Long Short-Term Memory Networks applied in Time Series

Fábio Silva

P.PORTO

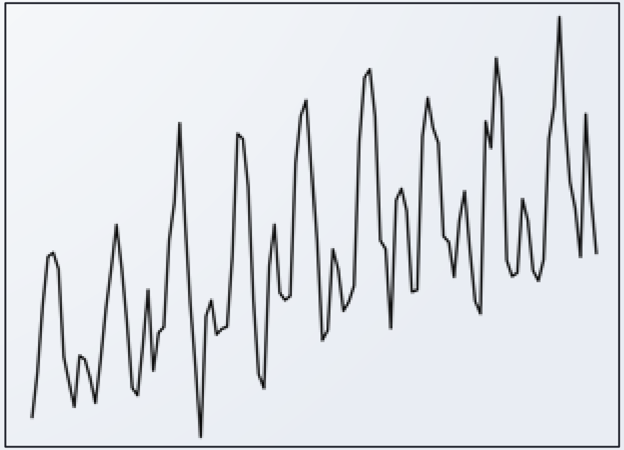
ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO

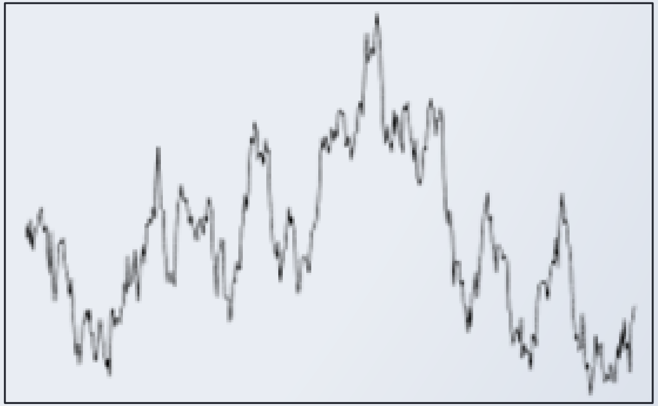
DSPT DATA SCIENCE PORTUGAL

Time Series

A sequence of ordered data points . . . .

Time Series





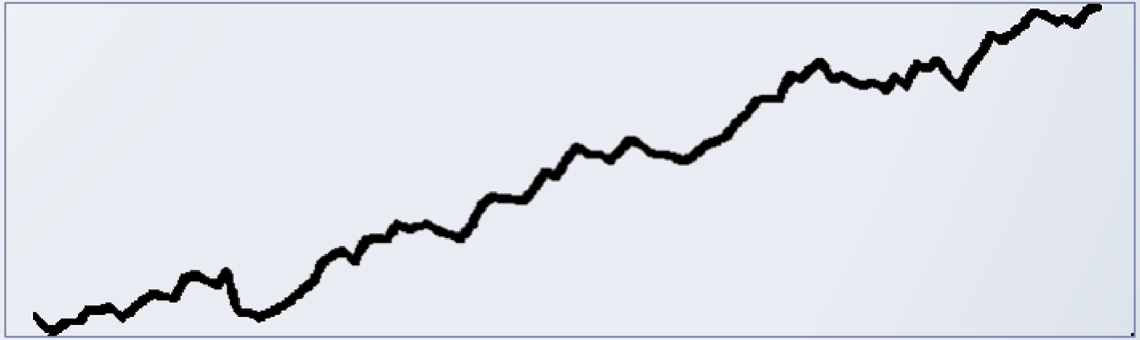
Time Series

Where can we find them?

Images source: https://www.flaticon.com/authors/freepik

Time Series

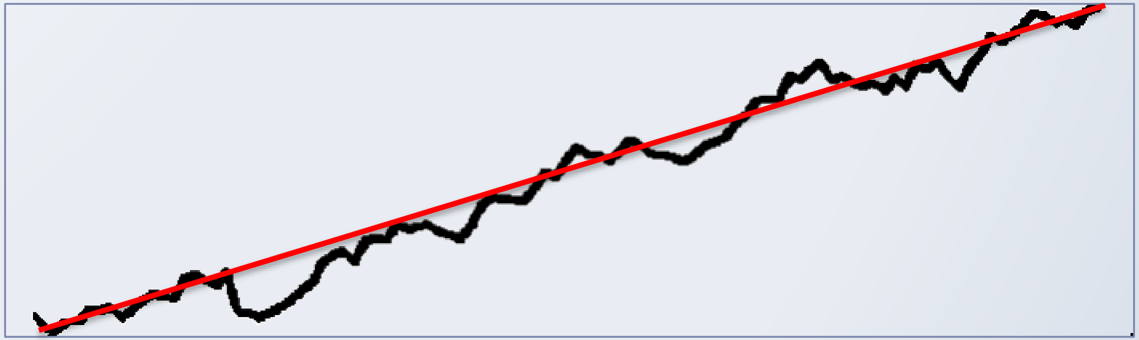
Properties in time series



Time Series

Properties in time series

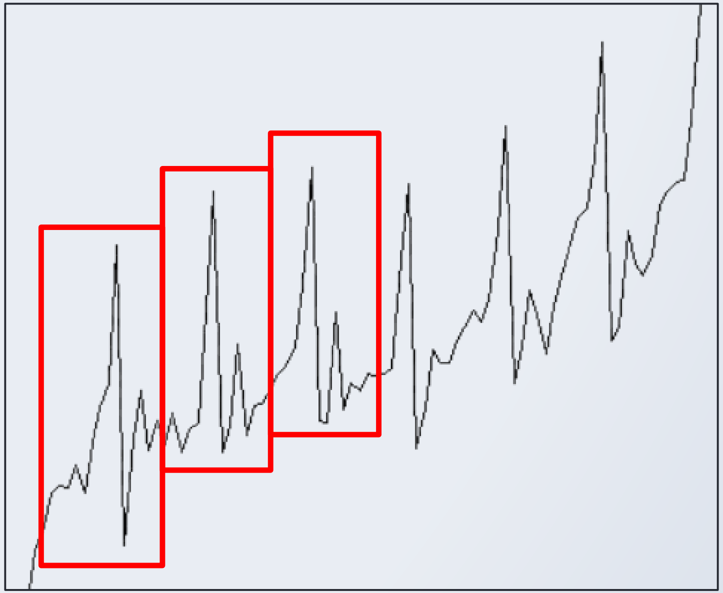
- Trend



Time Series

Properties in time series

- Seasonality



Time Series

Other Properties

- Noise
- Level

Time Series

Univariate

Multivariate

- Endegenous
- Exogenous

Forecast ≠ Prediction

Time Series

How to?

Predict Direct Values?

Data Transformations?

Data Transformations

Data Transformations:

- Power Transform
- Difference Transform
- Standardization
- Normalization

Data Transformations

Making data stationary

Why?

Data Transformations

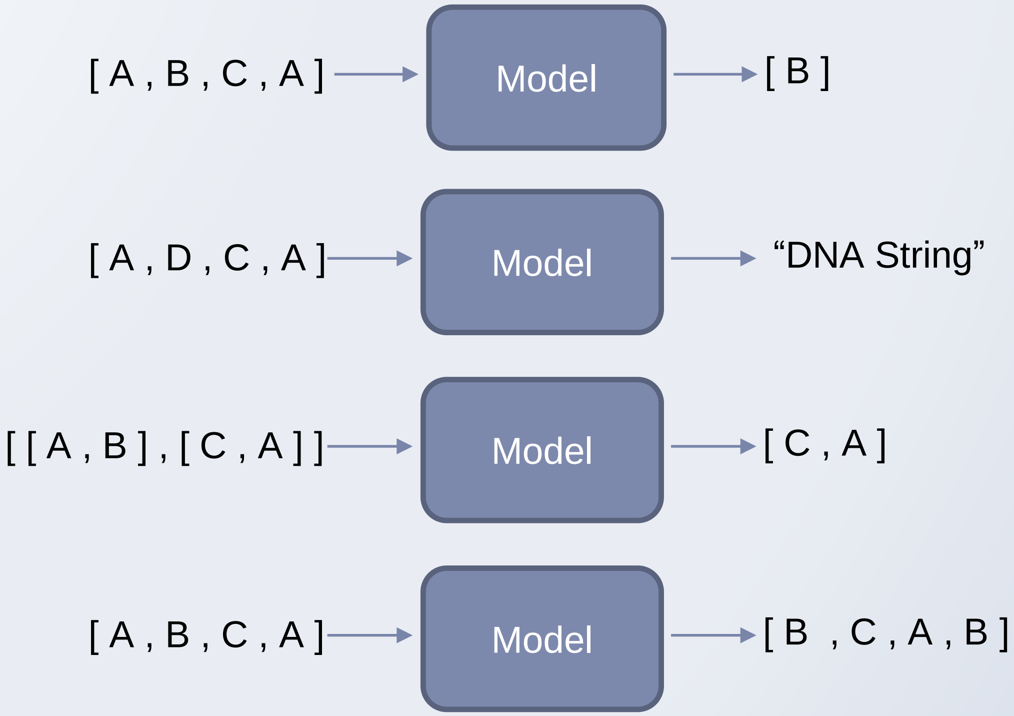
Source: https://www.kdnuggets.com/2019/05/machine-learning-time-series-forecasting.html

Data Transformations

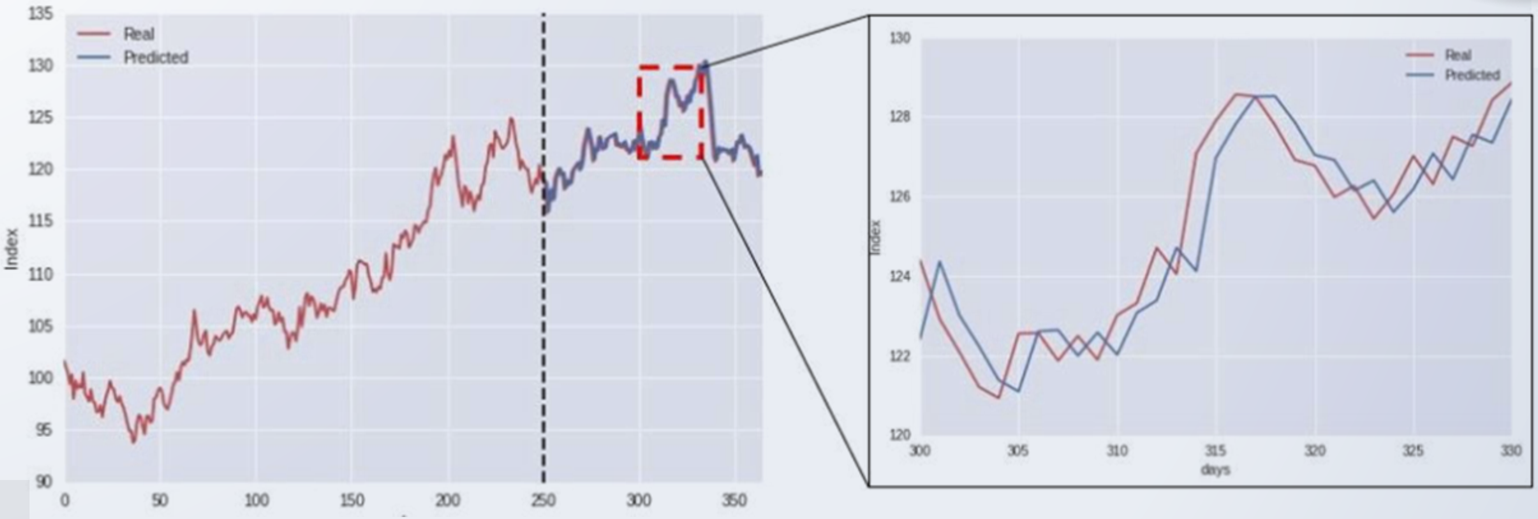
- Can we convert a time series problem into a classical machine learning problem?

Time series Models

Different
Approaches

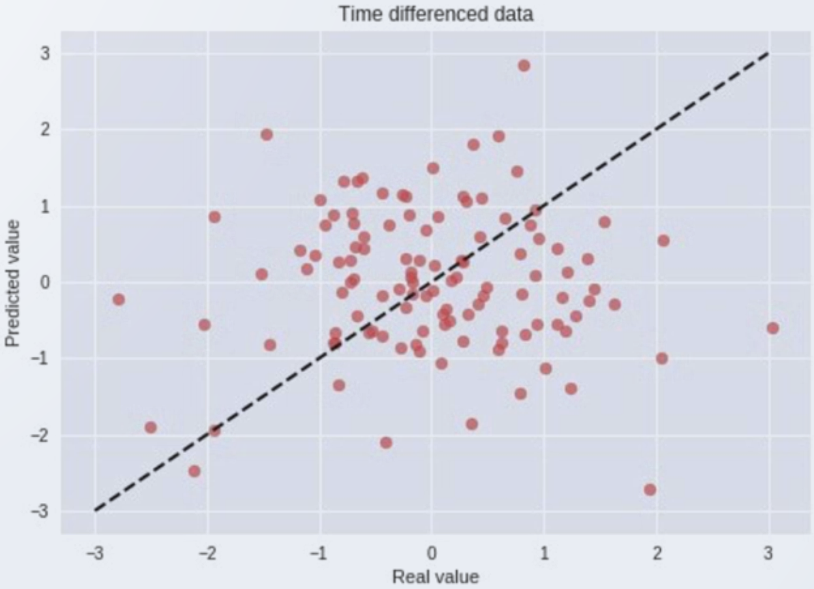


Time series Models



Source: https://www.kdnuggets.com/2019/05/machine-learning-time-series-forecasting.html

Time series Models



Source: https://www.kdnuggets.com/2019/05/machine-learning-time-series-forecasting.html

Time series Models

Different Approaches

- Regression
  - AR
  - ARMA
  - ARIMA
  - SARIMA
  - …
- Neural Networks
  - MLP
  - RNN
  - LSTM
  - …

Regression

Next time step as a function of some number of past (or lag) observations.

This is a common approach for classical statistical time series forecasting.

Regression

AutoRegressive (AR)

$$\check{Y}_t = \Phi_1 \Upsilon_{t-1} + \ldots + \Phi_1 \Upsilon_{t-p} + \varepsilon_t + c$$

AR(p) :

- p is the order (number of time lags)

Regression

AutoRegressive Moving Average (ARMA)

$$\check{Y}_t = c + \varepsilon_t + \Phi_1 \Upsilon_{t-1} + \ldots + \Phi_p \Upsilon_{t-p} + \Theta_1 \varepsilon_{t-1} \ldots + \Theta_q \varepsilon_{t-q}$$

ARMA(p,q) :

- p is the order (number of time lags)
- q is the order of the moving-average model

Regression

AutoRegressive Integrated Moving Average (ARIMA)

$$\check{Y}_t = \Phi_1 \Upsilon_{t-1} \ldots + \Phi_1 \Upsilon_{t-p} + a_t - \Theta_1 a_{t-1} \ldots - \Theta_p a_{t-q}$$

ARIMA(p,d,q) :

- p is the order (number of time lags)
- d s the degree of differencing
- q is the order of the moving-average model

Regression

Seasonal Autoregressive Integrated Moving Average (SARIMA)

An extension of ARIMA

Supports seasonal data

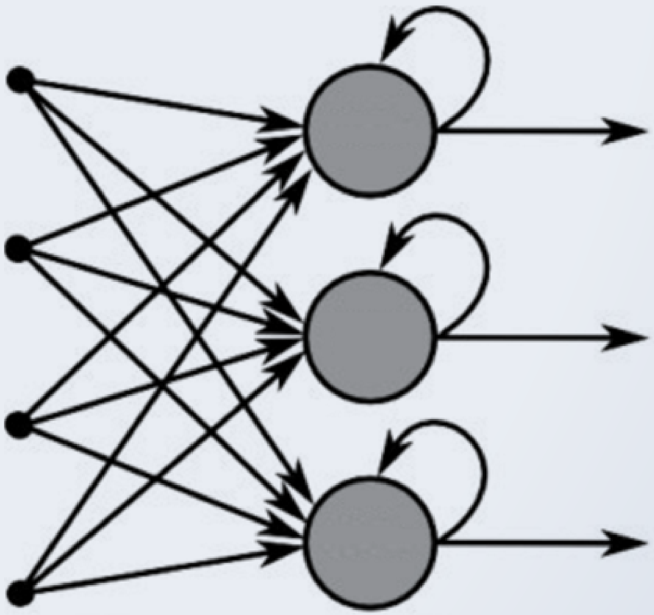
SARIMA(p,d,q) (P,D,Q)m :

- p is the order (number of time lags)
- d s the degree of differencing
- q is the order of the moving-average model

- P: Seasonal autoregressive order.
- D: Seasonal difference order.
- Q: Seasonal moving average order.
- m: The number of time steps for a single seasonal period.

Long Short Time Memory

LSTM uses an artificial RNN architecture



Long Short Time Memory

With LSTM we can have some temporal information “stored” while:

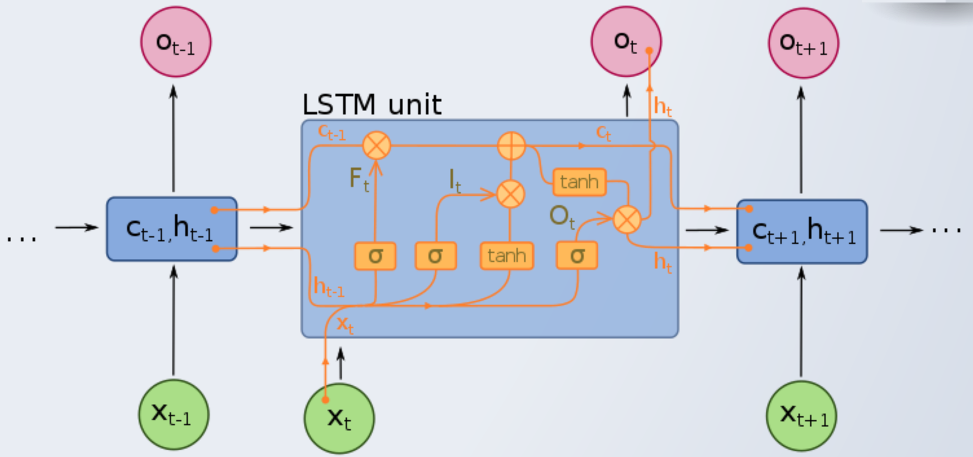
- we train the model
- we produce predictions

Deals with vanishing gradients and address exploding gradients

Long Short Time Memory

LSTM Unit:

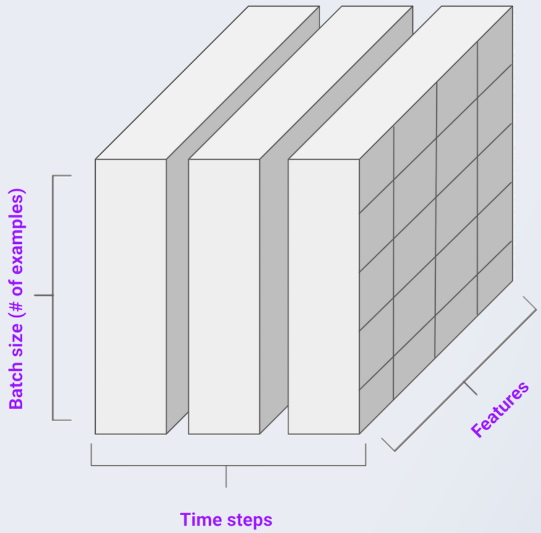
- cell
- input gate
- output gate
- forget gate



Source: https://en.wikipedia.org/wiki/Long_short-term_memory#/media/File:The_LSTM_cell.png

Long Short Time Memory

Data Input

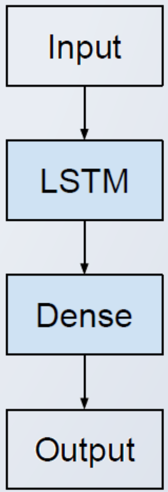


Source: https://www.tensorflow.org/tutorials/structured_data/time_series

Long Short Time Memory

Can be applied in different architectures

- Vanilla
- Stacked
- Encoder-Decoder
- Bidirectional
- ...



Practical Case

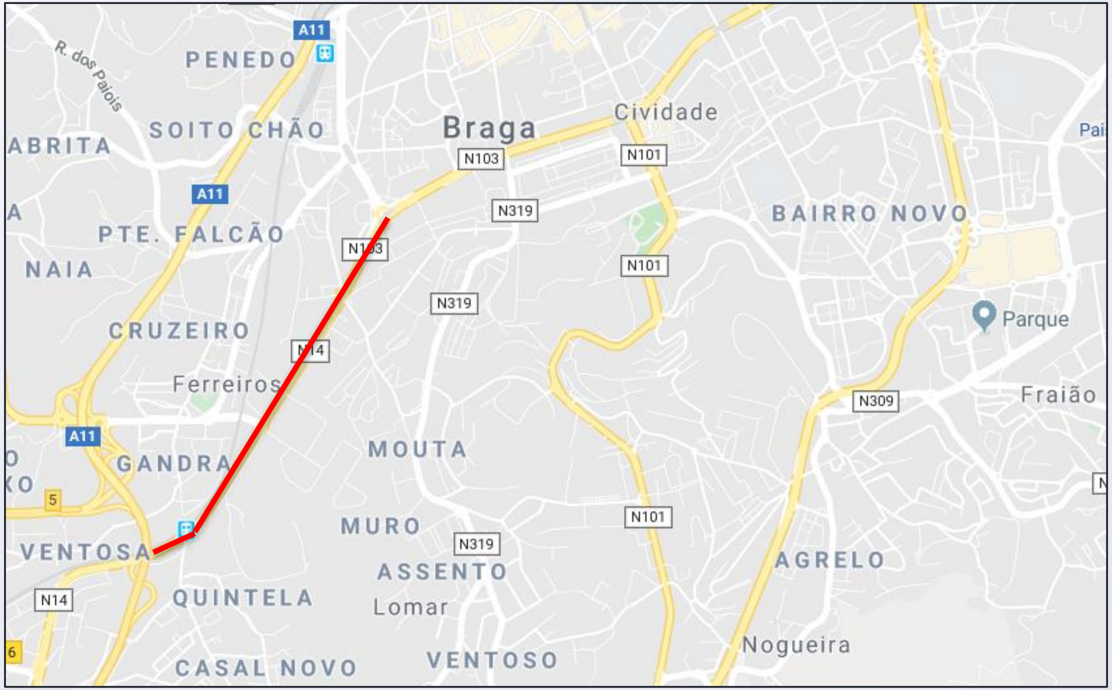
Traffic Analysis

- Smart City problem
- Discovering safer routes

Tile: Traffic Flow Forecasting on Data-Scarce Environments using ARIMA and LSTM Networks
Doi: 10.1007/978-3-030-16181-1_26

Images source: https://www.flaticon.com/authors/freepik

Practical Case



Adapted from : https://www.google.com/maps/

Practical Case

#	Features
1	City Name
2	Road Number
3	Road Name
4	Road Category
5	Current Speed
6	Free Flow Speed
7	Speed Diference (#6 - #5)
8	Current Travel Time
9	Free Flow Travel Time
10	Time Diff (#9 - #8)
11	Creation Date

#	Features
1	City Name
2	Wheather Description
3	Temperature
4	Atmospheric Pressure
5	Humidity
6	Wind Speed
7	Cloudiness
8	Precipition
9	Current Luminosity
10	Sunrise
11	Sunset
12	Creation Date

Images source: https://www.flaticon.com/authors/freepik

Practical Case

Assess performance of time series models:

- LSTM
- ARIMA

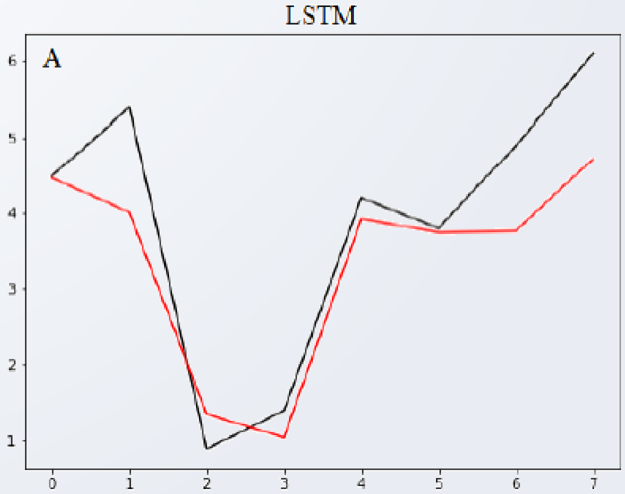
Grid search for parameter optimization

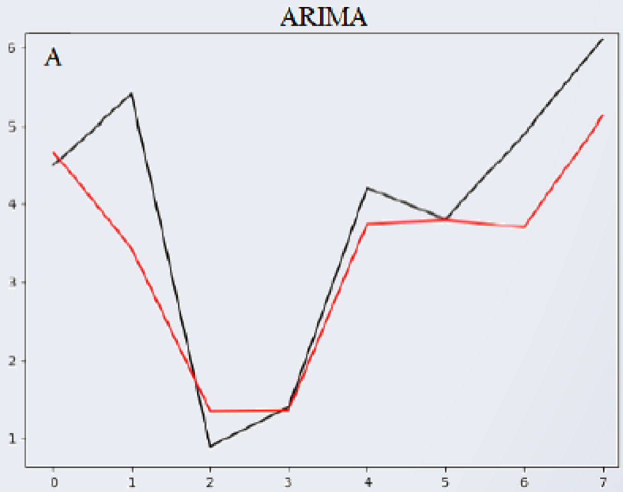
Walk Forward Validation

Practical Case – Results

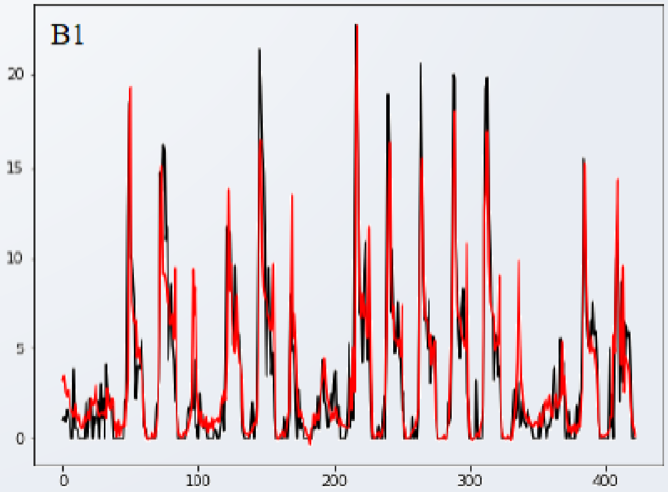
Dataset	Model	MAE	RMSE	Parameters
A	ARIMA	0.662	0.919	(10,1,1)
A	LSTM	0.703	0.905	Window Size=7 Batch Size = 20
B0	ARIMA	1.845	2.797	(3,1,1)
B0	LSTM	1.410	2.169	Window Size=24 Batch Size = 20
B1	ARIMA	1.858	2.659	(7,0,1)
B1	LSTM	1.479	2.304	Window Size=24 Batch Size = 50
B2	ARIMA	1.853	2.683	(3,0,1)
B2	LSTM	1.407	1.912	Window Size=48 Batch Size = 30

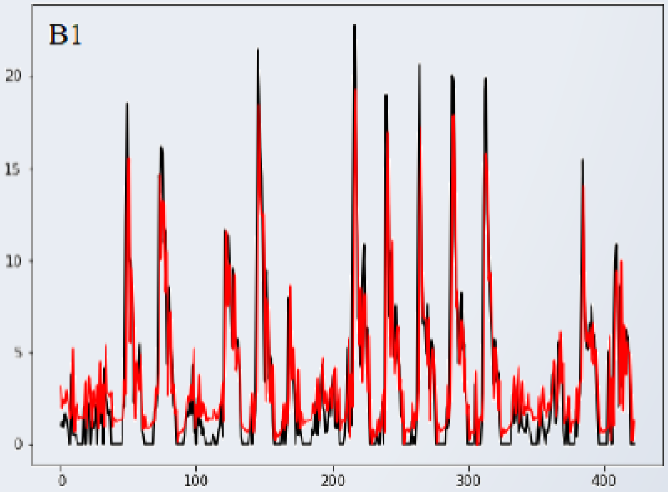
Practical Case – Results





Practical Case – Results





Practical Case – Results

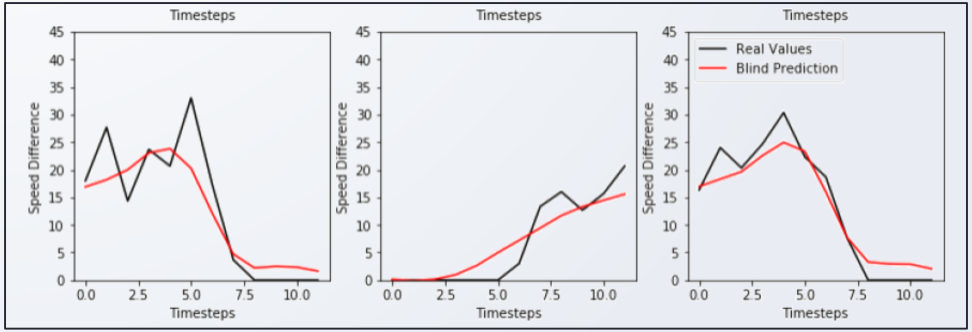
› LSTM Univariante vs multivariate

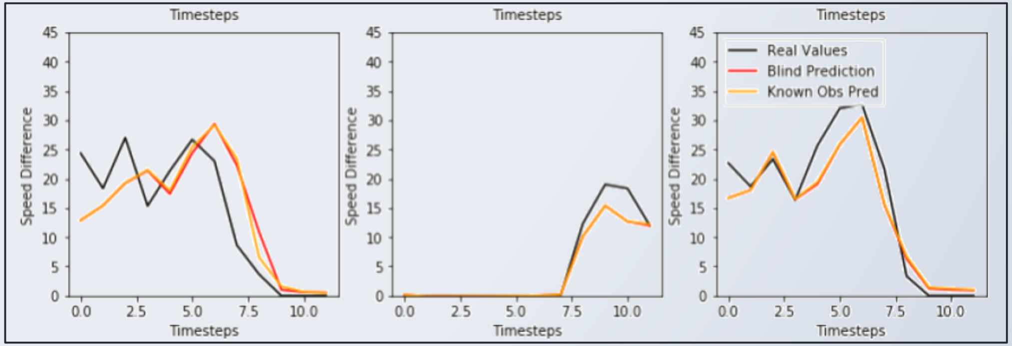
#	Timesteps	Batch	Layers	Neurons	Dropout	Act.	RMSE	MAE
209	**96**	**672**	**5**	**64**	**0.2**	**relu**	**3.496**	**1.567**
188	96	672	4	64	0.2	tanh	3.518	1.567
95	96	252	5	32	0.2	relu	3.555	1.592
195	96	672	4	128	0.5	tanh	3.583	1.598
12	48	252	3	64	0.5	relu	3.649	1.629

#	Timesteps	Batch	Layers	Neurons	Dropout	Act.	RMSE	MAE
53	**24**	**672**	**4**	**64**	**0.5**	**tanh**	**2.907**	**1.346**
24	24	672	4	32	0.5	tanh	3.006	1.412
16	24	252	4	32	0.5	tanh	3.031	1.419
17	24	672	5	64	0.5	tanh	3.037	1.402
37	48	252	2	64	0.5	tanh	3.038	1.425

Practical Case – Results

› LSTM Univariante vs multivariate





Practical Case – Results

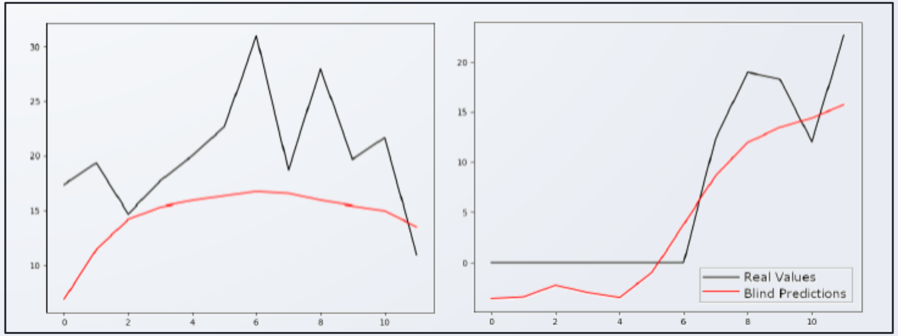
› ARIMA univariante vs multivariate

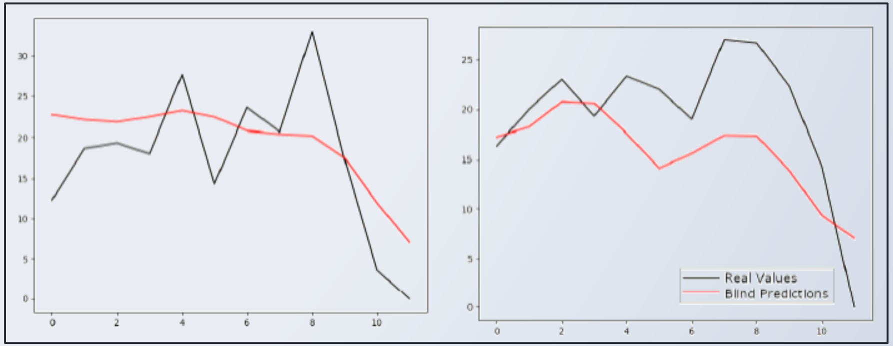
p	d	q	RMSE	MAE
8	1	1	8.869	7.474
8	1	2	7.967	6.275
12	1	1	6.452	5.200
12	**1**	**2**	**6.336**	**5.088**

p	d	q	RMSE	MAE
8	1	1	8.325	6.736
8	1	2	8.301	6.822
12	1	1	6.171	4.972
12	**1**	**2**	**6.110**	**4.918**

Practical Case – Results

› ARIMA Univariante vs multivariate





Practical Case – Results

Autoregression:

- depend on mathematical formulas
- less parameters
- less time to train
- smaller lag

LSTM:

- restrictions on data format
- more parameters to optimize
- more time to train

Practical Case – Results

Regression based modes

- some parametrization are impossible to compute valid models
- often requires data transformations
- overfit does not occur as often
- need constant model update for further predictions

Practical Case – Results

LSTMs:

- generally take longer to train
- require more memory and time to train
- easy to overfit
- handle bigger lags better
- desirable better hardware (GPU) to run efficiently
- models can be trained once

Pratical Case - Results

The findings on the practical case have been similar in other fields such as:

- Retail Sales
- Driving Patterns

Final Notes

On the literature contrary results can be found where LSTM performance is lower than other counterparts

Final Notes

About Me

- Teaching at ESTG, Polytechnique Institute of Porto
- Researcher at CIICESI and collaborator at ALGORITMI Centre
- Participation in several research projects and collaborations with international researchers

- Research Interests:
  - Intelligent Systems
  - Ubiquitous computing
  - Smart City
  - Industry 4.0

THANKS!

Any questions?

Long Short-Term Memory Networks applied in Time Series

Fábio Silva