Fast Retrieval of Large Hydroelectric Interaction Models Based on Long-Sequence Temporal Embeddings

Authors: Fu-Hai Chen, Zhang Feng, Chen Fuhai

Date: 2025-08-14T00:00:00+00:00

Abstract

Safety early warning and optimal dispatching of hydropower systems rely on efficient analysis of long-term high-precision time-series monitoring data (e.g., minute-level dam water level, inflow rate, etc.). However, existing monitoring modes still primarily rely on manual retrieval, suffering from issues such as response lag (missing critical event handling windows) and insufficient accuracy (relying on empirical judgment). Moreover, traditional time-series embedding models (e.g., TS2Vec) are constrained by fixed context windows of limited scope, making it difficult to capture long-cycle characteristics of hydropower systems, resulting in low recall rates for critical events such as rainstorms and failing to meet intelligent monitoring requirements. To address this, this paper proposes HTE, a long-sequence time-series embedding retrieval framework that integrates positional encoding, aiming to construct a large-model-based hydropower interactive data monitoring system that enables efficient and precise data retrieval through natural language dialogue. The core designs of this framework include: (1) constructing the first hydropower time-series retrieval benchmark dataset, HydroT-Bench, encompassing 12k sequence-query pairs (11,293 multi-parameter time-series data and 5,000 engineering-level queries); (2) innovating a hydropower cycle-aware dynamic NTK interpolation mechanism, combined with Rotary Position Embedding (RoPE), to extend the context window to 32k points, balancing long-cycle feature capture with local detail preservation; (3) designing a text-time-series cross-modal embedding space to support direct association between natural language queries and long time-series data. Experimental results demonstrate that: in 32k sequence retrieval, the critical event Recall@10 reaches 91%, representing a 68% improvement over TS2Vec; the RMSE for long-cycle pattern matching decreases to 0.13; and the inference latency for a single full-year 1425-point sequence is less than 100 seconds. This framework provides minute-level and even second-level retrieval capabilities for

intelligent hydropower monitoring, and the open-source dataset and pre-trained models can support research in the field.

Full Text

Fast Retrieval for Hydropower Interactive Large Models Based on Long-Sequence Temporal Embedding

Cao Chunlan[1], Zheng Zhiwen[1], Zhang Feng[2], Liu Jiarong[2], Jiang Hao[3], Zheng Wenhao[2], Chen Fuhai[2*]

Abstract

The safety early warning and optimal scheduling of hydropower systems depend on efficient analysis of long-term, high-precision temporal monitoring data (e.g., minute-level dam water levels, inflow rates, etc.). However, existing monitoring modes still rely primarily on manual retrieval, suffering from response lag (missing critical event handling windows) and insufficient accuracy (dependence on empirical judgment). Traditional temporal embedding models such as TS2Vec are constrained by fixed context windows of limited scope, making it difficult to capture the long-period characteristics of hydropower systems and resulting in low recall rates for critical events like rainstorms, which fails to meet intelligent monitoring requirements. To address this, we propose HTE, a long-sequence temporal embedding retrieval framework that fuses positional encoding, aiming to construct a hydropower interactive data monitoring system based on large models that enables efficient and precise data retrieval through natural language dialogue. The core designs of this framework include: (1) constructing the first hydropower temporal retrieval benchmark dataset, HydroT-Bench, encompassing 12k sequence-query pairs (11,293 multi-parameter temporal sequences and 5,000 engineering-level queries); (2) innovating a hydropower period-aware dynamic NTK interpolation mechanism combined with Rotary Positional Encoding (RoPE) to extend the context window to 32k points, balancing long-period feature capture with local detail preservation; and (3) designing a text-temporal cross-modal embedding space that supports direct association between natural language queries and long temporal data. Experimental results demonstrate that in 32k sequence retrieval, the critical event Recall@10 reaches 91%, representing a 68% improvement over TS2Vec; long-period pattern matching RMSE drops to 0.13; and inference latency for a single annual sequence of 1,425 points is less than 100 seconds. This framework provides minute-level and even second-level retrieval capabilities for intelligent hydropower monitoring, with the open-source dataset and pre-trained models offering support for domain research.

Keywords: Large Models; Temporal Embedding; Hydropower Data Monitoring

1 Introduction

Hydropower engineering serves as a core pillar of the clean energy system, and its safe and stable operation directly impacts watershed flood control, energy supply, and ecological protection. Intelligent monitoring of such engineering heavily depends on long-term, high-frequency, multi-dimensional temporal monitoring data—from dam water levels and inflow rates updated every minute to real-time changes in power generation flow and spillage discharge. These data constitute the "nerve endings" for perceiving hydropower system status. For instance, a sudden increase in inflow rate may indicate flood risks from upstream rainstorms, while minute fluctuations in dam water levels may imply structural safety hazards. Capturing and interpreting these signals requires precise correlation analysis of continuous temporal data spanning months or even years.

However, current hydropower monitoring still predominantly relies on traditional manual modes: dispatchers must manually screen and compare massive temporal curves across multiple terminal systems, judging abnormal patterns through experience. This approach has significant limitations. On one hand, a single hydropower station generates tens of thousands of temporal data points daily, making manual processing difficult to achieve minute-level or even second-level response, easily missing the golden handling window for sudden critical events like rainstorms or equipment failures. On the other hand, manual analysis depends on individual experience, resulting in inconsistent capture of cross-period correlation features (such as the implicit coupling between wet-season spillage and dry-season power generation), which severely constrains the intelligent upgrading of hydropower systems.

To address these pain points, this paper aims to construct a hydropower interactive data monitoring system based on large models, achieving an efficient retrieval mode of "replacing operations with dialogue" through cross-modal interaction between natural language and temporal data. Specifically, users can directly locate target information from massive temporal data through natural language queries such as "Dates in the 2023 flood season with daily maximum inflow exceeding 500m^3/s" or "Time periods where upstream water level rose continuously for 3 hours." This mode breaks the passive situation of "people adapting to the system" in traditional monitoring, enabling the system to actively understand human intentions, significantly improving the efficiency and accuracy of anomaly identification and trend prediction, and providing "ask-and-answer" intelligent support for dispatching decisions.

The core challenge in achieving this goal lies in efficient retrieval and key information extraction from long-sequence temporal data. Hydropower system temporal data exhibits distinct characteristics of "long-period, strong coupling, and scattered events."In terms of period, it contains multi-scale patterns ranging from intra-day (e.g., power generation flow fluctuations with electricity load), intra-month (e.g., reservoir storage and release regulation), intra-year (e.g., wet and dry season alternation), to inter-annual scales. Regarding critical event distribution, important signals such as water level surges caused by rainstorms or flow mutations caused by equipment failures are often scattered across long sequences of tens of thousands of points, with low distinguishability from normal fluctuations.

Existing temporal embedding models struggle to handle such scenarios. Taking the typical model TS2Vec as an example, its fixed 512-point context window can only cover about 8 hours of minute-level data, making it impossible to capture cross-month or even cross-season long-period correlations, resulting in a retrieval recall rate of only 28% for critical events like rainstorms—numerous abnormal signals hidden in the middle of long sequences are missed due to being "outside the window" [**?**]. Other methods such as piecewise aggregation can extend processing length but lose local details, while sparse attention mechanisms struggle to balance computational efficiency and feature retention, all failing to meet the dual demands of hydropower monitoring for "long-period coverage + fine-grained identification."

To break through this bottleneck, this paper proposes the first long-sequence temporal embedding retrieval framework for the hydropower domain (Hydro-Temporal Embedding, HTE). The framework' s core design philosophy is "extending context windows, combining period-aware dynamic NTK interpolation with RoPE positional encoding, and achieving cross-modal interaction." To address data scale and retrieval requirements, we construct the first hydropower temporal retrieval benchmark dataset, HydroT-Bench, covering 11,293 multi-parameter long temporal sequences and 5,000 engineering-level queries, providing standardized support for model training and evaluation. We innovatively introduce a period-aware dynamic NTK interpolation mechanism combined with Rotary Positional Embeddings (RoPE) to extend the context window from 512 points to 32k points, capturing both inter-annual long-period patterns and local features of minute-level data. Through text-temporal cross-modal embedding design, we enable direct association between natural language queries and long temporal data, allowing engineering problems such as "retrieving spillage flow peaks in a certain period" or "comparing power generation efficiency for the same period across two years" to be quickly solved through conversational interaction.

The core contributions of this study are fourfold: (1) **Pioneering a new domain problem**: We are the first to apply long-sequence temporal embedding retrieval technology to hydropower monitoring scenarios, explicitly defining the technical requirements for "natural language-temporal data" cross-modal interaction, filling a domain gap; (2) **Constructing a benchmark dataset**: We release HydroT-Bench, containing 12k sequence-query pairs covering 8 core monitoring parameters and 8 types of engineering query templates, providing a reusable evaluation benchmark for subsequent research; (3) **Proposing the innovative HTE framework**: By fusing dynamic NTK interpolation with RoPE encoding, we break through context limitations in long-sequence retrieval, enabling efficient processing of 32k-point data; and (4) **Verifying practical effectiveness**: Experiments show that the framework achieves a critical event retrieval Recall@10 of 91% (a 68% improvement over TS2Vec), reduces long-period pattern matching RMSE to 0.13, and maintains inference latency for a single annual sequence under 100 seconds, fully demonstrating its practicality in industrial scenarios.

2 Related Work

2.1 Time Series Representation Learning

Time series representation learning aims to transform raw temporal data into low-dimensional, information-rich vector representations. General temporal embedding models such as TS2Vec [?] generate segment-level embeddings through hierarchical contrastive learning, but their maximum context window is limited to 512 points, making it difficult to capture important dependencies spanning thousands of points in ultra-long sequences and unable to model long-period dependencies exceeding 8k points in hydropower scenarios (such as multi-year water level-power generation correlations). Similarly, TS-TCC [?] is based on temporal consistency contrast, and CoST [?] combines frequency domain decomposition to improve robustness, but both are constrained by inherent short-context modeling mechanisms and have not broken through the short-context bottleneck. Attempts for long sequences include: piecewise aggregation methods [?] compress sequences for dimensionality reduction but lose critical temporal details and local patterns; sparse attention mechanisms like Longformer [?] significantly reduce computational complexity, but when processing ultra-long sequences exceeding 4k points, their sparse patterns still struggle to effectively maintain the ability to capture fine-grained, localized features.

2.2 Long-Context Position Encoding Technology

Traditional absolute position encoding such as Sinusoidal [?] suffers from severe extrapolation degradation (positional representation distortion beyond training length). Relative position encoding schemes like T5 Bias [?] and ALiBi [?] improve extrapolation by introducing bias terms proportional to relative distance, but their preset static distance decay patterns are insufficiently adaptable to periodic temporal fluctuations such as daily/quarterly cycles. Rotary Position Embedding (RoPE) [?] introduces rotation matrices to maintain linear dependencies of relative positional relationships and has been verified to possess excellent extrapolation capabilities in Transformer-based language models, but its effectiveness in long-range modeling of complex multivariate industrial time series dominated by periodic patterns still lacks systematic verification.

2.3 Context Window Extension Methods

Positional Interpolation (PI) [?] achieves window extension by compressing position indices but introduces high-frequency information loss. NTK-aware Scaling [?] dynamically adjusts base frequencies based on neural tangent kernel theory, significantly outperforming fixed interpolation, but its scaling factor is a static hyperparameter that cannot adapt to the time-varying periodic demands determined by wet and dry seasons in hydropower data. Such methods have not been validated in hydropower monitoring scenarios. PI uniformly compresses position indices to achieve window extension but introduces significant high-frequency component loss due to excessive downsampling. NTK-aware Scaling extends frequency domain distribution based on neural tangent kernel theory, demonstrating significantly superior performance compared to fixed interpolation; however, its preset global scaling factor is a static hyperparameter that struggles to adapt to the dynamically changing periodic demands driven by strong seasonality and determined by wet and dry seasons in hydropower data. The specific effectiveness of such advanced context extension methods in complex hydropower monitoring scenarios remains unverified.

2.4 Hydropower Time Series Intelligent Analysis

Traditional methods rely on statistical features such as mean and variance and machine learning models like LSTM [**?**] and TCN [**?**] for prediction or anomaly detection, but they mostly involve single-task independent modeling (e.g., some literature only predicts water level [**?**]), lacking end-to-end long-sequence semantic retrieval capabilities. More critically, the hydropower domain lacks publicly available large-scale retrieval datasets, with existing work only validated on private small data, constraining algorithm reproducibility and fairness. Our research positioning is shown in Table 1 . Existing work has three major limitations: (1) **Model limitations**: General embedding models like TS2Vec are constrained by 512-point windows, while long-sequence methods like NTK interpolation ignore hydropower periodic characteristics; (2) **Task limitations**: Hydropower research focuses on prediction/classification without dedicated retrieval frameworks; and (3) **Data limitations**: No publicly available long-sequence retrieval benchmarks exist. HTE pioneers the fusion of dynamic NTK interpolation with RoPE encoding and constructs the first hydropower retrieval dataset, filling these gaps.

3 Methodology

The core workflow of the HTE framework consists of three interconnected components, as shown in Figure 1 [Figure 1: see original paper]. First, the input module receives multivariate long temporal hydrological data, typically collected by various sensors in watersheds or water conservancy projects, forming a set of time series signals with long-term dependencies. Typical variables include but are not limited to: upstream water level (reflecting inflow storage conditions), downstream water level (reflecting river conveyance capacity or backwater effects), inflow rate (total water volume entering the reservoir), outflow rate (total water volume released from the reservoir, possibly including flood discharge and water transfer), power generation flow (flow specifically used for turbine power generation), etc. These multi-source heterogeneous but highly correlated time series are then fed into the temporal adaptive encoder module, which is the core processor of the framework. It uses an adaptive encoder to learn the complex spatiotemporal dynamic associations and nonlinear features among multivariate sequences with different frequencies, magnitudes, and change patterns, extracting high-dimensional, information-rich hidden states. Finally, the features refined by the adaptive encoder are mapped and output to the unified semantic embedding space module. This space serves as a shared, low-dimensional vector representation environment that generates dense, semantically rich embedding vectors for all input variables and their interactions. These embeddings not only capture each variable' s own historical evolution patterns but, more importantly, encode their cross-variable, cross-timestep intrinsic relationships and hydrological physical meanings, providing a powerful, unified feature foundation for downstream tasks such as forecasting, anomaly detection, and pattern recognition.

3.1 HTE Overall Framework

Leveraging the strong semantic understanding capability of the text embedding model E5, we introduce RoPE positional encoding adapted to hydropower temporal features to construct a cross-modal retrieval bridge from text to temporal data. The text encoder adopts the E5-base pre-trained model (12-layer Transformer) as the backbone. At the input layer, natural language queries q such as "Maximum inflow rate in July 2023" are converted into token sequences. We replace the original positional encoding with the RoPE basic framework (Equation (1)) and integrate hydropower period-aware dynamic NTK interpolation to achieve window extension, enhancing positional awareness for long queries exceeding 512 tokens. A projection network is added at the temporal alignment layer to map text embeddings into the temporal space.

The optimization objective is to minimize the cosine distance between query embeddings and positive sample temporal embeddings (Equation (2)), where the positive sample temporal data vector represents the correctly retrieved temporal data embedding vector. In the HTE framework, E5-RoPE serves as the universal encoder for text queries, providing embedding representations for hydropower semantic queries.

3.2 Parallel Context Windows (PCW)

For hydropower data, the substantial volume accumulated over long periods and multiple ranges requires system-level data segmentation before feeding into large models. We adopt the Parallel Context Windows (PCW) method on structured data to extend the embedding model' s context window without training, enabling effective data segmentation. When processing this "long text" data, PCW divides long documents into multiple short segments, processes each segment in parallel, and finally aggregates their results. Specifically, document D is split into segments of length Lo tokens, and the embedding vectors of each segment are averaged to represent document D. For simplicity, we set the overlap between adjacent segments to 0, except for the last segment to ensure it contains Lo tokens.

The core advantage of this context window extension strategy lies in achieving a leap in long text processing capability at near-zero cost: it requires no retraining or fine-tuning of the model. Instead, it directly reuses existing weights through algorithmic innovation (such as position reorganization, interpolation, or block parallelism) to extend the context window several-fold (e.g., 512 tokens → 32K), significantly saving the massive computational resources (e.g., hundred-level GPU clusters) and time costs (weeks/months) required for training. Simultaneously, it maintains the model' s original performance on short texts and is compatible with efficient inference.

3.3 Position ID Reorganization

Segmenting input into blocks and processing them separately sacrifices interaction between blocks. In contrast, position reorganization accommodates longer contexts by reusing original position identifiers. Specifically, we experimented with two simple strategies: grouped position and cyclic position. The former groups original position identifiers as fgp(pid)→, while the latter cyclically assigns position identifiers according to the formula: frp(pid)→(pid mod pid/s).

3.4 Period-Aware Dynamic NTK Interpolation

Addressing the defect that existing window extension methods (such as NTK-aware Scaling [?]) cannot adapt to hydropower's time-varying periodicity due to static scaling factors, this paper proposes a dynamic NTK interpolation mechanism. The hydropower period parameter is extracted from the training set through Fourier transform (e.g., typical wet/dry season periods are 180/365 days), with learnable parameters α,β optimized through backpropagation. Initial values are set according to the "Hydrological Forecasting Specification" (GB/T 22482) [?] as α=1.2, β=0.8, consistent with typical reservoir operation curve characteristics.

The RoPE rotation angle base frequency (Equation (4)) achieves adaptive frequency adjustment when extending the window to 32k by combining the dynamic NTK interpolation of Equation (3). This design inherits RoPE' s extrapolation advantages [?] while introducing hydropower prior knowledge, significantly mitigating high-frequency information loss during interpolation, with comparative experiments shown in Section 4.2. NTK-aware interpolation significantly improves long text retrieval accuracy by non-uniformly adjusting the frequency distribution of positional encoding: traditional linear interpolation (PI) compresses all positional frequencies proportionally, causing high-frequency detail loss (e.g., keyword position offset), while NTK specifically reduces compression amplitude in low-frequency dimensions (preserving long-range dependencies) and simultaneously reduces scaling ratio in high-frequency dimensions (maintaining local position precision), thereby balancing global and local position resolution in the RoPE architecture. For example, at 32k context, E5-Mistral+NTK achieves 93.8% Passkey retrieval accuracy (4% improvement over PI), with critical information capture capability approaching original short text levels, and zero training cost for maintaining long document semantic coherence.

4 Experiments

4.1 HydroT-Bench Dataset Construction

To fill the gap in long-sequence retrieval benchmarks for the hydropower domain, this paper constructs the first dedicated dataset, HydroT-Bench. Its design follows two principles:

Multi-parameter long temporal coverage: We collected 11,293 temporal sequences from 3 hydropower stations spanning October 30, 2016 to July 22, 2024 (15-minute intervals), covering 8 parameters: upstream water level, downstream water level, inflow rate, outflow rate, power generation flow, spillage discharge, total plant output, and areal rainfall. The time span is 2016–2024, with average sequence length exceeding the general BEIR dataset by 40 times.

Engineering semantic query generalization: We defined 8 query templates as shown in Table 2 to generate 5,000 natural language queries. For example, the cross-category comparison template "{date}' s {category1} and {category2} which has higher value at hour {hour}?" forces the model to learn multivariate physical coupling relationships, breaking through traditional single-task prediction limitations [**?**][**?**].

4.2 Experimental Setup

Experiments were conducted on the HydroT-Bench dataset with 12k sequence-query pairs. The training set adopted 9,035 long temporal samples from 2016–2023, while the test set contained 2,258 temporal sequences and 1,000 engineering-level queries from 2024. Core evaluation metrics include: critical event detection capability (Recall@10 and F1@10), long-sequence similarity matching accuracy (0-1 normalized RMSE), and inference latency for processing a single 1,425-point sequence on NVIDIA V100 GPU, comprehensively validating model detection accuracy, matching quality, and real-time performance in industrial scenarios.

4.3 Results and Analysis

Analysis of Table 3 reveals that HTE significantly improves critical event detection performance, achieving the highest F1@10 and Recall@10 metrics. At the efficiency level, when processing 32k-point long sequences, memory usage increases by only 23%, far lower than Performer' s 105% increase. HTE innovatively achieves stable compression of annual data 1,425-point retrieval latency to under 100 seconds—through segmented parallel computation (8,192 points per segment) fused with the FlashAttention-2 acceleration engine, reducing latency by 40% compared to TS2Vec' s traditional segmented serial strategy, breakthroughly solving the real-time bottleneck of industrial-level long-sequence retrieval.

Analysis of Table 4 shows that using static NTK interpolation only yields marginal improvements in F1@10 and Recall@10 metrics, while the hydropower period-aware dynamic NTK interpolation technology significantly enhances critical event detection performance, improving F1@10 and Recall@10 by 33% and 9.6%, respectively.

4.4 Multimodal Retrieval Performance

Based on the query-type-specific test results in Table 5 , we derive the following analysis: The HTE model demonstrates significantly leading Recall@10% performance in three task types–single-hour value (92.9%), extreme value localization (93.1%), and trend analysis (91.7%)–outperforming fine-tuned E5-RoPE by 10.8%, 7.5%, and 13.4%, respectively, highlighting its advantages in hydropower temporal scenario modeling. However, in cross-category comparison tasks, HTE (69.4%) shows narrowed improvement margins compared to previous items, only exceeding fine-tuned E5-RoPE by 3.2%, indicating that cross-dimensional correlation analysis remains a common challenge for all models. This demonstrates that HTE further breaks through performance bottlenecks in professional scenarios through temporal optimization.

5 Conclusion

This paper proposes the HTE long-sequence temporal embedding retrieval framework, which breakthroughly extends the temporal embedding context window to 32k points by fusing RoPE positional encoding with hydropower period-aware NTK interpolation technology, solving the bottleneck of long-period feature modeling in hydropower systems. Empirical evidence shows that the framework achieves 91% critical event recall in 32k sequence retrieval, a 68% improvement over TS2Vec; reduces long-period pattern matching RMSE to 0.13, decreasing error magnitude by 66%; and maintains inference latency for a single 1,425-point annual sequence under 100 seconds. This study open-sources the first hydropower temporal retrieval dataset HydroT-Bench and pre-trained models, covering 8 years (2016–2024) of data across 8 parameters (upstream water level, downstream water level, inflow rate, outflow rate, power generation flow, spillage discharge, total plant output, areal rainfall), providing minute-level and even second-level long-sequence retrieval capabilities for hydropower systems. This work drives the transformation of intelligent monitoring paradigms from passive response to active prediction. Future work will expand to meteorological and video multimodal fusion retrieval and lightweight deployment on edge devices.

References

[1] Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., & Xu, B. (2022). TS2Vec: Towards Universal Representation of Time Series. Proceedings of the AAAI Conference on Artificial Intelligence, 36(4), 8980–8987.

[2] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C. K., Li, X., & Guan, C. (2021). Time-Series Representation Learning via Temporal and Contextual Contrasting. 2021 International Joint Conference on Neural Networks (IJCNN), 1–8.

[3] Woo, G., Liu, C., Sahoo, D., Kumar, A., & Hoi, S. (2022). CoST: Contrastive Learning Disentangled Seasonal-Trend Representations for Time Series Forecasting. International Conference on Learning Representations (ICLR).

[4] Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra, S. (2001). Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowledge and Information Systems, 3(3).

[5] Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The Long-Document Transformer. arXiv preprint arXiv:2004.05150.

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ⋯& Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008.

[7] Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., ⋯& Embeddings by Weakly-Supervised Contrastive Pre-training. arXiv preprint arXiv:2212.03533.

[8] Press, O., Smith, N. A., & Lewis, M. (2021). Train Short, Test Long: Attention with Linear Biases Enables Input Length Extrapolation. arXiv preprint arXiv:2108.12409.

[9] Su, J., Lu, Y., Pan, S., Murtadha, A., Wen, B., & Liu, Y. (2021). RoFormer: Enhanced Transformer with Rotary Position Embedding. arXiv preprint arXiv:2104.09864.

[10] Chen, S., Wong, S., Chen, L., & Tian, Y. (2023). Extending Context Window of Large Language Models via Positional Interpolation. arXiv preprint arXiv:2306.15595.

[11] Peng, B., & Quesnelle, J. (2023). NTK-Aware Scaled RoPE allows LLaMA models to have extended (8k+) context size without any fine-tuning and minimal perplexity. arXiv preprint arXiv:2306.15595.

[12] Zhang, J., Zhu, Y., Zhang, W., & Ye, X. (2020). Water level prediction based on LSTM for hydropower station. Journal of Physics: Conference Series, 1634(1), 012093.

[13] Li, X., Li, Z., & Wang, H. (2021). Short-term load forecasting of hydropower based on TCN. Journal of Physics: Conference Series, 2005(1), 012021.

[14] Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., & Gurevych, I. (2021). BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. Proceedings of the 44th International ACM SIGIR Conference.

[15] GB/T 22482–2008, Hydrological Information Forecasting Specification [S]. Beijing: China Standards Press, 2008.

Note: Figure translations are in progress. See original paper for figures.