A Time Series is Worth 64 Words: Long-Term Forecasting with Transformers

Yuqi et al addresses the challenges and effectiveness of Transformer-based models in time series forecasting. They propose PatchTST (Channel-Independence Patch Time Series Transformer) [1] and introduce two key designs, patching and channel-independence. Patching involves aggregating time steps into subseries-level patches to enhance the extraction of local semantic information since a single time step does not have semantic meaning like a word in a sentence. Channel-independence, a novel concept for Transformer-based models, ensures that each input token contains information from a single channel, offering a different perspective from previous approaches.

PatchTST method offers reduction in time and space complexity achieved through patching of the Transformer model by addressing the primary bottleneck in computation time and memory, the capability of learning from a longer look-back window since PatchTST can capture more important information without significantly increasing memory and computational usage, and the effectiveness of representation learning by capturing abstract representations and achieving state-of-the-art MSE in forecasting. Yuqi et al asserts that PatchTST significantly reduces computational complexity, allows for a longer look-back window without sacrificing efficiency, and demonstrates superior performance in terms of Mean Squared Error (MSE) compared to existing models.

[1] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2022). A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.