NANYANG TECHNOLOGICAL UNIVERSITY

School of Computer Science and Engineering

# Machine Learning Techniques for Prediction of Time Series Data

## Interim Report

**Student**

Andrew Ng Yong Kuan

**Supervisor**

A/P YEO CHAI KIAT

# 1. Interim Report

This document serves as a comprehensive progress report, encompassing achieved milestones, projected tasks, and an updated project timeline informed by the current progress.

# 2. Scope and Objective of Project

This project aims to propose and evaluate innovative machine learning techniques to enhance the accuracy of time series data prediction. The primary technique to be evaluated in this project is the transformation of time series data using Stationary Wavelet Transform (SWT). Coupled with a transformer network, this project compares the prediction capabilities of the SWT-Transformer with other existing models.

# 3. Summary of Work Completed

The following section describes all the work that has been completed for the project.

## 3.1. Project Planning

The project planning phase is dedicated to grasping the precise specifications and expectations outlined by the project supervisor.

The key accomplishments of this phase are the submission of the Final Year Project (FYP) plan and schedule, and a clearly defined research direction for this project.

Stationary Wavelet Transform (SWT) was introduced as the primary machine learning technique to be analyzed. In addition, source codes of models in published papers are provided for comparison. The project begins by performing a literature review of the published papers and the analysis of source codes. This step uncovers versioning errors in the source code and other complications such as finding and preparing the dataset similarly to what was discussed in the published papers.

## 3.2. SWT-Transformer Development

As the SWT-Transformer model is a key requirement for this project, it is crucial that the model functions correctly without error or biases. These steps are made to validate the model:

1. SWT-Transformer model was referenced from a GitHub page but was riddled with basic and versioning errors. These errors were rectified to obtain similar results to what was discussed in a published paper found to be coherent to this GitHub page.

2. The data preparation of datasets and source code of other models are done using the Pytorch library. The code sample for SWT-Transformer relies on the tensorflow keras library. For the convenience of reusing the same data preparation and fair comparison, the code sample for SWT-Transformer was carefully migrated to the Pytorch library. The code was repeatedly run during the migration to note any errors that could explain the differing results in the Pytorch library and the tensorflow keras.

3. Data from the datasets undergo SWT to produce transformed values. Due to the difference in the structure of the prepared data in Pytorch, the code to transform the data and the code to train the model require some tweaks.

4. Two different approaches were made in transforming the data using SWT and training the model.
    - Original method where a sliding window algorithm was used to convert time series data to waves.
    - Alternate method where the whole data for a time series are converted into a wave. It takes in the lookback weeks as input data to predict the next x number of days.

5. Current results show the sliding window algorithm predicts significantly better. The alternate method predicts worse than other models but the current method predicts 10x better.

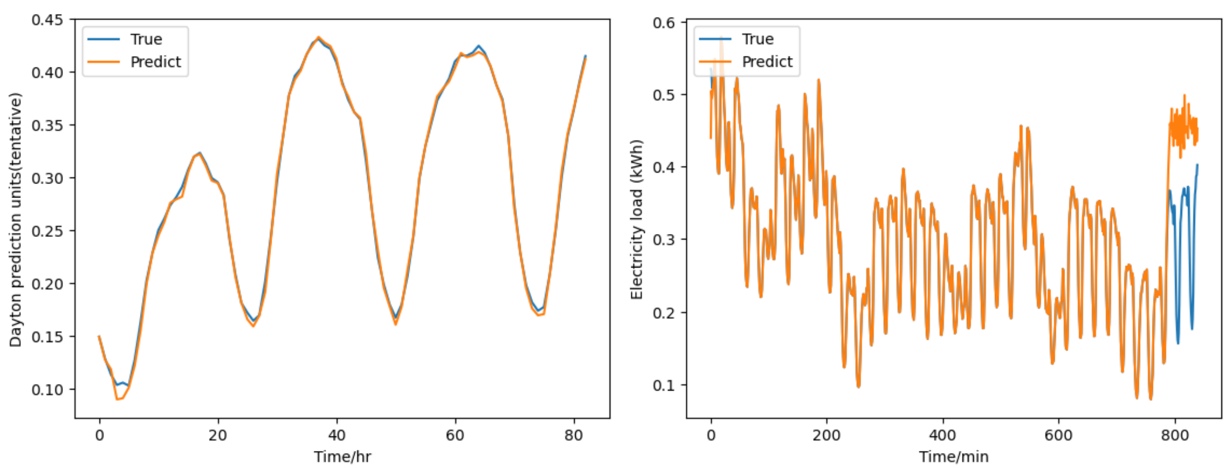


Figure 1: Sliding window algorithm (left) vs SWT on full length time series (right) for Dayton dataset

## 3.3. Literature Review

The following findings were yielded after performing a literature review for papers related to Stationary Wavelet Transform (SWT) and Transformers.

Literature Review in SWT:

- Time series features extraction using Fourier and Wavelet transforms on ECG data (Marisa Faraggi, 2019)
    - https://blog.octo.com/time-series-features-extraction-using-fourier-and-wavelet-transforms-on-ecg-data
- Time Series Forecasting Using Wavelet Denoising an Application to Saudi Stock Index (Rumaih M. Alrumaih, Mohammad A. Al-Fawzan, 2009)
    - https://www.sciencedirect.com/science/article/pii/S1018363918307554

Literature Review in Wavelet Transformers

- A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy (Erick Giovani Sperandio Nascimento, Talison A.C. de Melo, and Davidson M. Moreira, 2023)
    - https://www.sciencedirect.com/science/article/pii/S0360544223010721
- Dwtformer: Wavelet decomposition Transformer with 2D Variation for Long-Term Series Forecasting (Yujie Cao, Xi Zhao, 2023)
    - https://ieeexplore.ieee.org/document/10082078
- W-Transformers : A Wavelet-based Transformer Framework for Univariate Time Series Forecasting (Lena Sasal, Tanujit Chakraborty, Abdenour Hadid, NIL )
    - https://ideas.repec.org/p/arx/papers/2209.03945.html

## 3.4. Interim Report Writing

The Interim Report (this report) was written as per the requirements of the Final Year Project.

# 4. Forecasted Work

The following section describes all the work that is sought to be completed by the end of the project.

## 4.1. Code logic correction

The results produced by SWT-Transformer for all datasets are suspiciously accurate. This prompted some investigations into the code for signs of leakage where the model accepts input values that give information about the target value.

These are the current two hypotheses that were formulated during the analysis:

1. SWT transformation in each window could leak the actual value to other values in the wave since the waves are created based on the input (types of wave decomposition)

2. Currently, the implementation of this sliding window results in data in the prediction zone being added in future time steps. For example, for the London dataset with seq_len = 50, in the second window, 49 wave values + 1 ground truth wave value of 1st prediction time step is used to predict the wave value of the 2nd time step (Boundary Problem)

While scouring through the internet for similar problems, several articles seem to address this issue too. Therefore, it is vital that the code addresses this pressing issue.

**Boundary Problem and types of wave decomposition** in SWT:

- Forecasting Natural Gas Prices Using Wavelets, Time Series, and Artificial Neural Networks (Junghwan Jin, Jinsoo Kim, 2015)
    - https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0142064
- Boundary problem and data leakage: A caveat for wavelet-based forecasting (Yuto Kajita, 2018)
    - https://www.researchgate.net/publication/329443818_Boundary_problem_and_data_leakage_A_caveat_for_wavelet-based_forecasting

## 4.2. Exploring other models

In recent years, new models have been used for time series forecasting. One such model is the Time-Series Generative Adversarial Networks (TimeGAN). Given the time remaining, this will be taken as consideration and a simpler model, such as Autoregressive Moving Average (ARMA), could be discussed instead.

## 4.3. Final Report Writing

A Final Report will be written as per requirements of the Final Year Project. It will contain a comprehensive documentation of all activities and findings of this project.

## 4.4. Oral Presentation

An Oral Presentation will be delivered to the grading committee as per requirements of the Final Year Project, presenting the project's activities and findings in a succinct and engaging manner.

# 5. Summary of Problems Faced

This section provides a brief listing of all problems faced, grouped by the component of the project in which the problem was faced.

## 5.1. SWT-Transformer Development

The challenges faced during the model's implementation are as follows:

1. Finding and cleaning the dataset used by the SWT-transformer research paper
    a. The dataset was not readily available, and the codebase referenced a different dataset from the one in the research paper. The dataset was found by looking through the

references made in the research paper. The data is then analyzed and transformed similarly to what was described in the research paper.

2. Correcting version and basic errors in the SWT-Transformer codebase
    a. The largest issue in this section is the transformation of the data using SWT. Due to versioning errors, the main functions in the PyWavelets library, pywt.swt and pywt.iswt, no longer functions as before. Multiple tests using self-created data helped to narrow the errors to the change of data types between these transformations.

3. Conversion of SWT-Transformer codebase from tensorflow Keras to Pytorch library
    a. Although both libraries aid the user in creating a machine learning model, there are distinct differences in the way the model is instantiated. Examples include the lack of requirement for build function in Pytorch, the need to specify the number of incoming and outgoing channels in convolution layers in Pytorch, the need to track batch number in Pytorch, etc. This was time-consuming as there are cases where the model could run but the prediction loss is vastly different from the original codebase.

4. Preparing other datasets for SWT-Transformer
    a. While the other datasets were well prepared, further data modification and changes to how SWT is implemented were required.

5. Questionable results by SWT-Transformer on other datasets
    a. While the results show that SWT-Transformer is functioning correctly, the prediction results were suspiciously accurate. The SWT-Transformer seems to be able to predict about 10 times better than previously known models. This raised questions and led me to narrow down a few possible explanations. While writing this interim report, it has come to my attention that there are papers that support the explanations.

## 5.2. Literature Review

The main challenge faced during the literature review was the understanding of mathematical notations for SWT. The explanations provided by the authors of the papers assisted in my understanding of SWT.

# 6. Amended Project Schedule

The following Gantt chart details the amended project schedule.