

基于LSTM-GBRT模型的省级区域碳排放预测方法

林成¹, 吕俊兵², 张裕³, 张忠²

(1. 贵州电网有限责任公司, 贵阳 550002; 2. 大连理工大学电气工程学院, 辽宁 大连 116024; 3. 贵州电网有限责任公司电网规划研究中心, 贵阳 550002)

摘要: 针对省级区域碳排放预测由于历史数据不足导致预测精度较低问题, 提出了一种基于电力数据的省级区域碳排放预测方法。首先, 分析了省级区域电力负荷数据与碳排放总量之间的相关性; 然后, 构建长短期记忆-梯度提升回归树(long short-term memory-gradient boosting regression tree, LSTM-GBRT)混合模型, 并基于电-碳相关性完成省级区域碳排放预测; 为进一步提升预测精度, 采用K-means聚类算法对各省省级区域能源消费结构进行聚类分析, 挖掘碳排放特征相似省份, 用于完善数据集与特征集; 最后, 基于我国各省级区域的电力和碳排放历史数据对某省碳排放进行预测, 结果表明, 该模型能够有效地预测碳排放趋势, 平均绝对百分误差、平均绝对误差和均方根误差指标均优于LSTM、卷积神经网络(convolutional neural network, CNN)和门控循环单元(gated recurrent unit, GRU)等模型, 展示出更好的泛化能力与鲁棒性。

关键词: 省级区域碳排放; 电-碳相关性; 碳排放预测; LSTM-GBRT混合模型

Provincial-Level Area Carbon Emission Forecasting Method Based on LSTM-GBRT Model

LIN Cheng¹, LÜ Junbing², ZHANG Yu³, ZHANG Zhong²

(1. Guizhou Power Grid Co., Ltd., Guiyang 550002, China; 2. School of Electrical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China; 3. Power Grid Planning Research Center, Guizhou Power Grid Co., Ltd., Guiyang 550002, China)

Abstract: To address the issue of low prediction accuracy in provincial-level area carbon emission forecasting caused by insufficient historical data, a novel prediction method based on electricity data is proposed. Firstly, the correlation between provincial-level area power load data and total carbon emissions is analyzed. Subsequently, a hybrid long short-term memory-gradient boosting regression tree (LSTM-GBRT) model is constructed to perform carbon emission forecasting, based on the electricity-carbon correlation. To further enhance prediction accuracy, a K-means clustering algorithm is employed to classify provincial-level area energy consumption structures, enabling the identification of provinces with similar carbon emission characteristics. This clustering aids in enriching both the dataset and feature set. Finally, based on historical electricity and carbon emission data across multiple provincial-level area in China, the model is applied to forecast emissions in a target province. The results demonstrate that the proposed model can effectively capture the carbon emission trend. It consistently outperforms benchmark models-including long short-term memory (LSTM), convolutional neural network (CNN), and gated recurrent unit (GRU)-in terms of mean absolute percentage error, mean absolute error, and root mean square error. These findings highlight the model's superior generalization performance and robustness across varying data scenarios.

Key words: provincial-level area carbon emission; electric-carbon correlation; carbon emission forecasting; LSTM-GBRT hybrid architecture

基金项目: 智能电网国家科技重大专项(2024ZD0800600)。

Foundation item: Supported by the Smart Grid National Science and Technology Major Project(2024ZD0800600)。

0 引言

随着人类进入工业社会,二氧化碳等温室气体排放量不断增加,造成全球气温升高,引起极端气候等一系列后果^[1]。为应对全球气候和能源问题,国际社会已达成减少碳排放、实现碳达峰、碳中和的共识。我国作为主要的碳排放国家之一在节能减排中发挥着巨大的作用,在第七十五届联合国大会上我国表明二氧化碳排放量力争在2030年前达峰,并且在2060年实现碳中和^[2]。在该背景下,准确预测省级区域及行业的碳排放趋势成为支撑政策制定的重要基础。

如何科学预测省级区域及行业的碳排放趋势已成为当前的一个研究热点。目前,国内外学者提出诸多预测方法,主要包括:传统统计模型、神经网络模型和混合模型等。传统统计模型主要包括差分自回归移动平均模型(autoregressive integrated moving average model, ARIMA)^[3-5]和基于人口、富裕度与技术的随机回归影响模型(stochastic impacts by regression on population, affluence, and technology, STIRPAT)^[6-8]等。文献[9]基于STIRPAT扩展模型和岭回归分析预测了不同情境下的碳达峰时间和峰值。文献[10]基于STIRPAT模型和空间杜宾模型预测了电力行业碳排放趋势。

随着机器学习和深度学习技术的发展,神经网络模型在省级区域及行业的碳排放预测中得到广泛应用。与传统模型相比,神经网络模型能更好地捕捉数据中的复杂关系。神经网络模型主要包括长短期记忆模型(long short-term memory, LSTM)^[11-13]和反向传播神经网络(backpropagation neural network, BP)模型^[14-16]等。文献[17]基于改进的BP神经网络模型预测了河北省2021—2030年碳排放趋势。此外,为进一步提升模型性能,研究者提出了多种混合模型,其中最常见的是ARIMA-LSTM混合模型^[18-20]。文献[21]构建了基于ARIMA和LSTM加权组合的二氧化碳排放预测模型,并采用随机森林特征筛选方法来分析影响因素。文献[22]提出了用U-LSTM-LASSO混合模型来预测我国二氧化碳排放。然而,这些模型在捕捉时间变化规律和解释外部因素影响方面仍存在局限性。因此,需要一种既能分析碳排放变化趋势又能解释影响因素的综合预测方法。

当前,省级区域碳排放预测还面临历史数据不足、质量低等挑战。相比之下,电力数据具备时序连续性强、覆盖范围广等优势,部分学者开始研究电-碳之间的关系^[23-25]。文献[26]提出了一种基于电力数据的岭回归赋权多元线性回归模型,实现了省级区域及行业碳排放测算。文献[27]构建了电碳计算模型,并结合电力大数据与机器学习算法对碳排放趋势进行预测。上述研究虽在碳排放预测领域取得了显著效果,但仍存在一定局限性,表现为在对省级区域间碳排放数据的空间关联性及能源消费结构相似性对预测趋势的影响缺乏系统考量,导致预测模型的时空特征挖掘能力受限。

尽管碳排放与电力负荷、新能源发电、电价等电力相关数据均为时间序列预测,但在预测原理上存在本质差异。电力相关数据主要受天气和市场等周期性因素驱动,具有明显的短期时序周期性,且数据齐全,一般利用周期性特征开展电力相关数据预测;而碳排放则更多受到产业结构、能耗结构等长期因素的影响,变化相对缓慢,年度的碳排放指标不具有周期性,且能耗数据获取难度大,为此本研究基于电-碳相关性特征开展碳排放预测。由于预测原理与数据完整程度不同,现有负荷、新能源、电价等预测方法难以直接迁移应用于碳排放预测。碳排放预测更适合采用相关性预测的方法,即在电-碳相关性的基础上构建符合碳排放指标特征的专门模型。

为此,本文基于电力负荷数据与省级区域碳排放之间的相关性构建了融合LSTM-GBRT模型与聚类数据增强框架的预测体系。首先采用LSTM捕捉碳排放时序数据中的非线性特征和长期动态演变规律,充分挖掘电力负荷数据与碳排放之间的关系,通过其独特的门控机制建立动态记忆单元;同时引入GBRT构建混合模型,通过残差逼近策略和基学习器加权融合,有效缓解LSTM对数据缺失和异常值的敏感性,提升低质量数据条件下的预测稳定性。进一步基于不同省级区域间碳排放关联特征提出聚类分析的数据增强方法,该方法通过引入跨省级区域同源数据扩展训练样本空间,在保持数据相似性的基础上增强模型泛化能力。该融合框架通过算法协同优化与数据空间拓展实现了预测精度与鲁棒性的双重提升。

1 省级区域碳排放核算模型

1.1 省级区域碳排放核算标准

为分析电-碳历史数据相关性,首先开展省级区域碳排放的核算工作。根据国际标准化组织(ISO)颁布的温室气体核算标准(ISO14064)^[28],碳排放核算体系将温室气体排放源分为3个层级:Ⅰ类直接碳排放:本地区内因燃料燃烧和工业活动所产生的排放,主要包括燃烧碳排放和过程碳排放,燃烧碳排放来源于煤炭、天然气、石油等化石燃料的燃烧,过程碳排放指工业生产过程中非燃烧反应产生的温室气体,如钢铁冶炼等;Ⅱ类间接碳排放:由外购能源(电力/热力)消耗引发的排放,实际产生于能源生产端;Ⅲ类价值链排放:全价值链间接排放,涵盖供应链上下游(如原料运输、产品使用及废弃物处理)等关联环节。

当前,我国多尺度碳排放核算体系主要涉及Ⅰ类直接碳排放与Ⅱ类间接碳排放。省级区域碳排放核算范围:本省级区域内一次能源消耗与生产过程产生的直接碳排放以及省级区域间电力、热力交易对应的Ⅱ类间接碳排放。具体核算范围如图1所示。

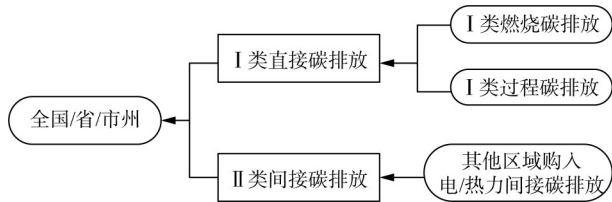


图1 省级区域碳排放核算范围示意图

Fig. 1 Diagrammatic representation of carbon emission accounting boundaries at the provincial administrative region level

基于相关能耗计算获得碳排放量是目前开展省级区域碳排放核算相对成熟的方法。基于IPCC国家温室气体清单指南^[29],碳计量方式可以分为以下两种:1)排放因子法:通过能源消费量(如化石燃料)与排放因子的耦合计算碳排放量;2)碳质量平衡法:基于物质守恒原理,通过系统输入碳总量(如燃料碳含量)与固定碳(如灰渣)及非二氧化碳的碳输出量的差值核算碳排放。相对而言,排放因子法凭借其较强的可操作性和较低的数据要求在省级区域碳排放核算和预测研究中得到了广泛应用,因

此,本文选用排放因子法进行碳排放核算。

1.2 碳排放因子及省级区域排放核算

碳排放因子是指单位活动量(如能源消耗、产品生产等)所产生的二氧化碳排放量。根据排放源的不同将碳排放因子分为直接碳排放因子和间接碳排放因子。直接碳排放因子又分为化石燃料直接燃烧碳排放因子和水泥等产品的生产过程排放因子两类。间接碳排放因子包括电力和热力碳排放因子^[30-31]。

化石燃料碳排放因子用于表征单位化石燃料消费量对应的碳排放,其计算公式如式(1)所示。

$$F_{Ci} = A_i \cdot V_i \cdot O_i \cdot \frac{M_{CO_2}}{M_C} \cdot N \quad (1)$$

式中: F_{Ci} 为第 i 类化石能源的碳排放因子,固体燃料单位为 kgCO_2/kg , 气体燃料单位为 kgCO_2/m^3 ; A_i 为第 i 类化石能源的平均低位发热量,固体燃料单位为 $\text{PJ}/10^4\text{t}$, 气体燃料单位为 $\text{PJ}/10^8\text{m}^3$; V_i 为第 i 类化石能源的单位热值含碳量, tC/TJ ; O_i 为第 i 类化石能源的碳氧化率; M_{CO_2} 为二氧化碳相对分子质量,取 44; M_C 为碳的相对原子质量,取 12; N 为单位换算乘子,当化石能源平均低位发热量单位为 $\text{PJ}/10^4\text{t}$ 时,取 0.1,当化石能源平均低位发热量单位为 $\text{PJ}/10^8\text{m}^3$ 时,取 0.01。

省级区域电力碳排放因子的计算公式如式(2)所示。

$$F_E = \frac{\sum_{i=1}^n C_{Bi}^G F_{Ci}}{E_G} = \frac{\sum_{i=1}^n C_{Bi}^G F_{Ci}}{E_{Gr} + E_{Gh}} \quad (2)$$

式中: F_E 为省级区域电力碳排放因子, kgCO_2/kWh ; C_{Bi}^G 为省级区域火力发电所用第 i 种化石能源的燃烧量,固体燃料单位为 kg , 气体燃料单位为 m^3 ; n 为火电发电燃料种类数量; E_G 为省级区域总发电量, kWh ; E_{Gr} 为省级区域一次能源发电量, kWh ; E_{Gh} 为省级区域火力发电量, kWh 。

省级区域热力碳排放因子的计算公式如式(3)所示。

$$F_H = \frac{\sum_{i=1}^n C_{Bi}^H F_{Ci}}{E_H} \quad (3)$$

式中: F_H 为省级区域热力碳排放因子, kgCO_2/MJ ; C_{Bi}^H 为省级区域制热所用第 i 种化石能源的燃烧量,固体燃料单位为 kg , 气体燃料单位为 m^3 ; E_H 为省级区域总制热量, MJ 。

省级区域碳排放量计算包括Ⅰ类直接碳排放与区间电热交换对应的Ⅱ类碳排放,计算公式如式(4)所示。

$$C = C_I + C_{II} \quad (4)$$

式中: C 为省级区域碳排放量; C_I 为Ⅰ类直接碳排放量,主要有17类化石能源燃烧产生的碳排放量与水泥过程碳排放量, kgCO_2 ; C_{II} 为Ⅱ类间接碳排放,主要来源于从其他地区购入电力和热力产生的碳排放量, kgCO_2 。

其中,Ⅰ类直接碳排放 C_I 的核算公式如式(4)一(5)所示。

$$C_I = \sum_{i=1}^m C_{Bi} F_{Ci} + C_P F_P \quad (5)$$

$$C_{Bi} = C_{Fi} + C_{Ei} + C_{Hi} - C_{NEi} \quad (6)$$

式中: m 为省级区域化石燃料消费种类数量; C_{Bi} 为省级区域第 i 类化石能源燃烧量,固体燃料单位为 kg ,气体燃料单位为 m^3 ; C_{Fi} 、 C_{Ei} 和 C_{Hi} 分别为第 i 类化石能源终端消费量、发电消费量和制热消费量; C_{NEi} 为第 i 类化石能源非燃烧用途消费量,固体燃料单位为 kg ,气体燃料单位为 m^3 ; C_P 为省级区域水泥产量, kg 。

Ⅱ类间接排放 C_{II} 的核算公式如式(7)所示。

$$C_{II} = C_{EA} F_E + C_{HA} F_H \quad (7)$$

式中: C_{EA} 为省级区域外净调入电力, kWh ; C_{HA} 为省级区域外净调入热力, MJ 。

2 电-碳相关性分析

电碳相关性特征可为省级区域碳排放核算与预测提供重要支撑。然而,在对电力负荷数据与碳排放的历史数据进行回归分析时,发现二者并不呈严格线性关系。随着电力负荷数据的增加,碳排放虽总体上升,但增长速率逐渐减缓,表现出显著的非线性特征。为更准确刻画这一关系,提出了一种幂函数形式的相关性分析方法,通过构造幂函数模型获取电-碳关联性更强的特征参数,以更精准地刻画省级区域电-碳相关性。该方法不仅在拟合精度上优于线性模型,也更契合当前我国电力结构逐步清洁化、单位用电碳强度下降的现实背景。幂函数模型如式(8)所示。

$$C = aE^\beta + b \quad (8)$$

式中: E 为省级区域电力负荷数据; β 为模型幂指

数; a 和 b 分别为指数项和常数项系数,两个系数通过回归分析得到。

描述相关性的指标有皮尔森相关系数、斯皮尔曼相关系数、灰色关联度、熵相关系数、Granger因果检验等。由于皮尔森相关系数能够有效衡量两个连续变量之间的线性相关性,且具有计算高效、结果直观等优点,本文采用皮尔森相关系数分析电力负荷数据与省级区域碳排放数据之间的相关性。皮尔森相关系数的数值在-1至1之间,其计算公式式(9)所示。

$$\theta = \frac{\sum_{k=1}^K (C_k - \bar{C}_k)(E_k^\beta - \bar{E}_k^\beta)}{\sqrt{\sum_{k=1}^K (C_k - \bar{C}_k)^2} \sqrt{\sum_{k=1}^K (E_k^\beta - \bar{E}_k^\beta)^2}} \quad (9)$$

式中: θ 为皮尔森相关性; k 为省级区域电力负荷数据和碳排放总量的样本序号, $k = 1, 2, \dots, K$; K 为年份样本总数量; C_k 为第 k 年省级区域碳排放量; $\bar{C}_k$ 为年碳排放平均量; E_k^β 为第 k 年省级区域电力负荷数据的 β 次方, $\beta = 0.2, 0.4, \dots, 2$; $\bar{E}_k^\beta$ 为电力负荷数据 β 次方的平均值, $\beta = 0.2, 0.4, \dots, 2$ 。

皮尔森相关系数数值可以分为: $0 \leq |\theta| < 0.2$ 表示极弱相关或者无相关, $0.2 \leq |\theta| < 0.4$ 表示弱相关, $0.4 \leq |\theta| < 0.6$ 表示中等强度相关, $0.6 \leq |\theta| < 0.8$ 表示强相关, $0.8 \leq |\theta| \leq 1$ 表示极强相关。

3 省级区域LSTM-GBRT碳排放预测模型

3.1 省级区域间时空相关特征构建

在进行碳预测时引入时序数据和电力消费数据对于揭示能源碳排放的动态变化规律及其时间依赖性具有重要作用。从时间维度来看,碳排放量呈现周期性波动,其受社会经济活动与政策干预的影响,通过融入时间特征可以有效刻画此类周期性规律,反映其演变趋势,提升预测精度。电力负荷数据作为能源需求的直接表征与碳排放主要来源的电力生产活动具有强相关性,其动态变化能够反映能源消费模式的变化。

利用单一省级区域数据集进行神经网络模型训练易导致过拟合现象,并影响模型泛化能力。原因是各省级区域能源消费结构存在明显差异性。为了弥补这一不足,对不同省级区域能源消费结构进行分类。不同省级区域可能处于电气化与脱碳过程的不同阶段,其电力负荷数据与碳排放的内在关系存

在一定共性和差异性特征。因此,通过整合多省级区域数据,模型能够学习具有迁移性的特征表示,增强跨场景预测能力。

为优化数据利用效率,提出了基于相似度的省级区域聚类方法,采用 K-means 聚类算法对各省级区域主要能源消费品类的占比数据进行分类,将能源消费结构相似的省级区域聚为一类,从而挖掘省级区域能源消费模型的潜在特征。通过聚类为各省份增加类型标签,使预测模型能够识别和区分各省不同能源结构下的电-碳关系特征。训练过程中,与目标省份同属一类省份的数据的借鉴程度更高,即对模型参数的影响程度更大;反之,对于与目标省份不属于一类省份的数据响应程度将相对较小。由此,模型能够有侧重地借鉴各省份数据集,减轻训练负担,提升预测精度和泛化能力。

3.2 基于LSTM-GBRT的碳排放模型

LSTM-GBRT 模型由 LSTM 特征处理模型与 GBRT 预测模型组成,融合了 LSTM 与 GBRT 的优势,属于机器学习领域中深度神经网络与集成学习方法的交叉创新。模型原理如图 2 所示,样本嵌入到 LSTM 特征处理模型从中提取时序相关性和特征内部相关性信息,并从最后一层 LSTM 中生成 GBRT 模型的嵌入特征。GBRT 模型结合初步嵌入特征进行预测,不断在前树的预测误差上构造新树,随着树的数量不断上升,模型误差逐渐下降。

LSTM-GBRT 模型通过多层次特征抽象与误差

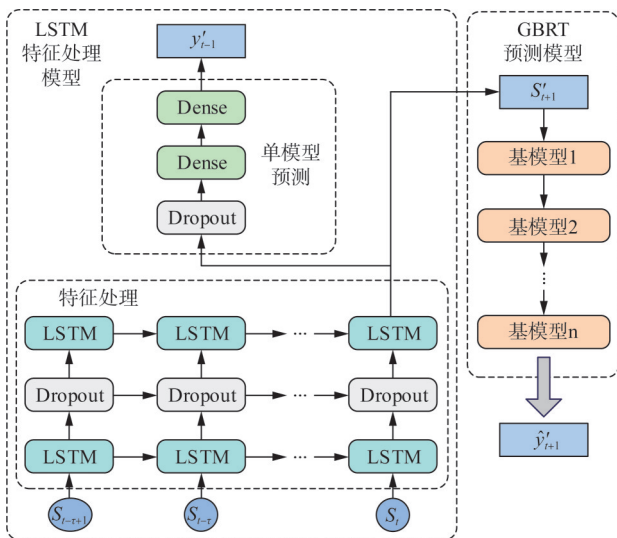


图2 LSTM-GBRT模型预测原理图

Fig. 2 Prediction principle diagram of LSTM-GBRT model

迭代优化机制能够有效捕捉电力负荷数据与碳排放之间的复杂非线性关联关系。具体而言, LSTM 模块擅长处理时间序列数据中的长期依赖特征,而 GBRT 模块则通过树结构集成策略强化对局部非线性模式的拟合能力。这种层级化特征学习框架不仅能够刻画电-碳系统的动态演变规律,还可通过多尺度特征融合提升模型对复杂能源结构的适应性。LSTM-GBRT 模型的预测流程如图 3 所示。该方法考虑到省级碳排放数据存在滞后性和数据不足等现实问题,选取连续性强、时效性高的电力负荷数据开展电-碳相关性分析,再基于电-碳相关性进行碳排放预测同时引入省级数据聚类来提高预测的准确性与稳定性。整体流程在当前碳排放历史数据不充足的条件下具有可操作性与预测准确性,是实现省级碳排放趋势预测的较优选择。

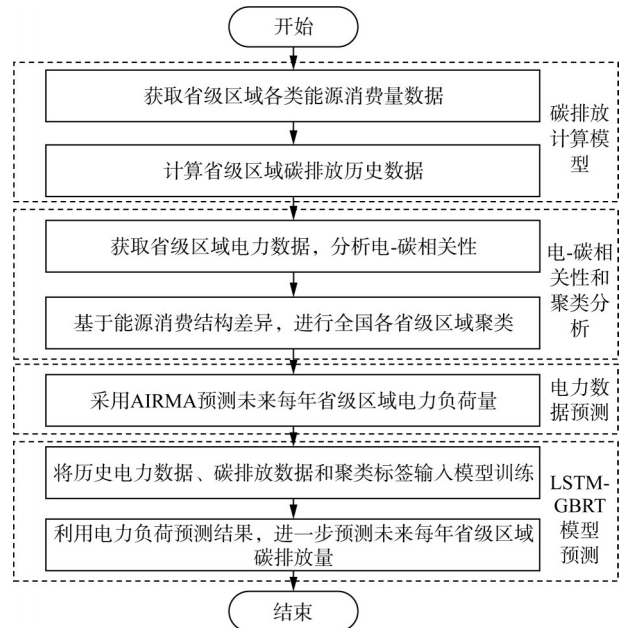


图3 基于LSTM-GBRT模型的省级区域碳排放预测流程图

Fig. 3 Prediction process diagram of provincial-level area carbon emissions based on the LSTM-GBRT model

采用 LSTM-GBRT 模型进行碳排放预测时,需先对电力负荷数据进行预测,作为省级区域碳排放预测的输入数据,从而完成省级区域碳排放量的预测。为此,采用 ARIMA 模型对电力负荷数据进行时间序列建模与预测,以捕捉其变化趋势和周期性特征。

ARIMA 模型的基本原理可以用式(10)表示。

$$\text{ARIMA}(p, d, q) = \text{AR}(p) + \text{I}(d) + \text{MA}(q) \quad (10)$$

式中: $AR(p)$ 为自回归模型; $I(d)$ 为差分模型; $MA(q)$ 为移动平均模型; p 、 d 、 q 分别为自回归阶数、差分阶数和移动平均阶数, 反映模型对历史观测值、差分处理次数以及随机误差项的考虑程度。

ARIMA 模型可以通过对时间序列数据进行分析 and 拟合估计出合适的模型参数, 从而进行数据预测和建模。

选取 3 个性能指标评估 LSTM-GBRT 模型的预测效果, 包括平均绝对百分误差(e_{MAPE})、平均绝对比例误差(e_{MAE})、均方根误差(e_{RMSE}), 各评价指标定义如式(11)–(13)所示。

$$e_{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{y}_t - y_t}{y_t} \right| \times 100\% \tag{11}$$

$$e_{MAE} = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t| \tag{12}$$

$$e_{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} \tag{13}$$

式中: T 为样本数量; $\hat{y}_t$ 为 t 时刻的预测值; y_t 为 t 时刻的实际值。

4 算例分析

4.1 碳排放核算

以某省级区域碳排放为研究对象, 开展省级区域的电-碳相关性分析, 碳排放因子计算数据来自中国碳核算数据库 (China Emission Accounts and Datasets, CEADs)^[32], 数据的时间步长为年度, 空间颗粒度为省级, 固体能源单位为万吨、气体能源单位为亿立方米和碳排放量单位为兆吨二氧化碳。该数据库提供了系统化的碳排放核算数据, 其中, 水泥生产过程碳排放因子 $F_p = 0.2906$ ^[33]。电力负荷数据来源于该省统计年鉴^[32], 数据单位为亿千瓦, 数据时间步长为年度。

该省级区域碳排放核算所需要参数包括 17 类一次能源消耗量与直接碳排放因子、区间电/热交易量及对应的电/热碳排放因子。直接碳排放因子按公式计算, 计算结果如表 1 所示。电/热碳排放因子按公式和计算得到, 结果如表 2 所示。该省级区域碳排放量按公式-核算获得, 其结果如表 3 所示。

将所构建的省级区域碳排放核算模型的结果与权威数据库 CEADs 及北京公众环境研究中心

表 1 各类能源直接碳排放因子

Tab. 1 Direct carbon emission factors for various energy sources

能源名称	碳排放因子	能源名称	碳排放因子
原煤	1.662	天然气	2.141
洗精煤	2.058	焦炉煤气	11.545
其他洗煤	1.187	其他煤气	5.952
煤制品	1.424	液化石油气	3.343
焦炭	2.867	炼厂干气	3.089
其他石油产品	3.088	煤油	3.036
汽油	2.927	柴油	3.058
其他焦化产品	2.508	燃料油	3.194
原油	3.039		

注: 焦炉煤气、其他煤气和天然气的碳排放因子单位为 kgCO_2/m^3 , 其余能源的碳排放因子单位为 kgCO_2/kg 。

表 2 2011 年到 2021 年电/热碳排放因子

Tab. 2 Carbon emission factors for electricity and heat production from 2011 to 2021

年份	电力碳排放因子/ ($\text{kgCO}_2 \cdot \text{kWh}^{-1}$)	热力碳排放因子/ ($\text{kgCO}_2 \cdot \text{MJ}^{-1}$)
2011	0.643 0	0.126 6
2012	0.531 5	0.120 6
2013	0.543 8	0.118 8
2014	0.454 7	0.230 2
2015	0.406 7	0.110 3
2016	0.458 1	0.106 9
2017	0.490 6	0.103 5
2018	0.473 1	0.109 3
2019	0.472 9	0.107 4
2020	0.431 1	0.136 6
2021	0.484 3	0.115 0

表 3 区域年度碳排放量

Tab. 3 Annual carbon emissions MtCO_2

年份	总碳排放量	年份	总碳排放量
2011	208	2017	256
2012	228	2018	248
2013	230	2019	254
2014	228	2020	247
2015	231	2021	259
2016	244		

(Institute of Public and Environmental Affairs, IPE) 的公开基准数据进行了对比验证, 如图 4 所示。所提模型计算结果与基准数据集之间呈现显著一致性特征, 3 组数据序列平均绝对百分比误差为

2.44 %, 最大相对误差不超过 6.56 %, 表明所提核算模型具有合理性和准确性。

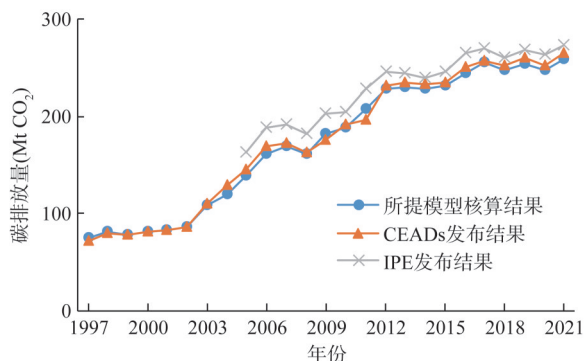


图4 总碳排放量核算结果比较

Fig. 4 Comparative analysis of total carbon emissions accounting results

4.2 电-碳相关性分析与特征构建

基于该省碳排放总量的核算结果进一步分析该省电力负荷数据的幂次函数与碳排放量的皮尔森相关系数, 分析结果如图5所示。结果表明, 电-碳相关性先随电力消费量函数中幂次指数的增加而上升, 幂次指数为0.6时相关性系数达到最大, 为0.917 0。当幂次指数为1时(即线性情况下)电-碳相关系数为0.916 8。同样, 选取部分省份进行电-碳相关性计算, 如表4所示。除青海省以外其他各省份的电-碳相关性函数中最大相关系数均在0.9以上, 表明了电-碳之间存在强相关性。

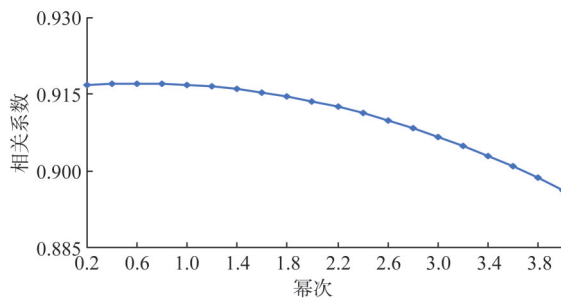


图5 电-碳相关性的变化趋势

Fig. 5 Variation trends of the electricity-carbon correlation

为挖掘更高的电-碳相关性指标用于优化该省基于电力数据的碳排放测算, 在预测前剔除各省份异常年份中含零值的异常数据, 以减少能源消费结构差异对分析结果的干扰。随后, 基于各省份主要能源消费比重开展聚类分析, 以更精确地刻画不同省级区域的能源利用特征, 为碳排放预测提供更稳

表4 部分省份电-碳最大相关系数

Tab. 4 Peak correlation coefficients of the electricity-carbon nexus in selected provinces

省份	最大相关系数	省份	最大相关系数
河北	0.991 2	安徽	0.990 6
内蒙古	0.999 0	山东	0.986 5
吉林	0.962 1	山西	0.995 0
河南	0.985 9	广东	0.984 1
湖北	0.956 5	广西	0.971 0
宁夏	0.991 9	青海	0.883 8

健的数据支撑。

由于该省的能源消费结构以原煤和电力为主, 因此选取“原煤终端消费量与总碳排放之比”及“电力终端消费量与总碳排放之比”两个指标, 对各省份进行聚类分析, 确定不同省份的能源-碳排放特征类别。所得的聚类类型标签作为模型训练特征之一, 以提升预测精度。

聚类过程分为两阶段: 先通过初步聚类, 仅保留与该省同组的省份数据, 形成包含21个省份的新数据集。再次基于相同的两个指标进行二次聚类, 以进一步细化各省份的分类, 聚类结果如图6所示。根据聚类结果将21个省份的数据集划分为3组, 并将其作为模型的输入特征, 具体分组编号如表5所示。

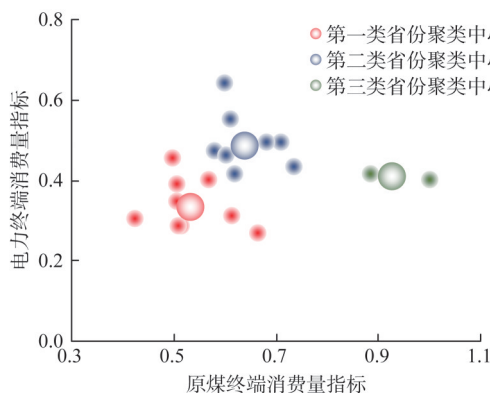


图6 各省份数据集第二次聚类结果

Fig. 6 Second-round clustering outcomes of provincial-level datasets

4.3 碳排放预测结果

基于表5聚类得到的数据集将年份、电力负荷数据、省份类型3个特征作为模型输入, 该省碳排放量作为输出, 开展神经网络训练。经过异常数据剔除后将该省1997—2008年的样本数据和其他30

表5 各省数据集分组结果

Tab. 5 Classification outcomes of provincial-level dataset groupings

第一组	第二组	第三组
云南、四川、广西、河南、甘肃、福建、重庆、陕西	内蒙古、吉林、宁夏、安徽、山东、山西、新疆、江西、河北、黑龙江	湖北、湖南、贵州

个省份 1997—2021 年的样本数据作为训练集, 该省 2009—2021 年的样本数据作为测试集。

为全面评估所提 LSTM-GBRT 模型的性能, 本文将其与 LSTM、卷积神经网络(convolutional neural network, CNN)和门控循环单元(gated recurrent unit, GRU)3 种常见深度学习模型进行了对比, 结果如图 7 所示, 预测误差指标如表 6 所示。结果表明, 相较于 LSTM、CNN、GRU 模型, LSTM-GBRT 模型的平均绝对百分误差最低, 仅为 4.16%。

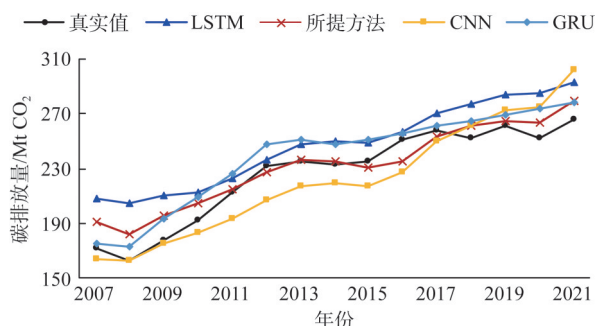


图7 4种模型的预测效果对比

Fig. 7 Prediction results comparison of four models

表6 4种模型预测误差指标对比

Tab. 6 Comparison of prediction error metrics among four models

预测模型	$e_{MAPE}/\%$	e_{RMSE}/MtCO_2	e_{MAE}/MtCO_2
LSTM-GBRT(所提模型)	4.16	11.15	9.22
LSTM	8.48	23.46	20.73
CNN	6.61	17.93	15.49
GRU	4.92	13.11	11.95

LSTM 模型擅长处理时间序列, 但易受复杂特征干扰; CNN 更适合提取局部模式, 其结构主要集中于提取短期特征, 对于挖掘时间序列中长跨度时间点的依赖关系, 该方法存在一定局限性; GRU 结构相对简单、训练快, 但对趋势变化响应不足。相比之下, LSTM-GBRT 模型结合了 LSTM 的时序

建模能力与 GBRT 对残差的精准拟合能力, 能够更全面地捕捉碳排放的动态变化特征。

为实现 2022—2030 年碳排放的预测, 首先采用 ARIMA 模型对该省该时期的电力负荷数据进行预测, 预测结果如图 8 所示, 电力负荷数据变化量近似线性增长, 符合历史变化规律。随后得到碳排放预测结果如图 9 所示。分析结果表明, LSTM-GBRT 模型相比其他预测模型能够更好地表现碳排放趋于平缓的趋势, 这是由于考虑了电力负荷数据的影响, 并且参考了其他省份的碳排放变化情况, 预测结果能够反映碳排放增长放缓并逐渐达峰的趋势。该预测结果显示该省碳排放将于 2029 年达到峰值, 峰值为 285.29 Mt CO₂。

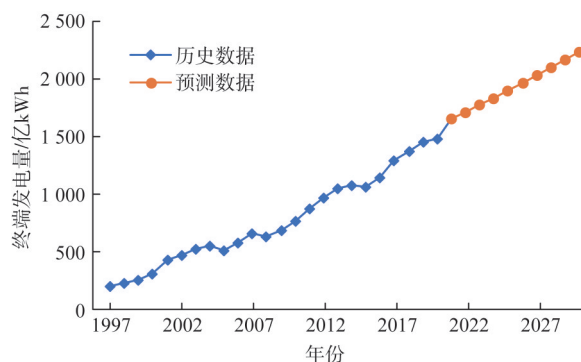


图8 电力负荷数据预测结果

Fig. 8 Prediction results of electric power data

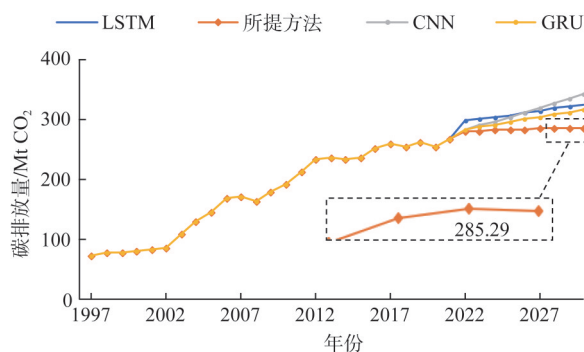


图9 4种模型的碳排放预测效果对比

Fig. 9 Comparison of carbon emission prediction performance among four models

5 结论

基于电力数据构建碳排放预测体系是支撑“双碳”战略实施的关键技术路径, 本文通过皮尔森相关系数分析发现省级区域电、碳间存在强非线性相关性, 提出了 LSTM-GBRT 碳排放预测模型, 该模

型利用深度神经网络挖掘省级区域电力负荷数据与碳排放量之间的复杂相关关系, 具有较好的通用性, 能够有效反映不同省级区域的碳排放变化趋势。

以某省份为测试分析对象, 核算了2011—2021年该省级区域历史碳排放量, 并融合21个能源相似省份的数据构建碳排放预测的训练数据集, 结果表明, 该模型的 e_{MAPE} 为4.16%, e_{MAE} 为9.22 MtCO₂, e_{RMSE} 为11.15 MtCO₂, 优于LSTM、CNN和GRU等模型。结合ARIMA电力数据预测结果捕捉到该省份碳排放将于2029年达峰的拐点, 为协助政府制定“双碳”政策提供了数据参考。

未来会继续对各行业开展电-碳相关性分析及碳排放预测, 重点建立黑色金属、有色金属和非金属等高耗能行业的碳排放预测模型, 并加入政策变量以完善模型。

参考文献

- [1] 温馨, 韩东. 全球变暖加速: 2024年再成最热年份[J]. 生态经济, 2025, 41(2): 1-4.
WEN Xin, HAN Dong. Global warming accelerates: 2024 projected to become the hottest year on record again [J]. Ecological Economy, 2025, 41(2): 1-4.
- [2] 习近平: 中国力争碳排放2030年前达到峰值, 努力争取2060年前实现碳中和[J]. 今日制造与升级, 2020(9): 10.
XI Jinping: China aims to peak carbon emissions before 2030 and strive to achieve carbon neutrality before 2060 [J]. Manufacture & Upgrading Today, 2020(9): 10.
- [3] HRITHIK P M, REHMAN Z M, DAR A A, et al. Forecasting CO₂ emissions in India: a time series analysis using ARIMA [J]. Processes, 2024, 12(12): 2699.1-2699.10.
- [4] DAR A A, JAIN A, MALHOTRA M, et al. Time series analysis with ARIMA for historical stock data and future projections [J]. Soft Computing, 2024(28): 12531-12542.
- [5] 刘吉雄, 邹瑞, 许思为. 基于ARIMA的风电发电量预测[J]. 现代信息科技, 2025, 9(4): 157-161, 166.
LIU Jixiong, ZOU Rui, XU Siwei. Power generation forecasting of wind-generated electricity based on ARIMA [J]. Modern Information Technology, 2025, 9(4): 157-161, 166.
- [6] 刘泽珩, 鞠美庭. 基于STIRPAT模型的北方某滨海城市碳达峰情景分析[J]. 环境科学, 2025, 46(9): 5524-5534.
LIU Zejun, JU Meiting. Scenario analysis of carbon peaking in a northern coastal city based on STIRPAT model [J]. Environmental Science, 2025, 46(9): 5524-5534.
- [7] 李志东, 杜烜锋. 基于STIRPAT模型的辽宁省碳排放预测[J]. 沈阳师范大学学报(自然科学版), 2024, 42(5): 398-402.
LI Zhidong, DU Xuanfeng. Carbon emission prediction in Liaoning province based on STIRPAT model [J]. Journal of Shenyang Normal University (Natural Science Edition), 2024, 42(5): 398-402.
- [8] ABIDEMI O S, HUSEYIN O, MEHDI S, et al. The determinants of CO₂ emissions in Brazil: the application of the STIRPAT model [J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2023, 45(4): 10843-10854.
- [9] 单葆国, 姚力, 张成龙, 等. 中国能源消费碳排放达峰情景预测: 基于STIRPAT扩展模型[J]. 环境科学, 2025, 46(7): 4052-4064.
SHAN Baoguo, YAO Li, ZHANG Chenglong, et al. Carbon peak prediction of energy consumption in China: based on STIRPAT extended model [J]. Environmental Science, 2025, 46(7): 4052-4064.
- [10] 胡旭东, 林振福, 黄彦璐, 等. 南方区域电力碳排放趋势与预测分析——基于STIRPAT模型的研究[J]. 技术经济与管理研究, 2024(9): 75-80.
HU Xudong, LIN Zhenfu, HUANG Yanlu, et al. Analysis of carbon emission trends and forecasts for electricity in Southern China: research based on STIRPAT model [J]. Journal of Technical Economics & Management, 2024(9): 75-80.
- [11] LI S, TONG Z, MAROON H. Estimation of transport CO₂ emissions using machine learning algorithm [J]. transportation research part d: transport and environment, 2024 (133): 104276.1-104276.11.
- [12] 钟炜. 基于LSTM模型的锂电池容量预测[J]. 汽车与新动力, 2025, 8(1): 6-10.
ZHONG Wei. Lithiumion battery capacity prediction based on LSTM models [J]. Automobile and New Powertrain, 2025, 8(1): 6-10.
- [13] 刘淳森, 曲建升, 葛钰洁, 等. 基于LSTM模型的中国交通运输业碳排放预测[J]. 中国环境科学, 2023, 43(5): 2574-2582.
LIU Chunsen, QU Jiansheng, GE Yujie, et al. LSTM model-based prediction of carbon emissions from China's transportation sector [J]. China Environmental Science, 2023, 43(5): 2574-2582.
- [14] QI Y, PENG W, YAN R, et al. Use of BP neural networks to determine China's regional CO₂ emission quota [J]. Complexity, 2021, 2021(1): 6659302.1-6659302.8.
- [15] 赵金元, 马振, 唐海亮. BP神经网络和多元线性回归模型对碳排放预测的比较[J]. 科技和产业, 2020, 20(11): 172-176.
ZHAO Jinyuan, MA Zhen, TANG Hailiang. Comparison of BP neural network and multiple linear regression models for carbon emissions prediction [J]. Science Technology and Industry, 2020, 20(11): 172-176.
- [16] 陈鸿林, 贾宪章, 贾宝柱, 等. 基于BP神经网络算法的光伏发电预测研究[J]. 中国设备工程, 2024(3): 153-155.
CHEN Honglin, JIA Xianzhang, JIA Baozhu, Research on photovoltaic power generation prediction using backpropagation neural network algorithm [J]. China Plant Engineering, 2024(3): 153-155.
- [17] 王永利, 李颀雯, 王欢, 等. 基于改进BP神经网络的河北省碳排放预测[J]. 生态经济, 2024, 40(6): 30-37.
WANG Yongli, LI Yiwen, WANG Huan, et al. Carbon emission prediction of Hebei province based on improved BP neural network [J]. Ecological Economy, 2024, 40(6): 30-37.
- [18] 朱少民, 夏虹, 吕新知, 等. 基于ARIMA和LSTM组合模型的核电厂主泵状态预测[J]. 核动力工程, 2022, 43(2): 246-253.

- ZHU Shaomin, XIA Hong, LÜ Xinzhi, et al. Condition prediction of reactor coolant pump in nuclear power plants based on the combination of ARIMA and LSTM [J]. Nuclear Power Engineering, 2022, 43(2): 246 – 253.
- [19] 陈蕪峰,王贺,李岩,等.组合两步分解和ARIMA-LSTM的短期风速预测研究[J].太阳能学报,2024,45(2):164 – 171.
- CHEN Hongfeng, WANG He, LI Yan, et al. Short-term wind speed prediction by combining two-step decomposition and ARIMA-LSTM [J]. Acta Energeiae Solaris Sinica, 2024, 45(2): 164 – 171.
- [20] 王英伟,马树才.基于ARIMA和LSTM混合模型的时间序列预测[J].计算机应用与软件,2021,38(2):291 – 298.
- WANG Yingwei, MA Shucai. Time series forecasting based on ARIMA-DLSTM hybrid model [J]. Computer Applications and Software, 2021, 38(2): 291 – 298.
- [21] WEN T, LIU Y, BAI Y, et al. Modeling and forecasting CO₂ emissions in China and its regions using a novel ARIMA-LSTM model [J]. Heliyon, 2023, 9(11): 21241.1 – 21241.10.
- [22] WANG C, XIE F, YAN J, et al. A U-MIDAS modeling framework for forecasting carbon dioxide emissions based on LSTM network and LASSO regression [J]. Energy Reports, 2025(13): 16 – 26.
- [23] CAI Y, FENG J, WANG Y, et al. The Optuna-light GBM-XGBoost model: a novel approach for estimating carbon emissions based on the electricity-carbon nexus [J]. Applied Sciences, 2024, 14(11): 4632.1 – 4632.10.
- [24] 沈亮,樊涛,慕群,等.基于“电—碳计算模型”碳排放核算方法评估标准体系研究[J].中国软科学,2023(S1):318 – 333,340.
- SHEN Liang, FAN Tao, MU Qun, et al. Study on evaluation standardization system of carbon emission accounting methodology based on “electricity-carbon emission calculation model” [J]. China Soft Science, 2023(S1): 318 – 333,340.
- [25] 张晨,张路争.基于电力大数据的城市“电-碳测算模型”构建与应用[J].资源节约与环保,2025(1):66 – 69.
- ZHANG Chen, ZHANG Luzheng. Development and application of an urban electricity-carbon measurement model utilizing power big data [J]. Resources Economization & Environmental Protection, 2025(1): 66 – 69.
- [26] 王成围,李沛,杨至元,等.基于电力数据的能耗及碳排放测算模型构建与应用[J/OL].南方电网技术,1 – 9[2025 – 02 – 19]. <http://kns.cnki.net/kcms/detail/44.1643.tk.20250219.1128.003.html>.
- WANG Chengwei, LI Pei, YANG Zhiyuan, et al. Construction and application of energy consumption and carbon emission calculation model based on electrical data [J/OL]. Southern Power System Technology, 1 – 9[2025 – 02 – 19]. <http://kns.cnki.net/kcms/detail/44.1643.tk.20250219.1128.003.html>.
- [27] 侯咎宇,蔡亦浓,刘博,等.基于电碳计算模型的区域与行业碳排放监测应用实践分析[J].数字技术与应用,2024,42(11):85 – 87.
- HOU Zanyu, CAI Yinong, LIU Bo, et al. Operational implementation analysis of carbon emission monitoring across regions and industries using an electricity-carbon computational model [J]. Digital Technology & Application, 2024, 42(11): 85 – 87.
- [28] 国际标准化组织.产品碳足迹量化与交流的要求与指导技术规范:ISO 14067[S].日内瓦:国际标准化组织,2019.
- [29] IPCC. 2006 IPCC guidelines for national greenhouse gas inventories [EB/OL]. [2022 – 08 – 19]. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html>.
- [30] 谢平平,陆秋瑜,王雪林,等.适应新型电力系统发展风光储联合发电站碳排放核算方法研究[J].广东电力,2025,38(6):1 – 10.
- XIE Pingping, LU Qiuyu, WANG Xuelin, et al. Research on carbon emission accounting method for wind-solar-storage combined power plant adapting to the development of new power system [J]. Guangdong Electric Power, 2025, 38(6): 1 – 10.
- [31] 王佳,吴任博,肖健,等.新型电力系统能源调度与碳排放计价优化方法研究[J].电测与仪表,2024,61(10):17 – 25.
- WANG Jia, WU Renbo, XIAO Jian, et al. Research on the optimization method of energy dispatching and carbon emission pricing for novel power system [J]. Electrical Measurement & Instrumentation, 2024, 61(10): 17 – 25.
- [32] 中国碳核算数据库(CEADs) [EB/OL]. [2024 – 03 – 20]. <https://www.ceads.net.cn/>.
- [33] SHAN Y, GUAN D, ZHENG H, et al. China CO₂ emission accounts 1997 – 2015 [J]. Scientific Data, 2018, 5(1):170201.1 – 170201.14.
- 收稿日期:2025-04-01;网络首发日期:2025-07-30
- 作者简介:
- 林成(1975),男,高级工程师,硕士,研究方向为电力系统规划,2606241251@qq.com;
- 吕俊兵(2001),男,通信作者,硕士研究生,研究方向为碳排放预测,abing427@mail.dlut.edu.cn;
- 张裕(1983),男,高级工程师,硕士,研究方向为碳排放核算,zhangyu830906@163.com。

征稿启事

《南方电网技术》是由中国南方电网有限责任公司主管、南方电网科学研究院有限责任公司主办的电力技术类科技期刊,向国内外公开发行人。刊登电力系统的科研、规划、设计、工程、生产运行、设备和系统维护等方面的成果、经验和动态,尤其是科研创新方面的论文。设有特约专稿、高压直流输电、高电压与绝缘、系统分析与运行、控制与保护、调度与通信、输变电技术、电力技术经济、智能电网、分布式电源与微电网、低碳电力等栏目。热烈欢迎国内外的电力研究人员、工程技术人员、大专院校师生踊跃投稿。

1 文稿要求

1)文稿应未在其他刊物上发表过,译文稿件须附上原作者的书面许可证明。内容要有科学性、创新性和实用性;论点明确、数据可靠、说理严谨、数学推导简明;语言流畅、文字简练、层次分明、重点突出。

2)学术论文请按GB/T 7713.2—2022《科学技术报告、学位论文和学术论文的编写格式》书写,篇幅以版面不超过9 000字(包括图表所占篇幅)为宜。同时请写明如下a、b、c、d、e五项:

a)作者署名(不宜超过7位)、作者单位、地址、电话、email信箱和邮政编码;

b)论文题目和作者单位英译文,作者姓名的汉语拼音;

c)作者的简历,包括姓名、出生年、性别、籍贯、职称、学位、当前从事的研究或工作等项;

d)350字左右的中文摘要,3~8个关键词,以及摘要和关键词相应的英文文本。摘要采用第三人称写法,内容包括论文的目的、方法、结果和结论;

e)若为国家基金资助项目或部、省级重大科研攻关项目,需提供项目名称及编号。

2 书写格式

1)要求文稿正文字号为5号,汉字字体为宋体,英数字体为Times New Roman,希腊字体为Symbol。文稿须采用可编辑的电子文档形式。

2)计量单位采用国家法定计量单位和符号,不能用“大气压”、“kgf/cm²”、“卡”、“ppm”等已废除的计量单位。

3)文中和公式中容易混淆的字符请注明文种(希文、英文、罗马字等)、大小写、上下标及上下标字母含义,表示矢量及矩阵的字符用黑斜体表示或特别注明。

4)文稿标题中不宜用缩略词(化学符号和公知公用者除外);摘要和正文中的缩略词在第一次出现时必须写出全称,后加括号附缩略词。

3 表格、插图及参考文献

1)表格尽量采用“三线表”。表格的上方写表序和表名。表名应有自明性且中文、英文表名并列。表注放在表底,缩入1个汉字以“注:”起头排版。

2)插图的下方面应有图序和图名。图名应有自明性且中英文图名并列。工程图、电气图和函数图采用AutoCAD、

Adobe Illustrator或Corel DRAW软件绘制,工程图和电气图要符合国家标准规定,函数图要标明曲线序号及其注释,坐标轴上要有标值、标目,注明物理量和单位;数码照片图具备200万总像素以上。

3)图片和表格中的文字、变量、单位和数字要齐全、清晰。

4)参考文献应选用公开发表的资料,按引用的先后次序列表于文后,以[1]、[2]…标识序号,且与正文中的指示序号一致。按《信息与文献 参考文献著录规则,GB/T 7714—2015》和《中国学术期刊(光盘版)检索与评价数据规范,CAJ-CD B/T 1—2006 修订版试行稿》的要求著录文后参考文献。中文参考文献后要并列其英文译文(如示例[6]所示)。

文献类型及其标识为:普通图书[M];会议论文[C];报纸文章[N];期刊文章[J];学位论文[D];报告[R];标准[S];专利[P];汇编[G];档案[B];古籍[O];参考工具[K];其他未说明的文献类型,例如可公开的政府行政部门编号文件、行业或大公司的技术规范或工作手册[Z]。网上期刊[J/OL];网上电子公告[EB/OL]。电子文献尚需在载体标记后加上发表或更新日期(加圆括号)、引用日期(加方括号)和电子文献网址。

4 投稿

1)请登录《南方电网技术》编辑部网站 <https://nfdwjs.csg.cn>,选择“作者登录”,登录或注册登录后选择投稿并按提示操作,登录方面出现问题请致电020-36625646咨询。

2)投稿时务必附上联系电话(手机为佳)。

3)编辑部不保证留存稿件,请作者自留底稿。稿件评审进程和结论可在《南方电网技术》编辑部网站上即时查询。作者也可致电020-36625642、36625643或发送电子邮件到nfdwjs@csg.cn向编辑部了解审稿情况。

4)文章一经发表,编辑部便取得作者的自动授权,可将文章收录入本刊网站,及与本刊有合作关系的数据库。

参考文献著录格式示例:

- [1] 刘国钧,陈绍业,王凤翥.图书馆目录[M].北京:高等教育出版社,1957:15-18.
- [2] 钟文发.非线性规划在可燃毒物配置中的应用[C]//赵玮.运筹学的理论与应用——中国运筹学会第五届大会论文集.西安:西安电子科技大学出版社,1996:468-471.
- [3] 冯西桥.核反应堆压力管道与压力容器的LBB分析[R].北京:清华大学核能技术设计研究院,1997.
- [4] 中国国家标准化管理委员会.信息与文献 参考文献著录规则:GB/T 7714—2015[S].北京:中国标准出版社,2015.
- [5] 邓一刚.全智能节电器:200610171314.3[P].2006-12-13.
- [6] 陈祥.无线技术在现代电网管理中的应用与展望[J].南方电网技术,2009,3(6):85-89.
CHEN Xiang. Development of wireless technology in modern power grid management and its perspective [J]. Southern Power System Technology, 2009, 3(6): 85-89.
- [7] 王明亮.关于中国学术期刊标准化数据库系统工程的进展[EB/OL].(1998-08-16)[1998-10-04].<http://www.cajcd.edu.cn/pub/wml.txt/980810-2.html>.

《南方电网技术》编辑部