

# 基于数据关联狄利克雷混合模型的电网 净负荷不确定性表征研究

李远征<sup>1,2</sup> 孙天乐<sup>1,2</sup> 刘云<sup>3</sup> 赵勇<sup>1,2</sup> 曾志刚<sup>1,2</sup>

**摘要** 针对电网净负荷时序数据关联的特点, 提出基于数据关联的狄利克雷混合模型 (Data-relevance Dirichlet process mixture model, DDPMM) 来表征净负荷的不确定性. 首先, 使用狄利克雷混合模型对净负荷的观测数据与预测数据进行拟合, 得到其混合概率模型; 然后, 提出考虑数据关联的变分贝叶斯推断方法, 改进后验分布对该混合概率模型进行求解, 从而得到混合模型的最优参数; 最后, 根据净负荷预测值的大小得到其对应的预测误差边缘概率分布, 实现不确定性表征. 本文基于比利时电网的净负荷数据进行检验, 算例结果表明: 与传统的狄利克雷混合模型和高斯混合模型 (Gaussian mixture model, GMM) 等方法相比, 所提出的基于数据关联狄利克雷混合模型可以更为有效地表征净负荷的不确定性.

**关键词** 狄利克雷混合模型, 净负荷, 不确定性表征, 时序序列, 预测误差

**引用格式** 李远征, 孙天乐, 刘云, 赵勇, 曾志刚. 基于数据关联狄利克雷混合模型的电网净负荷不确定性表征研究. 自动化学报, 2022, 48(3): 747-761

**DOI** 10.16383/j.aas.c210668

## Uncertainty Characterization of Power Grid Net Load of Dirichlet Process Mixture Model Based on Relevant Data

LI Yuan-Zheng<sup>1,2</sup> SUN Tian-Le<sup>1,2</sup> LIU Yun<sup>3</sup> ZHAO Yong<sup>1,2</sup> ZENG Zhi-Gang<sup>1,2</sup>

**Abstract** Considering the time sequence correlation of the net load in power grids, this paper proposes a non-parametric Bayesian framework based on the data-relevance Dirichlet process mixture model (DDPMM). First, the Dirichlet process mixture model is used to fit the observation data and the forecast data of net load, in order to obtain a joint mixed probability model. Then, we propose a modified variational Bayesian inference that considers the correlation of the time sequence to acquire optimal parameters of the mixture model. Afterwards, we obtain the corresponding net load forecast error probabilistic distribution according to different net load forecast values. Finally, this paper uses actual data from the Belgian power grid for verification. Compared with the traditional Dirichlet process mixture model and Gaussian mixture model (GMM), our proposed approach takes into account the correlation of time series. Therefore, it can better characterize the forecast error of the net load, and obtain the optimal number of component clusters that would capture the uncertainty of the net load with more efficiency.

**Key words** Dirichlet process mixture model (DPMM), net load, uncertain characterization, time sequence, forecast error

**Citation** Li Yuan-Zheng, Sun Tian-Le, Liu Yun, Zhao Yong, Zeng Zhi-Gang. Uncertainty characterization of power grid net load of Dirichlet process mixture model based on relevant data. *Acta Automatica Sinica*, 2022, 48(3): 747-761

收稿日期 2021-07-16 录用日期 2021-11-02

Manuscript received July 16, 2021; accepted November 2, 2021  
国家自然科学基金 (62073148), 腾讯—犀牛鸟基金 (RAGR20210102) 资助

Supported by National Natural Science Foundation of China (62073148) and Tencent Rhinoceros Foundation of China (RAGR20210102)

本文责任编辑 杨涛

Recommended by Associate Editor YANG Tao

1. 华中科技大学人工智能与自动化学院, 图像信息处理与智能控制教育部重点实验室 武汉 430074 2. 中国-测控技术“一带一路”联合实验室 武汉 430074 3. 华南理工大学电力学院 广州 510641

1. School of Artificial Intelligence and Automation, Key Laboratory on Image Information Processing and Intelligent Control of Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074 2. China-Belt and Road Joint Laboratory on Measurement and Control Technology, Wuhan 430074 3. School of Electrical Power Engineering, South China University of Technology, Guangzhou 510641

电网净负荷为电力用户用电需求与新能源出力之间的有功功率差值. 随着新能源的高速发展, 风电和光伏发电大规模接入电网, 在一定程度上减少了化石能源消耗, 降低了环境污染. 然而, 受新能源功率的波动性和间隙性影响, 特别是随着风光等接入比例越来越高, 电网净负荷的精准预测变得十分困难, 这给电网运行的安全性和经济性带来了一定的挑战<sup>[1-3]</sup>. 因此, 如何有效表征净负荷的不确定性, 从而在一定程度上合理安排电力备用资源, 对确保电网安全经济运行具有十分重要的意义.

目前, 国内外关于电网净负荷不确定性表征的相关研究, 主要分为以下两类. 第 1 类是通过概率

预测或者区间预测方法来直接表征净负荷的不确定性<sup>[4]</sup>. 第 2 类是通过研究确定性预测值的预测误差, 分析其统计特性, 从而表征净负荷的不确定性<sup>[5]</sup>. 由于确定性预测相对于概率/区间预测简单易行, 因此基于第 2 类的净负荷不确定性表征方法受到国内外学者的密切关注. Bhandari 等使用 Logistic 分布来描述净负荷的预测误差, 相比于正态分布, Logistic 分布能更好地描述其偏峰特征<sup>[6-7]</sup>. 然而, 当分布式可再生能源的渗透率较高时, Logistic 分布则无法描述其尖峰或者多峰的特征<sup>[8]</sup>. 为解决这一问题, 文献 [9-11] 使用高斯混合模型 (Gaussian mixture model, GMM) 来表征负荷与可再生能源的不确定性, 通过利用 GMM 对净负荷预测误差进行拟合, 可以更好地描述尖峰的特性. 然而, GMM 需要事先指定概率分布的混合模型组数, 目前主要依靠经验来设定, 一般会导致较大的误差. 为了解决 GMM 上述的缺点, Sun 等<sup>[12]</sup> 使用狄利克雷混合模型 (Dirichlet process mixture model, DPMM) 来表征负荷的不确定性, 但是该工作假设观测数据是独立的. 事实上, 在实际电网中, 净负荷表现为时间序列数据, 具有较强的数据关联特点, 即数据关联性<sup>[13]</sup>. 然而, 传统的 DPMM 没有考虑这一关联性, 导致其无法较好地描述净负荷信息, 存在一定的偏差.

针对 DPMM 的这一缺陷, 本文提出基于数据关联的狄利克雷混合模型 (Data-relevance Dirichlet process mixture model, DDPMM) 的贝叶斯框架, 使用考虑数据关联的改进后验概率分布, 并结合变分贝叶斯推断来获取最优的混合模型个数及其参数, 来更好地表征净负荷不确定性. 首先, 构造基于狄利克雷过程的混合模型, 其中主要包括每个混合模型的权重比例及其对应的参数; 其次, 利用净负荷数据具有关联性的特点, 基于变分推断方法提出一种改进的后验分布, 来描述各个模型与观测数据之间的关系. 进一步, 使用改进的期望最大 (Expectation-maximum, EM) 算法对混合模型进行迭代求解, 得到其在不同预测值下的预测误差边缘概率分布, 从而实现净负荷的不确定性表征. 最后, 根据比利时电网的数据进行实验验证. 本文的主要贡献包含以下 3 个方面:

1) 提出了基于数据关联的狄利克雷过程混合模型, 考虑净负荷时序关联性来改进 DPMM 的后验概率分布, 并描述其不确定性. 相比于 DPMM 算法, DDPMM 可以更为有效地利用净负荷数据.

2) 本文根据改进的 DPMM 后验概率分布, 构造出新的证据下界. 基于 EM 算法, 利用新构造证据下界的更新信息证明了 DDPMM 的收敛性.

3) 通过比利时电网的实际数据验证了 DDPMM

的效果. 算例结果表明, 与传统的 DPMM 相比, 考虑净负荷数据关联信息的 DDPMM 会大幅度加快收敛速度, 混合模型精度也得以进一步提升.

## 1 基于数据关联狄利克雷混合模型的贝叶斯框架

本章首先通过对净负荷数据进行分析, 揭示净负荷时间序列的数据关联性; 其次, 通过该数据关联性来改进传统的 DPMM, 提出考虑数据关联的狄利克雷混合模型的贝叶斯框架.

### 1.1 净负荷时序序列关联性

净负荷数据的表现形式为时序序列, 其当前状态与历史时刻相关, 因此本文使用滞后相关性来描述净负荷的数据关联程度<sup>[14]</sup>. 具体计算为

$$\text{corr}(\Gamma, \Gamma_T) = \frac{\sum_{t=T+1}^{N-T} \zeta_t \times \zeta_{t+T}}{\sqrt{\sum_{t=1}^{N-T} \zeta_t^2} \times \sqrt{\sum_{t=1}^{N-T} \zeta_{t+T}^2}} \quad (1)$$

针对式 (1), 以比利时 2019 年全年净负荷观测数据  $\Gamma = \{\zeta_1, \zeta_2, \dots, \zeta_t, \dots, \zeta_N\}$  为例, 其中,  $\zeta_t$  为第  $t$  个观测数据, 数据集  $\Gamma$  的时间分辨率为 15 min,  $N$  为  $96 \times 365 = 35040$  代表一年的数据量<sup>[15]</sup>.  $\Gamma_T$  代表滞后  $T$  步的净负荷时序序列,  $\zeta_{t+T}$  为第  $(t+T)$  个观测数据. 图 1 表示净负荷的关联程度, 横纵坐标代表时间滞后的尺度, 共 10 步, 每步步长为 15 min, 右侧色阶表示其数据关联程度大小. 在图 1 中, 可以观察到净负荷的数据关联程度随着平移时间的增加而逐渐减少. 以滞后 10 步为例 ( $T = 10$ ), 坐标 (0, 10) 对应的右纵坐标数值代表 150 min 后的净负荷  $\{\zeta_1, \zeta_2, \dots, \zeta_{N-10}\}$  与当前数据  $\{\zeta_{11}, \zeta_{12}, \dots, \zeta_N\}$  的关联程度大小, 即 0.82. 这表明当前数据与 150 min 后的净负荷关联性依然很强.

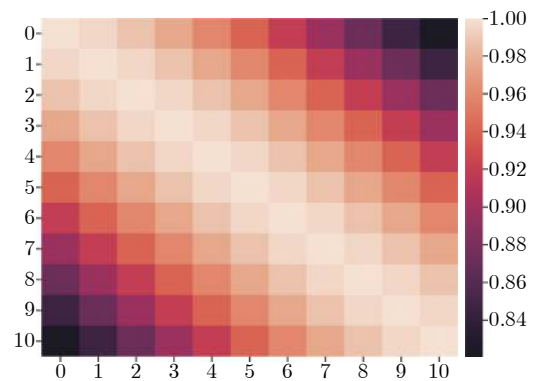


图 1 净负荷数据关联图

Fig.1 The data-relevance of net load

因此, 在具体使用 GMM 或 DPMM 对净负荷数据进行拟合时, 需要考虑历史数据对当前时刻的影响. 然而, 传统的 GMM 与 DPMM 在处理时序序列时会假设各个观测数据之间是独立的, 没有考虑净负荷数据关联这一特点, 可能导致 GMM 与 DPMM 生成的概率混合模型含有更多的组份, 容易引起过拟合, 影响净负荷不确定性的表征精度<sup>[16]</sup>.

为此, 本文通过考虑净负荷数据关联的特点来改进传统的 DPMM, 提出考虑数据关联的狄利克雷混合模型贝叶斯框架.

## 1.2 基于数据关联狄利克雷混合模型贝叶斯框架

本文所提出的基于数据关联的狄利克雷混合模型贝叶斯框架包括 3 个部分, 如图 2 所示.

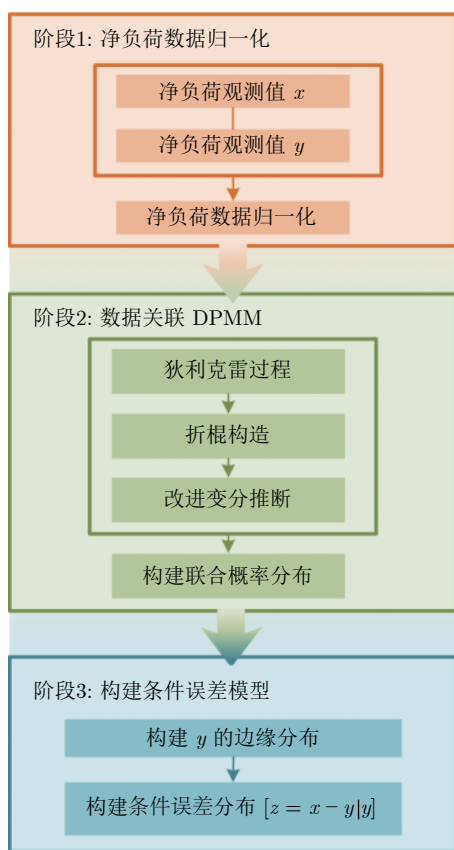


图 2 数据关联狄利克雷混合模型变分贝叶斯框架  
Fig.2 The framework of DDPMM variational Bayes

第 1 阶段进行净负荷数据归一化. 以  $x$  和  $y$  分别表示净负荷的观测值和预测值, 它们的数值在本文研究的比利时实际电网数据集中已经给出. 然后, 对  $x$  和  $y$  分别进行归一化操作<sup>[17]</sup>. 第 2 阶段, 使用提出的 DDPMM 得到  $x$  与  $y$  的联合概率分布. 首先, 使用折棍构造过程<sup>[18]</sup>来表示狄利克雷过程, 得到该联合概率分布的混合模型; 其次采用考虑数据关联

的改进变分贝叶斯推断来计算出混合模型的均值矩阵与协方差矩阵, 以及混合组数. 最后, 在第 3 阶段中, 根据  $x$  与  $y$  的联合概率分布, 得到关于预测值的边缘概率分布  $P(z|y)$ , 其中  $z = x - y$  代表净负荷的预测误差. 通过上述基于 DDPMM 的非参数贝叶斯框架来描述净负荷的不确定性, 其中主要包括折棍构造, 狄利克雷过程和改进变分推断这几个步骤, 下面分别进行阐述.

## 1.3 狄利克雷过程与折棍构造表示法

DDPMM 是基于 DPMM 的改进, 其中 DPMM 是一种非参数的贝叶斯的软聚类方法, 即它是将数据  $E = \{\epsilon_1, \epsilon_2, \dots, \epsilon_N\}$  以一定概率归于某一高斯分布<sup>[19]</sup>, 其中  $N$  表示数据个数. DPMM 可以从数据中得到最优的概率分布来描述该数据集  $E$ , 而无需指定混合模型的具体参数以及混合组数.

考虑一个连续的概率分布  $H(\Omega)$ , 它的每个样本点  $\Omega_k = \{\mu_k, \sigma_k\}$  都表示高斯分布, 其中  $\mu_i, \sigma_i$  为其均值与方差, 因此  $H(\Omega)$  为一个混合模型, 即它是由无穷个高斯分布加权得到, 每个数据被分类到某个  $\Omega_k$  中. 然而, 当  $H(\Omega)$  为连续分布时, 从  $H(\Omega)$  采样得到两个相同高斯分布的概率为 0, 即任意两个不同的数据  $\epsilon_i, \epsilon_j$  ( $i \neq j$ ) 被分类至不同的高斯分布, 将失去聚类的意义. 因此需要将  $H(\Omega)$  离散化, 使不同的数据  $\epsilon_i, \epsilon_j$  ( $i \neq j$ ) 划分到相同的类里, 该离散化过程称为狄利克雷过程, 它是用于贝叶斯建模和分析的随机过程<sup>[20]</sup>. 定义  $H(\Omega)$  为基分布, 离散化之后的分布记为  $G$ . 通过 DP 采样得到的分布  $G$  表示为

$$G \sim DP(\phi, H(\Omega)) \quad (2)$$

其中,  $\phi$  为浓度参数, 表示的是  $H(\Omega)$  离散化的程度. 当  $\phi \rightarrow 0$  时, 分布  $G$  达到最离散的状态, 即成为样本空间只有一个样本的离散分布列; 当  $\phi \rightarrow \infty$ , 分布  $G$  与  $H$  相等, 成为连续分布, 即  $G = H$ . 在这两极端的情况之间, 浓度参数  $\phi$  越大,  $G$  越接近基分布  $H(\Omega)$ <sup>[21]</sup>.

狄利克雷过程的构造方法主要通过折棍构造, 其可以直观地理解混合模型的构造过程. 折棍构造就是将一根单位长度的棍子, 通过多次折断, 得到不同长度的棍子, 该过程被证明服从狄利克雷过程<sup>[22]</sup>. 折棍构造可以用以下两个表达式得到:

$$v_m \sim \text{Beta}(v_m|1, \phi) \quad (3)$$

$$\pi_m = v_m \prod_{s=1}^{m-1} (1 - v_s) \quad (4)$$

考虑一根单位长度的棍子, 如图 3 所示, 每次选取当前长度的一定比例  $v_m$ , 其中  $v_m$  服从

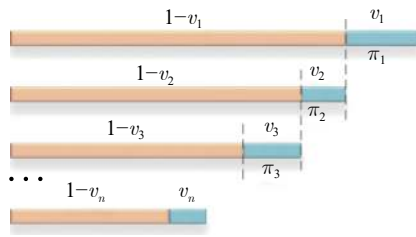


图3 折棍构造过程

Fig.3 The process of stick-breaking

$Beta(v_m|1, \phi)$ , 且  $v_m \in (0, 1)$ , 如式 (3) 所定义. 每次选取出来的棍子长度  $\pi_m$  为  $G$  中样本的权重, 即每个混合模型的比例, 且它是一个概率测度, 满足以下条件:

$$\sum_{m=1}^{\infty} \pi_m = 1 \quad (5)$$

因此, 每次选取出来的棍子其长度总和为 1, 即分布  $G$  的每个混合组份的权重之和为 1.

通过每次对  $H(\Omega)$  采样出一个高斯分布, 以及折棍构造计算出其权重系数, 进而得到分布  $G$ , 其中  $G$  的表达式为

$$G = \sum_{m=1}^{\infty} \pi_m \delta_{\Omega_m} \quad (6)$$

其中,  $\delta_{\Omega_m}$  为 Kronecker 函数, 表达式为

$$\delta_{\Omega_m} = \begin{cases} 1, & w = \Omega_m \\ 0, & \text{其他} \end{cases} \quad (7)$$

图 4 为基分布示意图, 其中位于横坐标之上的是基分布  $H(\Omega)$ , 而横坐标之下的是离散化之后的基分布  $G$ , 其中  $\Omega_i$  代表从  $H(\Omega)$  采样出的高斯分布, 而  $\pi_i$  表示每个高斯分布的权重.

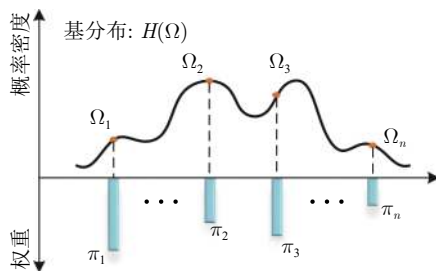


图4 基分布示意图

Fig.4 The base distribution

图 5 主要表示基分布  $G$  的形成过程. 首先,  $\Omega_i$  由  $H(\Omega)$  采样得到, 随后通过折棍构造得到其权重  $\pi_i$ . 重复此操作, 得到离散化的基分布  $G$ , 式 (6) 为其对应的数学表达式.

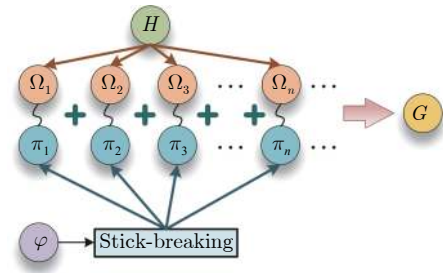


图5 狄利克雷混合模型

Fig.5 Dirichlet process mixture model

#### 1.4 非参数狄利克雷混合模型

事实上, 通过上述基于折棍构造得到的 DPMM, 结合数据的统计性质, 能够更新出最优的混合组数, 以及每个混合模型的权重  $\pi_m$  与  $\Omega_m$ , 且无需先验知识. 对于一个二维净负荷数据  $\epsilon_n = [x_n, y_n]$ , 可以采用 DPMM 将  $\epsilon_n$  的概率密度函数表示为

$$p(\epsilon_n|\pi, \mathcal{U}) = \sum_{m=1}^{\infty} \pi_m p_m(\epsilon_n|\Omega_m) \quad (8)$$

其中,  $x_n, y_n$  分别表示第  $n$  个净负荷数据的实际净负荷观测数据与预测净负荷,  $\pi = \{\pi_m\}_{m=1}^{\infty}$  和  $\mathcal{U} = \{\Omega_m\}_{m=1}^{\infty}$  分别表示  $\pi_m$  和  $\Omega_m$  的集合,  $p_m(\cdot)$  表示混合模型的第  $m$  个混合组份. 一般认为其先验为高斯分布<sup>[23]</sup>, 因此, 可以将 DPMM 表示为多元高斯分布的组合

$$p(\epsilon_n|\pi, \mathcal{U}) = \sum_{m=1}^{\infty} \pi_m N_m(\epsilon_n|\mu_m, \Lambda_m^{-1}) \quad (9)$$

其中,  $N_m(\epsilon_n|\mu_m, \Lambda_m^{-1})$  为混合模型的第  $m$  个组份, 为高斯分布, 其中  $\mu_m$  与  $\Lambda_m$  为其均值向量与协方差矩阵. 式 (9) 表示数据  $\epsilon_m$  由混合模型生成的条件概率.

在此基础上, 考虑数据集  $E = \{\epsilon_n\}_{n=1}^N$ , 其中, 对于  $\forall i, j \in \{1, 2, \dots, N\}$ ,  $\epsilon_i$  与  $\epsilon_j$  为互相独立的数据. 因此基于数据集  $E$  得到的 DPMM 可以表示为

$$p(E|\pi, \mathcal{U}) = \prod_{n=1}^N \left\{ \sum_{m=1}^{\infty} \pi_m N_m(\epsilon_n|\mu_m, \Lambda_m^{-1}) \right\} \quad (10)$$

为了将各数据与每个混合组份联系起来, 引入一个隐变量集合  $Z = \{z_m^{[n]}\}_{n=1}^N$ , 若第  $n$  个数据  $\epsilon_n$  与第  $m$  个混合部件关联, 则  $z_m^{[n]} = 1$ ; 否则,  $z_m^{[n]} = 0$ . 因此 DPMM 可以进一步表示为

$$p(E|Z, \mathcal{U}) = \prod_{n=1}^N \prod_{m=1}^{\infty} N_m(\epsilon_n|\mu_m, \Lambda_m^{-1})^{z_m^{[n]}} \quad (11)$$

图 6 表示 DPMM 各变量之间的关系. 首先混

合模型权重  $\pi$  由浓度系数  $\phi$  生成, 表示各混合模型的权重;  $\Omega_k$  表示从基分布  $H(\Omega)$  采样出的高斯分布. 另外, 混合模型权重  $\pi$  与净负荷数据是通过隐变量  $Z$  间接联系, 其中混合模型权重表示了各混合组份占混合模型的比例. 因此, 图 6 中隐变量  $Z$  是由  $\pi$  生成的. 可以得到  $Z$  在  $\pi$  上的条件概率分布表示为

$$p(Z|\pi) = \prod_{n=1}^N \prod_{m=1}^{\infty} \pi_m^{z_m^{[n]}} \quad (12)$$

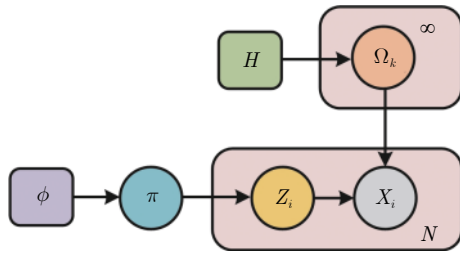


图 6 DPMM 概率图模型  
Fig.6 DPMM probability graph model

通过变分贝叶斯推断 (Variational Bayesian inference, VBI) 以及上述的非参数狄利克雷混合模型, 可以估计出较优的参数  $\pi$  和  $\Omega$ , 进而最终得到 DDPMM. VBI 将在第 2 节进行详细介绍.

## 2 基于 VBI 的数据关联的狄利克雷混合模型

本节提出了考虑数据关联的狄利克雷混合模型, 通过使用数据关联的信息来改进变分推断, 得到考虑数据关联之后的后验分布. 首先, 通过变分分布来近似净负荷数据的似然函数, 将变分推断问题转化为优化问题; 然后通过考虑净负荷数据关联特点, 建立新的变分后验分布; 最后通过构建新的上下界来证明 DDPMM 是收敛的.

### 2.1 非参数狄利克雷混合模型

VBI 用于估计参数  $\theta = \{\pi, \Omega\}$  以及隐变量  $Z$  的后验概率密度, 其中  $\pi = \{\pi_m\}_{m=1}^{\infty}$ . 在数据  $E$  给定的情况下, 目的是求解混合模型  $p$ , 使数据  $E$  在给定参数  $\theta$  上的后验分布  $p(E|\theta)$  最大:

$$\arg \max_{\theta} p(E|\theta) \quad (13)$$

由于  $p(E|\theta)$  难以计算出解析表达式, 本文利用 VBI 使用变分分布  $q$  来近似原分布  $p$ . VBI 假设变分分布  $q$  来自于变分族  $\mathcal{Q}$ , 通过最小化 KL 散度来近似后验分布. 因此, VBI 把贝叶斯推断问题转化为一个优化问题. 根据定义,  $p(E|\theta)$  的对数似然

可以分解为如下形式:

$$\begin{aligned} \ln p(E|\theta) &= \int \ln p(E|\theta) \times q(Z) dZ = \\ &= \int \ln \frac{p(E, Z|\theta)}{p(Z|E, \theta)} \times q(Z) dZ = \\ &= \int \ln \frac{p(E, Z|\theta)}{q(Z)} \times q(Z) dZ + \\ &= \int \ln \frac{q(Z)}{p(Z|E, \theta)} \times q(Z) dZ = \\ &= ELBO + KL(q(Z)||p(Z|E, \theta)) \end{aligned} \quad (14)$$

其中,  $ELBO$  与  $KL(q(Z)||p(Z|E, \theta))$  可表示为如下形式:

$$ELBO = \int \ln \frac{p(E, Z|\theta)}{q(Z)} \times q(Z) dZ \quad (15)$$

$$KL(q(Z)||p(Z|E, \theta)) = \int \ln \frac{q(Z)}{p(Z|E, \theta)} \times q(Z) dZ \quad (16)$$

因此, 极大化  $p(E|\theta)$  就需要优化  $Z$  与  $\theta$  两个参数, 分两步优化, 第 1 步优化参数  $Z$ , 使变分分布  $q(Z)$  接近  $p$ , 称为期望步 (Expectation step, E step), 即得到期望的变分分布; 第 2 步优化参数  $\theta$ , 称为最大化步 (Maximization step, M step), 通过优化  $\theta$  来极大化  $\ln p(E|\theta)$ .

本文着重考虑净负荷的数据关联性, 图 7 表示考虑数据关联的 DPMM 概率图, 与图 6 不同之处在于当前时刻数据参考了上一时刻数据的信息, 即  $Z_i$  的部分信息来自于  $Z_{i-1}$ . 这一部分由变分分布  $q(Z)$  决定, 因此 DDPMM 主要考虑变分分布的改进, 使得  $q(Z)$  去近似  $p(E|\theta)$ . 通过使用净负荷的关联数据的特性可得到改进后的变分分布  $q(Z)$ , 即极小化

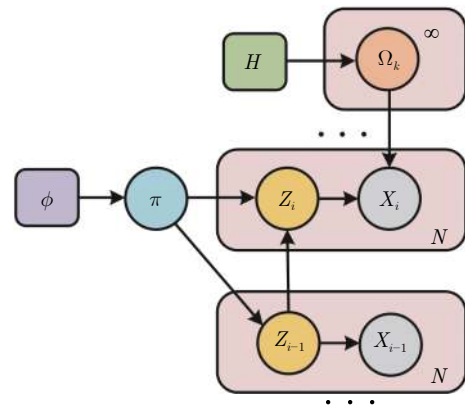


图 7 考虑数据关联性的 DPMM 概率模型图  
Fig.7 DPMM probability model graph considering data-relevance

$KL(q(Z)||p(Z|E, \theta))$ , 这对应于 EM 算法中的 E 步, 即得到期望的变分分布. 其中  $KL(q(Z)||p(Z|E, \theta))$  为 KL 散度, 可视为两个分布之间的距离, KL 散度越小两个分布越相似. 由 KL 散度的非负性可知,  $\ln p(E|\theta) \geq ELBO$ , 其中,  $ELBO$  为证据下界 (Evidence lower bound, ELBO). 当观测数据和预测数据已知, 混合模型固定时,  $p(E|\theta)$  为常数, 因此 KL 散度的极小化问题可以转化为如下问题:

$$\begin{aligned} \arg \min KL(q(Z)||p(Z|E, \theta)) &\Leftrightarrow \\ \arg \max (ELBO) &= \\ \int \ln \frac{p(E, Z|\theta)}{q(Z)} \times q(Z) dZ &= \\ E_{q(Z)} \left[ \ln \frac{p(E, Z|\theta)}{q(Z)} \right] &= \\ E_{q(Z)} [\ln p(E|Z, \theta)] + E_{q(Z)} [\ln p(Z|\pi)] - & \\ E_{q(Z)} [q(Z)] & \end{aligned} \quad (17)$$

VBI 通常假设变分分布来自场变分族, 隐变量之间相互独立, 为了更好地描述变分分布, 本文将折棍构造截断, 取一个稍大的值作为混合组数上限.

$$q(Z) = \prod_{i=1}^K q_i(z_i) \quad (18)$$

其中,  $K$  为混合模型成份的上限值. 通过使用信息传递算法进行循环坐标优化, 得到每个隐变量的更新公式. 对于第  $k$  个隐变量, 将与  $z_k$  无关的变量都视为常数. 因此, 式 (14) 的目标函数可以转化为如下形式:

$$\begin{aligned} E_{q(Z)} \left[ \ln \frac{p(E, Z|\theta)}{q(Z)} \right] &= \\ - \int \prod_{j=1}^K q_j \ln \frac{\prod_{j=1}^K q_j}{p} dZ &= \\ - \int \prod_{j=1}^K q_j \ln \prod_{j=1}^K q_j dZ + \int \prod_{j=1}^K q_j \ln p dZ &= \\ - \int \prod_{j=1}^K q_j (\ln q_k - \ln p) dZ - \int \prod_{j=1}^K q_j \sum_{i \neq k} \ln q_i dZ &= \\ - \int q_k \int \prod_{j \neq k} q_j (\ln q_k - \ln p) dZ_{-k} dz_k - \sum_{i \neq k} \ln q_i &= \\ \int q_k \{ E_{q(z-k)} \ln p - \ln q_k \} dz_k + \text{constant} & \end{aligned} \quad (19)$$

根据式 (17) 得到  $q_k$  的更新公式如下:

$$\begin{aligned} \frac{\partial E_{q(Z)} \left[ \ln \frac{p(E, Z|\theta)}{q(Z)} \right]}{\partial q_k} &= 0 \Rightarrow \\ E_{q(z-k)} \ln p - 1 + \text{constant} &= \ln q_k \Rightarrow \\ q_k^* &\propto \exp \{ E_{q(z-k)} \ln p(E, Z|\theta) \} \end{aligned} \quad (20)$$

因此, 最优的变分分布为  $q_k^*$ , 其中,  $E_{q(z-k)}$  表示除第  $k$  个隐变量, 对其他隐变量进行期望运算.

另外,  $E_{q(Z)} \left[ \ln \frac{p(E, Z|\theta)}{q(Z)} \right]$  对  $q_k$  的二阶偏导通过计算可表示为如下形式:

$$\begin{aligned} \frac{\partial^2 E_{q(Z)} \left[ \ln \frac{p(E, Z|\theta)}{q(Z)} \right]}{\partial q_k^2} &= \\ \frac{\partial \left\{ \int (E_{q(z-k)} \ln p - \ln q_k - 1) dz_k \right\}}{\partial q_k} &= \\ - \int \frac{1}{q_k} dz_k & \end{aligned} \quad (21)$$

通过式 (20) 迭代更新每个隐变量的变分分布, 可以得到最优的变分分布. 此外, 由式 (21) 可知  $E_{q(Z)} \left[ \ln \frac{p(E, Z|\theta)}{q(Z)} \right]$  对每个  $q_k$  的二阶偏导  $-\int (1/q_k) dz_k$  都小于 0, 因此其对于每个  $q_k$  都是凸的. 通过式 (20) 更新得到的变分分布, 可保证 ELBO 的更新单调不减. 最优变分分布的对数表达式  $\ln q_k^*(Z)$  可表示为

$$\begin{aligned} \ln q_k^*(Z) &= E_{q_j} \ln p(E, Z|\theta) + \text{constant} = \\ E_{q_j} \{ \ln p(E|Z, \theta) + \ln p(Z|\pi) \} &+ \text{constant} \end{aligned} \quad (22)$$

将式 (11) 与式 (12) 代入式 (22), 并将与  $Z$  无关的变量处理到常数项中, 得到

$$\ln q^*(Z) = \sum_{n=1}^N \sum_{k=1}^K z_k^{[n]} \ln \rho_k^n + \text{constant} \quad (23)$$

其中,  $\rho_k^n$  表示第  $k$  个混合模型在数据  $\epsilon_n$  处的概率密度值, 通过对式 (23) 简化, 可以得到最优变分分布  $q^*(Z)$ :

$$q^*(Z) = \prod_{n=1}^N \prod_{k=1}^K (r_k^n)^{z_k^{[n]}} \quad (24)$$

其中,  $r_k^n$  为第  $n$  个数据  $\epsilon_n$  对第  $k$  个混合组份的响应度, 表示  $\epsilon_n$  属于第  $k$  个混合组份的概率, 通过对  $\rho_k^n$  单位化计算得出:

$$r_k^n = p(z_k^{[n]} = 1) = \frac{\rho_k^n}{\sum_{j=1}^K \rho_j^n} \quad (25)$$

### 2.2 考虑数据关联的变分后验分布

传统的 VBI 假设  $\mathcal{Z}$  隐变量之间相互独立, 然而净负荷是时序序列, 其前后数据具有关联性. 因此, 考虑关联性的最优变分分布可以表示为

$$\begin{aligned}
 q^*(z_k) &= q(z_k|z_{k-1}) \Rightarrow \\
 p(z_k^{[n]} = 1) &= p(z_k^{[n]} = 1|z_{k'}^{[n-1]} = 1) = \\
 \frac{p(z_k^{[n]} = 1, z_{k'}^{[n-1]} = 1)}{p(z_{k'}^{[n-1]} = 1)} & \quad (26)
 \end{aligned}$$

其中,  $k'$  表示第  $n-1$  个数据最大响应度的混合组份. 考虑第  $n$  个数据与第  $n-1$  个数据的混合组份的信息, 可以将  $p(z_k^{[n]} = 1, z_{k'}^{[n-1]} = 1)$  定义为

$$\begin{aligned}
 p(z_k^{[n]} = 1, z_{k'}^{[n-1]} = 1) &= \\
 (w \times p(z_k^{[n]} = 1) + w' \times p(z_{k'}^{[n]} = 1)) \times \\
 (w \times p(z_k^{[n-1]} = 1) + w' \times p(z_{k'}^{[n-1]} = 1)) &= \\
 (w \times r_k^n + w' \times r_{k'}^n) \times (w \times r_k^{n-1} + w' \times r_{k'}^{n-1}) & \quad (27)
 \end{aligned}$$

其中,  $w = \frac{\pi}{\pi + \pi'}$  和  $w' = \frac{\pi'}{\pi + \pi'}$  表示第  $k$  和  $k'$  个组份所占的比例. 因此, 通过改进响应度值来改进变分分布可以表示为

$$\begin{aligned}
 \rho_{k'}^n &= \frac{p(z_k^{[n]} = 1, z_{k'}^{[n-1]} = 1)}{p(z_{k'}^{[n-1]} = 1)} = \\
 \frac{(w \times r_k^n + w' \times r_{k'}^n) \times (w \times r_k^{n-1} + w' \times r_{k'}^{n-1})}{r_{k'}^{n-1}} &= \\
 \left( w' + w \frac{r_k^{n-1}}{r_{k'}^{n-1}} \right) \times (w \times r_k^n + w' \times r_{k'}^n) & \quad (28)
 \end{aligned}$$

对  $\rho_{k'}^n$  归一化, 得到归一化之后的响应度  $r_k'^n$ :

$$\begin{aligned}
 r_k'^n &= \frac{\rho_k^n}{\sum_{j=1}^K \rho_j^n} = \\
 \frac{\left( w' + w \frac{r_k^{n-1}}{r_{k'}^{n-1}} \right) \times (w \times r_k^n + w' \times r_{k'}^n)}{\sum_{k=1}^K \left( w' + w \frac{r_k^{n-1}}{r_{k'}^{n-1}} \right) \times (w \times r_k^n + w' \times r_{k'}^n)} & \quad (29)
 \end{aligned}$$

因此,  $r_k'^n$  的上下界可表示为

$$\begin{aligned}
 r_k'^n &\geq \frac{w' \times \min(r_k^n, r_{k'}^n)}{\sum_{k=1}^K \max(r_k^n, r_{k'}^n)} \\
 r_k'^n &\leq \frac{\max(r_k^n, r_{k'}^n)}{w' \times \sum_{k=1}^K \min(r_k^n, r_{k'}^n)} \quad (30)
 \end{aligned}$$

通过式 (25) 可以得到改进之后的响应度  $r_k'^n$  与  $r_k^n$  之间的差值为

$$\begin{aligned}
 r_k'^n - r_k^n &\geq \frac{w' \times \min(r_k^n, r_{k'}^n)}{\sum_{k=1}^K \max(r_k^n, r_{k'}^n)} - r_k^n = LB_k^n \\
 r_k'^n - r_k^n &\leq \frac{\max(r_k^n, r_{k'}^n)}{w' \times \sum_{k=1}^K \min(r_k^n, r_{k'}^n)} - r_k^n = HB_k^n \quad (31)
 \end{aligned}$$

因此, 根据式 (15) 和式 (28) 可以得到改进之后的 MELBO (Modified evidence lower bound) 与原始的 ELBO 差值如下:

$$\begin{aligned}
 MELBO-ELBO &= \\
 E_{q(Z)} [\ln p(E|Z, \mathcal{U})] + E_{q(Z)} [\ln' p(Z|\pi)] - \\
 E_{q(Z)} [q'(Z)] - E_{q(Z)} [\ln p(E|Z, \mathcal{U})] - \\
 E_{q(Z)} [\ln p(Z|\pi)] + E_{q(Z)} [q(Z)] &= \\
 \left( \sum_{n=1}^N \sum_{k=1}^K (r_k'^n - r_k^n) \times \ln \pi_k \right) + \\
 \left( \sum_{n=1}^N \sum_{k=1}^K (r_k^n \ln r_k^n - r_k'^n \ln r_k'^n) \right) &= \\
 \sum_{n=1}^N \sum_{k=1}^K ((r_k'^n - r_k^n) (\ln \pi_k + 1 + \ln \xi_k^n)) & \quad (32)
 \end{aligned}$$

其中,  $\xi_k^n$  介于  $r_k'^n$  与  $r_k^n$  之间, 因此可以得到式 (32) MELBO-ELBO 的上下界:

$$\begin{aligned}
 \sum_{n=1}^N \sum_{k=1}^K ((r_k'^n - r_k^n) (\ln \pi_k + 1 + \ln \xi_k^n)) &\leq \\
 \sum_{n=1}^N \sum_{k=1}^K HB_k^n (1 + \ln [\pi_k \times \max(r_k'^n, r_k^n)]) &= MHB \quad (33)
 \end{aligned}$$

$$\begin{aligned}
 \sum_{n=1}^N \sum_{k=1}^K ((r_k'^n - r_k^n) (\ln \pi_k + 1 + \ln \xi_k^n)) &\geq \\
 \sum_{n=1}^N \sum_{k=1}^K LB_k^n (1 + \ln [\pi_k \times \min(r_k'^n, r_k^n)]) &= MLB \quad (34)
 \end{aligned}$$

MELBO-ELBO 的上下界为 MHB (Modified

high bound) 与 MLB (Modified low bound), 式 (33) 和式 (34) 中, *MHB* 与 *MLB* 分别表示改进上界与改进下界.

然后考虑净负荷数据关联的特征, 采用 EM 算法找到最优的  $q^*(Z)$ . 具体首先求解出变分分布的具体形式. 随后, 固定  $q^*(Z)$ , 使 MELBO 的改变只与参数  $\theta$  有关, 进而通过优化  $\theta$  可以得到最优的对数似然函数  $\int \frac{p(E, Z|\theta)}{q(Z)} \times q(Z)dZ$ , 从而计算出最优的  $\ln p(E|\theta)$ . 其具体的表达式以及收敛性证明可见附录 A.

EM 算法结合净负荷数据关联性的改进后验分布对应的迭代过程如图 8 所示. 首先, 通过式 (14) 得到对数似然值以及证据下确界 ELBO,  $KL(q||p)$  表示变分分布  $q(Z)$  与  $p(Z|E, \theta)$  之间的距离, 如图 8(a) 所示. 经过 E 步, 下界 ELBO 得到更新. 图 8(b) 表示 ELBO 单调递增导致  $KL(q||p)$  减小, 变分分布  $q(Z)$  与目标分布  $p(Z|E, \theta)$  距离缩短. 在得到 ELBO 之后, 计算 DPMM 后验分布时考虑了净负荷数据关联的特点, 因此使用改进之后的变分后验 Modified  $q(Z)$ . 该后验分布会导致图 8(b) 的 ELBO 产生变化, 所以引入改进上界 MHB 与改进下界 MLB 来得到改进之后的证据下界 MELBO 相对于 ELBO 的变化范围, 其中 MELBO 位于 MHB 与 MLB 之间, 如图 8(c) 所示. 经过后验分布的改进, 可得到最终的 MELBO. 最后, 通过 M 步优化参数  $\theta$  来极大化对数似然函数, 如图 8(d) 所示. 需要指出的是由于改进之后的 EM 算法得到的 MELBO 位于 MHB 与 MLB 之间, 会导致其与改进之前的 ELBO 存在差值, 进而在迭代时会发生震荡. 但是考虑了

数据关联的 DDPMM 能够提高收敛速度, 详细的结果将在第 4 节展示.

### 3 算例分析

#### 3.1 算例描述

为了验证提出的数据关联狄利克雷混合模型优越性, 本文根据比利时电网 2019 ~ 2020 年的净负荷数据进行验证, 比利时电网中风电与光伏装机容量为 3 157 MW 和 3 369 MW, 分别占总装机容量的 16.0% 与 17.0%. 图 9 为 2020 年 7 月 1 日的净负荷与净负荷预测曲线以及负荷与负荷预测曲线, 其中需要指出的是预测值由数据集直接给出.

可以看到负荷预测误差较小, 而净负荷误差较大, 因为其包括随机性较强的可再生能源. 为了验证所提出 DDPMM 的优越性, 本文使用 2019 年的净负荷数据作为训练集, 2020 年的净负荷数据作为测试集, 时间分辨率为 15 min. 在本文提出的基于 DDPMM 的非参数贝叶斯框架中, 首先归一化观测数据  $[z, y]^T$ , 其中  $z = x - y$ . 然后, 使用 DDPMM 得到  $[z, y]$  的联合概率分布. 最后根据净负荷的预测值得到预测误差在各预测值下的边缘概率密度, 并通过相关指标来评价 DDPMM 的优劣.

结果部分首先验证了 DDPMM 的收敛性; 进一步, 通过拟合净负荷预测误差, 得到 DDPMM 的拟合效果; 最后通过区间指标来评价 DDPMM 对净负荷预测误差的整体表征效果.

#### 3.2 DDPMM 收敛性分析

DDPMM 利用了时序序列数据关联的信息, 使

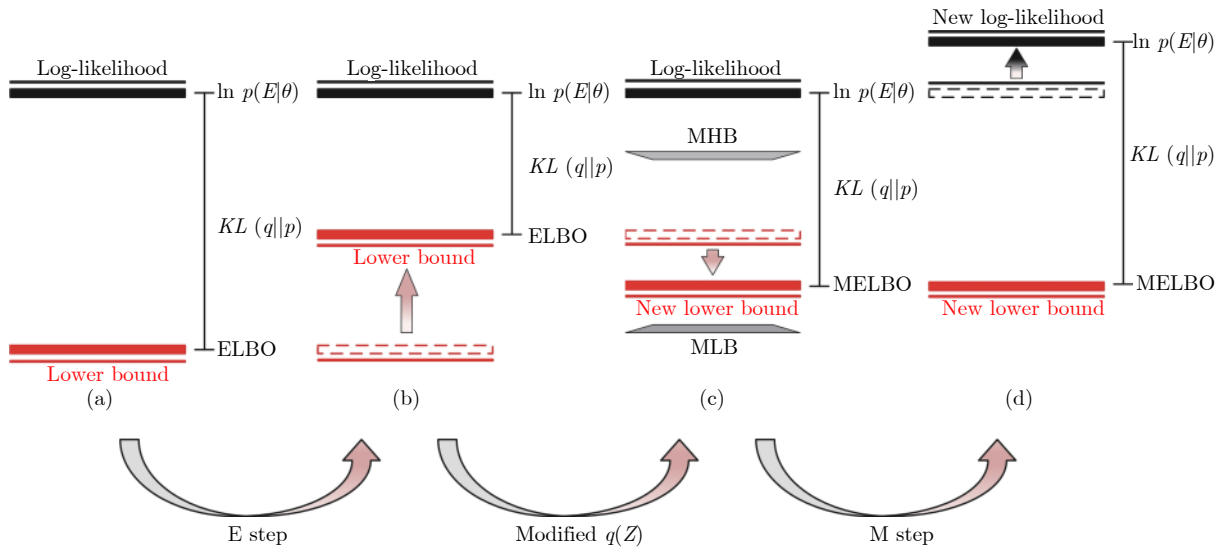


图 8 考虑数据关联的 EM 迭代图  
Fig. 8 EM iterative graph considering data-relevance



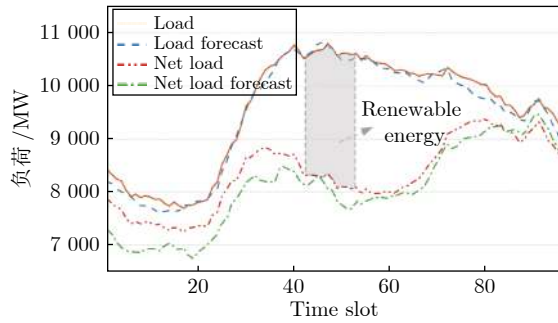


图9 2019年7月1日净负荷与负荷曲线

Fig.9 Net load and load curve on July 1, 2019

收敛速度加快。图10表示了随着迭代次数的增加, DDPMM与DPMM逐渐趋于收敛, 其中纵坐标代表了每次迭代的变化值 $p(E|\theta)$ , 若变化值趋于0, 则可认为已收敛。为了详细地对比DDPMM与DPMM的收敛速度, 本文采用了对数纵坐标轴, 对纵坐标轴进行以10为底的对数运算。从图10中可以观察到DDPMM与DPMM在3000代之后逐渐收敛。其中对于DPMM每代 $p(E|\theta)$ 的变化略大于 $10^{-10}$ , 而DDPMM的变化则更小。

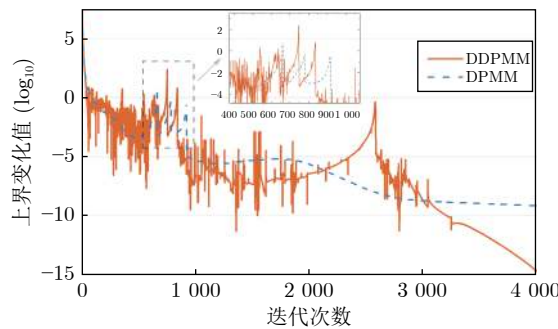


图10 基于2019年净负荷数据的DDPMM与DPMM迭代曲线

Fig.10 DDPMM and DPMM iterative curves based on net load data in the year of 2019

另外, 还可以观察到DDPMM在收敛时反复振荡。这是因为DDPMM在更新的时候考虑了净负荷数据关联的特点, 这会导致DDPMM对应的 $p(E|\theta)$ 变化相较于DPMM是不均匀的, 因此会产生振荡。另外DDPMM算法是基于DPMM改进的, 所以在迭代的前期, DDPMM与DPMM对应的变化趋势大致相同, 如图10中局部图所示。

随着迭代次数的增加, DDPMM由于获取了更多的净负荷数据信息, 其更新会加快。在大概迭代至3000代时, DDPMM的 $p(E|\theta)$ 变化逐渐减小, 而DPMM基本保持不变, 这意味着DDPMM相较于DPMM收敛更快。因此, 从图10的收敛曲线可以看出DDPMM在前期进行迭代的时候, 会牺牲

一部分稳定性去寻找更多优秀的解来加快迭代速度。

### 3.3 DDPMM拟合效果分析

为了比较DDPMM的条件概率分布误差拟合效果, 本文使用GMM与DPMM对其进行对比。其中GMM的组数通过赤池信息准则(Akaike information criterion, AIC)以及贝叶斯信息准则(Bayesian information criterion, BIC)来确定<sup>[24]</sup>, AIC与BIC的具体计算为

$$AIC = 2k - 2\ln(L) \quad (35)$$

$$BIC = k \ln(n) - 2\ln(L) \quad (36)$$

其中,  $k$ 表示模型参数个数,  $L$ 表示模型的似然函数,  $n$ 为净负荷样本数量。AIC与BIC为衡量统计模型拟合优良性的指标。通过计算GMM混合组数为1, 2, ...,  $m$ 个时的AIC与BIC数值(其中 $m$ 表示一个较大的数), 获得 $m_1$ 与 $m_2$ 使其对应得到的AIC与BIC数值最小, 将其记为GMM-AIC与GMM-BIC。

通过计算得到GMM-AIC的组数为20, GMM-BIC的组数为13。另外, 由DPMM得到的最优组数为15, 而DDPMM得到的最优组数为5。这是因为DDPMM拟合净负荷数据 $[x, y]^T$ 时, 考虑了更多的信息, 进而合并相似的高斯分布。但是, DPMM与GMM的混合组数多于DDPMM, 表明其考虑了更多局部的信息, 容易导致过拟合。

本文根据数据样本数目, 将其均分为10份, 每份样本条目的平均值作为每个条件概率分布的净负荷预测值。图11为DDPMM, DPMM与GMM在各预测值下的条件概率分布图, 通过与净负荷数据直方图的比较, 得到各模型在测试集上的拟合效果。

通过图11可以观察到, 当净负荷预测值较小或者较大时, 如图11(a)与图11(j)所示, 净负荷的预测误差较大。GMM模型无法很好的表征其不确定性, 而DDPMM与DPMM在这一方面表现较好。

为了进一步定量比较不同模型的净负荷预测误差拟合效果, 本文采用对数似然(Log-likelihood, Log-L)与卡方拟合优度值(Chi-square goodness of fit, Gof)两个指标, 其中对数似然以 $\ln p(E|\theta)$ 的值为指标, 该指标数值越大代表对应模型拟合效果越好; Gof衡量了两个概率分布之间的差异, Gof数值越小则表示拟合效果差异越小。

表1为4个模型的对数似然值, 在2020年净负荷数据上DDPMM有更大的对数似然值, 表明DDPMM拟合效果更好。表2为4个模型的Gof值, 可以观察到DDPMM表现比较良好, 在图11(a)~(d), 11(g)以及11(j)中Gof最小, 而在其他情况下表现

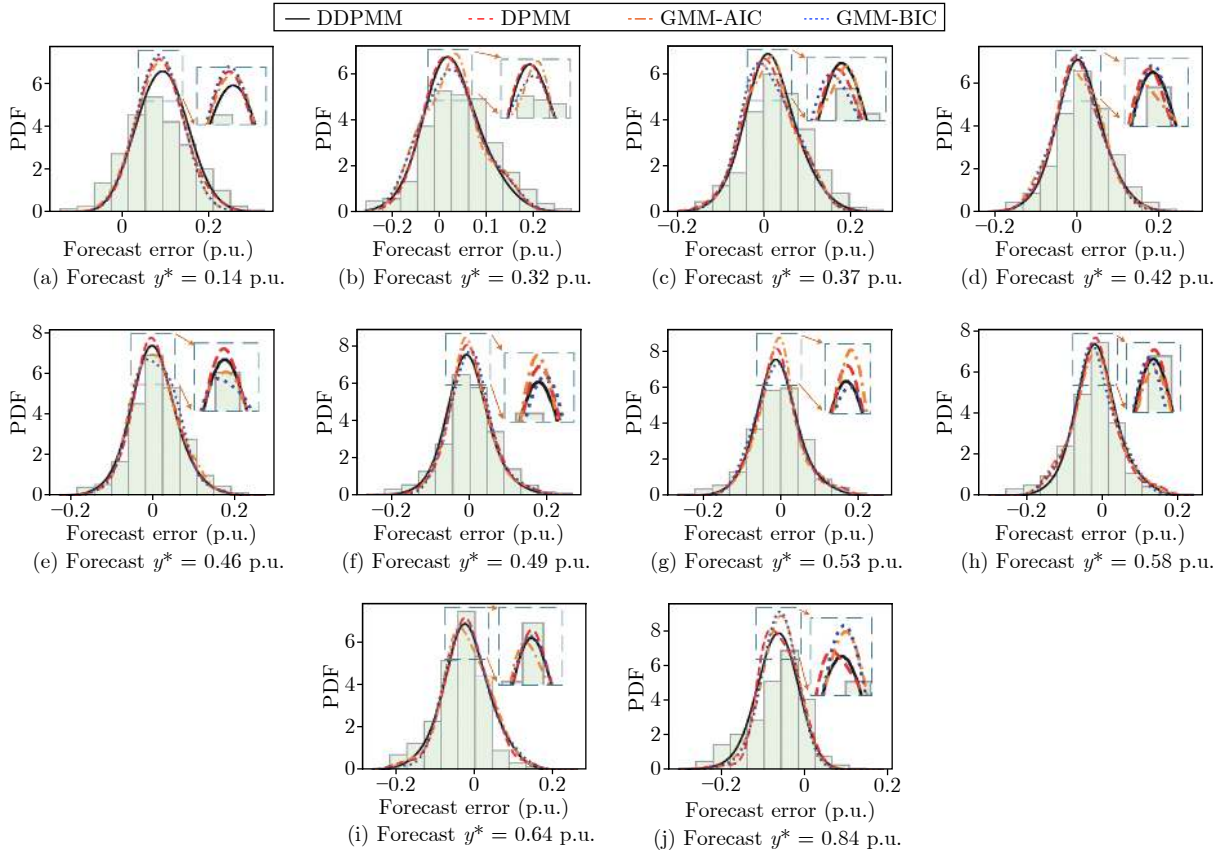


图 11 不同预测值下的净负荷预测误差条件概率分布

Fig. 11 The PDF of net load forecast error conditions under different forecast values

表 1 对数似然比较

Table 1 Comparison of log-likelihood

模型	Log-L/ $10^3$ (test)
DDPMM (5)	1.639
DPMM (15)	1.625
GMM-AIC (20)	1.612
GMM-BIC (13)	1.611

稍弱于 DPMM. 这是因为 Gof 在本文中用来评估图 11 中 10 种预测值下的条件概率分布拟合效果, 主要描述局部信息.

### 3.4 DDPMM 区间指标分析

因此, 为了更好地衡量 DDPMM 表征净负荷

整体不确定性的能力, 本文进一步使用区间指标来评价净负荷预测误差的整体信息. 具体是首先设置预测误差分位数  $\alpha$ , 在图 11 中根据不同预测值所属的预测误差条件概率分布, 得出预测值对应的上下分位数  $X_{\frac{\alpha}{2}}$  与  $X_{-\frac{\alpha}{2}}$ , 从而获取净负荷的预测上下界. 然后通过该预测上下界与实际观测数据  $x_i$  比较, 从而获得净负荷的整体信息, 如区间宽度, 区间覆盖率等指标. 预测值的区间上下界  $U^\alpha(x_i)$  和  $L^\alpha(x_i)$  表示为

$$\begin{aligned}
 U^\alpha(x_i) &= x_i + X_{\frac{\alpha}{2}} \\
 L^\alpha(x_i) &= x_i + X_{-\frac{\alpha}{2}}
 \end{aligned}
 \tag{37}$$

为了评价通过净负荷预测误差概率分布所得到的区间质量, 本文引入 5 个指标来进行评估, 分别

表 2 卡方拟合优度比较

Table 2 Comparison of goodness of fit of Chi-square

Test	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
DDPMM	4.36	3.03	3.58	1.63	1.54	1.96	1.59	1.96	5.32	5.11
DPMM	4.87	3.15	4.46	2.46	1.37	1.79	1.98	1.31	4.92	7.08
GMM-AIC	5.32	3.54	4.20	2.87	2.01	2.01	2.42	1.95	5.94	5.58
GMM-BIC	5.64	3.46	3.95	2.63	2.34	2.53	2.07	2.84	4.99	5.98

为 Winkler 分数、PICP (Predict interval coverage probability)、CWC (Coverage width-based criterion)、MPICD (Mean predict interval center deviation)<sup>[25]</sup>. 其中, PICP 主要衡量区间的覆盖率; Winkler, CWC 以及 AIS 描述了区间覆盖率与区间宽度之间的关系; MPICD 主要描述了预测值处于区间的位置的情况, 如果接近区间的中线则代表效果更好. 其中 Winkler、CWC 和 MPICD 越小越好, 而 PICP 和 AIS 越大越好. 这 5 个指标具体的数学表达式参见附录 B.

本文选择 0.95 的置信度来构造置信区间, 以 3 月、6 月、9 月以及 12 月为典型月份作为测试, 各模型得到的结果如图 12 所示. 可以观察到在 0.95 的置信度下, DDPMM 的指标表现优秀, 其指标基本优于其他三个模型. 表 3 列出了图 12(a) 对应的具体指标值, 可以得出 DDPMM 的 Winkler 分数相比 DPMM, GMM-AIC 与 GMM-BIC 分别提高了 9.7%, 14.2% 和 8.9%, 相对于 DPMM 提升明显. 另外, 还可以观察到 GMM-AIC 的 Winkler 分数最大, 而其混合组数为 20 个, 这说明 GMM-AIC 发生

了过拟合, 而其他三个模型相对来说拟合较好, 其中 DDPMM 拟合效果最好. 另外, CWC 与 AIS 综合考虑了预测区间覆盖率与预测区间的宽度, 可以看到 DDPMM 相较 DPMM 分别提高了 81.7% 与 34.0%, 提升效果显著. PICP 为区间覆盖率, 在 0.95 的置信度下 DDPMM 达到了 0.80, 而 DPMM 为 0.70, 说明 DDPMM 可以更好地表征净负荷不确定性. MPICD 描述了模型是否均匀, 可以看到在这个指标上 DDPMM 与其他三个模型差异并不十分明显, 但是仍优于其余三个模型.

表 4~6 给出了图 12(b)~(d) 的具体区间指标数值, 显然 DDPMM 在 Winkler 分数表现最优, 而 GMM-BIC 模型的 Winkler 分数优于 DPMM. 另外在 PICP 指标上, DDPMM 的优越性较为明显, 相对于 DPMM 在 6 月、9 月和 12 月分别提高了 8.1%, 11% 和 18%. 在 CWC 与 AIS 指标中, DDPMM 在 9 月与 12 月相对于其余三个模型表现更为优秀. 此外, 在 MPICD 指标的对比中, 3 月和 9 月 DDPMM 的表现更好, 说明 DDPMM 可以更好地表征净负荷的不确定性, 并且得到的净负荷预

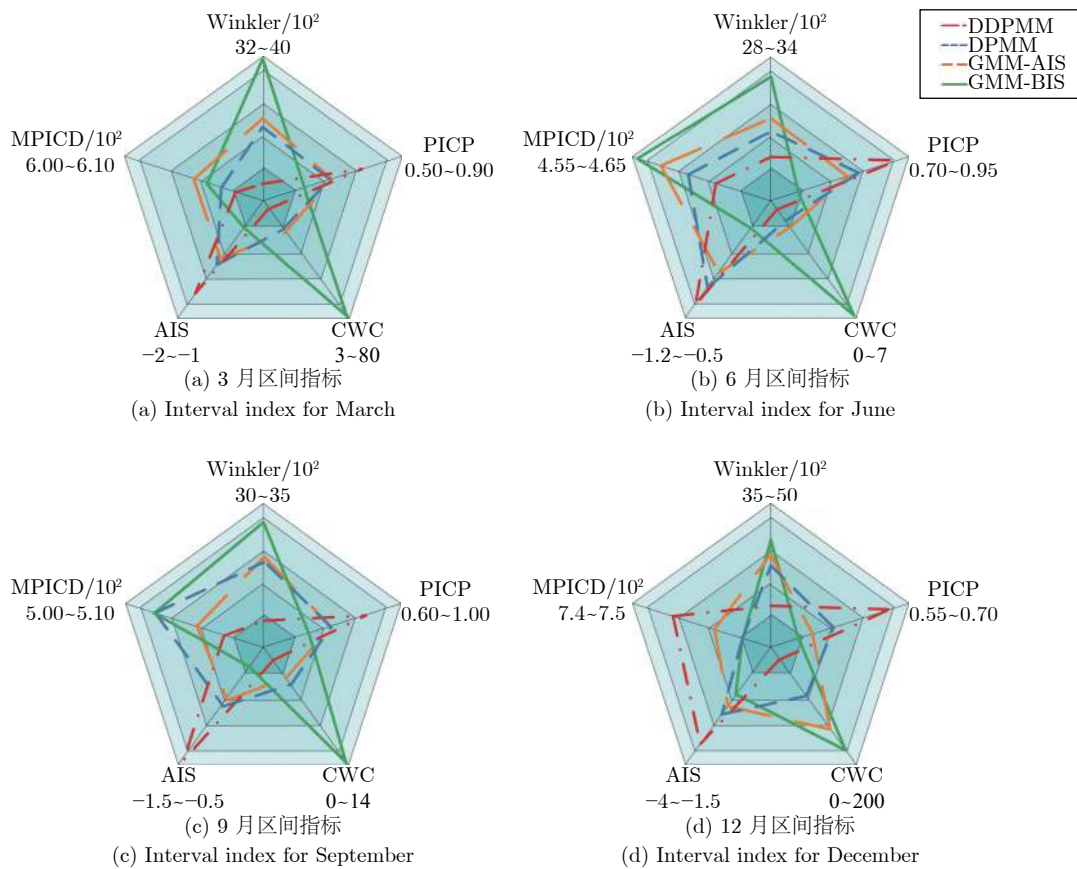


图 12 0.95 置信区间雷达图  
Fig.12 Interval radar chart with 0.95 confidence

表 3 0.95 置信度下 2020 年 3 月区间指标  
Table 3 Interval index for March 2020 with 0.95 confidence level

模型	Winkler/10 <sup>2</sup>	PICP	CWC	AIS	MPICD/10 <sup>2</sup>
DDPMM	33.57	0.80	4.14	-1.03	6.03
DPMM	37.18	0.70	22.58	-1.56	6.05
GMM-AIC	39.11	0.63	78.71	-1.84	6.07
GMM-BIC	36.86	0.69	26.93	-1.51	6.08

表 4 0.95 置信度下 2020 年 6 月区间指标  
Table 4 Interval index for June 2020 with 0.95 confidence level

模型	Winkler/10 <sup>2</sup>	PICP	CWC	AIS	MPICD/10 <sup>2</sup>
DDPMM	29.84	0.93	0.50	-0.58	4.59
DPMM	30.91	0.86	1.18	-0.68	4.61
GMM-AIC	33.19	0.75	6.68	-1.02	4.65
GMM-BIC	31.46	0.84	1.56	-0.77	4.63

表 5 0.95 置信度下 2020 年 9 月区间指标  
Table 5 Interval index for September 2020 with 0.95 confidence level

模型	Winkler/10 <sup>2</sup>	PICP	CWC	AIS	MPICD/10 <sup>2</sup>
DDPMM	30.60	0.89	0.93	-0.70	5.02
DPMM	32.59	0.80	3.71	-0.96	5.03
GMM-AIC	34.90	0.72	13.76	-1.32	5.03
GMM-BIC	32.85	0.79	3.68	-1.01	5.02

表 6 0.95 置信度下 2020 年 12 月区间指标  
Table 6 Interval index for December 2020 with 0.95 confidence level

模型	Winkler/10 <sup>2</sup>	PICP	CWC	AIS	MPICD/10 <sup>2</sup>
DDPMM	39.18	0.72	20.4	-1.97	7.45
DPMM	44.51	0.61	138.23	-2.77	7.44
GMM-AIC	46.02	0.59	173.46	-3.00	7.42
GMM-BIC	43.31	0.64	82.32	-2.59	7.42

表 7 0.8 置信度下 2020 年 6 月区间指标  
Table 7 Interval index for June 2020 with 0.8 confidence

模型	Winkler/10 <sup>2</sup>	PICP	CWC	AIS
DDPMM	32.76	0.77	0.41	-0.91
DPMM	34.70	0.68	1.34	-1.20
GMM-AIC	37.53	0.56	9.86	-1.66
GMM-BIC	35.41	0.66	1.90	-0.32

测上下界更均匀地分布于观测值两侧.

由表 7 和表 8 可以观察到, 在不同的置信度下,

表 8 0.5 置信度下 2020 年 6 月区间指标  
Table 8 Interval index for June 2020 with 0.5 confidence

模型	Winkler/10 <sup>2</sup>	PICP	CWC	AIS
DDPMM	39.84	0.49	0.16	-1.93
DPMM	41.73	0.38	0.73	-2.30
GMM-AIC	44.26	0.31	2.19	-2.73
GMM-BIC	42.33	0.38	0.66	-2.40

DDPMM 都表现优异, 并且它的 PICP 指标最接近置信度, 表明 DDPMM 的预测区间具有良好的区间覆盖度. 更进一步可以观察到, DPMM 与 GMM-BIC 指标相差不大, 可以认为在当前数据集下, DPMM 与 GMM-BIC 得到的混合组数相近, 其中 DPMM 为 15, GMM-BIC 为 13. 然而 DDPMM 只使用 5 个混合组份描述了更多的信息, 说明了 DDPMM 的优势.

通过以上的结果验证了 DDPMM 表征净负荷不确定性问题的优越性. 同 DPMM, GMM-AIC 与 GMM-BIC 相比, DDPMM 得到的混合模型组份数目更少, 获得的有效信息更多, 减少了过拟合对于表征净负荷不确定性的不利影响.

## 4 结论

本文提出了基于数据关联的狄利克雷混合模型, 通过变分推断方法改进了后验分布, 使后验分布考虑了更多的净负荷数据关联信息. 从而构造出改进证据的下界, 使得 DDPMM 通过这一下界得到合适的变分分布, 并结合 EM 算法证明其收敛性.

最后, 通过将比利时电网作为算例, 使用不同的指标来衡量 DDPMM 表征净负荷不确定性的局部与整体信息. 验证结果表明, DDPMM 具有更快的收敛速度, 更优的精度, 可以更好地表征电网负荷的不确定性.

## 附录 A EM 算法收敛性证明

EM 算法主要是通过迭代来逐步近似极大化  $\ln p(E|\theta)$ . 设第  $i$  次迭代之后  $\theta$  的估计值为  $\theta^{(i)}$ , 则希望新的估计值  $\theta$  会使  $\ln p(E|\theta) > \ln p(E|\theta^{(i)})$ , 并逐步达到其极大值. 考虑两者的差, 并利用 Jensen 不等式得到其下界:

$$\begin{aligned} \ln p(E|\theta) - \ln p(E|\theta^{(i)}) &= \\ \ln \int q(Z) \frac{p(E|Z, \theta)p(Z, \theta)}{q(Z)} dZ - \ln p(E|\theta^{(i)}) &\geq \\ \int q(Z) \ln \frac{p(E|Z, \theta)p(Z, \theta)}{q(Z)} dZ - \ln p(E|\theta^{(i)}) &= \\ \int q(Z) \ln \frac{p(E|Z, \theta)p(Z, \theta)}{q(Z)p(E|\theta^{(i)})} dZ &\quad (A1) \end{aligned}$$

令

$$B(\theta, \theta^{(i)}) \ln p(E|\theta^{(i)}) + \int q(Z) \ln \frac{p(E|Z, \theta)p(Z, \theta)}{q(Z)p(E|\theta^{(i)})} dZ \quad (A2)$$

由式 (A1) 和式 (A2) 可以得到  $\ln p(E|\theta) > B(\theta, \theta^{(i)})$ , 故  $B(\theta, \theta^{(i)})$  为  $\ln p(E|\theta)$  的下界,  $B(\theta, \theta^{(i)})$  增大会导致  $\ln p(E|\theta)$  增大. 通过选择合适的  $\theta^{(i+1)}$  使  $B(\theta, \theta^{(i)})$  达到极大值, 从而使  $\ln p(E|\theta)$  增大, 即

$$\theta^{(i+1)} = \arg \max_{\theta} B(\theta, \theta^{(i)}) =$$

$$\arg \max_{\theta} \left\{ \ln p(E|\theta^{(i)}) + \int q(Z) \ln \frac{p(E|Z, \theta)p(Z, \theta)}{q(Z)p(E|\theta^{(i)})} dZ \right\} = \arg \max_{\theta} Q(\theta, \theta^{(i)}) \quad (A3)$$

式 (A3) 为 EM 算法中的 M 步. 通过式 (A3) 可以得到一系列参数估计序列  $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(i)}, \dots\}$ , 并且  $\ln p(E|\theta^{(i)})$  是单调递增的.

证明.

$$\ln p(E|\theta) = \ln p(E, Z|\theta) - \ln p(Z|E, \theta) \quad (A4)$$

令

$$Q(\theta, \theta^{(i)}) = \int q(Z|E, \theta^{(i)}) \ln p(E, Z|\theta) dZ$$

$$H(\theta, \theta^{(i)}) = \int q(Z|E, \theta^{(i)}) \ln p(Z|E, \theta) dZ \quad (A5)$$

因此, 式 (A4) 可以表示为如下形式:

$$\ln p(E|\theta) = Q(\theta, \theta^{(i)}) - H(\theta, \theta^{(i)}) \quad (A6)$$

对式 (A6) 中  $\theta$  取  $\theta^{(i)}$  与  $\theta^{(i+1)}$ , 并相减得:

$$\ln p(E|\theta^{(i+1)}) - \ln p(E|\theta^{(i)}) = [Q(\theta^{(i+1)}, \theta^{(i)}) - Q(\theta^{(i)}, \theta^{(i)})] - [H(\theta^{(i+1)}, \theta^{(i)}) - H(\theta^{(i)}, \theta^{(i)})] \quad (A7)$$

由式 (A3) 可知  $Q(\theta^{(i+1)}, \theta^{(i)}) > Q(\theta^{(i)}, \theta^{(i)})$ , 只需证明  $H(\theta^{(i+1)}, \theta^{(i)}) - H(\theta^{(i)}, \theta^{(i)}) < 0$ .

$$H(\theta^{(i+1)}, \theta^{(i)}) - H(\theta^{(i)}, \theta^{(i)}) = \int q(Z|E, \theta^{(i)}) \ln \frac{p(Z|E, \theta^{(i+1)})}{p(Z|E, \theta^{(i)})} dZ \leq \ln \int q(Z|E, \theta^{(i)}) \frac{p(Z|E, \theta^{(i+1)})}{p(Z|E, \theta^{(i)})} dZ \quad (A8)$$

由于 E 步是对变分分布  $q(Z)$  逐步更新, 当迭代更新的次数变多时, 变分分布与  $p(Z|E, \theta)$  的距离足够接近, 可以认为两者近似相等. 因此, 式 (A8) 可以表达为以下形式:

$$\ln \int q(Z|E, \theta^{(i)}) \frac{p(Z|E, \theta^{(i+1)})}{p(Z|E, \theta^{(i)})} dZ = \ln \int p(Z|E, \theta^{(i+1)}) dZ = 0 \quad (A9)$$

即当迭代次数足够大时, DDPMM 总是向似然函数增大的方向进行更新, 即  $\ln p(E|\theta^{(i+1)}) > \ln p(E|\theta^{(i)})$ . 与此同时, 因为  $p(E|\theta)$  为概率密度函数, 存在上界, 根据单调收敛的性质,  $p(E|\theta)$  会收敛至极大值.  $\square$

## 附录 B 区间指标

### B.1 PICP

PICP 为真实值  $x_i$  落在预测区间的比率, 称为覆盖率. 因为得到预测区间是根据边缘概率分布的上下分位数确定的, 所以无法包含所有的数据点. PICP 可以表示为

$$PICP = \frac{1}{n} \sum_{i=1}^n c_i \quad (B1)$$

当  $x_i \in [L^\alpha(x_i), U^\alpha(x_i)]$  时,  $c_i = 1$ , 否则  $c_i = 0$ .

### B.2 Winkler 分数

Winkler 分数是对预测区间的一种综合度量类型, 它可以同时测量区间覆盖率和区间的宽度, Winkler 分数可以表示为

$$Winkler = \begin{cases} \frac{2(L^\alpha(x_i) - x_i)}{\alpha} + \delta, & x_i < L^\alpha(x_i) \\ \frac{2(x_i - U^\alpha(x_i))}{\alpha} + \delta, & x_i > U^\alpha(x_i) \\ \delta, & \text{其他} \end{cases} \quad (B2)$$

当实际观测值大于上区间或者小于下区间时, Winkler 分数就会得到一个惩罚. 惩罚由  $\alpha$  决定, 令  $\alpha = 0.1$ ,  $\delta = 1$ . 从 Winkler 分数的定义可以看出, Winkler 分数越小说明该模型所受到的惩罚越小, 表明该模型有更好的区间预测效果.

为了更好地衡量各模型的区间覆盖率与区间宽度, 通常情况下希望在分位数确定的情况下, 区间覆盖率越大越好, 而区间宽度越窄越好. 因此, 文中引入 CWC 和 AIS 两个指标来描述这两个特征.

### B.3 CWC

CWC 综合考虑了预测区间的覆盖率与狭窄程度, 其定义为

$$CWC = PINAW \times (1 + \gamma e^{-\eta(PICP - \mu)}) \quad (B3)$$

$$PINAW = \frac{1}{n} \sum_{i=1}^n (U_i^\alpha - L_i^\alpha) \quad (B4)$$

其中,  $PINAW$  为区间平均宽度,  $\gamma$  由  $PICP$  决定:

$$\gamma = \begin{cases} 1, & PICP < \mu \\ 0, & PICP \geq \mu \end{cases} \quad (B5)$$

其中,  $\mu$  为置信水平,  $\eta$  为惩罚系数.  $\eta$  越大,  $PICP$  没有达到置信水平时的惩罚就越大. 本文将  $\eta$  设置为 5.

### B.4 AIS

AIS 为另一个综合考虑覆盖率与区间宽度的指标, AIS 对应的第  $i$  个数据点预测区间的区间分数  $S^\alpha(x_i)$  的定义为

$$S^\alpha(x_i) = \begin{cases} -2\alpha\xi_i^{(\alpha)} - 4(L^\alpha(x_i) - x_i), & p_i < L^\alpha(x_i) \\ -2\alpha\xi_i^{(\alpha)}, & PICP \geq \mu \\ -2\alpha\xi_i^{(\alpha)} - 4(x_i - U^\alpha(x_i)), & p_i > U^\alpha(x_i) \end{cases} \quad (B6)$$

其中,  $\xi_i^{(\alpha)}$  为第  $i$  个预测区间的宽度, 当目标不在预测区间覆盖范围时, 会给予一定的惩罚. AIS 的数值为平均区间得分, 可以通过下式计算:

$$\bar{S}^{(\alpha)} = \frac{1}{n} \sum_{i=1}^n S^{(\alpha)}(x_i) \quad (\text{B7})$$

AIS 值越大说明预测区间的品质越好.

## B.5 MPICD

MPICD 表示观测数据偏离预测区间中线的程度, 其计算式为

$$MPICD = \frac{1}{n} \sum_{i=1}^n \left| \frac{U^\alpha(x_i) + L^\alpha(x_i)}{2} - x_i \right| \quad (\text{B8})$$

MPICD 越小, 表明  $x_i$  越接近预测区间中线.

## References

- Huang Bo-Nan, Wang Yong, Li Yu-Shuai, Liu Xin-Rui, Yang Chao. Multi-objective optimal scheduling of integrated energy systems based on distributed neurodynamic optimization. *Acta Automatica Sinica*, DOI: [10.16383/j.aas.c200168](https://doi.org/10.16383/j.aas.c200168) (黄博南, 王勇, 李玉帅, 刘鑫蕊, 杨超. 基于分布式神经动态优化的综合能源系统多目标优化调度. 自动化学报, DOI: [10.16383/j.aas.c200168](https://doi.org/10.16383/j.aas.c200168))
- Tang Hao, Liu Chang, Yang Ming, Tang Bi-Qiang, Xu Dan, Lv Kai. Learning-based optimization of active distribution system dispatch in industrial park considering the peak operation demand of power grid. *Acta Automatica Sinica*, 2021, **47**(10): 2449–2463 (唐昊, 刘畅, 杨明, 汤必强, 许丹, 吕凯. 考虑电网调峰需求的工业园区主动配电系统调度学习优化. 自动化学报, 2021, **47**(10): 2449–2463)
- Jiang Zhao-Yu, Jia Qing-Shan, Guan Xiao-Hong. A review of multi-temporal-and-spatial-scale wind power forecasting method. *Acta Automatica Sinica*, 2019, **45**(1): 51–71 (姜兆宇, 贾庆山, 管晓宏. 多时空尺度的风力发电预测方法综述. 自动化学报, 2019, **45**(1): 51–71)
- Sun M Y, Zhang T Q, Wang Y, Strbac G and Kang C Q. Using Bayesian deep learning to capture uncertainty for residential net load forecasting. *IEEE Transactions on Power Systems*, 2019, **35**(1): 188–201
- Wang Y, Zhang N, Chen Q X, Kirschen D S, Li P, Xia Q. Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV. *IEEE Transactions on Power Systems*, 2017, **33**(3): 3255–3264
- Rajbhandari N, Li W F, Du P W, Sharma S, Blevins B. Analysis of net-load forecast error and new methodology to determine non-spin reserve service requirement. In: Proceedings of the 2016 IEEE Power and Energy Society General Meeting, Boston, MA, USA: IEEE, 2016: 1–5
- Alvarado-Barrios L, del Nozal A R, Valerino J B, Vera I G, Martinez-Ramos J L. Stochastic unit commitment in microgrids: Influence of the load forecasting error and the availability of energy storage. *Renewable Energy*, 2020, **146**: 2060–2069
- Tang C H, Xu J, Sun Y Z, Liu J, Li X, Ke D P, Yang J, Peng X T. A versatile mixture distribution and its application in economic dispatch with multiple wind farms. *IEEE Transactions on Sustainable Energy*, 2017, **8**(4): 1747–1762
- Dong B, Li Z X, Rahman S M M, Vega R. A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings*, 2016, **117**: 341–351
- Zhao Shu-Qiang, Zhang Ting-Ting, Li Zhi-Wei, Li Dong-Xu, Xu Xiao-Yan, Liu Jin-Shan. Distribution model of day-ahead photovoltaic power forecasting error based on numerical characteristic clustering. *Automation of Electric Power Systems*, 2019, **43**(13): 36–45 (赵书强, 张婷婷, 李志伟, 李东旭, 许晓艳, 刘金山. 基于数值特性聚类的日前光伏出力预测误差分布模型. 电力系统自动化, 2019, **43**(13): 36–45)
- Yang Y, Wu W C, Wang B, Li M J. Analytical reformulation for stochastic unit commitment considering wind power uncertainty with Gaussian mixture model. *IEEE Transactions on Power Systems*, 2020, **35**(4): 2769–2782
- Sun W G, Zamani M, Hesamzadeh M R, Zhang H T. Data-driven probabilistic optimal power flow with nonparametric Bayesian modeling and inference. *IEEE Transactions on Smart Grid*, 2020, **11**(2): 1077–1090
- Brockwell P J, Davis R A, Berger J O, Fienberg S E, Singer B. *Time Series: Theory and Methods*. Berlin: Springer-Verlag, 2015. 48–59
- Kun Il Park. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*, Berlin: Springer, 2018. 171–183
- Han L, Jing H T, Zhang R C, Gao Z Y. Wind power forecast based on improved long short-term memory network. *Energy*, 2019, **189**: 116300
- Blei D M, Jordan M I. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 2006, **1**(1): 121–143
- Li Z L, Li Y Z, Liu Y, Wang P, Lu R Z, Gooi H B. Deep learning based densely connected network for load forecasting. *IEEE Transactions on Power Systems*, 2021, **36**(4): 2829–2840
- McLachlan G J, Peel D. *Finite Mixture Models*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc. 2000, 521–541
- Bishop C M. Pattern recognition and machine learning. *Journal of Electronic Imaging*, 2006, **16**(4): 140–155
- Bouguila N, Ziou D. A Dirichlet process mixture of generalized Dirichlet distributions for proportional data modeling. *IEEE Transactions on Neural Networks*, 2010, **21**(1): 107–122
- Sethuraman J. A constructive definition of the Dirichlet prior. *Statistica Sinica*, 1994, **4**(2): 639–650
- Dong H, Dwk B, Csp C. Prior selection method using likelihood confidence region and Dirichlet process Gaussian mixture model for Bayesian inference of building energy models. *Energy and Buildings*, 2020, **224**: 110293
- Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1976, **19**(6): 716–723
- Wang Y J, Liu Q. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock-recruitment relationships. *Fisheries Research*, 2006, **77**(2): 220–225
- Khosravi A, Nahavandi S, Creighton D, Atiya A F. Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Transactions on Neural Networks*, 2011, **22**(3): 337–346



**李远征** 华中科技大学人工智能与自动化学院副教授. 主要研究方向为人工智能及其在智能电网中的应用, 深度学习, 强化学习, 大数据分析.  
E-mail: yuanzheng\_li@hust.edu.cn  
(**LI Yuan-Zheng** Associate professor at the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. His research interest covers artificial intelligence and its application in smart grid, deep learning, reinforcement learning, and big data analysis.)

His research interest covers artificial intelligence and its application in smart grid, deep learning, reinforcement learning, and big data analysis.)



**孙天乐** 华中科技大学人工智能与自动化学院硕士研究生. 主要研究方向为新能源不确定性表征, 在线预测.

E-mail: stl\_221@163.com

**(SUN Tian-Le** Master student in system engineering at the School of Artificial Intelligence and Automation,

Huazhong University of Science and Technology. His research interest covers random characterization of renewable energy and online prediction.)



**刘云** 博士, 华南理工大学电力学院副教授. 主要研究方向为高比例新能源分布式协同控制, 主动配电网 P2P 能量交易, 电力信息安全与隐私防护, 人工智能在能源系统的应用.

E-mail: liuyun19881026@gmail.com

**(LIU Yun** Ph.D., associate professor

at South China University of Technology. His research interest covers distributed cooperative control of renewable generation, peer-to-peer energy trading, security of cyber physical system, and application of artificial intelligence in energy systems.)



**赵勇** 华中科技大学人工智能与自动化学院教授. 主要研究方向为决策理论、方法及应用, 大型工程项目管理, 社会经济系统的建模与仿真, 系统分析与集成.

E-mail: zhiwei98530@hust.edu.cn

**(ZHAO Yong** Professor at the

School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. His research interest covers decision-making theories, methods and applications, large-scale engineering project management, modeling and simulation of social economic systems, and system analysis and integration.)



**曾志刚** 华中科技大学人工智能与自动化学院院长, 长江学者特聘教授, 国家杰出青年基金获得者. 主要研究方向为切换系统控制理论与应用, 计算智能, 系统稳定性, 联想记忆. 本文通信作者.

E-mail: zgzeng@hust.edu.cn

**(ZENG Zhi-Gang** Dean and professor of the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. His research interest covers switching system control theory and application, computational intelligence, system stability, and associative memory. Corresponding author of this paper.)