连接点：用图神经网络进行多元时间序列预测

摘要

多元时间序列建模一直是一个吸引了经济学、金融学和交通学等多个领域研究者的课题。多元时间序列预测背后的一个基本假设是，其变量相互依赖，但仔细观察，可以公平地说，现有方法未能充分利用变量对之间潜在的空间依赖性。近年来，图形神经网络（GNNs）在处理关系依赖方面表现出了很强的能力。GNNs需要定义良好的图结构来进行信息传播，这意味着它们不能直接应用于依赖关系事先未知的多变量时间序列。本文提出了一种针对多元时间序列数据的通用图神经网络框架。该方法通过一个图学习模块自动提取变量间的单向关系，将外部知识如变量属性等集成到图学习模块中。提出了一种新的混合跳传播层和扩展的起始层来捕获时间序列中的空间和时间相关性。图学习、图卷积和时间卷积模块在端到端框架中联合学习。实验结果表明，我们提出的模型在4个基准数据集中的3个数据集上优于最新的基线方法，在提供额外结构信息的两个数据集上的性能与其他方法相当。

关键词

图神经网络，图结构学习，多元时间序列预测，时空图

1 介绍

现代社会得益于各种各样的传感器来记录温度、价格、交通速度、用电量和许多其他形式的数据。从不同传感器记录的时间序列可以形成多元时间序列数据，并且可以相互连接。例如，每日气温的升高可能会导致用电量的增加。为了捕捉一组动态变化的变量的系统趋势，多元时间序列预测问题已经研究了至少六十年。它在经济、金融、生物信息学和交通等领域有着巨大的应用。

多元时间序列预测方法内在地假设变量之间的相互依赖性。换句话说，每个变量不仅取决于它的历史值，还取决于其他变量。然而，现有的方法并不能有效地利用变量间潜在的相互依赖关系。统计方法，如向量自回归模型（VAR）和高斯过程模型（GP），假设变量之间存在线性依赖关系。统计方法的模型复杂度随变量个数的增加呈二次增长。他们面临着大量变量的过度拟合问题。最近开发的基于深度学习的方法，包括LSTNet和TPA-LSTM，对于捕捉非线性模式非常有效。LSTNet利用一维卷积神经网络将短期局部信息编码成低维向量，并通过递归神经网络对向量进行解码。TPA-LSTM通过递归神经网络处理输入，并采用卷积神经网络计算多个步骤的注意分数。LSTNet和TPA-LSTM没有显式地对变量间的成对依赖关系进行建模，这削弱了模型的可解释性。

图是描述不同实体之间关系的一种特殊形式的数据。近年来，图形神经网络由于其排列不变性、局部连通性和组合性，在处理图形数据方面取得了巨大的成功。通过结构传播信息，图神经网络允许图中的每个节点知道其邻域上下文。多元时间序列预测可以自然地从图形的角度来看待。多元时间序列中的变量可以看作是图中的节点，它们之间通过隐藏的依赖关系相互联系。因此，利用图神经网络对多变量时间序列数据进行建模是一种很有前途的方法，可以在充分利用时间序列之间的相互依赖性的同时保持其时间轨迹。

对于多元时间序列，最适合的图神经网络类型是时空图神经网络。时空图神经网络以多元时间序列和外部图形结构为输入，以预测多元时间序列的未来值或标签为目标。与不利用结构信息的方法相比，时空图神经网络取得了显著的改进。然而，由于以下挑战，这些方法仍然无法对多变量时间序列进行建模：

挑战1：未知的图形结构。现有的GNN方法在很大程度上依赖于预定义的图结构来进行时间序列预测。在大多数情况下，多元时间序列没有明确的图形结构。变量之间的关系必须从数据中发现，而不是作为基本真理知识提供。

挑战2：图形学习和GNN学习。即使有一个图结构可用，大多数GNN方法只关注于消息传递（GNN学习），而忽略了一个事实，即图结构不是最优的，应该在训练期间更新。接下来的问题是如何在端到端的框架中同时学习时间序列的图结构和GNN。

在本文中，我们提出了一种新的方法来克服这些挑战。如图1所示，我们的框架由三个核心组件组成：图学习层、图卷积模块和时间卷积模块。对于挑战1，我们提出了一种新的图学习层，该层基于数据自适应地提取稀疏图邻接矩阵。此外，我们开发了一个图卷积模块来处理变量之间的空间依赖关系，给出了由图学习层计算的邻接矩阵。这是专门为有向图设计的，避免了图卷积网络中经常出现的过平滑问题。最后，我们提出一个时间卷积模组，利用修正的一维卷积来撷取时间型态。它既能发现具有多个频率的时间模式，又能处理很长的序列。

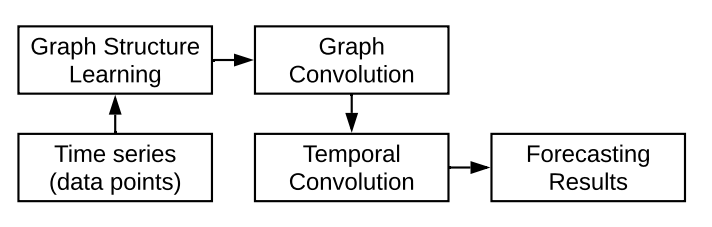


图1 我们提出的框架的概念图

由于所有参数都可以通过梯度下降来学习，因此该框架能够以端到端的方式同时对多变量时间序列数据建模和学习内部图结构（挑战2）。为了降低求解高度非凸优化问题的难度和减少处理大型图时的内存占用，提出了一种学习算法，该算法采用课程学习策略来寻找更好的局部最优解，并在训练过程中将多变量时间序列分解成子组。这里的优点是，我们提出的框架通常适用于大小图，短时间序列和长时间序列，有和没有外部定义的图结构。总之，我们的主要贡献如下：

据我们所知，这是第一次研究多元时间序列数据一般从基于图形的角度与图形神经网络。

我们提出了一个新的图形学习模块来学习变量之间隐藏的空间依赖关系。我们的方法为GNN模型在没有显式图结构的情况下处理数据打开了一扇新的大门。

我们提出了一个多变量时间序列数据建模和图形结构学习的联合框架。我们的框架比任何现有的时空图神经网络更通用，因为它可以处理有或没有预定义的图形结构的多变量时间序列。

实验结果表明，在4个基准数据集中的3个数据集上，我们的方法优于最先进的方法，在提供额外结构信息的两个流量数据集上，我们的方法的性能与其他GNNs相当。

2 背景

2.1 多元时间序列预测

时间序列预测的研究由来已久。现有的大多数方法都遵循统计方法。自回归综合移动平均（ARIMA）[1]推广了一系列线性模型，包括自回归（AR）、移动平均（MA）和自回归移动平均（ARMA）。向量自回归模型（VAR）扩展了AR模型来捕捉多个时间序列之间的线性相关性。同样，向量自回归滑动平均模型（VARMA）被提出作为ARMA模型的一个多变量版本。高斯过程（GP）作为一种贝叶斯方法，对多元变量在函数上的分布进行建模。GP可以自然地应用于多变量时间序列数据的建模[5]。尽管统计模型由于其简单性和可解释性在时间序列预测中得到广泛应用，但它们对平稳过程做出了很强的假设，并且不能很好地扩展到多元时间序列数据。基于深度学习的方法摆脱了静态假设，是捕捉非线性的有效方法。Lai等人[12]和Shih等人[19]是前两个基于深度学习的多变量时间序列预测模型。他们使用卷积神经网络来捕捉变量之间的局部依赖关系，使用递归神经网络来保持长期的时间依赖关系。卷积神经网络将变量间的相互作用封装成一个全局隐藏状态。因此，它们不能充分利用变量对之间的潜在依赖关系。

2.2 图神经网络

图神经网络在处理网络中实体间的空间依赖关系方面取得了巨大的成功。图神经网络假设一个节点的状态依赖于它的邻居的状态。为了捕捉这种空间相关性，人们通过消息传递[7]、信息传播[11]和图卷积[10]开发了各种各样的图神经网络。它们共享相似的角色，本质上通过将信息从节点的邻居传递到节点本身来捕获节点的高级表示。最近，我们看到了一种称为时空图神经网络的图神经网络的出现。这种形式的神经网络最初是为了解决交通预测[3,13,21,23,26]和基于骨架的动作识别[18,22]问题而提出的。时空图神经网络的输入是具有外部图形结构的多元时间序列，描述了多元时间序列中变量之间的关系。对于时空图神经网络，节点之间的空间依赖性通过图卷积来捕获，而历史状态之间的时间依赖性通过递归神经网络[13，17]或一维卷积来保持[22，23]。虽然现有的时空图神经网络与不使用图结构的方法相比有了很大的改进，但是由于缺乏一个预定义的图和一个通用的框架，它们不能有效地处理纯多元时间序列数据。

3 问题表述

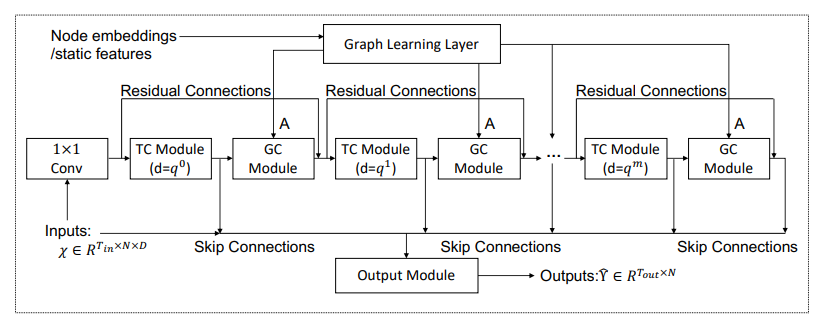
本文主要研究多元时间序列预测问题。设zt∈rn表示N维多元变量在时间步长t处的值，其中zt[i]∈R表示第i个变量在时间步长t处的值。给定一个多元变量X={zt1，zt2，···，ztP}的历史P时间步长序列，我们的目标是预测Y={ztP+Q}的Q步距值，或者未来值的序列Y={ztP+1，ztP+2，···，ztP+Q}。更一般地，输入信号可以与诸如一天中的时间、一周中的一天和季节中的一天之类的其他辅助特征耦合。将输入信号与辅助特征串接，假设输入为X={St1，St2，···，StP}，其中Sti∈rn×D，D为特征维数，Sti的第一列等于zti，其余为辅助特征。我们的目标是通过l2正则化最小化绝对损失来建立从X到Y的映射f（·）。

图描述网络中实体之间的关系。下面我们给出了图相关概念的形式化定义。定义3.1（图表）。图的形式为g=（V，E），其中V是节点集，E是边集。我们用N来表示图中的节点数。定义3.2（节点邻域）。设v∈v表示节点，e=（v，u）∈e表示从u到v的边。节点v的邻域定义为N（v）={u∈v |（v，u）∈e}。定义3.3（邻接矩阵）。邻接矩阵是图的一种数学表示形式，当（vi，vj）∈E和Aij=0时，用a∈rn×N表示，当（vi，vj）∈E时，Aij=0。我们使用图邻接矩阵来描述节点之间的关系。在大多数情况下，图邻接矩阵不是由多元时间序列数据给出的，而是由我们的模型来学习的。

4 MTGNN框架

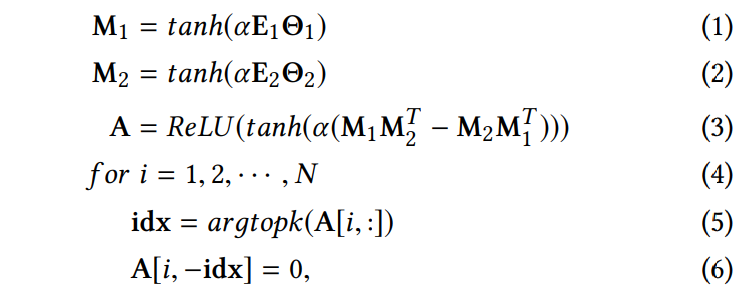
4.1 模型结构

我们首先阐述了模型的总体框架。如图2所示，最高级别的MTGNN由一个图学习层、m个图卷积GC模块、m个时间TC卷积模块和一个输出模块组成。为了发现节点之间隐藏的关联，图学习层计算一个图邻接矩阵，该矩阵随后被用作所有图卷积模块的输入。图卷积模块与时间卷积模块交织，分别捕获空间和时间依赖关系。图3演示了时间卷积模块和图卷积模块如何相互协作。为了避免梯度消失的问题，从时间卷积模块的输入到图卷积模块的输出中加入剩余连接。在每个时间卷积模块之后添加跳过连接。为了获得最终输出，输出模块将隐藏的特征投影到所需的输出维度。更详细地说，我们模型的核心组件如下所示：



4.2 图形学习层

图学习层自适应地学习一个图邻接矩阵来捕捉时间序列数据之间的隐藏关系。为了构造一个图，现有的研究通过一个距离度量来度量节点对之间的相似性，例如点积和欧氏距离[13]。这必然导致O（n2）的高时空复杂度问题。这意味着计算和存储成本随图形大小的增加呈二次增长。这限制了模型处理较大图形的能力。为了解决这个限制，我们采用了一种抽样方法，它只计算节点子集之间的成对关系。这就切断了每个迷你批处理中的计算和内存瓶颈。更多细节见第4.6节。另一个问题是现有的距离度量通常是对称的或双向的。在多变量时间序列预测中，我们期望一个节点状态的变化会引起另一个节点状态的变化，如交通流的变化。因此，学习的关系应该是单向的。我们提出的图形学习层专门用于提取单向关系，如下所示：



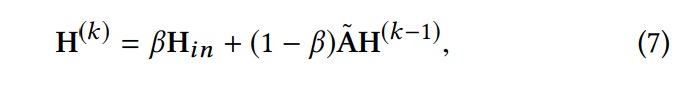
其中E1，E2表示在训练期间可学习的随机初始化的节点嵌入，Θ1，Θ2是模型参数，α是用于控制激活函数的饱和率的超参数，arДtopk（·）返回向量的前k个最大值的索引。我们提出的图邻接矩阵的非对称性质由方程3得到。减法项和ReLU激活函数使邻接矩阵正则化，使得Avu为正时，其对角对应的Auv为零。方程5-6是一种使邻接矩阵稀疏的策略，同时降低了后续图卷积的计算成本。对于每个节点，我们选择其前k个最近的节点作为其邻居。在保留连通节点权值的同时，将非连通节点的权值设为零。

合并外部数据。图形学习层的输入不限于节点嵌入。在已知每个节点属性的外部知识的情况下，我们还可以设置E1=E2=Z，其中Z是静态节点特征矩阵。一些工作考虑捕捉动态空间依赖[8，18]。换句话说，它们根据时间输入动态调整两个连接节点的权重。然而，当我们需要同时学习图的结构时，假设动态的空间依赖使得模型很难收敛。我们的方法的优点是我们可以在训练数据集期间学习稳定和可解释的节点关系。一旦模型在一个在线学习版本中被训练，我们的图邻接矩阵也可以随着新的训练数据更新模型参数而改变。

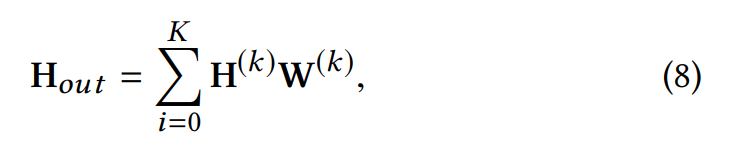
4.3 图卷积模块

图卷积模块的目的是将节点的信息与其邻居的信息进行融合，以处理图中的空间依赖关系。图卷积模块由两个mixhop传播层组成，分别处理通过每个节点的流入和流出信息。通过将两个混合跳传播层的输出相加得到净流入信息。图4显示了图卷积模块和混合跳传播层的体系结构。

混合跳跃传播层。在给定图邻接矩阵的情况下，我们提出了混合跳传播层来处理空间相关节点上的信息流。提出的混合跳传播层由信息传播和信息选择两个步骤组成。我们首先给出这两个步骤的数学形式，然后说明我们的动机。信息传播步骤定义如下：



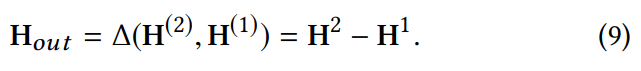
其中β是一个超参数，它控制保持根节点原始状态的比率。信息选择步骤定义如下：



其中K是传播深度，Hin表示前一层输出的输入隐藏状态，Hout表示当前层的输出隐藏状态，H（0）=Hin，A=D−1（A+I），DI=1+Íj Aij。在图4b中，我们展示了所提出的混合跳传播层中的信息传播步骤和信息选择步骤。它首先水平传播信息，然后垂直选择信息。

信息传播步骤沿着给定的图结构递归地传播节点信息。图卷积网络的一个严重限制是当图卷积层数趋于无穷大时，节点隐藏状态收敛到一个点。这是因为无论初始节点状态如何，多层图卷积网络都达到了随机游动的极限分布。为了解决这个问题，受Klicpera等人[11]的启发，我们在传播过程中保留了一部分节点的原始状态，这样传播的节点状态既可以保持局部性，又可以探索一个较深的邻域。然而，如果我们只应用公式7，一些节点信息将丢失。在不存在空间依赖的极端情况下，聚集邻域信息只会给每个节点增加无用的噪声。因此，引入了信息选择步骤，过滤掉每一跳产生的重要信息。根据等式8，参数矩阵W（k）用作特征选择器。当给定的图结构不包含空间依赖性时，方程8仍然能够通过将所有k>0的W（k）调整为0来保持原始节点自身信息。

与现有工程的连接。[9]和[2]探讨了混合跳跃的概念。Kapoor等人[9]将来自不同跃点的信息串联起来。Chen等人[2]提出了一种不同啤酒花间重量信息的注意机制。它们都应用GCN进行信息传播。然而，由于GCN面临着过度平滑的问题，来自更高跳数的信息可能不会对整体性能产生影响或产生负面影响。为了避免这种情况，我们的方法在本地信息和邻域信息之间保持平衡。此外，Kapoor等人[9]表明，他们提出的具有两个混合跳层的模型能够表示两个连续跳之间的差值。我们的方法可以达到同样的效果，只有一个混合跳传播层。假设K=2，W（0）=0，W（1）=1，W（2）=1，则



从这个角度来看，与级联方法相比，使用求和更有效地表示不同跳数的所有线性交互。

4.4 时间卷积模块

时间卷积模块采用一组标准的扩展一维卷积滤波器来提取高级时间特征。该模块由两个扩展的初始层组成。一个扩张的起始层后接一个正切双曲激活函数，起到滤波器的作用。另一层后面是一个sigmoid激活函数，它作为一个门来控制滤波器可以传递给下一个模块的信息量。图5显示了时间卷积模块和扩展的初始层的体系结构。

扩张的起始层。时间卷积模块通过一维卷积滤波器捕获时间序列数据的序列模式。为了得到一个既能发现不同范围的时间模式又能处理很长序列的时间卷积模块，我们提出了扩展的起始层，它结合了卷积神经网络中两种广泛应用的策略，即使用多种尺寸的滤波器[20]和应用扩展的卷积[24]

首先，对于卷积网络来说，选择合适的内核大小是一个具有挑战性的问题。滤波器的大小可能太大而不能精确地表示短期信号模式，或者太小而不能充分地发现长期信号模式。在图像处理中，一种广泛使用的策略称为初始（inception），它将具有三种不同核尺寸（1×1、3×3和5×5）的二维卷积滤波器的输出串联起来。从二维图像到一维时间序列，1×1、1×3和1×5滤波器大小的集合不适合时间信号的性质。由于时间信号趋向于具有若干固有周期，例如7、12、24、28和60，滤波器大小为1×1、1×3和1×5的初始层的堆栈不能很好地包含这些周期。或者，我们提出了一个由四个滤波器大小组成的时间起始层，即。1×2、1×3、1×6和1×7。上述时间段都可以通过这些过滤器尺寸的组合来覆盖。例如，为了表示周期12，模型可以将输入通过来自第一时间起始层的1×7滤波器，然后通过来自第二时间起始层的1×6滤波器。

第二，卷积网络的感受野大小随着网络的深度和滤波器的内核大小呈线性增长。考虑一个核大小为c的m 1D卷积层的卷积网络，卷积网络的感受野大小为，R=m(c-1)+1

要处理很长的序列，它需要非常深的网络或非常大的过滤器。我们采用扩展卷积来降低模型的复杂度。扩展卷积对具有特定频率的下采样输入操作标准卷积滤波器。例如，当膨胀因子为2时，它对每两步采样的输入应用标准卷积。在[14]之后，我们让每一层的膨胀因子以q的速率指数增加（q>1）。假设初始扩张因子为1，核大小为c的m层扩张卷积网络的感受野大小为

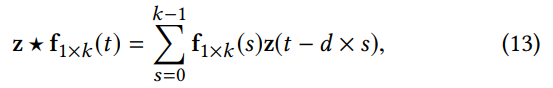


这表明网络的感受野大小也随着隐藏层数量的增加以q的速率呈指数增长。因此，使用这种扩张策略可以捕获比不使用它更长的序列。

形式上，结合初始和扩张，我们提出了扩张初始层，如图5b所示。给定一维序列输入z∈rt和由f1×2∈r2、f1×3∈r3、f1×6∈r6和f1×7∈r7组成的滤波器，扩张初始层的形式为，



其中，四个滤波器的输出根据最大滤波器被截断为相同的长度，并在整个信道维上串联，并且由z⋆f1×k表示的扩展卷积被定义为

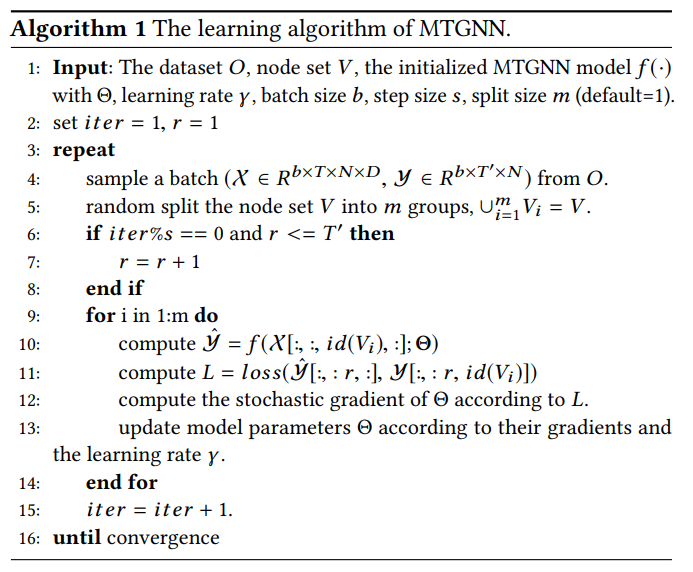


4.5 跳过连接层和输出模块

跳过连接层本质上是1×Li标准卷积，其中Li是到第i个跳过连接层的输入的序列长度。它将跳转到输出模块的信息标准化，使其具有相同的序列长度1。输出模块由两个1×1标准卷积层组成，将输入的通道维数转换为所需的输出维数。如果我们只想预测未来的某一步，那么期望的输出维是1。当我们想要预测Q个连续的步骤时，期望的输出维数是Q。

4.6 提出的学习算法

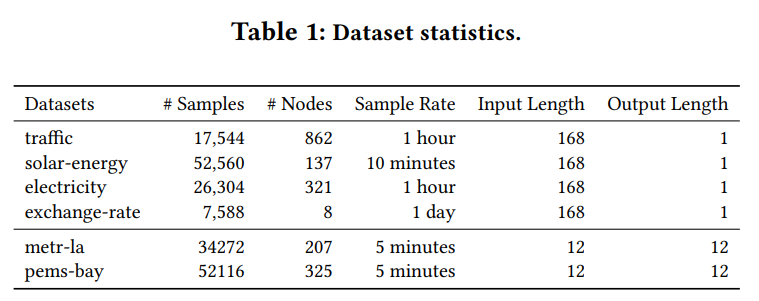
我们提出了一个学习算法来提高我们的模型处理大型图的能力和稳定在一个更好的局部最优。图上的训练通常需要将所有节点中间状态存储到内存中。如果一个图很大，它将面临内存溢出的问题。蒋等人[4]提出了一种子图训练算法来解决内存瓶颈问题。他们应用图聚类算法将图划分成子图，并在划分的子图上训练图卷积网络。在我们的问题中，基于节点的拓扑信息进行聚类是不现实的，因为我们的模型同时学习了潜在的图结构。或者，在每次迭代中，我们将节点随机分成若干组，并让算法根据采样的节点学习子图结构。这使得每个节点完全有可能被分配给一个组中的另一个节点，这样就可以计算和更新这两个节点之间的相似性得分。另一个好处是，如果我们将节点分成s组，我们可以在每次迭代中将图学习层的时间和空间复杂度从O（n2）降低到（N/s）2。在训练之后，由于所有的节点嵌入都经过良好的训练，因此可以构造一个全局图来充分利用空间关系。虽然邻接矩阵的计算代价很高，但在进行预测之前，可以并行地预先计算邻接矩阵。



我们所提出的算法的第二个考虑是使我们的模型稳定在一个更好的局部最优。在多步预测的任务中，我们观察到长期预测往往比短期预测在模型性能方面取得更大的改进。我们认为原因是我们的模型预测了多个步骤，而长期预测产生的损失要比短期预测高得多。因此，为了使总体损失最小化，该模型更注重提高长期预测的准确性。为了解决这个问题，我们提出了一个多步骤预测任务的课程学习策略。算法从解决最简单的问题开始，只预测下一步。为模型寻找一个良好的起点是非常有利的。随着迭代次数的增加，模型的预测长度逐渐增加，使得模型能够一步一步地学习硬任务。覆盖所有这些，我们的算法在算法1中给出。我们模型的进一步复杂性分析见附录A.1。

5 实验研究

我们在两个任务上验证了MTGNN-单步和多步预测。首先，我们在四个多变量时间序列预测的基准数据集上比较了MTGNN与其他多变量时间序列模型的性能，其目的是预测未来的一步。此外，为了说明MTGNN的性能，与其他使用预定义图形结构信息的时空图神经网络相比，我们在两个时空图神经网络的基准数据集上评估了MTGNN，其目的是预测未来的多个步骤。参数研究的进一步结果见附录A.4。



5.1 实验装置

在表1中，我们总结了基准数据集的统计数据。有关数据集的更多详细信息，请参见附录A.2。我们使用五个评估指标，包括平均绝对误差Mean Absolute Error（MAE）、均方根误差Root Mean Squared Error（RMSE）、平均绝对百分比误差Mean Absolute Percentage Error（MAPE）、相对平方根误差Root Relative Squared Error（RRSE）和经验相关系数Empirical Correlation Coefficient（CORR）。对于RMSE、MAE、MAPE和RRSE，值越低越好。对于CORR，值越高越好。其他实验装置见附录A.3。

5.2 比较基准方法

MTGNN和MTGNN+抽样是我们要评估的模型。MTGNN是我们提出的模型。MTGNN+抽样是我们提出的模型，在每次迭代中训练在一个图的抽样子集上。基线方法总结如下：

5.2.1 单步预测

•AR：自回归模型。

•VAR-MLP：多层感知（MLP）和自回归模型（VAR）的混合模型[25]。

•GP：高斯过程时间序列模型[6，16]。

•RNN-GRU：具有完全连接的GRU隐藏单元的递归神经网络。

•LSTNet：一种深度神经网络，结合了卷积神经网络和递归神经网络[12]。

•TPA-LSTM：一种注意力循环神经网络[19]。

5.2.2 多步预测

•DCRNN：一种扩散卷积递归神经网络，它将扩散图卷积与递归神经网络相结合[13]。

•STGCN：一种时空图形卷积网络，将图形卷积与一维卷积结合起来[23]。

•图形波网络：一种时空图形卷积网络，将扩散图卷积与一维扩展卷积相结合[21]。

•ST MetaNet：一种序列到序列架构，它利用元网络生成参数[15]。

•GMAN：一个具有时空注意力的图形多注意力网络[26]。

•MRA-BGCN：一种多范围注意双成分GCN[3]。

5.3 主要结果

表2和表3给出了MTGNN和MTGNN+取样的主要实验结果。我们观察到，MTGNN在大多数任务上都取得了最先进的结果，当它对子图进行抽样训练时，MTGNN的性能只会下降到很低的水平。接下来，我们分别讨论了单步预测和多步预测的实验结果

5.3.1单步预测

在这个实验中，我们比较了MTGNN与其他多元时间序列模型。表2显示了单步预测任务的实验结果。总的来说，我们的MTGNN在太阳能、交通和电力数据的几乎所有领域都取得了最先进的成果。特别是在流量数据方面，MTGNN在RSE方面的改进是显著的。MTGNN在流量数据的3、12、24区间内，RSE分别下降了7.24%、3.88%、4.83%。MTGNN之所以能显著地改善交通数据的结果，主要是因为交通数据的性质更适合我们的模型关于时空相关性的假设。显然，一条道路的未来交通占用率不仅取决于它的过去，还取决于它所连接的道路的占用率。MTGNN未能对汇率数据进行改进，这可能是由于图的大小较小，汇率数据的训练样本较少。

5.3.2多步预测

在这个实验中，我们比较了MTGNN与其他时空图神经网络模型。表3显示了多步预测任务的实验结果。MTGNN的意义在于，它在不使用预定义图的情况下实现了与最先进的时空图神经网络相当的性能，而DCRNN、STGCN和MRA-BGCN则完全依赖预定义图。图Wavenet提出了一种自适应邻接矩阵，但它需要与一个预定义的图相结合才能获得最佳的性能。ST MetaNet采用注意机制来调整预定义图的边权重。GMAN利用node2vec算法在执行注意机制的同时保留节点结构信息。当一个图没有定义时，这些方法不能有效地对多变量时间序列数据进行建模。

5.4 烧蚀研究

我们对METR-LA数据进行了消融研究，以验证有助于改进我们提出的模型结果的关键组件的有效性。我们将MTGNN命名为以下不含不同组件的名称：

•不带GC:MTGNN，不带图形卷积模块。我们用一个线性层来代替图卷积模块。

•w/o Mix-hop：在Mix-hop传播层中没有信息选择步骤的MTGNN。我们将信息传播步骤的输出直接传递给下一个模块。

•无起始层：扩张起始层中无起始层的MTGNN。在保持输出通道数相同的情况下，我们只使用一个1×7的滤波器。

•无CL:MTGNN，无课程学习。我们在不逐渐增加预测长度的情况下训练MTGNN。

我们将每个实验重复10次，每次重复50个周期，并在表4的验证集上报告MAE、RMSE、MAPE的平均值，标准偏差超过10次。图卷积模块的引入显著地改善了结果，因为它使信息在孤立但相互依赖的节点之间流动。混合跳的效果也很明显：验证了混合跳的使用有助于在混合跳传播层的每个信息传播步骤中选择有用的信息。初始效应对RMSE的影响显著，而对MAE的影响较小。这是因为在放大起始层的输出通道数不变的情况下，使用单个1×7滤波器比使用1×2、1×3、1×5、1×7滤波器的组合多出一半的参数。最后，我们的课程学习策略被证明是有效的。它使我们的模型能够快速收敛到一个最适合最简单任务的最优值，并随着学习难度的增加一步一步地微调参数。

5.5 图形学习层的研究

为了验证我们提出的图学习层的有效性，我们进行了一项研究，实验中使用了不同的方法来构造图邻接矩阵。表5显示了不同形式的A，在10次运行的METR-LA数据的验证集上测试了实验结果。预定义A由路网距离构成[13]。Global-A假设邻接矩阵是一个包含n2个参数的参数矩阵。受[21]的启发，无向A和有向A由节点嵌入的相似度得分计算。在[8,18]的激励下，Dynamic-A假设每个时间步的空间依赖性依赖于其节点输入。单向A是我们提出的方法。根据表5，我们提出的单向A达到最低的平均MAE、RMSE和MAPE。它比预定义A、无向A和动态A有显著的改进。我们的单定向A在MAE和MAPE方面比无定向A和定向A略有提高，但由于RMSE较低，证明更为稳健。

通过一个实例，我们进一步研究了学习的图邻接矩阵。在图6a中，我们绘制了节点55及其预定义的前3个邻居的原始时间序列。在图6b中，我们绘制了节点55及其学习到的前3个邻居的原始时间序列。图6c显示了这些节点的地理位置，绿色节点表示中心节点学习的前3个邻居，黄色节点表示中心节点预定义的前3个邻居。我们观察到中心节点预定义的前3个邻居更接近地图上的节点本身。结果，它们的时间序列同时更加相关，如图6a中的红色圆圈所示。相反，中心节点学习的前3个邻居分布得离它更远，但仍然位于它遵循的同一条道路上。根据图6b，所学习的前3个邻居的时间序列更能够提前指示中心节点的极端业务状况。

6 总结

本文提出了一种新的多元时间序列预测框架。据我们所知，我们是第一个解决多元时间序列预测问题，通过基于图形的深度学习方法。我们提出了一种有效的方法来利用多个时间序列之间的内在依赖关系。我们的方法在多变量时间序列预测任务中表现出了优异的性能，为使用GNNs处理各种非结构化数据打开了一扇新的大门。