

# TEMPO:基于提示的生成预训练 用于时间序列预测的 Transformer

曹德福<sup>1</sup>, 芙蓉佳<sup>1</sup>, 塞尔坎·奥·阿里克<sup>2</sup>, 托马斯·菲斯特<sup>2</sup>, 郑以翔<sup>1</sup>, 文也<sup>1</sup>, 刘岩<sup>1</sup>  
<sup>1</sup> 南加州大学  
<sup>2</sup> 谷歌云人工智能研究  
 {defucao, florajia, yixiangzheng, yewen, yanliu.cs}@usc.edu {soarik, tpfister}@google.com,

## 抽象的

过去十年见证了深度学习时间序列建模的重大进步。虽然实现了最先进的结果,但性能最佳的架构在不同的应用程序和领域之间存在很大差异。与此同时,对于自然语言处理,生成式预训练变压器 (GPT)通过跨各种文本数据集训练一个通用模型,展示了令人印象深刻的性能。探索 GPT 类型的架构是否能够有效地处理时间序列、捕获内在的动态属性并显着提高准确性是很有趣的。在本文中,我们提出了一种新颖的框架 TEMPO,它可以有效地学习时间序列表示。

我们专注于利用预训练模型的时间序列任务的两个基本归纳偏差: (i)趋势、季节性和残差成分之间复杂相互作用的分解; (ii) 引入基于选择的提示,以促进非平稳时间序列中的分布适应。TEMPO 扩展了根据不同领域内的数据对现实世界时间现象进行动态建模的能力。我们的实验证明了 TEMPO 在许多时间序列基准数据集上的性能优于最先进的方法。这种性能增益不仅在标准监督学习设置中观察到,而且在涉及以前未见过的数据集的场景以及具有多模式输入的场景中也观察到。这一引人注目的发现凸显了 TEMPO构成基础模型构建框架的潜力。

## 1简介

时间序列预测,即根据历史观察预测未来数据,在现实世界中具有广泛的应用,例如健康、交通、金融等。在过去的十年中,许多深度神经网络架构已应用于时间序列建模,包括卷积神经网络 (CNN) (Bai et al., 2018)、循环神经网络 (RNN) (Siarni-Namini et al., 2018)、图神经网络 (GNN) (Li et al., 2018; Cao et al., 2020)和 Transformers (Liu et al., 2021; Wu et al., 2021; Zhou et al., 2021; Wu et al., 2023; Zhou et al., 2022; Woo et al., 2022; Kitaev et al., 2020; Nie et al., 2023),得出了最先进的结果。在实现强大的预测性能的同时,之前关于时间序列的工作主要受益于序列建模 (从 RNN 和 GNN 到 Transformer)的进步,它捕获了时间依赖性,但忽略了时间序列数据中的一系列复杂模式,例如季节性、趋势,和残差。

但这些成分是时间序列与经典序列数据的关键区别因素 (Fildes 等, 1991)。因此,最近的研究表明,基于深度学习的架构可能不像之前想象的那么强大,甚至可能在某些基准上被浅层神经网络甚至线性模型超越 (Zeng 等人, 2023; Zhang 等人, 2022b;吴等人, 2023;聂等人, 2023)。

与此同时,自然语言处理 (NLP) 和计算机视觉 (CV) 领域基础模型的兴起,例如 LLaMA (Touvron 等人, 2023)、CLIP (Radford 等人, 2021)和 ChatGPT,标志着有效的里程碑表征学习。探索具有大量数据的基础时间序列模型的预训练路径,促进下游任务的性能改进是非常有趣的。最近的一些工作揭示了建造通用变压器的可能性

预印本。

时间序列 (Zhou et al., 2023; Sun et al., 2023; Xue & Salim, 2022)。此外, LLM 中的提示技术(例如 InstructGPT (Ouyang et al., 2022))提供了一种在预训练期间利用模型现有表示的方法,而不需要从头开始学习。然而,语言模型中现有的主干结构和提示技术并不能完全捕捉时间模式的演变和相互关联的动态随时间的进展,而这些是时间序列建模的基础。

在本文中,我们试图解决这些及时的挑战,并开发一种基于提示的时间序列生成预训练变换器,即 TEMPO (Time sEries proMpt POol)。受建立人脑长期记忆的记忆巩固理论 (Squire et al., 2015)的启发,TEMPO 包含两个用于有效时间序列表示学习的关键分析组件:一个专注于对特定时间序列模式进行建模,例如趋势和时间序列。季节性,另一个专注于从过去的的数据序列中获得更普遍和可转移的见解。具体来说,TEMPO首先通过局部加权散点图平滑将时间序列输入分解为三个附加成分,即趋势、季节性和残差 (Cleveland et al., 1990)。这些时间输入中的每一个随后都被映射到其相应的隐藏空间,以构建生成预训练变压器 (GPT)的时间序列输入嵌入。我们进行了理论分析,将时间序列域与频域联系起来,以强调分解这些组件进行时间序列分析的必要性。此外,我们从理论上揭示了注意力机制很难自动实现分解。其次,TEMPO 利用提示池来有效地调整 GPT (Radford 等人, 2019),通过指导重复使用一组可学习的连续向量表示,这些向量表示对趋势和季节性的时间知识进行编码。该过程通过将相似的时间序列实例映射到相似的提示上,允许在不断变化的时间分布上进行自适应知识整合,从而随着生成过程的发展保持预测能力。

此外,我们利用时间序列数据的三个关键加性成分 趋势、季节性和残差 构建广义加性模型 (GAM) (Hastie, 2017)。这使我们能够提供一个可解释的框架来理解输入组件之间的交互,这对于 Autoformer (Wu et al., 2021)、Fedformer (Zhou et al., 2022)和LaST (Wang et al., 2022a)来说实现起来具有挑战性。)由于训练阶段固有分解过程的设计。在七个基准数据集上进行的实验表明,与预测长度为 96 和 192 的时间序列预测的最先进模型相比,TEMPO 在 MAE 上实现了超过62.87%和35.59%的改进。重要的是,跨域预训练的强有力的实验结果TEMPO (所有预测长度的 MAE 平均提高30.8%)为时间序列的基础模型铺平了道路。此外,还引入了具有多模式时间序列观察的金融应用中的新数据集,即 TETS (TExt for Time Series),并将与社区共享,以促进时间序列分析预训练模型的进一步研究课题。当考虑多模态输入时, TEMPO为我们提出的 TETS 数据集带来了超过32.4%的SMAPE 改进。

总之,我们论文的主要贡献包括:(1)我们引入了一种可解释的基于提示调整的生成变压器 TEMPO,用于时间序列表示学习。它进一步推动了时间序列预测的范式转变 从传统的深度学习方法到预先训练的基础模型。(2) 我们通过关注两个基本归纳偏差来调整时间序列的预训练模型:首先,我们利用分解的趋势、季节性和残差信息。其次,我们采用即时选择策略来适应非平稳时间序列数据的动态特性。(3) 通过对七个基准数据集和一个建议数据集的广泛实验,我们的模型表现出了卓越的性能。值得注意的是,我们对跨域预训练的稳健结果显示,所有预测长度的平均 MAE 提高了 30.8%,凸显了基础模型在时间序列预测领域的潜力。

## 2相关作品

预训练的时间序列大型语言模型。大型语言模型 (LLM)的最新发展为时间序列建模开辟了新的可能性。法学硕士,如 T5 (Raffel et al., 2020)、GPT (Radford et al., 2018)、GPT-2 (Radford et al., 2019)、GPT-3 (Brown et al., 2020)、GPT-4 (OpenAI, 2023)、LLaMA (Touvron et al., 2023)已经展示了理解异构文本数据的复杂依赖关系并提供合理生成的强大能力。

最近,人们对将语言模型应用于时间序列任务越来越感兴趣。例如,

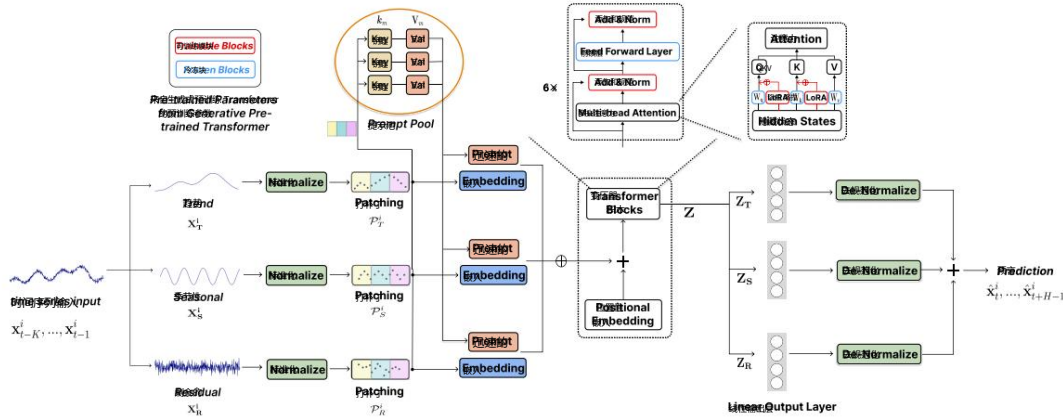


图 1: 提出的 TEMPO-GPT 的架构。趋势XT季节性XS和剩余XR分量被视为不同的语义归纳偏差,以馈送到预先训练的变压器中。基于选择的提示池可以促进基于查询P\*和键 k 检索相似的时间序列模式。

Xu 和 Salim 天真地将时间序列数据转换为文本序列输入,并取得了令人鼓舞的结果。孙等人。提出文本原型对齐嵌入,使法学硕士能够处理时间序列数据,并使法学硕士在不影响语言能力的情况下更容易接受嵌入,这尚未成功超越其他最先进的 (SOTA) 模型。此外,于等人。提出了一种利用法学硕士进行可解释的金融时间序列预测的创新方法。然而, (Yu et al., 2023) 提出的方法的一个显著局限性是它需要跨不同领域的样本使用不同的模板,从而降低了其灵活性。(Zhou et al., 2023) 和 (Chang et al., 2023) 中的工作与我们的工作最相关,因为它们都通过战略性地利用和微调法学硕士来引入时间序列分析方法。

然而,这些研究直接利用时间序列数据来构建嵌入,没有充分区分时间序列数据的固有特征,这对于在法学硕士中解耦此类信息具有挑战性 (Shin et al., 2020)。此外,针对具有时间序列的多模态数据的 LLM 工作仍然非常有限。METS (Li et al., 2023) 是追求这一方向的早期作品之一。虽然实验结果令人鼓舞,但由于时间序列和文本之间的嵌入对齐是特定的,因此很难将 METS 扩展到其他模式。

及时调整。快速调整是一种高效、低成本的方法,可以使预先训练的基础模型适应新的下游任务,该模型已适应跨各个领域的下游任务。

在 NLP 领域,通过提示调整 (Lester et al., 2021) 或前缀调整 (Li & Liang, 2021) 来使用具有可训练表示的软提示。提示技术也已扩展到 CV 任务,例如对象检测 (Li et al., 2022) 和图像字幕 (Zhang et al., 2022a) 等。

多模态作品,例如 CLIP (Radford et al., 2021),使用文本提示来执行图像分类并实现 SOTA 性能。此外,我们认可我们在基于检索的提示设计研究领域的工作。先前的相关工作包括 L2P (Wang 等人, 2022c),它展示了存储在共享池中的可学习提示的潜力,可以在没有排练缓冲区的情况下实现持续学习;以及 Dualprompt (Wang 等人, 2022b),它引入了双提示空间提示架构、为一般知识和专家信息维护单独的提示编码等。我们的研究建立在这些概念的基础上,探索使用基于检索的提示选择,专门用于跨时间序列预测问题的时间推理和知识共享。

### 3 方法论

在我们的工作中,我们采用了一种混合方法,将统计时间序列分析的稳健性与数据驱动方法的适应性结合起来。如图1所示,我们提出了一种将 STL (Cleveland 等人, 1990) 的季节和趋势分解集成到预训练 Transformer 中的新颖方法。这种策略使我们能够利用统计和机器学习方法的独特优势,增强我们的模型有效处理时间序列数据的能力。此外,引入提示池来帮助减少非平稳时间序列预测中分布变化的影响。提示池对时间效果进行编码,可以根据

预印本。

输入实例查询和键之间的相似性,使模型能够专注于适当地回忆相关的过去知识。

3.1问题定义

给定前K 个时间戳的观测值,多元时间序列预测的任务旨在预测下一个 H 时间戳的值。那是,

$$x^i_{t_1}, \dots, x^i_{t+H-1} = F(x^{n}_{t-1}; t-K, i, x, \text{维}; \phi) \tag{1}$$

其中 $x^i_{t_1}, \dots, x^i_{t+H-1}$  是根据对应通道*i*的时间戳*t*进行 H 步估计的向量,可以通过模型*F*推断出到第 *i* 个特征。给定历史值 $x^{n}_{t-K}, \dots, x^{n}_{t-1}$ ,

参数 $\phi$ 和提示 $V_i$

3.2时间序列输入表示

通过将复杂的输入数据分解为有意义的组件 (例如考虑法学硕士时的文本标记)来表示复杂的输入数据,有助于通过对这些组件的单独检查和建模来最佳地提取信息。对于时间序列,由于其在现实世界应用程序中的常见用法及其对从业者的可解释性,我们考虑趋势-季节性分解。

给定输入 $X \in \mathbb{R}^{n \times L}$ ,其中*n*是特征 (通道)大小, L是时间序列的长度,加法 STL 分解可以表示为:

$$X = \text{趋势} + \text{季节} + \text{残差} \tag{2}$$

这里, *i*是多元时间序列*k*输入的通道索引 (对应于某个协变量),趋势 $XT \in \mathbb{R}^{n \times L}$ 捕获数据中潜在的长期模式 $j = -k$ ,其中 $m = 2k + 1$ 和*k*是平均步长。季节性分量 $XS \in \mathbb{R}^{n \times L}$ 封装了重复的短期周期,可以在应用 Loess 平滑器消除趋势后进行估计 (Cleveland 等 1990)。残差分量 $XR \in \mathbb{R}^{n \times L}$ 表示提取趋势和季节性后数据的剩余部分。

此外,这种分解明确地能够识别异常观察结果以及季节性模式或趋势的变化。从理论角度来看,我们在附录E中建立了时间序列预测和频域预测之间的联系,其中我们的研究结果表明分解显著简化了预测过程。请注意,这种分解在基于电流互感器的方法中更为重要,因为从理论上讲,注意力机制可能无法自动解开不正交趋势和季节信号:

定理3.1假设我们有时间序列信号 $X = XT + XSt + XRt$ ,  $t \in [t_1, t_n]$ 。设 $E = \{e_1, e_2, \dots, e_n\}$ 表示一组正交基。设 $ES \subseteq E$ 表示 $XSt$ 上有非零特征值的*E*子集,  $ET \subseteq E$ 表示 $XT$ 上有非零 $X_i T t X_i = 0$ 的*E*子集,则 $ET \cap ES = \emptyset$ ,即*E*特征值。如果 $XSt$ 和 $XT$ 不正交,即不能将两个信号分解到两个不相交的碱基组上。

$$\sum_{i=1}^n \text{我} = 1 \quad \text{英石}$$

证明可以在附录E 中找到。对于方法论部分的其余部分,我们将使用趋势组件*XT*作为示例案例。分解后,我们分别对每个组件应用反向实例归一化 (Kim et al., 2022),以促进知识转移并最大限度地减少分布变化带来的损失。也就是说,对于来自*XT*的每个样本*x*

时间*t*,  $x^i$ 的第 *i* 个通道  $x^i_{\text{时间}} = \gamma T x^i_{\text{时间}} - Et x^i_{\text{时间}} / \text{变量 } x^i_{\text{时间}} + T + \beta T$ ,其中 $Et x^i_{\text{时间}}$ 和 $\text{Var } x^i_{\text{时间}}$ 是实例特定的平均值和标准差;  $\gamma T$ 和 $\beta T$ 是趋势分量的可训练仿射*T t*参数向量。然后,按照 (Nie et al., 2023),我们将时间列修补与时间编码相结合,通过将相邻时间步聚合到标记中来提取局部语义,显着增加历史范围,同时减少冗余。具体来说,我们得到 $X^i$ 第 *i* 个归一化趋势分量的修补标记,其中*LP*是修补程序长度,  $N = + 2$ 是修补程序数量, *S*是步幅。我们

$$\text{我} \text{与} P_{\text{时间}} \in \mathbb{R}^{LP \times N},$$

$$\frac{(L-LP)}{S}$$

获得补丁代币 $P_S$  我和 $P_R$  右 以同样的方式。然后,我们将修补后的时间序列标记提供给

预印本。

表示  $P) \in R$  架构,以有效地将其语言能力转移到新颖的顺序模式,其中  $L_E$  是嵌入  $P \times L_E$  用于语言模型嵌入层  $f$  以获得大小。

3.3提示设计

通过将特定于任务的知识编码为指导模型预测的提示,提示方法在许多应用中显示出了有希望的结果。以前的工作主要集中在利用固定提示通过微调来提高预训练模型的性能。考虑到现实世界时间序列数据具有分布变化的典型非平稳性质 (Huang et al., 2020),我们引入了一个存储为不同键值对的共享提示池。理想情况下,我们希望模型能够利用相关的过去经验,其中相似的输入时间序列倾向于从池中检索同一组提示 (Wang 等人, 2022c)。这将允许模型在单个时间序列实例输入级别有选择地调用最具代表性的提示。此外,这种方法可以提高建模效率和预测性能,因为该模型能够通过共享表示池更好地识别和应用跨不同数据集的学习模式。池中的提示可以编码与不同时间段相关的时间依赖性、趋势或季节性影响。具体来说,定义了提示键值对池

作为:

$$VK = \{(k1, V1), (k2, V2), \dots, (kM, VM)\}, \text{其中 } M \text{ 是提示池的长度, } V_m \in R^{L_p \times L_E} \text{ 是令牌长度为 } L_p \text{ 的单个提示,其中 } R^{L_E} \text{ 的形状和与 } P \text{ 相同的嵌入大小 } L_E \text{ 和 } k_m \in K =$$

$\{k_m\}$  是时间中  $k_m = 1$  公里=

分数匹配过程可以用分数匹配函数  $\gamma: P \times R^{L_E} \rightarrow R$  来表示,其中  $\gamma: R^{L_E} \rightarrow R$  模型以端到端的方式进行训练,以根据提示优化预测。用于检索前  $K$  个对应值的查询  $P$  来自修补的时间序列输入。因此,相似的时间序列可以分配给相似的提示。将  $\{s_j\}$  表示为所选前  $K$  个提示的  $K$  索引的子集,我们的输入  $K$   $j=1$

趋势嵌入如下:

$$xT = [Vs1; \dots; VSK; PT], \quad 1 \leq K \leq M, \tag{4}$$

我们沿着时间长度维度连接所有标记,例如  $xS$ 、 $xR$ 。每个实例都可以分配给多个提示,这些提示可以联合编码与预测任务相关的知识,例如时间序列表现出的周期性模式、流行趋势或季节性效应。

3.4生成式预训练Transformer架构

我们使用基于解码器的生成预训练变压器 (GPT) 作为骨干来构建时间序列表示的基础。为了以数据有效的方式利用分解的语义信息,我们选择将提示和不同的组件连接在一起并将它们放入 GPT 块中。具体来说,我们的时间序列嵌入的输入可以表示为:  $x = xT \oplus xS \oplus xR$ , 其中  $\oplus$  对应于连接操作,  $x^*$  可以被视为不同的句子。请注意,另一种替代方法是构建单独的 GPT 块来处理不同类型的时间序列组件。在 GPT 块内,我们采用 (Zhou et al., 2023) 中使用的策略,在训练期间冻结前馈层。同时,我们选择更新位置嵌入层和层归一化层的梯度。此外,我们采用 LORA (低秩适应) (Hu et al., 2021) 来有效地适应不同的时间序列分布,因为它用更少的参数来执行适应。

总体预测结果应该是各个组成部分预测的相加组合。最后, GPT 块中  $n$  个特征的输出  $Z$  可以根据它们在输入顺序中的位置分为  $ZT$ 、 $ZS$ 、 $ZR \in n \times P \times L_E \times R$  (对应于趋势、季节性和残差的输出)。然后,每个  $Z$  分量被输入到全连接层以生成预测  $Y^* \in R^{n \times L_H}$ , 其中  $L_H$  是预测长度。之后,我们根据归一化步骤中使用的相应统计数据对  $Y^*$  进行去归一化:  $Y^*$  通过重新组合这些加性元素,我们的方法旨在重建完整的时间轨迹

$$= \text{Var } x^* t^* t^* + \epsilon^* \cdot \frac{Y^* t^* - B^* Y^*}{Y^* t^* - B^* Y^*} + \text{乙醚 } x^* t^*。$$



预印本。

最能代表分解后的  $Y = Y_T + Y_S + Y_R$  捕获的不同时间尺度的潜在动态。输入表示。预测结果可表述为：

可解释性。由于我们假设趋势、季节和残差分量可能与最终输出存在非线性关系,因此我们可以基于 GPT 的输出构建可解释的广义加性模型(GAM) (Hastie, 2017; Enouen & Liu, 2022)来学习三个分量如何相互作用,即:  $g(Y) = F_0 + F_{It}(x_{It})$ , 其中  $F_0$  是归一化常数,脚注  $i$  对应于趋势、季节和残差分量。 $\{It\}$  是一组多个交互组件。然后,我们可以计算一阶灵敏度指数 (Sobol , 1990) 或 SHapley Additive exPlanations (SHap) (Lundberg & Lee, 2017) 来衡量每个组件的灵敏度。

## 4实验

我们使用七个流行的时间序列基准数据集 (Zhou et al., 2021), 包括用于长期预测的 ETTm1、ETTm2、ETTh1、ETTh2、天气、电力和交通,以及我们提出的短期预测数据集 TETS。我们使用 GPT-2 (Brown 等人, 2020) 作为我们的骨干来构建如图 1 所示的模型。为了全面展示我们模型的性能,我们将 TEMPO 与以下 14 种方法进行长期预测和短期预测进行比较期限预测: (1)预训练的基于LLM的模型,包括Bert、GPT2 (GPT4TS)、T5和LLaMA。(2)基于Transformer的模型,包括PatchTST、Autoformer、FEDformer、Informer、ETSformer、非稳态Transformer (Non-Stat.)和Reformer。(3)基于线性的模型的变体,包括NLinear、DLinear和LightTS模型。(4)通用2D变异模型,包括TimesNet。按照传统的预测工作,我们在本节中报告均方误差 (MSE) 和平均绝对误差 (MAE) 结果。详细的实验设置和基线请参阅附录B和D。

### 4.1长期预测结果

表1对不同预测长度的 MSE 和 MAE 指标的多个时间序列预测模型进行了定量比较,得分越低表明预测越准确。我们提出的模型 TEMPO 在所有数据集的所有预测范围内平均超过了现有基线,凸显了 TEMPO 的广泛适用性。我们的模型获得了最高的平均性能得分。具体来说,与之前最先进的模型 PatchTST 相比,它在 MAE 中分别将天气和 ETTm1 数据集提高了49.41%和54.30%。它还确保了众多单独数据集预测长度配置的最低错误率。例如,我们的模型在 7 个数据集上平均预测长度为 96 和 192 时,比PatchTST 分别高出62.87%和35.59%。与其他预训练的预测模型相比, TEMPO 在不同的时间序列数据集中始终提供最佳结果。与此同时,T5 表现最好的法学硕士之一 只能对电力需求数据产生准确的预测。

这些结果表明,将 LLM 与提示池相结合并实施时间序列分解可以显著提高时间序列预测的准确性和效率。

### 4.2基础模型训练

鉴于TEMPO在每个特定领域的实验设置下表现出的强大性能,从跨领域基础模型的角度来看,我们进一步研究使用在跨领域数据集上训练的单个TEMPO模型是否仍然可以在未见过的领域上获得可比的性能 (如反对为每个特定领域训练一个模型并在同一领域的数据上测试它)。我们在 5 个数据集上训练了一个模型,并在 ETTm2 和 Traffic 上进行了测试,这些数据在训练过程中是模型看不到的。对来自未知领域的数据进行预测本质上更具挑战性,因此我们预计 MSE 和 MAE 会比表1 更高。有关详细的实验设置,请参阅附录。表2提供了我们的模型与其他基线模型在训练期间模型看不到的两个多元时间序列数据集 (即 ETTm2 和流量)上的全面比较。由于其独特的特征,我们选择ETTh2 和流量数据集进行测试。ETTh2数据集与模型的训练数据有一些相似之处,其中包括其他 ETT 数据,但也

预印本。

表 1:时间序列基准数据集的长期预测结果。我们使用预测长度  $O \in \{96, 192, 336, 720\}$ 。MSE 越低表明性能越好。此后,对于表格,最佳结果用粗体标记,次优结果用下划线标记,分别用 MSE/MAE 表示。

地平线	模型	天气 MSE/	ETTh1	ETT2	ETTM1	ETTM2	电话记录	交通
		MAE MSE/MAE MSE/MAE MSE/MAE MSE/MAE MSE/MAE						
96	TEMPO 0.008/0.048 GPT4TS	0.201/0.268	0.085/0.166	<b>0.221/0.293</b>	0.015/0.083	0.010/0.066		
	0.162/0.212 T5 0.152/0.201 Bert	0.376/0.397	0.139/0.238	<b>0.389/0.422</b>	0.292/0.346	0.173/0.262		
	0.150/0.197 FEDformer	0.411/0.425	0.123/0.228	<b>0.365/0.352</b>	0.291/0.346	0.186/0.277		
	0.217/0.296 Autoformer	0.459/0.443	0.130/0.228	<b>0.376/0.363</b>	0.291/0.344	0.177/0.263		
	0.266/0.336 Informer 0.300/0.384	0.376/0.419	0.193/0.308	<b>0.350/0.366</b>	0.379/0.419	0.203/0.287		
	PatchTST 0.149/0.198 Reformer	0.449/0.459	0.201/0.310	<b>0.463/0.388</b>	0.505/0.475	0.255/0.339		
	0.689/0.596 LightTS 0.182/0.242	0.865/0.713	0.274/0.368	<b>0.753/0.529</b>	0.672/0.571	0.365/0.453		
	DLinear 0.176/0.237 TimesNet	0.370/0.399	0.129/0.228	<b>0.276/0.389</b>	0.290/0.342	0.165/0.255		
	0.172/0.220 非统计。 0.173/0.223	0.837/0.728	0.312/0.402	<b>0.752/0.423</b>	0.538/0.528	0.658/0.619		
	ETSformer 0.197/0.281	0.424/0.432	0.207/0.308	<b>0.357/0.309</b>	0.140/0.233	<b>0.049/0.022</b>		0.615/0.391
				.289/0.353	.299/0.343	.167/0.269		
				0.384/0.402	0.168/0.278	<b>0.469/0.291</b>		
				0.513/0.491	0.169/0.278	<b>0.761/0.588</b>		
				0.494/0.479	0.187/0.308	<b>0.407/0.382</b>		
192	TEMPO 0.027/0.082 GPT4TS	0.349/0.387	0.315/0.355	0.118/0.207	0.115/0.184	0.125/0.214	0.350/0.310	
	0.204/0.248 T5 0.196/0.242 Bert	0.416/0.418	0.354/0.389	0.332/0.372	0.229/0.301	0.153/0.251	0.407/0.290	
	0.196/0.240 FEDformer	0.457/0.447	0.396/0.418	0.342/0.379	0.249/0.319	0.149/0.240	0.385/0.259	
	0.276/0.336 Autoformer	0.548/0.492	0.415/0.433	0.339/0.374	0.243/0.305	0.149/0.240	0.387/0.261	
	0.307/0.367 Informer 0.598/0.544	0.420/0.448	0.429/0.439	0.426/0.441	0.269/0.328	0.201/0.315	0.604/0.373	
	PatchTST 0.194/0.241 Reformer	0.500/0.482	0.456/0.452	0.553/0.496	0.281/0.34	0.222/0.334	0.616/0.382	
	0.752/0.638 LightTS 0.227/0.287	1.008/0.792	5.602/1.931	0.795/0.669	0.533/0.563	0.296/0.386	0.696/0.379	
	DLinear 0.220/0.282 TimesNet	0.413/0.421	0.339/0.379	0.332/0.369	0.220/0.292	0.157/0.240	0.379/0.256	
	0.219/0.261 非统计。 0.245/0.285	0.923/0.766	11.120/2.979	0.658/0.592	1.078/0.827	0.348/0.433	0.733/0.420	
	ETSformer 0.237/0.312	0.475/0.462	0.520/0.504	0.400/0.407	0.311/0.382	0.213/0.316	0.601/0.382	
		.405/0.416	0.383/0.418	.335/0.365	224/0.303	.153/0.249	0.423/0.287	
		0.436/0.429	0.402/0.414	0.374/0.387	0.249/0.309	0.184/0.289	0.617/0.336	
		0.534/0.504	0.512/0.493	0.459/0.444	0.280/0.339	0.182/0.286	0.613/0.340	
		0.538/0.504	0.430/0.439	0.408/0.41	0.253/0.319	0.199/0.315	0.621/0.399	
336	TEMPO 0.111/0.170 GPT4TS	0.408/0.425	0.393/0.406	0.254/0.319	0.214/0.283	0.152/0.254	0.388/0.311	
	0.254/0.286 T5 0.249/0.285 Bert	0.442/0.433	0.373/0.407	0.366/0.394	0.286/0.341	0.169/0.266	0.412/0.294	
	0.247/0.280 FEDformer	0.482/0.465	0.430/0.443	0.374/0.399	0.308/0.358	0.166/0.258	0.398/0.267	
	0.339/0.38 Autoformer	0.576/0.511	0.414/0.437	0.374/0.395	0.299/0.346	0.165/0.256	0.402/0.271	
	0.359/0.395 Informer 0.578/0.523	0.459/0.465	0.496/0.487	0.445/0.459	0.325/0.366	0.214/0.329	0.621/0.383	
	PatchTST 0.245/0.282 模型	0.521/0.496	0.482/0.486	0.621/0.537	0.339/0.372	0.231/0.338	0.622/0.337	
	0.639/0.596 LightTS 0.282/0.334	1.107/0.809	4.721/1.835	1.212/0.871	1.363/0.887	0.300/0.394	0.777/0.420	
	DLinear 0.265/0.319 TimesNet	0.422/0.436	0.329/0.380	0.366/0.392	0.274/0.329	0.163/0.259	0.392/0.264	
	0.280/0.306 非统计。 0.321/0.338	1.097/0.835	9.323/2.769	0.898/0.721	1.549/0.972	0.350/0.433	0.742/0.420	
	ETSformer 0.298/0.353	0.518/0.488	0.626/0.559	0.438/0.438	0.442/0.466	0.230/0.333	0.613/0.386	
		.439/0.443	.448/0.465	.369/0.386	.281/0.342	.169/0.267	0.436/0.296	
		0.491/0.469	0.452/0.452	0.410/0.411	0.321/0.351	0.198/0.300	0.629/0.336	
		0.588/0.535	0.552/0.551	0.495/0.464	0.334/0.361	0.200/0.304	0.618/0.328	
		0.574/0.521	0.485/0.479	0.435/0.428	0.314/0.357	0.212/0.329	0.622/0.396	
720	TEMPO 0.251/0.282 GPT4TS	0.504/0.493	0.425/0.449	0.381/0.400	0.329/0.362	0.189/0.189	0.449/0.335	
	0.326/0.337 T5 0.324/0.336 Bert	0.477/0.456	0.406/0.441	0.417/0.421	0.378/0.401	0.285/0.297	0.450/0.312	
	0.324/0.334 FEDformer	0.643/0.553	0.440/0.463	0.427/0.428	0.391/0.408	0.204/0.291	0.433/0.288	
	0.403/0.428 Autoformer	0.665/0.563	0.461/0.470	0.421/0.426	0.401/0.410	0.210/0.293	0.434/0.290	
	0.419/0.428 Informer 1.059/0.741	0.506/0.507	0.463/0.474	0.543/0.490	0.421/0.415	0.246/0.355	0.626/0.382	
	PatchTST 0.314/0.334 Reformer	0.514/0.512	0.515/0.511	0.671/0.561	0.433/0.432	0.254/0.361	0.660/0.408	
	1.130/0.792 LightTS 0.352/0.386	1.181/0.865	3.647/1.625	1.166/0.823	3.379/1.338	0.373/0.439	0.864/0.472	
	DLinear 0.333/0.362 TimesNet	0.447/0.466	0.379/0.422	0.416/0.420	0.362/0.385	0.197/0.290	0.432/0.286	
	0.365/0.359 非统计。 0.414/0.41	1.257/0.889	3.874/1.697	1.102/0.841	2.631/1.242	0.340/0.420	0.755/0.423	
	ETSformer 0.352/0.288	0.547/0.533	0.863/0.672	0.527/0.502	0.675/0.587	0.265/0.360	0.658/0.407	
		.472/0.490	.605/0.551	.425/0.421	.397/0.421	.203/0.301	0.466/0.315	
		0.521/0.500	0.462/0.468	0.478/0.450	0.408/0.403	0.220/0.320	0.640/0.350	
		0.643/0.616	0.562/0.56	0.585/0.516	0.417/0.413	0.222/0.321	0.653/0.355	
		0.562/0.535	0.5/0.497	0.499/0.462	0.414/0.413	0.233/0.345	0.632/0.396	
平均。	TEMPO 0.099/0.146 GPT4TS	0.366/0.393	0.326/0.361	0.192/0.252	0.167/0.224	0.138/0.230	0.351/0.292	
	0.237/0.270 T5 0.230/0.266 Bert	0.427/0.426	0.354/0.394	0.352/0.383	0.266/0.326	0.167/0.263	0.414/0.294	
	0.229/0.263 FEDformer	0.498/0.473	0.399/0.427	0.358/0.388	0.284/0.340	0.161/0.253	0.395/0.267	
	0.309/0.360 Autoformer	0.562/0.502	0.417/0.436	0.356/0.385	0.280/0.331	0.163/0.253	0.397/0.268	
	0.338/0.382 Informer 0.634/0.548	0.440/0.460	0.437/0.449	0.448/0.452	0.305/0.349	0.214/0.327	0.610/0.376	
	PatchTST 0.225/0.264 Reformer	0.496/0.487	0.450/0.459	0.588/0.517	0.327/0.371	0.227/0.338	0.628/0.379	
	0.803/0.656 LightTS 0.261/0.312	1.040/0.795	4.431/1.729	0.961/0.734	1.410/0.810	0.311/0.397	0.764/0.416	
	DLinear 0.248/0.300 TimesNet	0.413/0.430	0.330/0.379	0.351/0.380	0.255/0.315	0.161/0.252	0.390/0.263	
	0.259/0.287 非统计。 0.288/0.314	1.029/0.805	6.736/2.191	0.799/0.671	1.479/0.915	0.338/0.422	0.741/0.422	
	ETSformer 0.271/0.334	0.491/0.479	0.602/0.543	0.435/0.437	0.409/0.436	0.229/0.329	0.622/0.392	
		.422/0.437	.431/0.446	.357/0.378	.267/0.333	.166/0.263	0.433/0.295	
		0.458/0.450	0.414/0.427	0.400/0.406	0.291/0.333	0.192/0.295	0.620/0.336	
		0.570/0.537	0.526/0.516	0.481/0.456	0.306/0.347	0.193/0.296	0.624/0.34	
		0.542/0.510	0.439/0.452	0.429/0.425	0.293/0.342	0.208/0.323	0.621/0.396	

表现出一定的独特分布。另一方面,流量数据集完全不同于模型之前遇到过的任何数据。TEMPO 优于所有基线模型,实现最低的 MSE 和 MAE。请注意,TEMPO 的平均 MSE 和 MAE 分别低30.8%和20.5% 分别比 ETTm2 数据集表现最佳的基线模型 (T5) 好。我们的模型还 MSE 和 MAE 的增加幅度最小,如误差增加所示 当从已见的数据集转移到未见的数据集时,表2中的行。误差增加表明

预印本。

表 2:我们的基础模型训练设置在两个未见过的数据集上的长期预测结果。对于每个数据集,我们显示每个预测长度的 MSE 和 MAE。此外,我们还报告了与最佳单域 MSE 和 MAE 相比的误差增加 (EI),如表1 所示。较低的 EI 表明精度降低较小,性能较好。

数据集长度		节奏 GPT4TS				T5				FEDformer PatchTST LightTS				D线性时代网			
		MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE				
ETTM2	96	0.059/0.157	0.196/0.275	0.187/0.269	0.31/0.389	0.26/0.334	0.19/0.28							0.2/0.291	0.213/0.294		
	192	0.14/0.228	0.241/0.305	0.243/0.306	0.388/0.439	0.281/0.341	0.277/0.347	0.284/0.36	0.24/0.306								
	336	0.23/0.293	0.292/0.336	0.294/0.339	0.497/0.517	0.323/0.364	0.402/0.42	0.413/0.441	0.293/0.342								
	720	0.335/0.363	0.382/0.392	0.38/0.393	0.481/0.478	0.409/0.413	0.804/0.612	0.622/0.556	0.410/0.419								
	平均。	0.191/0.26	0.278/0.327	0.276/0.327	0.419/0.456	0.318/0.363	0.418/0.415	0.38/0.412	0.289/0.34								
	EI	0.014/0.031	0.1/0.098	0.099/0.098	0.242/0.227	0.141/0.134	0.241/0.186	0.203/0.183	0.112/0.111								
交通	96	0.431/0.364	0.529/0.388	0.519/0.376	1.076/0.668	0.873/0.575	0.528/0.374	0.585/0.41	0.596/0.404								
	192	0.468/0.363	0.525/0.38	0.519/0.369	2.616/1.255	0.889/0.568	0.525/0.378	0.59/0.413	0.57/0.388								
	336	0.495/0.365	0.544/0.389	0.545/0.379	1.79/0.945	0.935/0.609	0.548/0.373	0.613/0.423	0.67/0.435								
	720	0.55/0.398	0.566/0.397	0.584/0.4	0.923/0.58	0.984/0.598	0.571/0.38	0.619/0.422	0.671/0.438								
	平均。	0.486/0.373	0.541/0.389	0.542/0.381	1.601/0.862	0.92/0.588	0.543/0.376	0.602/0.417	0.627/0.416								
	EI	0.135/0.109	0.19/0.126	0.191/0.118	1.25/0.6	0.569/0.324	0.19/0.113	0.251/0.154	0.276/0.153								

与分别在 ETTm2 和 Traffic 上训练和测试的最佳性能模型 (TEMPO) 相比,MSE 和 MAE 的增加量。令人惊讶的是,TEMPO 甚至优于表1中的一些基线模型,这些基线分别在 ETTm2 或 Traffic 上进行训练。这一发现揭示了 TEMPO 的强大通用性,并表明其作为基础时间序列预测模型的潜力,在未知领域保持稳健的性能。

4.3数据高效适应

对于数据高效的适应评估,也称为少镜头设置 (Wu et al., 2023; Zhou et al., 2023) ,我们仅采用了一小部分 (例如,5%、10%)训练时间步长。这种方法使我们能够探索模型在训练数据有限的情况下生成准确预测的能力,这在冷启动或非平稳场景中可能特别重要。结果如表3和表4所示。与TimesNet、DLinear、PatchTST 和 GPT2 等成熟模型相比,我们提出的模型在所有数据集上都表现出优越性。特别是,与 TimesNet和 DLinear 相比,我们的模型的 MSE 分别降低了约 20% 和 25%。这些结果凸显了我们模型的稳健性和数据效率。

4.4使用背景信息进行短期预测

数据集和指标。在本节中,我们向社区介绍 TETS,这是一个基于 S&P 500 数据集构建的新基准数据集,结合了上下文信息和时间序列。根据 (Papadimitriou 等人, 2020) ,我们选择 SMAPE 作为我们的指标。详细的数据集设置请参见附录B.2 ,建议的数据集收集流程请参见附录F。

上下文信息。为了将上下文信息合并到我们提出的TEMPO 中,我们利用生成式预训练转换器的内置标记化功能来派生输入文本的嵌入。然后,我们利用这些文本嵌入 Text来构造具有可学习参数的软提示,并将它们连接在输入嵌入的开头。即 $x = T \text{ ext} \oplus xT \oplus xS \oplus xR$ 。该方法并不严格限于我们提出的模型,而是可以切实应用于类似的工作中,以增强其处理上下文信息并从中受益的能力。附录C.3提供了与其他上下文信息设计策略的比较,以供进一步参考。

使用不同基线模型和我们的模型在 TETS 数据集上的 SMAPE 结果列于表5 (域内扇区)和表6 (跨域扇区,也称为零样本设置,作为来自这些扇区的数据样本在训练阶段看不到) 。检查所有部门的结果,我们提出的模型将时间序列数据与补充摘要 (上下文)数据相结合,在域内显着优于所有基线方法



预印本。

表 3:5% 数据的小样本学习结果。我们使用预测长度  $O \in \{96, 192, 336, 720\}$ 。A 较低的 MSE/MAE 表示更好的性能,最佳结果以粗体突出显示。 - 方法  
5% 的时间序列不足以构成训练集。

地平线	模型	天气 ETTm2 ECL MSE/MAE	WAE MSE/MAE	WAE MSE/MAE	交通
96	TEMPO 0.013/0.057 GPT4TS	0.108/0.218	0.064/0.157	0.113/0.208	0.303/0.268
	0.175/0.230 FEDformer 0.229/0.309	0.386/0.405	0.199/0.280	0.143/0.241	0.419/0.298
	TS 0.203/0.246 Autoformer	0.628/0.544	0.229/0.320	0.235/0.322	0.670/0.421
	0.227/0.299 Informer	0.438/0.424	0.225/0.305	0.148/0.245	0.421/0.290
	0.497/0.497 PatchTST 0.171/0.224	0.726/0.578	0.232/0.322	0.297/0.367	0.795/0.481
	Reformer 0.406/0.435 Light TS	1.130/0.775	3.599/1.478	1.265/0.919	1.557/0.821
	0.230/0.285 DLinear 0.184/0.242	0.399/0.414	0.206/0.288	0.145/0.244	0.404/0.286
	TimesNet 0.207/0.253 固定	1.234/0.798	3.883/1.545	1.414/0.855	1.586/0.841
	0.215/0.252 ETSformer 0.218 /	1.048/0.733	1.108/0.772	0.639/0.609	1.157/0.636
	0.295 0.207/0.295 伯特	0.332/0.374	0.236/0.326	0.427/0.304	0.150/0.251
		0.606/0.518	0.315/0.385	0.272/0.354	0.492
		0.823/0.587	0.238/0.316	0.484/0.518	1.468/0.821
		1.031/0.747	0.404/0.485	0.697/0.638	1.643/0.855
		0.477/0.443	0.229/0.305	0.146/0.242	0.427/0.300
192	TEMPO 0.108/0.173 GPT4TS	0.262/0.334	0.201/0.269	0.141/0.242	0.390/0.300
	0.227/0.276 FEDformer 0.265/0.317	0.440/0.438	0.256/0.316	0.159/0.255	0.434/0.305
	TS 0.265/0.330 Autoformer	0.666/0.566	0.394/0.361	0.247/0.341	0.653/0.405
	0.278/0.333 Informer	0.446/0.428	0.265/0.330	0.166/0.263	0.434/0.298
	0.620/0.545 PatchTST 0.230/0.277	0.750/0.591	0.291/0.357	0.308/0.375	0.837/0.503
	Reformer 0.446/0.450 Light TS	1.150/0.788	3.578/1.475	1.298/0.939	1.596/0.834
	0.274/0.323 DLinear 0.228/0.283	0.441/0.436	0.264/0.324	0.163/0.260	0.412/0.294
	TimesNet 0.272/0.307 0.290/0.307	1.287/0.839	3.553/1.484	1.240/0.919	1.602/0.844
	0.294/0.331 0.262/0.297	1.097/0.756	1.317/0.850	0.772/0.678	1.688/0.848
		0.358/0.390	0.306/0.373	0.163/0.263	0.447/0.315
		.681/0.539	.311/0.361	.318/0.396	0.894/0.517
	固定式	0.844/0.591	0.298/0.349	0.501/0.531	1.509/0.838
	教育考试服务中心	1.087/0.766	0.479/0.521	0.718/0.648	1.856/0.928
	伯特	0.534/0.464	0.295/0.344	0.165/0.261	0.450/0.312
336	TEMPO 0.231/0.287 GPT4TS	0.383/0.412	0.291/0.335	0.175/0.270	0.440/0.321
	0.286/0.322 FEDformer 0.353/0.392	0.485/0.459	0.318/0.353	0.179/0.274	0.449/0.313
	TS 0.361/0.388 Autoformer	0.807/0.628	0.378/0.427	0.267/0.356	0.707/0.445
	0.351/0.393 Informer	0.540/0.484	0.361/0.388	0.188/0.286	0.464/0.313
	0.649/0.547 PatchTST 0.294/0.326	0.851/0.659	0.478/0.517	0.354/0.411	0.867/0.523
	Reformer 0.465/0.459 Light TS	1.198/0.809	3.561/1.473	1.302/0.942	1.621/0.841
	0.318/0.355 DLinear 0.279/0.322	0.499/0.467	0.334/0.367	0.183/0.281	0.439/0.310
	TimesNet 0.313/0.328 0.353/0.348	1.288/0.842	3.446/1.460	1.253/0.921	1.668/0.868
	固定 ETSformer 0.359 /0.398	1.147/0.775	1.415/0.879	0.901/0.745	1.826/0.903
	0.325/0.340 伯特	0.402/0.416	0.380/0.423	0.175/0.278	0.478/0.333
		.786/0.597	.338/0.366	.340/0.415	0.853/0.471
		0.870/0.603	0.353/0.38	0.574/0.578	1.602/0.86
		1.138/0.787	0.552/0.555	0.758/0.667	2.08/0.999
		0.580/0.490	0.375/0.392	0.187/0.280	0.475/0.329
720	TEMPO 0.351/0.371 GPT4TS	0.521/0.485	0.675/0.523	0.228/0.315	-/-
	0.366/0.379 FEDformer 0.391/0.394	0.577/0.499	0.460/0.436	0.233/0.323	-/-
	TS 0.494/0.456 Autoformer	0.822/0.633	0.523/0.510	0.318/0.394	-/-
	0.387/0.389 Informer	0.636/0.539	0.494/0.456	0.238/0.325	-/-
	0.570/0.522 PatchTST 0.384/0.387	0.857/0.655	0.553/0.538	0.426/0.466	-/-
	Reformer 0.471/0.468 Light TS	1.175/0.794	3.896/1.533	1.259/0.919	-/-
	0.401/0.418 D线性 0.364/0.388	0.767/0.587	0.454/0.432	0.233/0.323	-/-
	TimesNet 0.400/0.385 固定	1.247/0.828	3.445/1.460	1.249/0.921	-/-
	0.452/0.407 0.461/0.461 ETSformer	1.200/0.799	1.822 / 0.984	1.200/0.871	-/-
	0.395/0.388 伯特	0.511/0.489	0.674/0.583	0.219/0.311	-/-
		.796/0.593	.509/0.465	.635/0.613	-/-
		0.893/0.611	0.475/0.445	0.952/0.786	-/-
		1.245/0.831	0.701/0.627	1.028/0.788	-/-
		0.686/0.553	0.525/0.461	0.239/0.323	-/-
平均。	TEMPO 0.175/0.222 GPT4TS	0.319/0.362	0.307/0.321	0.164/0.259	0.378/0.296
	0.263/0.301 FEDformer 0.309/0.353	0.472/0.450	0.308/0.346	0.178/0.273	0.434/0.305
	TS 0.331/0.355 Autoformer	0.730/0.592	0.381/0.404	0.266/0.353	0.676/0.423
	0.310/0.353 Informer	0.515/0.469	0.336/0.370	0.185/0.280	0.440/0.300
	0.584/0.527 PatchTST 0.269/0.303	0.796/0.620	0.388/0.433	0.346/0.404	0.833/0.502
	Reformer 0.447/0.453 Light TS	1.163/0.791	3.658/1.489	1.281/0.929	1.591/0.832
	0.305/0.345 DLinear 0.263/0.308	0.526/0.476	0.314/0.352	0.181/0.277	0.418/0.296
	TimesNet 0.298/0.318 固定	1.264/0.826	3.581/1.487	1.289/0.904	1.618/0.851
	0.327/0.328 ETSformer 0.333 /	1.123/0.765	1.415/0.871	0.878/0.725	1.557/0.795
	0.371 伯特 0.297/0.320	0.400/0.417	0.399/0.426	0.176/0.275	0.450/0.317
		.717/0.561	.344/0.372	.402/0.453	0.867/0.493
		0.857/0.598	0.341/0.372	0.627/0.603	1.526/0.839
		1.125/0.782	0.534/0.547	0.8/0.685	1.859/0.927
		0.569/0.488	0.356/0.376	0.184/0.277	0.451/0.314

部门 (能源除外)和跨领域部门。TEMPO 实现了30%以上的减少  
与除能源 (Ene) 之外的所有部门的最佳基线结果相比,SMAPE 误差  
和消费者周期性 (CC)。特别值得注意的是医疗保健 (Hea) 部门  
域内数据集和跨域数据集中的房地产 (RE) 部门,其中  
分别减少了51.2%和57.6%。在域内部门和跨域部门

预印本。

表 4:10% 数据的小样本学习结果。我们使用预测长度 $O \in \{96, 192, 336, 720\}$ 。  
MSE/MAE 越低表示性能越好,最佳结果以粗体突出显示。

地平线	模型	天气	ETTM1	ETTM2	ECL	MSE/MAE	MSE/	交通
		MAE MSE/MAE	MSE/MAE MSE/MAE					
96	TEMPO	0.028/0.084	GPT4TS	0.028/0.084	0.05/0.139	0.108/0.241	0.29/0.262	
		0.163/0.215	FEDformer	0.188/0.253	0.39/0.404	0.188/0.269	0.139/0.237	0.414/0.297
	TS	0.191/0.238	Autoformer	0.221/0.297	0.578/0.518	0.291/0.399	0.231/0.323	0.639/0.4
		0.374/0.401	PatchTST	0.165/0.215	0.407/0.413	0.208/0.289	0.143/0.24	0.41/0.287
		Reformer	0.335/0.38	0.217/0.269	0.774/0.614	0.352/0.454	0.261/0.348	0.672/0.405
		0.171/0.224	0.184/0.23	0.192/0.234	1.162/0.785	3.203/1.407	1.259/0.919	1.557/0.821
		0.199/0.272	0.185/0.232		0.41/0.419	0.191/0.274	0.14/0.238	0.403/0.289
	光TS				1.442/0.847	4.195/1.628	0.993/0.784	1.527/0.815
	DL线性				0.921/0.682	0.813/0.688	0.35/0.425	1.157/0.636
	时代网				0.352/0.392	0.213/0.303	0.15/0.253	0.419/0.298
	固定式				0.583/0.501	0.212/0.285	0.299/0.373	0.719/0.416
	教育考试服务中心				0.761/0.568	0.229/0.308	0.42/0.466	1.412/0.802
192	伯特				0.911/0.688	0.331/0.43	0.599/0.587	1.643/0.855
					0.478/0.449	0.219/0.295	0.143/0.239	0.419/0.298
	速度	0.085/0.15			0.232/0.307	0.182/0.251	0.143/0.24	0.382/0.3
	GPT4TS	0.21/0.254			0.429/0.423	0.251/0.309	0.156/0.252	0.426/0.301
	美联储前任	0.25/0.304			0.617/0.546	0.307/0.379	0.261/0.356	0.637/0.416
	TS	0.23/0.271			0.459/0.443	0.265/0.327	0.161/0.256	0.421/0.29
	自动成型机	0.27/0.322			0.754/0.592	0.694/0.691	0.338/0.406	0.727/0.424
	告密者	0.552/0.478			1.172/0.793	3.112/1.387	1.16/0.873	1.454/0.765
	补丁TST	0.21/0.257			0.437/0.434	0.252/0.317	0.16/0.255	0.415/0.296
	塑身机	0.522/0.462			1.444/0.862	4.042/1.601	0.938/0.753	1.538/0.817
	光TS	0.259/0.304			0.957/0.701	1.008/0.768	0.376/0.448	1.207/0.661
	DL线性	0.215/0.263			0.382/0.412	0.278/0.345	0.164/0.264	0.434/0.305
336	时代网	0.245/0.283			0.63/0.528	0.27/0.323	0.305/0.379	0.748/0.428
	固定式	0.269/0.295			0.781/0.574	0.291/0.343	0.411/0.459	1.419/0.806
	教育考试服务中心	0.279/0.332			0.955/0.703	0.4/0.464	0.62/0.598	1.641/0.854
	伯特	0.234/0.272			0.522/0.471	0.27/0.327	0.162/0.256	0.433/0.302
	TEMPO	0.192/0.239	GPT4TS	0.407/0.408	0.261/0.321	0.171/0.267	0.419/0.312	
		0.256/0.292	FEDformer	0.312/0.346	0.469/0.439	0.307/0.346	0.175/0.27	0.434/0.303
	TS	0.279/0.304	Autoformer	0.32/0.351	0.998/0.775	0.543/0.559	0.36/0.445	0.655/0.427
		PatchTST	0.259/0.297	塑身机	0.531/0.471	0.325/0.364	0.184/0.279	0.439/0.299
		0.715/0.535	0.303/0.334		0.869/0.677	2.408/1.407	0.41/0.474	0.749/0.454
		0.258/0.299	0.305/0.321	0.37/0.357	1.227/0.908	3.255/1.421	1.157/0.872	1.521/0.812
		0.356/0.386	0.289/0.312		0.476/0.454	0.306/0.353	0.18/0.276	0.426/0.304
	光TS				1.45/0.866	3.963/1.585	0.925/0.745	1.55/0.819
720	DL线性				0.998/0.716	1.031/0.775	0.428/0.485	1.334/0.713
	时代网				0.419/0.434	0.338/0.385	0.181/0.282	0.449/0.313
	固定式				0.725/0.568	0.323/0.353	0.319/0.391	0.853/0.471
	教育考试服务中心				0.803/0.587	0.348/0.376	0.434/0.473	1.443/0.815
	伯特				0.991/0.719	0.469/0.498	0.662/0.619	1.711/0.878
					0.593/0.496	0.347/0.374	0.185/0.28	0.443/0.307
	TEMPO	0.312/0.332	GPT4TS	0.734/0.555	0.441/0.416	0.231/0.313	0.483/0.338	
		0.321/0.339	FEDformer	0.387/0.393	0.569/0.498	0.426/0.417	0.233/0.317	0.487/0.337
	TS	0.353/0.359	Autoformer	0.39/0.396	0.693/0.579	0.712/0.614	0.53/0.585	0.722/0.456
		PatchTST	0.332/0.346	Reformer	0.686/0.548	0.452/0.436	0.241/0.326	0.476/0.32
		0.611/0.5	0.377/0.382	0.32/0.346	0.81/0.63	1.913/1.166	0.715/0.685	0.847/0.499
		0.381/0.371	0.441/0.405	0.437/0.448	1.207/0.797	3.909/1.543	1.203/0.898	1.605/0.846
平均。		0.373/0.369			0.681/0.556	0.433/0.427	0.241/0.323	0.474/0.331
	光TS				1.366/0.85	3.711/1.532	1.004/0.79	1.588/0.833
	DL线性				1.007/0.719	1.096/0.791	0.611/0.597	1.292/0.726
	时代网				0.49/0.477	0.436/0.44	0.223/0.321	0.484/0.336
	固定式				0.769/0.549	0.474/0.449	0.369/0.426	1.485/0.825
	教育考试服务中心				0.844/0.581	0.461/0.438	0.51/0.521	1.539/0.837
	伯特				1.062/0.747	0.589/0.557	0.757/0.664	2.66/1.157
					0.672/0.535	0.457/0.432	0.243/0.324	0.485/0.331
	TEMPO	0.154/0.201	0.238/0.275	0.35/0.339	0.234/0.282	0.163/0.265	0.394/0.303	
		FEDformer	GPT4TS	0.464/0.441	0.293/0.335	0.176/0.269	0.44/0.31	
		0.284/0.324	TS	0.263/0.293	0.722/0.605	0.463/0.488	0.346/0.427	0.663/0.425
		Autoformer	0.3/0.342	Informers	0.521/0.469	0.312/0.354	0.182/0.275	0.436/0.299
		0.597/0.495	PatchTST	0.242/0.279	0.802/0.628	1.342/0.93	0.431/0.478	0.749/0.446
		塑身机	0.546/0.469	0.289/0.322	1.192/0.821	3.37/1.44	1.195/0.891	1.534/0.811

预印本。

表 5:根据历史数据预测未来 4 季度 EBITDA 的短期预测结果  
域内数据集的 20 个季度的数据。我们过滤掉 80%/90% 的异常 SMAPE 值  
Papadimitriou 等人提出的阈值。(2020).(BM:基础材料;CS:通讯  
服务;Ene:能量;FS:金融服务;Hea:医疗保健;技术:技术;Uti:实用程序。)

	BM	CS	埃内	FS	希亚	技术	乌蒂
节奏	12.38/10.82	11.43/11.43	21.13/21.09	12/9.31	5.49/5.49	10.91/11.19	7.45/7.45
GPT4TS	24.26/24.95	26.91/26.68	43.4/44.96	22.55/22.73	15.33/15.08	26.73/26.73	17.98/17.98
伯特	27.35/27.35	26.52/27.14	3.51/45.01	23.01/23.19	14.49/14.49	26.84/26.98	18.27/18.27
T5	29.11/30.35	33.6/33.6	48.53/52.58	26.53/26.71	20.08/19.53	32.28/32.52	23.64/24.03
美洲驼	25.22/25.88	28.52/29.04	44.73/47.65	22.98/23.15	14.8/14.77	29.11/29.01	17.84/17.84
自动成型机	36.94/37.5	34.6/34.6	33.85/33.85	25.25/25.25	22.39/22.39	36.97/37.45	17.88/17.88
举报人	33.73/35.28	38.66/38.66	31.56/31.56	27.7/27.7	16.87/16.87	26.75/26.91	16.54/16.54
补丁TST	32.35/32.91	18.72/18.72	26.92/27.5	16.63/16.63	13.22/13.22	19.86/20.05	15.43/15.43
塑身机	31.61/32.17	20.91/21.43	23.79/23.79	15.9/15.9	11.24/11.24	18.87/19.04	17.07/17.44
FED前	49.97/53.19	60.82/61.71	65.37/66.69	57.53/58.49	55.84/57.14	71.45/73.27	34.88/35.59
光TS	29.62/30.71	18.58/18.58	18.77/19.98	16.08/16.08	13.96/13.96	21.65/22.35	13.07/13.07
线性	28.48/29.04	17.76/18.81	21.42/20.46	16.89/16.89	16.0/16.0	25.22/25.56	13.7/13.7
非线性	27.99/29.14	23.32/23.69	25.92/25.16	20.19/20.19	19.27/19.27	30.6/30.75	13.67/14.04
时代网	29.12/29.68	17.63/17.63	16.62/16.62	14.39/14.74	11.4/11.6	18.8/19.34	13.65/13.65
ESTformer	29.12/29.7	39.52/39.52	37.73/37.73	24.36/24.36	22.66/22.66	27.04/27.21	15.78/15.78

表 6:根据历史数据预测未来 4 季度 EBITDA 的短期预测结果  
跨域数据集的20个季度的数据。我们过滤掉 80%/90% 的异常 SMAPE 值  
阈值。(CC:周期性消费者;CD:防御性消费者;lnd:工业;RE:房地产。)

	CCRE	CD	工业
节奏	10.21/10.67	10.09/10.16	8.25/8.24
GPT4TS	23.98/24.63	19.01/19.01	19.48/19.54
伯特	24.86/25.29	19.26/19.26	19.6/19.88
T5	32.33/33.09	22.72/22.83	24.38/24.63
美洲驼	25.35/26.31	20.01/19.97	20.45/20.72
自动成型机	20.98/21.22	21.26/21.89	19.41/19.95
举报人	39.06/39.27	62.78/71.58	36.66/37.0
补丁TST	15.28/15.48	18.93/19.69	14.41/14.9
塑身机	15.78/15.98	18.77/18.64	14.89/15.26
FED前	49.76/50.86	49.69/53.83	47.22/47.98
光TS	16.8/16.79	19.51/20.31	14.62/14.93
线性	14.72/14.66	16.04/16.94	11.89/12.29
非线性	15.76/15.79	19.7/19.83	14.03/14.34
时代网	12.56/12.73	15.94/16.5	11.78/12.03
ESTformer	14.1/14.27	47.16/51.13	18.16/18.45

与\_\_\_\_\_

最高基线结果。Abs-SMAPE结果示于表15和表16中。

5分析

5.1解释模型预测

SHAP (SHapley Additive exPlanations)值作为特征重要性的综合度量,量化每个特征对所有可能的预测输出的平均贡献

特征组合。如图2(a)和图2(b)所示,当应用于我们的季节性和趋势分解(STL),广义加性模型(GAM)的SHAP值表明

趋势成分对模型预测的主要影响,意味着显着

模型对数据内整体方向变化的依赖性。季节性成分,它体现了重复出现的模式,在某些时间间隔也表现出巨大的贡献。

相反,残差部分,解释了不规则的数据波动,似乎发挥了影响相对较小。“错误”列中不断上升的值,表示差异

预印本。

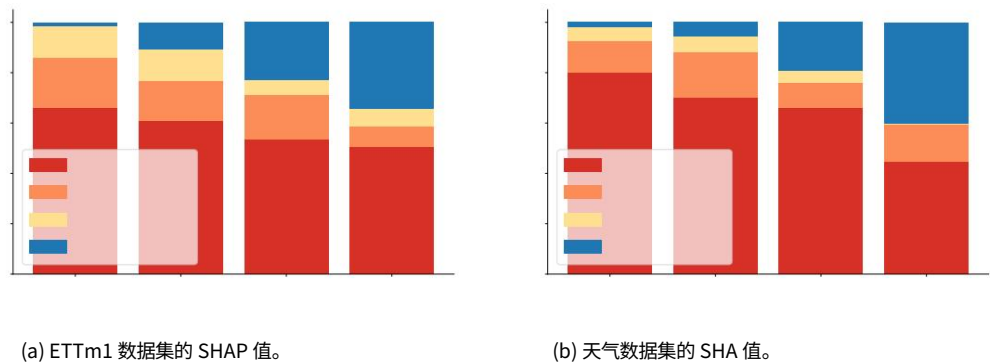


图 2:TEMPO 分解成分的 SHAP (Shapley Additive exPlanations)值。

模型的预测和真实情况之间的差异表明,随着预测长度的增加,模型的准确性可能会下降,这在大多数实验运行中确实观察到。这可能是由于数据波动性增加或模型潜在的过度拟合所致,这凸显了定期模型评估和调整的必要性。在这种情况下,STL 分解被证明是非常有价值的,因为它使我们能够识别和量化每个组件对整体预测的单独贡献,如 SHAP 值所示。这种详细的理解可以产生关于预训练变压器如何解释和利用分解预处理步骤的关键见解,从而为模型优化和增强提供坚实的基础。

5.2消融研究

表 7:针对即时池和时间序列分解的长期预测的平均 MSE/MAE 消融研究。最佳结果以粗体突出显示。

	TEMPO 不带提示,不分解			
	MSE/MAE	MSE/MAE	MSE/MAE	MSE/MAE
天气	0.099/0.146	0.107/0.154		0.228/0.263
ETTh1	0.192/0.252	0.196/0.258		0.386/0.403
ETTh2	0.177/0.229	0.179/0.235		0.278/0.331
ETTh1	0.366/0.393	0.435/0.439		0.469/0.454
ETTh2	0.326/0.361	0.351/0.374		0.369/0.403

所提供的消融研究 (表7)提供了关于提示和分解组件对我们模型性能的影响的重要见解。在此表中,报告了三种情况下各种数据集的 MSE 和 MAE:原始模型配置 (“我们的”)、没有提示池的模型 (“w/o 提示”)和没有分解的模型操作 (“不分解”)。

如果没有提示组件,所有数据集的 MSE 和 MAE 值都会增加,表明预测精度会下降。这表明提示有助于提高模型的预测性能。如果没有分解组件,性能下降甚至更加明显,MSE 和 MAE 值进一步增加就证明了这一点。

这意味着分解组件对于模型的性能也至关重要。例如,在 “ETTh1”数据集上,由于缺少提示和缺少 STL 分解,MSE 分别上升了 18.8% 至 28.1%。请注意,模型仅应用提示池而不进行分解会对主干模型的性能产生不利影响,请参见表1。

这可能归因于从提示池中检索纯时间序列数据的挑战,因为未经分解的可传输信息可能有限。这些观察结果揭示了提示和分解组件对于模型预测准确性和预测能力的至关重要性。

5.3即时池案例研究

如图3 所示,在我们的案例研究中,我们首先将 ETTm2 数据集中具有不同输入分布的三个时间序列实例: x1、 x2和x3分解为其趋势、季节性和残差分量。分解后, x1和x2的趋势分量显示出惊人的相似性

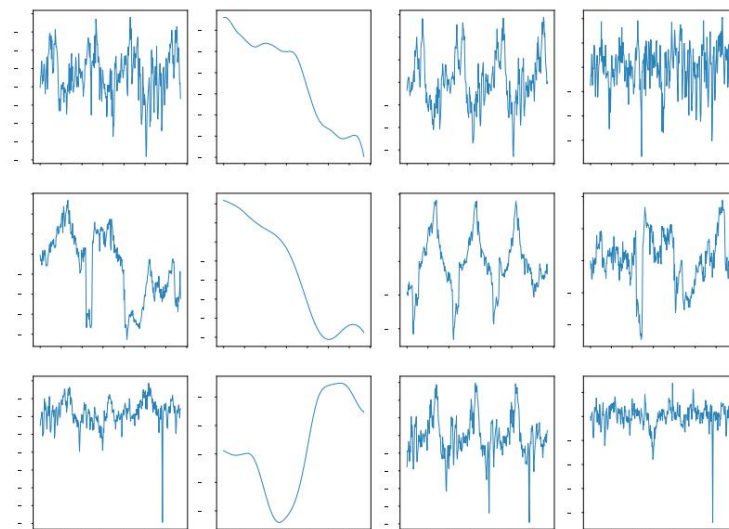


图 3:针对 ETTm2 数据集的三个不同实例的案例研究,其中 $x_1$ 和 $x_2$  (前两个)具有相似的趋势模式,  $x_2$ 和 $x_3$  (后两个)具有相似的季节性模式。

$x_2$ 和 $x_3$ 的季节性分量也相似。因此,当这些分量输入到TEMPO时,  $x_1$ 和 $x_2$ 的趋势分量从趋势信息经常选择的提示池中检索相同的提示ID,而 $x_2$ 和 $x_3$ 的季节分量检索相同的提示ID,通常与季节性因素相关。这一发现验证了该模型成功地识别并利用了潜在趋势和季节性水平上的代表性相似性,即使完整的实例有所不同,这也符合在不断变化的模式中巩固知识的目标。至关重要的是,这种分解过程使不同的组件能够处理语言模型的不同语义,从而简化任务复杂性。该案例演示了所提出的基于分解的提示调整如何能够发现并应用相关时间序列实例之间的共享结构信息,同时还通过组件分离简化预测问题。我们对附录 C.4 中提出的提示池进行了更多的案例研究和分析。

## 六,结论

本文提出了一种基于即时选择的生成变压器 TEMPO,它在时间序列预测方面实现了最先进的性能。我们在基于 Transformer 的预训练主干中引入了提示池和季节性趋势分解的新颖集成,使模型能够根据时间序列输入相对于不同时间语义组件的相似性,专注于适当地回忆过去相关时间段的知识。此外,我们还证明了 TEMPO 与多模型输入的有效性,有效利用时间序列预测中的上下文信息。最后,通过大量的实验,我们强调了TEMPO 在准确性、数据效率和通用性方面的优越性。值得进一步研究的一个潜在限制是,具有更好数字推理能力的优秀法学硕士可能会产生更好的结果。此外,根据我们的跨领域实验,这项工作的潜在未来轨迹涉及进一步开发时间序列分析的基础模型。

预印本。

---

## 参考

- 白绍杰, J Zico Kolter 和 Vladlen Koltun. 用于序列建模的通用卷积和循环网络的实证评估。 arXiv 预印本 arXiv:1803.01271, 2018。
- 汤姆·B·布朗、本杰明·曼、尼克·莱德、梅兰妮·苏比亚、贾里德·卡普兰、普拉芙拉·达里瓦尔、阿文德·尼拉坎坦、普拉纳夫·希亚姆、吉里什·萨斯特里、阿曼达·阿斯科尔、桑迪尼·阿加瓦尔、阿里尔·赫伯特·沃斯、格雷琴·克鲁格、TJ·赫尼汉、Rewon Child、阿迪亚 Ramesh、Daniel M. Ziegler、Jeff Wu、Clemens Winter、Christopher Hesse、Mark Chen、Eric Sigler、Mateusz Litwin、Scott Gray、Benjamin Chess、Jack Clark、Christopher Berner、Sam McCandlish、Alec Radford、Ilya Sutskever 和 Dario Amodei。语言模型是小样本学习者。神经信息处理系统的进展, abs/2005.14165, 2020。
- 曹德福, 王玉静, 段娟勇, 张策, 朱夏, 黄从瑞, 佟云海, 徐必雄, 白静, 童杰, 等。用于多元时间序列预测的谱时间图神经网络。神经信息处理系统的进展, 33:17766–17778, 2020。
- 张青、彭文智、陈天福。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。
- 罗伯特·B·克利夫兰、威廉·S·克利夫兰、让·E·麦克雷和伊尔玛·特彭宁。Stl: 季节性趋势分解。J. 关。统计, 6 (1): 3-73, 1990。
- 雅各布·德夫林、张明伟、肯顿·李和克里斯蒂娜·图塔诺娃。BERT: 用于语言理解的深度双向转换器的预训练。计算语言学协会北美分会 2019 年会议记录: 人类语言技术 (NAACL-HLT), 美国明尼苏达州明尼阿波利斯, 2019 年 6 月 2-7 日, 第 4171-4186 页, 2019 年。
- 詹姆斯·埃努恩和刘岩。通过特征交互检测和稀疏选择的稀疏交互附加网络。神经信息处理系统的进展, 35:13908–13920, 2022 年。
- 罗伯特·菲尔德斯、安德鲁·哈维、迈克·韦斯特和杰夫·哈里森。预测、结构时间序列模型和卡尔曼滤波器。《运筹学杂志》, 42:1031, 11 1991。doi: 10.2307/2583225。
- 特雷弗·J·哈斯蒂。广义加性模型。S 的统计模型, 第 249-307 页。劳特利奇, 2017 年。
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang 和 Weizhu Chen。Lora: 大型语言模型的低阶适应。 arXiv 预印本 arXiv:2106.09685, 2021。
- 黄碧伟、张坤、张吉吉、约瑟夫·拉姆齐、鲁本·桑切斯-罗梅罗、克拉克·格利摩和伯恩哈德·斯科普夫。从异构/非平稳数据中发现因果关系。机器学习研究杂志, 21(1):3482–3534, 2020。
- Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi 和 Jaegul Choo。可逆实例归一化, 用于针对分布变化进行准确的时间序列预测。国际学习表征会议, 2022 年。
- 尼基塔·基塔耶夫、卢卡斯·凯撒和安塞姆·列夫斯卡娅。改革者: 高效的变压器。第八届学习表征国际会议 (ICLR), 埃塞俄比亚的斯亚贝巴, 2020 年 4 月 26-30 日, 2020 年 4 月 26-30 日。
- 布莱恩·莱斯特、拉米·艾尔富和诺亚·康斯坦特。规模的力量可实现参数高效的提示调整。 arXiv 预印本 arXiv:2104.08691, 2021。
- 李军、刘军、程思博、Rossella Arcucci 和 Shenda Hong。冻结语言模型有助于心电图零样本学习, 2023 年。
- Liunian Harold Li, Pengchuan 张, Haotian 张, Jianwei Yang, Chunyuan Li, Yiwuzhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang 等。扎根的语言图像预训练。IEEE/CVF 计算机视觉和模式识别会议论文集, 第 10965–10975 页, 2022 年。



预印本。

---

李翔丽莎和梁珀西。前缀调优:优化生成的连续提示。 arXiv 预印本 arXiv:2101.00190, 2021。

李亚光、于玫瑰、赛勒斯·沙哈比和刘岩。扩散卷积循环神经网络:数据驱动流量预测。在国际学习表征会议 (ICLR 18), 2018 年。

刘世展、于航、廖聪、李建国、林伟耀、Alex X Liu 和 Schahram Dust-dar。Pyraformer:用于长期时间序列建模和预测的低复杂度金字塔注意力。学习表征国际会议,2021 年。

刘勇,吴海旭,王建民,龙明生。非平稳变压器:探索时间序列预测中的平稳性。神经信息处理系统的进展, 2022 年。

斯科特·M·伦德伯格 (Scott M Lundberg) 和李秀仁 (Su-In Lee)。解释模型预测的统一方法。进展神经信息处理系统,2017 年 30 月。

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong 和 Jayant Kalagnanam。时间序列值 64 个字:使用 Transformer 进行长期预测。国际学习表征会议 (ICLR 23), 2023 年。

开放人工智能。Gpt-4 技术报告, 2023 年。

欧阳龙、吴杰夫、徐江、迪奥戈·阿尔梅达、卡罗尔·L·温赖特、帕梅拉·米什金、张冲、桑迪尼·阿加瓦尔、卡塔琳娜·斯拉马、亚历克斯·雷·约翰·舒尔曼、雅各布·希尔顿、弗雷泽·凯尔顿、卢克·E·米勒、麦迪·西蒙斯、阿曼达·阿斯克尔、彼得·韦林德、保罗·弗朗西斯·克里斯蒂安诺、扬·雷克和瑞安·J·洛。训练语言模型遵循人类反馈的指令。

ArXiv, abs/2203.02155, 2022。URL <https://api.semanticscholar.org/CorpusID:246426909>。

Antony Papadimitriou, Urjitkumar Patel, Lisa Kim, Grace Bang, Azadeh Nematzadeh 和刘晓默。大规模财务预测的多方面方法。首届 ACM 国际金融人工智能会议论文集, 第 1-8 页, 2020 年。

亚历克·雷德福、卡蒂克·纳拉西姆汉、蒂姆·萨利曼斯、伊利亚·苏茨克弗等。通过生成预训练提高语言理解。2018。

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者。OpenAI 博客, 1(8):9, 2019。

亚历克·雷德福、金钟旭、克里斯·哈拉西、阿迪亚·拉梅什、加布里埃尔·吴、桑迪尼·阿加瓦尔、吉里什·萨斯特里、阿曼达·阿斯克尔、帕梅拉·米什金、杰克·克拉克等。从自然语言监督中学习可迁移的视觉模型。在国际机器学习会议上, 第 123 页。8748-8763。PMLR, 2021。

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li 和 Peter J Liu。使用统一的文本到文本转换器探索迁移学习的局限性。机器学习研究杂志, 21 (1) :5485-5551, 2020。

泰勒·辛·亚萨曼·拉泽吉、罗伯特·L·洛根四世、埃里克·华莱士和萨米尔·辛格。自动提示:通过自动生成的提示从语言模型中提取知识。2020 年自然语言处理经验方法 (EMNLP) 会议记录, 第 4222-4235 页, 2020 年。

西玛·西亚米·纳米尼、内达·塔瓦科利和阿克巴·西亚米·纳明。arima 和 lstm 在预测时间序列方面的比较。2018 年第 17 届 IEEE 机器学习与应用国际会议 (ICMLA), 第 1394-1401 页。IEEE, 2018。

伊利亚·米罗维奇·索博尔。非线性数学模型的灵敏度估计。数学模型, 2(1):112-118, 1990。

拉里·R·斯奎尔、丽莎·根泽尔、约翰·T·维克斯泰德和理查德·G·莫里斯。记忆巩固。寒冷的 Spring Harbor 生物学观点, 7(8):a021766, 2015。

预印本。

---

孙晨曦,李亚良,李红艳,洪申达。测试:文本原型对齐嵌入

激活llm的时间序列能力。 arXiv 预印本 arXiv:2308.08241, 2023。

雨果·图夫龙、蒂博·拉夫里尔、戈蒂埃·伊扎卡尔、泽维尔·马丁内特、玛丽-安妮·拉肖、蒂莫西·拉克鲁瓦、巴蒂斯特·罗齐埃、纳曼·戈亚尔、埃里克·汉布罗、费萨尔·阿扎尔、奥雷利安·罗德里格斯、阿曼德·朱兰、爱德华·格雷夫和纪尧姆·兰普尔。 Llama :开放高效的基础语言模型。

ArXiv,abs/2302.13971,2023,URL <https://api.semanticscholar.org/CorpusID:257219404>。

王志远、Xovee Xu、张伟峰、Goce Trajcevski、Tingzhong 和 Fan Zhou。学习时间序列预测的潜在季节性趋势表示。神经信息处理系统的进展,2022a。

王子峰、张子钊、Sayna Ebrahimi、孙若曦、张涵、李振宇、任晓琪、苏国龙、文森特·佩罗、Jennifer Dy 等。双提示:补充提示,实现免排练的持续学习。欧洲计算机视觉会议,第 631-648 页。

施普林格,2022b。

王子峰、张子钊、李晨宇、张涵、孙若曦、任晓琪、苏国龙、文森特·佩罗、Jennifer Dy 和 Tomas Pfister。学习提示持续学习。 IEEE/CVF 计算机视觉和模式识别会议论文集,第 139-149 页,2022c。

Gerald Woo、Chenghao Liu、Doyen Sahoo、Akshat Kumar 和 Steven Hoi。 Etsformer:用于时间序列预测的指数平滑变体。 arXiv 预印本 arXiv:2202.01381, 2022。

吴海旭、徐杰辉、王建民、龙明生。 Autoformer:具有自相关性的分解变体,用于长期序列预测。神经信息处理系统进展(NeurIPS),第 101-112 页,2021 年。

吴海旭、胡腾革、刘勇、周航、王建民、龙明生。 Timesnet:用于一般时间序列分析的时间二维变化建模。载于2023 年第十一届学习表征国际会议。网址[https://openreview.net/forum?id=ju\\_Uqw384Oq](https://openreview.net/forum?id=ju_Uqw384Oq)。

薛浩和弗洛拉·D·萨利姆。 Promptcast:一种新的基于提示的时间序列学习范例预测。 2022 年。

于新丽,陈正,凌远,董树静,刘宗义,陆彦斌。时间数据符合llm 可解释的金融时间序列预测。 arXiv 预印本 arXiv:2306.11025, 2023。

曾爱玲、陈木希、张雷、徐强。变体对时间序列有效吗预测? 2023 年。

张浩天、张彭川、胡小伟、陈彦春、李柳年、戴喜阳、王丽娟、袁璐、黄振能和高剑峰。 Glipv2:统一本地化和视觉语言理解。神经信息处理系统的进展,35:36067-36080,2022a。

张天平、张一卓、曹伟、边江、易小涵、郑顺和李健。少即是多:使用面向轻采样的 MLP 结构进行快速多元时间序列预测。 arXiv预印本 arXiv:2207.01186, 2022b。

周浩一、张尚航、彭杰琪、张帅、李建新、熊辉和张万才。告密者:超越用于长序列时间序列预测的高效变体。 AAAI会议记录,2021 年。

周田、马自清、温青松、王雪、孙亮、金荣。 FEDformer:用于长期序列预测的频率增强分解变体。在过程中。第 39 届国际机器学习会议 (ICML 2022),2022 年。

田周、牛培松、王雪、孙亮、金荣。一应俱全:通过预训练的 lm 进行通用时间序列分析。神经信息处理系统的进展,2023 年。

预印本。

展示柜

图 4:长期预测结果的可视化。我们的模型 TEMPO 和GPT4TS 在 ETTh1 数据集上的比较

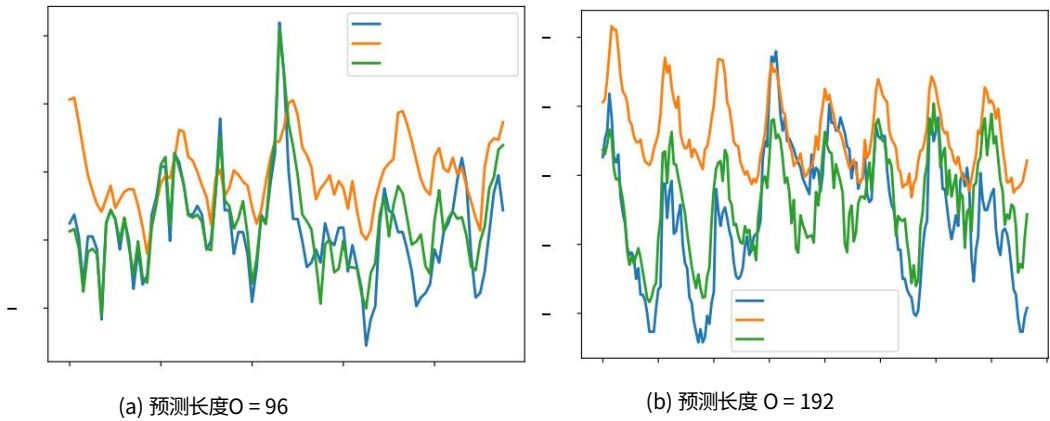
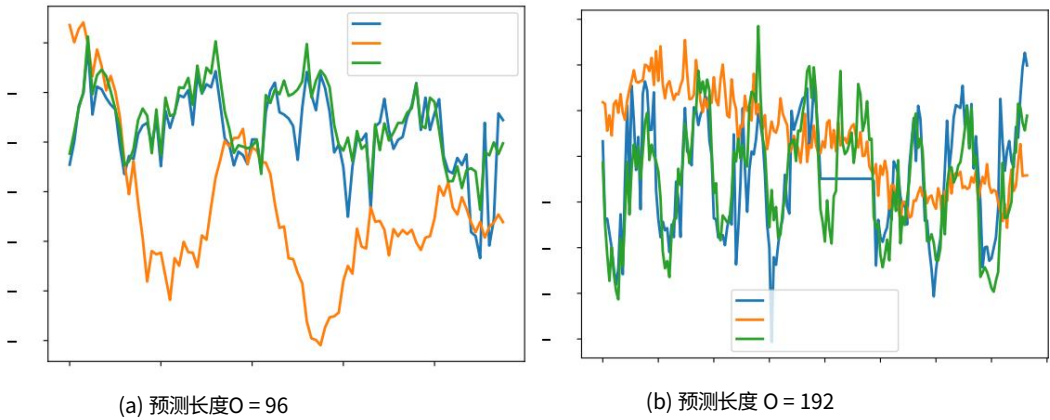


图 5:长期预测结果的可视化。我们的模型 TEMPO 和GPT4TS 在 ETTh2 数据集上的比较



在图4、5、6、7、8中,我们绘制了在给定回溯窗口的情况下我们的模型和 GPT4TS 模型的预测值的比较。如数据集中所示,我们能够预测接近真实情况,这也通过我们优于表1中其他模型的性能得到了证明。我们在不同的预测长度下选择具有不同特征的时间序列 $O \in \{96, 192\}$ :具有高变异性的时间序列 (图4a)、周期性 (图4b、5b、6b)、趋势变化的非周期性 (图6a、图8a)

B实验设置

B.1特定领域的实验和基础模型实验  
细节

众所周知,通道独立性对于时间序列数据集效果很好,因此我们将每个多元时间序列视为多个独立的单变量时间序列。我们使用七个流行的时间序列基准数据集 (Zhou et al., 2021): ETTm1、ETTm2、ETTh1、ETTh2、天气、电力和交通。 1) ETTm1、ETTm2、ETTh1、ETTh2 包含两个电站按 15 分钟级别和每小时级别的用电负荷。 2)天气数据集包含21个气象

图 6:长期预测结果的可视化。我们的模型 TEMPO 和GPT4TS 在 ETTm1 数据集上的比较

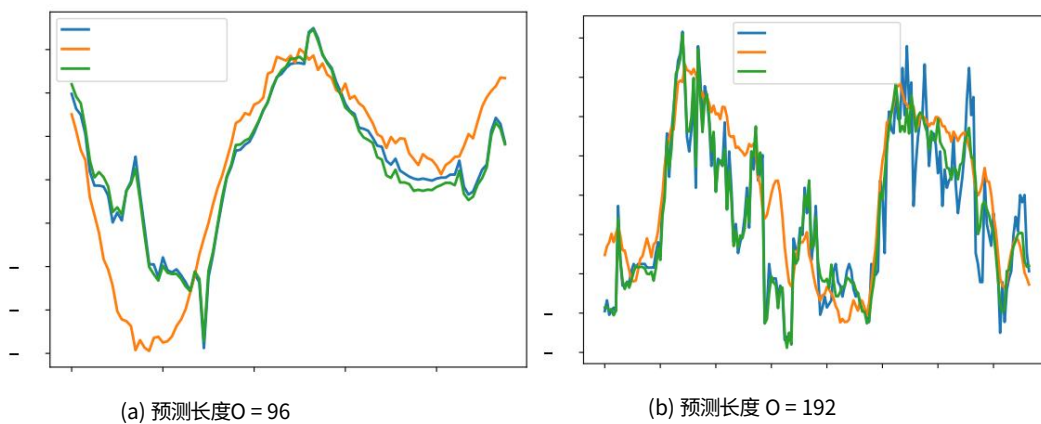


图 7:长期预测结果的可视化。我们的模型 TEMPO 和GPT4TS 在 ETTm2 数据集上的比较

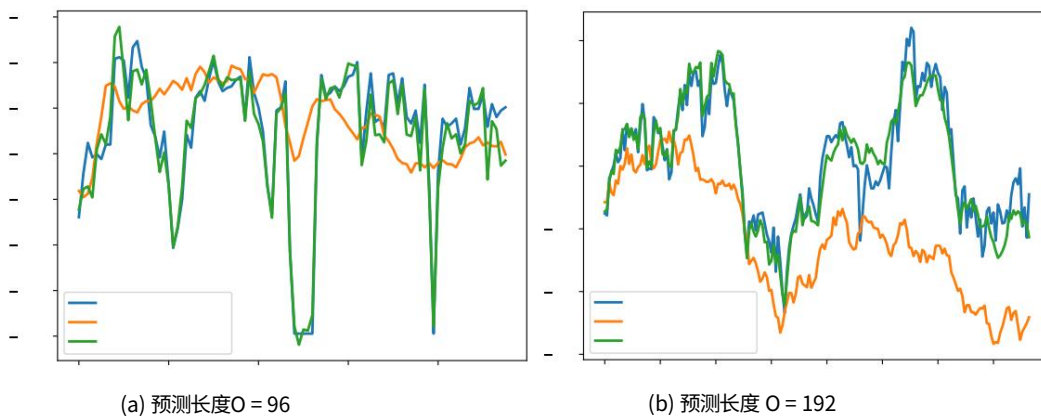
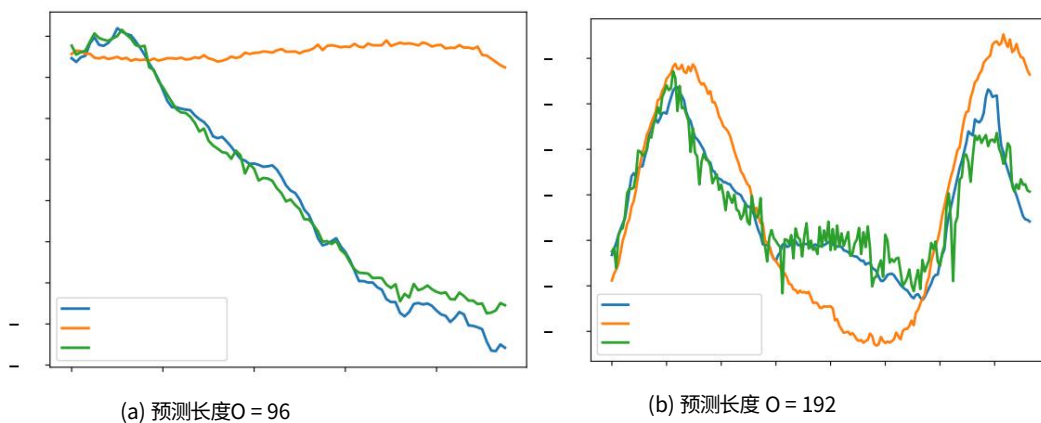


图 8:长期预测结果的可视化。我们的模型 TEMPO 和GPT4TS 在天气数据集上的比较



预印本。

德国1年内的指标； 3)电力数据集包含用电量； 4)交通数据集包含整个加利福尼亚州高速公路系统的占用率。如同传统实验设置,每个时间序列 (ETTh1,ETTh2,ETTM1,天气、电力、ETTM2,Traffic)分为三部分:训练数据、验证数据和测试数据,如下7:1:2 的比例 (Zhou et al., 2022) 。回溯窗口L如下 (Zhou et al., 2023) ,并且预测长度O设置为{96,192,336,720}。

表 8:基准数据集的数据集详细信息。

数据集	长度协变量采样周期	
以太坊 17420	7	1小时
ETTM 69680	7	15分钟
天气 52696	22	10分钟
电 26304 17544	321	1小时
交通	862	1小时

特定领域的实验  
数据集和预测长度的每种组合,我们训练一个模型  
一个特定领域的训练数据  
并在同一领域的测试数据上测试模型。在少  
样本实验设置中,我们限制了数量  
  
训练数据分别减少到 5% 和 10%。

迈向基础模型  
每个预测长度,我们训练一个  
基于混合训练数据的模型

来自不同领域的数据,并在两个未见过的领域的的数据上测试模型。我们构造组合的  
通过汇集来自 ETTh1,ETTh2,ETTM1、天气和电力的训练数据来训练数据集  
并彻底洗牌。我们在组合训练数据集上训练每个模型。为防止不当行为  
偏差并确保组合训练数据中每个域的数据的公平表示,我们选择  
来自每个领域的训练数据的相同数量的训练示例。我们注意到这个数字  
与 ETTh1 和 ETTh2 相比,ETTh1 和 ETTh2 的训练样本数量级要小得多  
其他三个训练数据集 (ETTM1、天气、电力) ,因此选择最小数量  
所有五个训练数据集集中的训练样本将导致 ETTm1 丢失过多数据,  
天气和电力。因此,我们将 ETTh1 和 ETTh2 中的所有训练示例包含在  
组合训练数据集。对于 ETTm1、天气和电力数据,示例数量  
选择要合并到组合训练数据集集中的样本的最小数量  
这三个训练数据集集中的训练示例。随后,我们在  
ETTM2和Traffic的测试数据。

B.2提议的TETS数据集设置

预测目标我们实验的主要目标是预测之前的收益  
标准普尔 500 指数上市公司的利息、税收、折旧和摊销 (EBITDA),以及  
我们的数据范围从 2000 年到 2022 年。遵循 中提出的多元时间序列框架  
(Papadimitriou et al., 2020) ,我们从损益表中选择基本财务指标为  
输入特征:销售成本 (COGS)、销售、一般和管理费用 (SG&A)、研发费用  
费用 (RD EXP)、EBITDA 和收入。与其他指标相比,所选指标  
包含与我们的预测目标更相关的信息。对于基于大型语言的模型,  
包括我们的模型 TEMPO、GPT4TS 和 T5,我们应用通道独立策略来执行  
单变量时间序列预测任务。所有五个特征都用于训练 (预测其未来  
值基于其过去的值) ,而在训练阶段只能获取 EBITDA。其他  
模型遵循多元时间序列预测设置,将五个特征视为多元  
在训练和测试阶段输入并预测目标 EBITDA。

我们根据过去 20 个季度的数据预测季度 EBITDA。然后使用该预测值  
预测下一季度的 EBITDA,迭代四次,从而得出年度预测。为了  
根据年度累计值 (4 个季度的总和)衡量这些预测的准确性,  
我们采用对称平均绝对百分比误差 (SMAPE)作为评估指标,其中  
将在B.2 中进一步阐述。

数据拆分对于每个板块下的公司,我们采用窗口法生成  
连贯的培训和测试实例。在我们分开的渠道独立设置下  
每个特征以获得单变量时间序列,我们从七个域内扇区获得 80,600 个样本,  
以及来自四个零样本扇区 (也称为跨域扇区)的 9,199 个样本,共 5 个

正如我们进入通道相关设置一样。扇区分割在F中进行了详细说明。在表5所示的实验中,我们使用70%的域内数据进行训练,10%的域内数据用于评估, 20%的域内数据用于域内测试,以及用于未见过的测试的所有零样本数据。

评估指标实际上,不同公司的财务指标的大小可能存在很大差异。因此,我们选择对称平均绝对百分比误差 (SMAPE) (一种基于百分比的准确性度量)作为我们的评估指标:

$$SMAPE = \frac{200\%}{n} \sum_{t=1}^n \frac{|英尺-在|}{英尺+在}, \quad (5)$$

此外,对于 EBITDA 而言,有许多负面结果可能会影响最终的 SMAPE。我们介绍另一种形式的 SMAPE-Abs SMAPE:

$$绝对SMAPE = \frac{200\%}{n} \sum_{t=1}^n \frac{|英尺-在|}{|英尺| + |在|}, \quad (6)$$

这里,  $F_t$ 代表真实值,  $A_t$ 代表我们系统中的预测值,  $n$ 代表我们需要预测的总时间步长。

SMAPE 对异常值特别敏感。具体来说,当真实数据和预测符号相反时,产生的误差可能高达200%,严重扭曲最终结果。按照 (Papadimitriou et al., 2020)中的方法,我们过滤掉 80% 和 90% 阈值的数据点,发现大多数异常值与并购 (M&A) 导致的重大财务变化有关,如图所示表 6 的标题。一个显着的例子是 Facebook在 2014 年以 190 亿美元收购 WhatsApp,这对结果产生了重大影响。

### B.3选择提示设置。

不同的池设置,包括池大小、前k个数和提示长度,将导致不同的结果。为了探索这一点,我们总共进行了 27 次实验,为 3 种设置中的每一种设置 3 个不同的值:(1) 池大小为 10、20 和 30。(2) 前k 个数字为 1、2 和 3。(3) 提示长度为 1、2 和 3。我们选择 TEMPO 设置效果最佳的组合。对于长期和短期预测实验,我们选择池大小  $M = 30$ ,  $K=3$ ,提示长度为 3。

## C更多分析

### C.1基准数据集的消融研究

所提供的消融研究 (表9)提供了关于提示和分解组件对我们模型性能的影响的重要见解。在此表中,报告了三种情况下各种数据集的 MSE 和 MAE :原始模型配置 (“我们的”)、没有提示池的模型 (“w/o 提示”)和没有分解操作的模型 (“不分解”)。如果没有提示组件,所有数据集的 MSE 和 MAE 值都会增加,表明预测精度会下降。这表明提示有助于提高模型的预测性能。如果没有分解组件,性能下降甚至更加明显,MSE 和 MAE值进一步增加就证明了这一点。这意味着分解组件对于模型的性能也至关重要。

例如,在 “ETTh1”数据集上,由于缺少提示和缺少 STL 分解,MSE 分别上升了 18.8% 至 28.1%。请注意,模型仅应用提示池而不进行分解会对主干模型的性能产生不利影响,请参阅表1。这可能归因于从提示池中检索纯时间序列数据的挑战,因为可转移的数据可能有限信息无需分解。这些观察结果揭示了提示和分解组件对于模型预测准确性和预测能力的至关重要性。



表 9:长期预测的消融研究,最佳结果以粗体突出显示。不带/不带意思是**没有**。

方法 指标		我们的	无需提示,无需分解	
		MSE/MAE MSE/MAE MSE/MAE		
R	96	0.008/0.048	0.013/0.057	0.151/0.2
	192	0.027/0.082	0.037/0.097	0.197/0.245
	336	0.111/0.17	0.123/0.179	0.249/0.286
	720	0.251/0.282	0.255/0.283	0.315/0.322
	平均。	0.099/0.146	0.107/0.154	0.228/0.263
LIFT3	96	0.015/0.083	0.016/0.085	0.315/0.356
	192	0.118/0.207	0.116/0.206	0.361/0.392
	336	0.254/0.319	0.258/0.326	0.408/0.415
	720	0.381/0.4	0.393/0.414	0.461/0.452
	平均。	0.192/0.252	0.196/0.258	0.386/0.403
SPLIT3	96	0.01/0.066	0.014/0.076	0.177/0.264
	192	0.115/0.184	0.125/0.194	0.232/0.301
	336	0.214/0.283	0.222/0.286	0.298/0.348
	720	0.369/0.384	0.354/0.384	0.404/0.413
	平均。	0.177/0.229	0.179/0.235	0.278/0.331
LIFT3	96	0.201/0.268	0.302/0.348	0.4/0.419
	192	0.349/0.387	0.393/0.414	0.432/0.431
	336	0.408/0.425	0.471/0.463	0.45/0.441
	720	0.504/0.493	0.574/0.529	0.596/0.526
	平均。	0.366/0.393	0.435/0.439	0.469/0.454
SPLIT3	96	0.173/0.235	0.187/0.245	0.261/0.333
	192	0.315/0.355	0.334/0.356	0.353/0.39
	336	0.393/0.406	0.419/0.424	0.432/0.439
	720	0.425/0.449	0.463/0.47	0.428/0.448
	平均。	0.326/0.361	0.351/0.374	0.369/0.403

C.2对拟议的TETS数据集的消融研究

为了凭经验验证我们提出的 TEMPO 四个不同模型的有效性和预测能力,每个模型都排除 TEMPO 的 ,我们构建特定组成部分:(1)没有上下文信息(摘要提示)。(2)没有STL。(3)没有时间序列池。(4)没有任何提示(时间序列池和摘要提示)。如表10和表11所示,

TEMPO 优于所有变体,突出了每个组件对模型的贡献整体表现。具体来说,当将 TEMPO 与没有 stl 的版本进行比较时,所有部门的平均 SMAPE 误差减少 - 阈值为 0.8 时高达 56.4%,阈值为 0.8 时高达 57.1% 0.9 - 强调将 STL 纳入 TEMPO 的重要性。另外,排除时间序列池和摘要提示使平均 SMAPE 误差增加 11%/10.6% 和 17.9%/20.5% 分别证明时间序列池和摘要提示都发挥了重要作用 时间序列数据中不存在的附加信息。

预印本。

表 10:TETS 数据集消融研究的 SMAPE 结果。w/o 的意思是没有。

	不带文本的扇区 TEMPO		不带 STL 不带 TS 池 不带提示	
骨髓	12.38/10.82 12.84/12.84 24.05/24.05 12.67/12.67 11.66/11.66			
船级社	11.43/11.43 14.37/15.54 25.25/25.25 14.17/13.5 13.01/13.62			
烯	21.13/21.0 22.35/24.02 45.46/46.84 21.69/21.69 21.01/21.01			
满量程	9.12/9.31 10.13/10.33 22.06/22.42 9.82/9.82 10.17/10.17			
高	5.49/5.49 7.35/7.35 14.08/14.08		6.4/6.13	5.83/5.83
技术	10.91/11.19 13.54/13.21 26.71/26.84 12.32/12.29 12.44/12.40			
乌蒂	7.45/7.45 8.87/8.87 17.48/17.48 8.27/8.27			7.47/7.47
CC	10.21/10.67 12.52/12.76 23.88/24.65 11.53/11.82 11.98/11.8			
CD	8.25/8.24	9.55/9.8	18.82/18.8	8.92/8.92 8.98/8.98
工业	8.05/8.03 9.48/9.51 19.17/19.47 8.93/8.96			8.53/8.64
RC	10.09/10.16 11.78/11.95 23.16/23.55 11.32/11.13 10.42/10.49			

表 11:对所提出的 TETS 数据集进行消融研究的 Abs-SMAPE 结果。w/o 的意思是没有。

	不带文本的扇区 TEMPO		不带 STL 不带 TS 池 不带提示	
骨髓	14.08/15.41 13.77/13.77 24.88/24.88 14.76/14.76 13.53/13.53			
中海	16.18/16.18 18.23/19.37 28.17/28.17 17.53/18.09 16.50/17.10			
烯	23.08/24.64 26.31/27.94 47.32/48.68 27.72/27.72 25.57/25.57			
FS	9.84/10.04 10.8/11.0 22.06/22.42 10.35/10.35 11.02/11.02			
平均	8.86/8.86 11.28/11.28 19.18/19.18 9.83/10.07 10.01/10.01			
技术	12.72/13.01 15.92/16.20 26.71/27.49 13.33/13.90 14.26/14.82			
乌蒂	7.45/7.45 8.87/8.87 17.48/17.48 8.27/8.27			7.47/7.47
CC	14.76/15.35 16.33/16.84 26.14/26.89 15.66/16.37 15.83/16.35			
CD	9.53/9.79 10.54/10.79 19.29/19.52 10.18/10.18 10.11/10.11			
工业	10.43/10.89 11.86/12.18 20.82/21.50 11.06/11.46 10.92/11.42			
区域委员会	10.68/10.93 12.53/12.70 23.49/23.88 11.63/11.79 11.11/11.36			

C.3上下文信息注入设计分析

合并上下文信息的不同方法将导致不同的结果。我们的方法训练软提示映射来提取摘要信息。为了证明其有效性，我们用 5 种不同的方法替换 TEMPO 中的软摘要提示，以融入上下文对比分析的信息：(1)汇总池：与时间序列池类似，其设计获取摘要即可得到对应的摘要提示。(2)硬总结提示：这种方法直接将摘要的平均序列与时间序列数据连接起来，而不使用软提示映射。(3)硬提示：我们手动设计了一个提示，称为“预测未来”给定“{时间序列数据类型}”的时间步长，适用于 3 个不同的时间序列（趋势、季节、残差）STL分解后。(4)软提示：这利用了硬提示的软提示映射(3)中提到。(5)对齐 (Li et al., 2023)：该方法利用余弦相似度来对齐趋势时间序列数据的摘要。如表12和13所示，

我们的 TEMPO 型号具有软提示映射，在所有其他设计中取得了最佳效果获取上下文信息。特别是，TEMPO 将平均 SMAPE 误差降低了 4.5%/5.3%与所有部门的第二个最优结果进行比较。

预印本。

表 12:注入上下文信息设计的 SMAPE 结果。

	扇区	TEMPO	摘要池	硬摘要提示	硬提示	软提示	对齐	-	-	-	-	
▪	骨髓	12.38/10.82	11.76/12.45					12.27/12.27		14.14/14.14	12.51/11.76	13.25/13.25
	CS	11.43/11.43	13.35/13.91					11.45/11.45		13.67/15.49	11.46/11.46	14.77/14.05
	烯	21.13/21.00	20.36/20.36					21.89/21.89		21.55/21.55	20.12/22.57	22.14/22.98
	FS	9.12/9.31		9.88/9.88				9.68/9.68		10.5/10.5	9.71/9.71	10.40/10.40
	健康指数	5.49/5.49		6.69/6.69				6.37/6.37		7.48/7.48	5.88/5.88	6.12/6.37
	技术	10.91/11.19	12.19/11.20					11.22/11.52		11.85/11.20	12.11/12.08	12.24/11.88
	乌蒂	7.45/7.45		8.27/8.27				7.92/7.92		8.89/8.89	8.14/8.14	8.15/8.15
▪	碳钢	10.21/10.67	11.55/11.64					10.85/11.25		11.5/11.88	11.27/11.22	12.66/12.69
	CD	8.25/8.24		8.77/8.77				8.64/8.63		8.94/8.94	9.10/9.10	9.50/9.50
	工业	8.05/8.03		8.87/8.96				8.6/8.54		8.92/8.88	8.32/8.50	9.55/9.44
	RC	10.09/10.16	10.91/10.91					10.26/10.26		11.01/10.91	10.32/10.32	10.87/10.96

表 13:注入上下文信息设计的 Abs-SMAPE 结果。

	扇区	TEMPO	摘要池	硬摘要提示	硬提示	软提示	对齐	-	-	-	-	
▪	BM	14.08/15.41	14.55/15.23					14.8/14.8		14.69/14.69	13.07/13.71	13.67/13.67
	CS	16.18/16.18	17.09/17.64					16.53/16.53		17.24/19.02	15.69/15.69	18.37/18.97
	烯	23.08/24.64		26.1/26.1				24.37/24.37		26.23/26.23	24.58/26.96	24.67/25.5
	FS	9.84/10.04		10.47/10.47				10.23/10.23		11.1/11.1	10.24/10.24	11.49/11.49
	健康指数	8.86/8.86		10.39/10.39				9.85/9.85		10.93/10.93	9.84/9.84	10.2/10.44
	技术	12.72/13.01		14.15/15.0				13.31/13.61		13.94/14.5	13.64/14.2	13.66/14.48
	乌蒂	7.45/7.45		8.27/8.27				7.92/7.92		8.89/8.89	8.14/8.14	8.15/8.15
▪		14.76/15.35	15.67/16.32					15.34/16.00		15.69/16.48	15.49/16.13	16.89/17.34
	CD	9.53/9.79		10.32/10.32				9.63/9.89		10.25/10.25	10.23/10.23	10.28/10.28
	工业	10.43/10.89	11.23/11.6					10.57/10.97		10.76/11.48	10.46/10.83	11.75/12.2
	参考文献	10.68/10.93	11.89/11.89					10.96/10.96		11.78/11.87	10.92/10.92	11.66/11.75

C.4即时池分析

C.4.1即时池案例研究

如图9 所示,在我们的案例研究中,我们首先分解三个时间序列实例: x1、 x2和 x3 ,具有来自 ETTm2 数据集的不同输入分布到其趋势、季节性和残差成分。分解后, x1和x2的趋势分量显示出惊人的相似性,并且 x2和x3的季节性分量也相似。因此,当输入这些组件时,在我们的模型中, x1和x2检索相同的提示 ID (1, 10, 28),通常对应于趋势分量,而x2和x3检索相同的提示 ID (3, 16, 28),通常关联与季节性成分。这一发现验证了该模型成功识别和利用潜在趋势和季节性水平上的代表性相似性,即使完整的实例有所不同 这与在不断变化的模式中巩固知识的目标是一致的。至关重要,这种分解过程使不同的组件能够处理不同的语义对于语言模型,简化任务复杂性。这个案例展示了所提议的基于分解的提示调整能够发现并应用共享的结构信息相关时间序列实例之间的关系,同时还通过以下方式简化预测问题

预印本。

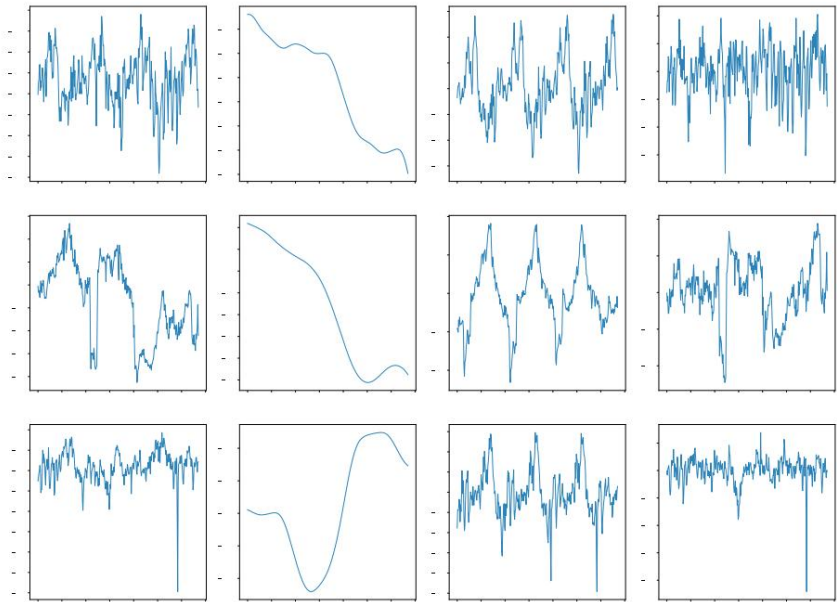


图 9:对 ETTm2 数据集中三个不同实例的案例研究,其中x1和x2 (顶部两个)具有相似的趋势模式,而x2和x3 (底部两个)具有相似的季节性模式。

成分分离。这些见解表明我们的方法有可能提供强大的时间系列建模。

此外,这些发现在不同数据集 (ETTh1、ETTh2、ETTm1、ETTm2)中是一致的,如图10、11、12和13所示,这些图描述了如何有选择地选择提示基于时间序列的基本趋势和季节性组成部分。分开的选择不同组件的提示使得仅在相似的实例之间共享知识通过提示组件。这不是分享整个时间序列的知识,而是使共享专业知识以改善知识转移,同时减少可能的冲突实例之间。

C.4.2留一分析

表 14:屏蔽提示选择

方法 指标	TEMPO 面膜所有提示 面膜前 3 名 趋势提示 面膜前 3 季提示 面膜前 3 名 剩余提示				
	MSE/MAE	MSE/MAE	硕士/硕士	硕士/硕士	硕士/硕士
ETT	96 0.015/0.083	0.618/0.535	0.119/0.19	0.049/0.122	0.053/0.13
	192 0.118/0.207	0.194/0.279	0.122/0.211	0.123/0.213	0.119/0.21
	336 0.254/0.319	0.374/0.406	0.295/0.353	0.273/0.335	0.284/0.345
	720 0.381/0.4	0.757/0.595	0.464/0.451	0.455/0.447	0.433/0.433
	平均 0.192/0.252	0.486/0.454	0.25/0.301	0.225/0.279	0.222/0.279
ETT	96 0.01/0.066	0.388/0.387	0.199/0.206	0.074/0.131	0.071/0.129
	192 0.115/0.184	0.17/0.255	0.121/0.202	0.118/0.194	0.118/0.197
	336 0.214/0.283	0.295/0.34	0.227/0.294	0.216/0.284	0.222/0.288
	720 0.369/0.384	1.244/0.74	0.765/0.638	0.398/0.398	0.493/0.445
	平均 0.177/0.229	0.524/0.431	0.328/0.335	0.202/0.252	0.226/0.265

我们遵循可解释性的留一法 (LOO)原则来评估提示的影响水池。如表14所示,我们将完整模型与留一法下的几种设置进行了比较。对于特征扰动,我们使用屏蔽策略,通过分配零来屏蔽提示

图 10:根据不同实例的分解组件之间的相似性进行提示选择。ETTh1 上的示例。

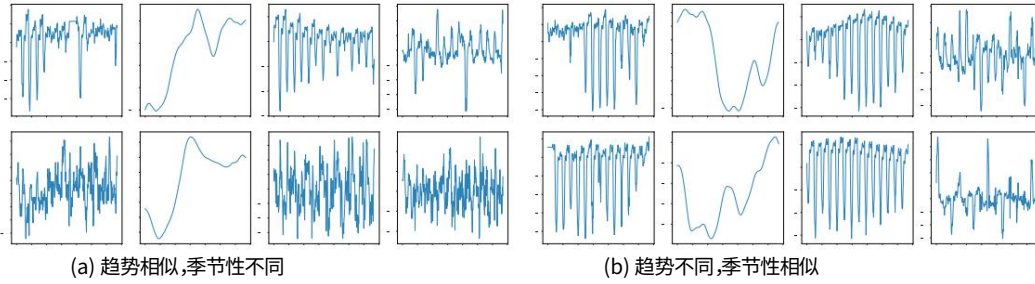
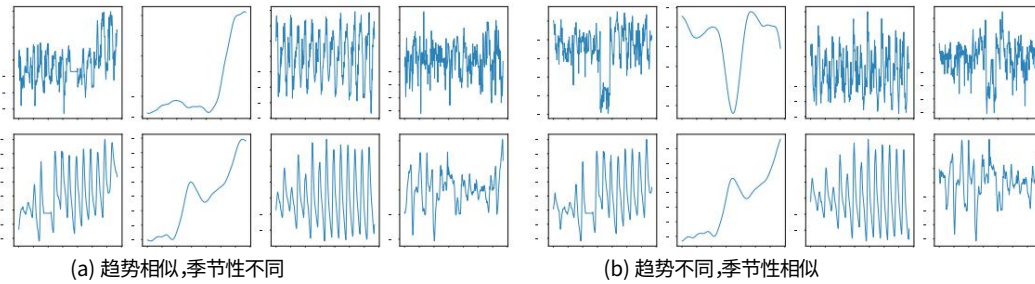


图 11:根据不同实例的分解组件之间的相似性进行提示选择。ETTh2 上的示例。



测试期间提示池中所有提示值的向量:  $V_i = 0 \in \text{RLp} \times \text{LD} : \forall i \in [M]$ 。  
在这里,我们考虑四种屏蔽设置:

- 屏蔽提示池中的所有提示
- 屏蔽趋势组件的前 3 个选定提示
- 屏蔽季节性成分的前 3 个选定提示
- 屏蔽掉前 3 个选定的残留成分提示

屏蔽所有提示后,我们可以观察到ETTh1 数据集下MSE 和 MAE 的误差增加了153.12%和79.95% ,  
ETTh2 数据集下MSE 和 MAE 的误差增加了195.56%和87.86%。这提供了对重要性的见解。模型中的提示池。

屏蔽提示后性能的大幅下降强调了它们在提高模型预测准确性方面的关键作用。掩盖趋势、季节性和剩余成分的前 3 个选定提示也会损害模型的性能。性能下降的不同程度表明趋势成分的重要性显著高于季节性和残留成分,这也与我们对分解成分的 SHAP 分析一致。

### C.5隐藏表示

图14演示了预训练语言模型的输出隐藏空间表示之间的差异。虽然从 GPT4TS 学习到的时间序列的表示是整体居中的,但从 TEMPO 中分解的分量的表示意味着三个分量之间存在一定的软边界。这演示了 TEMPO 如何分别学习趋势、季节性、剩余部分的表示,这有助于我们的模型 TEMPO 的卓越性能。

预印本。

图 12:根据不同实例的分解组件之间的相似性进行提示选择。 ETTm1 示例。

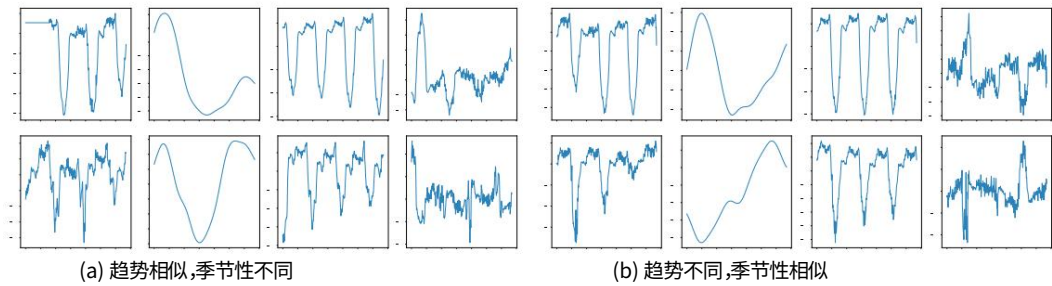
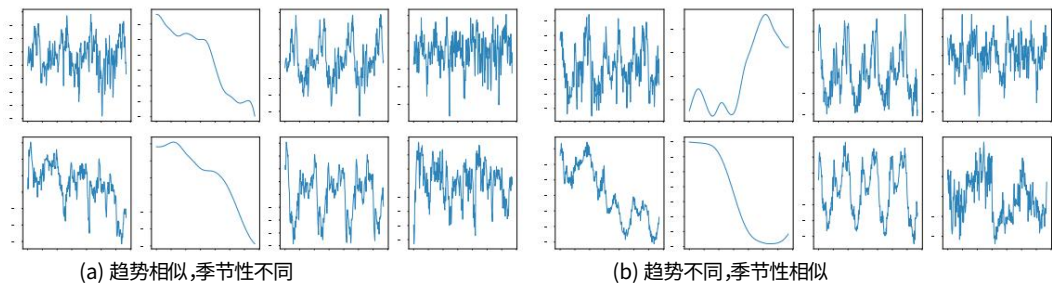


图 13:根据不同实例的分解组件之间的相似性进行提示选择。 ETTm2 示例。



C.6模型结构讨论

如第 3.4 节所述,我们的 GPT 块有两种方式利用趋势、季节性和残差信息。一种有效的方法将这三个元素连接到单个 GPT 块的单个输入中。或者,这些信息可以通过三个单独的 GPT 块单独处理,其参数可以与 MSE 损失函数同时更新。总的来说,第二种具有多个 GPT 块的方式可以有更准确的预测性能。然而,考虑到数据效率和训练时间,表1 中记录的我们的结果是基于我们观察到的相当有效和高效的策略,其中涉及连接信息。天气数据是个例外,我们发现所有预测长度上分离GPT 块更加准确,其中使用{96, 192, 336, 720}中单个 GPT 块的天气数据集的 MSE/MAE为 0.011/0.049; 0.038/0.099; 0.124/0.184; 分别为0.255/0.286 ,平均 MSE/MAE (0.107/0.154) 比使用三个单独的GPT 块低 8.1%/6.0%。

C.7模型训练时间比较

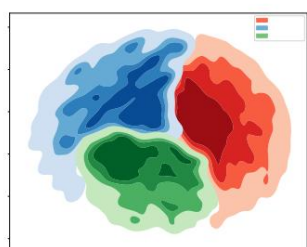
图15显示了其他基线模型与我们的模型 TEMPO 的训练时间对比。每个模型的训练时间以相对于 TEMPO 训练时间的比率表示。值小于1 表示模型训练速度快于 TEMPO,而值大于 1 则表示相反。我们使用水平条直观地表示每个模型的相对训练时间,条形分别根据它们比我们的模型 TEMPO 快还是慢而延伸到中心垂直线的左侧或右侧。

D基线模型解释

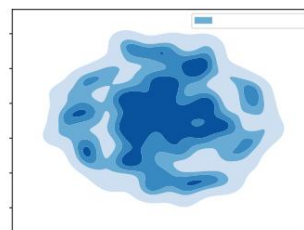
我们在下面的实验中展示了我们比较的 14 个基线模型：



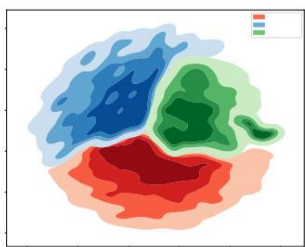
图 14:使用 TSNE 预测长度 $O = 96$  时 GPT4TS 表示与 TEMPO 表示的比较。红色为趋势,蓝色为季节性,绿色为残差



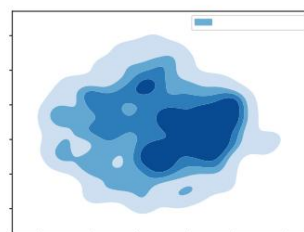
(a) TEMPO-ETTh1



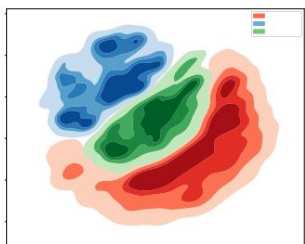
(b) GPT4TS-ETTh1



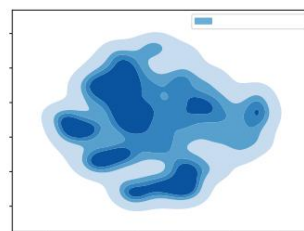
(c) TEMPO-ETTh2



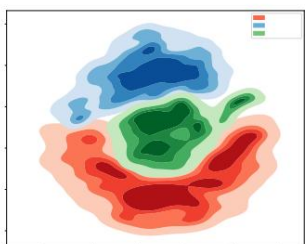
(d) GPT4TS-ETTh2



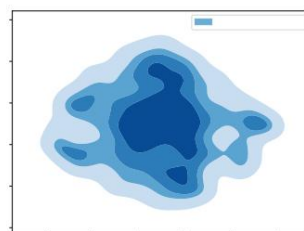
(e) TEMPO-ETTm1



(f) GPT4TS-ETTm1



(g) TEMPO-ETTm2



(h) GPT4TS-ETTm2

Response Category	2006 (%)	2008 (%)
U.S. should take action to reduce greenhouse gas emissions	~10	~85
U.S. should take action to reduce greenhouse gas emissions	~10	~95
U.S. should take action to reduce greenhouse gas emissions	~10	~10
U.S. should take action to reduce greenhouse gas emissions	~10	~10
U.S. should take action to reduce greenhouse gas emissions	~10	~10

28

·**非固定变压器**(Liu et al., 2022):非平稳变换器通过将数据标准化的系列平稳化与去平稳注意力相结合来增强时间序列预测,以重新引入固有的时间变化并对抗过度平稳化。

·**时代网** (Wu 等人, 2023): TimesNet 将 1D 时间序列转换为 2D 张量,捕获周期内和周期期间的变化,并使用带有初始块的 TimesBlock 来提取复杂的时间模式,在多个时间序列任务中表现出色。

· **GPT-2** (Radford et al., 2019): GPT-2 是 OpenAI 开发的基于解码器的语言模型,旨在根据给定的提示生成连贯且多样化的文本内容。在我们的工作中,我们使用 6 层的 GPT-2 作为主干,它改编自 GPT4TS (Zhou et al., 2023)。

·**伯特** (Devlin 等, 2019): BERT (Trans-formers 的双向编码器表示)是一种基于编码器的深度学习模型,利用 Google 设计的 Transformer 架构,通过双向分析文本来理解句子中单词的上下文。在我们的工作中,我们使用前 6 层的 Bert 作为基线。· **T5** (Raffel et al., 2020): T5 (文本到文本传输转换器)是一种最先进的神经网络模型,具有由 Google 设计的基于编码器-解码器的架构,可将每个语言问题转换为文本到文本的问题-文本格式。在我们的工作中,我们使用前 6 层的 T5 作为基线。·**美洲驼** (Touvron 等人, 2023): LLaMA (大型语言模型元 AI)是最先进的基础语言模型的集合,参数范围从 7B 到 65B,可提供卓越的性能,同时显著降低所需的计算能力和资源。在我们的工作中,我们使用 7B LLaMA 的前 6 层。

## 理论分析

### E.1 定理 3.1 的证明

定理 E.1 假设我们有时间序列信号  $Y(t) = S(t) + T(t) + R(t)$ ,  $t \in [t_1, t_n]$ , 其中  $S(t)$  是季节性信号 (周期信号),  $T(t)$  是趋势信号 (非周期),  $R(t)$  是残差信号。设  $E = \{e_1, e_2, \dots, e_n\}$  表示一组正交基。设  $E_S \subseteq E$  表示  $E$  的子集, 其中  $S(t)$  具有非零特征值,  $E_T \subseteq E$  表示  $E$  的子集, 其中  $T(t)$  具有非零特征值。如果  $S(t)$  和  $T(t)$  不正交, 即

$$\sum_{i=1}^n S(t_i)T(t_i) \neq 0, \text{ 则}$$

$E_T \cap E_S \neq \emptyset$ , 即  $E$  无法将两个信号分解到两个不相交的碱基集上。

证明 1 我们将  $S(t)$  和  $T(t)$  分解到  $E$  上, 得到  $S(t) = \sum_{i \in E_S} a_i e_i$  和  $T(t) = \sum_{i \in E_T} b_i e_i$ 。

那么显然  $e_i \in E_S \Leftrightarrow a_i \neq 0$  且  $e_i \in E_T \Leftrightarrow b_i \neq 0$ 。现在, 让我们考虑  $S(t)$  和  $T(t)$  的内积:

$$\sum_{i=1}^n S(t_i)T(t_i) = S(t) \cdot T(t) = \left( \sum_{i \in E_S} a_i e_i \right) \cdot \left( \sum_{j \in E_T} b_j e_j \right) = \sum_{i,j} a_i b_j e_i e_j \quad (7)$$

请注意, PCA 找到的分量是一组正交基。因此, 对于任何  $i \neq j$ , 我们有  $e_i e_j = 0$ 。因此, 我们有:

$$\sum_{i=1}^n S(t_i)T(t_i) = S(t) \cdot T(t) = \left( \sum_{i \in E_S} a_i e_i \right) \cdot \left( \sum_{j \in E_T} b_j e_j \right) = \sum_{i \in E_S} a_i b_i \|e_i\|^2 \quad (8)$$

$= 0$ 。因此,  $\sum_{i=1}^n S(t_i)T(t_i) = 0$ 。因此, 必须至少有一个  $i$  使得  $a_i \neq 0$  且  $b_i \neq 0$ 。因此, 请注意, 换句话说,  $E_T \cap E_S \neq \emptyset$  且  $E_T \cap E_S \neq \emptyset$ 。

上述定理证明, 如果  $T(t)$  和  $S(t)$  不正交, 则不存在将  $S(t)$  和  $T(t)$  分解为两个不相交的基集的正交基集。请注意, 周期信号与非周期信号不正交是很常见的。这是因为周期信号的频谱是离散的, 而非周期信号的频谱是连续的。

因此, 周期性信号的那些非零频率很可能存在重叠。

请注意,PCA 还旨在学习一组数据的正交基。我们可以很快得出一个推论,即 PCA 无法将两个信号分解为两个不相交的碱基集。基于 (Zhou et al., 2023) 的定理 1, 我们可以揭示预训练大型模型中的自注意力学习执行与 PCA 密切相关的功能。因此,自注意力机制无法自动将时间序列分解为其趋势和季节分量,除非我们手动执行此操作。

## E.2 从频域解释模型预测

除了第5.1节给出了分解为何有助于预测结果的实验视角之外,我们还提供了谱域的理论分析。具体来说,时间序列信号可以表示为谱域中不同频率的组合。

预测具有挑战性,因为现实世界的序列包含具有重叠周期性的复杂变化混合物。然而,通过将我们的观点转移到频域,我们可以通过 STL 分解来识别不同的分量,其中包含从频谱的其余部分中明显突出的孤立频率。主要周期性模式的这种分离至关重要,因为预测未来值相当于预测这些潜在频率如何随时间演变:

命题E.2 (时域预测和频域预测的等价性)

假设  $x_0, x_1, \dots, x_{N-1}$  和  $x^0, x^1, \dots, x^{N-1}$  是频率模型的输入和输出序列。那么,从频域转来的  $x^N$  就是时间戳  $N$  处的预测值。

给定输入序列  $\{x_t | t = 0, 1, \dots, N-1\}$ , 其中  $N$  是离散时间戳的数量,在时域中,离散傅里叶变换(DFT,  $F$ )和逆离散傅里叶变换(IDFT,  $f$ )获得频域的运算可以定义为:

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{-\frac{i2\pi ux}{N}}, \quad u = 0, 1, \dots, N-1, \quad (9)$$

$$f(x) = \sum_{u=0}^{N-1} F(u) e^{\frac{i2\pi ux}{N}}, \quad x = 0, 1, \dots, N-1. \quad (10)$$

根据命题E.2, 假设  $F(u)$  的下一个值, 可以预测为  $F(N)$ , 时域和频域中的其他未知变量, 包括第  $(N+1)$  个离散样本  $f(N)$  和新的 DFT 的结果  $F(N+1)$  由给定的  $F$  决定

$$F'(u), u = 0, 1, 2, \dots, \quad F'(N).$$

证明2设

$$A = \sum_{x=0}^{N-1} f(x) \left( \frac{e^{-\frac{i2\pi ux}{N}}}{N} - \frac{e^{-\frac{i2\pi ux}{N+1}}}{N+1} \right), \quad (11)$$

$$B = \frac{1}{N+1} \sum_{x=0}^{N-1} f(x) e^{-\frac{i2\pi Nx}{N+1}}, \quad (12)$$

那么我们有:

$$f(N) = (N+1) \left( F'(N) - B \right) e^{-\frac{i2\pi N^2}{N+1}}, \quad (13)$$

$$F'(u) = A + (F'(N) - B) e^{\frac{i2\pi(N-u)N}{N+1}}. \quad (14)$$

对于  $u = 0, 1, 2, \dots, N-1$ ,  $F$  的值  $F'(u) - F(u)$  可以表示为:

$$F'(u) - F(u) = A + f(N) e^{-\frac{i2\pi uN}{N+1}}. \quad (15)$$

当  $u = N$  时,  $F$  的值  $F'(N)$  可表示为

$$F'(N) = B + f(N) e^{-\frac{i2\pi N^2}{N+1}} \quad (16)$$

预印本。

给定  $F'(N)$ , 我们可以推论  $F'(u)$  通过:

$$F'(u) = A + (F'(N) - B)e^{\frac{j2\pi(N-u)N}{N+1}}, u = 0, 1, 2, \dots, N - 1. \tag{17}$$

和  $f(N)$ :

$$f(N) = (N + 1)(F'(N) - B)e^{-\frac{j2\pi N^2}{N+1}}, \tag{18}$$

因此,唯一需要预测的变量是  $F$

这个命题表明,如果很容易预测频域中的模式,我们就可以更容易地预测时间序列的未来值。预测相当于预测构成时间序列信号的基础频率的演变。STL 分解通过分离具有不同主导周期模式的组件来显着帮助完成此任务。使用 STL,每个组件呈现的需要解开的相互交织的周期性影响要少得多,这显着简化了预测问题。例如,趋势成分可能会表现出明显主导其频谱的单独年度周期。仅专注于准确估计此孤立频率的进展的有针对性的预测模型可以生成准确的预测。同样,季节性元素巧妙地隔离了每天或每周重复出现的频率。专门为这些已知周期性定制模型可以进行高度可预测的外推。相比之下,直接对具有大量混合周期分量的原始数据的压缩频谱进行建模会产生不令人满意的近似值。重叠的频率很难独立区分和预测。

通过频域透视镜概念化预测揭示了 STL 如何将复杂的频谱混合分解为可区分的基于频率的子问题。这允许实施优化的预测策略来发现每个组件中的模式,从而显着改进时间序列预测。从本质上讲,STL 通过将频谱内容分解为更简单的可预测形式来促进准确的未来预测。

F TETS数据集的详细信息

时间序列数据分析和预测公司未来的盈利能力和生存能力对于其发展和投资策略至关重要。财务评估和预测是数据驱动的,主要依赖于公司报告等多种数据类型的组合。

在这个项目中,我们的主要来源是公司的财务报表:资产负债表、损益表和现金流量表。

标准普尔 500 指数 (S&P 500) 代表衡量美国最大 500 家公司股票表现的股票市场指数。我们的数据集中包含 S&P500 中的 11 个行业:基础材料 (21 家公司)、通信服务 (26 家公司)、能源 (22 家公司)、金融服务 (69 家公司)、医疗保健 (65 家公司)、科技 (71 家公司)、公用事业 (30 家公司)、周期性消费品 (58 家公司)、消费者防御性 (36 家公司)、工业 (73 家公司)、房地产 (32 家公司)。在数据集划分方面,我们将数据集中的扇区分开,以实现域内任务设置和零样本任务设置。

前七个扇区被视为训练和评估扇区,而后四个扇区被保留为零样本预测任务的未见扇区。

为了解决标准普尔 500 指数中缺乏 2010 年之前数据的公司的数字信息缺失问题,我们在尝试了各种方法后应用了线性插值。线性插值是一种使用两个已知端点值估计某个范围内的值的技术。对于研发费用缺失值,我们采取了零填充策略。这是因为这些报表中的空条目通常表明该公司没有在该领域进行任何投资。

上下文数据收集自然语言处理领域大规模预训练模型 (LLM)的兴起为其在时间序列分析中的应用提供了新的可能性。事实证明,法学硕士对于分析和学习复杂的关系以及跨不同时间序列进行推断非常有用。然而,大多数现有方法主要将时间序列数据转换为直接输入到法学硕士中,忽略了法学硕士是专门针对自然语言进行预训练的事实,因此忽略了上下文数据的合并。

此外,时间序列数据中包含的信息是有限的,尤其是在金融领域。金融领域的时间序列数据,如公司报表,主要反映财务数字

预印本。

变化基于公司的历史战略和更广泛的宏观经济变化。这些数据包含公司的内部历史信息。然而,更广阔的市场环境,即外部信息,对公司未来的发展也起着重要作用。

例如,医药和医疗保健公司在COVID-19 爆发之前经历了稳定增长。但2019年至2020年间,疫情爆发后,此类企业的财务报表受到了较大影响。因此,我们认识到将新闻和报告整合为外部数据源以补充时间序列数据中包含的内部信息的价值。外部数据包含的信息主要包括3部分:(i).跨地区的政策转变(ii).全球发生的重大事件 (iii).公众对公司产品的反应。

这些元素共同提供了时间序列数据 (内部数据)中缺失的补充信息,从而增强了我们的预测能力。

从不同来源提取新闻和报告等上下文数据是一项重大挑战。在当今的数字时代,众多新闻网站和应用程序提供广泛的世界新闻,从影响整个行业的有影响力的新闻到琐碎的小报道。因此,过滤和总结信息、区分关键新闻和次要新闻至关重要。幸运的是,Open AI 最近发布的 ChatGPT API<sup>1</sup>提供了在指定时间内收集和总结新闻和报告的功能。

通过将所有相关详细信息 (查询、季度、年度背景、公司信息和具体要求)合并到用户消息中,并设置响应的令牌上限为 110 个令牌,我们可以从 ChatGPT API 高效地获取所需的上下文信息。为了便于说明,图16显示了 A 公司的示例,展示了ChatGPT 3.5 中设计的提示和相应的响应。如果无法生成上下文信息,API 通常会返回带有“不幸”和“抱歉”等关键字的消息。我们检测并用术语“无”替换它们,代表中立的上下文信息。此外,图17和19提供了我们的数据集的说明,包括时间序列数据和相应的上下文文本。上下文文本的详细视图如图18和20所示

**提示:**

假设你生活在(年份:2000年),你能帮我总结(年份:2000年)的(季度:第三季度)的(公司名称:A公司)的:新闻和报道吗? (公司:科技公司) 请直接给我2句话以内的答案,无需道歉。which is an {company sector: Technology} company. Please directly give me the answer limited to 2 sentences without apology.

**GPT 响应:**

第三季度: A公司报告净利润为0.33亿美元,高于上年同期的0.23亿美元,这得益于其X计算机和Product Y的强劲销售。quarter of the previous year, driven by strong sales of its X computers and Products Y. 然而,在警告称由于销售放缓,第三季度利润将低于预期后,该公司股价下跌。d-quarter profits would be below expectations due to slower sales.

图 16:使用 OPENAI ChatGPT-3.5 API 设计提示的示例。

<sup>1</sup><https://platform.openai.com/docs/guides/gpt>



预印本。

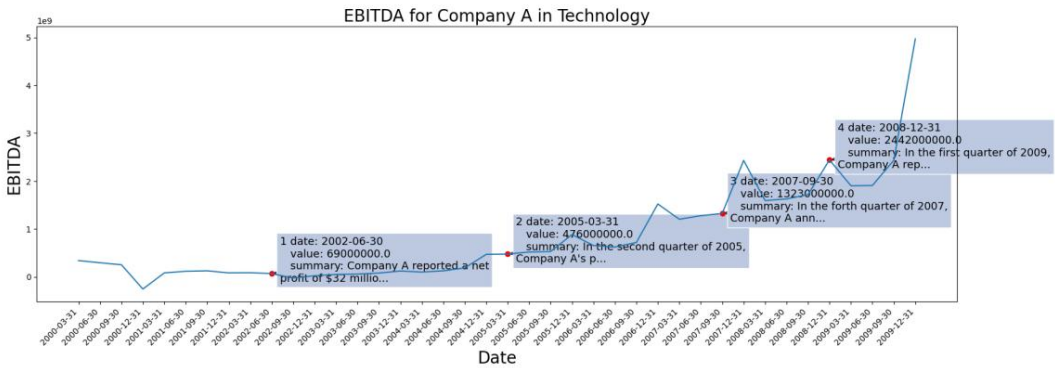


图 17:A 公司的 EBITDA 及背景信息

1

2002年第三季度ter

2

2005年第二季度arter

3

2007年第四季度arter

4

2008年第一季度er

图 18:为17中标记的公司 A 生成的上下文信息示例

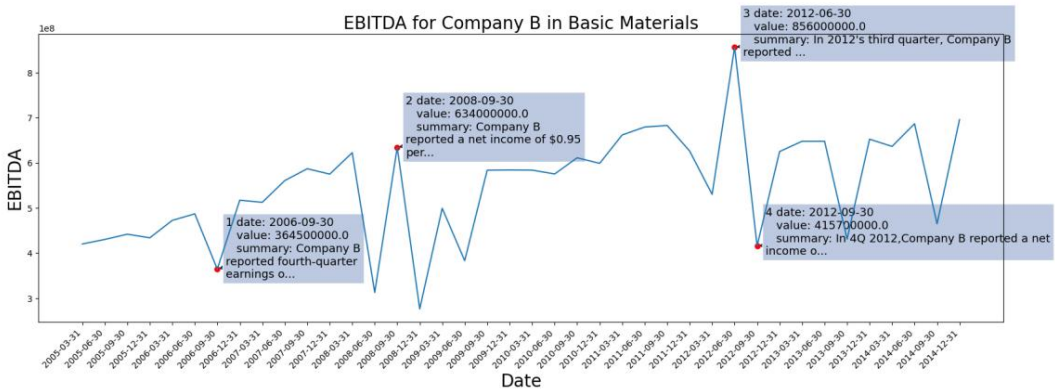


图 19:B 公司的 EBITDA 及相关信息

预印本。

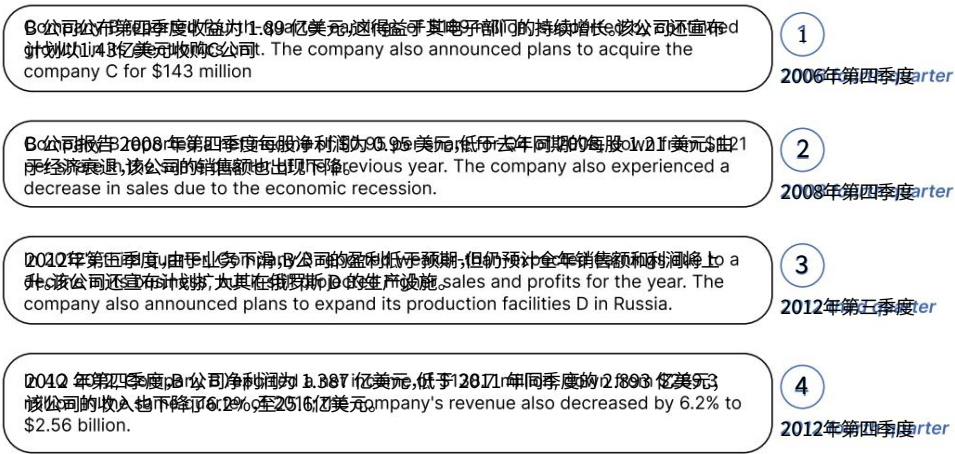


图 20:为19中标记的公司 B 生成的上下文信息示例

表 15:EBITDA 的 Abs-SMAPE 结果与基线和我们提出的域内方法数据集。我们在Papadimitriou中过滤掉 80%/90% 阈值的异常 SMAPE 值等人。(2020).(BM:基础材料;CS:通信服务;Ene:能源;FS:金融服务;Hea:医疗保健;技术:技术; Uti:实用程序。)

部门	BM	CS	埃内	FS	希亚	技术	乌蒂
我们的	14.08/15.41 16.18/16.18 23.08/24.64 9.84/10.04 8.86/8.86 12.72/13.01 7.45/7.45						
GPT4TS 24.26/24.95 27.09/28.24 44.37/47.8 22.55/22.73 19.68/19.88 27.12/27.12 17.98/17.98							
伯特	27.35/27.35 28.2/28.78 44.14/45.63 23.01/23.19 20.02/20.02 26.84/27.59 18.27/18.27						
T5	29.93/31.17 33.6/33.6 48.53/52.58 26.53/26.71 26.38/26.8 32.28/32.52 23.64/24.03						
自动成型机 36.94/37.5 34.6/34.6 36.87/36.87 25.25/25.25 22.39/22.39 36.97/37.45 17.88/17.88							
举报人 33.73/35.28 38.66/38.66 31.56/31.56 27.7/27.7 16.87/16.87 26.75/26.91 16.54/16.54							
补丁TST 32.35/32.91 18.72/18.72 26.98/27.57 16.63/16.63 13.22/13.22 19.86/20.05 15.43/15.43							
塑身机 31.61/32.17 20.91/21.43 26.23/26.23 15.9/15.9 11.24/11.24 18.87/19.04 17.07/17.44							
光TS 29.62/30.71 19.12/19.12 21.41/22.6 16.08/16.08 13.96/13.96 21.65/22.35 13.07/13.07							
线性 28.48/29.04 19.82/20.86 23.16/24.96 16.89/16.89 16.0/16.0 25.38/25.72 13.89/13.89							
时代网 29.12/29.68 18.74/18.74 22.11/22.11 14.39/14.74 11.4/11.6 18.8/19.34 14.19/14.19							
美洲鸵 25.22/25.88 29.9/30.41 45.57/48.47 22.98/23.15 19.96/20.4 29.11/29.6 17.84/17.84							
FED前 49.97/53.19 63.73/64.6 66.57/69.4 57.53/58.49 55.87/57.17 72.07/73.88 35.37/36.07							
东部时间 29.12/29.7 39.87/39.87 39.1/39.1 24.36/24.36 22.66/22.66 27.04/27.21 15.78/15.78							
非线性 27.99/29.14 24.78/26.29 28.92/29.47 20.19/20.19 19.27/19.27 30.69/30.84 14.1/14.47							

预印本。

表 16:EBITDA 的 Abs-SMAPE 结果与基线和我们提出的跨域方法数据集。我们过滤掉 80%/90% 阈值的异常 SMAPE 值。（抄送 :周期性消费者; CD:消费者防御 ;行业 :工业 ;回复 :房地产。）

部门	抄送	光盘	工业	关于
我们的	14.76/15.35 9.53/9.79	10.43/10.89	10.68/10.93	
GPT4TS 26.16/27.08	19.75/19.75	21.1/21.65	24.71/24.94	
伯特	27.17/28.04 19.99/19.99	21.34/22.2	24.14/24.75	
T5	34.3/35.19 22.99/23.1	25.78/26.32	31.38/31.67	
自动成型机 21.29/21.53	44.47/46.95	19.73/20.28	45.81/47.8	
举报人 39.06/39.27	62.78/71.58	86.66/37.0	44.8/47.69	
补丁TST 15.29/15.49	34.99/36.49	14.61/15.16	44.75/46.1	
塑身机 15.79/16.0	36.93/38.32	15.08/15.6	46.33/48.18	
光TS 16.84/16.93	41.54/44.02	14.84/15.3	32.14/32.9	
线性 14.82/14.95	31.94/33.56	12.21/12.61	34.04/35.0	
时代网12.65/12.82	32.09/33.12	12.01/12.47	35.64/36.6	
美洲驼 27.81/29.33	20.73/20.97	21.88/22.73	24.61/24.84	
FED前 50.7/51.9	70.47/78.91	47.55/48.46	51.02/55.64	
ESTformer 14.62/14.79	47.41/51.55	18.27/18.78	36.22/37.57	
非线性 16.19/16.31	40.89/43.59	14.29/14.74	32.64/33.47	

表 17:TEMPO 主要符号表

符号	描述
$X_t^i$	我时间步 t 的通道预测
$X_t^h$	我时间步 t 的通道回溯窗口/历史值
$\Phi$	型号参数
$V$	提示池中的提示值
$X$	输入数据可分解为XT XS XR
$X_t, X_{St}, X_{Rt}$	趋势、季节、时间 t 内设置的剩余分量
$X_{时间}^{th}$	我通道 x 的时间步长 $x_{时间}^{th}$
$X_{时间}^R$	趋势分量的预测值
磷	输入数据补丁
公里	提示池中的第 m 个键
电压	提示池中的第 m 个值
电压	提示池
K	超参数,提示数量选择
中号	超参数,提示池长度
$Z^*$	* 的 GPT 输出 (趋势、季节性、残差)
具体要素	预测长度
LE	嵌入向量长度
$Y^*$	反归一化前的最终预测值
$Y^{**}$	最终预测值