HIFOR

HYBRID HIERARCHICAL DEMAND FORECASTING FRAMEWORK FOR SUPPLY CHAINS

Ümit Kaan Usta
umitkaan19@marun.edu.tr

Süleyman Emirhan Uslu
suleymanemirhanuslu@icloud.com

Gökberk Köksoy
gokberkkoksoy@icloud.com

Supervisor: Prof.Dr. Çiğdem Eroğlu Erdem

Research Mentorship by Amazon

INTRODUCTION

Background

- Forecasting: Predict future from past
- Hierarchy: Parent-child relationship btwn time-series
- **Accuracy → $$ savings, meet demand, less idle capacity, no out-of-stock** [1]

Challenges

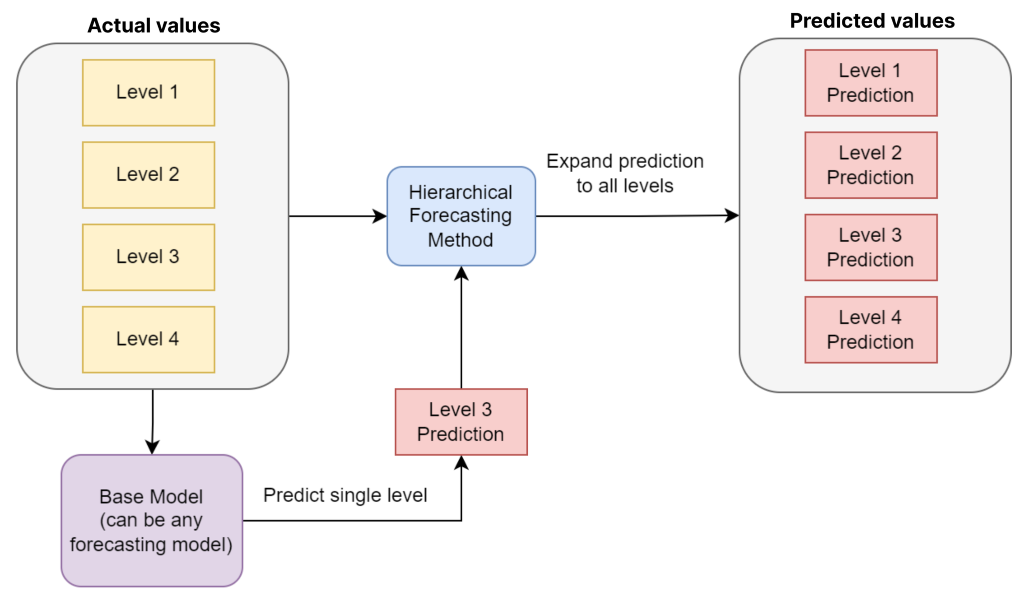
- **1.** Combine HF methods on a base model
- **2.** Combine base models on an HF method

Our solution

- Combine base models & different HF methods
- Customizable "forecasting units" for each level
- RNN-like approach
  - Traverse the hierarchy top-down
  - Output of previous level = feature of next level (hidden state)
- Post-process to ensure coherency

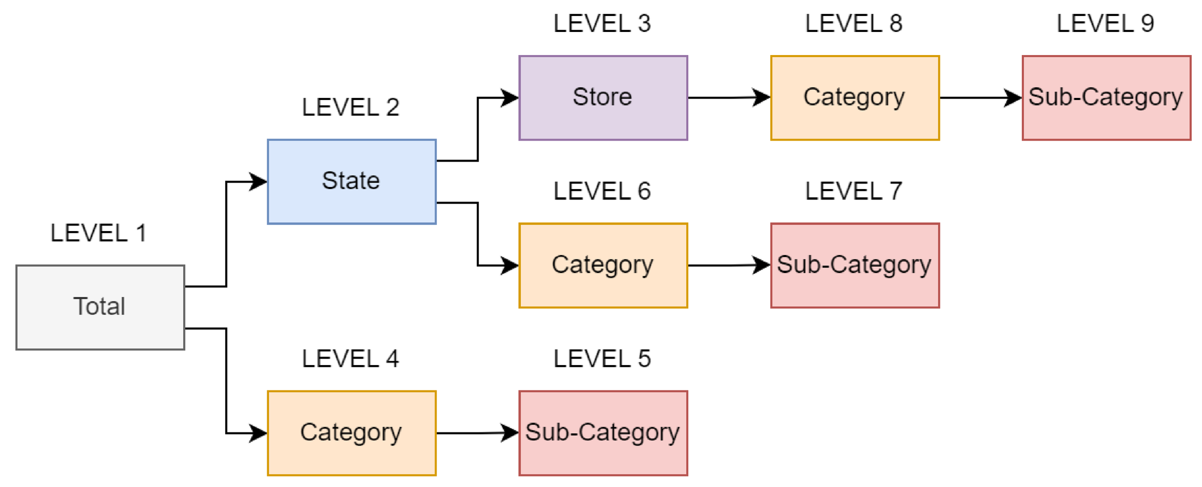
How hierarchical forecasting (HF) works

Idea: Take prediction at single/multiple levels, expand into all levels of the hierarchy



DATASET

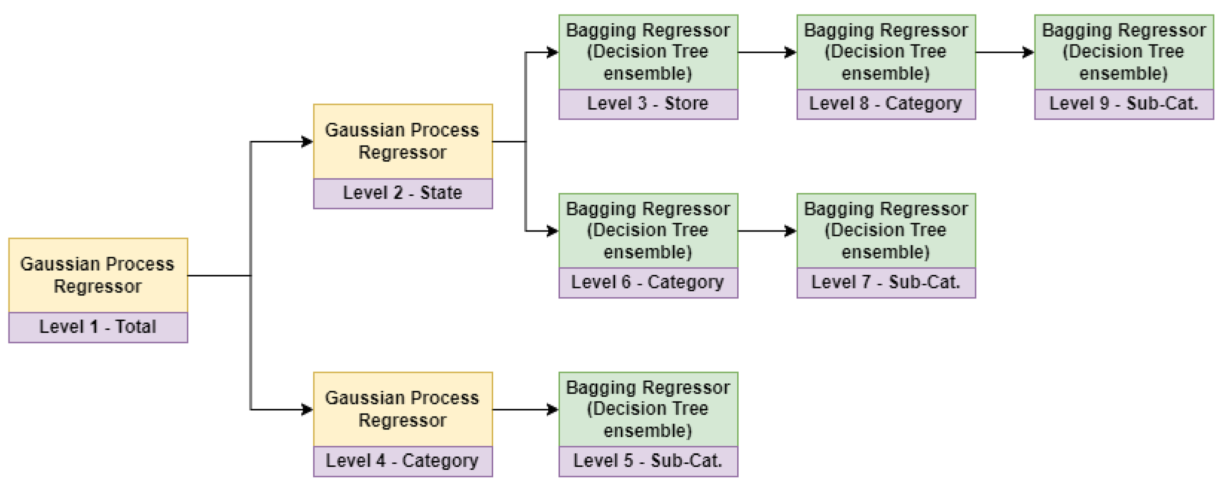
- Subset of Kaggle M5 Competition Dataset [2], A well-known & representative benchmark [3]
- Hierarchical demand forecasting competition on Walmart's daily sales data
- 5 years of daily sales (2011-2016), 9 levels, 154 time-series
- Forecast 28 days ahead



METHODOLOGY

Experiment settings

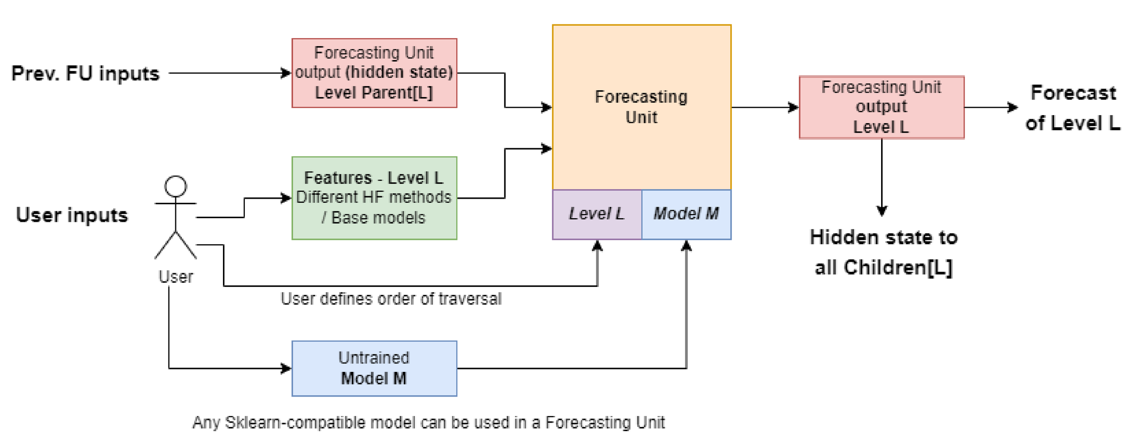
- Experiments for Task 1 (Reconciler) & Task 2 (Ensembler)
- Base models (on levels 3, 8, 9)
  - Statistical: AutoARIMA
  - LGBM + auto-feat-eng: LGBM+catch22, LGBM+TSFresh
  - M5 top solutions: M5-winner, M5-2nd, M5-3rd
- HF methods
  - Simple: BottomUp, MiddleOut
  - Model-based: MinTrace (MinT), Weighted Least Squares (WLS)
- HiFor settings
  - Traversal order: Level 1→ Level 9
  - Coherence step: MiddleOut on Level 8
  - Model choices - see below:



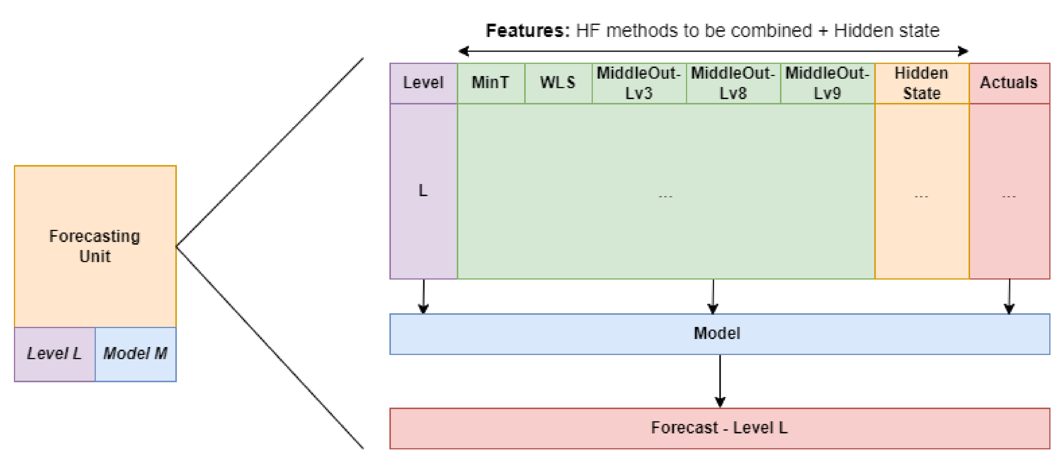
HOW HIFOR WORKS

Forecasting Unit - building block of HiFor

Each level L is forecasted by a "Forecasting Unit (FU)"

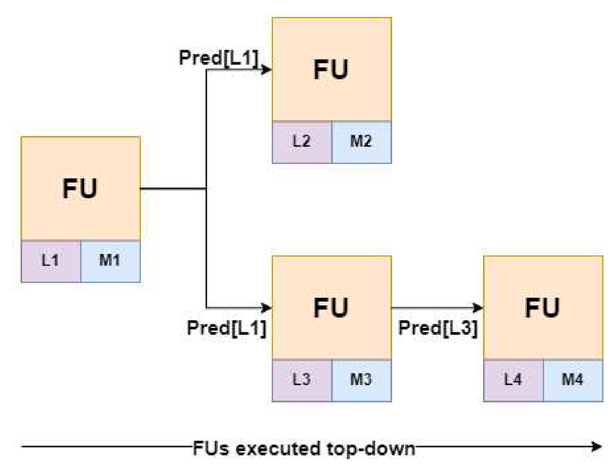


A deeper dive - How HF methods are combined



A network of forecasting units

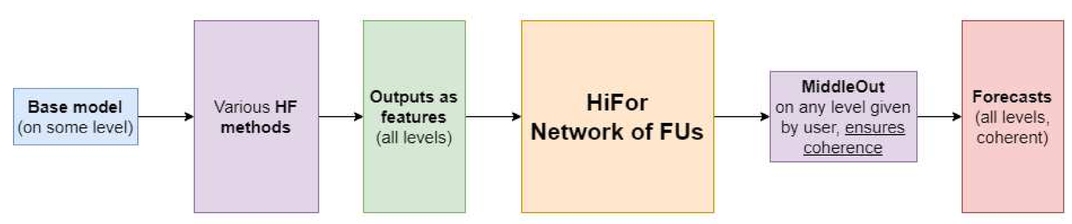
An example hierarchy with 4 levels



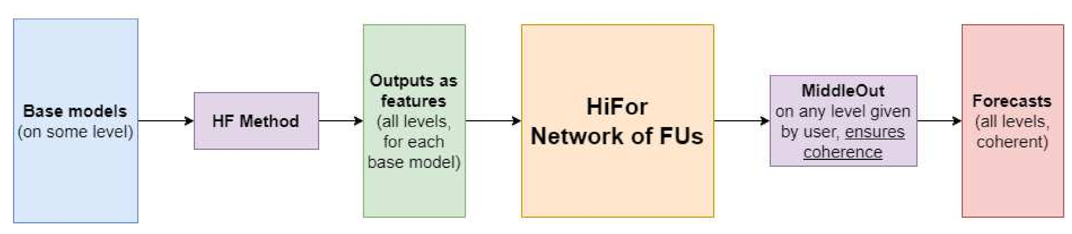
Unlike MiddleOut, HiFor can directly work with non-chain-like hierarchies

Summary

Task 1 - HiFor as Reconciler: Combine HF methods on some base model



Task 2 - HiFor as Ensembler: Ensemble base models on some HF method



EXPERIMENT RESULTS

Loss Function - WRMSSE (Weighted Scaled RMSE)

- "How much does the MSE improve on Naive forecast?"
- Lower is better
- Naive forecast: Taking day-1 value as forecast

$$RMSSE = \sqrt{\frac{\text{MSE}}{\text{MSE of Naive Forecast}}}$$

$$\text{WRMSSE}_{\text{level}} = \sum_{\text{ts}} (\text{AvgSalesPrice}_{\text{level,ts}})(\text{RMSSE}_{\text{level,ts}})$$

- 'ts' stands for each time-series in a given level
- WRMSSE > 1 → worse than Naive forecast
- Avg WRMSSE = Equal weighted avg of all levels **used as the metric to compare models**

Results - improvements shown in bold

HiFor as Reconciler: Single base model, combine mult. HF methods

Base Model	HiFor (ours)	Best HF Method	Improvement%
LightGBM+TSFresh	**0.543**	0.582 (MiddleOut on Level8)	**7.16**
AutoARIMA	**0.644**	0.680 (MiddleOut on Level8)	**5.71**
LightGBM+Catch22	**1.651**	1.672 (MiddleOut on Level8)	**1.29**
M5-Winner	**0.995**	1.003 (MiddleOut on Level8)	**0.80**
M5-2ndPlace	**0.984**	0.988 (MiddleOut on Level8)	**0.41**
M5-3rdPlace	**0.995**	0.999 (MiddleOut on Level8)	**0.39**

HiFor as Ensembler: Single HF method, combine mult. base models

HF Method	HiFor (ours)	Best Base Model	Improvement%
MiddleOut on Level3	**0.656**	0.673 (LightGBM+TSFresh)	**2.54**
MiddleOut on Level8	**0.57**	0.582 (LightGBM+TSFresh)	**2.18**
MiddleOut on Level9	0.526	0.515 (LightGBM+TSFresh)	-2.01

*Experiments with HF methods MinT, WLS pruned due to no base model having WRMSSE < 1 with them

Code available on GitHub: github.com/umitkaanusta/hifor-experiments

Runtime of HiFor: ~3min for this config on AWS EC2 c4.xlarge (4 vCPUs, 8GB RAM)

CONCLUSION & FUTURE WORK

- **HiFor overperforms all widely used HF methods in reconciliation, and performs well in ensembling**
- Can work in non-chainlike hierarchies (unlike MiddleOut)
- Can work with only a single level of base model forecasts (unlike MinT, WLS)
- Experiment with different datasets & domains
- Compare with newer but less known HF methods
- Publicize HiFor and gather user feedback

TECHNOLOGIES USED

pandas NumPy kaggle learn LightGBM

REFERENCES

[1] R. J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice (2nd ed). Melbourne: Otexts, 2018.

[2] Kaggle M5 Forecasting Competition Dataset. [Online]. Available: https://www.kaggle.com/c/m5-forecasting-accuracy (Date of access: 26/11/2022).

[3] E. Theodorou, S. Wang, Y.Kang, E. Spiliotis, S. Makridakis, V. Assimakopoulos, "Exploring the representativeness of the M5 competition data", International Journal of Forecasting, vol. 38, no 4, pp. 1500- 1506, 2022.