

基于元模式挖掘与提示引导的时空预测方法

庞双双^{1),3)} 邓 攀^{1),3)} 赵 宇^{2),3)} 刘俊廷^{1),3)} 王子昂^{1),3)} 李思睿^{1),3)}

¹⁾(北京航空航天大学软件学院 北京 100191)

²⁾(北京航空航天大学计算机学院 北京 100191)

³⁾(北京航空航天大学复杂关键软件环境全国重点实验室 北京 100191)

摘要 时空预测在城市规划、环境监测等众多现实场景中有着至关重要的作用。现有时空预测方法过于关注全局平均特性,忽略了时空数据中固有的异质性,难以有效捕捉不同空间区域或时间段内时空模式的局部差异和动态变化,从而导致预测性能不佳。为此,本文提出了一种新颖的基于元模式挖掘与提示引导的时空预测方法 ST-MP2G。本文方法基于简单的多层感知机作为预测网络,通过生成时空提示捕获时空异质性,结合自监督学习挖掘历史数据中的元模式以增强时空提示,引导时空建模过程,从而克服时空异质性的干扰,实现准确而鲁棒的预测。具体而言,首先构建了一个时空提示生成模块,利用时间先验嵌入和记忆网络生成多尺度的时间提示以捕获时间异质性,利用可学习的空间嵌入矩阵和表征学习生成语义化的空间提示以捕获空间异质性,通过构建时空交互张量建模复杂的时空交互作用,并利用反向张量分解方法实现时空提示的深度融合;接着构建了一个自监督元模式挖掘模块,通过自监督学习在特征空间挖掘历史数据中有判别能力的元模式,将其参数冻结并存储为元模式库,用于后续检索;最后构建了一个时空提示引导模块,通过检索元模式库得到与输入时空序列最相关的元模式,以补充和增强时空提示,通过逐层适配器将时空提示映射到每一层的特征空间,利用哈达玛积实现时空表征的细粒度增强,从而引导时空建模过程以克服异质性的干扰。在六个真实世界数据集上进行的大量实验表明,本文所提时空预测方法能够有效克服时空异质性的干扰,显著增强模型的准确性和鲁棒性,使其在各个时空场景数据集上的预测误差平均降低了 4.53%,进一步的综合研究证明了本文方法具备良好的可解释性、效率、泛化性和鲁棒性。

关键词 时空预测;异质性;深度学习;提示学习;自监督学习

中图法分类号 TP18

DOI号 10.11897/SP.J.1016.2025.02561

Spatio-Temporal Prediction Method Based on Meta-Pattern Mining and Prompt Guidance

PANG Shuang-Shuang^{1),3)} DENG Pan^{1),3)} ZHAO Yu^{2),3)} LIU Jun-Ting^{1),3)}
WANG Zi-Ang^{1),3)} LI Si-Rui^{1),3)}

¹⁾(School of Software, Beihang University, Beijing 100191)

²⁾(School of Computer Science and Engineering, Beihang University, Beijing 100191)

³⁾(State Key Laboratory of Complex Critical Software Environment, Beihang University, Beijing 100191)

Abstract Spatio-temporal prediction plays a critically indispensable role in numerous real-world scenarios, such as traffic forecasting, urban planning, environmental monitoring, and disaster prevention; however, a significant limitation persists in existing spatio-temporal prediction methodologies. These approaches tend to focus excessively on global average characteristics while systematically neglecting the inherent heterogeneity embedded within spatio-temporal data struc-

收稿日期:2024-12-16;在线发布日期:2025-07-17。本课题得到城市智慧管网“观、管、防”一体化关键技术的研发与示范应用项目(No. 25100002024121001)资助。庞双双,博士研究生,主要研究领域为时空数据挖掘、交通流预测、持续学习。E-mail: ireland@buaa.edu.cn。邓 攀(通信作者),博士,副教授,主要研究领域为时空数据挖掘、交通流预测、因果性学习。E-mail: pandeng@buaa.edu.cn。赵 宇,博士研究生,主要研究领域为时空数据挖掘、交通流预测、因果性学习。刘俊廷,博士研究生,主要研究领域为时空数据挖掘、交通流预测、因果性学习。王子昂,博士研究生,主要研究领域为时空数据挖掘、交通流预测、因果性学习。李思睿,硕士研究生,主要研究领域为时空数据挖掘、交通流预测、因果性学习。

tures. This oversight makes it fundamentally challenging to effectively capture nuanced local differences and dynamic temporal variations in spatio-temporal patterns across diverse spatial regions or evolving time periods, ultimately culminating in suboptimal prediction performance that fails to address real-world complexity. To address this pervasive issue, this paper proposes a novel spatio-temporal prediction method named STMP2G (Spatio-Temporal Meta-Pattern Prompt Guidance), which innovatively integrates meta-pattern mining with prompt guidance mechanisms. The framework employs a deliberately simplified multi-layer perceptron (MLP) as its core prediction network, thereby prioritizing adaptability over architectural complexity. Crucially, STMP2G captures spatio-temporal heterogeneity by dynamically generating specialized spatio-temporal prompts. These prompts undergo enhancement through a dedicated self-supervised learning phase designed to mine universally applicable meta-patterns from rich historical data repositories. Ultimately, the refined prompts guide the entire spatio-temporal modeling process, strategically overcoming heterogeneity interference to achieve both accurate and robust predictions even in highly non-stationary environments. Specifically, the method first constructs a spatio-temporal prompt generation module. This component generates multi-scale temporal prompts—spanning hourly, daily, and weekly granularities—using temporal prior embedding combined with a memory network to comprehensively capture temporal heterogeneity. Simultaneously, it produces semantic spatial prompts via learnable spatial embedding matrices optimized through representation learning objectives, thereby encoding region-specific attributes reflecting spatial heterogeneity. To model intricate interdependencies, the module constructs a high-order spatio-temporal interaction tensor and achieves deep fusion of spatial and temporal prompts using advanced inverse tensor decomposition techniques, ensuring cohesive cross-dimensional feature integration. Subsequently, STMP2G incorporates a self-supervised meta-pattern mining module. This stage employs contrastive self-supervised learning to discover transferable, discriminative meta-patterns from extensive historical sequences within a unified feature space. These foundational patterns—once identified—have their parameters permanently frozen and are systematically archived within a structured meta-pattern library, creating a reusable knowledge base for efficient real-time retrieval during inference. Finally, the architecture features a spatio-temporal prompt guidance module. During prediction, this component retrieves the top- k most relevant meta-patterns from the meta-pattern library based on similarity to the input spatio-temporal sequence, using these to contextually supplement and enhance the initial prompts. The augmented prompts are then mapped to the feature space of each MLP layer via lightweight, layer-specific adapters. Through Hadamard product operations, the module achieves fine-grained, adaptive enhancement of evolving spatio-temporal representations, thereby guiding the modeling process to dynamically counteract heterogeneity interference at multiple hierarchical levels. Extensive experiments on six real-world datasets demonstrate our spatio-temporal prediction method effectively counters heterogeneity, boosting accuracy and robustness. It reduces prediction errors by 4.53% on average. The method excels in interpretability, efficiency, generalization, and robustness, adapting to diverse patterns and conditions like data sparsity. Its lightweight design ensures real-time efficiency, while interpretable prompts show dynamic adjustments based on learned patterns. This versatility overcomes limitations of traditional methods in handling real-world variations, enabling more reliable predictions for applications from smart cities to climate modeling.

Keywords spatiotemporal prediction; heterogeneity; deep learning; prompt learning; self-supervised learning

1 引言

时空预测的目的是通过分析和建模历史时空数据,从而预测未来在特定时空维度上的变化趋势。准确的未来趋势预测能够为各类决策提供科学依据,推动社会管理和资源配置的优化。因此,时空预测在交通管理、气象预报、环境监测等众多领域中具有重要的应用价值,近年来受到了广泛关注^[1-6]。

深度学习技术具有强大的特征学习能力,能够自动建模观测数据中复杂的时空依赖关系,因此被广泛应用于时空预测任务中。基于深度学习的时空预测方法^[7-13]通常将时空数据中的空间特征建模为图结构,其中节点和边分别表示空间单元及其相互依赖关系,将时间特征建模为图上节点的时间序列,通过循环神经网络^[14]及其变体^[15-16]或卷积神经网络^[17]及其变体^[18-19]建模时间维度上的动态依赖关系,通过图卷积网络^[20]建模空间维度上的局部依赖关系,或者利用注意力机制^[21-23]捕获时空相关性。尽管研究者们为实现准确的时空预测做出了巨大的努力,但现有模型过于关注全局平均特性,假设所有空间区域或时间段内的模式变化是均匀的,忽略了时空数据中固有的异质性,导致模型无法有效捕捉时空模式的局部差异和动态变化,预测性能大幅下降。

时空异质性是指在不同的空间区域或不同的时间段内,数据的分布和模式表现出的差异和不均匀性。以交通场景为例,图1(a)展示了三条北京市通州区内具有不同时空属性的道路,分别为城市快速路A、商业区道路B和住宅区道路C。在空间上,如图1(b)所示,不同类型道路的速度分布具有显著差异,然而现有模型往往忽略了这种空间异质性,为所有空间区域采用统一的建模方式,因此难以准确捕捉不同区域之间差异化的时空模式,导致预测表现不佳。一些研究通过为每个空间区域分配独立的参数空间来捕获空间异质性^[24-25],但其庞大的参数量容易导致优化困难和过拟合问题,且高昂的计算和内存成本也使其难以高效处理大规模的时空数据。此外,元学习技术被用于快速适应不同空间区域的特征^[26-28],但其有效性严重依赖于辅助的空间信息,例如兴趣点和传感器位置,这种依赖可能限制模型的泛化能力,尤其是在缺乏足够辅助信息的区域或数据场景中。

在时间上,如图1(c)所示,三种不同类型的道

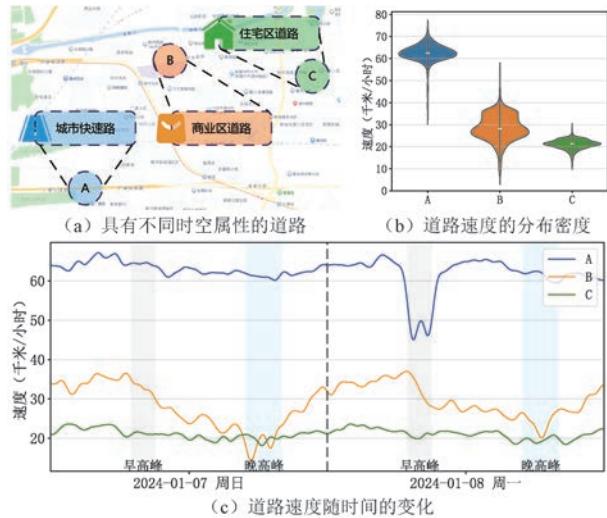


图1 时空异质性图示

路在周日与周一的速度变化均明显不同。城市快速路A的速度受通勤模式的影响,在工作日的早高峰时段骤降,但在周末则基本没有波动;商业区道路C在周末的晚高峰时段出现购物高峰,导致速度大幅下降,但在工作日的对应时段则表现不明显;住宅区道路E由于道路规划和交通管理始终保持低速平稳的状态,在早晚高峰时段的速度变化都不显著。因此,同一道路在不同时间段的模式差异揭示了多尺度的时间异质性,然而现有模型通常以全局共享的参数静态地建模随时间不断变化的相互依赖关系,这使其难以准确捕捉各时间段差异化的时空模式,从而降低预测性能。一些研究通过为不同时间段建立独立的邻接矩阵,或者为每个时间步独立学习依赖关系来捕获时间异质性^[29-31],但这些方法要么只能建模固定时间周期的差异性,难以有效捕捉多尺度的时间异质性,要么计算成本过高且容易发生过拟合,从而在大规模时空场景中应用受限。此外,一些研究通过设计额外的自监督学习任务来捕获异质性^[32-33],但其完全端到端的联合优化为时空预测带来了困难。

为克服时空数据异质性对预测精度的影响,本文提出了一种新颖的基于元模式挖掘与提示引导的时空预测方法STMP2G(Spatio-Temporal Meta-Pattern Mining and Prompt Guidance)。该方法通过生成时空提示捕获时空异质性,并与离线挖掘的元模式共同引导时空建模过程以克服时空异质性的干扰。STMP2G包含三个主要组成部分:(1)时空提示生成模块:该模块分为时间提示生成和空间提示生成两部分。在时间上,将多尺度的时间先验信息从离散映射到连续空间,得到时间融合嵌入;通过

记忆网络存储长期时间记忆,并结合嵌入查询生成时间提示,以捕获多尺度时间异质性。在空间上,利用可学习的节点嵌入矩阵捕获空间异质性,并通过表征学习生成语义化的空间提示;最后通过反向张量分解建模复杂的时空交互作用,实现时空提示的深度融合,从而引导时空建模过程。(2)自监督元模式挖掘模块:该模块通过自监督学习从历史数据中挖掘具有判别能力的元模式,并通过掩码、加噪和对比学习等策略增强元模式的判别能力和鲁棒性,最后将其存储为参数冻结的元模式库,用于后续检索。元模式能够提供与当前输入时空序列最相关的局部趋势信息,与时空提示的全局背景信息互补,从而更全面地解决时空异质性问题。(3)时空提示引导模块:该模块首先从元模式库中检索与输入时空序列最相关的元模式,以增强时空提示;随后通过逐层适配器将提示信息映射到不同层次的特征空间中;最后利用哈达玛积对时空表征进行细粒度增强,从而引导时空建模过程,有效克服异质性的干扰。

综上所述,本文的主要贡献总结如下:

(1)本文针对时空预测任务中的异质性挑战,提出了一种新颖的基于元模式挖掘与提示引导的时空预测方法 STMP2G,它是最早结合自监督学习与提示学习以克服时空异质性干扰的工作之一;

(2)设计了灵活高效的时空提示生成与引导方法,通过捕获时空异质性为时空预测模型提供全局背景信息,并对时空表征进行逐层适配与提示引导,从而克服异质性对时空建模的干扰,显著提升模型的预测精度与适应能力;

(3)构建了离线的自监督学习任务,通过从历史数据中挖掘具有判别能力的元模式来补充和增强时空提示,该学习过程独立于端到端训练的预测过程,能够有效提升元模式的质量且易于优化;

(4)在六个真实世界的时空数据集上进行了大量实验,结果表明 STMP2G 在多样化的时空场景中均显著优于主流时空预测方法,验证了通过时空提示捕获时空异质性的正确性,以及利用提示引导时空建模过程从而克服异质性干扰的有效性。此外,进一步的综合实验评估证明了本文方法具有较强的可解释性、效率、泛化性与鲁棒性。

2 相关工作

为了实现准确的时空预测,研究者们已提出多种不同的方法和模型。早期的时空预测方法主要分

为统计学习方法与机器学习方法两类。统计学习方法^[34-36]的优势在于模型简单且可解释性较强,但其通常假设时间序列数据是线性且平稳的,因此只适合建模简单的时间依赖关系,而对复杂的非线性关系建模能力不足。机器学习方法^[37-38]能够处理非线性问题,并在高维特征数据中表现出较强的灵活性,然而它们通常依赖大量的手工特征工程且对数据的质和量要求较高。此外,两类方法通常将时空预测问题简化为单变量或多变量的时间序列预测问题,忽略了动态复杂的空间依赖性,导致预测偏差较大,从而限制其在真实时空数据上的表现。

随着深度学习技术的不断发展,其强大的自动特征提取和非线性建模能力使得越来越多的研究者将深度学习应用于时空预测,探索如何全面建模时空数据中的时间与空间依赖关系。基于深度学习的早期模型^[7-9]通常利用循环神经网络^[14]及其变体^[15-16]建模时间依赖关系,同时使用卷积神经网络^[17]建模空间依赖关系。这类方法在捕获规则网格结构的时空关系时表现较好,但受限于传统卷积操作对网格结构的依赖,难以泛化到交通网络等非欧几里得结构。之后,图卷积网络^[20]的提出为非欧几里得结构数据的建模提供了解决方案,其通过在图结构数据上直接应用卷积操作,有效地捕获了节点间的空间依赖关系。研究者们在此基础上进一步结合图卷积网络与时间序列建模方法,发展出多种时空图神经网络^[11-13]。其中,结合图卷积网络与一维卷积网络的模型能够高效提取局部时空特征,在短期依赖建模中表现优异,但在处理长期依赖关系时存在一定的局限性。结合图卷积网络与循环神经网络的模型^[10-24]则采用递归方式建模时间步长,能够更好地捕获长期依赖关系,但由于递归计算的限制,其训练效率较低且难以并行化。此外,Transformer^[22]强大的全局依赖建模能力和高效的并行计算优势为时空预测任务带来了新的思路,一系列基于 Transformer 的时空模型^[22-23]被提出,并展现出卓越的时空依赖建模能力。然而这些方法在计算上存在较高的成本,尤其是在处理大规模数据时,其计算复杂度和内存消耗成为实际应用中的主要瓶颈。尽管上述模型在时空预测任务中取得了显著进展,但大多数方法对所有节点采用统一的建模方式,并依赖全局共享的参数空间来建模动态的时间模式,忽略了时空数据在空间维度和时间维度上的异质性特征。这种局限性可能导致局部动态模式未被充分建模,从而产生偏差的预测结果。因此,如何有

效捕捉并利用时空数据中的异质性,成为优化时空预测模型的重要研究方向。

近年来,许多异质性建模方法被提出,以充分挖掘时空数据中固有的异质性特征。这些方法大致分为三类:仅针对时间异质性的方法、仅针对空间异质性的方法以及综合处理时空异质性的方法。

(1)单一的时间异质性建模方法主要集中于捕捉时间维度上的异质性,通过识别某种周期性对不同时间段进行动态建模。DMSTGCN^[29]针对一天中的每个时间步学习独立的邻接矩阵,虽然缓解了短期的时间异质性,但是忽略了长范围(如周末与工作日)的异质性特征。TGCNN^[30]通过时间感知图结构动态捕捉空间关联的趋势性和周期性变化,虽然有助于解决时间异质性,但时间差异学习的约束条件过于简单导致学习不充分,且为每个时间步学习动态图结构并计算图卷积的计算压力过大。这类方法在时间异质性建模方面有一定的优势,但由于忽略了空间异质性的影响,导致模型无法适应不同空间区域的局部特征,从而限制了对全局复杂时空依赖关系的捕获能力,降低了预测精度,特别是在区域差异显著的场景中表现不佳。

(2)单一的空间异质性建模方法专注于捕捉空间维度上的异质性,通常为不同节点采用差异性的参数空间或建模方式。AGCRN^[24]为每个节点分配了独立的参数空间以捕获节点具体的模式,并通过矩阵分解的方式从预设的权重池中生成参数。这是对解决空间异质性的一次有效尝试,然而其单一的邻接矩阵构建方式限制了对空间依赖的充分捕获,同时参数池规模随着模型的复杂化不断增加,导致冗余计算和优化困难问题。TESTAM^[39]使用混合专家模型为不同节点灵活地采用不同的空间建模方式,包括静态、自适应以及基于注意力机制的空间建模方式,能够有效缓解空间异质性,但其直接对模型输出结果进行选择的方式可能导致节点间特征空间的不平衡问题,同时其计算复杂度和内存成本较高。这类方法对异质性空间节点的局部模式具有良好的适应性,但由于忽略了时间异质性的影响,导致模型无法适应不同时间范围的模式变化,从而难以准确预测时空序列的动态变化过程,尤其在具有强时间依赖性的场景中表现受限。

(3)此类方法通常采用复杂的模型架构以同时捕捉时间和空间维度的异质性。MegaCRN^[40]通过记忆网络构建元节点库学习时空图,初步捕获了时空异质性,但其使用唯一的元节点库同时处理时间

和空间维度,容易导致特征平均化,从而削弱单一特性的建模能力。STSSL^[32]通过自适应图增强来捕获节点间的语义关系,并在此基础上以自监督的方式建模时空异质性,但其过于强调时间的差异性而忽视了共性特征,且对所有空间区域提取全局模式作为参照可能导致表征学习趋同,从而在捕获时间异质性的过程中削弱对空间异质性的捕获。HimNet^[25]为空间、时间、时空分别开辟了一组独立的参数空间,并通过嵌入查询为异质性建模寻找最合适的参数来缓解时空异质性,但其参数量随着时空模型的复杂化成倍增加,带来了极大的资源消耗,难以在大规模的时空场景中广泛应用。此外,STSSL、HimNet 完全端到端的联合优化也为充分捕获时空异质性带来了困难。

尽管前人为了实现准确的时空预测做出了许多努力,但仍然缺乏一种高效、全面的时空异质性捕获与利用方法。因此,本文通过灵活的时空提示生成捕获时空异质性;构建离线的自监督学习任务以挖掘元模式,用于补充和增强时空提示;最终利用提示高效地引导时空建模过程,从而克服时空异质性的干扰,实现准确而鲁棒的预测。

3 问题定义

给定历史时空数据 $\mathcal{X} \in \mathbb{R}^{N \times P \times C}$, 其中 N 为空间节点数量, P 为历史时间步长, C 为特征维度, 时空预测的目标是训练一个含参数集 θ 的映射函数 $\mathcal{F}_\theta: \mathbb{R}^{N \times P \times C} \rightarrow \mathbb{R}^{N \times Q \times C}$, 使得未来 Q 个时间步的预测值 $\hat{y} = \mathcal{F}_\theta(\mathcal{X})$ 与真实值 y 的差异最小。给定一个损失函数 $\mathcal{L}(\cdot)$, 时空预测问题可以形式化定义为

$$\theta^* = \operatorname{argmin}_\theta \mathcal{L}(\mathcal{F}_\theta(\mathcal{X}), y) \quad (1)$$

4 方法

本文提出了一种新颖的基于元模式挖掘与提示引导的时空预测方法 STMP2G, 包含三个组成模块: 时空提示生成模块、自监督元模式挖掘模块、时空提示引导模块, 整体架构如图 2 所示。与过分关注捕获时空依赖关系的深度时空预测方法不同, STMP2G 更侧重于对各层网络编码得到的时空表征进行提示引导, 增强其在时空异质性干扰下的表征能力, 从而提高预测精度和鲁棒性。其中, 时空提示生成模块分别在时间和空间

维度上生成多尺度的时间提示和语义化的空间提示,以区分和建模全局时空异质性,进而通过反向张量分解建模复杂的时空交互作用,实现时空提示的深度融合。自监督元模式挖掘模块构建独立的自监督学习任务,从历史时空数据中挖掘具有判别能力的元模式,精准捕获局部时空异

质性,并将其参数冻结存储为元模式库。时空提示引导模块首先检索元模式库得到局部趋势信息,用于补充和增强时空提示;接着将全面的时空提示输入各层适配器,灵活映射到不同层次的特征空间中;最后对各层时空表征进行提示引导,有效克服异质性对时空建模的干扰。

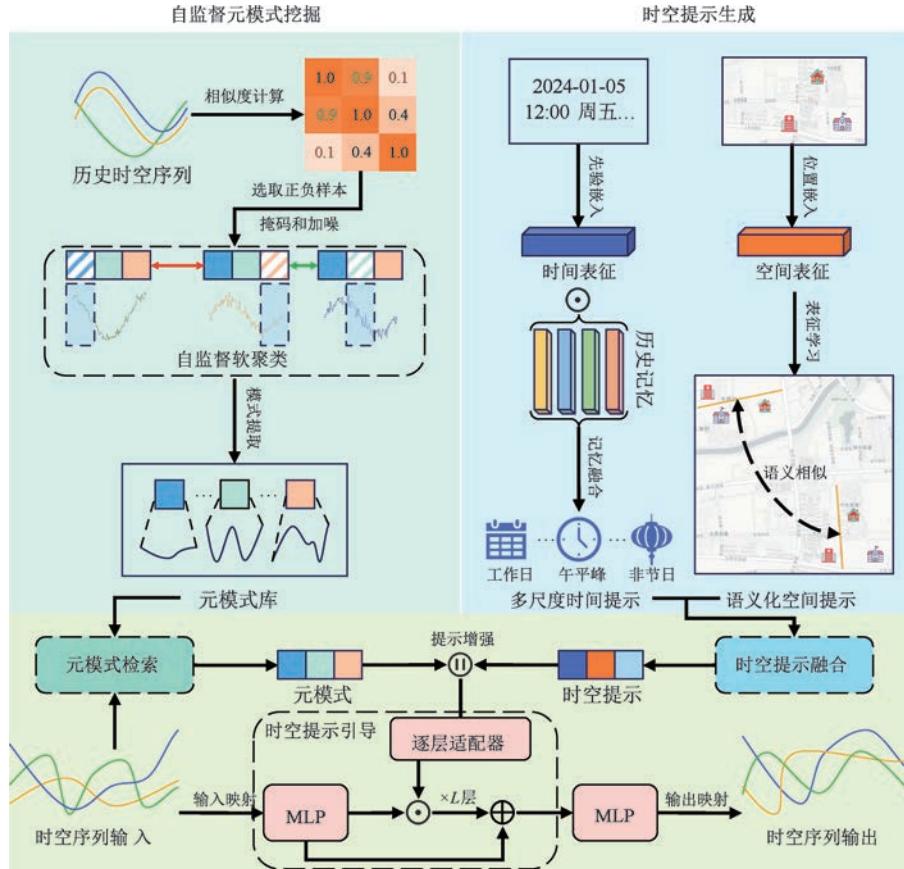


图 2 STMP2G 方法架构图

4.1 时空提示生成模块

传统深度时空预测方法往往过于关注全局平均特性而忽略时空数据中固有的异质性,导致模型难以捕捉局部区域或特定时间段内的细粒度差异与动态变化。本模块通过时间提示生成捕获时间异质性,通过空间提示生成捕获空间异质性,并通过建模时空交互作用来融合时空提示,用于引导预测模型的时空建模过程,从而有效克服异质性的干扰。下文将对时间与空间提示的生成与融合过程分别进行详细阐述。

4.1.1 时间提示生成

与仅依靠短期历史序列进行预测相比,多尺度的时间先验能够提供全局信息,有助于模型识别当前时间段内的特定模式,并充分利用从历史数据中积累的知识进行针对性预测。因此,本部分利用先

验嵌入与记忆网络生成多尺度的时间提示,捕获潜在的时间异质性,从而提高对不同时间段变化模式的预测精度。

首先,本部分将多尺度的时间先验知识(包括月份、日期、时间步、周、是否为工作日、是否为节假日等信息)进行从离散到连续的映射。对于不同的时空场景,时间先验知识的重要性可能不同。例如,在交通场景中,工作日与节假日的区分显著影响人类通勤模式,因此在交通预测中具有重要作用;但由于人类行为模式对气象的影响较弱,在气象场景中区分工作日或节假日并不会为时空预测带来明显的性能提升。因此,本部分设计了一种灵活的时间先验嵌入方式,基于领域知识与实验验证,仅选择对当前时空场景具有显著影响的 T_m 个时间先验知识 $t \in$

$\mathbb{R}^{T_m \times 1}$ 进行嵌入:

$$e = \text{Embedding}(t) \in \mathbb{R}^{T_m \times d_i} \quad (2)$$

其中, $\text{Embedding}(\cdot)$ 表示嵌入操作, T_m 和 d_i 分别为时间先验知识的数量和嵌入维度。以一个小批次 B 的输入数据 $\mathbf{X}_{t-P+1:t} \in \mathbb{R}^{B \times N \times P \times C}$ 为例, 为了充分利用时间先验信息, 本文为历史窗口 P 和未来窗口 Q 的每个时间步都生成相应的嵌入 $\mathbf{E} \in \mathbb{R}^{B \times (P+Q) \times D_T}$, 其中 $D_T = T_m \times d_i$ 为先验嵌入拼接的特征维度。在此基础上, 本部分通过卷积特征融合捕捉不同时间步间的局部依赖关系, 并结合适应性最大池化提取关键特征, 从而充分发挥时间先验知识对当前预测任务的指导能力, 如下式:

$$\mathbf{E}_t = \text{MaxPool1D}(\text{Conv1D}_2(\sigma(\text{Conv1D}_1(\mathbf{E}^\top)))) \quad (3)$$

其中, $\mathbf{E}_t \in \mathbb{R}^{B \times d_t}$ 为时间先验嵌入的融合表征, d_t 为其特征维度, $\text{Conv1D}_1(\cdot)$ 和 $\text{Conv1D}_2(\cdot)$ 分别表示两层卷积操作, $\sigma(\cdot)$ 为 GELU 激活函数, $\text{MaxPool1D}(\cdot)$ 表示适应性最大池化。

之后, 从现有关于记忆网络的研究^[39-41] 中获得启发, 本部分使用记忆网络 $\mathcal{M} \in \mathbb{R}^{m \times d_t}$ 存储从历史数据中学习的长期时间记忆, 从而全面建模复杂的长期时间依赖关系, 并进一步增强时间融合表征。其中, m 表示 \mathcal{M} 中记忆单元的数量。与使用记忆网络同时存储时空特征导致建模干扰的研究不同, 本部分专注于时间记忆的存储, 可以有效捕捉复杂的长期依赖关系。记忆网络的查询过程如下式:

$$\left\{ \begin{array}{l} \mathbf{Q}_t^{(i)} = \mathbf{E}_t^{(i)} \mathbf{W}_q + \mathbf{b}_q \\ a_{i,j} = \frac{\exp(\mathbf{Q}_t^{(i)} \mathcal{M}[j]^\top)}{\sum_{j=1}^m \exp(\mathbf{Q}_t^{(i)} \mathcal{M}[j]^\top)} \\ \mathbf{V}_t^{(i)} = \sum_{j=1}^m a_{i,j} \mathcal{M}[j] \\ \mathbf{O}_t^{(i)} = \mathbf{V}_t^{(i)} \mathbf{W}_o + \mathbf{b}_o \end{array} \right. \quad (4)$$

其中, i 为批次中的样本索引, $\mathbf{Q}_t \in \mathbb{R}^{B \times d_t}$ 表示由时间嵌入 \mathbf{E}_t 映射得到的查询向量, $a_{i,j}$ 为第 i 个时间嵌入与第 j 个记忆单元 $\mathcal{M}[j] \in \mathbb{R}^{d_t}$ 之间的注意力分数, $\mathbf{V}_t \in \mathbb{R}^{B \times d_t}$ 表示 \mathcal{M} 中记忆单元的加权和, $\mathbf{O}_t \in \mathbb{R}^{B \times d_t}$ 表示记忆增强的时间融合表征, 即生成的多尺度的时间提示, 其捕获了潜在的时间异质性。

4.1.2 空间提示生成

传统时空预测方法通常忽略不同空间区域的差异性, 并假设空间依赖关系是同质且静态的, 使得模

型难以有效适应空间异质性, 导致在异质性强的时空场景下产生有偏的预测结果。空间先验有助于揭示空间异质性的潜在驱动因素, 使模型能够更深入理解不同空间区域的特性, 从而在预测时针对异质性的空间区域进行差异化建模。然而, 与时间先验不同, 空间先验往往存在获取困难、易引入噪声、人工成本高昂等问题, 这使得基于静态知识的建模方式难以有效捕获隐式的空间依赖。因此, 本部分引入了可学习的空间嵌入矩阵 $\mathbf{E}_s \in \mathbb{R}^{N \times d_s}$, 其中 d_s 为分配给每个空间位置嵌入向量的特征维度。经过模型训练, \mathbf{E}_s 将自动学习语义化的空间关系, 从而捕获不同空间位置潜在的异质性, 提高对个性模式的预测精度, 具有更强的灵活性和通用性。

4.1.3 时空提示融合

只关注时间异质性或空间异质性可能会限制模型对时空数据的理解能力, 且对复杂时空交互的捕捉和解释不足。受前人工作的启发^[42], 本部分利用可学习的张量建模二者之间复杂的交互作用, 并将时空提示的融合过程视为张量分解的反向过程, 整个时空提示融合过程如图 3 所示。

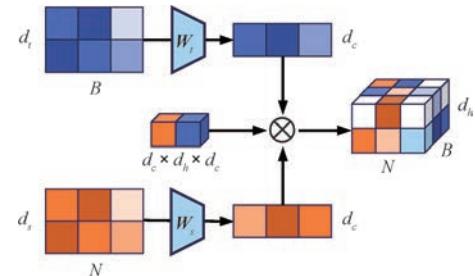


图 3 时空提示融合

具体来说, 本部分结合前文生成的时间提示 \mathbf{O}_t 与空间提示 \mathbf{E}_s , 通过可学习的张量全面地建模了时空交互作用, 实现了时空提示的深度融合:

$$\mathbf{E}_{st} = (\mathbf{O}_t \mathbf{W}_t) \cdot \mathbf{W}_c \cdot (\mathbf{E}_s \mathbf{W}_s)^\top \quad (5)$$

其中, $\mathbf{E}_{st} \in \mathbb{R}^{B \times N \times d_h}$ 表示融合后的时空提示, $\mathbf{W}_c \in \mathbb{R}^{d_c \times d_h \times d_c}$ 为可学习的时空交互张量, d_c 为时空交互维度, d_h 为隐藏层维度, $\mathbf{W}_t \in \mathbb{R}^{d_t \times d_c}$ 和 $\mathbf{W}_s \in \mathbb{R}^{d_s \times d_c}$ 分别为映射时间和空间提示到交互空间的权重。

4.2 自监督元模式挖掘模块

传统时空预测方法通常假设时空模式是全局均匀变化的, 忽略了局部区域中蕴含的具体差异与动态性, 导致模型难以准确捕捉局部的时空异质性, 从而限制了预测性能。例如, 在交通场景中, 商业区在

早晚高峰时段出现的车流高峰和节假日景区周边的车流模式是典型的局部时空异质性表现。这些元模式反映了特定区域和时段的独特规律,是具有判别能力的潜在特征模式,能够为预测模型提供关键的局部趋势信息。因此,本文通过从历史数据中挖掘元模式,将其与捕获全局时空异质性的时空提示相互补充,进一步增强了时空提示对建模过程的引导能力。通过结合元模式的局部动态信息与时空提示的全局背景信息,本文方法能够更全面地克服时空异质性的干扰,显著提升模型的预测精度与适应性。

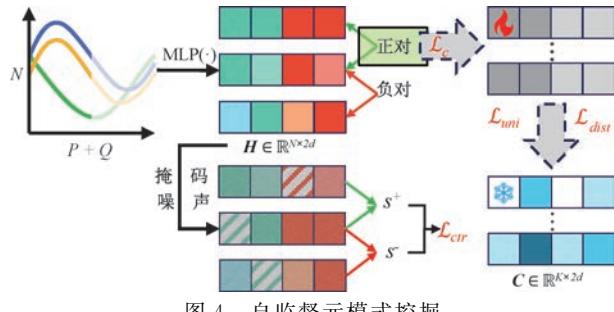


图 4 自监督元模式挖掘

如图 4 所示,本模块首先将时空序列从数据空间映射到特征空间,以期增强表达能力,并拼接历史窗口与未来窗口对应的特征向量:

$$H = \text{concat}(MLP_P(\mathbf{X}_{t-P+1:t}), MLP_Q(\mathbf{X}_{t+1:t+Q})) \quad (6)$$

其中, $MLP_P(\cdot), MLP_Q(\cdot)$ 分别为历史序列 $\mathbf{X}_{t-P+1:t}$ 和未来序列 $\mathbf{X}_{t+1:t+Q}$ 的特征映射网络, $\mathbf{H} \in \mathbb{R}^{N \times 2d_h}$ 为映射后的特征向量, $\text{concat}(\cdot)$ 表示特征拼接操作。

之后,本模块基于特征空间上的软聚类过程挖掘历史预测片段中的元模式,使用 K 个可训练的原型 $\mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_K\} \in \mathbb{R}^{K \times 2d_h}$ 作为聚类中心,聚类过程表示为样本到原型的分配概率:

$$p_{i,k} = \frac{\exp(\text{sim}(\mathbf{h}_i, \mathbf{c}_k) / \gamma)}{\sum_{j=1}^K \exp(\text{sim}(\mathbf{h}_i, \mathbf{c}_j) / \gamma)} \quad (7)$$

其中, $\text{sim}(\cdot, \cdot)$ 表示相似性度量(本文使用余弦相似度), γ 为温度参数,用于控制分布的平滑性。

构建自监督学习任务需要自监督信号,本模块基于动态空间相似度选取正、负样本对,用正样本对的分配概率计算多分类交叉熵作为优化目标。具体而言,本模块对每个小批量的样本计算平均空间相似度矩阵,从而捕获样本间动态的空间相关性:

$$A = \frac{1}{B} \sum_{b=1}^B \left(\frac{\mathbf{H}_b}{\|\mathbf{H}_b\|_2} \cdot \frac{\mathbf{H}_b^T}{\|\mathbf{H}_b\|_2} \right) \quad (8)$$

其中, $\mathbf{H}_b \in \mathbb{R}^{N \times 2d_h}$ 表示批次中第 b 个样本特征。相较于计算全部样本的相关性,该方式更有利于捕获局部的动态关系且效率更优。接着,本模块对 \mathbf{A} 进行平移归一化 $\mathbf{A} = (\mathbf{A} + 1) / 2$ 与对角掩蔽 $\text{diag}(\mathbf{A}) = -\infty$,并通过预设的阈值 τ 来选取正、负样本对:

$$j_i^+ = \begin{cases} \text{Rand}(\{j | A_{ij} > \tau\}), & \text{if } \exists j, A_{ij} > \tau \\ \text{argmax}_j A_{ij}, & \text{otherwise} \end{cases} \quad (9)$$

其中, $\text{Rand}(\cdot)$ 表示从集合中随机选取, j_i^+ 为样本 i 对应的正样本索引。负样本的选取同理,将阈值改为 $1 - \tau$,从而选取低相似度的样本。这种随机选取的方式有助于捕获多样化的语义关系,通过阈值 τ 限制随机范围能够有效避免随机性过大导致的噪声问题,同时最近邻的选取方式确保了数据不平衡情况下(如高相似度样本稀疏)的训练稳定性。最后,本模块以正样本到原型的分配概率作为自监督信号,构建完整的自监督任务:

$$\mathcal{L}_{ce} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K p_{i,k}^+ \log p_{i,k} \quad (10)$$

此外,基于选取的正、负样本 \mathbf{h}^\pm ,本模块通过添加随机掩码和高斯噪声来增强原型的鲁棒性,从而减少其对某些特定特征的过度依赖:

$$\mathbf{h}_{aug}^\pm = \mathbf{h}^\pm \odot (1 - \mathbf{M}) + \epsilon \quad (11)$$

其中, \mathbf{h}_{aug}^\pm 表示增强后的特征向量, \odot 表示哈达玛积, \mathbf{M} 为掩码矩阵, ϵ 为高斯噪声。之后,本模块采用对比学习策略,在特征空间中拉近正样本对、推远负样本对,从而增强元模式的判别能力,提高聚类过程的鲁棒性:

$$\mathcal{L}_{ctr} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(s_i^+ / \tau)}{\exp(s_i^+ / \tau) + \exp(s_i^- / \tau)} \quad (12)$$

其中, s_i^+ 和 s_i^- 分别表示样本 i 与其正样本和负样本之间的余弦相似度。

最后,为了防止样本到原型的分配过于不平衡而导致模型崩溃,本模块添加了均匀正则化项,以鼓励样本均匀分配到各原型中,同时添加了距离正则化项,以提升原型的区分性和表达能力:

$$\mathcal{L}_{uni} = KL\left(\frac{1}{K} \mathbf{1}_K, \frac{1}{N} \sum_{i=1}^N p_i\right) \quad (13)$$

$$\mathcal{L}_{dist} = \frac{1}{K(K-1)} \sum_{i=1}^K \sum_{j=1, j \neq i}^K (\text{sim}(\mathbf{c}_i, \mathbf{c}_j))^2 \quad (14)$$

其中, $KL(\cdot, \cdot)$ 表示 KL 散度^[43],用于衡量两个概率分布之间的差异。因此,本模块通过以下优化目

标实现了自监督的元模式挖掘:

$$\mathcal{L}_{ssl} = \mathcal{L}_{ce} + \mathcal{L}_{ctr} + \lambda_{reg} (\mathcal{L}_{uni} + \mathcal{L}_{dist}) \quad (15)$$

其中, \mathcal{L}_{ce} 表示交叉熵损失, 用于自监督软聚类任务; \mathcal{L}_{ctr} 表示对比损失, 用于增强元模式的判别能力; \mathcal{L}_{uni} 和 \mathcal{L}_{dist} 分别表示均匀性损失和分散性损失, 用于正则化元模式的分布; λ_{reg} 为正则化项的系数, 用于平衡各项损失的权重。

4.3 时空提示引导模块

传统时空预测方法存在对时空异质性的建模不充分、计算和内存成本过高等问题, 难以在真实的大规模时空场景中有效应用。因此, 本模块通过元模式检索提取与输入时空序列最相关的元模式, 用以增强生成的时空提示, 并输入逐层适配器将其映射到不同层次的特征空间中, 最后对各层网络的时空表征进行高效的提示引导, 从而使模型克服时空异质性的干扰, 适应动态复杂的时空变化, 提高预测精度和鲁棒性。

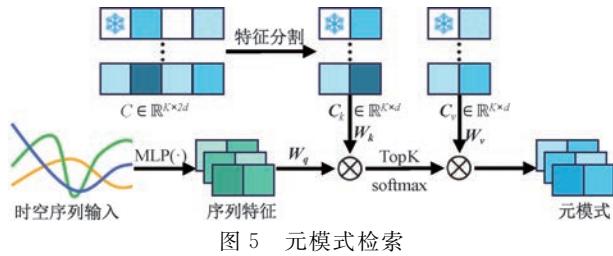


图 5 元模式检索

如图 5 所示, 本模块首先对元模式库 \mathbf{C} 中的每个元模式进行特征分割, 得到分别对应于历史窗口和未来窗口的元模式组 $\mathbf{C}_k \in \mathbb{R}^{K \times d_h}$, $\mathbf{C}_v \in \mathbb{R}^{K \times d_h}$; 接着用输入时空序列检索元模式库, 得到其与各个元模式的匹配程度 $\mathbf{S} \in \mathbb{R}^{N \times K}$, 但只保留最相关的 $topk$ 个; 最后对元模式进行加权融合, 提取与输入时空序列最相关的元模式, 如下式:

$$\left\{ \begin{array}{l} \mathbf{S} = \text{MLP}_P(\mathbf{X}_{t-P+1:t}) \mathbf{W}_q \cdot (\mathbf{C}_k \mathbf{W}_k)^T \\ \alpha = \text{softmax}(\text{TopK}(\mathbf{S}, k = topk)) \\ \mathbf{E}_p = \alpha \cdot \mathbf{C}_v \mathbf{W}_v \end{array} \right. \quad (16)$$

其中, $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{d_h \times d_h}$ 为可学习的映射权重, $\text{TopK}(\cdot, k = topk)$ 表示 TopK 选择操作, 只保留每个查询对应的 $topk$ 个最高分数, $\alpha \in \mathbb{R}^{N \times K}$ 是最终的注意力权重, $\mathbf{E}_p \in \mathbb{R}^{B \times N \times d_h}$ 表示融合的元模式。

时空提示捕获了全局的时空异质性, 检索得到的融合元模式捕获了局部的时空异质性, 因此, 本模块将元模式 \mathbf{E}_p 与时空提示 \mathbf{E}_{st} 进行特征拼接, 得到增强的时空提示 $\mathbf{E}_c \in \mathbb{R}^{B \times N \times 2d_h}$, 旨在全面引导时空建模过程, 如下式:

$$\mathbf{E}_c = \text{concat}(\mathbf{E}_p, \mathbf{E}_{st}) \quad (17)$$

最后, 本模块实现了逐层适配器, 将增强的时空提示 \mathbf{E}_c 灵活映射到不同层次的特征空间中, 从而增强其适应性和表征能力, 并借助哈达玛积和残差连接实现细粒度、稳定的提示引导, 如下式:

$$\text{MLP}_A^{(l)}(\mathbf{E}_c) = \sigma(\mathbf{E}_c \mathbf{W}_0^{(l)}) \mathbf{W}_1^{(l)} \quad (18)$$

$$\mathbf{H}^{(l)} = \mathbf{H}^{(l)} \odot (1 + \text{MLP}_A^{(l)}(\mathbf{E}_c)) \quad (19)$$

其中, $\mathbf{W}_0^{(l)} \in \mathbb{R}^{2d_h \times d_h}$ 和 $\mathbf{W}_1^{(l)} \in \mathbb{R}^{d_h \times d_h}$ 为第 l 层适配器的映射权重, $\mathbf{H}^{(l)} \in \mathbb{R}^{d_h \times d_h}$ 表示第 l 层网络编码得到的时空表征, $\mathbf{H}^{(l)'} \in \mathbb{R}^{d_h \times d_h}$ 表示提示引导后的时空表征。

4.4 优化策略

STMP2G 采用二阶段的训练策略, 利用反向传播和随机梯度下降算法进行优化。第一阶段优化式 (15) 中的 \mathcal{L}_{ssl} , 挖掘历史数据中的元模式, 并将其存储到元模式库中, 随后冻结其参数, 避免在后续阶段优化其他部分时产生干扰。与其他将自监督学习任务与主任务联合优化的研究不同, STMP2G 将元模式挖掘过程与主任务独立进行, 能够减少任务间的干扰, 从而提升学习效率与特征质量。

第二阶段聚焦于时空预测模型的优化, 目标是最小化预测值 \hat{y} 与真实值 y 之间的平均绝对误差 (MAE), 损失函数为:

$$\mathcal{L}_{mae} = \frac{1}{NQC} \sum_{n=1}^N \sum_{q=1}^Q \sum_{c=1}^C |\hat{y}_{n,q,c} - y_{n,q,c}| \quad (20)$$

5 实验

5.1 实验设置

5.1.1 数据集描述

本文在三个交通数据集和三个空气质量数据集上进行了大量试验, 以充分评估本文所提方法的有效性, 数据集的详细信息如表 1 所示:

TZ2024 数据集包含北京市通州区 2024 年 1 月 5 日至 3 月 31 日的城市道路速度数据, 由 645 个道路传感器采集, 采样间隔为 5 分钟。PeMS04 和 PeMS08 数据集^[22] 由加州交通性能管理系统提供, 分别包含 307 和 170 个传感器的高速公路流量数据, 采样间隔同样为 5 分钟。KnowAir 数据集^[5] 涵盖 2015 年 1 月 1 日至 2018 年 12 月 31 日四年的 PM2.5 浓度数据, 覆盖 184 个城市, 采样间隔为 3 小时。BJAQI 和 GDAQI 数据集^[4] 包含 2014 年 5 月 1 日至 2015 年 4 月 30 日一年的空气质量数据,

表 1 数据集概要

数据集	时间跨度	时间间隔	时间步数	空间节点数量
TZ2024	2024/01/05—2024/03/31	5min	25056	645
PeMS04	2018/01/01—2018/02/28	5min	16992	307
PeMS08	2016/07/01—2016/08/31	5min	17856	170
KnowAir	2015/01/01—2018/12/31	3h	11688	184
BJAQI	2014/05/01—2015/04/30	1h	8760	274
GDAQI	2014/05/01—2015/04/30	1h	8760	163

分别覆盖 274 和 163 个空气质量站,采样间隔为 1 小时。

5.1.2 数据预处理与划分

对于 TZ2024、PeMS04 和 PeMS08 三个交通数据集,本文设置历史窗口和未来窗口大小均为 12,即利用历史 1 小时的数据预测未来 1 小时的时空序列,训练集、验证集和测试集按 7 : 1 : 2 的比例划分。对于 KnowAir 数据集,历史窗口设为 8,未来窗口设为 4,实现历史 24 小时到未来 12 小时的预测,训练集、验证集和测试集按原文推荐的 2 : 1 : 1 划分。BJAQI 和 GDAQI 数据集的历史窗口和未来窗口分别设为 12 和 6,以实现短期时空预测,数据集划分沿用 7 : 1 : 2 的比例。此外,本文对所有数据集进行了 Z 分数标准化,以统一数据尺度,提升训练效率和模型稳定性。

5.1.3 实验细节

实验针对本文方法在每个数据集上进行了超参数搜索,并将隐藏层维度设为 $d_h = 64$,时间记忆单元数设为 $m = 64$,时间嵌入维度设为 $d_t = 64$,空间嵌入维度设为 $d_s = 64$,时空交互维度设为 $d_c = 32$,元模式数设为 $K = 16$, $topk = 4$,正则化项系数设为 $\lambda_{reg} = 0.1$ 。对于三个交通数据集,选取月份、日期、时间步、周、是否为周末、是否为节假日作为时间先验,对于三个空气质量数据集,仅选取月份、日期、时间步和周作为时间先验。

此外,实验采用 AdamW^[44]作为优化器,学习率设为 0.001。损失函数选用 Huber loss^[45],因其较 MAE 更为鲁棒,并结合早停策略(耐心设为 20)选取最优权重进行测试,最大训练轮次设为 200 次。所有实验均在 NVIDIA Tesla V100S GPU 上进行,使用 PyTorch 框架^[46]。

5.1.4 评估指标

本文采用时空预测领域通用的度量方式,即平均绝对误差(Mean Absolute Error, MAE),均方根误差(Root Mean Square Error, RMSE),以及平均绝对百分比误差(Mean Absolute Percentage Er-

ror, MAPE)作为评估指标来衡量预测模型的性能,计算式如下:

$$\begin{aligned} MAE &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ MAPE &= \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \end{aligned} \quad (21)$$

其中, n 表示样本数量, y_i 和 \hat{y}_i 分别表示真实值和预测值。

5.2 对比实验与结果分析

为了评估本文所提方法的整体表现,将其与 9 种现有的时空预测模型进行对比,包括广泛采用的基线和最先进的模型:

(1) STGCN^[11] 通过时空卷积块整合图卷积和门控时间卷积来捕获时空依赖;

(2) GraphWaveNet^[12] 提出了自适应的邻接矩阵来捕获数据中隐藏的空间依赖,并通过堆叠的膨胀一维卷积来捕获时间依赖;

(3) AGCRN^[24] 通过节点自适应参数学习来捕获节点特定的模式,用统一的节点嵌入捕获潜在的空间依赖,并结合循环网络自动捕获时空相关性;

(4) DMSTGCN^[29] 通过构造以天为周期的自适应空间邻接矩阵来建模变化的空间相关性,将现有的图卷积扩展到动态图卷积;

(5) STSSL^[32] 在数据层面和拓扑结构层面分别对交通流图数据进行自适应增强,并通过辅助的自监督学习任务捕获时空异质性;

(6) MegaCRN^[40] 通过时空元图学习来捕获时空异质性,并结合元图上的图卷积和循环网络捕获时空相关性;

(7) TESTAM^[39] 通过混合专家模型为每个节点选取特定的空间建模方式,并结合时间增强的注意力模型捕获时空依赖;

(8) TGCRN^[30] 学习了能够感知周期性和趋势性的时变图结构,并结合门控循环单元联合捕获动态的时空依赖;

(9) HimNet^[25] 通过时空嵌入隐式地捕获时空异质性,并从元参数池中学习时空特定的参数。

为了保证实验的公平性,本文参照对比方法的官方代码库进行参数调优,基于原始论文的默认设置,在合理的参数空间内进行网格搜索,选择验证集上表现最优的参数组合,并以此参数设置在测试集上进行评估,确保所有对比方法在其最佳超参数设置下运

行。本文在六个真实世界数据集上进行了三次独立实验,并计算各项评估指标的均值作为模型最终预测

效果。如表2所示,STMP2G在所有数据集和评估指标上均显著优于所有对比方法,充分证明了其有效性。

表2 对比实验结果

数据集	评估指标	STGCN	GraphWaveNet	AGCRN	DMSTGCN	STSSL	MegaCRN	TESTAM	TGCRN	HimNet	STMP2G
TZ2024	MAE	3.3673	3.3644	3.4135	3.3989	3.4515	<u>3.3463</u>	3.422	3.3747	3.3536	3.2569
	RMSE	5.0924	5.0854	5.142	5.1743	5.1823	5.0772	5.2199	5.0804	<u>5.0746</u>	4.9735
	MAPE	10.2942	10.2829	10.402	10.4401	10.7414	<u>10.22</u>	10.4349	10.3091	10.257	9.9724
PeMS04	MAE	19.3812	18.9261	19.2033	20.3561	19.8567	19.0814	19.4594	19.8066	<u>18.6424</u>	18.0219
	RMSE	29.8266	<u>28.9847</u>	29.675	31.3527	30.0943	29.1666	29.4728	30.0197	29.206	28.1032
	MAPE	12.8503	12.6798	12.8752	13.8767	13.1662	12.5157	12.9235	13.6532	<u>12.3569</u>	11.9054
PeMS08	MAE	16.8966	15.6352	16.0796	15.3847	16.6355	15.6374	15.635	16.781	<u>15.383</u>	13.8364
	RMSE	25.5645	24.0486	24.329	24.0552	25.042	23.7418	24.3125	25.2231	<u>23.6824</u>	22.1409
	MAPE	11.0082	10.3578	10.7471	9.9208	10.1982	<u>9.8438</u>	11.3711	11.3653	10.2204	9.0568
KnowAir	MAE	12.615	12.9254	13.1515	13.3816	12.5545	<u>12.528</u>	14.1526	13.8408	12.7136	12.1016
	RMSE	19.8722	20.1974	20.0765	20.6932	19.6164	<u>19.5352</u>	21.7387	21.4599	19.9273	19.0377
	MAPE	<u>32.9993</u>	35.328	38.4106	37.5283	34.4903	35.7253	38.6702	41.2108	35.9345	31.195
BJAQI	MAE	20.3012	<u>19.5245</u>	21.5468	20.3283	20.7732	20.1075	22.2716	22.4652	20.8381	18.3357
	RMSE	33.881	<u>33.1811</u>	34.5915	34.1822	34.4595	33.4858	36.8001	35.8109	34.5837	32.159
	MAPE	<u>34.1698</u>	34.5461	39.1352	37.1613	37.5391	37.2237	40.0052	40.7145	37.7068	32.3516
GDAQI	MAE	7.7395	6.8103	8.4797	<u>6.7793</u>	7.8471	6.8816	8.3049	7.9618	7.4905	6.602
	RMSE	12.5682	11.6488	13.8879	<u>11.6135</u>	12.8863	11.6459	13.1728	12.8378	12.1789	11.3695
	MAPE	28.2064	23.9046	31.5456	<u>23.5801</u>	28.1894	24.6053	28.9435	30.201	26.845	22.4561

注:加粗表示最优解,其他表同。

其中,STGCN 基于预定义图卷积捕获空间依赖,但其固定图结构难以建模节点间的潜在关联。GraphWaveNet 能够通过自适应的邻接矩阵来学习隐藏关系,但其学习到的空间依赖仍然是静态的,无法随不同时间步发生变化。DMSTGCN 为一天中的每个时间步都学习一个邻接矩阵来建模动态的空间关系,但其忽略了长范围的时间周期与特殊事件的影响。TGCRN 通过时间感知图结构学习动态捕捉了空间依赖的趋势性和周期性变化,但其忽略了节点特定模式的差异性,从而降低了预测精度。AGCRN 为每个空间节点分配了独立的参数空间以学习节点特定模式,TESTAM 通过混合专家模型为每个空间节点选取了特定的空间建模方式,但它们都无法适应时空关系的动态变化。尽管以上方法在特定场景下表现较好,但由于未能有效区分和建模时空异质性,导致在复杂的时空场景中性能受限。

为了克服时空异质性的干扰,STSSL 构建了辅助的自监督学习任务以增强交通模式表示,但其对时间异质性的绝对假设以及参照所有空间区域共性模式的方式削弱了其建模能力,且自监督学习与主任务完全端到端的联合优化存在困难。MegaCRN 利用记忆网络构建时空元节点库来捕获时空异质

性,但由于未对时间和空间做出有效的区分性约束,导致建模单一特性的能力被削弱。HimNet 通过时空嵌入隐式地捕获了时空异质性,并借助元参数池学习时空特定模式,但由于引入了大量的冗余参数,导致优化困难和过高的计算和内存成本,限制了其在大规模时空场景中的应用。此外,上述方法未能充分关注元模式中的局部差异与动态变化,导致在捕捉局部时空异质性方面表现不足。

与上述方法相比,STMP2G 通过提示学习生成了多尺度的时间提示与语义化的空间提示,并将其有机融合,充分捕获了全局时空异质性;同时构建了独立的自监督学习任务,从历史时空序列中挖掘具有判别能力的元模式,以增强对局部时空异质性的建模能力,二者相互补充共同引导了时空建模过程,从而克服时空异质性的干扰。因此,STMP2G 展现出更优越的预测表现,充分证明了其在实际应用中的有效性和鲁棒性。

5.3 组件有效性研究

为了验证 STMP2G 中关键组件的有效性,本文设计了如下变体模型并在交通数据集 TZ2024 和空气质量数据集 KnowAir 上进行了详细的实验分析:

(1) base model: 该模型删除了时空提示引导模块,代表仅捕获时空依赖但无法捕获时空异质性的

基础模型,以验证本文提出的时空提示引导方法能否有效克服时空异质性的干扰从而提升模型性能;

(2) w/o \mathbf{O}_t : 该模型删除了时间提示生成模块,用全为 1 的张量替换时间提示张量,以验证多尺度的时间提示能否有效捕获时间异质性从而提升模型性能;

(3) w/o \mathbf{E}_s : 该模型删除了空间提示生成模块,用全为 1 的张量替换空间提示张量,以验证语义化空间提示能否有效捕获空间异质性从而提升模型性能;

(4) w/o \mathbf{W}_c : 该模型删除了时空提示融合模块,用全为 1 的张量替换时空交互张量,以验证自适应的时空提示融合能否有效建模时空交互作用从而提

升模型性能;

(5) w/o \mathbf{E}_p : 该模型删除了自监督元模式挖掘模块,用全为 1 的张量替换元模式张量,以验证元模式能否有效增强时空提示对时空建模的引导能力从而提升模型性能;

(6) joint: 该模型将元模式的挖掘过程和预测网络的优化过程同时进行,以验证采用二阶段的优化策略是否能够提高学习效率和特征质量。

(7) w/o Adapter: 该模型删除了逐层适配器,用相同的时空提示对各层时空表征进行引导,以验证逐层适配器能否有效增强时空提示在不同层次特征空间的适应性和表征能力从而提升模型性能。

表 3 STMP2G 及变体模型的预测结果

模型	TZ2024			KnowAir		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
base model	3. 7402	5. 3741	11. 3809	14. 2917	21. 8661	43. 6437
w/o \mathbf{O}_t	3. 3285	5. 0642	10. 1987	12. 2419	19. 4523	32. 7956
w/o \mathbf{E}_s	3. 3419	5. 0995	10. 1912	12. 3147	19. 2891	32. 4178
w/o \mathbf{W}_c	3. 2911	5. 0216	10. 1864	12. 2168	19. 2039	33. 8724
w/o \mathbf{E}_p	3. 2812	5. 0329	10. 0845	12. 3926	19. 4762	32. 4741
joint	3. 2983	5. 0546	10. 0929	12. 4411	19. 5715	32. 4912
w/o Adapter	3. 3626	5. 0833	10. 3891	12. 7952	19. 6547	36. 2915
STMP2G	3. 2569	4. 9735	9. 9724	12. 1016	19. 0377	31. 195

表 3 展示了上述变体与本文所提 STMP2G 在两个数据集上的实验结果对比。从表中可以看出,STMP2G 的性能相较于 base model 有明显提升,这充分证明了本文所提方法能够利用时空提示引导建模过程,克服时空异质性的干扰,从而提升时空预测的准确性。

w/o \mathbf{O}_t 与 w/o \mathbf{E}_s 都出现了明显的性能下降,证明了时间提示生成模块通过多尺度的先验嵌入与记忆网络查询有效捕获了时间异质性,提高了预测模型对不同时间段变化模式的预测精度,同时,空间提示生成模块通过自适应的节点嵌入与空间表征学习克服了空间异质性的干扰,提高了预测模型对不同空间区域个性模式的预测精度。此外,w/o \mathbf{W}_c 的性能下降也表明时空提示融合模块能够捕捉到时空异质性的耦合特性,有效建模复杂的时空交互作用。

与 STMP2G 相比,w/o \mathbf{E}_p 的性能不佳充分证明了本文挖掘得到的元模式能够有效捕捉时空数据中隐含的局部趋势和动态变化特性,从而增强时空提示对时空建模的引导能力。同时,从 joint 的实验结果可以看出,混合训练方式的效果相较于两阶段训练存在明显下降,这是由于元模式挖掘过程与预

测网络优化过程的目标不同,混合训练方式会带来任务之间的干扰,使模型学习到次优特征,从而降低学习效率和特征质量。

不同网络层次的特征空间具有差异性,逐层适配器能够将时空提示与当前层的特征分布对齐,平衡不同层次的学习目标。去除该组件将显著降低时空提示的适应性和表征能力,因此 w/o Adapter 的性能急剧下降。以上实验分析充分证明,STMP2G 是一个完整且不可分割的整体,各组件相互协作共同实现了精准的时空预测。

5.4 超参数敏感性研究

为了验证 STMP2G 对关键超参数的敏感性,本文在 TZ2024、PeMS04、PeMS08 和 KnowAir 数据集上针对六个关键超参数进行了系统的实验分析,包括隐藏层维度 d_h 、时间嵌入维度 d_t 、空间嵌入维度 d_s 、时空交互维度 d_c 、元模式数量 K 以及时间记忆数量 m 。实验中,在修改一个特定超参数的同时,保持其他超参数的默认值,实验结果如图 6 所示。其中,预测误差随着隐藏层维度、时间和空间嵌入维度的增加而降低,这表明复杂的时空数据需要更大的特征空间来捕获

其固有的异质性特征。同时,适当地增加时空交互维度有助于充分建模时空交互作用从而有效融合时空提示,但其过大会引入较多的噪声从而降低模型性能。此外,时间记忆单元数反映了模型存储长期记忆的能力,其稳定的误差曲线说明模型已对当前数据集具备足够的记忆能力。最后,元模式数量反映了模型区分和建模局部时空

异质性的能力,过小的值将导致模型无法捕捉数据中的重要模式,过大的值可能导致模型捕捉到数据中的噪声或无关模式,从而降低泛化能力。总体而言,模型性能受超参数变化的影响较小,这表明本文方法具有较强的鲁棒性,能够在不同设置下保持稳定性能,从而在显著降低人工调参成本的同时保证实际应用中的高可用性和可靠性。

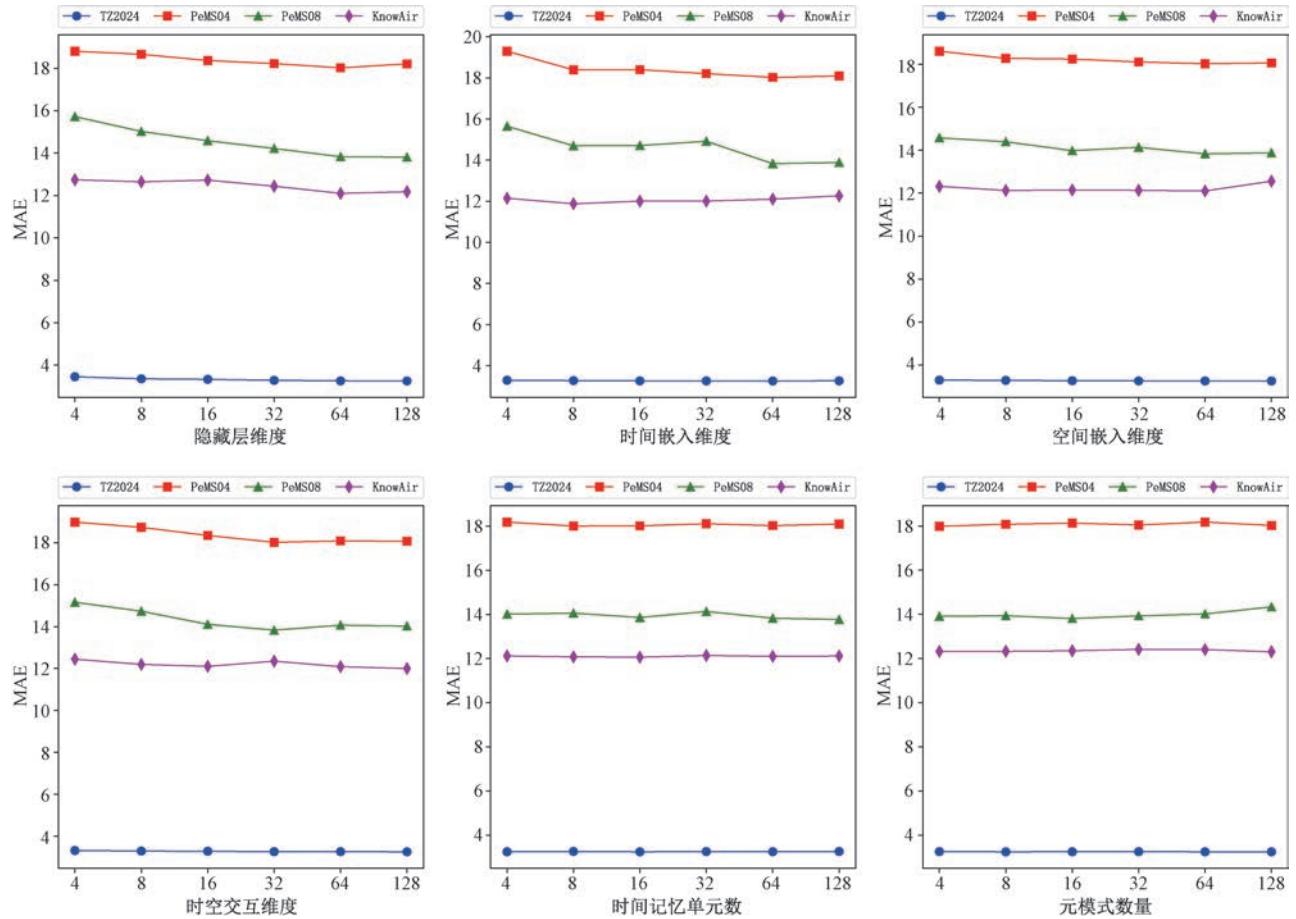


图 6 超参数敏感性研究

5.5 可解释性研究

为了验证 STMP2G 的可解释性,本文针对 TZ2024 数据集中异质性的时间段和空间区域,将生成的时空提示进行了详细的可视化分析,并与其他先进的时空预测方法进行了对比。

首先,本文基于 t-SNE^[47]降维方法对一天中不同时间段生成的时间提示进行了可视化分析,如图 7(a)所示,凌晨(红色)、早高峰(橙色)、日间平峰(黄色)、晚高峰(绿色)和深夜(紫色)等时间段形成了紧密而有显著区分性的簇,并且凌晨与深夜时段对应的簇较为接近,高峰与平峰时段对应的簇明显分离,这与实际的行为规律高度一致。图 7(c)展示了周末与工作日对应的降维可视化结果,其中左下方对

应周六和周日,右上方对应周一至周五,它们各自紧密成簇同时相互清晰分离,这充分反映了周末与工作日之间的显著差异。同时,图 7(b)和图 7(d)分别展示了各对比模型在不同时间段和一周内的预测表现,从图中可以看出,STMP2G 能够在各种异质性的时间范围内始终保持优越的预测性能,且具有较高的稳定性。此外,图 7(f)展示了 2024 年 2 月每一天的时间提示之间的余弦相似度,从图中可以看出,周末与工作日的区分非常明显,并且相邻日期之间较为相似,同时中间区域(即 2 月 10 日至 2 月 17 日)整体呈现出高度的相似性,而这一时期恰好对应春节假期,与真实世界的交通模式相一致。与本文方法相比,如图 7(e)所示,HiNet 仅能够捕获一周

内的时间差异。上述结果充分证明, STMP2G 能够通过生成多尺度的时间提示有效捕获复杂的时间异质性,甚至揭示一个月内的宏观时间趋势以及节假日带来的特殊变化,从而确保预测的准确性和鲁棒性。

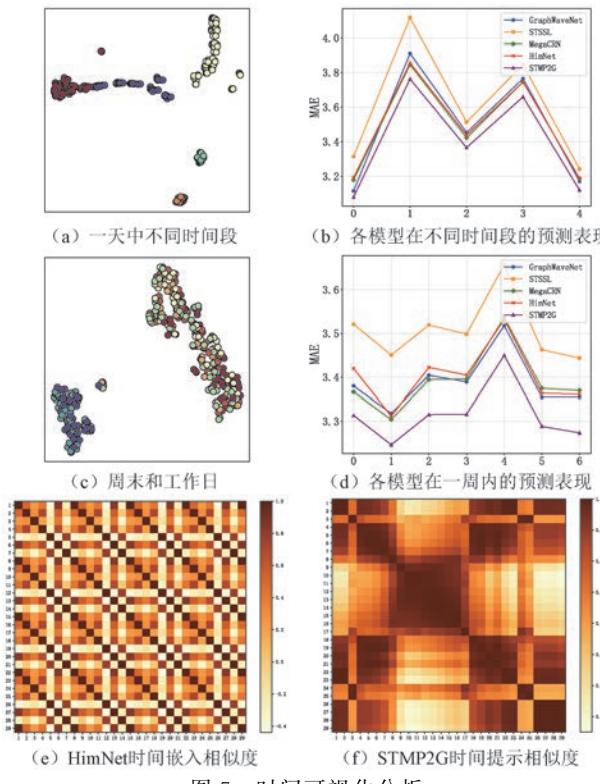


图 7 时间可视化分析

之后,本文对不同的空间区域生成的空间提示进行了聚类降维和可视化分析,如图 8(a)所示,不同簇的空间提示之间具有显著的区分性,并在同一簇内紧密聚集。图 8(c)绘制了各簇节点在一周内的平均速度曲线,不同簇的节点呈现出显著的模式差异,与聚类分离效果高度一致。此外,如图 8(b)所示,STMP2G 在不同簇的预测表现始终优于各对比模型。上述结果充分证明,STMP2G 能够通过生成语义化的空间提示有效捕获复杂的空间异质性,从而确保预测的准确性和鲁棒性。

由于元模式是特征空间中具有判别能力的潜在特征模式,直接对其可视化不利于展示其效果。因此,本文通过训练一个简单的自编码器,将 TZ2024 数据集中挖掘得到的元模式张量(以 $K = 32$ 为例)映射回原始的时空序列,并进行了可视化。如图 9 所示,各元模式反映了具体的局部趋势信息,通过元模式检索得到与输入时空序列最相关的元模式,能够预知未来时空序列的局部趋势性变化,从而与捕获全局时空异质性的时空提示相互补充,并进一步

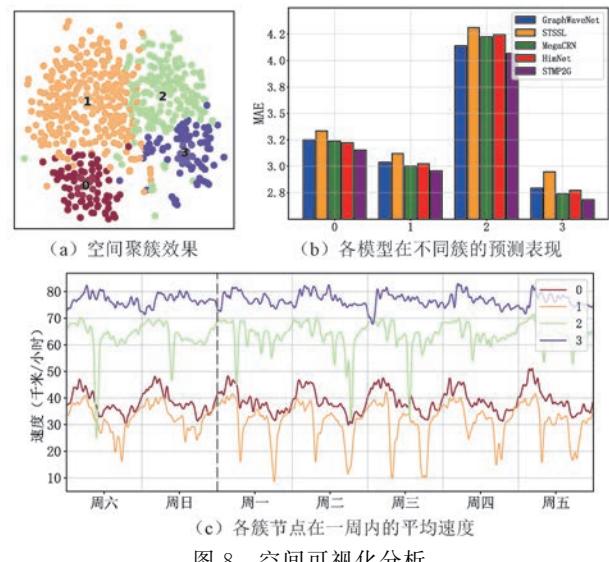


图 8 空间可视化分析

增强其引导能力。尽管各元模式有一定的区分,但仍能发现有些序列区分性不明显,这说明过高的 K 值导致模型捕获了重复的局部趋势信息,应针对不同数据集的规律特性谨慎设置 K 值。

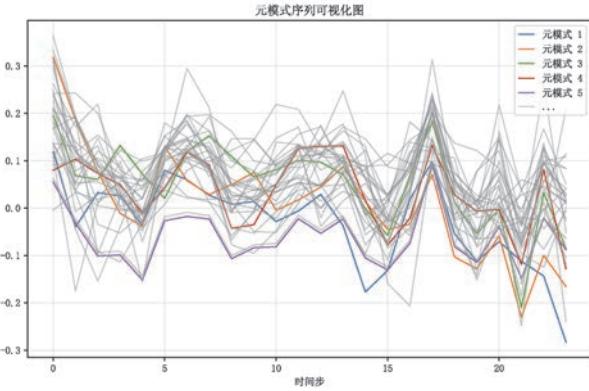


图 9 元模式可视化分析

5.6 效率研究

为了验证 STMP2G 在实际时空预测任务中的有效性和效率,本文在规模最大的 TZ2024 数据集上对所有方法进行了效率研究,系统对比了各模型的推理时间和内存占用情况。如图 10 所示,图中圆圈的大小直观地反映了各模型在推理过程中的内存占用情况。实验结果表明,本文提出的方法不仅在精度上达到了最低的误差水平,同时在推理效率和内存消耗方面也显著优于所有对比方法。在推理速度上,本文方法(1.47s)比 HimNet(73s)快了近 50 倍;在内存占用上,本文方法(2550MB)也比 HimNet(16990MB)、TGCRN(18160MB)、TESTAM(15462MB)等最先进的模型低得多。这一方面得益于本文方法使用简单的 MLP 模型作为骨干网

络,不包含图卷积、注意力或序列模型等复杂的计算单元;另一方面得益于轻量化的时空提示引导模块帮助预测模型克服了时空异质性的干扰,从而在最小的内存占用和时间消耗下展现出最优的性能。

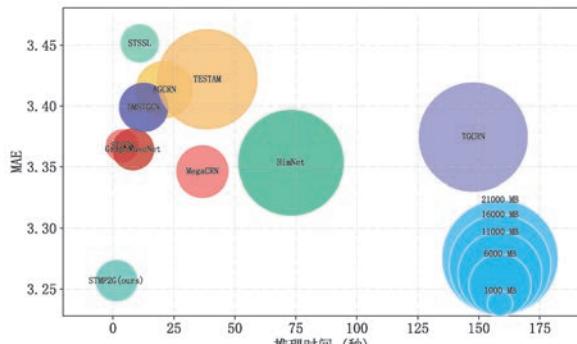


图 10 各模型的推理时间和内存占用

此外,本文考虑历史窗口长度 P 、未来窗口长度 Q 以及空间节点数量 N 等与问题紧密相关的参数,分析了 STMP2G 各模块的时间复杂度,以从理论上证明其计算复杂性优势。时空提示生成模块的时间复杂度为 $O(N(P+Q))$, 自监督元模式挖掘模块的时间复杂度为 $O(N^2+P+Q)$, 时空提示引导模块的时间复杂度为 $O(N(P+Q))$, 因此, STMP2G

的总体时间复杂度为 $O(N^2+P+Q+N(P+Q))$, 而推理时间复杂度为 $O(N(P+Q))$, 具有线性复杂度,显著优于主流的时空预测方法。

5.7 泛化性与鲁棒性研究

为了进一步验证 STMP2G 的泛化性与鲁棒性,本文基于 TZ2024 数据集构造了多种数据稀疏场景,以模拟实际应用中可能遇到的数据缺失和不完整性。具体而言,本文设计了两种稀疏化策略:一是随机删除部分时间步数据,以模拟时间维度上的数据缺失;二是随机删除部分空间节点数据,以模拟空间维度上的数据不完整性。通过这两种稀疏化操作,本文构建了不同稀疏比例(从 0% 到 30%,间隔 5%)的实验场景,以全面评估模型在数据稀疏情况下的性能表现。

在实验设计上,本文将 STMP2G 与现有方法(HimNet^[25]、TESTAM^[39]、MegaCRN^[40])进行了对比分析。如图 11 所示,随着数据缺失率的增加,STMP2G 与其他方法的性能差距不断变大,充分证明了本文方法在处理缺失数据时的强大泛化能力与鲁棒性,为实际应用中的复杂数据场景提供了可靠的技术支持。

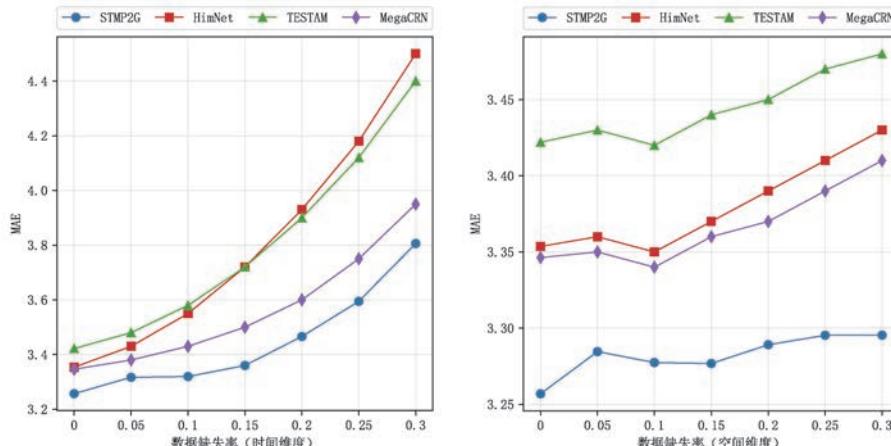


图 11 稀疏数据场景下的表现评估

6 结 论

本文针对时空数据中固有的异质性对时空建模带来的挑战,提出了一种基于元模式挖掘与提示引导的时空预测方法 STMP2G。该方法通过生成多尺度时间提示信息和语义化的空间提示信息来捕获时空异质性,并通过挖掘历史数据中具有判别能力的元模式来增强时空提示,最后利用逐层适配器对时空提示进行适应性映射,从而增强了时空预测模

型的表征能力,有效克服了异质性干扰,实现了更准确且更鲁棒的预测。在多个真实世界的交通和空气质量数据集上进行的大量实验表明,STMP2G 在不同时空场景下都展现了卓越的性能,验证了通过提示引导时空建模过程解决时空异质性问题的有效性,这为时空预测领域提供了新的研究视角。进一步的综合研究证明,STMP2G 在可解释性、效率、泛化性和鲁棒性上表现优异,在复杂多变的异质性时空场景中依然能够保持高效稳定的预测性能。未来我们将结合持续学习技术,开发自适应的提示信息

调整机制,使模型能够在时空数据异质性特征随时间变化时动态调整,从而进一步提升在复杂、变化环境中的预测能力。

参 考 文 献

[1] Rahmani S, Baghbani A, Bouguila N, et al. Graph neural networks for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(8): 8846-8885

[2] Lin H, Gao Z, Xu Y, et al. Conditional local convolution for spatio-temporal meteorological forecasting//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtual, 2022: 7470-7478

[3] Ji J, He J, Lei M, et al. Spatio-temporal transformer network for weather forecasting. *IEEE Transactions on Big Data*, 2025, 11(2): 372-387

[4] Zheng Y, Yi X, Li M, et al. Forecasting fine-grained air quality based on big data//Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015: 2267-2276

[5] Wang S, Li Y, Zhang J, et al. Pm2. 5-gnn: a domain knowledge enhanced graph neural network for pm2. 5 forecasting//Proceedings of the 28th International Conference on Advances in Geographic Information Systems. Seattle, USA, 2020: 163-166

[6] Zhao Y, Deng P, Liu J, et al. Generative causal interpretation model for spatio-temporal representation learning//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach, USA, 2023: 3537-3548

[7] Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017: 1655-1661

[8] Yao H, Wu F, Ke J, et al. Deep multi-view spatial-temporal network for taxi demand prediction//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 2588-2595

[9] Yao H, Tang X, Wei H, et al. Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019: 5668-5675

[10] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018: 1-16

[11] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 3634-3640

[12] Wu Z, Pan S, Long G, et al. Graph wavenet for deep spatial-temporal graph modeling//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019: 1907-1913

[13] Wu Z, Pan S, Long G, et al. Connecting the dots: multivariate time series forecasting with graph neural networks//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual, USA, 2020: 753-763

[14] Elman J L. Finding structure in time. *Cognitive science*, 1990, 14(2): 179-211

[15] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*, 1997, 9(8): 1735-1780

[16] Cho K, van Merriënboer B, Güneş Ç, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014: 1724-1734

[17] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324

[18] Van den Oord A, Dieleman S, Zen H, et al. Wavenet: a generative model for raw audio//Proceedings of the 9th ISCA Speech Synthesis Workshop. Sunnyvale, USA, 2016: 125-125

[19] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*: 1803.01271, 2018

[20] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering//Proceedings of the 30th Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 3844-3852

[21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Proceedings of the 31st Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 5998-6008

[22] Guo S, Lin Y, Feng N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019: 922-929

[23] Zheng C, Fan X, Wang C, et al. Gman: a graph multi-attention network for traffic prediction//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, USA, 2020: 1234-1241

[24] Bai L, Yao L, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting//Proceedings of the 34th Advances in Neural Information Processing Systems. Held virtually, 2020: 17804-17815

[25] Dong Z, Jiang R, Gao H, et al. Heterogeneity-informed meta-parameter learning for spatiotemporal time series forecasting//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain, 2024: 631-641

[26] Pan Z, Liang Y, Wang W, et al. Urban traffic prediction from spatio-temporal data using deep meta learning//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 1720-1730

[27] Pan Z, Zhang W, Liang Y, et al. Spatio-temporal meta learning for urban traffic prediction. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(3): 1462-1476

[28] Ye X, Fang S, Sun F, et al. Meta graph transformer: a novel framework for spatial-temporal traffic prediction. *Neurocomputing*, 2022, 491: 544-563

[29] Han L, Du B, Sun L, et al. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Virtual, Singapore, 2021: 547-555

[30] Ma M, Hu J, Jensen C S, et al. Learning time-aware graph structures for spatially correlated time series forecasting//Proceedings of the 2024 IEEE 40th International Conference on Data Engineering. Utrecht, The Netherlands, 2024: 4435-4448

[31] Wang Y, Xu Y, Yang J, et al. Fully-connected spatial-temporal graph for multivariate time-series data//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Washington, USA, 2024, 38(14): 15715-15724

[32] Ji J, Wang J, Huang C, et al. Spatio-temporal self-supervised learning for traffic flow prediction//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA, 2023: 4356-4364

[33] Gao H, Jiang R, Dong Z, et al. Spatial-temporal-decoupled masked pre-training for spatiotemporal forecasting//Proceedings of the 33rd International Joint Conference on Artificial Intelligence. Jeju, Republic of Korea, 2024: 3998-4006

[34] Hamilton J D, Susmel R. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 1994, 64(1-2): 307-333

[35] Stock J H, Watson M W. Vector autoregressions. *Journal of Economic perspectives*, 2001, 15(4): 101-115

[36] Williams B M, Hoel L A. Modeling and forecasting vehicular traffic flow as a seasonal arima process: theoretical basis and empirical results. *Journal of Transportation Engineering*, 2003, 129(6): 664-672

[37] Vanajakshi L, Rilett L R. A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed//Proceedings of the IEEE Intelligent Vehicles Symposium, 2004: 194-199

[38] Davis G A, Nihan N L. Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering*, 1991, 117(2): 178-188

[39] Lee H, Ko S. Testam: a time-enhanced spatio-temporal attention model with mixture of experts//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria, 2024: 1-19

[40] Jiang R, Wang Z, Yong J, et al. Spatio-temporal meta-graph learning for traffic forecasting//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA, 2023: 8078-8086

[41] Deng J, Deng J, Jiang R, et al. Learning gaussian mixture representations for tensor time series forecasting//Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Macao, China, 2023: 2077-2085

[42] Tucker L R. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966, 31(3): 279-311

[43] Kullback S, Leibler R A. On information and sufficiency. *The annals of mathematical statistics*, 1951, 22(1): 79-86

[44] Loshchilov I, Hutter F. Decoupled weight decay regularization// Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA, 2019: 1-10

[45] Huber P J. Robust estimation of a location parameter// *Breakthroughs in statistics: Methodology and distribution*. New York: Springer, 1992: 492-518

[46] Paszke A, Gross S, Chintala S, et al. Automatic differentiation in pytorch//Proceedings of the 31st Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 1-4

[47] Van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2008, 9 (11): 2579-2605



PANG Shuang-Shuang, Ph. D. candidate. His research interests include spatio-temporal data mining, traffic flow prediction and continual learning.

DENG Pan, Ph. D. ,associate professor. Her research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

ZHAO Yu, Ph. D. candidate. His research interests include spatio-temporal data mining, traffic flow prediction

and causal learning.

LIU Jun-Ting, Ph. D. candidate. His research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

WANG Zi-Ang, Ph. D. candidate. His research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

LI Si-Rui, M. S. candidate. Her research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

Background

Spatio-temporal prediction is a critical research area that focuses on forecasting future trends across specific spatial and temporal dimensions by analyzing historical spatio-temporal data. Accurate predictions provide scientific support for decision-making processes, facilitating improvements in social management and resource allocation. Due to its significant applications in fields such as traffic management, weather forecasting, and environmental monitoring, spatio-temporal prediction has garnered extensive attention in recent years.

Deep learning techniques, with their strong feature learning capabilities, have become the foundation for many spatio-temporal prediction methods. These methods typically model spatial dependencies using graph structures, where nodes represent spatial units and edges encode their relationships. Temporal dependencies are often captured as time series on graph nodes, employing recurrent neural networks and their variants or convolutional neural networks and their extensions. Additionally, graph convolutional networks are utilized to model local spatial dependencies, while attention mechanisms further enhance the capture of spatio-temporal correlations. Despite these advancements, existing methods primarily focus on global average characteristics and assume uniform patterns across spatial regions and temporal periods, overlooking the inherent heterogeneity of spatio-temporal data. This limitation significantly reduces the models' ability to capture localized variations and dynamic changes, leading to suboptimal predictive performance.

Spatio-temporal heterogeneity refers to the uneven and distinctive distribution and patterns of data across different spatial regions and temporal periods. For instance, in traffic

scenarios, various types of roads (e.g., highways, commercial streets, residential roads) exhibit significantly different speed distributions. Existing models often ignore these spatial heterogeneities, treating all regions uniformly, which undermines their ability to capture differentiated spatio-temporal patterns effectively. Methods attempting to address spatial heterogeneity, such as assigning independent parameter spaces to regions or leveraging meta-learning, face challenges like high computational costs, overfitting, and dependence on auxiliary spatial information.

Similarly, temporal heterogeneity is characterized by distinct patterns in different time periods, such as weekday rush hours or weekend shopping peaks. Existing methods struggle with temporal heterogeneity, as they rely on globally shared parameters or static modeling of dynamic dependencies, making them incapable of capturing multi-scale variations. Techniques like dynamic adjacency matrices or self-supervised tasks have been proposed but are limited by fixed periodicity assumptions, high computational overhead, or optimization challenges.

To address these challenges, this paper introduces ST-MP2G (Spatio-Temporal Meta-Pattern Mining and Prompt Guidance), a novel framework designed to capture spatio-temporal heterogeneity comprehensively. STMP2G generates spatio-temporal prompts to capture heterogeneity, mines meta-patterns to enhance the prompts, and leverages them to guide the spatio-temporal modeling process. The proposed method achieves accurate predictions across various spatio-temporal datasets and maintains robust performance in heterogeneous temporal periods and spatial regions.