

基于数据驱动的非监测用户用电模式识别方法

李凯¹ 杨大伟¹ 张建业² 马崇瑞³ 李德高² 王慧³

¹(国网新疆电力有限公司信息通信公司 新疆 乌鲁木齐 830000)

²(国网新疆电力有限公司 新疆 乌鲁木齐 830000)

³(北京中电普华信息技术有限公司 北京 100000)

摘要 借助于终端用户侧安装的智能电表能够有效地分析其异常用电行为和用电模式,为填补在过渡期可能存在的用户数据缺失,提出一种基于数据驱动的非监测用户用电模式识别方法。利用装有智能电表用户的典型日负荷曲线历史数据提取典型用电模式;对多时间尺度机器学习模型进行训练来估计用户用电量;采用递归贝叶斯学习和支路电流状态估计残差法,从非监测用户月度电费账单中获得日负荷曲线。采用实际系统的量测数据进行算例验证,仿真结果表明所提出方法能够快速而准确地识别出非监测用户的用电模式。

关键词 用电模式识别 频谱聚类 递归贝叶斯学习

中图分类号 TP3 文献标志码 A DOI:10.3969/j.issn.1000-386x.2024.05.016

DATA-DRIVEN UNMONITORED USER ELECTRICITY CONSUMPTION PATTERN RECOGNITION METHOD

Li Kai¹ Yang Dawei¹ Zhang Jianye² Ma Chongrui³ Li Degao² Wang Hui³

¹(Information and Communication Company, State Grid Xinjiang Electric Power Co., Ltd., Urumqi 830000, Xinjiang, China)

²(State Grid Xinjiang Electric Power Co., Ltd., Urumqi 830000, Xinjiang, China)

³(Beijing Zhongdian Puhua Information Technology Co., Ltd., Beijing 100000, China)

Abstract Smart meters installed on the end user side can effectively analyze their abnormal power consumption behavior and power consumption patterns. To fill the user data missing that may exist in the transition period, a data-driven unmonitored user power pattern recognition method is proposed. We used the typical daily load curve historical data of users with smart meters to extract the typical power consumption patterns, and trained multi-time scale machine learning models to estimate the monthly consumption of users. The recursive Bayesian learning and branch current state estimation residual method were used to obtain the daily load curve from the monthly electricity bill of the unmonitored user. The simulation results on measurement data from actual systems show that the proposed method can identify the power consumption mode of unmonitored users quickly and accurately.

Keywords Electricity consumption pattern recognition Spectrum clustering Recursive Bayes learning

0 引言

先进计量设备保障了电力公司能够开展用电监测、需求侧控制和窃电行为识别等工作^[1]。但考虑到经济和通信系统的约束,目前配电网中的智能电表数

量仍然较为有限,这就很大程度上限制了配电网的可观测性。因此电力公司迫切需要在智能电表未完全普及前,识别非监测用户的用电行为模式,估计其典型日负荷曲线^[2]。

现有针对用户用电模式识别的研究根据其在聚类中所使用数据的时间尺度可分为基于小时级数据和基

于月度电费账单两类估计方法。文献[3]提出了一种基于 K 均值的负荷估算方法,利用半小时尺度的用电数据来估算缺失的用电量测量值。文献[4]基于概率神经网络,利用用户的外部信息(例如周末的营业时间和经济法规)在特定组之间分配各自的负荷曲线。文献[5]通过将日均用电量与从可监测用户处获得的典型负荷曲线所对应的日均用电量进行比较,从而将典型负荷曲线与相应的无监测用户匹配。根据每月的用电量水平,可以从居民、商业和工业用户的历史数据中统计分析得到无监测用户的典型日负荷曲线^[6]。上述研究一方面对于实时量测单元的覆盖率有较高要求,而另一方面又假定用户日均用电量相等,因此难以获得无监测用户准确的用电模式。

由此,论文提出了一种基于频谱聚类的多时间尺度学习架构,实现在仅已知用户月度电费账单情况下,识别用户用电模式,获得无监测用户的典型日负荷曲线估计值。首先利用装有智能电表用户处采集的历史量测数据进行频谱聚类,得到典型日负荷曲线,从而构建用电模式库;然后利用每一类中的用户负荷曲线,训练所提出的多时间尺度学习模型将用户月度用电量经过逐层机器学习模型进行不同时间尺度的细化分解,从月到周再到日最后到小时;最后考虑到缺乏无监测用户的实时数据,采用支路电流状态估计法来识别该用户可能的用电模式,并将状态估计的残差进行递归贝叶斯学习从而获得该用户属于所有类别的概率,选择其中概率最高的类别作为该无监测用户的用电模式。

1 无监测用户用电模式识别框架

所提出的无监测用户用电模式识别框架包含以下三个阶段,如图 1 所示。

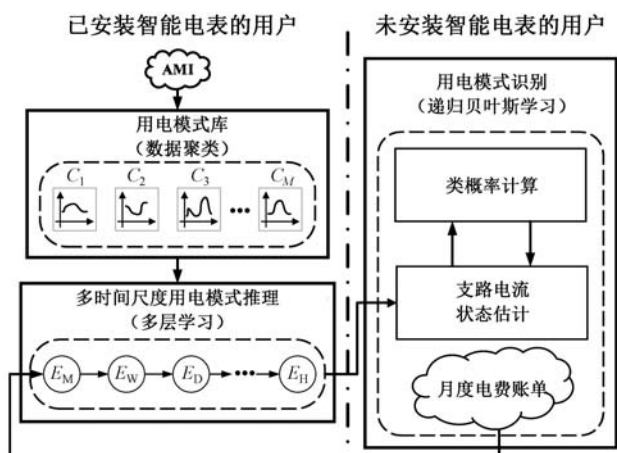


图 1 用电模式识别框架

第一阶段:用电模式库构建。采用频谱聚类算法

检测各典型日负荷曲线的相似性。聚类结果 $\{C_1, C_2, \dots, C_M\}$ 根据用户类型存储在特定的用电模式库中,每个类代表一个典型日负荷曲线。

第二阶段:多时间尺度用电模式推理。针对每个类别的用户,使用可监测用户的智能电表数据对单独的多层多时间尺度学习模型进行了训练,将月度电费数据转换为小时级用电量数据。在不同时间尺度上的用电量被定义为每月用电量 E_M ,每周用电量 E_W ,每日用电量 E_D 和每小时用电量 E_H 。每月数据被视为模型第一层的输入,每小时用电量变量为最后一层的输出。

第三阶段:用电模式匹配。为了从用电模式库中确定最匹配无监测用户的用电模式,使用基于支路电流状态估计的递归贝叶斯学习方法来识别这些用户的日度负荷情况。观察每个用电模式的测量残差,并将其用于无监测用户与相匹配的日常用电模式之间建立联系。基于可监测残差,采用递归贝叶斯学习方法为每个无监测用户递归地将概率值分配给每个典型用电模式。然后,将具有最高概率的模型识别为“匹配的”用户用电模式。

2 用电模式聚类

根据图拉普拉斯算子的性质,频谱聚类算法采用图矩阵的特征向量进行数据重建。这种重建过程增强了数据中的聚类属性,因此可以很容易地从重建数据集中检测聚类,降低了聚类过程对异常值的敏感性,基于频谱聚类的用电典型模式识别方法的主要步骤如下:

步骤 1 将量测数据集转换为相似度图 $G = (V, E)$,该相似度图由一组顶点 V 和一组连接不同顶点的边 E 组成。论文中,顶点 V 是由可监测用户的平均日负荷曲线确定的, V_i 是第 i 个用户一天内每个小时的平均用电量。每个顶点 V_i 的局部 $\alpha_i = \|V_i - V_k\|$, V_k 是顶点 V_i 的第 k 个邻居,这使得顶点与顶点之间距离能够自调整。

步骤 2 利用高斯核函数,图 $W = (w_{ij})$, $i, j = 1, 2, \dots, n$ 的加权邻接矩阵计算如下:

$$\omega_{ij} = \exp\left(-\frac{\|V_i - V_j\|^2}{\alpha_i \alpha_j}\right) \quad (1)$$

步骤 3 建立加权邻接矩阵后,频谱聚类算法就将聚类过程转换为图分区问题。通过移除连接两组的边将图分为 k 个不相交的顶点集。当不同集合之间的边缘权重较低,而集合内的边缘权重较高时,则可以得到最佳的图形分区。因此,图分区的目标功能是最大

化不同聚类之间的差异以及每个类内的总相似度:

$$N(G) = \min_{A_1, A_2, \dots, A_\eta} \sum_{i=1}^n \frac{c(A_i, \bar{A}_i)}{d(A_i)} \quad (2)$$

式中: η 是顶点数; A_i 是属于 V 的子集; $c(A_i, \bar{A}_i)$ 是顶点 A_i 与其余子集顶点间的权重和; $d(A_i)$ 是 A_i 中顶点权重的总和。

$N(G)$ 的最小值在拉普拉斯矩阵的第二最小特征向量上取到。图拉普拉斯矩阵是频谱聚类算法的主要元素,它是使用邻接矩阵 W 和对角矩阵 D 构造的,对角矩阵 D 的第 (i, i) 个元素是 W 第 i 行的总和。归一化的图拉普拉斯算子为:

$$L = D^{-1/2} W D^{-1/2} \quad (3)$$

步骤 4 计算拉普拉斯矩阵的特征向量 $[y_1, y_2, \dots, y_n]$, 并选择对应于最小特征值的特征向量。提取前 k 个特征向量建立新矩阵 $Y \in \mathbf{R}^{n \times k}$, 由于图拉普拉斯算子的性质, 顶点 V_i 由 Y 矩阵的第 i 行表示。这种表示形式的改变增强了数据聚类属性, 然后利用 k -means 聚类算法来获取原始顶点 V_i 的 k 个对应簇。

步骤 5 为了找到最佳的分区, 论文采用 Davies-Bouldin 验证索引 (Davies-Bouldin Validation Index, DBI), 该方法通过测量集群内和集群间相似度之比来校准频谱聚类算法。使用不同的 k 值重复步骤 4, 并记录每个 k 的对应 DBI 值。选择 DBI 最小的 k 值作为最佳群集数^[7]。对其余数据子集执行同样的过程, 从而确定典型用电模式的数量。

3 用电模式推理

利用第 2 节中得到的模式库中的典型用电模式数据, 将每月用电量数据映射到每个类别中的用户每小时用电量。虽然不能直接在月度水平上观察到每小时负荷的变化, 但是多层结构 (其中每个层对应于不同时间尺度的总用电量) 能够在月度和小时数据之间建立关联联系。因此, 以保持不同层的输入-输出之间的高相关水平来维持逐层估计精度的方式构造多时间尺度学习。为了识别具有高相关系数水平的变量以设计多时间尺度学习的结构, 对量测数据集进行了基本统计分析, 如表 1 所示。时间尺度定义为每月用电量 E_M , 每周用电量 E_W , 工作日用电量 E_{Dw} , 周末用电量 E_{Dnw} , 工作日每小时用电量 E_{Hw} 和周末每小时用电量 E_{Hnw} , 并通过智能电表的历史数据来获得。对于不同类型的用户, 相关系数值见表 1, 并确定如下:

$$\rho(X, Y) = \left| \frac{\sigma_{X,Y}^2}{\sigma_X \sigma_Y} \right| \quad (4)$$

式中: X 和 Y 是在特定时间范围内可监测用户的用电量水平, 例如每月或每周的用电量; $\sigma_{X,Y}^2$ 是 X 和 Y 的协方差; σ_X 为标准差。

表 1 多时间尺度用电模式统计分析

层数	相关系数	工业负荷	商业负荷	居民负荷
第一层	$\rho(E_M, E_W)$	0.974 4	0.992 1	0.961 3
	$\rho(E_W, E_W)$	0.960 0	0.984 3	0.930 9
第二层	$\rho(E_W, E_{Dw})$	0.976 4	0.987 5	0.940 0
	$\rho(E_{Dw}, E_{Dw})$	0.967 7	0.986 2	0.898 3
	$\rho(E_W, E_{Dnw})$	0.923 4	0.950 0	0.928 1
	$\rho(E_{Dnw}, E_{Dnw})$	0.924 1	0.977 1	0.887 1
第三层	$\rho(E_{Dw}, E_{Hw})$	0.949 8	0.942 9	0.774 7
	$\rho(E_{Hw}, E_{Hw})$	0.983 8	0.979 3	0.788 2
	$\rho(E_{Dnw}, E_{Hnw})$	0.957 3	0.966 7	0.772 8
	$\rho(E_{Hnw}, E_{Hnw})$	0.988 1	0.983 3	0.796 0

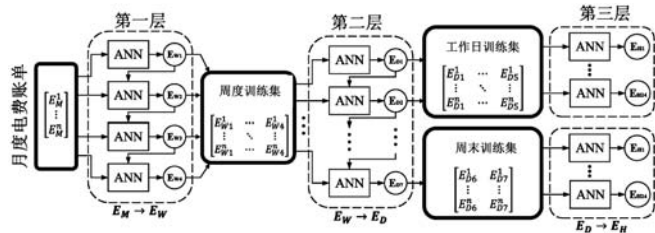


图 2 多时间尺度学习模型结构

通过相关性分析, 为每类用户开发了一个三层多时间尺度学习模型结构, 如图 2 所示。第一层使用串联的 ANN 将每月总用电量 E_M 转换为的周用电量集合 $E_W = \{E_{W1}, E_{W2}, \dots, E_{W4}\}$ 。为获得连续几周的用电量之间的时间相关性, 将每周的估计用电量反馈送到与下一周的用电量相对应的 ANN, 即:

$$E_{W_i} = ANN(E_M, E_{W(i-1)}) \quad (5)$$

利用第一层所得到的周度训练集, 在第二层中通过 ANN 将每周用电量 E_W 转换为日用电量 $E_D = \{E_{D1}, E_{D2}, \dots, E_{D7}\}$ 。然后, 根据工作日和周末不同的用户行为, 对第三层进行了训练, 以将每日总用电量映射到每小时用电量 $E_H = \{E_{H1}, E_{H2}, \dots, E_{H24}\}$ 。论文使用 Levenberg-Marquardt (LM) 反向传播方法来更新网络权重和偏差变量^[8], 由此训练目标函数和更新方程为:

$$\min_b F(b) = \sum_{i=1}^Q v_i^2(b) = \mathbf{v}^T(b) \mathbf{v}(b) \quad (6)$$

$$\Delta b_l = -[J^T(b_l) J(b_l) + \mu_l I]^{-1} J^T(b_l) \mathbf{v}(b_l) \quad (7)$$

式中: μ_l 是迭代 l 的组合参数; b 是学习参数的集合; J 是训练目标函数的雅可比行列式; I 是识别矩阵; \mathbf{v} 是误差向量。

在每次迭代中,基于近似性能指标 $F(b)$ 的变化更新 μ_i 的值。如果获得较小的值,则将 μ_i 除以某个因子 $\theta > 1$ 。否则,将 μ_i 乘以 θ 以进行下一次迭代。对于每个 ANN,数据集被随机分为三个单独的子集进行训练(占总数据的 70%)、验证(占总数据的 15%)和测试(占总数据的 15%)。为了校准每个 ANN 的超参数,我们使用网格搜索方法来获得 LM 的四个重要参数:隐藏层的数量,神经元的数量,增加因子 θ 和减少因子 $1/\theta$ 的值。而为了克服过拟合问题,论文采用了两种措施:1) 早期停止机制。在验证过程中,当验证误差开始增大时就提前终止本次迭代训练从而尽早避免 ANN 出现过拟合。2) 噪声注入机制。在测试过程中,向训练集注入干扰数据从而提高 ANN 的鲁棒性。

4 用电模式识别

由于配电网中普遍存在的实时数据源是有限数量的馈线级量测数据,例如 SCADA 电压和电流。为了仅使用馈线级量测数据来识别和分配相应的日典型用电模式和长时间尺度学习给无监测用户,论文提出了基于支路电流状态估计的递归贝叶斯学习方法。该学习算法使用支路电流状态估计算法的残差为无监测用户计算每种典型负荷模式的概率,然后根据概率值将最可能的类别确定为无监测用户用电模式。

4.1 支路电流状态估计

支路电流状态估计通过采用支路电流作为状态变量来提高计算效率和存储要求,论文采用应用广泛的加权最小二乘法来获得系统节点负荷的估计,其目标函数定义如下:

$$\min_x J = (\mathbf{z} - h(\mathbf{x}))^T \Sigma (\mathbf{z} - h(\mathbf{x})) \quad (8)$$

式中: \mathbf{z} 是测量向量; $\mathbf{x} = [I_r, I_x]$ 是状态向量; I_r, I_x 分别为支路电流的实部和虚部; h 是与测量 \mathbf{z} 相关的非线性测量函数。支路电流状态估计的残差矢量定义为实测值与估计值之差 $\mathbf{r} = \mathbf{z} - h(\mathbf{x})$, Σ 表示测量精度的权重矩阵。通常,使用测量误差的方差 ϕ^2 来构建 Σ , 论文中 $\Sigma = \text{diag} \{ \phi_1^{-2}, \phi_2^{-2}, \dots, \phi_s^{-2} \}$, s 为 \mathbf{z} 的基数。

采用高斯-牛顿法来求解这一非凸优化问题,算法的迭代过程如下:

$$\mathbf{G}(\mathbf{x}) = \mathbf{H}^T(\mathbf{x}) \Sigma \mathbf{H}(\mathbf{x}) \quad (9)$$

$$[\mathbf{G}(\mathbf{x}^m)] \Delta \mathbf{x}^m = \mathbf{H}^T(\mathbf{x}^m) \Sigma (\mathbf{z} - h(\mathbf{x}^m)) \quad (10)$$

$$\mathbf{x}^{m+1} = \mathbf{x}^m + \Delta \mathbf{x}^m \quad (11)$$

式中: \mathbf{H} 是测量函数 $h(\mathbf{x})$ 的雅可比矩阵; \mathbf{G} 是增益矩阵; m 是迭代次数。

4.2 基于递归贝叶斯学习的用电模式匹配

为了确定无监测用户的用电模式,引入递归贝叶斯学习,其步骤如下:

步骤 1 从用电模式库中为无监测用户 j 选择一个类别 i 。

步骤 2 使用所选类别的多时间尺度学习从用户的月度电费账单估计其小时级负荷曲线。

步骤 3 利用日负荷曲线估计值进行支路电流状态估计。通过将实际测量值与估计值进行比较,获得残差。

步骤 4 定义概率 $p_{i,j}$: 类别 i 是用户 j 的准确日均用电量的概率。 $p_{i,j}$ 的初始值为 $1/N$, N 是特定用户类型的多时间尺度学习模型数量。应用贝叶斯定理并假设测量误差为高斯分布,可以得到概率更新的递归表达式:

$$p_{ij}^o = \frac{\exp\left(-\frac{1}{2} \mathbf{r}_{ij}^{oT} \Phi \mathbf{r}_{ij}^o\right) p_{ij}^{o-1}}{\sum_{i=1}^N \exp\left(-\frac{1}{2} \mathbf{r}_{ij}^{oT} \Phi \mathbf{r}_{ij}^o\right) p_{ij}^{o-1}} \quad (12)$$

式中: o 是迭代次数; $\mathbf{r}_{i,j}^o$ 是第 i 类相对于第 j 个用户的残差向量; $\Phi = \text{diag} \{ \sigma_{ri,j,R}^2, \sigma_{ri,j,I}^2 \}$ 是与残差分量相对应的方差对角矩阵; $\sigma_{ri,j,R}^2, \sigma_{ri,j,I}^2$ 是支路电流实部残差和虚部残差的方差。

步骤 5 返回步骤 1。

步骤 6 将无监测用户 i^* 的日负荷曲线标识为最可能的类别: $i^* = \text{argmax}_i p_i^j$ 。

步骤 7 对所有无监测用户重复上述过程,直到确定所有用户的典型日负荷曲线。

步骤 8 利用无监测用户对应类的小时级负荷估计值来进行在线支路电流状态估计,从而实现对该系统的实时监测。

5 算例仿真与分析

本节将所提出的无监测用户用电模式识别方法应用于某实际配网馈线系统中,并与文献[5]中的方法进行对比。该系统包含三类负荷:工业负荷(3%)、商业负荷(20%)和居民负荷(77%),共计 6 000 多个用户,小时级负荷(用电量)数据范围为 2015 年 1 月至 2018 年 5 月。对数据进行预处理除去由于通信错误而导致的丢失数据,并将量测数据集划分为六个子集,

分别对应于居民、商业和工业用户在工作日和周末的负荷曲线。

5.1 训练参数校准

为了校准频谱聚类法和 ANN 的参数,利用 DBI 索引和网格搜索来找到最佳参数。采用最小 DBI 值获得频谱聚类方法的最优簇数 k ,采用网格搜索方法确定 ANN 的最优超参数集,如图 3 所示。

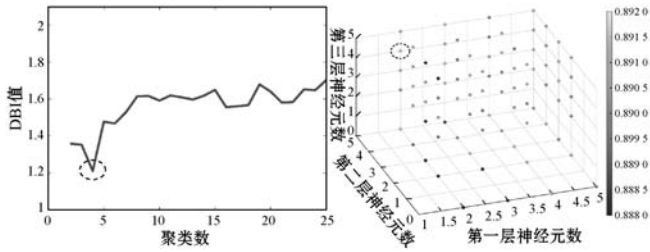


图 3 频谱聚类(左)和 ANN(右)的校准结果

5.2 频谱聚类算法性能

基于量测数据集,应用频谱聚类算法获得用户用电模式库。图 4 - 图 5 显示了平日和周末不同类型用户的典型负荷模式。可以看出,工作日的典型负荷曲线数量要少于周末,这与周末的各种活动相比复杂多样有关。而同样的,由于居民负荷行为的变化较大,因此居民用户的负荷模式比工业和商业用户更多。

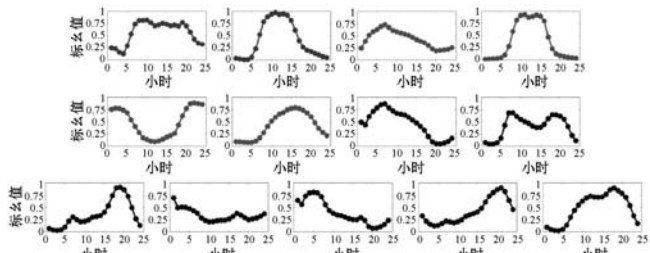


图 4 工业、商业和居民用户工作日用电模式库

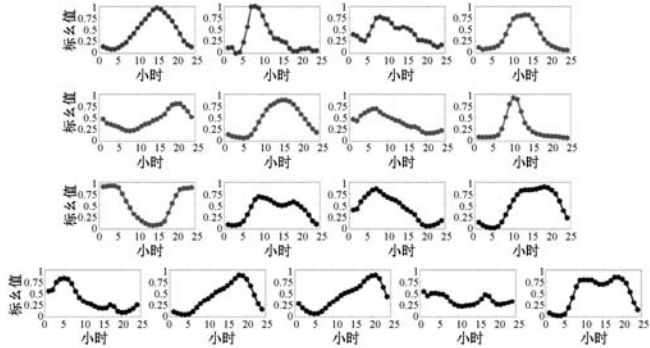


图 5 工业、商业和居民用户周末用电模式库

5.3 伪量测数据模拟

通过将用电模式库应用于多时间尺度学习模型的训练之后,利用测试集数据进行验证。图 6 显示了从月度电费账单获得的一个用户的小时级用电量曲线估计值与实际值之间的比较。可以看出,负荷估计曲线能够准确跟踪用户的实际用电量。

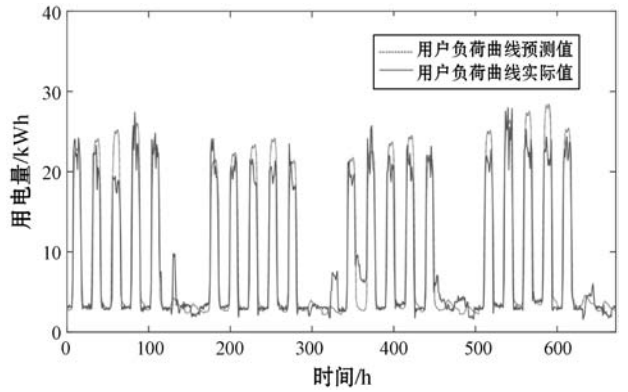
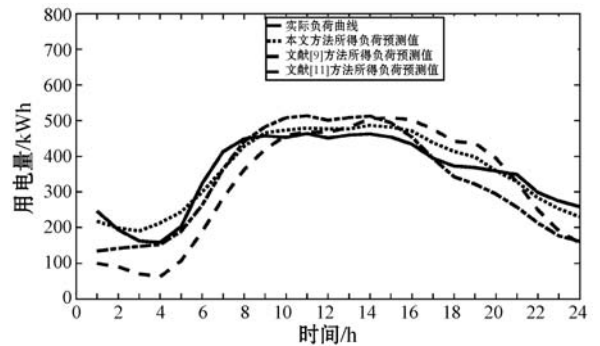


图 6 每小时负荷推断与实际负荷曲线的比较

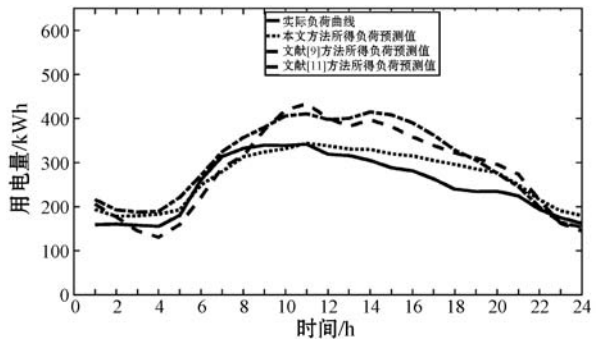
图 7 中给出论文所提出方法与现有方法在用户日负荷曲线估计上的结果对比。平均绝对百分比误差 (MAPE) 准则用于评估估计方法的准确性:

$$M = \frac{100\%}{n_s} \sum_{t=1}^{n_s} \left| \frac{A(t) - E\{A(t)\}}{A(t)} \right| \quad (13)$$

式中: A 是实际负荷值; $E\{A(t)\}$ 是估计值。



(a) 馈线工作日平均负荷预测值



(b) 馈线周末平均负荷预测值

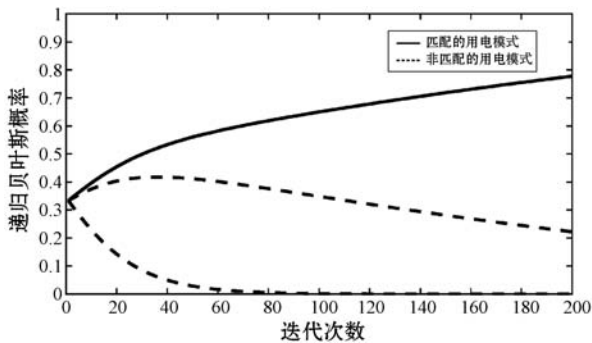
图 7 用户用电模式推理结果对比

由图 7 可知,对于工作日和周末,所提出的方法估计结果的 MAPE 值分别为 7.40%、10.02%,而文献 [9] 中方法的平均 MAPE 值分别为 19.47%、20.32% 和 13.79%、21.16%。因此,与以前的工作相比,所提出的方法在小时负荷推断方面显示出更高的准确性。

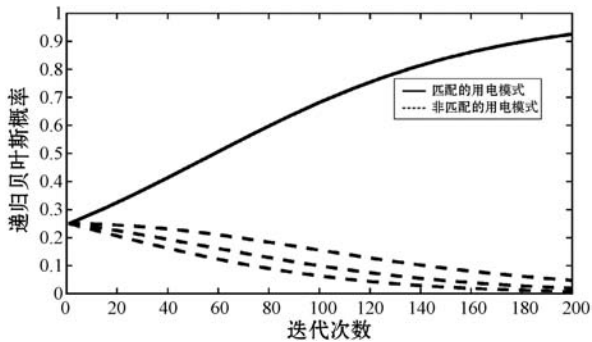
5.4 用电模式识别性能

针对工业负荷用户、商业负荷用户和居民负荷用户,分别应用论文所提出的无监测用户用电模式识别方法,采用其馈线出口出的量测数据,构建日用电量数

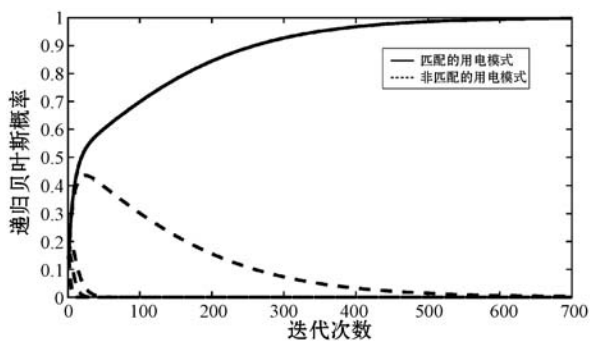
数据库,从无监测用户的月度电费账单中提取了小时级负荷估计值。图 8 中显示了递归贝叶斯学习算法所确定的典型日负荷曲线库中正确和不正确负荷模式的概率。以工业负荷用户的结果为例,由图 8(a)中可以看出,随着迭代过程的进行,与该无监测用户用电模式相匹配的那个多时间尺度学习模型的渐近概率会逐渐趋近于 1,而其他模型的概率则会逐渐降为零,因此这个递归贝叶斯概率趋近于 1 的学习模型所代表的用电模式,即为与该无监测用户相匹配的目标用电模式。由图 8(b)和图 8(c)可以看出,同样的情况在商业负荷用户和居民负荷用户中也是相同的。因此,这表明所提出的基于数据驱动的无监测用户用电模式识别方法的有效性,其能够依据无监测用户的月度电费账单,借助于基于频谱聚类的多时间尺度学习模型,准确地从已有的用电模式库中识别出与该用户用电模式相匹配的典型日负荷曲线。



(a) 工业负荷用户辨识



(b) 商业负荷用户辨识



(c) 居民负荷用户辨识

图 8 支路电流状态估计辅助的递归贝叶斯学习每日概况识别方法对三种类型用户的性能

6 结 语

本文提出了一种数据驱动的负荷用电模式识别方法,以提高智能电表未完全覆盖配电网的可观察性。所提出的方法能够从所有类型的用户(居民、商业和工业用户)的月度电费账单中提取其小时级的用电量估计值,并通过基于支路电流状态估计的多时间尺度概率学习方法来识别无监测用户的平均日负荷模式。

参 考 文 献

- [1] 曹颖爽. 基于大数据方法的用户用电模式分析[D]. 上海:上海交通大学,2016.
- [2] 彭显刚,赖家文,陈奕. 基于聚类分析的客户用电模式智能识别方法[J]. 电力系统保护与控制,2014,42(19): 68-73.
- [3] Al-Wakeel A, Wu J Z, Jenkins N. K-means based load estimation of domestic smart meter measurements[J]. Applied Energy,2017,194:333-342.
- [4] Gerbec D, Gasperic S, Smon I, et al. Allocation of the load profiles to consumers using probabilistic neural networks[J]. IEEE Transactions on Power Systems,2005,20(2): 548-555.
- [5] Gahrooei Y R, Khodabakhshian A, Hooshmand R A. A new pseudo load profile determination approach in low voltage distribution networks[J]. IEEE Transactions on Power Systems,2018,33(1):463-472.
- [6] Jardini J A, Tahan C M V, Gouvea M R, et al. Daily load profiles for residential commercial and industrial low voltage consumers[J]. IEEE Transactions on Power Delivery,2000,15(1):375-380.
- [7] McLoughlin F, Duffy A, Conlon M. A clustering approach to domestic electricity load profile characterization using smart metering data[J]. Applied Energy,2015,141:190-199.
- [8] Lv C, Xing Y, Zhang J Z, et al. Levenberg-Marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical system[J]. IEEE Transactions on Industrial Informatics,2018,14(8):3436-3446.
- [9] Wang S J, Zhao Y X. Online Bayesian tree-structured transformation of HMMs with optimal model selection for speaker adaptation[J]. IEEE Transactions on Speech and Audio Processing,2001,9(6):663-677.