



DOI:10.12404/j.issn.1671-1815.2403932

引用格式:朱磊,周璇,陈城,等.基于TICC的建筑用电时间序列自适应季节性分割方法[J].科学技术与工程,2025,25(11):4689-4697.

Zhu Lei, Zhou Xuan, Chen Cheng, et al. Adaptive seasonal segmentation method of building electricity consumption time series based on TICC[J]. Science Technology and Engineering, 2025, 25(11): 4689-4697.

建筑科学

基于TICC的建筑用电时间序列自适应季节性分割方法

朱磊^{1,2}, 周璇^{1,2,3*}, 陈城^{1,2,3}, 何敏^{2,3}, 闫军威^{1,2,3}

(1. 华南理工大学机械与汽车工程学院, 广州 510640; 2. 广州现代产业技术研究院, 广州 511458;

3. 人工智能与数字经济广东省实验室(广州), 广州 511442)

摘要 建筑用电时间序列(building electricity consumption time series, BECTS)的季节性分割对于准确的电力负荷预测与模式挖掘意义重大。针对传统定时分割、定温分割和自适应候温分割方法难以实现准确的BECTS季节性分割问题,提出了一种基于Toeplitz逆协方差聚类(Toeplitz inverse covariance-based clustering, TICC)的BECTS自适应季节性分割方法。该方法基于建筑逐时用电负荷与室外干球温度二元时间序列,利用TICC算法进行实时分割与聚类。夏热冬暖地区某大型公共建筑真实用电数据的分析结果表明,该方法增强了同类样本之间的相似性和异类样本之间的差异性,与定时分割、定温分割和自适应候温分割方法相比,TICC分割后各季节的平均动态时间规整(dynamic time warping, DTW)距离分别提高46.54%、35.73%和7.59%。该方法可作为数据预处理,为单体建筑数据挖掘分析如建筑用电模式挖掘和负荷预测提供数据支撑。

关键词 时间序列; 自适应季节性分割; Toeplitz逆协方差聚类; 动态时间规整

中图分类号 TU17;

文献标志码 A

Adaptive Seasonal Segmentation Method of Building Electricity Consumption Time Series Based on TICC

ZHU Lei^{1,2}, ZHOU Xuan^{1,2,3*}, CHEN Cheng^{1,2,3}, HE Min^{2,3}, YAN Jun-wei^{1,2,3}

(1. School of Mechanical & Automotive Engineering, South China University of Technology, Guangzhou 510640, China;

2. Guangzhou Institute of Modern Industrial Technology, Guangzhou 511458, China;

3. Artificial Intelligence and Digital Economy Guangdong Province Laboratory(Guangzhou), Guangzhou 511442, China)

[Abstract] Seasonal segmentation of building electricity consumption time series (BECTS) is of great significance for accurate load forecasting and pattern mining. Aiming at the problem that accurate BECTS seasonal segmentation is difficult to be realized by traditional timing segmentation, fixed-temperature segmentation and adaptive five-days temperature segmentation methods, a new adaptive seasonal segmentation method for BECTS based on Toeplitz inversed covariance-based clustering (TICC) was proposed. The method was based on the binary time series of building hourly electricity load and outdoor dry bulb temperature, and the TICC algorithm was used for real-time segmentation and clustering. A large public building electricity load case in a hot summer and warm winter area was analyzed, and the result showed that the similarity between samples of the same type and the difference between samples of different types were enhanced by the method. Compared with the timing segmentation, fixed-temperature segmentation and adaptive five-days temperature segmentation methods, the average dynamic time warping (DTW) distance of each category after TICC segmentation was improved respectively by 46.54%, 35.73% and 7.59%. This method can be used as data preprocessing to provide data support for single building data mining analysis, such as building electricity consumption pattern mining and load forecasting.

[Keywords] time series; adaptive seasonal segmentation; toeplitz inversed covariance-based clustering; dynamic time warping

收稿日期: 2024-05-27 修订日期: 2024-09-25

基金项目: 广东省自然科学基金(2022A1515011128)

第一作者: 朱磊(1999—),男,汉族,四川眉山人,硕士研究生。研究方向:建筑节能,数据挖掘。E-mail:zhulei090390@163.com。

*通信作者: 周璇(1976—),女,汉族,江西上饶人,博士,教授级高级工程师。研究方向:建筑节能,数据挖掘。E-mail:zhouxuan@scut.edu.cn。

投稿网址:www.stae.com.cn

海量时序数据的产生不断推动着时间序列数据挖掘 (time series data mining, TSDM) 技术的发展。时间序列分割是时间序列数据挖掘的重要任务之一^[1], 作为 TSDM 的预处理步骤, 将时间序列分割为更简单、更短的子序列, 为提高时间序列趋势预测、模式挖掘以及异常检测的准确性奠定基础。目前, 时间序列分割已经被广泛应用于多个领域。在水文学、气候学以及环境计量学等相关学科领域, 广泛用于水文^[2]、海温^[3]等时间序列数据变异点检测或风况类型^[4]划分中; 在金融领域, 用于金融时间序列的预处理^[5]、趋势和波动分析^[6]; 在工业^[7-8]、交通^[9]等其他领域, 广泛用于模式挖掘和负荷序列的合理分段。

“十三五”期间, 中国建筑节能与绿色建筑发展取得了重大进展, 中国 37 个省市已建成建筑能耗监管平台, 为“十四五”期间进一步提高建筑能效提供了海量用电时间序列数据, 其中蕴含着大量的建筑用电的模式和特征信息。建筑用电时间序列 (building electricity consumption time series, BECTS) 与季节因素密切相关, 具有明显的周期性和季节性特征, 准确的季节特征提取与分割能够帮助管理者深入理解 BECTS 的季节性模式, 提高用电负荷的预测精度, 为建筑能源管理、电力需求侧管理提供理论依据。

时间序列分割方法可以分为基于规则的方法、基于模型的方法和基于重要点检测的方法。基于规则的方法利用专家经验预先定义规则对时间序列进行分割, 如滑动窗口分割^[10]和人工经验分割^[11], 这类方法实施简单, 但规则的制定存在一定的主观性, 导致方法的泛化能力较弱, 难以处理复杂的时间序列结构。基于模型的方法将误差作为时间序列分割依据, 李军等^[12]利用分段线性表示 (piecewise linear representation, PLR) 来拟合风电机组实发功率时间序列, 使得原始序列和拟合序列的残差平方和最小。江艺美等^[13]针对线性拟合可能在现实生活中带来较大的拟合误差的问题, 提出基于重要点和灰色 GM(1,1) 模型的分段方法, 以曲线形式进行拟合, 减小了拟合误差。徐琛等^[11]利用自回归移动平均 (autoregressive moving average, ARMA) 模型对交通流时间序列进行预测, 根据预测值和真实值的差值确定时间序列的分段点。这类算法复杂度普遍较高, 且需要满足一定的假设前提^[14]。基于重要点检测的方法通过检测时间序列中的对序列走势有重要影响的点来进行分割, 包括极值点^[14]、趋势突变点^[15]和边缘点^[16]等, 这类方法可以保留整个序列中重要的趋势情况, 但需要准确对重要点进行定义^[17]。

现有 BECTS 的季节性分割方法主要是基于规

则的方法, 包括定时分割和定温分割。定时分割以月份为分割依据, 苏适等^[18]按月份将全年分为四个季节: 春季 (3—5 月), 夏季 (6—8 月), 秋季 (9—11 月), 冬季 (12 月—次年 2 月)。这种分割方法是一种静态定长分割方法, 仅考虑时间特征, 对于不同气候区可能会带来较大的分割误差。定温分割以候温 (即连续五天日平均气温的平均值) 为分割依据, 如《气候季节划分》(GB/T 42074—2022)^[19]设定 10 °C 和 22 °C 的候温为季节划分的指标阈值。这种分割方法依托大量历史气象数据, 是一种动态分割方法, 但仅以固定温度值进行分割, 不能保证不同地区季节划分的准确性。彭显刚等^[20]以地区负荷聚类结果为分类目标, 以候温为分类特征量, 构建了季节性负荷分类决策树, 实现了地区季节性负荷自适应划分。这种分割方法在地区整体季节性负荷分割上效果可能较好, 但由于没有考虑由建筑本身用电特征引起的负荷变化, 对于单体建筑而言, 季节性分割效果可能不佳。

时间序列分割也可以看作具有时序结构的特殊聚类问题, 这种问题被称为时序聚类问题^[21]。时序聚类通过将时间序列分割为不重叠的子序列片段并对子序列进行聚类来实现时间序列的分割, 是解决时间序列分割问题的一种强有力的工具。TICC (Toeplitz inverse covariance-based clustering) 是一种经典的时序分割聚类算法, 已被广泛用于海上 NWP (numerical weather prediction) 数据^[4]、变压器油线谱数据^[22]和风电机组 SCADA (supervisory control and data acquisition) 数据^[23]等多元时序数据的分割。

BECTS 逐时特征提取是其季节性分割准确性的前提。室外气象条件是影响建筑电力负荷的重要因素之一^[24]。为了更好地实现准确的 BECTS 自适应季节性分割, 拟在建筑逐时用电负荷与室外全球温度二元时间序列的基础上, 尝试采用 TICC 算法提取 BECTS 季节性特征, 探讨 BECTS 自适应季节性分割方法, 为建筑能源管理和优化提供依据。

1 TICC 算法原理

1.1 问题描述

TICC 算法是 2017 年 Hallac 等^[25]提出的一种经典的时序分割聚类算法, 它通过将多元时间序列分割为一定长度的子序列, 并依据子序列的结构相似性来实现子序列的聚类^[25]。设已有长度为 T 的多元时间序列

$$\mathbf{X}_{\text{orig}} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i, \dots, \mathbf{X}_T] \quad (1)$$

式 (1) 中: \mathbf{X}_i 为 n 维向量, 包含 n 个特征; \mathbf{X}_{orig} 为 $n \times T$ 维矩阵。

TICC 算法通过在每个向量前面截取若干个向量堆叠后形成若干个长度为 w 的子序列,以这些子序列为聚类对象进行聚类。堆叠后的多元时间序列为

$$\mathbf{X}_{\text{new}} = [\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_i, \dots, \mathbf{X}'_T] \quad (2)$$

$$\mathbf{X}'_i = [\mathbf{X}_{i-w-1}, \mathbf{X}_{i-w}, \dots, \mathbf{X}_i] \quad (3)$$

式中: \mathbf{X}'_i 为 $\mathbf{X}_{i-w-1}, \dots, \mathbf{X}_i$ 叠而成的长度为 w 的子序列,是 $n \times w$ 维向量; \mathbf{X}_{new} 为堆叠后的多元时间序列,是 $nw \times T$ 维矩阵。

不同于传统聚类算法, TICC 通过为每个簇定义一个依赖网络或马尔可夫随机场 (Markov random field, MRF) 来描述不同聚类簇中不同时间戳数据之间的相互关联性^[23], TICC 中用 Toeplitz 结构的逆协方差矩阵 Θ_i 来表示每个聚类簇的 MRF, 该矩阵的形式为

$$\Theta_i = \begin{bmatrix} \mathbf{A}^{(0)} & (\mathbf{A}^{(1)})^T & \dots & (\mathbf{A}^{(w-1)})^T \\ \mathbf{A}^{(1)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & (\mathbf{A}^{(1)})^T \\ \mathbf{A}^{(w-1)} & \dots & \mathbf{A}^{(1)} & \mathbf{A}^{(0)} \end{bmatrix} \quad (4)$$

式(4)中: $\Theta_i \in \mathbf{R}^{nw \times nw}$, 为分块 Toeplitz 矩阵; $\mathbf{A}^{(0)}, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(w-1)} \in \mathbf{R}^{n \times n}$ 为不同时间 n 个特征之间的逆协方差矩阵, 代表时间内的偏相关性; $\mathbf{A}_{i,j}^{(0)}$ 为时刻 t 特征 i 与时刻 t 特征 j 之间的相关关系; $\mathbf{A}_{i,j}^{(1)}$ 表示时刻 t 特征 i 与时刻 $t+1$ 特征 j 之间的相关关系。

TICC 的目标就是要求解这些逆协方差矩阵 $\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_i, \dots, \Theta_K\}$, 以及由此产生的聚类分配结果 $\mathbf{P} = \{P_1, P_2, \dots, P_i, \dots, P_K\}$ 。整体优化问题为

$$\operatorname{argmin}_{\Theta \in T} \sum_{i=1}^K \{ \|\lambda^\circ \Theta_i\|_1 + \sum_{X_i \in P_i} [-\ell\ell(X_i, \Theta_i) + \beta l\{X_{i-1} \notin P_i\}] \} \quad (5)$$

式(5)中: $\lambda \in \mathbf{R}^{nw \times nw}$ 是正则化参数; $\|\lambda^\circ \Theta_i\|_1$ 为 Hadamard 积的 1 范数惩罚; β 为强制时间一致性参数, 即当相连子序列不属于同一簇时施加的惩罚, 当 $\beta = 0$ 时, 各子序列独立聚类, 当 β 逐渐增大时, 相邻子序列越来越趋向于聚到一类, 当 $\beta = \infty$ 时, 所有子序列聚为一类; $l\{X_{i-1} \notin P_i\}$ 是一个检查相邻点是否分配到同一聚类的指标函数。 $\ell\ell(X_i, \Theta_i)$ 为 X_i 来自聚类簇 i 的对数似然, 其展开式为

$$\ell\ell(X_i, \Theta_i) = -\frac{1}{2} (X_i - \mu_i)^T \Theta_i (X_i - \mu_i) + \frac{1}{2} \lg \det(\Theta_i) - \frac{n}{2} \lg(2\pi) \quad (6)$$

式(6)中: μ_i 为聚类簇 i 的经验平均值。

1.2 问题求解

式(5)所示问题属于混合组合连续优化问题,

包含 Θ 和 \mathbf{P} 两个变量, 变量之间的耦合导致问题高度非凸, 没有易于处理的方法来求解全局最优解^[25]。在 TICC 中将该问题分解为为聚类分配点和更新聚类参数两个子问题。

1.2.1 为聚类分配点

该问题可以描述为: 给定 Θ_i , 通过最小化其负对数似然和最大化时间一致性求解 P_i 。将子序列 X_i 划分到聚类簇 i 的代价为

$$\min \sum_{i=1}^K \sum_{X_i \in P_i} [-\ell\ell(X_i, \Theta_i) + \beta l\{X_{i-1} \notin P_i\}] \quad (7)$$

可以采用动态规划法来求取这个问题的全局最优解^[25], 得到聚类分配结果 \mathbf{P} 。

1.2.2 更新聚类参数

该问题可以描述为: 给定 P_i , 通过最小化其负对数似然总和求解 Θ_i 。该问题可以转化为以下优化问题, 即

$$\min -\lg \det(\Theta_i) + \operatorname{tr}(S_i \Theta_i) + \frac{1}{|P_i|} \|\lambda^\circ \Theta_i\|_1 \quad \text{subject to } \Theta_i \in \Gamma \quad (8)$$

式(8)中: $\operatorname{tr}(\cdot)$ 为矩阵的迹; S_i 为 P_i 中所有子序列的经验协方差; Γ 为分块 Toeplitz 矩阵。可以采用交替方向乘法 (alternating direction method of multipliers, ADMM)^[25] 来求解该问题。通过迭代求解这两个子问题, 直到达到预设最大迭代数或误差, 输出聚类分配结果 \mathbf{P} 和对应的 Toeplitz 矩阵 Θ 。

1.3 模型参数设置

TICC 算法有 4 个超参数需要设置, 包括正则化参数 λ , 惩罚因子 β , 子序列长度 w 和聚类数 K 。其中 λ 和 β 通常设定为一个常数, 以减小搜索空间^[25]。

聚类数 K 是 TICC 的重要参数, 一般需要综合考虑聚类结果的可解释性, 结合实际场景来确定, 也可以通过贝叶斯信息准则 (Bayesian information criterion, BIC) 或交叉验证确定。

子序列长度 w 决定算法学习特征间跨时间相关性的尺度^[23], w 越大, 所跨的时间尺度越大, 但是, 若 w 过大, 则会导致在边界点上分类错误。因此, 一般设置 w 较小。它的确切值可以根据实际需求来选择, 也可以通过 BIC 或交叉验证来选择。

2 基于 TICC 的 BECTS 自适应季节性分割方法

本文中提出的基于 TICC 的 BECTS 自适应季节性分割方法主要包括数据预处理和模型建立两个步骤, 具体实施流程如图 1 所示。

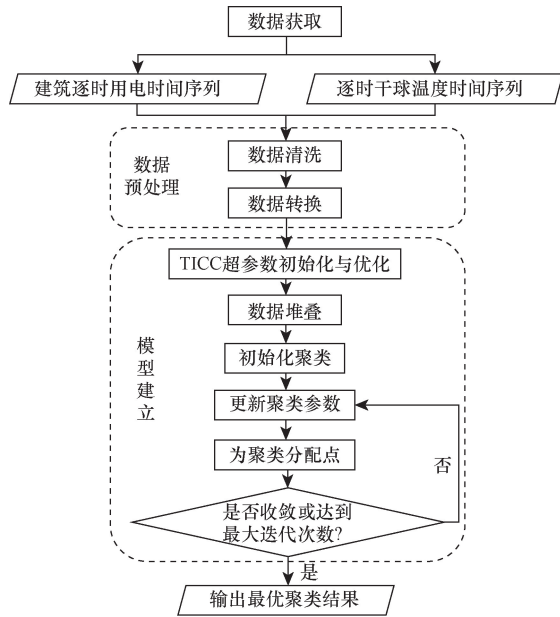


图1 基于TICC的BECTS自适应季节性分割方法流程图

Fig. 1 Flowchart of TICC-based self-adaptive seasonal segmentation method for BECTS

2.1 数据预处理

数据预处理主要包括数据清洗和数据转换两个过程。

2.1.1 数据清洗

由于数据在采集、通信、传输、存储等各个环节容易受到噪声干扰、传感器失效、通信中断等多种因素影响,产生大量缺失、突变或非正常零值等各种异常^[26],为保证时序数据的合理性、规范性和连续性,本文中对BECTS中常见的四种数据异常分别进行识别和插补。

(1) 针对单个缺失值:采用其前后最邻近的两个数据点的均值作为其插补值。

(2) 针对连续缺失值:采用其前后五个相同日期属性和时刻值的数据点的均值作为其插补值。

(3) 针对单个突变值:分时刻利用拉依达准则检测突变值,然后将其视作单个缺失值进行替换。

(4) 针对累积突变值:一般表现为连续出现多个异常零值并以累积突变值作为结尾的形式,通过单个突变值检测条件和连续异常零值检测进行识别,其插补方法如下。

设 D_1, D_2, \dots, D_k 为连续 k 个时刻的异常零值, D_{k+1} 为这 k 个时刻的异常零值所对应的累积突变值, $\varepsilon_i (i = 1, 2, \dots, k+1)$ 为这 $k+1$ 个异常值的初始替换值,采用连续缺失值的插补方法得到,实际插补值根据下式进行修正。

$$w_i = \frac{\varepsilon_i}{\sum_{t=1}^{k+1} \varepsilon_t} \quad (9)$$

$$D'_i = w_i D_{k+1} \quad (10)$$

式中: w_i 为各个初始替换值的权重; D'_i 为这 t 个时刻负荷的修正替换值。

2.1.2 数据转换

数据转换包括数据平滑和数据归一化两个部分。

数据平滑是一种统计技术或数据处理方法,通过对原始数据进行处理,可以改善数据质量,使其更加平缓、稳定和具有较低的噪声或变动性,更能凸显出长期变化趋势,便于观察和总结^[27]。本文中参考气象学“候温”的概念,采用滑动平均值滤波法对建筑逐时用电负荷和逐时室外干球温度时间序列进行平滑处理,设置滑动窗口大小为24,即将当前时刻及前面连续23h的数据的滑动平均值作为当前时刻新的数据特征,具体方式为

$$L'_t = \frac{\sum_{i=0}^{23} L_{t-i}}{24} \quad (11)$$

$$T'_t = \frac{\sum_{i=0}^{23} T_{t-i}}{24} \quad (12)$$

式中: L'_t 为 t 时刻滑动平均负荷值; L_t 为 t 时刻实际用电负荷值; T'_t 为 t 时刻滑动平均干球温度值; T_t 为 t 时刻实际干球温度值。

数据归一化是消除数据量纲影响的一种常用方法,本文中使用的归一化方法为

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (13)$$

式(13)中: X_{norm} 为归一化后的数据; X 为原始数据; X_{\max} 为原始数据中的最大值; X_{\min} 为原始数据中的最小值。

2.2 模型建立

利用预处理后的数据建立BECTS自适应季节性分割模型,主要步骤如下:

(1) TICC超参数初始化与优化:根据本文1.3节分析,为减小搜索空间,本文参考文献^[22,25],设定正则化参数 $\lambda = 0.11$,惩罚因子 $\beta = 500$;此外,本文中采用网格搜索法优化TICC聚类数 K 和子序列长度 w ,结合实际季节性划分的需求,初始化 $K = 3$,设定其变化区间为 $[3, 6]$;初始化 $w = 1$,设定其变化区间为 $[1, 4]$,以DBI(Davies-Bouldin index)和CHI(Calinski-Harabasz index)为评价指标,通过五折交叉验证对超参数寻优结果进行验证,最终将优化后的TICC最优超参数组合用于后续建模,DBI和CHI计算方式为

$$\text{DBI} = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \frac{d(C_i) + d(C_j)}{d(\mu_i, \mu_j)} \quad (14)$$

$$\text{CHI} = \frac{B/(K-1)}{W/(N-K)} \quad (15)$$

$$W = \sum_{i=1}^K \sum_{x \in C_i} (x - \mu_i)^2 \quad (16)$$

$$B = \sum_{i=1}^K n_i (\mu_i - \mu)^2 \quad (17)$$

式中: K 为聚类簇数; C_i 为第 i 个聚类簇; $d(C_i)$ 为聚类簇 C_i 中各个点距离簇质心的平均距离; μ_i 为聚类簇 C_i 的质心; $d(\mu_i, \mu_j)$ 为聚类簇 C_i 质心和 C_j 质心之间的距离; W 为类内紧密度; B 为类间分离度; n_i 为聚类簇 C_i 中样本点的个数; μ 为全局簇中心; N 为样本点的总数量; DBI 计算所有类的最大相似度的均值, 其值越小意味着类内距离越小且类间距离越大, 即聚类效果越好。CHI 综合考虑类间分离度和类内紧密度, 其值越大, 表明聚类效果越好。

(2)数据堆叠:根据步骤(1)所确定的 w , 对原始数据进行堆叠, 将堆叠后得到的长度为 w 的子序列作为聚类对象。

(3)初始化聚类:利用 K -means 算法将堆叠后的数据初始化聚为 K 类, 得到初始化聚类结果 P 以及原始均值、堆叠均值等聚类参数, 各聚类参数计算方式见文献[23]。

(4)更新聚类参数:将原始均值、堆叠均值、堆叠数据集等参数作为输入, 利用 ADMM (alternating direction method of multipliers) 算法求解式(8)所示问题, 得到 Toeplitz 矩阵 Θ 。

(5)为聚类分配点:将 Toeplitz 矩阵 Θ 等参数作为输入, 利用动态规划算法求解式(7)所示问题, 从而更新聚类结果 P 。

(6)迭代终止条件:重复步骤(4)和(5)直到聚类结果收敛即聚类结果 P 不再改变或达到最大迭代次数时终止迭代, 输出最优聚类结果 P 。

3 案例分析

本文研究对象为夏热冬暖地区某大型公共建筑, 建筑总面积约为 198 560.7 m², 该建筑于 2020 年 12 月建成能耗监测平台, 数据采集频率为 1 h, 选用 2021 年 6 月 1 日 00:00—2022 年 12 月 31 日 23:00 的用电量数据、室外干球温度数据进行分析。

3.1 数据预处理

文中选定的数据样本共 13 896 个, 通过分析检测出异常数据 281 个, 占总样本的 2.02%, 如表 1 所示。采用 2.1.1 节的数据清洗方法处理异常数据。

采用 2.1.2 节的数据转换方法对数据进行处理, 最终将数据格式转化为 $2 \times 13 896$ 的矩阵, 作为 TICC 算法的输入进行后续的分割。

表 1 异常数据检测结果

Table 1 Abnormal data detection results

异常类型	异常数据数量/个	异常数据占比/%
单个缺失	5	0.036 0
连续缺失	76	0.547 0
单个突变	126	0.906 7
累积突变	74	0.532 5

3.2 超参数优化结果

利用式(14)~式(17)计算不同 K 和 w 组合下的 DBI 和 CHI 来选取最优超参数组合, 结果如表 2 所示, 当 $K=3$ 、 $w=1$ 时对应 DBI 最小, 对应的 CHI 最大, 为此, 采用该参数组合作为 TICC 算法的优化参数。

表 2 不同 K 和 w 组合下的 DBI 和 CHI 计算结果

Table 2 DBI and CHI calculations for different combinations of K and w

K	w	DBI	CHI	K	w	DBI	CHI
3	1	0.95	10 411.56	5	1	2.21	5 309.41
3	2	1.03	10 401.27	5	2	3.63	5 418.17
3	3	1.03	10 389.59	5	3	1.58	5 567.30
3	4	1.03	10 378.83	5	4	1.71	5 666.83
4	1	1.95	7 070.28	6	1	3.28	2 579.45
4	2	1.95	7 063.13	6	2	1.02	4 195.02
4	3	1.41	7 341.18	6	3	2.07	4 337.54
4	4	1.29	7 609.88	6	4	3.28	4 600.43

3.3 基于 TICC 的 BECTS 自适应季节性分割结果

3.3.1 TICC 方法与其他分割方法的定性比较

为验证基于 TICC 的 BECTS 自适应季节性分割方法的有效性, 将 TICC 方法与定时、定温和自适应候温三种分割方法对 BECTS 进行季节性分割结果进行比较。

其中定时分割方法参考文献[18], 采用北半球通用的气候季节划分方式, 将一年划分为春夏秋冬四个季节。

定温分割方法参考《气候季节划分》(GB/T 42074—2022)^[19], 利用日平均气温和候温划分四季, 分割结果如图 2 所示。

自适应候温分割方法参考文献[20], 设置聚类数 $K=3$, 候温 $H_s \in [3 \text{ } ^\circ\text{C}, 31 \text{ } ^\circ\text{C}]$, 变化步长为 0.5, 得到季节性负荷分类决策树如图 3 所示。

四种方法下 BECTS 季节性分割结果如图 4 所示。

如图 4(a)和图 4(b)所示, 面向本文研究对象, 定时分割和定温分割将 BECTS 分为了四类季节, 对应一年中春夏秋冬四个季节; 如图 4(c)和图 4(d)所示, 自适应候温分割和 TICC 将 BECTS 分为夏季、过渡季和冬季三个季节, 其中, 过渡季相当于将春季

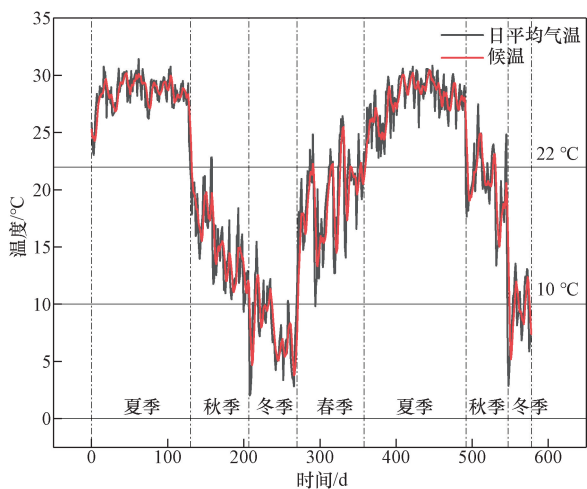


图2 定温分割结果示意图

Fig. 2 Schematic diagram of fixed-temperature segmentation result

和秋季合并。

从分割后各季节片段形态来看,定时分割方法在夏季和秋季片段分割点前后没有明显的形态特征变化,分割误差较大,相比于这种方法,其余三种方法分割得到的各季节片段更具有独立性。三种方法分割点信息统计如表3所示。

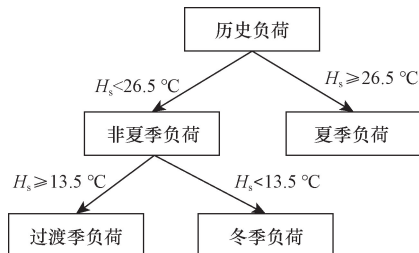
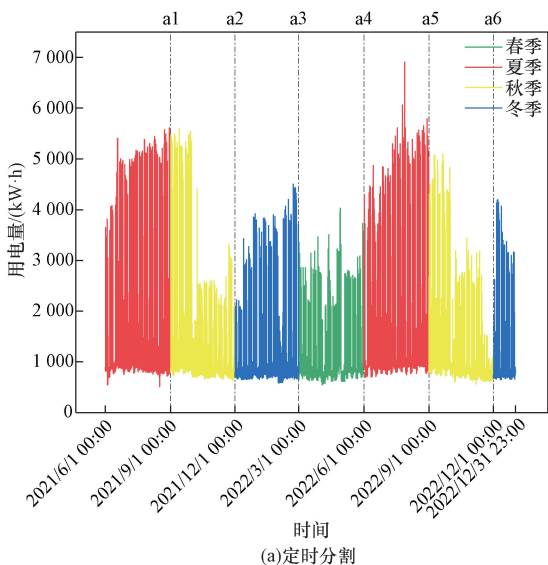
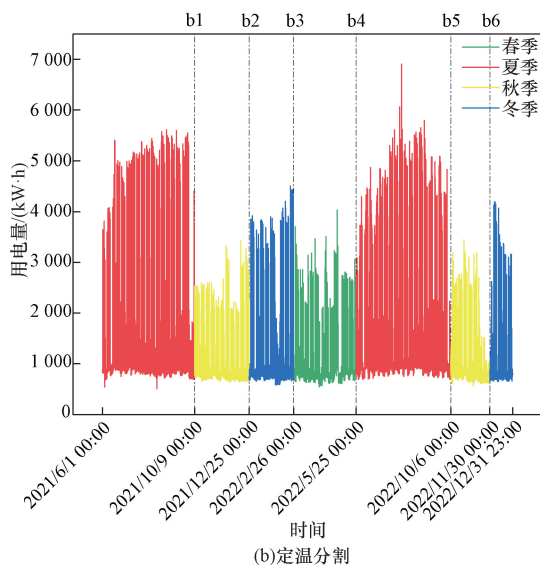


图3 季节性负荷分类决策树(自适应候温分割)

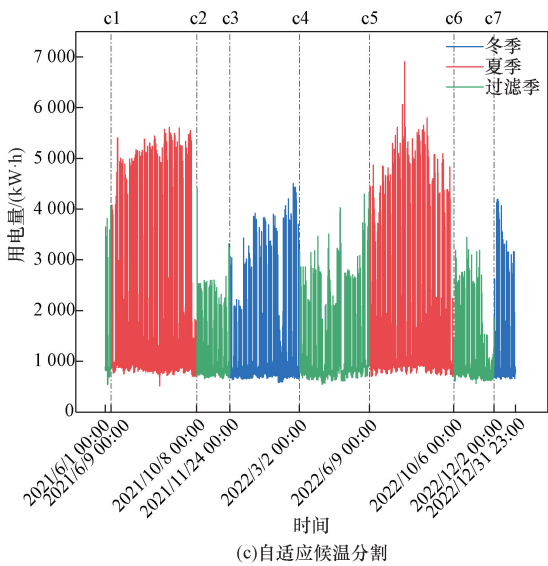
Fig. 3 Seasonal load classification decision tree (adaptive five-days temperature segmentation)



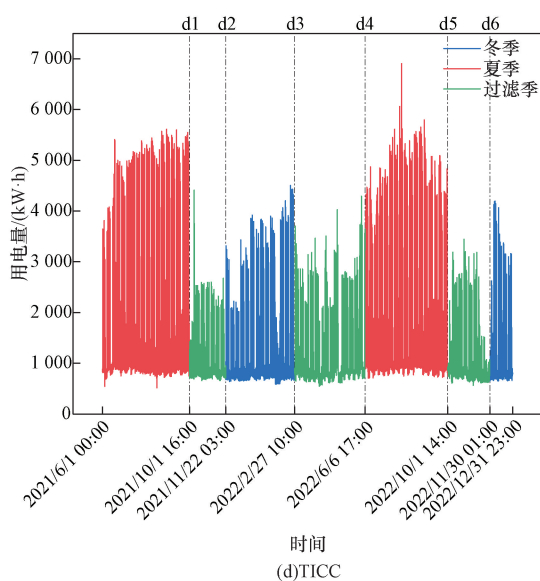
(a)定时分割



(b)定温分割



(c)自适应候温分割



(d)TICC

图4 BECTS 季节性分割结果

Fig. 4 BECTS seasonal segmentation results

表 3 三种方法分割点信息统计

Table 3 Segmentation point information statistics for three methods

时间	分割位置	分割点	最大时间差异/d
2021年6月1日00:00—2021年12月31日23:00	过渡季(春季)和夏季	c1	
	夏季和过渡季(秋季)	b1、c2、d1	>7
	过渡季(秋季)和冬季	b2、c3、d2	>7
2022年1月1日00:00—2022年12月31日23:00	冬季和过渡季(春季)	b3、c4、d3	<7
	过渡季(春季)和夏季	b4、c5、d4	>7
	夏季和过渡季(秋季)	b5、c6、d5	<7
	过渡季(秋季)和冬季	b6、c7、d6	<7

如表3所示,自适应候温分割方法相比于其他方法增加了分割点c1,从图4(c)中可以看出该分割点的用电量与其他过渡季用电量存在一定差异。

2021年夏季和过渡季(秋季)的分割上,分割点{b1,c2,d1}代表建筑从持续供冷到结束供冷平均用电水平断崖式下降的过程。分割点b1和c2之间的时间差异性较小,与d1的时间差异位于当年10月1日—10月9日这段时间之内,这段时间属于国庆假期,虽然室外温度较高,但由于节假日建筑用能活动减少,其用电水平位于较低水平,TICC综合考虑室外温度和建筑用电特征的影响,将其归入负荷水平相近的过渡季片段以提高分割效果。

2021年过渡季(秋季)和冬季以及2022年过渡季(春季)与夏季的分割上,分割点{b2,c3,d2}和{b4,c5,d4}均代表建筑从过渡季较低用电水平到开始供暖或供冷后用电水平显著提升的转变过程。在同一分割位置上,分割点c3和d2之间、c5和d4之间的时间差异性较小,分别与b2和b4的时间差异均达到7d以上,表明自适应候温分割和TICC方法能够从候温变化或建筑用电特征变化的角度识别出由于气候变化建筑开始供暖或供冷的时间,而仅以固定候温分割的方式难以准确描述这种转变。

2022年冬季和过渡季(春季)、夏季和过渡季(秋季)以及过渡季(秋季)和冬季的分割上,三种方法涉及的分割点b3、c4和d3之间、b5、c6和d5之间以及b6、c7和d6之间的时间差异均小于7d,分割结果相近。

总的来说,自适应候温分割能够准确识别候温变化,取得了优于定温分割的效果;TICC能同时捕捉建筑自身用电特征行为和室外气象条件变化,并将其作为季节性知识纳入BECTS季节性分割过程中,进一步提升了自适应季节性分割效果。

3.3.2 TICC方法评价

(1)评价指标。时间序列分割效果可以通过时间序列之间的相似度来进行量化评价,评价指标主要包括欧式距离和动态时间规整(DTW)距离^[28]。由于时间序列季节性分割通常并非定长分割,即分

割片段的长度并不相等,传统的欧式距离难以适用。DTW距离采用时间规整来对齐两个不同长度的时间序列,从而计算出两序列之间的距离,有效解决了不等长时间序列相似性度量的问题,DTW距离越大,表明分割效果越好。

(2)评价结果。将TICC与定时、定温和自适应候温三种分割方法计算得到的平均DTW距离进行比较,结果如表4所示。

表4 四种分割方法计算的平均DTW距离

Table 4 Average DTW distance calculated by four segmentation methods

	夏季	冬季	过渡季		综合
			春季	秋季	
定时分割	1 707.64	1 724.15	956.58	1 160.00	1 387.09
定温分割	2 380.94	1 688.79	903.48	1 018.22	1 497.86
自适应候温分割	2 309.38	2 065.31	1 293.32		1 889.34
TICC	2 461.55	2 198.47	1 438.04		2 032.69

如表4所示,TICC所得到的各季节平均DTW距离均大于其余三种方法。

夏季,TICC的平均DTW距离相比定时、定温和自适应候温分割分别提高44.15%、3.39%和6.59%,定温、自适应候温和TICC三种分割方法的分割效果显著优于定时分割,表明气象条件是影响夏季分割的关键因素。由于TICC综合考虑了气象条件和建筑用电特征的变化,进一步提高了夏季片段的独立性。

冬季,TICC的平均DTW距离比其他方法分别提高27.51%、30.18%和6.45%,分割效果与自适应候温分割相近,且显著优于其余两种方法,表明自适应确定负荷或候温的变化点的方式均能较准确识别冬季负荷季节性分割点。

过渡季,TICC的平均DTW距离相比其他方法分别提高35.88%、49.66%和11.19%,分割效果与自适应候温分割相近且显著优于其余两种方法,表明在春秋两季,由于气候条件的相似性,建筑用电特征并没有发生显著改变,合并这两个季节可以有效提高季节片段的独立性。

综上所述,TICC 分割后综合平均 DTW 距离相比定时、定温和自适应候温分割分别提高 46.54%、35.73% 和 7.59%,表明 TICC 能对 BECTS 进行准确的自适应季节性分割,且有效增强分割效果。

4 结论

(1)利用 TICC 算法将 BECTS 季节性分割问题转化为时序聚类问题,提高了 BECTS 季节性分割的准确性,并能根据不同对象进行自适应分割。

(2)本文方法在聚类过程中综合考虑室外气象条件和建筑特征用电行为的影响,摆脱了地域位置和物候景象的限制,可实现小时级 BECTS 的自适应季节性分割。

(3)利用夏热冬暖地区某大型公共建筑真实用电数据,将 TICC 分割方法与定时、定温和自适应候温分割方法进行比较,结果显示 TICC 分割方法的综合平均 DTW 距离分别提高 46.54%、35.73% 和 7.59%,取得了最优的分割效果。

(4)本文方法可作为数据预处理步骤,为单体建筑数据挖掘分析如建筑用电模式挖掘、异常用电模式检测和负荷预测提供有效的支持。

参 考 文 献

[1] Fu T. A review on time series data mining[J]. Engineering Applications of Artificial Intelligence, 2011, 24(1): 164-181.

[2] 王玲,朱慧. 基于 KPCA 和 G-G 聚类的多元时间序列模糊分段[J]. 控制与决策, 2021, 36(1): 115-124.

Wang Ling, Zhu Hui. Fuzzy segmentation of multivariate time series with KPCA and G-G clustering[J]. Control and Decision, 2021, 36(1): 115-124.

[3] 鲍家勇,赵月旭. 基于 Floyd 算法的海温时间序列分割[J]. 数理统计与管理, 2019, 38(2): 326-333.

Bao Jiayong, Zhao Yuxu. Sea surface temperature time series segmentation based on Floyd algorithm[J]. Journal of Applied Statistics and Management, 2019, 38(2): 326-333.

[4] 周海,刘建锋,周健,等. 基于改进降噪自编码器和多元时序聚类的海上风电功率预测[J]. 太阳能学报, 2023, 44(3): 129-138.

Zhou Hai, Liu Jianfeng, Zhou Jian, et al. Offshore wind power prediction based on improved denoising auto-encoder and multivariate time series clustering[J]. Acta Energetica Solaris Sinica, 2023, 44(3): 129-138.

[5] 桑夏夏,李旭伟. 一种金融时间序列区域分割方法的研究[J]. 四川大学学报(自然科学版), 2018, 55(6): 1189-1196.

Sang Xiaxia, Li Xuwei. Research on the region based segmentation method of financial time series[J]. Journal of Sichuan University (Natural Science Edition), 2018, 55(6): 1189-1196.

[6] Si Y W, Yin J. OBST-based segmentation approach to financial time series[J]. Engineering Applications of Artificial Intelligence, 2013, 26(10): 2581-2596.

[7] 解初,王建东,韩邦磊,等. 基于趋势特征聚类的多元相似时

间序列的提取[J]. 科学技术与工程, 2020, 20(7): 2786-2793.

Xie Chu, Wang Jiandong, Han Banglei, et al. Extraction of multivariate similar time series based on trend feature clustering[J]. Science Technology and Engineering, 2020, 20(7): 2786-2793.

[8] 周威豪,庞勇,宋学官. 基于非线性自回归的盾构机时间序列数据分割[J]. 机械设计, 2022, 39(S1): 13-16.

Zhou Weihao, Pang Yong, Song Xueguan. Time series segmentation of tunnel boring machine based on nonlinear autoregression[J]. Journal of Machine Design, 2022, 39(S1): 13-16.

[9] 王本超,李丹,秦攀,等. 基于 Gamma 分布的交通流时间序列分割模型[J]. 大连理工大学学报, 2020, 60(3): 293-299.

Wang Benchao, Li Dan, Qin Pan, et al. Gamma distribution based traffic flow time series segmenting model[J]. Journal of Dalian University of Technology, 2020, 60(3): 293-299.

[10] 李正欣,刘畅,吴诗辉,等. 特征点分段提取的时间序列模式匹配方法[J]. 北京航空航天大学学报, 2023, 49(7): 1593-1599.

Li Zhengxin, Liu Chang, Wu Shihui, et al. Segmentation extraction of feature points for time series pattern matching[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(7): 1593-1599.

[11] 徐琛,董德存,欧冬秀. 传感网中数据驱动的多时段控制方法优化研究[J]. 计算机工程与应用, 2020, 56(15): 235-241.

Xu Chen, Dong Decun, Ou Dongxiu. Time-of-day control optimization of data-driven urban road constant-peak-type intersections in sensor networks[J]. Computer Engineering and Applications, 2020, 56(15): 235-241.

[12] 李军,刘恩仁,王敏,等. 基于时间序列分段的风电机组调节速率估计[J]. 科学技术与工程, 2021, 21(27): 11609-11614.

Li Jun, Liu Enren, Wang Min, et al. Ramp rate estimation of wind power generation units based on time series segmentation[J]. Science Technology and Engineering, 2021, 21(27): 11609-11614.

[13] 江艺羨,张岐山. 基于重要点与灰色 GM(1,1)模型的时间序列分段算法[J]. 统计与决策, 2016(24): 28-30.

Jiang Yixian, Zhang Qishan. The time series segmentation based on important points and gray GM(1,1) Model[J]. Statistic & Decision, 2016(24): 28-30.

[14] 罗宇成,张明恩,刘飞,等. 基于分段特征提取的仿真模型结果验证方法[J]. 系统仿真学报, 2024, 36(1): 272-281.

Luo Yucheng, Zhang Mingen, Liu Fei, et al. Result validation method of simulation models based on piecewise feature extraction[J]. Journal of System Simulation, 2024, 36(1): 272-281.

[15] 王塞,王思诗,樊风雷. 基于时间序列分割算法的雅鲁藏布江流域 NDVI(1985—2018)变化模式研究[J]. 生态学报, 2020, 40(19): 6863-6871.

Wang Sai, Wang Sishi, Fan Fenglei. Change patterns of NDVI (1985—2018) in the Yarlung Zangbo River Basin of China based on time series segmentation algorithm[J]. Acta Ecologica Sinica, 2020, 40(19): 6863-6871.

[16] 殷炜宏,王若愚,段倩倩,等. 基于时态边缘算子的时间序列自主分段表示法[J]. 计算机工程与科学, 2021, 43(6): 1104-1111.

- Yin Weihong, Wang Ruoyu, Duan Qianqian, et al. An autonomous segmental representation of time series based on temporal edge operator[J]. *Computer Engineering & Science*, 2021, 43(6): 1104-1111.
- [17] 孙志伟,董亮亮,马永军. 一种基于重要点的时间序列分段算法[J]. *计算机工程与应用*, 2018, 54(18): 250-255.
- Sun Zhiwei, Dong Liangliang, Ma Yongjun. Algorithm for time series based on important points[J]. *Computer Engineering and Applications*, 2018, 54(18): 250-255.
- [18] 苏适,李康平,严玉廷,等. 基于密度空间聚类和引力搜索算法的居民负荷用电模式分类模型[J]. *电力自动化设备*, 2018, 38(1): 129-136.
- Su Shi, Li Kangping, Yan Yuting, et al. Classification model of residential power consumption mode based on DBSCAN and gravitational search algorithm[J]. *Electric Power Automation Equipment*, 2018, 38(1): 129-136.
- [19] 中国气象局. 气候季节划分: GB/T 42074—2022[S]. 北京: 中国标准出版社, 2022.
- China Meteorological Administration. Division of climatic seasons: GB/T 42074—2022 [S]. Beijing: Standards Press of China, 2022.
- [20] 彭显刚,潘可达,张丹,等. 基于季节性负荷自适应划分及重要点分割的多分段短期负荷预测[J]. *电网技术*, 2020, 44(2): 603-613.
- Peng Xiangang, Pan Keda, Zhang Dan, et al. Piecewise short-term load forecasting based on adaptive seasonal load category and important point segment[J]. *Power System Technology*, 2020, 44(2): 603-613.
- [21] Abonyi J, Feil B, Nemeth S, et al. Modified Gath-Geva clustering for fuzzy segmentation of multivariate time-series[J]. *Fuzzy Sets and Systems*, 2005, 149(1): 39-56.
- [22] 邓春宇,吴克河,谈元鹏,等. 基于多元时间序列分割聚类的异常值检测方法[J]. *计算机工程与设计*, 2020, 41(11): 3123-3128.
- Deng Chunyu, Wu Kehe, Tan Yuanpeng, et al. Outlier detection method based on multivariate time series segmentation clustering[J]. *Computer Engineering and Design*, 2020, 41(11): 3123-3128.
- [23] 肖削,邓杰文,刘晓明,等. 基于运行规律和 TICC 算法的风电 SCADA 高维时序数据聚类方法[J]. *机械工程学报*, 2022, 58(23): 196-207.
- Xiao Zhao, Deng Jiewen, Liu Xiaoming, et al. Clustering method of high-dimensional time series SCADA data from wind turbines based on operational laws and TICC algorithm[J]. *Journal of Mechanical Engineering*, 2022, 58(23): 196-207.
- [24] 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 Zengping, Zhao Bing, Ji Weijia, et al. Short-term load forecasting method based on GRU-NN model[J]. *Automation of Electric Power System*, 2019, 43(5): 53-58.
- [25] Hallac D, Vare S, Boyd S, et al. Toeplitz inverse covariance-based clustering of multivariate time series data[C]//*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Halifax NS Canada: ACM, 2017: 215-223.
- [26] 周璇,崔少伟,周裕东. 办公建筑逐时能耗异常数据在线插补方法[J]. *建筑科学*, 2018, 34(6): 82-90.
- Zhou Xuan, Cui Shaowei, Zhou Yudong. Online interpolation method for abnormal data of hourly energy consumption of office buildings[J]. *Building Science*, 2018, 34(6): 82-90.
- [27] 董是,龙志友,毕洁夫,等. 基于 Savitzky-Golay 平滑-小波降噪处理的桥梁结构监测数据分析方法[J]. *测绘通报*, 2023(9): 100-106.
- Dong Shi, Long Zhiyou, Bi Jiefu, et al. Analysis method for bridge structure monitoring data based on Savitzky-Golay smoothing-wavelet noise reduction processing[J]. *Bulletin of Surveying and Mapping*, 2023(9): 100-106.
- [28] 宋维凯,张洪海,万俊强,等. 基于动态时间规整和长短期记忆的空中交通流量短期预测[J]. *科学技术与工程*, 2022, 22(25): 11256-11261.
- Song Weikai, Zhang Honghai, Wan Junqiang, et al. Short-term prediction of air traffic flow based on dynamic time warping-long short-term memory[J]. *Science Technology and Engineering*, 2022, 22(25): 11256-11261.