

基于多任务学习的短期风能发电功率预测研究

杜明山, 张海清, 李代伟, 唐 鑫

(成都信息工程大学 软件工程学院, 四川 成都 610200)

摘 要: 为提高短期风能发电功率预测的准确性, 提出一种硬参数共享的多任务学习模型, 分为参数共享层和任务指定层。在参数共享层中, 一方面, 由于时序数据中不仅有时间上的依赖, 而且有跨维度的依赖, 有助于提升预测精度, 因此提出跨维度—时间依赖混合编码模块(CD-TDMEM); 另一方面, 鉴于不同尺度的信息对预测未来结果所起到的不同作用, 针对多层次的高低频分量分别进行编码, 构建多尺度信息编码模块(MSIEM)。在任务指定层中, 对于每个子任务, 首先对多尺度模块的编码信息和跨维度—时间依赖混合编码信息进行融合, 经过多层解码器解码, 最终使用Transformer生成式预测模型的一步解码得到每个子任务的预测输出。实验结果表明, 对于短期风能发电功率预测任务, 相比其他模型, 该模型在传统的单任务学习上MSE分别降低了13.97%、69.78%、75.36%、2.30%、7.18%, MAE分别降低了7.33%、53.82%、54.29%、-2.07%、4.32%。而利用多任务学习机制之后, 相比单任务学习, MSE进一步降低了1.94%, MAE进一步降低了7.00%, 表明多任务学习的预测效果更好。另外, 相比其他几个多任务学习模型, MSE分别降低了9.97%、8.27%、3.34%, MAE分别降低了16.43%、10.74%、7.96%。消融实验也证明了各模块的有效性。

关键词: 风能发电功率预测; 多尺度信息; 跨维度—时间依赖; 多任务学习

DOI: 10.11907/rjdk.241830

开放科学(资源服务)标识码(OSID):

中图分类号: TP301.6

文献标识码: A

文章编号: 1672-7800(2025)004-0032-10



Research on Short-Term Wind Power Prediction Based on Multi-Task Learning

DU Mingshan, ZHANG Haiqing, LI Daiwei, TANG Xin

(School of Software Engineering, Chengdu University of Information Technology, Chengdu 610200, China)

Abstract: In order to improve the accuracy of short-term wind power prediction, a multi-task learning model with hard parameter sharing is proposed. It is divided into parameter sharing layer and task specifying layer. In the shared layer, on the one hand, since there are not only time dependencies in the time series data, cross-dimensional dependencies also help to improve the prediction accuracy. Therefore, a cross-Dimensional-temporal Dependency Mixing Encoding Module (CD-TDMEM) is proposed. On the other hand, in view of the different effects of different scales of Information on predicting future results, multi-scale Information Encoding Module (MSIEM) is constructed for multi-level high and low frequency components. In the task specification layer, for each subtask, the coding information of the multi-scale module and the cross-dimension-time-dependent hybrid coding information are fused, and then decoded by the multi-layer decoder. Finally, the one-step decoding of the Transformer generative prediction model is used to obtain the prediction output of each subtask. The experimental results show that for the short-term wind power prediction task, the MSE of the model constructed in this paper decreases by 13.97%, 69.78%, 75.36%, 2.30%, 7.18%, and MAE decreases by 7.33%, 53.82%, 54.29%, -2.07%, 4.32%, respectively, in the traditional single-task learning task. After using the multi-task learning mechanism, MSE and MAE are further reduced by 1.94% and 7.00% compared with single-task learning. The results show that the prediction effect of multi-task learning is better. Compared with some other multi-task learning models, MSE decreased by 9.97%, 8.27%, 3.34%, and MAE decreased by 16.43%, 10.74%, 7.96%, respectively. The ablation experiment also shows the effectiveness of each module.

Key Words: wind power prediction; multi-scale information; cross dimensional-temporal dependency; multi-task learning

收稿日期: 2024-09-27

扫描二维码阅读全文:



基金项目: 国家自然科学基金项目(61602064); 四川省科技厅项目(2021YFH0107, 2022NSFSC0571)

作者简介: 杜明山(1994-), 男, 成都信息工程大学软件工程学院硕士研究生, 研究方向为人工智能、时序预测; 张海清(1986-), 女, 博士, 成都信息工程大学软件工程学院教授、博士生导师, 研究方向为数据挖掘、机器学习、人工智能; 李代伟(1976-), 男, 博士, 成都信息工程大学软件工程学院副教授, 研究方向为数据挖掘、机器学习、粗糙集; 唐鑫(1998-), 男, 成都信息工程大学软件工程学院硕士研究生, 研究方向为人工智能、时序预测。本文通讯作者: 张海清。

0 引言

随着能源消耗和环境污染的日益加剧,全球都在寻求可替代化石能源的清洁能源,其中风能作为一种可再生能源引起了大量关注,在过去的几十年间得到了迅速发展。但是,由于风能发电功率的强波动性和间歇性,给电网的平稳、安全运行带来严峻挑战。一方面,在风电功率大波动情况下,为应对机组发电量和负荷需求的不平衡,电力调度部门需要预留大量的旋转备用容量,维持电力和用户两端的供需平衡^[1]。另一方面,风电电场需要向电力调度部门提前上报发电计划,当上报电量偏差较大时,风电场需承担电网考核费用、备用采购费以及上报电量过低产生的主动弃风损失^[2]。准确的风能发电功率预测能够大大降低风电对电网造成的不确定性干扰,提高对风电的消纳能力,节约运行成本。

风能发电功率预测按照预测时长分为超短期预测(从几秒钟到30 min)、短期预测(从30 min到6 h)、中期预测(从6 h到1天)和长期预测(从1天到1周或更长时间)^[3]。常用的风能发电功率预测方法分为:基于统计的方法、基于物理的方法、基于机器学习的方法3类。基于统计的方法通过分析过去的观测数据,找出其中的规律和相关性,从而对未来进行预测,其主要的缺点是对数据平稳性的假设并不能保证真实世界中数据的平稳性^[4-5]。基于物理的方法利用详细的物理特征对风能发电进行建模,其主要的缺点是所建立的模型受到地域限制以及计算量大等。当前更多研究使用深度学习方法或将深度学习与其他方法相结合,基于通用近似理论,深度学习方法可以任意精度来拟合一个函数,有着优异的非线性拟合能力,并能够动态学习和捕获数据中隐藏的模式。以上所提及的都是只对未来的风能功率进行预测,也即单任务学习(Single Task Learning, STL)方法。通常使用该方式能够获得可接受的结果,但是仅仅聚焦于单一任务上,可能会忽略有助于提升所关注指标的相关信息。具体来说,这些信息来源于相关任务的训练信号,通过共享相关任务之间的表示,可以使模型在原任务上获得更好的泛化性,也即多任务学习(Multi Task Learning, MTL)方法^[6]。

多任务学习在自然语言处理^[7-8]、计算机视觉^[9-11]、智能驾驶^[12]、医疗^[13-14]等领域都得到了广泛应用。多任务学习也适用于多变量时序预测,因为任务本身是相关的,并且用于不同任务的内部表示也是相关的^[15]。

文献[15]使用单任务学习和多任务学习对机器人行进过程中1~8 m远距离的声纳和相机图片进行预测,实验结果表明,多任务学习除在1 m距离预测结果的准确率略低外,在其他距离上都取得了比单任务学习更高的准确率。值得指出的是,相比单任务学习,其在更远距离上预测的准确性提升更多,也即是说,多任务学习有助于学习

更难、范围更大的预测任务。一个合理的解释为单个模型进行短时间的预测可能难以发现长期的时序过程,反之亦然。使用多任务学习在一个模型中进行不同时间尺度的训练,由于模型可以同时学习短期和长期两方面的过程,提升了将短期和长期过程结合的可能性,从而有助于进行预测。

近年来,多项研究也将多任务学习应用于风能发电功率预测。文献[16]提出一种硬参数共享的多任务学习预测模型,其将风能功率预测作为主任务,将风速预测作为附属任务。该模型使用长短期记忆网络(Long Short-Term Memory, LSTM)作为参数共享层,再将参数共享层所提取的特征分别与局部风能特征、局部风速特征进行融合,之后分别对风能和风速进行预测,结果表明预测效果优于单任务学习模型,但其缺少对复杂历史风能和风速信息的特征抽取。文献[17]针对新建风电场不能提供充足的源域数据(source domain data),从而无法利用传统的迁移学习对目标涡轮机进行建模的问题,构建了基于共享—私有(shared-private)机制的多任务学习模型。该模型由基于对偶空洞卷积的自注意力(Dual dilated convolution-based self-attention)子模型和自回归(Autoregressive)子模型组成,分别用于学习非线性成分和线性成分,最终将两者结合,输出预测结果。文献[18]为了有效捕捉与利用时间序列中的复杂模式和趋势,将WaveNet和MMoE(Multigate mixture-of-experts)结构进行结合提出了WaveNet-MMoE模型。WaveNet利用其可扩展的卷积网络对预处理后的数据进行特征抽取与长期时序信息捕获。然后将这些信息作为MMoE的输入,并将多步的预测视为多个独立但是相互关联的任务。实验对比了WaveNet-MMoE和多个单任务学习模型,WaveNet-MMoE都取得了更好的预测结果。值得指出的是,WaveNet-MMoE预测的准确性也高于Transformer。

根据上述分析,多任务学习有助于提升风能发电功率预测效果。然而,上述针对风能发电功率预测的研究主要在于挖掘时序上的依赖,而缺乏与跨维度特征交互信息的融合。另外,本文模型融合了多尺度信息,综合了不同粒度信息对未来预测序列的影响。

本文的主要工作如下:

(1)提出跨维度—时间依赖混合编码模块(Cross Dimension-Temporal Dependency Mixing Encoding Module, CD-TDMEM),先对每个特征进行扩展,进而分别对时间依赖信息和跨维度特征交互信息进行抽取,并利用路由机制对两方面信息进行融合。

(2)利用小波变换对特征进行多层次分解,所构建的多尺度信息编码模块(Multi Scale Information Encoding Module, MSIEM)分别针对多层次不同尺度的信息进行编码,以针对性地挖掘所要预测的未来时间序列在短时间和长时间的影响。

(3)相关性分析结果表明,风速和风能发电功率之间存在强相关性。将风能发电功率预测作为主任务,风速预测作为附属任务,构建了一种硬参数共享的多任务学习模型。

1 相关工作

1.1 相关性分析

对多任务学习来说,多个任务共享底层的网络参数,而在任务指定层配置每个任务特有的参数,虽然不同任务之间存在联系,但是联合训练不相关甚至冲突的任务,会导致任务之间出现负迁移现象,影响最终效果。因此,有必要针对风能发电功率与特征之间的相关性进行分析。本文利用 Pearson 相关系数进行了相关性分析。

Pearson 相关性分析是指分析两个变量之间的线性相关性,也即是说,当一个变量增大,另一个也会随之增大或减小。使用相关系数来度量两个变量之间的相关程度,计算公式如下:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (1)$$

该值位于 $[-1, 1]$ 区间内,根据不同的取值范围,将相关程度划分为5个等级,具体见表1。

Table 1 Rank of correlation

表1 相关性等级表

lr取值范围	相关程度
0.00~0.19	极低相关
0.20~0.39	低度相关
0.40~0.69	中度相关
0.70~0.89	高度相关
0.90~1.00	极高相关

相关性热力图如图1所示。

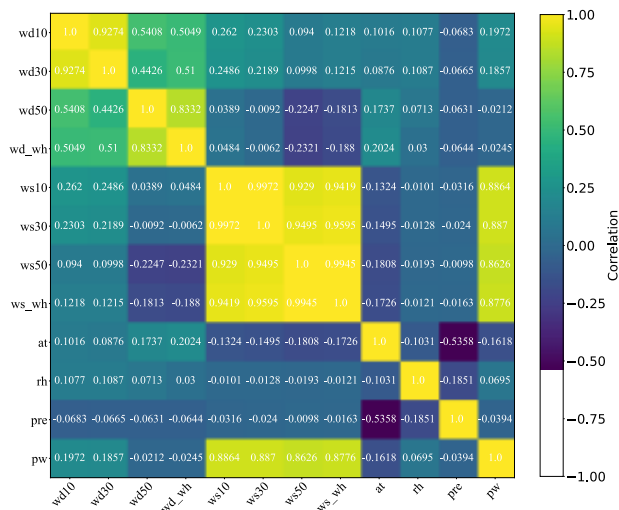


Fig. 1 The heat map of correlation

图1 相关性热力图

1.2 周期性分析

风能发电功率预测通常被定义为时序预测问题,周期性是时序数据的一个重要特性。为了发现时间序列中潜在的周期,需要计算功率谱密度。由于周期是频率的倒数,通过识别携带大部分能量的频率,可以发现主要的周期。循环自相关是估计功率谱密度的方法之一,通过计算自相关函数(AutoCorrelation Function ACF)估计一个时间序列的周期。具体公式如下:

$$ACF(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(\tau) \cdot x(n + \tau) \quad (2)$$

公式(2)计算了一个序列和不同滞后步序列的相似性,其中 τ 表示滞后的时间步。为了降低计算的复杂性,将其转化为计算频域的点积:

$$ACF = F^{-1} < X, X^* > \quad (3)$$

其中, X^* 表示 X 的复共轭, F^{-1} 表示逆傅里叶变换。由此得到风向、风速、发电功率的自相关函数图,如图2—图4所示。

通过各特征的自相关函数图可知,随着滞后时间步的增长,各特征的自相关性快速下降。温度、相对湿度在某些滞后时间步出现了明显的峰值,表明其存在易于发现的周期性,而气压、离地各高度处的风向与风速及发电功率并没有呈现出明显的周期性。

2 方法介绍

2.1 问题定义

对于单任务的风能发电功率多步预测,以给定的一个历史时间序列 $X = (x_1, x_2, \dots, x_T)$ 作为输入,预测未来时间序列 $Y = (y_1, y_2, \dots, y_H)$ 。其中, $x_t (t = 1, 2, \dots, T) \in R^D$, T 表示输入的历史时间序列长度, $y_h (h = 1, 2, \dots, H) \in R^M$, H 表示需要预测的时间序列长度。构建模型的目标就是找到一个输入序列和输出序列之间的非线性映射: $\bar{Y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_H) = F(X, \Theta)$,满足 $\min_{\{\Theta \in \Omega\}} (L(Y - \bar{Y}))$ 。其中, F 表示映射关系, Θ 表示模型参数, Ω 表示整个参数空间, L 表示损失函数, Y 表示预测目标的真实值, $\bar{Y}$ 表示预测结果。

对于多任务的风能发电功率多步预测,以给定的一个历史时间序列 $X = (x_1, x_2, \dots, x_T)$ 作为输入,预测多个未来时间序列 $Y = (y_1, y_2, \dots, y_K)$ 。其中, $x_t (t = 1, 2, \dots, T) \in R^D$, T 表示输入的历史时间序列长度, $y_k (k = 1, 2, \dots, K) \in R^{H \times M}$ 表示对应子任务的预测目标, K 表示子任务数量。所构建的多任务模型的目标是要找到一个输入序列和对应各子任务之间的非线性映射关系。具体定义为:

$$\bar{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix} = \begin{bmatrix} F_1(\Theta_1, \Theta_{sh}, X) \\ F_2(\Theta_2, \Theta_{sh}, X) \\ \vdots \\ F_k(\Theta_k, \Theta_{sh}, X) \end{bmatrix} \quad (4)$$

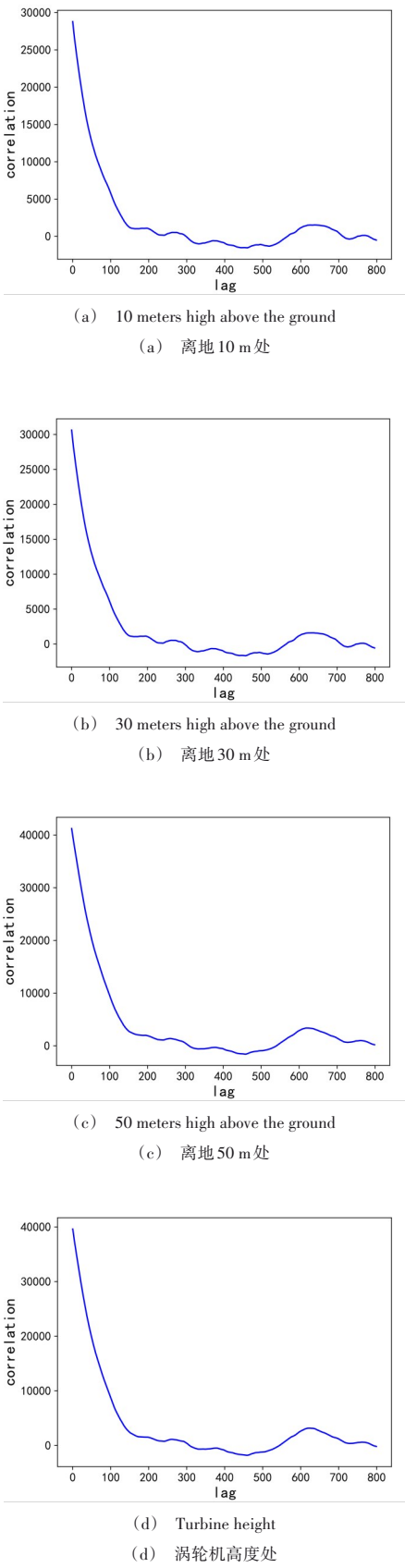


Fig. 2 Autocorrelation function plot of wind speed at different altitudes above the ground
图2 离地不同高度处风速自相关函数图

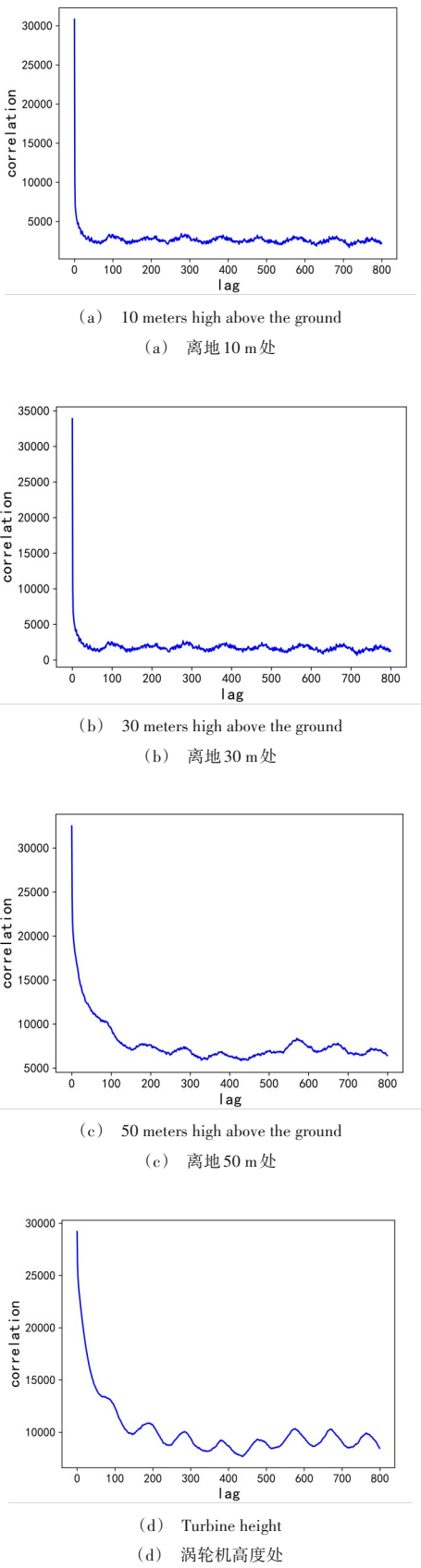


Fig. 3 Autocorrelation function plot of wind direction at different altitudes above the ground
图3 离地不同高度处风向自相关函数图

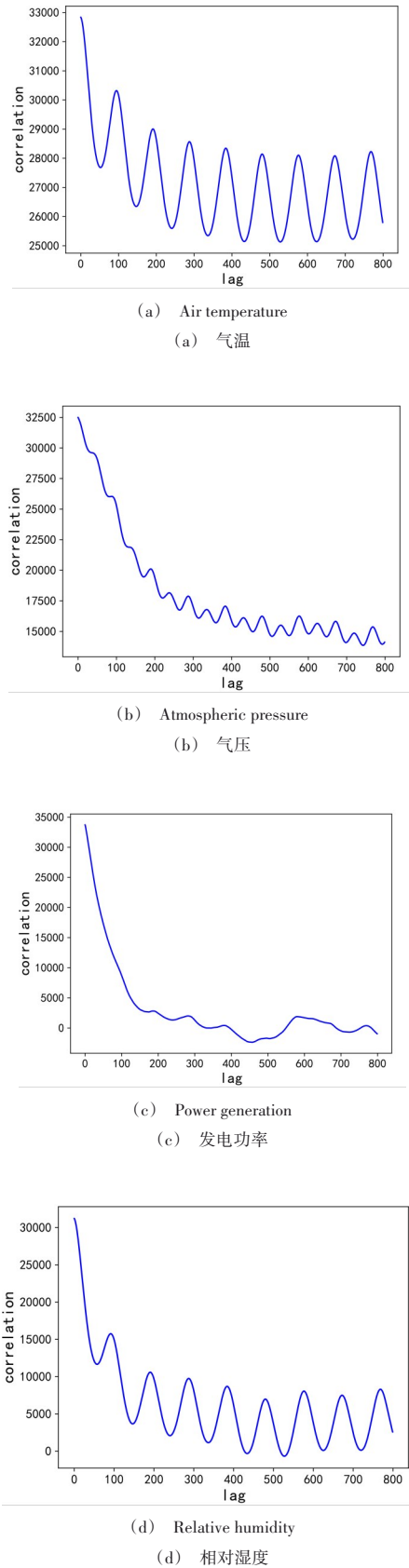


Fig. 4 Autocorrelation function plot of air temperature, atmospheric pressure, power generation and relative humidity
图4 气温、气压、发电功率、相对湿度自相关函数图

其中, $\min_{\theta \in \Omega} (W_i L_i (Y - \bar{Y}))$, $F_i (i = 1, 2, \dots, k)$ 表示对应子任务的映射关系, Θ_{sh} 表示共享层参数, $\Theta_i (i = 1, 2, \dots, k)$ 表示任务指定层参数。 $\Theta = \{\Theta_{sh}, \Theta_1, \Theta_2, \dots, \Theta_k\}$, W_i 表示子任务所对应的损失权重, L_i 表示子任务所对应的损失函数。

2.2 小波分解

小波分解是一个被广泛用于在时域和频域对时间序列进行多尺度分析的方法,其基本原理是通过使用一组高通和低通滤波将原始序列分解成若干个高频子序列和一个低频子序列。除第一层是对原始序列进行分解外,其余每一层都对其中的低频子序列作进一步分解。图5展示了小波分解过程。

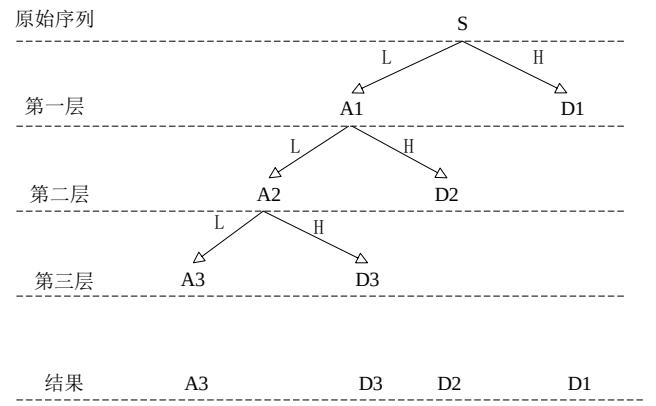


Fig. 5 Example diagram of wavelet decomposition

图5 小波分解示例图

其中, $A_i (i = 1, 2, 3)$ 代表每层中的低频分量, $D_i (i = 1, 2, 3)$ 代表每层中的高频分量, L 代表低通滤波器, H 代表高通滤波器。在每一层中,对母波进行不同尺度的缩放和平移,得到所使用的小波函数。通过将小波函数与所要分解的序列进行卷积操作,得到下一个水平的分量。具体公式如下:

$$r_{j,k} = \int_{-\infty}^{\infty} x(\tau) \Psi_{j,k}(\tau) d\tau \quad (5)$$

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \Psi\left(\frac{t - k \cdot 2^j}{2^j}\right) \quad (6)$$

其中, Ψ 表示母波, j, k 表示缩放和平移参数。

3 模型构建

相比单任务学习,多任务学习通过共享部分模型参数完成不同任务之间的信息迁移,更深层次地挖掘主任务统计特性变化与其他任务数据模式潜在的联系,提升了对复杂动态模式的捕捉能力和对变化趋势的适应能力,有助于提升主任务预测精度。由相关性分析结果可知,风速与风能发电功率之间存在强相关性,因此所构建的模型采用底层共享的模型框架。这种框架便于构建模型,并且适合处理相关性较强的任务,能有效降低过拟合的风险^[16]。模型

可分为参数共享层和任务指定层,其整体结构如图6所示。以下对模型的各个模块进行详细说明。

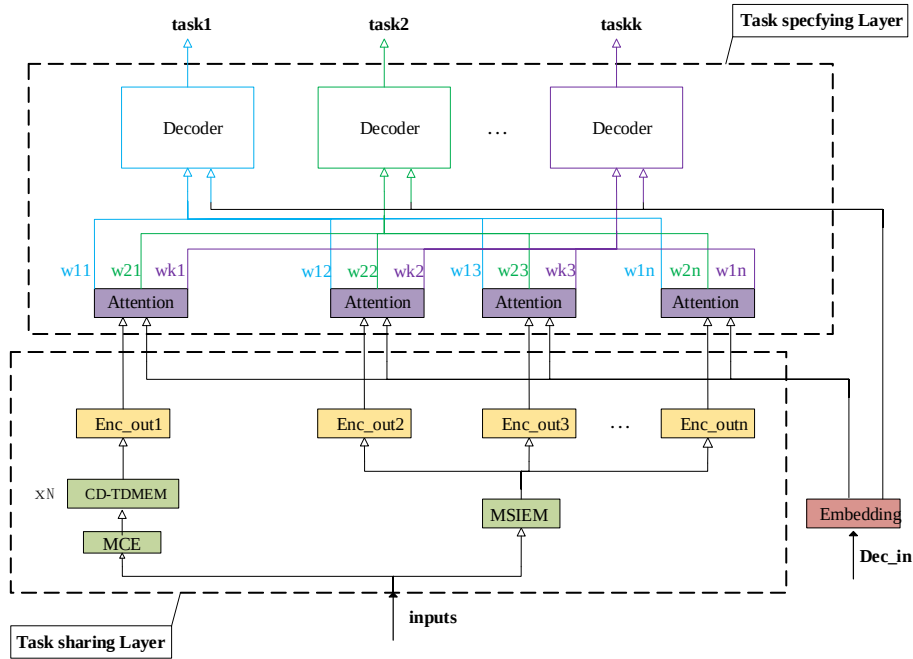


Fig. 6 The overall overview of the model
图6 模型整体概览图

3.1 参数共享层

3.1.1 多通道扩展

周期性分析结果表明,各特征呈现了复杂的周期性。为便于进行更细致的分析,进一步对特征进行了扩展,使每个特征都有单独的子模型,每个子模型分别学习序列中不同的特征信息。示例图如图7所示。针对每个特征使

用一维卷积,将原来的单通道序列转化为多个通道的表示。一维卷积实现了局部特征的自动提取,多通道的输出实现了表示扩展。其实现了捕获时间序列中不同模式(可能是短期趋势、周期性变化或特定的信号特征等)的目标,所形成的多通道输出可以更全面地展现时间序列的内在结构。

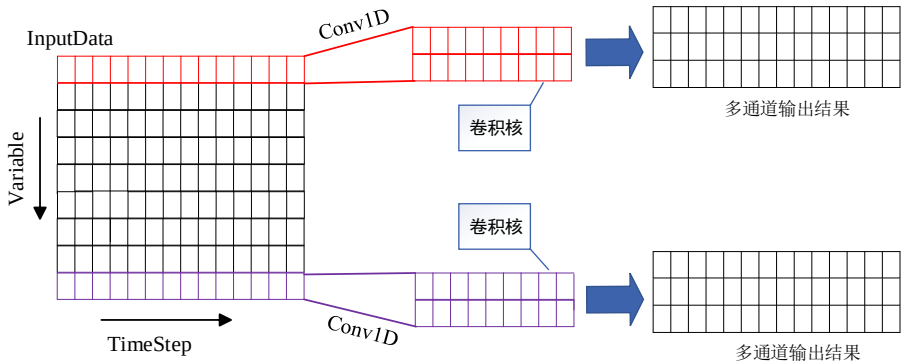


Fig. 7 Example plot of single-feature multichannel extension
图7 单特征多通道扩展示例图

3.1.2 跨维度—时间依赖混合编码模块

对于多变量时序预测,除时间上的依赖外,跨维度上的依赖也非常重要。对于一个特定的维度,来自其他维度的关联序列信息可能提升预测效果^[19]。文献[20]中使用数学表达式揭示了风速、风向、温度、气压、湿度等气象因素相互之间以及与风能发电功率之间复杂的物理关系。因此,除考虑在时间维度上的顺序依赖性外,也要考虑特征之间的交互。本文设计了一个跨维度—时间依赖混合

编码模块,包括时间依赖注意力(Temporal Dependency Attention, TDA)和跨维度依赖注意力(Cross-Dimension Dependency Attention, CDDA)两部分。其中,TDA用于捕捉混合特征在时间维度上的依赖关系,CDDA用来发现多变量之间混合的潜在关系。对于多变量时间序列而言,在单个时间点上,单个数值只能提供少量信息^[19]。如图8所示,使用多层的卷积操作,随着卷积层数的增加,上层的特征图中具有更大的感受野,通过逐层的特征提取,可以使

模型更好地挖掘潜藏在数据中复杂的特征模式。

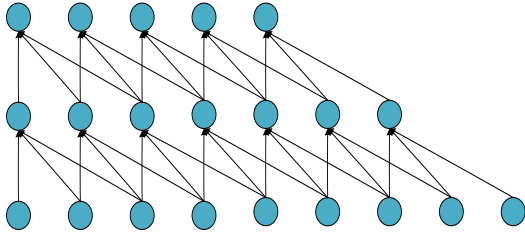


Fig. 8 Example plot of convolution

图8 卷积示例图

对于所提取的时间依赖和特征交互信息,借鉴文献[19]中的路由机制,分别使用一个可学习的矩阵作为路由对这两者的信息进行汇集,将所汇集的信息进行相加与融合即是所谓的跨维度—时间依赖的混合编码。其整体结构如图9所示。

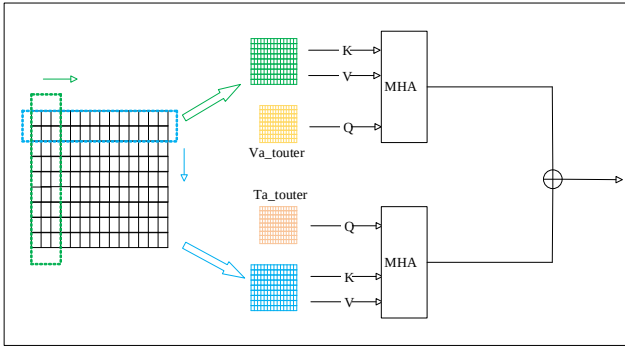


Fig. 9 Example diagram of cross dimension-temporal dependency mixing encoding module

图9 跨维度—时间依赖混合编码模块示例图

为避免网络深度增加导致的梯度消失或爆炸问题,提高模型的训练速度,在每层之间使用了残差连接和归一化操作。具体公式如下:

$$TDA = MHA(ConvlD(X), ConvlD(X), Ta_router) \quad (7)$$

$$CDDA = MHA(ConvlD(X.permute(0, 2, 1)), ConvlD(X.permute(0, 2, 1)), Va_router) \quad (8)$$

$$Module_out = LayerNorm(Add(X, TDA + CDDA)) \quad (9)$$

其中, X 表示模块的输入, $ConvlD$ 表示一维卷积, MHA 表示多头注意力机制, $permute$ 表示维度转换操作, Add 表示残差连接, $LayerNorm$ 表示层间归一化操作。

3.1.3 多尺度信息编码模块

时间序列既表现出粗粒度的时间模式,又表现出细粒度的时间模式,这两类信息通过动态交互的方式相互影响。粗粒度模式反映了塑造细粒度模式的宏观背景,而细粒度模式提供了细化和调整粗粒度模式的微观反馈^[22]。通过离散小波变换,各特征的时间序列被分解成多个层次不同频域的分量。其中,短期时间序列预测更依赖于高频模式,而长期预测更依赖于低频模式^[23],因此,在该模块中挖掘不同层次的分量对于所预测未来时间序列不同程度的影响。该模块的总体概览图如图10所示。

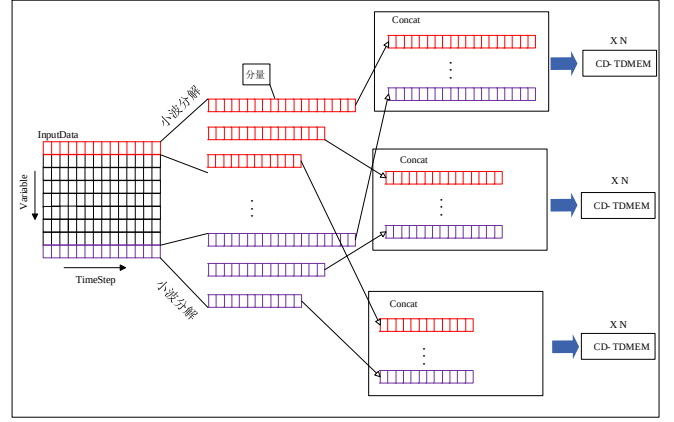


Fig. 10 Example diagram of the multiscale information encoding module

图10 多尺度信息编码模块示例图

3.2 任务指定层

在任务指定层中,将预测未来的风能发电功率设置为主任务,将预测未来的风速作为附属任务。对于每个子任务来说,需要预测多个未来时间步的时间序列。为避免累计误差,本文使用Transformer生成式预测模型普遍使用的一步解码操作实现未来时间序列的预测输出。具体公式如下:

$$X_{dec} = Concat(Y_{lable}, Z_H) \quad (10)$$

$$X_{dec_in} = Embedding(X_{dec}) \quad (11)$$

其中, $Concat$ 表示拼接操作, Y_{lable} 表示已知序列, $lable$ 表示已知序列的时间步长, Z_H 表示填零的序列, H 表示预测序列的时间步长, X_{dec_in} 表示解码器的输入, $Embedding$ 表示嵌入操作。

从参数共享层不仅得到了跨维度—时间依赖混合信息,而且得到了多尺度信息,多种信息的互补、融合才能充分展现时间序列的整体样貌^[24]。在任务指定层中,首先以解码器输入作为自注意力机制的查询矩阵,以共享层各编码器的输出作为自注意力机制的键矩阵和值矩阵,分别筛选了各信息对未来预测有重要影响的部分。之后通过一个动态加权机制对各信息进行融合,再使用多层解码器模块对融合的编码信息进行解码。

$$X_i = MHA(X_{dec_in}, enc_out, enc_out) \quad (12)$$

$$X_{fusion} = \sum w_i * X_i \quad (13)$$

$$X_{dec_out} = MHA(X_{dec_in}, X_{fusion}, X_{fusion}) \quad (14)$$

其中, enc_out 表示参数共享层中编码器的输出, X_{dec_in} 表示解码器的输入, w_i 表示可学习的权重参数。

最终通过一个全连接层(Fully Connected NetWork, FCN),经过一步解码,得到未来时间序列的预测结果。公式如下:

$$output = FCN(X_{dec_out}) \quad (15)$$

$output \in R^{(lable+H) \times M}$ 表示一步解码的输出结果,截取 $output[lable:lable+H]$ 部分数据即为最终的输出 y_h 。

4 实验与结果分析

4.1 数据集与数据预处理

实验数据来源于文献[25],由电力工程多相流动与传热上海市重点实验室所收集整理。数据记录间隔为 15 min,记录了从 2019.01.01–2020.12.31 两年间的 70 177 条数据,包括历史风能发电量、空气湿度、温度、气压、离地 10 m/30 m/50 m 高处以及涡轮机所在高度的风速和风向等特征信息。根据《风电功率预测系统功能规范 NB_T 31046–2022》等行业技术规范文件进行了数据填充、异常值订正、合规性检验等。由于各变量之间的量纲不完全相同,因此对各变量进行了标准化处理。具体公式如下:

$$X_{norm} = \frac{X - X_{mean}}{X_{std}} \quad (16)$$

其中, X_{norm} 表示归一化结果, X 表示进行归一化的变量, X_{mean} 表示变量的均值, X_{std} 表示变量的标准差。

4.2 实验环境

实验设备为 NVIDIA Tesla A100 40G, 操作系统为 Ubuntu, 实验代码基于 Pytorch 深度学习框架实现。输入为 96 个历史时间点的时间序列数据, 由于本文是针对短期风能发电进行预测, 每个子任务的输出序列长度为未来 24 个时间点的预测数据。

4.3 评价指标

为验证模型精度, 本文使用均方误差 (Mean Square Er-

ror, MSE) 以及平均绝对误差 (Mean Absolute Error, MAE) 作为评价指标。

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2 \quad (17)$$

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

其中, x_i 表示真实值, $\hat{x}_i$ 表示预测值, N 表示测试集样本个数。

4.4 对比实验

4.4.1 单任务学习模型对比实验

将本文提出的多任务学习模型 (以 proposed_mtl 表示) 和多个当前在时序预测领域较为先进的单任务模型以及采用本文模型进行单任务学习 (以 proposed_stl 表示) 的预测精度进行实验对比。结果见表 2, 其中加粗字体表示最优的实验结果。图 11 展示了随机截取的 5 个连续 24 个时间点的预测结果。

Table 2 Comparative experimental results of single-task learning models

表 2 单任务学习模型对比实验结果

Model	MSE	MAE
Transformer ^[26]	0.168 1	0.260 3
Informer ^[27]	0.478 6	0.522 3
PatchTST ^[28]	0.586 8	0.527 7
iTransformer ^[29]	0.148 0	0.236 3
Non-stationary Transformer ^[30]	0.155 8	0.252 1
Proposed_stl	0.144 6	0.241 2
Proposed_mtl	0.141 8	0.224 3

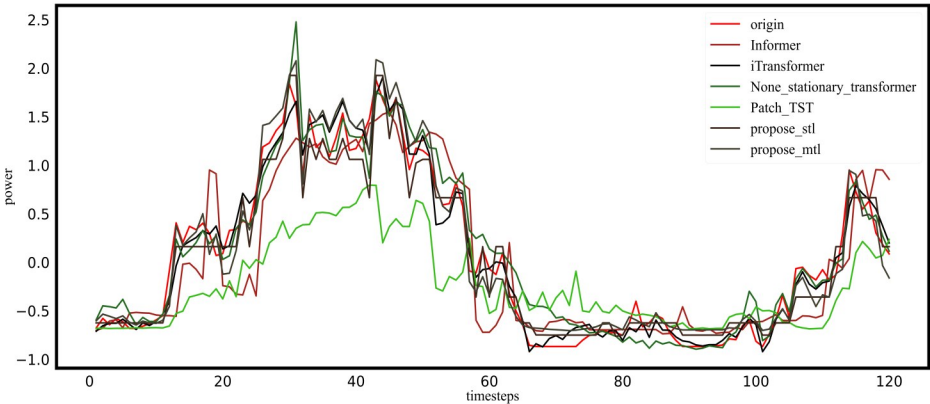


Fig. 11 Comparison of prediction performance of single-task learning models

图 11 单任务学习模型预测效果比较

由以上实验结果可知, 针对风能发电功率预测任务, 使用本文模型进行传统的单任务学习, 相比 Transformer、Informer、PatchTST、iTransformer、Non-stationary Transformer 模型, 本文模型 MSE 分别降低了 13.97%、69.78%、75.36%、2.30%、7.18%, MAE 分别降低了 7.33%、53.82%、54.29%、-2.07%、4.32%。本文模型进行多任务学习, 相比其他模型, MSE 分别降低了 15.64%、70.37%、75.83%、4.19%、8.98%、1.94%, MAE 分别降低了 13.83%、57.05%、

58.04%、5.08%、11.02%、7.00%。相比单任务学习, 多任务学习有效提升了预测精度。一个可能的解释是, 多任务学习作为一种利用相关任务包含领域特定信息的归纳迁移方法, 对于附属任务的学习, 经过反向传播, 将附属任务所挖掘的特征在参数共享层进行表示, 同时这些特征又可以被主任务所使用。另外, 多任务学习不仅挖掘有助于一个任务的特征, 而且能够挖掘在单任务学习中得不到体现的有助于多个任务的特征。多任务学习的这种机制有效强

化了风能发电功率预测任务的学习效果。

4.4.2 多任务学习模型对比实验

将本文模型与其他相关的多任务模型进行实验对比,实验结果见表3,其中加粗字体表示最优的实验结果。图12展示了随机截取的5个连续24个时间点的预测结果。

由以上实验结果可知,针对风能发电功率预测任务,使用本文提出的多任务学习模型,相比 MTL-LSTM、MTL-LSTM、WaveNet-MmoE 模型, MSE 分别降低了 8.27%、

3.34%、9.97%, MAE 分别降低了 10.74%、7.96%、16.43%。

Table 3 Comparative experimental results of multi-task learning models

表3 多任务学习模型对比实验结果		
Model	MSE	MAE
MTL-LSTM ^[16]	0.154 6	0.251 3
D2SA-AR ^[17]	0.146 7	0.243 7
WaveNet - MmoE ^[18]	0.157 5	0.268 4
proposed_mtl	0.141 8	0.224 3

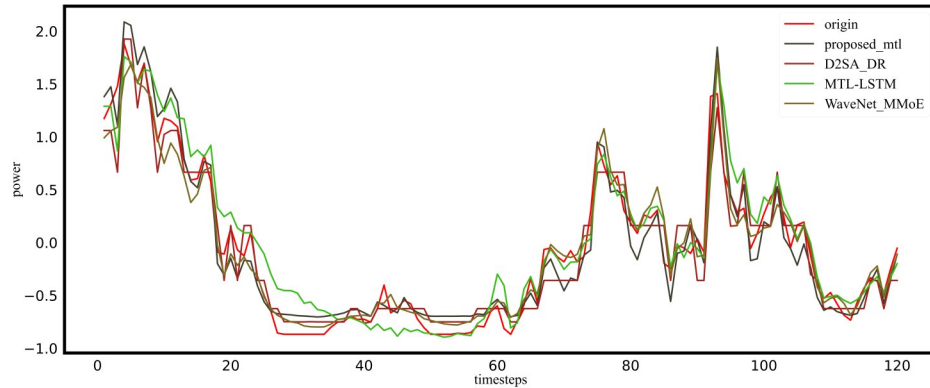


Fig. 12 Comparison of prediction performance of multi-task learning models

图12 多任务学习模型预测效果比较

4.5 消融实验

为验证模型中各个模块的有效性,本文分别针对4个模型的消融变体进行了一系列消融实验,包括:①Model-V1:去除模型的多尺度模块;②Model-V2:去除模型的跨维度—时间依赖混合编码模块和特征的多通道扩展模块;③Model-V3:去除模型中特征的多通道扩展模块;④Model-V4:去除跨维度—时间依赖混合编码模块,使用Transformer中的原始编码器模块。

实验结果如表4所示。由消融实验结果可知,对应于每个消融实验, MSE 分别提升了 4.93%、44.99%、2.53%、3.52%, MAE 分别提升了 12.92%、49.88%、3.07%、7.71%。其中,去除模型的跨维度—时间依赖混合模块和特征的多通道扩展模块导致预测精度下降最多,仅移除特征多通道扩展模块导致的模型预测精度下降最少。从整体来看,去除任何子模块都会导致模型性能降低,说明每一个模块的存在都是有意义的。移除多尺度信息模块相比移除跨维度—时间依赖混合编码模块和特征的多通道扩展模块对于模型性能的影响要小得多,说明跨维度和时间依赖混合信息相比多尺度信息对于预测的准确性更为重要。

Table 4 Results of ablation experiments

表4 消融实验结果

Model	MSE	MAE
Model-V1	0.148 8	0.253 3
Model-V2	0.205 6	0.336 2
Model-V3	0.145 4	0.231 2
Model-V4	0.142 6	0.228 6

5 结语

为了提升短期风能发电功率预测精度,根据对各特征的周期性分析,发现离地 10 m、30 m、50 m 以及涡轮高度处的风速和风能发电功率未呈现明显的周期性,而温度、气压、湿度等特征呈现了较为明显的周期性。受相关文献启发,针对每一个特征都进行了特征扩展,并且提出跨维度—时间依赖的混合交叉注意力模块,对时间上的依赖和跨维度的依赖两方面信息进行融合。另外,利用小波变换将特征分解为不同频域的多个分量。鉴于不同粒度的信息对于预测的不同影响,分别针对不同粒度的信息进行编码。各特征之间的相关性分析结果表明,10 m、30 m、50 m、涡轮机高度处风速以及风能发电功率都呈现了较强的相关性;多任务学习通过窃听、表征偏好等多种机制,实现了不同任务之间的信息共享,有助于挖掘各子任务之间的潜在数据模式。因此,本文以风能发电功率预测作为主任务,以风速预测作为附属任务,提出一个硬参数共享的多任务学习模型。实验结果表明,在风电功率预测任务上,本文模型相比其他单任务模型性能有所提升,但是利用多任务学习机制构建模型的预测性能得到了进一步提升,表明多任务学习在风能发电功率预测研究领域具有广阔前景。由于气象时间序列和风能发电功率序列为非平稳性的时间序列,会发生分布偏移,因此其统计量会随着时间窗口的滑动而发生变化。目前已有相关工作对该领域进行了探讨,提升模型对于分布偏移的建模能力也值得作进

一步研究。

参考文献:

- [1] HAN Z F, JING Q M, ZHANG Y K, et al. Review of wind power forecasting methods and new trends [J]. Power System Protection and Control, 2019, 47(24): 178-187.
韩自奋, 景乾明, 张彦凯, 等. 风电预测方法与新趋势综述[J]. 电力系统保护与控制, 2019, 47(24): 178-187.
- [2] DAI Q B, HUANG N T. Review of research on error correction of short-term wind power prediction [J]. Journal of Northeast Electric Power University, 2023, 43(2): 1-7.
戴千斌, 黄南天. 短期风电功率预测误差修正研究综述[J]. 东北电力大学学报, 2023, 43(2): 1-7.
- [3] MA Z J, MEI G. A hybrid attention-based deep learning approach for wind power prediction [J]. Applied Energy, 2022, 323: 119608.
- [4] COX D R. Statistical analysis of time series: some recent developments [J]. Scandinavian Journal of Statistics, 1981, 8(2): 93-115.
- [5] GOMES P T, CASTRO R. Wind speed and wind power forecasting using statistical models: autoregressive moving average (ARMA) and artificial neural networks (ANN) [J]. International Journal of Sustainable Energy Development, 2012(2): 41-50.
- [6] RUDER S. An overview of multi-task learning in deep neural networks [DB/OL]. <http://arxiv.org/abs/1706.05098v1>.
- [7] FENG X, ZHAO Y, ZONG W, et al. Adaptive multi-task learning for speech to text translation [EB/OL]. <https://asmp-eurasipjournals.springeropen.com/articles/10.1186/s13636-024-00359-1>.
- [8] WANG C, SHEN X Z. Feature-enhanced multi-task learning for speech emotion recognition using decision trees and LSTM [J]. Electronics, 2024, 13(14): 2689.
- [9] WU Q H, LI W T. Multi-focus image fusion algorithm based on multi-task learning and PS-ViT [J]. IEICE Transactions on Information and Systems, 2024, 107: 1422-1432.
- [10] ROHANI M, FARSI H, MOHAMADZADEH S. Facial feature recognition with multi-task learning and attention-based enhancements [J]. Iranica Journal of Energy and Environment, 2025, 16(1): 136-144.
- [11] WALHA R, DRIRA F, BEDHIEF R. A multi-task learning framework for image restoration using a novel generative adversarial network [C]//International Conference on Agents and Artificial Intelligence, 2024: 928-935.
- [12] JIANG Z H. Improving multi-task learning in autonomous driving perception with dynamic loss weights and individual encoders [J]. KTH Royal Institute of Technology, 2024, 105: 39.
- [13] MIAO X X, YU W W. ARGENT: multi-task learning model for predicting autism-related genes and drug targets using heterogeneous graph convolutional network [J]. Future Generation Computer Systems, 2024, 160: 942-950.
- [14] ISIAM M S, ADNAN T, FREYBERG J, et al. Accessible, at-home detection of parkinson's disease via multi-task video analysis [DB/OL]. <https://arxiv.org/abs/2406.14856>.
- [15] CARUANA R. Multitask learning [J]. Machine Learning, 1997, 28: 41-75.
- [16] WEI J Q, WU X J, YANG T M, et al. Ultra-short-term forecasting of wind power based on multi-task learning and LSTM [J]. International Journal of Electrical Power & Energy Systems, 2023, 149: 109073.
- [17] TANG Y G, YANG K, ZHANG S J, et al. Wind power forecasting: a hybrid forecasting model and multi-task learning-based framework [J]. Energy, 2023, 278: 127864.
- [18] WANG H, PENG C, LIAO B L, et al. Wind power forecasting based on WaveNet and multitask learning [J]. Sustainability, 2023, 15(14): 10816.
- [19] ZHANG Y H, YAN J C. Crossformer: transformer utilizing cross-dimension dependency for multivariate time series forecasting [C]//International Conference on Learning Representations, 2023: 1-21.
- [20] WANG Q, LAI K K, NIU D X, et al. A multivariate wind power forecasting model based on LS-SVM [C]//2012 Fifth International Joint Conference on Computational Sciences and Optimization, 2012: 127-131.
- [21] GUI H Y, LI G J, TANG X H, et al. CATodyNet: cross-attention temporal dynamic graph neural network for multivariate time series classification [J]. Knowledge-based Systems, 2024, 300: 112210.
- [22] HU Y F, LIU P Y, ZHU P, et al. Adaptive multi-scale decomposition framework for time series forecasting [DB/OL]. <https://arxiv.org/abs/2406.03751>.
- [23] GUO Q, FANG L X, WANG R, et al. Multivariate time series forecasting using multiscale recurrent networks with scale attention and cross-scale guidance [J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 36(1): 540-554.
- [24] CHEN Z K, E J Z, ZHANG X, et al. Multi-task time series forecasting with shared attention [C]//2020 International Conference on Data Mining Workshops (ICDMW), 2020: 917-925.
- [25] CHEN Y B, XU J J. Solar and wind power data from the Chinese state grid renewable energy generation forecasting competition [J]. Scientific Data, 2022, 9(1): 577.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [DB/OL]. <https://arxiv.org/abs/1706.03762>.
- [27] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 11106-11115.
- [28] NIE Y Q, NGUYEN N H, SINTHONG P, et al. A time series is worth 64 words: long-term forecasting with transformers [DB/OL]. <https://arxiv.org/abs/2211.14730>.
- [29] LIU Y, HU T G, ZHANG H R, et al. Ittransformer: inverted transformers are effective for time series forecasting [DB/OL]. <https://arxiv.org/abs/2310.06625>.
- [30] LIU Y, WU H X, WANG J M, et al. Non-stationary transformers: exploring the stationarity in time series forecasting [J]. Advances in Neural Information Processing Systems, 2022, 35: 9881-9893.

(责任编辑:黄健)