On performance of temporal aggregation in time series forecasting

Bahman Rostami-Tabar (@Bahman_R_T)

Website www.bahmanrt.com
Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?

- How does temporal aggregation approaches perform on M4 competition data?

- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (Working paper 1)?

- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (Working paper 2)?
Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?

- How does temporal aggregation approaches perform on M4 competition data?

- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (Working paper 1)?

- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (Working paper 2)?
Using time series forecasting to inform decisions

Data and forecast time granularity

• Forecasting time granularity level and its horizon are determined by decisions made in the light of forecast.

• One common assumption is that time series granularity matches forecast requirement, i.e. to produce daily forecasts, we use daily time series.

• However, the level of time series granularity does not necessarily match the level of forecast granularity.

• The level of temporal granularity in the forecast might be lower than the existing time series granularity. For instance, while a forecast might be required at the annual level, a monthly time series is available. With advances in IT, data is often recorded at the finest temporal granularity (e.g. arrival time)
Time series forecasting problem

- We consider a time series forecasting problem where an original time series has a higher temporal granularity (e.g. monthly) than the required forecast (e.g. annual).

- We aim to generate a forecast of the total value over a number of time periods ahead, forecast horizon aggregation or forecast over the leadtime period.

A key question then to be answered is:

should the original series be used to generated the forecast for the required horizon and then sum them up to obtain the forecast horizon aggregation (lead-time), i.e. Aggregate Forecast (AF) or should we first aggregate time series to match the forecast requirement granularity and then extrapolate directly at that level, i.e. Aggregate Data (AD).

I will illustrate these approaches using a simple example.

There is no disaggregation to the original time granularity
Terminology

One time series

- Data time granularity (e.g. daily, monthly, annual)
- Forecast time granularity (e.g. daily, monthly, annual)
- Forecast horizon (e.g. 12 months ahead)
- Forecast horizon aggregation / leadtime (e.g. 1 week, 1 quarter, 1 year)
- Temporal aggregation
  - Aggregate Forecast (or Bottom-Up)
  - Aggregate Data
    - Non-overlapping temporal aggregation (NOA)
    - Overlapping temporal aggregation (OA)
Forecast horizon aggregation: an example

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>21</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>
Temporal aggregation: aggregate forecast

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>21</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Forecast 3 periods-ahead, then add them up

Forecast: 3, 3, 3
Temporal aggregation: aggregate forecast

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>21</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>

Forecast 3 periods-ahead, then add them up

Aggregate Forecast (AF) or Bottom-up (BU)
Non-overlapping temporal aggregation: aggregate data

Original time granularity

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>21</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Non-overlapping temporally aggregated series

Aggregated time granularity

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q1</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>17</td>
<td>33</td>
<td>10</td>
<td>?</td>
</tr>
</tbody>
</table>

Forecast 3 periods-ahead, then add them up

Aggregate Forecast (AF)
Overlapping temporal aggregation
Using information at multiple levels of time granularity instead of a single level - MAPA

Using information at multiple levels of time granularity instead of a single level- temporal hierarchies

It is often recommended to aggregate data and then forecast when a time series history is recorded at a higher frequency time granularity (e.g. monthly) and forecast is required at a lower level (e.g. annual).

For an example, please refer to page 153 of Profit from Your Forecasting Software, by Paul Goodwin.

Let's examine the performance of aggregating data versus aggregating forecast approaches using M4 competition dataset.
Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?

- How does temporal aggregation approaches perform on M4 competition data?

- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (Working paper 1)?

- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF? (Working paper 2)
Time series data

- M4 competition data time series
  - 24,000 Quarterly
  - 48,000 monthly
  - 4,227 daily

- Time series features
  - 42 features
  - Extract features using `tsfeatures::tsfeatures()` in R

- Forecasting methods: Exponential Smoothing State Space (ETS) (ARIMA is also considered).
- Point forecast accuracy measure: Mean Absolute Scaled Error (MASE), Root Mean Squared Scaled Error (RMSSE), and more.
- Time series cross validation is performed.

M4 Monthly time series features
M4 Monthly time series features
Percentage of series for which each approach was more accurate (using MASE)
Performance of AF vs. AD (based on non-overlapping temporal aggregation)
Questions

Given the comparative performance of temporal aggregation approaches:

- Whether combining forecasts generated by Bottom-Up (BU), Non-overlapping (NOA) and Overlapping approaches (OA) improves the forecast accuracy? how to combine?

- How data temporal aggregation changes time series features and is there any association between time series features and the forecasting performance of AD versus AF?
Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?

- How does temporal aggregation approaches perform on M4 competition data?

- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (Working paper 1)?

- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (Working paper 2)?
Experiment design - 1

1. For a given time series
   - Temporally aggregate series by a non-overlapping approach, for a given aggregation level M.
   - Temporally aggregate series by an overlapping approach, for a given aggregation level M.
   - Use forecasting method to generate a one-step-ahead forecast.
   - Forecast generated for aggregated time series (non-overlapping Aggregation: NOA) using Long Short-Term Memory (LSTM).
   - Forecast generated for aggregated time series (overlapping Aggregation: OA) using Overlapping Aggregation (OA).
   - Combine forecasts using simple average.
   - Combine forecasts using Multilayer Perceptron (MLP).

2. Forecasting Performance Evaluation.
Combining algorithm

Algorithm 1 MLP combining rule

Initialization:
- set the vector of learning rates of individual approaches \((\eta_0^{BU}, \eta_0^{NOA}, \eta_0^{OA})\)
- set the vector of regrets of individual approaches \((R_0^{BU}, R_0^{NOA}, R_0^{OA}) = (0, 0, 0)\)

repeat

At each time (aggregate horizon) in the out-of-sample

1. compute the learning rates \(\eta_{t-1}^k\) according to Equation (1)
2. calculate the combining weights of each individual method by
\[ p_t^k = \frac{\eta_{t-1}^k \max(0, R_{t-1}^k)}{\sum_{k=1}^{K} \eta_{t-1}^k \max(0, R_{t-1}^k)} \]
3. obtain the loss vector \(\ell_t = (\ell_t^{BU}, \ell_t^{NOA}, \ell_t^{OA})\) and the weighted loss \(\hat{\ell}_t = p_t^{BU} \ell_t^{BU} + p_t^{NOA} \ell_t^{NOA} + p_t^{OA} \ell_t^{OA}\)
4. update the regret \(R_t^k = R_{t-1}^k + (\hat{\ell}_t - \ell_t^k)\)

until End of the out-of-sample;

### Mean (median) MASE for M4 monthly series with ETS forecasting method

<table>
<thead>
<tr>
<th>Aggregation level</th>
<th>Pattern</th>
<th>Approach</th>
<th>MLP</th>
<th>Average</th>
<th>Overlapping</th>
<th>Non-overlapping</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N, N)</td>
<td></td>
<td>6.675 (4.456)</td>
<td>7.958 (5.38)</td>
<td>8.491 (5.723)</td>
<td>9.822 (6.939)</td>
<td>7.673 (5.590)</td>
</tr>
<tr>
<td></td>
<td>(T, N)</td>
<td></td>
<td>6.483 (4.299)</td>
<td>8.785 (5.834)</td>
<td>11.161 (7.643)</td>
<td>11.472 (7.66)</td>
<td>7.603 (5.495)</td>
</tr>
<tr>
<td></td>
<td>(T, S)</td>
<td></td>
<td>6.215 (4.394)</td>
<td>8.064 (5.957)</td>
<td>10.23 (7.96)</td>
<td>9.417 (6.820)</td>
<td>7.396 (5.671)</td>
</tr>
<tr>
<td>Semi-annual</td>
<td>(N, N)</td>
<td></td>
<td>3.023 (2.500)</td>
<td>3.236 (2.635)</td>
<td>3.398 (2.800)</td>
<td>3.642 (2.871)</td>
<td>3.170 (2.637)</td>
</tr>
<tr>
<td></td>
<td>(T, N)</td>
<td></td>
<td>2.440 (1.809)</td>
<td>2.893 (2.153)</td>
<td>3.502 (2.640)</td>
<td>3.511 (2.607)</td>
<td>2.714 (2.045)</td>
</tr>
<tr>
<td></td>
<td>(N, S)</td>
<td></td>
<td>3.421 (2.803)</td>
<td>3.875 (3.167)</td>
<td>5.161 (3.992)</td>
<td>4.025 (3.315)</td>
<td>3.82 (3.156)</td>
</tr>
<tr>
<td></td>
<td>(T, S)</td>
<td></td>
<td>2.835 (2.290)</td>
<td>3.366 (2.695)</td>
<td>5.105 (3.833)</td>
<td>3.442 (2.768)</td>
<td>3.156 (2.548)</td>
</tr>
<tr>
<td>4-monthly</td>
<td>(N, N)</td>
<td></td>
<td>1.886 (1.620)</td>
<td>1.951 (1.668)</td>
<td>2.041 (1.751)</td>
<td>2.155 (1.819)</td>
<td>1.916 (1.635)</td>
</tr>
<tr>
<td></td>
<td>(T, N)</td>
<td></td>
<td>1.430 (1.063)</td>
<td>1.574 (1.181)</td>
<td>1.845 (1.421)</td>
<td>1.837 (1.420)</td>
<td>1.499 (1.115)</td>
</tr>
<tr>
<td></td>
<td>(N, S)</td>
<td></td>
<td>2.288 (1.937)</td>
<td>2.538 (2.115)</td>
<td>3.501 (2.752)</td>
<td>2.551 (2.146)</td>
<td>2.457 (2.056)</td>
</tr>
<tr>
<td></td>
<td>(T, S)</td>
<td></td>
<td>1.824 (1.497)</td>
<td>2.121 (1.693)</td>
<td>3.330 (2.354)</td>
<td>2.057 (1.601)</td>
<td>1.948 (1.578)</td>
</tr>
<tr>
<td>Quarterly</td>
<td>(N, N)</td>
<td></td>
<td>1.358 (1.179)</td>
<td>1.382 (1.198)</td>
<td>1.446 (1.259)</td>
<td>1.496 (1.292)</td>
<td>1.366 (1.167)</td>
</tr>
<tr>
<td></td>
<td>(T, N)</td>
<td></td>
<td>0.998 (0.731)</td>
<td>1.054 (0.772)</td>
<td>1.198 (0.909)</td>
<td>1.195 (0.917)</td>
<td>1.012 (0.733)</td>
</tr>
<tr>
<td></td>
<td>(N, S)</td>
<td></td>
<td>1.722 (1.471)</td>
<td>1.862 (1.567)</td>
<td>2.515 (1.986)</td>
<td>1.854 (1.553)</td>
<td>1.807 (1.529)</td>
</tr>
<tr>
<td></td>
<td>(T, S)</td>
<td></td>
<td>1.338 (1.093)</td>
<td>1.524 (1.198)</td>
<td>2.372 (1.644)</td>
<td>1.448 (1.181)</td>
<td>1.394 (1.121)</td>
</tr>
<tr>
<td>Bi-monthly</td>
<td>(N, N)</td>
<td></td>
<td>0.881 (0.758)</td>
<td>0.886 (0.764)</td>
<td>0.931 (0.799)</td>
<td>0.926 (0.799)</td>
<td>0.877 (0.745)</td>
</tr>
<tr>
<td></td>
<td>(T, N)</td>
<td></td>
<td>0.615 (0.423)</td>
<td>0.627 (0.429)</td>
<td>0.684 (0.470)</td>
<td>0.679 (0.493)</td>
<td>0.608 (0.412)</td>
</tr>
<tr>
<td></td>
<td>(N, S)</td>
<td></td>
<td>1.191 (1.030)</td>
<td>1.218 (1.050)</td>
<td>1.504 (1.241)</td>
<td>1.223 (1.043)</td>
<td>1.208 (1.031)</td>
</tr>
<tr>
<td></td>
<td>(T, S)</td>
<td></td>
<td>0.895 (0.717)</td>
<td>0.951 (0.745)</td>
<td>1.329 (0.951)</td>
<td>0.919 (0.737)</td>
<td>0.900 (0.713)</td>
</tr>
</tbody>
</table>
Outline

• Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?

• How does temporal aggregation approaches perform on M4 competition data?

• Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (Working paper 1)?

• How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (Working paper 2)?
Experiment design - 2

1. Temporally aggregated series
   - Non-overlapping Temporal aggregation (e.g. quarterly \( m = 4 \), annual \( m = 12 \))
   - Extract time series features

2. Original series (e.g. monthly)

3. Forecast aggregated series (AD approach)
   - Forecasting method: ETS
   - Aggregate forecasts (AF approach)

4. Forecast original series

5. Calculate forecast accuracy

6. Forecast accuracy for AF and AD

7. Time series features

8. Machine learning
How does non-overlapping TA change time series features?
How does non-overlapping TA change time series features (continue)?
Features relationship and AD/AF performance
MCB test for all classifiers

We also use missclassification error, F-statistics and Area under the Curve (AUC).
Important features
Partial dependence plot
Probability of AF performing better
Partial dependence plot (continue)
Probability of AF performing better
Summary and conclusions (continue)

• Although aggregating time series seems to be intuitive, it might not always improve forecast accuracy. Our results indicate that Aggregate Forecast is a competitive approach, but neither of them dominate. They both have a merit.

• Combining aggregate data (non-overlapping and overlapping) and aggregate forecast approaches improve forecast accuracy. Combination again works here.

• Aggregate data using temporal aggregation changes the features of time series. The magnitude of the change varies for different features. In particular, we observe that with increase in the aggregation level, the strength of seasonality, the autocorrelation, coefficient of variation, linearity, curvature and KPSS unitroot statistic decrease. However, non-linearity, mean, variance, ARCH.LM, trend, unitroot pp statistics increase. Entropy is the only measure that both increases and decreases based on its initial value.
Summary and conclusions

- Random Forest model is the most accurate classifier among ML algorithm in predicting which approach provides more accurate forecast given a set of time series features as input.

- The most important features for predicting whether AF or AD should be used for a given monthly time series in M4 competition include curvature, nonlinearity, seas_pacf, unitroot_up, mean, ARCH.MLM, Coefficient of Variation, stability, linearity and max_level_shift.

- Increasing trend, ARCH.LM, hurst, autocorrelation lag 1 and unitroot_pp and seas_pacf may increases the chance of AF performing better.

- Increasing lumpiness, entropy, no-linearity, curvature, strength of seasonality may increase the chance of AD performing better, so the strong presence of these features may favorite AD over AF.
Wrok in progress

- Rostami-Tabar B., Goltsos T. Wang S. (2022), Forecasting for lead-time period by temporal aggregation: Whether to combine and how

- Rostami-Tabar B., Mercetic D. (2022), On time series features and the performance of temporal aggregation

Published recently


References for temporal aggregation forecasting

- Demand forecasting by temporal aggregation, Naval Research Logistics
- Forecasting with temporal hierarchies, European Journal of Operational Research
• Slides and papers: www.bahmanrt.com
• Check out also www.f4sg.org

Say hello: @Bahman_R_T

Connect: Bahman ROSTAMI-TABAR