



新型电力系统中一种基于LSTM和CNN的倾斜样本预测算法

张琦, 李志浩, 范叶平, 马剑波
(安徽继远软件有限公司, 安徽 合肥 230093)

摘要: 基于电力数字空间技术体系, 提出了一种基于特征关联性分析与改进卷积神经网络 (convolutional neural network, CNN)-长短期记忆 (long short-term memory, LSTM) 网络的预测算法。首先, 在注意力层, 对用户关联性进行分析, 将与电力负荷用户强相关的影响因素进行加权求平均, 调整这些类别在用户电力负荷预测中的权重大小, 加强少且重要的电力样本特征对电力负荷算法的影响并避免冗余特征的干扰, 实现特征选择; 其次, 在CNN层, 通过构建2层神经网络, 提取样本的空间特征, 并将该空间特征输入LSTM网络中; 最后, 通过LSTM网络层提取序列的时间特征, 输出预测结果。实验基于北方某城市供应电力资源的6个重要用户的保密数据进行预测, 实验结果表明, 基于关联性分析与改进CNN-LSTM模型的MAPE在5类用户测试的误差均低于LSTM、随机森林和BP神经网络的误差。

关键词: 样本倾斜; LSTM; 电力负荷预测; CNN; 关联性分析

中图分类号: TP183

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025182

An unbalanced sample prediction algorithm based on LSTM and CNN in the new power systems

ZHANG Qi, LI Zhihao, FAN Yeping, MA Jianbo
Anhui Jiyuan Software Co., Ltd., Hefei 230093, China

Abstract: A prediction algorithm based on feature correlation analysis and improved CNN (convolutional neural network)-LSTM (long short-term memory) network was proposed based on the power digital space technology system. Firstly, in the attention layer, the user correlation was analyzed, and the influencing factors strongly related to power load users were weighted and averaged. The weight of these categories in user power load prediction was adjusted to strengthen the influence of few and important power sample features on power load calculation and avoid the interference of redundant features, achieving feature selection. Secondly, in the CNN layer, by constructing a two-layer neural network, the spatial features of the samples are extracted and inputted into LSTM. Finally, LSTM extract the temporal features of the sequence through the LSTM layer and output the prediction results. The experiment was based

on confidential data from six important users supplying power resources in a northern city for prediction. The experimental results show that the MAPE based on correlation analysis and improved CNN-LSTM model was lower than the error of LSTM, random forest, and BP neural network in five types of user tests.

Key words: unbalanced sample, LSTM, power load forecasting, CNN, correlation analysis

0 引言

2021年3月15日,习近平总书记提出“构建以新能源为主体的新型电力系统”的要求,而新型电力系统的构建离不开数字化技术的支撑。“双碳”目标引导下,中国电力市场正在经历一场前所未有的时代变革,以“安全高效、清洁低碳、柔性灵活、智慧融合”为核心的新型电力系统正成为能源转型的重要抓手,将给电力及新能源行业带来更多机遇。随着数字化新技术的深度应用,国网信息通信产业集团有限公司首创性地提出了电力数字空间理论架构与技术体系,其中以人工智能、数字孪生为主体的电力数字空间智慧中枢是支撑能源生态的重要桥梁。

在新能源源网荷储一体化建设中的电力负荷预测场景中,样本倾斜广泛发生在电力负荷预测中。样本倾斜是指预测样本数据集类之间呈现出不均等的分布。由于影响电力负荷预测的因素具有很强的不确定性,传统电力负荷预测模型预测精度难以得到保证,因此,需要对这些影响因素进行深入分析和研究。

本文围绕新型电力系统源网荷储一体化建设过程,在业务调研中发现,在新能源为主体的新型电力系统建设过程中,很多场景数据预测精度不高是因为样本数据少,无法有效地获取数据短期的变化特征。调研还发现,目前较常见的影响电力负荷的因素主要包括自然环境因素(气温、湿度、风速及降水量)、社会因素(城市人口数量、生活方式、电力供给)、电力供给因素(燃料供应量、发电厂容量、输电线路)。在实际进行电力负荷预测时,由天气因素及社会因素获取

的样本类别较常见,而部分重要的样本类别如燃料供应量、发电容量、电力供给线路等由于电力数据保密等因素,很难完全获取到,这就造成电力负荷预测的数据量存在倾斜,进而导致分布式新能源发电功率预测、用电负荷预测偏差大,难以支撑有源配电网的调控。

为了解决以上难题,本文基于电力数字空间技术体系,主要研究如何在倾斜样本数据中有效学习短期数据变化特征,构建组件以实现用户用电负荷的精准预测,支撑智慧中枢应用的建设落地。本文提出了一种基于卷积神经网络(convolutional neural network, CNN)+长短期记忆(long short-term memory, LSTM)网络+全连接层的预测算法。首先,该预测算法在LSTM网络模型特征提取阶段加入CNN层,用于获取数据样本深层次特征,在CNN层中设置池化层和卷积层。其次,采用Gumbel Copula函数对电力负荷数据集进行相关性分析,依据概率分布曲线的特性描述各电力用户在外围影响因素下的关联性,将与电力负荷用户强相关的影响因素进行加权求平均,调整这些类别在用户电力负荷预测中的权重大小,从而有效获取数据变化特征。最后,针对倾斜数据进行数据预处理,降低倾斜样本对预测精度产生的影响。实验结果表明,该预测算法能有效促进电力平衡,优化电力系统资源配置,化解电力供求矛盾。

1 研究背景

电力系统的短期电力负荷预测是电网管理中的一个重要组成部分,是规划电网结构、营销、交易、安排调度计划及交易计划的前提和基础,能有效减少资源浪费,合理配置资源,促进新型



电力系统的安全平稳运行,为电网运行方式安排提供决策支持。电力负荷预测精度的高低直接关系到电网能否安全、稳定、经济地运行。因此,本文选取电力负荷场景进行数据预测研究。

1.1 电力负荷预测技术路线

近年来,随着深度学习技术的发展和算力的提升,电力负荷预测受到企业和学术界的广泛关注,并开始尝试将深度学习CNN、循环神经网络(recurrent neural network, RNN)、LSTM网络等算法应用在电力负荷研究中,深度学习模型对电力预测精度有一定的促进作用。文献[1]使用深度学习对不同层次负荷进行研究,其平均绝对百分比误差(mean absolute percentage error, MAPE)降低至23%,在测试集中,与经典RNN预测模型相比较,均方根误差(root mean square error, RMSE)降低了6.5%。文献[2]就电表级(meter-level)负荷预测展开研究,在输入设备能量消耗序列的情况下,LSTM网络的MAPE最低:将LSTM网络应用于短期负荷预测时,与传统的自回归积分滑动平均(autoregressive integrated moving average, ARIMA)模型相比,MAPE从9.74%下降到5.76%。文献[3]提出了一种DeepEnergy模型来预测负荷,该模型结构由3个一维的卷积层和3个池化层组成,最后由1个全连接层连接到输出层,经过测试,该模型的MAPE和均方根误差累积变异(cumulative variation of root mean square error, CV-RMSE)分别为9.77%和11.66%,展示出较好的精度。文献[4]基于LSTM与XGBoost组合模型的超短期电力负荷预测,针对电力负荷数据,首先建立了LSTM预测模型和XGBoost预测模型,然后使用误差倒数法将LSTM与XGBoost组合起来进行预测,实验结果表明,组合预测模型的MAPE低于单一预测模型。文献[5]提出了一种基于双向LSTM(bidirectional LSTM, Bi-LSTM)网络和特征关联性分析的日尖峰负荷预测方法,通过描述类、曲线类指

标分析不同行业下的用户日峰值负荷特性,并基于Copula函数定量分析多维时序数据之间的关联度,构建基于Bi-LSTM网络的幅值预测模型,对高位负荷的预测具有较高的精度。文献[6]提出了一种结合LSTM网络和自注意力机制的图卷积网络(graph convolutional network, GCN)的短期电力负荷预测(short term load forecasting, STLF)模型GCNLS-STLF,该模型基于LSTM网络和自注意力机制将原始多维时间序列数据转化为包含序列间关联关系的电力负荷图,并通过GCN、LSTM网络和图傅里叶变换(graph Fourier transform, GFT)对电力负荷图进行特征提取,以增强原始电力负荷数据的表达能力。文献[7]提出了一种将模糊聚类以及随机森林回归算法进行组合的电力系统负荷预测方法,通过对电力系统负荷的周期性、天气相关性等特征进行分析,利用C均值模糊聚类算法对历史样本进行聚类,在进行随机森林回归预测时,使用聚类后同类数据作为训练集样本构建决策树,有效降低了实际负荷的平均绝对误差(mean absolute error, MAE)。文献[8]提出了一种基于注意力机制的CNN门控循环单元(gated recurrent unit, GRU)短期电力负荷预测方法,该方法将历史负荷数据作为输入,搭建由一维卷积层和池化层等组成的CNN架构,提取反映负荷复杂动态变化的高维特征:将所提特征向量构造为时间序列形式,作为GRU网络的输入,建模学习特征内部动态变化规律,并引入注意力机制通过映射加权和学习参数矩阵赋予GRU隐含状态不同的权重,减少历史信息的丢失并加强重要信息的影响,最后完成短期负荷预测。文献[9]提出了一种改进的支持向量机(support vector machine, SVM)电力负荷预测方法,该方法引入实时电价这一影响因素,同时将加权灰色关联投影算法应用于节假日的历史负荷序列的选择,并采用改进的粒子群算法优化模型参数,最终实现电力负荷实时预测,与传统反向

传播 (back propagation, BP) 神经网络对比, 该方法具有较高的精确度和稳定性。文献[10]提出了一种基于改进BP神经网络的短期电力负荷预测方法, 该方法将猫群优化 (cat swarm optimization, CSO) 算法和BP神经网络相结合用于短期负荷预测, 模型的输入因子是负荷数据和气象信息等, 利用猫群算法对BP神经网络的权值和阈值进行优化, 得到BP神经网络预测模型的最优解, 从而建立短期预测模型。实例验证结果表明, 改进模型能够有效降低BP神经网络模型的预测误差, 提高预测精度。文献[11]提出了一种基于聚类经验模态分解的CNN和LSTM网络混合预测算法, 该算法首先通过经验模态分解法将负荷数据分解为平稳性好、规律性强的若干本征模态函数和残差, 其次为了简化后续模型的计算体量, 运用 k 均值聚类方法对分解所得的各分量进行分组集成, 同时分析不同聚类数对应的预测效果, 选取最优聚类标签构造神经网络输入数据, 之后将各组数据分别输入CNN-LSTM混合神经网络中, 利用CNN挖掘数据间的特征以形成特征向量, 并将其输入LSTM网络中进行预测, 最后将所有预测结果进行线性相加得到完整预测负荷, 通过在真实负荷上进行验证并与现有模型进行比较, 所提方法展现出更高的预测精度。

以上这些方法^[1-11]均有效提高了电力场景中的电力负荷预测算法的精确度, 但针对电力负荷样本中存在的倾斜样本研究较少, 而样本不均衡对预测算法的训练和预测会产生较大影响, 导致预测结果产生偏差。因此, 本文针对样本倾斜这一问题展开研究, 首先, 通过增加CNN层挖掘倾斜样本中的重要特征, 其次, 通过关联性分析调节倾斜样本在训练中的权重大小, 并通过数据预处理操作平衡倾斜数据, 以提高电力负荷预测算法的精确度。

1.2 电力负荷预测研究难点

电力负荷预测可分为4类: 超短期预测、短

期预测、中期预测和长期预测^[12]。超短期预测主要以小时或者分钟为单位, 预测未来几个小时内的负荷变化。短期预测以天为单位, 预测未来几天的负荷变化。中期预测主要预测未来几周或者几个月内的变化。长期预测以年为单位, 通常对未来5年内的电力负荷进行预测, 需要考虑的外界因素更多。

当前的电力数据普遍存在预测难、预测精度不足的难题。例如, 在农忙时节、生产用电高峰期, 用电负荷激增造成数据变化波动大, 数据易受外界因素 (如环境、天气等) 的影响。目前针对用电量预测的传统方法有自回归模型、多元回归、主成分分析法等, 但这些方法对于短期不确定变化规律的数据预测效果较差。

为了更好地处理短期变化数据的预测, 提升模型性能, 本文结合某地区电力负荷数据案例, 提出一种基于注意力机制的改进CNN-LSTM的深度学习算法, 并将之与原始LSTM网络算法、随机森林算法、BP神经网络算法3种算法进行对比, 对未来一天内的超短期电力负荷进行研究。

2 电力负荷预测算法研究

为构建电力负荷预测模型组件, 将处理后的数据输入电力模型进行实验, 输出预测结果, 模型实验流程如图1所示, 本文主要使用深度学习技术针对电力负荷进行实验, 实验发现传统的LSTM网络、随机森林算法、BP神经网络算法的预测精度不高, 且无法学习到长距离数据特征, 预测结果只与最近数据相关。为了提高超短期电力负荷预测精度, 并更好地学习以往数据特征, 本文提出改进CNN-LSTM算法模型, 并将此模型与其他深度学习模型进行对比研究。

2.1 数据预处理

在实际业务生产中, 电力数据存在数据量大但有效数据少的特点, 并且样本数据类别之间倾斜, 称为倾斜数据。例如, 在电力负荷预测场景

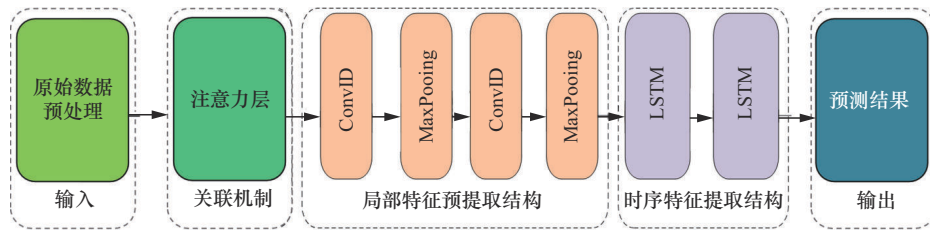


图1 模型实验流程

中,部分样本数据量少,传统的深度学习模型因训练不足导致精度缺失,这进而使得倾斜样本数据对模型造成较大的精度误差。

2.1.1 异常值处理

某个时刻的数据与一定时间序列内其他数据相差较大,称为异常数据。对于异常数据的识别,本文采用数学统计学 3σ 原则进行判断,将一天内的序列数据表示为 $\{X_1, X_2, \dots, X_{96}\}$, X_i 为第*i*个时刻的电力负荷值,计算这些数据的数学期望和标准差进行分析,计算式分别为:

$$E(i) = \frac{1}{N} \sum_{k=1}^N X(N, i) \quad (1)$$

$$\delta(i) = \sqrt{\frac{1}{N} \sum_{k=1}^N (X(N, i) - E(i))^2} \quad (2)$$

当电力负荷值不在 $[E(i) - 3\delta(i), E(i) + 3\delta(i)]$ 范围内时,判断为异常数据。对于异常数据,使用该点的前一刻数据和后一刻数据相加后的均值点替代,即 $X_i = (X_{i-1} + X_{i+1})/2$ 。

2.1.2 稀疏值处理

针对电力负荷场景中,部分样本类别数据量较少,如发电站以6h为周期统计一次发电量,而在实际电力负荷预测中,需要每30min输出预测结果,这导致重要输入数据量较少,从而使预测值产生较大误差,因此,对这部分稀疏数据采用线性插值处理。假设某一时刻发电站的发电量与电力负荷值为 $\{x_0, y_0\}$,以及后一时刻发电站的发电量与电力负荷值为 $\{x_1, y_1\}$,则根据线性插值理论,在 x_0 与 x_1 之间的第*i*个时刻,电力负

荷值 x_i 与发电量值 y_i 之间的关系表示为:

$$\frac{y_i - y_0}{y_1 - y_0} = \frac{x_i - x_0}{x_1 - x_0} \quad (3)$$

即:

$$y_i = y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x_i - x_0) \quad (4)$$

其中, y_i 为近似得到的两个时刻之间的电力负荷值。

2.1.3 标准化处理

电力负荷数据分布区间较大,存在的数据量纲也不同,为了使实验结果的精度更高,需要对电力负荷数据进行标准化处理,将所有数据缩放到一定区间中。本次实验使用零—均值(Z-score)标准化,将原始电力负荷数据 X 使用式(5)缩放到 $[-1, 1]$,经标准化得到的电力负荷数据 Z 均值为0,方差为1。

$$Z = \frac{X - \mu}{\sigma} \quad (5)$$

其中, μ 为原始电力负荷数据均值, σ 为原始电力负荷数据标准差。

2.2 评价指标

深度学习通常使用“准确率”指标来评价模型效果,预测结果与实际结果完全相同即为预测正确,而在电力实际预测场景中存在一定的合理误差区间。鉴于此,为了估算模型预测精度,本文使用MAE、MAPE、准确率(Accuracy)这3个指标对电力预测实验结果进行评价。假设真实值为 $y = \{y_1, y_2, y_3, \dots, y_n\}$,预测值为 $\hat{y} = \{\hat{y}_1, \hat{y}_2, \hat{y}_3, \dots, \hat{y}_n\}$,那么MAE满足:当预测值与真实值完全吻合时,等于0,即完美模型;误差越大,该值越大,MAE的值越小,说明预测模型拥有

更好的精确度^[13]，其计算式为：

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \hat{y}_i - y_i \right| \quad (6)$$

其中， \hat{y}_i 为负荷预测值， y_i 为实际值， n 为预测点个数。

MAPE表示真实值与预测结果之间的平均误差。MAPE为0%时表示完美模型，MAPE大于100%时表示劣质模型。因此，MAPE数值越小越好，表示为：

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (7)$$

2.3 基于改进CNN-LSTM算法的电力负荷预测研究

2.3.1 注意力层

注意力层关注目标特征和相关特征之间的关联紧密程度。首先，将不同用户数据送入注意力层，并通过Gumbel Copula函数对不同用户数据进行关联分析，以概率分布曲线特性描述每个用户在外围因素影响下的关联性；其次，根据Gumbel Copula函数分析结果，调整不同外围因素对各用户的影响权重，将该权重系数与相应的输入变量相乘，得到考虑外围因素贡献率的关联特征。Copula函数定义为：设 $F_X(x)=u$ 、 $F_Y(y)=v$ 为二元随机变量 (X,Y) 的边缘累积分布函数，则二维Copula函数式表示为：

$$F(x,y) = C_\theta(F_X(x), F_Y(y)) = C_\theta(u,v) \quad (8)$$

其中， $F(x,y)$ 为随机变量 (X,Y) 在 (x,y) 处的联合分布函数， C_θ 为Copula函数， θ 为Copula参数， $u=F_X(x)$ 、 $v=F_Y(y)$ 分别为随机变量 X 和 Y 的边缘分布。

本文采用Gumbel Copula函数计算不同用户之间的相关系数，Gumbel Copula分布函数表达式为：

$$C(u,v) = \exp \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\} \theta \in [1, \infty) \quad (9)$$

基于Gumbel Copula函数建立随机变量间相

关度量指标。本文相关度量指标为Spearman秩相关系数 ρ 。若连续随机向量 (X,Y) 的边缘分布为 $F(X)$ 和 $F(Y)$ ，相应的Copula函数为 $C(u,v)$ ，则 $C(u,v)$ 与Spearman秩相关系数 ρ 值有如下关系：

$$\rho = 12 \int_0^1 \int_0^1 C(u,v) du dv - 3 \quad (10)$$

其中， $u=F(x) \sim U(0,1)$ ， $v=F(y) \sim U(0,1)$ 。Spearman秩相关系数 ρ 值的绝对值越接近1，说明序列一致性相关度越高。

由Gumbel Copula函数求解得到参数 θ ，并将其与相应的特征变量相乘，得到考虑特征变量贡献率的关联特征，关联特征集合表达式为：

$$S_{obj} = [\theta_t^{(1)} x_t^{(1)}, \theta_t^{(2)} x_t^{(2)}, \dots, \theta_t^{(D)} x_t^{(D)}] \quad (11)$$

其中， S_{obj} 为关联特征集合， $x_t^{(m)}$ 表示 t 时间步下第 n 个特征变量， $\theta_t^{(m)}$ 表示对应 t 时间步下第 n 个特征变量的权重系数。

输入数据经过特征注意力层处理，该处理过程考虑了输入特征与外部影响因素之间的关系，提高了强相关性的外部因素权重，并降低了弱相关性的外部因素权重，据此对用户特征的重要程度进行权重分配。

2.3.2 特征提取层

CNN模型采用局部连接和权重共享的方式，对原始数据进行再次处理，能够自动提取数据中的内部特征^[14]。为了使模型更好地学习到负荷数据的特征，本文采用CNN模型提取特征，CNN由2个卷积层及2个池化层组成，在控制参数量的基础上有效提取深层特征。首先，利用卷积层获取输入数据的空间特征，并采用ReLU激活函数进行非线性变换；其次，使用池化层对特征进行下采样，生成压缩特征。输入特征经过卷积及池化后映射为高级特征，输入LSTM网络中进行时序特征提取。特征提取层结构如图2所示。

特征提取层结构包括：输入层、卷积层、池化层、全连接层和输出层。其中，输入层中将历



史数据作为模型的输入，一天内每 15 min 一个数据，将 96 维数据输入卷积层中；卷积层主要对输入的历史序列进行特征提取，设置 2 个卷积层和 2 个池化层，选择最大池化策略；数据经过卷积层和池化层的处理后，经全连接层输出，并设置激活函数为 ReLU；输出层输出结果为对次日 96 个时刻负荷的预测值。

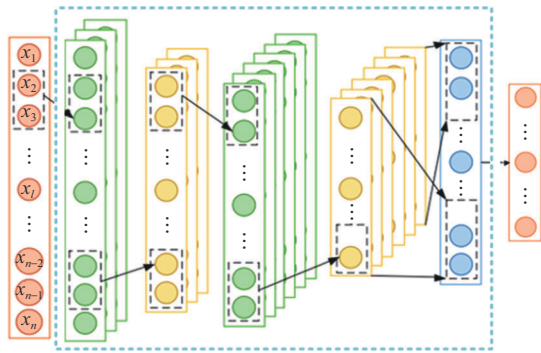


图2 特征提取层结构

图2 特征提取层结构

2.3.3 LSTM 网络层

LSTM 网络^[15]主要解决 RNN 的长期依赖问题。由于 RNN 在参数求导更新时，最主要的问题是存在梯度消散，一旦出现较小数值时，会导致在该处较远时刻的梯度几乎为 0，使长期依赖更新时无法学习特征。LSTM 网络则能够解决这一问题，在 LSTM 网络的结构中，通过在 RNN 的隐藏层中加入一个新的状态信息来进行线性传递，并且 LSTM 网络使用门控机制来进行信息筛选和记忆，确保只有需要被记忆的信息才能被传递下去，这种结构使得即使在较远的时间步长，早期的信息也能对当前状态产生影响，从而改善长期依赖问题。LSTM 网络的工作原理如图 3 所示。

LSTM 网络模型是一种特殊的 RNN 结构，在每一个计算节点中加入了遗忘门、输入门和输出门。其中，输入门和输出门设置为非线性的激活函数，而权重由遗忘门控制，通过门的控制来调节序列数据对实验结果的影响。

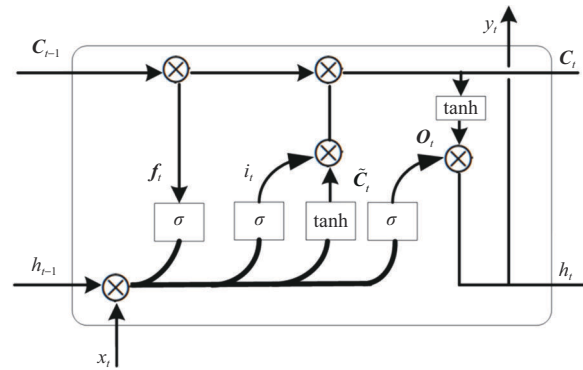


图3 LSTM 网络的工作原理

LSTM 网络中遗忘门用于决定哪些信息将被保留，哪些信息将被舍弃。在遗忘门中包含一个 sigmoid 神经网络层，包含权重参数 W_f 和偏移参数 b_f ，遗忘门接收 t 时刻的输入信号 x_t 和 $t-1$ 时刻 LSTM 网络的上一个输出信号 h_{t-1} ，则状态候选向量 f_t 表示为：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (12)$$

其中， σ 表示 sigmoid 激活函数。

LSTM 网络中输入门的作用与遗忘门相反，它将决定新输入的信息 x_t 和 h_{t-1} 中哪些信息将被保留。输入门包含两部分：sigmoid 神经网络层，包含权重 W_i 和偏移 b_i ；tanh 神经网络层，包含权重 W_c 和偏移 b_c 。sigmoid 层输出信号 i_t 与 tanh 层状态候选向量 \tilde{C}_t 分别表示为：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (13)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (14)$$

通过遗忘门和输入门，本文可以确定细胞更新状态 C_t ：

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (15)$$

C_t 将继续传递到 $t+1$ 时刻的 LSTM 网络中，作为新的细胞状态传递下去。

LSTM 网络中输出门用来控制单元状态 C_t 有多少输入 LSTM 网络的当前输出 h_t ，则状态候选向量 o_t 表示为：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (16)$$

则下一时刻的输出信号 h_t 表示为：

$$h_t = O_t \cdot \tanh C_t \quad (17)$$

3 实验设置与结果分析

3.1 数据简介

本文选取中国北方某城市供应电力资源的 6 个重要用户的保密电力负荷数据，每 15 min 记录 1 个数据，24 h 共记录 96 个数据，单位为 kW，获取 2018 年 10 月 11 日至 2021 年 10 月 11 日共 3 年的实时数据共 105 120 个。电力负荷数据按照行业划分为银行、医院、事业单位、化工厂、中学、水利。外部影响因素包括日平均气温、季节、节假日、输电线路及发电厂容量。本文将此数据集划分为训练集、验证集和测试集，比例为 7:2:1，将最后 34 天的数据作为测试数据。

3.1.1 数据预处理

本文针对该数据集中存在的孤立异常点进行数据清洗，包括异常数据处理和线性插值处理。对于存在的部分连续异常值使用缩放、归一化的方式，将数据规范到正常数据范围内，减少异常值对数据集的影响。针对某一时段内孤立异常数据的处理前后对比如图 4 所示。

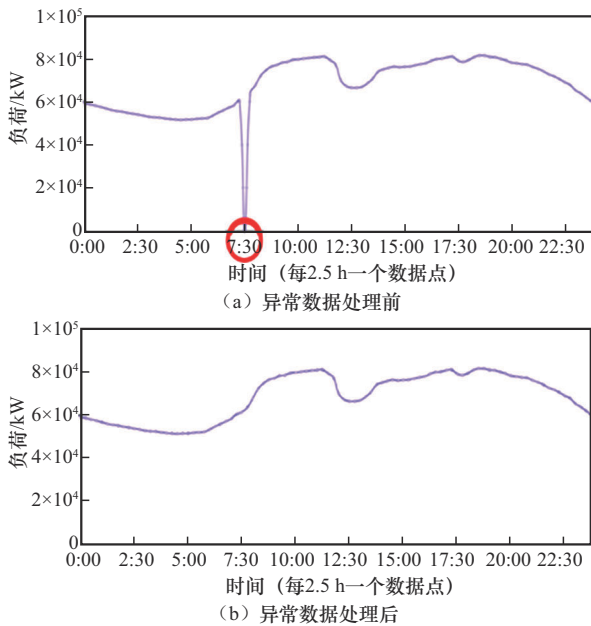


图4 异常数据的处理前后对比

3.1.2 关联性分析

本文采用 Gumbel Copula 函数对电力负荷数据集与外部影响因素数据（日平均气温、季节、节假日、输电线路、发电厂容量）进行相关性分析，通过概率分布曲线的特性描述各电力用户在外围影响因素下的关联性，根据分析结果调整不同行业的外部因素权重大小。Gumbel Copula 函数选择银行、事业单位、医院、化工厂、中学以及水利共 6 个电力用户的电力负荷数据，外部影响因素包括日平均气温、季节、节假日、输电线路、发电厂容量。不同电力负荷用户的相关性分析如图 5 所示。图 5 展示了 6 个电力用户的电力负荷数据与各影响因素在不同时间段下的 Gumbel Copula 参数 θ 的平均值。

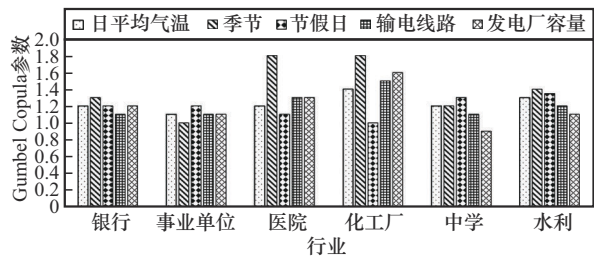


图5 不同电力负荷用户的相关性分析

由 Gumbel Copula 函数可知，Gumbel Copula 的参数值越偏离 1，则表明该用户电力负荷与该类别的正相关性越强。由此可知，化工厂、医院和水利的电力负荷对于季节、输电线路和发电量变化较敏感，这与企业的营业性质具有强关联性；中学对于节假日、日平均气温和季节因素较为关注，这可能与学校的作息时间挂钩；事业单位与银行用户负荷波动则与季节和日平均气温因素的关联较强。因此，提高季节、输电线路和发电量对化工厂、医院和水利这 3 个特征的权重贡献，能够有效减少因样本倾斜而导致的模型训练误差。本文将与电力负荷用户强相关的影响因素进行加权求平均，调整这些类别在用户电力负荷预测中的权重大小，将该特征集合 S_{obj} 输入改进的 CNN-LSTM 算法中。



3.2 实验参数设置

3.2.1 改进CNN-LSTM模型参数设置

本实验中,改进CNN-LSTM模型参数设置与Bi-LSTM^[5]类似,其中,输入层维数等于输入变量的个数,即包含所有类别的强关联因素特征个数,输出层数值为1,模型层数设为5,学习率设置为0.01,使用最大池化层,训练轮数为200,批处理大小为6,激活函数使用ReLU函数,优化器为Adam。改进CNN-LSTM实验参数见表1。

表1 改进CNN-LSTM实验参数

模型参数	属性值
LSTM层数	5
隐藏层维数	96
优化器	Adam
学习率	0.01
Dropout	0.2
训练轮数	200

实验训练过程的损失函数如图6所示。

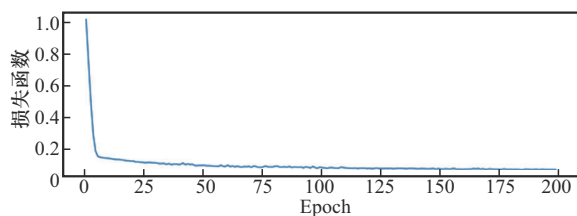


图6 实验训练过程的损失函数

3.2.2 基于随机森林算法的电力负荷预测实验

随机森林电力负荷预测是利用分类和回归树(classification and regression tree, CART)决策树集进行集成计算。在使用决策树算法时,决策树的个数对实验结果有较大的影响。当决策树个数设置为很大时,计算量增大且负荷预测的泛化误差不断减小;当决策树个数设置为很小时,实验结果误差较大。因此,综合考虑,本次实验设置决策树大小为300,实验时选用Sklearn的随机森林回归模型,输入预定处理后的某地区电力负荷实验数据,输出电力负荷某天96个时刻的预测结

果,再对实验结果分别从MAE、MAPE、Accuracy这3个评价指标上进行分析。随机森林算法的参数设置见表2。

表2 随机森林算法的参数设置

模型参数	属性值
批处理大小	6
迭代次数	200
最大树深度	7
子节点最小权重和	2
样本采样率	0.9
特征采样率	0.6
评估器数量	50

3.2.3 基于BP神经网络算法的电力负荷预测实验

BP神经网络算法是基于数学和统计类型的学习方法,它具有较强的学习能力,可以通过误差反向传递进行模型监督^[16]。BP神经网络的原理:对于输入样本训练数据集,采用误差逆传播算法对神经元之间的权重进行调整,当输出层的误差小于指定误差或训练次数大于迭代次数时,训练完成。BP神经网络算法的参数设置见表3。

表3 BP神经网络算法的参数设置

模型参数	属性值
初始权重	0.5
隐藏层神经元数量	180
激活函数	Softmax
学习率	0.05
最小性能梯度	1×10^{-6}
最大迭代次数	300

BP神经网络由输入层、1个或多个隐藏层以及1个输出层组成,其隐含层可以逼近任意一个多项式,由前向网络和反前向网络构成。输入预定处理后的某地区电力负荷实验数据,输出电力负荷某天96个时刻的预测结果,再对实验结果分别从MAE、MAPE、Accuracy这3个评价指标上进行分析。

3.3 实验结果分析

对本文所提出的预测模型进行评估，将改进 CNN-LSTM 模型与 LSTM 网络、随机森林和 BP 神经网络 3 种算法对某地区电力负荷的同一数据集进行实验，使用 MAE、MAPE、Accuracy 这 3 个指标对电力预测实验结果进行评价，模型评估结果见表 4。

由表 4 可知，首先，对于幅值范围较小的银行、医院、水利、中学和事业单位的电力用户，CNN 的引入在任一指标上均能获得最佳的预测效果，反映出 CNN 对非线性数据的特征提取能力和模型对行业输入的敏感性，这使得 CNN 能有效提取输入数据的特征，发掘出时间细粒度较小但重要性较高的数据；其次，联合用户关联性分析，将与电力负荷用户强相关的影响因素进行加权求平均，调整这些类别在用户电力负荷预测中的权重大小，有效提高了电力负荷算法的预测准确性。对于用电负荷较大的化工行业，采用 BP

神经网络算法的效果优于基于改进 CNN-LSTM 的结果，这可能是因为这类用户的负荷突变性较强、用电需求变化较大、工厂检查及开机以及需求订单变化的影响较大。因此，加入 CNN 特征提取层及特征关联性分析能够使预测模型有效地获取不同用户之间的关联影响因素，提取深层次数据特征，从而提高了电力负荷预测算法的鲁棒性及 Accuracy。

基于改进 CNN-LSTM 模型对银行某天 96 个时刻预测结果与真实数据对比的电力负荷预测如图 7 所示。

实验结果表明，基于改进 CNN-LSTM 模型的预测算法能够准确预测银行某天的电力负荷趋势，为电力负荷预警及低碳减排提供有效参考。综上，在实际电力负荷预测中，电网企业可以根据用电负荷高峰趋势进行预警，提前处置用电负荷险情，并将相关数据输入预测模型中，通过重点提取对应时段的用电变化信息，不断迭代训

表 4 模型评估结果

用户	改进 CNN-LSTM			LSTM 网络			随机森林			BP 神经网络		
	MAE	MAPE	Accuracy	MAE	MAPE	Accuracy	MAE	MAPE	Accuracy	MAE	MAPE	Accuracy
银行	0.53	0.83	94.54	0.75	0.89	93.20	0.60	0.81	92.60	0.62	0.63	92.20
医院	0.72	0.96	73.12	0.87	1.02	71.40	0.75	1.37	81.56	0.79	0.91	73.04
水利	1.09	1.13	71.29	0.41	1.24	70.40	1.24	1.59	66.84	1.21	1.73	70.10
中学	0.58	1.24	71.46	0.63	1.37	69.25	1.35	1.77	64.71	1.38	1.65	68.40
事业单位	0.47	0.58	73.68	0.73	0.87	71.45	0.93	1.41	72.51	1.27	1.64	70.78
化工厂	1.95	2.28	72.13	1.44	1.58	68.30	1.82	2.41	50.31	1.25	1.73	74.07

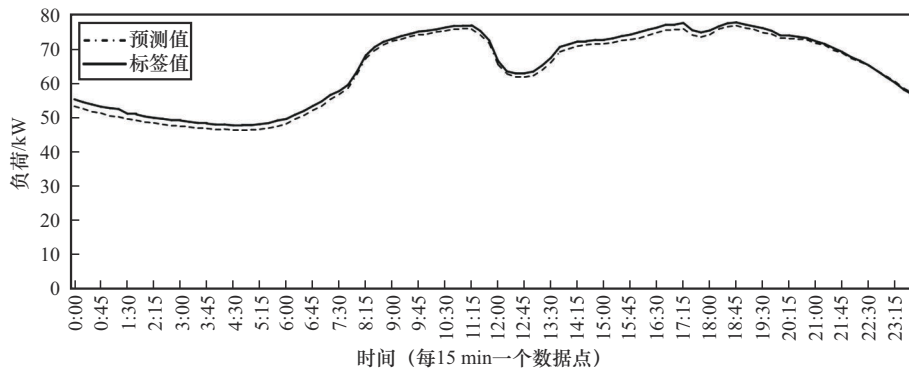


图 7 银行某天 96 个时刻预测结果与真实数据对比的电力负荷预测



练, 提高对用电负荷高峰期数据的敏感度, 从而进一步提高电力负荷预测算法的预测精度。

4 结束语

本文针对倾斜电力样本数据预测偏差问题提出一种基于注意力机制的改进CNN-LSTM算法。本文首先采用Gumbel Copula函数对各用户与外部影响因素之间进行关联性分析, 并调节各外部因素对用户电力负荷预测的权值, 改善倾斜样本对算法预测精度的影响; 其次, 构建CNN层提取用户数据空间特征以及构建LSTM网络层提取用户数据时序特征。实验结果表明, 与其他数据预测技术在电力负荷场景中进行实验对比, 基于改进的CNN-LSTM算法实验精度对于短期不确定性数据的预测精度高, 效果较好。未来将不断深化研究, 完善组件在其他电力业务场景中的应用。例如, 预测能源价格趋势以支撑对市场的精准研判, 通过价格变化回溯对能源储备做出预警, 提高风险防控的技术手段, 实现对未来能源结构的精准预测。

参考文献:

- [1] SHI H, XU M H, LI R. Deep learning for household load forecasting: a novel pooling deep RNN[J]. IEEE Transactions on Smart Grid, 2017, 9(5): 5271-5280.
- [2] KONG W C, DONG Z Y, HILL D J, et al. Short-term residential load forecasting based on resident behaviour learning[J]. IEEE Transactions on Power Systems, 2017, 33(1): 1087-1088.
- [3] KUO P H, HUANG C J. A high precision artificial neural networks model for short-term energy load forecasting[J]. Energies, 2018, 11(1): 213.
- [4] 陈振宇, 刘金波, 李晨, 等. 基于LSTM与XGBoost组合模型的超短期电力负荷预测[J]. 电网技术, 2020, 44(2): 614-620.
CHEN Z Y, LIU J B, LI C, et al. Ultra short-term power load forecasting based on combined LSTM-XGBoost model[J]. Power System Technology, 2020, 44(2): 614-620.
- [5] 李玉志, 刘晓亮, 邢方方, 等. 基于Bi-LSTM和特征关联性分析的日尖峰负荷预测[J]. 电网技术, 2021, 45(7): 2719-2730.
LI Y Z, LIU X L, XING F F, et al. Daily peak load prediction based on correlation analysis and bi-directional long short-term memory network[J]. Power System Technology, 2021, 45(7): 2719-2730.
- [6] 史含笑, 王雷春. 结合LSTM和自注意力机制的图卷积网络短期电力负荷预测[J]. 计算机应用, 2024, 44(1): 311-317.
SHI H X, WANG L C. Short-term power load forecasting by graph convolutional network combining LSTM and self-attention mechanism[J]. Journal of Computer Applications, 2024, 44(1): 311-317.
- [7] 李焱, 贾雅君, 李磊, 等. 基于随机森林算法的短期电力负荷预测[J]. 电力系统保护与控制, 2020, 48(21): 117-124.
LI Y, JIA Y J, LI L, et al. Short term power load forecasting based on a stochastic forest algorithm[J]. Power System Protection and Control, 2020, 48(21): 117-124.
- [8] 赵兵, 王增平, 纪维佳, 等. 基于注意力机制的CNN-GRU短期电力负荷预测方法[J]. 电网技术, 2019, 43(12): 4370-4376.
ZHAO B, WANG Z P, JI W J, et al. A short-term power load forecasting method based on attention mechanism of CNN-GRU[J]. Power System Technology, 2019, 43(12): 4370-4376.
- [9] 赵佩, 代业明. 基于实时电价和加权灰色关联投影的SVM电力负荷预测[J]. 电网技术, 2020, 44(4): 1325-1332.
ZHAO P, DAI Y M. Power load forecasting of SVM based on real-time price and weighted grey relational projection algorithm[J]. Power System Technology, 2020, 44(4): 1325-1332.
- [10] 王克杰, 张瑞. 基于改进BP神经网络的短期电力负荷预测方法研究[J]. 电测与仪表, 2019, 56(24): 115-121.
WANG K J, ZHANG R. Research on short-term power load forecasting method based on improved BP neural network[J]. Electrical Measurement & Instrumentation, 2019, 56(24): 115-121.
- [11] 刘亚璋, 赵倩. 基于聚类经验模态分解的CNN-LSTM超短期电力负荷预测[J]. 电网技术, 2021, 45(11): 4444-4451.
LIU Y Z, ZHAO Q. Ultra-short-term power load forecasting based on cluster empirical mode decomposition of CNN-LSTM[J]. Power System Technology, 2021, 45(11): 4444-4451.
- [12] 陈卓, 孙龙祥. 基于深度学习LSTM网络的短期电力负荷预测方法[J]. 电子技术, 2018, 47(1): 39-41.
CHEN Z, SUN L X. Short-term electrical load forecasting based on deep learning LSTM networks[J]. Electronic Technology, 2018, 47(1): 39-41.
- [13] 李冲. 随机森林模型预测岩溶区酸性煤矿井水锰污染[J]. 中国煤炭地质, 2021, 33(3): 43-47, 59.
LI C. Prediction of karst region acidic coalmine water manganese pollution based on random forest[J]. Coal Geology of China, 2021, 33(3): 43-47, 59.
- [14] 魏健, 赵红涛, 刘敦楠, 等. 基于注意力机制的CNN-LSTM短期电力负荷预测方法[J]. 华北电力大学学报(自然科学版), 2021, 48(1): 42-47.
WEI J, ZHAO H T, LIU D N, et al. Short-term power load fore-

casting method by attention-based CNN-LSTM[J]. Journal of North China Electric Power University (Natural Science Edition), 2021, 48(1): 42-47.

[15] SRIVASTAVA A K, PANDEY A S, SINGH D. Notice of violation of IEEE publication principles: short-term load forecasting methods: a review[C]//Proceedings of the 2016 International Conference on Emerging Trends in Electrical Electronics & Sustainable Energy Systems (ICETEESES). Piscataway: IEEE Press, 2016: 130-138.

[16] 王红亚, 王旭红, 孙俊敏, 等. 混合蛙跳算法优化神经网络的同步电机转子故障检测[J]. 微电机, 2017, 50(9): 22-26.

WANG H Y, WANG X H, SUN J M, et al. Synchronous machine rotor inter-turn short circuit fault detection based on ISFLA-BP neural network[J]. Micromotors, 2017, 50(9): 22-26.

[作者简介]



张琦 (1995-), 男, 安徽继远软件有限公司助理工程师, 主要研究方向为深度学习与图像识别。



李志浩 (1981-), 男, 安徽继远软件有限公司工程师, 主要研究方向为电力数字孪生技术、电力物联网及量子通信。



范叶平 (1979-), 男, 安徽继远软件有限公司高级工程师, 主要研究方向为电力数字建模、数字孪生基础技术平台及大数据分析。



马剑波 (1998-), 男, 安徽继远软件有限公司工程师, 主要研究方向为电力人工智能及大数据分析。