

DOI: 10.13718/j.cnki.xdzk.2020.10.002

基于时间序列多尺度分解的 建筑用电负荷预测方法

鞠亚轩, 张春雨, 朱仁敬, 朱习军

青岛科技大学 信息科学技术学院, 山东 青岛 266100

摘要: 随着近年来建筑能耗的迅速增长, 建筑节能成为可持续发展战略的一个重要问题, 因此, 构建一个可以快速精准预测建筑能耗的模型成为实现建筑节能的关键一步. 本文结合集合经验模态分解, 将 LSSVR 与 ARIMA 相融合, 构建了一种基于时间序列多尺度分解的预测建筑用电负荷数据的 EEMD-LSSVR-ARIMA 混合模型. 该模型通过 EEMD 将建筑用电负荷数据分解为多个频率不同的分量, 使用 LSSVR 模型预测高频分量以及用电负荷数据与各分量之和的差值序列, 使用 ARIMA 模型预测低频分量, 最后将各分量的预测结果以及差值序列的预测结果叠加得到最终的预测结果. 并通过某建筑的用电数据进行实验分析, 通过与传统的 ARIMA 和 EEMD-ARIMA 模型以及基于残差的 ARIMA-LSTM 模型进行对比, 实验结果表明, 本文提出的模型预测精度达到了 98% 以上, 与其他模型相比预测精度提升了将近 2%.

关键词: 能耗预测; EEMD; LSSVR; ARIMA

中图分类号: TM71 **文献标志码:** A **文章编号:** 1673-9868(2020)10-0008-09

电力作为一种特殊能源, 无法实现大规模存储, 因此, 保持电力的供需平衡是至关重要的, 供电过量会造成资源的浪费, 供电不足则会导致经济损失, 甚至影响社会稳定. 我国建筑面积庞大, 若是能够建立准确预测建筑用电负荷的模型, 对于我国加快节能减排政策的落实将会有重大的意义. 而如何提高用电负荷预测的准确率一直以来都是国内外学者研究的热点问题.

李婉华^[1]以随机森林和极限学习机为基础, 通过引入粒子群优化算法以及变异算子, 改进了算法的泛化性能以及稳定性, 提升了电力数据的预测性能; Fazil Kaytez^[2]结合最小二乘支持向量机与自回归移动平均对土耳其的电力数据进行了预测, 并将结果与其他模型进行比较, 证明了模型的有效性; 王月强等^[3]提出了基于多时空的用电负荷预测模型, 通过小波分解方法分解负荷数据的各时空分量, 并对不同分量选择不同模型进行分析预测, 大大提升了模型预测的精确度; 魏华栋等^[4]通过小波变换细化负荷数据的伸缩变换, 并结合长短期记忆神经网络的梯度计算, 有效提升了负荷预测结果的精度; 罗澍忻等^[5]提出了一种堆叠长短期记忆网络的模型, 将不同时间尺度的历史负荷数据以及地区经济和气候数据导入模型, 通过实验验证该模型提升了负荷预测的准确性; 邢伟等^[6]通过集合干扰重构经验模态分解算法将用电负荷数据分解为多个平稳分量, 并提出了改进多尺度散布熵算法刻画原始用电数据的内在特征, 从而提升了模型预测的准确性.

尽管用电负荷数据预测的研究很广泛, 但仍存在以下几个问题:

1) 特征的选择不够深入. 现有的预测研究方法没有深入分析用电负荷数据的影响因素, 往往直接将影响因素作为输入进行预测, 很大程度上影响了模型的预测精度.

收稿日期: 2020-09-30
基金项目: 山东省重点研发计划基金(2015GSF119016).
作者简介: 鞠亚轩(1995—), 男, 硕士研究生, 主要从事数据挖掘方面的研究.

2) 数据量要求大. 目前的研究方法主要以神经网络为主, 需要庞大的数据量进行训练, 才能取得理想的预测精度, 而我国的能源监测系统近年来才逐渐完善, 针对单一建筑的用电负荷数据有限, 很难获取如此大的数据量进行训练.

3) 对波动性强的用电负荷数据预测精度较差. 现有的研究方法对于用电规律明显的数据预测精度较高, 但当用电负荷数据波动性较大时, 预测的精度显著下降.

针对以上问题, 本文提出了一种基于时间序列多尺度分解的混合预测模型(EEMD-LSSVR-ARIMA 模型). 经研究发现, 用电量与建筑面积、楼层高度以及各气象因素密切相关, 而本次实验针对的是单一建筑, 建筑面积与楼层高度固定, 因此本次实验主要分析气象数据对用电量的影响, 经综合考量, 本文选择空气温度、露点温度、海平面气压作为影响因素, 使用集合经验模态分解(EEMD)^[7]分别将用电负荷数据和 3 种影响因素数据按其内在的时间尺度特征自适应地分解为多个本征模态分量, 保证原始数据信息不丢失, 在深入分析用电负荷数据分量与各影响因素分量之间关系的基础上, 使用最小二乘支持向量回归(LSSVR)^[8-9]模型和差分自回归移动平均(ARIMA)^[10]模型分别对用电负荷数据的各分量进行预测, 从而可以最大程度地拟合各分量数据, 再将预测结果叠加, 得到最后的预测结果. 并通过实验分析验证模型的有效性.

1 模型设计过程

1.1 EEMD-LSSVR-ARIMA 模型

考虑到建筑用电负荷数据具有很强的波动性和非线性特征, 本文提出了一种基于时间序列多尺度分解的 EEMD-LSSVR-ARIMA 混合预测模型对用电负荷数据进行预测, 通过 EEMD 将数据分解为多个频率不同的分量, 使用线性模型 ARIMA 预测低频分量, 使用非线性模型 LSSVR 预测高频分量, 充分发挥不同模型的预测性能, 从而达到提升预测精度的目的. 模型流程图如图 1 所示.

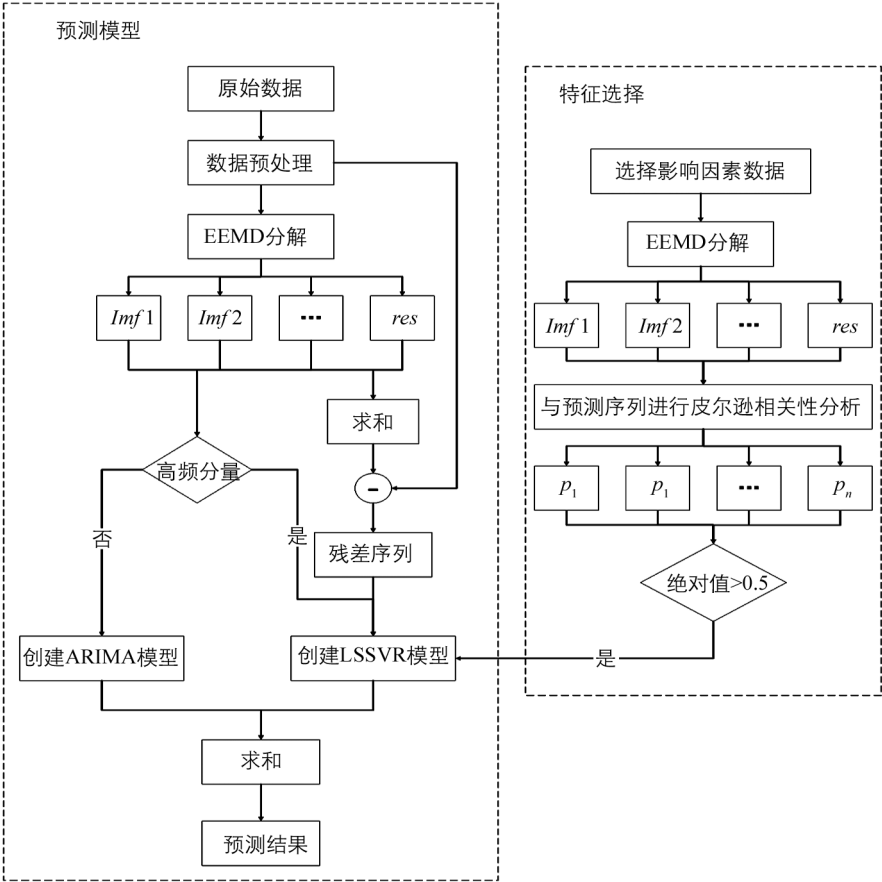


图 1 模型流程图

模型设计思路如下：

- 1) 使用 EEMD 将预处理后的用电负荷数据分解为多个本征模态分量；
- 2) 计算原数据序列与经 EEMD 分解后的各分量之和的差值，得到差值序列；
- 3) 选择影响用电负荷数据因素，使用 EEMD 分解影响因素序列，分解为多个频率不同的分量；
- 4) 使用 LSSVR 预测用电负荷数据的高频分量和差值序列，通过皮尔逊关联分析法分析影响因素数据的各分量序列与要预测的数据序列的相关性，选择相关系数绝对值大于 0.5 的分量作为 LSSVR 模型的输入数据，输出要预测的数据；
- 5) 使用 ARIMA 模型预测用电负荷数据中的低频分量；
- 6) 将各分量预测结果和差值序列预测结果等权值求和，得到最终的预测结果。

1.2 EEMD 分解数据序列

使用 EEMD 分解经数据处理后的用电负荷数据序列和影响因素数据序列(空气温度、露点温度和海平面气压)。每次分解添加 50 次白噪声，各数据序列的分解结果如图 2—图 5 所示。

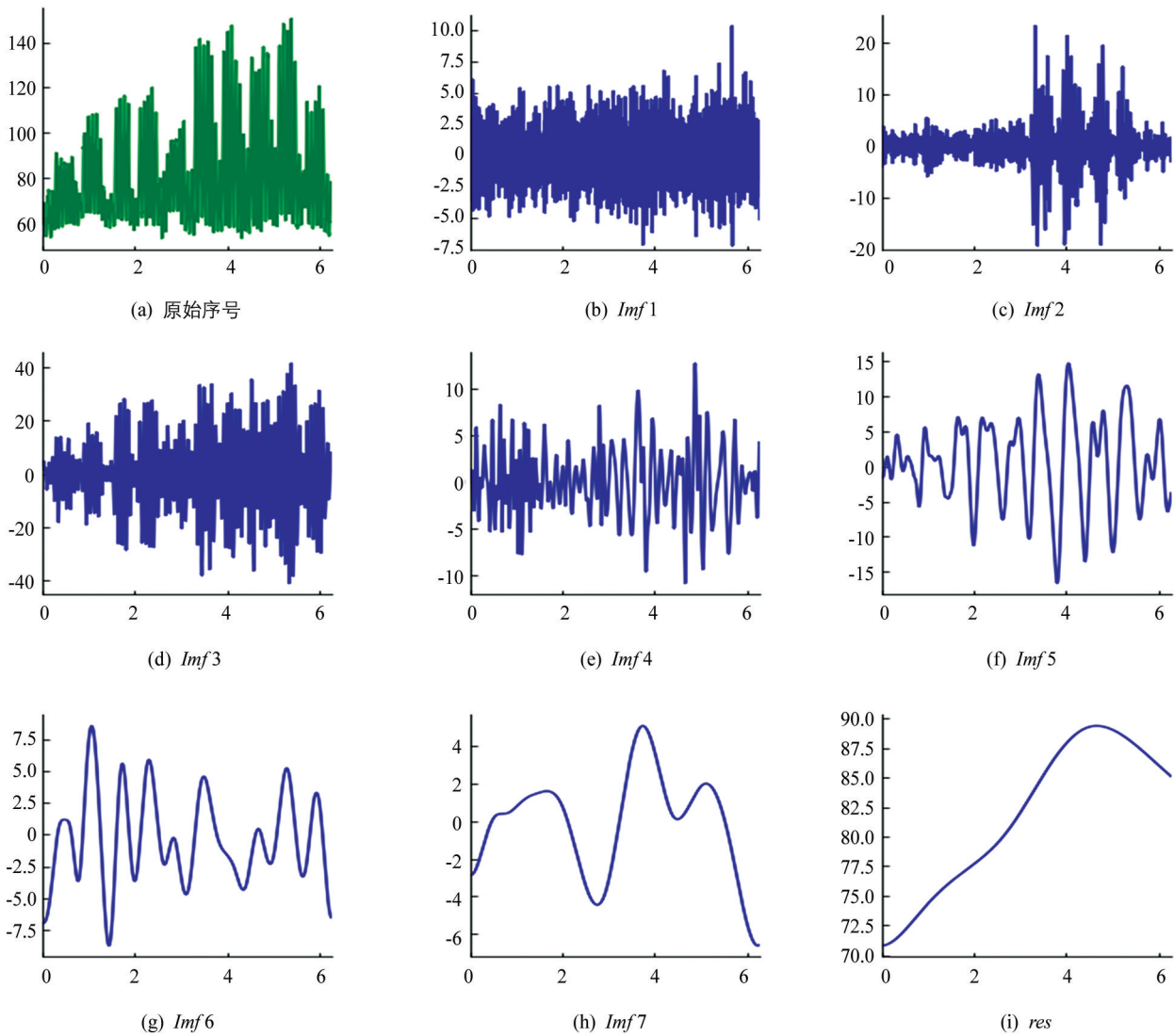


图 2 用电负荷分解结果

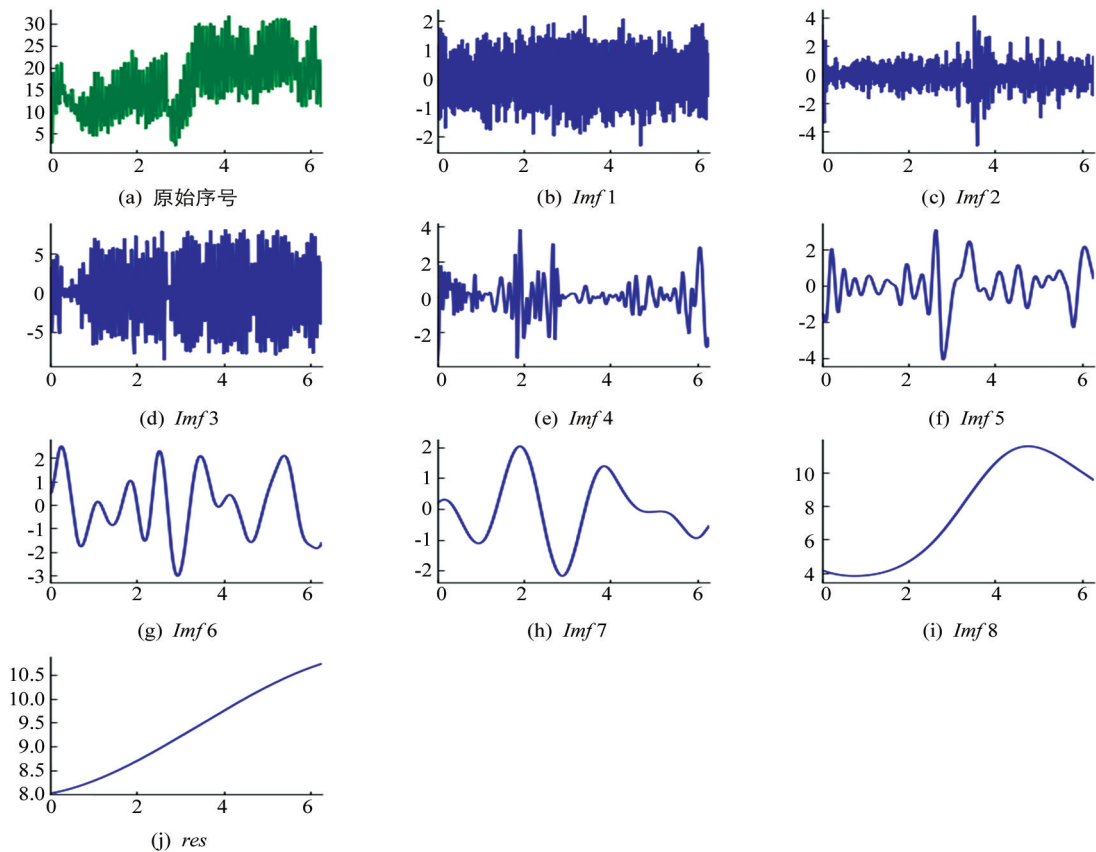


图 3 空气温度分解结果

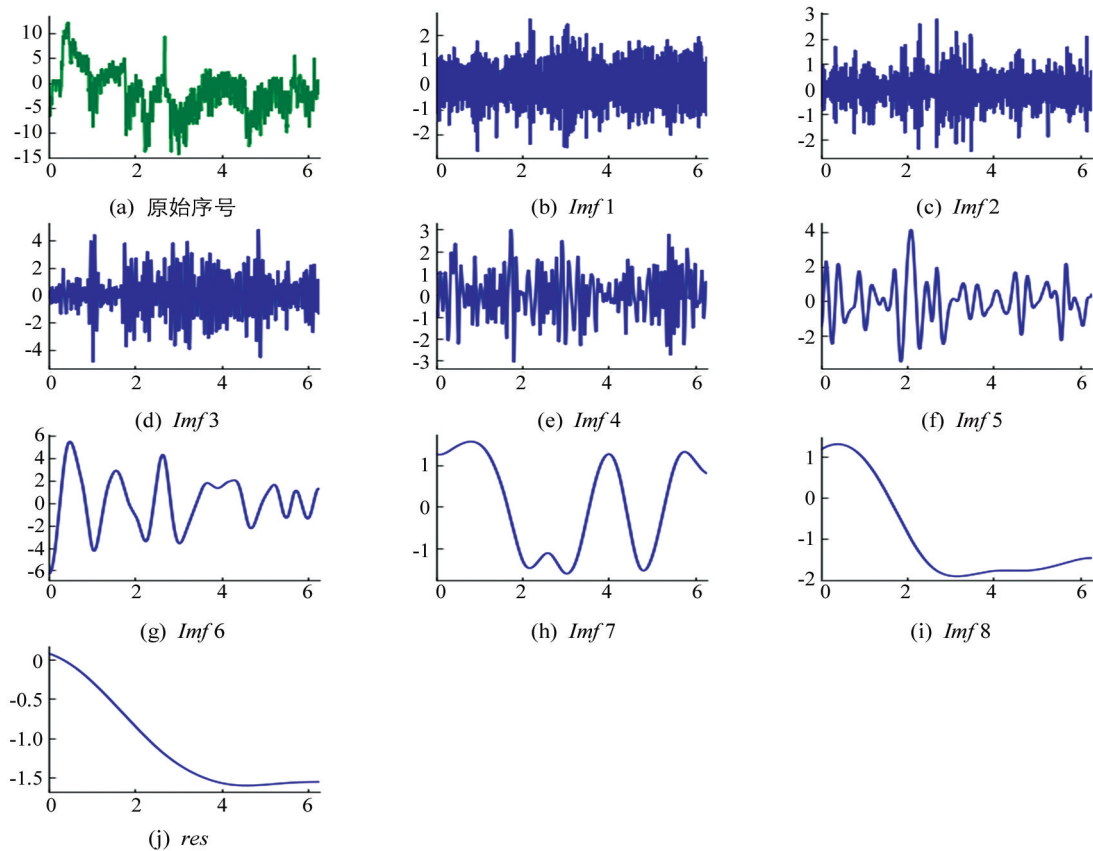


图 4 露点温度分解结果

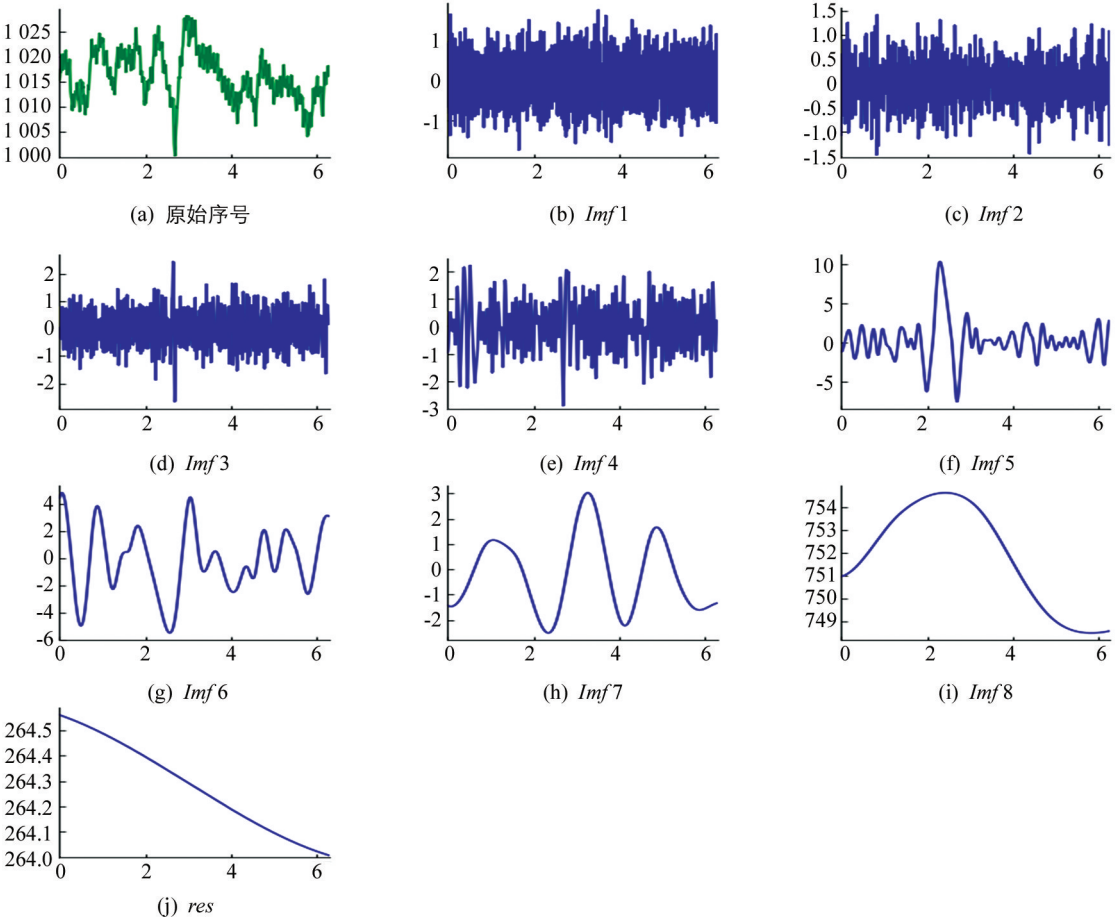


图 5 海平面气压分解结果

从图 2—图 5 可以看出，分解过程中，各分量的振幅逐次降低，周期变长，各分量保留原数据序列的局部时间特征信号。

1.3 基于皮尔逊相关性进行特征选择的 LSSVR 建模流程

通过分析用电负荷数据经 EEMD 分解后的分量，本文选择使用构建的 LSSVR 模型预测用电负荷数的第一个 *Imf* 分量和差值序列，其中差值序列通过用电负荷数据与 EEMD 分解后的各分量之和的差值来表示。首先通过皮尔逊相关性分析要预测的分量以及各影响因素数据分解后的分量序列。图 6、图 7 是其分析结果图。

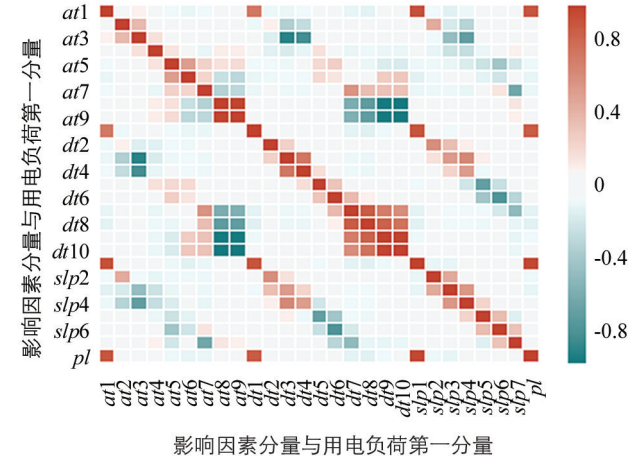


图 6 *Imf1* 相关系数热力图

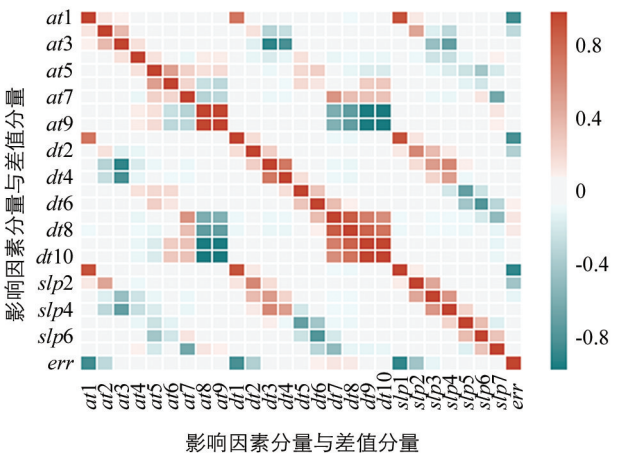


图 7 差值相关系数热力图

其中 $p1$ 表示用电负荷数据的第一分量序列, err 表示差值序列, $at1$ 表示空气温度数据的第一分量, $dt1$ 表示露点温度数据的第一分量, $slp1$ 表示海平面气压的第一分量, 依此类推.

通过皮尔逊相关分析, 本文选择的 LSSVR 模型的预测序列和输入分量如表 1 所示.

表 1 LSSVR 模型输入及输出

预测序列	输入分量	预测序列	输入分量
$p1$	$at1, dt1, slp1$	err	$at1, dt1, slp1$

1.4 ARIMA 建模流程

通过 ARIMA 模型对用电负荷数据的低频分量进行预测, 首先通过单位根检验(ADF)验证模型的平稳性, 对非平稳序列通过差分处理进行平稳化. 将最终满足平稳序列的数据通过自相关函数(ACF)和偏自相关函数(PACF)进行定阶, 最终选择的 ARIMA 模型参数 p, d, q 如表 2 所示.

表 2 ARIMA 模型参数

预测分量	参 数			预测分量	参 数		
	p	d	q		p	d	q
$Imf2$	5	0	3	$Imf6$	3	0	1
$Imf3$	4	0	1	$Imf7$	0	2	0
$Imf4$	4	0	4	$Imf8$	0	2	0
$Imf5$	5	0	0	res	1	2	0

最后将构建的 LSSVR 模型和 ARIMA 模型预测的各分量及差值序列进行叠加求和, 得到最后的预测结果.

2 实验结果与分析

2.1 数据来源及特点

为验证 EEMD-LSSVR-ARIMA 混合模型的有效性, 本文对模型进行了实验分析, 实验数据为某建筑 2016 年的实际用电负荷数据共计 8 784 条, 本文使用其中 80% 的数据作为训练集, 20% 的数据作为测试集. 经分析, 用电负荷数据具有以下几个特征:

- 1) 连续性: 用电负荷数据呈现出连续变化趋势, 即每 2 个采样点之间是连续变化的;
- 2) 周期性: 用电负荷数据呈现出周期性变化, 每隔一定时间会达到用电峰值, 随后开始下降;
- 3) 波动性: 由于受到各种因素的影响, 数据在周期变化的基础上呈现出小幅波动性.

本次实验所用软件为 Anaconda3, 编程语言为 python.

2.2 数据处理方法

由于在数据采集过程中, 受到人为因素或者外界因素干扰, 导致部分数据采集不准确, 会影响模型的训练效果. 因此需要对数据序列中的异常数据进行处理, 首先分析用电负荷数据的分布区间, 结果如图 8 所示.

从图 8 可知, 数据近似于正态分布, 因此本文选择使用 3σ 准则判断数据中的异常值. 对于检测出的异常数据本文选择使用异常值的前 5 个采样点和后 5 个采样点的均值来替换.

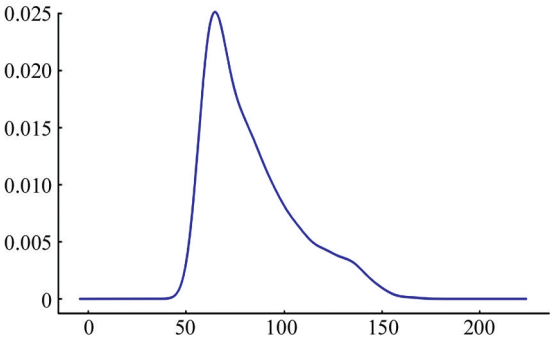


图 8 数据分布区间图

2.3 结果分析

通过 EEMD-LSSVR-ARIMA 模型对经数据处理后的用电负荷数据进行预测, 部分预测结果如图 9 所示, 其横坐标表示预测的采样点, 纵坐标表示预测值。

为有效分析模型, 本文分别通过经典的 ARIMA 模型和不考虑影响因素分量的 EEMD-ARIMA 模型, 以及基于残差的 ARIMA-LSTM 模型作为对比实验对同样的数据集进行了预测, 部分预测结果如图 10、图 11 和图 12 所示。

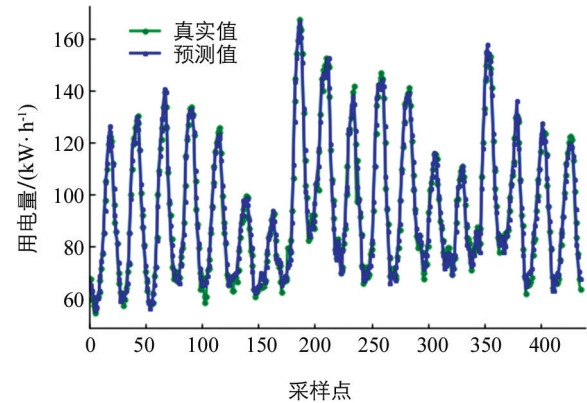


图 9 EEMD-LSSVR-ARIMA 模型预测结果

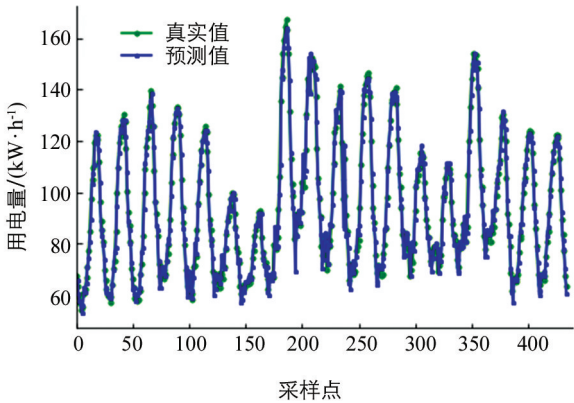


图 10 ARIMA 模型预测结果

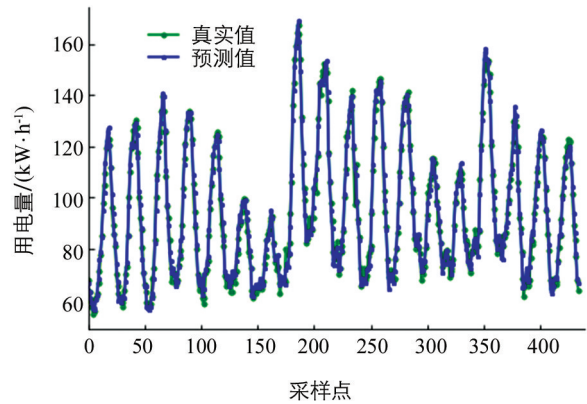


图 11 EEMD-ARIMA 模型预测结果

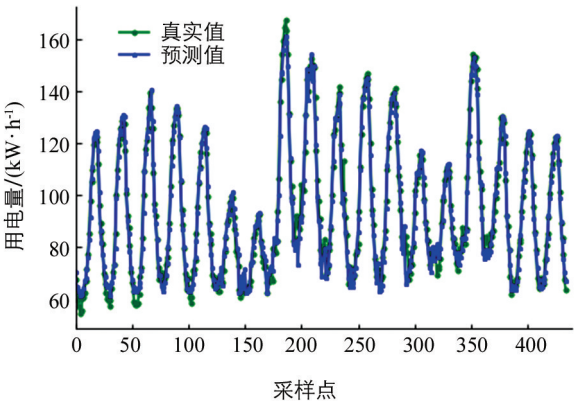


图 12 ARIMA-LSTM 模型预测结果

表 3 为 4 种模型的部分预测结果与真实值的比较. 本文通过分析模型预测误差的平均绝对误差和均方根误差, 以及验证模型预测数据与真实数据拟合程度的 R^2 来分析比较 4 种模型, 分析结果如表 4 所示。

表 3 模型对比结果

编号	真实值	预 测 值			
		EEMD-LSSVR-ARIMA	ARIMA-LSTM	EEMD-ARIMA	ARIMA
1	122.499	121.065	122.040	118.186	120.826
2	122.164	121.494	122.840	118.862	121.757
3	121.298	116.215	118.406	119.930	119.118
4	113.349	112.597	111.973	113.274	117.545
5	101.079	104.252	98.684	103.974	105.823

表 4 模型对比结果

评价指标	EEMD-LSSVR-ARIMA	ARIMA-LSTM	EEMD-ARIMA	ARIMA
MAE	2.414 8	2.779 1	2.564 3	3.346 8
RMSE	2.997 4	3.183 8	3.286 9	4.734 6
R^2	0.986 4	0.978 4	0.976 3	0.966 1

从上述对比实验可以发现, 单一的 ARIMA 模型在所有预测模型中表现最差, 基于时间序列多尺度分解但未考虑影响因素的 EEMD-ARIMA 模型表现优于单一的 ARIMA 模型, 而基于残差的 ARIMA-LSTM 模型表现与 EEMD-ARIMA 模型相近, 拟合度略高于 EEMD-ARIMA 模型, 而基于时间序列多尺度分解并且考虑影响因素的 EEMD-LSSVR-ARIMA 模型在所有模型中表现最佳, 与真实数据的拟合程度达到了 98% 以上, 误差在 4 种模型中最低.

3 总 结

我国作为一个能源消耗大国, 对于能耗数据尤其是电力消耗数据的研究起步相对较晚, 用电问题一直都是困扰我国经济发展的关键问题. 本文首先整理和分析了近年来对于用电负荷预测的研究现状, 针对目前的研究情况提出了自己的见解. 并针对某建筑用电负荷数据特征, 选择空气温度等气象数据作为影响因素, 拟合建筑用电规律, 提出了一种新型的基于时间序列多尺度分解的 EEMD-LSSVR-ARIMA 混合预测模型, 并通过实验验证了模型的有效性, 实现了对建筑用电负荷数据的精准预测.

参考文献:

[1] 李婉华. 面向电力大数据的数据挖掘算法研究及应用 [D]. 福州: 福州大学, 2017.

[2] KAYTEZ F. A Hybrid Approach Based on Autoregressive Integrated Moving Average and Least-square Support Vector Machine for Long-term Forecasting of Net Electricity Consumption [J]. Energy, 2020, 197: 117200.

[3] 王月强, 王 煥, 李卫彬, 等. 长兴岛产业园区电力需求多时序预测技术研究 [J]. 电力与能源, 2020, 41(2): 160-166, 171.

[4] 魏华栋, 陶 媛, 蔡昌春, 等. 基于改进长短期记忆神经网络的短期负荷预测 [J/OL]. 电测与仪表. (2020-09-12) [2020-09-25]. <https://kns.cnki.net/kcms/detail/23.1202.TH.20200911.1902.002.html>.

[5] 罗澍忻, 麻敏华, 蒋 林, 等. 考虑多时间尺度数据的中长期负荷预测方法 [J/OL]. 中国电机工程学报. (2020-07-01) [2020-09-25]. <https://kns.cnki.net/kcms/detail/11.2107.TM.20200701.0954.008.html>.

[6] 邢 伟, 母卓元. 基于 EIEMD-IMDE-XGBoost 模型的短期电力负荷预测 [J]. 山西电力, 2020(2): 9-14.

[7] 李 旋, 周清锋, 戴吾蛟, 等. 经验模式分解与小波分解在变形分析应用中的对比分析 [J]. 工程勘察, 2009, 37(7): 67-71.

[8] 蔡新闻. 基于 EN-LSSVR 模型的短时交通流预测方法研究 [D]. 镇江: 江苏大学, 2019.

[9] 夏邓成. 基于 LSSVR-HSMM 的滚动轴承剩余使用寿命预测研究 [D]. 武汉: 华中科技大学, 2019.

[10] 涂钊颖. 广东省博罗县中长期电力负荷预测的研究 [D]. 长沙: 湖南大学, 2019.

A Method for Forecasting Building Power Load Based on Time Series Multi-scale Decomposition

JU Ya-xuan, ZHANG Chun-yu, ZHU Ren-jing, ZHU Xi-jun

*School of Information Science and Technology, Qingdao University of Science and Technology,
Qingdao Shandong 266100, China*

Abstract: With the rapid growth of building energy consumption in recent years, building energy conservation has become an important issue in the sustainable development strategy. Therefore, building a model that can quickly and accurately predict building energy consumption has become a key step to achieve building energy conservation. In a study reported in this paper, combined with empirical mode decomposition, LSSVR and ARIMA were integrated to construct a mixed model of EEMD-LSSVR-ARIMA based on time series multi-scale decomposition to predict building electricity data. With this model, the electricity load data of the building was decomposed into a number of different-frequency components by EEMD, the LSSVR model was used to predict the high-frequency component, and the difference between load data and the sum of the components in the sequence, and the ARIMA model was used to predict the low-frequency component. Lastly, the predicted results of the components and those of the difference sequences were superimposed, and the final results were obtained. This method was employed to analyze the data of power consumption of a certain building, and was compared with the traditional ARIMA and EEMD-ARIMA models and the residual-based ARIMA-LSTM model. The experimental results showed that the prediction accuracy of the model proposed in this paper was more than 98%, which was nearly 2% higher than that of the other models.

Key words: energy consumption prediction; EEMD(Ensemble Empirical Mode Decomposition); LSSVR (least square support vector regression); ARIMA(Auto-regressive Integrated Moving Average)

责任编辑 崔玉洁