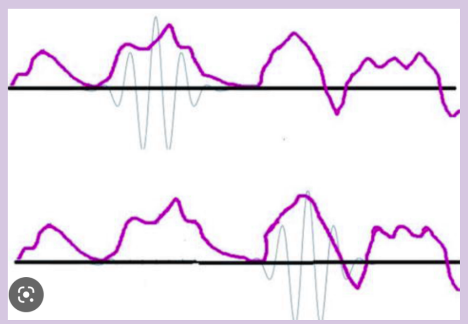
Wavelet

Wavelet is a wave-like oscillation localized in time. It can capture the time and the frequency of a signal. In our work, we focus on DWT that represents a signal using an orthonormal basis representation.This decomposition will give us a set of time series where each time series has a coefficient describing the evolution in time of the signal in a specific frequency band.

To compute the wavelet decomposition, we have to choose the mother wavelet that will convolve the time series. For our study, we choose the Haar wavelet because it helps to capture sudden change between two time step.

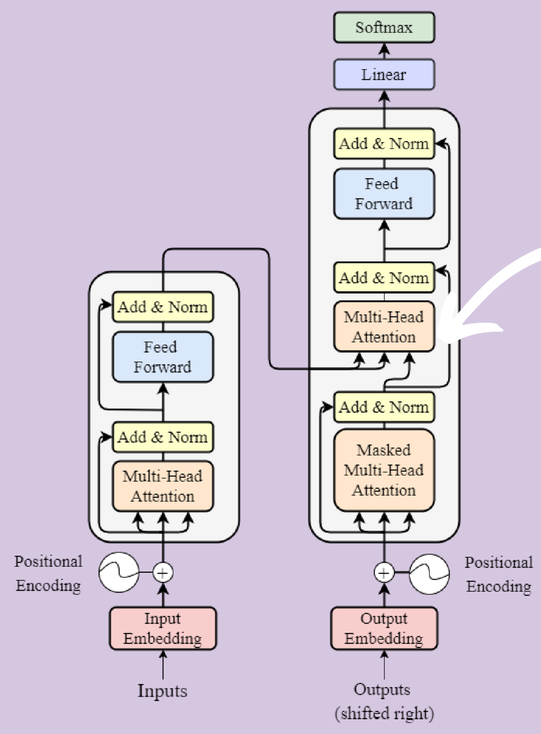


Haar wavelet

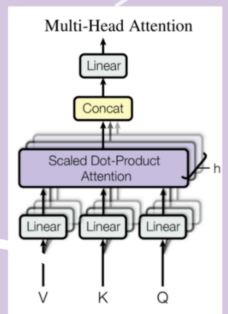


Example of how the wavelet convolve the signal

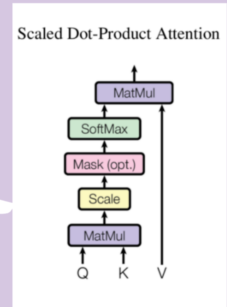
Transformers



$$\text{MultiHead}(Q,K,V) = \text{Concat}(head_1, \dots, head_m)W^O$$

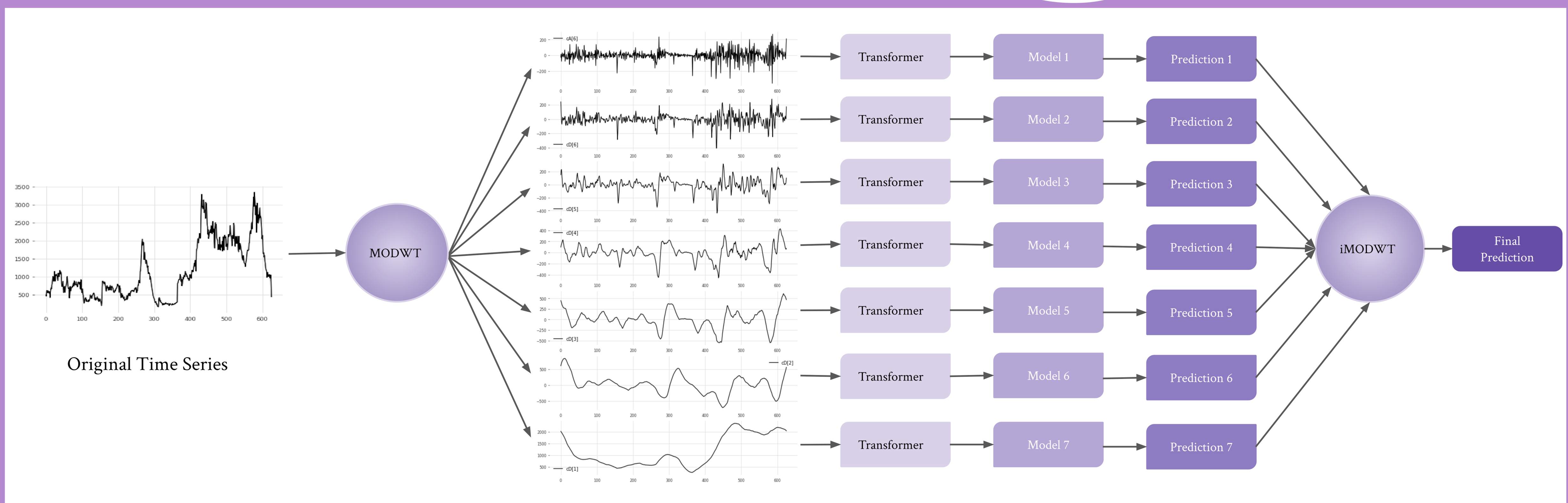


$$head_p = Attn(QW_p^Q, KW_p^K, VW_p^V)$$



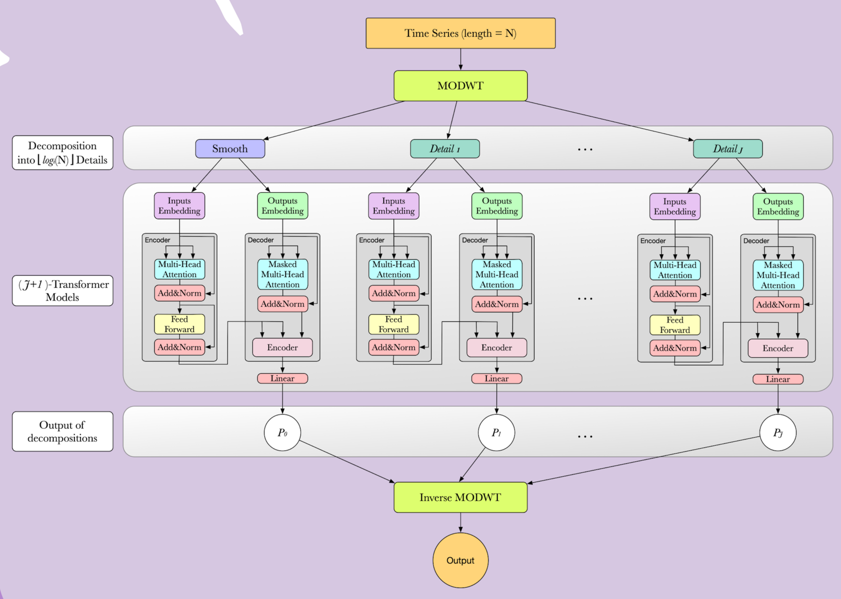
$$Attn(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

Transformers model the long-range dependencies and interactions in time series The multi-headed attention mechanism of Transformers makes them suitable for time series data analysis: they concurrently represent each input sequence element by considering its context, while multiple attention heads can consider different representation subspaces, i.e., multiple aspects of relevance between input elements. This may correspond to multiple periodicities in the signal



Framework

We compute the time series with the MODWT algorithm to get the decomposition, the results give us log(N) time series that have the same length as the original time series with N the length of the time series. The log(N)-1 time series are called details while the last one is the smooth. Then we fit the log(N) time series into separate transformers to get log(N) model. Each model will give the prediction of one details time series that we recombine with the inverse MODWT to get the final forecasting. To assess the performance of the model we then use 4 different metrics, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (sMAPE), and Mean Absolute Scaled Error (MASE) .



Datasets

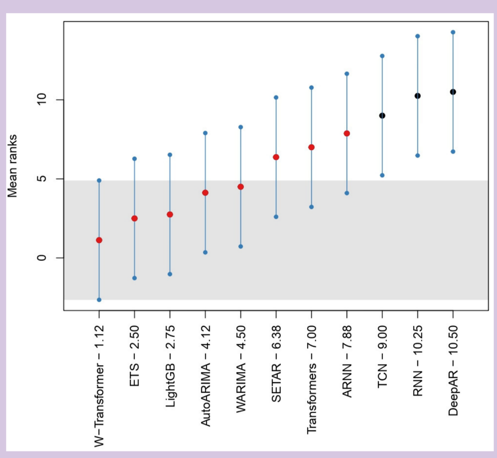
Data Name	Frequency	Observations	Test Size	Behaviour
NFLX	Daily	254	30/90	N-S, L
Website Traffic	Daily	2167	30/90	N-S, N-L
Sunspot	Daily	2663	30/90	N-S, L
Colombia Dengue	Weekly	626	26/52	N-S, N-L
Japan Flu	Weekly	964	26/52	S, N-L
Bangkok Dengue	Monthly	180	12/24	N-S, N-L
Network Analytics	5-min	25631	576/1152	N-S, N-L

Characteristics of the datasets
(N-S : Non-Stationary, S : Stationary, N-L : Non-Linear, L : Linear)

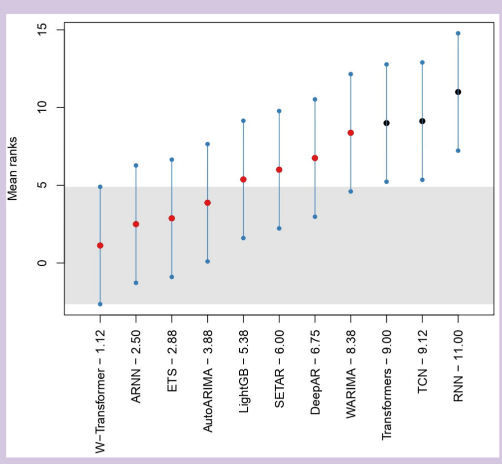
Results

The results of our proposed W-Transformers are compared with 10 baseline models for short and long term horizon. W-Transformers is more efficient for long-term forecasting because wavelets handle nonstationarity while Transformers handle long-term forecasting. The overall performance of our W-Transformers is superior compared to all forecasters considered in this study. For small sample-sized datasets, the performance of statistical methods is comparatively better than that of W-Transformers. However, the proposal performed better than all the baseline models for large temporal datasets

Short-Term average rank of the models



Long-Term average rank of the models



References

- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” CoRR, vol. abs/1706.03762, 2017
- D. B. Percival and A. T. Walden, Wavelet methods for time series analysis. Cambridge university press, 2000, vol. 4
- M. Aminghafari and J.-M. Poggi, “Forecasting time series using wavelets,” International Journal of Wavelets, Multiresolution and Information Processing, vol. 5, no. 05, pp. 709–724, 2007.

Paper

GitHub