

DOI:10.19951/j.cnki.1672-9331.20230215001

文章编号:1672-9331(2023)06-0128-12

引用格式:岳首志,洪海生,邓祺,等.基于曲线特征聚类与信息聚合的电力负荷预测[J].长沙理工大学学报(自然科学版),2023,20(6):128-139.

Citation: YUE Shouzhi, HONG Haisheng, DENG Qi, et al. Power load forecasting based on curve feature clustering and information aggregation [J]. J Changsha Univ Sci Tech (Nat Sci), 2023, 20(6): 128-139.

# 基于曲线特征聚类与信息聚合的电力负荷预测

岳首志<sup>1,2</sup>, 洪海生<sup>2</sup>, 邓祺<sup>2</sup>, 许陈德<sup>2</sup>, 罗锋<sup>2</sup>

(1. 长沙理工大学 电气与信息工程学院, 湖南 长沙 410114;

2. 广东电网有限责任公司 广州供电局, 广东 广州 510620)

**摘要:**【目的】为获得准确可靠的超短期电力负荷预测结果以满足电力系统快速响应和实时调度的需要,考虑到电网负荷数据非线性、时序性等特征,提出一种基于曲线特征聚类与信息聚合的电力负荷超短期区间预测方法。【方法】首先,考虑负荷曲线的局部波动特征和整体趋势特征,将电力负荷曲线分为不同类别;然后,将高斯过程回归模型作为表征负荷整体趋势的预测模型,并将基于分位数的双向长短期记忆神经网络作为表征负荷局部波动的预测模型;最后,引入聚合思想,将Choquet积分算法作为聚合函数,对上述两种预测模型的结果进行聚合。【结果】所提预测方法有效实现了考虑多种特征的日负荷曲线的聚类;对单一模型的预测结果进行聚合,得到了不同场景下各置信度的区间预测结果。通过算例分析,所提预测方法的可靠性指标比上述两个单一预测模型的分别平均提高了14.70%、10.81%,综合性能分别平均提高了3.14%、15.55%。【结论】算例结果表明,与常见负荷概率预测方法和单一预测模型相比,所提方法在预测精度和可靠性上均有显著提高。此外,聚类方法和信息聚合思想的引入有助于预测模型精度的提升。

**关键词:**负荷区间预测;长短期记忆神经网络;K-means聚类;高斯过程回归;信息聚合

中图分类号:TM73

文献标志码:A

## 0 引言

随着科技与经济的不断进步,我国年耗电量逐年增长,这对电网的稳定、可靠运行提出了更高的要求<sup>[1]</sup>。准确的电力负荷预测是电力系统稳定运行的前提,提高电力负荷预测精度有利于合理安排电力系统运行模型及制定维护计划,从而降低电力系统运行成本,提高电力系统运行效益<sup>[2]</sup>。

根据预测形式的不同可将负荷预测分为确定性预测和概率性预测。确定性预测的结果为预测对象在未来时刻的単点期望值,而概率性预测则以概率分布或置信区间的形式给出未来时刻的预测结果,可有效实现对预测结果不确定性的量化分析<sup>[3]</sup>。与确定性预测相比,概率性预测能够为电力系统提供更丰富、准确的不确定性信息。

近年来,国内外众多学者对负荷预测问题展开了广泛研究。罗凤章等<sup>[4]</sup>构建了基于卷积神经网络(用于提取数据的隐含特征)与支持向量回归的负荷预测模型,并得出了较为精确的预测结果。李丹等<sup>[5]</sup>提出了基于约束并行长短期记忆(long and short-term memory, LSTM)神经网络的分位数回归(quantile regression, QR)预测方法,其将LSTM神经网络与分位数相结合,并考虑分位数预测值之间的约束关系,得到了较好的概率预测结果。张淑清等<sup>[6]</sup>结合优化变分模态分解与双向长短期记忆(bidirectional long and short-term memory, BiLSTM)神经网络,提出了一种利用BiLSTM神经网络挖掘数据时序特征的两阶段负荷预测方法,并通过算例验证了该方法的有效性。黄南天等<sup>[7]</sup>在预测前引入场景生成方法,使场景特征表达得更为充分,从而进一步提高了预测模

收稿日期:2023-02-15;修回日期:2023-03-10;接受日期:2023-03-10

基金项目:中国南方电网资助项目(GZHKJXM20200037)

通信作者:岳首志(1993—)(ORCID:0009-0006-1477-7832),男,助理工程师,主要从事电力系统大数据分析方面的研究。E-mail:1411342870@qq.com

投稿网址: <http://csjgxbzk.csust.edu.cn/cslgdxxbzk/home>

型的精度。肖白等<sup>[8]</sup>提出了基于多级聚类与支持向量机的负荷预测方法,该方法根据各元胞的属性对元胞负荷进行多级聚类,得到了较精确的预测结果。上述研究成果说明,聚类分析可以将不同规律的数据区分开,进而有助于提高预测精度。牛东晓等<sup>[9]</sup>提出了一种有效的组合预测模型,该模型能够较精确地实现电力负荷的短期预测,算例表明,与组合前的单一模型相比,组合模型的预测精度得到了进一步提高。刘升伟等<sup>[10]</sup>通过K-means算法提取数据特征,并利用改进高斯过程回归算法建立负荷区间预测模型,得到了较精确的区间预测结果。

在预测问题中,预测误差是不可避免的,且具有一定的不确定性,而后处理方法作为改进初始预测的常用方法,其通过某种方法对初始预测进行调整来提高总体预测精度<sup>[11]</sup>。此外,不同预测模型表现出的特性有所不同,且各模型预测结果存在差异,若能利用不同模型的优势,使相互冲突的预测结果之间达到最大的一致性,则可进一步提高预测精度。因此,信息聚合思想非常适合被引入后处理方法中,并运用到本文的预测模型中,与此同时将基于模糊测度的Choquet积分作为聚合函数。

待聚合的预测模型应具有各自的预测优势。本文将电力负荷曲线的特征分为局部波动和整体趋势两种,需要选择合适的预测模型对这两种不同特征进行预测。BiLSTM神经网络模型擅长挖掘时间序列的特征规律,可以用来学习负荷序列的变化规律并预测负荷曲线的局部波动。而高斯过程回归(Gaussian process regression, GPR)模型可以学习历史负荷的分布规律,实现对负荷曲线整体趋势特征的预测。同时,电力负荷日曲线具有随机性和波动性,仅用一种预测模型映射所有负荷曲线的变化规律会导致模型训练不够理想、预测结果精度较低等问题。因此,应考虑先将负荷曲线按一定规律进行分类,这样可使某个聚类场景下的负荷曲线具有一致的变化趋势。

综上所述,本文提出基于曲线特征聚类与信息聚合的电力负荷超短期区间预测方法,通过整体趋势及局部波动特征对日负荷曲线加以描述,并通过基于动态弯曲(dynamic time warping, DTW)距离的K-means算法进行聚类。对于某个聚类场景,通过GPR模型拟合该场景日内各采样

点的负荷分布规律,将GPR模型作为侧重负荷曲线整体趋势的预测模型,并用日气象数据作为输入特征。同时,将擅长处理时间序列问题,能够深度挖掘序列规律的基于分位数的双向长短期记忆(QR-BiLSTM)神经网络模型作为侧重负荷曲线局部波动特征的预测模型。将以上两种具有不同侧重的模型通过Choquet积分算法进行信息聚合,得到最终的区间预测结果。与传统的预测模型相比,本文所提方法在预测精度及可靠性等方面均有提高。

本文所作的主要贡献如下:

1) 用3种表征曲线不同特性的相似度距离代替K-means算法中原始的欧式距离,有效实现了基于曲线特征的电力负荷日曲线的聚类。

2) 将信息聚合思想引入预测方法中,建立了本文的区间预测组合模型,充分发挥了各单一模型聚合前的优势。

3) 通过算例分析,验证了曲线聚类以及信息聚合方法的有效性;将本文所提方法与其他预测模型做对比分析,证明本文所提方法具有较高的精度。

## 1 兼顾多特征的日负荷曲线的聚类

K-means算法作为聚类问题的常用方法,有着原理简单、易于实现、收敛速度快等优点<sup>[12]</sup>。传统的K-means算法采用欧式距离来衡量两样本之间相似度的大小,但对于本文的负荷曲线聚类问题,欧式距离只能反映两样本所对应采样点的数值差异,不足以全面描述两样本之间的相似度。因此,本文引入DTW距离来更有效地衡量两样本间的相似度,从而描述曲线的多种特征。

### 1.1 DTW距离

DTW距离是一种相似度距离,可以衡量两个序列形态的相似性。其运用动态规划技术,将两个序列进行相似性匹配,并通过欧式距离得到度量结果。对任意两个时间序列 $X_D = \{x_{D_1}, x_{D_2}, \dots, x_{D_n}\}$ 与 $Y_D = \{y_{D_1}, y_{D_2}, \dots, y_{D_m}\}$ ,其DTW距离的计算步骤如下:

构建尺寸为 $n \times m$ 的矩阵 $D$ ,其元素 $d_{ij}$ 为:

$$d_{ij} = d_{ED}(x_i, y_j), 1 \leq i \leq n, 1 \leq j \leq m \quad (1)$$

式中: $d_{ED}()$ 为距离计算函数,本文采用欧式距离。

对于序列 $X_D, Y_D$ ,当 $n=m$ 时,其欧式距离的计

算公式为:

$$d_{ED}(X_D, Y_D) = \sqrt{\sum_{i=1}^n (x_{D_i} - y_{D_i})^2} \quad (2)$$

动点最初在  $d_{11}$  元素的位置,通过移动使其到达与其相邻元素的位置,如  $d_{12}$ 、 $d_{22}$ 、 $d_{21}$ ,并构建集合  $P_D$  来存放该动点所有移动路径的点坐标,记为  $P_D = \{p_{D_1}, p_{D_2}, \dots, p_{D_e}, \dots, p_{D_v}\}$ 。对于在元素  $d_{ij}$  位置的动点,其下一次移动到达的元素的位置只可能是  $d_{i+1,j}$ 、 $d_{i-1,j+1}$ 、 $d_{i,j+1}$  元素中某一个元素的位置。重复移动该动点直至动点到达元素  $d_{n,m}$  为止。

在动点移动过程中,需要满足以下目标函数,使其总的移动路径最短:

$$d_{DTW}(X_D, Y_D) = \min \sum_{e=1}^v D(p_{D_e}) \quad (3)$$

式中:  $D(p_{D_e})$  为第  $e$  个点的路径距离;  $d_{DTW}(X_D, Y_D)$  为时间序列  $X_D$  与  $Y_D$  的 DTW 距离。

## 1.2 电力负荷的日曲线特征

影响电力负荷的因素较多,如用户行为习惯、天气因素等,因此日负荷曲线有着较强的规律性,同时也伴随着随机波动,其曲线特征主要包括整体趋势特征以及局部波动特征<sup>[13]</sup>。本文使用欧式距离来衡量两样本所对应采样点之间距离的大小,通过式(2)进行计算,记为  $D_1(X_D, Y_D)$ 。负荷曲线的整体趋势特征与其规律性相匹配,能够反映曲线在时间维度上的总体变化趋势。本文采用 DTW 距离作为两序列的相似度距离,并通过式(3)进行计算,记为  $D_2(X_D, Y_D)$ 。局部波动特征可以反映曲线在各采样点上的波动程度。本文将序列  $X_D$ 、 $Y_D$  通过前项差分的方法得到能够映射波动情况的两序列  $X'_D$ 、 $Y'_D$ ,再通过式(3)计算其 DTW 距离来描述局部波动特征,并记为  $D_3(X_D, Y_D)$ 。

欧式距离只能计算两曲线相同采样点的数值距离,对于衡量负荷曲线的相似度,DTW 距离能够很好地弥补欧式距离的缺陷。图 1 所示为日负荷曲线的特征对比,图中,  $A'$ 、 $B'$ 、 $C'$  为 3 条归一化后的日负荷曲线,每条曲线有 96 个采样点。从图 1 可以看出,  $A'$  曲线的形态与  $B'$  曲线的非常相似,而  $C'$  曲线与  $A'$ 、 $B'$  曲线在形态与数值上都相差较大。经计算,  $A'$  曲线与  $B'$  曲线、 $A'$  曲线与  $C'$  曲线的欧式距离分别为  $D_1(A', B')=0.689 0$ 、 $D_1(A', C')=0.667 7$ ,两者相差无几,可见,欧式距离不能明显区分一些形态差别大的曲线;其 DTW 距离分别为  $D_2(A', B')=0.355 9$ 、 $D_2(A', C')=2.141 3$ ,两者相差较大,说明

DTW 距离对形态变化较为敏感。因此,DTW 距离可以有效度量负荷曲线间动态特性的差异。

为了有效区分负荷曲线,本文充分考虑曲线的多种特征,采用了一种兼顾曲线数值分布特征、形态趋势特征以及局部波动特征的相似度距离,计算公式如下:

$$D_{total}(X_D, Y_D) = r_1 D_1(X_D, Y_D) + r_2 D_2(X_D, Y_D) + r_3 D_3(X_D, Y_D) \quad (4)$$

式中:  $r_1$ 、 $r_2$ 、 $r_3$  分别为  $D_1(X_D, Y_D)$ 、 $D_2(X_D, Y_D)$ 、 $D_3(X_D, Y_D)$  对应的权重,可采用熵权法<sup>[11]</sup>确定。

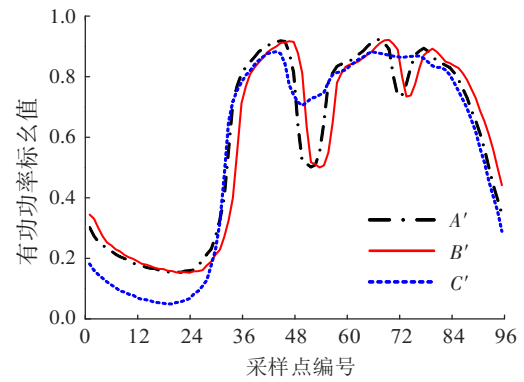


图 1 日负荷曲线特征对比

Fig. 1 Comparison of daily load curve characteristics

## 1.3 K-means 聚类算法

在使用用于曲线聚类的 K-means 算法时需要输入向量集合与聚类个数  $K$ 。由于人为给定的  $K$  值有较强的主观性,本文采用戴维森堡丁指标 (Davies-Bouldin index, DBI) 来衡量聚类效果,记为  $I_{DBI}$ ,并给出  $I_{DBI}$  最好时的聚类个数作为  $K$  值,  $I_{DBI}$  的计算公式为:

$$I_{DBI} = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left( \frac{\bar{S}_i + \bar{S}_j}{\|H_i - H_j\|_2} \right) \quad (5)$$

式中:  $H_i$  为第  $i$  个聚类中心;  $\|\cdot\|_2$  表示二范数;  $\bar{S}_i$  为第  $i$  个类内样本到聚类中心的平均距离,代表类内各曲线的分散程度,计算公式为:

$$\bar{S}_i = \sqrt{\frac{1}{R_i} \sum_{l=1}^{R_i} |G_{i,l} - H_i|^2} \quad (6)$$

式中:  $R_i$  为第  $i$  类样本的个数;  $G_{i,l}$  为第  $i$  类中第  $l$  个样本。

基于  $D_{total}$  距离的 K-means 算法的步骤如下:

- 1) 输入需要聚类的等长曲线样本数据集  $S = \{s_1, s_2, \dots, s_i, \dots, s_h\}$ , 其中  $s_i$  为第  $i$  条曲线样本。
- 2) 在  $S$  中随机选取  $K$  个不同的样本作为  $K$ -



means的初始聚类中心。

3) 遍历所有样本,通过式(4)计算第*i*个样本与所有聚类中心之间的距离,并将其归入与该样本之间距离最小的一类中。

4) 遍历所有类别,重新计算该类别的聚类中心,若当前聚类中心相较于上一次聚类有所改变,则重复步骤3)、4),直至聚类中心不再发生改变,则输出聚类结果。

## 2 日内局部波动概率预测模型

LSTM神经网络模型有着独特的循环架构和存储单元,可以捕捉时间序列自身的变化规律,在长时间序列预测中有着较好的效果<sup>[14]</sup>。而负荷曲线具有时间序列的特点,因此,本文采用基于LSTM的深度学习模型来预测负荷的局部波动规律。该预测模型作为待聚合预测模型的一部分,能够侧重于负荷时间序列的局部波动来进行预测。

### 2.1 BiLSTM神经网络的基本原理

LSTM神经网络是由循环神经网络(recurrent neural network, RNN)变化而来的,可以有效储存网络的前序输入信息,从而更好地实现时间序列预测。另外,其特殊的门控单元的引入还有效避免了RNN在反向传播时的梯度消失问题。

图2为LSTM神经网络原理图,图中 $x_t$ 、 $h_t$ 和 $C_t$ 分别为*t*时刻的网络输入、隐含层输出和单元状态, A为LSTM神经网络单元,  $V$ 为解释变量,  $U$ 表示在时间维度上的循环过程。*t*时刻LSTM神经网络的输入除了 $x_t$ 外,还有*t*-1时刻的隐含层输出 $h_{t-1}$ 和单元状态 $C_{t-1}$ ,由此可实现利用前序信息的目的。LSTM神经网络单元主要由遗忘门( $f_t$ )、输入门( $i_t$ )与输出门( $o_t$ )构成,遗忘门负责剔除过去相关性不强的信息,输入门将当前输入的 $x_t$ 的部分信息更新到单元状态中,输出门可以控制下一时刻隐含层的输出 $h_t$ 。3种门的共同作用可以使LSTM神经网络有效学习序列规律。其计算公式分别为:

$$i_t = \sigma(W_i[x_t, h_{t-1}]^T + b_i) \quad (7)$$

$$f_t = \sigma(W_f[x_t, h_{t-1}]^T + b_f) \quad (8)$$

$$o_t = \sigma(W_o[x_t, h_{t-1}]^T + b_o) \quad (9)$$

$$\tilde{C}_t = \tanh(W_c[x_t, h_{t-1}]^T + b_c) \quad (10)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (11)$$

$$h_t = o_t \odot \tanh C_t \quad (12)$$

式(7)~(12)中: $W_i$ 、 $W_f$ 、 $W_o$ 分别为输入门、遗忘门、输出门的输入权重矩阵; $b_i$ 、 $b_f$ 、 $b_o$ 分别为输入门、遗忘门、输出门的偏置向量; $b_c$ 为中间偏置向量; $\tilde{C}_t$ 为中间单元状态; $\sigma()$ 为sigmoid激活函数; $\odot$ 表示矩阵对应的元素相乘。

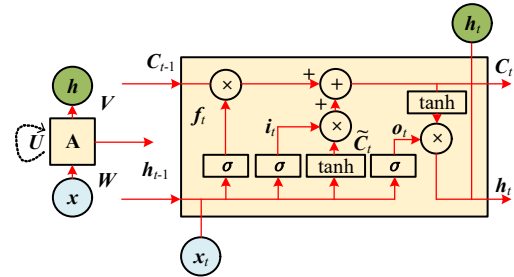


图2 LSTM神经网络原理图

Fig. 2 LSTM neural network schematic diagram

在LSTM神经网络的训练过程中,数据的输入顺序是从前向后的,这会出现难以最大化利用数据序列前后变化规律的问题。而BiLSTM神经网络基于LSTM神经网络原理构建了一个双向循环神经网络,即前向LSTM神经网络与后向LSTM神经网络,它们的共同作用使BiLSTM神经网络能够同时学习序列的前后变化规律,进而使得LSTM神经网络的性能得到有效提升。BiLSTM神经网络的结构如图3所示。

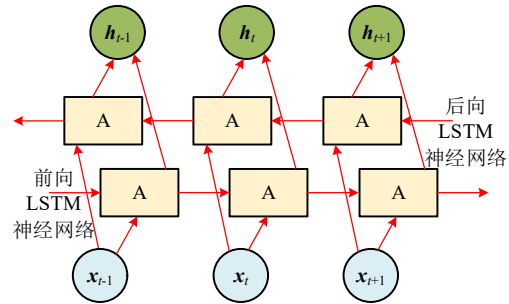


图3 BiLSTM神经网络结构图

Fig. 3 BiLSTM neural network structure diagram

### 2.2 基于分位数回归的BiLSTM神经网络概率预测模型

分位数回归方法能够确定解释变量 $V$ 与响应变量 $M$ 的条件分位数之间的关系,其回归模型为:

$$Q_M(\tau|V) = V'\beta(\tau) = \beta_0(\tau) + \sum_{i=1}^c \beta_i(\tau)V_i \quad (13)$$

式中: $c$ 为解释变量 $V$ 的个数; $V'$ 为 $\tau$ 对应的解释变量; $Q_M(\tau|V)$ 为响应变量 $M$ 的第 $\tau$ 个条件分位数;

$\beta(\tau)$ 为第 $\tau$ 个条件分位数对应的回归系数向量,可以通过求解下式的优化问题得到:

$$\min_{\beta(\tau)} \sum_{i|M_i \geq V'\beta(\tau)} \tau |M_i - V'\beta(\tau)| + \sum_{i|M_i < V'\beta(\tau)} (1 - \tau) |M_i - V'\beta(\tau)| \quad (14)$$

由于需要通过 BiLSTM 神经网络得到概率预测结果,因此在网络最后加入分位数回归层,其 $\tau$ 分位数对应的响应变量为:

$$M(\tau) = f\left(\sum_{j=1}^H W_j(\tau) h_j(\tau) + b(\tau)\right) \quad (15)$$

式中: $H$ 为隐含层的单元个数; $f$ 为输出层激活函数; $h_j(\tau)$ 为 LSTM 神经网络隐含层的输出; $W_j(\tau)$ 、 $b(\tau)$ 为输出层的权重与偏置项。BiLSTM 神经网络的训练过程实质上是其网络参数 $W_j(\tau)$ 、 $b(\tau)$ 的求解更新过程。

### 3 日内整体趋势概率预测模型

日内整体趋势概率预测模型作为待聚合预测模型的一部分,能够侧重于负荷时间序列整体变化趋势的预测。在通过本文的聚类方法计算得到 $K$ 类日负荷曲线后,由于日内整体趋势特征可以反映某聚类场景下负荷的整体变化规律,因此利用 GPR 模型学习输入(日气象数据)与输出(日内每个采样点的真实负荷数值)间的映射关系,可得到能反映该聚类场景下日负荷的整体变化趋势,及与日气象数据相匹配的区间预测结果。

GPR 模型是一种基于核的非线性非参数回归方法,能够通过训练得到一个描述输入特征数据与输出数据之间关系的映射函数,在处理高维度非线性问题时具有易于实现、参数自适应及输出结果为概率形式等优点<sup>[15]</sup>。

对于包含 $n$ 个样本的训练集,输入特征矩阵为 $X=[x_1, x_2, \dots, x_n]^T$ , $x_i$ 为第 $i$ 个样本的输入向量,其长度为 $d$ ,对应的输出响应为 $y=[y_1, y_2, \dots, y_n]^T$ , $y_i$ 为第 $i$ 个输出响应值。定义 $f(X)$ 为 $x_i$ 对应的函数 $f(x_1), f(x_2), \dots, f(x_n)$ 所构成的随机变量的集合,且服从联合高斯分布。一个高斯过程可表示为:

$$f(x) \sim G(m(x), k(x, x')) \quad (16)$$

式中: $x, x'$ 均为随机变量; $m(x)$ 为均值函数; $k(x, x')$ 为协方差函数。 $m(x)$ 和 $k(x, x')$ 的计算公式为:

$$m(x) = E(f(x)) \quad (17)$$

$$k(x, x') = E((f(x) - m(x)) \cdot (f(x') - m(x'))) \quad (18)$$

为方便推导,设 $m(x)=0$ ,即在没有任何观测值时函数期望为0。协方差函数用于量化样本间的相似度,当输入样本的相似度越高时,输出值可能越相近<sup>[16]</sup>。

回归问题的数学模型可表示为:

$$y_i = f(x_i) + \varepsilon_i \quad (19)$$

式中: $\varepsilon_i \sim N(0, \sigma_d^2)$ ,即 $\varepsilon_i$ 服从均值为0、方差为 $\sigma_d^2$ 的高斯白噪声。

由此可以得到输出观测值 $y$ 的先验分布:

$$y \sim G(0, K(X, X) + \sigma_n^2 I_n) \quad (20)$$

式中: $K(X, X)$ 为 $n$ 阶协方差矩阵,其元素 $k_{ij}=k(x_i, x_j)$ ;  $I_n$ 为 $n$ 阶单位矩阵。

已知测试集的输入 $X^*=[x_{n+1}, \dots, x_{n+m}]$ ,将第 $i$ 个对应向量记为 $x_i^*$ ,测试集输出的预测值为 $y^*$ ,则训练集观测值 $y$ 与预测值 $y^*$ 的联合先验分布为:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X, X) + \sigma_n^2 I_n & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right) \quad (21)$$

由此可计算预测值 $y^*$ 的后验分布为:

$$y^*|X, y, X^* \sim N(\bar{y}^*, \text{cov}(\bar{y}^*)) \quad (22)$$

式中: $\bar{y}^*$ 、 $\text{cov}(\bar{y}^*)$ 分别表示测试集输出的预测值的均值与方差。该均值对应其确定性预测结果,结合方差可确定置信区间。

## 4 基于信息聚合的区间预测模型

单个模型的预测性能在一定程度上取决于其预测方式及输入数据的特征。每个模型都有自身的优势和局限性,本文所提出的基于信息聚合的区间预测模型对日内局部波动概率预测模型和日内整体趋势概率预测模型的预测结果进行聚合,并通过后处理方式有效减小预测误差,从而提升预测性能。

### 4.1 Choquet 积分聚合函数

信息聚合是组合多个来源提供的不同信息片段以获得统一决策的过程<sup>[17]</sup>。在本文的预测过程中,通过聚合函数将两个模型的预测区间聚合为一个区间。聚合函数使用基于模糊测度的 Choquet 积分,模糊测度代表待聚合元素间的相关程度。

对于由待聚合的 $u$ 个元素组成的集合 $A=\{a_1,$

$a_2, \dots, a_u\}$ , 其聚合后的数值  $v_c$  为:

$$v_c = g(a_1, a_2, \dots, a_u) \quad (23)$$

式中:  $g(\cdot)$  表示聚合函数, 即  $g: [0, 1]^u \rightarrow [0, 1]$ , 当且仅当同时满足以下条件时成立:

1) 边界条件:  $g(0_1, 0_2, \dots, 0_u) = 0$ , 且  $g(1_1, 1_2, \dots, 1_u) = 1$ ;

2) 单调性: 若  $x_{c_i} \leq y_{c_i}, \forall i \in \{1, \dots, u\}$  有  $g(x_{c_1}, \dots, x_{c_u}) \leq g(y_{c_1}, \dots, y_{c_u})$ .

定义1(模糊测度)<sup>[17]</sup>: 对于参考集  $Q = \{1, \dots, u\}$ ,  $2^Q$  为  $Q$  的幂集,  $J, E$  为两个集合, 若函数  $\mu: 2^Q \rightarrow [0, 1]$  满足以下性质, 则称其为模糊测度。

$\forall J, E \in Q$ :

1) 边界条件:  $\mu(\emptyset) = 0, \mu(X_c) = 1$ ;

2) 单调性: 若  $J \subseteq E$  有  $\mu(J) \leq \mu(E)$ 。

式中:  $X_c$  为模糊测度输入变量的最大值。

本文使用聚合值集基数的幂均值作为模糊测度(参考文献[18], 本文  $q$  值取为 2), 其计算公式为:

$$\mu(P) = \left( \frac{|P|}{u} \right)^q, q > 0 \quad (24)$$

式中:  $P$  为模糊测度  $u$  的输入变量。

对于  $A = \{a_1, a_2, \dots, a_u\}$ , 基于模糊测度  $m$  的离散 Choquet 积分  $F: [0, 1]^u \rightarrow [0, 1]$ , 其计算公式为:

$$F_\mu(a) = \sum_{i=1}^u (a'_i - a'_{i-1}) \cdot \mu(P_i) \quad (25)$$

式中:  $a'_i$  为集合  $\{a'_1, a'_2, \dots, a'_u\}$  中的第  $i$  个元素, 该集合是通过  $\{a_1, a_2, \dots, a_u\}$  按递增规律重新排列得到的新集合;  $P_i = \{x'_i, \dots, x'_u\}$  为  $J$  中最大的  $u-i+1$  个元素。

#### 4.2 区间预测模型

为了综合考虑负荷曲线的多种特征及各预测模型的优势, 本文提出一种基于曲线特征聚类与信息聚合的电力负荷区间预测方法。图4为本文所提方法的预测流程图。首先, 将归一化后的负荷曲线数据通过基于多距离的  $K$ -means 算法进行聚类, 得到  $K$  类曲线的集合结果, 对于第  $i$  类结果, 将其与所对应的气象数据分为训练集与测试集, 并分别输入日内局部波动概率预测模型与日内整体趋势概率预测模型进行训练, 得到参数更新完毕的两种模型。然后, 将两种模型的测试集特征值分别输入, 得到  $t$  分位数下两种模型的区间上下限, 日内局部波动模型预测区间上下限分别记为

$P_{1t}^{(\bar{\tau})}, P_{1t}^{(\underline{\tau})}$ , 日内整体趋势模型预测区间上下限分别记为  $P_{gt}^{(\bar{\tau})}, P_{gt}^{(\underline{\tau})}$  ( $t$  表示预测日的采样时刻点)。最后, 对两模型的预测区间上限与区间下限分别进行信息聚合, 计算出聚合后的预测区间上限  $y_t^{(\bar{\tau})}$  和区间下限  $y_t^{(\underline{\tau})}$ , 计算公式如下:

$$\begin{cases} y_t^{(\bar{\tau})} = F_\mu([P_{1t}^{(\bar{\tau})}, P_{gt}^{(\bar{\tau})}]) \\ y_t^{(\underline{\tau})} = F_\mu([P_{1t}^{(\underline{\tau})}, P_{gt}^{(\underline{\tau})}]) \end{cases} \quad (26)$$

式中:  $F_\mu(\cdot)$  为 Choquet 积分的计算函数。

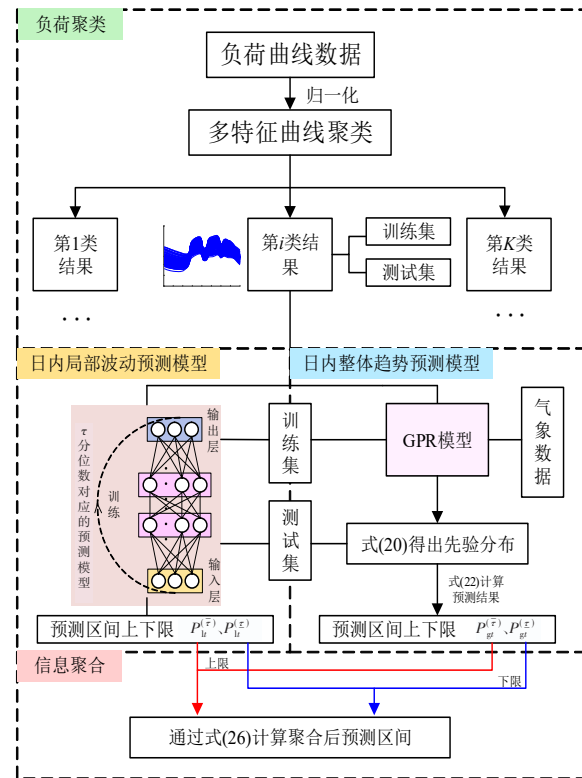


图4 本文所提方法的预测流程图

Fig. 4 Prediction flow chart of the method in this paper

## 5 算例分析

### 5.1 仿真数据及评价指标

本文采用某地区2012年整年的电力负荷数据, 其分辨率为15 min, 即每天包含96个采样点。除负荷数据外, 可用的气象数据包括日最高温度( $^{\circ}\text{C}$ )、日最低温度( $^{\circ}\text{C}$ )、日平均温度( $^{\circ}\text{C}$ )、日相对湿度、日降雨量(mm), 以上数据的分辨率均为24 h, 即每天只有一个数据点。在将数据输入模型之前, 先对其中的空值进行插值替换, 并对所有数据进行归一化处理, 计算公式为:



$$X'_{data,i} = \frac{X_{data,i} - \min(X_{data})}{\max(X_{data}) - \min(X_{data})} \quad (27)$$

式中:  $X_{data}$  为待归一化数据向量;  $X_{data,i}$  为第  $i$  个待归一化数据。

评价指标是衡量预测精度的统一标准。对于区间预测结果,用于确定性预测的两点做差式评估方法难以实现对区间预测精度的有效评估,区间预测需考虑可靠性、综合性能等<sup>[19]</sup>。因此,本文用以下3种指标对模型的预测效果进行量化评估:

1) 区间经验覆盖率(empirical coverage probability, ECP)  $I_{ECP}^{(\beta)}$ 。其代表真实值出现在预测区间的概率大小,其值越大说明模型预测区间的可靠性越高。 $I_{ECP}^{(\beta)}$ 的计算公式为:

$$I_{ECP}^{(\beta)} = \frac{1}{T} \sum_{t=1}^T \left[ \frac{1}{1-\bar{\alpha}} (y_t - y_t^{(\tau)}) I(y_t - y_t^{(\tau)}) + \frac{1}{\bar{\alpha}} (y_t^{(\tau)} - y_t) I(y_t^{(\tau)} - y_t) + (y_t^{(\tau)} - y_t^{(\tau)}) \right] \quad (30)$$

式中:  $\bar{\alpha}$  为置信区间对应的分位数值。

3) 平均宽度  $I_A^{(\beta)}$ 。在保证可靠性一定的情况下,平均宽度越小,则锐度性能越好。

$$I_A^{(\beta)} = \frac{1}{T} \sum_{t=1}^T (y_t^{(\tau)} - y_t^{(\tau)}) \quad (31)$$

## 5.2 算例仿真与分析

### 5.2.1 单一模型的区间预测特性分析

通过式(5)对不同聚类个数计算  $I_{DBI}$ ,得出在聚类个数为3时的  $I_{DBI}$ 最优。不同聚类个数对应的  $I_{DBI}$ 见图5,根据图5可设置  $K$  值为3。3种距离对应的权重  $r_1, r_2, r_3$  分别为0.453 8、0.324 8、0.221 4。对所有归一化后的日负荷曲线数据进行聚类,并得到3种聚类结果,如图6所示。将每类负荷曲线的前80%作为训练集用于训练模型,后20%作为测试集用于评估模型性能。

对于QR-BiLSTM模型,其优势在于对负荷序列前后时序具有较强的记忆性。因此,本文用QR-BiLSTM模型作为侧重局部波动规律的负荷区间预测模型,并用其训练输入特征以作为预测时刻的前1、2、3个采样点的负荷值,从而更好地捕捉负荷曲线的局部动态特征。QR-BiLSTM模型和GPR模型在不同置信度下的评价指标见表1。图7所示为QR-BiLSTM模型在测试集中某日的负荷区间的预测结果。结合评价指标可知,QR-BiLSTM模型的预测结果在各置信度下的区间宽度整体较窄,但其  $I_{ECP}^{(\beta)}$  值比GPR模型的小,说明QR-BiLSTM模型的预测结果有着较高的锐度性能和相对较低的可靠性,能够体现出该预测模型对

$$I_{ECP}^{(\beta)} = \frac{1}{T} \sum_{t=1}^T I(y_t^{(\tau)} - y_t) I(y_t - y_t^{(\tau)}) \quad (28)$$

$$I(\theta) = \begin{cases} 1, & \theta < 0 \\ 0, & \theta \geq 0 \end{cases} \quad (29)$$

式中:  $T$  为预测区间采样点的总数;  $y_t^{(\tau)}$  为预测区间下限对应的  $t$  采样点的数值,  $y_t^{(\tau)}$  为预测区间上限对应的  $t$  采样点的数值;  $y_t$  为真实负荷曲线对应的  $t$  时刻的数值;  $\beta$  为预测区间对应的置信度。

2) Winker分数  $I_W^{(\beta)}$ 。其作为区间预测的常用评价指标,能综合考虑区间特性来反映预测结果的综合性能,其值越小表示预测结果综合性能越好。 $I_W^{(\beta)}$ 的计算公式为:

于负荷曲线的波动较为敏感,且预测区间能够匹配负荷的波动规律。

GPR模型的优势在于对负荷数据整体分布规律的提取,因此,本文用其作为侧重负荷数据整体趋势的区间预测模型。GPR模型的训练输入特征为预测日当天的归一化后的气象数据。图8所示为GPR模型在不同置信度下的区间预测结果(与图7为同一日)。从图8可以看出,与QR-BiLSTM模型相比,GPR模型的  $I_{ECP}^{(\beta)}$  值普遍较高,但该模型预测区间的宽度除了在95%置信度下与QR-BiLSTM模型的相近,在其余置信度下均大幅超过了QR-BiLSTM模型的,说明该预测结果有着较高的可靠性和相对较低的锐度性能。因此,GPR预测模型能够根据所输入的气象数据来确定映射关系,以拟合相应场景下整体负荷数据的分布规律,因而适合作为表征负荷整体趋势的预测模型。

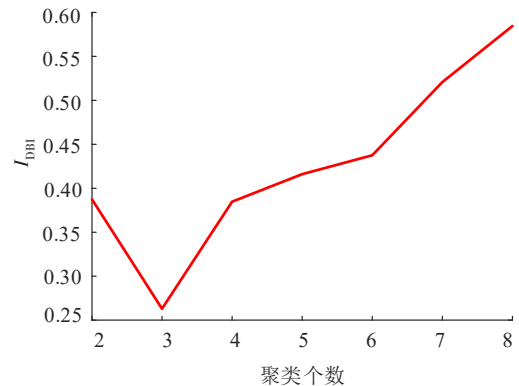
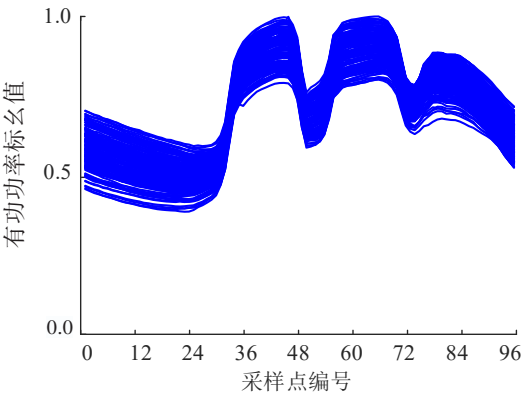
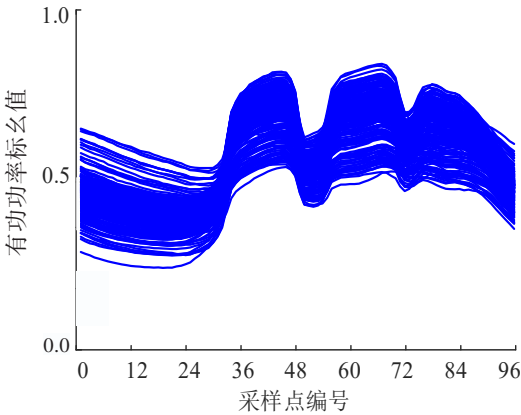


图5 不同聚类个数的  $I_{DBI}$  对比

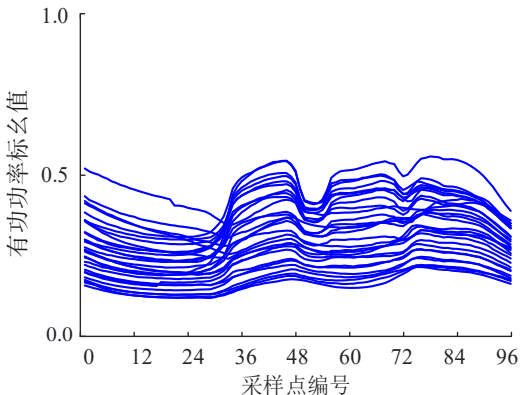
Fig. 5 Comparison of  $I_{DBI}$  with different cluster numbers



(a) 第1类日负荷曲线聚类结果



(b) 第2类日负荷曲线聚类结果



(c) 第3类日负荷曲线聚类结果

图6 日负荷曲线聚类结果

Fig. 6 Daily load curve clustering results

表1 QR-BiLSTM模型和GPR模型的评价指标

Table 1 Evaluation index of QR-BiLSTM model and GPR model

置信度/%	QR-BiLSTM模型		GPR模型	
	$I_{ECP}^{(\beta)}$	$I_A^{(\beta)}$	$I_{ECP}^{(\beta)}$	$I_A^{(\beta)}$
95	0.968 8	1 625.2	1.000 0	1 590.4
80	0.791 7	786.1	0.989 6	1 007.2
70	0.718 8	535.4	0.895 8	814.6
50	0.531 3	322.6	0.572 9	530.1

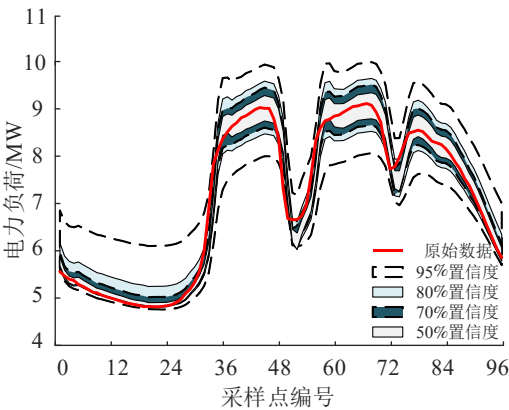


图7 不同置信度下QR-BiLSTM模型的预测区间

Fig. 7 Prediction interval of QR-BiLSTM model under different confidence levels

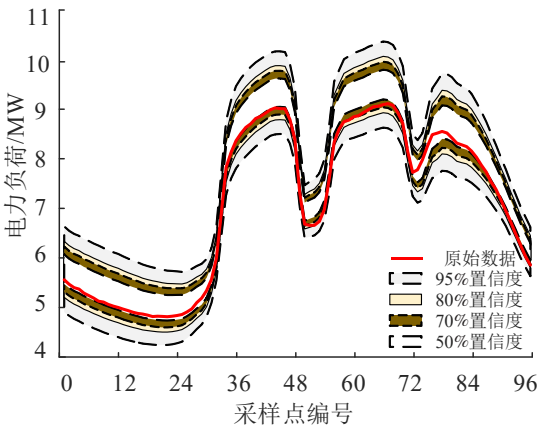


图8 不同置信度下GPR模型的预测区间

Fig. 8 Prediction interval of GPR model under different confidence levels

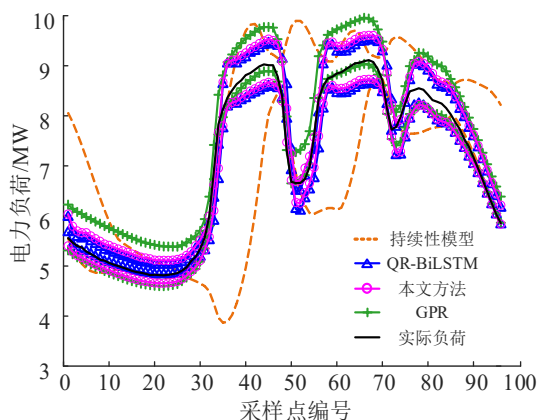
同时,对于不同的置信度对应的预测区间,置信度高的可靠性较高,区间宽度也较大,置信度低的则与之相反,这导致预测结果的参考价值较小。为了避免出现置信度过高或过低导致的区间预测结果不够理想的情况,在模型间做横向对比时统一使用70%置信度下的区间预测结果。

5.2.2 本文方法区间预测的有效性验证

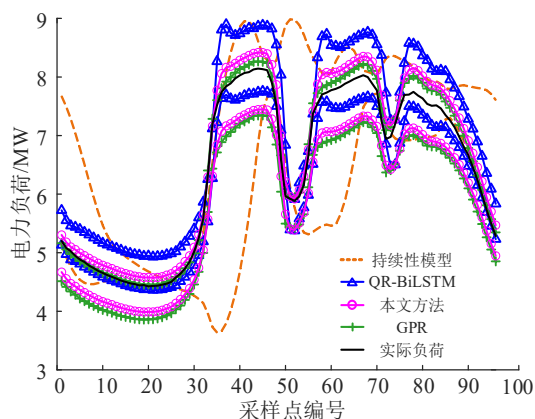
为了验证本文方法的预测精度,从测试集中选取不同聚类场景下3日(记为预测日1、预测日2、预测日3)的数据进行预测,得出本文方法的区间预测结果,并选用两种单一模型QR-BiLSTM、GPR以及持续性区间预测模型<sup>[20]</sup>做指标对比。持续性区间预测模型作为预测对比时常用的参考模型,其预测时刻的概率分布由该时刻前4 h内的实际概率分布决定。图9所示为70%置信度下不同场景对应预测日中各模型预测区间对比,各模型



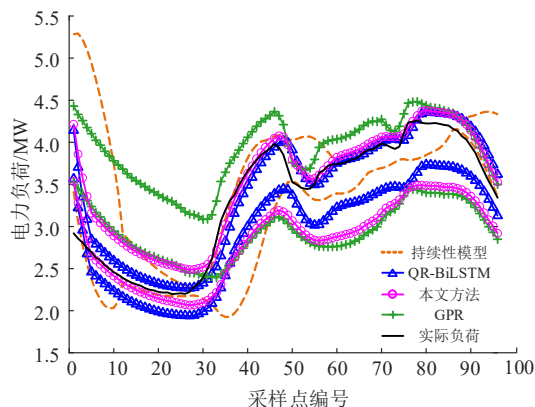
的评价指标见表2。由表2可知,与另外3个模型相比,本文方法在3种场景所对应的预测日中的评价指标最优,有着更高的可靠性与综合性能。对于预测日3,由于对应场景下的聚类个数较少,本文方法中的基础单一模型的训练学习不够充分,导致信息聚合后 $I_{ECP}^{(\beta)}$ 值偏低,但仍然比另外3个单一模型的高,说明相比于其他模型,本文方法有着



(a) 预测日1中各模型预测区间对比



(b) 预测日2中各模型预测区间对比



(c) 预测日3中各模型预测区间对比

图9 不同场景对应预测日中各模型预测区间对比

Fig. 9 Comparison of prediction intervals of various models under different scenarios corresponding to prediction days

表2 本文方法与其他单一模型评价指标对比

Table 2 The proposed method is compared with other single model evaluation indicators

预测日	区间预测模型	$I_{ECP}^{(\beta)}$	$I_W^{(\beta)}$
1	QR-BiLSTM	0.770 8	711.9
	GPR	0.895 8	818.7
	持续性模型	0.708 3	2 336.7
	本文方法	0.916 7	693.6
2	QR-BiLSTM	0.947 9	835.1
	GPR	0.875 0	779.3
	持续性模型	0.697 9	2 113.3
	本文方法	0.979 2	778.4
3	QR-BiLSTM	0.666 7	772.1
	GPR	0.687 5	1 122.4
	持续性模型	0.687 5	937.4
	本文方法	0.812 5	771.7

较高的预测精度。相比于信息聚合前的两个单一模型(QR-BiLSTM模型和GPR模型),本文方法的区间预测结果也最优,其可靠性指标比这两个单一模型的分别平均提高了14.70%、10.81%,综合性能分别平均提高了3.14%、15.55%,这也验证了本文方法的有效性。

### 5.2.3 曲线聚类的有效性验证

为了验证曲线聚类对于本文方法的作用,以下对比3种模型在预测日1中的预测表现,结果见表3。其中,模型1为本文预测方法,模型2为在本文方法中不加入曲线聚类步骤,模型3为在本文方法中不引入DTW距离。

表3 3种模型在预测日1中的评价指标对比

Table 3 Comparison of evaluation indicators of three models in forecast day 1

预测模型	$I_{ECP}^{(\beta)}$	$I_W^{(\beta)}$
模型1	0.916 7	693.6
模型2	0.718 8	1 946.9
模型3	0.875 0	728.0

预测偏差是衡量模型预测精度的一种常用方法<sup>[21]</sup>,在区间形式的预测中,使用真实值在预测区间的越限情况作为预测偏差,当真实值处于区间内部时,偏差值为0。图10对比了3种模型的预测偏差。

从图10可以看出,模型1的越限次数最少,越限幅值最小,模型2的越限次数最多。通过表3的评价指标对比及图10中模型1与模型2的预测结果对比,说明曲线聚类的引入较好地提高了预测精度。本文方法与不引入聚类方法的预测模型相

比,可靠性提高了27.53%,综合性能提高了64.37%;模型1与模型3的对比结果说明DTW距离的引入使负荷曲线的聚类规律更适合预测模型的学习,并进一步提高了预测精度。

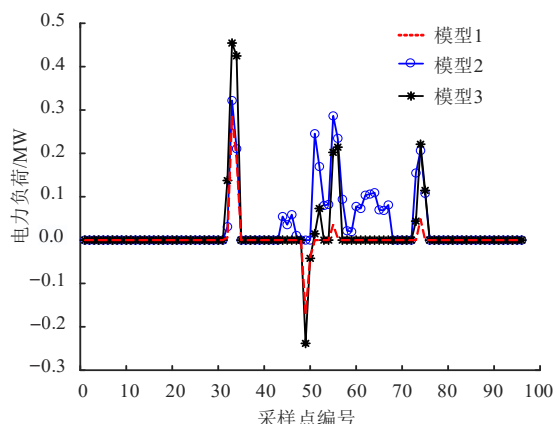


图10 3种模型的预测偏差对比

Fig. 10 Comparison of the prediction bias of the three models

## 6 结论

为了有效提取负荷曲线特征,综合考虑负荷曲线日内整体趋势及局部波动,实现预测模型互补,本文提出了一种基于曲线特征聚类与信息聚合的电力负荷超短期区间预测方法。主要结论如下:

1) 通过引入基于DTW距离的K-means聚类算法实现了负荷曲线的聚类,有效区分了不同类别曲线的特征。

2) 负荷曲线聚类有助于提高预测精度,与不引入聚类方法的预测模型相比,本文方法的可靠性提高了27.53%,综合性能提高了64.37%。

3) 信息聚合方法的引入有效提高了预测精度,充分结合了两个单一预测模型(QR-BiLSTM模型和GPR模型)的优势,本文方法的可靠性指标比这两个单一模型的分别平均提高了14.70%、10.81%,综合性能分别平均提高了3.14%、15.55%。

综上所述,本文方法能够利用负荷曲线的自身特性,发挥各单一模型的优势,从而有效实现精确的电力负荷的区间超短期预测,有利于电力系统安全稳定运行。在本文方法的基础上,提高待聚合模型的预测精度、选取更精细的气象数据并提取有效特征是本文后续工作的研究重点。

## [参考文献]

- [1] 唐贤伦,陈洪旭,熊德意,等.基于极端梯度提升和时间卷积网络的短期电力负荷预测[J].高电压技术,2022,48(8):3059-3067. DOI: 10.13336/j.1003-6520.hve.20211771.  
TANG Xianlun, CHEN Hongxu, XIONG Deyi, et al. Short-term power load forecasting based on extreme gradient boosting and temporal convolutional network [J]. High Voltage Engineering, 2022, 48(8): 3059-3067. DOI: 10.13336/j.1003-6520.hve.20211771.
- [2] HOU H, LIU C, WANG Q, et al. Review of load forecasting based on artificial intelligence methodologies, models, and challenges [J]. Electric Power Systems Research, 2022, 210: 108067. DOI: 10.1016/j.epsr.2022.108067.
- [3] 万灿,崔文康,宋永华.新能源电力系统概率预测:基本概念与数学原理[J].中国电机工程学报,2021,41(19):6493-6509. DOI: 10.13334/j.0258-8013.pcsee.210931.  
WAN Can, CUI Wenkang, SONG Yonghua. Probabilistic forecasting for power systems with renewable energy sources: basic concepts and mathematical principles [J]. Proceedings of the CSEE, 2021, 41(19): 6493-6509. DOI: 10.13334/j.0258-8013.pcsee.210931.
- [4] 罗凤章,张旭,杨欣,等.基于深度学习的综合能源配电网系统负荷分析预测[J].高电压技术,2021,47(1):23-32. DOI: 10.13336/j.1003-6520.hve.20201477.  
LUO Fengzhang, ZHANG Xu, YANG Xin, et al. Load analysis and prediction of integrated energy distribution system based on deep learning [J]. High Voltage Engineering, 2021, 47(1): 23-32. DOI: 10.13336/j.1003-6520.hve.20201477.
- [5] 李丹,张远航,杨保华,等.基于约束并行LSTM分位数回归的短期电力负荷概率预测方法[J].电网技术,2021,45(4):1356-1364. DOI: 10.13335/j.1000-3673.pst.2020.1124.  
LI Dan, ZHANG Yuanhang, YANG Baohua, et al. Short time power load probabilistic forecasting based on constrained parallel-LSTM neural network quantile regression mode [J]. Power System Technology, 2021, 45(4): 1356-1364. DOI: 10.13335/j.1000-3673.pst.2020.1124.
- [6] 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. DOI: 10.13335/j.1000-3673.pst.2021.0969.  
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. DOI: 10.13335/j.1000-3673.pst.2021.0969.

- [7] 黄南天, 刘德宝, 蔡国伟, 等. 基于多相关日场景生成的电动汽车充电负荷区间预测[J]. 中国电机工程学报, 2021, 41(23): 7980-7990. DOI: 10.13334/j.0258-8013.pcsee.201906.  
HUANG Nantian, LIU Debao, CAI Guowei, et al. Interval prediction of electric vehicle charging load based on scene generation with multiple correlation days [J]. Proceedings of the CSEE, 2021, 41(23): 7980-7990. DOI: 10.13334/j.0258-8013.pcsee.201906.
- [8] 肖白, 聂鹏, 穆钢, 等. 基于多级聚类分析和支持向量机的空间负荷预测方法[J]. 电力系统自动化, 2015, 39(12): 56-61. DOI: 10.7500/AEPS20140520001.  
XIAO Bai, NIE Peng, MU Gang, et al. A spatial load forecasting method based on multilevel clustering analysis and support vector machine [J]. Automation of Electric Power Systems, 2015, 39(12): 56-61. DOI: 10.7500/AEPS20140520001.
- [9] 牛东晓, 魏亚楠. 基于FHNN相似日聚类自适应权重的短期电力负荷组合预测[J]. 电力系统自动化, 2013, 37(3): 54-57. DOI: 10.7500/AEPS201202139.  
NIU Dongxiao, WEI Ya'nan. Short-term power load combinatorial forecast adaptively weighted by FHNN similar-day clustering [J]. Automation of Electric Power Systems, 2013, 37(3): 54-57. DOI: 10.7500/AEPS201202139.
- [10] 刘升伟, 王星华, 鲁迪, 等. 基于改进高斯过程回归的短期负荷概率区间预测方法[J]. 电力系统保护与控制, 2020, 48(1): 18-25. DOI: 10.19783/j.cnki.pspc.190179.  
LIU Shengwei, WANG Xinghua, LU Di, et al. Electric load probabilistic interval prediction method based on improved Gaussian process regression [J]. Power System Protection and Control, 2020, 48(1): 18-25. DOI: 10.19783/j.cnki.pspc.190179.
- [11] YANG D Z, VAN DER MEER D. Post-processing in solar forecasting: ten overarching thinking tools [J]. Renewable and Sustainable Energy Reviews, 2021, 140: 110735. DOI: 10.1016/j.rser.2021.110735.
- [12] 赵书强, 要金铭, 李志伟. 基于改进K-means聚类和SBR算法的风电场景缩减方法研究[J]. 电网技术, 2021, 45(10): 3947-3954. DOI: 10.13335/j.1000-3673.pst.2020.2013.  
ZHAO Shuqiang, YAO Jinming, LI Zhiwei. Wind power scenario reduction based on improved K-means clustering and SBR algorithm [J]. Power System Technology, 2021, 45(10): 3947-3954. DOI: 10.13335/j.1000-3673.pst.2020.2013.
- [13] 宋军英, 崔益伟, 李欣然, 等. 基于欧氏动态时间弯曲距离与熵权法的负荷曲线聚类方法[J]. 电力系统自动化, 2020, 44(15): 87-94. DOI: 10.7500/AEPS20191016008.  
SONG Junying, CUI Yiwei, LI Xinran, et al. Load curve clustering method based on euclidean dynamic time warping distance and entropy weight [J]. Automation of Electric Power Systems, 2020, 44(15): 87-94. DOI: 10.7500/AEPS20191016008.
- [14] 王振浩, 王翀, 成龙, 等. 基于集合经验模态分解和深度学习的光伏功率组合预测[J]. 高电压技术, 2022, 48(10): 4133-4142. DOI: 10.13336/j.1003-6520.hve.20210971.  
WANG Zhenhao, WANG Chong, CHENG Long, et al. Photovoltaic power combined prediction based on ensemble empirical mode decomposition and deep learning [J]. High Voltage Engineering, 2022, 48(10): 4133-4142. DOI: 10.13336/j.1003-6520.hve.20210971.
- [15] 梁智, 孙国强, 俞娜燕, 等. 基于高斯过程回归和粒子滤波的短期风速预测[J]. 太阳能学报, 2020, 41(3): 45-51. DOI: 10.19912/j.0254-0096.2020.03.007.  
LIANG Zhi, SUN Guoqiang, YU Nayan, et al. Short-term wind speed forecasting based on Gaussian process regression and particle filter [J]. Acta Energetica Sinica, 2020, 41(3): 45-51. DOI: 10.19912/j.0254-0096.2020.03.007.
- [16] 张逸, 刘必杰, 邵振国, 等. 基于高斯过程回归的谐波源不确定性通用模型[J]. 中国电机工程学报, 2022, 42(3): 992-1002. DOI: 10.13334/j.0258-8013.pcsee.210767.  
ZHANG Yi, LIU Bijie, SHAO Zhenguo, et al. General and uncertain harmonic source model based on Gaussian process regression [J]. Proceedings of the CSEE, 2022, 42(3): 992-1002. DOI: 10.13334/j.0258-8013.pcsee.210767.
- [17] ABDEL-NASSER M, MAHMOUD K, LEHTONEN M. Reliable solar irradiance forecasting approach based on choquet integral and deep LSTMs [J]. IEEE Transactions on Industrial Informatics, 2021, 17(3): 1873-1881. DOI: 10.1109/TII.2020.2996235.
- [18] MAHMOUD K, HUSSEIN M M, ABDEL-NASSER M, et al. Optimal voltage control in distribution systems with intermittent PV using multiobjective grey-wolf-leve optimizer [J]. IEEE Systems Journal, 2020, 14(1): 760-770. DOI: 10.1109/JSYST.2019.2931829.
- [19] 万灿, 宋永华. 新能源电力系统概率预测理论与方法及其应用[J]. 电力系统自动化, 2021, 45(1): 2-16. DOI: 10.7500/AEPS20200811008.  
WAN Can, SONG Yonghua. Theories, methodologies and applications of probabilistic forecasting for power systems with renewable energy sources [J]. Automation of Electric Power Systems, 2021, 45(1): 2-16. DOI: 10.7500/AEPS20200811008.
- [20] SIDERATOS G, HATZIARGYRIOU N D. Probabilistic wind power forecasting using radial basis function neural networks [J]. IEEE Transactions on Power Systems, 2012, 27(4): 1788-1796. DOI: 10.1109/



TPWRS.2012.2187803.  
[21] 周金应,徐磊,程前,等.基于负荷跟随阈值变化规则的并联混合动力汽车能量管理策略[J].公路与汽运, 2021(4): 1-7. DOI: 10.3969/j.issn.1671-2668.2021.04.002.

ZHOU Jinying, XU Lei, CHENG Qian, et al. Energy management strategy of parallel hybrid electric vehicle based on load following threshold change rule [J]. Highways & Automotive Applications, 2021(4): 1-7. DOI: 10.3969/j.issn.1671-2668.2021.04.002.

## Power load forecasting based on curve feature clustering and information aggregation

YUE Shouzhi<sup>1,2</sup>, HONG Haisheng<sup>2</sup>, DENG Qi<sup>2</sup>, XU Chende<sup>2</sup>, LUO Feng<sup>2</sup>

(1. School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China;

2. Guangzhou Power Supply Bureau, Guangdong Power Grid Co., Ltd., Guangzhou 510620, China)

**Abstract:** [Purposes] In order to obtain accurate and reliable ultra-short term power load forecasting results to meet the needs of rapid response and real-time dispatching of power system, considering the characteristics of non-linear and timely ordering of power grid load data, this paper proposed a method of ultra-short term interval forecasting of power load based on curve feature clustering and information aggregation. [Methods] Firstly, considering the local fluctuation characteristics and overall trend characteristics of the load curve, the power load curve was divided into different categories. Then, the Gaussian process regression model was used as the prediction model to characterize the overall trend of load, and the quantile based bidirectional long-term and short-term memory neural network was used as the prediction model to characterize the local fluctuation of load. Finally, the aggregation idea was introduced, and the Choquet integral algorithm was used as the aggregation function to aggregate the results of the two prediction models. [Findings] This method effectively realized the clustering of daily load curves considering multiple characteristics. Meanwhile, the prediction results of a single model were aggregated, which obtained the interval prediction results of each confidence level under different scenarios. Under example analysis, the reliability index of the prediction method in this paper was 14.70% and 10.81% higher than that of the two single models, and the comprehensive performance was 3.14% and 15.55% higher than that of the two single models. [Conclusions] The results of the example showed that the forecasting accuracy and reliability of the model proposed in this paper are significantly improved comparing with the common load probability forecasting methods and the single forecasting model. In addition, the introduction of clustering method and information aggregation idea is useful to improve the accuracy of the forecasting model.

**Key words:** load interval prediction; long and short-term memory neural network; K-means clustering; Gaussian process regression; information aggregation

Manuscript received: 2023-02-15; revised: 2023-03-10; accepted: 2023-03-10

Foundation item: Project (GZHKJXM20200037) supported by China Southern Power Grid

Corresponding author: YUE Shouzhi (1993—) (ORCID: 0009-0006-1477-7832), male, assistant engineer, research interest: power system big data analysis. E-mail: 1411342870@qq.com

(责任编辑:石月珍;校对:刘平;英文编辑:彭卓寅)