文章编号:1674-2974(2024)02-0057-11 DOI:10.16339/j.cnki.hdxbzkb.2024226

基于时空图注意力的短期电力负荷预测方法

李文英[1],杨高才[1],文明[1],罗姝晨[1],于宗超[1],姜羽[2†],王鼎湘[2]
(1. 国网湖南省电力有限公司经济技术研究院,湖南 长沙 410000;
2. 湖南大学 电气与信息工程学院,湖南 长沙 410082)

摘 要:准确的电力负荷预测对现代电力系统的安全经济运行至关重要.电力负荷预测可以表述为一个具有一定潜在空间依赖性的多变量时序预测问题.然而,大多数现有的电力负荷预测工作未能探索这种空间依赖关系.基于此,本文提出了一种基于时空图注意网络的短期电力负荷预测方法.提出一种基于时空图注意网络模块,该模块使用图注意层实现自适应的捕捉各用户间的潜在空间依赖性,同时使用门控卷积注意力层对各用户用电量在时间维度上进行自适应拟合,以提高网络的预测精度.实际数据实验表明,本文提出的模型整体预测精度提高明显,特别是在一定程度上缓解了长程预测精度恶化的问题,验证了所提方法的有效性与可行性.

关键词:电力负荷预测;小世界网络;时空图注意力;门控扩张因果卷积
中图分类号:TM73 **文献标志码**:A

A Short-term Power Load Forecasting Method Based on Spatiotemporal Graph Attention

LI Wenying[1], YANG Gaocai[1], WEN Ming[1], LUO Shuchen[1], YU Zongchao[1], JIANG Yu[2†], WANG Dingxiang[2]
(1. Economic and Technological Research Institute of Hunan Electric Power Company, Changsha 410000, China;
2. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: Accurate power load forecasting is crucial to the safe and economic operation of modern power systems. Power load forecasting can be expressed as a multivariable time series forecasting problem with certain potential spatial dependence. However, most existing power load forecasting work fails to explore this spatial dependency relationship. Based on this, this paper proposes a short-term power load forecasting method based on the spatiotemporal graph attention network. A spatiotemporal graph-based attention network module is proposed, which uses a graph attention layer to adaptively capture potential spatial dependencies between users. At the same time, a gated convolutional attention layer is used to adaptively fit the electricity consumption of each user in the time dimension to improve the prediction accuracy of the network. Actual data experiments show that the overall

* **收稿日期**:2023-06-28
基金项目:国家自然科学基金青年资助项目(62106072),National Natural Science Foundation of China(62106072);能源互联网供需运营湖南省重点实验室(2019TP1053),Hunan Provincial key Laboratory of Engrgy Internet Supply and Demand Operation(2019TP1053);国网湖南省电力有限公司科技项目(5216A221N008),Hunan Electric Power Company Technology Projects(5216A221N008)
作者简介:李文英(1989—),女,湖南娄底人,国网湖南省电力有限公司经济技术研究院副高级工程师,硕士
† 通信联系人,E-mail:jy199548@hnu.edu.cn

prediction accuracy of the model proposed is significantly improved, especially in alleviating the problem of deteriorating long-range prediction accuracy to a certain extent, verifying the effectiveness and feasibility of the proposed method.

Key words: electric load forecasting; small-world network; spatiotemporal graph attention; gated convolutional attention

电力系统需要满足负荷需求,不间断地稳定供应电力.通过分析电力负荷的历史行为、趋势及影响因素,预测未来需求,称为电力负荷预测.近年来,电力系统整合了多种电力资源(如可再生能源),其中具有较强波动性的风能和太阳能将负荷预测现象推向了一个更具挑战性的阶段.从电力系统的角度来看,可靠的预测结果可以有效减少风能和太阳能发电不确定性的不利影响,帮助调度部门及时制定和调整调度计划,降低运行成本.负荷预测是实现电力供需平衡的前提,在电力规划中起着至关重要的作用.

电力负荷的预测方法按时间尺度可分为四种:超短期预测[1]、短期预测[2-4]、中期预测[5]和长期预测[6].其中,短期负荷预测结果可为电力系统调度方案的制定提供依据.

现有的负荷预测方法可分为两类:经典的基于模型的方法和基于机器学习的方法.基于模型的负荷预测是一种非黑箱模型[7].常用的方法有回归分析[8]、指数平滑[9]、加权最小二乘[10]等.负荷预测结果与其驱动因素之间的对应关系是明确的,但在预测前需要了解负荷特性并明确模型类型,要根据场景选择预测方法,保证预测的准确性.基于模型的预测方法在使用时具有极大的局限性,并且模型主要是捕获线性关系,在处理较为复杂的时序数据时达不到理想的预测精度.

除了经典的基于模型的方法外,由于机器学习方法能拟合时序数据各时间戳的非线性关系,基于机器学习的方法近年来也越来越受到关注.该种方法主要使用支持向量回归、基于核的方法和前馈神经网络等,通过特征工程构造对预测结果具有因果的特征集合,最终得到时序预测模型.然而,这些模型中的大多数仅实现了静态建模预测,在单独的时间尺度上考虑信息,而不能挖掘多变量时序数据间的动态依赖关系.

深度神经网络具有较强的非线性拟合能力,能够学习复杂的多变量时序数据表示.目前,学者们已开发出不同的深度神经网络来解决时间序列预测问题,例如使用循环神经网络及其变体,自编码器、时间卷积和Transformer等.

由于同一街区内的住宅可能受到相似因素的影响,各住宅的电力负荷具有一定的潜在空间依赖性.然而,现有的大多数基于机器学习的方法仅利用电力负荷和天气信息等历史时间值,未能探索这种依赖关系,不同单位之间的空间相关性[11]通常被忽略[12].图神经网络(GNN)能够挖掘单元间的潜在依赖关系,目前在交通流预测和风速预测等任务中的应用已经证明了其优越性.与深度神经网络不同,GNN擅长对类似图的数据结构进行建模,这使得它们非常适合捕获电网中不同住宅之间复杂的潜在空间依赖性.然而,基于GNN的时空模型却很少用于电力负荷预测问题研究,主要由于其在使用过程中可能遇到下述问题:

1) GNN需要图结构作为基础,而电力负荷数据并没有显示的图结构.目前一些研究使用k-近邻来创建正则的图结构.然而当使用不恰当的k构建图结构时,会产生过平滑或冗余的问题.使用较小的k构建出来的图结构,为了聚合和更新图结构上的全部节点信息,需要增加后续GNN的层数,然而GNN是一种低通滤波器[13],当GNN的层数过多时,会产生过平滑的问题.使用较大的k构建出来的图结构,会将不存在的关系引入图结构,为后续耦合关系的提取带来误差[14].

2) 变量之间的耦合关系随着时间的推移而变化,如何动态地捕捉变量之间的关系是一个值得考虑的问题.

3) 用于捕捉时序特征的RNN(循环神经网络)/GRU(门控循环单元)/LSTM(长短时记忆网络)虽然擅长描述非线性关系,但没有考虑到因果时滞效应,对于时序数据的记忆有限,难以捕捉长期依赖关系.

针对上述问题,本文提出了一种基于小世界理

论的多变量时间序列预测方法——小世界时空图注意力网络(Small-World Spatio-Temporal Graph Attention Network, SWST-GAT).在电力负荷预测中自适应的建模与捕捉同一街区内的住宅用电量间的潜在空间依赖性,解决了现有多变量多步负荷预测技术中的问题.主要贡献具体如下:

1) 提出一种基于小世界特性的k-近邻网络构建技术,解决了GNN在多变量时序分析中图建模时存在的过平滑或冗余的问题.

2) 提出了一个基于节点重要性引导的注意力机制,与图注意层相结合,动态地捕捉变量之间的耦合关系.

3) 使用门控扩张因果卷积层提取时间序列数据的因果时滞特征,捕捉长期依赖关系,并提出基于动态系统的复杂性和耗散性质的超参数时间跨度和扩张率估计,帮助门控扩张因果卷积层的超参数选取.

1　小世界时空图卷积网络模型

本文提出的SWST-GAT模型的任务是对多变量时序$X\in\mathbb{R}^{N\times n\times d}$进行图结构构建后,将图结构与多变量时序数据$X\in\mathbb{R}^{N\times n\times d}$输入到ST-GCM模块中,使用空间卷积层提取各个变量间的耦合关系,使用时间卷积层提取每一个变量在时间维度的因果时滞关系,时间与空间卷积层交替作用,以同时提取多变量时序数据的时空特征,对变量时序数据$X\in\mathbb{R}^{N\times n\times d}$进行拟合,预测下一时刻的目标变量$y_{i,t+\mathrm{TS}+1}$或下一段时间的目标变量$Y=[y_{i,t+\mathrm{TS}+1}, y_{i,t+\mathrm{TS}+2},\cdots, y_{i,t+\mathrm{TS}+\mathrm{pr}}]$,使用最小正则化$L_2$损失来训练网络,以构建从$X$到$Y$映射$f(\cdot)$.

1.1　模型介绍

本文提出的多变量时序多步预测模型SWST-GAT的总体框架如图1所示,具体概括如下.

1) 图结构构建模块:提出一种基于小世界特性的k-近邻网络构建技术,构建多变量时间序列的静态图结构$\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{E})$,其中$\boldsymbol{V}$为节点集合,每一个节点代表一个变量,$\boldsymbol{E}$是边集,每条边表示边两端的节点间存在耦合关系.

2) ST-GCM:ST-GCM为"三明治"结构,由NA、GCN和GDCCN组成.因此,NA将度中心性引导的注意力机制(Degree Centrality Guided Attention, DCGA)

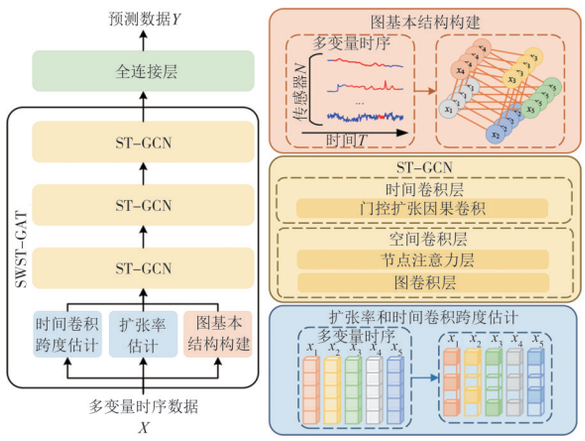


图1　SWST-GAT的网络架构

Fig.1　Network structure of SWST-GAT

和压缩激励模块(Squeeze and Excitation Networks, SENet)相结合来自动学习各个节点和连边的相对重要性,增加重要的节点和连边的权重,抑制无用的节点和连边的权重,同时使用GDCCN捕捉变量的时间维度的因果时滞特征.

3) 为了确定GDCCN层的两个关键超参数的选择,提出了基于动态系统的复杂性和耗散性质的TS估计和dr估计算法.

1.2　图结构构造模块

由于k的选择关系到网络的拓扑结构是否能够反映多变量时间序列的分布,因此,邻近数k在网络构建过程中起着非常重要的作用[15].目前讨论如何选择邻近数k的研究成果较少.当选择较小的k时,我们将得到一个极其稀疏的随机网络,此时图结构不连通,产生信息孤岛,多变量间的信息并不能共享.当k逐渐增大时,网络将开始向全连接网络演进,网络连接会逐渐变密,导致网络极度冗余.图结构的传播效率和稀疏性可通过复杂网络的网络平均一致性h和网络平均聚类系数ACC进行描述,h用于衡量复杂网络的全局传输容量;ACC用于量化复杂网络中节点之间的连接程度.其计算公式如下:

$$\mathrm{GE}(\boldsymbol{G})=\frac{1}{N(N-1)}\sum_{i,j\in V,i\neq j}\frac{1}{\mathrm{dis}_{ij}^{\min}} \tag{1}$$

$$h=\frac{1}{\mathrm{GE}(\boldsymbol{G})} \tag{2}$$

其中,h为全局效率(Global Efficiency, GE)的逆,GE与网络中节点之间的距离成反比.$\mathrm{dis}_{ij}^{\min}$表示节点v_i到节点v_j之间的最短距离.

$$\mathrm{ACC}(\boldsymbol{G})=\frac{1}{N}\sum_{i}^{N}\mathrm{CC}(i)=\frac{1}{N}\sum_{i}^{N}\frac{E(i)}{k_i(k_i-1)/2} \tag{3}$$

其中,$k_i(k_i-1)/2$是k_i相邻节点之间可能的边数,$E(i)$是节点v_i的k_i相邻节点之间的实际边数.

因此,本文拟通过h和ACC对图结构传播效率和稀疏性的定量计算,寻找两者之间的平衡,从而确定邻近数k.基于小世界网络性质的自适应k-近邻选择算法的流程如图2所示.

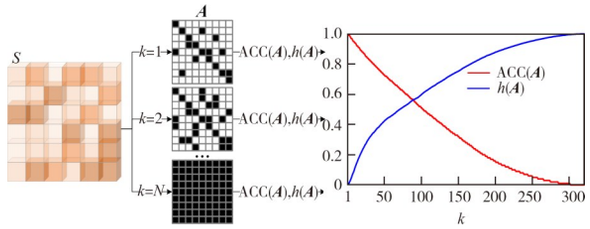


图2 基于小世界网络性质的自适应k-近邻选择算法

Fig.2 Adaptive k-nearest neighbor selection algorithm based on small-world network properties

根据小世界网络生成的过程,提出基于小世界网络的邻接矩阵生成算法,如算法1所示.具体步骤如下:首先,根据多变量时序数据$\boldsymbol{X}$计算任意两个变量间的相似性$\boldsymbol{S}=\boldsymbol{X}\cdot\boldsymbol{X}^{\mathrm{T}}$,并进行归一化,得到相似性矩阵$\boldsymbol{S}$.随后,将邻近数$k$设为0,得到一个全断开的图结构;最后,逐渐增大邻近数k至N,在此过程中,图结构会逐渐变得稠密,此时ACC逐渐变大,h逐渐变小,分别计算对应的$h(\boldsymbol{A})$和ACC$(\boldsymbol{A})$,最终形成如图2右侧所示的$h(\boldsymbol{A})$和ACC$(\boldsymbol{A})$曲线。选择$h(\boldsymbol{A})$和ACC$(\boldsymbol{A})$曲线的交(ratio, k)作为邻近数k,根据选择到的邻近数k,进行图结构$\boldsymbol{A}$的构建,具体定义公式如下:

$$\boldsymbol{A}(v_i,v_j)=\begin{cases}1, & \text{if } v_j\in \mathrm{Neigh}_k(v_i)\\ 0, & \text{if } v_j\notin \mathrm{Neigh}_k(v_i)\end{cases} \tag{4}$$

其中,$\mathrm{Neigh}_k(v_i)$为距离节点v_i最近的k个节点的集合.最后针对每一个节点,随机选择两个未与其相连接的节点,以概率p进行连接.

1.3 时空图卷积模块

由于多变量时序数据具有强耦合和大时滞的特点,变量之间的时空关系是复杂且时变的,如图3所示.因此我们提出一种新的ST-GCM针对多变量时序数据的时空特性进行处理.

1.3.1 时间卷积层

GDCCN的架构如图4所示,包含一个扩张因果卷积网络(Dilated Causal Convolution Network, DCCN)[16]、一个门控线性单元(Gated linear units, GLU)[17]和一个残差模块[18],GDCCN层的计算公

算法1:生成邻接矩阵$\boldsymbol{A}$

输入:相关性矩阵$\boldsymbol{R}\in\mathbb{R}^{N\times N}$.

输出:邻接矩阵$\boldsymbol{A}\in\mathbb{R}^{N\times N}$.

初始化:$\boldsymbol{A}\in\mathbb{R}^{N\times N}:=0$.

S ← Normalize(S)

for k in 0 : 1 : N do

Set $a_{ij}\leftarrow 1$ if $v_j\in\mathrm{Neigh}_k(v_i)$

$h_k\leftarrow h(\boldsymbol{A})$

$\mathrm{ACC}_k\leftarrow \mathrm{ACC}(\boldsymbol{A})$

end for

get intersection point (ratio, k) of curve h and ACC

set k

if $v_j\in\mathrm{Neigh}_k(v_i)$ then

$a_{ij}\leftarrow 1$

else

$a_{ij}\leftarrow 0$

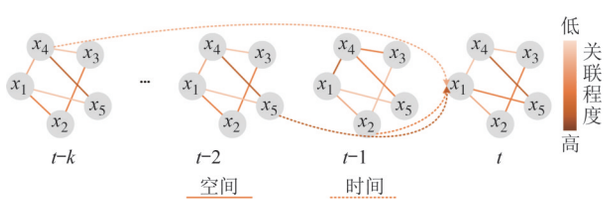


图3 多变量时序数据的时空特性

Fig.3 The spatiotemporal characteristics of multivariate temporal data

式为:

$$\begin{aligned}\boldsymbol{X}^{\mathrm{GDCCN}} &= F(\boldsymbol{X})*\Gamma = \\ &(F(\boldsymbol{X})*_{\mathrm{G}}\boldsymbol{W}_{\mathrm{G}}+b_{\mathrm{G}})\otimes\sigma(F(\boldsymbol{X})*_{\mathrm{G}}\boldsymbol{V}_{\mathrm{G}}+c_{\mathrm{G}}) = \\ &((X*_{\mathrm{D}}f)*_{\mathrm{G}}\boldsymbol{W}_{\mathrm{G}}+b_{\mathrm{G}})\otimes\sigma((X*_{\mathrm{D}}f)*_{\mathrm{G}}\boldsymbol{V}_{\mathrm{G}}+c_{\mathrm{G}})\end{aligned} \tag{5}$$

其中,$\boldsymbol{X}$和$\boldsymbol{X}^{\mathrm{GDCNN}}$为GDCCN的输入和输出,$\boldsymbol{W}_{\mathrm{G}}$、$\boldsymbol{V}_{\mathrm{G}}$、$c_{\mathrm{G}}$和$b_{\mathrm{G}}$为GDCCN层的可学习参数,$*_{\mathrm{G}}$和$*_{\mathrm{D}}$为门控卷积和扩展卷积核,$\sigma$为ReLU函数,$\otimes$为哈达马积.

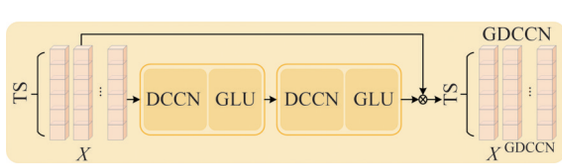


图4 GDCCN的架构

Fig.4 Network structure of GDCNN

1)扩张因果卷积网络

不同于传统非线性动态过程系统建模包含较多的历史数据,本文使用DCCN解决了长期预测结果精度降低的问题.在使用CNN(卷积神经网络)或者RNN进行长距离时间序列预测时,不可避免地需要增加CNN的层数或RNN的卷积核大小,此时网络参数和计算量将大大增加.DCCN使用具有空洞的

CNN,可以在不增加网络参数的情况下扩大卷积核的视野,如图 5 所示.扩张卷积的核心思想是将原本紧密相连的卷积核参数按照设定的比值进行扩张.扩张卷积的计算公式为

$$F(t)=(x*_{D}f)(t)=\sum_{i=0}^{k-1}f(i)x_{t-\mathrm{dr}\cdot i} \tag{6}$$

其中,dr 为膨胀比,k 为卷积核的大小,$x_{t-\mathrm{dr}\cdot i}$ 为历史数据.

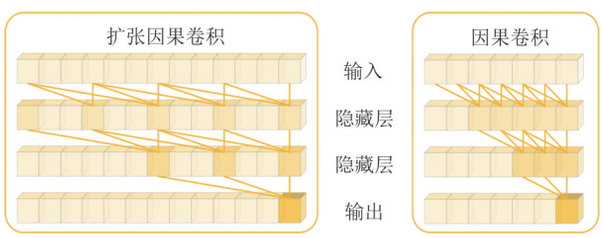


图 5　扩张因果卷积

Fig.5　Dilated causal convolution network

2) 门控线性单元

GLU 是一种简化的门控机制,可以减少梯度分散问题,保持层的非线性,加速收敛,其结构如图 6 所示.

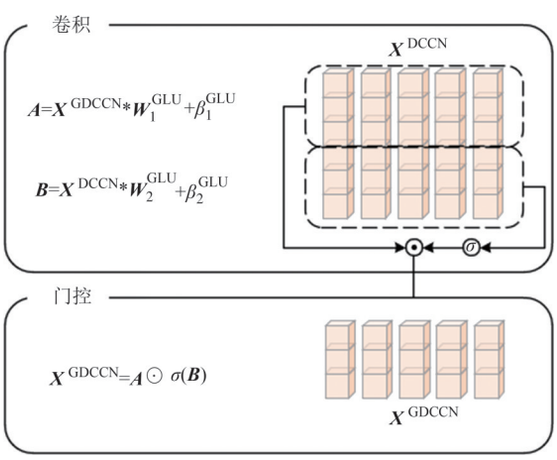


图 6　GLU 架构

Fig.6　Structure of GLU

GLU 利用 CNN 和并行处理时序数据的门控机制来实现 RNN 的功能.GLU 在处理时序数据时,严格按照时序位置保存信息,以提高性能,并通过并行处理结构加快运算速度.GLU 的公式如下:

$$X^{\mathrm{GDCCN}}=(X^{\mathrm{DCCN}}*W_1^{\mathrm{GLU}}+\beta_1^{\mathrm{GLU}})\otimes\sigma(X^{\mathrm{DCCN}}*W_2^{\mathrm{GLU}}+\beta_2^{\mathrm{GLU}}) \tag{7}$$

其中,X^{DCCN} 和 X^{GDCCN} 是 GLU 的输入和输出矩阵,W_1^{GLU}、W_2^{GLU}、β_1^{GLU} 和 β_2^{GLU} 是可学习的参数,σ 为 ReLU 函数,$\otimes$ 为哈达马积.

1.3.2　空间卷积层

1)　节点注意力层

NA 包含一个 DCGA 和一个 SENet[19].NA 层的计算公式为:

$$X^{\mathrm{NA}}=F_{\mathrm{scale}}(X^{\mathrm{GDCCN}},\mathrm{Att}_{\mathrm{DC}}\cdot F_{\mathrm{ex}}(F_{\mathrm{sq}}(X^{\mathrm{GDCCN}}),W^{\mathrm{SENet}})) \tag{8}$$

其中,X^{NA} 为 NA 层的输出数据,W^{SENet} 为 SENet 模块的可学习参数,F_{scale},F_{ex} 和 F_{sq} 为 SENet 的缩放、激励和压缩操作,$\mathrm{Att}_{\mathrm{DC}}\in\mathbb{R}^{N\times N}$ 是根据节点嵌入计算出的度中心性引导的注意力:

$$\mathrm{Att}_{\mathrm{DC}}=\mathrm{softmax}(\mathrm{DC}(\mathrm{ReLu}(W_1^{\mathrm{DC}}E_1(W_2^{\mathrm{DC}}E_2)^{\mathrm{T}}+d^{\mathrm{DC}}))) \tag{9}$$

其中,DC(·)为度中心性计算函数,矩阵 E_1 和 E_2 为随机初始化的节点嵌入,在模型训练期间可以学习,W_1^{DC} 和 W_2^{DC} 为可训练的模型参数.

作为软注意力机制的一种应用,SENet 不同于传统的通过添加特征卷积层来细化特征提取任务的方法,而是根据特征响应显式,建模特征图通道之间的相互依赖性,自适应地重新校准通道.由于多变量时序数据是典型的时空数据,因此变量之间的耦合关系会随着时间而改变.因此,本文将使用 SENet 自动学习不同通道特征的相对重要性,然后可以通过赋予不同的权重来增强或抑制相应的特征,从而提高网络模型的识别精度.SENet 的结构如图 7 所示.

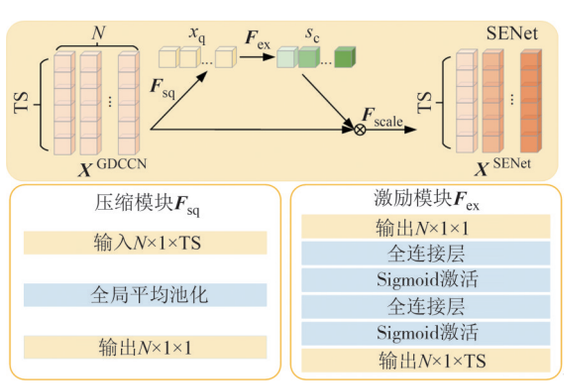


图 7　SENet 的架构

Fig.7　Structure of SENet

SENet 分为挤压和激励两步操作.首先,SENet 对输入节点变量执行挤压操作,将特征形状从 N×1×TS 压缩为 N×1×1,得到该层输入节点变量的分布,其数学描述如下:

$$x_{\mathrm{q}}=F_{\mathrm{sq}}(X^{\mathrm{GDCCN}})=\frac{1}{1\times\mathrm{TS}}\sum_{t=1}^{\mathrm{TS}}x(1,t) \tag{10}$$

其中,$x(1,t)\in\mathbb{R}^{N\times1}$ 是 t 时刻 N 个节点的数据,TS 为网络输入数据的时间长度.压缩操作通常采用全局最

大池化或全局平均池化两种算法,最后将全局空间信息压缩成一个通道维度描述符 x_q,本文选择全局平均池化.

激励操作定义如下:

$$\mathbf{sc} = F_{ex}(x_q, \boldsymbol{W}^{\mathrm{SENet}}) = \sigma(\boldsymbol{W}_2^{\mathrm{SENet}}\sigma(\boldsymbol{W}_1^{\mathrm{SENet}}x_q)) \tag{11}$$

其中,$\boldsymbol{W}_1^{\mathrm{SENet}} \in \mathbb{R}^{(\mathrm{TS}/\beta \times \mathrm{TS})}$ 和 $\boldsymbol{W}_2^{\mathrm{SENet}} \in \mathbb{R}^{\mathrm{TS} \times (\mathrm{TS}/\beta)}$,$\sigma$ 是 ReLU 函数,β 是维度变换率.

最后将得到的缩放矩阵 $\mathbf{sc}$ 的元素与图数据的通道一一相乘,SENet 的输出定义为:

$$\boldsymbol{X}^{\mathrm{SENet}} = F_{\mathrm{scale}}(\boldsymbol{X}^{\mathrm{GDCCN}}, \mathbf{sc}) = \mathbf{sc}_i \cdot x_i \tag{12}$$

其中,sc_i 是第 i 个节点向量 x_i 的缩放系数.SENet 的运行充分利用了注意力机制,使模型在训练过程中会关注更重要的节点,从而提高模型的整体性能.

2) 图卷积层

使用 Chebyshev 多项式改进的 GCN 提取空间特征[20],其结构如图 8 所示.GCN 的卷积核为 $g_\theta = \sum_{i=1}^{K} \boldsymbol{W}_i T_i(\tilde{\Lambda})$,其中 $\tilde{\Lambda} = 2\boldsymbol{\Lambda}/\lambda_{\max} - \boldsymbol{I}$,$T_i(\tilde{\Lambda})$ 是 i 阶多项式,Λ 为邻接矩阵 $\boldsymbol{A}$ 的特征值.Chebyshev 多项式 $T_i(\tilde{\Lambda})$ 通过递归计算得到,其计算公式如下:

$$T_k(\tilde{\Lambda}) = 2\tilde{\Lambda}T_{k-1}(\tilde{\Lambda}) - T_{k-2}(\tilde{\Lambda}) \tag{13}$$

其中,$T_0(\tilde{\Lambda}) = 1$,$T_1(\tilde{\Lambda}) = \tilde{\Lambda}$,$k$=3.图卷积核的卷积操作的计算式如下:

$$\begin{aligned} \boldsymbol{X}^{\mathrm{GCN}} &= \boldsymbol{X} *_{\mathrm{G}} g_\theta = \boldsymbol{U}\left(\sum_{i=1}^{K} \boldsymbol{W}^{\mathrm{GCN}} T_k(\tilde{\Lambda})\right)\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{NA}} \\ &= \sum_{i=1}^{K} \boldsymbol{W}^{\mathrm{GCN}} T_i(\tilde{L})\boldsymbol{X}^{\mathrm{NA}} \end{aligned} \tag{14}$$

其中,$\boldsymbol{X}^{\mathrm{NA}}$ 和 $\boldsymbol{X}^{\mathrm{GCN}}$ 是 GCN 层的输入和输出矩阵,$\boldsymbol{W}^{\mathrm{GCN}}$ 是可学习的参数,$*_{\mathrm{G}}$ 为 Chebyshev 多项式构造出的卷积核.

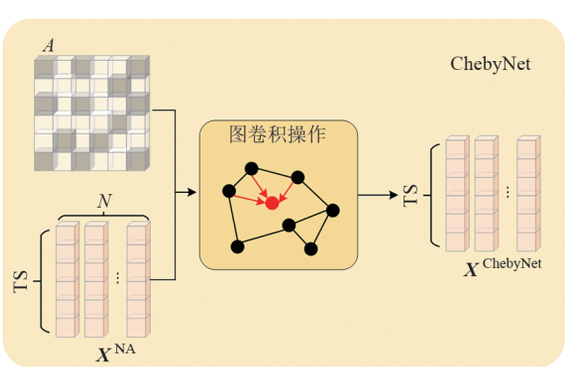


图 8 GCN 的架构

Fig.8 Structure of GCN

1.4 超参数估计

1.4.1 时间跨度估计

时间跨度 TS 是时间卷积层 GDCCN 的卷积核的大小.在前期研究准备过程中,我们发现现有文献中对 TS 的选择是基于经验的.例如,交通流量预测中一般选择 24(一天 24 小时)×7(一周 7 天),并且这种经验选择仍然用于文献[21]中的所有多变量时序数据集.但是对于不同领域内的数据集,例如工业或金融数据集,使用这个经验值是不合适的.为了解决这个问题,提出了一个时间跨度估计算子以估计 GDCCN 的卷积核的大小.

考虑到更有效地利用时间序列中包含的信息,使用信息论中的经典指标:CD 和 LS,估计能够等效包含原始动态信息的时间维数的最小值[22],将其作为时间跨度 TS.CD 反映了动态系统中不规则变化的程度.系统变化越复杂,CD 越大,表示系统演化所需的数据维数也越高.此外,任何耗散系统的 LS 中至少有一个维度应小于 0,并且所有维度之和应小于 0.因此,当选择的 TS 不能覆盖时间序列数据的时间动态时,LS 各项之和大于零.基于此,本文提出了一种基于动态系统分析的时间跨度估计算法.伪代码如算法 2 所示.

算法 2: 时间跨度估计算法

输入: 原始数据 $\boldsymbol{X} \in \mathbb{R}^{N \times n \times d}$.

输出: 时间跨度 TS.

for every variable x_i in X do

$D_2(i) \leftarrow \mathrm{CD}(x_i)$

$\mathrm{TS} \leftarrow \lceil \min(D_2(X_i)) \rceil$

for TS in $\min(D_2(X_i))$: 1: $\max(D_2(X_i))$ do

if LS$\leftarrow \sum_{k=1}^{\mathrm{TS}} \lambda_k > 0$

break

1.4.2 扩张率估计

时间序列中的每一组收集到的多变量时序数据都是各种影响因素的综合结果,这些影响因素大致可以分为三种:趋势性、周期性和随机性.每一时刻多变量的实际变化通常是由若干影响因素的叠加或组合引起的.因为工业领域的变量具有不同的物理意义,所以每个变量的趋势和周期变化很可能是不同的.本节将变量的趋势和周期变化定义为时间节律.

基于此,本文采用基于动态系统分析的方法对节律时钟进行估计,以指导扩张率的选择.选择互信息(Mutual information, MI)来估计每个变量 x_i 的时间节律 $T(i, i)$.$T(i, i)$ 的计算公式如下:

$$I(x_i) = 2H(x_i) - H(x_i, x_i) \tag{15}$$

$$H(x_i) = -\sum_{p=1}^{d} P_{x_i}(x_{i,p}) \log_2 P_{x_i}(x_{i,p}) \tag{16}$$

$$H(x_i,x_i)=-\sum_{p=1}^{d}P_{x_i,x_i}(x_{i,p},x_{i,p})\log_2 P_{x_j,x_i}(x_{i,p},x_{i,p}) \tag{17}$$

$T(i,\ i)$是$I(x_i)$的第一个局部最小值,表示系统获取信息的时间尺度.同时,我们计算$T(i,j)$得到变量x_i和变量x_j之间的延迟时间,并进行数据对齐.扩张卷积数据对齐流程图如图9所示,右上角的子图是变量之间的对齐过程.x_i是要预测的变量.x_j的$t+T(x_i,\ x_j)$时刻和x_k的$t+T(x_i,\ x_k)$时刻与x_i的t时刻对齐.右下角的子图显示了不同扩张率的扩张卷积对不同变量的处理.

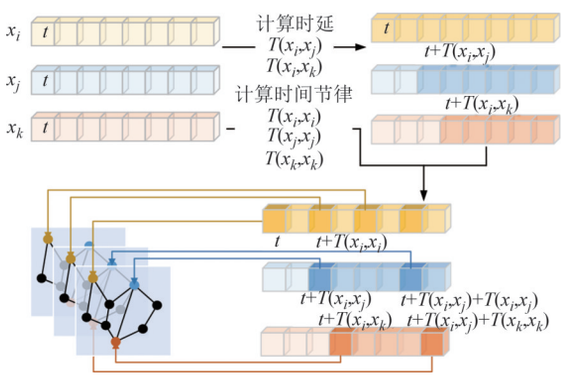


图9　扩张卷积数据对齐流程图

Fig.9　Data alignment flowchart using DCCN

1.5　模型训练与测试

均方误差损失函数用于模型训练优化,损失函数的定义式为:

$$\text{Loss}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Y}-\hat{\boldsymbol{Y}})^2 \tag{18}$$

其中,n是时间序列中训练数据的总数,$\boldsymbol{Y}$是真实值,$\hat{\boldsymbol{Y}}$是预测值.

根据上述SWST-GAT框架,利用Adam优化算法在反向传播中对每个权重参数求网络误差的梯度,通过参数更新过程得到新的权重.迭代计算模型权重,直到达到预定的小损失,并获得最佳预测值。选择Adam作为优化算法的原因是它可以为不同的参数设计独立的自适应学习率,计算效率较高.训练算法具体细节在算法3中给出.

算法3 训练算法

输入:多变量时序数据$\boldsymbol{X}\in\mathbb{R}^{N\times n\times d}$,邻接矩阵$\boldsymbol{A}$ and 初始化模型f.

输出:可训练的权重矩阵$\boldsymbol{W}$和偏置矩阵$\boldsymbol{\beta}$.

初始化:确定$\hat{\boldsymbol{Y}}$学习率γ,批尺寸b,迭代次数iter,历史数据长度h,预测步数p.随机初始化权重矩阵$\boldsymbol{W}$和偏置矩阵$\boldsymbol{\beta}$.

Repeat:

Forward Propagation:

Do

get batch $\boldsymbol{X}^b$;

$\hat{\boldsymbol{A}}\leftarrow\boldsymbol{A}+\boldsymbol{I}$;

$D\leftarrow\deg\text{ree}(\hat{\boldsymbol{A}})$;

$X^0\leftarrow\boldsymbol{X}^b$;

for $l\leftarrow 1$ to L

Do

$\hat{\boldsymbol{y}}\leftarrow f(\boldsymbol{X}^0,\boldsymbol{W},\boldsymbol{\beta})$

$\text{Loss}\leftarrow 1/n\sum_{i=1}^{n}\left\|\boldsymbol{Y}-\hat{\boldsymbol{Y}}\right\|^2$

update $\boldsymbol{W}$ and $\boldsymbol{\beta}$.

end for

iter = iter + 1.

Until convergence

2　电力负荷预测实验分析

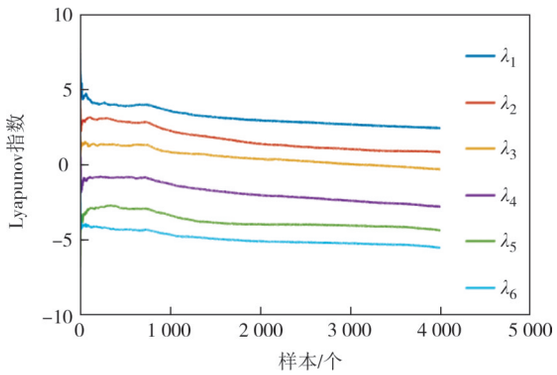
本节将在电力负荷数据EL上验证所提出的网络SWST-GAT的预测性能,EL数据集是新加坡某电站2012~2014年间321个客户的用电量(kWh),采样频率为15 min.数据集存放有26 304个变量数为321的样本,采样率为1 h.数据对应Lyapunov谱如图10(a)所示,根据基于小世界网络性质的自适应k-近邻选择算法,计算得到的h和ACC($\boldsymbol{G}$)曲线如图10(b)所示,因此,选定数据集时间跨度TS=4、近邻数k=83.本节所有实验均在Windows系统上编译和测试(CPU:Intel(R) Core(TM) i9-10900K CPU @ 3.70 GHz,GPU:NVIDIA GeForce RTX 3 090).

针对模型的预测精度,本文使用五个评价指标对预测性能进行评价,包括平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)和相关系数(Correlation Coefficient, CC).MAE,RMSE,MAPE的值越小,CC的值越大,所评价模型的预测精度越高.

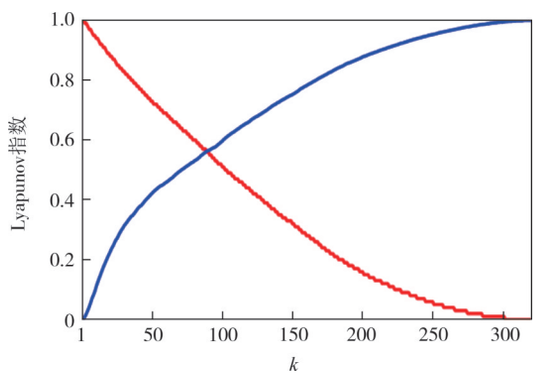
$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i| \tag{19}$$

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i)^2} \tag{20}$$

$$\text{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i}{\boldsymbol{Y}}\right|\times 100\% \tag{21}$$



(a) 电力负荷数据的李雅普诺夫谱



(b) 电力负荷数据的近邻数选择

图 10 新加坡某电站2012~2014年间电力负荷情况

Fig.10 The power load of a power station in Singapore from 2012 to 2014

$$CC = \frac{\mathrm{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{Y}}_i)}{\sqrt{\mathrm{Var}(\boldsymbol{Y})\mathrm{Var}(\hat{\boldsymbol{Y}})}} \times 100\% \tag{22}$$

2.1 实验结果

为了证明SWST-GAT的有效性,将SWST-GAT与五个多变量时序预测模型进行比较,模型的简要描述在表1中列出.五个对比模型分别为TPA-LSTM[23]、LSTNet[24]、DCGNet[25]、Graph WaveNet[26]和MTGNN[21].TPA-LSTM使用一组滤波器来提取时不变的时间模式.在此基础上,提出了一种新的注意机制来选择相关的时间序列,进行预测;LSTNet使用CNN来提取变量之间的短期局部依赖模式,并使用RNN发现时间序列趋的长期模式.同时,利用传统的自回归模型来解决神经网络模型的尺度不敏感问题;DCGNet使用深度CNN提取多变量的局部耦合特征和门控循环单元网络提取时间维度的非线性动态特征;Graph WaveNet通过建立一个新的自适应依赖矩阵,对节点嵌入进行学习,捕获数据中隐藏的空间依赖,同时使用一维的CNN进行长序列的特征提取;MTGNN提出一个单向图结构生成算法,并结合DCCN和GCN对多变量数据进行预测.

表1 对比模型的简要描述

Tab.1 Brief description of the comparative model

性质	模型名	模型描述
RNN	TPA-LSTM	结合注意力机制的RNN
	LSTNet	结合了CNN和RNN的混合模型
	DCGNet	结合CNN和GRU的混合模型
GNN	Graph WaveNet	将扩散图卷积与一维扩张卷积相结合的ST-GCN模型
	MTGNN	集成了图学习、GCN和TCN的ST-GCN模型

所有比较的方法都使用滑动窗口输入,表2总结了六种模型的评估结果,最优结果用红色字体加粗显示,次优结果用加粗显示.六种预测方法的真实值与预测值对比如图11所示,对应的预测误差曲线如图12所示.从评估结果可以看出本节提出的SWST-GAT模型预测精度最好,RMSE和MAE最小,CC最大,表明真实值和预测值之间变化趋势基本一致.

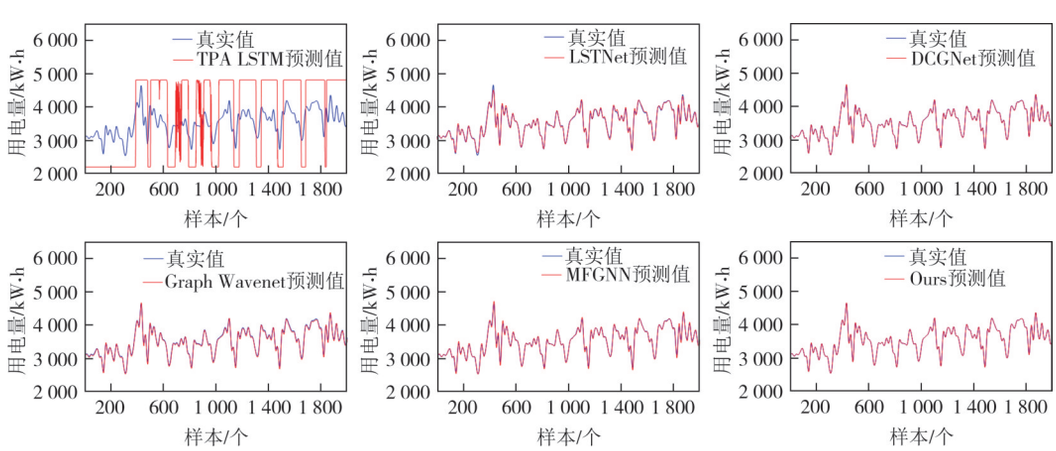


图 11 不同模型的预测值与真实值

Fig.11 Predicted and true values of different models

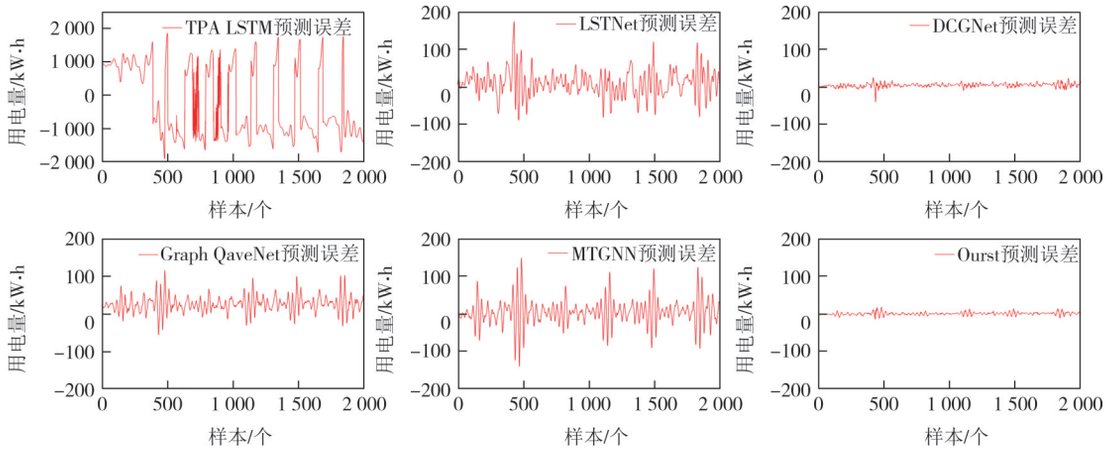


图 12　不同模型的预测值与真实值的误差

Fig.12　Error between predicted and true values of different models

基于 LSTM 的 TPA-LSTM 和 LSTNet 模型的 MAE、RMSE 和 MAPE 明显高于其他模型,预测精度最低,表明这类模型不能轻易地适应实际值的局部趋势.与上述两种模型相比,结合动态 RNN 的 DCGNet模型的预测精度有了显著提高,在部分数据集上取得了次优结果,如表 2 所示,在混合模型 DCGNet 模型的短程预测中,预测曲线符合实际趋势.基于 GCN 的 Graph WaveNet 和 MTGNN 与表 2 中的静态和动态循环网络相比,进一步验证了 GNN 在考虑变量间耦合关系的后处理多变量时间序列方面的有效性,尤其在长程预测方面.与所提出模型的预测性能相比,这些网络的预测性可以进一步提高.

表 2　模型评估结果

Tab.2　Model evaluation results

模型	指标	预测步数				模型	指标	预测步数			
		3	6	12	24			3	6	12	24
TPA-LSTM	MAE	1 029.63	1 279.65	1 020.84	1 614.65	LSTNet	MAE	7.839 8	12.163 8	10.461 7	14.331 9
	RMSE	1 479.56	1 327.92	1 486.26	1 703.96		RMSE	10.039 2	16.380 1	19.105 8	19.105 8
	MAPE	12.34	18.48	20.19	17.59		MAPE	0.233 0	0.370 1	0.307 7	0.421 6
	CC	66.781 5	69.691 0	67.757 4	64.617 1		CC	99.976 3	99.918 2	99.989 8	99.886 8
DCGNet	MAE	**3.488 2**	**4.930 8**	**7.548 2**	20.306 6	Graph WaveNet	MAE	13.221 3	13.192 0	10.618 9	11.027 9
	RMSE	**4.992 8**	**6.120 2**	**9.704 3**	30.086 8		RMSE	19.479 1	15.614 7	13.956 6	14.350 2
	MAPE	**0.102 3**	**0.143 7**	**0.220 6**	0.612 2		MAPE	0.403 5	0.400 0	0.319 9	0.318 5
	CC	**99.993 1**	**99.993 7**	**99.978 6**	99.729 7		CC	99.892 1	99.970 7	99.944 3	99.937 7
MTGNN	MAE	7.003 1	9.530 6	10.833 4	**10.190 8**	SWST-GAT	MAE	**1.468 5**	**2.466 2**	**1.264 6**	**4.423 7**
	RMSE	9.454 3	13.404 2	20.829 7	**14.765 7**		RMSE	**2.101 3**	**3.061 8**	**1.697 2**	**6.137 4**
	MAPE	0.205 5	0.294 2	0.311 9	**0.308 9**		MAPE	**0.044 7**	**0.071 9**	**0.037 6**	**0.132 4**
	CC	99.974 6	99.948 6	99.860 0	**99.932 3**		CC	**99.998 6**	**99.998 4**	**99.999 1**	**99.988 4**

此外,为了进一步评价不同预测模型的性能,引入了误差概率分布曲线.图 13 为六种不同预测模型下预测残差的概率分布.与其他模型相比,本文模型得到的预测值的误差分布更接近于 0,变化更小,进一步说明了本文方法的可靠性.

2.2　模块有效性分析

为了证明邻接矩阵构建的近邻数 k 选择和 GDCCN 的时间跨度选扩张率的选择的有效性,进行了消融实验.首先定义不同的模型如下.

1) our-0.1:选择连接最近的 0.1N 个邻居.

2) our-0.8:选择连接最近的 0.8N 个邻居.

3) our-nTS:不使用时间跨度估计模块.

4) our-nDCGA:不使用基于 DCGA.

5) our-nSENet:不使用 SENet 模块.

6) our-ndr:不使用扩张率估计模块.

对于上述模型,评价指标测得的测试结果如表 3 所示.

从这些实验结果得出的几个结论总结如下.

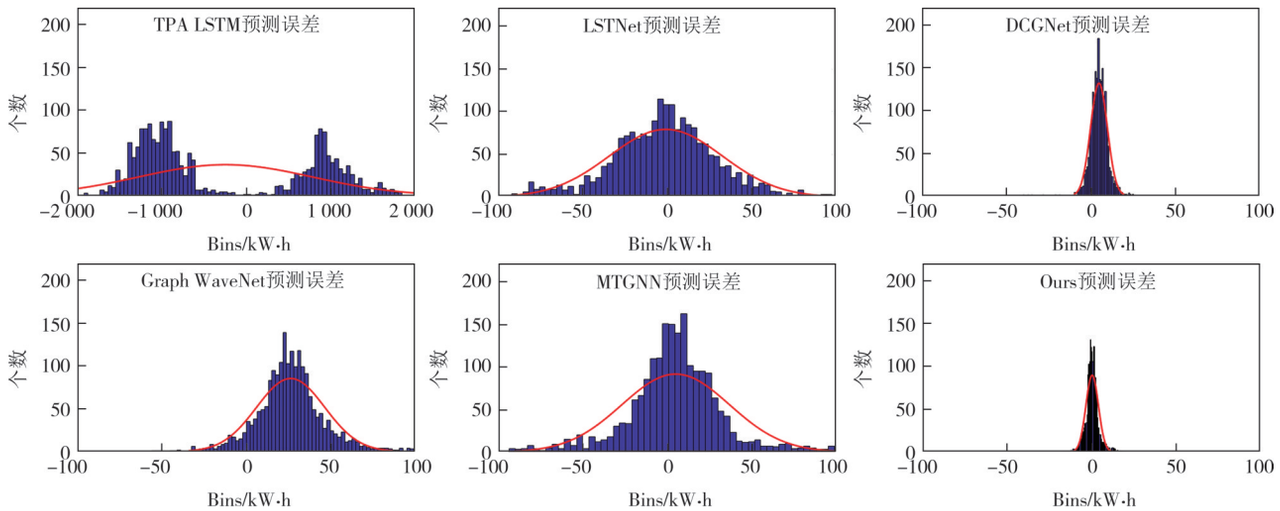


图 13 不同模型的预测值与真实值的误差

Fig.13 Error between predicted and true values of different models

表 3 消融实验结果

Tab.3 Results of ablation experiments

模型	MAE	RMSE	MAPE	CC
Complete model	5.938 3	8.058 6	0.178 9	99.981 0
our-0.1	13.260 0	9.751 5	0.290 9	98.909 0
our-0.8	6.066 9	8.394 9	0.181 8	99.400 0
our-nTS	25.938 3	28.058 6	0.678 9	99.781 0
our-nDCGA	6.066 9	8.394 9	0.181 8	99.400 0
our-nSENet	20.403 0	23.279 0	0.609 3	99.840 0
our-ndr	22.796 5	23.279 2	0.684 8	99.691 0

1)本文提出的基于小世界网络的邻接矩阵生成算法确实可以在保证预测精度的同时减少计算量.通过与 our-0.1 和 our-0.8 的结果相对比,可以看出当近邻数 k 足够大时,即使增大 k,也不会提高整体算法的性能,有时反而会造成性能的衰退.造成这种情况的原因可能是大量无关的边相连后导致的图结构的扭曲,与真实的多变量间的耦合情况相去甚远,造成后续利用耦合关系进行预测的 GCN 的性能的下降.若选择的近邻数 k 较小,则无法通过图结构详细的表示出多变量间的耦合关系,导致 GCN 提取的网络耦合关系不足以完成高精度的预测.

2) SENet 和 NA 确实可以增加对预测影响较大的变量的权重,抑制对预测影响较小的变量的权重,提高预测精度;

3) GDCCN 的超参数选择十分重要,当沿用经验设定值时,模型的效果大打折扣.造成这种情况的原因可能是时间序列中的每一组收集到的多变量时序数据都是各种影响因素的综合结果.每一时刻多变量的实际变化通常是由若干影响因素的叠加或组合引起的.因为来自不同系统的变量具有不同的物理意义,所以每个变量的趋势和周期变化很可能是不同的.

3 结论

本文针对电力负荷预测中多变量间的非线性耦合特性难建模问题,提出了一种基于小世界理论的模型 SWST-GAT,用于多变量时间序列的动态长程预测.针对多变量时序数据没有清晰图结构的问题,设计了权衡网络全局效率和冗余度的图基本结构构建算法;针对多变量时序变量间存在时滞的问题,构建了一个全新的时间卷积层以捕捉较长历史数据内的非线性关联关系;针对网络超参数选取凭经验的问题,提出了一种基于复杂动态系统的超参数估计方法,避免经验值导致长程预测效果的减弱.得到结论如下:

1) 相比于使用 CNN 捕获多变量间关系的方法,多变量时序预测方法的精度更高;

2) 使用 GDCCN 能捕捉更加长的历史数据的非线性特征,一定程度上缓解了长程预测精度的下降;

3) GDCCN 的超参数选择十分重要,当沿用经典设置值时,模型的效果大打折扣;

4) 使用 DCGA 与 SENet 相结合,从图数据的结构和属性两个层面进行动态拟合的效果要比单纯使用 GCN 的效果好.

参考文献

[1] HSIAO Y H. Household electricity demand forecast based on context information and user daily schedule analysis from meter

data [J]. IEEE Transactions on Industrial Informatics, 2015, 11(1): 33-43.

[2] CHEN Y, XU P, CHU Y, et al. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings [J]. Applied Energy, 2017, 195: 659-670.

[3] CAI M, PIPATTANASOMPORN M, RAHMAN S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques [J]. Applied Energy, 2019, 236: 1078-1088.

[4] JURADO S, NEBOT À, MUGICA F, et al. Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques [J]. Energy, 2015, 86: 276-291.

[5] KUSTER C, REZGUI Y, MOURSHED M. Electrical load forecasting models: A critical systematic review [J]. Sustainable Cities and Society, 2017, 35: 257-270.

[6] ZHANG D, AHMAD T, SHAH W A J I A. Efficient energy planning with decomposition-based evolutionary neural networks[J]. 2020, 8: 134880-134897.

[7] ZHANG N, LI Z, ZOU X, et al. Comparison of three short-term load forecast models in Southern California [J]. Energy, 2019, 189: 116358.

[8] LEE Y-S, TONG L-I. Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming [J]. Knowledge-Based Systems, 2011, 24(1): 66-72.

[9] 叶宗斌, 周步祥, 林楠, 等. 基于等维新息指数平滑法模型的中长期负荷预测 [J]. 电力系统保护与控制, 2012, 40(18): 47-51.
YE Z B, ZHOU B X, LIN N, et al. Mid-long term load forecasting based on equal dimension new information exponential smoothing model[J]. Power System Protection and Control, 2012, 40(18): 47-51.(in Chinese)

[10] 赵超, 戴坤成. 自适应加权最小二乘支持向量机的空调负荷预测 [J]. 重庆大学学报, 2016, 39(1): 55-64.
ZHAO C, DAI K C. Air conditioning load forecasting based on adaptive weighted least squares support vector machine [J]. Journal of Chongqing University, 2016, 39 (1): 55-64. (in Chinese)

[11] CORIZZO R, CECI M, FANAEE-T H, et al. Multi-aspect renewable energy forecasting [J]. Information Sciences, 2021, 546: 701-722.

[12] TASCIKARAOGLU A. Evaluation of spatio-temporal forecasting methods in various smart city applications [J]. Renewable and Sustainable Energy Reviews, 2018, 82: 424-435.

[13] BO D Y, WANG X, SHI C, et al. Beyond low-frequency information in graph convolutional networks[DB/OL]. http://arxiv.org/abs/2101.00797.

[14] SILVA T C. Machine learning in complex networks: modeling, analysis, and applications [D]. Beijing: Institute of Computing, Chinese Academy of Social Sciences, 2012.

[15] MAIER M, LUXBURG U V, HEIN M, et al. Influence of graph construction on graph-based clustering measures [J]. NIPS'08: Proceedings of the 21st International Conference on Neural Information Processing Systems, 2008, 12: 1025-1032.

[16] MARIANI S, RENDU Q, URBANI M, et al. Causal dilated convolutional neural networks for automatic inspection of ultrasonic signals in non-destructive evaluation and structural health monitoring [J]. Mechanical Systems and Signal Processing, 2021, 157: 107748.

[17] DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks [C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, NSW, Australia, 2017.

[18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016.

[19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 2018.

[20] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain, 2016.

[21] WU Z, PAN S, LONG G, et al. Connecting the Dots: multivariate time series forecasting with graph neural networks [C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA, 2020.

[22] BROWN R, BRYANT P, ABARBANEL H D I. Computing the Lyapunov spectrum of a dynamical system from an observed time series [J]. Physical Review A, 1991, 43(6): 2787-2806.

[23] SHIH S, SUN F, LEE H. Temporal pattern attention for multivariate time series forecasting [J]. Machine Learning, 2019, 108(8-9): 1421-1441.

[24] LAI G, CHANG W C, YANG Y, et al. Modeling long-and short-term temporal patterns with deep neural networks [C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York, NY, USA, 2018.

[25] ZHANG X, LEI Y, CHEN H, et al. Multivariate time-series modeling for forecasting sintering temperature in rotary kilns using DCGNet [J]. IEEE Transactions on Industrial Informatics, 2021, 17(7): 4635-4645.

[26] 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. Macao, China, 2019.