Forecast with forecasts: diversity matters

— Feng Li, School of Statistics and Mathematics, Central University of Finance and Economics

Outline

1. Forecasting large collections of time series
2. Using forecast diversity to forecast
3. Diversity for intermittent demand

Modern time series

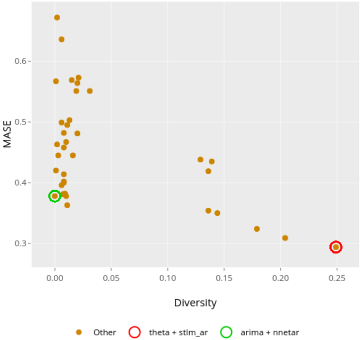
- Retail: daily sales for thousands of categories spanning over two years.
- Electricity: hourly electricity load data to make ultra forecast (6 months).
- Smart meter: noisy household data (in 15 minutes) to forecast next day's gas demand.
- Web traffic: Wikipedia hourly page views.

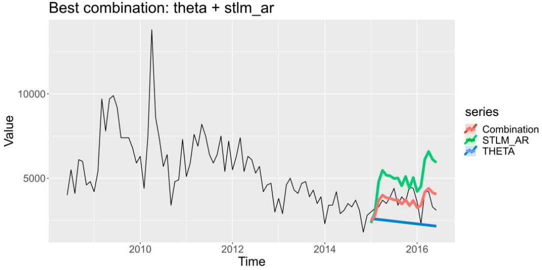
There is no free lunch in forecasting algorithm selections

- *"If we know that learning algorithm A is superior to B averaged over some set of targets F , then the No Free Lunch theorems tell us that B must be superior to A if one averages over all targets not in F . This is true even if algorithm B is the algorithm of purely random guessing."* — Wolpert & Macready (1997)
- There is no single model that will always do better than any other model.
- Using measurable features of the problem instances to predict which algorithm is likely to perform best. (Smith-miles 2009)
- **Forecasting algorithm selections**: inspect a large collection of time series, understand their characteristics, and select an appropriate method with minimal computational costs for each series.

Using forecast diversity to forecast

- Input: the forecasts ($\hat{y}_{t+h}$) from a pool of models.
- Measure their diversity — a feature that has been identified as a decisive factor in forecast combination.
- Through meta-learning, we link the diversity of the forecasts with their out-of-sample performance to fit combination models based on diversity (Kang et al. 2022).


MASE
0.6
0.5
0.4
0.3
0.00
0.05
0.10
0.15
0.20
0.25
Diversity
Other
theta + stlm_ar
arima + nnetar

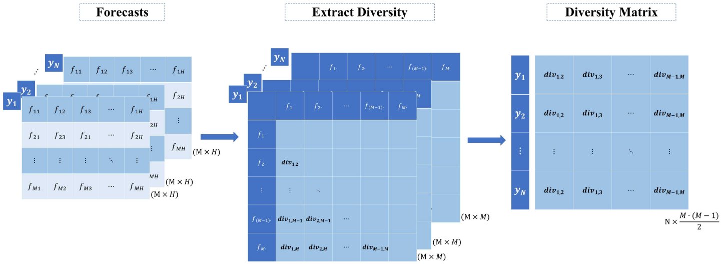

Best combination: theta + stlm_ar
Value
10000
5000
2010
2012
2014
2016
Time
series
Combination
STLM_AR
THETA

Using forecast diversity to forecast
↬ Why it works?

$$
\begin{aligned}
MSE_{comb} &= \frac{1}{H}\sum_{i=1}^{H}\left(\sum_{i=1}^{M} w_i f_{ih} - y_{T+h}\right)^2 \\
&= \frac{1}{H}\sum_{i=1}^{H}\left[\sum_{i=1}^{M} w_i (f_{ih} - y_{T+h})^2 - \sum_{i=1}^{M} w_i (f_{ih} - f_{ch})^2\right] \\
&= \frac{1}{H}\sum_{i=1}^{H}\left[\sum_{i=1}^{M} w_i (f_{ih} - y_{T+h})^2 - \sum_{i=1}^{M-1}\sum_{j=1, j>i}^{M} w_i w_j (f_{ih} - f_{jh})^2\right] \\
&= \sum_{i=1}^{M} w_i MSE_i - \sum_{i=1}^{M-1}\sum_{j=1, j>i}^{M} w_i w_j Div_{i,j},
\end{aligned}
$$

where H is the forecasting horizon, M is the number of forecasting methods, and T is the historical length.

Diversity for forecast combination



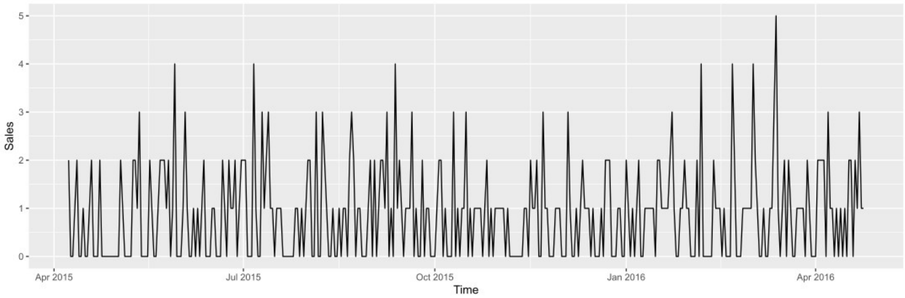
- Two diversity measures are used. $Div_{i,j} = \frac{1}{H}\sum_{i=1}^{H}(f_{ih} - f_{jh})^2$ and $sDiv_{i,j} = \frac{\sum_{h=1}^{H}(f_{ih}-f_{jh})^2}{\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\left[\sum_{h=1}^{H}(f_{ih}-f_{jh})^2\right]}$.
- The following optimization problem is then solved to obtain the combination weights

$$\text{argmin}_w \sum_{n=1}^{N}\sum_{i=1}^{M} w(Div_n)_i \times \text{Err}_{ni}$$

Forecasting 100,000 series with various frequencies

Method	Overall	Yearly	Quarterly	Monthly	Weekly	Daily	Hourly
			MASE				
SA	1.9040	3.6907	1.2432	0.9813	6.3826	5.8921	3.3319
FFORMA	1.5586	3.0842	1.1220	0.8980	**2.2309**	3.2464	0.8822
Diversity	**1.5478**	**3.0670**	**1.1095**	**0.8915**	2.2744	**3.2296**	**0.8540**
FD	**1.5507**	**3.0615**	**1.1096**	0.8997	2.2639	**3.2345**	**0.8574**
			MSIS				
SA	17.5077	42.0776	9.9248	8.3012	22.4778	31.5910	11.4214
FFORMA	14.5934	32.0185	9.2388	7.8189	**16.0496**	**27.7694**	6.6161
Diversity	**14.0197**	**30.3312**	**8.7805**	**7.6385**	16.4015	28.0220	**6.3587**
FD	**14.0254**	**30.3980**	**8.7995**	**7.6248**	16.0936	27.8723	**6.3145**

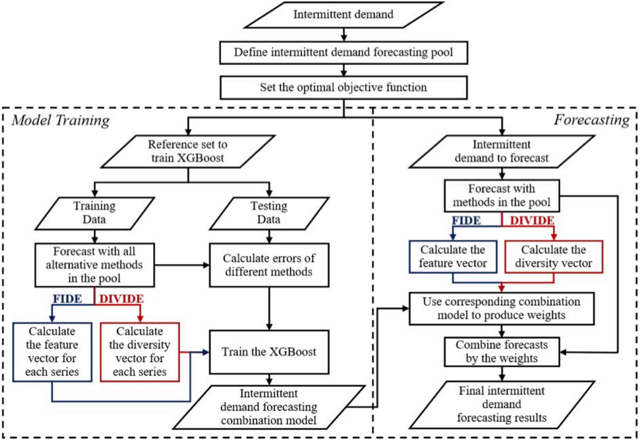
Diversity for intermittent demand



- Several periods of zero demand.
- Ubiquitous in practice - retailing, aerospace, etc.
- Two sources of uncertainty: sporadic demand occurrence and demand arrival timing.

Forecast combination for intermittent demand

- **FIDE**: Feature-based Intermittent DEmand forecasting
- **DIVIDE**: DIVersity-based Intermittent DEmand forecasting (Li et al. 2022)



Application to Royal Air Force (RAF) data

- 5000 monthly time series with high intermittence.

Method	sMAE	sMSE	MASE	sMAPIS	Average rank
Naive	1.4569	66.7696	0.8491	108.5635	10.25
sNaive	1.6546	69.3216	0.9706	94.2955	13.00
SES	1.6615	59.9634	0.9657	95.7405	13.00
MA	1.6780	59.6904	0.9865	99.2663	14.25
ARIMA	1.6650	57.5444	0.9474	77.7958	8.75
ETS	1.6693	57.8585	0.9493	79.5946	11.50
CRO	1.7504	57.2558	0.9971	79.5944	11.75
optCro	1.7483	57.2632	0.9957	79.4407	11.25
SBA	1.7394	**57.2542**	0.9907	79.0423	9.75
TSB	1.6923	57.3153	0.9655	78.0195	9.50
ADIDA	1.5832	57.6764	0.9094	77.7058	6.50
IMAPA	1.6266	57.2964	0.9238	74.8702	4.75
SA	1.6551	57.7310	0.9509	79.1409	10.00
Median	1.6283	57.3461	0.9251	75.6345	6.50
FFORMA	1.1406	57.4036	0.6396	80.8106	7.00
FIDE	0.9342	57.3391	0.5285	73.2410	3.25
DIVIDE	**0.9323**	57.3097	**0.5281**	**69.4176**	**2.00**

Thank you!

https://feng.li
feng.li@cufe.edu.cn
@f3ngli

References

Kang, Y., Cao, W., Petropoulos, F. & Li, F. (2022), 'Forecast with forecasts: Diversity matters', *European Journal of Operational Research* **31**(1), 180–190.

Li, L., Kang, Y., Petropoulos, F. & Li, F. (2022), 'Feature-based intermittent demand forecast combinations: bias, accuracy and inventory implications', *arXiv preprint arXiv:2204.08283* .

Smith-miles, K. (2009), 'Cross-disciplinary perspectives on meta-learning for algorithm selection', *ACM Computing Surveys* **41**(1), 6.

Wolpert, D. H. & Macready, W. G. (1997), 'No free lunch theorems for optimization', *IEEE Transactions on Evolutionary Computation* **1**(1), 67–82.