



计算机科学

COMPUTER SCIENCE

基于AR与DNN联合模型的地理传感器时间序列预测

董红斌, 韩爽, 付强

引用本文

董红斌, 韩爽, 付强. 基于AR与DNN联合模型的地理传感器时间序列预测[J]. 计算机科学, 2023, 50(11): 41-48.

DONG Hongbin, HAN Shuang, FU Qiang. [Geo-sensory Time Series Prediction Based on Joint Model of Auto Regression and Deep NeuralNetwork](#) [J]. Computer Science, 2023, 50(11): 41-48.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于纳什竞价的空间众包任务定价算法](#)

Spatial Crowdsourcing Task Pricing Algorithm Based on Nash Bidding

计算机科学, 2023, 50(10): 184-192. <https://doi.org/10.11896/jsjcx.220900130>

[融合粗粒度代价体及双边网格的轻量级多视图三维重建](#)

Lightweight Multi-view Stereo Integrating Coarse Cost Volume and Bilateral Grid

计算机科学, 2023, 50(8): 125-132. <https://doi.org/10.11896/jsjcx.220600046>

[基于动态时空神经网络的城市交通流量预测方法](#)

City Traffic Flow Prediction Method Based on Dynamic Spatio-Temporal Neural Network

计算机科学, 2023, 50(6A): 220600266-7. <https://doi.org/10.11896/jsjcx.220600266>

[基于深度多尺度卷积稀疏编码的图像去噪算法](#)

Image Denoising Algorithm Based on Deep Multi-scale Convolution Sparse Coding

计算机科学, 2023, 50(4): 133-140. <https://doi.org/10.11896/jsjcx.220100090>

[基于图像颜色随机变换的对抗样本生成方法](#)

Adversarial Examples Generation Method Based on Image Color Random Transformation

计算机科学, 2023, 50(4): 88-95. <https://doi.org/10.11896/jsjcx.211100164>

基于 AR 与 DNN 联合模型的地理传感器时间序列预测

董红斌 韩爽 付强

哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001

(donghongbin@hrbeu.edu.cn)

摘要 地理传感器时间序列具有复杂动态的语义时空相关性和地理时空相关性。尽管已经开发了各种深度学习模型用于时间序列预测,但很少有模型能专注于捕捉地理传感器时间序列内的多类型时空相关性。此外,同时预测多个传感器在未来某一时间步的值非常具有挑战性。为了解决上述问题,提出了一种自回归模型与深度神经网络的联合模型(Joint model of Autoregression and Deep Neural Network, J-ARDNN),用于处理地理传感器时间序列的多目标预测任务。在该模型中,空间模块用于捕捉不同序列间多类型空间的相关性,时间模块采用时间卷积网络来提取单个序列内的时间依赖关系。此外,还引入自回归模型来提高预测模型的鲁棒性。为了验证 J-ARDNN 模型的有效性和优越性,在不同领域的真实时间序列数据集上进行了充分的实验,结果表明, J-ARDNN 模型的预测性能优于对比方法。

关键词: 地理传感器时间序列;多目标预测;时空相关性;自回归模型;深度神经网络

中图法分类号 TP391

Geo-sensory Time Series Prediction Based on Joint Model of Auto Regression and Deep Neural Network

DONG Hongbin, HAN Shuang and FU Qiang

College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Abstract Geo-sensory time series contain complex and dynamic semantic spatio-temporal correlations and geographic spatio-temporal correlations. Although a variety of existing deep learning models have been developed for time series prediction, few of them focus on capturing multi-type of spatial-temporal correlations within geo-sensory time series. In addition, it is challenging to simultaneously predict the future values of multiple sensors at a certain time step. To address these issues and challenges, this paper proposes a joint model of autoregression and deep neural network (J-ARDNN) to achieve the multi-objective prediction task of geo-sensory time series. In this model, the spatial module is proposed to capture the multi-type spatial correlations between different series, the temporal module introduces the temporal convolutional network to extract the temporal dependencies within a single series. Moreover, the autoregression model is introduced to improve the robustness of the J-ARDNN prediction model. To prove the superiority and effectiveness of the J-ARDNN model, the proposed model is evaluated in three real-world datasets from different fields. Experimental results show that the proposed model can achieve better prediction performance than state-of-the-art contrast models.

Keywords Geo-sensory time series, Multi-objective prediction, Spatio-Temporal correlation, Autoregression model, Deep neural network

1 引言

人类的生活范围逐步扩大,而单个传感器的监测范围有限,不能代表整个区域或者城市的环境信息。因此,需要部署多个位于不同地理空间位置的传感器同时监测整个区域的环境。这些具有地理空间相关性的传感器收集的时间序列,被称为地理传感器时间序列^[1]。地理传感器时间序列的应用

范围广泛,如交通预测^[2]、天气预测^[3]、空气质量预测^[4]、城市供水预测^[5]等。

多目标预测是地理传感器时间序列预测的一个重要分支,越来越受到研究人员的关注。不同于预测一个目标序列未来值的预测任务,多目标时间序列预测设定所有的序列为目标序列,其目的是同时预测所有目标序列的未来值。与依次为每个目标序列建模的做法相比,多目标预测模型具有

到稿日期:2023-05-30 返修日期:2023-08-31

基金项目:国家自然科学基金(61472095);黑龙江省自然科学基金(LH2020F023)

This work was supported by the National Natural Science Foundation of China(61472095) and Natural Science Foundation of Heilongjiang Province, China(LH2020F023).

通信作者:韩爽(hshanshuang@hrbeu.edu.cn)

效率高的特点。它在现实生活中起着至关重要的作用,例如,根据未来多支股票的趋势预测来设计投资策略,通过预测一个城市不同路段的交通流量来规划出行路线^[6]。

多目标预测需要同时考虑不同序列之间的复杂的空间相关性和单个序列内的时间相关性。其中,序列之间存在多类型空间相关性,即地理空间相关性和语义空间相关性。地理传感器时间序列收集于实际的地理空间背景下,因此不同序列之间的相关性受到地理因素的影响。例如,相邻区域的交通流往往由于道路连接的关系,具有相似的趋势。语义空间相关性指地理空间不相邻的序列因相似外部环境因素而具有相似趋势。图1给出了6个观测站点的气温变化。从图中观察到,站点1与站点4和站点5具有高度相似的涨幅趋势,说明它们之间存在着较强的相关性,并且这种相关性随时间不断变化。例如,在第21—29个时间步,站点1与站点5的气温相似度高于站点4,如图1的左侧方框所示;第89—97个时间步,站点1与站点4的气温相似度高于站点5,如图1的右侧方框所示。此外,站点1对站点4的影响与站点4对站点1的影响是不同的,因为存在与站点4更相关的站点6,这说明站点之间的相关性是单向的。因此,对不同序列之间多类型相关性建模对多目标预测具有重要意义。

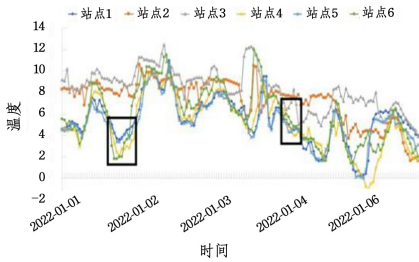


图1 6个站点的历史气温

Fig. 1 Historical temperature at six stations

深度学习模型由于在捕捉非线性关系方面具有灵活性,在时间序列预测领域引起了极大的关注。LSTNet^[7]和TPA^[8]等深度学习模型在多目标预测领域取得了令人瞩目的成就。但是,在处理具有动态周期模式或非周期模式的数据时,它们的性能会显著下降。MTNet^[9]模型利用记忆网络的思想来处理多目标预测任务,并且不受限于数据类型。然而,LSTNet,TPA-LSTM和MTNet都采用卷积神经网络将变量之间的相关性封装在一个隐藏状态中,没有显式地对变量之间的依赖性进行建模,削弱了模型的可解释性。为了解决这个问题,一些模型采用注意力机制衡量变量之间的相关性,例如,DSANet^[10]使用了一个self-attention模块来建模不同序列之间的依赖关系。MTGNN^[11]模型引入图卷积神经网络来融合节点与其邻居信息,以处理空间相关性。尽管这些方法在时间序列预测方面取得了巨大的成功,但是它们并没有考虑地理传感器时间序列中的地理空间因素。

针对上述挑战和现有工作中存在的问题,本文提出了一种AR与深度神经网络的联合模型,旨在显式地对不同序列之间的空间相关性建模以及捕捉单个序列内的时间相关性。研究表明,线性模型和非线性模型的混合模型在多目标预测任务中比单一网络更准确^[12-13]。深度学习神经网络由空间模块

和时间模块组成,其中空间模块包括可解释的多变量LSTM网络和混合空间注意力机制,以捕捉序列之间的多类型空间相关性,时间模块的时间卷积网络(Temporal Convolutional Network, TCN)针对单个序列内的时间相关性进行建模。将AR与深度神经网络并行集成,以解决非线性模型对输入尺度不敏感的问题。

本文研究的主要贡献如下:

- 1)提出了一种联合网络,用于建模地理传感器时间序列数据。所提网络能够显式地捕捉不同序列之间复杂的空间相关性和时间相关性。
- 2)设计了一种混合空间注意力机制,同时考虑序列之间不同类型的空间相关性。
- 3)引入AR模型,利用其优势进一步提高了模型的鲁棒性。
- 4)在3个真实数据集上进行了实验,实验结果表明,J-ARDNN模型优于对比方法。

2 相关工作

现有的时间序列预测方法可分为统计模型、机器学习模型和深度学习模型。统计模型是用于预测稳定的自相关时间序列数据的经典模型。虽然向量自回归模型(Vector Autoregressive, VAR)^[14]早期用于处理多目标预测任务,但是它只考虑了稳定且自相关时间序列之间的关系,导致时空信息有限。机器学习模型在处理非线性和非平稳数据方面具有优势。其中,神经网络^[15]、高斯过程回归^[16]和支持向量回归^[17]等模型在小样本时间序列预测任务中取得了不错的效果。例如,文献^[18]提出了最小二乘支持向量回归模型,用于实现股票指数和债券指数预测。虽然统计模型和机器学习模型被广泛应用于时间序列预测,但由于地理传感器时间序列中存在多类型时空相关性,它们不能很好地扩展到该类预测中。

随着深度学习的出现,优秀的预测模型被提出。近年来,基于循环神经网络(Recurrent Neural Network, RNN)^[19]的编码-解码结构已被证明对序列预测问题非常有效。在编码阶段,RNN作为编码器将输入序列编码为固定长度的向量。在解码阶段,RNN作为解码器将固定长度的向量解码并输出预测序列。但随着输入序列长度的增加,编码-解码结构的性能会迅速恶化^[20]。针对这一问题,研究人员提出了基于注意力机制的编码-解码结构,利用注意力机制选择相关序列,以捕捉时空相关性。文献^[21]提出了一种基于self-attention和inter-attention的混合注意力网络,self-attention捕捉编码器和解码器中信息量丰富的潜在特征,而inter-attention能够强化编码器和解码器之间的联系。为推断变量之间的因果关系,文献^[22]在输入和LSTM之间建立注意力机制进行报警根源诊断。上述研究提出的注意力机制,多用于捕捉单目标预测中非预测序列在不同时间阶段对目标序列的影响,并不能区分序列之间的多种类型空间相关性。

卷积神经网络(Convolutional Neural Network, CNN)因为多线程GPU计算而流行。尽管传统上CNN是为图像数据开发的,但是也可用于建模多元时间序列任务^[23]。然而,CNN只能利用时间窗口内的观测数据进行预测。因此,需要

仔细调整时间窗口的大小,以确保模型可以利用所有相关的历史观测值。

3 AR 与深度神经网络联合模型

3.1 问题定义

假设在某个区域的不同地理位置部署了 n 个传感器,得到一组完全可观测的地理传感器时间序列 $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)=(x_1^1, x_1^2, \dots, x_1^n), \dots, (x_T^1, x_T^2, \dots, x_T^n)$, 其中 $\mathbf{x}_t=(x_t^1, x_t^2, \dots, x_t^n) \in \mathbb{R}^n$ 表示时间步 t 的 n 个序列, x_t^i 表示时间步 t 的第 i 个变量值, T 表示时间窗口大小。本文旨在进行多目标预测任务,问题描述为:

$$\hat{\mathbf{x}}_{T+1}=f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T) \quad (1)$$

其中, f 是本文要学习的非线性映射函数。

3.2 总体框架

图 2 给出了 J-ARDNN 的总体框架,该模型主要由 4 部分组成。1)空间模块,显式地捕捉序列之间的多类型空间相关性。首先,采用 IMV-LSTM 网络将输入序列转换为高层次特征,以获取足够的表达能力。然后,利用混合空间注意力机制计算不同序列对之间的权重系数,构建注意力权重矩阵,该矩阵随后作用于所有序列的高层次特征。2)时间模块,利用 TCN 提取单个序列内的时间相关性。3)线性模块,采用 AR 模型捕捉不同时间粒度的时间依赖关系。4)融合模块,集成深度神经网络模块和 AR 模块的输出,并将其作为最终预测。

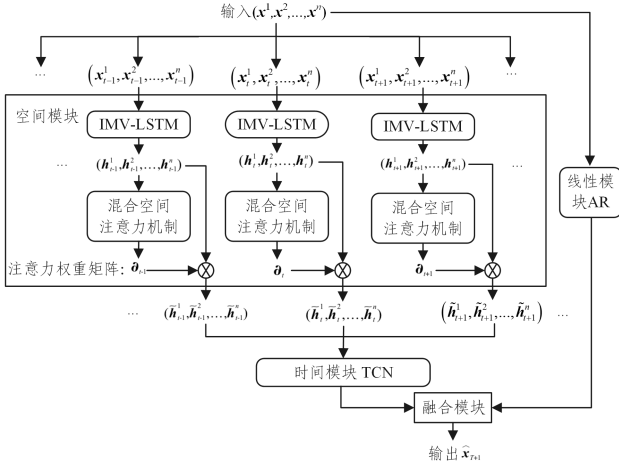


图 2 J-ARDNN 模型的框架

Fig. 2 Framework of J-ARDNN model

3.3 空间模块

由于 RNN 生成不透明的隐藏状态,缺乏对多变量数据的可解释性。具体来说,当输出多个变量的观测值并将其作为输入时,RNN 盲目地将所有的信息融合到用于预测的隐藏状态中。因此,通过隐藏状态难以区分单个变量对预测的贡献。在多目标预测任务中,单个变量通常呈现不同的动态,这一信息混合在多变量数据的隐藏状态中而被隐地忽略,从而潜在地阻碍了预测性能。为此,本文采用了可解释的多变量 LSTM 网络 (Interpretable Multi-Variable LSTM, IMV-LSTM)^[24],其思想是利用隐藏矩阵的每个元素封装输入的某个变量的信息,以获取足够的表达能力。

例如,给定一个具有两个变量的输入序列,IMV-LSTM 可以将输入特征编码为大小为 4×2 的隐藏矩阵,即每个变量

具有一个 4 维的隐藏状态。IMV-LSTM 作为一个标准的 LSTM 神经网络,包含一个输入门 i_t 、忘记门 f_t 、输出门 o_t 和记忆单元 s_t 。对于 n 个变量,IMV-LSTM 单元的迭代更新过程如下:

$$i_t = \sigma(\mathbf{W}_i \otimes \mathbf{H}_{t-1} + \mathbf{U}_i \otimes \mathbf{x}_t + \mathbf{b}_i) \quad (2)$$

$$f_t = \sigma(\mathbf{W}_f \otimes \mathbf{H}_{t-1} + \mathbf{U}_f \otimes \mathbf{x}_t + \mathbf{b}_f) \quad (3)$$

$$o_t = \sigma(\mathbf{W}_o \otimes \mathbf{H}_{t-1} + \mathbf{U}_o \otimes \mathbf{x}_t + \mathbf{b}_o) \quad (4)$$

$$j_t = \tanh(\mathbf{W}_j \otimes \mathbf{H}_{t-1} + \mathbf{U}_j \otimes \mathbf{x}_t + \mathbf{b}_j) \quad (5)$$

$$s_t = f_t \odot s_{t-1} + i_t \odot j_t \quad (6)$$

$$\mathbf{H}_t = o_t \odot \tanh(s_t) \quad (7)$$

其中, $\mathbf{H}_t = [\mathbf{h}_t^1, \mathbf{h}_t^2, \dots, \mathbf{h}_t^n]$ 表示当前时间步隐藏状态矩阵,行向量 $\mathbf{h}_t^i \in \mathbb{R}^m$ 表示 t 时刻第 i 个变量的隐藏状态, $\mathbf{x}_t \in \mathbb{R}^n$ 是当前的输入。更新矩阵 $j_t \in \mathbb{R}^{n \times m}$ 用于更新记忆单元。过度张量 $\mathbf{W} \in \mathbb{R}^{n \times m \times m}$ 和 $\mathbf{U} \in \mathbb{R}^{n \times m}$ 是可学习参数。 \otimes 表示张量点运算,为两个张量沿 n 轴的乘积,例如, $\mathbf{W}_j * \mathbf{H}_{t-1} = [\mathbf{W}_j^1 \mathbf{h}_{t-1}^1, \mathbf{W}_j^2 \mathbf{h}_{t-1}^2, \dots, \mathbf{W}_j^n \mathbf{h}_{t-1}^n] \in \mathbb{R}^{n \times m}$, 其中 $\mathbf{W}_j^i \mathbf{h}_{t-1}^i \in \mathbb{R}^m$ 。 σ 和 \odot 分别是 sigmoid 函数和哈达玛积。基于式 (2) — 式 (5), IMV-LSTM 网络也可以被认为是一组并行的 LSTM 网络,每个 LSTM 处理一个变量序列,然后融合隐藏状态。

一旦获得每个变量的特征,就将其作为混合空间注意力机制的输入。图 3 给出了混合空间注意力机制的计算过程。利用点积计算查询序列 i 与序列 $j \in N_{-i}$ 之间在时间步 t 的相似性得分。点积注意力机制计算效率高,因为它可以使用高度优化的矩阵乘法代码来实现^[25]。此外,序列之间的相关性受实际地理空间位置的影响。不同序列之间的地理空间相关性使用实际距离的倒数表示,记为矩阵 $\mathbf{G} \in \mathbb{R}^{n \times n}$, 其中元素 G^{ij} 表示序列 i 和 j 之间的地理空间相关性。地理空间相关性矩阵 \mathbf{G} 是固定不变的,可以被认为是先验知识。因此,通过混合空间注意力机制计算时间步 t 中序列 i 与序列 $j \in N_{-i}$ 之间的相似度得分。

$$e_i^j = \mathbf{W}_i^j (\mathbf{h}_i^t)^T \mathbf{h}_j^t + \mathbf{W}_G^j G^{ij} \quad (8)$$

其中, $N_{-i} = \{1, \dots, i-1, i+1, \dots, n\}$ 是除变量 i 以外其他变量编号的集合, \mathbf{W}_i^j 和 \mathbf{W}_G^j 是用于调节不同类型相关性的可学习权重系数,并且 $\mathbf{W}_i^j + \mathbf{W}_G^j = 1$ 。如果 $\mathbf{W}_i^j < \mathbf{W}_G^j$, 则说明序列之间的相关性受地理空间因素的影响更大。

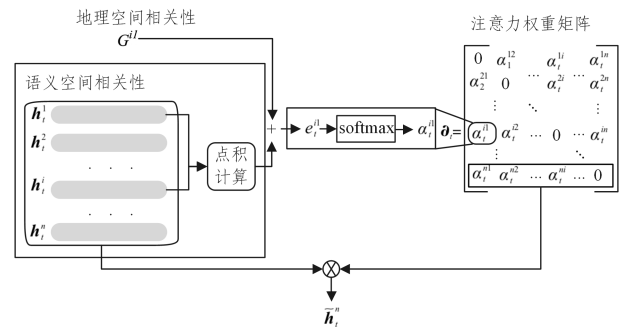


图 3 混合空间注意力机制的计算过程

Fig. 3 Computational process of hybrid spatial attention mechanism

为了使不同变量之间的权重系数易于比较,使用 softmax 函数对所有 i 的选择进行归一化:

$$\alpha_i^j = \text{softmax}(e_i^j) = \frac{\exp(e_i^j)}{\sum_{k \in N_{-i}} \exp(e_k^j)} \quad (9)$$

其中,权重 α_t^{ij} 用于衡量变量 j 对 i 的重要性。组合不同变量之间的注意力权重,得到如下矩阵:

$$\alpha_t = \begin{bmatrix} 0 & \alpha_t^{12} & \cdots & \alpha_t^{1i} & \cdots & \alpha_t^{1n} \\ \alpha_t^{21} & 0 & \cdots & \alpha_t^{2i} & \cdots & \alpha_t^{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha_t^{i1} & \alpha_t^{i2} & \cdots & 0 & \cdots & \alpha_t^{in} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_t^{n1} & \alpha_t^{n2} & \cdots & \alpha_t^{ni} & \cdots & 0 \end{bmatrix}$$

因为序列之间的相关性是单向的,所以 $\alpha_t^{ij} \neq \alpha_t^{ji}$ 。

注意力权重矩阵直接作用于隐藏状态矩阵 \mathbf{H}_t 得到如下上下文矩阵:

$$\tilde{\mathbf{H}}_t = \alpha_t \mathbf{H}_t \quad (10)$$

其中, $\tilde{\mathbf{H}}_t = (\tilde{h}_t^1, \tilde{h}_t^2, \dots, \tilde{h}_t^j)$ 是具有空间相关性的新特征。每个变量的新特征计算为所有相关特征的加权和:

$$\tilde{h}_t^j = \sum_{j \in N_{t-1}} \alpha_t^{ij} h_t^j \quad (11)$$

3.4 时间模块

时间模块采用 TCN 提取单个序列的时间依赖关系。TCN 应用于时间序列存在两个优势:1)网络架构中的卷积是具有因果关系的,以保证未来的信息不影响过去信息;2)网络架构具有与 RNN 一样的特性,可以接受任意长度的序列输入并将其映射到相同长度的输出数据上。TCN 采用一维全卷积网络和残差模块训练模型来避免梯度消失问题。残差模块由两层扩张因果卷积和非线性模型组成,其中使用 ReLU 作为激活函数,如图 4 所示。为了避免过拟合,在每个扩张因果卷积之后添加 Dropout 层进行正则化。此外,添加额外 1×1 卷积以考虑不同的输入和输出长度^[26]。与 RNN 按顺序处理时间特征不同,TCN 的每一层网络都使用相同的滤波器,卷积操作可以并行完成。因此,在训练和评估中,一个很长的输入序列在 TCN 中可以作为一个整体进行处理。

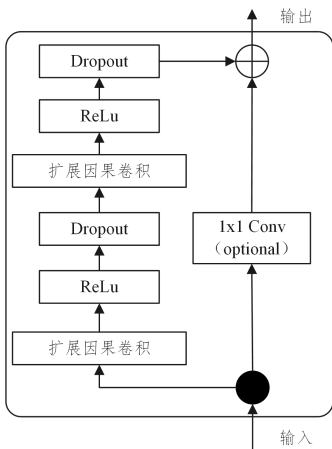


图 4 TCN 的残差模块

Fig. 4 TCN residual block

扩张因果卷积能够实现指数级大的感受野,以处理非常长的序列^[27]。假设给定一个一维的输入序列 x 和一个滤波器 $f: \{0, 1, \dots, k-1\}$, 序列元素 s 上的扩张卷积运算 F 定义为:

$$F(s) = (x * {}_l f)(s) = \sum_{i=0}^{k-1} f(i) x_{s-l+i} \quad (12)$$

其中, l 是扩张因子, k 是滤波器大小, $s-l \cdot i$ 表示过去的方向, $*$ 表示卷积操作。图 5 给出了扩张因果卷积结构,扩张因子 $l = \{1, 2, 4\}$, 滤波器大小 $k = 3$ 。从图中可以观察到,增大滤波器大小 k 和扩张因子 l 可以扩大网络的感受野。感受野越大,TCN 可以越长远地提取历史的信息。最终,TCN 模型生成与输入维度相同的输出。

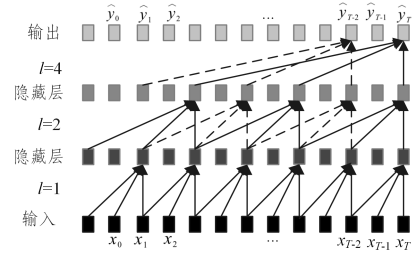


图 5 扩展因果卷积

Fig. 5 Dilated causal convolution

3.5 线性模块

由于 RNN 和 TCN 具有非线性的特性,导致其输出的尺度对输入尺度不敏感^[7]。在真实数据集中,输入数据的尺度以非周期的方式不断变化,大大降低了神经网络模型的预测精度。为了解决这个问题,J-ARDNN 模型添加了一个额外的 AR 模型作为线性模块,主要关注局部缩放问题。

AR 模型可以分析和表征单个变量数据内部的相互依赖性和相关性。具体来说,预测值由时间步 t 之前的 T_{AR} 个观测值加权而成。假设给定序列的第 i 个序列,AR 模型的定义如下:

$$\hat{x}_{t,i}^{AR} = \sum_{k=0}^{T_{AR}-1} W_k^{AR} x_{t-k}^{AR} + b^{AR} \quad (13)$$

其中, W_k^{AR} 和 b^{AR} 是 AR 模型的系数, T_{AR} 表示时间窗口大小。在 AR 模型中,所有维度共享同一组线性参数。针对多目标预测,AR 模型的最终预测结果记为 $\hat{x}_t^{AR} \in \mathbb{R}^n$ 。

3.6 预测和目标函数

深度神经网络预测结果通过对时空特征做线性映射得到:

$$\hat{\mathbf{x}}_t^D = \mathbf{v}_y^T (\mathbf{W}_y^T \tilde{\mathbf{H}}_t + \mathbf{b}_w^T)^T + \mathbf{b}^T \quad (14)$$

其中,参数 $\mathbf{W}_y \in \mathbb{R}^n$ 和 $\mathbf{b}_w \in \mathbb{R}^m$ 将新特征映射到维度为 m 。最后,使用线性变换 $\mathbf{v}_y \in \mathbb{R}^m$ 和 $\mathbf{b} \in \mathbb{R}$ 生成预测结果。

J-ARDNN 模型的最终预测结果通过集成深度神经网络模块和 AR 模块的输出得到:

$$\hat{\mathbf{x}}_t = \mathbf{W}_D \hat{\mathbf{x}}_t^D + \mathbf{W}_{AR} \hat{\mathbf{x}}_t^{AR} \quad (15)$$

其中, $\hat{\mathbf{x}}_t$ 表示模型在时间步 t 生成的最终预测结果, \mathbf{W}_D 和 \mathbf{W}_{AR} 表示权重系数,且 $\mathbf{W}_D + \mathbf{W}_{AR} = 1$ 。

本文方法的所有参数使用 MSE 作为目标函数来学习,这是一种常见的用于时间序列预测任务的损失函数,其表达式为:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N (x_i^j - \hat{x}_i^j)^2 \quad (16)$$

其中, θ 表示 J-ARDNN 所有可学习参数, N 表示样本数目, x_i^j 和 \hat{x}_i^j 分别表示时间步 t 的真实值和预测值。

4 实验

4.1 实验数据

为了评估 J-ARDNN 的性能,使用 3 个真实的地理传感器时间序列数据集:北京空气质量数据集、交通流量 (PEMS08)数据集和爱尔兰天气数据集,如表 1 所列。

1)北京空气质量数据集:该数据集收集了北京 12 个空气质量监测点的 PM2.5 浓度。该数据集每一小时采样一次,涵盖了 2013 年 3 月 1 日—2017 年 2 月 28 日的数据。

2)交通流量 (PEMS08)数据集:该数据集由从加州交通局绩效测量系统中收集的 170 个传感器的数据组成,这些数据从原始数据聚合为 5 分钟窗口^[28]。数据集的范围为 2016 年 7 月 1 日—2016 年 8 月 31 日。受实验环境的限制,我们决定简化实验数据,选择数据集中的前 20 个传感器数据来验证 J-ARDNN 模型,以解决机器的内存溢出问题。

3)爱尔兰天气数据集:该数据集记录了 2018 年 1 月 1 日—2022 年 2 月 1 日的爱尔兰 23 个气象站的每小时气温数据。

在现实世界中,由于传感器电源中断或通信错误,以上数据集中都存在缺失数据。在本实验中,采用线性插值法填补缺失值。3 个数据集均按照 6:2:2 的比例划分为训练集、验证集和测试集。

表 1 数据集的详细信息

Table 1 Dataset details

数据集	传感器数	时间间隔	时间跨度	样本数
北京 PM2.5 数据集	12	1 h	2013-03-01— 2017-02-28	35065
交通流量数据集	20	5 min	2016-07-01— 2016-08-31	17857
爱尔兰天气数据集	23	1 h	2018-01-01— 2022-02-01	35810

4.2 对比方法

为了验证所提出的 J-ARDNN 模型在多目标预测中的有效性,本实验选择统计模型和性能优异的深度学习模型作为对比方法。对比方法的介绍如下:

1)VAR-MLP^[29]:该模型是向量自回归和多层感知器的混合模型。

2)GP^[30]:GP 是用于时间序列建模的高斯过程。

3)RNN_GRU:GRU 是 RNN 的一种变体,在使用门控函数方面类似于 LSTM,但与 LSTM 不同的是,它没有存储单元。它不仅解决了 RNN 的梯度爆炸和消失问题,还可以捕捉时间序列中的长期依赖关系。

4)LSTNet^[7]:该模型结合了 CNN 和 RNN 的神经网络,以提取多目标预测任务中变量之间短期局部依赖关系,并发现时间序列趋势的长期模式。

5)TPA-LSTM^[8]:该模型首先使用 CNN 提取时间模型,然后采用注意力机制选择相关的时间序列,并利用其频域信息进行多目标预测。

6)MTNet^[9]:该模型利用记忆网络的思想来处理多目标预测任务,旨在计算输入数据和存储在记忆组件中的

数据之间的相似度。

7)MTGNN^[11]:该模型是专门为多目标序列预测任务设计的通用图神经网络框架。首先,图学习模块自动提取变量之间的单向关系;然后,图卷积模块与时间卷积模块交错,分别捕捉空间和时间相关性。

4.3 实验设置和评价指标

本实验中,执行网格搜索策略,为 J-ARDNN 模型中的超参数选择最优值。具体来说,J-ARDNN 和对比方法共享时间窗口 T 的搜索范围,即 $\{5, 10, 15, 20, 25\}$ 。对于具有 RNN 模块和 CNN 模块的对比模型,设置隐藏单元大小的范围为 $\{16, 32, 64, 128, 256\}$ 。对于 VAR-MLP 模型,密集层的大小在 $\{32, 50, 100\}$ 上执行网格搜索。对于 GP 模型,设置径向基函数 (Radial Basis Function, RBF) 的核宽度 σ 和噪声级别 α 的搜索范围为 $\{2^{-10}, 2^{-8}, \dots, 2^8, 2^{10}\}$ 。在训练过程中,epoch 设置为 100, batch size 设置为 128, 初始学习率被设置为 0.001, 每 10 个 epoch 衰减 10%。

为了评估 J-ARDNN 和其他预测模型的性能,考虑了 3 种不同的评价指标。其中平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Squared Error, RMSE) 两种评价指标在时间序列预测中被广泛使用,用于度量预测值与观测值之间的误差。MAE 和 RMSE 的值越小,说明预测误差越小,模型的预测精度越高。此外,本文利用决定系数 (R-Square, R^2) 判断模型的拟合效果。 R^2 的取值范围确定为 $[0, 1]$, R^2 值越接近 1, 说明模型的拟合效果越好。假设 y_t 是 t 时刻的真实值, \hat{y}_t 是 t 时刻的预测值, \bar{y} 是真实值的平均值, N 是样本数,评价指标的定义如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (17)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{\sum_{i=1}^N |y_i - \bar{y}|} \quad (19)$$

4.4 实验结果分析

本节总结了本文方法 J-ARDNN 和对比方法在 3 个数据集上获得的实验结果,并使用 3 个指标对预测性能进行了评估。表 2 列出了所有模型的预测性能,黑体标记的数据为每个数据集的最佳预测结果。在 3 个数据集上,J-ARDNN 的 3 个评价指标明显优于所有对比模型。具体来说,可以得出以下结论:

1)从表中可以观察到,传统基线方法 VAR-MLP 和 GP 表现不佳。以北京 PM2.5 数据集为例,J-ARDNN 的 MAE 值比 VAR-MLP 和 GP 降低了 58.1% 和 45.2%, 并且 J-ARDNN 的 R^2 值更接近 1, 说明模型的拟合效果更好。这是因为 VAR-MLP 对具有复杂时空相关性的多元时间序列数据的建模能力有限。

2)与传统基线方法相比,基于深度学习的模型取得了更好的效果,这说明它们可以解决复杂的预测任务。GRU 通过门控制单元捕捉长期的时间依赖关系,因此预测性能相比 VAR-MLP 和 GP 得到显著提高。然而,GRU 不能同时对

空间相关性和时间依赖性进行建模。

3)与GRU相比,LSTNet,TPA-LSTM和MTNet等混合模型结合了CNN和LSTM学习时间步间和不同时间序列之间的复杂非线性依赖关系,并且预测性能得到进一步提高。例如,在交通流数据集上,LSTNet,TPA-LSTM和MTNet比GRU分别降低了22.2%,31.6%和30.7%。LSTNet模型引入了跳跃循环神经网络,用于捕捉具有周期性的数据的长期依赖性,因此,在处理本章中的非周期性数据集时,性能略有

下降。此外,LSTNet,TPA-LSTM和MTNet只通过CNN提取局部空间特征,且没有显式地对不同序列之间的空间相关性建模,因此,它们的性能比MTGNN和J-ARDNN差。

4)MTGNN通过图学习层对不同序列之间的空间相关性进行抽样学习,导致无法考虑多类型的空间相关性,因此其性能逊于J-ARDNN。在北京PM2.5数据集、交通流量数据集和爱尔兰天气数据集上,J-ARDNN的MAE值比MTGNN降低了2.4%,0.7%和3.2%。

表2 在地理传感器时间序列数据集上的预测结果

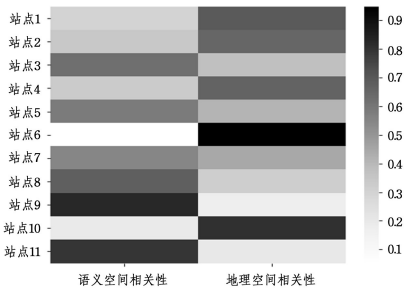
Table 2 Prediction results on geo-sensory time series datasets

methods	北京PM2.5数据集			交通流量数据集			爱尔兰天气数据集		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
VAR-MLP	40.1060	22.9476	0.8058	42.7400	32.3216	0.9104	2.3498	1.7683	0.7288
GP	32.1587	17.5391	0.8742	39.6772	30.8984	0.8794	1.8239	1.3974	0.8777
RNN_GRU	24.0183	13.0627	0.9312	33.4994	24.0523	0.9452	1.3886	1.0460	0.9292
LSTNet	19.4641	10.1711	0.9438	27.9205	18.7227	0.9536	0.7556	0.5552	0.9720
TPA-LSTM	18.3986	9.9748	0.9503	27.0170	18.1389	0.9561	0.6832	0.4987	0.9773
MTNet	18.4028	10.0850	0.9477	27.6302	18.4691	0.9538	0.7181	0.5256	0.9810
MTGNN	18.3735	9.8448	0.9526	27.0472	18.0009	0.9590	0.6846	0.4908	0.9827
J-ARDNN	17.9176	9.6080	0.9569	26.1712	17.8777	0.9618	0.6615	0.4751	0.9838

总体而言,J-ARDNN考虑了不同序列之间复杂的空间相关性和单个序列内的时间相关性。因此,J-ARDNN能够在多目标预测任务中取得最佳的性能。

4.5 可解释性分析

在混合空间注意力机制中,通过可学习调节权重 W_{ij}^s 和 W_{ij}^g 来确定语义空间相关性和地理空间相关性的重要性。图6可视化了北京PM2.5数据集中站点0与站点1—站点11之间的语义空间相关性和地理空间相关性的调节权重 W_{ij}^s 和 W_{ij}^g ,颜色越深表示权重值越大。



影响更大,即 $W_{ij}^g > W_{ij}^s$;站点0与站点3、站点5、站点7—站点9、站点11因外部条件相似而具有更强的语义空间相关性,即 $W_{ij}^s > W_{ij}^g$ 。上述实验结果表明,具有相关性调节机制的混合空间注意力机制成功地捕捉到了多类型时空相关性。

4.6 消融实验

为了进一步说明不同模块的有效性,本文针对3种变体在北京PM2.5数据集、交通流量数据集和爱尔兰天气数据集上进行了详细的消融实验。具体来说,依次从本文所提的J-ARDNN模型中删除每个组件,然后对比变体模型和完整模型的性能。首先,将变体模型命名为:

1)J-ARDNNw/oAttn,该变体删除了J-ARDNN模型中的混合空间注意力机制。

2)J-ARDNNw/oTCN,该变体删除了J-ARDNN模型中的时间模块TCN。

3)J-ARDNNw/oAR,该变体删除了J-ARDNN模型中的线性模块AR。

表3列出了在3个数据集上的消融实验的预测结果,可以得出如下结论:

1)与其他变体模型相比,J-ARDNN在每个数据集上取得了最佳的多目标预测结果。

2)变体模型J-ARDNNw/oAttn在数据集上的预测性能显著下降,说明混合空间注意力机制捕捉不同序列之间的多类型空间相关性,对提升多目标预测效果有至关重要的作用。

图6 在北京PM2.5数据集上混合空间注意力机制的调节权重热力图

Fig. 6 Adjustment weight heatmap of hybrid spatial attention mechanism on Beijing PM2.5 dataset

从图中可以明显观察到,因站点0与1,2,4,6,10之间具有地理空间临近性,导致序列之间的相关性受地理空间因素

表3 消融实验的预测结果

Table 3 Prediction results of ablation experiments

methods	北京PM2.5数据集			交通流量数据集			爱尔兰天气数据集		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
J-ARDNNw/oAttn	18.8417	10.4288	0.9400	29.1372	19.2819	0.9502	0.8344	0.6747	0.9751
J-ARDNNw/oTCN	18.6096	10.3731	0.9432	28.9250	18.8938	0.9516	0.8263	0.6702	0.9752
J-ARDNNw/oAR	18.4708	10.1273	0.9486	27.8853	18.6317	0.9522	0.7632	0.5792	0.9718
J-ARDNN	17.9176	9.6080	0.9569	26.1712	17.8777	0.9618	0.6615	0.4751	0.9838

3) 变体模型 J-ARDNNw/oTCN 和 J-ARDNNw/oAR 的预测性能下降,表明采用 TCN 和 AR 模型能够改善模型的性能损失。这是因为 TCN 能够捕捉单个序列内的长期依赖性,而 AR 通常对数据的尺度变化具有鲁棒性。

综上所述,将所有组件组建在一起的 J-ARDNN 模型在所有数据集上都取得了鲁棒的预测性能。因此,在建立多元时间序列预测模型时,有必要考虑不同模型的特点和优劣,选择合适的模型结构来提高模型的预测性能。

结束语 本文构建了一种 AR 与深度神经网络的联合模型,用于地理传感器时间序列的多目标预测任务。为了捕捉时空相关性,分别设计了捕捉序列间不同类型空间相关性的空间模块和学习单个序列内时间相关性的时间模块。此外,将线性模型和深度神经网络结合可以提高模型的鲁棒性。在 3 个数据集上的实验结果表明,本文模型明显优于目前先进的方法。

未来的工作主要包括两个方面。1) 由于传感器故障、数据传输失败或存储损坏等原因,本文收集的数据是不完整的。缺失的读数不仅会影响实时监测,而且在紧急情况下,还会影响预测和推断等进一步数据分析的性能。目前,本文采用了线性插补法补充缺失数据。然而,线性插补法考虑临近两个数据来进行数值估计,不能捕捉变量之间的空间相关性和复杂的时间模式来进行插补。为了更好地估计缺失数据,考虑全局时空动态关系插补缺失的时间序列数据将是下一步的研究重点。2) 本文方法在未来将扩展为长期多目标时间序列,为一段时间内的资源调度和决策提供指导。

参 考 文 献

- [1] LIANG Y, KE S, JUNZ, et al. GeoMAN: Multi-level Attention Networks for Geo-sensory Time Series Prediction[C]// Twenty-Seventh International Joint Conference on Artificial Intelligence. 2018:3428-3434.
- [2] YIN C, DAI Q. A deep multivariate time series multi-step forecasting network[J]. Applied Intelligence, 2022, 52(8): 8956-8974.
- [3] CHEN D, CHEN L, ZHANG Y, et al. A Multiscale Interactive Recurrent Network for Time-Series Forecasting[J]. IEEE Transactions on Cybernetics, 2022, 52(9): 8793-8803.
- [4] CANDANEDO L, FELDHEIM V. Data driven prediction models of energy use of appliances in a low-energy house[J]. Energy and Buildings, 2017, 140(4): 81-97.
- [5] MA Z, LIU S, GUO G, et al. Hybrid Attention Networks for Flow and Pressure Forecasting in Water Distribution Systems[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [6] DU S D, LI T R, YANG Y, et al. A Sequence-to-Sequence Spatial-Temporal Attention Learning Model for Urban Traffic Flow Prediction[J]. Journal of Computer Research and Development, 2020, 57(8): 1715-1728.
- [7] LAI G, CHANG W, YANG Y, et al. Modeling long- and short-term temporal patterns with deep neural networks[C]// Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: Association for Computing Machinery, 2018:95-104.
- [8] SHIH S, SUN F, LEE H. Temporal pattern attention for multivariate time series forecasting[J]. Machine Learning, 2019, 108: 1421-1441.
- [9] CHANG Y, SUN F, WU Y, et al. A Memory-Network Based Solution for Multivariate Time-Series Forecasting[J]. arXiv: 1809.02105, 2018.
- [10] HUANG S, WANG D, WU X, et al. DSANet: Dual Self-Attention Network for Multivariate Time Series Forecasting[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: Association for Computing Machinery, 2019:2129-2132.
- [11] WU Z, PAN S, LONG G, et al. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks[C]// Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2020:753-763.
- [12] JAIN A, KUMAR A M. Hybrid neural network models for hydrologic time series forecasting[J]. Applied Soft Computing, 2007, 7(2): 585-592.
- [13] HUANG G, WANG L. Hybrid Neural Network Models for Hydrologic Time Series Forecasting Based on Genetic Algorithm[C]// Proceedings of the 4th International Joint Conference on Computational Sciences and Optimization. New York: IEEE Press, 2011:1347-1350.
- [14] FENG X, CHEN J, ZHANG Z, et al. State-of-charge estimation of lithium-ion battery based on clockwork recurrent neural network[J]. Energy, 2021, 236: 1-10.
- [15] ZHANG Y, PENG N, DAI M, et al. Memory-Gated Recurrent Networks[C]// Proceedings of the 35th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2021: 10956-10963.
- [16] MA Q, LIN Z, CHEN E, et al. Temporal pyramid recurrent neural network[C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2020: 5061-5068.
- [17] WANG X, ZHANG M, REN F. Sparse Gaussian Conditional Random Fields on top of Recurrent Neural Networks[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2018: 4219-4226.
- [18] MALDONADO S, GONZALEZ A, CRONE S. Automatic time series analysis for electric load forecasting via support vector regression[J]. Applied Soft Computing, 2019, 83(3): 1-10.
- [19] LIU J X, CHEN S. Non-Stationary Multi-variate Time Series Prediction with MIX Gated Unit[J]. Journal of Computer Research and Development, 2019, 56(8): 1642-1651.
- [20] LI Z X, LIU H Y. Combining Global and Sequential Patterns for Multivariate Time Series Forecasting[J]. Chinese Journal of Computers, 2023, 46(1): 70-84.
- [21] WANG J, SUN T, LIU B Y, et al. CLVSA: A Convolutional

- LSTM Based Variational Sequence-to-Sequence Model with Attention for Predicting Trends of Financial Markets[C] // Proceedings of International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2021: 3705-3711.
- [22] OZDEMIR A, BULUS K, ZOR K. Medium-to long-term nickel price forecasting using LSTM and GRU networks[J]. Resources Policy, 2022, 78: 1-10.
- [23] LIM B, ZOHREN S. Time Series Forecasting With Deep Learning: A Survey[J]. Philosophical Transactions of the Royal Society A, 2021, 379: 1-12.
- [24] GUO T, LIN T, ANTULOV-FANTULIN N. Exploring Interpretable LSTM Neural Networks over Multi-Variable Data [C] // Proceedings of the 36th International Conference on Machine Learning. New York: ACM, 2019: 4424-4440.
- [25] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5999-6009.
- [26] OORD A, DIELMAN S, ZEN H, et al. WaveNet: A Generative Model for Raw Audio[J]. arXiv:1609.03499v2, 2016.
- [27] GAN Z, LI C, ZHOU J, et al. Temporal convolutional networks interval prediction model for wind speed forecasting[J]. Electric Power Systems Research, 2021, 191: 1-10.
- [28] LI M, ZHU Z. Spatial-Temporal Fusion Graph Neural Networks for Traffic Flow Forecasting [C] // Proceedings of the 35th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2021: 4189-4196.
- [29] ZHANG G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50(6): 159-175.
- [30] FRIFOLA R. Bayesian Time Series Learning with Gaussian Processes[D]. Cambridgeshire: University of Cambridge, 2015.



DONG Hongbin, born in 1963, Ph.D., professor, Ph.D supervisor, is a senior member of China Computer Federation. His main research interests include natural computation, multi-agent systems, machine learning, and data mining.



HAN Shuang, born in 1990, postgraduate. Her main research interests include time series forecasting and deep learning.

(责任编辑:李亚辉)