This is the author's accepted version of the article. The final version published by IEEE is Pablo Fernández Pérez, Claudio Fiandrino, Marco Fiore and Joerg Widmer, "Dissecting Advanced Time Series Forecasting Models with AIChronoLens," *IEEE INFOCOM Posters 2024 - IEEE Conference on Computer Communications Posters*, Vancouver, Canada, 2024, pp. TBD, doi: TBD.

Dissecting Advanced Time Series Forecasting Models with AIChronoLens

Pablo Fernández Pérez, Claudio Fiandrino, Marco Fiore and Joerg Widmer
IMDEA Networks Institute, Madrid, Spain
Email: {pablo.fernandez, claudio.fiandrino, marco.fiore, joerg.widmer}@imdea.org

***Abstract*—Mobile traffic forecasting is instrumental in efficiently managing network resources. In this poster paper, we dissect the behavior of advanced time series forecasting techniques, namely DLinear and PatchTST, when applied to the problems of predicting future mobile traffic volumes. Being black-box models hard to interpret, we ground our analysis on EXplainable Artificial Intelligence (XAI) by using AIChronoLens, a new tool that links legacy XAI explanations with the temporal properties of the input sequences. We find that the DLinear significantly improves the prediction accuracy over PatchTST and state-of-the-art techniques like Long-Short Term Memory (LSTM). The analysis with AIChronoLens shows that, unlike PatchTST, DLinear is capable of focusing its prediction decisions on a few key samples of the input sequences, which makes it possible for DLinear to match the ground truth closely.**

I. Introduction

Recently, time series forecasting has received significant attention from the Artificial Intelligence (AI) community. In this poster paper, we focus on two very recent methods, namely DLinear [1] and PatchTST [2], which were shown to improve significantly over well-known techniques like LSTM or AutoRegressive Integrated Moving Average (ARIMA) that are widely adopted for mobile traffic forecasting [3]. Specifically, transformer-based models like PatchTST learn better long-range dependencies than LSTM that are limited to capture dependencies in fixed windows of time. PatchTST uses attention mechanisms that provide a competitive advantage over LSTM or ARIMA when the data is noisy or where there are multiple patterns in the data. Unfortunately, both PatchTST and DLinear are not self-explainable like Decision Trees (DTs), which may lower their chances of being actually deployed in production networks. XAI techniques mitigate this problem by shedding light on models' operation.

In this context, we make the following contributions: *(i)* we are the first to apply DLinear and PatchTST for the problem of mobile traffic forecasting and *(ii)* we dissect their behavior with XAI and benchmark their performance. For this purpose, we use AIChronoLens [3], a tool that links legacy XAI explanations with the temporal properties of the input sequences. Indeed, the explanations provided by the legacy XAI techniques are only relevance scores of the input that do not have a unique relationship with the temporal characteristics of the input sequences. The lack of such a relationship suggests that the legacy XAI techniques are either not effective in capturing the salient characteristics of the model or that the model itself is not adequate for the job. AIChronoLens addresses precisely this shortcoming, thus making it possible to perform a one-to-one comparison of different AI models when applied to the very same dataset.

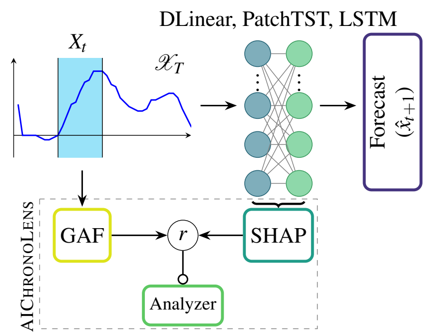


Fig. 1. The AIChronoLens architecture

II. Methodology

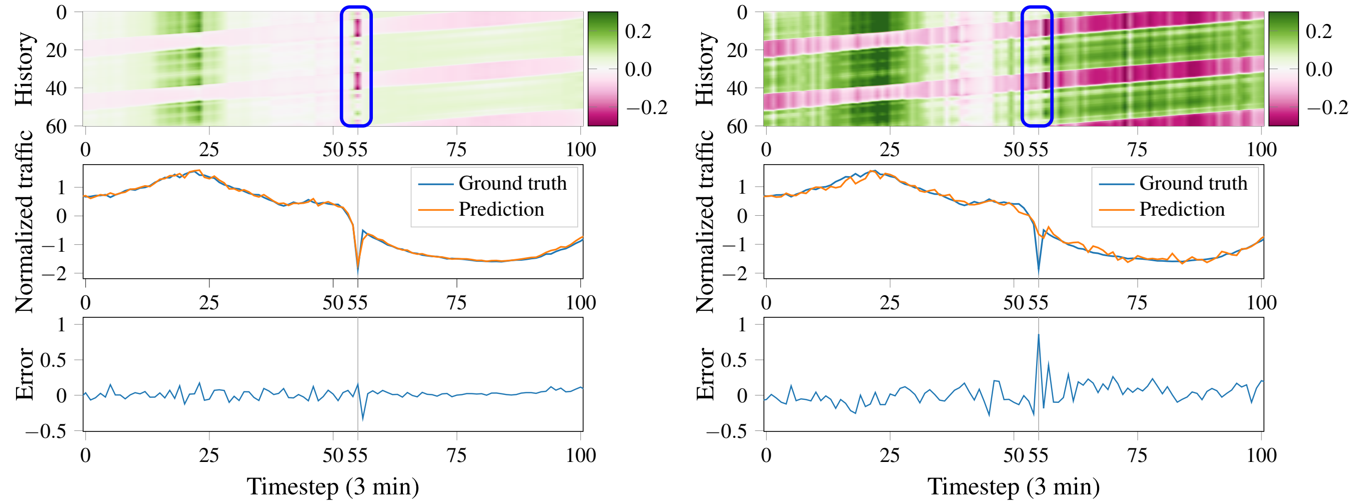
Problem. Given a sequence of values $\mathcal{X}_T = \{x_1, x_2, \dots, x_T\}$ at time $t = \{1, 2, \dots, T\}$, the problem of time series forecasting is to predict the future value having observed $X_t = \{x_{t-n+1}, x_{t-n+2}, \dots, x_t\}$, a set of historical n past values at time t. AI models compute the forecast $\hat{x}_{t+1}$ at time $t+1$ as $\hat{x}_{t+1} = F(X_t)$, where F is a generic prediction function, trained by evaluating at each iteration a loss function to fulfill a specific objective, *e.g.*, minimizing the Mean Absolute Error (MAE) or Mean Square Error (MSE).

Dataset. We use a dataset that contains measurements of traffic volumes recorded in a production 4G network that serves a large metropolitan region in Europe. The dataset provides fine-grained information at 3 minute granularity about the traffic volumes at each Base Station (BS) for 3 months. We use 28 541 samples for training and 7 121 for testing (80:20 split ratio).

AI Forecasting Models. We focus on prominent AI models for time series forecasting:

- DLinear [1] is an incredibly simple model that consists of a single linear layer that is applied to a decomposed input sequence in the form of trend and seasonal components.
- PatchTST [2] is a transformer-based model that builds on two methods: patching and channel independence. Patching aggregates sub-sequences from the input sequence to better extract local and long-range dependencies. In the case of multivariate time series, PatchTST processes independently each variate to learn their unique patterns.
- LSTM is the state-of-the-art model for forecasting [3]. The architecture we used consists of 50 cells and a linear layer.

The three models are trained for 50 epochs using the ADAM optimizer, a batch size of 15 samples, and MSE as the



(a) DLinear

(b) PatchTST

Fig. 2. Analysis with AICHRONOLENS of the advanced forecasting models when applied to the problem of mobile traffic forecasting

TABLE I
ACCURACY OF THE ANALYZED FORECASTING MODELS

METRIC	MODELS		
	DLinear	PatchTST	LSTM
MAE	0.0454	0.1323	0.0812
MSE	0.0046	0.0341	0.0231

loss function. We set the learning rate to 0.001 and use the `ReduceLROnPlateau` scheduler to reduce the learning rate by an order of magnitude every time the train loss stopped improving. The lookback window is 300 time steps. We set the horizon to 60 timesteps and retain as a prediction the first item, *i.e.*, the future value at $t+1$.

AICHRONOLENS. The tool, shown in Fig. 1, uses legacy XAI techniques to define the contribution of each element of the input sequence X_t to the forecast $\hat{x}_{t+1}$. Here, we use SHapely Additive exPlanations (SHAP) [4] as legacy XAI method. Then, it probes for linear correlation (with the Pearson coefficient, r) between SHAP and the Gramian Angular Field (GAF) computed on X_t to assess whether the temporal characteristics of the input sequence match with what the model deems relevant for the prediction. The “Analyzer” module monitors over time such relationship to synthesize explanations.

III. RESULTS AND INSIGHTS

Table I summarizes the prediction accuracy of the analyzed models. We observe that DLinear performs comparatively better than PatchTST and LSTM for both MAE and MSE. Fig. 2 portrays a comparative analysis of the operation of the two models under analysis, *i.e.*,, DLinear (in Fig. 2(a)) and PatchTST (in Fig. 2(b)). The plots on the top are the correlation coefficients generated by AICHRONOLENS, in the center we show the ground truth and prediction (scaled with a standard scaler), and in the bottom the error. We observe that:

• O_1: The correlations between input traffic history and the output forecast are stronger in PatchTST, with values closer to -1 or 1, than for DLinear. The reason lies in the way the two models exploit the input sequences for the predictions: while in PatchTST the SHAP values indicate that the model focuses on the entire sequence because there are no scores considerably higher than others, DLinear focuses on few samples whose relevance scores are much higher than those of the others.

• O_2: The predictions of DLinear are less noisy than PatchTST and, unlike PatchTST, the model is able to capture with no delay the load drop in timestep $t = 55$. As highlighted in O_1, DLinear focuses on a few key samples to make its predictions. At $t = 55$, the model reacts to the sudden change by assigning very high relevance to a few more samples than usual and this is captured in the change of correlation coefficients (see the blue highlighted area in both plots).

ACKNOWLEDGMENT

This work is partially supported by bRAIN project PID2021-128250NB-I00 funded by MCIN/ AEI /10.13039/501100011033/ and the European Union ERDF “A way of making Europe”; by Spanish Ministry of Economic Affairs and Digital Transformation, European Union NextGeneration-EU/PRTR projects MAP-6G TSI-063000-2021-63, RISC-6G TSI-063000-2021-59 and AEON-ZERO TSI-063000-2021-52; C. Fiandrino is a Ramón y Cajal awardee (RYC2022-036375-I), funded by MCIU/AEI/10.13039/501100011033 and the ESF+, and P. Fernandéz received funding from “Programa Investigo” grant 2022-C23.I01.P03.S0020-0000038, funded by the European Union NextGeneration-EU/PRTR and MITES/SEPE.

REFERENCES

[1] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are transformers effective for time series forecasting?” in *Proc. of the AAAI Conference on AI*, 2023.

[2] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, “A time series is worth 64 words: Long-term forecasting with transformers,” in *Proc. of ICLR*, 2023.

[3] C. Fiandrino, E. Pérez Gómez, P. Fernández Pérez, H. Mohammadalizadeh *et al.*, “AIChronoLens: advancing explainability for time series AI forecasting in mobile networks,” in *Proc. of IEEE INFOCOM*, 2024, Available online: https://git2.networks.imdea.org/wng/aichronolens.

[4] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proc. of NIPS*, 2017, pp. 4768–4777.