时间-LLM:时间序列预测通过重新编程大型语言模型

明金1 , Shiyu Wang2 , 临洮马2 , Zhixuan Chu2 , James Y. Zhang2 , Xiaoming Shi2 , Pin-Yu Chen3 , Yuxuan Liang6 , Yuan-Fang Li1 , Shirui Pan4 †Qingsong Wen5 †

1 莫纳什大学2 蚂蚁集团 IBM 研究院4 格里菲斯大学5阿里巴巴集团6 香港科技大学（广州） {ming.jin,yuanfang.li}@monash.edu, pin-yu.chen@ibm.com yuxliang@outlook.com, s.pan@griffith.edu.au, qingsongedu@gmail.com {weiming.wsy,lintao.mlt,chuzhixuan.czx,james.z,peter.sxm}@antgroup.com

抽象的

时间序列预测在许多现实世界的动态系统中都具有重要意义,并且已被广泛研究。与自然语言处理 (NLP) 和计算机视觉 (CV) 中单个大型模型可以处理多项任务不同,时间序列预测模型通常是专门化的,需要针对不同的任务和应用进行不同的设计。虽然预训练的基础模型在 NLP 和 CV 领域取得了令人瞩目的进步,但它们在时间序列领域的发展却受到数据稀疏性的制约。最近的研究表明,大型语言模型 (LLM) 对复杂的标记序列拥有强大的模式识别和推理能力。然而,如何有效地协调时间序列数据和自然语言的模态以充分利用这些能力,仍然是一个挑战。在本研究中,我们提出了TIME-LLM,这是一个可重编程框架,用于在保留主干语言模型不变的情况下,将 LLM 重新用于通用时间序列预测。我们首先使用文本原型对输入时间序列进行重编程,然后将其输入到冻结的 LLM 中,以协调两种模态。为了增强 LLM 推理时间序列数据的能力,我们提出了 Prompt-as-Prefix (PaP),它可以丰富输入上下文并指导重新编程的输入块的转换。最终,对来自 LLM 的转换后的时间序列块进行投影以获得预测结果。我们的综合评估表明， TIME-LLM 是一个强大的时间序列学习器,其性能优于最先进的专业预测模型。此外， TIME-LLM 在少样本学习和零样本学习场景中均表现出色。代码可在https://github.com/KimMeen/Time-LLM获取。

1引言

时间序列预测是许多现实世界动态系统（Jin et al., 2023a)中的关键能力,其应用范围广泛,从需求规划（Leonard, 2001）和库存优化（Li et al., 2022）,到能源负荷预测（Liu et al., 2023a)和气候建模（Schneider & Dickinson, 1974）。每个时间序列预测任务通常都需要广泛的领域专业知识和针对特定任务的模型设计。这与 GPT-3（Brown et al., 2020）、GPT-4（OpenAI, 2023）、Llama（Touvron et al., 2023)等基础语言模型形成了鲜明对比,这些模型在少量甚至零样本的环境下,可以在各种自然语言处理任务中表现出色。

大型语言模型 (LLM) 等预训练基础模型推动了计算机视觉 (CV) 和自然语言处理 (NLP) 领域的快速发展。虽然时间序列建模尚未获得同样重大的突破,但 LLM 的卓越能力激发了其在时间序列预测中的应用 (Jin et al., 2023b)。利用 LLM 推进预测技术有几个关键要素:泛化能力。LLM已展现出卓越的少样本和零样本迁移学习能力 (Brown et al., 2020)。这表明它们

同等贡献 †通讯作者

跨领域通用预测的潜力,无需每个任务从头开始重新训练。相比之下,当前的预测方法通常严格地针对特定领域。数据效率。通过利用预先训练的知识,LLM 已展示出仅用少量示例即可执行新任务的能力。这种数据效率可以使预测在历史数据有限的环境中成为可能。相比之下,当前的方法通常需要大量的领域内数据。

推理。LLM展现出复杂的推理和模式识别能力 (Mirchandani et al., 2023; Wang et al., 2023; Chu et al., 2023)。利用这些技能,可以通过利用已学的高级概念做出高精度预测。现有的非 LLM 方法主要基于统计,缺乏太多先天推理能力。多模态知识。随着 LLM 架构和训练技术的改进,它们获得了涵盖视觉、语音和文本等多种模态的更多样化知识 (Ma et al., 2023)。利用这些知识可以实现融合不同数据类型的协同预测。传统工具缺乏联合利用多个知识库的方法。易于优化。LLM只需在海量计算上训练一次,即可应用于预测任务,无需从头开始学习。优化现有的预测模型通常需要大量的架构搜索和超参数调整 (Zhou et al., 2023b)。总而言之,与目前专门的建模范式相比,法学硕士 (LLM) 提供了一条充满希望的途径,使时间序列预测更加通用、高效、协同且易于理解。因此,将这些强大的模型应用于时间序列数据可以释放巨大的未开发潜力。

上述优势的实现取决于时间序列数据和自然语言模态的有效匹配。然而,由于LLM操作的是离散标记,而时间序列数据本质上是连续的,因此这极具挑战性。此外,LLM的预训练模型中并不天然具备解释时间序列模式的知识和推理能力。

因此,解锁法学硕士中的知识,以准确、数据高效和任务无关的方式激活其进行一般时间序列预测的能力仍然是一个开放的挑战。

在本研究中,我们提出了TIME-LLM,这是一个可重编程框架,旨在将大型语言模型应用于时间序列预测,同时保持其骨干模型的完整性。其核心思想是将输入的时间序列重编程为更自然地适应语言模型功能的文本原型表示。为了进一步增强模型对时间序列概念的推理能力,我们引入了 Prompt-as-Prefix (PaP),这是一种新颖的思路,它通过附加上下文来丰富输入的时间序列,并以自然语言的模态提供任务指令。这为应用于重编程输入的所需转换提供了声明式指导。语言模型的输出随后被投影以生成时间序列预测。我们的全面评估表明,通过这种重编程方法,大型语言模型可以充当有效的少样本和零样本时间序列学习器,其性能优于专门的预测模型。通过利用 LLM 的推理能力,同时保持模型的完整性,我们的工作为在语言和序列数据任务中均表现优异的多模态基础模型指明了方向。我们提出的重编程框架提供了一种可扩展的范例,能够为大型模型注入超越其原始预训练能力的新功能。我们的主要贡献可以概括如下:

- 我们引入了一种新颖的概念,即在不改变预先训练的骨干模型的情况下,对大型语言模型进行重新编程,以进行时间序列预测。通过这种方式,我们表明预测可以被视为另一种“语言”任务,可以通过现成的LLM有效解决。· 我们提出了一个新框架TIME-LLM,它包含将输入时间序列重新编程为对LLM更自然的文本原型表示,并使用声明性提示(例如,领域专家知识和任务指令)来增强输入上下文,以指导LLM推理。我们的技术旨在构建在语言和时间序列方面均表现优异的多模态基础模型。

- TIME-LLM 在主流预测任务中,尤其是在少样本和零样本场景下,始终超越最先进的性能。此外,这种卓越的性能是在保持优异的模型重编程效率的同时实现的。因此,我们的研究是释放 LLM 在时间序列以及其他序列数据方面尚未开发的潜力的坚实一步。

2相关工作

特定任务学习。大多数时间序列预测模型都是针对特定任务和领域(例如,交通预测)而设计的,并在小规模数据上进行端到端训练。如图所示

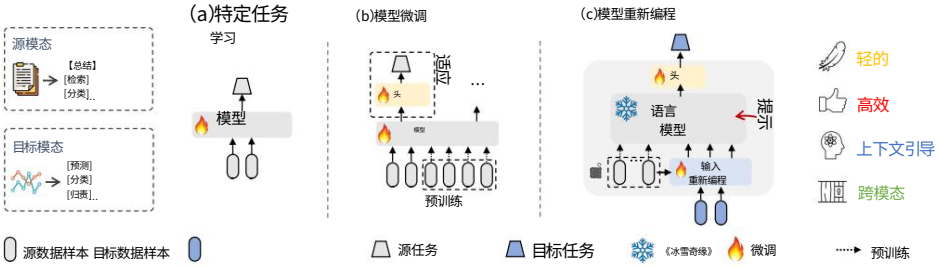


图 1:重新编程大型语言模型 (LLM) 的示意图,对比了(a)任务特定学习和(b)模型微调。我们的提案研究并演示了(c)在缺乏成熟的时间序列预训练模型的情况下,如何有效地将开源 LLM 重新编程为强大的时间序列学习器。

图 1(a)。例如,ARIMA 模型专为单变量时间序列预测而设计（Box 等人,2015）,LSTM 网络专为序列建模而设计（Hochreiter & Schmidhuber,1997）,而时间卷积网络（Bai 等人,2018)和 Transformer（Wen 等人,2023)则是为了处理更长的时间依赖关系而开发的。虽然这些模型在特定任务上取得了良好的性能,但它们缺乏对多样化时间序列数据的通用性和泛化能力。

模态内自适应。计算机视觉和自然语言处理领域的相关研究已经证明了预训练模型的有效性,这些模型可以针对各种下游任务进行微调,而无需从头开始进行昂贵的训练 (Devlin 等人,2018;Brown 等人,2020;Touvron 等人,2023)。受这些成功的启发,近期的研究集中于时间序列预训练模型 (TSPTM) 的开发。其中第一步是使用不同的策略对时间序列进行预训练,例如监督学习 (Fawaz 等人,2018) 或自监督学习 (Zhang 等人,2022b;Deldari 等人,2022;Zhang 等人,2023)。这使得模型能够学习表示各种输入时间序列。预训练完成后,可以在相似领域进行微调,以学习如何执行特定任务 (Tang et al., 2022)。示例如图 1(b) 所示。TSPTM 的开发借鉴了预训练和微调在自然语言处理 (NLP) 和计算机视觉 (CV) 领域的成功经验,但由于数据稀疏性,其在较小规模上仍然受到限制。

跨模态自适应。在模态内自适应的基础上,近期研究进一步探索了如何通过多模态微调（Yin et al., 2023)和模型重编程（Chen, 2022)等技术,将知识从自然语言处理（NLP)和计算机视觉（CV)领域强大的预训练基础模型迁移到时间序列建模。我们的方法与此类似;然而,关于时间序列的相关研究有限。例如,Voice2Series（Yang et al., 2021）,它通过将时间序列编辑成适合声学模型 (AM) 的格式,将语音识别中的声学模型 (AM) 适配到时间序列分类。最近,Chang et al. (2023) 提出了 LLM4TS,用于使用 LLM 进行时间序列预测。它在 LLM 上设计了一个两阶段的微调过程　首先对时间序列进行监督预训练,然后进行特定于任务的微调。Zhou et al. (2023) (2023a) 利用预训练语言模型,无需改变其自注意力机制和前馈层。该模型在各种时间序列分析任务上进行了微调和评估,并通过迁移自然语言预训练知识展现出相当甚至最先进的性能。与这些方法不同,我们既不直接编辑输入时间序列,也不对主干 LLM 进行微调。相反,如图 1(c) 所示,我们建议使用源数据模态对时间序列进行重新编程,并结合提示,以释放 LLM 作为高效时间序列机器的潜力。

3方法论

我们的模型架构如图 2 所示。我们专注于重新编程一个嵌入可见的语言基础模型,例如 Llama (Touvron 等人,2023) 和 GPT-2 (Radford 等人,2019),使其能够进行通用的时间序列预测,而无需对主干模型进行任何微调。具体来说,我们考虑以下问题:给定一个历史观测序列 X ∈ R,它由 N 个不同的一维变量组成,跨越 T 个时间步长。我们的目标是重新编程一个大型语言模型 f(·),使其理解输入的时间序列并准确预测未来 H 个时间步长的读数,记为 Y　∈ R N×H,总体目标是最小化真实值 Y 与预测值之间的均方误差 ||Y　h − Yh||2 ,即:

$$\frac{1}{\text{哈}} \quad \text{哈}_{h=1} \qquad \text{F.}$$

我们的方法包含三个主要部分：(1)输入转换；(2)预训练并冻结的 LLM;以及 (3)输出投影。首先,将一个多变量时间序列划分为 N 个

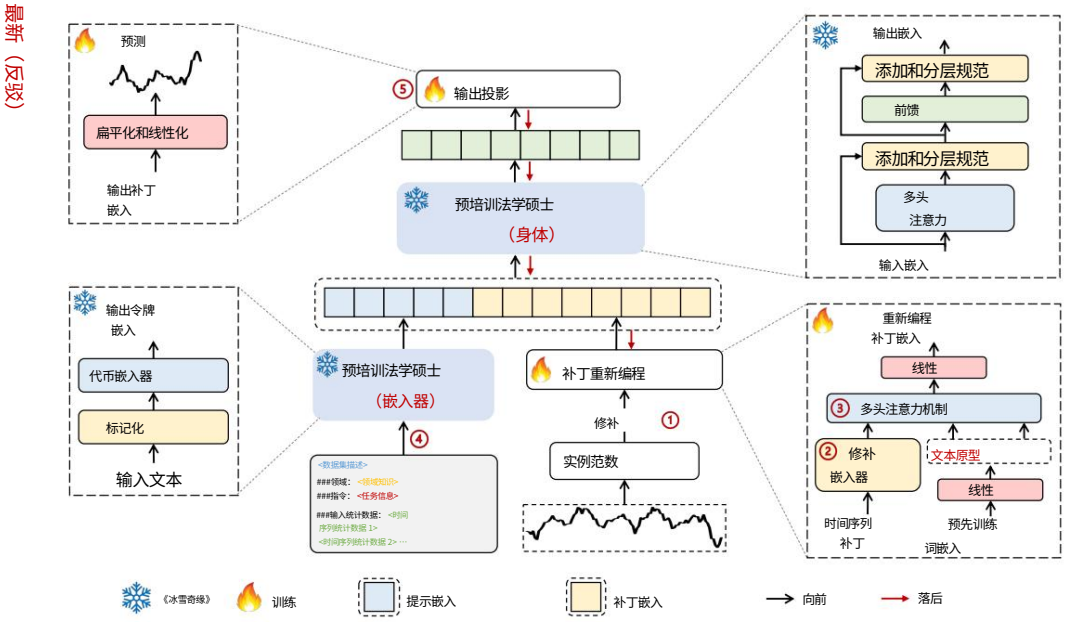


图 2： TIME-LLM的模型框架。给定一个输入时间序列，我们首先通过1补丁和2自定义嵌入层对其进行标记和嵌入。3然后，这些补丁嵌入使用精简的文本原型重新编程，以对齐两种模态。为了增强 LLM 的推理能力，在输入中添加了4额外的提示前缀，以指导输入补丁的转换。5 LLM的输出补丁被投影以生成预测。

单变量时间序列，随后进行独立处理（Nie 等人，2023）。第 i 个序列表示为X(i) ∈ R，经过归一化、修补和嵌入处理后，再使用学习到的文本原型进行重新编程，以对齐源模态和目标模态。然后，我们增强 LLM 的时间序列推理能力，将其与重新编程的补丁一起输入，生成输出表示，并将其投影到最终预测 Y (i) ∈ R 1×H。 1×T

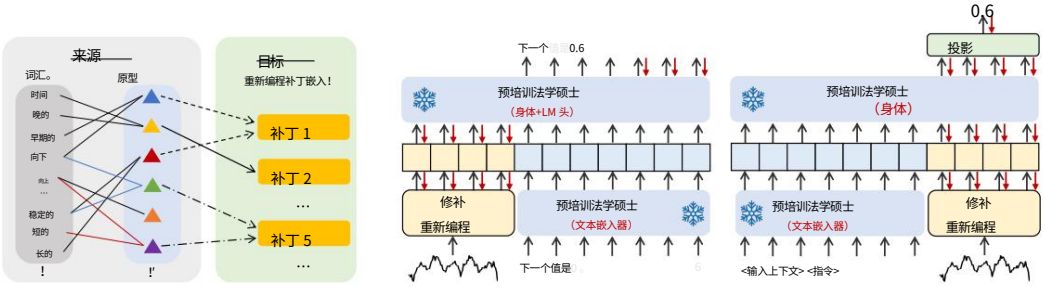
我们注意到，只有轻量级输入变换和输出投影的参数会被更新，而骨干语言模型则被冻结。与通常使用成对的跨模态数据进行微调的视觉语言模型和其他多模态语言模型不同， TIME-LLM 是直接优化的，只需少量时间序列和少量训练周期即可轻松上手。与从头构建大型领域特定模型或对其进行微调相比，TIME-LLM 保持了高效率，且资源限制更少。为了进一步减少内存占用，可以无缝集成各种现成的技术（例如量化）来精简TIME-LLM。

3.1模型结构

输入嵌入。首先，通过可逆实例归一化 (RevIN) 将每个输入通道X(i)单独归一化为零均值和单位标准差，以减轻时间序列分布偏移 (Kim et al., 2021)。然后，我们将X(i)划分为多个连续的、重叠或不重叠的 patch (Nie et al., 2023)，长度为Lp；因此，输入 patch 的总数(T −Lp)为 P = + 2，其中S表示水平滑动步幅。S 的潜在动机有两个方面：(1) 通过将局部信息聚合到每个 patch 中来更好地保留局部语义信息；(2) 用作标记化，形成紧凑的输入 token 序列，从而减少(i)

计算负担。给定这些块 X ∈ R，采用简单的线性层作为块嵌入器来创建维度dm磷 Lp ∈ R × ，我们将它们嵌入为 X 磷 (i) P × dm，

补丁重编程。我们将补丁嵌入重新编程到源数据表示空间中，以匹配时间序列和自然语言的模态，从而激活主干网络的时间序列理解和推理能力。一种常见的做法是学习一种“噪声”，当将其应用于目标输入样本时，允许预训练的源模型生成所需的目标输出，而无需更新参数。这对于桥接数据在技术上是可行的。



(a)补丁重编程　　(b)补丁作为前缀和提示作为前缀

图 3：(a)补丁重新编程和(b)补丁作为前缀与提示作为前缀的图示。

相同或相似的模态。例如，重新利用视觉模型来处理跨域图像（Misra 等人，2023），或重新编程声学模型来处理时间序列数据（Yang 等人，2021）。在这两种情况下，源数据和目标数据之间都存在显式的、可学习的转换，从而允许直接编辑输入样本。然而，时间序列既不能直接编辑，也不能用自然语言进行无损描述，这给直接引导 LLM 理解时间序列（无需进行资源密集型的微调）带来了重大挑战。

最新（照相就绪）

为了弥补这个差距，我们建议使用预先训练的词嵌入 E ∈ R 对X 磷 (i)进行重新编程 高×深
在主干网络中，V 是词汇量。然而，没有先验知识表明哪些源标记是直接相关的。因此，仅仅利用 E 将导致巨大且可能密集的重编程空间。一个简单的解决方案是维护一个小的文本集合V ′×D，其中 V 通过线性探测 E 进行原型设计，表示为E′ ∈ R ≪ V。

如图 3(a) 所示。文本原型学习连接性语言

线索，例如“短促上扬”（红线）和“稳定下落”（蓝线），然后将它们组合起来表示局部块信息（例如，“短促上扬然后稳定下落”用于表征块 5），而无需离开语言模型预训练的空间。这种方法高效，并且允许自适

应地选择相关的源信息。为了实现这一点，我们采用了一个多头交叉注意力层。具体而言，对于每个头 k = {1, · · · = E′WK (i)和值矩阵 V k

, K}，我们定义查询矩阵 Q (我) =X (i) (i)WQ k，密
千 磷 千
= E′WV k， 其中WQ 千 钥矩阵 K ∈ R dm×d且WK 千 k，
千 ，西弗吉尼亚州 ∈R 直径×直径 。具体来说，
D 是骨干模型的隐藏维度，d = 重新编程每个注意力头中的时间序列补丁，定义 分米 。然后，我们有操作来
钾
为：

从k (—) =注意（Q (—) (—) 钾 (—) (—) 在) = SOFTMAX(问 千 (—) k (—))V k。 (1)
k，k，k √ dk

###领域知识：我们通常观察到，用电量通常在中午达到峰值，变压
器负荷明显增加……

通过聚合每个投影的 Z 来将隐千 (我) ∈R 功率×d 在每个头中，我们得到Z (i) ∈ R P ×dm。然后线性
藏维度与主干模型对齐，得到O(i) ∈ R P ×D。

<BOS>电力变压器温度（ETT）指示电力的长期部署。每个数据点包含目标油温和 6 个电力负荷特征……
该数据集包含来自……的每小时采样数据。
###任务说明：根据前面<t>步骤信息预测接下来的<t>步骤

###输入统计数据：输入的最小值为<min value>，最大值为<max value>，中值为<median value>。总体趋势为<upward or down>，前五个滞后值为<lag values>。 <EOS>

提示作为前缀。提示是一种直接而有效的方法，可以激活特定任务的 LLM（Yin 等人，2023）。然而，将时间序列直接翻译成自然语言面临着相当大的挑战，这既阻碍了指令遵循数据集的创建，也阻碍了实时数据的有效利用。

每个数据点由目标油温和 6 个电力负荷特征组成……
###领域知识：我们通常观察到，用电量通常在中午达到峰值，变压
器负荷明显增加……

###任务说明：根据前面<t>步骤信息预测接下来的<t>步骤

飞行提示而不会影响性能（Xue&Salim，2022）。最近的进展表明，其他数据模态（例如图像）可以无缝集成为提示的前缀，从而促进基于这些输入的有效推理（Tsimpoukelli等人，2021）。受这些发现的启发，为了使我们的方法直接适用于现实世界的时间序列，我们提出了另一个问题：提示是否可以充当前缀来丰富输入上下文并指导重新编程的时间序列补丁的转换？我们将此概念称为Prompt-as-Prefix（PaP），并观察到它显着增强了LLM对下游任务的适应性，同时补充了补丁重新编程（参见后面的4.5节）。

###输入统计数据：输入的最小值为<min value>，最大值为<max value>，中值为<median value>，总体趋势为<upward or down>，前五个滞后值为<lag values>。

电力变压器温度 (ETT) 指示电力的长期部署。每个数据点包含目标油温和 6 个电力负荷特征……

以下是有关输入时间序列的信息：

[开始数据]
***
[领域]：我们通常观察到中午用电高峰，变压器负荷显著增加
***
[说明]：根据前面<t>步骤的信息预测接下来的<t>步骤
***
[统计]：输入的最小值为<min_val>，最大值为<max_val>，中值为<median_val>，总体趋势为<向上或向下>，前五个滞后值为<lag_val>。

[结束数据]

图 4：提示示例。 <>和<>是特定于任务的配置和计算的输入统计数据。

图 3(b) 展示了两种提示方法。在 Patch-as-Prefix 中,语言模型被提示预测时间序列中的后续值,并用自然语言表达。这种方法存在一些限制:(1) 在没有外部工具的帮助下,语言模型在处理高精度数字时通常会表现出较低的灵敏度,因此在准确处理长期实际预测任务方面面临巨大挑战;(2) 由于不同的语言模型是在不同的语料库上进行预训练的,并且可能使用不同的标记化类型来精确高效地生成高精度数字,因此需要对其进行复杂的、定制化的后处理。这导致预测结果以不同的自然语言格式表示,例如 [0 , . , 6 , 1] 和 [0 , . , 61] (表示小数 0.61)。

另一方面,提示作为前缀巧妙地避开了这些限制。在实践中,我们确定了构建有效提示的三个关键组成部分:(1)数据集上下文,(2)任务指导,以及 (3)输入统计数据。提示示例如图 4 所示。数据集上下文为 LLM 提供了有关输入时间序列的重要背景信息,而输入时间序列通常在不同领域表现出不同的特征。任务指导是 LLM 转换特定任务的块嵌入的关键指南。我们还通过趋势和滞后等其他关键统计数据来丰富输入时间序列,以促进模式识别和推理。

输出投影。如图 2 所示,将提示和块嵌入O(i)打包并前馈到冻结的 LLM 中后,我们丢弃前缀部分并获得输出表示。之后,我们对其进行扁平化和线性投影,以得出最终预测 Y (i)

。

4主要结果

TIME-LLM 在多个基准测试和设置中始终大幅超越最先进的预测方法,尤其是在少样本和零样本场景中。我们将我们的方法与一系列最新模型进行了比较,其中包括一项近期研究,该研究针对时间序列分析对语言模型进行了微调 (Zhou et al., 2023a)。为确保比较公平,我们在所有基准测试中均遵循 (Wu et al., 2023) 中的实验配置,并使用统一的评估流程 1。除非另有说明,否则我们使用 Llama-7B (Touvron et al., 2023) 作为默认骨干网络。

基准。我们与 SOTA 时间序列模型进行了比较,并在适用的情况下引用了 (Zhou et al., 2023a) 中它们的性能。我们的基准包括一系列基于 Transformer 的方法:PatchTST (2023)、ESTformer (2022)、Non-Stationary Transformer (2022)、FEDformer (2022)、Aut-oformer (2021)、Informer (2021) 和 Reformer (2020)。我们还选择了一组近期的竞争模型,包括 GPT4TS (2023a)、LLMTime (2023)、DLinear (2023)、TimesNet (2023) 和 LightTS (2022a)。在短期预测方面,我们进一步将我们的模型与 N-HiTS (2023b) 和 N-BEATS (2020) 进行了比较。更多详细信息请参见附录 A。

4.1长期预测

设置。我们评估了 ETTh1、ETTh2、ETTm1、ETTm2、天气、电力 (ECL)、交通和 ILI 等数据,这些数据已被广泛用于对长期预测模型进行基准测试 (Wu et al., 2023)。实现细节和数据集可在附录 B 中找到。输入时间序列长度 T 设置为 512,我们使用四个不同的预测范围 H ∈ {96, 192, 336, 720}。

评估指标包括均方误差 (MSE)和平均绝对误差 (MAE)。

结果。我们的简要结果如表 1 所示, TIME-LLM 在大多数情况下都优于所有基线模型,并且在大多数情况下表现显著。与 GPT4TS (Zhou et al., 2023a) 的比较尤其值得关注。GPT4TS 是一项近期的研究,它涉及对骨干语言模型进行微调。我们注意到,与 GPT4TS 和 TimesNet 相比,TIME-LLM 的平均性能分别提升了12%和20% 。与目前为止最成功的、针对特定任务的 Transformer 模型 PatchTST 相比,通过对最小的 Llama 模型进行重新编程, TIME-LLM 实现了平均 MSE 降低 1.4%。

相对于其他模型,例如 DLinear,我们的改进也很显著,超过12%。

4.2短期预测

设置。我们选择 M4 基准 (Makridakis et al., 2018) 作为测试平台,其中包含不同采样频率的营销数据集合。更多详细信息请参见附录 B。本例中的预测范围相对较小,参见 [6, 48]。输入长度

[1] https://github.com/thuml/Time-Series-Library

表 1:长期预测结果。所有结果均取自四个不同的预测范围的平均值:H ∈
ILI 为 {24, 36, 48, 60},其他为 {96, 192, 336, 720}。值越低,性能越好。
红色:最佳,蓝色:次佳。完整结果见附录 D。

方法	时间-LLM (我们的)	GPT4TS (2023a)			线性 (2023)	PatchTST (2023)	TimesNet (2023)	FEDformer (2022)	Autoformer (2022)	Stationary (2022)	ETSformer (2022)	LightTS (2022)	DLinear (2023)	FEDformer (2022)	Autoformer (2022)	告密者(2021)	改革者 (2020年)
公制	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ET h1	0.408	0.423	0.465	0.455	0.422	0.437	0.413	0.430	0.458	0.450	0.440	0.460	0.496	0.487	0.570	0.537	0.542	0.510	0.491	0.479	1.040	0.795	1.029	0.805						
ET h2	0.334	0.383	0.381	0.412	0.431	0.446	0.330	0.379	0.414	0.427	0.437	0.449	0.450	0.459	0.526	0.516	0.439	0.452	0.602	0.543	4.431	1.729	6.736	2.191						
ET Tm1	0.329	0.372	0.388	0.400	0.357	0.378	0.351	0.380	0.400	0.406	0.448	0.452	0.588	0.517	0.481	0.456	0.429	0.425	0.435	0.437	0.961	0.734	0.799	0.671						
ET Tm2	0.251	0.313	0.284	0.339	0.267	0.333	0.255	0.315	0.291	0.333	0.305	0.349	0.327	0.371	0.306	0.347	0.293	0.342	0.409	0.436	1.410	0.810	1.479	0.915						
天气	0.225	0.257	0.237	0.270	0.248	0.300	0.225	0.264	0.259	0.287	0.309	0.360	0.338	0.382	0.288	0.314	0.271	0.334	0.261	0.312	0.634	0.548	0.803	0.656						
预期亏损	0.158	0.252	0.167	0.263	0.166	0.263	0.161	0.252	0.192	0.295	0.214	0.327	0.227	0.338	0.193	0.296	0.208	0.323	0.229	0.329	0.311	0.397	0.338	0.422						
交通量	0.388	0.264	0.414	0.294	0.433	0.295	0.390	0.263	0.620	0.336	0.610	0.376	0.628	0.379	0.624	0.340	0.621	0.396	0.622	0.392	0.764	0.416	0.741	0.422						
或	1.435	0.801	1.925	0.903	2.169	1.041	1.443	0.797	2.139	0.931	2.847	1.144	3.006	1.161	2.077	0.914	2.497	1.004	7.382	2.003	5.137	1.544	4.724	1.445						
1 stCount	7		0		0		5		0		0		0		0		0		0		0		0		0		0		0	

表 2:M4 的短期时间序列预测结果。预测范围参见 [6, 48],
提供的三行数据是不同采样间隔下所有数据集的加权平均值。较低的值
表示性能更佳。红色表示最佳,蓝色表示次佳。更多结果请参见附录 D。

方法	TIME-LLM (我们的)	GPT4TS (2023a)	TimesNet (2023)	PatchTST (2023)	N-HiTS (2023)	N-BEATS (2023b)	ETSformer (2022)	LightTS (2022)	DLinear (2022a)	FEDformer (2023)	Stationary (2022)	Autoformer (2022)	Informer (2021)	Reformer (2021年)	(2020年)
平均	SMAPE 11.983	MASE	12.69	12.88	12.059	12.035	1.625	12.25	14.718	13.525	13.639	2.111	13.16	12.780	12.909	14.086	18.200
	1.595	OWA 0.859	1.808	1.836	1.623	0.869		1.698	2.408	2.095	1.051	1.051	1.775	1.756	1.771	2.718	4.223
			0.94	0.955	0.869			0.896	1.172				0.949	0.930	0.939	1.230	1.775

是预测范围的两倍。评估指标是对称平均绝对百分比
误差（SMAPE）、平均绝对缩放误差（MSAE)和总体加权平均值（OWA）。

结果。表 2 列出了我们所有方法使用统一种子的简要结果。TIME -LLM 持续超越所有基线,比 GPT4TS 高出8.7%。TIME -LLM 仍然保持竞争力
即使与 SOTA 模型、N-HiTS（Challu 等人,2023b）、mase 和 owa 相比。

4.3小样本预测

设置。法学硕士 (LLM) 最近展示了卓越的小样本学习能力 (Liu et al.,
2023b)。在本节中,我们评估重新编程的 LLM 是否保留了这种预测能力
任务。我们遵循（Zhou et al., 2023a)中的设置进行公平比较,并评估
训练数据有限的场景（即,≤前 10% 的训练时间步骤）。

结果。我们简短的 10% 和 5% 小样本学习结果分别列于表 3 和表 4。TIME- LLM在所有基线方法中表现优异,我们将其归因于成功的知识
在我们重新编程的 LLM 中激活。有趣的是,我们的方法和 GPT4TS 都一致
超越其他竞争基线,进一步凸显了语言模型的潜在实力
作为熟练的时间序列机器。

在10 % 小样本学习的范围内,我们的方法相比于
到 GPT4TS,无需对 LLM 进行任何微调。关于最近的 SOTA 模型

表 3:基于 10% 训练数据的少样本学习。我们使用与表 1 相同的协议。所有结果均为平均值
来自四个不同的预测范围:H ∈ {96, 192, 336, 720}。完整结果见附录 E。

方法	时间-LLM (我们的)	GPT4TS (2023a)			线性 (2023)	PatchTST (2023)	TimesNet (2023)	FEDformer (2022)	Autoformer (2022)	Stationary	ETSformer (2022)	LightTS (2022)	DLinear (2021)	FEDformer (2022a)	Autoformer (2022)	告密者 (2021)	改革者 (2020年)
公制	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ET h1	0.556	0.522	0.590	0.525	0.691	0.600	0.633	0.542	0.869	0.628	0.639	0.561	0.702	0.596	0.915	0.639	1.180	0.834	1.375	0.877	1.199	0.809	1.249	0.833						
ET h2	0.370	0.394	0.397	0.421	0.605	0.538	0.415	0.431	0.479	0.465	0.466	0.475	0.488	0.499	0.462	0.455	0.894	0.713	2.655	1.160	3.872	1.513	3.485	1.486						
ET Tm1	0.404	0.427	0.464	0.441	0.411	0.429	0.501	0.466	0.677	0.537	0.722	0.605	0.802	0.628	0.797	0.578	0.980	0.714	0.971	0.705	1.192	0.821	1.426	0.856						
ET Tm2	0.277	0.323	0.293	0.335	0.316	0.368	0.295	0.343	0.320	0.353	0.463	0.488	1.342	0.930	0.332	0.366	0.447	0.487	0.987	0.756	3.370	1.440	3.978	1.587						
天气	0.234	0.273	0.238	0.275	0.241	0.283	0.242	0.279	0.279	0.301	0.284	0.324	0.300	0.342	0.318	0.323	0.318	0.360	0.289	0.322	0.597	0.495	0.546	0.469						
预期毒性	0.175	0.270	0.176	0.269	0.180	0.280	0.180	0.273	0.323	0.392	0.346	0.427	0.431	0.478	0.444	0.480	0.660	0.617	0.441	0.489	1.195	0.891	0.965	0.768						
交通量	0.429	0.306	0.440	0.310	0.447	0.313	0.430	0.305	0.951	0.535	0.663	0.425	0.749	0.446	1.453	0.815	1.914	0.936	1.248	0.684	1.534	0.811	1.551	0.821						
1 stCount	7		1		0		1		0		0		0		0		0		0		0		0		0		0		0	

表 4:基于 5% 训练数据的少样本学习。我们使用与表 1 相同的协议。所有结果均为平均值
来自四个不同的预测范围:H ∈ {96, 192, 336, 720}。完整结果见附录 E。

方法	时间-LLM (我们的)	GPT4TS (2023a)	线性 (2023)	PatchTST (2023)	TimesNet (2023)	FEDformer (2022)	Autoformer (2022)	Stationary (2022)	ETSformer (2022)	LightTS (2022)	DLinear (2023)	ETSformer (2022)	LightTS (2022)	DLinear (2023)	告密者(2021)	改革者 (2020年)
公制	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ET h1	0.627	0.543	0.681	0.560	0.750	0.611	0.694	0.569	0.925	0.647	0.658	0.562	0.722	0.598	0.943	0.646	1.189	0.839	1.451	0.903	1.225	0.817	1.241	0.835				
ET h2	0.382	0.418	0.400	0.433	0.694	0.577	0.827	0.615	0.439	0.448	0.463	0.454	0.441	0.457	0.470	0.489	0.809	0.681	3.206	1.268	3.922	1.653	3.527	1.472				
ET Tm1	0.425	0.434	0.472	0.450	0.400	0.417	0.526	0.476	0.717	0.561	0.730	0.592	0.796	0.620	0.857	0.598	1.125	0.782	1.123	0.765	1.163	0.791	1.264	0.826				
ET Tm2	0.274	0.323	0.308	0.346	0.399	0.426	0.314	0.352	0.344	0.372	0.381	0.404	0.388	0.433	0.341	0.372	0.534	0.547	1.415	0.871	3.658	1.489	3.581	1.487				
天气	0.260	0.309	0.263	0.301	0.263	0.308	0.269	0.303	0.298	0.318	0.309	0.353	0.310	0.353	0.327	0.328	0.333	0.371	0.305	0.345	0.584	0.527	0.447	0.453				
预期亏损	0.179	0.268	0.178	0.273	0.176	0.275	0.181	0.277	0.402	0.453	0.266	0.353	0.346	0.404	0.627	0.603	0.800	0.685	0.878	0.725	1.281	0.929	1.289	0.904				
交通量	0.423	0.298	0.434	0.305	0.450	0.317	0.418	0.296	0.867	0.493	0.676	0.423	0.833	0.502	1.526	0.839	1.859	0.927	1.557	0.795	1.591	0.832	1.618	0.851				
1 stCount	5		2		1		1		0		0		0		0		0		0		0		0		0		0	

例如 PatchTST、DLinear 和 TimesNet,我们的平均增强幅度分别超过8%、 12%和33%
相对 MSE。在 5% 的小样本学习场景中也可以看到类似的趋势,其中我们的
与 GPT4TS 相比,平均进步超过5%。与 PatchTST、DLinear 和
TimesNet、 TIME-LLM 表现出超过 20% 的惊人平均进步。

4.4零样本预测

设置。除了小样本学习之外,
LLM 具有潜力成为有效
零样本推理机（Kojima 等人,
2022）。在本节中,我们评估
零样本学习能力
重新编程的法学硕士
跨域自适应框架。具体来说,我们研究如何

表 5:零样本学习结果。红色:最佳,蓝色:次佳。附录 E 展示了我们的详细结果。

方法	TIME-LLM (我们的)		GPT4TS (2023a)		LLMTime (2023)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)	
公制	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ET h1 → ET h2	0.353	0.387	0.406	0.422	0.992	0.708	0.493	0.488	0.380	0.405	0.421	0.431		
蒸散时间 h1 → 蒸散时间 Tm2	0.273	0.340	0.325	0.363	1.867	0.869	0.415	0.452	0.314	0.360	0.327	0.361		
ET h2 → ET h1	0.479	0.474	0.757	0.578	1.961	0.981	0.703	0.574	0.565	0.513	0.865	0.621		
蒸散时间 T h2 → 蒸散时间 Tm2	0.272	0.341	0.335	0.370	1.867	0.869	0.328	0.386	0.325	0.365	0.342	0.376		
蒸散时间 Tm1 → 蒸散时间 T h2	0.381	0.412	0.433	0.439	0.992	0.708	0.464	0.475	0.439	0.438	0.457	0.454		
蒸散时间 Tm1 → 蒸散时间 Tm2	0.268	0.320	0.313	0.348	1.867	0.869	0.335	0.389	0.296	0.334	0.322	0.354		
ET Tm2 → ET T h2	0.354	0.400	0.435	0.443	0.992	0.708	0.455	0.471	0.409	0.425	0.435	0.443		
蒸散时间 Tm2 → 蒸散时间 Tm1	0.414	0.438	0.769	0.567	1.933	0.984	0.649	0.537	0.568	0.492	0.759	0.567		

模型在数据集上的表现
♣ 当它在另一个上优化时
数据集♠,其中模型尚未
遇到数据集中的任何数据样本♣。与小样本学习类似,我们使用长期
预测协议并利用 ETT 数据集对各种跨域场景进行评估。

结果。我们的简要结果见表 5。TIME -LLM 的表现始终优于最具竞争力的
基线大幅领先,MSE 降低率超过14.2% ,排名第二。考虑到
从少量结果来看,我们观察到重新编程 LLM 往往会产生更好的结果
在数据稀缺的情况下。例如,在 10% 的小样本训练中,我们相对于 GPT4TS 的总体错误率降低了
预测、5% 少量预测和零样本预测逐渐增加： 7.7%,
8.4%和22%。即使与该领域最新的方法 LLMTime 相比,
与同等规模（7B)的主干 LLM 相比， TIME-LLM 显示出了显著的改进
超过75%。我们认为这是因为我们的方法能够更好地激活法学硕士的知识
在执行时间序列任务时以资源高效的方式传输和推理能力。

4.5模型分析

语言模型变体。我们比较了两种具有不同容量的代表性主干模型
（表 6 中的A.1-4 ）。我们的结果表明,在 LLM 重新编程后,缩放定律仍然保持不变。我们
默认采用 Llama-7B 的全部容量,其性能明显优于其 1/4 容量版本
（A.2;包含前 8 个 Transformer 层)降低了14.5%。平均 MSE 降低了14.7% ,
在 GPT-2 (A.3) 上观察到,其性能略优于其变体 GPT-2 (6) (A.4) 2.7%。

跨模态比对。表 6 中的结果表明,消除任何一种斑块重编程
或以提示为前缀会损害 LLM 重新编程以获得有效时间序列的知识转移
预测。在缺乏表征对齐（B.1)的情况下,我们观察到平均性能显著下降9.2%,这在小样本任务中变得更加明显（超过
17%） 。
在TIME-LLM 中,提示行为是发挥 LLM 能力的关键要素
用于理解输入和任务。消除这一部分（B.2)会导致超过8%的
标准预测任务和小样本预测任务分别下降了19% 。我们发现,移除
输入统计数据（C.1)的影响最大,导致平均MSE增加10.2% 。这是

表 6:ETTh1 和 ETTm1 的消融预测结果分别提前 96 步和 192 步（报告的 MSE）。红色:最佳。

变体	长期预测				小样本预测			
	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192
A.1羊驼（默认； 32）	0.362	0.398	0.272	0.310	0.448	0.484	0.346	0.373
A.2火焰 (8)	0.389	0.412	0.297	0.329	0.567	0.632	0.451	0.490
A.3 GPT-2（12）	0.385	0.419	0.306	0.332	0.548	0.617	0.447	0.509
A.4 GPT-2（6）	0.394	0.427	0.311	0.342	0.571	0.640	0.468	0.512
B.1无补丁重新编程B.2无提示符作为前缀	0.410	0.412	0.310	0.342	0.498	0.570	0.445	0.487
	0.398	0.423	0.298	0.339	0.521	0.617	0.432	0.481
C.1无数据集上下文C.2无任务说明C.3无统计上下文	0.402	0.417	0.298	0.331	0.491	0.538	0.392	0.447
	0.388	0.420	0.285	0.327	0.476	0.529	0.387	0.439
	0.391	0.419	0.279	0.347	0.483	0.547	0.421	0.461

表 7： TIME-LLM 对 ETTh1 进行不同步骤预测的效率分析。

长度	ETTh1-96			ETTh1-192			ETTh1-336			ETTh1-512		
公制	参数 (M)	内存 (MiB)	速度 (s/iter)	参数 (M)	内存 (MiB)	速度 (s/iter)	参数 (M)	内存 (MiB)	速度 (s/iter)	参数(M)	内存 (MiB)	速度 (s/iter)
D.1看涨期权 (32) 3404.53 D.2看涨期权 (8) 975.83 6.39 D.3 （无法学硕士）	32136	0.517	3404.57	33762	0.582	3404.62	37988	0.632	3404.69	39004	0.697	
	11370	0.184	975.87	12392	0.192	975.92	13188	0.203	976.11	13616	0.217	
	3678	0.046	6.42	3812	0.087	6.48	3960	0.093	6.55	4176	0.129	

参与,因为外部知识可以通过提示自然地融入,以促进学习
和推理。此外,为法学硕士提供清晰的任务指导和输入上下文（例如，
数据集字幕)也很有用（即C.2和C.1;分别获得超过7.7%和9.6%）。

重编程解释。我们提供了一个关于 ETTh1 的案例
研究,该研究对 48 个时间序列块进行了重编程，
其中包含 100 个文本
图 5 中的原型。前 4 个子图
可视化重编程的优化
空间从随机初始化（a)到良好优化（d）。我们发现只
有一小部分原型（列)参与了子图（e)中输入块（行）
的重新编程。

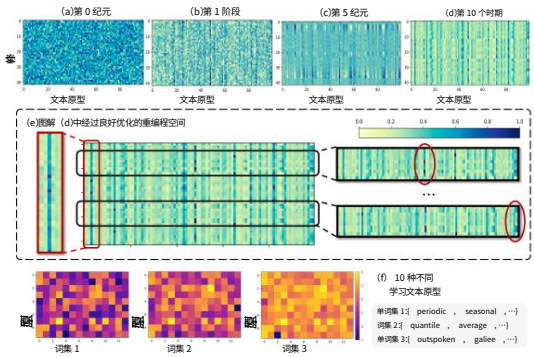


图 5:补丁重编程的展示。

此外,补丁会经历不同的表现形式
通过原型的不同组合。
这表明：(1)文本原型学会总结语言线索,并且少数
与表示信息高度相关
局部时间序列补丁,我们将其可视化为
在子图(f)中随机选择 10 个。我们的结果表明,描述
时间序列属性（即词集 1 和 2）；(2)块通常具有不同的底层语义,需要不同的原型来表示。

重编程效率。表 7 提供了TIME-LLM的整体效率分析，
并且没有主干 LLM。我们提出的重编程网络本身(D.3)是轻量级的
在激活法学硕士的时间序列预测能力方面（即,可训练的人数不足 660 万
参数;仅占 Llama-7B 总参数的0.2%左右），并且整体效率
TIME-LLM 实际上受到杠杆主干课程（例如D.1和D.2)的限制。这是有利的
甚至与参数高效的微调方法（例如 QLoRA (Dettmers et al., 2023))相比
平衡任务绩效和效率。

5结论和未来工作

TIME-LLM 有望通过以下方式适应冻结的大型语言模型进行时间序列预测
将时间序列数据重新编程为文本原型,对于法学硕士来说更自然,并提供自然
通过提示作为前缀来增强推理能力的语言指导。评估表明,适应
LLM 的表现可以超越专门的专家模型,这表明它们有潜力成为有效的时间序列
机器。我们的研究结果还提供了一个新颖的见解:时间序列预测可以被视为另一种“语言”任务,可以通过现成的LLM
来解决,并通过我们的Time-LLM框架实现最先进的性能。进一步的研究应该探索最佳的重新编程

通过持续的预训练,用明确的时间序列知识丰富 LLM，
并建立跨时间序列、自然语言和
其他模式。此外,应用重新编程框架,使法学硕士能够更广泛地
还应考虑时间序列分析能力或其他新功能。

参考

Shaojie Bai、J Zico Colter 和 Vladlen Koltun。通用卷积和循环网络在序列建模中的实证评估。arXiv 预印本 arXiv:1803.01271，

George EP Box、Gwilym M Jenkins、Gregory C Reinsel 和 Greta M Ljung。时间序列分析：预测与控制。John Wiley & Sons，2015年。

Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell 等人。语言模型是小样本学习器。《神经信息处理系统进展》，33:1877–1901，2020 年。

Cristian Challu、Kin G Olivares、Boris N Oreshkin、Federico Garza、Max Mergenthaler 和 Artur Dubrawski。N-hits：用于时间序列预测的神经分层插值。AAAI 人工智能会议论文集，2023a。

Cristian Challu、Kin G Olivares、Boris N Oreshkin、Federico Garza Ramirez、Max Mergenthaler Canseco 和 Artur Dubrawski。Nhits：用于时间序列预测的神经分层插值。载于《AAAI 人工智能会议论文集》，第 37 卷，第 6989-6997 页，2023b。

Ching Chang、Wen-Chih Peng 和 Tien-Fu Chen。Llm4ts：时间序列的两阶段微调使用预先训练的 llms 进行预测。arXiv 预印本 arXiv:2308.08469，2023 年。

Pin-Yu Chen。模型重编程：资源高效的跨域机器学习。arXiv 预印本 arXiv:2202.10629，2022 年。

Zhixuan Chu, Hongyan Hao, Xin Ouyang, Simeng Wang, Yan Wang, Yue Shen, Jinjie Gu, Qing Cui, Longfei Li, Siqiao Xue, et al. Leveraging large language models for pre-trained recommender systems. arXiv preprint arXiv:2308.10837, 2023.

Shohreh Deldari、Hao Xue、Aaqib Saeed、Jiayuan He、Daniel V Smith 和 Flora D Salim。《超越视觉：多模态和时间数据的自监督表征学习综述》。arXiv 预印本 arXiv:2206.02353，2022 年。

蒂姆·德特默斯、阿蒂多罗·帕尼奥尼、阿里·霍尔兹曼和卢克·泽特尔莫耶。 Qlora：高效微调量化法学硕士。神经信息处理系统进展，2023年。

Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。Bert：用于语言理解的深度双向 Transformer 预训练。载于 2019 年计算语言学协会北美分会会议论文集：人类语言技术，2018 年。

Hassan Ismail Fawaz、Germain Forestier、Jonathan Weber、Lhassane Idoumghar 和 Pierre-Alain Muller。时间序列分类的迁移学习。IEEE 大数据国际会议，第 1367-1376 页。IEEE，2018 年。

Nate Gruver、Marc Anton Finzi、Shikai Qiu 和 Andrew Gordon Wilson。大型语言模型是零样本时间序列预测器。《神经信息处理系统进展》，2023 年。

Julien Herzen、Francesco Lassig、Samuele Giuliano Piazzetta、Thomas Neuer、Leo Tafti、Guillaume Raille、Tomas Van Pottelbergh、Marek Pasieka、Andrzej Skrodzki、Nicolas Huguenin 等。 Darts：用户友好的现代时间序列机器学习。机器学习研究杂志，23(1):5442–5447，2022。

塞普·霍赫赖特和尤尔根·施米德胡贝尔。长短期记忆。 神经计算，9(8):1735–1780，1997 年。

Ming Jin、Huan Yee Koh、Qingsong Wen、Daniele Zambon、Cesare Alippi、Geoffrey I Webb、Irwin King 和 Shirui Pan。时间序列图神经网络综述：预测、分类、插补和异常检测。arXiv 预印本 arXiv:2307.03759, 2023a。

Ming Jin, Qingsong Wen, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, et al. Large models for time series and spatio-temporal data: A survey and outlook. arXiv preprint arXiv:2310.10196, 2023b.

Taesung Kim、Jinhee Kim、Yunwon Tae、Cheonbok Park、Jang-Ho Choi 和 Jaegul Choo。可逆实例归一化，用于针对分布偏移进行精确的时间序列预测。2021 年国际学习表征会议。

Diederik P. Kingma 和 Jimmy Ba。《Adam：一种随机优化方法》。国际学习表征大会，2015 年。

Nikita Kitaev、 ukasz Kaiser 和 Anselm Levskaya。Reformer：高效的Transformer。2020年国际学习表征大会。

Takeshi Kojima、Shixiang Shane Gu、Machel Reid、Yutaka Matsuo 和 Yusuke Iwasawa。大型语言模型是零样本推理器。《神经信息处理系统进展》，35:22199–22213，2022 年。

Michael Leonard，需求规划的促销分析与预测：实用的时间序列
方法。附有展品，1，2001 年。

李娜、唐纳德·M·阿诺德、道格拉斯·G·唐、丽贝卡·巴蒂、约翰·布莱克、蒋飞、汤姆·考特尼、玛丽安·怀托、里克·特里富诺夫和南希·M·赫德尔。通过整合机器学习、统计建模和库存优化，实现红细胞从需求预测到库存订购决策。《输血》，62(1):87–99，2022年。

刘恒波，马自清，杨林晓，周天，夏睿，王毅，温青松，孙良。Sadi：一种用于极端事件下电力负荷预测的自适应分解可解释框架。IEEE 声学、语音和信号处理国际会议，2023a。

Xin Liu、Daniel McDuff、Geza Kovacs、Isaac Galatzer-Levy、Jacob Sunshine、Jiening Zhan、Ming-Zher Poh、Shun Liao、Paolo Di Achille 和 Shwetak Patel。大型语言模型是少量样本的健康学习器。arXiv 预印本 arXiv:2305.15525，2023b。

刘勇, 吴海旭, 王建民, 龙明生. 非平稳Transformer：探索时间序列预测中的平稳性. 神经信息处理系统进展, 35:9881–9893, 2022.

Ziyang Ma, Wen Wu, Zhisheng Zheng, Yiwei Guo, Qian Chen, Shiliang Zhang, and Xie Chen.
利用语音 ptm、文本 llm 和情感 tts 进行语音情感识别。arXiv 预印本 arXiv:2309.10294，2023 年。

Spyros Makridakis 和 Michele Hibon。m3 竞赛：结果、结论和影响。
国际预测杂志，16（4）：451-476，2000。

Spyros Makridakis、Evangelos Spiliotis 和 Vassilios Assimakopoulos。M4 竞赛：结果、发现、结论及未来发展方向。《国际预测杂志》，34(4):802–808，2018 年。

Igor Melnyk、Vijil Chenthamarakshan、Pin-Yu Chen、Payel Das、Amit Dhurandhar、Inkit Padhi 和 Devleena Das。重新编程预训练的语言模型以填充抗体序列。国际机器学习会议，2023 年。

Suvir Mirchandani、Fei Xia、Pete Florence、Danny Driess、Montserrat Gonzalez Arenas、Kanishka Rao、Dorsa Sadigh、Andy Zeng 等人。大型语言模型作为通用模式机器。载于2023年第七届机器人学习年会论文集。

Diganta Misra、Agam Goyal、Bharat Runwal 和 Pin Yu Chen。约束条件下的重新编程：重新审视彩票的高效可靠可转让性。arXiv 预印本 arXiv:2308.14969，2023 年。

Yuqi Nie、Nam H Nguyen、Phanwadee Sinthong 和 Jayant Kalagnanam。“一个时间序列胜过 64 个字：使用 Transformer 进行长期预测”。2023 年国际学习表征大会。

OpenAI。GPT-4 技术报告，2023 年。

Boris N Oreshkin、Dmitri Carpov、Nicolas Chapados 和 Yoshua Bengio。N-beats：用于可解释时间序列预测的神经基础扩展分析。2020 年国际学习表征大会。

Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga 等人。Pytorch：一种命令式高性能深度学习库。《神经信息处理系统进展》，第 32 卷，2019 年。

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever 等。语言模型是无监督的多任务学习器。OpenAI 博客，1(8):9，2019。

Stephen H Schneider 和 Robert E Dickinson。气候建模。《地球物理评论》，12(3):447–493，1974年。

唐一红、屈敖、周海峰、林伟、黄世昌和马伟。领域对抗性时空网络：一个可迁移的跨城市短期交通预测框架。载于第31届ACM信息与知识管理国际会议论文集，第1905-1915页，2022年。

Hugo Touvron、Thibaut Lavril、Gautier Izacard、Xavier Martinet、Marie-Anne Lachaux、Timothee´ Lacroix、Baptiste Roziere、Naman Goyal、Eric Hambro、Faisal Azhar 等人。Llama：开放高效的基础语言模型。arXiv 预印本 arXiv:2302.13971，2023 年。

Maria Tsimpoukelli、Jacob L Menick、Serkan Cabi、SM Eslami、Oriol Vinyals 和 Felix Hill。基于冻结语言模型的多模态小样本学习。《神经信息处理系统进展》，34:200–212，2021 年。

Ashish Vaswani、Noam Shazeer、Niki Parmar、Jacob Uszkoreit、Llion Jones、Aidan N Gomez、 ukasz Kaiser 和 Illia Polosukhin。您所需要的就是关注。神经信息处理系统的进展，2017 年 30 月。

Ria Vinod、Pin-Yu Chen 和 Payel Das。用于分子表征学习的重编程语言模型。2020 年神经信息处理系统年会。

Yan Wang, Zhixuan Chu, Xin Ouyang, Simeng Wang, Hongyan Hao, Yue Shen, Jinjie Gu, Siqiao Xue, James Y Zhang, Qing Cui, et al. Enhancing recommender systems with large language model reasoning graphs. arXiv preprint arXiv:2308.10835, 2023.

Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. 时间序列中的 Transformer：一项综述。国际人工智能联合会议，2023 年。

Gerald Woo、Chenghao Liu、Doyen Sahoo、Akshat Kumar 和 Steven Hoi。等效项：时间序列预测的指数平滑变换。arXiv 预印本 arXiv:2202.01381。

吴海旭，徐杰辉，王建民，龙明胜。Autoformer：用于长期序列预测的自相关分解变换器。《神经信息处理系统进展》，34:22419–22430，2021年。

吴海旭、胡腾格、刘勇、周航、王建民和龙明胜。Timesnet：用于通用时间序列分析的时间二维变异建模。国际学习表征会议，2023 年。

Hao Xue 和 Flora D Salim。基于提示的时间序列预测：一项新任务和数据集。arXiv 预印本 arXiv:2210.08964，2022 年。

Chao-Han Huck Yang、Yun-Yun Tsai 和 Pin-Yu Chen。Voice2series：重新编程用于时间序列分类的声学模型。国际机器学习会议，第 35 页。11808–11819。PMLR，2021 年。

Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on 多模态大型语言模型。arXiv 预印本 arXiv:2306.13549，2023 年。

曾爱玲、陈慕熙、张磊和徐强。Transformer 对时间序列预测有效吗?载于《AAAI 人工智能会议论文集》,第 37 卷,第 11121–11128 页,2023 年。

Kexin Zhang, Qingsong Wen, Chaoli Zhang, Rongyao Cai, Ming Jin, Yong Liu, James Zhang, Yuxuan Liang, Guansong Pang, Dongjin Song, et al. Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. arXiv preprint arXiv:2306.10125, 2023.

Tianping Zhang, Yizhuo Zhang, Wei Cao, Jiang Bian, Xiaohan Yi, Shun Zheng, and Jian Li. Less is more: Fast multivariate time series forecasting with light sampling-oriented mlp structures. arXiv preprint arXiv:2207.01186, 2022a.

张翔,赵子远,Theodoros Tsiligkaridis 和 Marinka Zitnik。基于时频一致性的自监督时间序列对比预训练。《神经信息处理系统进展》,35:3988–4003,2022b。

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer:超越高效 Transformer 的长序列时间序列预测。载于《AAAI 人工智能会议论文集》,第 35 卷,第 11106-11115 页,2021 年。

周天、马自清、温青松、王雪、孙良和金蓉。Fedformer:用于长期序列预测的频率增强分解变压器。国际机器学习会议,第 27268-27286 页。PMLR,2022 年。

周天,牛培松,王雪,孙亮,金蓉。一刀切:基于预训练流式细胞术的功率通用时间序列分析。神经信息处理系统进展,36,2023a。

周云逸、储志轩、阮一佳、金戈、黄雨辰和李胜。ptse:一种用于概率时间序列预测的多模型集成方法。第32届国际人工智能联合会议,2023b。

更相关的工作

任务特定学习。我们对任务特定学习的相关工作进行了扩展,尤其关注与我们进行比较的最相关的模型。近期的研究通过结合信号处理原理（例如补丁、指数平滑、分解和频率分析）,改进了 Transformer（Vaswani 等人,2017）,使其能够进行时间序列预测。

例如,PatchTST（Nie et al., 2023)将时间序列分割成patch,作为Transformer的输入token。这保留了局部语义,减少了注意力机制所需的计算/内存,并允许更长的历史记录。与其他Transformer模型相比,它提高了长期预测的准确性。它在自监督预训练和迁移学习中也取得了优异的表现。ETSformer（Woo et al., 2022)将指数平滑原理融入Transformer的注意力机制中,以提高准确性和效率。它使用指数平滑注意力机制和频率注意力机制来取代标准的自注意力机制。FEDformer（Zhou et al., 2022)将Transformer与季节性趋势分解相结合。分解捕捉全局特征,而Transformer捕捉细节结构。

它还利用频率增强进行长期预测。这比标准 Transformer 提供了更好的性能和效率。Autoformer（Wu 等人,2021)使用具有自相关的分解架构,以实现对复杂序列的渐进分解能力。

自相关基于序列周期性设计,用于进行依赖关系发现和表征聚合。其效率和准确率均优于自注意力机制。

虽然这些方法与原生 Transformer 相比提高了效率和准确性,但它们大多是针对特定领域内的窄带预测任务进行设计和优化的。这些模型通常在小型、特定领域的数据集上进行端到端训练。虽然这些专门的模型在目标任务上取得了优异的表现,但它们牺牲了对现实世界中遇到的各种时间序列数据的多功能性和泛化能力。这种狭窄的关注点限制了它们对新数据集和新任务的适用性。为了推进时间序列预测,需要更灵活、更广泛适用的模型,这些模型能够适应新的数据分布和任务,而无需进行大量的再训练。理想的模型应该学习能够跨领域迁移知识的鲁棒时间序列表示。开发这种具有广泛应用能力的预测模型仍然是一个悬而未决的挑战。

根据我们对相关前期工作的讨论,近期研究已开始通过预训练和架构创新来探索模型的多功能性。然而,要实现我们在本研究中推进的真正通用的预测系统,我们还需要进一步的努力。

跨模态自适应。我们对跨模态自适应的相关工作进行了扩展概述,特别关注时间序列和其他数据模态模型重编程的最新进展。模型重编程是一种资源高效的跨领域学习方法,它涉及将一个领域（源)中成熟的预训练模型调整到处理不同领域（目标)中的任务,而无需进行模型微调,即使这些领域差异显著,正如 Chen (2022) 所述。在时间序列数据的背景下,Voice2Series (Yang et al., 2021) 通过变换时间序列以拟合模型并将输出重新映射到新标签,调整了语音识别中的声学模型以进行时间序列分类。

类似地,LLMTime（Gruver 等人,2023)将 LLM 调整为零样本时间序列预测,专注于对主干 LLM 的输入时间序列进行有效的标记化,然后自回归地生成预测。与这些方法不同， TIME-LLM 不直接编辑输入时间序列。相反,它建议使用源数据模态对时间序列进行重新编程,并促使 LLM 充分发挥其在标准、小样本和零样本场景中作为多功能预测器的潜力。该领域的其他值得注意的研究（主要在生物学领域)包括 R2DL（Vinod 等人,2020)和 ReproBert（Melnyk 等人,2023）,它们使用词向量嵌入对氨基酸进行重新编程。我们的块重编程方法的一个关键区别在于,与完整的氨基酸集合不同,时间序列块并不构成完整的集合。因此,我们建议优化一小组文本原型及其到时间序列补丁的映射,而不是直接优化两个完整集合（例如词汇和氨基酸)之间的大型转换矩阵。

B实验细节

B.1实施

我们主要遵循 (Wu et al., 2023) 中的实验配置,在https://github.com/thuml/Time-Series-Library中的统一评估流程中对所有基线进行

公平比较。我们使用 Llama-7B (Touvron et al., 2023) 作为默认主干模型，除非
除非另有说明。我们所有的实验都重复了三次，并报告平均结果。
我们的模型实现在 PyTorch (Paszke et al., 2019) 上，所有实验均在
NVIDIA A100-80G GPU。我们的详细模型配置见附录 B.4，我们的代码如下：
可在https://github.com/KimMeen/Time-LLM 上获取。

技术细节。我们从三个方面提供了 TIME-LLM 的更多技术细节：（1）
学习文本原型；（2）计算时间序列中的趋势和滞后以用于提示；
以及（3）输出投影的实现。识别一小组文本原型
$E' \in R^{V' \times D}$ 从 $E \in R^{V \times D}$，我们学习一个矩阵 $W \in R^{V' \times V}$ 作为中介。描述
为了用自然语言描述整体时间序列趋势，我们计算连续时间步长之间的差异总和。总和大于 0 表示上升趋势，总和小于 0 表示下降趋势。
下降趋势。此外，我们计算时间序列的前 5 个滞后值，通过使用快速傅里叶变换计算自相关性并选择具有最高

相关值。在我们打包并前馈提示和补丁嵌入 $O(i) \in R^{功率 \times 功率}$
通过冻结的 LLM，我们丢弃前缀部分并获得输出表示，表示为
作为 $O \in R^{P \times D}$。随后，我们遵循 PatchTST (Nie et al., 2023) 并展平 O_i 变成一维
长度为 $P \times D$ 的张量，然后线性投影为 $Y_i \in R^H$。

B.2数据集详细信息

数据集统计数据总结在表 8 中。我们评估了
八个不同的基准，包括四个 ETT 数据集 (Zhou et al., 2021)（即，
ETTh1、ETTh2、ETTm1 和 ETTm2）、天气、电力、交通和 ILI 来自（Wu 等人，2023 年）。
此外，我们评估了 M4 基准（Makri-dakis 等人，2018)和 M3 基准（Makridakis & Hibon，2000)中的季度数据集的短期
预测性能。

表 8:数据集统计数据来自 (Wu et al., 2023)。维度表示时间
系列（即通道），数据集大小按（训练、验证、测试)进行组织。

任务	数据集	尺寸系列长度		数据集大小	频域	
	气管插管	7 {96, 192, 336, 720} (34465, 11521, 11521) 15分钟 温度				
长期 ETTm2		7 {96, 192, 336, 720} (34465, 11521, 11521) 15分钟 温度				
预测 ETTh1		7 {96, 192, 336, 720} (8545, 2881, 2881) 7 {96,			1小时	温度
	ETTh2	192, 336, 720} (8545, 2881, 2881) 321 {96,			1小时	温度
	电	192, 336, 720} (18317, 2633, 5261) 862 {96, 192,			1小时	电
	交通	336, 720} (12185, 1757, 3509) 21 {96, 192, 336,			1小时交通	
	天气	720} (36792, 5271, 10540)			10分钟	天气
	或者	7	{24, 36, 48, 60}	(617, 74, 170)	1周	疾病
	M3-季度 1		8	(756, 0, 756)	季刊	多种的
	M4-每年	1	6	(23000, 0, 23000)	年度人口统计	
	M4-季度 1		8	(24000, 0, 24000)	季刊	金融
短期M4月度		1	18	(48000, 0, 48000)	每月	行业
预测 M4-Weakly		1	十三	(359, 0, 359)	弱地	宏
	M4-每日	1	14	(4227, 0, 4227)	日常的	微
	M4-每小时	1	四十八	(414, 0, 414)	每小时	其他

电力变压器温度（ETT;反映长期电力的指标
部署)基准由两年的数据组成，数据来源于中国两个县，
并细分为四个不同的数据集，每个数据集具有不同的采样率:ETTh1 和 ETTh2，
每 1 小时采样一次，ETTm1 和 ETTm2 每 15 分钟采样一次
级别。ETT 数据集中的每个条目都包含六个电力负荷特征和一个目标变量，
称为“油温”。电力数据集包含以下日期的电力消耗记录：
321 位客户，以 1 小时采样率进行测量。天气数据集包含一年的记录

来自德国 21 个气象站,采样频率为 10 分钟。交通
数据集包括 862 个传感器记录的高速公路系统占用率数据
覆盖加利福尼亚州,采样率为 1 小时。流感样疾病 (ILI) 数据集
包含患有严重流感并伴有并发症的患者的记录。

M4 基准包含 100K 个时间序列,收集自各种常见的领域
商业、金融和经济预测。这些时间序列被划分为六个
独特的数据集,每个数据集的采样频率各不相同,从每年到每小时不等。
M3-Quarterly 数据集包含 M3 基准中的 756 个季度采样时间序列。这些
该系列分为五个不同的领域:人口、微观、宏观、行业和金融。

B.3评估指标

对于评估指标,我们利用均方误差(MSE)和平均绝对误差(MAE)来
长期预测。对于 M4 基准的短期预测,我们采用对称平均绝对百分比误差 (SMAPE)、平均绝对尺度误差 (MASE) 和总体

加权平均值 (OWA),类似于 N-BEATS (Oreshkin 等人,2020)。请注意,OWA 是一个特定的指标
在M4比赛中使用。这些指标的计算如下:

$$\text{均方误差} = \frac{1}{\text{哈}}\sum_{h=1}^{T}(Yh - Y\ \ h^{2}), \qquad \text{IS} = \frac{1}{\text{哈}}\sum_{h=1}^{\text{哈}}|Yh - Y\ \ h|,$$

$$\text{平均孔径比} = \frac{200}{\text{哈}}\sum_{h=1}^{\text{哈}}\frac{|Yh - Y\ \ h|}{|Yh| + |Y\ \ h|}, \qquad \text{地图} = \frac{100}{\text{哈}}\sum_{h=1}^{\text{哈}}\frac{|Yh - Y\ \ h|}{|是|},$$

$$\text{质量} = \frac{1}{\text{哈}}\sum_{h=1}^{\text{哈}}\frac{|Yh - Y\ \ h|}{\frac{1}{\text{氢键}}\sum_{j=s+1}^{\text{哈}}|Yj - Yj-s|}, \qquad \text{OWA} = \frac{1}{2}\left(\frac{\text{斯玛佩}}{\text{SMAPENa¨ ve2}} + \frac{\text{群众}}{\text{MASENa¨ ve2}}\right),$$

其中 s 是时间序列数据的周期。H 表示数据点的数量(即预测
在我们的案例中是地平线)。Yh和Y h是第 h 个基本事实和预测,其中 h ∈{1, · · · , H}。

B.4模型配置

我们的模型针对不同任务和数据集的配置汇总在表 9 中。
默认情况下,所有实验均采用 Adam 优化器(Kingma & Ba,2015)。
具体来说,文本原型的数量 V ,短期保持不变为 100 和 1000,
长期预测任务。我们充分利用了 Llama-7B 模型,并将所有任务的主干模型层数保持在 32 层作为标准。输入长度 T 表示

原始输入时间序列数据中存在的时间步数。块维度dm表示重新编程之前嵌入时间序列块的隐藏维度。最后,

头 K 与用于块重编程的多头交叉注意机制相关。在四个
表 9 最右侧列详细说明了与模型训练相关的配置。

表 9:TIME-LLM 实验配置概览。“LTF”和“STF”表示
分别为长期预测和短期预测。

任务数据集/配置	模型超参数					训练过程		
	文本原型 V	骨干层输入长度T补丁尺寸dm头K LR损失批量大小Epochs						
LTF-ETTh1	1000	=+=	512	16	8	10−3均方误差	16	50
LTF-ETTh2	1000	=+=	512	16	8	10−3均方误差	16	50
LTF-ETTm1	1000	=+=	512	16	8	10−3均方误差	16	100
LTF-ETTm2	1000	=+=	512	16	8	10−3均方误差	16	100
LTF - 天气	1000	=+=	512	16	8	10−2均方误差	8	100
LTF - 电力	1000	=+=	512	16	8	10−2均方误差	8	100
LTF - 交通	1000	=+=	512	16	8	10−2均方误差	8	100
LTF-ILI	100	=+=	96	16	8	10−2均方误差	16	50
STF - M3-季度	100	=+=	2 × H †	=+=	8	10−4 SMAPE 32		50
STF-M4	100	=+=	2 × H †	=+=	8	10−4 SMAPE 32		50

† H 代表 M4 和 M3 数据集的预测范围。
LR表示初始学习率。

C超参数敏感度

我们针对 TIME-LLM 中的四个重要超参数进行了超参数敏感性分析:即主干模型层数、文本原型数量 V、时间序列输入长度 T 和重编程交叉注意头的数量 K。相关结果如图 6 所示。通过分析,我们得出以下观察结果:(1)主干 LLM 中的 Transformer 层数与TIME-LLM 的性能呈正相关,表明在 LLM 重编程后,缩放定律得以保留;(2)通常,获取更多文本原型可提高性能。我们假设有限数量的原型 V 可能会在聚合语言线索时引入噪声,从而阻碍有效学习对于表征输入时间序列补丁至关重要的高度代表性原型;(3)输入时间长度 T 与预测准确性呈直接关系,在预测扩展范围时尤为明显。这一观察结果是合乎逻辑的,并且与传统的时间序列模型一致; (4)在重新编程输入块期间增加注意力头的数量被证明是有利的。

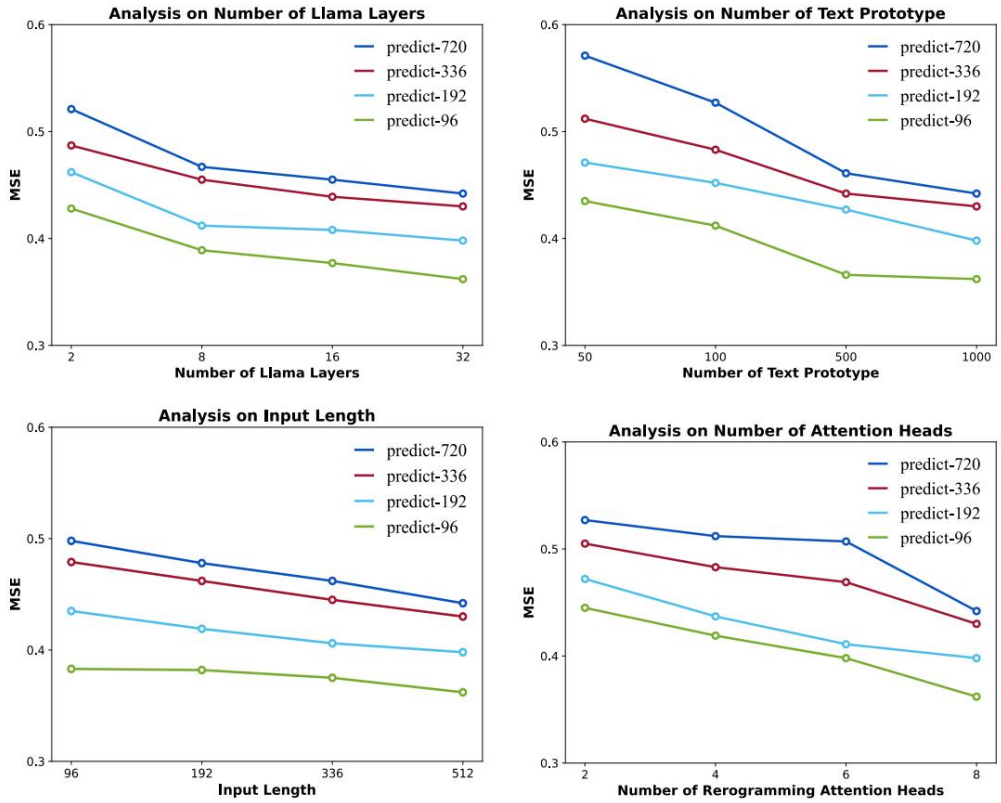


图 6:ETTh1 数据集上的超参数敏感性分析。

D长期和短期预测

D.1长期预测

TIME-LLM仅通过对最小的 Llama 模型进行重新编程并保持其完整性,便在 8 个时间序列基准测试的 40 个实例中36 个实例中达到了 SOTA 性能。这凸显了 LLM 作为稳健可靠的时间序列预测器的巨大潜力。此外,我们将所提出的方法与表 11 中的其他成熟基准进行了比较。比较对象包括三种著名的统计方法(AutoARIMA、AutoTheta 和 AutoETS)(Herzen 等人,2022)以及两种近期的时间序列模型 N-HiTS(Challu 等人,2023b)和 N-BEATS(Ore-shkin 等人,2020)。值得注意的是, TIME-LLM 在所有情况下都确保了 SOTA 性能,在 MSE 和 MAE 方面分别以超过22%和16%的显著优势超越第二好的结果。

表 10:完整的长期预测结果。我们为 ILI 设置预测范围 H ∈ {24, 36, 48, 60},为其他设置预测范围 {96, 192, 336, 720}。值越低,性能越好。红色:最佳,蓝色:次佳。

方法	TIME-LLM		GPT4TS			线性			PatchTST			TimesNet			FEDformer			Autoformer			Stationary			ETSformer			LightTS				告密者		改革者	
公制	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
电	96	0.362	0.392	0.376	0.397	0.375	0.399	0.370	0.399	0.384	0.402	0.376	0.419	0.449	0.459	0.513	0.491	0.494	0.479	0.424	0.432	0.865	0.713	0.837	0.728	0.192	0.398	0.418	0.416	0.418	0.405	0.416	0.413	0.421	0.436	0.429
	0.420	0.448	0.500	0.482	0.534	0.504	0.538	0.504	0.475	0.462	1.008	0.792	0.923	0.766	336	0.430	0.427	0.442	0.433	0.439	0.443	0.422	0.436	0.491	0.469	0.459	0.465	0.521	0.496	0.588	0.535	0.574	0.521	0.518	0.488	1.107
	0.809	1.097	0.835	720	0.442	0.457	0.477	0.456	0.472	0.490	0.447	0.466	0.521	0.500	0.506	0.507	0.514	0.512	0.643	0.616	0.562	0.535	0.547	0.533	1.181	0.865	1.257	0.889	平均	0.408	0.423	0.465	0.455	0.422	0.437	0.413
	0.430	0.458	0.450	0.440	0.460	0.496	0.487	0.570	0.537	0.542	0.510	0.491	0.479	1.040	0.795	1.029	0.805																			
ETTh1	96	0.368	0.398	0.375	0.399	0.376	0.398	0.374	0.400	0.391	0.374	0.358	0.397	0.346	0.388	0.476	0.458	0.340	0.391	0.397	0.437	3.755	1.525	2.626	1.317	192	0.329	0.375	0.354	0.389	0.383	0.418	0.339	0.379	0.402	0.414
	0.429	0.439	0.456	0.452	0.512	0.493	0.430	0.439	0.520	0.504	5.602	1.931	11.12	2.979	336	0.368	0.409	0.373	0.407	0.448	0.465	0.329	0.380	0.452	0.452	0.496	0.487	0.482	0.486	0.552	0.551	0.485	0.479	0.626	0.559	4.721
	1.835	9.323	2.769	720	0.372	0.420	0.406	0.441	0.605	0.551	0.379	0.422	0.462	0.468	0.463	0.474	0.515	0.511	0.562	0.560	0.500	0.497	0.863	0.672	3.647	1.625	3.874	1.697	平均	0.334	0.383	0.381	0.412	0.431	0.446	0.330
	0.379	0.414	0.427	0.437	0.449	0.450	0.459	0.526	0.516	0.439	0.452	0.602	0.543	4.431	1.729	6.736	2.191																			
ETTm1	96	0.272	0.334	0.292	0.346	0.299	0.343	0.290	0.342	0.338	0.375	0.379	0.419	0.505	0.475	0.386	0.398	0.375	0.398	0.374	0.400	0.672	0.571	0.538	0.528	192	0.310	0.358	0.332	0.372	0.335	0.365	0.332	0.369	0.374	0.387
	0.426	0.441	0.553	0.496	0.459	0.444	0.408	0.410	0.400	0.407	0.795	0.669	0.658	0.592	336	0.352	0.384	0.366	0.394	0.369	0.386	0.366	0.392	0.410	0.411	0.445	0.459	0.621	0.537	0.495	0.464	0.435	0.428	0.438	0.438	1.212
	0.871	0.898	0.721	720	0.383	0.411	0.417	0.421	0.425	0.421	0.416	0.420	0.478	0.450	0.543	0.490	0.671	0.561	0.585	0.516	0.499	0.462	0.527	0.502	1.166	0.823	1.102	0.841	平均	0.329	0.372	0.388	0.403	0.357	0.378	0.351
	0.380	0.400	0.406	0.448	0.452	0.588	0.517	0.481	0.456	0.429	0.425	0.435	0.437	0.961	0.734	0.799	0.671																			
ETTm2	96	0.161	0.253	0.173	0.262	0.167	0.269	0.165	0.255	0.187	0.267	0.203	0.287	0.255	0.339	0.192	0.274	0.189	0.280	0.209	0.308	0.365	0.453	0.658	0.619	192	0.219	0.293	0.229	0.301	0.224	0.303	0.220	0.292	0.249	0.309
	0.269	0.328	0.281	0.340	0.280	0.339	0.253	0.319	0.311	0.382	0.533	0.563	1.078	0.827	336	0.271	0.329	0.286	0.341	0.281	0.342	0.274	0.329	0.321	0.351	0.325	0.366	0.339	0.372	0.334	0.361	0.314	0.357	0.442	0.466	1.363
	0.887	1.549	0.972	720	0.352	0.379	0.378	0.401	0.397	0.421	0.362	0.385	0.408	0.403	0.421	0.415	0.433	0.432	0.417	0.413	0.414	0.413	0.675	0.587	3.379	1.338	2.631	1.242	平均	0.251	0.313	0.284	0.339	0.267	0.333	0.255
	0.315	0.291	0.333	0.305	0.349	0.327	0.371	0.306	0.347	0.293	0.342	0.409	0.436	1.410	0.810	1.479	0.915																			
天气	96	0.147	0.201	0.162	0.212	0.176	0.237	0.149	0.198	0.172	0.220	0.217	0.296	0.266	0.336	0.173	0.223	0.197	0.281	0.182	0.242	0.300	0.384	0.689	0.596											
	192	0.189	0.234	0.204	0.248	0.220	0.282	0.194	0.241	0.219	0.261	0.276	0.336	0.307	0.367	0.245	0.285	0.237	0.312	0.227	0.287	0.598	0.544	0.752	0.638											
	336	0.262	0.279	0.254	0.286	0.265	0.319	0.245	0.282	0.280	0.306	0.339	0.380	0.359	0.395	0.321	0.338	0.298	0.353	0.282	0.334	0.578	0.523	0.639	0.596											
	720	0.304	0.316	0.326	0.337	0.333	0.362	0.314	0.334	0.365	0.359	0.403	0.428	0.419	0.428	0.414	0.410	0.352	0.288	0.352	0.386	1.059	0.741	1.130	0.792											
	平均值	0.225	0.257	0.237	0.270	0.248	0.300	0.225	0.264	0.259	0.287	0.309	0.360	0.338	0.382	0.288	0.314	0.271	0.334	0.261	0.312	0.634	0.548	0.803	0.656											
ECL	96	0.131	0.224	0.139	0.238	0.140	0.237	0.129	0.222	0.168	0.272	0.193	0.308	0.201	0.317	0.169	0.273	0.187	0.304	0.207	0.307	0.274	0.368	0.312	0.402											
	192	0.152	0.241	0.153	0.251	0.153	0.249	0.157	0.240	0.184	0.289	0.201	0.315	0.222	0.334	0.182	0.286	0.199	0.315	0.213	0.316	0.296	0.386	0.348	0.433											
	336	0.160	0.248	0.169	0.266	0.169	0.267	0.163	0.259	0.198	0.300	0.214	0.329	0.231	0.338	0.200	0.304	0.212	0.329	0.230	0.333	0.300	0.394	0.350	0.433	720	0.192	0.298	0.206	0.297	0.203	0.301	0.197	0.290	0.220	0.320
	0.246	0.355	0.254	0.361	0.222	0.321	0.233	0.345	0.265	0.360	0.373	0.439	0.340	0.420	平均	0.158	0.252	0.167	0.263	0.166	0.263	0.161	0.252	0.192	0.295	0.214	0.327	0.227	0.338	0.193	0.296	0.208	0.323	0.229	0.329	0.311
	0.397	0.338	0.422																																	
交通	96	0.362	0.248	0.388	0.282	0.410	0.282	0.360	0.249	0.593	0.321	0.587	0.366	0.613	0.388	0.612	0.338	0.607	0.392	0.615	0.391	0.719	0.391	0.732	0.423	192	0.374	0.247	0.407	0.290	0.423	0.287	0.379	0.256	0.617	0.336
	0.604	0.373	0.616	0.382	0.613	0.340	0.621	0.399	0.601	0.382	0.696	0.379	0.733	0.420	336	0.385	0.271	0.412	0.294	0.436	0.296	0.392	0.264	0.629	0.336	0.621	0.383	0.622	0.337	0.618	0.328	0.622	0.396	0.613	0.386	0.777
	0.420	0.742	0.420	720	0.430	0.288	0.450	0.312	0.466	0.315	0.432	0.286	0.640	0.350	0.626	0.382	0.660	0.408	0.653	0.355	0.632	0.396	0.658	0.407	0.864	0.472	0.755	0.423	平均	0.388	0.264	0.414	0.294	0.433	0.295	0.390
	0.263	0.620	0.336	0.610	0.376	0.628	0.379	0.624	0.340	0.621	0.396	0.622	0.392	0.764	0.416	0.741	0.422																			
病毒	24	1.285	0.727	2.063	0.881	2.215	1.081	1.319	0.754	2.317	0.934	3.228	1.260	3.483	1.287	2.294	0.945	2.527	1.020	8.313	2.144	5.764	1.677	4.400	1.382	36	1.404	0.814	1.868	0.892	1.963	0.963	1.430	0.834	1.972	0.920	2.679
	1.080	3.103	1.148	1.825	0.848	2.615	1.007	6.631	1.902	4.755	1.467	4.783	1.448	48	1.523	0.807	1.790	0.884	2.130	1.024	1.553	0.815	2.238	0.940	2.622	1.078	2.669	1.085	2.010	0.900	2.359	0.972	7.299	1.982	4.763	1.469
	4.832	1.465	60	1.531	0.854	1.979	0.957	2.368	1.096	1.470	0.788	2.027	0.928	2.857	1.157	2.770	1.125	2.178	0.963	2.487	1.016	7.283	1.985	5.264	1.564	4.882	1.483	平均	1.435	0.801	1.925	0.903	2.169	1.041	1.443	0.797
	2.139	0.931	2.847	1.144	3.006	1.161	2.077	0.914	2.497	1.004	7.382	2.003	5.137	1.544	4.724	1.445																				
1 stCount					0			1			17			0			0			0			0			0			0			0			0	

D.2短期预测

表 12 展示了我们完整的短期预测结果。TIME -LLM 在大多数情况下始终优于大多数基线模型。值得注意的是,我们大幅超越了 GPT4TS（例如,总体提升8.7% ,M4-Yearly 提升13.4% ,M4-Hourly、M4-Daily 和 M4-Weekly 平均提升21.5% ），也超越了 TimesNet（例如,总体提升10% ， M4-Yearly 提升14.1% ,M4-Hourly、M4-Daily 和 M4-Weekly 平均提升30.1% ）。与近期最先进的预测模型 N-HiTS 和 PatchTST 相比， TIME-LLM 在主干 LLM 无需任何参数更新的情况下表现出相当甚至更优的性能。

此外,我们在 M3-Quarterly 数据集上对 TIME-LLM 与表现最佳的模型进行了比较分析,结果如表 13 所示。除了 M3 竞赛中使用的默认 SMAPE 之外,我们还提供了额外的指标,即 MRAE 和 MAPE。在该数据集上， TIME-LLM 的性能与 TimesNet 和 PatchTST 相当,并且大幅超越 GPT4TS,SMAPE、MRAE 和 MAPE 分别降低了23%、 35%和26%以上。

少量样本和零样本预测

E.1小样本预测

我们在小样本预测任务中的完整结果详见表 14 和表 15。在 10% 小样本学习的范围内， TIME-LLM 在35 个案例中的32 个案例中取得了 SOTA 性能,涵盖了七个不同的时间序列基准。在 5% 小样本学习的场景下,我们方法的优势更加明显,在32 个案例中的21 个案例中取得了 SOTA 性能。我们将此归功于我们重新编程的 LLM 中成功的知识激活。

表 11:长期预测任务中与其他基线的额外比较。我们将预测
对于 ILI,视界 H ∈ {24, 36, 48, 60},对于其他视界,视界 H ∈ {96, 192, 336, 720}。值越低,表示
性能。红色:最佳,蓝色:第二好。 _____

方法		TIME-LLM		N-BEATS		N-HiTS		AutoARIMA		AutoTheta		AutoETS			
公制		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE		
厄尔	96	0.362	0.392	0.496	0.475	0.392	0.407	0.933	0.635	1.266	0.758	1.264	0.756		
	192	0.398	0.418	0.544	0.504	0.442	0.438	0.868	0.621	1.188	0.749	1.181	0.745		
	336	0.430	0.427	0.592	0.533	0.497	0.471	0.964	0.663	1.310	0.799	1.292	0.792		
	720	0.442	0.457	0.639	0.588	0.559	0.533	1.043	0.705	1.510	0.882	1.405	0.842		
	平均值	0.408	0.423	0.568	0.525	0.473	0.462	0.952	0.656	1.319	0.797	1.286	0.784		
厄尔	96	0.268	0.328	0.384	0.431	0.321	0.368	0.390	0.417	0.461	0.430	0.444	0.403		
	192	0.329	0.375	0.496	0.493	0.398	0.421	0.545	0.492	0.754	0.537	0.771	0.461		
	336	0.368	0.409	0.585	0.542	0.453	0.459	0.697	0.562	1.355	0.683	1.526	0.522		
	720	0.372	0.420	0.792	0.651	0.775	0.609	0.907	0.658	3.971	1.061	5.183	0.633		
	平均值	0.334	0.383	0.564	0.529	0.487	0.464	0.635	0.532	1.635	0.678	1.981	0.505		
埃特姆1	96	0.272	0.334	0.393	0.412	0.327	0.368	1.091	0.661	1.211	0.704	1.519	0.768		
	192	0.310	0.358	0.425	0.427	0.376	0.400	1.119	0.682	1.237	0.724	1.535	0.784		
	336	0.352	0.384	0.464	0.454	0.407	0.423	1.125	0.698	1.231	0.735	1.472	0.782		
	720	0.383	0.411	0.521	0.488	0.471	0.456	1.243	0.745	1.394	0.801	1.591	0.825		
	平均值	0.329	0.372	0.451	0.445	0.395	0.412	1.145	0.697	1.268	0.741	1.529	0.790		
埃特姆2	96	0.161	0.253	0.204	0.302	0.188	0.273	0.435	0.375	0.245	0.316	0.359	0.333		
	192	0.219	0.293	0.282	0.358	0.274	0.338	0.995	0.494	0.413	0.401	0.756	0.396		
	336	0.271	0.329	0.378	0.425	0.384	0.406	2.324	0.648	0.790	0.528	1.747	0.467		
	720	0.352	0.379	0.555	0.523	0.501	0.488	9.064	1.020	2.451	0.847	6.856	0.639		
	平均值	0.251	0.313	0.355	0.402	0.337	0.376	3.205	0.634	0.975	0.523	2.430	0.459		
天气	96	0.147	0.201	0.185	0.244	0.160	0.222	0.255	0.273	0.279	0.266	0.331	0.277		
	192	0.189	0.234	0.225	0.282	0.202	0.265	0.390	0.353	0.337	0.316	0.498	0.345		
	336	0.262	0.279	0.274	0.323	0.253	0.303	0.775	0.457	0.472	0.385	0.898	0.423		
	720	0.304	0.316	0.340	0.373	0.323	0.354	2.898	0.707	0.818	0.526	2.820	0.580		
	平均	0.225	0.257	0.256	0.306	0.235	0.286	1.080	0.448	0.477	0.373	1.137	0.406		
电	96	0.131	0.224	0.233	0.327	0.184	0.275	0.520	0.466	0.653	0.532	0.650	0.526		
	192	0.152	0.241	0.246	0.340	0.190	0.282	0.581	0.499	0.713	0.561	0.704	0.549		
	336	0.160	0.248	0.262	0.355	0.205	0.298	0.602	0.515	0.797	0.603	0.766	0.577		
	720	0.192	0.298	0.296	0.383	0.239	0.330	0.685	0.558	1.023	0.688	0.901	0.628		
	平均值	0.158	0.252	0.259	0.351	0.205	0.296	0.597	0.510	0.797	0.596	0.755	0.570		
交通	96	0.362	0.248	0.608	0.447	0.410	0.329	1.068	0.694	3.207	1.219	3.254	1.221		
	192	0.374	0.247	0.605	0.448	0.414	0.330	1.380	0.775	3.407	1.262	3.569	1.264		
	336	0.385	0.271	0.618	0.454	0.428	0.337	1.448	0.790	3.473	1.274	3.971	1.275		
	720	0.430	0.288	0.650	0.467	0.456	0.354	1.481	0.799	3.952	1.382	6.784	1.379		
	平均值	0.388	0.264	0.620	0.454	0.427	0.338	1.344	0.765	3.510	1.284	4.395	1.285		
疾病	24	1.285	0.727	6.809	1.870	2.675	1.080	4.909	1.329	5.991	1.510	4.869	1.315		
	36	1.404	0.814	6.850	1.890	3.081	1.194	5.079	1.440	5.922	1.539	4.917	1.422		
	48	1.523	0.807	6.788	1.876	2.973	1.176	4.276	1.339	4.637	1.329	3.966	1.301		
	60	1.531	0.854	6.908	1.893	3.259	1.232	3.855	1.276	4.378	1.345	3.540	1.229		
	平均	1.435	0.801	6.839	1.882	2.997	1.171	4.530	1.346	5.232	1.431	4.323	1.317		
1 stCount		40		0		1		0		0		0			

表 12:完整的短期时间序列预测结果。预测范围参见[6, 48],最后
在不同采样间隔下,所有数据集的三行加权平均。较低的值表示
表现更佳。红色:最佳,蓝色:次佳。 _____

方法		TIME-LLM		GPT4TS		TimesNet		PatchTST		N-HiTS		N-BEATS		ETSformer		LightTS		DLinear		FEDformer		Stationary		Autoformer		Informer		Reformer	
每年	SMAPE	13.419	MASE	15.11	15.378	13.477	13.422	13.487	18.009	14.247	16.965	3.109	14.021	13.717	13.974	14.727	16.169												
		3.005		3.565	3.554	3.019	3.056	3.036	4.487	4.283			3.036	3.078	3.134	3.418	3.800												
	海外资产负债表	0.789		0.911	0.918	0.792	0.795	0.795	1.115	0.827	1.058		0.811	0.807	0.822	0.881	0.973												
季度	SMAPE	10.110	MASE	10.597	10.465	1.253	10.38	10.185	10.564	13.376	11.364	12.145	1.328	11.1	10.958	11.338	11.360	13.313											
	OWA	0.889		1.227			1.233	1.18	1.252	1.906	1.520	1.106		1.35	1.325	1.365	1.401	1.775											
		1.178		0.938	0.923		0.921	0.893	0.936	1.302	1.000			0.996	0.981	1.012	1.027	1.252											
每月	平均每英里	12.980		13.258	13.513		12.959	13.059	13.089	14.588	14.014	13.514		14.403	13.917	13.958	14.062	20.128											
	MASE	0.963	OWA	1.003	1.039		0.97	1.013	0.996	1.368	1.053	1.037		1.147	1.097	1.103	1.141	2.614											
		0.903		0.931	0.957		0.905	0.929	0.922	1.149	0.981	0.956		1.038	0.998	1.002	1.024	1.927											
每周	平均孔隙率	4.795		6.124	6.913	4.952	4.711	6.599	7.267	15.880	6.709	11.434		7.148	6.302	5.485	24.460	32.491											
	质量	3.178		4.116	4.507	3.347	3.054	4.43	5.240	4.953				4.041	4.064	3.865	20.960	33.355											
	OWA	1.006		1.259	1.438	1.049	0.977	1.393	1.591	3.474	1.487			1.389	1.304	1.187	5.879	8.679											
每日	SMAPE	11.983	MASE	12.69	12.88	12.059	12.035	12.25	14.718	13.525	13.639	2.111	13.16	12.780	12.909	14.086	18.200												
		1.595		1.808	1.836	1.623	1.625	1.698	2.408	2.095			1.775	1.756	1.771	2.718	4.223												
	海外资产负债表	0.859		0.94	0.955	0.869	0.869	0.896	1.172	1.051	1.051		0.949	0.930	0.939	1.230	1.775												

E.2零样本预测

表 16 总结了零样本预测的完整结果。TIME -LLM 在零样本适应性方面显著超越了六个最具竞争力的时间序列模型。总体而言,我们观察到
所有基线模型的 MSE 和 MAE 平均降低了23.5%和12.4% 。在典型的跨域场景中（例如 ETTh2 → ETTh1）,我们的改进始终非常显著。
和 ETTm2 → ETTm1）,平均 MSE 和 MAE 分别超过20.8%和11.3% 。值得注意的是,
与 LLMTime 相比,TIME-LLM 表现出了更优异的性能提升 (Gruver et al., 2023),
它采用了类似规模的主干法学硕士（7B）,是利用法学硕士进行

表 13:M3（季度)的附加短期时间序列预测结果。预测期为
8. 数值越低,性能越好。红色:最佳,蓝色:次佳。

方法TIME-LLM GPT4TS TimesNet PatchTST N-HiTS N-BEATS DLinear FEDformer							
SMAPE 11.171	14.453	10.410	12.380	12.616	18.640	15.028	12.927
MRAE 3.282	5.035	3.310	2.401	4.271	4.612	2.793	3.653
地图0.151	0.203	0.140	0.154	0.168	0.247	0.196	0.174

零样本时间序列预测。我们将这一成功归功于我们的重新编程框架
更好地激活法学硕士的知识转移和推理能力,以资源高效的方式
执行时间序列任务时的方式。

F消融研究

完整的消融结果见表 17。我们还比较了
重编程和微调（使用 QLoRA Dettmers 等人（2023））方案。我们的结果表明
与 QLoRA 变体（A.5)相比,我们的方法的性能明显提高了平均19% 。

G效率与模型微调的比较

设置。我们比较了模型微调的效率（与 QLoRA Dettmers 等人（2023））
以及我们在本节中提出的使用两种不同骨干的模型重编程,即 Llama
1/4 容量（前 8 个 Transformer 层)和满容量。我们坚持长期
ETTh1 上的预测协议可以预测两个不同的步骤（在本例中为 96 和 336）
领先。对于评估指标,我们报告了可训练参数的总数（百万）,
GPU 内存（以兆字节为单位)和运行时间（每次迭代的秒数）。

结果。结果列于表 18。我们发现模型重编程显著地
与 QLoRA 的参数高效微调 (PEFT) 相比,远程

表 14:基于 10% 训练数据的完整小样本学习结果。我们使用与表 1 相同的协议。

方法TIME-LLM GPT4TS			线性		贴片TST		TimesNet FEDformer Autoformer 固定式 ETSformer LightTS						告密者		改革者	
公制 MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE MSE MAE																
和	96 0.448 0.460 0.458 0.456 0.492 0.495 0.516 0.485 0.861 0.628 0.512 0.499 0.613 0.552 0.918 0.639 1.112 0.806 1.298 0.838 1.179 0.792 1.184 0.790															
	192 0.484 0.483 0.570 0.516 0.565 0.538 0.598 0.524 0.797 0.593 0.624 0.555 0.722 0.598 0.915 0.629 1.155 0.823 1.322 0.854 1.199 0.806 1.295 0.850															
	336 0.589 0.540 0.608 0.535 0.721 0.622 0.657 0.550 0.941 0.648 0.691 0.574 0.750 0.619 0.939 0.644 1.179 0.832 1.347 0.870 1.202 0.811 1.294 0.854															
	720 0.700 0.604 0.725 0.591 0.986 0.743 0.762 0.610 0.877 0.641 0.728 0.614 0.721 0.616 0.887 0.645 1.273 0.874 1.534 0.947 1.217 0.825 1.223 0.838															
	平均值0.556 0.522 0.590 0.525 0.691 0.600 0.633 0.542 0.869 0.628 0.639 0.561 0.702 0.596 0.915 0.639 1.180 0.834 1.375 0.877 1.199 0.809 1.249 0.833															
和	96 0.375 0.326 0.331 0.374 0.357 0.411 0.353 0.389 0.378 0.409 0.382 0.416 0.413 0.451 0.389 0.411 0.678 0.619 2.022 1.006 3.837 1.508 3.788 1.533															
	192 0.374 0.373 0.402 0.411 0.569 0.519 0.403 0.414 0.490 0.467 0.478 0.474 0.474 0.477 0.473 0.455 0.785 0.666 2.329 1.104 3.856 1.513 3.552 1.483															
	336 0.406 0.429 0.406 0.433 0.671 0.572 0.426 0.441 0.537 0.494 0.504 0.501 0.547 0.543 0.507 0.480 0.839 0.694 2.453 1.122 3.952 1.526 3.395 1.526															
	720 0.427 0.449 0.449 0.464 0.824 0.648 0.477 0.480 0.510 0.491 0.499 0.509 0.516 0.523 0.477 0.472 1.273 0.874 3.816 1.407 3.842 1.503 3.205 1.401															
	平均值0.370 0.394 0.397 0.421 0.605 0.538 0.415 0.431 0.479 0.465 0.466 0.475 0.488 0.499 0.462 0.455 0.894 0.713 2.655 1.160 3.872 1.513 3.485 1.486															
和h1	96 0.346 0.388 0.390 0.404 0.352 0.392 0.410 0.419 0.583 0.501 0.578 0.518 0.774 0.614 0.761 0.568 0.911 0.688 0.921 0.682 1.162 0.785 1.442 0.847															
	192 0.373 0.416 0.429 0.423 0.382 0.412 0.437 0.434 0.630 0.528 0.617 0.546 0.754 0.592 0.781 0.574 0.955 0.703 0.957 0.701 1.172 0.793 1.444 0.862															
	336 0.413 0.426 0.469 0.439 0.419 0.434 0.476 0.454 0.725 0.568 0.998 0.775 0.869 0.677 0.803 0.587 0.991 0.719 0.998 0.716 1.227 0.908 1.450 0.866															
	720 0.485 0.476 0.569 0.498 0.490 0.477 0.681 0.556 0.769 0.549 0.693 0.579 0.810 0.630 0.844 0.581 1.062 0.747 1.007 0.719 1.207 0.797 1.366 0.850															
	平均值0.404 0.427 0.464 0.441 0.411 0.429 0.501 0.466 0.677 0.537 0.722 0.605 0.802 0.628 0.797 0.578 0.980 0.714 0.971 0.705 1.192 0.821 1.426 0.856															
和h2	96 0.177 0.261 0.188 0.269 0.213 0.303 0.191 0.274 0.212 0.285 0.291 0.399 0.352 0.454 0.229 0.308 0.331 0.430 0.813 0.688 3.203 1.407 4.195 1.628															
	192 0.241 0.314 0.251 0.309 0.278 0.345 0.252 0.317 0.270 0.323 0.307 0.379 0.694 0.691 0.291 0.343 0.400 0.464 1.008 0.768 3.112 1.387 4.042 1.601															
	336 0.274 0.327 0.307 0.346 0.338 0.385 0.306 0.353 0.323 0.353 0.543 0.559 2.408 1.407 0.348 0.376 0.469 0.498 1.031 0.775 3.255 1.421 3.963 1.585															
	720 0.417 0.390 0.426 0.417 0.436 0.440 0.433 0.427 0.474 0.449 0.712 0.614 1.913 1.166 0.461 0.438 0.589 0.557 1.096 0.791 3.909 1.543 3.711 1.532															
	平均值0.277 0.323 0.293 0.335 0.316 0.368 0.296 0.343 0.320 0.353 0.463 0.488 1.342 0.930 0.332 0.366 0.447 0.487 0.987 0.756 3.370 1.440 3.978 1.587															
气	96 0.161 0.210 0.163 0.215 0.171 0.224 0.165 0.215 0.184 0.230 0.188 0.253 0.221 0.297 0.192 0.234 0.199 0.272 0.217 0.269 0.374 0.401 0.335 0.380															
	192 0.204 0.248 0.210 0.254 0.215 0.263 0.210 0.257 0.245 0.283 0.250 0.304 0.270 0.322 0.269 0.295 0.279 0.332 0.259 0.304 0.552 0.478 0.522 0.462															
	336 0.261 0.302 0.256 0.292 0.258 0.299 0.259 0.297 0.305 0.321 0.312 0.346 0.320 0.351 0.370 0.357 0.356 0.386 0.303 0.334 724 0.541 0.715 0.535															
	720 0.309 0.332 0.321 0.339 0.320 0.346 0.332 0.346 0.381 0.371 0.387 0.393 0.390 0.396 0.441 0.405 0.437 0.448 0.377 0.382 0.739 0.558 0.611 0.500															
	平均值0.234 0.273 0.238 0.275 0.241 0.283 0.242 0.279 0.279 0.301 0.284 0.324 0.300 0.342 0.318 0.323 0.318 0.360 0.289 0.322 0.597 0.495 0.546 0.469															
电	96 0.139 0.241 0.139 0.237 0.150 0.253 0.140 0.238 0.299 0.373 0.231 0.323 0.261 0.348 0.420 0.466 0.599 0.587 0.350 0.425 1.259 0.919 0.993 0.784															
	192 0.151 0.248 0.156 0.252 0.164 0.264 0.160 0.255 0.305 0.379 0.261 0.356 0.338 0.406 0.411 0.459 0.620 0.598 0.376 0.448 1.160 0.873 0.938 0.753															
	336 0.169 0.270 0.175 0.270 0.181 0.282 0.180 0.276 0.319 0.391 0.360 0.445 0.410 0.474 0.434 0.473 0.662 0.619 0.428 0.485 1.157 0.872 0.925 0.745															
	720 0.240 0.322 0.233 0.317 0.223 0.321 0.241 0.323 0.369 0.426 0.530 0.585 0.715 0.685 0.510 0.521 0.757 0.664 0.611 0.597 1.203 0.898 1.004 0.790															
	平均值0.175 0.270 0.176 0.269 0.180 0.280 0.180 0.273 0.323 0.392 0.346 0.427 0.431 0.478 0.444 0.480 0.660 0.617 0.441 0.489 1.195 0.891 0.965 0.768															
澳	96 0.418 0.291 0.414 0.297 0.419 0.298 0.403 0.289 0.719 0.416 0.639 0.400 0.672 0.405 1.412 0.802 1.643 0.855 1.157 0.636 1.557 0.821 1.527 0.815															
	192 0.414 0.296 0.426 0.301 0.434 0.305 0.415 0.296 0.748 0.428 0.637 0.416 0.727 0.424 1.419 0.806 1.641 0.854 1.207 0.661 1.454 0.765 1.538 0.817															
	336 0.421 0.311 0.434 0.303 0.449 0.313 0.426 0.304 0.853 0.471 0.655 0.427 0.749 0.454 1.443 0.815 1.711 0.878 1.334 0.713 1.521 0.812 1.550 0.819															
	720 0.462 0.327 0.487 0.337 0.484 0.336 0.474 0.331 1.485 0.825 0.722 0.456 0.847 0.499 1.539 0.837 2.660 1.157 1.292 0.726 1.605 0.846 1.588 0.833															
	平均值0.429 0.306 0.440 0.310 0.447 0.313 0.430 0.305 0.951 0.535 0.663 0.425 0.749 0.446 1.453 0.815 1.914 0.936 1.248 0.684 1.534 0.811 1.551 0.821															
1 stCount	2	9	3	3	0	0	0	0	0	0	0	0	0	0	0	0

表 15:基于 5% 训练数据的完整小样本学习结果。我们使用与表 1 相同的协议。“-”表示 5%的时间序列不足以构成训练集。

方法		TIME-LLM		GPT4TS		线性		贴片TST		TimesNet		FEDformer		Autoformer		固定		ETSformer		轻量级		告密者		改革者	
公制		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
和1	96	0.483	0.464	0.543	0.506	0.547	0.503	0.557	0.519	0.892	0.625	0.593	0.529	0.681	0.570	0.952	0.650	1.169	0.832	1.483	0.91	1.225	0.812	1.198	0.795
	192	0.629	0.540	0.748	0.580	0.720	0.604	0.711	0.570	0.940	0.665	0.652	0.563	0.725	0.602	0.943	0.645	1.221	0.853	1.525	0.93	1.249	0.828	1.273	0.853
	336	0.768	0.626	0.754	0.595	0.984	0.727	0.816	0.619	0.945	0.653	0.731	0.594	0.761	0.624	0.935	0.644	1.179	0.832	1.347	0.87	1.202	0.811	1.254	0.857
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	平均值	0.627	0.543	0.681	0.560	0.750	0.611	0.694	0.569	0.925	0.647	0.658	0.562	0.722	0.598	0.943	0.646	1.189	0.839	1.451	0.903	1.225	0.817	1.241	0.835
[illegible]	96	0.336	0.397	0.376	0.421	0.442	0.456	0.401	0.421	0.409	0.420	0.390	0.424	0.428	0.468	0.408	0.423	0.678	0.619	2.022	1.006	3.837	1.508	3.753	1.518
	192	0.406	0.425	0.418	0.441	0.617	0.542	0.452	0.455	0.483	0.464	0.457	0.465	0.496	0.504	0.497	0.468	0.845	0.697	3.534	1.348	3.975	1.933	3.516	1.473
	336	0.405	0.432	0.408	0.439	1.424	0.849	0.464	0.469	0.499	0.479	0.477	0.483	0.486	0.496	0.507	0.481	0.905	0.727	4.063	1.451	3.956	1.520	3.312	1.427
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	平均值	0.382	0.418	0.400	0.433	0.694	0.577	0.827	0.615	0.439	0.448	0.463	0.454	0.441	0.457	0.470	0.489	0.809	0.681	3.206	1.268	3.922	1.653	3.527	1.472
[illegible]	96	0.316	0.377	0.386	0.405	0.332	0.374	0.399	0.414	0.606	0.518	0.628	0.544	0.726	0.578	0.823	0.587	1.031	0.747	1.048	0.733	1.130	0.775	1.234	0.798
	192	0.450	0.464	0.440	0.438	0.358	0.390	0.441	0.436	0.681	0.539	0.666	0.566	0.750	0.591	0.844	0.591	1.087	0.766	1.097	0.756	1.150	0.788	1.287	0.839
	336	0.450	0.424	0.485	0.459	0.402	0.416	0.499	0.467	0.786	0.597	0.807	0.628	0.851	0.659	0.870	0.603	1.138	0.787	1.147	0.775	1.198	0.809	1.288	0.842
	720	0.483	0.471	0.577	0.499	0.511	0.489	0.767	0.587	0.796	0.593	0.822	0.633	0.857	0.655	0.893	0.611	1.245	0.831	1.200	0.799	1.175	0.794	1.247	0.828
	平均值	0.425	0.434	0.472	0.450	0.400	0.417	0.526	0.476	0.717	0.561	0.730	0.592	0.796	0.620	0.857	0.598	1.125	0.782	1.123	0.765	1.163	0.791	1.264	0.826
和m2	96	0.174	0.261	0.199	0.280	0.236	0.326	0.206	0.288	0.220	0.299	0.229	0.320	0.232	0.322	0.238	0.316	0.404	0.485	1.108	0.772	3.599	1.478	3.883	1.545
	192	0.215	0.287	0.256	0.316	0.306	0.373	0.264	0.324	0.311	0.361	0.394	0.361	0.291	0.357	0.298	0.349	0.479	0.521	1.317	0.850	3.578	1.475	3.553	1.484
	336	0.273	0.330	0.318	0.353	0.380	0.423	0.334	0.367	0.338	0.366	0.378	0.427	0.478	0.517	0.353	0.380	0.552	0.555	1.415	0.879	3.561	1.473	3.446	1.460
	720	0.433	0.412	0.460	0.436	0.674	0.583	0.454	0.432	0.509	0.465	0.523	0.510	0.553	0.538	0.475	0.445	0.701	0.627	1.822	0.984	3.896	1.533	3.445	1.460
	平均值	0.274	0.323	0.308	0.346	0.399	0.426	0.314	0.352	0.344	0.372	0.381	0.404	0.388	0.433	0.341	0.372	0.534	0.547	1.415	0.871	3.658	1.489	3.581	1.487
天气	96	0.172	0.263	0.175	0.230	0.184	0.242	0.171	0.224	0.207	0.253	0.229	0.309	0.227	0.299	0.215	0.252	0.218	0.295	0.230	0.285	0.497	0.497	0.406	0.435
	192	0.224	0.271	0.227	0.276	0.228	0.283	0.230	0.277	0.272	0.307	0.265	0.317	0.278	0.333	0.290	0.307	0.294	0.331	0.274	0.323	0.620	0.545	0.446	0.450
	336	0.282	0.321	0.286	0.322	0.279	0.322	0.294	0.326	0.313	0.328	0.353	0.392	0.351	0.393	0.353	0.348	0.359	0.398	0.318	0.355	0.649	0.547	0.465	0.459
	720	0.366	0.381	0.366	0.379	0.364	0.388	0.384	0.387	0.400	0.385	0.391	0.394	0.387	0.389	0.452	0.407	0.461	0.461	0.401	0.418	0.570	0.522	0.471	0.468
	平均值	0.260	0.309	0.263	0.301	0.263	0.308	0.269	0.303	0.298	0.318	0.309	0.353	0.310	0.353	0.327	0.328	0.333	0.371	0.305	0.345	0.584	0.527	0.447	0.453
电	96	0.147	0.242	0.143	0.241	0.150	0.251	0.145	0.244	0.315	0.389	0.235	0.322	0.297	0.367	0.484	0.518	0.697	0.638	0.639	0.609	1.265	0.919	1.414	0.855
	192	0.158	0.241	0.159	0.255	0.163	0.263	0.163	0.260	0.318	0.396	0.247	0.341	0.308	0.375	0.501	0.531	0.718	0.648	0.772	0.678	1.298	0.939	1.240	0.919
	336	0.178	0.277	0.179	0.274	0.175	0.278	0.183	0.281	0.340	0.415	0.267	0.356	0.354	0.411	0.574	0.578	0.758	0.667	0.901	0.745	1.302	0.942	1.253	0.921
	720	0.224	0.312	0.233	0.323	0.219	0.311	0.233	0.323	0.635	0.613	0.318	0.394	0.426	0.466	0.952	0.786	1.028	0.788	1.200	0.871	1.259	0.919	1.249	0.921
	平均值	0.179	0.268	0.178	0.273	0.176	0.275	0.181	0.277	0.402	0.453	0.266	0.353	0.346	0.404	0.627	0.603	0.800	0.685	0.878	0.725	1.281	0.929	1.289	0.904
交通	96	0.414	0.291	0.419	0.298	0.427	0.304	0.404	0.286	0.854	0.492	0.670	0.421	0.795	0.481	1.468	0.821	1.643	0.855	1.157	0.636	1.557	0.821	1.586	0.841
	192	0.419	0.291	0.434	0.305	0.447	0.315	0.412	0.294	0.894	0.517	0.653	0.405	0.837	0.503	1.509	0.838	1.856	0.928	1.688	0.848	1.596	0.834	1.602	0.844
	336	0.437	0.314	0.449	0.313	0.478	0.333	0.439	0.310	0.853	0.471	0.707	0.445	0.867	0.523	1.602	0.860	2.080	0.999	1.826	0.903	1.621	0.841	1.668	0.868
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	平均值	0.423	0.298	0.434	0.305	0.450	0.317	0.418	0.296	0.867	0.493	0.676	0.423	0.833	0.502	1.526	0.839	1.859	0.927	1.557	0.795	1.591	0.832	1.618	0.851
1 stCount		21		6		7		6		0		1		0		0		0		0		0		0	

预测任务的可训练参数总数、GPU 内存开销和
训练速度。量化来看,可训练参数在四年内平均减少了71.2%。
场景,内存消耗减少23.1% ,训练速度提高25.3% 。

误差线

所有实验都进行了三次,我们给出了我们的标准差
模型和第二名模型。我们的方法与第二名的
方法 PatchTST (Nie 等,2023)在长期预测任务上的表现如表 19 所示。
下表报告了四个 ETT 数据集的平均 MSE 和 MAE,完整
标准差。此外,表 20 对比了我们的方法与
第二佳方法 N-HiTS (Challu et al., 2023a) 采用不同的 M4 数据集进行
比较。

可视化

在本部分中,我们将 TIME-LLM 的预测结果与最先进的代表性方法(例如 GPT4TS (Zhou et al., 2023a)、PatchTST (Nie et al., 2023)、
和 Autoformer (Wu et al., 2021)) 在各种场景下展示了
TIME-LLM。

图 7 和图 8 将各种方法的长期(输入 96,预测 96)和短期(输入 36,预测 36)预测与真实值进行了比较。图中,TIME-LLM 展示了
预测准确率明显优于 GPT4TS、PatchTST 和经典
基于变压器的方法,自动变压器。

我们还提供了少量和零次预测场景下预测结果的视觉比较,
如图 9 和图 10 所示。我们坚持长期(输入 96-预测 96)预测设置

表 16:ETT 数据集上的完全零样本学习结果。数值越低,表示性能越好。红色:
最好,蓝色:第二好。

方法		TIME-LLM		LLMTime		GPT4TS			线性		PatchTST		TimesNet		自动成型机	
公制		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE				
并且 T h1 → 并且 T h2	96	0.279	0.337	0.510	0.576	0.335	0.374	0.347	0.400	0.304	0.350	0.358	0.387	0.469	0.486		
	192	0.351	0.374	0.523	0.586	0.412	0.417	0.447	0.460	0.386	0.400	0.427	0.429	0.634	0.567		
	336	0.388	0.415	0.640	0.637	0.441	0.444	0.515	0.505	0.414	0.428	0.449	0.451	0.655	0.588		
	720	0.391	0.420	2.296	1.034	0.438	0.452	0.665	0.589	0.419	0.443	0.448	0.458	0.570	0.549		
	平均值	0.353	0.387	0.992	0.708	0.406	0.422	0.493	0.488	0.380	0.405	0.421	0.431	0.582	0.548		
并且 T h1 → 并且 Tm2	96	0.189	0.293	0.646	0.563	0.236	0.315	0.255	0.357	0.215	0.304	0.239	0.313	0.352	0.432		
	192	0.237	0.312	0.934	0.654	0.287	0.342	0.338	0.413	0.275	0.339	0.291	0.342	0.413	0.460		
	336	0.291	0.365	1.157	0.728	0.341	0.374	0.425	0.465	0.334	0.373	0.342	0.371	0.465	0.489		
	720	0.372	0.390	4.730	1.531	0.435	0.422	0.640	0.573	0.431	0.424	0.434	0.419	0.599	0.551		
	平均	0.273	0.340	1.867	0.869	0.325	0.363	0.415	0.452	0.314	0.360	0.327	0.361	0.457	0.483		
并且 T h2 → 并且 T h1	96	0.450	0.452	1.130	0.777	0.732	0.577	0.689	0.555	0.485	0.465	0.848	0.601	0.693	0.569		
	192	0.465	0.461	1.242	0.820	0.758	0.559	0.707	0.568	0.565	0.509	0.860	0.610	0.760	0.601		
	336	0.501	0.482	1.328	0.864	0.759	0.578	0.710	0.577	0.581	0.515	0.867	0.626	0.781	0.619		
	720	0.501	0.502	4.145	1.461	0.781	0.597	0.704	0.596	0.628	0.561	0.887	0.648	0.796	0.644		
	平均值	0.479	0.474	1.961	0.981	0.757	0.578	0.703	0.574	0.565	0.513	0.865	0.621	0.757	0.608		
并且 T h2 → 并且 Tm2	96	0.174	0.276	0.646	0.563	0.253	0.329	0.240	0.336	0.226	0.309	0.248	0.324	0.263	0.352		
	192	0.233	0.315	0.934	0.654	0.293	0.346	0.295	0.369	0.289	0.345	0.296	0.352	0.326	0.389		
	336	0.291	0.337	1.157	0.728	0.347	0.376	0.345	0.397	0.348	0.379	0.353	0.383	0.387	0.426		
	720	0.392	0.417	4.730	1.531	0.446	0.429	0.432	0.442	0.439	0.427	0.471	0.446	0.487	0.478		
	平均值	0.272	0.341	1.867	0.869	0.335	0.370	0.328	0.386	0.325	0.365	0.342	0.376	0.366	0.411		
并且 Tm1 → 并且 T h2	96	0.321	0.369	0.510	0.576	0.353	0.392	0.365	0.415	0.354	0.385	0.377	0.407	0.435	0.470		
	192	0.389	0.410	0.523	0.586	0.443	0.437	0.454	0.462	0.447	0.434	0.471	0.453	0.495	0.489		
	336	0.408	0.433	0.640	0.637	0.469	0.461	0.496	0.494	0.481	0.463	0.472	0.484	0.470	0.472		
	720	0.406	0.436	2.296	1.034	0.466	0.468	0.541	0.529	0.474	0.471	0.495	0.482	0.480	0.485		
	平均值	0.381	0.412	0.992	0.708	0.433	0.439	0.464	0.475	0.439	0.438	0.457	0.454	0.470	0.479		
并且 Tm1 → 并且 Tm2	96	0.169	0.257	0.646	0.563	0.217	0.294	0.221	0.314	0.195	0.271	0.222	0.295	0.385	0.457		
	192	0.227	0.318	0.934	0.654	0.277	0.327	0.286	0.359	0.258	0.311	0.288	0.337	0.433	0.469		
	336	0.290	0.338	1.157	0.728	0.331	0.360	0.357	0.406	0.317	0.348	0.341	0.367	0.476	0.477		
	720	0.375	0.367	4.730	1.531	0.429	0.413	0.476	0.476	0.416	0.404	0.436	0.418	0.582	0.535		
	平均值	0.268	0.320	1.867	0.869	0.313	0.348	0.335	0.389	0.296	0.334	0.322	0.354	0.469	0.484		
并且 Tm2 → 并且 T h2	96	0.298	0.356	0.510	0.576	0.360	0.401	0.333	0.391	0.327	0.367	0.360	0.401	0.353	0.393		
	192	0.359	0.397	0.523	0.586	0.434	0.437	0.441	0.456	0.411	0.418	0.434	0.437	0.432	0.437		
	336	0.367	0.412	0.640	0.637	0.460	0.459	0.505	0.503	0.439	0.447	0.460	0.459	0.452	0.459		
	720	0.393	0.434	2.296	1.034	0.485	0.477	0.543	0.534	0.459	0.470	0.485	0.477	0.453	0.467		
	平均值	0.354	0.400	0.992	0.708	0.435	0.443	0.455	0.471	0.409	0.425	0.435	0.443	0.423	0.439		
并且 Tm2 → 并且 Tm1	96	0.359	0.397	1.179	0.781	0.747	0.558	0.570	0.490	0.491	0.437	0.747	0.558	0.735	0.576		
	192	0.390	0.420	1.327	0.846	0.781	0.560	0.590	0.506	0.530	0.470	0.781	0.560	0.753	0.586		
	336	0.421	0.445	1.478	0.902	0.778	0.578	0.706	0.567	0.565	0.497	0.778	0.578	0.750	0.593		
	720	0.487	0.488	3.749	1.408	0.769	0.573	0.731	0.584	0.686	0.565	0.769	0.573	0.782	0.609		
	平均值	0.414	0.438	1.933	0.984	0.769	0.567	0.649	0.537	0.568	0.492	0.769	0.567	0.755	0.591		

表 17:完全消融 ETTh1 和 ETTm1,预测未来 96 步和 192 步（报告的 MSE）。

变体	长期预测				小样本预测			
	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192
A.1羊驼（默认； 32）	0.362	0.398	0.272	0.310	0.448	0.484	0.346	0.373
A.2火焰 (8)	0.389	0.412	0.297	0.329	0.567	0.632	0.451	0.490
A.3 GPT-2（12）	0.385	0.419	0.306	0.332	0.548	0.617	0.447	0.509
A.4 GPT-2（6）	0.394	0.427	0.311	0.342	0.571	0.640	0.468	0.512
A.5火焰（QLoRA;32）	0.391	0.420	0.310	0.338	0.543	0.611	0.578	0.618
B.1无补丁重新编程B.2无提示符作为前缀	0.410	0.412	0.310	0.342	0.498	0.570	0.445	0.487
	0.398	0.423	0.298	0.339	0.521	0.617	0.432	0.481
C.1无数据集上下文C.2无任务说明C.3无统计上下文	0.402	0.417	0.298	0.331	0.491	0.538	0.392	0.447
	0.388	0.420	0.285	0.327	0.476	0.529	0.387	0.439
	0.391	0.419	0.279	0.347	0.483	0.547	0.421	0.461

在两种情况下。TIME -LLM 在有限数据预测方面表现出显著优势　事实上
与 GPT4TS 相比,这一点尤为突出。

表18:模型重编程与参数高效微调（PEFT)的效率比较
使用 QLoRA（Dettmers 等人,2023)在 ETTh1 数据集上预测未来的两个不同步骤。

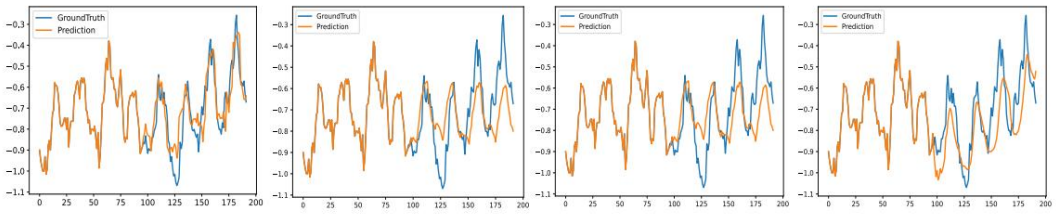
长度		ETTh1-96			ETTh1-336		
公制		可训练参数 (M)	内存 (MiB)	速度 (s/iter)	可训练参数 (M)	内存 (MiB)	速度 (s/iter)
火焰 (8)	QLoRA	12.60	14767	0.237	12.69	15982	0.335
	重新编程	5.62	11370	0.184	5.71	13188	0.203
火焰 (32)	QLoRA	50.29	45226	0.697	50.37	49374	0.732
	重新编程	6.39	32136	0.517	6.48	37988	0.632

表 19:我们的方法和第二佳方法（PatchTST)在所有时间序列上的标准差
用于长期预测的数据集。

模型	时间法学硕士		PatchTST (2023)	
数据集	均方误差	有	均方误差	有
ETTh1	0.408±0.011	0.423±0.012	0.413±0.001	0.430±0.002
ETTh2	0.334±0.005	0.383±0.009	0.330±0.002	0.379±0.007
ETTm1	0.329±0.006	0.372±0.007	0.351±0.006	0.380±0.002
ETTm2	0.251±0.002	0.313±0.003	0.255±0.003	0.315±0.002
天气	0.225 ± 0.009	0.257 ± 0.008	0.225 ± 0.001	0.264 ± 0.001
电力	0.158±0.004	0.252±0.007	0.161±0.001	0.252±0.001
交通量	0.388±0.001	0.264±0.006	0.390±0.003	0.263±0.003
或	1.435±0.011	0.801±0.008	1.443±0.012	0.797±0.002

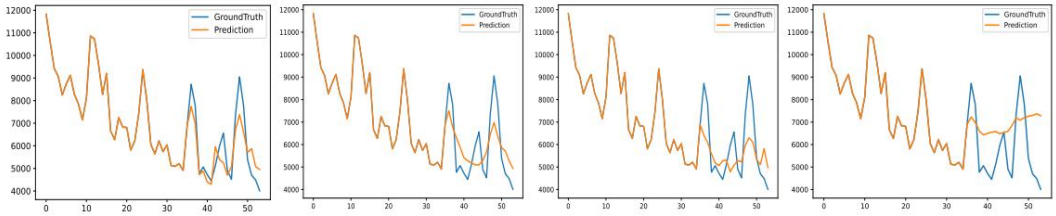
表 20:我们的TIME-LLM 和第二佳方法（N-HiTS)在 M4 数据集上的标准差
短期预测。

模型	时间法学硕士			N-HiTS (2023a)		
数据集	SMAPE	地图	海外工作办公室	斯玛佩	地图	海外工作办公室
每年	13.419 ± 0.117	3.005 ± 0.011	0.789 ± 0.003	13.422 ± 0.009	3.056 ± 0.017	0.795 ± 0.010
季度	10.110 ± 0.107	1.178 ± 0.009	0.889 ± 0.007	10.185 ± 0.107	1.180 ± 0.007	0.893 ± 0.001
每月	12.980 ± 0.102	0.963 ± 0.005	0.903 ± 0.001	13.059 ± 0.101	1.013 ± 0.007	0.929 ± 0.005
其他	4.795±0.117	3.178±0.012	1.006±0.009	4.711±0.117	3.054±0.011	0.997±0.012
平均	11.983 ± 0.011	1.595 ± 0.021	0.859 ± 0.002	12.035 ± 0.111	1.625 ± 0.012	0.869 ± 0.005



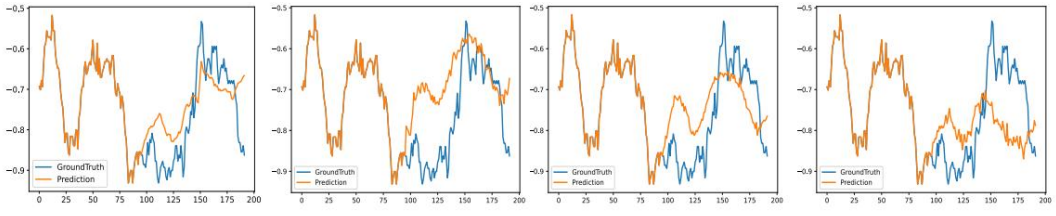
(a)时间-LLM (b)GPT4TS (c)PatchTST (d)自耦变压器

图 7:在输入 96 个样本、预测 96 个样本的设置下,不同模型对 ETTh1 的长期预测案例。蓝线表示真实值,橙线表示模型预测值。



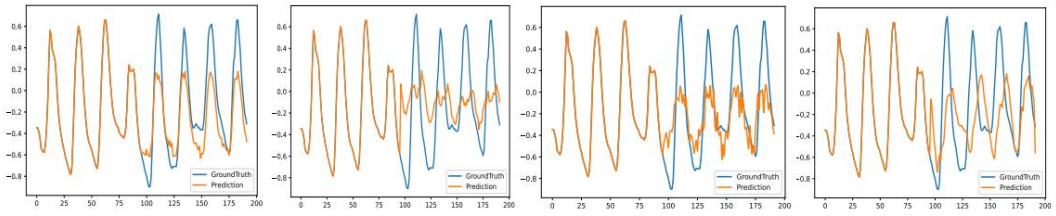
(a)时间-LLM (b)GPT4TS (c)时代网 (d)自耦变压器

图 8:在输入 36 预测 18 设置下,不同模型对 M4 数据集进行的短期预测。



(a)时间-LLM (b)GPT4TS (c)PatchTST (d)自耦变压器

图 9:在输入 96 个样本、预测 96 个样本的设置下,不同模型对 ETTm1 进行的小样本预测。蓝线表示真实值,橙线表示模型预测值。



(a)时间-LLM (b)GPT4TS (c)PatchTST (d)自耦变压器

图 10:在输入 96 个样本、预测 96 个样本的设置下,不同模型对 ETTh1→ETTh2 的零样本预测案例。蓝线表示真实值,橙线表示模型预测值。