Probabilistic LTSF: Investigating a DMS-IMS trade-off

Kai Reffert

January 8, 2025

1 Introduction

Point Long-term Time Series Forecasting (LTSF) was transformed by the introduction of Transformer architectures, leveraging global attention and parallel computation. However, recent research [20] questions whether the improvements are due to these innovations or the use of Direct Multi-Step (DMS) forecasting, which optimizes predictions across the entire horizon, unlike Iterated Multi-Step (IMS) methods that suffer from error accumulation. In addition to point forecasting, a substantial body of research has focused on Probabilistic Short-term Time Series Forecasting (STSF). While probabilistic forecasts are considered superior for decision-making [6], progress in this area has not necessarily translated into advancements in probabilistic LTSF. Consequently, Zhang et al. [21] highlight that probabilistic LTSF remains a significant challenge today. Altogether, this study aims to investigate the DMS-IMS trade-off and evaluate probabilistic forecasting in LTSF, addressing key gaps to improve the reliability and applicability of these models.

2 Background

Time series forecasting (TSF) is an important unsolved task, encompassing various domains. Similar to other fields, the research interest shifted from using traditional statistical models towards deep learning based approaches [2, 8, 11]. Previous research on developing deep learning models for time series forecasting has frequently focused on specific aspects, such as point long-term TSF (LTSF) or probabilistic short-term TSF (STSF) [21].

In point LTSF, Convolutional and recurrent neural networks have been implemented successfully, e.g. TCN [1] or DeepAR [15]. While there are some limitations to these approaches, e.g. limited receptive field size and vanishing gradient problem, they are still important in LTSF today [17, 5]. Due to the large influence of the self-attention mechanism, the Transformer architecture resolves some issues of previous methods, as it enables global access to all the past history and is parallelizable. However, the memory and time complexity of basic Transformers grows quadratically $O(L^2)$ with the input length L. Hence, many of the first Transformer-based models for time series focused on creating novel architectures to reduce this bottleneck. Following good performances of these Transformer-based models over previous SOTA LTSF models [12, 24, 23, 13, 18], Zeng et al. [20] criticize that those methods were solely compared to autoregressive or iterated multi-step (IMS) forecasting methods, i.e. where a single-step predictor is sequentially applied to produce multi-step forecasts. IMS methods are known to exhibit error accumulation problems, which is especially prevalent for the LTSF task. Therefore, they hypothesized that the improvement of those novel Transformer-based models can be mainly attributed to their direct multi-step (DMS) forecasting typology, in which the multi-step prediction task is optimized directly. To investigate this, Zeng et al. [20] introduce DLinear, a simple linear DMS model, which was able to outperform the Transformer-based methods on multiple different benchmarks. While many subsequent approaches were able to outperform DLinear [14, 22, 7, 3, 4], all of them employ a DMS strategy without investigating the reasons behind better performances of DMS methods. Hence, raising several questions, why and how DMS and IMS models are related.

Probabilistic STSF focuses on capturing the complex data distribution of future time series, they are regarded superior for decision-making, as they quantify the uncertainty of time series forecasts [6]. It includes multiple different techniques, such as distributional forecasting [15, 3], quantile regression [16, 9] or advanced deep generative models (e.g. GANs [10] or VAEs [19]).

Overall, Zhang et al. [21] state that the combination of the two branches, i.e. probabilistic LTSF, remains a significant challenge.

3 Goals and Work Plan

In the following, I list the main goals as well as a work plan for my master thesis:

G1: Literature study At first, I will look into related works to attain a solid understanding of related work. I specifically want to focus on probabilistic forecasting methods. On top of that, I plan on reading more about the relationship between DMS and IMS methods as well. A rough time estimate for this would be one month, although the general reading phase will probably continuously span over the entire 6 months.

G2: Probabilistic LTSF. I am trying to answer the question, which probabilistic methods are reasonable to adapt SOTA LTSF-models (or SOTA short-term probabilistic TSF models) to the probabilistic LTSF setting. One possible way of implementing this is through distributional forecasting (see the DeepAR model [15]), in which parameters of a specified distribution for each time step are predicted instead of the values itself. This methodology should be straightforward to apply to current SOTA models, since it primarily involves a change in the prediction head as well as a modified loss. Alternatively, quantile forecasts are also an option and are similarly straightforward to implement, see [16]. However, when combined with DMS models, the probabilistic predictions are time step based and there is no (auto-)correlation between them, as it would be natively the case for IMS models. Therefore, some possible directions of obtaining probabilistic time series models include model-based approaches (IMS methods), post-hoc approaches (for example Gaussian Processes on top of DMS models) and hybrid approaches (for example IMS on top of DMS). Since this goal is the main technical contribution, I would assume this takes about two to three months. In total, the completion of this stage would be the first milestone.

G3: Experimental study After obtaining the appropriate methods in G2, I plan to perform a multitude of experiments to analyze their behavior. First, I am interested in evaluating the differences of probabilistic time step models versus probabilistic time series models. This could involve analyzing how the methods express and learn the correlation between different time steps via curated synthetic data sets. Subsequently, this also requires a measurement/metric to evaluate correlation effects in time series. Additionally, as mentioned above, the shift from IMS to DMS methods (and anything in between) remains largely underexplored, especially in a probabilistic forecasting setting. Hence, I want to try to answer the question of how and why error accumulation problems arise. Potentially, applying methods to circumvent the problem (nice-to-have). Generally, a nice-to-have would be synthetically generated datasets that illustrate points of the analysis. Overall, I would assume this stage requires most of the time ($\sim$ 3 months).

References

[1] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, April 2018. arXiv:1803.01271 [cs].

[2] Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, François-Xavier Aubet, Laurent Callot, and Tim Januschowski. Deep Learning for Time Series Forecasting: Tutorial and Literature Survey. *ACM Computing Surveys*, 55(6):121:1–121:36, December 2022.

[3] Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan O. Arik, and Tomas Pfister. TSMixer: An All-MLP Architecture for Time Series Forecasting, September 2023. arXiv:2303.06053.

[4] Luo Donghao and Wang Xue. ModernTCN: A Modern Pure Convolution Structure for General Time Series Analysis. October 2023.

[5] Xiaojing Fan, Chunliang Tao, and Jianyu Zhao. Advanced Stock Price Prediction with xLSTM-Based Models: Improving Long-Term Forecasting, August 2024.

[6] Tilmann Gneiting and Matthias Katzfuss. Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151, January 2014.

[7] Zeying Gong, Yujin Tang, and Junwei Liang. PatchMixer: A Patch-Mixing Architecture for Long-Term Time Series Forecasting, October 2024. arXiv:2310.00655.

[8] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. *International Journal of Forecasting*, 37(1):388–427, January 2021. arXiv:1909.00590.

[9] Shayan Jawed and Lars Schmidt-Thieme. GQFormer: A Multi-Quantile Generative Transformer for Time Series Forecasting. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 992–1001, December 2022.

[10] Alireza Koochali, Andreas Dengel, and Sheraz Ahmed. If You Like It, GAN It—Probabilistic Multivariate Times Series Forecast with GAN. *Engineering Proceedings*, 5(1):40, 2021. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[11] Pedro Lara-Benítez, Manuel Carranza-García, and José C. Riquelme. An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *International Journal of Neural Systems*, 31(03):2130001, March 2021.

[12] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[13] Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X. Liu, and Schahram Dustdar. Pyraformer: Low-Complexity Pyramidal Attention for Long-Range Time Series Modeling and Forecasting. October 2021.

[14] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers, March 2023. arXiv:2211.14730.

[15] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, July 2020.

[16] Ruofeng Wen, Kari Torkkola, Balakrishnan Narayanaswamy, and Dhruv Madeka. A multi-horizon quantile recurrent forecaster, 2018.

[17] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. September 2022.

[18] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting, January 2022. arXiv:2106.13008 [cs].

[19] Ye Yuan and Kris Kitani. Diverse Trajectory Forecasting with Determinantal Point Processes, December 2019. arXiv:1907.04967 [cs].

[20] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are Transformers Effective for Time Series Forecasting?, August 2022. arXiv:2205.13504.

[21] Jiawen Zhang, Xumeng Wen, Zhenwei Zhang, Shun Zheng, Jia Li, and Jiang Bian. ProbTS: Benchmarking Point and Distributional Forecasting across Diverse Prediction Horizons, October 2024. arXiv:2310.07446 [cs].

[22] Yitian Zhang, Liheng Ma, Soumyasundar Pal, Yingxue Zhang, and Mark Coates. Multi-resolution Time-Series Transformer for Long-term Forecasting. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, pages 4222–4230. PMLR, April 2024. ISSN: 2640-3498.

[23] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, March 2021. arXiv:2012.07436.

[24] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting, June 2022. arXiv:2201.12740.