Hierarchical Forecasting and Optimal Reconciliation with Immutable Forecasts

Feng Li

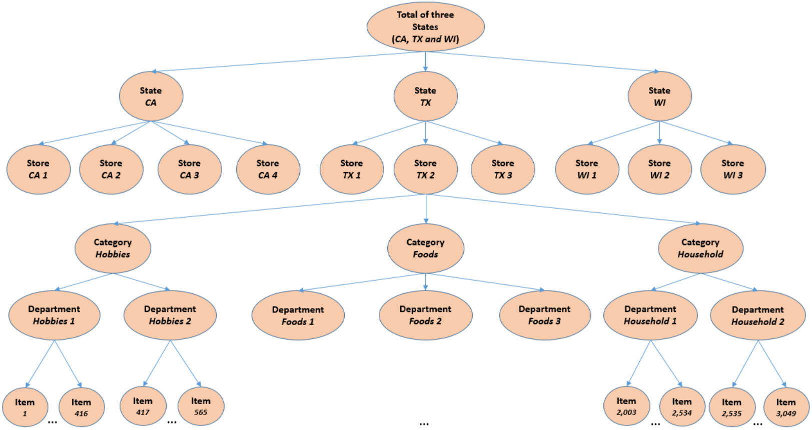
School of Statistics and Mathematics
Central University of Finance and Economics

Outline

1. Hierarchical time series
2. Top-down alignment method
3. Reconciliation method

Hierarchical time series

- Hierarchical time series are special network data.
- Constants exist that the forecasts on each hierarchy should be **coherent** with the aggregation structure.



Forecasting hierarchical time series
↬ Challenging for standard models

- Hard to use multivariate time series models, like VARs, because the bottom level is highly noisy or intermittent.
- Vine copula models are hard to scale up.
- Pure deep learning models fail to meet the constants.

Forecasting hierarchical time series

↪ How can we make these coherent?

- Having base forecasts at all nodes.
- **Bottom Up** (Schwarzkopf et al., 1988) ignores top nodes because the noise in the bottom level transferred to the top level.
- **Top Down** (Gross & Sohl, 1990) ignores bottom nodes. It discards information and can bias even unbiased base forecasts.

Business scenarios

- The lowest level of the hierarchy exhibits a strong **intermittent pattern**
- The upper hierarchy levels contain forecastable components such as the trend or seasonality aggregated by the bottom level.
- **Business scenarios:**
  - Managers at the headquarter or the investors may not be interested in the forecasts of a particular product in a specific store, but they are more concerned with the upper level's forecasts for all products, which affect the overall revenue (Anderer & Li, 2022).
  - Primer wanted to retain the nation's economy steady by keeping forecasts of the six provinces fixed (Zhang, Kang, Panagiotelis & Li, 2023).

Top-down alignment method
↪ Improved N-BEATS ensembles for upper levels

- The N-BEATS (Oreshkin et al., 2020) is a pure deep learning approach with a deep neural architecture based on backward and forward residual links.
- The performance of the N-BEATS model relies on the ensembling step, a forecasting combination technique that combines the forecasts from different models.
- The learning accuracy is very sensitive to the learning rate, and the convergence rate is slow. The lookahead optimizer (Zhang et al., 2019) is adopted to the N-BEATS model trainer further to improve the training accuracy and learning stability.

Top-down alignment method

↪ Bias-adjusted LightGBM model for the bottom level

- At the bottom level, we use a bias-adjusted LightGBM to model the intermittent time series.
- The root mean squared error (RMSE) loss is used in the LightGBM as follows,

$$RMSE = \sqrt{\frac{1}{h}\sum_{t=n+1}^{n+h}(Y_t - \hat{Y}_t)^2}.$$

- We further use a customized gradient for the RMSE loss as follows

$$\text{gradient} = \begin{cases} -2e_t & e_t < 0, \\ -2\lambda e_t & e_t \geqslant 0 \end{cases}$$

where $\lambda > 0$ is a tuning parameter named to allow for an asymmetric loss.

Top-down alignment method

↪ Aligning top level and aggregated bottom level forecasts

- Given the independent forecasting results for the top level from stable N-BEATS, the final step is to find the optimal bottom level forecasts produced by the LightGBM model. We tune the loss multiplier (λ) introduced for the LightGBM model.
- Objective: the RMSE between N-BEATS forecasts and aggregated bottom level forecasts reaches a minimum by the **hierarchical alignment**,

$$\arg\min_{\lambda} \left\{ \sqrt{\frac{1}{h} \sum_{t=n+1}^{n+h} \left(\hat{Y}_t^{(\mathrm{top})} - Agg(\hat{Y}_t^{(\mathrm{bottom})}(\lambda)) \right)^2} \right\},$$

where $Agg(\cdot)$ is the aggregating method used at the bottom level.

Top-down alignment method

↬ Hierarchical structure for M5 competition data

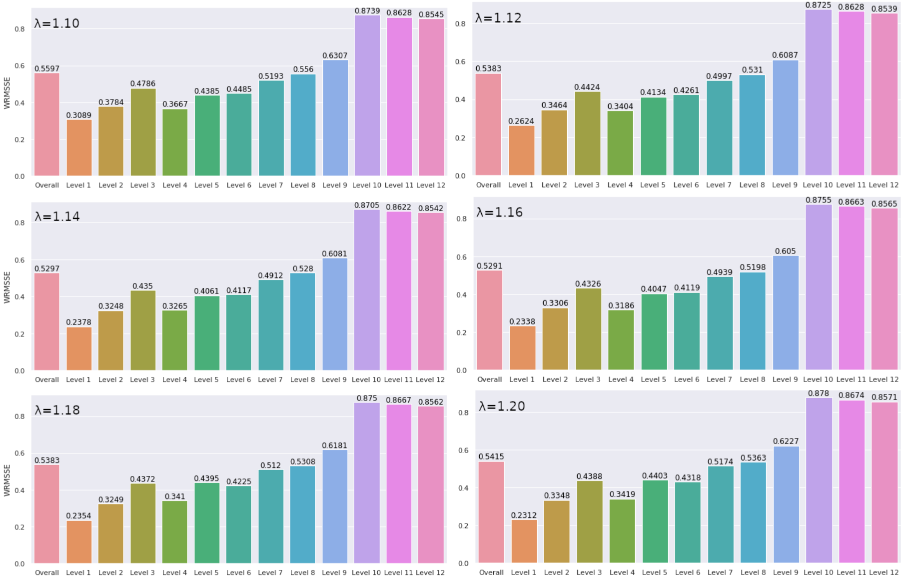
- The data consisting of a hierarchical structure of daily sales data of total $42,840$ series spanning 1,941 days.

Hierarchy level	Description	Number of series
1	All products, all stores, all states	1
2	All products by states	3
3	All products by store	10
4	All products by category	3
5	All products by department	7
6	Unit sales of all products, aggregated for each State and category	9
7	Unit sales of all products, aggregated for each State and department	21
8	Unit sales of all products, aggregated for each store and category	30
9	Unit sales of all products, aggregated for each store and department	70
10	Unit sales of product x, aggregated for all stores/states	3,049
11	Unit sales of product x, aggregated for each State	9,147
12	Unit sales of product x, aggregated for each store	30,490
Total		42,840

Top-down alignment method
↛ Features for intermittent time series (bottom level)

Feature	Description
sell_price	Price of item in store for given date.
event_type	108 categorical events, e.g. sporting, cultural, religious.
event_name	157 event names for event_type, e.g. super bowl, valentine's day, president's day.
event_name_2	Name of event feature as given in competition data.
event_type_2	Type of event feature as given in competition data.
snap_CA, TX, WI	Binary indicator for SNAP information in CA, TX, WI.
release	Release week of item in store.
price_max, min	Maximum, minimum price for item in store in the train data.
price_mean, std, norm	Mean, standard deviation, and normalized price for item in store in the train data.
item, price_nunique	Number of unique items, prices for item in store.
price_diff_w	Weekly price changes for items in store.
price_diff_m	Price changes of item in store compared to its monthly mean.
price_diff_y	Price changes of item in store compared to its yearly mean.
tm_d	Day of month.
tm_w	Week in year.
tm_m	Month in year.
tm_y	Year index in the train data.
tm_wm	Week in month.
tm_dw	Day of week.
tm_w_end	Weekend indicator.

Top-down alignment method
↬ Forecasting errors and bias multiplier

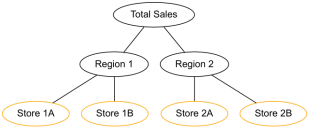


Top-down alignment method
↪ Conclusions

- We focus on the point forecast in this paper.
- N-Beats relies on efficient ensembles and LightGBM needs feature engineering.
- Probabilistic forecast with the presented scheme should be straightforward to implement with a corresponding probabilistic loss function.

Reconciliation method
↛ What is reconciliation?

- Base forecasts stacked in an n-vector as $\hat{\boldsymbol{y}}$. Reconciled forecasts $\tilde{\boldsymbol{y}}$ given by
$$\tilde{\mathbf{y}} = \mathbf{S}\left(\mathbf{S}'\mathbf{W}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{W}\hat{\mathbf{y}}$$
- Different choices of $\mathbf{W}$ give rise to different methods, e.g., MinT to minimize $E\left[(\boldsymbol{y}-\tilde{\boldsymbol{y}})'(\boldsymbol{y}-\tilde{\boldsymbol{y}})\right]$ (Wickramasuriya et al., 2019).
- $\boldsymbol{S}$ is the summing matrix containing zeros and ones.



$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{b}$$

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Reconciliation method
↪ Reconciliation with imitable series

- Reconciliation finds $\tilde{\boldsymbol{y}}$ to minimize

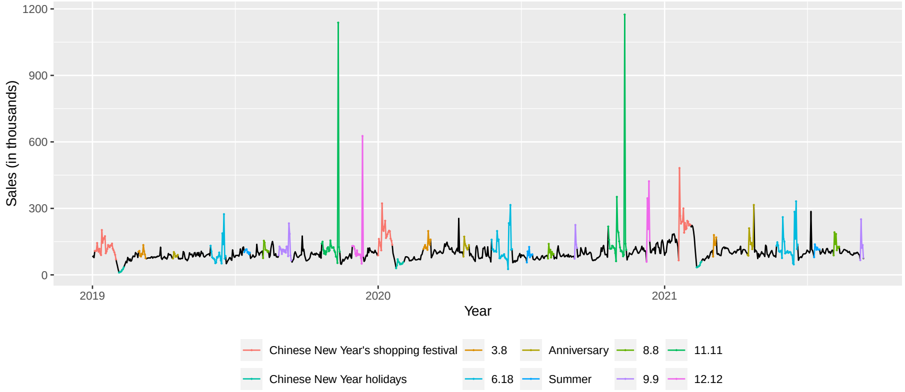
$$(\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})'\boldsymbol{W}(\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})$$

- Subject to $\tilde{\boldsymbol{y}}$ being coherent, immutability constraints and non-negativity constants.
- Solved by quadratic programming.

Reconciliation method

↪ Application

- Sales data from Chinese online retailer
- Consider "Food" sales as top level.
- There are 40 "middle" level categories, 1905 "bottom" level categories.
- Many series are intermittent. Promotions are very important



Reconciliation method

↪ Base models

- Intermittent Series (more than 60% zeroes) use simple exponential smoothing.
- Regression (predictors are strength of promotion) with ARIMA errors and Box Cox transformation for other series.
- Immutable series are on top level when series with longer histories (more than 1 year)

Table: Out of sample forecasting accuracy (RMSE) for the Chinese online retailer data. Subcolumns "C" and "C+NN" show forecasting accuracies of reconciliation with immutability constraints without and imposing non-negativity constraints, respectively. Subcolumns "U" and "U+NN" show forecasting accuracy of unconstrained reconciliation without and with imposing non-negativity constraints, respectively. The accuracy of base forecasts is also shown for comparison. For columns "C" and "C+NN", the top level forecast as well as forecasts for intermittent series (Bottom-2) and series with long histories (Bottom-3) are immutable, while forecasts for the middle level and remaining bottom level (Bottom-1) are mutable.

Level	Base	OLS				WLS_s				WLS_v			
		C	C+NN	U	U+NN	C	C+NN	U	U+NN	C	C+NN	U	U+NN
Top	2.94	2.94	2.94	2.93	2.92	2.94	2.94	2.72	2.72	2.94	2.94	2.75	2.77
Middle	2.66	9.31	4.94	272.83	48.84	6.41	4.83	16.09	6.50	2.43	2.47	2.39	2.40
Bottom-1	2.04	8.98	4.31	3.98	2.70	7.19	3.71	2.96	2.32	1.97	1.88	1.86	1.83
Bottom-2	0.11	0.11	0.11	42.66	15.43	0.11	0.11	26.99	8.34	0.11	0.11	1.52	1.52
Bottom-3	1.08	1.08	1.08	1.64	1.48	1.08	1.08	1.36	1.25	1.08	1.08	1.58	1.19

Reconciliation method
↪ Main findings

- Imposing immutability constraints still leads to improvements over base forecasts.
- Immutability stabilizes forecasting performance of intermittent series in particular.
- Imposing immutability does not lead to better accuracy in all series.
- Imposing non-negativity constraints generally improves performance.

Next step

- Hierarchical forecasting at scale.
- Discrete hierarchical forecasting.
- Probabilistic hierarchical forecasting.
- Forecast combination on infinite models.

References I

- Zhang, B., Kang, Y., Panagiotelis, A. & Li, F. (2023). "Optimal Reconciliation with Immutable Forecasts". *European Journal of Operational Research* 308.(1), pp. 650–660.
- Anderer, M. & Li, F. (2022). "Hierarchical Forecasting with a Top-down Alignment of Independent-Level Forecasts". *International Journal of Forecasting* 38.(4), pp. 1405–1414.
- Oreshkin, B. N., Carpov, D., Chapados, N. & Bengio, Y. (2020). "N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting". In: *Procedings of The 2020 International Conference on Learning Representations*.
- Wickramasuriya, S. L., Athanasopoulos, G. & Hyndman, R. J. (2019). "Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series through Trace Minimization". *Journal of the American Statistical Association* 114.(526), pp. 804–819.

References II

Zhang, M., Lucas, J., Ba, J. & Hinton, G. E. (2019). "Lookahead Optimizer: k Steps Forward, 1 Step Back". In: *Advances in Neural Information Processing Systems*. Vol. 32.

Gross, C. W. & Sohl, J. E. (1990). "Disaggregation Methods to Expedite Product Line Forecasting". *Journal of Forecasting* 9.(3), pp. 233–254.

Schwarzkopf, A. B., Tersine, R. J. & Morris, J. S. (1988). "Top-down versus Bottom-up Forecasting Strategies". *International Journal of Production Research* 26.(11), pp. 1833–1843.

Thank you!

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