

# Paper List

## 1. Time Series

### 1.1 Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting

核心贡献：提出基于Prob稀疏自注意力和蒸馏机制的高效Transformer变体，显著降低长序列预测的计算复杂度 ( $O(L \log L)$ )。

意义：解决了传统Transformer在长序列预测中的内存和计算瓶颈，成为时间序列领域的经典基线模型。

paper link: <https://cdn.aaii.org/ojs/17325/17325-13-20819-1-2-20210518.pdf>

Supplementary material: <https://github.com/zhouhaoyi/Informer2020?tab=readme-ov-file>

### 1.2 Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting

核心贡献：引入自相关机制（替代传统注意力）和序列分解模块（趋势-季节分离），提升长期预测能力。

意义：通过结合时序分解先验知识，显著提高了预测的稳定性和可解释性。

paper link: <https://arxiv.org/abs/2106.13008>

Supplementary material: <https://github.com/thuml/Autoformer>

### 1.3 RWKV-TS: Beyond Traditional Recurrent Neural Network for Time Series Tasks

核心贡献：将RNN与Transformer结合的RWKV架构（线性注意力）应用于时间序列，兼顾效率与长程建模。

意义：在保持RNN推理效率的同时，获得接近Transformer的性能。

paper link: <https://arxiv.org/abs/2401.09093>

Supplementary material: <https://github.com/howard-hou/RWKV-TS>

## 1.4 TimeMachine: A Time Series is Worth 4 Mambas for Long-term Forecasting

核心贡献：将Mamba架构扩展为多分支版本，结合频率分析和分解策略，优化长期预测。

意义：首次将Mamba应用于时间序列预测，实现SOTA性能。

paper link: <https://arxiv.org/abs/2403.09898>

Supplementary material: <https://github.com/Atik-Ahamed/TimeMachine>

## 1.5 A decoder-only foundation model for time-series forecasting (TimesFM)

核心贡献：Google提出的Decoder-only基础模型，通过大规模预训练实现零样本泛化能力（类似时间序列GPT）。

意义：推动时间序列预测进入“基础模型”时代，支持开放域任务。

paper link: <https://arxiv.org/abs/2310.10688>

Supplementary material: <https://github.com/google-research/timesfm?tab=readme-ov-file>

# 2. NextGen AI Models

## 2.1 Schema-learning and rebinding as mechanisms of in-context learning and emergence

核心贡献：提出“模式学习”（schema-learning）和“重新绑定”（rebinding）作为上下文学习（in-context learning）的核心机制。模型通过动态绑定已有模式到新任务，实现快速适应和涌现能力。

意义：解释了大型语言模型（LLMs）的上下文学习能力，为构建更高效的小样本学习模型提供理论支持。

paper link: <https://arxiv.org/pdf/2307.01201>

## 2.2 Learning to (Learn at Test Time): RNNs with Expressive Hidden States

核心贡献：设计了一种具有表达性隐藏状态的循环神经网络（RNN），能够在测试时动态调整内部状态以适应新任务，无需额外训练。

意义：提升了模型在动态环境中的在线适应能力，适用于时间序列预测和实时决策任务。

paper link: <https://arxiv.org/abs/2407.04620>

## 2.3 Inductive Moment Matching

核心贡献：提出一种基于矩匹配（moment matching）的归纳学习方法，通过匹配数据分布的统计矩实现高效的知识迁移和模型压缩。

应用：被Luma Labs用于轻量化模型部署，支持在资源受限设备上运行复杂模型。

意义：提供了一种参数高效的知识蒸馏方法，平衡模型性能与计算开销。

paper link: <https://arxiv.org/abs/2503.07565>

## 3.Multi Modality

### 3.1 BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models

核心贡献：BLIP-2通过冻结图像编码器和大型语言模型，减少了训练成本，同时保持了强大的多模态理解能力。它通过引导式预训练（bootstrapping）方法，逐步提升模型性能。

意义：提供了一种高效的预训练策略，能够在减少计算资源的情况下实现高性能的多模态理解。

paper link: <https://arxiv.org/abs/2301.12597>