Time Series Forecasting using Global Models

Presented by

Rakshitha Godahewa

Supervised by

Professor Geoff Webb
Dr. Christoph Bergmeir

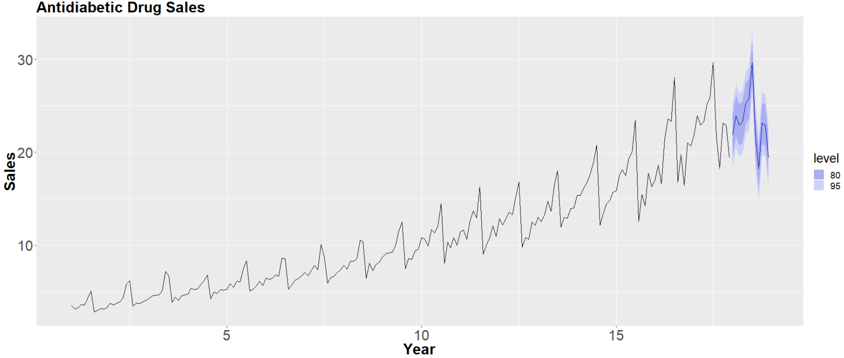
Faculty of Information Technology, Monash University

Overview

Introduction

Univariate Time Series Forecasting

- Predicting future values of a particular quantity based on its own past values and possibly some other time varying exogenous series which influence the variable under study



Accurate Forecasting

- Many businesses and industries routinely collect **massive amounts of related time series** that require accurate forecasts
- Recently, **Global Forecasting Models (GFM)** have shown a huge potential in providing accurate forecasts for collections of time series compared to the local models

Local Vs Global Time Series Forecasting (Januschowski et al., 2020)

Local Models	Global Models
Build one forecasting model per each series	Build one forecasting model across many series
Fit local parameters separately on each series	Fit global parameters that are same across many series
Improve the performance of models over individual series	Improve the performance of the model over many series

Global Ensemble Models for Forecasting

- Ensembling is now very important for global models, as it can be a solution to heterogeneous data
- Aggregates predictions provided by a series of prediction models
- Mitigates the problem of providing poor forecasts by single forecasting models
- Addresses the data, model and parameter uncertainties
- Solutions used ensemble models have won many forecasting competitions
  e.g. Exponential Smoothing-Recurrent Neural Network (ES-RNN, Smyl, 2020)

Our Achievements with Global Models and Ensembling

- **M5 Forecasting Competition 2020 (Accuracy Stream) - Global 17th Position**
  - Ensemble model containing a global Pooled Regression (PR) model and a global Light Gradient Boosting (LightGBM) model
- **IEEE CIS Technical Challenge 2020 - Global 4th Position**
  - Ensemble model containing multiple global models: Neural Networks, Random Forests and PR
- **Decathlon Forecasting Competition 2019 - 1st Position**
  - Ensemble model containing a global Recurrent Neural Network (RNN) and an Auto-Regressive Integrated Moving Average (ARIMA) model

Project 1: Motivation

- There are no proper benchmarking archives available in time series forecasting space containing sets of time series which are required to train and evaluate GFMs
  - Datasets with many series
  - Datasets with long series
  - Datasets with diverse characteristics
- The performance of GFMs can be evaluated across a large number of datasets

Monash Time Series Forecasting Repository

- Repository available at `https://forecastingdata.org/`
- Implementations available at `https://github.com/rakshitha123/TSForecasting`
- Contains 26 publicly available univariate time series datasets

Summary of Datasets

	Dataset	Domain	No: of Series	Min. Length	Max. Length	No: of Freq.	Missing	Competition
1	M1	Multiple	1001	15	150	3	No	Yes
2	M3	Multiple	3003	20	144	4	No	Yes
3	M4	Multiple	100000	19	9933	6	No	Yes
4	Tourism	Tourism	1311	11	333	3	No	Yes
5	NN5	Banking	111	791	791	2	Yes	Yes
6	CIF 2016	Banking	72	34	120	1	No	Yes
7	Web Traffic	Web	145063	803	803	2	Yes	Yes
8	Solar	Energy	137	52560	52560	2	No	No
9	Electricity	Energy	321	26304	26304	2	No	No
10	London Smart Meters	Energy	5560	288	39648	1	Yes	No
11	Wind Farms	Energy	339	6345	527040	1	Yes	No
12	Car Parts	Sales	2674	51	51	1	Yes	No
13	Dominick	Sales	115704	28	393	1	No	No
14	FRED-MD	Economic	107	728	728	1	No	No
15	San Francisco Traffic	Transport	862	17544	17544	2	No	No
16	Pedestrian Counts	Transport	66	576	96424	1	No	No
17	Hospital	Health	767	84	84	1	No	No
18	COVID Deaths	Nature	266	212	212	1	No	No
19	KDD Cup	Nature	270	9504	10920	1	Yes	Yes
20	Weather	Nature	3010	1332	65981	1	No	No
21	Sunspot	Nature	1	73931	73931	1	Yes	No
22	Saugeen River Flow	Nature	1	23741	23741	1	No	No
23	US Births	Nature	1	7305	7305	1	No	No
24	Electricity Demand	Energy	1	17520	17520	1	No	No
25	Solar Power	Energy	1	7397222	7397222	1	No	No
26	Wind Power	Energy	1	7397147	7397147	1	No	No

Format of Datasets

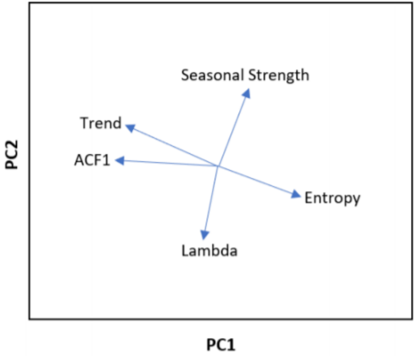
- .tsf format

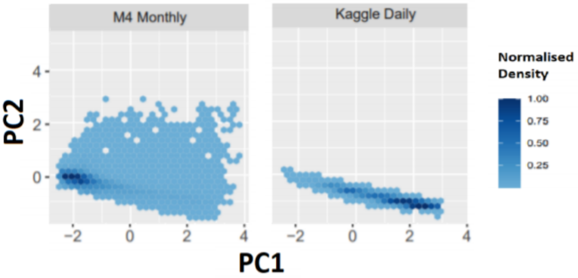
```
# Dataset Information
# This dataset was used in the NN5 forecasting competition.
# It contains 111 daily time series from the banking domain.
# The goal is predicting the daily cash withdrawals from ATMs in UK.
#
# For more details, please refer to
# Taieb, S.B., Bontempi, G., Atiya, A.F., Sorjamaa, A., 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competit
# Neural Forecasting Competitions, 2008. NN5 forecasting competition for artificial neural networks and computational intelligence. Accessed: 2020-05-10. URL http://www.neural-foreca
#
@relation NN5
@attribute series_name string
@attribute start_timestamp date
@frequency daily
@horizon 56
@missing true
@equallength true
@data
T1:1996-03-18 00-00-00:13.4070294784581,14.7250566893424,20.5640589569161,34.7080498866213,26.6298185941043,16.609977324263,15.3202947845805,11.6071428571429,19.8837868480726,23.7670
T2:1996-03-18 00-00-00:11.5504535147392,13.5912698412698,15.0368480725624,21.5702947845805,19.4444444444444,0,9.72222222222222,12.2448979591837,15.5045351473923,18.9342403628118,26.0
T3:1996-03-18 00-00-00:5.640589569161,14.3990929705215,24.4189342403628,28.7840136054422,20.6207482993197,13.8038548752834,11.5362811791383,10.7426303854875,14.8242630385488,25.21258
T4:1996-03-18 00-00-00:13.1802721088435,8.44671201814059,19.515306122449,28.8832199546485,19.4727891156463,0,7.35544217687075,10.827664399093,15.6179138321995,21.1592970521542,33.489
T5:1996-03-18 00-00-00:9.77891156462585,10.8134920634921,21.6128117913832,38.5204081632653,24.7448979591837,12.3299319727891,12.9960317460317,11.0402494331066,7.95068027210884,19.515
T6:1996-03-18 00-00-00:9.24036281179138,11.6354875283447,12.1031746031746,21.4143990929705,24.6740362811791,5.22959183673469,11.3803854875283,9.55215419501134,15.717120181 4059,14.866
T7:1996-03-18 00-00-00:14.937641723356,16.2840136054422,16.6666666666667,23.5685941043084,26.3038548752834,14.8951247165533,16.0430839002268,13.8888888888889,20.3939909297052,19.3452
T8:1996-03-18 00-00-00:2.89115646258503,12.3582766439909,16.3832199546485,30.1587301587302,31.1791383219955,19.812925170068,17.4886621315193,8.71598639455782,15.5045351473923,21.4852
T9:1996-03-18 00-00-00:7.34126984126984,9.15532879818594,10.5867346938776,12.5,7.15702947845805,5.640589569161,7.69557823129252,5.72562358276644,5.51303854875283,8.41836734693878,13.
T10:1996-03-18 00-00-00:10.2891156462585,12.7125850340136,14.4416099773243,19.4019274376417,21.5419501133787,15.2210884353741,11.3520408163265,11.1394557823129,13.9314058956916,17.34
T11:1996-03-18 00-00-00:13.9314058956916,12.953514739229,19.0192743764172,27.5368480725624,24.1780045351474,21.499433106576,19.06179138322,12.5425170068027,15.0651927437642,21.357709
T12:1996-03-18 00-00-00:10.1473922902494,15.4620181405896,14.5691609977324,19.189342403628 1,21.9812925170068,9.43877551020408,12.7551020408163,8.94274376417234,11.5646258503401,15.39
T13:1996-03-18 00-00-00:12.0890022675737,11.3803854875283,18.4807256235828,24.702380952381,30.031179138322,18.43820861678,13.9030612244898,11.6496598639456,15.0368480725624,21.031746
T14:1996-03-18 00-00-00:11.281179138322,12.7125850340136,17.4886621315193,23.3843537414966,22.6473922902494,11.8480725623583,14.4557823129252,10.8843537414966,13.1660997732426,18.197
T15:1996-03-18 00-00-00:8.23412698412698,13.109410430839,18.3531746031746,25.921201814059,19.6286848072562,11.1111111111111,10.1048752834467,9.35374149659864,10.1757369614512,18.2681
T16:1996-03-18 00-00-00:8.51757369614512,14.4982993197279,30.5697278911565,45.422335600907,48.7386621315193,0,15.249433106576,13.2794784580499,16.7233560090703,29.8185941043084,51.62
T17:1996-03-18 00-00-00:15.0368480725624,15.6320861678005,17.6587301587302,29.3509070294785,35.7851473922902,9.42460317460317,18.7641723356009,17.7295918367347,20.464852607709 8,27.65
T18:1996-03-18 00-00-00:11.621315192743 8,13.2936507936508,27.4518140589569,33.843537414966,32.2420634920635,14.1156462585034,18.0839002267574,15.1218820861678,17.7437641723356,31.221
T19:1996-03-18 00-00-00:15.858843537415,17.0209750566893,30.6547619047619,44.7137188208617,30.3146258503401,18.8492063492063,21.4002267573696,16.1848072562358,17.5028344671202,32.114
T20:1996-03-18 00-00-00:13.2227891156463,12.6133786848073,12.953514739229,21.3577097505669,22.108843537415,15.4761904761905,12.5283446712018,12.0323129251701,14.3565759637188,17.3894
T21:1996-03-18 00-00-00:13.3786848072562,13.9597505668934,14.2290249433107,24.1921768707483,21.8112244897959,?,10.3741496598639,12.1598639455782,14.4274376417234,17.4886621315193,29.
T22:1996-03-18 00-00-00:11.9331065759637,10.0481859410431,20.3231292517007,22.7040816326531,5.99489795918367,?,16.0289115646259,10.0907029478458,11.5504535147392,19.8270975056689,23.
```

Feature Analysis

- Extracted 42 tsfeatures (Hyndman et al., 2019) and catch22 features (Lubba et al., 2019) from each time series
- **Consider 5 important features for feature analysis**
  - First order autocorrelation (ACF1)
  - Trend
  - Entropy
  - Seasonal strength
  - Lambda
- Apply Principal Component Analysis (PCA, Jolliffe, 2011) to reduce the feature dimensionality

Feature Analysis Plots





Feature Analysis Conclusions

- M competition datasets show strong trend and ACF1 levels
- Intermittent datasets such as Kaggle Web Traffic show high degree of entropy
- The monthly datasets generally show high seasonal strengths compared to the other frequencies
- The datasets with high frequencies such as weekly, daily and hourly show low seasonal strengths

Evaluation

- **Benchmarks**
  - Exponential Smoothing State Space Model (**ETS**, Hyndman, 2008)
  - **ARIMA** (Box and Jenkins, 1990)
  - Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (**TBATS**, Livera et al., 2011)
  - Dynamic Harmonic Regression ARIMA (**DHR-ARIMA**, Hyndman et al., 2015)
  - **Theta** (Assimakopoulos and Nikolopoulos, 2000)
  - Simple Exponential Smoothing (**SES**, Hyndman, 2008)
  - A globally trained **PR** model (Trapero et al., 2015)
- **Performance Measures**
  - Symmetric Mean Absolute Percentage Error (**SMAPE**, Suilin, 2017)
  - Mean Absolute Scaled Error (**MASE**, Hyndman and Koehler, 2006)
  - Mean Absolute Error (**MAE**, Sammut and Webb, 2010)
  - Root Mean Squared Error (**RMSE**)

Evaluation

- **Benchmarks**
  - Exponential Smoothing State Space Model (**ETS**, Hyndman, 2008)
  - **ARIMA** (Box and Jenkins, 1990)
  - Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (**TBATS**, Livera et al., 2011)
  - Dynamic Harmonic Regression ARIMA (**DHR-ARIMA**, Hyndman et al., 2015)
  - **Theta** (Assimakopoulos and Nikolopoulos, 2000)
  - Simple Exponential Smoothing (**SES**, Hyndman, 2008)
  - A globally trained **PR** model (Trapero et al., 2015)
- **Performance Measures**
  - Symmetric Mean Absolute Percentage Error (**SMAPE**, Suilin, 2017)
  - Mean Absolute Scaled Error (**MASE**, Hyndman and Koehler, 2006)
  - Mean Absolute Error (**MAE**, Sammut and Webb, 2010)
  - Root Mean Squared Error (**RMSE**)

PR Results

Dataset	mean sMAPE	median sMAPE	mean MASE	median MASE	mean MAE	median MAE	mean RMSE	median RMSE
NN5 Daily	22.23	21.27	0.873	0.835	3.78	3.51	5.31	4.91
NN5 Weekly	12.34	10.66	0.908	0.808	15.69	14.50	18.87	17.70
CIF 2016	12.32	8.43	1.019	0.746	563205.57	95.13	648890.31	109.09
Tourism Yearly	45.84	16.62	3.457	2.356	80263.27	4141.82	86326.86	4652.24
Tourism Quarterly	14.83	12.23	1.505	1.254	8007.72	863.48	10624.97	1039.61
Tourism Monthly	19.16	16.46	1.485	1.349	1939.54	435.35	2429.51	540.99
Traffic Hourly	5.49	4.96	0.349	0.343	0.02	0.01	0.02	0.02
Electricity Hourly	25.58	20.41	0.490	0.474	428.08	111.59	581.62	148.86
M3 Yearly	17.43	11.62	2.948	2.032	1024.75	662.79	1182.58	746.46
M3 Quarterly	9.55	5.40	1.227	0.889	509.04	300.88	590.15	354.57
M3 Monthly	15.08	10.29	1.003	0.818	687.76	474.80	824.20	571.76
M4 Yearly	14.26	8.99	3.459	2.480	856.88	437.45	979.82	504.41
M4 Quarterly	10.81	6.33	1.309	1.030	608.32	294.62	707.32	346.20
M4 Monthly	13.71	8.17	1.078	0.844	595.19	280.39	719.34	332.37
M4 Weekly	9.29	5.16	0.605	0.501	349.59	214.65	426.26	269.56
M4 Daily	3.07	2.02	1.164	0.871	182.42	93.15	213.50	108.74
M4 Hourly	15.34	5.25	2.250	1.378	299.17	19.94	363.29	26.07
Car Parts	43.23	30.30	0.755	0.375	0.41	0.25	0.73	0.58
Hospital	17.56	16.12	0.782	0.740	19.24	6.67	23.48	8.25

Project 2: Overview & Contributions

Conduct an extensive empirical study on the performance of the ensembled GFMs that contain sub-models trained on different parts of the datasets at a time to address localisation and heterogeneity issues.

- Systematically compare the existing GFM localisation approaches
- Develop new ensembled localised GFMs which can provide more accurate forecasts for heterogeneous time series datasets compared to non-ensemble localisation approaches

Base GFMs

- Feed-Forward Neural Networks (FFNN, Goodfellow et al., 2016)
- RNN (Hewamalage et al., 2020)
- PR (Trapero et al., 2015)

Ensemble Models

- **Clustering**

Name	Details	Techniques
Feature clustering	Extract features from series. Cluster features. Cluster time series corresponding with feature clusters. Train a GFM per each cluster.	Kmeans (Lloyd, 1982) Kmeans++ (Arthur and Vassilvitskii, 2007)
Distance-based Clustering	Directly cluster time series. Train a GFM per each cluster.	K-medoids (Jin and Han, 2010) with Dynamic Time Warping (DTW) distance

- **Ensemble of Specialists** (Smyl, 2020)
- **Forecast Combinations with Global and Local Models**

Ensemble Models

- **Clustering**

Name	Details	Techniques
Feature clustering	Extract features from series. Cluster features. Cluster time series corresponding with feature clusters. Train a GFM per each cluster.	Kmeans (Lloyd, 1982) Kmeans++ (Arthur and Vassilvitskii, 2007)
Distance-based Clustering	Directly cluster time series. Train a GFM per each cluster.	K-medoids (Jin and Han, 2010) with Dynamic Time Warping (DTW) distance

- **Ensemble of Specialists** (Smyl, 2020)
- **Forecast Combinations with Global and Local Models**

Ensemble Models

- **Clustering**

Name	Details	Techniques
Feature clustering	Extract features from series. Cluster features. Cluster time series corresponding with feature clusters. Train a GFM per each cluster.	Kmeans (Lloyd, 1982) Kmeans++ (Arthur and Vassilvitskii, 2007)
Distance-based Clustering	Directly cluster time series. Train a GFM per each cluster.	K-medoids (Jin and Han, 2010) with Dynamic Time Warping (DTW) distance

- **Ensemble of Specialists** (Smyl, 2020)
- **Forecast Combinations with Global and Local Models**

New Clustered Ensemble Models

- **GFM.Cluster.Number**
  - Use different cluster numbers (e.g. 2 to 7) in multiple iterations and create multiple cluster sets
  - Train a GFM per each cluster
  - Average the predictions provided by each cluster
- **GFM.Cluster.Seed**
  - Find the optimal number of clusters to be used using Elbow method (Kaufman and Rousseeuw, 1990)
  - Create cluster sets with different seeds in multiple iterations
  - Train a GFM per each cluster
  - Average the predictions provided by each cluster

New Clustered Ensemble Models Cont.

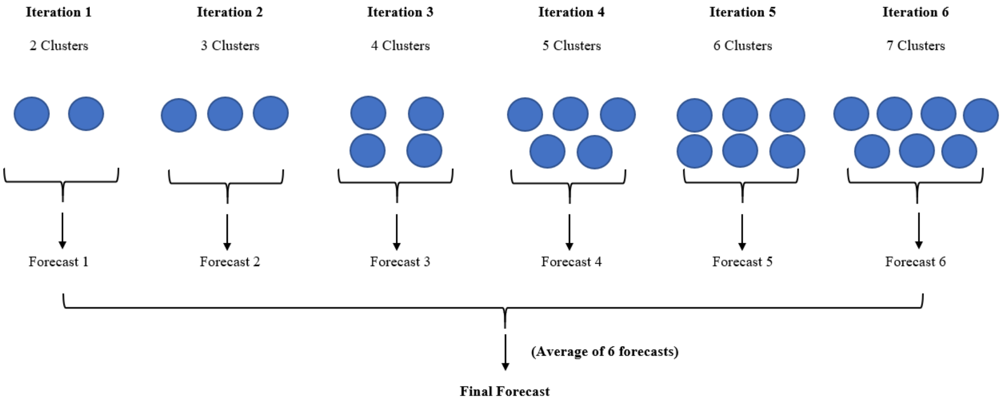


Figure 1: Ensemble Model based on Clustering

Non-Ensemble Global Models

- Baseline GFMs
- **Non-ensemble models using feature clustering with optimal number of clusters**
  - Kmeans clustering with Elbow method (Kaufman and Rousseeuw, 1990)
  - Kmeans++ clustering with Elbow method
  - Xmeans clustering (Pelleg and Moore, 2000)

Datasets

Table 1: Datasets for the Experiments

Dataset Name	No. of Time Series	Forecasting Horizon	Frequency	Min. Length	Max. Length
M3	1428	18	Monthly	48	126
M4	48000	18	Monthly	42	2794
CIF 2016	72	6, 12	Monthly	22	108
Kaggle Web Traffic	997	59	Daily	550	550
Ausgrid Half-Hourly	300	96	Half-Hourly	1488	1488
Ausgrid Monthly	1405	12	Monthly	38	88

Evaluation

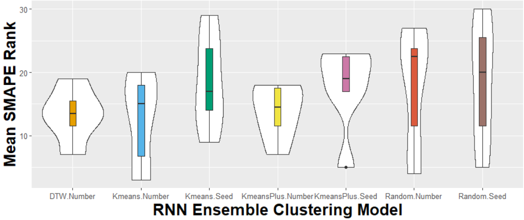
- **Benchmarks**
  - ETS
  - ARIMA
  - TBATS
  - DHR-ARIMA
  - Non-ensemble GFMs
  - Feature-based Forecast Model Averaging (FFORMA, Montero-Manso et al., 2020)
- **Performance Measures**
  - SMAPE
  - MASE

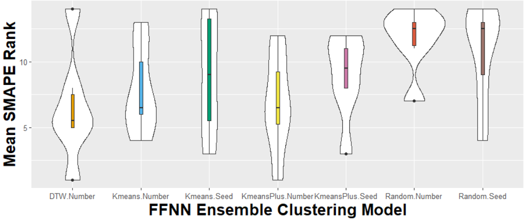
Evaluation

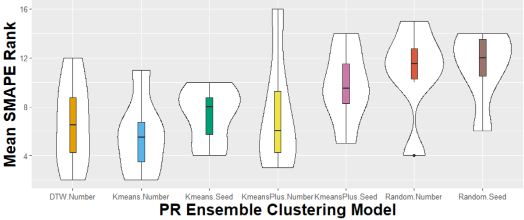
- **Benchmarks**
  - ETS
  - ARIMA
  - TBATS
  - DHR-ARIMA
  - Non-ensemble GFMs
  - Feature-based Forecast Model Averaging (FFORMA, Montero-Manso et al., 2020)
- **Performance Measures**
  - SMAPE
  - MASE

Relative Performance of Clustered Ensemble Models - Mean SMAPE

DTW.Number is the best on RNNs and FFNNs
Kmeans.Number is the best on PR







Projects 2: Conclusions

- **Clustered ensemble models** show an overall better performance than clustered non-ensemble models across all base GFMs
- **Clustered ensemble models** show an overall better performance than ensemble of specialists across all base GFMs and FFORMA
- **GFM.Cluster.Number** is a better ensembling approach than GFM.Cluster.Seed
- **GFMs** show an overall better performance than local models
- **RNNs and PR models** show a competitive performance over FFNNs
- **Forecast combinations** of global and local models show the best performance across many datasets as they incorporate the strengths of both global and local models

Projects 2: Publications

Ensembles of Localised Models for Time Series Forecasting
Available at: https://arxiv.org/abs/2012.15059
Authors: Rakshitha Godahewa, Kasun Bandara, Geoffrey I. Webb, Slawek Smyl, Christoph Bergmeir

Project 3: Overview & Contributions

Develop a new automated ensemble time series forecasting framework which can provide accurate forecasts for weekly time series.

- Identify suitable base models to develop a weekly forecasting framework
- Design a fully automated forecast combination approach which can provide accurate forecasts for weekly time series (Available at https://github.com/rakshitha123/WeeklyForecasting)

Project 3: Motivation

- Many businesses are now operating on a weekly time scale
- Long and non-integer seasonal cycles presented in weekly series should be carefully handled
- Even though there are accurate weekly forecasting frameworks (Darin and Stellwagen, 2020), they require a significant amount of manual intervention and domain knowledge

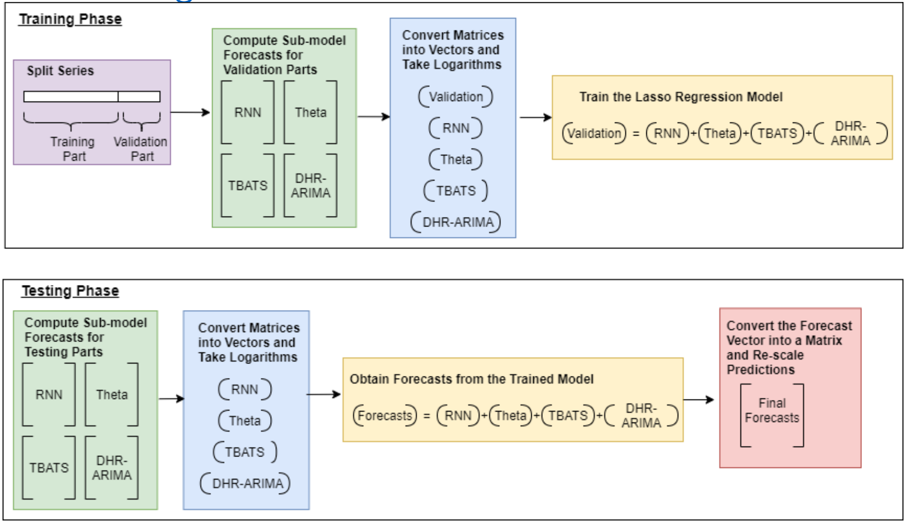
Base Models

Model	Reason to Use	Method of Handling Seasonality
TBATS (Livera et al., 2011)	Capability of modelling long/non-integer periodic effects Current state-of-the-art in weekly forecasting space	Fourier terms
DHR-ARIMA (Hyndman et al., 2015)	Capability of modelling long/non-integer periodic effects Current state-of-the-art in weekly forecasting space	Fourier terms
Theta (Assimakopoulos and Nikolopoulos, 2000)	Adding diversity to the pool of models Winner of the M3 forecasting competition	Non-seasonal model
Global RNN (Hewamalage et al., 2020)	Adding diversity to the pool of models Learning cross-series information Ability to deal with short series	Fourier terms Seasonal lags

Forecast Combination Method

- **Approach:** Stacking (Wolpert, 1992)
- **Meta-learner:** Lasso regression (Tibshirani, 1994)
  - Prevents model overfitting by adding a regularisation term (L1) to the cost function of linear regression
  - Coefficients corresponding with the worst performing models shrink towards zero
  - A global model

Training and Testing Phases



Datasets

Table 2: Summary of the Datasets Used to Evaluate Our Proposed Weekly Forecasting Model

Dataset	**No. of Series**	**Forecast Horizon**	**Min. Length**	**Max. Length**
M4	359	13	80	2597
NN5	111	8	105	105
Web Traffic	1000	8	106	106
Ausgrid	299	8	148	148
Traffic	862	8	96	96
Solar	137	5	44	44

Evaluation

- **Benchmarks**
  - 4 base models: TBATS, DHR-ARIMA, Theta and RNN
  - Simple average of the base models
  - FFORMA (Montero-Manso et al., 2020)
  - Modified version of FFORMA that use our 4 base models instead of its original base models
- **Performance Measures**
  - SMAPE (Suilin, 2017)
  - MASE (Hyndman and Koehler, 2006)

Evaluation

- **Benchmarks**
  - 4 base models: TBATS, DHR-ARIMA, Theta and RNN
  - Simple average of the base models
  - FFORMA (Montero-Manso et al., 2020)
  - Modified version of FFORMA that use our 4 base models instead of its original base models
- **Performance Measures**
  - SMAPE (Suilin, 2017)
  - MASE (Hyndman and Koehler, 2006)

Variants of Our Proposed Model

- **Different meta-learners instead of lasso regression**
  - Linear regression
  - eXtreme Gradient Boosting (XGBoost, Chen and Guestrin, 2016)
- Adding more base models
- Train separate meta-learners per each horizon

Project 3: Conclusions

- **Overall, our proposed weekly forecasting model significantly outperforms all considered benchmarks and variants**
  - Lasso regression is a better option to be used as a meta-learner compared to linear regression and XGBoost due to its model selection capability
  - When the base models are more suitable for weekly forecasting, the final forecasts get more accurate
  - Training a global meta-learner across all horizons is a better option compared to training multiple meta-learners per each horizon
- Our proposed model provides the best forecasts for the M4 weekly dataset based on mean SMAPE
- Hence, our proposed model is a fully automated, accurate and strong baseline in weekly time series forecasting space

Project 3: Publications

A Strong Baseline for Weekly Time Series Forecasting
Available at: https://arxiv.org/abs/2010.08158
Authors: Rakshitha Godahewa, Christoph Bergmeir, Geoffrey I. Webb, Pablo Montero-Manso

Concluding Remarks & Future Research Directions

- Global modelling and ensembling support to get more accurate forecasts
- Probabilistic forecasting
- Multivariate forecasting

Thank you

rakshitha.godahewa@monash.edu

References I

Arthur, D., Vassilvitskii, S., 2007. K-means++: The advantages of careful seeding, in: Proc. of the Annu. ACM-SIAM Symp. on Discrete Algorithms, pp. 1027–1035. doi:10.1145/1283383.1283494.

Assimakopoulos, V., Nikolopoulos, K., 2000. The theta model: a decomposition approach to forecasting. International Journal of Forecasting 16, 521–530. doi:https://doi.org/10.1016/S0169-2070(00)00066-2. the M3- Competition.

Box, G., Jenkins, G., 1990. Time Series Analysis, Forecasting and Control. HoldenDay Inc.

References II

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 785–794. URL: https://doi.org/10.1145/2939672.2939785, doi:10.1145/2939672.2939785.

Darin, S.G., Stellwagen, E., 2020. Forecasting the m4 competition weekly data: Forecast pro's winning approach. International Journal of Forecasting 36, 135 – 141. doi:https://doi.org/10.1016/j.ijforecast.2019.03.018. m4 Competition.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. http://www.deeplearningbook.org.

References III

Hewamalage, H., Bergmeir, C., Bandara, K., 2020. Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting doi:https://doi.org/10.1016/j.ijforecast.2020.06.008.

Hyndman, R., 2008. Forecasting with exponential smoothing: the state space approach. Springer.

Hyndman, R., Kang, Y., Talagala, T., Wang, E., Yang, Y., 2019. tsfeatures: Time series feature extraction. r package version 1.0.1. https://pkg.robjhyndman.com/tsfeatures/.

Hyndman, R.J., Athanasopoulos, G., S. Razbash, D.S., Zhou, Z., Khan, Y., Bergmeir, C., Wang, E., 2015. forecast: Forecasting functions for time series and linear models. R package version 6, 7.

References IV

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688.

Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., Callot, L., 2020. Criteria for classifying forecasting methods. International Journal of Forecasting 36, 167–177.

Jin, X., Han, J., 2010. Encyclopedia of Machine Learning. Springer US, Boston, MA.

Jolliffe, I., 2011. Principal Component Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 1094–1096. URL: https://doi.org/10.1007/978-3-642-04898-2_455, doi:10.1007/978-3-642-04898-2_455.

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley.

References V

Livera, A.M.D., Hyndman, R.J., Snyder, R., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. J. Am. Stat. Assoc. 106, 1513–1527.

Lloyd, S.P., 1982. Least squares quantization in pcm. Information Theory, IEEE Transactions 28, 129–137.

Lubba, C.H., Sethi, S.S., Knaute, P., Schultz, S.R., Fulcher, B.D., Jones, N.S., 2019. catch22: Canonical time-series characteristics, in: Data Min Knowl Disc, p. 1821–1852. doi:https://doi.org/10.1007/s10618-019-00647-x.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R., Talagala, T., 2020. Fforma: Feature-based forecast model averaging. International Journal of Forecasting 36, 86–92.

References VI

Pelleg, D., Moore, A., 2000. X-means: Extending k-means with efficient estimation of the number of clusters, in: Proceedings of the Seventeenth International Conference on Machine Learning, p. 727–734.

Sammut, C., Webb, G.I. (Eds.), 2010. Mean Absolute Error. Springer US, Boston, MA. pp. 652–652. URL: https://doi.org/10.1007/978-0-387-30164-8_525, doi:10.1007/978-0-387-30164-8_525.

Smyl, S., 2020. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. International Journal of Forecasting 36, 75–85. doi:https://doi.org/10.1016/j.ijforecast.2019.03.017. m4 Competition.

Suilin, A., 2017. kaggle-web-traffic. https://github.com/Arturus/kaggle-web-traffic/blob/master/how_it_works.md. Accessed: 2019-11-29.

References VII

Tibshirani, R., 1994. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.

Trapero, J.R., Kourentzes, N., Fildes, R., 2015. On the identification of sales forecasting models in the presence of promotions. Journal of the Operational Research Society 66, 299 – 307. doi:https://doi.org/10.1057/jors.2013.174.

Wolpert, D.H., 1992. Stacked generalization. Neural networks 5, 241–259.