

Research Article

LLM-ABBA: Understanding Time Series via Symbolic Approximation

Erin Carson¹, Xinye Chen², Cheng Kang³

1. Department of Numerical Mathematics, Charles University, Prague, Czech Republic; 2. CNRS, LIP6, Sorbonne Université, France; 3. Department of Cybernetics, Czech Technical University in Prague, Prague, Czechia

The success of large language models (LLMs) for time series has been demonstrated in previous work. Utilizing a symbolic time series representation, one can efficiently bridge the gap between LLMs and time series. However, the remaining challenge is to exploit the semantic information hidden in time series by using symbols or existing tokens of LLMs, while aligning the embedding space of LLMs according to the hidden information of time series. The symbolic time series approximation (STSA) method called adaptive Brownian bridge-based symbolic aggregation (ABBA) shows outstanding efficacy in preserving salient time series features by modeling time series patterns in terms of amplitude and period while using existing tokens of LLMs.

In this paper, we introduce a method, called LLM-ABBA, that integrates ABBA into large language models for various downstream time series tasks. By symbolizing time series, LLM-ABBA compares favorably to the recent state-of-the-art (SOTA) in UCR and three medical time series classification tasks. Meanwhile, a fixed-polygonal chain trick in ABBA is introduced to avoid obvious drifting during prediction tasks by significantly mitigating the effects of cumulative error arising from misused symbols during the transition from symbols to numerical values. In time series regression tasks, LLM-ABBA achieves the new SOTA on Time Series Extrinsic Regression (TSER) benchmarks. LLM-ABBA also shows competitive prediction capability compared to recent SOTA time series prediction results. We believe this framework can also seamlessly extend to other time series tasks.

I. Introduction

Time series are fundamental mathematical objects with applications across diverse disciplines such as classification^[1], regression^[2], and prediction^[3]. Recently, the power of large language models (LLMs) in time series applications has been recognized. One recent review^[4] concludes that there are three main LLM-based approaches to learn intricate semantic and knowledge representations from time series to perform various tasks. The first approach is to patch and tokenize numerical signals and related text data, followed by fine-tuning on time series tasks^{[5][6][7]}; the second one is preprocessing time series data to fit LLM input spaces by adding a customized *tokenizer*^[8]; the last one is to build foundation models from scratch, and this approach aims to create large, scalable models, both generic and domain-specific^{[9][10]}.

These three techniques each come with their own limitations. Patching and tokenizing time series segments can build the mapping between time series and the latent embedding of LLMs, instead of discrete language tokens. When outputting the numerical value, this method should generate the digit one by one, which eventually reduces the generation speed^[6]. Furthermore, by adding a customized tokenizer, LLMs can handle positions of time series patterns and reproduce the internal logic of given time series signals^[11]. Because LLM tokenizers, not designed for numerical values, separate continuous values and ignore the temporal relationship of time series, this method should convert tokens into flexible continuous values^[12]. It inevitably requires token transitions from time series feature space to the latent embedding space of LLMs and cannot avoid the risk of semantic loss. Building foundational time series models from scratch can essentially solve these problems. But considering that one should balance the high development costs and their applicability, the challenge of expensive training persists should be tackled^[4].

By aligning time series and native language, large language models and specialized time series models constitute a new paradigm, where the LLMs are prompted with both time series and text-based instructions^[4]. Following this paradigm, time series and textual information provide essential contexts, LLMs contribute to internal knowledge and reasoning capabilities, and time series models offer fundamental pattern recognition assurances. This novel integration is depicted in Figure 1, where a successful combination of these components showcases the potential for a general-purpose, unified system in next-generation time series analysis. Therefore, the challenge is to develop one tool that can transform the

internal patterns of time series to the contents that LLMs can recognize (Step 1 of Fig. 1). Moreover, this tool should also transform the generated contents back to the time series domain so as to aid the time series analysis (Step 2 of Fig. 1).

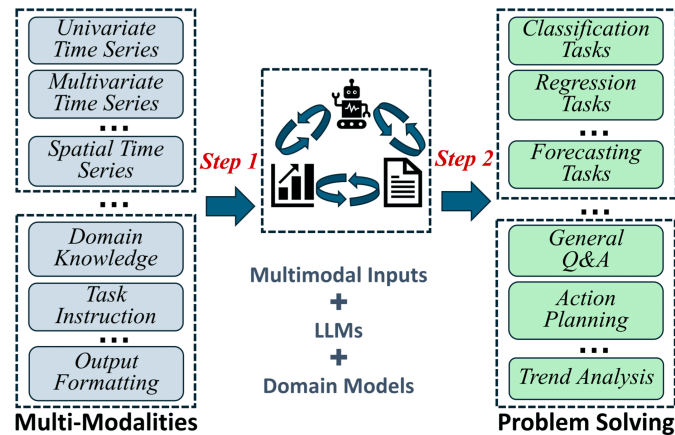


Figure 1. The integration of time series and LLM demonstrates potential in solving complex real-world problems.

Symbolic time series approximation (STSA) is a method that converts time series into symbols. It establishes a bridge between strings and numerical time series, which enables the chain-of-pattern (COP) of strings to be as informative as possible compared to raw data. Utilizing the symbolic representation of time series, one can model time series as native languages by encoding time series as a sequence of strings and performing efficient text analysis techniques upon it rather than manipulating raw numerical values, e.g., converting time series forecasting to next-token prediction in text. STSA could both implicitly and explicitly align the time series features with symbols, which enables the manipulation of natural language processing and learning on time series. If possible, there is no necessity to (1) patch and tokenize time series segments, (2) add an extra customized tokenizer set, or (3) build foundational time series models from scratch. Symbolic representations obtained from transformed numerical time series can potentially reveal the linguistic logic hidden inside time series signals, and this technology roadmap is able to provide LLMs with the ability to understand temporal patterns. Therefore, the time series semantic information can be well exploited in LLMs. Inspired by this idea, it is desirable to obtain a method that can efficiently transform numerical time series into symbols, and fine-tune LLMs on time series analysis tasks (e.g., classification, regression, and prediction).

However, the technique to integrate STSA methods with LLMs is lacking. Applying LLMs on symbolic time series representations is tricky. First, we need to address the symbolic consistency issues that exist in STSA methods, as the information of the same symbols across different time series under the same symbolization scheme should be identical. It is also unclear whether LLMs will learn consistent knowledge from the transformed symbols that contain the time series pattern logic. Second, LLMs can generate text contents from given information, but could they also generate symbolic series and reconstruct the time series pattern logic via STSA methods? These considerations bring us to ABBA^[13] (incl. its accelerated variant fABBA^[14]), the most recent STSA method, which shows a competitive advantage in the shape capturing of time series over existing STSA methods. Compared to other STSA methods, ABBA enables users to specify customized strings for symbolization and provides open-source software with easy-to-use APIs¹. Each ABBA symbol is associated with a unique real-valued cluster center, which enables a natural word embedding for symbols as a native language. A straightforward way to see how much information the STSA methods can capture is via the visualization of their symbolic reconstruction. A comparison of reconstruction using Symbolic Aggregate appRoXimation (SAX)^[15] and fABBA^[14] is as illustrated in Figure 2. It is clear that SAX fails to capture the trend of time series in both figures (also noted in^[16]), and the peak information in figure (b) is missing in the SAX reconstruction. Figure 2 also shows that fABBA is better at capturing the essential information of time series patterns compared to SAX.

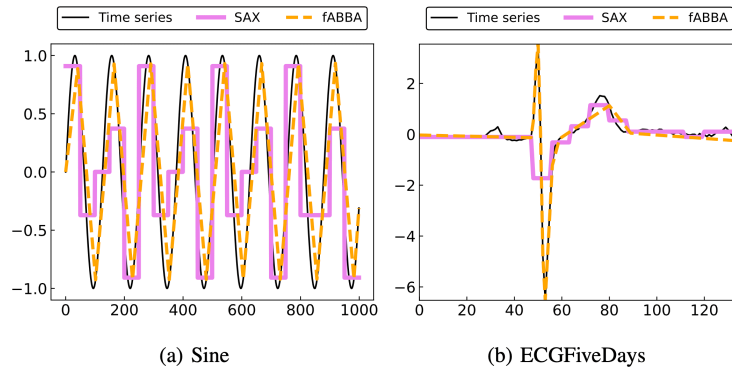


Figure 2. Plot (a) shows a sine function with 1,000 points, and (b) shows the ECGFiveDays time series from the UCR Archive. We first perform fABBA with $\text{tol} = 0.1$ and $\alpha = 0.1$ and perform SAX with approximately the same length of symbolic representation and the number of distinct symbols. In plot (a), fABBA generates symbols “aBbCbCbCbCbCbCA” while SAX generates symbols “aACBbaACBbaACBbaAABb”; in figure (b), fABBA generates symbols “fAcaDECEdbbF” while SAX generates symbols “AAAAAABbCcDaaaAaa”.

In this paper, we propose LLM-ABBA, which can help LLMs to understand time series by using an ABBA method and transforming numerical time series signals into symbolic series. Concretely, LLM-ABBA first transforms time series signals to compressed representations by adaptively compressing numerical inputs. Next, it digitizes the compressed representation with given symbols or pretrained tokens. Then, LLM-ABBA gives LLMs a series of symbols (or pretrained tokens) that LLMs can recognize from the beginning, and these symbols (or pretrained tokens) essentially contain the COP of time series signals. Classification tasks only need to identify symbolic series, but for forecasting or regression tasks, an additional step is taken to predict the future time series values. By using the QLoRA fine-tuning method (a recent, frequently-used adaption fine-tuning method)^[47], LLM-ABBA exhibits a trade-off between task performance and efficiency, which also ensures the adaptability on various domains. Therefore, the LLM is capable of incorporating the COP of time series and diving into the analysis of time series on a macroscopic view along with the domain knowledge from prompting instructive commands.

Our contributions include:

1. We propose a unified and improved ABBA approach for efficiently symbolizing multiple time series and mitigating the accumulated shift in time series reconstruction, enabling an effective inference task over out-of-sample data.
2. For time series regression tasks, LLM-ABBA achieves SOTA performance, and it also achieves comparable performance on medical time series classification tasks. To the best of our knowledge, this is the first work to practically combine LLMs with STSA. We believe our work can be easily extended to other STSA methods.
3. LLM-ABBA can retain language semantics and learn the COPs of time series by adapter fine-tuning methods in time series forecasting tasks.
4. The universality and convenience of LLMs’ multi-modality on time series tasks obtains a valuable improvement.

The rest of the paper is structured as follows. Section II discusses related work in applications of LLMs to time series. Section III lays the foundation of the ABBA method and proposes our LLM-ABBA framework. Section IV presents the simulations of our method as well as the comparisons between our method and SOTA methods. Section V discusses the limitations of our method and future work. Section VI concludes the paper.

II. Related work

LLMs for time series methods have seen significant advances in recent years. The work^[18] argues that this success stems from the ability of LLMs to naturally represent multimodal distributions of time series. By framing a time series forecasting task as a sentence-to-sentence task, AutoTimes^[18] minimizes the tunable parameters needed to generate time series embeddings while freezing the parameters of the LLM, and FPT^[19] fine-tunes LLM parameters to serve as a general representation extractor for various time series analysis tasks. These approaches

maximize the use of inherent token transitions, leading to improved model efficiency. In terms of multivariate time series forecasting, UniTime^[20] trains and fine-tunes a language model to provide a unified forecasting framework across multiple time series domains. Leveraging advanced prompting designs and techniques, PromptCast^[21] transforms time series data into text pairs, and TEMPO^[22] models specific time series patterns, such as trends and seasonality, by using weighted scatterplot smoothing^[23].

Tuning-based predictors use accessible LLM parameters, typically involving pre-processing and tokenizing numerical signals and related prompt text, followed by fine-tuning on time series tasks^[4]. In summary, there are four steps needed to adapt LLM to time series:

1. \mathcal{T}_{inp} = Pre-processing(\mathcal{T}): With a Patching operation^{[5][18]} or a weighted scatterplot smoothing processing^[22], the time series set \mathcal{T} is pre-processed to specific knowledge-contained inputs \mathcal{T}_{inp} ;
2. \mathcal{M}_{inp} = Tokenizer(Prompt, \mathcal{T}_{inp}): An additional option is to perform a Tokenizer operation on time series \mathcal{T}_{inp} and related prompt text to form text sequence tokens \mathcal{M}_{inp} ;
3. $\mathcal{M}_{outp} = f_{LLM}^{\Delta}(\mathcal{M}_{inp})$: With the instruction prompt Prompt, time series processed tokens and optional text tokens are fed into $f_{LLM}^{\Delta}(\cdot)$ with partial unfreezing or additional adapter layers. \mathcal{M}_{outp} can be either a fine-tuned result or an intermediate result;
4. $\hat{Y} = \text{Task}(\mathcal{M}_{outp})$: To generate or output required label \hat{Y} , an extra task operation, denoted as Task(\cdot), is finally introduced to perform different analysis tasks.

III. Methodologies

A. ABBA symbolic approximation

Our research is inspired by the observation that speech signals often contain a plethora of semantic information^[24], which enables the language model to perform extremely well across a multitude of tasks; see^[4] and references therein. However, directly applying language models to time series is not possible due to the fact that time series are made up of numerical values and lack useful embedding patterns; further, the high dimensionality of time series makes it difficult for the sequential and recurrent model to capture the dependencies of time series features. Thus learning an informative symbolic time series representation while having dimensionality reduced is a practical yet challenging problem. ABBA—a symbolic approximation method—is designed to address this as it compresses the time series to a symbolic presentation in terms of amplitude and period, and each symbol describes the oscillatory behavior of time series during a specific period.

ABBA utilizes adaptive polygonal chain approximation followed by mean-based clustering to achieve symbolization of time series. The reconstruction error of the representation can be modeled as a *Brownian bridge* with pinned start and end points. ABBA symbolization contains two dominant procedures, namely *compression* and *digitization*, to aggregate a time series $T = [t_1, t_2, \dots, t_n] \in \mathbb{R}^n$ into its symbolic representation $A = a_1 a_2 \dots a_N$ where $N \ll n$ and a_i is an element in a specific letter set \mathcal{L} , which is referred to as a *dictionary* in the ABBA procedure.

1. Compression

The ABBA compression is performed to compute an adaptive piecewise linear continuous approximation (APCA) of T . The ABBA compression plays a critical role in dimensionality reduction in ABBA symbolic approximation—a user-specific tolerance, denoted by tol , is given to determine the degree of the reduction. The ABBA compression proceeds by adaptively selecting $N + 1$ indices $i_0 = 0 < i_1 < \dots < i_N = n$ given a tolerance tol such that the time series T is well approximated by a polygonal chain going through the points (i_j, t_{i_j}) for $j = 0, 1, \dots, N$. This leads to a partition of T into N pieces $p_j = (\text{len}_j, \text{inc}_j)$ that represents cardinality and increment of $T_{i_{j-1}:i_j} = [t_{i_{j-1}}, t_{i_{j-1}+1}, \dots, t_{i_j}]$, which is calculated by $\text{len}_j \in \mathbb{N} := i_j - i_{j-1} \geq 1$ and $\text{inc}_j \in \mathbb{R} := t_j - t_{j-1}$. As such, each piece p_j is represented by a straight line connecting the endpoint values $t_{i_{j-1}}$ and t_{i_j} . Given an index i_{j-1} and starting with $i_0 = 0$, the procedure seeks the largest possible i_j such that $i_{j-1} < i_j \leq n$ and

$$\sum_{i=i_{j-1}}^{i_j} \left(t_{i_{j-1}} + (t_{i_j} - t_{i_{j-1}}) \cdot \frac{i - i_{j-1}}{i_j - i_{j-1}} - t_i \right)^2 \leq (i_j - i_{j-1} - 1) \cdot \text{tol}^2. \quad (1)$$

This means that this partitioning criterion indicates that the squared Euclidean distance of the values in p_j from the straight polygonal line is upper bounded by $(\text{len}_j - 1) \cdot \text{tol}^2$.

Following the above, the whole polygonal chain can be recovered *exactly* from the first value t_0 and the tuple sequence $[p_1, p_2, \dots, p_N]$ in the sense that the reconstruction error of this representation is with pinned start and end points and can be naturally modeled as a Brownian bridge. In terms of (1), a lower tol value is required to ensure an acceptable compression of time series with a great variety of features such as trends, seasonal and nonseasonal cycles, pulses and steps. As indicated in [13], the error bound between the reconstruction and original time series is upper bounded by $(n - N) \cdot \text{tol}^2$.

2. Digitization

The ABBA compression is followed by a reasonable digitization that leads to a *symbolic representation*. Prior to digitizing, the tuple lengths and increments are separately normalized by their standard deviations σ_{len} and σ_{inc} , respectively. After that, further scaling is employed by using a parameter scl to assign different weights to the length of each piece p_i , which denotes the importance assigned to its length value in relation to its increment value. Hence, the clustering is effectively performed on the *scaled tuples* $p_i^s = (\text{scl} \frac{\text{len}_i}{\sigma_{\text{len}}}, \frac{\text{inc}_i}{\sigma_{\text{inc}}})$, $i = 1, \dots, N$. In particular, if $\text{scl} = 0$, then clustering will be only performed on the increment values of p_i^s , while if $\text{scl} = 1$, the lengths and increments are treated with equal importance.

Algorithm 1: Greedy sorting-based aggregation

- 1) Scale and sort data points, and assume they are denoted p_1^s, \dots, p_n^s . Label all of them as “unassigned”.
 - 2) For $i \in \{1, \dots, n\}$ let the first unassigned point p_i^s as *starting point* and set $j := i$. If there are no unassigned points left, go to Step 6.
 - 3) If $\|p_i^s - p_j^s\|_2 \leq \alpha$,
 - assign p_j^s to the same group as p_i^s
 - increase $j := j + 1$
 - 4) If $j > n$ or termination condition is satisfied, go to Step 2. Otherwise go to Step 3.
 - 5) For each group computed, compute the center of the group as the mean of all its points.
-

The next step after normalization works with a mean-based clustering technique in Euclidean space. In the ABBA setting, letting the input of N vectors be $P^s = [p_1^s, \dots, p_N^s] \in \mathbb{R}^{2 \times N}$, one seeks a codebook of k vectors, i.e., $C = [c_1, \dots, c_k] \in \mathbb{R}^{2 \times k}$ ($k \ll N$) where each c_i is associated with a unique cluster S_i such that k clusters from P^s minimize the sum of Euclidean distances SSE constructed by C . The obtained codebook vectors are known as cluster centers. A quality codebook produces k clusters $\{S_1, S_2, \dots, S_k\} \subseteq P^s$ such that the sum of squared errors $\text{SSE} = \sum_{i=1}^k \sum_{p^s \in S_i} \|p^s - c_i\|_2^2$ is small enough to an optimal level. However, this is a suboptimal solution to minimizing SSE. The k-means problem aims to find k clusters within data in d -dimensional space, to minimize the SSE. As the iterations proceed, the mean value $\mu_i := \frac{1}{|S_i|} \sum_{p^s \in S_i} p^s$ is always chosen for updating centers c_i in k-means algorithms [25] to ensure SSE decreases. However, attaining the global minimum SSE is NP-hard even if k is restricted to 2 [26][27] or in the plane [28]. Typically, the sub-optimal k-means problem in digitization can also be solved by a greedy sorting-based aggregation [44], which achieves a speedup by orders of magnitude compared to the one using k-means (in practice, k-means++ is employed). The principle of the aggregation is to achieve an error-controlled clustering by greedily selecting the *starting points* according to a precomputed sorting. An efficient algorithm to symbolize time series in a large-scale way is desired, and thus sorting-based aggregation (See Algorithm 1 for a description) is preferred. The number of symbols generated by Algorithm 1 is determined by the parameter α ; see [44] for details. The SSE achieved by Algorithm 1 is upper bounded by $\alpha^2(N - k)$, and the expected SSE value is $\frac{\alpha^2(N-k)}{2}$.

In the context of symbolic approximation, we refer to the cluster centers as *symbolic centers*, and each symbolic center is associated with an identical symbol. Then each p_i^s is assigned to the closest symbolic center c^i associated with its symbol $c^i = \text{argmin}_{c \in C} (\|p^s - c\|)$. After that, each p_i^s is associated with a unique center, which is assigned as a label. We use a symbol to correspond to the label. The symbols can be represented by text characters, which are not limited to English alphabet letters—e.g., ASCII codes or any of its combinations. As such, the ABBA symbolization can be flexibly adapted to LLMs’ pretrained tokens.

3. Inverse symbolization

The *inverse symbolization* step converts the symbolic representation A back to the reconstructed series \hat{T} , which is key for some value prediction tasks in time series. The inverse symbolization is followed by a *inverse-digitization* that uses the k representative elements $c_i \in C$ to replace the symbols in A and denormalize them separately, thus resulting in a 2-by- N array \tilde{P} —an approximation of P . Each $\tilde{p}_i \in \tilde{P}$ is the closest symbolic center $c^i \in C$ to $p_i^s \in P^s$ (in contrast to P) after denormalization. However, the inverse digitization often leads to non-integer values for the reconstructed length len , so a rounding method is used to align the accumulated lengths with the closest integers. The first length is rounded to an integer value, i.e., $\hat{\text{len}}_1 := \text{round}(\tilde{\text{len}}_1)$ and the rounding error $e := \tilde{\text{len}}_1 - \hat{\text{len}}_1$ is computed. The error is then added to the rounding of $\tilde{\text{len}}_2$, i.e., $\hat{\text{len}}_2 := \text{round}(\tilde{\text{len}}_2 + e)$, and the new error e' is calculated as $\hat{\text{len}}_2 + e - \tilde{\text{len}}_2$. Then e' is similarly involved in the next rounding. After all rounding is computed, we obtain

$$\{(\hat{\text{len}}_1, \hat{\text{inc}}_1), (\hat{\text{len}}_2, \hat{\text{inc}}_2), \dots, (\hat{\text{len}}_N, \hat{\text{inc}}_N)\} \in \mathbb{R}^2, \quad (2)$$

where the increments inc are unchanged, i.e., $\hat{\text{inc}} = \tilde{\text{inc}}$. The last step is to recover \hat{P} exactly from the initial time value t_0 and the tuple sequence (2), resulting in the reconstructed time series \hat{T} .

B. Error analysis reconstruction

We are concerned with the reconstruction error of ABBA's symbolization since a symbolic representation with a higher reconstruction error is a less informative representation. It is worth noting that the reconstruction of time series from the compression procedure proceeds by establishing a polygonal chain \tilde{T} going through the chosen tuples $\{(i_j, t_{i_j})\}_{j=0}^N$ from the original time series T and $\text{len}_j = i_{j+1} - i_j$. As indicated in [13], a polygonal chain \hat{T} stitching together $\{(i_j, \hat{t}_{i_j})\}_{j=0}^N$ via a tuple sequence \hat{P} is reconstructed by the inverse symbolization, thus we have Theorem III.1.

Theorem III.1. Let $(\mu_i^{\text{len}}, \mu_i^{\text{inc}}) = \frac{1}{|S_i|} \sum_{(\text{len}, \text{inc}) \in S_i} (\text{len}, \text{inc})$, we denote the mean set for len and inc by $\mathcal{U}_{\text{len}} = \{\mu_i^{\text{len}}\}_{i=1}^k$ and $\mathcal{U}_{\text{inc}} = \{\mu_i^{\text{inc}}\}_{i=1}^k$, respectively. Since $i_0 = 0$, the reconstruction indices and size of time series values are given by

$$\{(i_j, \hat{t}_{i_j}) = (\sum_{\ell=1}^j \hat{\text{len}}_\ell, t_0 + \sum_{\ell=1}^j \hat{\text{inc}}_\ell), \text{for } j = 0, \dots, N\}, \quad (3)$$

where $(\hat{\text{len}}_\ell, \hat{\text{inc}}_\ell)$ are the computed cluster centers, i.e., $\hat{\text{len}}_\ell \in \mathcal{U}_{\text{len}}$ and $\hat{\text{inc}}_\ell \in \mathcal{U}_{\text{inc}}$.

Theorem III.1 shows the accumulated deviations from the true lengths and increments are canceled out (as analyzed in [13]) at the right endpoint of the last piece p_N , thus $(\hat{i}_N, \hat{t}_{i_N}) = (i_N, t_{i_N}) = (n, t_n)$, which indicates the start and ending point between \hat{T} , \tilde{T} and T are identical. We thus have the following result.

We now denote the local deviation of the increment and length by

$$d_\ell^{\text{inc}} := \hat{\text{inc}}_\ell - \tilde{\text{inc}}_\ell, \quad d_\ell^{\text{len}} := \hat{\text{len}}_\ell - \tilde{\text{len}}_\ell. \quad (4)$$

Theorem III.2 ([13]).

$$\sum_i \sum_{(\text{len}, \text{inc}) \in S_i} (d_\ell^{\text{len}}, d_\ell^{\text{inc}}) = (0, 0).$$

Theorem III.3. Assume that ABBA is performed with hyperparameter α , and it results in k clusters S_1, \dots, S_k . Then we have

$$\max_\ell \{(d_\ell^{\text{inc}})^2 + (d_\ell^{\text{len}})^2\} \leq \alpha^2, \quad (5)$$

and further

$$\sigma = \max_{i=1, \dots, k} \frac{1}{|S_i|} \sum_{(\text{len}, \text{inc}) \in S_i} (|\text{len} - \mu_i^{\text{len}}|^2 + |\text{inc} - \mu_i^{\text{inc}}|^2) \leq \alpha^2,$$

Following Theorem III.3, the σ is explicitly controlled by α , and thus we remove the need to estimate the additional parameter tol_s that is used in [13] by directly relating it to the hyperparameter α .

Given the N data points selected by adaptive polygonal approximation chain, letting $e_j^{\text{len}} := \sum_{\ell=1}^j d_\ell^{\text{len}}$ and $e_j^{\text{inc}} := \sum_{\ell=1}^j d_\ell^{\text{inc}}$, it is obvious that $e_j^{\text{inc}} = \hat{t}_{i_j} - t_{i_j}$ if $e_j^{\text{len}} = 0$ for $j = 1, \dots, N$. This leads to Theorem III.4 and Theorem III.5 below.

Theorem III.4.

$$|e_j^{\text{inc}}| \leq j\sqrt{\alpha^2 - (d_\ell^{\text{len}})^2} \leq j|\alpha|,$$

where $j = 0, \dots, N$.

Similarly, the shift of the time series has $e_j^{\text{len}} \leq j\sqrt{\alpha^2 - (d_\ell^{\text{inc}})^2} \leq j|\alpha|$ for $j = 0, \dots, N$.

Theorem III.5.

$$\mathbb{P}(|e_j^{\text{inc}}| \geq h) \leq \exp\left(-\frac{h^2}{2j\alpha^2}\right) \quad \text{and} \quad \mathbb{P}(|e_j^{\text{len}}| \geq h) \leq \exp\left(-\frac{h^2}{2j\alpha^2}\right).$$

for all $h > 0$.

Proof of Theorem III.5. From Theorem III.2, we can easily obtain

$$(e_0^{\text{len}}, e_0^{\text{inc}}) = (0, 0), \quad (e_N^{\text{len}}, e_N^{\text{inc}}) = (0, 0)$$

associated with expectation $E(e_j^{\text{len}}) = E(e_j^{\text{inc}}) = 0$.

For $j = 1, \dots, N$, since $d_j^{\text{len}}, d_j^{\text{inc}} \in [-\alpha, \alpha]$, using (5) and Hoeffding's inequality,

$$\begin{aligned} & \mathbb{P}\left(\left|\sum_{\ell=1}^j (d_\ell^{\text{inc}} - \mathbb{E}[d_\ell^{\text{inc}}])\right| \geq h\right) \\ &= \mathbb{P}\left(|e_j^{\text{inc}} - \mathbb{E}[e_j^{\text{inc}}]| \geq h\right) \\ &\leq \exp\left(-\frac{h^2}{2j\alpha^2}\right). \end{aligned}$$

Therefore, $\mathbb{P}(|e_j^{\text{len}}| \geq h) \leq \exp\left(-\frac{h^2}{2j\alpha^2}\right)$ and $\mathbb{P}(|e_j^{\text{inc}}| \geq h) \leq \exp\left(-\frac{h^2}{2j\alpha^2}\right)$ for all $t > 0$. \square

This means that a decrease of α is prone to result in a smaller reconstruction error e_j . This phenomenon was mentioned in [\[31\]](#). The growth of j increases the possibility of larger errors since the errors coming from the previous reconstruction will be accumulated into the subsequent reconstruction by the principle of inverse symbolization.

C. ABBA to LLM

In the following, we write a single time series containing n data points as T , and use $\mathcal{T} = \{T_i\}_{i=1}^q$ to denote a set of time series of cardinality q , associated with its corresponding symbolic representation set $\mathcal{A} = \{A_i\}_{i=1}^q$.

1. Fixed-point adaptive polygonal chain

In time series prediction settings, the value-based prediction is converted into a token-based prediction using STSA. However, it is very desirable to mitigate the negative effect of the preceding mistakenly predicted symbol on the subsequent time series recovery since the recovery proceeds from front to back. However, APCA and the symbolic recovery often lead to a cumulative error for symbolic prediction, that is, an incorrect replacement of a previous symbol will influence the subsequent reconstruction. A *fixed-point polygonal chain* trick is introduced to mitigate this issue. We still partition the time series into pieces following (1) while $p_j = (\text{len}_j, \text{inc}_j)$ is replaced with $p_j = (\text{len}_j, t_{i_j})$ before normalization. We call the new approximation method fixed-point adaptive piecewise linear continuous approximation (FAPCA). The resulting tuples p_i will be normalized and one can be recovered from the other since $\text{inc}_j = t_{i_j} - t_{i_{j-1}}$. Figure 4 shows that FAPCA eliminates the cumulative errors arising from the preceding mistaken symbol and improves the recovery.

2. Symbolizing multiple time series

Existing work on symbolic approximation focuses on converting a single time series; it can not convert another time series with consistent symbolic information (i.e., each symbol is associated with a unique symbolic center). To allow the manipulation of co-evolving time series or multiple time series, it is necessary to keep consistent symbolic information for multiple symbolic time series representations.

We illustrate a unified approach towards a consistent symbolic approximation for multiple time series.

- Step 1: Use APCA or FAPCA to compress each time series T_i into P_i for $i = 1, \dots, q$

- Step 2: Compute normalized P_i^s and concatenate P_i^s to form $\mathcal{P}^s := [P_i^s]_{i=1}^q$
- Step 3: Perform digitization on \mathcal{P}^s
- Step 4: Allocate symbols to each time series (the number of symbols for T_i is equal to $|P_i^s|$)

3. Symbolizing out-of-sample data

Symbolizing out-of-sample time series data with consistent symbols is essential for various time series downstream tasks, e.g., inference tasks. To symbolize $\mathcal{T}^t = \{T_i^t\}_{i=1}^{q'}$, we perform the following steps:

- Step 1: Compress each time series T_i^t into P_i^t for $i = 1, \dots, q'$
- Step 2: Assign a symbol to $p \in P_i^t$ for $i = 1, \dots, q'$ following the rule of digitization

4. Feeding the LLM

ABBA can transform numerical time series to symbolic series and keep the internal logic chain from which LLMs can learn temporal knowledge. In other words, by ensuring the precondition that the input symbolic series inherits the polygonal chain of numerical time series and then represents this chain via symbolic series (or LLMs' tokens) that can be recognized by LLMs, LLMs can **reconstruct** the embedding space without the use of any new tokens via adapting fine-tuning methods. As seen in Figure 3, the left panel is the traditional setting in terms of corresponding tasks, such as classification, regression, and prediction. The right panel is the instruction setting that contains these three tasks. Instructions (the right panel of Figure 3) only guide LLMs to understand the tasks, and this smart design is equal to the left panel (without LLMs' Instructions) of Figure 3.

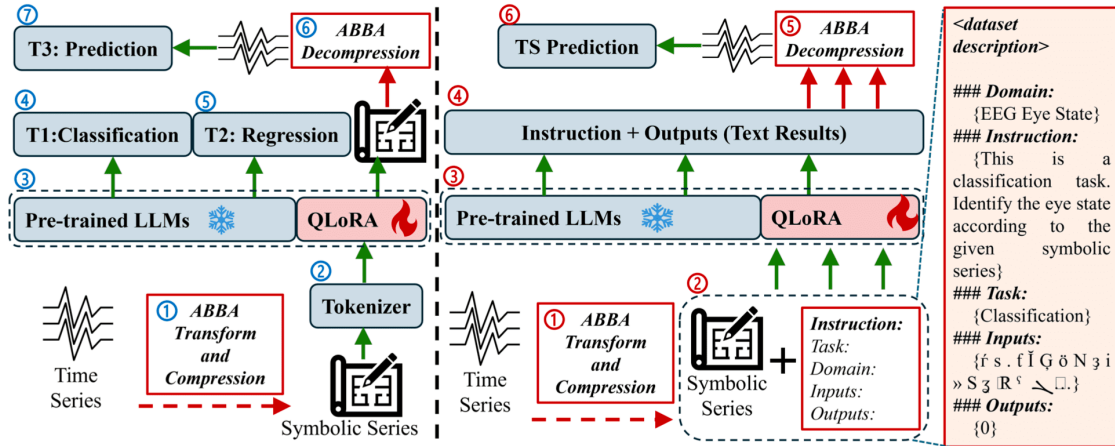


Figure 3. The model framework of LLM-ABBA: Given an input time series, we first transform and compress the time series to a symbolic series via ① and ②. These symbolic series will be tokenized by the LLM's tokenizer ②. The designed instruction that contains the symbolic series also will be tokenized by the LLM's tokenizer ②. Additionally, by only fine-tuning the pretrained LLM, the QLoRA with inhibition mechanism is utilized both in ③ and ③. To implement the corresponding tasks, ④ and ⑤ loads the LLM according to the type of task. However, ④ loads the LLM on the generation task. Moreover, to inverse symbolic series back to numerical time series, ⑥ and ⑥ utilizes ABBA to decompress the generated symbolic series. Lastly, in ⑦ and ⑦ the output time series from LLM-ABBA are projected to generate the forecasts.

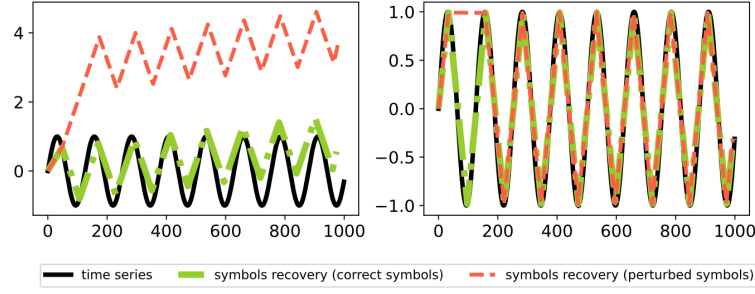


Figure 4. We generate a synthetic trigonometric sine series of 1,000 points, and separately perform symbolic approximation with 4 symbols using APCA (left) and FAPCA (right) on the time series. ABBA with APCA and FAPCA generate symbols “aBbBbBbBbBbBbBbBA” and “abBbBbBbBbBbBbBbBA”, respectively, associated with their respective perturbed symbols, “abbBbBbBbBbBbBbBA” and “aBBbBbBbBbBbBbBbBA”. The symbol recovery is performed on correct symbols and perturbed symbols, respectively.

For the consistency of related tuning-based methods, \mathcal{T} is referred to as the input in the time series dataset, \mathcal{A} is the symbolic representation generated by ABBA; $\phi: \mathcal{T} \rightarrow \mathcal{A}$ denotes the symbolization of ABBA, and $\phi^{-1}: \mathcal{A} \rightarrow \mathcal{T}$ is referred to as the inverse symbolization of ABBA. We formulate the framework of LLM-ABBA:

- i. $\mathcal{A} = \phi(\mathcal{T})$: The input \mathcal{T} is converted to its symbolic representation \mathcal{A} .
- ii. $\mathcal{M}_{inp} = \text{Tokenizer}(\text{Prompt}, \mathcal{A})$: Tokenizing the symbolic representation \mathcal{A} ; here, the *Tokenizer* is the default Tokenizer for LLMs.
- iii. $\mathcal{M}_{outp} = f_{LLM}^{\Delta}(\mathcal{M}_{inp})$: Feed the tokenized input to LLM model.
- iv. $\hat{Y} = \text{Task}(\mathcal{M}_{outp})$: If this is a classification task, \hat{Y} is a generated label. If the task is a regression or prediction task, \hat{Y} is an ABBA-transformed numerical value or sequence produced by the inverse symbolization process of ABBA:

$$\begin{cases} \hat{Y} = \mathcal{M}_{outp}, & \text{Classification task,} \\ \hat{Y} = \phi^{-1}(\mathcal{M}_{outp}), & \text{Regression / prediction task} \end{cases}$$

D. Linguistics investigation: Zipf's law

In nearly all corpora, the most common word appears approximately twice as frequently as the next common word; this phenomenon is explained by Zipf's law^[29]. Zipf's law asserts that the frequencies of certain events are inversely proportional to their rank, and further, the rank-frequency distribution is an inverse power law relation.

In Figure 5, we can see unigrams generated by ABBA symbolization from 7 different time series datasets from the UCR Archive coarsely meet Zipf's law. This showcases an appealing alignment between ABBA symbols and the native language words.

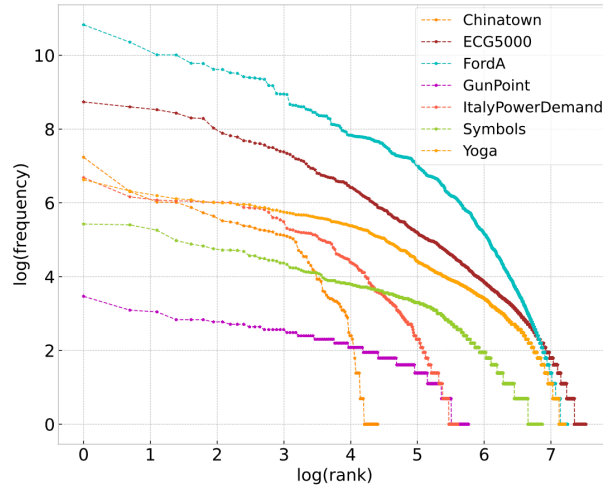


Figure 5. Frequency and rank of symbols in various UCR datasets.

IV. Experiments

In this section, we study three time series tasks to validate the efficiency of ABBA in LLM. We also fine-tune three language models on the training data using QLoRA^[17] with inhibition^[30]. All experiments are simulated in PyTorch with a single NVIDIA A100 40GB GPU. The benefits of LLM-ABBA include (1) avoiding the need for LLMs to learn time series from scratch, and (2) only utilizing compression and decompression without the need for the training of extra embedding layers^[6]. For a fair comparison, we evaluate our models on the same settings for each task. In the following, unless otherwise stated, we assume that the greedy aggregation is used for the ABBA digitization.

A larger dataset needs more symbols or LLM tokens, as a larger time series dataset contains more information and symbolic semantics. RoBERTa_{Large} is based on BERT^[31] which considers two directions of the input language sentence, and Llama2-7B and Mistral-7B originate from the GPT architecture^[32] that only takes one direction (from left to right) into account. Causality analysis which is frequently used to compute the contextual of each signal has been widely used to analyze multichannel EEG signals. However, ECG signals mostly rely on sequential features. Thus, we infer that when using LLM-ABBA to analyze medical time series, the properties and characteristics should be analyzed first. For some datasets, we could not find or reproduce SOTA performance numbers. For a comprehensive analysis, we test ABBA with LLMs on three main time series analysis tasks. In this section, three LLMs are used to process the COP in symbolic series. M1 is RoBERTa_{Large}^[33], M2 is Llama2-7B^[34], and M3 is the Mistral-7B^[35].

A. Hyperparameters

1. Hyperparameters of ABBA

There are four interactive parameters that establish the transition of time series when integrating ABBA into LLMs. The tolerance tol is chosen from $\{1 \times 10^{-2}, 1 \times 10^{-4}, 1 \times 10^{-6}\}$ to control the degree of the compression and dimension reduction, and the digitization parameter α is chosen from $\{1 \times 10^{-2}, 1 \times 10^{-4}, 1 \times 10^{-6}\}$ to determine the number of distinct symbols. \mathcal{L} is a finite letter set that can be specified as the LLMs' tokens, and $\text{scl} \in \{1, 2, 3\}$ is used as the normalized scaling for the length of each piece.

2. Hyperparameters of LLMs

LLM-ABBA on Classification Tasks												
Models	Quant.	Tokens	Metric	LoRA					Optim.	Epochs	LR	Batch Size
	4-bits	Length		alpha	low rank r	dropout	inhib.	Embed.				
RoBERTa _{Large}	True	512	Acc.	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	5e-7	4
Llama2-7B	True	4,096	Acc.	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	5e-7	4
Mistral-7B	True	4,096	Acc.	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	5e-7	4

Table I. Hyperparameters of Classification tasks. Quant. is the model quantization process. Inhib. is the inhibition threshold in QLoRA. Embed. means to save tuned embeddings. Optim. is the optimization method. LR is the learning rate. Acc. is the accuracy rate (%).

LLM-ABBA on Regression Tasks												
Models	Quant.	Tokens	Metric	LoRA					Optim.	Epochs	LR	Batch Size
	4-bit	Length		alpha	low rank r	dropout	inhib.	Embed.				
RoBERTa _{Large}	True	512	RMSE	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	2×10^{-6}	4
Llama2-7B	True	4,096	RMSE	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	2×10^{-6}	4
Mistral-7B	True	4,096	RMSE	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	2×10^{-4}	4

Table II. Hyperparameters of Regression tasks. Quant. is the model quantization process. Inhib. is the inhibition threshold in QLoRA. Embed. means to save tuned embeddings. Optim. is the optimization method. RMSE is the root-mean-square-error.

LLM-ABBA on Prediction Tasks												
Models	Quant.	Tokens	Metric	LoRA					Optim.	Epochs	LR	Batch Size
	4-bit	Length		alpha	low rank r	dropout	inhib.	Embed.				
RoBERTa _{Large}	True	512	MAE, MSE	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	2×10^{-6}	4
Llama2-7B	True	4,096	MAE, MSE	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	2×10^{-6}	4
Mistral-7B	True	4,096	MAE, MSE	16	16, 64, 256	0.05	0.3	Save	adamw_8bit	10	2×10^{-6}	4

Table III. Hyperparameters of Prediction tasks. Quant. is the model quantization process. Inhib. is the inhibition threshold in QLoRA. Embed. means to save tuned embeddings. Optim. is the optimization method. MAE is the mean-absolute-error, and MSE is the mean-square-error.

There are three time series analysis tasks: classification, regression, and prediction. We quantize LLMs by 4-bits using the bitsandbytes package². In order to fine-tune LLMs, the shunting inhibition mechanism^[30] is utilized during the QLoRA adapter fine-tuning progress. The modified embedding layer is also saved after fine-tuning on the corresponding task. For the classification task, the metric is accuracy rate (%). Root-mean-square-error is used as the metric for regression tasks. Mean-square-error and mean-absolute-error are used as the metrics for prediction tasks,

and we also visualize the correlation coefficient of prediction tasks on ETTh1 data in terms of their seven features. We control the fine-tuning epoch and apply a small batch size on every task. The alpha of QLoRA is set to 16.

B. Compression and Recovery

To transform the numerical time series to symbolic time series, we use tokens of LLMs as the initial dictionary of ABBA for the symbolic representation, and there are no extra tokens that will be used to represent the numerical input. ABBA shows a strong symbolic transition on time series signals (See Figure 6 and Table IV).

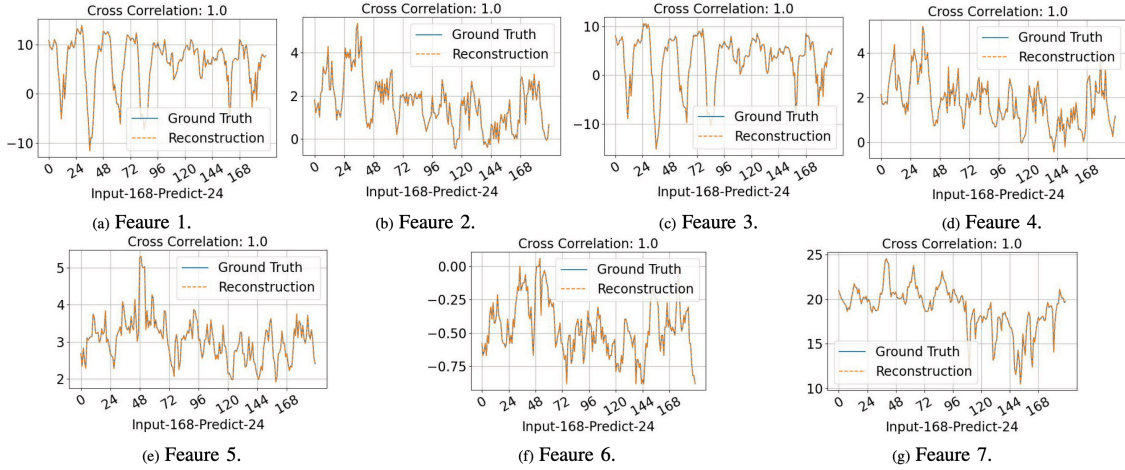


Figure 6. Visualization of reconstructed input-168-predict-24 results on ETTh1 data by using ABBA symbolic approximation, where $\text{tol} = 0.01$, $\alpha = 0.01$ and $\text{scl} = 3$.

To visualize the performance of ABBA on time series transition processes, we employ ETTh1 time series data to compute the correlation coefficient and reconstruction error of ABBA. This multivariate data has seven features, and in terms of these seven features, the average of mean-absolute-error (MSE), mean-square-error (MAE), and correlation coefficient between original time series input and reconstructed outputs is computed.

ABBA Settings		Number of Symbols	Reconstructed Time Series		
tol and α	scl	Used LLM's tokens	MSE	MAE	Correlation Coefficient
$1 \times 10^{-2}, 1 \times 10^{-2}$	3	846	2.5×10^{-7}	1×10^{-2}	1.0
$1 \times 10^{-4}, 1 \times 10^{-4}$	3	2,713	4.2×10^{-8}	1.4×10^{-4}	1.0
$1 \times 10^{-6}, 1 \times 10^{-6}$	3	2,789	3.2×10^{-8}	1.3×10^{-4}	1.0

Table IV. Symbolic approximation performance on ETTh1 data using ABBA. ABBA describes a time series sample by using symbolic approximation, and the number of used symbols depends on the complexity of the data. If the time series sample is a regular wave (for example, a sine wave), the number of used symbols is small; otherwise, ABBA needs more symbols.

In this section, we observe which ABBA settings better suit time series characteristics. The default scl is set to 3, which is used in other LLM tasks. tol and α are set to be the same. Table IV reports the input-168-predict-96 results when using ABBA to reconstruct ETTh1 data in terms of seven features. Setting smaller tol and α in ABBA can reduce the MSE and MAE scores, but more symbols or LLM tokens will be used. Under all above conditions, the correlation coefficient is 1.0.

C. Time Series Classification Tasks

For the classification task, we evaluate these three pretrained LLMs on UCR Time Series Archive datasets^[36], EEG eye state^[37], and MIT-BIH^[38]^[39] which have been extensively adopted for benchmarking time series classification models. We utilize cross-entropy loss for the classification training. Details of the implementation and datasets can be found in Table I. The evaluation metric is accuracy rate (%).

The UCR Archive contains 128 datasets already partitioned into train and test sets, although the ratio of the train set and test set is not always consistent³. These datasets have varying numbers of labels and feature dimensions. Also, there can be uneven numbers of labels, which often results in overfitting. Therefore, classifying time series in the UCR Archive is a challenging task. Table V reports the full time series classification results on UCR2018. J1 refers to results of using k-means in the digitization process, and J2 refers to the results of using greedy aggregation (Algorithm 1) in the digitization process. We find that Algorithm 1 outperforms k-means symbolization time series transition progress in most cases.

Data	Classes	Symbols	RoBERTa _{Large}			Llama2-7B			Mistral-7B			V2Sailor	
	Number	Number	Para.	J1	J2	Para.	J1	J2	Para.	J1	J2	Para.	SOTA
BME	3	836	2.65M	34.0	60.2	12.7M	41.3	84.7	9.56M	43.3	77.3	0.3M	-
BeetleFly	2	731	2.65M	65.0	95.0	12.7M	50.0	65.0	9.56M	55.0	75.0	0.3M	-
BirdChicken	2	424	2.65M	55.0	70.0	12.7M	60.0	65.0	9.56M	55.0	75.0	0.3M	-
ChinaTown	2	585	2.65M	72.0	72.6	12.7M	58.3	84.3	9.56M	61.5	89.2	0.3M	-
Coffee	2	701	2.65M	50.0	89.3	12.7M	60.7	96.5	9.56M	78.6	89.3	0.3M	100
DistalPhalanxOutlineAgeGroup	3	1,444	2.65M	68.3	68.3	12.7M	71.2	73.4	9.56M	67.6	74.8	0.3M	-
DodgerLoopWeekend	2	143	2.65M	72.6	73.9	12.7M	70.3	64.5	9.56M	69.6	71.7	0.3M	-
ECG200	2	1,781	2.65M	70.0	68.0	12.7M	63.0	64.0	9.56M	66.8	68.0	0.3M	87.4
ECG5000	5	10,334	2.65M	81.2	76.0	12.7M	75.7	74.7	9.56M	75.4	73.4	0.3M	94.0
ECGFiveDays	2	2,463	2.65M	52.6	56.9	12.7M	53.3	63.9	9.56M	49.5	68.8	0.3M	-
Earthquakes	2	940	2.65M	52.7	74.8	12.7M	77.7	76.3	9.56M	79.1	76.3	0.3M	78.4
FordA	2	9,759	2.65M	68.9	68.9	12.7M	58.7	61.1	9.56M	62.7	60.9	0.3M	100
FordB	2	9,352	2.65M	68.9	58.1	12.7M	56.1	58.9	9.56M	55.1	57.0	0.3M	100
FreezerRegularTrain	2	2,663	2.65M	61.9	74.5	12.7M	64.1	76.1	9.56M	63.2	75.4	0.3M	-
FreezerSmallTrain	2	2,593	2.65M	62.3	74.1	12.7M	63.8	67.8	9.56M	63.3	67.5	0.3M	-
GunPoint	2	791	2.65M	51.4	73.3	12.7M	54.0	82.7	9.56M	48.0	80.0	0.3M	96.7
GunPointAgeSpan	2	2,057	2.65M	83.5	94.3	12.7M	69.9	84.5	9.56M	67.1	85.5	0.3M	-
GunPointMaleVersusFemale	2	2,057	2.65M	57.9	76.3	12.7M	59.8	71.2	9.56M	55.7	74.1	0.3M	-
GunPointOldVersusYoung	2	2,057	2.65M	66.7	97.5	12.7M	62.9	85.1	9.56M	67.9	80.0	0.3M	-
HandOutlines	2	7,572	2.65M	66.5	77.0	12.7M	63.5	68.6	9.56M	65.1	71.6	0.3M	93.2
Herring	2	982	2.65M	59.4	65.6	12.7M	62.5	62.5	9.56M	54.7	60.9	0.3M	68.8
HouseTwenty	2	1,385	2.65M	50.8	67.1	12.7M	69.7	89.1	9.56M	75.6	93.3	0.3M	-
ItalyPowerDemand	2	1,759	2.65M	59.7	70.4	12.7M	55.7	73.4	9.56M	53.4	73.2	0.3M	97.1
Lightning2	2	2,175	2.65M	67.2	65.6	12.7M	68.9	65.6	9.56M	67.2	62.3	0.3M	100
Meat	3	161	2.65M	55.0	70.0	12.7M	68.3	70.0	9.56M	66.7	70.0	0.3M	-
MelbournePedestrian	10	1,081	2.65M	34.6	68.5	12.7M	27.1	76.8	9.56M	29.2	74.4	0.3M	-
MiddlePhalanxOutlineCorrect	2	1,700	2.65M	59.8	67.4	12.7M	58.1	69.8	9.56M	61.2	67.7	0.3M	91.1
MiddlePhalanxTW	6	1345	2.65M	53.9	54.5	12.7M	53.9	48.7	9.56M	51.9	46.8	0.3M	84.9
OliveOil	4	150	2.65M	66.7	46.7	12.7M	76.7	70.0	9.56M	73.3	73.3	0.3M	-
PhalangesOutlinesCorrect	2	2,785	2.65M	62.2	65.4	12.7M	63.9	67.5	9.56M	62.7	67.5	0.3M	-
Plane	7	1,424	2.65M	33.3	81.0	12.7M	39.0	78.1	9.56M	38.1	83.8	0.3M	-
PowerCons	2	2,007	2.65M	77.8	79.0	12.7M	72.8	81.1	9.56M	77.8	80.6	0.3M	-
ProximalPhalanxOutlineCorrect	2	1,298	2.65M	71.5	82.8	12.7M	73.9	85.6	9.56M	72.9	83.9	0.3M	-
ProximalPhalanxTW	6	1,101	2.65M	67.8	80.0	12.7M	69.8	80.0	9.56M	68.8	74.1	0.3M	-
SemgHandGenderCh2	4	2,840	2.65M	49.1	54.7	12.7M	59.5	67.2	9.56M	58.3	73.3	0.3M	-

Data	Classes	Symbols	RoBERTa _{Large}			Llama2-7B			Mistral-7B			V2S ^[40]	
	Number	Number	Para.	J1	J2	Para.	J1	J2	Para.	J1	J2	Para.	SOTA
SmallKitchenAppliances	2	2,207	2.65M	66.2	69.3	12.7M	60.8	63.2	9.56M	57.6	61.6	0.3M	83.5
SonyAIBORobotSurface1	2	2,558	2.65M	54.2	60.4	12.7M	64.1	71.7	9.56M	68.2	78.5	0.3M	-
StarLightCurves	3	27,131	2.65M	67.8	72.9	12.7M	68.6	72.6	9.56M	67.6	70.1	0.3M	-
Strawberry	2	3,593	2.65M	71.2	85.1	12.7M	69.5	84.9	9.56M	69.5	88.4	0.3M	97.6
ToeSegmentation2	2	2,714	2.65M	79.7	73.1	12.7M	69.2	59.2	9.56M	77.7	80.0	0.3M	-
Trace	4	870	2.65M	49.5	88.0	12.7M	54.0	90.0	9.56M	47.0	77.0	0.3M	100
TwoLeadECG	2	2,487	2.65M	59.6	69.1	12.7M	53.2	64.6	9.56M	53.2	63.9	0.3M	97.8
Wafer	2	4,805	2.65M	94.6	96.8	12.7M	91.3	93.5	9.56M	90.9	95.2	0.3M	100
Wine	2	171	2.65M	53.6	57.4	12.7M	59.3	63.0	9.56M	63.0	55.6	0.3M	90.7
Worms	5	5,377	2.65M	62.6	67.5	12.7M	57.1	64.9	9.56M	54.5	63.6	0.3M	83.1
WormsTwoClass	2	5377	2.65M	74.3	81.8	12.7M	62.3	70.1	9.56M	61.0	79.2	0.3M	98.7

Table V. Full comparison of results for time series classification tasks(%) on UCR datasets.

In Table V, we report the classification performance on a partial dataset of UCR2018. In most cases, although LLM-ABBA cannot outperform the SOTA in terms of time series classification tasks, ABBA with LLMs can reach an acceptable application requirement in some practical cases (such as “Coffee”, “Earthquakes”, “Herring”, “Strawberry”, “Trace”, “Wafer”, “WormsTwoClass”). Compared to V2S^[40] which is the SOTA, although these three LLMs with the use of QLoRA occupies more memory, the multi-modality of LLMs, especially on time series analysis tasks, achieves a noticeable improvement.

In the medical domain (for example, identifying the eye state using EEG signals, distinguishing abnormal ECG signals, and classifying the “normal beats”, “supraventricular ectopy beats”, “ventricular ectopy beats”, “fusion beats”, and “unclassifiable beats” of ECG signals), we report the performance of LLM-ABBA on three medical time series datasets. We set $\text{tol} = \alpha = 0.01$. In Table VI, compared to CNN^[41] on the PTB-DB data set, LLM-ABBA achieves performance almost equivalent to the SOTA. In the aspect of distinguishing MIT-BIH, CNN^[41] and BiRNN^{[38][42]} performs the best, but LLM-ABBA slightly outperforms LSTM^{[43][44]}.

Data	Classes	Symbols	RoBERTa _{Large}			Llama2-7B			Mistral-7B			CNN ^[41]	BiRNN ^[42]	LSTM ^[44]
	Number	Number	r=16	r=64	r=256	r=16	r=64	r=256	r=16	r=64	r=256			
EEG	2	938	60.1	66.0	64.4	55.9	57.4	57.5	58.5	58.0	60.1	53.1	55.3	50.7
ptb-db	2	2,179	89.5	90.6	89.3	99.0	98.6	98.3	98.9	98.7	98.6	99.4	97.0	90.7
mit-bih	5	2,926	86.4	86.4	86.3	89.6	89.4	89.1	89.3	89.7	89.3	93.4	96.5	88.1

Table VI. Full comparison of results on medical time series classification tasks(%) on EEG eye states, ptb-db, and MIT-BIH.

D. Time Series Regression Tasks

For the regression task, we evaluate these three pretrained LLMs on the Time Series Extrinsic Regression (TSER) benchmarking archive^[2], which contains 19 time series datasets from 5 application domains, including Health Monitoring, Energy Monitoring, Environment Monitoring,

Sentiment Analysis, and Forecasting⁴. To use as few symbols as possible, we initialize the setting of $\text{tol} = 0.01$ and $\alpha = 0.01$. We also utilize the L2 loss for the regression training. Details of the implementation and datasets can be found in Table II. The evaluation metric is root-mean-square-error (RMSE).

Experimenting on the TSER benchmark archive^[2], the empirical results are shown in Table VII, in which for 15 out of 19 use-cases, LLM-ABBA outperforms the machine learning SOTA results. We believe that LLM-ABBA can exploit the semantic information hiding beneath the time series in the task of time series regression. ABBA is able to provide COPs to LLMs by compressing and digitizing time series to symbols, which finally results in the change of embedding space by using adaption fine-tuning methods.

Data	Symbols	RoBERTa _{Large}			Llama2-7B			Mistral-7B			SOTA ^[2]
		r=16	r=64	r=256	r=16	r=64	r=256	r=16	r=64	r=256	
	Number	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
AppliancesEnergy	778	1.73	2.09	1.74	2.43	2.43	2.43	2.34	2.02	2.11	2.29
HouseholdPowerConsumption1	1717	377.02	377.20	377.20	398.01	398.05	398.05	228.83	228.78	228.67	132.80
HouseholdPowerConsumption2	1717	27.64	27.71	27.73	36.63	36.71	36.69	24.54	24.56	24.51	32.61
BenzeneConcentration	3037	4.01	4.00	4.00	5.57	5.56	5.56	4.03	4.03	4.03	0.64
BeijingPM10Quality	970	66.16	66.07	66.07	93.25	93.26	93.26	65.25	65.25	65.24	93.14
BeijingPM25Quality	970	54.16	54.16	54.16	76.75	76.73	76.73	53.50	53.49	53.49	59.50
LiveFuelMoistureContent	5689	20.56	20.56	20.56	29.32	29.33	29.32	20.94	20.88	20.85	29.41
FloodModeling1	969	0.00	0.00	0.00	0.05	0.05	0.05	0.37	0.36	0.36	0.00
FloodModeling2	979	0.00	0.00	0.00	0.05	0.04	0.04	0.40	0.39	0.39	0.01
FloodModeling3	948	0.00	0.00	0.00	0.06	0.05	0.05	0.41	0.37	0.39	0.00
AustraliaRainfall	4740	4.36	4.36	4.36	6.05	6.01	6.02	4.31	4.28	4.30	8.12
PPGDalia	12298	9.32	9.32	9.32	12.54	12.50	12.52	9.04	9.02	9.03	9.92
IEEEPPG	8971	17.06	17.00	17.04	22.59	22.53	22.55	17.15	17.12	17.16	23.90
BIDMC32HR	9423	6.73	6.98	6.71	12.02	11.98	12.04	8.24	8.21	8.23	9.42
BIDMC32RR	9412	1.77	1.74	1.76	2.64	2.61	2.62	2.09	2.06	2.08	3.02
BIDMC32SpO2	5537	2.90	2.85	2.89	3.82	3.79	3.81	2.95	2.91	2.93	4.45
NewsHeadlineSentiment	5537	0.07	0.07	0.07	0.13	0.13	0.13	0.11	0.11	0.11	0.14
NewsTitleSentiment	5537	0.07	0.07	0.07	0.13	0.13	0.13	0.11	0.11	0.11	0.14
Covid3Month	227	0.02	0.02	0.02	0.11	0.11	0.11	0.45	0.44	0.44	0.04

Table VII. Full comparison of results on the regression task on 19 Monash Time Series Regression datasets.

E. Time Series Forecasting Tasks

For time series forecasting, we experimented on 4 well-established benchmarks: ETT datasets (including 4 subsets: ETTh1, ETTh2, ETTm1, ETTm2) ^{[45][46]}. Details of the implementation and datasets can be found in Table III. The input length of the time series is 168, and we use three different prediction horizons $H \in \{24, 96, 168\}$. The evaluation metrics include MSE and MAE.

Although LLM-ABBA cannot obtain a new SOTA on time series forecasting tasks, it compares favorably to the Informer architecture which is trained from scratch. The congenital defect of ABBA is that the symbolization tends to be affected by the fluctuation and oscillation of time series

signals, which eventually leads to higher MSE and MAE scores. Because LLM-ABBA utilizes a totally different technical roadmap to existing methods, it only remodels the construction of the LLM's tokens. However, remodeling pretrained tokens inevitably brings the previous pretrained semantics to the LLM-ABBA design. Thus, we discussed the semantic consistency of LLM-ABBA using extra symbols or tokens to overcome this problem.

F. QLoRA Fine-Tuning

Because the low rank of adapter fine-tuning will influence the efficiency of passing information^{[17][30]} from the previous layer, we use different low rank settings of QLoRA on the corresponding tasks during the fine-tuning progress. But for time series regression and prediction tasks, we select $r \in \{16, 46, 256\}$ for the corresponding data input. We find that there is no obvious over-fitting problem, and more tunable parameters are not able to improve the performance of LLM-ABBA.

In medical time series domains, ptb-db and MIT-BIH arrhythmia data sets are mostly used. EEG eye state data set has two categories, and because of its high complexity, the accuracy always stays at around 60%. EEG eye state data and MIT-BIH has more than one channel, which indicates that LLM-ABBA might have the ability to process complicate features across channels. Table VI presents the full medical time series classification results using LLM-ABBA.

LLM-ABBA achieves comparable time series prediction results to the SOTAs, and there is no over-fitting in these tasks when using different low rank r . Because ABBA tends to symbolize trends and altitudes of the time series signals, LLM-ABBA always strengthens the vibration of predicted time series segments which can be seen in Figure 7.

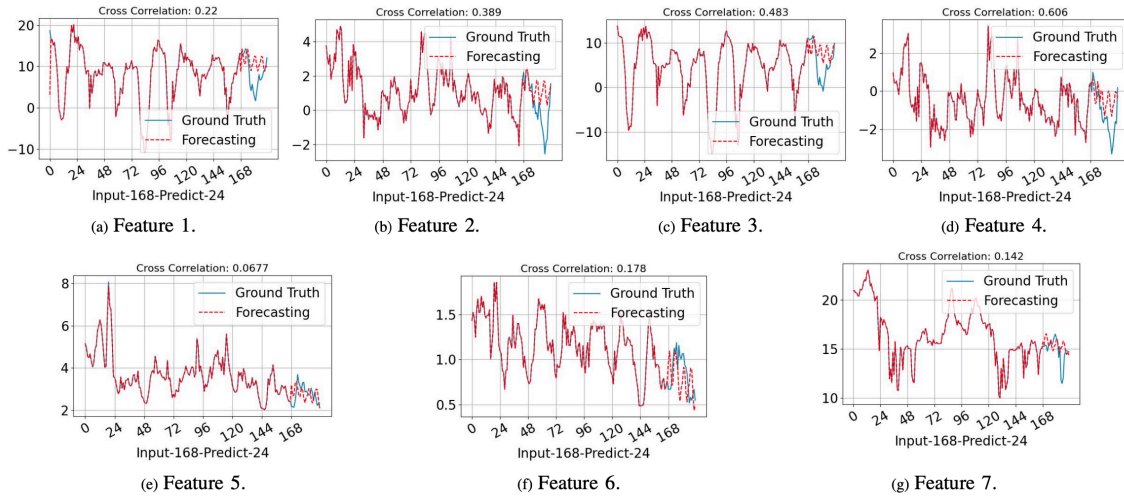


Figure 7. Visualization of input-168-predict-24 results on ETTh1 using LLM-ABBA.

G. Semantic consistency

When using pretrained tokens as the input symbols, fine-tuning on no language content (such as time series signals) will generally bring semantic loss to LLMs. Therefore, we use ASCII codes to generate new symbols by adding more digits and expanding the used alphabet table. Following the same fine-tuning process to the above experiment settings, we compute the forecasting performance by fine-tuning on Mistral-7B. Compared to Table VIII, Table IX shows that the difference is not noticeable.

Data	Predictor	Symbols	Llama2-7B						Mistral-7B						Informer ^[2]		Time-LLM ^[5]		TimeMixer ^[2]	
			r=16		r=64		r=256		r=16		r=64		r=256							
	Length	Number	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	168/24	2,789	0.689	0.653	0.647	0.696	0.658	0.677	0.631	0.681	0.622	0.631	0.626	0.677	0.577	0.549	-	-	-	-
ETTh2	168/24	5,383	0.798	0.788	0.784	0.761	0.789	0.772	0.776	0.787	0.759	0.761	0.762	0.771	0.720	0.665	-	-	-	-
ETTh1	168/24	3,170	0.403	0.397	0.386	0.364	0.392	0.385	0.457	0.422	0.401	0.387	0.407	0.397	0.323	0.369	-	-	-	-
ETTh2	168/24	6,878	0.224	0.209	0.201	0.198	0.215	0.207	0.251	0.237	0.214	0.203	0.218	0.209	-	-	-	-	-	-
ETTh1	168/96	2,789	0.762	0.786	0.754	0.752	0.759	0.60	0.792	0.804	0.773	0.782	0.7811	0.788	-	-	0.362	0.392	0.375	0.440
ETTh2	168/96	5,383	0.912	0.885	0.892	0.881	0.907	0.876	0.899	0.887	0.871	0.866	0.878	0.872	-	-	0.268	0.328	0.289	0.341
ETTh1	168/96	3,170	0.542	0.537	0.531	0.528	0.538	0.520	0.541	0.533	0.524	0.517	0.529	0.520	-	-	0.272	0.233	0.320	0.357
ETTh2	168/96	6,878	0.302	0.286	0.288	0.267	0.293	0.278	0.289	0.302	0.276	0.281	0.280	0.285	-	-	0.161	0.253	0.175	0.258
ETTh1	168/168	2789	1.161	1.010	1.087	0.964	1.096	0.989	1.182	1.217	1.174	1.968	1.179	1.992	0.931	0.752	0.398	0.418	0.429	0.421
ETTh2	168/168	5,383	4.103	2.675	3.975	2.101	4.086	2.537	4.092	2.626	3.898	2.134	3.910	2.245	3.489	1.515	0.329	0.375	0.372	0.392
ETTh1	168/168	3,170	0.989	0.962	0.974	0.952	0.979	0.959	1.001	0.986	0.966	0.958	0.972	0.966	0.678	0.614	0.310	0.358	0.361	0.381
ETTh2	168/168	6,878	0.616	0.583	0.576	0.544	0.580	0.561	0.592	0.541	0.521	0.503	0.532	0.509	-	-	0.219	0.293	0.237	0.299

Table VIII. Full comparison of results for the prediction task on 4 time series prediction datasets.

Data	Predictor	Symbols	Mistral-7B					
			r=16		r=64		r=256	
	Length	Number	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	168/24	2,789	0.636	0.692	0.626	0.632	0.629	0.681
ETTh2	168/24	5,383	0.779	0.788	0.761	0.763	0.763	0.777
ETTh1	168/24	3,170	0.457	0.402	0.402	0.387	0.407	0.399
ETTh2	168/24	6,878	0.253	0.238	0.215	0.203	0.219	0.209

Table IX. The performance of LLM-ABBA with extra new tokens (symbolic ASCII codes) on ETTh1 data in terms of time series forecasting tasks.

V. Limitations

ABBA is assessed carefully via performance profiles with respect to its reconstruction evaluated via 2-norm, DTW, and their respective differenced measures, which shows competitive performance against the SOTA STSA methods (e.g., SAX). Previous STSA methods have been applied in various data mining applications, e.g., EEG signal analysis^[42] and Internet of Things^[48]. ABBA also shows improved performance in the anomaly detection method TARZAN by simply replacing SAX methods^{[13][14]}.

LLMs can understand generated symbols of ABBA. Each data sample can be approached by symbols, and each used symbol has a specific meaning which presents one node of the internal COP of time series data. LLM-ABBA performs well not only on time series classification tasks, but also on time series regression tasks (as seen in Table V and Table VII). Because these symbolic series have a logic chain that can represent the trend of time series samples, LLMs are able to learn the trend of time series via adapter fine-tuning methods. As shown in Figure 7, by using the inverse

symbolization process of ABBA, LLMs can predict the trend of time series signals, and these predicted parts have a smaller drift. Therefore, time series forecasting tasks also can demonstrate these findings. ABBA perfectly approximates time series via symbolic series, but because LLMs are born with hallucination, more generated contents would contain more “hallucination knowledge”. Therefore, LLM-ABBA performs better on short-term time series prediction tasks and time series regression tasks.

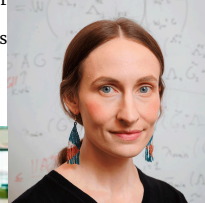
Our proposed FAPCA strategy for ABBA still cannot fully guarantee a complete removal of cumulative error arising from the previous mistaken symbols from the recovery, because an incorrect len_i might leads to a minor shift if certain symbols are replaced improperly. Additionally, as a congenital defect of LLMs, hallucination cannot be addressed in this work, and the vibration or adverse response of predicted sequences can still have negative effects on final performance. Moreover, after using ABBA to transform time series, most LLMs only can support up to 4,096 tokens, which fundamentally prohibits long-term time series analysis tasks.

VI. Conclusion

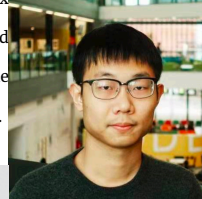
In this paper, we propose LLM-ABBA for time series classification, regression, and forecasting tasks. We discuss how to seamlessly integrate time series symbolization with LLMs and enhance its performance. In theory, we analyze the reconstruction error of ABBA symbolization, how it relates to the dominant parameters, and the congenital defect of LLM-ABBA. To mitigate the drift phenomenon of time series, we introduce the FAPCA method to improve ABBA symbolization. The empirical results demonstrate our method achieves a performance comparable to the SOTA on classification and regression tasks. In terms of convenience and universality, LLM-ABBA improves the multi-modality of LLMs on time series analysis. We believe the potential of ABBA extends to other time series applications, which will be left as future work.

About the Authors

Erin Carson is an Assistant Professor at the Department of Numerical Mathematics, Charles University. She obtained her Ph. D. at the University of California Berkeley. Her research focuses on numerical linear algebra and parallel algorithms underlying computational science, machine learning, and data science applications.



Xinye Chen is a postdoctoral researcher at Sorbonne University and a member of the NumPEX and PEQUAN team at LIP6 laboratory. Before that, He worked at Univerzita Karlova and obtained his Ph. D. in at The University of Manchester. His research interests include the scientific computing, algorithm analysis, high performance computing, and machine Learning.



Cheng Kang is PhD student at Czech Technical University. His research interests mainly focus on neuroscience, nature language processing, and medical image processing. He has published 12 scientific papers in IEEE TFS, IEEE TETCI, IEEE TNSRE, IEEE/ACM TCBB, etc., and served as a reviewer for journals and conferences, including IEEE TMI, IEEE TIP, IEEE JBHI, IEEE TFS, IEEE TETCI, etc.



Acknowledgements

The first author acknowledges funding from the European Union (ERC, inEXASCALE, 101075632), and additionally acknowledges funding from the Charles University Research Centre program No. UNCE/24/SCI/005. The second author acknowledges funding from the France 2030 NumPEX Exa-MA (ANR-22-EXNU-0002) project managed by the French National Research Agency (ANR).

Footnotes

¹ <https://github.com/nla-group/fABBA>

² <https://github.com/bitsandbytes-foundation/bitsandbytes>

³ The UCR Archive 2018 is available at https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

⁴ Monash regression data is available at <http://tseregression.org/>.

- 20

22. ²² Cao D, Jia F, Arik SO, Pfister T, Zheng Y, Ye W, Liu Y (2023). "Tempo: Prompt-based generative pre-trained transformer for time series forecasting." *arXiv preprint arXiv:2310.04948*. Available from: <https://arxiv.org/abs/2310.04948>.
23. ²³ Cleveland RB, Cleveland WS, McRae JE, Terpenning I et al. (1990). "STL: A seasonal-trend decomposition." *Journal of Official Statistics*. 6 (1): 3–73.
24. ²⁴ van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K. "WaveNet: A Generative Model for Raw Audio." In: *Proc. 9th ISCA Workshop on Speech Synthesis Workshop (SSW 9)*. 2016. p. 125.
25. ²⁵ Lloyd S. "Least squares quantization in PCM." *IEEE Transactions on Information Theory*. 28 (2): 129–137, 1982.
26. ²⁶ Drineas P, Frieze A, Kannan R, Vempala S, Vinay V (2004). "Clustering large graphs via the singular value decomposition". *Machine Learning*. 56 (1–3): 9–33.
27. ²⁷ Dasgupta S, Freund Y. "Random projection trees and low dimensional manifolds." In: *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, ser. STOC '08. ACM; 2008. p. 537–546.
28. ²⁸ Mahajan M, Nimbhorkar P, Varadarajan K (2012). "The planar k-means problem is NP-hard." *Theoretical Computer Science*. 442: 13–21. Special Issue on the Workshop on Algorithms and Computation (WALCOM 2009).
29. ²⁹ Powers DMW (1998). "Applications and explanations of Zipf's law." In: *New Methods in Language Processing and Computational Natural Language Learning*.
30. ³⁰ Kang C, Prokop J, Tong L, Zhou H, Hu Y, Novak D (2024). "InA: Inhibition adaption on pre-trained language models". *Neural Networks*. p. 106410.
31. ³¹ Devlin J, Chang M-W, Lee K, Toutanova K (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, pp. 4171–4186.
32. ³² Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. (2019). "Language models are unsupervised multitask learners." *OpenAI blog*. 1 (8): 9.
33. ³³ Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019). "RoBERTa: A robustly optimized BERT pretraining approach." *arXiv preprint arXiv:1907.11692*. Available from: <https://arxiv.org/abs/1907.11692>.
34. ³⁴ Touvron H, Lavril T, Izacard G, Martinet X, Lachaux M-A, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, et al. "LLaMA: Open and efficient foundation language models." *arXiv preprint arXiv:2302.13971*. 2023.
35. ³⁵ Jiang AQ, Sablayrolles A, Mensch A, Bamford C, Chaplot DS, Casas DdL, Bressand F, Lengyel G, Lample G, Saulnier L, et al. (2023). "Mistral 7b." *arXiv preprint arXiv:2310.06825*. Available from: <https://arxiv.org/abs/2310.06825>.
36. ³⁶ Dau HA, Bagnall A, Kamgar K, Yeh CCM, Zhu Y, Gharghabi S, Ratanamahatana CA, Keogh E (2019). "The ucr time series archive." *IEEE/CAA Journal of Automatica Sinica*. 6 (6): 1293–1305.
37. ³⁷ Seyfi A, Rajotte JF, Ng R (2022). "Generating multivariate time series with Common Source Coordinated GAN (COSCI-GAN)." *Advances in Neural Information Processing Systems*. 35: 32777–32788.
38. ³⁸ Mousavi S, Afghah F (2019). "Inter-and intra-patient ECG heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach." In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. pp. 1308–1312.
39. ³⁹ Liu Z, Zhang X (2021). "ECG-based heart arrhythmia diagnosis through attentional convolutional neural networks." In: *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTIS)*. IEEE. pp. 156–162.
40. ⁴⁰ Yang C-HH, Tsai YY, Chen PY (2021). "Voice2series: Reprogramming acoustic models for time series classification." In: *International Conference on Machine Learning*. PMLR. pp. 11808–11819.
41. ⁴¹ Kachuee M, Fazeli S, Sarrafzadeh M (2018). "ECG heartbeat classification: A deep transferable representation." *IEEE International Conference on Healthcare Informatics*. pp. 443–444.
42. ⁴² Shashikumar SP, Shah AJ, Clifford GD, Nemati S (2018). "Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks." In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. pp. 715–723.
43. ⁴³ Singh S, Pandey SK, Pawar U, Janghel RR (2018). "Classification of ECG arrhythmia using recurrent neural networks." *Procedia Computer Science*. 132: 1290–1297.
44. ⁴⁴ Saadatnejad S, Oveis M, Hashemi M (2019). "LSTM-based ECG classification for continuous monitoring on personal wearable devices". *IEEE Journal of Biomedical and Health Informatics*. 24 (2): 515–523.

45. [△]Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W (2021). "Informer: Beyond efficient transformer for long sequence time-series forecasting." *Proceedings of the AAAI Conference on Artificial Intelligence*. 35 (12): 11106–11115.
46. [△]Wu H, Xu J, Wang J, Long M (2021). "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting." *Advances in Neural Information Processing Systems*. 34: 22419–22430.
47. [△]Jain D, Ranjan R, Sharma A, Sharma SN, Jain A (2024). "Fast and accurate ECG signal peaks detection using symbolic aggregate approximation". *Multi media Tools and Applications*. 83 (30): 75033–75059.
48. [△]Jha V, Tripathi P (2024). "Probabilistic SAX: A cognitively-inspired method for time series classification in cognitive IoT sensor network." *Mobile Networks and Applications*.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.