

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2022.0399

周期规律增强的多视角短期电力负荷预测

苏伟*, 肖小龙, 史明明, 方鑫, 司鑫尧

(国网江苏省电力有限公司 电力科学研究院, 南京 210000)

摘 要: 短期电力负荷预测对电力系统的可靠运行具有重要意义。现有方法存在如下问题: 缺乏对特征之间依赖关系的挖掘; 忽略了电力负荷变化的周期性规律。为此, 提出一种周期规律增强的多视角短期电力负荷预测网络 (EPISODE) 方法。EPISODE 方法主要包括 2 个核心组件: 多视角特征学习组件和周期规律增强的电力负荷预测组件。前者旨在有效提取电力负荷数据中的静态特征与时序特征, 以得到增强的特征表示; 后者则是对电力负荷数据进行一般性时序挖掘和周期性时序挖掘, 从而得到全面的电力负荷历史数据表征。基于后期融合的方式, 实现短期电力负荷预测。在真实公开的电力负荷预测数据集上进行了大量实验。实验结果证明了所提方法相比现有基准方法的先进性。

关 键 词: 周期规律增强; 短期负荷预测; 特征压缩与激励; 门控循环单元; 注意力神经网络

中图分类号: TP399

文献标志码: A

文章编号: 1001-5965(2024)02-0477-10

电力系统是中国经济发展的重要基石。其中, 电力能源也是重要的能源之一。准确的电力负荷预测不仅可以指导电力生产部门制定合理的发电计划, 降低发电成本, 还有利于提高电力系统运行的安全稳定性^[1]。根据负荷预测的时间跨度, 电力负荷预测可大致分为: 长期负荷预测、中长期负荷预测、短期负荷预测和超短期负荷预测。其中, 长期负荷预测通常指未来若干年的负荷预测; 中长期负荷预测是未来若干月或若干年的负荷预测; 短期负荷预测是对未来若干日或若干周的负荷预测; 超短期负荷预测是指对未来若干秒或若干分钟的电力预测。由于交流电不易存储的特点, 电力系统需要及时调整生产调度。此时, 调度的主要依据来源于短期电力负荷预测的结果。因此, 短期电力负荷预测对于电力系统的重要性不言而喻。

传统的短期电力负荷预测方法主要分为基于统计理论的方法和基于机器学习的方法。基于统

计理论的方法主要包括多元线性回归 (multiple linear regression, MLR)^[2]、自回归移动平均值 (autoregressive moving average, ARMA)^[3]、差分整合移动平均自回归 (autoregressive integrated moving average, ARIMA)^[4] 等。这类方法简单、效率高, 但对非线性因素处理的能力较弱, 很难取得好的实际应用效果。基于机器学习的方法主要包括人工神经网络 (artificial neural network, ANN)^[5]、支持向量机 (support vector machine, SVM)^[6-7]、随机森林 (random forest, RF)^[8] 等。这类方法虽然考虑了负荷的时序性和其他的非线性因素, 但其只能对少量数据及特征进行处理, 而且训练时间长, 对原始数据的平稳性要求也较高。

近年来, 随着深度学习算法在诸多领域的成功应用, 国内外许多学者都开始尝试利用深度学习算法解决短期电力负荷预测任务^[9]。经典的深度学习算法包括卷积神经网络 (convolutional neural net-

收稿日期: 2022-05-21; 录用日期: 2022-07-02; 网络出版时间: 2022-12-14 15:58

网络出版地址: link.cnki.net/urlid/11.2625.V.20221213.1218.004

基金项目: 国网江苏省电力有限公司科技项目 (J2021165)

*通信作者. E-mail: susw2011@163.com

引用格式: 苏伟, 肖小龙, 史明明, 等. 周期规律增强的多视角短期电力负荷预测 [J]. 北京航空航天大学学报, 2024, 50 (2): 477-486.

SU W, XIAO X L, SHI M M, et al. Periodic pattern-enhanced multi-view short-term load prediction [J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50 (2): 477-486 (in Chinese).

work, CNN)^[10] 和循环神经网络(recurrent neural network, RNN)^[11] 等。其中, CNN 凭借其优秀的局部特征提取能力, 被广泛用于计算机视觉领域; RNN 及其变种模型, 如长短期记忆(long short-term memory, LSTM)^[12] 网络和门控循环单元(gated recurrent unit, GRU)^[13] 网络, 则更加适用于挖掘序列数据, 因而被广泛用于自然语言处理和时间序列预测等领域。为了增强模型对电力负荷历史的表征, 研究学者联合多种深度神经网络来实现电力负荷的有效预测^[14]。虽然这些基于深度学习算法的工作已经取得了显著的短期电力负荷预测效果, 但是这些工作主要存在 2 个局限: ①缺乏对模型所学各维度潜在特征之间的依赖关系的挖掘, 难以达到最佳的特征提取效果; ②忽略了电力负荷变化的周期性规律。比如, 电力负荷往往在晚间或周末较为突出。

为了解决上述局限, 本文提出了一种周期规律增强的多视角短期电力负荷预测网络(periodic pattern-enhanced multi-view short-term power load prediction networks, EPISODE)方法, 该预测方法包括 2 个关键的核心组件: 多视角特征学习组件和周期规律增强的电力负荷预测组件。其中, 多视角特征学习组件旨在从多个视角提取电力数据中的特征信息。分别利用设计的 1D-CNN 网络与 GRU 网络, 提取电力数据中包含的静态特征与时序特征, 并利用上下文特征融合单元对 2 种特征进行融合, 以增强特征的表示学习。而在周期规律增强的电力负荷预测组件中, 除了利用 GRU 进行一般性时序历史编码外, 还使用 GRU-Skip 来进行周期性时序历史编码, 并利用后期融合的方式对二者进行融合, 实现最终的电力负荷预测。为了验证模型的有效性, 本文在公开数据集上开展了大量实验。结果表明相比现有方法, 本文方法不仅在预测精度而且在预测效率上都具有明显的优越性。

1 相关工作

早期的电力负荷预测主要使用统计学理论或机器学习方法来进行电力负荷预测。例如, 彭鹏和彭佳红^[9] 根据 GDP、人口和全社会用电量等变量的历史数据建立多元线性回归模型, 以进行电力负荷预测; 李龙等^[15] 利用传统的人工神经网络对电力负荷进行预测。虽然这类方法取得了一定的预测效果, 但这类方法普遍不能很好地处理非线性因素对电力负荷值的影响, 导致其预测值与真实值的差异较大。

随着深度学习技术的蓬勃发展, 越来越多学者

开始采用深度学习作为主要技术路径, 以解决电力负荷预测任务。陆继翔等^[16] 提出 CNN-LSTM 这一混合神经网络模型, 其先通过 CNN 来提取可能影响电力负荷的特征, 再将其作为 LSTM 网络的输入, 从而实现短期电力负荷预测。李泽文等^[17] 提出多因素短期负荷预测模型, 其先运用主成分分析法对原始多维输入变量的时间序列进行主成分提取, 实现负荷数据的降维, 然后通过双向 LSTM 网络, 最终得到预测结果。杨洋和谷震浩^[18] 提出 XGBoost-DNN, 其首先利用 XGBoost 学习到从数据到目标的交叉特征, 然后利用这份预训练学习到的交叉特征结合原数据通过 DNN 的特性去学习中期电力负荷的时序关系, 最后得到预测结果。

虽然上述方法取得了一定的成功, 但所提出的模型对时序特征和静态特征的利用仍然不充分, 这导致模型挖掘数据中重要信息的能力下降。同时, 上述工作忽略了电力负荷变化存在一定的周期性规律。因此, 为了提升电力负荷预测的精度, 本文方法引入了特征压缩与激励机制, 增强对重要特征的筛选, 利用 GRU-Skip 实现了对电力负荷数据中存在的周期性信息进行建模, 同时基于后期融合方式进行电力负荷预测, 进而实现了更加精准的电力负荷预测效果。

2 多视角短期电力负荷预测框架

2.1 问题定义

为能够及时地指导电力调度, 节约能源, 本文重点研究短期电力负荷预测, 尤其是预测未来一天内的电力负荷。假设有一段时期(如几年)内某区域电力负荷的采样序列为 $\{x_1, x_2, \dots, x_Q\}$ 。该序列的采样间隔设为 D , 单位为 min, 其中 D 可以被 60 整除。这样每天的电力负荷一共有 $60/D \times 24$ 个采样点。与此同时, 电力负荷采样序列各时刻相对应的天气相关数据序列, 设为 $\{x'_1, x'_2, \dots, x'_Q\}$, 其中 $x'_i \in \mathbf{R}^d$ ($i = 1, 2, \dots, Q$) 表示每一时刻的天气相关多变量数据, 包括温度和湿度等。 d 为该天气多变量的维度。由于天气相关因素大大影响电力负荷的大小, 这些多变量数据可以用于辅助预测未来的电力负荷。

为了构建训练样本集合, 对上述采样序列和天气相关数据按照滑动窗口的方式进行分割, 其中, 每个窗口覆盖 M 天的历史电力负荷数据和天气相关数据, 即 $M(60/D \times 24)$ 个采样点数据。最终, 假设获得一个包含 N 个训练数据样本的集合 $S = \{(X^1, y^1), (X^2, y^2), \dots, (X^N, y^N)\}$ 。其中, $X^j = [x'_1, x'_2, \dots, x'_T]^T \in \mathbf{R}^{(d+1) \times T}$ ($j = 1, 2, \dots, N$) 表示 $T = M(60/D \times 24)$ 个连续历史时刻的电力负荷和天气信息历史数据,

$\mathbf{x}_t^j = [X_{(j-1)*T+t}; \mathbf{x}'_{(j-1)*T+t}] \in \mathbf{R}^{d+1} (j = 1, 2, \dots, T)$ 表示第 t 个时刻观察到的电力负荷采样值和相关天气多变量数据。 $\mathbf{y}^j = [X_{j*T+1}, X_{j*T+2}, \dots, X_{j*T+\tau}] \in \mathbf{R}^\tau (j = 1, 2, \dots, N)$ 表示对应的未来一天内 τ 个时刻的真实电力负荷采样值, 其中, $\tau = 60/D \times 24$ 。本文旨在利用这些训练数据优化一个模型 F , 使其可以捕捉未来短期电力负荷变化与历史负荷和天气等数据之间的潜在联系, 即能够根据历史 T 个时刻的负荷历史和天气信息即 $\mathbf{X}$, 预测未来 τ 个时刻的电力负荷 $\mathbf{y}$ 。模型的表达式为

$F(\mathbf{X}, \boldsymbol{\theta}) \rightarrow \mathbf{y}$

(1)

式中: $\boldsymbol{\theta}$ 为待学习的参数。

2.2 模型结构

如图 1 所示, EPISODE 方法包含 2 个组件: 多视角特征学习组件和周期规律增强的电力负荷预测组件。不同于传统的单一特征提取方式, 为了充分利用 CNN 的静态局部特征提取能力和门控循环单元的动态时序信息挖掘能力, 在特征学习组件中, 同时采用基于一维 CNN 和基于 GRU 的 2 种特征提取手段。其中, 将基于一维 CNN 提取的特征当作上下文信息, 用于增强基于 GRU 的特征提取。值得一提的是, 除了使用一维 CNN, 本文还引入了特征压缩与激励(squeeze-and-excitation)技术,

该技术在一定程度上弥补了 CNN 无法很好捕捉各特征之间的依赖关系的不足, 促进该组件的特征提取能力。同时, 为了深入挖掘电力负荷使用数据的变化趋势, 本文联合使用基于 GRU 的一般性时序历史编码和基于 GRU-Skip 的周期性时序历史编码, 以增强模型对电力负荷历史的表征能力。借助注意力机制, 对基于 GRU-Skip 挖掘到的周期性时序历史编码表示, 进行动态融合, 以实现更加准确的周期性时序历史表征。最后, 采用后期融合的方式, 实现对电力负荷的预测。

2.3 多视角特征学习

如图 1 所示, 本文设计的多视角特征学习组件主要由若干特征提取层构成。假设一共有 K 层特征提取层。每个特征提取层主要由 2 个模块构成: 基于 1D-CNN 的静态特征提取模块和基于 GRU 的时序特征提取模块。

2.3.1 基于 1D-CNN 的静态特征提取

如图 2 所示, 本文设计的基于 1D-CNN 的静态特征提取模块由一维卷积层、批量归一化层和压缩激励层组成。

该模块可定义为

$$\begin{cases} \hat{\mathbf{X}}^0 = \mathbf{X} \\ \hat{\mathbf{X}}^{k+1} = G(\hat{\mathbf{X}}^k) \end{cases} \quad k = 1, 2, \dots, K$$

(2)

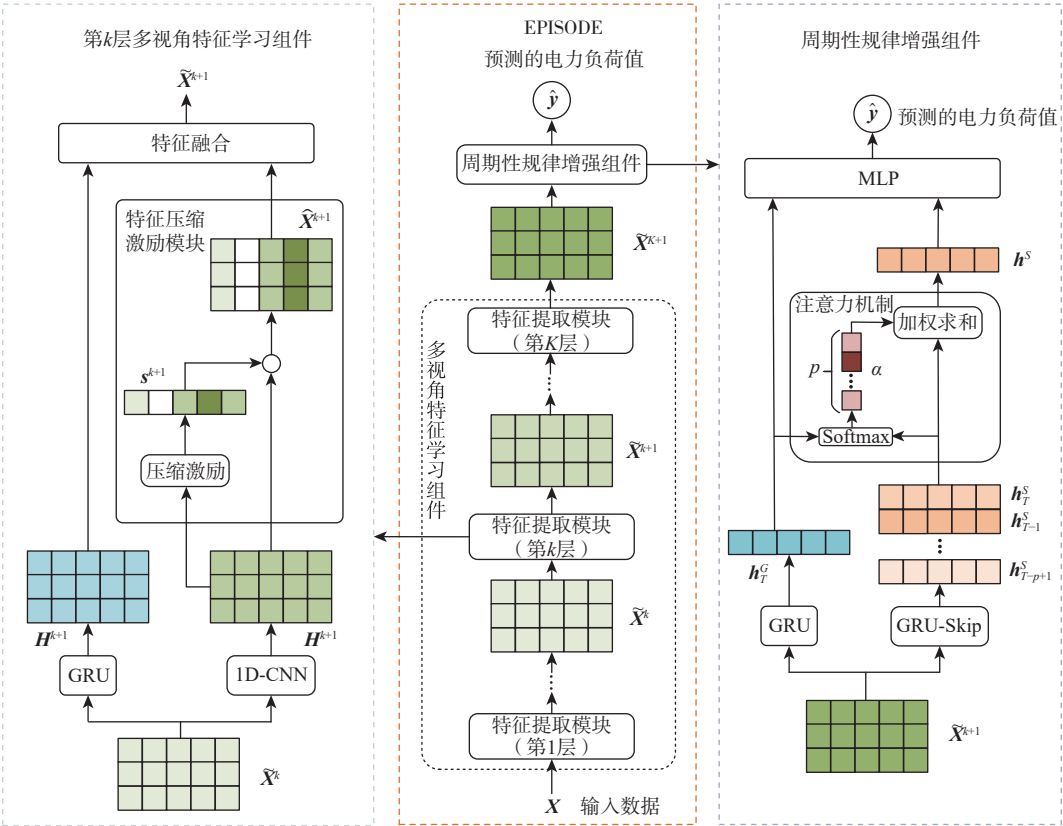


图 1 EPISODE 整体网络结构图

Fig. 1 Overall network structure diagram of EPISODE

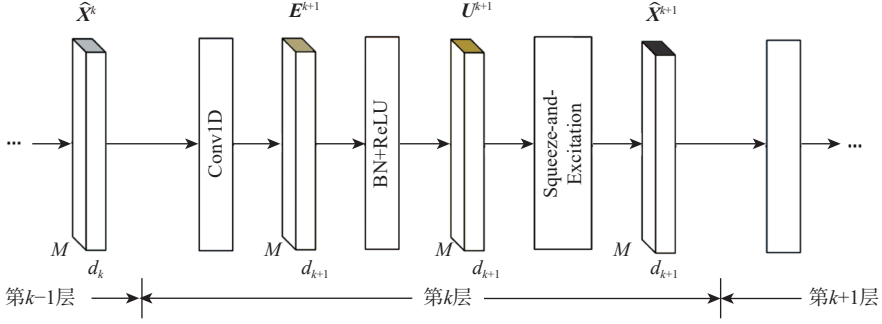


图2 基于一维卷积神经网络的静态特征提取模块

Fig. 2 1D-CNN based static feature extraction module

式中: $\hat{X}^k$ 表示第 $k-1$ 层的静态特征提取操作的输出,也是第 k 层静态特征提取操作的输入。 $\hat{X}^0 = X \in \mathbf{R}^{(d+1) \times T}$,表示原始的输入数据。 G 表示基于一维卷积的静态特征提取网络。该网络主要包含3个部分:一维卷积、批量归一化和特征的压缩与激励。

传统的CNN主要应用于图像等二维数据,其卷积核可以沿2个方向进行滑动,而一维卷积操作的卷积核则只按照一个维度进行滑动。在本任务背景下,将卷积核沿着时间的维度进行滑动,以此获得电力负荷及其相关天气变量的时序变化规律。为了方便后续说明,将第 k 层的一维卷积操作定义为

$$E^{k+1} = \text{CNN1}(\hat{X}^k) \quad k = 1, 2, \dots, K \quad (3)$$

式中: $E^{k+1} = [e_1^{k+1}, e_2^{k+1}, \dots, e_T^{k+1}] \in \mathbf{R}^{d_{k+1} \times T}$ 为 $\hat{X}^k$ 通过第 k 层1D-CNN得到的潜在表示特征。

本文引入一层批量归一化层以防止梯度爆炸,并将输出结果表示为 $U^{k+1} = [u_1^{k+1}, u_2^{k+1}, \dots, u_T^{k+1}] \in \mathbf{R}^{d_{k+1} \times T}$,具体计算如式(4)所示:

$$U^{k+1} = \delta(\text{BN}(E^{k+1})) \quad k = 1, 2, \dots, K \quad (4)$$

式中:BN表示批量归一化层; δ 表示ReLU激活函数。

虽然上述运算一定程度上可以将特征的重要信息融合至局部空间中,但其不能很好地捕获特征之间的关联性,因而无法有选择地突出重要的信息特征和抑制不太重要的特征。因此,为了更好地捕捉数据特征,本文采用压缩激励技术^[19],如图3所示。本模块中的压缩激励操作主要涉及2个子操作:特征压缩(squeeze)操作 F_{sq} 和特征激励(excitation)操作 F_{ex} 。

特征压缩操作旨在捕获每维特征的全局性代表信息,其可通过使用全局平均池化等操作来实现。具体而言,对于通过批量归一化层获得的输出数据 U^{k+1} ,将其在时间维度上进行压缩,以获得每维特征的全局性表示。假设该全局信息表示向量

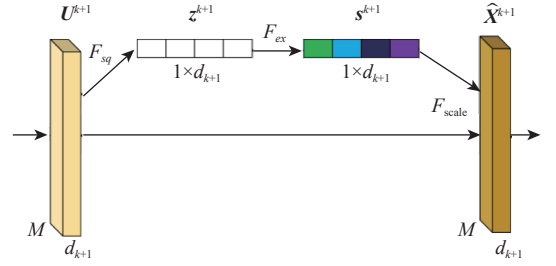


图3 压缩激励模块结构

Fig. 3 Squeeze-and-excitation module structure

为 $z^{k+1} = [z_1^{k+1}, z_2^{k+1}, \dots, z_{d_{k+1}}^{k+1}] \in \mathbf{R}^{d_{k+1}}$,其中, z^{k+1} 的第 j 个元素 z_j^{k+1} 计算式为

$$z_j^{k+1} = F_{\text{sq}}(U^{k+1}[j,:]) = \text{avg}(U^{k+1}[j,:]) \quad j = 1, 2, \dots, d_{k+1} \quad (5)$$

特征激励操作通过挖掘特征之间的依赖关系,以更好地激励对电力负荷预测较重要的特征,同时抑制不太重要的特征。以第 k 个特征提取层为例,该特征激励操作可以通过使用如下的门控机制实现:

$$s^{k+1} = F_{\text{ex}}(z^{k+1}) = \sigma(W_2^k \delta(W_1^k z^{k+1})) \quad (6)$$

式中: σ 为Sigmoid激活函数; δ 为ReLU激活函数; $W_1^k \in \mathbf{R}^{\frac{d_{k+1}}{r_k} \times d_{k+1}}$ 和 $W_2^k \in \mathbf{R}^{d_{k+1} \times \frac{d_{k+1}}{r_k}}$ 为第 k 层特征提取单元中该操作的相关参数, W_1^k 和 W_2^k 主要用于提升泛化能力,并使输出的向量 $s^{k+1} = [s_1^{k+1}, s_2^{k+1}, \dots, s_{d_{k+1}}^{k+1}] \in \mathbf{R}^{d_{k+1}}$ 与 z^{k+1} 的维度保持一致; r_k 为第 k 层特征提取单元中该操作的超参数,代表缩减比例,主要用于减少模型的计算复杂度,取值范围是正整数; $s^{k+1} \in \mathbf{R}^{d_{k+1}}$ 表示各潜在表示特征的重要性。最后,根据 s^{k+1} 中的权重信息对潜在表示特征 U^{k+1} 进行相应的激励,具体如式(7)所示:

$$\hat{X}^{k+1}[j,:] = F_{\text{scale}}(U^{k+1}[j,:], s_j^{k+1}) = s_j^{k+1} U^{k+1}[j,:] \quad j = 1, 2, \dots, d_{k+1} \quad (7)$$

式中: s_j^{k+1} 为向量 s^{k+1} 的第 j 个元素; $\hat{X}^{k+1} = [\hat{x}_1^{k+1}, \hat{x}_2^{k+1}, \dots, \hat{x}_T^{k+1}] \in \mathbf{R}^{d_{k+1} \times T}$ 为第 k 层特征提取单元对于输

入数据提取的最终特征表示。

2.3.2 基于GRU的时序特征提取

针对短期电力负荷预测任务, 本文选用GRU作为时序特征提取网络。其原因在于传统RNN在实际中, 较难捕获时间序列中长期时序依赖关系。如图4所示, 本文设计的基于GRU的时序特征提取模块由GRU单元和上下文特征融合单元 F_{cd} 组成。GRU用于挖掘特征潜在的时序性依赖关系, 而上下文特征融合单元 F_{cd} 用于将基于ID-CNN捕获的静态特征作为上下文信息与GRU提取的时序特征进行融合, 进而增强特征的表示学习。

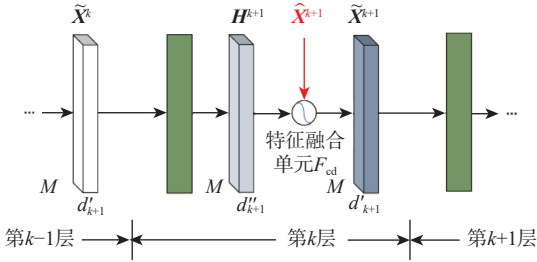


图4 基于GRU的时序特征提取模块结构

Fig. 4 GRU based time series feature extraction module structure

整体而言, 第 k 层特征提取单元中该模块的定义, 如式(8)所示:

$$\begin{cases} \tilde{X}^0 = X \\ \tilde{X}^{k+1} = H(\tilde{X}^k) \end{cases} \quad k = 0, 1, \dots, K \quad (8)$$

式中: $\tilde{X}^k = [\tilde{x}_1^k, \tilde{x}_2^k, \dots, \tilde{x}_T^k] \in \mathbf{R}^{d_k \times T}$ 表示通过第 $k-1$ 层时序特征提取模块提取到的潜在特征表示; $\tilde{X}^0 = X \in \mathbf{R}^{(d+1) \times T}$ 表示原始的输入数据; H 表示基于GRU的特征提取操作。

具体而言, 第 k 层特征提取单元中的GRU操作可定义为

$$H^{k+1} = \text{GRUf}(\tilde{X}^k) \quad k = 1, 2, \dots, K \quad (9)$$

式中: $H^{k+1} = [h_1^{k+1}, h_2^{k+1}, \dots, h_T^{k+1}] \in \mathbf{R}^{d_{k+1} \times T}$ 为第 k 层GRU输出的整体隐藏状态表示, 蕴含输入数据的时序特征; $h_t^{k+1} \in \mathbf{R}^{d_{k+1}}$ 为第 k 层GRU对应第 $t(t=1, 2, \dots, T)$ 个时刻的隐藏状态输出。上述公式可进一步详细写为

$$\begin{cases} r_t^{k+1} = \sigma(\tilde{x}_t^k W_{xr}^{k+1} + h_{t-1}^{k+1} W_{hr}^{k+1} + b_r^{k+1}) \\ v_t^{k+1} = \sigma(\tilde{x}_t^k W_{xv}^{k+1} + h_{t-1}^{k+1} W_{hv}^{k+1} + b_v^{k+1}) \\ z_t^{k+1} = \tanh(\tilde{x}_t^k W_{xz}^{k+1} + r_t^{k+1} \odot (h_{t-1}^{k+1} W_{hz}^{k+1}) + b_z^{k+1}) \\ h_t^{k+1} = (1 - v_t^{k+1}) \odot h_{t-1}^{k+1} + v_t^{k+1} \odot z_t^{k+1} \end{cases} \quad (10)$$

式中: $W_{xr}^{k+1}, W_{hr}^{k+1}, W_{xv}^{k+1}, W_{hv}^{k+1}, W_{xz}^{k+1}, W_{hz}^{k+1}, b_r^{k+1}, b_v^{k+1}, b_z^{k+1}$ 为待学习的相关参数。

最后, 将每一层基于GRU提取的时序特征表

示 H^{k+1} 与基于一维卷积神经网络提取的静态特征 $\hat{X}^{k+1}$ 通过特征融合单元 F_{cd} , 具体操作是将二者级联, 再通过一层全连接层:

$$\begin{aligned} \tilde{X}^{k+1} &= F_{cd}(H^{k+1}, \hat{X}^{k+1}) = \\ &[H^{k+1}, \hat{X}^{k+1}] W_{hx}^k + b_{hx}^k \quad k = 1, 2, \dots, K \end{aligned} \quad (11)$$

式中: $W_{hx}^k \in \mathbf{R}^{(d_{k+1} + d_{k+1}') \times T}$ 和 $b_{hx}^k \in \mathbf{R}^T$ 为相关待学习参数, 主要用于控制特征连接后的维度, 并进一步提升泛化能力。

最后, $\tilde{X}^{K+1} = [\tilde{x}_1^{K+1}, \tilde{x}_2^{K+1}, \dots, \tilde{x}_T^{K+1}] \in \mathbf{R}^{d_{K+1} \times T}$ 即为通过 K 层特征提取单元得到的最终数据表示特征。

2.4 周期规律增强的电力负荷预测

本文设计的周期规律增强的电力负荷预测组件包含3个模块: 基于GRU的一般性时序历史编码、基于GRU-Skip的周期性时序历史编码和基于后期融合的电力负荷预测。

2.4.1 基于GRU的一般性时序历史编码

基于GRU的一般性时序历史编码旨在基于通过多视角特征学习组件得到的最终数据特征表示, 对已知的 T 个时刻的电力负荷数据进行一般性编码。具体而言, 该模块可定义为

$$[h_1^G, h_2^G, \dots, h_T^G] = \text{GRUg}(\tilde{X}^{K+1}) \quad (12)$$

式中: $h_t^G \in \mathbf{R}^{d_G}$ 表示基于GRUg获得的第 t 个时刻的隐藏状态表示。本文采用最后一个时刻的隐藏状态表示来表征时序历史信息。

2.4.2 基于GRU-Skip的周期性时序历史编码

由于电力负荷往往具有周期性特点, 比如电力负荷往往在晚间或周末达到峰值。因此, 电力负荷数据之间存在周期性的依赖关系, 而这些关系难以被GRU捕获。为解决该问题, 本文引入了GRU-Skip模块, 该模块已被文献[20]证明了其对周期性规律编码的有效性。该模块的具体操作可以通过式(13)定义:

$$\begin{cases} r_t^S = \sigma(\tilde{x}_t^K W_{xr}^S + h_{t-p}^S W_{hr}^S + b_r^S) \\ v_t^S = \sigma(\tilde{x}_t^K W_{xv}^S + h_{t-p}^S W_{hv}^S + b_v^S) \\ z_t^S = \tanh(\tilde{x}_t^K W_{xz}^S + r_t^S \odot (h_{t-p}^S W_{hz}^S) + b_z^S) \\ h_t^S = (1 - v_t^S) \odot h_{t-p}^S + v_t^S \odot z_t^S \end{cases} \quad (13)$$

式中: $h_t^S \in \mathbf{R}^{d_S}$ 表示基于GRU-Skip获得的第 $t(t=p+1, \dots, T)$ 个时刻的隐藏状态表示。从式(13)可以看出, 相较于传统的GRU网络, GRU-Skip使用 $p(1 < p < T)$ 个时刻之前的隐藏状态, 而不是上一时刻的隐藏状态, 作为当前时刻的输入。 p 为超参数, 表示电力负荷变化规律的周期。不同于一般性的时序历史编码(其默认时序变化周期为1), 在周期性时序历史编码中, 选用最后 p 个时刻(最后一个周期)

的隐藏层状态 $[\mathbf{h}_{T-p+1}^S, \mathbf{h}_{T-p+2}^S, \dots, \mathbf{h}_T^S] \in \mathbf{R}^{p \times d_s}$, 来全面表征整个历史 (T 个时刻) 中包含的周期性信息。

由于这最后一个周期的每一时刻所包含的周期性信息对于整个电力负荷历史采样序列的表征, 贡献程度不同, 采用了注意力机制 (attention mechanism, AM), 将 p 个时刻的周期性历史信息进行动态聚合。如式 (14) 所示:

$$\begin{cases} f(\mathbf{h}_j^S, \mathbf{h}_T^G) = \tanh(\mathbf{h}_j^S \mathbf{W}_{SG} \mathbf{h}_T^G) \\ \alpha_j = \frac{\exp(f(\mathbf{h}_j^S, \mathbf{h}_T^G))}{\sum_{l=T-p+1}^T \exp(f(\mathbf{h}_l^S, \mathbf{h}_T^G))} \\ j = T-p+1, T-p+2, \dots, T \end{cases} \quad (14)$$

式中: $\mathbf{W}_{SG} \in \mathbf{R}^{d_s \times d_g}$ 为待学习的参数。该设计的整体思想是参考一般性时序历史信息 $\mathbf{h}_T^G$, 对各个时刻的周期性历史表征分配相应权重。接着, 根据式 (14) 获得的注意力值 α_j , 对基于 GRU-Skip 获得的一系列周期性历史信息表征进行如下的加权融合:

$$\mathbf{h}^S = \sum_{j=T-p+1}^T \alpha_j \mathbf{h}_j^S \quad (15)$$

式中: $\mathbf{h}^S \in \mathbf{R}^{d_s}$ 表示基于 GRU-Skip 获得的最终表征向量。

2.4.3 基于后期融合的电力负荷预测

本文分别通过线性运算得到基于一般性时序历史和基于周期性时序历史的预测结果, 然后采用后期融合的方式, 对 2 种预测结果进行融合。具体如下如式 (16) 所示:

$$\hat{\mathbf{y}} = \mathbf{W}_G \mathbf{h}_T^G + \mathbf{W}_S \mathbf{h}^S \quad (16)$$

式中: $\mathbf{W}_G \in \mathbf{R}^{r \times d_g}$; $\mathbf{W}_S \in \mathbf{R}^{r \times d_s}$; $\hat{\mathbf{y}} \in \mathbf{R}^r$ 表示最终对于未来 r 个时刻的电力负荷预测结果。

针对模型优化, 本文采用均方误差作为损失函数, 即

$$L = \frac{1}{N} \sum_{i=1}^N \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|^2 + \frac{\lambda}{2N} \|\mathbf{w}\|^2 \quad (17)$$

式中: $\mathbf{y}_i \in \mathbf{R}^r$ 为第 i 个样本未来电力负荷的真实值; $\hat{\mathbf{y}}_i \in \mathbf{R}^r$ 为模型针对第 i 个样本未来电力负荷的预测值; N 为训练样本总个数; $\frac{\lambda}{2N} \|\mathbf{w}\|^2$ 为 L_2 正则化项, 防止过拟合问题, 提高模型的泛化能力, $\lambda (\lambda > 0)$ 为超参数, $\mathbf{w}$ 为权重参数向量。

3 实验结果与分析

3.1 数据集

本文选取的数据集为第九届“中国电机工程学会杯”全国大学生电工数学建模竞赛提供的标准数

据集。该数据集包含某一地区从 2012 年 1 月 1 日至 2015 年 1 月 10 日的电力负荷数据 (每 15 min 采样 1 次, 即 $D = 15$, 每日共计 96 个时刻)、气象因素数据 (包括日最高温度、日最低温度、日平均温度、日相对湿度及日降雨量)。与现有工作相同, 本文采纳的具体任务为: 根据一段时间的电力负荷和气象历史数据, 预测未来 1 天共计 96 个时刻的电力负荷瞬时值。针对模型验证, 本文通过随机采样的方式将该数据集进行如下分割: 80% 作为训练集, 10% 作为验证集, 10% 作为测试集。

3.2 实验设置

超参数设置: 本文将多视角特征学习组件中的特征提取层数 K 设为 5。因而, 有 5 个一维卷积层, 其卷积核数量分别设为 256、128、256、256、128; 卷积核大小分别设为 5、5、3、3、3; 对于 5 个 SE 模块, 其中的缩减比例 r_k 均设为 16。对于 5 层 GRUf, 其神经元个数分别设为 64、128、128、128、64; 而对于 5 层特征融合单元中全连接层的神经元个数, 分别设为 96、128、128、128、96。

周期规律增强的电力负荷预测组件中 GRUg 和 GRU-Skip 模块中的神经元个数均设为 64, GRU-Skip 模块中的超参数 p 设为 7。式 (16) 中的超参数 λ 设为 0.01。由于本文定义的电力负荷历史的时间窗口 M 是一个重要的超参数。本文采用网格搜索策略, 将该超参数分别设置为 7、14、21、28、30, 并开展了不同超参数设置下的实验。根据实验结果, 设 M 为 21。本文基于 Pytorch 框架实现的 EPISODE 方法。具体使用 RMSProp 优化器优化网络参数。学习速率被设置为 0.001。

评价指标: 由于电力负荷预测本质上是回归任务。因此, 本文采用回归任务中常用的评价指标: 平均绝对百分比误差指标 (mean absolute percentage error, MAPE) M_{MAPE} 和均方根误差 (root mean square error, RMSE) R_{MSE} 来评价模型。这 2 个评价指标的定义分别为

$$M_{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (18)$$

$$R_{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

式中: n 为待预测的电力负荷采样点总个数; y_i 为每一时刻电力负荷的真实值; $\hat{y}_i$ 为每一时刻电力负荷的预测值。

3.3 针对时间窗口大小的实验

将用于数据划分的时间窗口大小 M 分别设置为 7、14、21、28 和 30, 进而训练模型, 并在测试集

上得到相应的模型表现。这样设置的依据是考虑到以一周、周的倍数及一月的数据进行预测符合实际生产周期规律。如表 1 所示, 当时间窗口大小 $M=21$ 时, 即根据前 21 天的电力负荷和气象历史数据预测未来一天电力负荷的方案预测效果最好, 因此, 接下来得实验均以 $M=21$ 的时间窗口对数据进行划分条件, 进而训练模型。

表 1 EPISODE 在不同时间窗口的预测精度对比
Table 1 Comparison of prediction accuracy of EPISODE in different window

时间窗口	MAPE/%	RMSE/MW
$M=7$	1.523 4	165.724 9
$M=14$	1.399 3	112.493 8
$M=21$	1.011 5	81.246 0
$M=28$	1.305 3	115.563 7
$M=30$	1.140 0	100.872 8

为了直观显示不同时间窗口大小条件下的预测效果, 本文在测试集中随机选取一天作为待预测对象, 并根据相应的实验结果绘制图 5。根据图 5 中的曲线对比, 当时间窗口大小 $M=21$ 时, 预测效果最佳。因此, 接下来的实验均以 $M=21$ 的时间窗口对数据进行划分条件, 进而训练模型。

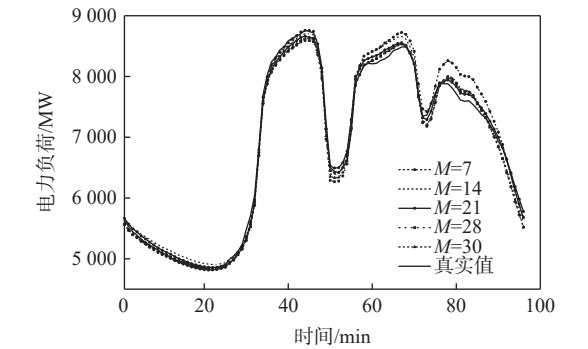


图 5 EPISODE 在不同时间窗口下的预测结果
Fig. 5 Prediction results of EPISODE in different windows

3.4 与基准方法对比试验

本文选择以下的基准方法进行比较, 来评估本文方法的有效性。本文选择的基准方法具体描述如下:

- 1) Attention-LSTM^[14]: 该方法将电力负荷历史数据输入 2 层 LSTM, 1 层注意力机制网络和 1 个全连接层, 对电力负荷进行预测。
- 2) FOMNN^[21]: 该方法分别使用多个卷积层和 GRU 模块来提取影响电力负荷的静态特征和时序特征, 然后通过注意力机制进行特征融合, 再通过 1 层 maxout 层和全连接层输出预测结果。
- 3) GRU-NN^[22]: 该方法先单独将历史负荷通过一层 GRU 网络, 再与其他特征拼接, 再通过多层深

度神经网络, 最后输出预测结果。

- 4) Dropout-ILSTM^[23]: 该方法通过将 2 层 LSTM 的输出与输入进行加权求和, 得到输出结果, 并利用 Dropout 方法防止过拟合。
- 5) Attention-BiLSTM-LSTM^[24]: 该方法先将所有输入依次通过 1 层双向 LSTM 和 1 层 LSTM, 然后将二者的输出通过注意力机制进行动态结合, 最后输出预测结果。
- 6) CNN-BiLSTM^[25]: 该方法将所有输入依次通过 2 层一维卷积层、1 层池化层和 1 层扁平层, 以提取静态特征, 再通过 1 层双向 LSTM 和全连接层, 最后输出预测结果。

表 2 为 EPISODE 方法和 6 个基准方法对于电力负荷预测的精度和测试时间对比结果。表格中汇报的时间为方法在测试集所有样本上运行的结果, 测试集共 91 组样本。

表 2 EPISODE 与基准方法在数据集上的预测精度和测试时间对比
Table 2 The comparison of EPISODE and baselines methods on prediction accuracy and test time

方法	MAPE/%	RMSE/MW	时间/s
EPISODE	1.011 5	81.246 0	0.295 3
Attention-LSTM	2.787 7	233.002 2	0.717 8
FOMNN	2.196 7	194.704 7	0.305 0
GRU-NN	6.244 9	472.842 9	0.762 4
Dropout-ILSTM	5.549 7	392.077 4	0.922 2
Attention-BiLSTM-LSTM	2.198 4	194.757 7	1.509 4
CNN-BiLSTM	4.179 2	422.531 1	2.718 1

由表 2 和图 6 可以看出 EPISODE 方法在数据集上的表现都取得了比其他方法更好的预测效果, 验证了 EPISODE 方法的有效性。特别地, EPISODE 性能优于同样挖掘静态特征与时序特征的基准方法 FOMNN, 这表明引入特征压缩与激励模块与周期性时序历史编码的重要性。

3.5 模型消融实验

为了验证 EPISODE 方法各个关键模块的有效性, 引入 3 个衍生模型: EPISODE_NoSE, EPISODE_NoGRU_Skip 和 EPISODE_NoAttention。

- 1) 对于模型 EPISODE_NoSE, 在 EPISODE 方法的基础上去掉了特征压缩与激励模块。
- 2) 对于模型 EPISODE_NoSE, 在 EPISODE 方法的基础上去掉了用于周期性时序历史编码的 GRU_Skip 模块, 并用传统的 GRU 代替。
- 3) 对于模型 EPISODE_NoAttention, 在 EPISODE 方法的基础上去掉了用于周期性时序历史表征学习的注意力机制, 转而使用全连接层代替。具体公

式如下:

$$\tilde{\mathbf{h}} = \mathbf{W}'_h \mathbf{h}_T^G + \left(\sum_{i=T-p+1}^T \mathbf{W}_i^S \mathbf{h}_i^S + \mathbf{b}_h \right) \quad (20)$$

式中: $\mathbf{W}'_h \in \mathbf{R}^{\tau \times d_G}$; $\mathbf{W}_i^S \in \mathbf{R}^{\tau \times d_S}$; $\mathbf{b}_h \in \mathbf{R}^{\tau}$; τ 为未来待预测的电力负荷采样点个数。

表3为EPISODE和3个衍生模型对电力负荷预测的精度对比结果。图7为EPISODE和3个衍生模型对测试集中某一天的电力负荷的预测结果。

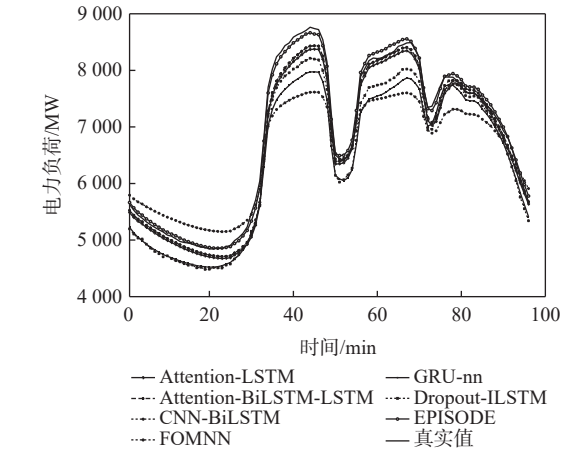


图6 EPISODE与基准方法的预测结果

Fig. 6 The load prediction results of EPISODE and baselines

表3 EPISODE与衍生模型的预测精度对比

Table 3 Comparison of prediction accuracy between EPISODE and derivations

模型	MAPE/%	RMSE/MW
EPISODE	1.011 5	81.246 0
EPISODE_NoSE	1.454 2	123.208 9
EPISODE_NoGRU_Skip	1.622 3	168.620 0
EPISODE_NoAttention	1.642 7	188.648 5

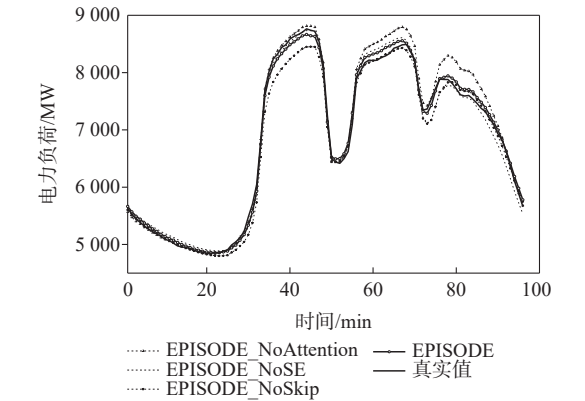


图7 EPISODE与衍生模型的预测结果

Fig. 7 The load prediction results of EPISODE and derivations

从表3和图7可以看出,EPISODE方法在数据集中的表现均高于3个衍生模型。这一实验结果

说明添加特征压缩与激励模块,GRU-Skip模块,以及基于注意力机制的周期性历史编码方法的必要性。这些模块设计有效增强了模型对输入数据的特征提取能力,有助于实现更加准确的电力负荷预测。

3.6 敏感度分析实验

对EPISODE方法进行了敏感度分析,主要对其的关键超参数:多视角特征提取单元层数和GRU-Skip模块中表征周期的参数 p ,进行了相应的敏感度分析实验。

3.6.1 多视角特征提取单元层数 K

本文将多视角特征提取单元层数 K 设置在[1,6]范围内,并以1为步长进行调节。如此设置是为了在保证模型尽可能轻量的前提下,实现精准的预测。图8为不同层数对电力负荷预测结果的影响。从图中可以看出,随着层数增加,预测精度也逐渐增加。但是当层数 $K \geq 5$ 时,预测精度开始下降,即当 $K=5$ 时,模型的预测精度最高。这说明,随着特征提取单元层数的增加,模型的泛化能力逐渐增强。但是当特征提取单元层数过大时,参数量过多,可能会发生过拟合现象。

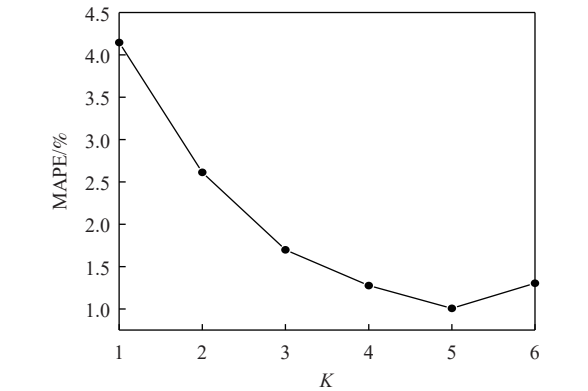


图8 特征提取单元层数的敏感性实验结果

Fig. 8 Sensitivity experiment results with different feature extraction layers

3.6.2 GRU-Skip模块中表征周期的参数 p

在该项敏感度实验中,将GRU-Skip模块中表征周期的参数 p (即跳过的时间步数)的取值分别设为5、7、10、14去训练模型。其中,取值为7和14是考虑到1周有7天,以7及7的倍数为1个周期符合现代生产规律。而取值为5和10则是分别对1周与2周天数的折中近似。图9为不同的 p 取值对电力负荷的预测结果的影响。从图中可以看出,模型的预测效果受 p 的取值影响较大,且在 p 为7时,模型预测值与真实值之间误差MAPE达到最小值,预测精度最高。这验证了结论:电力负荷数据具有一定的周期性特征,且利

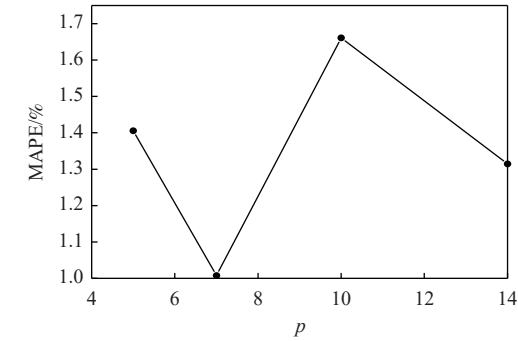


图 9 GRU-Skip 模块中 p 取值的敏感性实验结果

Fig. 9 Sensitivity experiments with different value of p in GRU-Skip module

用好周期性特征,将有助于提升预测的精度。

4 结 论

- 1) 本文方法可实现优异的短期电力负荷预测效果,可实现 1.0115% 的 MAPE 预测误差,远高于同类的基准方法。
- 2) 消融实验证明本文方法中的各个组件对精准的电力负荷预测效果起到重要作用。
- 3) 本文方法目前只考虑了电力负荷的历史数据和相关气象数据,未来拟引入更多影响因素,例如节假日数据、当地经济数据、以及配电网拓扑结构等,以实现更加准确的电力负荷预测。

参考文献 (References)

[1] 康重庆,夏清,张伯明. 电力系统负荷预测研究综述与发展方向的探讨[J]. 电力系统自动化, 2004, 28(17): 1-11.

KANG C Q, XIA Q, ZHANG B M. Review of power system load forecasting and its development[J]. Automation of Electric Power Systems, 2004, 28(17): 1-11 (in Chinese).

[2] KEMP F. Applied multiple regression/correlation analysis for the behavioral sciences[J]. Journal of the Royal Statistical Society: Series D (the Statistician), 2003, 52(4): 691-691.

[3] DI CAPRIO U, GENESIO R, POZZI S, et al. Short term load forecasting in electric power systems: A comparison of ARMA models and extended Wiener filtering[J]. Journal of Forecasting, 1983, 2(1): 59-76.

[4] LEE Y S, TONG L I. Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming[J]. Knowledge-Based Systems, 2011, 24(1): 66-72.

[5] H ERNÁNDEZ L, BALADRÓN C, AGUIAR J, et al. Artificial neural network for short-term load forecasting in distribution systems[J]. Energies, 2014, 7(3): 1576-1598.

[6] LDIN W M, TU C S, YANG R F, et al. Particle swarm optimisation aided least-square support vector machine for load forecast with spikes[J]. IET Generation, Transmission & Distribution, 2016, 10(5): 1145-1153.

[7] 吴倩红,高军,侯广松,等. 实现影响因素多源异构融合的短期负荷预测支持向量机算法[J]. 电力系统自动化 2016, 40(15): 67-72.

WU Q H, GAO J, HOU G S, et al. Short-term load forecasting sup-

port vector machine algorithm based on multi-source heterogeneous fusion of load factors[J]. Automation of Electric Power System, 2016, 40(15): 67-72 (in Chinese)

[8] 李婉华,陈宏,郭昆,等. 基于随机森林算法的用电负荷预测研究[J]. 计算机工程与应用, 2016, 52(23): 236-243.

LI W H, CHEN H, GUO K, et al. Research on electrical load prediction based on random forest algorithm[J]. Computer Engineering and Applications, 2016, 52(23): 236-243(in Chinese).

[9] 彭鹏,彭佳红. 基于多元线性回归模型的电力负荷预测研究[J]. 中国安全生产科学技术, 2011, 7(9): 158-161.

PENG P, PENG J H. Research on the prediction of power load based on multiple linear regression model[J]. Journal of Safety Science and Technology, 2011, 7(9): 158-161 (in Chinese).

[10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[11] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088), 533-536.

[12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

[13] CHUNG J, GULCEHRE C, CHO K, et al. Gated feedback recurrent neural networks[C]//International Conference on Machine Learning. Lile: PMLR, 2015: 2067-2075.

[14] 彭文,王金睿,尹山青. 电力市场中基于 Attention-LSTM 的短期负荷预测模型[J]. 电网技术, 2019, 43(5): 1745-1751.

PENG W, WANG J R, YIN S Q. Short-term load forecasting model based on attention-LSTM in electricity market[J]. Power System Technology, 2019, 43(5): 1745-1751 (in Chinese).

[15] 李龙,魏靖,黎灿兵,等. 基于人工神经网络的负荷模型预测[J]. 电工技术学报, 2015, 30(8): 225-230.

LI L, WEI J, LI C B, et al. Prediction of load model based on artificial neural network[J]. Transactions of China Electrotechnical Society, 2015, 30(8): 225-230 (in Chinese).

[16] 陆继翔,张琪培,杨志宏,等. 基于 CNN-LSTM 混合神经网络模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(8): 131-137.

LU J X, ZHANG Q P, YANG Z H, et al. Short-term load forecasting method based on CNN-LSTM hybrid neural network model[J]. Automation of Electric Power Systems, 2019, 43(8): 131-137 (in Chinese).

[17] 李泽文,胡让,刘湘,等. 基于 PCA-DBILSTM 的多因素短期负荷预测模型[J]. 电力系统及其自动化学报, 2020, 32(12): 32-39.

LI Z W, HU R, LIU X, et al. Multi-factor short-term load prediction model based on PCA-DBILSTM[J]. Proceedings of the CSU-EPSA, 2020, 32(12): 32-39 (in Chinese).

[18] 杨洋,谷震浩. 基于 XGBOOST-DNN 的中期电力负荷预测[J]. 计算机系统应用, 2021, 30(9): 186-191.

YANG Y, GU Z H. Mid-term power load forecasting based on XG-Boost-DNN[J]. Computer Systems & Applications, 2021, 30(9): 186-191 (in Chinese).

[19] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[EB/OL]. (2019-05-16) [2022-05-18]. <https://arxiv.org/abs/1709.01507.pdf>.

[20] LAI G K, CHANG W C, YANG Y M, et al. Modeling long- and short-term temporal patterns with deep neural networks[EB/OL].

(2018-04-18) [2022-05-18] .<https://arxiv.org/abs/1703.07015.pdf>.

[21] 庞昊, 高金峰, 杜耀恒. 基于多神经网络融合的短期负荷预测方法[J]. 电力自动化设备, 2020, 40(6): 37-43.
PANG H, GAO J F, DU Y H. Short-term load forecasting method based on fusion of multiple neural networks[J]. Electric Power Automation Equipment, 2020, 40(6): 37-43 (in Chinese).

[22] 王增平, 赵兵, 纪维佳, 等. 基于 GRU-NN 模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(5): 53-58.
WANG Z P, ZHAO B, JI W J, et al. Short-term load forecasting method based on GRU-NN model[J]. Automation of Electric Power Systems, 2019, 43(5): 53-58 (in Chinese).

[23] 刘皓琪, 高飞, 王耀力, 等. 基于 Dropout-ILSTM 网络的短期电力负荷预测[J]. 电测与仪表, 2021, 58(5): 105-111.
LIU H Q, GAO F, WANG Y L, et al. Short-term power load forecasting based on Dropout-ILSTM network[J]. Electrical Measurement & Instrumentation, 2021, 58(5): 105-111 (in Chinese).

[24] 龚飘怡, 罗云峰, 方哲梅, 等. 基于 Attention-BiLSTM-LSTM 神经网络的短期电力负荷预测方法[J]. 计算机应用, 2021, 41(S1): 81-86.
GONG P Y, LUO Y F, FANG Z M, et al. Short-term power load forecasting method based on Attention-BiLSTM-LSTM neural network[J]. Journal of Computer Applications, 2021, 41(S1): 81-86 (in Chinese).

[25] 朱凌建, 荀子涵, 王裕鑫, 等. 基于 CNN-Bi LSTM 的短期电力负荷预测[J]. 电网技术, 2021, 45(11): 4532-4539.
ZHU L J, XUN Z H, WANG Y X, et al. Short-term power load forecasting based on CNN-BiLSTM[J]. Power System Technology, 2021, 45(11): 4532-4539 (in Chinese).

Periodic pattern-enhanced multi-view short-term load prediction

SU Wei*, XIAO Xiaolong, SHI Mingming, FANG Xin, SI Xinyao

(State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing 210000, China)

Abstract: Short-term load prediction is essential to ensure the proper operation of the power system. The existing efforts have two limitations: lack of mining the dependencies between features and ignore the periodic pattern of load changes. To solve the above limitations, we propose periodic Pattern-enhanced Multi-view Short-term power load prediction nEtworks, dubbed EPISODE. The framework includes two core components: a multi-view feature learning component and a periodic pattern-enhanced load prediction component. The former aims to effectively extract static features and time series features to obtain enhanced feature representation; the latter is to perform general time series mining and periodic time series mining to obtain comprehensive historical feature representation. The combination of the two aforementioned qualities results in the realization of the short-term load forecast. Extensive experiments have been conducted on real-world datasets, and the experimental results demonstrate the superiority of our proposed method.

Keywords: periodic pattern-enhanced; short-term load prediction; feature squeeze and excitation; gated recurrent unit; attention mechanism