Final Assignment: Research Project

(N-HiTS: Neural Hierarchical Interpolation for Time Series Forecasting)

Honggang YAN, Ananya AGRAWAL, Jiashuo HAN, Xiang LI, Chuang LU

Master Applied Data Science, Frankfurt School of Finance & Management, Frankfurt, Germany

Deep Learning

Prof. Dr. Florian Ellsäßer

https://github.com/Blair-Lu-c/MADS/tree/main/Deep_Learning_Project

Introduction

The research paper focuses on advancing long-horizon time series forecasting with optimal computational efficiency. Long-horizon forecasting, which involves predicting future data points over a significant number of time steps, is crucial in various real-world applications. However, it poses challenges due to increased forecast volatility and computational complexity.

The paper introduces the N-HiTS model, building upon the successful N-BEATS framework, to address these challenges. By incorporating innovative techniques such as multi-rate data sampling and hierarchical interpolation, N-HiTS aims to significantly improve both the accuracy and computational efficiency of long-horizon forecasting.

The findings showcase substantial advancements, outperforming state-of-the-art baselines and providing interpretable forecast decompositions. This research not only represents a significant breakthrough in time series forecasting but also holds implications for broader applications of deep learning in business contexts.

Summary

2.1. Main objectives and motivation of the research

Long-horizon forecasting with optimal computational complexity.
The primary motivation for this research is to improve the effectiveness of long-horizon time series forecasting, which is a critical task in various real-world applications such as risk management, infrastructure planning, and healthcare monitoring. At the same time, to optimize the time and cost of computation. Long-horizon forecasting involves making predictions for a significant number of time steps into the future, which can be highly challenging due to the potential for increased forecast volatility and computational complexity.

The paper acknowledges that recent progress in neural time series forecasting has led to improvements in the performance of large-scale forecasting systems. This progress has been marked by the adoption of advanced architectural elements, including the attention mechanism and Transformer-inspired approaches. These developments have made it easier to capture long-range dependencies and to efficiently predict long sequences in a single forward pass. These advancements have significantly enhanced the forecasting capabilities of neural networks.

Prior research in the field of neural long-horizon forecasting has primarily focused on improving the efficiency of the attention mechanism. Various techniques, such as making self-attention sparse, redefining attention through locality-sensitive hashing, and utilizing FFT, have been explored to reduce computational costs and enhance accuracy. While these efforts have led to incremental improvements in computational efficiency and forecast accuracy, a comprehensive solution for long-horizon forecasting has not yet been discovered.

The paper emphasizes the need for a breakthrough in long-horizon time series forecasting that can substantially reduce computational costs while simultaneously improving forecast accuracy. This research seeks to provide such a breakthrough by introducing the N-HiTS model, which redefines existing forecasting architectures.

2.2. Key Methodologies and Techniques

N-BEATS and N-HiTs

N-HiTS builds upon the architectural principles of N-BEATS and introduces innovative techniques such as multi-rate data sampling and hierarchical interpolation. These techniques collectively enable more efficient and accurate long-horizon time series forecasting while maintaining interpretability.

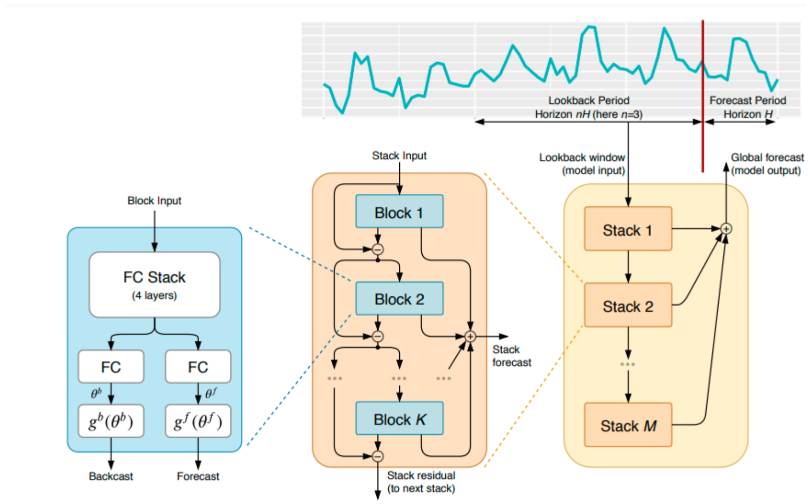


Figure 1: N-BEATS Architecture

To introduce the methodologies and techniques employed in the N-HiTS research, we begin by exploring its basis in the N-BEATS framework which is depicted in Fig.1. N-BEATS was designed with three key principles in mind: simplicity, generality, and extendibility. The architecture was deliberately kept simple yet expressive, avoiding time-series-specific components and ensuring interpretability. N-BEATS introduces a multi-stack structure, with each stack consisting of blocks, and these blocks contain fully connected layers. The model generates both forecasts and backcasts, making it possible to assess the model's predictions against actual data. Residual connections facilitate capturing information that might have been missed by previous blocks, and the combination of these stacks results in the final forecast.

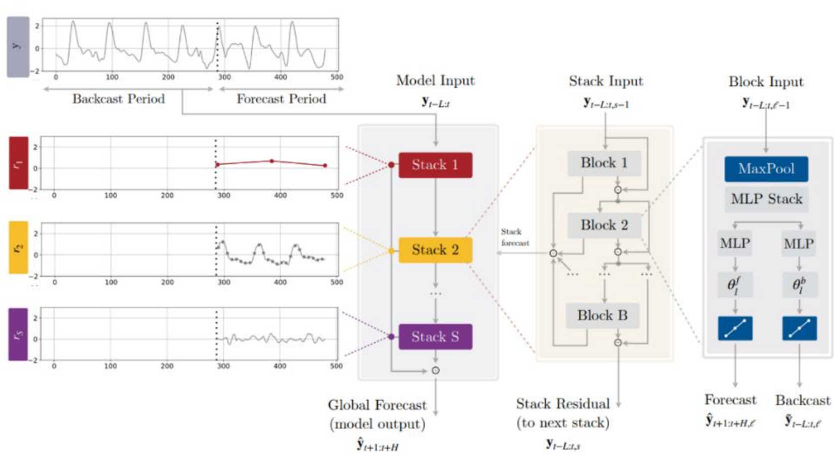


Figure 2: N-HiTS Architecture

Building upon this foundation, the N-HiTS architecture depicted in Fig.2 comprises multiple stacks, each containing blocks with multilayer perceptrons (MLPs). These MLPs generate coefficients for both backcast and forecast outputs. The backcast output is used to clean inputs for subsequent blocks, while forecasts are combined to produce the final prediction. This hierarchical organization of blocks allows specialization in forecasting different frequency bands of the time-series signal.

Multi-rate data sampling

The N-HiTS incorporates several methodologies and techniques to address the specific challenges of long-horizon time series forecasting. One of the fundamental techniques used in the N-HiTS model is multi-rate data sampling. This approach introduces a new layer in front of fully-connected blocks, which allows for subsampling the input time series data. The key idea here is to reduce the memory footprint and computational requirements while maintaining the ability to model long-range dependencies. By using MaxPool layers with varying kernel sizes, the model focuses on different components of the input signal at different scales. Larger kernel sizes tend to emphasize the analysis of low-frequency, large-scale components, which are critical for making consistent long-horizon forecasts. This technique effectively limits the width of the input for most blocks, reducing memory requirements and the risk of overfitting while preserving the overall receptive field.

Hierarchical interpolation

To ensure the smoothness of multi-step predictions, the research introduces the concept of hierarchical interpolation, which serves as a performance enhancer, since it enables the model to keep the number of parameters reasonably low and computationally less difficult. It is designed to handle the dimensionality of the forecast more efficiently, particularly for long forecasting horizons. Unlike conventional approaches where the neural network's prediction dimension equals the forecasting horizon length, the research proposes a novel approach.

It defines the dimensionality of the interpolation coefficients in terms of an expressiveness ratio, which controls the number of parameters per unit of output time. In the hierarchical N-HiTS structure, blocks closer to the input employ smaller expressiveness ratios and larger kernel sizes, enabling more aggressive interpolation for generating low-granularity signals from heavily subsampled and smoothed data. The hierarchical forecast combines outputs from all blocks, effectively integrating interpolations at different time-scale hierarchy levels. Each block specializes in handling distinct scales of input and output signals, creating a well-structured hierarchy of interpolation granularity that accounts for various time-scale levels. To manage a wide range of frequency bands while controlling the number of parameters, the model adopts exponentially increasing expressiveness ratios. Notably, each stack can specialize in modeling different known time-series cycles, like weekly or daily patterns, by using matching expressiveness ratios, enhancing the model's adaptability to diverse temporal patterns.

To provide a theoretical foundation for the hierarchical interpolation technique, the research introduces the Neural Basis Approximation Theorem. This theorem mathematically supports the model's ability to approximate infinitely long or dense forecasting horizons as long as the interpolation function follows certain properties, and the forecast relationships are smooth.

2.3. Results and findings

Higher Accuracy and faster

The N-HiTS research presents significant advancements in long-horizon time series forecasting. When compared to state-of-the-art multivariate baselines, it outperforms with a 14% improvement in Mean Absolute Error (MAE) and a 16% reduction in Mean Squared Error (MSE) across various datasets and horizons. Even for the longest forecasting horizon, N-HiTS reduces MAE by 11% and MSE by 17%. Univariate baselines are also surpassed. N-HiTS is highly efficient, being 45 times faster than Autoformer and requiring fewer parameters.

Interpretability

N-HiTS develops the ability to produce interpretable forecast decomposition providing valuable information about trends and seasonality in separate channels. This novel interpretable decomposition can provide insights to users, improving their confidence in high-stakes applications.

The research also discusses rigorous training and hyperparameter optimization, resulting in optimal configurations. Ablation studies confirm the benefits of hierarchical interpolation and multi-rate sampling, with their combined use outperforming other models. Further studies reveal key choices impacting performance, including pooling methods and interpolation techniques. To conclude, N-HiTS offers complementarity and effectiveness in long-horizon forecasting, suggesting promising directions for future research. It excels in providing both accuracy and interpretability. The findings question the effectiveness of existing multivariate forecasting approaches and hint at the potential of integrating N-HiTS with Transformer-inspired architectures. The hierarchical interpolation component opens up avenues for multiresolution analysis.

2.4. Implications and significance of the study in the broader context of deep learning.

The N-HiTS research marks a breakthrough in long-horizon time series forecasting. Its innovative techniques, like hierarchical interpolation and multi-rate data sampling, redefine deep learning approaches for complex forecasting tasks. This not only enhances forecast accuracy but also reduces computational costs, making it highly valuable in real-world applications. Moreover, N-HiTS introduces a new level of interpretability, crucial for trust in critical areas. Its efficiency gains, with significantly faster computations and fewer parameters, hold promise for resource-constrained environments in businesses.

The hierarchical interpolation component not only advances time series forecasting but also offers a novel approach to multiresolution analysis, applicable in diverse fields from image processing to audio analysis.

In essence, the superior performance of N-HiTS compared to state-of-the-art multivariate baselines challenges conventional wisdom in the field of forecasting. This suggests that established methods may not always be the most effective, and innovative approaches like N-HiTS should be considered viable alternatives.

Analysis

3.1. Strengths

Multi-Rate Data Sampling

This paper incorporates MaxPool(with different kernel sizes) in front of fully connected blocks, significantly reducing the memory footprint and the amount of computation needed while maintaining the ability to model long-range dependencies. The blocks focus on analyzing large-scale components that will produce consistent and more accurate long-horizon forecasts.

Hierarchical Interpolation

This paper enforces the smoothness of the multi-step predictions by reducing the dimensionality of the neural network's prediction and matching its time scale with that of the final output via multi-scale hierarchical interpolation, which serves as a performance enhancer since it enables the model to keep the number of parameters reasonably low and computationally less difficult.

3.2. Weaknesses

Pooling Configuration

This paper conducts a study to compare the accuracy effects of different pooling alternatives based on five datasets, which are ETTm2, Exchange, ECL, TrafficL, and Weather. The MaxPool and AveragePool configurations depicted in Fig.3 are considered. The authors conclude the MaxPool operation consistently outperforms the AveragePool alternative, with MAE improvements of up to 15% and MSE up to 8% in the most extended horizon. On average, the forecasting accuracy favors the MaxPool method across the datasets and horizons. However, we think using this experimental way to decide which pooling configuration to implement is a bit dogmatic and overlooks the nature of the data and the specific characteristics of the problem we are trying to solve. Only using empirical evaluation is not convincing enough. Actually, in the result of comparing the pooling configuration, the long-term weather forecast and mid-term exchange forecast favor AveragePool.

		MaxPool		AveragePool	
		MSE	MAE	MSE	MAE
ETTm2	96	**0.185**	0.265	0.186	**0.262**
	192	**0.244**	**0.308**	0.257	0.315
	336	**0.301**	**0.347**	0.312	0.356
	720	**0.429**	**0.438**	0.436	0.447
ECL	96	**0.152**	**0.257**	0.181	0.290
	192	**0.172**	**0.275**	0.212	0.320
	336	**0.197**	**0.304**	0.238	0.343
	720	**0.248**	**0.347**	0.309	0.400
Exchange	96	**0.109**	**0.232**	0.112	0.238
	192	0.280	0.375	**0.265**	**0.371**
	336	**0.472**	0.504	0.501	**0.502**
	720	**1.241**	**0.823**	1.610	0.942
TrafficL	96	**0.405**	**0.286**	0.468	0.332
	192	**0.421**	**0.297**	0.490	0.347
	336	**0.448**	**0.318**	0.531	0.371
	720	**0.527**	**0.362**	0.602	0.400
Weather	96	**0.164**	**0.199**	0.167	0.200
	192	**0.224**	**0.255**	0.226	0.255
	336	0.285	0.311	**0.284**	**0.297**
	720	0.366	0.359	**0.360**	**0.352**
P. Diff.	96	**-8.911**	**-6.251**	0.000	0.000
	192	**-7.544**	**-6.085**	0.000	0.000
	336	**-8.740**	**-4.575**	0.000	0.000
	720	**-15.22**	**-8.318**	0.000	0.000

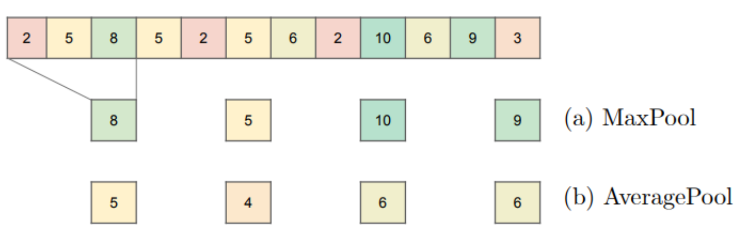
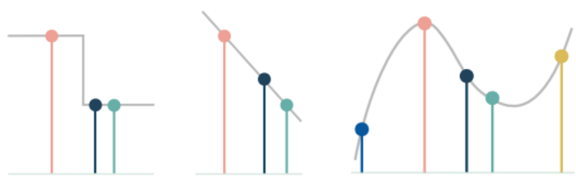


Figure 3: Pooling Configuration

Interpolation Configurations

This paper conducts an empirical evaluation of long multi-horizon multivariate forecasts for N-HiTS with different interpolation configurations, which are nearest neighbor, linear, and cubic, depicted in Fig.4, over the same five datasets. In general, the linear interpolation and the cubic interpolation are better. But when comparing between linear and cubic the results are inconclusive as different datasets and horizons have slight performance differences. This paper states: On average across the datasets both the forecasting accuracy and computational performance favor the linear method, with which we conducted the main experiments of this work with this technique. In terms of computational performance, yes, linear interpolation is favored but we cannot overlook that in some cases cubic interpolation has higher accuracy which means cubic interpolation may interpret data better.

		Linear		Cubic		N.Neighbor	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.185	0.265	**0.179**	**0.256**	0.180	0.259
	192	0.244	0.308	**0.241**	**0.303**	0.252	0.315
	336	**0.301**	**0.347**	0.314	0.358	0.302	0.351
	720	**0.429**	**0.438**	0.439	0.450	0.442	0.455
ECL	96	0.152	0.257	**0.149**	**0.252**	0.151	0.255
	192	**0.172**	**0.275**	0.174	0.279	0.175	0.279
	336	0.197	0.304	**0.190**	**0.295**	0.211	0.318
	720	**0.248**	**0.347**	0.256	0.353	0.263	0.358
Exchange	96	**0.109**	**0.232**	0.1307	0.254	0.126	0.248
	192	0.280	0.375	**0.247**	**0.357**	0.357	0.416
	336	**0.472**	**0.504**	0.625	0.560	0.646	0.560
	720	**1.241**	**0.823**	1.539	0.925	1.740	0.973
TrafficL	96	0.405	0.286	**0.402**	**0.282**	0.405	0.359
	192	0.421	0.297	**0.417**	**0.295**	0.419	0.201
	336	0.448	0.318	**0.446**	**0.315**	0.445	0.253
	720	0.527	0.362	0.540	0.366	**0.525**	**0.318**
Weather	96	0.164	0.199	0.162	0.203	**0.161**	0.360
	192	0.224	0.255	0.225	0.257	**0.218**	0.928
	336	**0.285**	**0.311**	0.285	0.304	0.298	0.988
	720	**0.366**	**0.359**	0.380	0.369	0.368	1.047
P. Diff.	96	**-0.907**	**-0.717**	0.146	1.61	0.000	0.000
	192	-5.582	-3.259.	**-7.985**	**-4.332**	0.000	0.000
	336	**-10.516**	**-4.199**	-2.108	-1.455	0.000	0.000
	720	**-15.800**	**-7.042**	-5.480	-1.579	0.000	0.000



(a) N.Neighbor (b) Linear (c) Cubic

Figure 4: Interpolation Configurations

Order of Hierarchical Representations

This paper uses an unconventional approach by employing a Top-Down hierarchical structure depicted in Fig.5. This means that we initially prioritize synthesizing predictions for low-frequency components, gradually incorporating higher-frequency details. The underlying idea is that this Top-Down hierarchy serves as a form of regularization, guiding the model to concentrate on the broader factors influencing predictions, rather than fixating on intricate details.

To validate these hypotheses, the paper conducted an experiment where it created a Bottom-Up hierarchy of prediction and then compared the performance on the validation dataset. The paper states: The Top-Down predictions consistently outperform the Bottom-Up counterpart. However, in terms of the weather dataset, it is inconclusive again.

		Top-Down		Bottom-Up	
		MSE	MAE	MSE	MAE
ETTm2	96	**0.185**	**0.265**	0.191	0.266
	192	**0.244**	**0.308**	0.261	0.320
	336	**0.301**	**0.347**	0.302	0.353
	720	**0.429**	**0.438**	0.440	0.454
ECL	96	**0.152**	**0.257**	0.164	0.270
	192	**0.172**	**0.275**	0.186	0.292
	336	**0.197**	**0.304**	0.217	0.327
	720	**0.248**	**0.347**	0.273	0.369
Exchange	96	**0.109**	**0.232**	0.114	0.242
	192	**0.280**	**0.375**	0.436	0.452
	336	**0.472**	**0.504**	0.654	0.574
	720	**1.241**	**0.823**	1.312	0.861
TrafficL	96	**0.405**	**0.286**	0.410	0.292
	192	**0.421**	**0.297**	0.427	0.305
	336	**0.448**	**0.318**	0.456	0.323
	720	**0.527**	**0.362**	0.557	0.379
Weather	96	0.164	**0.199**	**0.163**	0.200
	192	0.224	0.255	**0.219**	**0.252**
	336	**0.285**	**0.311**	0.288	0.311
	720	0.366	0.359	**0.365**	**0.355**
P. Diff.	96	**-2.523**	**-2.497**	0.000	0.000
	192	**-12.296**	**-6.793**	0.000	0.000
	336	**-11.176**	**-5.507**	0.000	0.000
	720	**-4.638**	**-3.699**	0.000	0.000

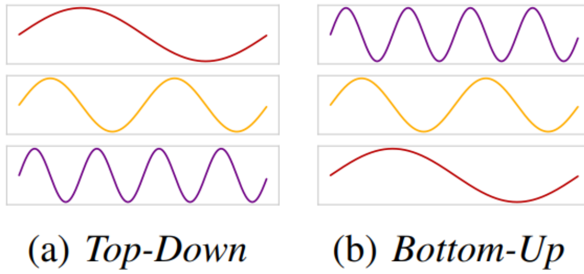


Figure 5: Order of Hierarchical Representations

3.3. Potential Improvements.

This paper demonstrates a commendable level of innovation and elegance in its approach to time series forecasting, utilizing straightforward fully connected neural networks alongside multi-rate sampling and hierarchical interpolation methodologies.

To refine this methodology, it's important to carefully consider the configurations for pooling, interpolation, and hierarchical representations. Instead of relying solely on experiments, we should select settings that align with the specific problem at hand.

For instance, when choosing between MaxPooling and average pooling, here are some considerations to help you decide: MaxPooling emphasizes crucial features, is robust to small shifts, and is computationally efficient. It retains essential information while reducing data dimensionality. Average Pooling smoothens data, maintains spatial information, and prevents overemphasis on outliers. It provides a broader sense of value distribution in a time period or region.

Implementation

4.1. Accuracy in Reproducing the Paper's Results

Reproduction of the N-HiTS empirical results compared to the N-BEATS, Autoformer, and Informer models for ETTm2, Exchange rate, and Weather datasets indicates that the N-HiTS performs the best in MAE & MSE. The ETTm2 dataset with NHITS and Transformers models as the sample in the original code (LongHorizon_with_NHITS.ipynb and LongHorizon_with_Transformers.ipynb). Based on this, we implemented the exchange rate and weather datasets at the end of the original model file.

Table 1 presents our reproduction results against the original paper, the difference between the results in our implementation and the original paper is because one of the original hyperparameters of the paper was missing, but the results are close enough. The results in bold are the best.

Dataset	Version	NHiTs		NBEATs		Autoformer		Informer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
EETM2	Reproduction	**0.183**	**0.261**	0.184	0.268	0.255	0.335	0.419	0.445
	Paper	**0.176**	**0.255**	0.184	0.263	0.255	0.339	0.365	0.453
Exchange rate	Reproduction	**0.087**	0.209	0.089	**0.208**	0.162	0.296	0.947	0.715
	Paper	**0.092**	**0.202**	0.098	0.206	0.197	0.323	0.847	0.752
Weather	Reproduction	**0.166**	**0.198**	0.178	0.208	0.231	0.290	0.258	0.320
	Paper	**0.158**	**0.195**	0.167	0.203	0.266	0.336	0.300	0.384

Table 1 Comparison of reproduction and the original paper

In these four models and three datasets, we achieved replication. But for other models and datasets, we did not reproduce the results because we are limited by memory and computational cost. The original code changes and additions are explained as follows:

- N-HiTS:
  Two hyperparameters we changed for the N-HiTS model were batch size (number of windows in the batch) and the frequency of the dataset according to the summary of the dataset in the original paper shown in Table 2, and we reduced the training steps to 100 due to computational cost.

Dataset	Frequency	Time Series	Total Observations	Test Observations	Rolled forecast evaluation data points	Horizon (H)
ETTm2	15 Minute	7	403,200	80,640	5.81e7	{96, 192, 336, 720}
Exchange	Daily	8	60,704	12,136	8.74e6	{96, 192, 336, 720}
ECL	Hourly	321	8,443,584	1,688,460	1.22e9	{96, 192, 336, 720}
TrafficL	Hourly	862	15,122,928	3,023,896	2.18e9	{96, 192, 336, 720}
Weather	10 Minute	21	1,106,595	221,319	1.59e8	{96, 192, 336, 720}
ILI	Weekly	7	6,762	1,351	9.73e5	{24, 36, 48, 60}

Table 2 Summary of the dataset

- NBEATS, Autoformer, and Informer
  We reproduced the comparing models using the same approach as N-HiTS.

4.2. Innovativeness and Code Enhancements

In this section of our report, we have undertaken the task of model implementation on the 'Airpassenger' dataset, which serves as a fundamental historical repository comprising monthly aggregates of international passenger counts over the temporal span from 1949 to 1960. In the pursuit of a rigorous and systematic approach to this implementation, we have closely adhered to the methodologies and procedures laid out by the model's author. Specifically, we have employed the N-HiTS model as the primary computational framework, utilizing this model to make predictive inferences concerning the monthly counts of international passengers for the final twelve months of the dataset, encompassing the time period from January 1960 to December 1960.

In evaluating the effectiveness and accuracy of our model, we have employed two prominent performance metrics, namely the Mean Absolute Error (MAE) and the Mean Squared Error (MSE). These metrics serve as essential quantitative indicators, allowing us to assess the extent to which our predictions align with the actual passenger counts. By using these performance metrics, we gain valuable insights into the predictive capabilities of the N-HiTS model, shedding light on its proficiency in capturing and forecasting the underlying trends and patterns within the 'Airpassenger' dataset.

In accordance with the default configuration of the model, as we stated in the Analysis section, the author conducted empirical investigations and determined that the optimal combination of pooling and interpolation configurations manifests as MaxPooling in conjunction with a linear interpolation mode.

With an emphasis on computational complexity, we have replicated the experimental methodology using the Airpassenger dataset. The ensuing outcomes are detailed in Table 3:

Interpolation	MaxPool: MAE	MaxPool: MSE	AvgPool MAE	AvgPool: MSE
cubic	**12.68**	**252.47**	14.27	318.59
nearest	13.81	295.46	**13.43**	**286.05**
linear	14.20	270.22	14.64	309.45

Table 3 Comparison of different interpolation mode and Pooling mode

In terms of pooling configurations, MaxPooling consistently delivers superior results due to its ability to effectively preserve sharp edges and subtle details within the dataset. This characteristic enables the neural network to prioritize and concentrate on the most distinctive and relevant information. The Airpassenger dataset, which contains spikes, is particularly suited for MaxPooling, given its proficiency in handling such distinctive characteristics.

When it comes to selecting an interpolation method, our decision to employ cubic interpolation for the Airpassenger dataset instead of linear and nearest interpolation is driven by the dataset's inherent complexities. The Airpassenger dataset contains time-series data that portrays varying passenger counts over time, often exhibiting curves and irregular spikes. These patterns may result from seasonal fluctuations, economic events, or other factors affecting air travel demand.

Cubic interpolation, renowned for its flexibility, is better equipped to capture these intricate and non-linear variations. Its ability to provide a smooth and continuous fit makes it an ideal choice for modeling such dynamic and complex datasets accurately. Unlike linear interpolation, which assumes a straight-line connection between data points, or nearest interpolation, which offers a less refined approximation, cubic interpolation can gracefully navigate the dataset's unique fluctuations and provide a more faithful representation of the underlying dynamics.

In summary, our preference for MaxPooling cubic interpolation in the case of the Airpassenger dataset is a strategic choice, tailored to the dataset's behavior and its need for a method that can effectively handle its complex and nonlinear patterns.

Conclusion

The N-HiTs model represents a significant advancement in long-horizon time series forecasting. The primary objective is to improve the accuracy and computational efficiency of forecasting over a large number of time steps. The N-HiTs builds upon the successful N-BEATS framework and incorporate innovative techniques like multi-rate data sampling and hierarchical interpolation. These techniques collectively enhance the model's ability to make accurate predictions while reducing computational complexity. The findings demonstrate substantial improvements over state-of-the-art baselines, showcasing the effectiveness of the N-HiTS model in providing both accuracy and interpretability in long-horizon forecasting.

We applied the N-HiTS model to the same datasets of the paper, and We conducted comparative experiments with other forecasting models like N-BEATS, Autoformer, and Informer. Overall, our reproduction is close to the results of the paper. In terms of predictive accuracy, the N-HiTS model has the best performance among all the models.

We also experimented N-HiTs with the 'Airpassenger' dataset. The implementation of the Airpassenger dataset involves the selection of optimal configurations for pooling and interpolation. The results obtained through implementation affirm the effectiveness of the chosen configurations, underscoring the importance of thoughtful parameter selection in achieving accurate forecasting results.

The N-HiTS model is a significant advancement in long-horizon time series forecasting, improving accuracy with efficient computation. Our replication experiments confirm its reliability. N-HiTS consistently outperforms other models in predictive accuracy, and the selection of configuration will impact the results. In essence, N-HiTS is a powerful tool for precise and efficient forecasting.

Reference

Challu, C., Olivares, K. G., Oreshkin, B. N., Ramirez, F. G., Canseco, M. M., & Dubrawski, A. (2023, June). Nhits: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 6, pp. 6989-6997).

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv preprint arXiv:1905.10437.

GitHub repository:

https://github.com/Nixtla/neuralforecast

https://github.com/cchallu/n-hits