

基于跨城市知识迁移的时空图小样本学习

夏志宇

摘要

时空图学习是城市计算任务中的一种重要方法，如在交通流量、出租车需求和空气质量预测等任务中均有应用。由于数据收集的成本很高，一些新兴城市的可用数据很少，这使得训练一个性能优异的模型变得不可行。为了应对这一挑战，近年来的一些工作提出了通过跨城市知识转移的相关方法进行学习的模型，即使用从数据充足的的城市学到的模型来帮助数据稀缺的城市的学習过程。然而，不同城市之间的时空图显示出不规则的结构和不同的特征，限制了此类小样本学习的相关方法的可用性。针对上述不足，此次课程复现研究的文章 *Spatio-Temporal Graph Few-Shot Learning with Cross-City Knowledge Transfer* 中，作者团队提出了一个与模型无关的时空图小样本学习框架 ST-GFSL。具体来说，为了通过转移跨城市的知识来增强特征提取，ST-GFSL 建议基于节点级的元知识来生成非共享参数。目标城市的节点通过参数匹配转移知识，从相似的时空特征中进行检索。此外，文章还提出了利用元知识重建图结构，从而限制元知识的学习方向。文章在四个交通速度预测数据集上进行了综合实验，结果表明 ST-GFSL 的性能优于当前最先进的方法。在文章算法基础上，本复现工作进一步优化 ST-GFSL 模型，针对原模型的一个具体模块 STNN 中的多步预测器加入了一个多层感知机，使得复现改进后的模型在 MAE 指标对比原文设计基础上再获得超过 5% 的性能提升。

关键词：时空数据；小样本学习；图神经网络；城市计算

1 引言

随着城市化的快速发展，城市中的人类、车辆和智能设备产生了相当多的时空数据，这些数据极大地改变了城市管理的方式，出现了许多城市相关的机器学习应用，如交通流量^[1-2]、出租车需求^[3-4]和空气质量预测^[5]任务等。然而，现有的机器学习算法需要大量的样本数据来学习有效的算法模型，这种条件可能不适用于没有足够数据的新兴城市。城市的相似性启发我们去考虑应用跨城市的知识转移方法来减轻数据收集的负担，从而提高智慧城市建设的效率。

目前，在解决少数场景下的城市计算任务方面已经取得了很多研究进展。Wang 等人^[6]将城市建模为网格，并首次提出在数据稀缺的城市促进时空预测。为了实现更好的类似区域间的匹配，他们引入了大规模的辅助数据（社交媒体签到数据）。而寻找和收集适当的辅助数据本身是有成本的，并面临着风险泄漏的可能性。在相关工作^[7]中，作者提出 FLORAL，通过从一个数据丰富的城市转移语义相关的字典来进行空气质量分类。然而，由于两个城市之间的巨大差异，从一个单一来源城市的知识转移面临着负面转移的风险。Yao 等人^[8]结合元学习方法，从目标领域的多个源城市学习一个良好的初始化模型，但是没有考虑到不同城市和城市内部的不同特征差异。更重要的是，上述方法只适用于基于网格的数据，但与基于图的建模不兼容。事实上，基于图的模型最近在学术界引起了广泛的关注，并已经在道路网络、地铁网络、传感器网络数据的时空学习方面取得了巨大的成功。

在复现的研究中，文章目标是在基于图的小样本学习场景中转移城市间的知识，同时探索知识转移对多个城市的影响。然而，该存在以下两个挑战：**(1) 如何通过来自多个源城市的知识在目标城市进行特征提取？**目前的元学习方法假定可转移的知识是在多个城市之间可全局共享的。然而，即使在一个城市的不同区域，其时空特征也可能有很大差异。现有的方法不能有效地处理复杂场景间的知识转移。**(2) 如何协调不同图结构对不同城市间知识转移的影响？**与基于网格的数据相比，基于图的建模反映出城市间不规则的结构信息。节点间的边明确地描述了各种特征的相互作用。现有的小样本学习方法在知识转移时往往会忽略结构的重要性，进而导致模型结果不稳定，甚至可能出现结构偏差。

为了解决上述挑战，该研究提出了一个新颖的、与模型无关的时空图小样本学习框架 ST-GFSL。该工作是第一个研究时空图学习中的小样本学习场景。为了适应多个城市的多样性，ST-GFSL 不再像往常一样学习全局共享的模型。而是根据节点级的元知识生成非共享的模型参数，从而加强特定特征的提取能力。城市间的知识转移是通过参数匹配来实现的，即对源城市和目标城市中具有相似时空特征的节点进行匹配。此外，文章提出了 ST-Meta 学习器，从地图结构和时间序列的相关性中学习节点级的元知识。在元学习的过程中，ST-GFSL 基于元知识重建不同城市的图结构。图重建的 Loss 被定义为指导结构学习的误差，从而避免在多个源城市之间出现图结构偏差。原文通过在四个公共城市数据集上应用交通速度预测来证明所提出的框架的优越性。研究在 METR-LA、PEMS-Bay、Didi-Chengdu 和 Didi-Shenzhen 数据集上进行了广泛的实验验证了所提出的方法比最先进的基准方法更有效、更通用。

在本次复现任务中，基于作者在原文给出的代码，我复现了 ST-GFSL 在上述相关数据集中的性能表现结果，并且进一步地通过实验探究了原作者对于元知识嵌入维度超参数取值设置的原因。此外，在原作工作的基础上，本次复现工作针对原论文所设计的 ST-GFSL 模型中一个时空神经网络 (STNN) 模块进行优化，为其中的多步预测器从原本的一个线性层 (Linear Layer) 加入了一个多层感知机 (MLP)，使得改进后的模型在 MAE 指标对比原模型再获得超过 5% 的性能提升，优化后的模型性能表现证明了多层感知机对于 ST-GFSL 模型在时空神经网络中针对预测任务的有效性。

2 相关工作

2.1 时空图学习

时空图学习是城市计算任务中一个基本的、被广泛研究的问题。早期，研究人员从时间序列分析的角度研究该问题，并提出了一系列方法，如 ARIMA、VAR 和 Kalman 滤波等^[9]。随着深度学习和图神经网络的兴起，图作为一种描述空间结构关系的有效数据结构，被应用于分析一系列城市问题。Bai 等人^[10]提出 STG2Seq，以图为基础建立多步骤的城市客运需求模型。Do 等人^[11]利用车辆上的物联网设备来感知城市空气质量，并通过变异图自动编码器估计未知的空气污染物。特别地，Pan 等人^[12]提出利用深度元学习，通过概括不同区域的学习能力来提高交通预测性能。

然而，上述论文中的研究都是基于一个拥有大规模训练数据的城市。数据稀缺的情况并不在研究范围内，但这是一个非常值得研究的问题。在复现的研究中，论文的目标是通过跨城市的知识转移来实现时空图小样本学习。此外，该文章也致力于提出一个与模型无关的架构，这样可以实现与最新的时空图学习模型相结合，从而进一步提高性能。

2.2 图小样本学习

小样本学习在计算机视觉和自然语言处理领域已经取得了重大突破。例如，MAML^[13]、原型网络^[14]和元迁移学习^[15]。然而，在非欧几里得度量域中的小样本学习，如图小样本学习，则还没有得到充分的探索。在最近的图小样本学习相关工作中，Meta-GNN^[16]是第一个将元学习范式 MAML 纳入图节点分类任务中。然而，它并没有完全描述节点之间的相互关系。Liu 等人^[17]提出了图节点的相对位置和绝对位置的概念，从而更好地捕捉节点之间的依赖关系。Yao 等人^[18]采用了原型网络的思想，通过寻找最相类似的类原型来进行小样本的节点分类。

通过对上述工作的分析，可以发现现有的方法主要集中在图小样本节点分类上，而许多城市计算问题是回归问题。此外，与一般的静态图相比，时空图具有更加复杂和动态的节点特征。直接将小样本学习方法与原始的 GNN 模型结合起来，对于捕捉复杂的节点关联特征是不可行的。

2.3 城市间的知识迁移

知识转移解决了数据匮乏场景下的机器学习问题。特别是在城市计算任务中，如何实现跨城市的知识转移，以减少数据收集的成本并提高学习效率是一个正在被学界研究得火热的问题。FLORAL^[7]是一项早期的工作，它通过从一个城市现有的足够多的多模态数据和标签知识迁移来实现对空气质量的分类。RegionTrans^[6]通过将城市划分为不同的网格进行时空特征匹配来进行跨城市的知识转移。Yao 等人^[8]首次提出了使用 MetaST 来转移多个城市之间的知识。

尽管如此，上述研究并不能直接应用于文章所针对的任务，其主要原因如下：(1) 正如前述，许多城市计算问题是回归问题，而 FLORAL 是为分类问题设计的。(2) RegionTrans 和 MetaST 是为基于网格的数据设计的，这与本任务中基于图的建模并不兼容。同时，在匹配过程中，RegionTrans 引入了额外的社交媒体打卡数据，这降低了其通用性。(3) FLORAL 和 RegionTrans 只关注单一源城市的知识转移。如何利用来自多个城市的数据并避免负面知识转移是一个值得研究的问题。在此研究中，作者旨在从多个基于图的数据集中学习跨城市的元知识，并在不引入辅助数据集的情况下转移到一个数据稀缺的城市。

3 本文方法

文章采用惯常的时空图表示方法： $\mathcal{G}_{ST} = (\mathcal{V}, \mathcal{E}, A, X)$ 。其中， $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ 表示节点集， $N = |\mathcal{V}|$ 是节点数。 $\mathcal{E} = \{e_{ij} = (v_i, v_j) \subseteq (\mathcal{V} \times \mathcal{V})\}$ 表示边集。 $A = \{a_{ij}\} \in \mathbb{R}^{N \times N}$ 是时空图的邻接矩阵， $a_{ij} = 1$ 表示节点 v_i 与 v_j 之间有一条边；否则， $a_{ij} = 0$ 。 X 是节点特征矩阵，指图上传递的信息，如一段时间内的交通速度、空气污染物浓度或出租车的客流量。我们把在时间上观察到的节点特征作为图信号 $X^t \in \mathbb{R}^{N \times d}$ ，其中 d 是节点特征的维度。

3.1 问题定义

问题 1：时空图预测

假设我们有 T 个历史时空图信号，我们想预测未来的 M 个图信号。预测任务可被表述为学习一个函数 $f(\cdot)$ ，在给定一个时空图 \mathcal{G}_{ST} ，可实现

$$[X^{t-T+1}, \dots, X^t; \mathcal{G}_{ST}] \xrightarrow{f(\cdot)} [X^{t+1}, \dots, X^{t+M}] \quad (1)$$

问题 2：城市时空图小样本学习

设有一组时空图数据丰富的源城市 $\mathcal{G}_{1:P}^{source} = \{\mathcal{G}_1^{source}, \dots, \mathcal{G}_P^{source}\}$ 和一个时空图数据稀缺的目标城市 \mathcal{G}^{target} 。在对 $\mathcal{G}_{1:P}^{source}$ 进行训练后，一个模型需要能够利用多个源城市的时空图元知识，完成对一个只有少量结构化时空图数据可用的目标城市 \mathcal{G}^{target} 的数据预测任务，这样的模型学习问题称为城市时空图小样本学习问题。

3.2 本文方法概述

本节将详细描述原文章所提出的 ST-GFSL 框架。此处首先给出 ST-GFSL 的框架，如图 1 所示。图的左边是 ST-GFSL 的输入。我们从多个城市进行知识迁移，而目标城市只有少量的训练样本。图的右边主要由两部分组成，即时空神经网络（STNN）模块和跨城市知识转移模块。具体来说，STNN 作为 ST-GFSL 的基础特征提取器，可以使用任何时空学习模型架构，如图神经网络（GNN）、循环神经网络（RNN）和其他最先进的时空图学习模型；而跨城市知识转移模块从多个源城市转移知识，这些城市在图 1 中以灰色虚线框出。针对这种跨城市的知识迁移，我们首先设计了 ST-Meta 学习器来获得空间和时间领域的节点级元知识，而后生成非共享的特征提取器参数 θ_{ST} 以便实现在源城市数据和目标城市数据之间的特征提取。ST-Meta 图的重建是通过重建不同城市的结构关系以进一步约束实现带有结构感知的元训练。ST-GFSL 的端到端学习过程遵循基于 MAML^[13] 的训练方法，即通过模仿目标城市中的少量场景，对成批的少量训练任务进行采样，以获得具有强大适应能力的模型。

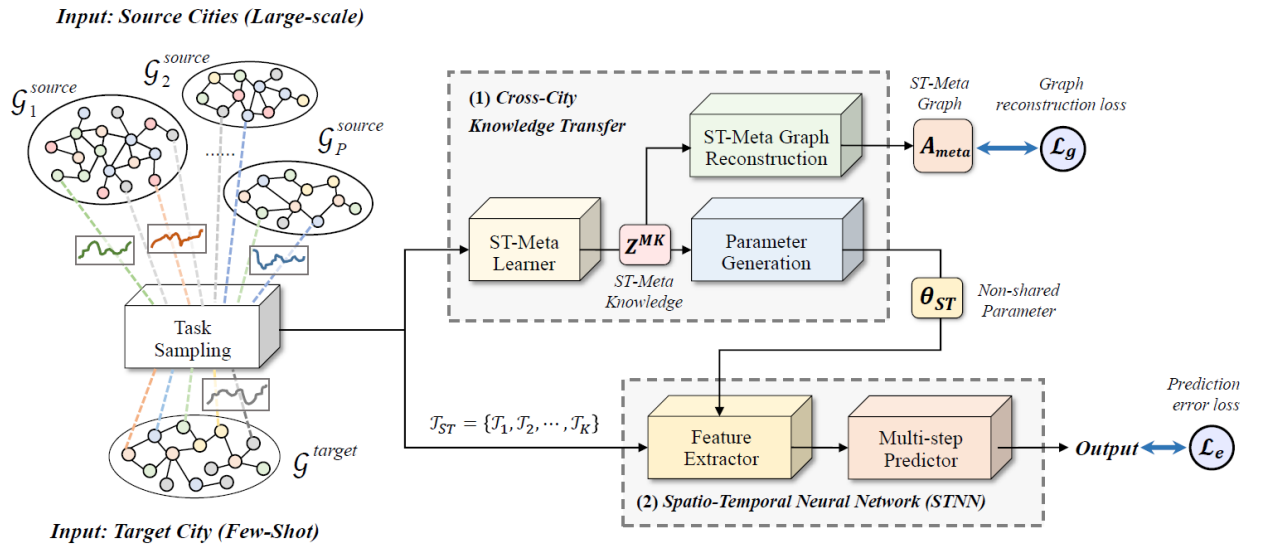


图 1: ST-GFSL 框架。左边是模型的输入，其中源城市有大规模的训练样本，而目标城市的样本很少。右边是 ST-GFSL 的两个主要部分：（1）跨城市知识转移和（2）时空神经网络（STNN）。

3.3 时空神经网络

时空神经网络（Spatio-Temporal Neural Networks, STNN）可分为特征提取器和多步预测器，如图 1 底部的虚线框所示。多步预测器在原文设计中是一层线性层。特征提取器是根据不同的任务和数据特征设计的，如基于 RNN、基于 CNN 和基于 GNN 的模型。例如，在后续复现实验中，本文选择了经典的时间序列分析网络 GRU^[19]，除此以外，原文的公开源代码还实现了 STGCN、GWN 等优越的时空图神经网络模型。由于 ST-GFSL 是为模型无关的框架设计的，参数生成方法根据相应的模型结构自适应地生成非共享特征提取器参数 θ_{ST} 。因此，原文所提出的 ST-GFSL 模型能够从最新的时空图学习技术突破中受益并且应用于数据稀缺的场景。

3.4 时空元知识学习器

时空元知识学习器（ST-Meta Learner）在空间和时间领域提取节点级的元知识。为了对时空图的时间动态进行编码，原文采用了广泛用于时间序列建模的门控递归单元（GRU）^[19]。与经典的 RNN 模型相比，GRU 保留了提取长序列的能力，减轻了梯度消失或梯度爆炸的问题。以节点 v_i 为例，节点级的时间元知识 z_i^{tp} 被表示为 $h_{i,t}$ 的最终状态。而 $h_{i,t}$ 的生成过程如

$$z = \sigma(U^z x_{i,t} + W^z h_{i,t-1}), \quad (2)$$

$$r = \sigma(U^r x_{i,t} + W^r h_{i,t-1}), \quad (3)$$

$$c = \phi(U^c x_{i,t} + W^c (h_{i,t-1} \circ r)), \quad (4)$$

$$h_{i,t} = (1 - z) \circ c + z \circ h_{i,t-1}, \quad (5)$$

其中， $x_{i,t} \in \mathbb{R}^d$ 是节点 v_i 在时间 t 的输入向量， $h_{i,t-1}$ 是时间 $t-1$ 的隐藏状态。 $U^z, U^r, U^c \in \mathbb{R}^{d \times d'}$ 和 $W^z, W^r, W^c \in \mathbb{R}^{d' \times d''}$ 是加权矩阵， \circ 是元素乘， σ 是非线性激活函数 *sigmoid*， ϕ 是 *tanh*。这样，可以通过 GRU 模型得出一个城市的时间元知识，表示为 $\mathbf{Z}^{tp} = \{z_1^{tp}, z_2^{tp}, \dots, z_N^{tp}\} \in \mathbb{R}^{N \times d'}$ 。

而对于时空图的空间编码，原文使用了基于空间的图注意网络（GAT）进行特征提取。GAT 被视作为一个信息传递过程，其中信息可以直接从一个节点沿边传递到另一个节点。具体地，原文把得到的一个城市的空间元知识表示为 $\mathbf{Z}^{sp} = \{z_1^{sp}, z_2^{sp}, \dots, z_N^{sp}\} \in \mathbb{R}^{N \times d'}$ 。

通过整合时空特征，可以得到元知识，表示为 $\mathbf{Z}^{MK} = \{z_1^{MK}, z_2^{MK}, \dots, z_N^{MK}\} \in \mathbb{R}^{N \times d_{MK}}$ 。而后通过一个可学习的比率 γ ($\gamma \in \mathbb{R}^{d'}$) 对时间和空间元知识进行加权求和，以数据驱动的方式从空间或时间领域学习影响。与经典的串联方法相比，它更容易通过可学习的比率来适应跨城市数据的时空特征。同时，它减少了数据生成的参数量。具体地，元知识 \mathbf{Z}^{MK} 的计算方式如下

$$\mathbf{Z}^{MK} = W^\gamma (\gamma \circ \mathbf{Z}^{tp} + (1 - \gamma) \circ \mathbf{Z}^{sp}), \quad (6)$$

其中 $W^\gamma \in \mathbb{R}^{d' \times d_{MK}}$ 是元知识输出层的加权矩阵， d_{MK} 是元知识嵌入的维度。

3.5 ST-Meta 图的重建

为了学习图的结构信息，减少不同源数据分布造成的图结构偏差，ST-Meta 图通过元知识来重建图结构，从而达到结构感知学习的目的。具体地，原文通过设计预测节点 v_i 和 v_j 之间存在的边的概率，从而来保证图的结构、减少偏差。其方法是通过将学到的元知识 \mathbf{z}_i^{MK} 和 \mathbf{z}_j^{MK} 相乘，形如

$$p(a_{ij} | \mathbf{z}_i^{MK}, \mathbf{z}_j^{MK}) = \text{sigmoid}((\mathbf{z}_i^{MK})^T \cdot \mathbf{z}_j^{MK}), \quad (7)$$

这样，ST-Meta 图 \mathbf{A}_{meta} 的重建就可以表示为

$$\mathbf{A}_{meta} = \text{sigmoid}[(\mathbf{Z}^{MK})^T \cdot \mathbf{Z}^{MK}], \quad (8)$$

其中， $(\cdot)^T$ 表示矩阵的转置。

这样，为了指导元知识的结构感知学习，原文在训练过程中引入了 ST-meta 图 \mathbf{A}_{meta} 和原始邻接矩阵 \mathbf{A} 之间的图重建损失 \mathcal{L}_g ，其计算方法如下

$$\mathcal{L}_g = \|\mathbf{A}_{meta} - \mathbf{A}\|^2, \quad (9)$$

3.6 ST-GFSL 模型训练

ST-GFSL 的学习过程遵循基于 MAML^[13]的小样本训练学习过程，具体训练过程及整体损失函数设计可参照 MAML 设计或翻阅本次复现的原论文^[20]，两份工作均有详尽说明。

4 复现细节

4.1 与已有开源代码对比

原论文中作者在 GitHub 上公开了模型的源码供公众学习与验证其性能结果。在本次复现任务中，我的主要工作如下：

1、首先使用原作提供的代码，复现作者在原文中声称的结果，并且把此结果作为本次复现优化工作的性能基准 baseline；

2、针对元学习的特点，通过调整代码给出的参数设定，探究元学习中特有的元知识嵌入这一超参数的嵌入维度对于模型性能的影响，进而说明原作者设定该参数数值的原因；

3、在原作模型设计的基础上，针对 ST-GFSL 中的 STNN 模块进行改进优化，把模型中的多步预测器组件从单层线性层改为多层感知机（MLP），并与原设计进行对比试验，通过更优的性能结果说明本复现工作中的优化设计的有效性。

下面将对上述的第 3 点复现优化工作进行详细展开以阐述本优化设计的工作原理。

原论文中，为了与其他 baseline 模型进行公平比较，作者在 STNN 模块中的特征提取网络（Feature Extractor）组件使用了 GRU 模型，在原论文给出的源码中也可印证这一点。时空图经过特征提取网络 GRU 后获取到的知识嵌入被接入到一个多步预测器（Multi-step Predictor）组件中进行图节点预测，进而获得节点交通流量预测值，时空图数据处理流如原文架构图 1 模块（2）所示。但在考察源代码过程中，我发现原作仅使用了一个简单的线性层来处理特征提取网络所提取出来的知识，而后便输出预测结果。即若设 \mathbf{m} 为特征提取网络获得的时空图知识嵌入，时空交通流量预测结果 \mathbf{p} 即表示为

$$\mathbf{p} = \text{Linear}_{\theta}(\mathbf{m}) \quad (10)$$

其中， θ 为线性层的参数。原论文的设计如图 2 上半部分所示。

在时空图知识的学习中，知识嵌入包含了所有时空图的特征信息，信息量很丰富，信息熵高。我认为仅用一个单层线性层作为知识解码器不足以充分理解嵌入中的所有时空信息。因此，在本次复现优化任务中，我针对此 STNN 模块中的多步预测器组件进行改进优化，从单层从单层线性层分别改为 1 层、2 层的多层感知机（MLP）网络，具体改进设计如图 2 下半部分所示。若设 \mathbf{m} 为特征提取网络获得的时空图知识嵌入，时空交通流量预测结果为 \mathbf{p} ，则 2 层的 MLP 网络可表示为

$$\mathbf{x}_1 = \text{MLP}_{\theta_1}(\mathbf{m}) \quad (11)$$

$$\mathbf{x}_2 = \text{MLP}_{\theta_1}(\mathbf{x}_1) \quad (12)$$

$$\mathbf{p} = \text{Linear}_{\theta}(\mathbf{x}_2) \quad (13)$$

其中， θ_1 、 θ_1 、 θ 为网络参数， \mathbf{x}_1 、 \mathbf{x}_2 为 MLP 网络的中间件。在本次优化实验中，每层 MLP 网络的维度被设定为 32。

在复现实验中，我将分别探究 ST-GFSL 原始设计（Original GRU）、在 STNN 模块多步预测器中

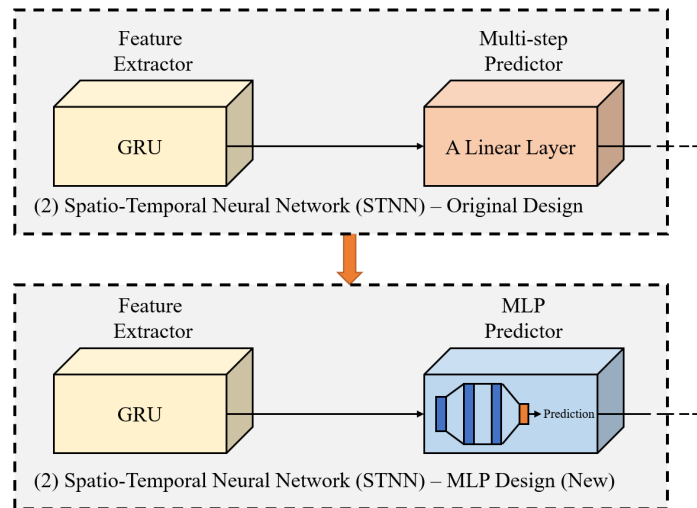


图 2: 对原模型设计 STNN 模块的改进（标记为 New）

加入 1 层 MLP 网络（GRU+1MLP）及在 STNN 模块多步预测器中加入 2 层 MLP 网络（GRU+2MLP）三个算法模型的性能表现。具体的性能测试结果与分析见小节 4.4。

4.2 实验环境搭建

从原作者提供的 GitHub 网址拉取模型代码至本地后，复现工作主要在一台配置如表 1 所示的机器中进行测试与优化。

表 1: 实验环境软硬件配置表

软硬件	配置信息
CPU	Intel i9-12900K@3.20 GHz × 16
内存	32 GB
GPU	NVIDIA GeForce RTX 3090
操作系统	Windows 11 version 22H2

文章的模型代码使用 Python 语言实现。原文所提供的开源代码中，原作者在 Readme.md 文件中详尽列出了模型复现所依赖的库，结合本机所搭建的环境，本次复现工作使用的语言及所依赖的主要库版本信息如下：

- Python = 3.9
- torch = 1.12.0
- numpy = 1.21.5
- scikit-learn = 1.0.2
- pytorch geometric = 2.1.0
- pyaml = 21.10.1

4.3 实验结果与分析

本节将对复现任务所完成的工作进行展述。各小节将分别展示 4.1 节所提到的各主要工作结果并对实验结果进行详尽分析。

4.3.1 原文模型复现结果与分析

本次复现工作首先对原作者在文章中所提出的模型进行模型性能的结果复现。实验结果如表 2 所示。本次复现过程中，我分别记录了 ST-GFSL 模型在原文 4 个不同的公开数据集（METR-LA、PEMS-Bay、Didi-Chengdu 和 Didi-Shenzhen）中未来 6 个时间片的交通流量数据预测 MAE 值与 RMSE 值。在获得实验结果后把实验所得的结果与原文声称的结果（如图 3 所示）作对比，其中，原文所展示的结果只有未来第 1、第 3、第 6 个时间片上的性能表现数据。通过与原文相同指标下的性能表现数据作对比可以发现，原作者所实现的模型确实可以达到文章所声称的性能表现，此结果说明了 ST-GFSL 模型在时空图小样本学习任务上的优越性能，展现出 ST-GFSL 设计针对时空图知识有着强大的知识迁移能力。

表 2: ST-GFSL 模型复现结果

Dataset		PEMS-BAY		METR-LA		Didi-Chengdu		Didi-Shenzhen	
STNN Model	Time Slot	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
GRU	1	1.301	2.100	2.664	4.181	2.230	3.234	1.926	2.793
	2	1.551	2.788	2.928	5.008	2.540	3.703	2.198	3.270
	3	1.761	3.398	3.265	5.639	2.683	3.922	2.342	3.550
	4	1.948	3.920	3.541	6.149	2.807	4.117	2.468	3.792
	5	2.119	4.375	3.790	6.588	2.925	4.305	2.586	4.007
	6	2.283	4.775	4.124	7.004	3.049	4.477	2.698	4.206

Baselines	PEMS-BAY Dataset						METR-LA Dataset					
	MAE (↓)			RMSE (↓)			MAE (↓)			RMSE (↓)		
	5 min	15 min	30 min	5 min	15 min	30 min	5 min	15 min	30 min	5 min	15 min	30 min
HA	4.373	4.373	4.373	6.745	6.745	6.745	6.021	6.021	6.021	9.483	9.483	9.483
ARIMA	2.019	2.307	2.429	3.929	4.648	5.360	2.900	3.058	3.369	4.179	5.279	7.670
Target-only	1.556	1.920	2.368	3.092	4.043	5.153	2.740	3.229	3.860	4.924	6.118	7.417
Fine-tuned (Vanilla)	1.823	2.166	2.590	3.434	4.280	5.276	2.757	3.277	3.900	4.883	6.123	7.413
Fine-tuned (ST-Meta)	1.371	1.791	2.277	2.699	3.747	4.920	2.647	3.188	3.800	4.368	5.759	7.110
AdaRNN [29]	1.248	1.928	2.749	2.084	3.796	5.725	2.513	2.897	3.312	4.298	5.567	6.732
MAML [14]	1.081	1.600	2.141	1.906	3.291	4.708	2.405	2.960	3.639	4.159	5.710	7.124
ST-GFSL (ours)	1.073	1.560	2.073	1.865	3.180	4.584	2.355	2.896	3.557	4.099	5.588	6.961

Baselines	Didi-Chengdu Dataset						Didi-Shenzhen Dataset					
	MAE (↓)			RMSE (↓)			MAE (↓)			RMSE (↓)		
	10 min	30 min	60 min	10 min	30 min	60 min	10 min	30 min	60 min	10 min	30 min	60 min
HA	3.438	3.438	3.438	4.879	4.879	4.879	2.955	2.955	2.955	4.342	4.342	4.342
ARIMA	2.825	3.305	4.317	3.889	4.253	5.597	2.888	3.056	3.596	4.489	4.764	5.575
Target-only	2.386	2.700	3.085	3.516	4.017	4.569	2.071	2.454	2.834	3.154	3.793	4.422
Fine-tuned (Vanilla)	2.586	2.877	3.246	3.746	4.213	4.751	2.117	2.490	2.867	3.196	3.831	4.442
Fine-tuned (ST-Meta)	2.240	2.693	3.083	3.249	3.956	4.519	2.033	2.454	2.850	2.989	3.719	4.385
AdaRNN [29]	2.260	2.724	3.036	3.231	3.942	4.324	2.107	2.473	2.807	3.041	3.674	4.231
MAML [14]	2.215	2.599	2.956	3.215	3.858	4.399	1.917	2.330	2.673	2.825	3.546	4.158
ST-GFSL (ours)	2.188	2.579	2.927	3.190	3.820	4.339	1.890	2.288	2.644	2.763	3.477	4.100

图 3: 原文中声称的模型性能（红色框高亮处）

4.3.2 元学习嵌入维度的选取探究

相比起其他深度学习模型，在元学习范式中，元学习特有的元知识嵌入是其他一般的机器学习范式中所未有的，它多用于指导具体的小样本学习模型的初始参数设置。好的元学习知识嵌入能够通过源知识域学习到的知识指导小样本学习模型快速训练至收敛，实现知识迁移目的，进而使得小样本学习模型在数据缺乏的条件之下仍然可以取得不错的性能表现。而如何恰当设置存储元知识的元学习嵌入维度这一超参数一直是困扰一线元学习方向研究人员的一个重要问题。基于 ST-GFSL 模型在元

学习上的出色性能表现，本节，我将通过设计步长实验探究原作者设置该超参数的原因。

在原作者提供的源码中，元学习嵌入维度被设定为 16。在本次复现探究过程中，我固定其他参数不变，调整元学习嵌入维度，范围分别为 {8, 16, 24, 32}，在 PEMS-BAY 数据集上开展实验，固定记录时间片为 3 的预测性能 MAE 值，获得的实验结果如图 4所示。

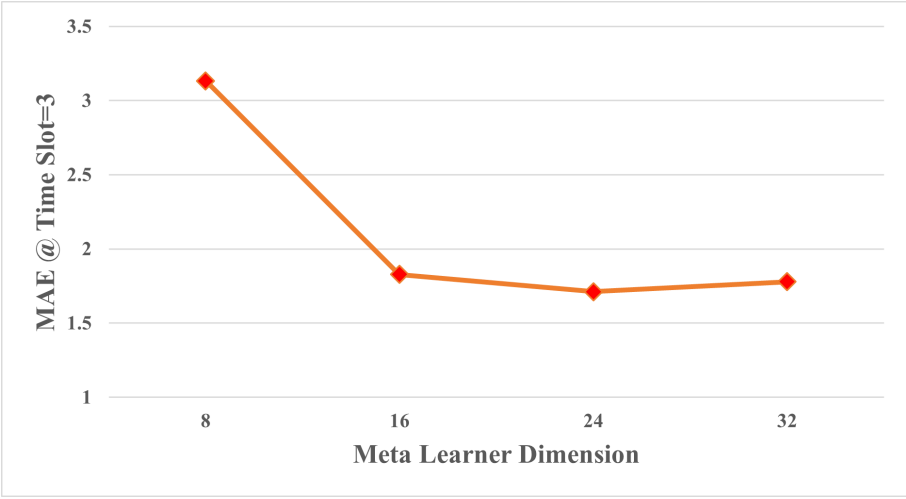


图 4: 元学习嵌入维度的选取探究实验结果

实验结果表明：在元学习嵌入维度为 8 时，低维向量可能不足以存储在源数据集上学习到的所有元知识，故最终的模型性能表现受到了一定影响，无法展现出模型的最佳性能；当元学习嵌入维度增大后，在 $Dimension = 16$ 时，嵌入维度足够存储元知识，故模型性能表现相比起在维度为 8 时能够取得更加的结果。但当模型元学习嵌入维度不断增加时，如 $Dimension > 16$ ，模型需要计算优化的参数会增多，但步长实验表明模型的性能没有明显的提高。故作者应该在平衡了性能与计算开销的利弊后，选择了元学习嵌入维度等于 16 作为模型的最终公开版本。

4.4 针对 ST-GFSL 模型的改进

如前述（小节 4.1），在 ST-GFSL 模型的基础之上，我针对其 STNN 模块进行改进优化，把模型中的多步预测器组件从单层线性层改为多层感知机（MLP），以求能够让其更好地处理特征提取网络在小样本数据上提取到的知识，最终使得改进后的模型在其基础上获得更优的性能表现。针对 STNN 模块的多步预测器组件，我分别设计了加入了 1 层 MLP（GRU+1MLP）与 2 层 MLP（GRU+2MLP）来处理特征提取网络喂入的嵌入，两个改进模型分别与 ST-GFSL 原始设计（Original GRU）作比较。本改进实验在 PEMS-BAY 数据集上开展，实验结果记录未来 6 个时间片的交通流量数据预测 MAE 值与 RMSE 值。实验结果如图 5 及表 3 所示。

对比图 5 中每个时间片簇的数据结果可以看出：ST-GFSL 及其改进版本在长时间片（如，时间片 = 6）预测上的性能表现不如最近的时间片（如，时间片 = 1），说明预测误差随着时间的推移，模型对于长时间的预测不如短时间内预测得准确，这可能与时间推移的误差累积或者近期知识不能很好地服务于未来长时间的预测任务有关。但总体而言，ST-GFSL 及改进版在未来一段时间内（时间片 $\in [1, 6]$ ）的预测性能还是十分优秀的，性能损失在可接受范围内。此外，在同一个时间片簇中，2 层 MLP 的多步预测器均能取得最佳的性能效果，说明在 ST-GFSL 模型的 STNN 模块中，针对组件多步预测器的改进给模型的整体性能表现带来了积极的影响。结合表 3 的改进后模型性能数据表明：在同一个时间片的预测任务中，2 层 MLP 的多步预测器的效果（加粗显示）相比起原始设计（下划线显示）在 MAE

性能指标下均有超过 5% 的性能提升。此结果说明 MLP 的改进设计能够更好地处理与理解特征提取网络提取到的时空图知识，从而进一步地优化原文模型 ST-GFSL 的预测性能。



图 5: 改进 ST-GFSL 模型的性能比较图

表 3: 改进 ST-GFSL 模型的性能数据结果

STNN Model Time Slot	Original GRU		GRU+1MLP		GRU+2MLP		Gains v.s. Origin
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	2.664	4.181	2.507	4.169	2.493	4.256	6.42%
2	2.928	5.008	2.802	5.052	2.776	5.083	5.19%
3	3.265	5.639	3.069	5.712	3.016	5.696	7.63%
4	3.541	6.149	3.261	6.230	3.257	6.227	8.02%
5	3.790	6.588	3.539	6.736	3.475	6.687	8.31%
6	4.124	7.004	3.661	7.109	3.658	7.069	11.30%

4.5 复现与改进的创新点

如前述，在本次复现任务中，主要工作与创新点体现在 1) 复现了原文所声称的模型性能结果，并将其作为改进模型的性能基准；2) 探究了原作者公开的源代码中元学习嵌入维度的设置原因与合理性；3) 针对原文模型中的 STNN 模块进行改进优化，把模型中的多步预测器组件从单层线性层改为多层感知机（MLP）以更好地处理与理解时空图知识，并设计相关实验，与原文所提出的模型进行比较，说明改进版模型在原模型的基础上进一步获得性能提升。

5 总结与展望

本部分对整个文档的内容进行归纳并分析目前实现过程中的不足以及未来可进一步进行研究的方 向。本次复现任务选取了在数据挖掘顶级会议 KDD 2022 上发表的一篇文章 Spatio-Temporal Graph Few-Shot Learning with Cross-City Knowledge Transfer 作为研究对象，文章针对小样本城市计算时空图学习的应用场景，提出了以元学习为基础的知识迁移小样本学习框架模型 ST-GFSL。在本次复现任务中，本报告复现了作者在论文中的模型性能实验结果，探究了元学习嵌入维度的设置原因与合理性，同时 在原作者设计的基础上，应用 MLP 网络改进了模型 STNN 模块中的多步预测器组件，改进后的模型性能与原设计对比在 MAE 指标再获得超过 5% 的性能提升。本次复现改进的工作表明 MLP 对理解特征提取网络所提取到的时空图知识有积极意义。

参考文献

- [1] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019: 1907-1913.
- [2] LU B, GAN X, JIN H, et al. Spatiotemporal adaptive gated graph convolution network for urban traffic flow forecasting[C]// Proceedings of the 29th ACM International conference on information & knowledge management. 2020: 1025-1034.
- [3] LIU Z, ZHANG H, OUYANG G, et al. Data-Driven Pick-Up Location Recommendation for Ride-Hailing Services[J]. IEEE Transactions on Mobile Computing, 2022.
- [4] LIU Z, GONG Z, LI J, et al. mT-Share: A Mobility-Aware Dynamic Taxi Ridesharing System[J]. IEEE Internet of Things Journal, 2021, 9(1): 182-198.
- [5] WANG C, ZHU Y, ZANG T, et al. Modeling inter-station relationships with attentive temporal graph convolutional network for air quality prediction[C]// Proceedings of the 14th ACM international conference on web search and data mining. 2021: 616-634.
- [6] WANG L, GENG X, MA X, et al. Cross-city transfer learning for deep spatio-temporal prediction[J]. arXiv preprint arXiv:1802.00386, 2018.
- [7] WEI Y, ZHENG Y, YANG Q. Transfer knowledge between cities[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1905-1914.
- [8] YAO H, LIU Y, WEI Y, et al. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction[C]// The World Wide Web Conference. 2019: 2181-2191.
- [9] MOREIRA-MATIAS L, GAMA J, FERREIRA M, et al. Predicting taxi-passenger demand using streaming data[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(3): 1393-1402.
- [10] BAI L, YAO L, KANHERE S S, et al. STG2seq: spatial-temporal graph to sequence model for multi-step passenger demand forecasting[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019: 1981-1987.
- [11] DO T H, TSILIGIANNI E, QIN X, et al. Graph-deep-learning-based inference of fine-grained air quality from mobile IoT sensors[J]. IEEE Internet of Things Journal, 2020, 7(9): 8943-8955.
- [12] PAN Z, LIANG Y, WANG W, et al. Urban traffic prediction from spatio-temporal data using deep meta learning[C]// Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 1720-1730.
- [13] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C]// International conference on machine learning. 2017: 1126-1135.
- [14] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[J]. Advances in neural information processing systems, 2017, 30.

- [15] SUN Q, LIU Y, CHUA T S, et al. Meta-transfer learning for few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 403-412.
- [16] ZHOU F, CAO C, ZHANG K, et al. Meta-gnn: On few-shot node classification in graph meta-learning [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 2357-2360.
- [17] LIU Z, FANG Y, LIU C, et al. Relative and absolute location embedding for few-shot node classification on graph[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 35: 5. 2021: 4267-4275.
- [18] YAO H, ZHANG C, WEI Y, et al. Graph few-shot learning via knowledge transfer[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 34: 04. 2020: 6656-6663.
- [19] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.
- [20] LU B, GAN X, ZHANG W, et al. Spatio-Temporal Graph Few-Shot Learning with Cross-City Knowledge Transfer[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 1162-1172.