Dynamic and Heterogeneous Ensembles for Time Series Forecasting

Vítor Cerqueira[1], Luís Torgo[1], Mariana Oliveira[1], and Bernhard Pfahringer[2]

INESCTEC / University of Porto, Portugal[1]

University of Auckland, New Zealand[2]

Objectives

The main goal of this work is to uncover new techniques for adaptively combining diverse base learners of an ensemble model for univariate time series forecasting.

Motivation and assumptions

- Different learning models have different areas of expertise across the input space ⇒ **ensemble heterogeneity**;
- Recurring structures are common and forecasting models have varying relative performance over time ⇒ **dynamic combination**.

Forecasting t+1

Each base model M_j, $\forall j \in \{1, ..., m\}$ is built to approximate a function $Y_{t+1} = f(Y_t, \cdots, Y_{t-K})$ using the available time series Y.

Weighting predictions according to recent performance:

- The MSE of each forecaster is tracked over the last Ω observations;
- The forecaster's weight is computed by applying the complementary Gaussian error function, **erfc**, to the MSE.

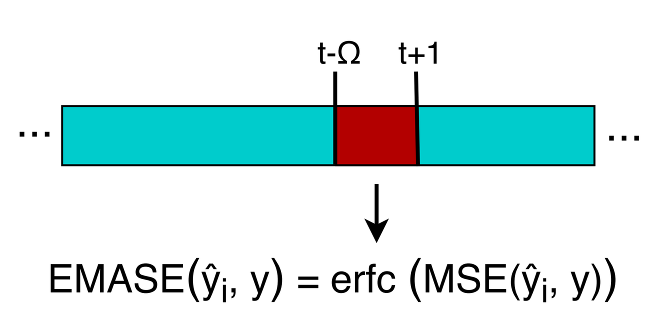

Figure 1: Computing EMASE – Erfc Moving Average Squared Error

Suspending models with poor recent performance:

- The ensemble has a dynamic composition – at each point, the top α% of models in the last Ω observations are weighted.

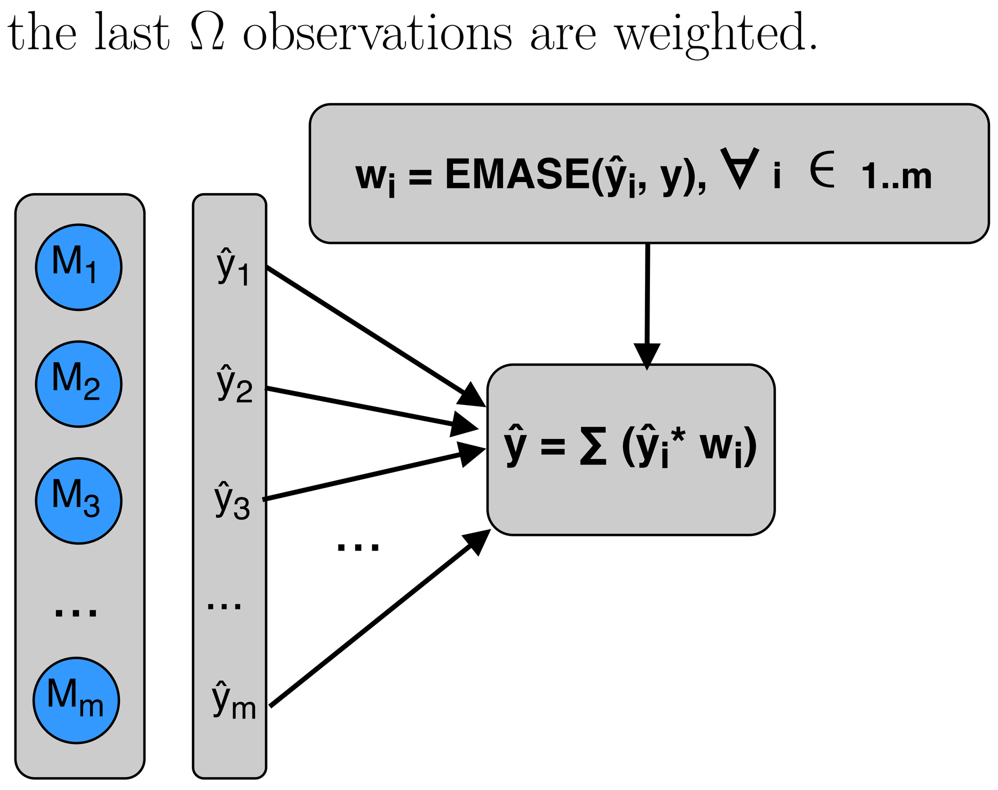

Figure 2: The prediction of each forecaster, $\hat{y}_i$ is weighted according to EMASE.

Acknowledgements

This work is financed by the ERDF âĂŞ European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT âĂŞ Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project Fotocatgraf - UTAP-ICDT/CTM-NAN/0025/2014. This work is also financed by the ERDF âĂŞ European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project ÂńPOCI-01-0145-FEDER-006961Âż, and by National Funds through the FCT âĂŞ Fundação para a Ciência e a Tecnologia as part of project UID/EEA/50014/2013.

Encouraging data diversity

$$Y_{[N,K]} = \begin{bmatrix} y_1 & y_2 & \cdots & y_{K-1} & y_K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{i-K+1} & y_{i-K+2} & \cdots & y_{i-1} & y_i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{N-K+1} & y_{N-K+2} & \cdots & y_{N-1} & y_N \end{bmatrix}$$

Subsetting embedding size (columns) and training window (rows)

- embedding size diversity: **K**, **K/2**, **K/4**
- training window diversity: **N**, **N/2**, **N/4**
  - Leading to 9 different data combinations available for training

Empirical Results

- Base models include SVR; Multi-layer perceptron; Gaussian Processes; Linear regression; Random Forest; Generalized Boosted regression; PPR; MARS, and Rule-based regression
- 10% of the models with top performance in the last 50 observations are weighted for predicting t+1;
- Testing on 16 time series from several domains

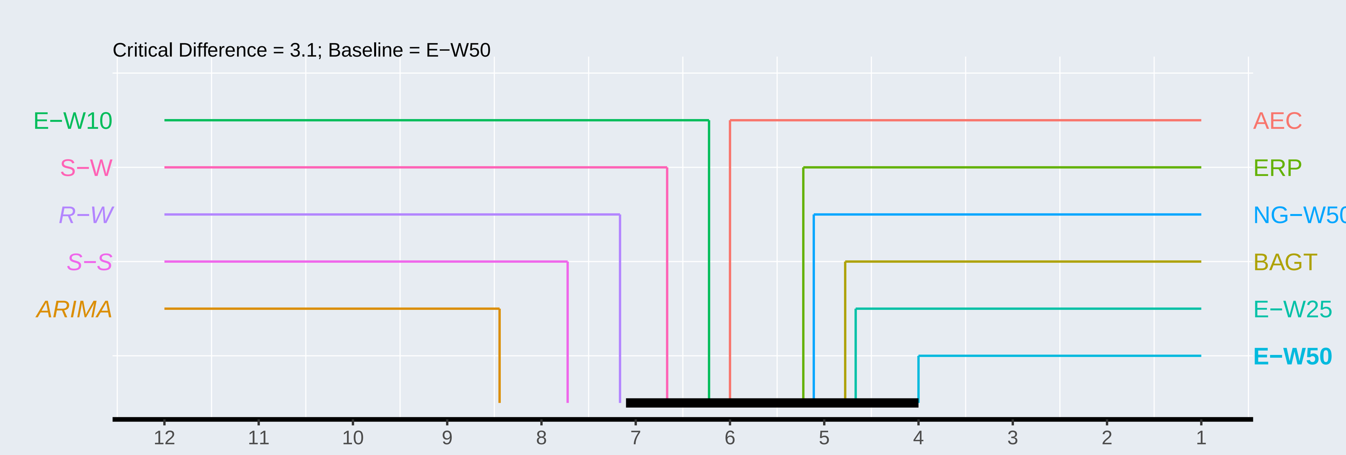

Figure 3: Comparing the proposed method **E-W50** to baselines & state of the art with a post-hoc Bonferroni-Dunn test. **NG-W50** is the state-of-art windowing combination approaches; **BAGT** is a bagging of trees specially designed for time series forecasting; **S-W** weights forecasters according to error in the training set; **S-S** combines models using a simple average; and **ERP**, **R-W**, and **AEC** are other dynamic combination approaches in the literature.

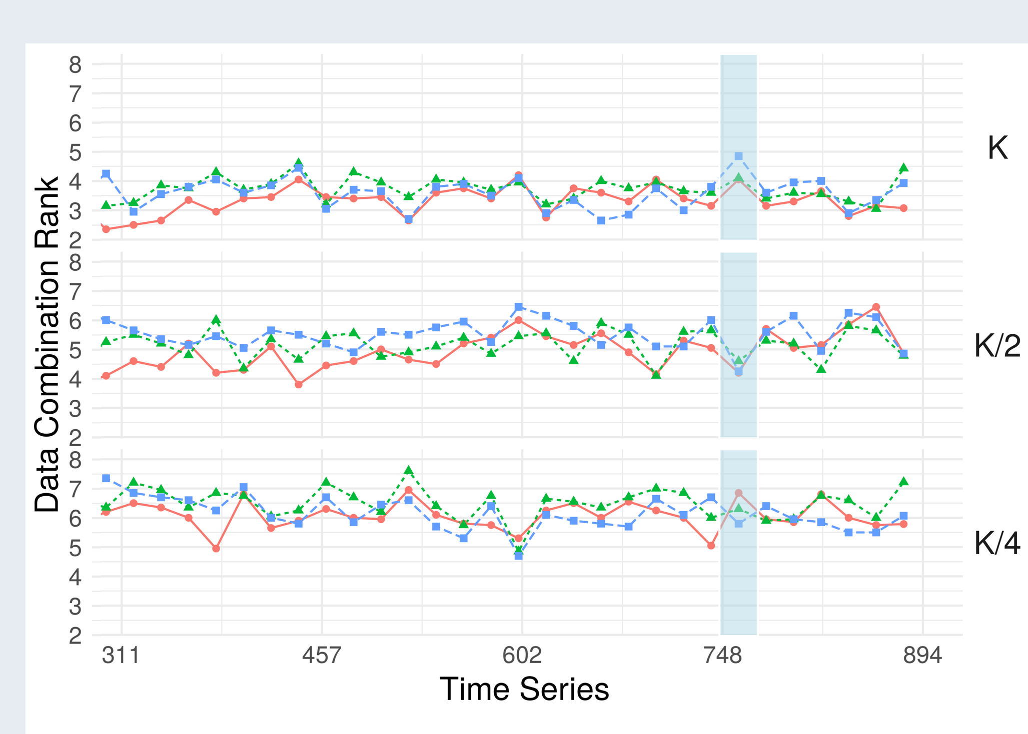

Figure 4: Squared error rank of each data combination strategy. There is systematic evidence that subsets of the original time series show better performance in some time series intervals.

Reproducibility

R package: **tsensembler**