

大语言模型赋能的气象 Agent 构建及其在短期负荷预测中的应用

蔡子龙¹, 林红娅¹, 沈赋¹, 毛肖², 王健¹, 杨宇林¹

(1. 昆明理工大学电力工程学院, 昆明市 650500; 2. 昆明理工大学城市学院, 昆明市 650500)

摘要:【目的】多元气象要素作为短期电力负荷预测的核心驱动因素,其精细化数据的获取和应用仍面临显著约束。基于此,本文对大语言模型(large language model, LLM)赋能的气象 Agent 进行研究,并将其应用于短期负荷预测中。【方法】首先,通过大语言模型开发平台构建融合 LLM 的工作流智能体,实现了结构化天气信息动态获取和解析,增强了预测模型的实时响应能力,降低了短期负荷预测系统的开发和使用难度。在此基础上,建立了融合气象要素、电力负荷时序数据、日期特征的电力系统短期负荷预测多模态输入体系,结合时间序列特性与梯度提升树算法构建电力系统短期负荷预测模型。最后,通过实验验证与案例分析验证模型和算法的可行性和有效性。【结果】所提模型和算法简化了气象信息获取流程,提高了预测精度。【结论】研究成果为电力系统中低成本获取和使用气象信息提供了新的研究思路。

关键词:大语言模型(LLM);短期负荷预测;气象 Agent;工作流

中图分类号: TM715

文献标志码: A

文章编号: 1000-7229(2026)09-0001-14

DOI: 10.12204/j.issn.1000-7229.2026.09.001

LLM-Empowered Meteorological Agent Construction and Its Application in Short-Term Power Load Forecasting

CAI Zilong¹, LIN Hongya¹, SHEN Fu¹, MAO Xiao², WANG Jian¹, YANG Yulin¹

(1. Faculty of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650500, China;

2. City College, Kunming University of Science and Technology, Kunming 650500, China)

ABSTRACT: [Objective] As the core driving factors for short-term power load forecasting, multivariate meteorological elements still confront prominent constraints in the acquisition and practical application of refined meteorological data. Accordingly, this paper investigates meteorological Agents empowered by large language models (LLMs) and applies them to short-term load forecasting. [Methods] First, a workflow agent integrated with large language models is constructed via the LLM development platform, which realizes the dynamic acquisition and intelligent parsing of structured meteorological information. This approach effectively improves the real-time response performance of the forecasting model and reduces the development and application complexity of short-term load forecasting systems. On this basis, a multimodal input framework for power system short-term load forecasting is established by integrating meteorological factors, power load time-series data, and calendar features. Combined with the inherent characteristics of time series and the gradient boosting tree algorithm, an optimized short-term power load forecasting model is further developed. Finally, the feasibility and effectiveness of the proposed model and algorithm are verified through experimental validation and case analysis. [Results] The proposed model and algorithm streamline the meteorological information-acquisition process and improve the forecasting accuracy. [Conclusions] This research provides a novel technical idea for the low-cost acquisition and efficient utilization of meteorological data in power system operation and analysis.

This work is supported by National Natural Science Foundation of China (No. 62563016), Yunnan Applied Basic Research Project (No. 202601AS070059, No. 202501AT070350), Yunnan Xingdian Talent Support Program (No. KKRD202204021) and High-level Platform Construction Project of Kunming University of Science and Technology (No. KKZ7202004004).

KEYWORDS: large language model (LLM); short-term power load forecasting; meteorological agent; workflow

基金项目:国家自然科学基金项目(62563016);云南省应用基础研究计划项目(202601AS070059,202501AT070350);云南省兴滇英才支持计划项目(KKRD202204021);昆明理工大学高层次人才平台建设项目(KKZ7202004004)

0 引言

短期负荷预测作为电力调度决策的核心支撑技术,其精度直接关系到机组最优组合、经济调度、最优潮流、电力市场、备用容量配置等关键环节。提高短期负荷预测精度对于新型电力系统的构建具有重要意义^[1-4]。

电力系统的短期负荷特性受气象条件、社会经济活动、居民用电习惯等众多因素的综合影响,这些因素之间相互关联且呈现出高度的非线性特征,传统的统计和机器学习模型如时间序列法、回归分析法、自回归积分移动平均模型 (autoregressive integrated moving average, ARIMA)^[5]、支持向量机 (support vector machine, SVM)^[6-7]、联邦学习^[8]以及以极端梯度提升 (eXtreme gradient boosting, XGBoost)^[9]和轻量级梯度提升机 (light gradient boosting machine, LightGBM)^[10]为代表的梯度提升决策树模型,这些模型虽具备可解释性优势,但其难以处理时间、天气、负荷之间复杂的非线性特征。以神经网络^[11]、循环神经网络 (recurrent neural network, RNN)^[12]、卷积神经网络 (convolutional neural network, CNN)^[13-16]、长短期记忆网络 (long short-term memory network, LSTM)^[17-19]、双向长短期记忆网络 (bidirectional long short-term memory network, BiLSTM)^[20]、时间卷积网络 (temporal convolutional network, TCN)^[21]等深度学习为代表的预测模型^[22-23],通过时序特征自动提取,能显著提升预测精度,但模型的可解释性差,且存在模型固化后无法在线更新的瓶颈。近年来发展起来的大语言模型 (large language model, LLM) 在提示词工程、多模态处理、模型微调等方面具有较强的技术优势与实际应用价值,在最优潮流计算、专业问答、电力设备智能运维等智能化领域已经展示出广泛的应用潜力^[24-27],然而基于 LLM 构建的短期负荷预测系统在算力要求高、参数规模大和部署成本过高的问题^[28]。

气象是影响短期负荷的重要因素^[29-31],传统气象数据主要从以下 3 个途径获取:一是通过专业气象预测平台从地面气象站获取气象监测数据,涵盖温度、湿度、风速、气压、降水及云量等关键指标;二是通过气象卫星获取高时空分辨率数据,可弥补地面气象站点在空间分布上的局限性;三是采用中国气象局全球预报系统发布的未来 72 h 温度、风速、相对湿度等数值天气预报数据。然而,上述传统的气象数据获取方式主要依托于应用程序接口 (application programming interface, API) 及文件传输协议 (file transfer protocol, FTP),这在处理非结构化语义信息

时存在明显的信息损耗与实时性滞后,且高度依赖高频的人工干预,对于缺乏专业气象接口和数据分析团队的工业园区、分布式电站及基层电力单位而言,这种高门槛的准入机制严重阻碍了其预测精度的提升。因此,如何降低获取难度、简化处理流程,已成为此类企业亟待解决的难题。

基于此,本文针对现有短期负荷预测中存在的多源气象数据获取能力不足、预测系统运维门槛高等问题,开展大语言模型赋能的气象 Agent 构建的研究工作,并将其应用于短期负荷预测中。本文所做的创新性工作如下:

1) 针对多源异构气象信息整合难、信息孤岛严重的技术难点,构建以大语言模型为逻辑中枢的集成框架。该框架改变了传统 API/FTP 硬编码的数据接入模式,通过语义映射机制实现多维气象信息的深度对齐,从架构底层解决了异构数据接入的技术难题。

2) 构建具备较强交互能力的气象 Agent。首先,该 Agent 通过自然语言接口将复杂的业务需求转化为自动化工作流,实现气象数据的全自动获取与持久化存储。其次,利用 LLM 具备的良好语义整合与知识推理能力,整合和处理多源异构气象信息。通过自然语言指令集,LLM 能实现 JSON (JavaScript object notation)、可扩展标记语言 (extensible markup language, XML)、CSV (comma-separated values) 等异构格式的自动化模态对齐,确保特征维度的全局一致性。最后,通过实验量化验证所构建的气象 Agent 在提升数据获取方面的时效性。

3) 针对工业园区、分布式电站以及基层电力单位等轻量化应用场景,构建基于梯度提升回归树 (gradient boosting regression tree, GBRT) 的预测模型。案例分析表明,在不显著增加模型算力消耗的前提下,GBRT 能够有效提升平稳气象场景下的预测精度。研究成果对 LLM 在电力系统中的轻量化、低成本落地具有重要的参考价值。

1 短期负荷预测系统总体架构设计

本文所提短期负荷预测系统集成气象 Agent、舆情监测 Agent、极端天气预警 Agent、能量管理系统 (energy management system, EMS) 以及需求响应平台等模块,在短期负荷预测系统中实现了多源数据的整合,其总体架构设计如图 1 所示。

在图 1 中,舆情监测 Agent 的主要作用是对舆情的的情感与活跃度进行量化^[32-33],将舆情热度与情感得分转化为数值特征,形成负荷波动舆情预警系数,将其输入短期负荷预测模型,以捕捉社会活动突变引发

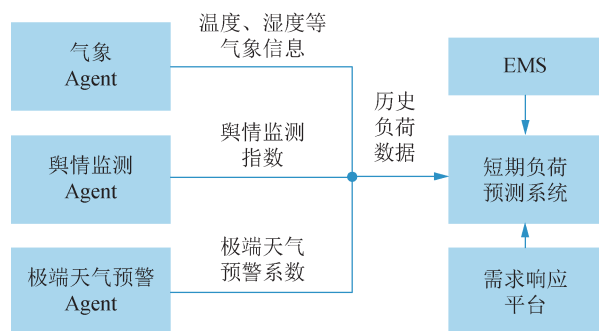


图1 短期负荷预测系统总体架构

Fig. 1 Overall architecture of short-term load forecasting system

的负荷异常波动。极端天气预警 Agent 的主要作用是对气象局发布的红色高温和台风等预警信号进行实时监控。该 Agent 将每次极端天气下的预测偏差自动存入向量数据库,形成长效记忆。当实际负荷与预测值出现较大偏差时,Agent 自动复盘其与历史极端天气的关联度,动态微调搜索空间参数,实现系统的闭环优化。最后,在预测系统的输入层增加极端天气预警指数,用于量化极端天气对系统预测值的影响。

考虑到电力负荷预测对数据隐私与系统稳定性的高要求,同时极端预警信号在统计学上属于稀疏样本,本文已有的预测数据集尚没有舆情监测数据和极端气象预警信息,基于以上考虑,本文的短期负荷预测原型系统仅对后2个模块接入系统的方案进行讨论,暂未将其作为闭环控制的直接判据。随着这两大模块的接入,系统能从单纯依赖物理气象数据转向“物理+社会”双驱动,显著提升模型在特殊时段的动态适应性。鉴于此,以下内容仅针对平稳气象信息获取的气象 Agent 的构建技术进行研究。

2 LLM 赋能的气象 Agent 构建

2.1 气象 Agent 架构设计

为构建一个具备高可用性、健壮性的气象信息获取 Agent,将气象 Agent 抽象为一个4层的分层架构,以解决系统访问中 API 访问限制、返回的数据格式异常以及地理位置解析失败等问题。气象 Agent 的架构设计如图2所示。

由图2可知,气象 Agent 由4层结构组成,分别为交互与编排层、逻辑执行与 workflow 层、容灾与冗余层、知识与存储层。

1)交互与编排层。

该层为 Agent 的接口,负责理解用户意图并驱动后台任务。在该层定义人设与回复逻辑,进行用户意图识别。此外,该层还负责进行输入的标准化工作,其利用 LLM 将模糊查询地址转化为标准经纬度

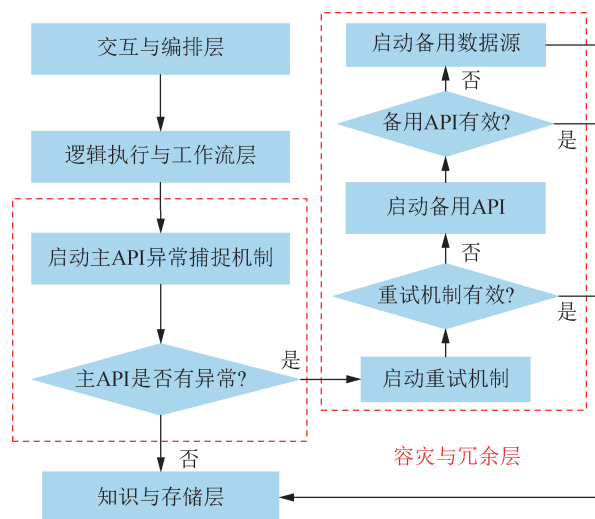


图2 气象 Agent 架构设计

Fig. 2 Architecture design of the meteorological agent

或城市 ID。

2)逻辑执行与 workflow 层。

该层为气象 Agent 的核心骨架,其主要作用是通过 workflow 来实现结构化气象信息的获取。在 workflow 中特别设计用于数据清洗的功能节点,使用 JSON 格式过滤掉 API 返回的冗余字段,仅提取与短期负荷预测相关的温度、湿度等关键参数,节省 Token 消耗。

3)容灾与冗余层。

该层提供异常捕捉机制、自动重试机制和备用数据源,以确保气象 Agent 不宕机。其中访问异常捕捉机制通过对访问返回的状态码进行异常监控,实时捕捉权限限制、访问限流或服务异常信号。在捕捉到访问异常后,启动自动重试机制。若自动重试机制失败,则自动切换到备用 API 进行处理。最后,该层提供备用数据源进行兜底,即若所有 API 均调用失效,则从历史数据库中检索并返回历史相似日的气象信息。

4)知识与存储层。

知识与存储层用于存储历史气象数据,包括各时间节点的温度和湿度等结构化气象信息。

2.2 气象 Agent 搭建流程

在 Coze 平台下,气象 Agent 是一个集成了 LLM、插件工具、知识库和预设 workflow 的任务导向型 AI 应用。它可以根据用户气象查询指令,自主调用模型、知识库、插件等技能并完成编排,最终完成用户气象信息的返回。通过 Coze 开发平台构建的气象 Agent 既可以单独使用,也可以集成到短期负荷预测系统中使用。具体搭建过程如图3所示。

在图3中,人设与回复逻辑节点主要完成气象 Agent 架构中的交互与编排层任务。其中人设是气象 Agent 的身份性格底色,定义回复的语气与立场,本文

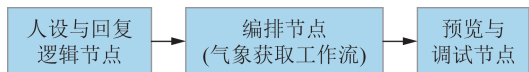


图3 气象 Agent 搭建过程示意图

Fig. 3 Schematic diagram of the meteorological Agent construction process

的人设定为“专业气象信息获取专家”。回复逻辑是回应的规则框架,用于规范回答内容的边界。编排节点主要完成气象 Agent 架构中第2层至第4层的工作。具体是通过 LLM 将回复逻辑输入到气象获取 workflow,以获取结构化气象信息;预览与调试节点的作用是通过与智能体进行交互,以验证智能体效果并修正其中的问题。在气象 Agent 中,可以通过该节点与 Agent 工作流交互获取气象预测信息。

2.3 气象获取 workflow 设计

气象获取 workflow 是气象 Agent 的核心组成部分。在 Coze 平台下,气象获取 workflow 执行过程如图 4 所示。本文从以下 3 个方面对气象获取 workflow 的可靠性进行针对性设计。

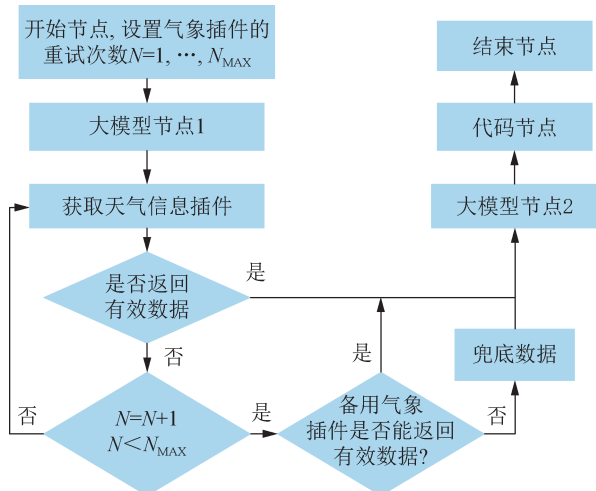


图4 气象信息获取 workflow 流程图

Fig. 4 Workflow flowchart of meteorological information acquisition

首先,开始节点用于处理用户的输入信息,此处输入所查询天气信息的城市名,设置气象插件的重试次数为 N_{MAX} 。大模型节点1用于整理输入信息,设置查询位置、查询时间等标准解析模块,将这些参数统一转换为插件所需的地点和时间戳格式,解决气象查询的时空精准性需求。

其次,根据短期负荷预测程序调用气象信息 workflow 的场景,设计带双分支校验的异常处理逻辑,实现思路为:在调用气象插件节点后,先校验返回数据是否有缺失的状态信息,如果返回状态正常,则直接进入数据格式整理环节,否则触发重试机制,重试机制失败则调用备用气象数据源节点,如果备用数据

源可用,则返回备用数据源节点查询结果,否则最终返回兜底数据,如采用历史均值数据或者最近同类型日对应时段气象数据。

最后,通过大模型节点2对查询到的天气信息进行处理,以便形成结构化的查询结果。本文对 LLM 节点设置了精确的提示词,以便 LLM 发挥其强大的自然语言处理能力,从而得到高质量的结构化气象信息,如附录 A 所示。代码节点作为适配处理节点,用于将 LLM 返回的结构化信息进行存储,最后在结束节点结束 workflow 的处理流程。

3 短期负荷预测模型

3.1 整体思路

为验证气象智能体与预测算法的耦合性能,本文构建了实证研究平台,通过典型场景下的闭环测试评估其工程可行性。该平台首先通过气象 Agent 获取结构化气象信息,其次对获取的气象信息进行数据预处理、特征选择和特征工程,最后将预测数据注入预测模型,实现从非结构化气象感知到负荷预测决策的闭环验证环境,其预测流程如图 5 所示。

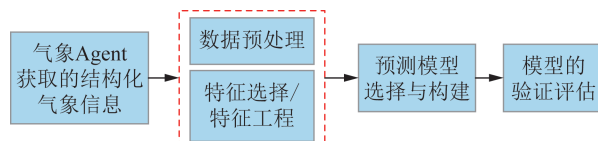


图5 预测流程

Fig. 5 Prediction process

由图 5 可以看出,气象 Agent 所获取的结构化气象数据并不能直接等同于预测模型的有效输入。从原始气象源到最终预测结果的生成,须通过数据预处理、特征选择、特征工程、预测模型的选择与构建以及模型评估等环节。预测模型的选择直接决定了系统对 Agent 所获取信息的利用深度。鉴于气象驱动因子与负荷之间存在复杂的映射关系,本文采用在处理高维特征和捕捉非线性规律方面具有显著优势的 GBRT 模型作为短期负荷预测模型。相对于传统模型,GBRT 捕捉负荷与气象、负荷与时间的复杂非线性关系;相对于深度学习模型和大语言模型,GBRT 具有求解效率高的优点。GBRT 是一种强大的集成学习算法,其核心原理是通过分阶段构建弱学习器的加权组合来逼近目标函数。相比于线性回归或基础的时序模型,该模型能更有效地挖掘 Agent 产出的多维气象特征间的耦合关系。在特征空间构建完成后,通过精细化的模型参数调优,并结合多维度的验证评估,确保了预测精度的稳健性。

3.2 数据预处理

数据预处理的作用是对原始数据进行清洗与标准化。针对本文提出的电力负荷预测机器学习模型,数据预处理过程主要包括以下步骤:首先,统一时间格式,排序去重,确保时间序列严格单调递增,以避免时间戳混乱。其次,进行缺失值处理。对缺失的温度数据进行时间插值,以保留温度随时间的连续性。由于湿度变化较缓慢,使用前向填充的方式进行插值;滞后特征会导致前几行缺失,生成时需删除缺失行。

3.3 特征工程

特征工程的意义是提取时间与物理意义特征,核心是捕捉负荷的时间自相关性与外部因素的非线性关系。GBRT模型特征包括:

1)滞后特征。基于电力负荷序列的自相关性理论,短期负荷具有显著的惯性效应,当前时刻负荷深度依赖于前序时刻的状态^[34-35]。为此,选取0.5、1.0、1.5 h滞后时间点负荷值,用于捕捉负荷的短期波动;选择3、6 h滞后时间点负荷值,用于反映日间负荷变化规律;选择24 h滞后时间点负荷值,对应完整日周期。滞后特征对应自回归模型AR(p)为:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \quad (1)$$

式中: y_t 为 t 时段的负荷值; c 为常数项,表示负荷的均值; p 为模型的滞后阶数,即模型使用目标变量自身过去 p 个时间点的历史观测值作为输入特征来预测当前值; ϕ_i 为自回归系数,表示过去 i 时刻的值对当前值的影响权重; ε_t 为白噪声误差项。

2)时间周期特征。基于社会生产活动的节奏性规律,通过正余弦转换将时间戳映射至单位圆,解决了传统整数编码(如23点到0点)的数值断层问题^[36-37]。时间的周期特征可表示为:

$$\begin{cases} X_{\sin} = \sin(2\pi t/\tau) \\ Y_{\cos} = \cos(2\pi t/\tau) \end{cases} \quad (2)$$

式中: $X_{\sin}=\sin(2\pi t/\tau)$ 为时间点 t 的正弦编码,用于表征周期 τ 内的波动特性; $Y_{\cos}=\cos(2\pi t/\tau)$ 为时间 t 的余弦编码,用于表征周期 τ 内的相位信息。两者共同构成时间的周期特征表示。

3)温度相关特征。选取3 h温度变化率差分信息,反映温度突变对空调等负荷的影响。

4)温度平方项特征。电力负荷与温度并非简单的线性关系,在极热和极冷天气下负荷均会激增。引入平方项是为了模拟这种非线性U型曲线特征^[38]。本文将温度平方用于捕捉负荷值与温度之间的非线性关系。

$$y_t = \beta_0 + \beta_1 T_t + \beta_2 T_t^2 + \omega_t \quad (3)$$

式中: β_0 为常数项; β_i ($i=1,2$)为影响系数,分别反映温度一次项、温度平方项对预测值的影响; T_t 为 t 时段温度值; ω_t 表示随机误差项。

5)交互特征。引入温度与湿度的交互特征,通过乘积形式表征温湿度的综合效应,体现其对负荷的叠加影响^[39],如式(4)所示。

$$I_{t,i} = T_{t,i} H_{t,i} \quad (4)$$

式中: $I_{t,i}$ 表示第 i 日 t 时间点的交互特征量; $T_{t,i}$ 、 $H_{t,i}$ 分别表示第 i 日 t 时间点的温度、相对湿度值。

6)日期特征。周末用电模式与工作日显著不同,例如周末商业用电减少,居民用电增加。为此将周六或周日标记为1,将周一至周五标记为0。

3.4 GBRT短期负荷预测模型构建

GBRT是加法模型,用决策树作为基学习器,通过迭代拟合残差优化目标函数。对于本文提出的基于GBRT短期负荷预测问题,其预测函数的形式为:

$$f_M(x) = \sum_{m=1}^M \gamma h_m(x) \quad (5)$$

式中: M 为决策树的数量; $f_M(x)$ 为最终的预测函数,由 M 棵决策树共同作用得到; $h_m(x)$ 为第 m 棵决策树,每棵树会对输入 x 给出一个预测结果; γ 为学习率,用于控制每棵树的贡献。

GBRT短期负荷预测模型的核心机制通过迭代集成多棵回归决策树实现预测精度提升。其建模过程可表述为:每棵新增的基学习器以前序模型预测结果与真实值之间的残差为优化目标,其中平方损失函数通过负梯度方向驱动模型参数更新。在第 m 次迭代中,残差项可表示为前 $m-1$ 次迭代模型预测误差的累加值。

$$r_{im} = y_i - f_{m-1}(x_i) \quad (6)$$

式中: r_{im} 为残差,指训练第 m 棵树时,前 $m-1$ 棵树组成的预测函数对样本 x_i 的预测值与该样本真实值 y_i 之间的误差。目标函数采用平方损失,其数学表达式为:

$$L[y, f(x)] = \frac{1}{2} [y - f(x)]^2 \quad (7)$$

将式(7)对 $f(x)$ 求导,可得:

$$\frac{dL}{df(x)} = -[y - f(x)] \quad (8)$$

从式(7)和式(8)可以看出,梯度提升的核心是用平方损失的负梯度来近似残差。

4 实验验证与案例分析

4.1 实验平台和环境配置

本研究的实验验证平台和平台环境配置情况见表1。

表 1 实验平台和环境配置
Table 1 Experimental platform and environment configuration

类别	参数项	详细配置/版本
硬件环境	处理器 CPU	11th Gen Intel(R) Core(TM) i7-1165G7@ 2.80 GHz
软件环境	开发语言/环境	Python 3.10/ PyCharm 2023
API 接口	主 API/备用 API	Coze Bot API(用于工作流编排与插件调用)/ SiliconCloud API(用于异构模型冗余备份)
数据源/插件	主气象插件/备用气象插件	byteDance Weather Service (getNowAndForecastWeather)/墨迹天气、高分天气、和风天气(via Coze Plugins)
模型基座	LLM 逻辑基座	DeepSeek-V3(基于 Coze 平台部署)
预测模型	对比基准模型/本文核心模型	SVM, LSTM, XGBoost, LightGBM/GBRT

4.2 气象 Agent 气象信息获取效能分析

为验证本文气象 Agent 数据处理的时效性,针对某城市未来 10 日天气预报信息进行连续 10 次迭代查询。该气象 Agent 每次查询能获取并解析 240 组结构化气象数据,每组包含 15 项关键气象特征,每次查询平均耗时仅为 70 s。实验结果表明,相较于人工维护 API/FTP 链路,气象 Agent 方案将多源异构数据的接入周期从天/小时级缩短至分钟级,且显著降低了对专业数据分析团队的依赖。同时,相较于易受反爬机制干扰的网络爬虫,气象 Agent 表现出更强的任务鲁棒性与数据获取可靠性。

4.3 预测数据来源与特征选择

1) 预测数据来源。

由 4.2 节可知,通过气象 Agent 可以获取每天逐小时的气象信息,该气象信息含有 15 个关键气象特征,接下来需要研究这些特征之间的关系,以便进行关键特征选择,最后将选择好的特征代入预测模型,才能完成如图 5 所示的从预测数据获取、特征选择、特征工程到预测模型的完整负荷预测流程。但由于缺乏完整的气象和负荷数据,本文将采取离线验证的方式完成上述负荷预测的闭环流程。

本文采用文献[22]提供的公开数据集进行实验,时间跨度自 2010 年 2 月 1 日至 5 月 2 日,共计 91 天。原始数据涵盖时间刻度、环境温度、相对湿度及电力负荷等核心维度,采样频率为 30 min/点。经数据清洗、异常值剔除及缺失值修复等预处理环节后,最终构建了每日 48 个采样点,包含 4368 条记录的标准化时序数据集,其中训练集为预测数据集的前

80%,测试集为预测数据集的后 20%。

2) 负荷特性分析。

对负荷数据进行傅里叶级数分解以确定负荷的周期,分解结果如图 6 所示。

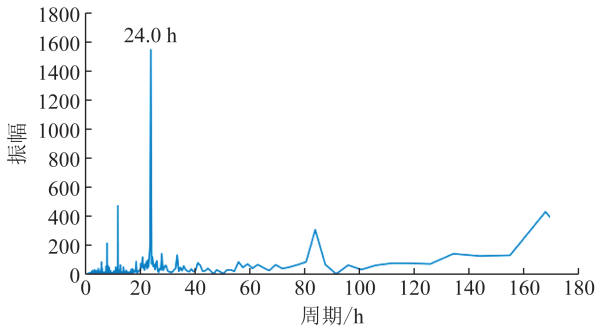


图 6 负荷值的频谱分析结果

Fig. 6 Spectral analysis results of load values

从图 6 可以看出,原始负荷具有明显的日周期特性,周期为 24.0 h 的归一化振幅值最大,因此在建模时设定负荷的周期为 24 h,对应于一天 48 个负荷值。图 7 展示了 2 月 1 日至 2 月 8 日的日电力负荷曲线。

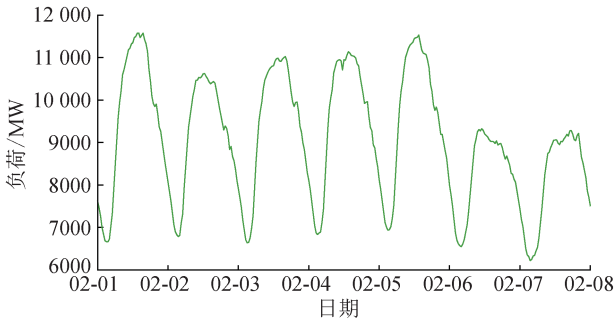


图 7 日负荷曲线

Fig. 7 Daily load curve

从图 7 可以看出,原始负荷具有明显的日周期特性,并且周末(2 月 6 日和 2 月 7 日)的负荷数据和工作日差异明显。因此,预测模型应能同时应对工作日和周末的负荷预测。

3) 特征选择。

选择 2 月 1 日至 5 月 2 日共 13 周的数据作为原始数据集。首先采用相关性分析进行特征选择。相关性分析用于测量 2 个变量之间的线性关联程度,其计算式为:

$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \tag{9}$$

式中: r_{xy} 为变量 x 、 y 之间的 Pearson 相关系数; $\text{cov}(x, y)$ 为变量 x 、 y 之间的协方差; σ_x 、 σ_y 分别为变量 x 、 y 的标准差。对前述原始数据集做相关性分析,结果如图 8 所示。

从图 8 中可以看出,温度和湿度与负荷的 3 个特征相关性均较大,因此这 2 个特征均初步保留。接下

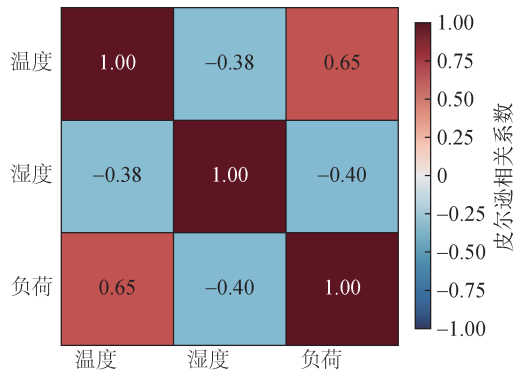


图 8 变量间的 Pearson 相关系数

Fig. 8 Pearson correlation coefficients between variables

来进一步检查温度、湿度各变量之间是否存在多重共线性关系,从而决定需要保留的特征变量。采用方差膨胀因子(variance inflation factor, VIF)来检测两特征与负荷之间是否存在多重共线性,计算式为:

$$\chi_{VIFk} = \frac{1}{1 - R_k^2} \tag{10}$$

式中: χ_{VIFk} 为第 k 个自变量的方差膨胀因子; R_k^2 为将第 k 个自变量作为因变量,与其他所有自变量进行线性回归时的决定系数。当 $VIF < 5$ 时,表示两变量存在低共线性;当 $5 \leq VIF \leq 10$ 时,表示两变量存在中度共线性;当 $VIF > 10$ 时,表示两变量之间存在严重共线性。对前述数据集的温度和湿度进行 VIF 分析,所得的结果如表 2 所示。

表 2 温度、湿度 VIF 分析结果
Table 2 VIF analysis results of temperature and humidity

变量	VIF
温度	1.169 52
湿度	1.169 52

从表 2 中可以看出,温度和湿度的 VIF 均为 1.169 52,因此,这 2 个变量之间存在较低的共线性。综上所述,温度和湿度 2 个特征均选择作为短期负荷预测的特征变量。

4.4 预测性能评价与结果分析

4.4.1 实验设置与参数优化

为提升 GBRT 模型的预测精度,增强其泛化能力,本文采用基于随机搜索结合 K 折交叉验证的自动化调优策略。相比传统的网格搜索穷举法,随机搜索在预设参数空间内进行随机采样,在重要参数的捕捉上往往优于网格搜索,能有效避免陷入局部次优解。同时为确保参数组合的稳健性,调优过程采用了 K 折交叉验证,以确保模型在异构子集上的性能一致性。具体实验设置和参数调优过程如下:

首先,进行样本切分。将原始训练集随机划分

为 3 个互斥子集。其次,对于每一组随机选取的参数配置,对模型进行 3 轮训练与验证。每一轮选取 2 个子集作为训练数据,剩余 1 个子集作为验证数据。最后,进行性能聚合。以 3 轮验证结果的平均负均方误差作为衡量该参数组合优劣的最终评价指标。

在优化参数选择上,针对 GBRT 模型的特点,重点优化了弱学习器数量(`n_estimators`)、最大树深度(`max_depth`)、学习速率(`learning_rate`)、子采样比例(`subsample`)、节点分裂最小样本数(`min_samples_split`)和最大特征选择数(`max_features`)6 个核心超参数。经过超参数寻优,得到最优的模型超参数如表 3 所示。

表 3 模型超参数寻优结果
Table 3 Hyperparameter optimization results of the model

超参数名	搜索范围	步长	选定值
<code>n_estimators</code>	[100, 500]	200	500
<code>max_depth</code>	[3, 8]	1 或 2	8
<code>learning_rate</code>	[0.01, 0.10]	非固定	0.05
<code>subsample</code>	[0.7, 0.9]	0.1	0.7
<code>min_samples_split</code>	[2, 10]	非固定	10
<code>max_features</code>	{'sqrt', None}	类别选择	None

注:None 表示全特征。

4.4.2 预测结果分析

在数据预处理和特征选择的基础上,将预测数据代入 GBRT 模型中进行模型训练和预测结果输出。对于预测结果,选择以下评价指标进行预测结果的评估。

- 1) 平均绝对误差(mean absolute error, MAE) δ_{MAE} , 反映预测的绝对偏差值的大小。
- 2) 均方根误差(root mean square error, RMSE) δ_{RMSE} ,用于惩罚大误差。
- 3) 决定系数(coefficient of determination) R^2 ,反映模型解释方差的比例。
- 4) 平均绝对百分比误差(mean absolute percentage error, MAPE) δ_{MAPE} ,用于衡量预测值与真实值之间的平均相对误差大小。

上述指标的计算式为:

$$\left\{ \begin{aligned} \delta_{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ \delta_{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ R^2 &= \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ \delta_{MAPE} &= \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \end{aligned} \right. \tag{11}$$

式中: n 为样本数量; y_i 为实际值; $\hat{y}_i$ 为预测值; $\bar{y}$ 为实际值的均值。

在相同数据集和相同特征工程的基础上,在型号为LAPTOP-E9RUK688、处理器为11th Gen Intel(R) Core(TM) i7-1165G7@2.80 GHz的笔记本电脑上,采用python3.11编程软件,将本文所提GBRT模型分别与基础模型SVM,集成模型XGBoost、LightGBM以及经典的深度学习模型LSTM进行对比实验,几种模型在测试集上的单步预测评估指标对比如表4所示。

表 4 测试集上单步预测评估指标对比
Table 4 Comparison of evaluation metrics for one-step prediction on the test set

模型	MAE/ MW	RMSE/ MW	R^2	MAPE/%	时间/s
SVM	207.95	289.21	0.926 7	2.69	2.140 0
LSTM	252.75	349.90	0.892 7	3.01	438.000 0
XGBoost	70.03	100.12	0.991 2	0.84	0.219 0
LightGBM	73.24	102.88	0.990 7	0.89	0.180 4
GBRT	66.37	92.35	0.992 5	0.81	11.075 7

注:表中的时间包括模型参数的寻优时间和程序运行时间,下同。

从表4中可以看出,本文所提GBRT预测模型在测试集上的整体表现均显著优于基础模型SVM和典型的深度学习模型LSTM。同时在MAE、RMSE、 R^2 、MAPE几个指标上优于XGBoost、LightGBM两个模型的表现,仅在模型的处理时间上高于后两者。主要原因有:1)相比于经典的机器学习模型SVM,GBRT通过迭代拟合残差的机制,能够比SVM等全局优化模型更精准地捕捉电力负荷中细微的气象波动。2)相比XGBoost和LightGBM模型,GBRT模型采用原生梯度提升框架,参数更直观,适合中小规模数据集。本文的负荷预测样本数据属于中小规模数据集,而GBRT在中小规模数据集上表现更为优秀。3)GBRT不引入过多的复杂正则化或近似计算,在特征维度较低时表现更稳健。本文选取的短期负荷预测特征以及特征的组合仅为6个,属于小规模特征维度的数据集。同时本文结合气象数据存在多重共线性的特点,对原始数据集进行了特征选择工作,这有效降低了无关气象噪声对MAPE指标的影响。4)GBRT模型作为标准的梯度提升算法,采用串行生成树,在处理较大数据集时计算速度慢,但本文的短期负荷预测的时间粒度为0.5 h,模型的处理时间在可接受范围之内。综合来看,从样本数量、评价指标、内存要求以及模型运行时间4个维度看,GBRT实现了计算成本与实时性的平衡,更适合用于本文的短期负荷预测场景。

为便于直观观察模型的预测效果和泛化能力,选择典型的SVM、GBRT和LSTM这3种模型,3种模型在测试集上最后一周的单步预测效果对比如图9所示。

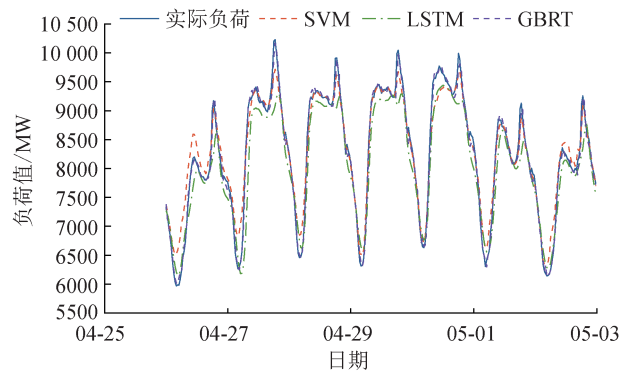


图 9 测试集最后一周的单步预测效果对比
Fig. 9 Single-step forecasting comparison for the last week on the test set

4.4.3 预测模型的多步预测应用

上述短期负荷预测模型能进行单步预测,即预测序列中紧接着当前时刻的下一个预测时刻点的负荷值。而实际使用时,经常要进行多步预测以支撑电力调度制定日前发电计划、预留备用容量及评估系统供需平衡,以确保电网安全稳定经济运行。目前有3种主流的预测策略能实现多步预测:第一种是直接预测策略。这种策略为每个未来时位独立构建映射模型(如48步需训练48个模型),这种方式虽然计算成本高但能有效避免误差累积的影响。第二种是递归预测策略。这种策略仅训练单步模型并采用预测值迭代回传的方式逐步推演,模型简洁但在长序列中易产生累积误差。第三种是多输出预测策略。这种策略通过单一模型构建高维输出向量,一次性映射未来全时段数值,兼顾了预测效率与时序相关性。

由于电力负荷在不同时位的气象敏感度与时间关联逻辑存在异质性,采用直接预测策略通常能捕获更精细的非线性映射关系,从而获得更高的预测精度。本文采用直接预测策略,对前述的SVM、LSTM、XGBoost、LightGBM、GBRT在多步预测上的性能指标进行比较。为确保实验的严谨性与可比性,所有模型在相同预测步长下均采用统一的特征集与特征工程流程,重点针对6步(3 h)、24步(12 h)及48步(24 h)3个典型预测维度的实验结果进行了定量对比,具体各项性能指标详见表5。

同样,为直观评估不同算法的泛化性能,选取SVM、GBRT及LSTM三类代表性模型,对其在测试集最后6个采样周期内的拟合效果进行深度分析。3种模型在测试集上最后6天的48步预测效果对比如图10所示。

表 5 测试集上多步预测评估指标对比

Table 5 Comparison of evaluation metrics for multi-step prediction on the test set

步长	模型	MAE/MW	RMSE/MW	R^2	MAPE/%	运行时间/s
6	SVM	245.90	349.00	0.902 9	2.99	3.86
	LSTM	333.23	441.29	0.845 2	4.05	9.56
	XGBoost	190.78	265.19	0.943 9	2.32	8.61
	LightGBM	195.06	267.33	0.943 0	2.37	45.80
	GBRT	206.68	286.82	0.934 4	2.49	57.18
24	SVM	330.55	463.82	0.829 0	4.04	5.73
	LSTM	376.80	487.86	0.810 8	4.70	9.15
	XGBoost	307.32	449.95	0.839 0	3.83	9.07
	LightGBM	308.29	439.34	0.846 5	3.81	29.85
	GBRT	305.19	446.44	0.841 5	3.77	66.22
48	SVM	373.26	526.94	0.778 3	4.53	4.23
	LSTM	280.74	381.37	0.883 9	3.41	12.22
	XGBoost	387.75	529.43	0.776 2	4.70	9.22
	LightGBM	346.64	466.65	0.826 2	4.22	40.78
	GBRT	358.62	504.17	0.797 1	4.31	68.07

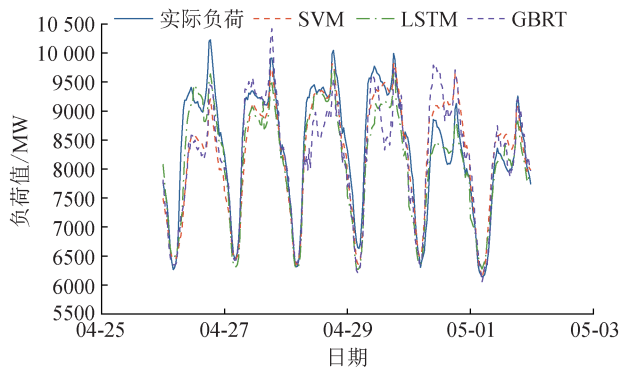


图 10 测试集上最后 6 天的 48 步预测效果对比

Fig. 10 48-step forecasting comparison for the last 6 days on the test set

从表 5、图 9、图 10 中可以看出,随着预测步长的增加,各模型的评价指标均出现不同程度的恶化,证实了预测精度随时间跨度拉长而衰减的普遍规律。在 48 步的日前预测中,LSTM 模型凭借其长序列时序特征捕获能力表现出最佳的鲁棒性。需要指出的是,本文主要聚焦于负荷短时波动规律与实时预测性能验证,暂未引入跨季节、跨年度数据以分析季节性影响。而在单步、3 步、6 步等短步长预测场景下,GBRT 模型的各项指标显著优于其他模型,这进一步印证了本文所提 GBRT 模型能更精准地刻画该地区电力负荷的细微波动特性,在超短期预测任务中具有更高的实用价值。

4.5 原型系统部署与可靠性设计

本文所提短期负荷预测的原型系统设计思路和

流程如附录 B 所示。为保证该原型系统能够真正适用于电网企业,需具有工程化实施的具体保障措施。具体采用边缘感知-云端推理-本地控制混合部署架构,该架构设计既利用了 LLM 的云端算力,又保证了负荷预测核心逻辑在本地的闭环运行。在该混合云部署架构中,将负荷数据采集与 GBRT 预测核心部署在电网内部服务器,而 LLM 语义推理与非关键气象插件通过 Coze/API 部署在云端。这样既保证核心负荷数据不出网,又符合安全规范,同时降低全链路时延。同时针对 API 与插件设置了三重可靠性保障机制:第一层异常捕获机制,针对 API 超时或插件报错,设置自动重试机制;第二层本地语义备选,当外部 LLM 不可用时,系统降级使用规则库处理气象信息;第三层数据质量监控机制,对插件返回的气象数值进行物理合理性检查,例如当 2 h 内气温跳变超过 10℃则视为无效,启用本地历史相似气象数据替代。

此外,本原型系统在开发阶段选用 Coze 平台及其插件,旨在快速验证 LLM 赋能的气象 Agent 在气象信息获取、多源数据融合等方面的潜在价值。在实际工程落地中,该架构可无缝迁移至企业级私有 Agent 平台,可有效规避 API 配额限制与数据隐私风险。

5 结 论

当前短期负荷预测系统存在模型复杂度、精度与效率兼顾不足,气象信息获取难度高、实时性差、

现场使用精度低等诸多挑战。本文对 LLM 赋能的气象 Agent 进行研究,并将其应用于短期负荷预测中,简化了气象信息获取流程,提高了预测精度,所得结论如下:

1)通过 Coze 开发平台开发了气象 Agent,在 Agent 中构建了气象信息获取 workflow,实现了 LLM 赋能的气象数据获取功能。同时在工作流中配置异常重试机制以及备用数据源,确保获取到高质量的温度、湿度等天气信息返回值,大大减轻了负荷预测工作量。该策略为短期负荷预测系统加入舆情监测、极端天气/事件监测模块提供了新的研究思路,现有负荷预测系统可根据该思路进一步完善。

2)在对负荷特性进行分析的基础上,构建了 GBRT 短期负荷预测模型。实现了负荷特征提取、特征变量选择、预测模型特征生成等功能,兼顾模型复杂度和效率。案例运行结果表明,该模型的各项指标满足单步和多步预测要求,并且运行时间短、预测精度高。

3)本文的研究成果表明,利用 LLM 作为逻辑中枢而非仅作为计算工具,是实现人工智能技术在电力基层单位低成本落地的关键。这不仅对后续构建更高级别的电力行业大模型 Agent 具有重要参考价值,也为能源互联网的智数融合转型提供了新的视角。

应该指出的是,当前大语言模型赋能气象智能体的研究仍处于探索阶段,尚未形成成熟统一的技术框架。同时受限于气象预测专用数据集稀缺、高质量标注数据不足等现实约束,LLM 直接参与端到端特征学习与预测推理仍存在较大技术瓶颈。本文采用 LLM 协同 GBRT 模型的轻量化落地路径,在保证预测精度与工程可行性的前提下,通过 LLM 实现多源气象数据整合、结构化特征规整与任务流程调度,该方案规避了纯 LLM 预测不稳定、泛化性不足的问题,为 LLM 参与特征工程和预测推理提供了坚实的研究基础,具备阶段性研究价值。未来将进一步打通 EMS 实测数据流,并深度集成舆情监测与极端天气预警 Agent,构建具备全时域感知能力的 Agent 集群。同时探索将 LLM 应用于自动化特征学习、特征选择、特征工程、超参数寻优以及预测推理之中,进一步降低人工运维成本,在复杂的实际运行环境中全面增强模型的鲁棒性、泛化能力及现场投运率。

利益冲突声明(Conflict of Interests):

所有作者声明不存在利益冲突。

作者贡献声明(Authors' Contributions):

蔡子龙提出论文总体研究思路、研究方案,对研究方案可行性进行调查分析,组织具体研究的实施;林红娅负责文献调研与整理、模型选择和程序编写;沈赋起草、修订、审核论文;毛肖负责实验数据收集,采集、清洗与分析数据;王健负责确定机器学习预测模型;杨宇林负责论文图片制作和修订。所有作者均阅读并同意了论文终稿内容。

6 参考文献

- [1] 韩富佳, 王晓辉, 乔骥, 等. 基于人工智能技术的新型电力系统负荷预测研究综述[J]. 中国电机工程学报, 2023, 43(22): 8569-8591.
Han Fujia, Wang Xiaohui, Qiao Ji, et al. Review on artificial intelligence based load forecasting research for the new-type power system[J]. Proceedings of the CSEE, 2023, 43(22): 8569-8591.
- [2] 张艺涵, 杨萌, 刘军会, 等. 基于时间图卷积网络的短期电力负荷时空预测方法[J]. 电网与清洁能源, 2024, 40(12): 27-35.
Zhang Yihan, Yang Meng, Liu Junhui, et al. A short-term spatial-temporal power load forecasting method based on temporal graph convolution network[J]. Power System and Clean Energy, 2024, 40(12): 27-35.
- [3] 冉启武, 张宇航. 基于模态分解及 GRU-XGBoost 短期电力负荷预测[J]. 电网与清洁能源, 2024, 40(4): 18-27, 34.
Ran Qiwu, Zhang Yuhang. Short-term power load forecasting based on modal decomposition and GRU-XGBoost[J]. Power System and Clean Energy, 2024, 40(4): 18-27, 34.
- [4] 张家安, 李凤贤, 王铁成, 等. 一种采用记忆神经网络和曲线形状修正的负荷预测方法[J]. 电力工程技术, 2024, 43(1): 117-126.
Zhang Jia'an, Li Fengxian, Wang Tiecheng, et al. A load prediction method using memory neural network and curve shape correction[J]. Electric Power Engineering Technology, 2024, 43(1): 117-126.
- [5] 麦鸿坤, 肖坚红, 吴熙辰, 等. 基于 R 语言的负荷预测 ARIMA 模型并行化研究[J]. 电网技术, 2015, 39(11): 3216-3220.
Mai Hongkun, Xiao Jianhong, Wu Xichen, et al. Research on ARIMA model parallelization in load prediction based on R language[J]. Power System Technology, 2015, 39(11): 3216-3220.
- [6] Jiang H G, Zhang Y C, Muljadi E, et al. A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization[J]. IEEE Transactions on Smart Grid, 2018, 9(4): 3341-3350.
- [7] 郑翔, 杜奇伟, 阮黎翔, 等. 基于 WOA-SVM 的智能变电站二次系统故障参数映射模型[J]. 浙江电力, 2024, 43(1): 36-44.
Zheng Xiang, Du Qiwei, Ruan Lixiang, et al. A WOA-based fault parameter mapping model for the secondary systems of intelligent substations[J]. Zhejiang Electric Power, 2024, 43(1): 36-44.
- [8] 冯昌森, 钱懿飞, 邵亮, 等. 考虑特征缺失的个性化居民短期负荷预测[J]. 电力系统自动化, 2025, 49(16): 75-84.
Feng Changsen, Qian Yifei, Shao Liang, et al. Personalized short-term residential load forecasting considering missing feature[J]. Automation of Electric Power Systems, 2025, 49(16): 75-84.
- [9] 姜建, 刘海琼, 李衡, 等. 基于 XGBoost 的配电网线路峰值负荷

- 预测方法[J]. 电力系统保护与控制, 2021, 49(16): 119-127.
- Jiang Jian, Liu Haiqiong, Li Heng, et al. Peak load forecasting method of distribution network lines based on XGBoost[J]. Power System Protection and Control, 2021, 49(16): 119-127.
- [10] 韩璟琳, 冯喜春, 胡平, 等. 基于 Hyperopt-LightGBM 的直流配电网短期负荷抗噪声预测[J]. 高电压技术, 2024, 50(11): 4902-4911.
- Han Jinglin, Feng Xichun, Hu Ping, et al. Hyperopt-LightGBM-based noise-resistant forecasting of short-term loads in the DC distribution network [J]. High Voltage Engineering, 2024, 50(11): 4902-4911.
- [11] 王炜, 冯斌, 黄刚, 等. 基于自注意力编码器和深度神经网络的短期净负荷预测[J]. 中国电机工程学报, 2023, 43(23): 9072-9083.
- Wang Wei, Feng Bin, Huang Gang, et al. Short-term net load forecasting based on self-attention encoder and deep neural network [J]. Proceedings of the CSEE, 2023, 43(23): 9072-9083.
- [12] Shi H, Xu M H, Li R. Deep learning for household load forecasting: a novel pooling deep RNN[J]. IEEE Transactions on Smart Grid, 2018, 9(5): 5271-5280.
- [13] 任建吉, 位慧慧, 邹卓霖, 等. 基于 CNN-BiLSTM-Attention 的超短期电力负荷预测[J]. 电力系统保护与控制, 2022, 50(8): 108-116.
- Ren Jianji, Wei Huihui, Zou Zhuolin, et al. Ultra-short-term power load forecasting based on CNN-BiLSTM-Attention [J]. Power System Protection and Control, 2022, 50(8): 108-116.
- [14] 朱凌建, 荀子涵, 王裕鑫, 等. 基于 CNN-BiLSTM 的短期电力负荷预测[J]. 电网技术, 2021, 45(11): 4532-4539.
- Zhu Lingjian, Xun Zihan, Wang Yuxin, et al. Short-term power load forecasting based on CNN-BiLSTM [J]. Power System Technology, 2021, 45(11): 4532-4539.
- [15] 凌佳凯, 章逸舟, 胡金峰, 等. 基于 CNN-LSTM-Attention 的配电网拓扑实时辨识方法[J]. 浙江电力, 2024, 43(3): 84-94.
- Ling Jiakai, Zhang Yizhou, Hu Jinfeng, et al. A real-time topology identification method of distribution networks based on CNN-LSTM-Attention[J]. Zhejiang Electric Power, 2024, 43(3): 84-94.
- [16] 范竞敏, 贺广林, 钟铭伟, 等. 考虑日期模式等效和多级并行卷积模块的短期负荷预测[J]. 电网技术, 2025, 49(6): 2554-2562.
- Fan Jingmin, He Guanglin, Zhong Mingwei, et al. Short-term load forecasting for date mode equivalent and multi-level parallel convolution modules is considered[J]. Power System Technology, 2025, 49(6): 2554-2562.
- [17] Kong W C, Dong Z Y, Jia Y W, et al. Short-term residential load forecasting based on LSTM recurrent neural network [J]. IEEE Transactions on Smart Grid, 2019, 10(1): 841-851.
- [18] Li J, Deng D Y, Zhao J B, et al. A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network[J]. IEEE Transactions on Industrial Informatics, 2021, 17(4): 2443-2452.
- [19] Lv L L, Wu Z Y, Zhang J H, et al. A VMD and LSTM based hybrid model of load forecasting for power grid security[J]. IEEE Transactions on Industrial Informatics, 2022, 18(9): 6474-6482.
- [20] 张淑清, 李君, 姜安琦, 等. 基于 FPA-VMD 和 BiLSTM 神经网络的新型两阶段短期电力负荷预测[J]. 电网技术, 2022, 46(8): 3269-3279.
- Zhang Shuqing, Li Jun, Jiang Anqi, et al. A novel two-stage model based on FPA-VMD and BiLSTM neural network for short-term power load forecasting[J]. Power System Technology, 2022, 46(8): 3269-3279.
- [21] 李加文, 孙永辉, 王森, 等. 计及异常场景数据缺失的负荷超短期预测[J]. 电力系统自动化, 2025, 49(15): 133-143.
- Li Jiawen, Sun Yonghui, Wang Sen, et al. Ultra-short-term load forecasting considering missing data in anomalous scenarios [J]. Automation of Electric Power Systems, 2025, 49(15): 133-143.
- [22] 蔡子龙, 李嘉棋, 沈赋, 等. 基于 SDAE-EEMD 降噪分解与改进 Informer-BiLSTM 模型的电力短期负荷预测方法[J]. 电网技术, 2025, 49(12): 5009-5018.
- Cai Zilong, Li Jiaqi, Shen Fu, et al. A short-term power load forecasting method based on SDAE-EEMD denoising decomposition and an improved informer-BiLSTM model [J]. Power System Technology, 2025, 49(12): 5009-5018.
- [23] Li Y, Wang R N, Yang Z. Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting [J]. IEEE Transactions on Sustainable Energy, 2022, 13(1): 159-169.
- [24] 牛泽原, 李嘉媚, 艾芊. 大语言模型在电力系统中的应用初探[J]. 电网技术, 2025, 49(4): 1327-1336.
- Niu Zeyuan, Li Jiamei, Ai Qian. Preliminary exploration of the application of large language models in power systems [J]. Power System Technology, 2025, 49(4): 1327-1336.
- [25] 丁俐夫, 陈颖, 肖谭南, 等. 基于大语言模型的新型电力系统生成式智能应用模式初探[J]. 电力系统自动化, 2024, 48(19): 1-13.
- Ding Lifu, Chen Ying, Xiao Tannan, et al. Exploration of generative intelligent application mode for new power systems based on large language models [J]. Automation of Electric Power Systems, 2024, 48(19): 1-13.
- [26] 曹祎, 张莉, 郭静, 等. 基于大语言模型的低碳电力市场发展应用前景[J]. 智慧电力, 2024, 52(2): 8-16.
- Cao Yi, Zhang Li, Guo Jing, et al. Prospects for development of low-carbon electricity markets based on large language models [J]. Smart Power, 2024, 52(2): 8-16.
- [27] 周远翔, 陈健宁, 赵亮, 等. 电力设备数智化发展历程与大语言模型智能体应用展望[J]. 高电压技术, 2025, 51(8): 4089-4109.
- Zhou Yuanxiang, Chen Jianning, Zhao Liang, et al. Evolution of digitization and intelligence of power equipment and prospects for application of large language model based agent [J]. High Voltage Engineering, 2025, 51(8): 4089-4109.
- [28] 李刚, 方鸿, 刘云鹏, 等. 新型电力系统中的大模型驱动技术: 现状、机遇与挑战[J]. 高电压技术, 2024, 50(7): 2864-2878.
- Li Gang, Fang Hong, Liu Yunpeng, et al. Large-model drive technology in new power system: status, challenges and prospects [J]. High Voltage Engineering, 2024, 50(7): 2864-2878.
- [29] 杨明, 李梦林, 王勃, 等. 面向新型电力系统的数值天气预报技术及应用综述[J]. 电力系统自动化, 2025, 49(17): 1-20.
- Yang Ming, Li Menglin, Wang Bo, et al. Review on technologies and applications of numerical weather prediction for new power system [J]. Automation of Electric Power Systems, 2025, 49(17): 1-20.
- [30] 李滨, 陆明珍. 考虑实时气象耦合作用的地区电网短期负荷预测建模[J]. 电力系统自动化, 2020, 44(17): 60-68.
- Li Bin, Lu Mingzhen. Short-term load forecasting modeling of

regional power grid considering real-time meteorological coupling effect[J]. Automation of Electric Power Systems, 2020, 44(17): 60-68.

[31] 陈仕启, 吴燕, 杨德昌, 等. 基于负荷二次分解与特征处理的电力系统短期负荷预测[J]. 高电压技术, 2025, 51(5): 2571-2581.

Chen Shiqi, Wu Yan, Yang Dechang, et al. Short-term load forecasting of power system based on secondary load decomposition and feature processing[J]. High Voltage Engineering, 2025, 51(5): 2571-2581.

[32] 李知艺, 韩旭涛, 许悦, 等. 面向新型配电系统的多主体协同运行及弹性提升探讨[J/OL]. 电力系统自动化, 2025-08-18. <https://link.cnki.net/urlid/32.1180.TP.20250818.0945.006>.

Li Zhiyi, Han Xutao, Xu Yue, et al. Discussion on multi-agent cooperative operation and resilience improvement for new-type power distribution systems[J/OL]. Automation of Electric Power Systems, 2025-08-18. <https://link.cnki.net/urlid/32.1180.TP.20250818.0945.006>.

[33] 王铁楠, 卢静, 陈星彤, 等. 考虑多重因素叠加情景的能源清洁低碳转型风险评估[J]. 中国电力, 2024, 57(3): 183-189.

Wang Yinan, Lu Jing, Chen Xingtong, et al. Risk assessment of clean and low-carbon energy transformation considering superimposed scenario of multiple factors[J]. Electric Power, 2024, 57(3): 183-189.

[34] 李丹, 孙光帆, 缪书唯, 等. 基于多维时序信息融合的短期电力负荷预测方法[J]. 中国电机工程学报, 2023, 43(S1): 94-106.

Li Dan, Sun Guangfan, Miao Shuwei, et al. A short-term power load forecasting method based on multidimensional temporal information fusion[J]. Proceedings of the CSEE, 2023, 43(S1): 94-106.

[35] 臧海洋, 陈玉伟, 程礼临, 等. 基于多尺度分量特征学习的用户级超短期负荷预测[J]. 电网技术, 2024, 48(6): 2584-2592.

Zang Haixiang, Chen Yuwei, Cheng Lilin, et al. User level ultra-short-term load forecasting based on multi-scale component feature learning[J]. Power System Technology, 2024, 48(6): 2584-2592.

[36] 张怀天, 贾东梨, 王帅, 等. 基于Transformer-集成学习的配电网短期负荷预测方法[J/OL]. 中国电力, 2026-01-29. <https://link.cnki.net/urlid/11.3265.TM.20260128.1616.004>.

Zhang Huaitian, Jia Dongli, Wang Shuai, et al. A transformer-ensemble learning-based method for short-term load forecasting in distribution[J/OL]. Electric Power, 2026-01-29. <https://link.cnki.net/urlid/11.3265.TM.20260128.1616.004>.

[37] 陈逸枫, 张大海, 于浩, 等. 基于Seq2seq模型的多特征短期母线负荷预测[J]. 电力系统及其自动化学报, 2023, 35(1): 1-6, 35.

Chen Yicong, Zhang Dahai, Yu Hao, et al. Short-term bus load forecasting of multi-feature based on Seq2seq model[J]. Proceedings of the CSU-EPSA, 2023, 35(1): 1-6, 35.

[38] 严通煜, 杨迪珊, 项康利, 等. 基于时间分解技术的中远期逐时负荷预测模型[J]. 电力系统保护与控制, 2019, 47(6): 110-117.

Yan Tongyu, Yang Dishan, Xiang Kangli, et al. Mid-long term hourly load forecasting model based on time decomposition[J]. Power System Protection and Control, 2019, 47(6): 110-117.

[39] 卢泓旭, 王仁浚, 高红均, 等. 考虑温度与分时电价对工业负荷基线影响的机器学习预测方法[J/OL]. 南方电网技术, 2025-09-16. <https://link.cnki.net/urlid/44.1643.TK.20250916.0942.002>.

Lu Hongxu, Wang Renjun, Gao Hongjun, et al. Machine learning forecasting method considering the influence of temperature and time-of-use electricity price on industrial load baseline[J/OL]. Southern Power System Technology, 2025-09-16. <https://link.cnki.net/urlid/44.1643.TK.20250916.0942.002>.

收稿日期: 2026-01-19 修回日期: 2026-04-15

作者简介:



蔡子龙

蔡子龙(1976),男,博士,副教授,硕士生导师,主要研究方向为电力系统机器学习、V2G、车网互动、负荷预测,E-mail:20040142@kust.edu.cn;

林红娅(1998),女,硕士研究生,主要研究方向为新型电力系统电力负荷预测技术,E-mail:834509371@qq.com;

沈赋(1988),男,博士,副教授,硕士生导师,通信作者,主要研究方向为新型电力系统建模、电网运行与调度,E-mail:shenfu@kust.edu.cn;

毛肖(1981),男,硕士,高级实验师,主要研究方向为电力系统分析、电力系统保护,E-mail:49005540@qq.com;

王健(1993),男,博士,硕士生导师,主要研究方向为电力系统可靠性评估与风险预测,E-mail:wangjian0924@kust.edu.cn;

杨宇林(2001),男,硕士研究生,主要研究方向为电动汽车自适应充电调度,E-mail:yangyulin0929@163.com。

(编辑 景贺峰)

附录 A

1)气象 Agent 中的 LLM 系统提示词。系统提示词通常在对话开始前设定,用于规定模型的角色、行为准则、知识边界及输出格式。它的作用是长期的、全局性的。系统提示词示例见附表 A1。

2)气象 Agent 中的 LLM 用户提示词。用户提示词是对话过程中的具体指令,用于向模型传递当前需要处理的数据、具体请求或即时问题。它的作用是具体的、瞬时的。用户提示词示例见附表 A2。

表 A1 系统提示词示例

Table A1 System prompt example

角色定位:你是一名精通电力气象学的资深数据工程师,专门负责将异构气象数据转化为电力负荷预测模型可用的“认知特征”。
处理逻辑:
1. 数据清洗:统一单位(温度:℃;湿度:%;风速:m/s)。
2. 风向矢量化:将方位描述(如“东北风”)转化为0°~360°的角度数值。
3. 语义增强:根据温度、湿度与风力计算“体感舒适度指数”,并评估当前天气对居民空调负荷或工业用电的潜在影响等级。
4. 异常处理:若原始数据包含缺失值,请根据上下文逻辑进行语义补偿(如标注为均值或推断值),而非返回空值。
输出约束:严禁返回任何解释性文字,仅输出标准的JSON对象。

表 A2 用户提示词示例
Table A2 User prompt example

请处理以下由气象 Agent 工作流获取的原始数据:

【原始数据输入】

{{Workflow_Output}}

【任务要求】

请根据上述数据,输出如下格式的 JSON 报文:

JSON

{

"timestamp": "YYYY-MM-DD HH:mm",

"temperature": "数值",

"humidity": "数值",

"wind_speed": "数值",

"wind_direction_degree": "0~360 数值"

"comfort_index": "0~100 (基于温湿度的体感综合得分)"

"load_impact_factor": "0~1.0 (气象对电力负荷的影响权重预估)",

"is_extreme_weather": "boolean",

"data_quality_note": "数据完整性说明"

}

附录 B

本文设计的短期负荷预测原型系统的界面如图 B1 所示。在图 B1 中,预测模型模块主要用于训练模型;短期预测模块用于对实时负荷进行预测;气象数据模块通过 API 从 Coze 平台的工作流中获取气象数据;历史负荷模块中历史负荷数据从 EMS 中获取;舆情监测和气象预警模块属于舆情监测 Agent 和极端天气预警两模块的具体实现环节。本文主要聚焦于气象 Agent 的构建,并将其集成到 GBRT 短期负荷预测模型,舆情监测 Agent 和极端气象预警模块还有待在后续的开发中逐步实现。



图 B1 LLM 赋能的短期电力负荷预测平台界面

Fig. B1 LLM-empowered short-term load forecasting platform interface

B.1 气象工作流的调用

气象信息既可以通过 Agent 中的预览与调试窗口获取,也可以通过 Agent 中的气象工作流获取,本文使用后者,所开发的短期负荷预测程序通过 API 调

用 Agent 中的气象工作流实现气象信息的获取,基本流程为:首先,进行环境准备与认证配置。在 Coze 平台创建气象数据获取工作流后,获取唯一的工作流 ID 和 API 访问密钥。其次,配置认证信息,设置 Authorization 请求头和指定 API 端点 URL。接着根据气象工作流的输入要求,构建包含查询城市、查询日期范围的 JSON 请求参数,并确保参数格式与工作流定义完全匹配,包括必填字段和可选参数的正确处理。然后,执行 API 请求。使用 python 的 requests 库发起 HTTP POST 请求,设置合适的重试和兜底策略。最后,进行响应解析与气象信息的提取。

B.2 气象工作流的调用示例

点击图 B1 中的气象数据按钮,进入图 B2 所示的天气信息获取界面。在图 B2 中,输入城市名,点击查询按钮,通过 API 从 Coze 平台的工作流中获取该地区的气象预测的时间、温度和湿度等信息,并以 xlsx 格式持久化保存于硬盘中。以昆明市呈贡区的天气信息获取为例,在对话框里输入呈贡区,点击查询按钮,结果如图 B3 所示。

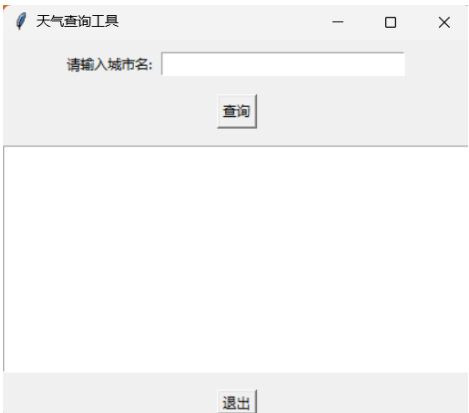


图 B2 气象信息预测界面

Fig. B2 Meteorological information forecasting interface



图 B3 气象信息获取结果示意

Fig. B3 Meteorological information acquisition results

从图 B3 中可以看出,由于获取气象数据的时间粒度为 1 h,而负荷采样序列时间粒度为 0.5 h,二者在时域上不匹配。为此,基于气象要素的时间连续性规律,引入线性插值算子进行数据增广,通过对相邻整点观测值进行加权内插,构造出中间时刻的准观测值。该处理有效模拟了气象物理场的平滑演变过程,实现了异构数据源在时空维度上的精准对齐,从而消除特征滞后效应,显著提升了 GBRT 模型的特征耦合精度。

B.3 短期负荷预测原型系统应用

将所得的模型进行保存后,接下来就可以使用所开发的 GBRT 模型进行相应地区的短期负荷预测工作。首先在图 B1 中点击历史负荷按钮,从 EMS 中获取历史负荷;其次在图 B1 中点击气象数据按钮,在图 B3 中设置查询城市,调用气象 workflow 获取查询城市的气象信息,将 EMS 数据和气象信息数据按时间戳对齐的方式进行合并,以获得预测所需要的数据集。点击图 B4 中的外推预测按钮,进入图 B5 所示的短期负荷预测界面。在图 B5 所示的界面中,依次选

择预测数据集、历史数据集、预测模型、结果保存路径和图表保存路径,其中历史数据用于计算模型中所需要的特征数据。预测结果展示如图 B6 所示。

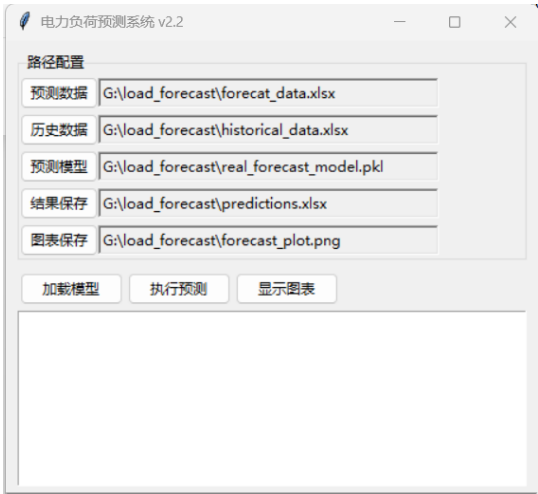


图 B5 短期负荷预测界面

Fig. B5 Short-term load forecasting interface

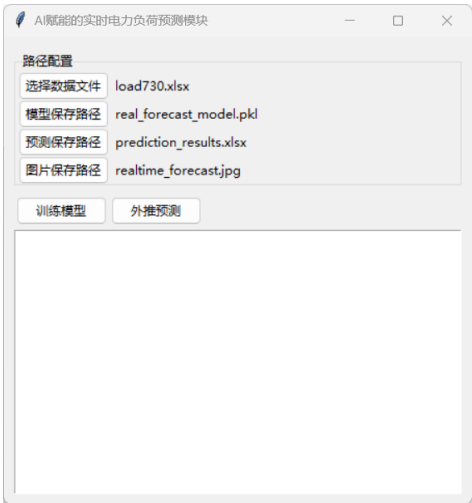


图 B4 短期负荷预测模块

Fig. B4 Short-term load forecasting module

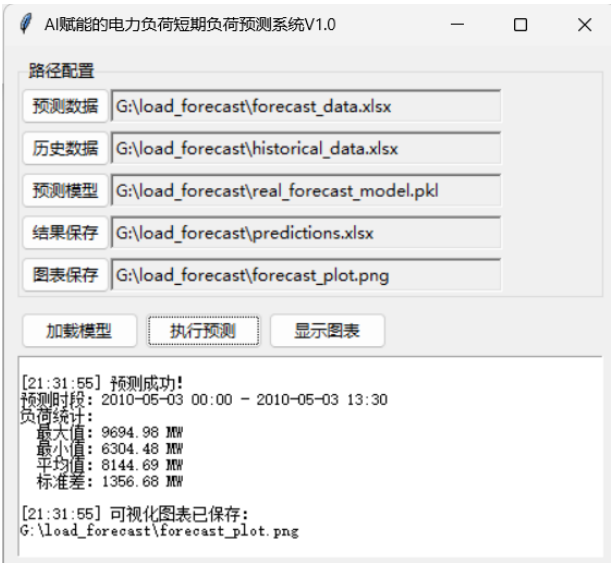


图 B6 短期负荷预测结果示意

Fig. B6 Short-term load forecasting results