Déjà vu: forecasting with similarity

Yanfei Kang, Evangelos Spiliotis, Fotios Petropoulos, Nikolaos Athiniotis, Feng Li, Vassilios Assimakopoulos

https://arxiv.org/abs/1909.00221

Sep 12, 2019

Outline

- **Introduction**
- **Methodology**
- **Application**

Introduction

Introduction

- Time series models: impossible to perfectly capture the actual DGP of the data because of the over-simplified assumptions.
- Model selection:
  - aggregate selection & individual selection
  - different models being selected when different criteria or cost functions are used (Billah et al., 2006)
- Model averaging:
  - determination of the pool of methods
  - identification of weights
  - estimation of methods
- Three sources of uncertainty exist in forecasting: model, parameter, and data (Petropoulos et al., 2018).

Introduction

- **This paper** argue that there is another way to avoid selecting a single model: to select no model at all.
- Forecasting with similarity.
  - No model extrapolation takes place
  - A data-centric approach
  - cross-learning
  - no explicit assumptions (DGP and distribution of residuals)

Methodology

Methodology

1. Remove seasonality
2. Smoothing
3. Scaling
4. Measuring similarity
5. Forecasting
6. Inverse scaling
7. Add seasonality

Methodology

1. Remove seasonality
   - Box-Cox transformation

     (additive & multiplicative seasonality)
   - seasonality test (ACF)
   - STL (Seasonal and Trend decomposition using Loess)
2. Smoothing
   - Loess method
   - decomposes the series into the trend and remainder components.
3. Scaling (divided by forecast origin)

6. Inverse scaling
7. Add seasonality

Methodology

4. Measuring similarity
   - $\mathcal{L}_1$ norm
   - $\mathcal{L}_2$ norm
   - DTW

$$d_{\mathcal{L}_1}(\tilde{y}, \tilde{Q}(i)_{1,\dots,n}) = \left\| \tilde{y}_t - \tilde{Q}(i)_t \right\|_1 ,$$

$$d_{\mathcal{L}_2}(\tilde{y}, \tilde{Q}(i)_{1,\dots,n}) = \left\| \tilde{y}_t - \tilde{Q}(i)_t \right\|_2 ,$$

$$d_{\mathrm{DTW}}(\tilde{y}, \tilde{Q}(i)_{1,\dots,n}) = D(n, n),$$

$$D(v, w) = |\tilde{y}_v - \tilde{Q}(i)_w| + \min \left\{ \begin{array}{c} D(\tilde{y}_v, \tilde{Q}(i)_{w-1}) \\ D(\tilde{y}_{v-1}, \tilde{Q}(i)_{w-1}) \\ D(\tilde{y}_{v-1}, \tilde{Q}(i)_w) \end{array} \right\}. \quad D(1,1) = |\tilde{y}_1 - \tilde{Q}(i)_1|.$$

DTW allows various matches among the points of the series being compared

5. Forecasting (Figure 1)
   - selecting the k most similar series, and then the median is calculated for each planning horizon (statistical aggregation).

Application

Application

- Reference dataset: yearly, quarterly, and monthly subsets of the M4 competition.

- Test dataset: yearly, quarterly, and monthly series of the M3 forecasting competition.

- Performance evaluation:

$$\mathrm{MASE} = \frac{1}{h} \frac{\sum_{t=n+1}^{n+h} |y_t - f_t|}{\frac{1}{n-s} \sum_{i=s+1}^{n} |y_t - y_{t-s}|}.$$

Application

Distance measure
($\mathcal{L}_1$, $\mathcal{L}_2$, DTW)

Number of aggregates
(k)

If prepocessing
(YES, NO)

1

Optimal settings

Sizes of the historical sample

2

Forecasting performance

Application

1 Optimal settings

Frequency	Distance Measure	Number of aggregated reference series (k)							M4 competition
		1	5	10	50	100	500	1000	
Yearly	$\mathcal{L}_1$	3.289	2.837	2.787	2.689	2.668	2.632	2.634	
	$\mathcal{L}_2$	3.333	2.866	2.785	2.703	2.684	2.638	2.639	2.980 (-11.98%)
	DTW	3.270	2.835	2.730	2.656	2.641	**2.623**	2.637	
Quarterly	$\mathcal{L}_1$	1.312	1.205	1.175	1.136	1.135	1.127	1.126	
	$\mathcal{L}_2$	1.336	1.199	1.162	1.138	1.134	1.126	1.127	1.111 (+0.36%)
	DTW	1.293	1.177	1.158	1.117	1.115	**1.115**	1.116	
Monthly	$\mathcal{L}_1$	1.004	0.908	0.887	0.871	0.870	0.867	0.869	
	$\mathcal{L}_2$	1.008	0.910	0.891	0.871	0.869	0.866	0.868	0.884 (-3.05%)
	DTW	1.001	0.895	0.875	0.861	0.861	**0.857**	0.857	
Total	$\mathcal{L}_1$	1.607	1.427	1.397	1.356	1.351	1.339	1.340	
	$\mathcal{L}_2$	1.626	1.433	1.395	1.360	1.354	1.339	1.341	
	DTW	1.597	1.413	1.373	1.339	1.335	**1.329**	1.332	

Application

1 Optimal settings

the Multiple Comparisons from the Best (MCB) test

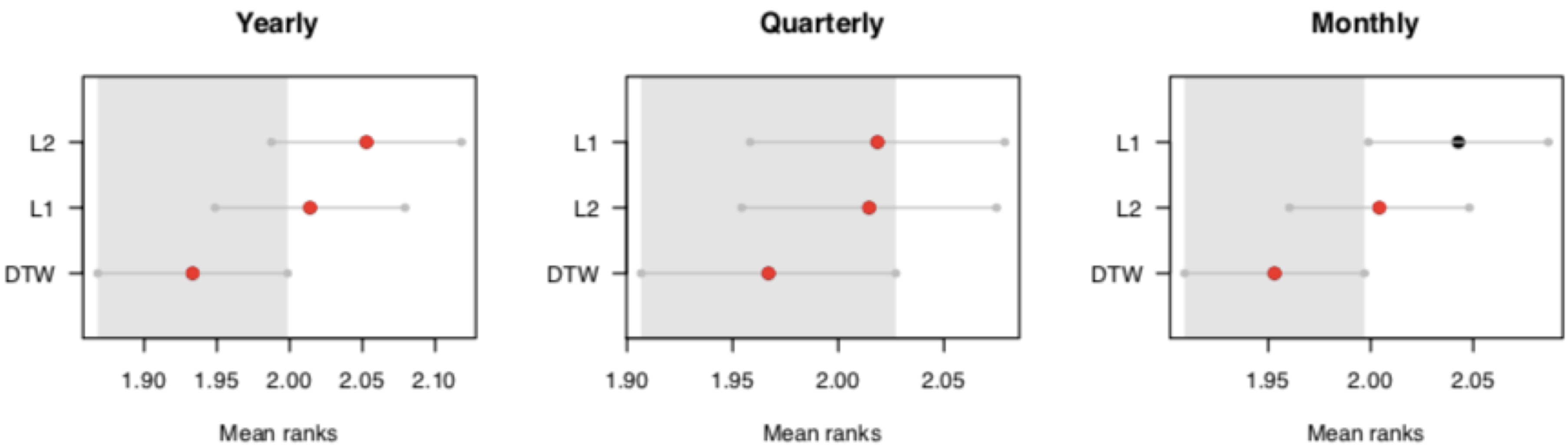

Figure 2: MCB significance tests for the three distance measures for each data frequency.

Application

1 Optimal settings

Table 4: The performance of forecasting with similarity, with and without preprocessing. The DTW distance measure is considered.

Frequency	Preprocessing	Number of aggregated reference series (k)						
		1	5	10	50	100	500	1000
Yearly	NO	3.544	2.821	2.735	2.644	2.639	2.626	2.641
	YES	3.270	2.835	2.730	2.656	2.641	**2.623**	2.637
Quarterly	NO	1.657	1.411	1.359	1.384	1.396	1.419	1.422
	YES	1.293	1.177	1.158	1.117	1.115	**1.115**	1.116
Monthly	NO	1.263	1.077	1.020	1.011	1.012	1.040	1.060
	YES	1.001	0.895	0.875	0.861	0.861	**0.857**	0.857
Total	NO	1.888	1.564	1.502	1.483	1.486	1.503	1.517
	YES	1.597	1.413	1.373	1.339	1.335	**1.329**	1.332

DTW, k=500, Preprocessing

Application

2 Historical sample sizes

- Benchmark methods:
  - optimally selected ETS by AICc (ETS)
  - simple combination of Simple Exponential Smoothing, Holt's linear trend Exponential Smoothing, and Damped trend Exponential Smoothing (SHD)
- Similarity (proposed method)
- ETS-Similarity
  - the simple forecast combination of ETS and Similarity

Conclusions:

There are benefits for both model-based and data-centric approaches for forecasting. (Figure 3, 4)