Novelty Assessment Report

Paper: MesaNet: Sequence Modeling by Locally Optimal Test-Time Training

PDF URL: https://openreview.net/pdf?id=xa3OnTb6c3

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-05

Abstract

Sequence modeling is currently dominated by causal transformer architectures that use softmax self-attention. Although widely adopted, transformers require scaling memory and compute linearly during inference. A recent stream of work linearized the softmax operation, resulting in powerful recurrent neural network (RNN) models with constant memory and compute costs such as DeltaNet, Mamba or xLSTM. These models can be unified by noting that their recurrent layer dynamics can all be derived from an in-context regression objective, approximately optimized through an online learning rule. Here, we join this line of work and introduce a numerically stable, chunkwise parallelizable version of the recently proposed Mesa layer (von Oswald et al., 2024), and study it in language modeling at the billion-parameter scale. This layer again stems from an in-context loss, but which is now minimized to optimality at every time point using a fast conjugate gradient solver. Through an extensive suite of experiments, we show that optimal test-time training enables reaching lower language modeling perplexity and higher downstream benchmark performance than previous RNNs, especially on tasks requiring long context understanding. This performance gain comes at the cost of additional flops spent during inference time. Our results are therefore intriguingly related to recent trends of increasing test-time compute to improve performance -- here by spending compute to solve sequential optimization problems within the neural network itself.

Core Task Landscape

This paper addresses: **Sequence Modeling with Efficient Recurrent Neural Networks**

A total of **50 papers** were analyzed and organized into a taxonomy with **20 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **RNN Architecture Design and Variants**
- **Efficiency and Scalability Optimization**
- **Training and Optimization Methodologies**
- **Long-Range Dependency Modeling**
- **Transformer-RNN Comparisons and Sequence-to-Sequence Models**
- **Application Domains**
- **Survey and Review Literature**
- **Frameworks and Tools**

Complete Taxonomy Tree

- Sequence Modeling with Efficient Recurrent Neural Networks Survey Taxonomy
- RNN Architecture Design and Variants
  - Gating Mechanisms and Memory Structures (7 papers)
  - [1] Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications (Ibomoiye Domor Mienye, 2024) View paper
  - [2] Sequence Modeling using Gated Recurrent Neural Networks (Pezeshki Mohammad, 2022) View paper
  - [3] Recurrent neural networks (RNNs): architectures, training tricks, and introduction to influential research (Susmita Das, 2023) View paper
  - [8] Long short-term memory recurrent neural network architectures for large scale acoustic modeling. (HaÅim Sak, 2014) View paper
  - [10] A review of recurrent neural network architecture for sequence learning: Comparison between LSTM and GRU (Shiva Nosouhian, 2021) View paper
  - [14] Hierarchically gated recurrent neural network for sequence modeling (Qin Zhen, 2023) View paper
  - [34] Griffin: Mixing Gated Linear Recurrences with Local Attention for Efficient Language Models (De, 2024) View paper
  - Alternative RNN Formulations (3 papers)
  - [4] Quantum recurrent neural networks for sequential learning (Yanan Li, 2023) View paper
  - [9] Symplectic Recurrent Neural Networks (Zhengdao Chen, 2022) View paper
  - [15] Weight-Space Linear Recurrent Neural Networks (Keshtmand, 2025) View paper
  - Hybrid Architectures (3 papers)
  - [6] Structured sequence modeling with graph convolutional recurrent networks (Seo, 2018) View paper
  - [29] Recurrent Attentive Neural Process for Sequential Data (Qin, 2022) View paper
  - [43] Recurrent Convolutional Neural Network for Sequential Recommendation (Chengfeng Xu, 2019) View paper
- Efficiency and Scalability Optimization
  - Model Compression and Sparsity (2 papers)
  - [26] LightRNN: Memory and computation-efficient recurrent neural networks (Xiang Li, 2016) View paper

- [37] Efficient and effective training of sparse recurrent neural networks (Shiwei Liu, 2021) View paper
  - Hardware Acceleration and Implementation (2 papers)
  - [18] E-RNN: Design optimization for efficient recurrent neural networks in FPGAs (Li Zhe, 2019) View paper
  - [50] ReRAM-based processing-in-memory architecture for recurrent neural network acceleration (Yun Long, 2018) View paper
  - Efficient Sequence Processing Strategies ★ (5 papers)
  - [0] MesaNet: Sequence Modeling by Locally Optimal Test-Time Training (Anon et al., 2026) View paper
  - [27] Dynamic beam width tuning for energy-efficient recurrent neural networks (Daniele Jahier Pagliari, 2019) View paper
  - [38] Segrnn: Segment recurrent neural network for long-term time series forecasting (Lin Shengsheng, 2023) View paper
  - [45] Dual-path rnn: efficient long sequence modeling for time-domain single-channel speech separation (Yi Luo, 2020) View paper
  - [46] Efficient sequence learning with group recurrent networks (Fei Gao, 2018) View paper
- Training and Optimization Methodologies
  - Regularization and Generalization (2 papers)
  - [11] Recurrent Neural Network Regularization (Zaremba, 2022) View paper
  - [33] Residual recurrent neural networks for learning sequential representations (Boxuan Yue, 2018) View paper
  - Distributed and Privacy-Preserving Training (1 papers)
  - [21] Tensor Recurrent Neural Network With Differential Privacy (Jun Feng, 2024) View paper
- Long-Range Dependency Modeling
  - State Space Models and Linear RNNs (2 papers)
  - [39] RWKV: Reinventing RNNs for the Transformer Era (Bo Peng, 2023) View paper
  - [42] Efficiently modeling long sequences with structured state spaces (Gu, 2021) View paper
  - Enhanced RNN Architectures for Long Sequences (4 papers)
  - [12] Multi-time-scale with clockwork recurrent neural network modeling for sequential recommendation (Nana Huang, 2025) View paper
  - [22] mLANet: An efficient recurrent neural network for long-term time series forecasting (Jihua Jiang, 2025) View paper
  - [23] Rwkv-ts: Beyond traditional recurrent neural network for time series tasks (Hou Haowen, 2024) View paper
  - [49] Resurrecting Recurrent Neural Networks for Long Sequences (Orvieto, 2023) View paper
- Transformer-RNN Comparisons and Sequence-to-Sequence Models (2 papers)
  - [5] A survey on recurrent neural network architectures for sequential learning (B. Shiva Prakash, 2018) View paper
  - [7] Sequence to Sequence Learning with Neural Networks (Ilya Sutskever, 2022) View paper
- Application Domains
  - Natural Language Processing and Speech (3 papers)
  - [16] Phone sequence modeling with recurrent neural networks (Nicolas Boulanger-Lewandowski, 2014) View paper
  - [25] Pixel Recurrent Neural Networks (Aaron van den Oord, 2022) View paper
  - [30] Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding. (GrÃ©goire Mesnil, 2013) View paper
  - Time Series Forecasting and Financial Modeling (3 papers)
  - [13] TRNN: An efficient time-series recurrent neural network for stock price prediction (M. Lu, 2024) View paper
  - [17] Investigating explainablity methods in recurrent neural network architectures for financial time-series data (Freeborough, 2022) View paper
  - [31] Investigating Explainability Methods in Recurrent Neural Network Architectures for Financial Time Series Data (Warren Freeborough, 2022) View paper
  - Sequential Recommendation Systems (3 papers)
  - [36] MV-RNN: A multi-view recurrent neural network for sequential recommendation (Qiang Cui, 2018) View paper
  - [41] Recurrent Attentive Neural Networks for Sequential Recommendation (Lei Tan, 2023) View paper
  - [47] CALRec: Contrastive Alignment of Generative LLMs for Sequential Recommendation (Yaoyiran Li, 2024) View paper
  - Biomedical and Healthcare Applications (2 papers)
  - [20] Recurrent neural network architectures for analysing biomedical data sets (Mohammed Khalaf, 2017) View paper
  - [24] Predicting the risk of heart failure with EHR sequential data modeling (Bo Jin, 2018) View paper
  - Security and Anomaly Detection (2 papers)
  - [19] Modeling an intrusion detection using recurrent neural networks (M. Ibrahim, 2023) View paper
  - [32] Recurrent Neural Network based Incremental model for Intrusion Detection System in IoT (H. Sharma, 2024) View paper
  - Computer Vision and Motion Modeling (1 papers)
  - [48] Anticipating in Vehicle Accident using Recurrent Neural Network (Waqar Riaz, 2022) View paper
  - Specialized Domain Applications (1 papers)
  - [40] Recurrent Neural Network Model for Constructive Peptide Design (Alex T. MÃ¼ller, 2018) View paper
- Survey and Review Literature (2 papers)
  - [28] Recurrent neural network architectures (F. Bianchi, 2017) View paper
  - [44] Recurrent neural network (Charles Chen, 2021) View paper
- Frameworks and Tools (1 papers)
  - [35] Deep-sdm: A unified computational framework for sequential data modeling using deep learning models (Nawa Raj Pokhrel, 2024) View paper

Narrative

Core task: sequence modeling with efficient recurrent neural networks. The field encompasses a broad spectrum of research directions organized around making RNNs more practical and powerful. At the highest level, the taxonomy distinguishes between foundational architecture design (exploring gating mechanisms, novel cell structures, and hybrid models), efficiency and scalability optimization (addressing computational bottlenecks through pruning, quantization, and hardware-aware strategies), training and optimization methodologies (tackling gradient issues and regularization), long-range dependency modeling (capturing temporal patterns over extended horizons), comparisons with Transformers and sequence-to-sequence frameworks, diverse application domains (from speech recognition to biomedical analytics), and supporting survey literature and tooling. Representative works such as RNN Comprehensive Review[1] and RNN Survey[5] provide overviews of architectural evolution, while studies like Gated Recurrent Networks[2] and RNN Architectures Training[3] delve into specific design and training challenges.

Several active lines of work reveal contrasting priorities and open questions. One thread focuses on architectural innovation to balance expressiveness and efficiency, exemplified by hierarchical gating schemes (Hierarchically Gated RNN[14]) and modular designs (Group

Recurrent Networks[46]). Another emphasizes direct efficiency gains through segmentation and streamlined processing, as seen in SegRNN[38] and Dual Path RNN[45], which reduce redundant computation without sacrificing accuracy. MesaNet[0] sits within this efficiency-oriented cluster, proposing strategies for efficient sequence processing that align closely with works like SegRNN[38] and Dynamic Beam Width[27]. Compared to SegRNN[38], which segments inputs to simplify recurrence, MesaNet[0] appears to explore complementary mechanisms for accelerating inference or training. Meanwhile, efforts such as RWKV Transformer Era[39] and Resurrecting RNN[49] investigate whether modern RNN variants can rival Transformer scalability, highlighting ongoing debates about the trade-offs between recurrent and attention-based paradigms in large-scale sequence modeling.

Related Works in Same Category

The following **4 sibling papers** share the same taxonomy leaf node with the original paper:

1. Dynamic beam width tuning for energy-efficient recurrent neural networks

Authors: Daniele Jahier Pagliari, Francesco Panini, D. J. Pagliari, Enrico Macii, F. Panini, et al. (8 authors total) | **Year/Venue**: 2019 | **URL**: View paper

Abstract

Recurrent Neural Networks (RNNs) are state-of-the-art models for many machine learning tasks, such as language modeling and machine translation. Executing the inference phase of a RNN directly in edge nodes, rather than in the cloud, would provide benefits in terms of energy consumption, latency and network bandwidth, provided that models can be made efficient enough to run on energy-constrained embedded devices. To this end, we propose an algorithmic optimization for improving the energy effici...

Relationship Analysis

Both papers belong to the Efficient Sequence Processing Strategies category, focusing on optimizing computational efficiency during sequence processing. While the original paper (MesaNet) addresses efficiency through chunkwise parallelization and optimal test-time training with conjugate gradient solvers for recurrent layers, the candidate paper focuses on dynamic beam width tuning during inference for encoder-decoder RNNs in machine translation tasks. The key difference is that MesaNet optimizes the internal recurrent layer computation through parallelization and adaptive solver iterations, whereas the candidate paper optimizes the decoding search strategy by dynamically adjusting beam width based on model confidence.

2. Segrnn: Segment recurrent neural network for long-term time series forecasting

Authors: Lin Shengsheng, Lin WeiWei, Shengsheng Lin, Weiwei Lin, Zhao Feiyu, et al. (11 authors total) | **Year/Venue**: 2023 | **URL**: View paper

Abstract

RNN-based methods have faced challenges in the Long-term Time Series Forecasting (LTSF) domain when dealing with excessively long look-back windows and forecast horizons. Consequently, the dominance in this domain has shifted towards Transformer, MLP, and CNN approaches. The substantial number of recurrent iterations are the fundamental reasons behind the limitations of RNNs in LTSF. To address these issues, we propose two novel strategies to reduce the number of iterations in RNNs for LTSF task...

Relationship Analysis

Both papers belong to the Efficient Sequence Processing Strategies category, focusing on reducing computational costs in sequence modeling. They overlap in addressing the challenge of processing long sequences efficiently through chunking and parallelization techniques. However, the original paper (MesaNet) achieves efficiency through optimal test-time training with conjugate gradient solvers and chunkwise parallel processing of linear attention, while the candidate paper (SegRNN) reduces recurrent iterations by replacing point-wise with segment-wise iterations and introducing parallel multi-step forecasting specifically for RNN architectures.

3. Dual-path rnn: efficient long sequence modeling for time-domain single-channel speech separation

Authors: Yi Luo, Takuya Yoshioka, Zhuo Chen | **Year/Venue**: 2020 | **URL**: View paper

Abstract

â¦ dual-path recurrent neural network (DPRNN), a simple yet effective method for organizing â¦ a deep structure to model extremely long sequences. DPRNN splits the long sequential input â¦

Relationship Analysis

Both papers belong to the Efficient Sequence Processing Strategies category, focusing on techniques to handle long sequences in recurrent models. The original paper (MesaNet) addresses long sequences through optimal test-time training with conjugate gradient solvers and chunkwise parallelization, while the candidate paper (Dual-path RNN) tackles long sequences by splitting them into overlapping chunks and applying intra-chunk and inter-chunk RNNs alternately. The key difference is that MesaNet optimizes a cumulative loss at each time step using an iterative solver, whereas Dual-path RNN uses a fixed dual-path architecture with standard RNN operations to achieve sublinear complexity.

4. Efficient sequence learning with group recurrent networks

Authors: Fei Gao, Lijun Wu, Li Zhao, Zhao Li, Tao Qin, et al. (8 authors total) | **Year/Venue**: 2018 | **URL**: View paper

Abstract

Recurrent neural networks have achieved state-of-the-art results in many artificial intelligence tasks, such as language modeling, neural machine translation, speech recognition and so on. One of the key factors to these successes is big models. However, training such big models usually takes days or even weeks of time even if using tens of GPU cards. In this paper, we propose an efficient architecture to improve the efficiency of such RNN model training, which adopts the group strategy for recu...

Relationship Analysis

Both papers belong to the Efficient Sequence Processing Strategies category, focusing on reducing computational complexity in recurrent neural networks. The original paper (MesaNet) introduces optimal test-time training via conjugate gradient solvers within a Mesa layer for sequence modeling, while the candidate paper proposes group recurrent networks that partition hidden states and inputs into disjoint groups with representation rearrangement between layers. The key difference is that MesaNet optimizes a cumulative loss at each time step through iterative solving, whereas the candidate achieves efficiency through architectural sparsity (block-diagonal weight matrices) and parameter reduction via grouping.

Contributions Analysis

Overall novelty summary. The paper introduces a numerically stable, chunkwise parallelizable version of the Mesa layer optimized through conjugate gradient solvers for in-context learning, and evaluates it at billion-parameter scale for language modeling. Within the taxonomy, it resides in 'Efficient Sequence Processing Strategies' under 'Efficiency and Scalability Optimization', alongside four sibling papers focused on segmentation, dual-path processing, and parallelization techniques. This leaf contains five papers total, representing a moderately active but not overcrowded research direction within the broader 50-paper taxonomy covering RNN sequence modeling.

The taxonomy reveals that 'Efficient Sequence Processing Strategies' sits adjacent to 'Model Compression and Sparsity' and 'Hardware Acceleration', forming a cluster addressing computational bottlenecks in RNN inference and training. Neighboring branches include 'State Space Models and Linear RNNs' (focused on structured recurrences) and 'Transformer-RNN Comparisons' (examining architectural trade-offs). The paper's emphasis on optimal test-time training through conjugate gradient solvers distinguishes it from siblings like SegRNN, which simplifies recurrence via segmentation, and from state space models that rely on linear recurrence formulations rather than iterative optimization.

Among 23 candidates examined across three contributions, none were flagged as clearly refuting the work. Contribution A (parallelizable Mesa layer) examined 3 candidates with 0 refutable; Contribution B (MesaNet architecture performance) examined 10 candidates with 0 refutable; Contribution C (comparative RNN analysis) examined 10 candidates with 0 refutable. This suggests that within the limited search scope—top-K semantic matches plus citation expansion—no prior work was found providing directly overlapping methods or results. The absence of refutable candidates across all contributions indicates potential novelty, though the search scale (23 papers) leaves open the possibility of relevant work outside this sample.

Based on the limited literature search, the work appears to occupy a distinct position combining optimal in-context learning with efficient RNN design. The taxonomy context shows it addresses efficiency concerns shared by neighboring papers but through a unique optimization-based approach. However, the analysis covers only top-23 semantic matches and does not exhaustively survey all RNN efficiency literature or recent state space model developments, which may contain related optimization strategies not captured here.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Parallelizable and numerically stable Mesa layer with adaptive forgetting

Description: The authors introduce a chunkwise parallelizable version of the Mesa layer that solves linear systems using conjugate gradient methods. This new formulation enables efficient training on modern accelerators, supports dynamic forgetting through gating mechanisms, and maintains numerical stability, overcoming limitations of the original sequential Mesa layer.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Recurrent neural networks for edge intelligence: a survey

URL: View paper

Brief Assessment

RNN Edge Intelligence[51] is a survey paper about recurrent neural networks for edge computing applications, not a research contribution proposing parallelizable recurrent layers with conjugate gradient methods. The candidate focuses on deployment of RNNs in resource-constrained environments rather than novel layer architectures.

---

2. A conjugate gradient learning algorithm for recurrent neural networks

URL: View paper

Brief Assessment

Conjugate Gradient RNN[52] focuses on integrating conjugate gradient methods into RTRL for recurrent networks, not on parallelizable Mesa layers with adaptive forgetting mechanisms for modern sequence modeling architectures.

---

3. Numerically Stable Recurrence Relations for the Communication Hiding Pipelined Conjugate Gradient Method

URL: View paper

Brief Assessment

Pipelined Conjugate Gradient[53] focuses on stabilizing recurrence relations in the conjugate gradient method for solving linear systems in parallel computing contexts, not on designing recurrent neural network layers for sequence modeling with adaptive forgetting mechanisms.

Contribution 2: MesaNet architecture achieving strong language modeling performance

Description: The authors train MesaNet models at 140M, 440M, and 1B parameter scales on the SlimPajama dataset. These models achieve lower validation perplexity compared to existing recurrent models like Mamba2, xLSTM, DeltaNet, and Gated DeltaNet, while matching or exceeding transformer performance on various benchmarks.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Bayesian Neural Network Language Modeling for Speech Recognition

URL: View paper

Brief Assessment

Bayesian Language Model[69] focuses on Bayesian learning frameworks for LSTM-RNN and transformer language models in speech recognition tasks, not on recurrent neural network architectures like MesaNet compared to transformers on perplexity benchmarks.

---

2. Antiviral Peptide-Generative Pre-Trained Transformer (AVP-GPT): A Deep Learning-Powered Model for Antiviral Peptide Design with High-Throughput â¦

URL: View paper

Brief Assessment

AVP-GPT[71] focuses on antiviral peptide generation using GPT models for biological sequences, not recurrent neural network language models or perplexity benchmarks on text datasets like SlimPajama.

---

3. You Do Not Fully Utilize Transformer's Representation Capacity

URL: View paper

Brief Assessment

Transformer Representation Capacity[66] focuses on representation collapse in transformers and proposes layer-integrated memory, not recurrent neural network architectures or language modeling comparisons with models like Mamba2, xLSTM, or DeltaNet.

---

4. Just read twice: closing the recall gap for recurrent language models

URL: View paper

Brief Assessment

Read Twice[70] focuses on improving recurrent LMs through prompt repetition and non-causal processing techniques, not on proposing a new recurrent architecture like MesaNet that uses optimal test-time training with conjugate gradient solvers.

---

5. Learning to (Learn at Test Time): RNNs with Expressive Hidden States

URL: View paper

Brief Assessment

Test Time Training[63] focuses on RNN layers with expressive hidden states using self-supervised learning at test time, not specifically on the MesaNet architecture or its comparative performance against Mamba2, xLSTM, DeltaNet on SlimPajama benchmarks.

---

6. Improving language model predictions via prompts enriched with knowledge graphs

URL: View paper

Brief Assessment

Knowledge Graph Prompts[68] focuses on enriching language model predictions via knowledge graphs, not on recurrent neural network architectures or perplexity benchmarks for sequence modeling.

---

7. On the predictive power of neural language models for human real-time comprehension behavior

URL: View paper

Brief Assessment

Neural Language Model Prediction[64] focuses on evaluating neural language models' ability to predict human reading behavior, not on proposing novel recurrent architectures or comparing their perplexity performance on language modeling benchmarks.

---

8. Combining RNN with Transformer for Modeling Multi-Leg Trips.

URL: View paper

Brief Assessment

RNN Transformer Multi Leg[72] focuses on combining RNN with Transformer for travel destination prediction tasks, not general language modeling or perplexity benchmarks on datasets like SlimPajama. The architectural combination and application domain differ fundamentally from MesaNet's contributions.

---

9. Do Transformer Interpretability Methods Transfer to RNNs?

URL: View paper

Brief Assessment

Transformer Interpretability RNN Transfer[65] focuses on evaluating interpretability methods across architectures (Mamba, RWKV), not on developing new RNN architectures or comparing language modeling perplexity. The paper does not challenge the novelty of MesaNet's architectural contributions or performance claims.

---

10. Does Transformer Interpretability Transfer to RNNs?

URL: View paper

Brief Assessment

Transformer Interpretability Transfer[67] focuses on interpretability methods for RNN architectures (Mamba, RWKV), not on developing new RNN architectures or comparing their language modeling performance against transformers on perplexity benchmarks.

Contribution 3: In-depth comparative analysis of modern RNN architectures

Description: The authors conduct comprehensive analyses revealing that RNN models and transformers reduce perplexity differently across sequence positions, with RNNs excelling early in sequences while transformers perform better on later tokens. They also disentangle downstream benchmarks by global versus local language modeling requirements using controlled Sliding-Window Attention ablations.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. RWKV: Reinventing RNNs for the Transformer Era

URL: View paper

Brief Assessment

RWKV Transformer Era[39] focuses on introducing a novel RNN architecture (RWKV) and comparing it to transformers on standard benchmarks, but does not provide the specific type of comparative analysis described in the original contribution—namely, analyzing how RNNs and transformers reduce perplexity differently across sequence positions or disentangling benchmarks by global vs. local language modeling requirements.

---

2. Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention

URL: View paper

Brief Assessment

Transformers are RNNs[54] focuses on linearizing transformer attention mechanisms and expressing them as RNNs, not on comparative analysis of RNN vs transformer performance across sequence positions in language modeling tasks.

---

3. A comparative study on transformer vs rnn in speech applications

URL: View paper

Brief Assessment

Transformer vs RNN Speech[59] focuses on comparing Transformer and RNN architectures in speech applications (ASR, TTS, ST), not on language modeling tasks. The paper does not analyze perplexity reduction across sequence positions in language modeling contexts, nor does it examine global versus local language modeling requirements through controlled attention ablations as described in the original contribution.

---

4. Evaluating long-context understanding via latent and positional structure queries in large language models

URL: View paper

Brief Assessment

Long Context Understanding[56] focuses on evaluating long-context understanding in LLMs through latent and positional structure queries, not on comparing RNN and transformer architectures across sequence positions in language modeling.

---

5. Segmented Recurrent Transformer: An Efficient Sequence-to-Sequence Model

URL: View paper

Brief Assessment

Segmented Recurrent Transformer[62] focuses on combining segmented attention with recurrent mechanisms for sequence-to-sequence tasks, not on comparative analysis of RNN versus transformer performance across sequence positions in language modeling.

6. Compact Recurrent Transformer with Persistent Memory

URL: View paper

Brief Assessment

Compact Recurrent Transformer[55] focuses on combining shallow transformers with RNNs for memory compression in language and video tasks, not on comparative analysis of RNN vs. transformer performance across sequence positions.

7. Recurrent Memory Transformer

URL: View paper

Brief Assessment

Recurrent Memory Transformer[61] focuses on memory-augmented transformers with segment-level recurrence for long sequence processing, not on comparative analysis of RNN architectures versus transformers across sequence positions.

8. Positional encoding helps recurrent neural networks handle a large vocabulary

URL: View paper

Brief Assessment

Positional Encoding RNN[60] focuses on improving RNN performance through positional encoding for handling large vocabularies, not on comparative analysis of RNNs versus transformers across sequence positions in language modeling tasks.

9. Back to recurrent processing at the crossroad of transformers and state-space models

URL: View paper

Brief Assessment

Recurrent Processing Crossroad[57] focuses on recurrent processing mechanisms in vision and language tasks, not on comparative analysis of RNN architectures versus transformers in language modeling across sequence positions.

10. Transformers are Multi-State RNNs

URL: View paper

Brief Assessment

Transformers Multi State RNNs[58] focuses on redefining transformers as multi-state RNNs and developing compression policies, not on comparative analysis of RNN vs. transformer performance across sequence positions in language modeling tasks.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] MesaNet: Sequence Modeling by Locally Optimal Test-Time Training View paper
- [1] Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications View paper
- [2] Sequence Modeling using Gated Recurrent Neural Networks View paper
- [3] Recurrent neural networks (RNNs): architectures, training tricks, and introduction to influential research View paper
- [4] Quantum recurrent neural networks for sequential learning View paper
- [5] A survey on recurrent neural network architectures for sequential learning View paper
- [6] Structured sequence modeling with graph convolutional recurrent networks View paper
- [7] Sequence to Sequence Learning with Neural Networks View paper
- [8] Long short-term memory recurrent neural network architectures for large scale acoustic modeling. View paper
- [9] Symplectic Recurrent Neural Networks View paper
- [10] A review of recurrent neural network architecture for sequence learning: Comparison between LSTM and GRU View paper
- [11] Recurrent Neural Network Regularization View paper
- [12] Multi-time-scale with clockwork recurrent neural network modeling for sequential recommendation View paper
- [13] TRNN: An efficient time-series recurrent neural network for stock price prediction View paper
- [14] Hierarchically gated recurrent neural network for sequence modeling View paper
- [15] Weight-Space Linear Recurrent Neural Networks View paper
- [16] Phone sequence modeling with recurrent neural networks View paper
- [17] Investigating explainablity methods in recurrent neural network architectures for financial time-series data View paper
- [18] E-RNN: Design optimization for efficient recurrent neural networks in FPGAs View paper
- [19] Modeling an intrusion detection using recurrent neural networks View paper
- [20] Recurrent neural network architectures for analysing biomedical data sets View paper
- [21] Tensor Recurrent Neural Network With Differential Privacy View paper
- [22] mLANet: An efficient recurrent neural network for long-term time series forecasting View paper
- [23] Rwkv-ts: Beyond traditional recurrent neural network for time series tasks View paper
- [24] Predicting the risk of heart failure with EHR sequential data modeling View paper
- [25] Pixel Recurrent Neural Networks View paper
- [26] LightRNN: Memory and computation-efficient recurrent neural networks View paper
- [27] Dynamic beam width tuning for energy-efficient recurrent neural networks View paper
- [28] Recurrent neural network architectures View paper
- [29] Recurrent Attentive Neural Process for Sequential Data View paper
- [30] Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding. View paper
- [31] Investigating Explainability Methods in Recurrent Neural Network Architectures for Financial Time Series Data View paper
- [32] Recurrent Neural Network based Incremental model for Intrusion Detection System in IoT View paper
- [33] Residual recurrent neural networks for learning sequential representations View paper
- [34] Griffin: Mixing Gated Linear Recurrences with Local Attention for Efficient Language Models View paper
- [35] Deep-sdm: A unified computational framework for sequential data modeling using deep learning models View paper
- [36] MV-RNN: A multi-view recurrent neural network for sequential recommendation View paper

- [37] Efficient and effective training of sparse recurrent neural networks View paper
- [38] Segrnn: Segment recurrent neural network for long-term time series forecasting View paper
- [39] RWKV: Reinventing RNNs for the Transformer Era View paper
- [40] Recurrent Neural Network Model for Constructive Peptide Design View paper
- [41] Recurrent Attentive Neural Networks for Sequential Recommendation View paper
- [42] Efficiently modeling long sequences with structured state spaces View paper
- [43] Recurrent Convolutional Neural Network for Sequential Recommendation View paper
- [44] Recurrent neural network View paper
- [45] Dual-path rnn: efficient long sequence modeling for time-domain single-channel speech separation View paper
- [46] Efficient sequence learning with group recurrent networks View paper
- [47] CALRec: Contrastive Alignment of Generative LLMs for Sequential Recommendation View paper
- [48] Anticipating in Vehicle Accident using Recurrent Neural Network View paper
- [49] Resurrecting Recurrent Neural Networks for Long Sequences View paper
- [50] ReRAM-based processing-in-memory architecture for recurrent neural network acceleration View paper
- [51] Recurrent neural networks for edge intelligence: a survey View paper
- [52] A conjugate gradient learning algorithm for recurrent neural networks View paper
- [53] Numerically Stable Recurrence Relations for the Communication Hiding Pipelined Conjugate Gradient Method View paper
- [54] Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention View paper
- [55] Compact Recurrent Transformer with Persistent Memory View paper
- [56] Evaluating long-context understanding via latent and positional structure queries in large language models View paper
- [57] Back to recurrent processing at the crossroad of transformers and state-space models View paper
- [58] Transformers are Multi-State RNNs View paper
- [59] A comparative study on transformer vs rnn in speech applications View paper
- [60] Positional encoding helps recurrent neural networks handle a large vocabulary View paper
- [61] Recurrent Memory Transformer View paper
- [62] Segmented Recurrent Transformer: An Efficient Sequence-to-Sequence Model View paper
- [63] Learning to (Learn at Test Time): RNNs with Expressive Hidden States View paper
- [64] On the predictive power of neural language models for human real-time comprehension behavior View paper
- [65] Do Transformer Interpretability Methods Transfer to RNNs? View paper
- [66] You Do Not Fully Utilize Transformer's Representation Capacity View paper
- [67] Does Transformer Interpretability Transfer to RNNs? View paper
- [68] Improving language model predictions via prompts enriched with knowledge graphs View paper
- [69] Bayesian Neural Network Language Modeling for Speech Recognition View paper
- [70] Just read twice: closing the recall gap for recurrent language models View paper
- [71] Antiviral Peptide-Generative Pre-Trained Transformer (AVP-GPT): A Deep Learning-Powered Model for Antiviral Peptide Design with High-Throughput â¦ View paper
- [72] Combining RNN with Transformer for Modeling Multi-Leg Trips. View paper