Forecasting Retail Client Flow with LSTMs on Inconsistent Time Series

Autores
Pedro Gusmão, IEETA, <pedrogusmao@ua.pt>
Ana Tomé, DETI, <ana@ua.pt>
José Moreira, DETI, <jose.moreira@ua.pt>

14 de Julho de 2021

Motivation and Context

- Stores have technologies that perform data collection in the form of **time series** (TSs)
- Stores require forecasting to plan ahead of time
  - **Optimize employee work schedules**
- Research on retail **high-frequency** TSs is lacking
- ML/DL models excelled in recent forecasting competitions (e.g. Makridakis competition)

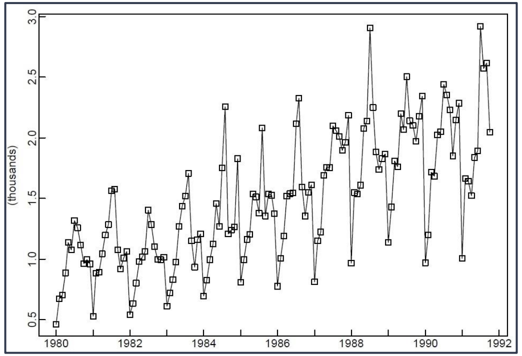

Figure 1. Monthly sales of australian wine between Jan. 1980 and October 1991

Contributes

- Explain how a neural network, namely **RNN-LSTMs**, can use TSs to perform forecasts
  - Two LSTM architectures
- Obtain a general view on how these models perform with long horizon forecasts on high-frequency TSs
- Investigate if a classical TS preprocessing method brings benefits to such models

Classical Preprocessing

Time Series Analysis

- With classical forecasting models:
  - Desired feature: **stationarity**.
  - To ease the forecasting task it is common to decompose the TSs.

A TS can be **decomposed** by separating it in:

- Trend-Cycle; $m(t)$
- Seasonal; $s(t)$
- Noise. $\varepsilon(t)$

Assuming it follows an additive model

$$\boxed{x(t) = m(t) + s(t) + \varepsilon(t)}$$

Stationarity

- A TS originates from a **stationary process** if it is not described by any trend or seasonality (mean and variance are constant).
- Seasonality is observable through plots (e.g. PACF, **FFT**).
- For trend-stationarity, various statistical tests were created. One of the most popular is the **Augmented Dickey-Fuller** test.
- If the TS is not stationary then it can be **transformed**.

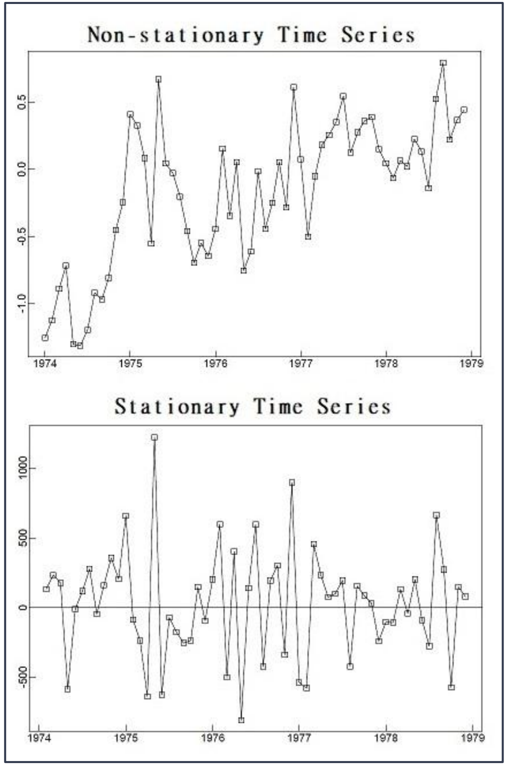

Figure 2. An example of stationary and a non-stationary TS.

Harmonic Regression

- Regression can handle trend and seasonality at the same time
- With Fourier terms it can take into account **multiple** seasonalities

Seasonal (Fourier Harmonic Terms)

$$\boxed{s(t) = \sum_{k=1}^{K} \left(a^{(k)} \cos\left(\frac{2\pi k t}{z}\right) + b^{(k)} \sin\left(\frac{2\pi k t}{z}\right) \right)}$$

$+$

Linear Trend

$$\boxed{m(t) = \beta_0 + \beta_1 t}$$

Example of complete model

$$\boxed{x(t) = \beta_0 + \beta_1 t + \beta_2 \cos\left(\frac{2\pi t}{z_1}\right) + \beta_3 \sin\left(\frac{2\pi t}{z_1}\right) + \beta_4 \cos\left(\frac{2\pi t}{z_2}\right) + \beta_5 \sin\left(\frac{2\pi t}{z_2}\right) + \varepsilon(t)}$$

Neural and Long Short-Term Memory Networks

Multilayer Perceptron

- Classified as a Feedforward Neural Network
- Hidden layers + nonlinear activation functions, theoretically, means that ANNs are **universal approximators**

$$h = \phi\left(W_1\, x + b_1 \right)$$

Trainable Parameters

$$\hat{y} = W_2\, h + b_2$$

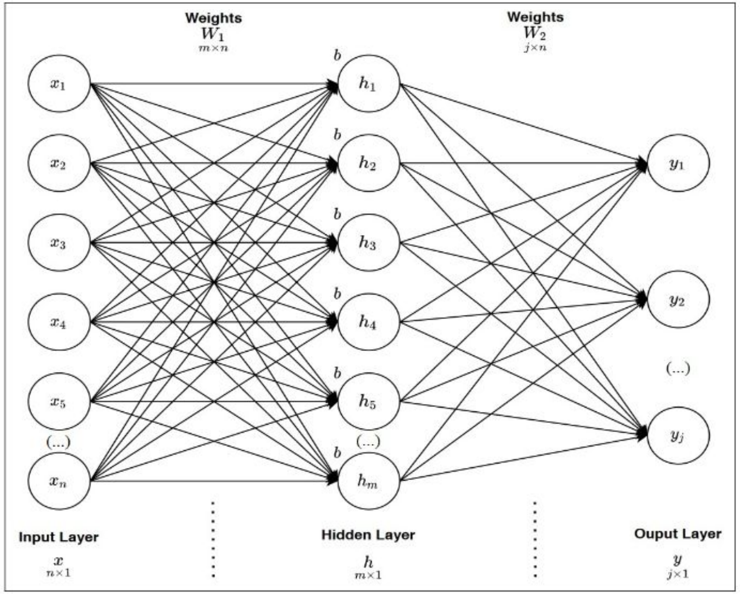

Figure 3. Multilayer Perceptron (FNN).

Recurrent Neural Networks

- Can handle sequences unlike feedforward networks
- Model temporal relationships by preserving an internal state
- The state is used to compute the output for next step (feedback loop)

Traditional RNNs have **training issues** so an extension was used - LSTM

- Each gate can be considered a FNN

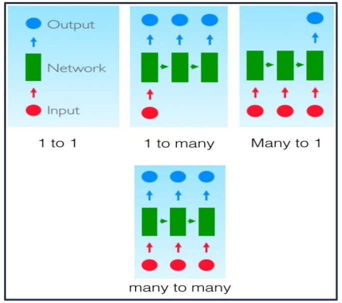

Figure 4. Sequence problems.

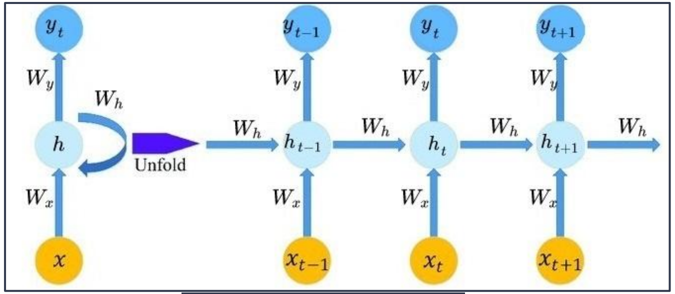

Figure 5. Many-to-many RNN.

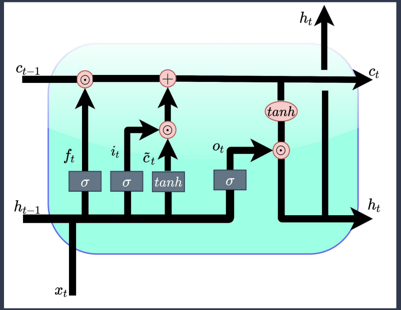

Figure 6. LSTM unit.

LSTM-1

- Only the last hidden state is connected to the output layer
- The LSTM layer follows a **many-to-one** philosophy

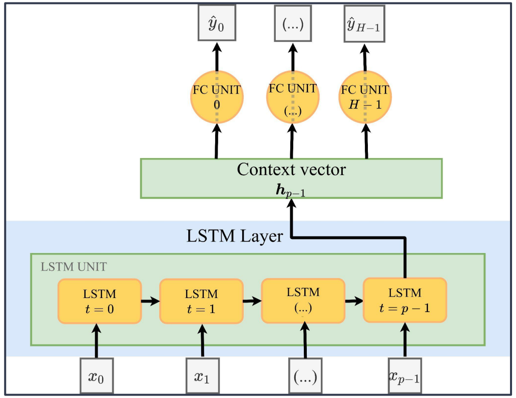

Figure 7. LSTM-1.

LSTM-2

- This architecture requires that p is equal to the forecast horizon H
- The hidden state is returned in all processing steps
- Has only one output unit and it serves all the returned states (i.e., it is time distributed)
- The LSTM layer follows a **many-to-many** philosophy

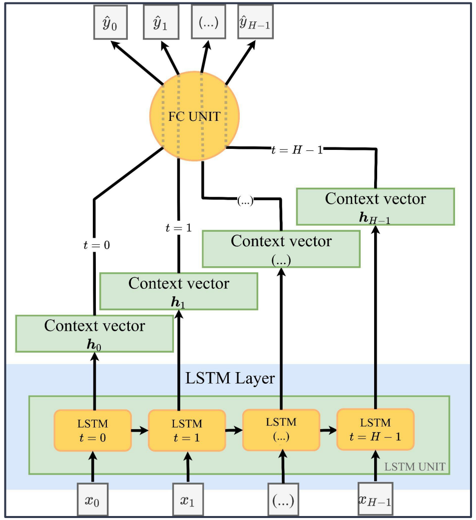

Figure 8. LSTM-2.

Sliding Window

- Time Series forecasting can be naturally considered a **supervised learning regression** problem
  - Inputs have associated target outputs that can be compared with the produced predictions

- It has three parameters:
  - The number of lagged values *p*
  - The forecast horizon *H*
  - The stride *s*

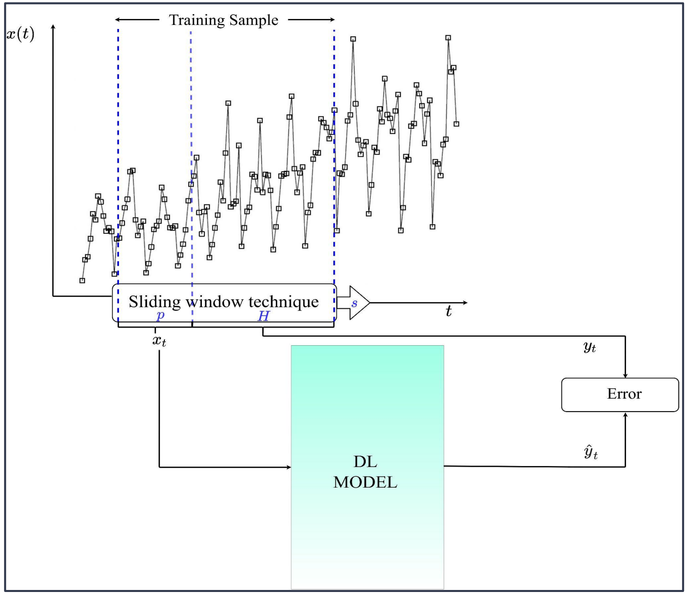

Figure 9. Sliding Window.

Preliminary Experiments

Problem Definition and Datasets

- There are two stores (A and B) that need month-ahead client entries predictions (30 days)
  - A: 1649 days, B: 976 days
- Schedules vary
  - A standardization method was developed
    - This is a requirement for HR

	1st MFS	2nd MFS	3rd MFS
Store A	*09:00-21:00*: **1006**	*09:00-21:30*: **142**	*09:30-14:00*: **104**

Figure 11. Most frequent schedules in store A.

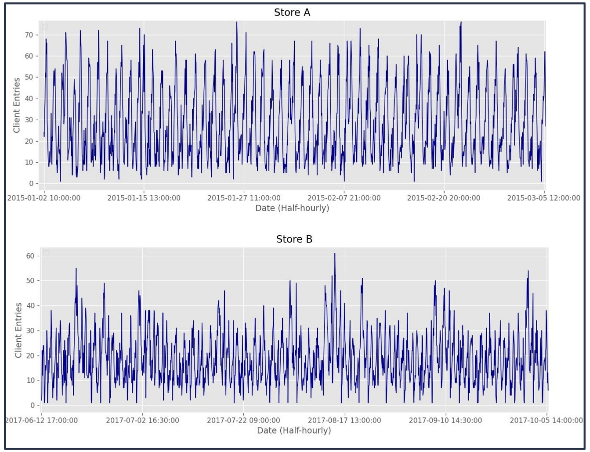

Figure 10. First 1400 observations of the TSs.

Summary

- The TSs clearly have multiple seasonalities (**FFT**)
- Thus, we have two datasets with three settings:
  - Raw (RAW)
  - Schedules standardized (STD)
  - STD + Harmonic Regression (STD+HR)
- LSTM-1 tested with *p* -> {10, 25, 50, 100, 150, 175}
  - LSTM-2 (*p=H*)
  - Both configured with the same internal parameters
- Compared against a baseline naive model
- **Metrics**: MAE, RMSE and MAAPE

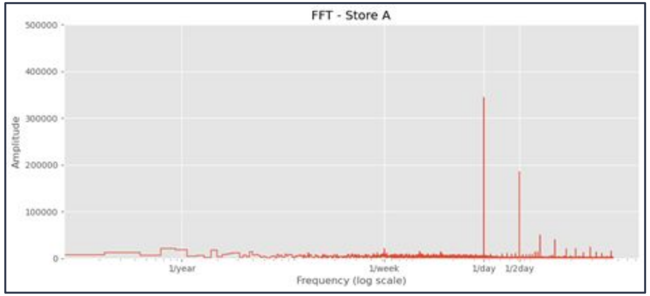

Figure 12. FFT of store A's training set.

Some Results and Discussion

- The HR preprocessing method did not show any notable benefit
- Schedule preprocessing greatly improved results (up to ~23% MAAPE)
- Both LSTMs were notably better than the baseline (up to ~14% MAAPE)
- LSTM-1 (Many-to-One) was superior to LSTM-2 (Many-to-Many)

RAW

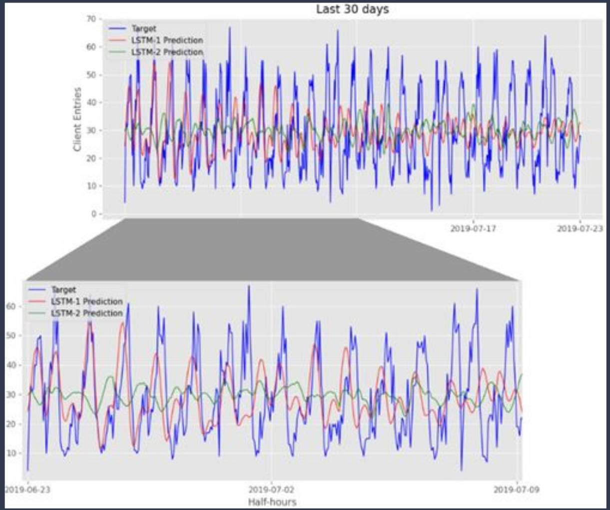

Figure 14. LSTM-1 (*p=150*) and LSTM-2 performance in store A's unstandardized TS (first 16 days zoomed in).

STD+HR

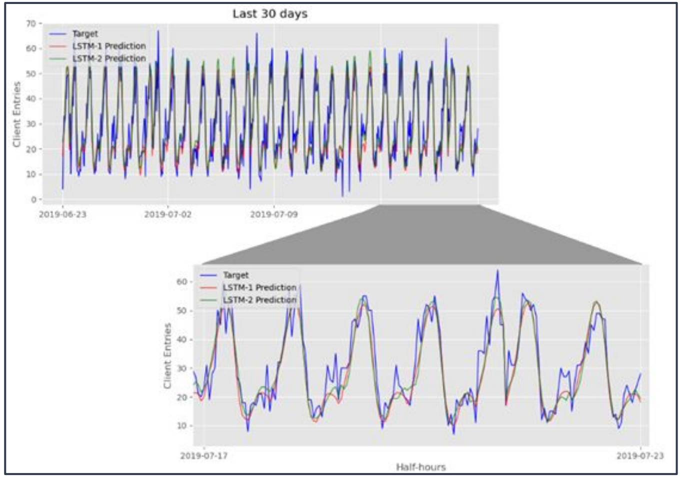

Figure 13. LSTM-1 (*p=25*) and LSTM-2 performance in store A's standardized TS that was previously deseasonalized and detrended.

Future Work

- Enhance the models by:
  - Adjusting the HR method (e.g. more/less seasonalities considered)
  - Hypertuning the LSTMs (internal state units, regularization, etc...)
  - Tuning the sliding window parameters (historical period, stride)
- Test with TSs that are not trend-stationary
- To overcome the drawback of schedule standardization is necessary for real-world application

CAPSI2021

Thank you

universidade de aveiro

ieeta