Transforming Time-Series Data for Improved LLM-based Forecasting through Adaptive Encoding

Vladimir Čeperić, PhD, Vladimir.Ceperic@Fer.Hr
Tomislav Marković, PhD

University of Zagreb, Croatia

Cambridge, 2024

Content

Introduction

- A large language model (LLM) is a language model notable for its ability to achieve general-purpose language generation and other natural language processing tasks such as classification.
- LLMs have revolutionized natural language processing, showcasing their potential to extend into areas beyond text, including time-series forecasting.
- Adapting LLMs to handle quantitative time-series data poses challenges, primarily due to their design for textual data processing, highlighting the need for innovative approaches to bridge this gap.

Why LLMs?


Time to Reach 100M Users
Months to get to 100 million global Monthly Active Users
78
70
61
55
41
30
9
2
Google Translate
Uber
Telegram
Spotify
Pinterest
Instagram
TikTok
ChatGPT
Source: UBS / Yahoo Finance
@EconomyApp
APP ECONOMY INSIGHTS

Challenges with LLMs and Time-Series Data

- **Inherent Challenges**
  - Text-based Large Language Models (LLMs) are primarily designed for natural language, making the forecasting of quantitative time-series data challenging.
  - Capturing Temporal Patterns: Time-series data requires understanding and leveraging patterns over time, a task for which LLMs are not inherently designed.
  - High Dimensionality and Variability: The complex nature of time-series data, including its variability and dimensionality, poses significant challenges for traditional LLM processing.
- This research aims to serve as a bridge, leveraging the textual analysis capabilities of LLMs to effectively interpret and predict the quantitative nature of time-series data, addressing these inherent challenges.

Similar, but not the same

- Foundation models
  - Garza&Mergenthaler-Canseco, October 2023, TimeGPT-1
    - zero-shot inference
    - was trained on the largest collection of publicly available time series, collectively encompassing over 100 billion data points.
  - Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting, February 2024
- Prompting/tokenization - type of modelsmodels
  - zero-zero shot inference?
  - Time-LLM [ICLR, May 2024], LLMTime, PromptCast

One LM to rule them all

- Multi-modal LLMs
  - text, images, video, sound etc.
- LLMs converting to LM
- Key find:
  - Improvement in one domain brings improvement in all domains
  - Data quantity matters!
- Examples: CLIP computer vision models got SOTA results because of the addition of language
  - CLIP (Contrastive Language-Image Pre-Training) is a neural network trained on a variety of (image, text) pairs.
    - Improved models DINOv2, SAM-CLIP etc.

The problem no one talks about

- LLMs are trained on huge quantities of data
- That that probably includes the most common time-series datasets
  - Time-LLM adding descriptions make things worse
- Testing will have to be on unseen data :)
  - Exceptionally hard to validate!

Objectives of the Study

- **Develop an Encoding Strategy:** Create a method to translate time-series data into a format interpretable by LLMs, maintaining its quantitative integrity.
- **Empirical Testing:** Evaluate the effectiveness of the encoding strategy across various datasets to ensure reliability and versatility.

Background and Related Work

- Studies like LLMTime [7] (NeurIPS 2023) and PromptCast [8] (IEEE Transactions on Knowledge and Data Engineering, 2023) underscore LLMs' versatility, extending their capabilities to the realm of time-series forecasting.
- Despite their potential, adapting LLMs for time-series forecasting presents unique challenges due to the specialized nature of temporal data.

Challenges in Time-Series Forecasting with LLMs

- The quantitative nature of time-series data poses a significant hurdle for text-centric LLMs, which excel in qualitative analysis.
- Accurately capturing temporal patterns and dependencies is crucial, yet it is a complex task for models trained on static text.
- LLMs are often challenged by the high dimensionality and variability inherent in time-series data, impacting their forecasting ability.

Proposed Encoding Strategy

- Our strategy utilizes adaptive segmentation and tokenization, tailoring LLM analysis to the dynamic nature of time-series data.
- Incorporates the fast Brownian bridge-based aggregation (fABBA) algorithm, enhancing the efficiency of data transformation for LLMs.
- Carefully prepares time-series data, ensuring crucial temporal patterns are maintained for accurate LLM interpretation and forecasting.

Algorithm 1 Time-Series Encoding Process

Input: Time series $X = \{x_1, x_2, \ldots, x_n\}$
Output: Tokenized representation $T = \{t_1, t_2, \ldots, t_m\}$
Use fABBA to partition X into segments $\{s_1, s_2, \ldots, s_k\}$
for each segment s_i **do**
  Standardize and discretize s_i into tokens $\{t_{i1}, t_{i2}, \ldots, t_{il}\}$
  Append tokens to T
end for
Format T for LLM input

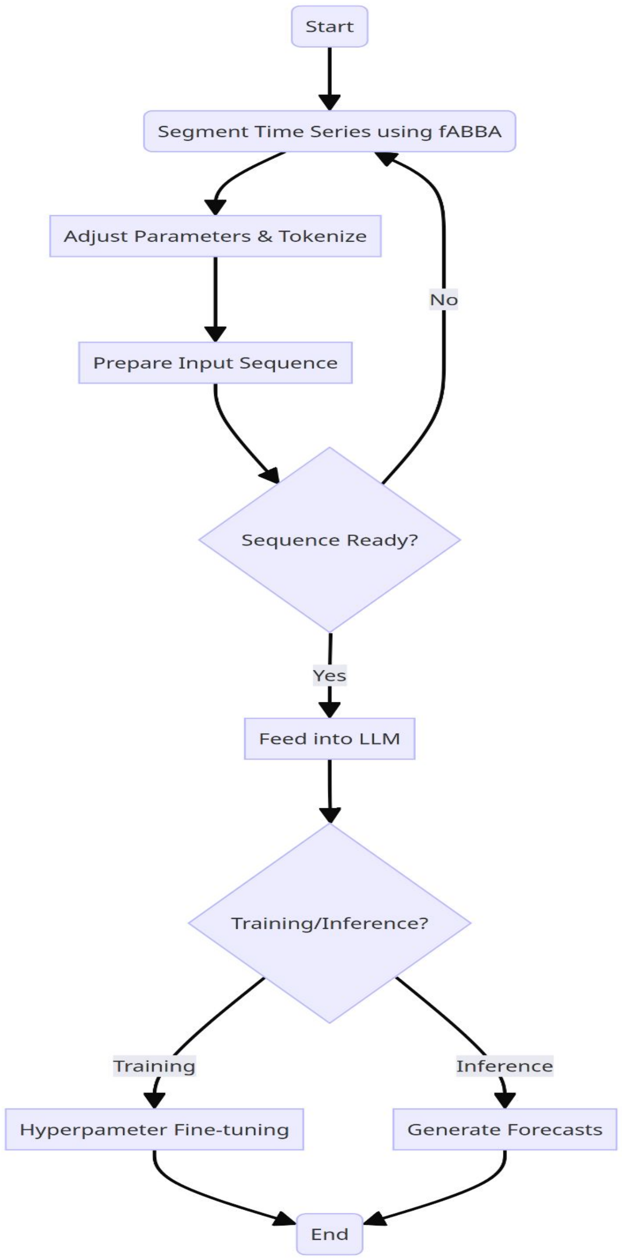
Example:

ts = [np.sin(0.05*i) for i in range(1000)]

a B b C b C b C b C b C b C b C A

Adaptive Segmentation and Tokenization for LLM Processing

- Process ensures that the time-series data is appropriately scaled and tokenized, facilitating its analysis by LLMs.
- Bridging the gap between the quantitative nature of time-series data and the textual processing strengths of LLMs, paving the way for improved forecasting models.



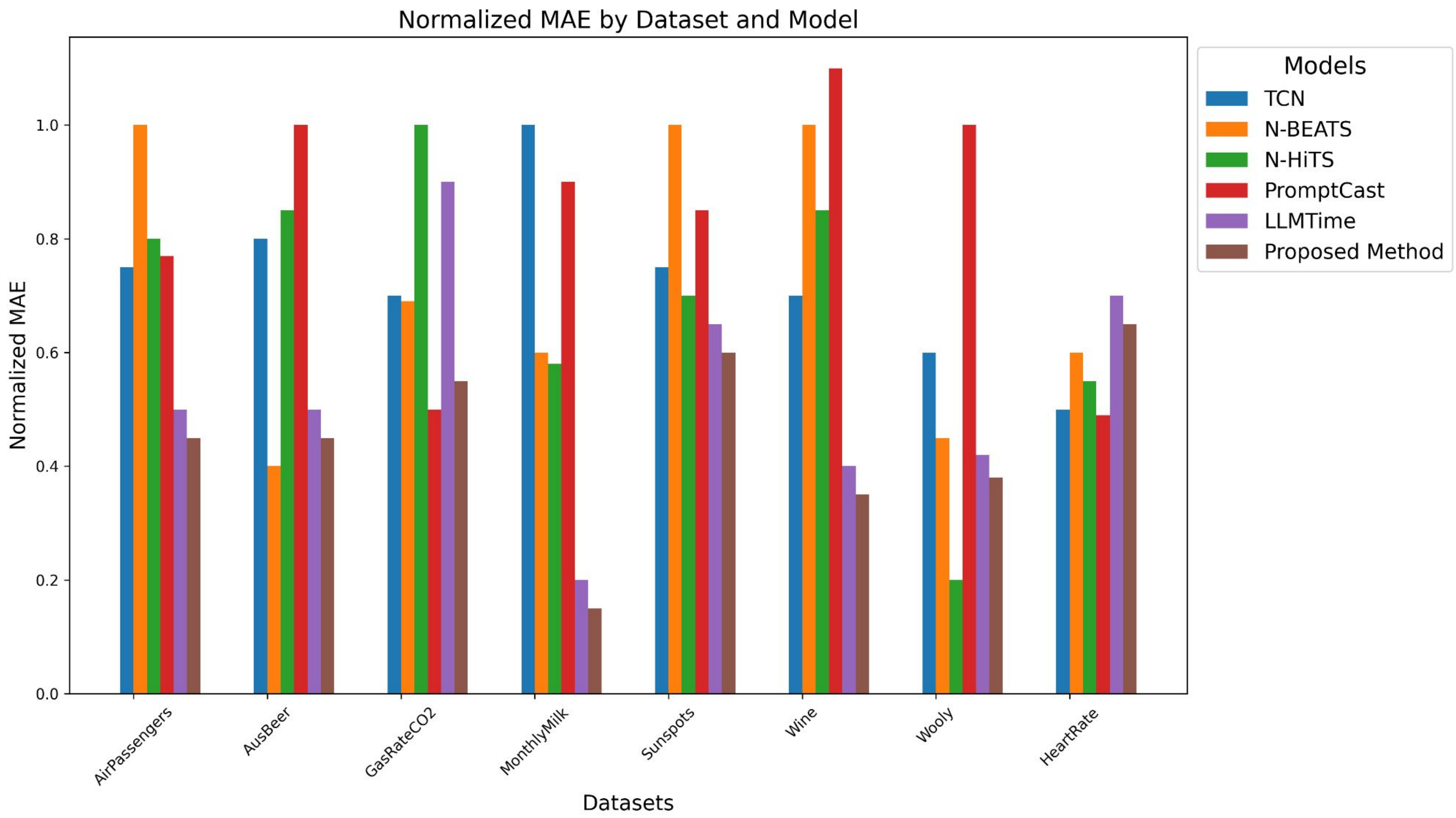
Benefits of the Proposed Method

- Augments forecasting accuracy with improved data representation that captures essential time-series dynamics.
- The method's adaptive nature ensures it effectively addresses the diverse characteristics and scales of time-series data.
- Optimizes LLM processing time by reducing data size, enabling a more efficient analysis without sacrificing detail.

Experimental Design

- Employed DARTS datasets, selected for their variety, to present a broad spectrum of forecasting challenges.
  - Evaluated across various sectors including aviation (Air Passengers), beverage consumption (Aus Beer), environmental data (Gas Rate CO2), among others.
  - Datasets present common forecasting complexities such as seasonal patterns, unpredictable trends, and sector-specific variability
- Strategically split data into 80% for model training and 20% for validation to ensure the integrity of the forecasting results.
- Compared the proposed method's performance with established models such as TCN, N-BEATS, and N-HiTS, to benchmark its efficacy.
- Employed standard metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to quantify forecasting accuracy.

Performance Comparison - Benchmarking Our Method



Performance Comparison - Benchmarking Our Method

TABLE I
FORECASTING PERFORMANCE ACROSS DATASETS

Dataset	Model	MAE	RMSE	MAPE
Air Passengers Dataset	TCN	54.960	65.535	11.667
	N-BEATS	97.886	118.072	20.756
	N-HiTS	59.160	79.032	12.546
	PromptCast	56.448	67.988	13.183
	LLMTime	34.373	41.210	7.738
	Proposed Method	**31.536**	**37.089**	**6.964**
Aus Beer Dataset	TCN	30.897	35.883	7.309
	N-BEATS	10.394	14.074	2.395
	N-HiTS	34.229	40.878	7.811
	PromptCast	62.221	75.611	14.602
	LLMTime	16.127	18.928	3.796
	Proposed Method	**14.515**	**17.035**	**3.416**
Gas Rate CO2 Dataset	TCN	2.641	2.981	4.837
	N-BEATS	2.628	3.085	4.806
	N-HiTS	3.854	4.503	7.227
	PromptCast	**2.093**	**2.416**	**3.842**
	LLMTime	3.496	4.214	6.235
	Proposed Method	2.198	2.495	3.900
Monthly Milk Dataset	TCN	70.859	88.781	7.893
	N-BEATS	33.641	40.258	4.064
	N-HiTS	32.726	39.349	3.849
	PromptCast	81.103	90.011	9.344
	LLMTime	9.677	11.938	1.112
	Proposed Method	**8.709**	**10.744**	**1.001**
Sunspots Dataset	TCN	51.816	70.214	264.954
	N-BEATS	73.151	91.815	96.035
	N-HiTS	49.933	68.827	196.231
	PromptCast	61.729	80.228	227.035
	LLMTime	47.339	66.677	136.676
	Proposed Method	**42.606**	**60.009**	**123.212**
Wine Dataset	TCN	3287.137	4559.931	13.980
	N-BEATS	4562.018	6059.011	16.366
	N-HiTS	3909.508	5498.956	14.924
	PromptCast	6789.694	7772.813	29.771
	LLMTime	1569.324	2055.041	6.558
	Proposed Method	**1413.392**	**1850.337**	**5.900**
Wooly Dataset	TCN	1158.795	1279.322	25.356
	N-BEATS	903.013	1034.849	19.535
	N-HiTS	**382.088**	**453.168**	**7.771**
	PromptCast	1949.667	2072.459	42.117
	LLMTime	808.731	877.411	17.322
	Proposed Method	728.258	789.270	15.600
Heart Rate Dataset	**TCN**	**5.493**	**6.599**	**5.953**
	N-BEATS	6.566	7.697	6.934
	N-HiTS	6.098	7.774	6.914
	PromptCast	5.511	7.166	6.143
	LLMTime	6.211	8.012	7.009
	Proposed Method	5.500	6.892	6.800

Limitations and Directions for Future Research

- **Scope of Evaluation:**
  - broader testing range is necessary
  - datasets beyond DARTS for a more thorough evaluation
- **Handling Volatility:**
  - further exploration is needed
- **Ongoing Refinement:**
  - ongoing improvement of the encoding techniques

Conclusion

- Presented an innovative encoding strategy that enhances time-series forecasting with LLMs.
- Confirmed improved forecasting accuracy on (some) diverse datasets.
- Additional benefit can be reduced prompt size (due to inherit fABBA compression) and hence reduced overall costs of running LLM inferece.
- Our approach to tokenizing time-series data significantly enhanced LLM comprehension, contributing to more accurate forecasts.
- Our method paves the way for expanded use of LLMs in time-series analysis, suggesting a promising future for LLM-driven forecasting, with various time-series encoding techniques.

Vladimir.Ceperic@fer.hr

Acknowledgment

- Special thanks to FWO for funding Project G088822N.