[文章编号] 1003－4684(2026)01-0012-06

基于 AK-means 聚类算法和 BLSTM 的中央空调负荷预测研究

丁 璐[1,3]，戚得众[1,2,3]，田 晨[1,3]，丁 坦[4]

(1 湖北工业大学农机工程研究设计院，湖北 武汉 430068；2 山东农业大学机械与电子工程学院，山东 泰安，271018；
3 湖北省农机装备智能化工程技术研究中心，湖北 武汉 430068；4 武昌首义学院机电学院，湖北 武汉 430068)

[摘 要] 精准的冷负荷预测是中央空调系统节能优化运行的关键前提。为进一步提高冷负荷预测精度，提出一种融合 AK-means 聚类算法与 BLSTM 的中央空调负荷预测模型。首先，改进 K-means 算法，引入动态调整 K 值策略，形成 AK-means 算法，以预处理输入数据，提升数据质量；其次，采用白头鹰搜索算法优化 LSTM 网络参数，构建 BLSTM 预测模型；最后，将 BLSTM 模型与 BP、CNN、LSTM 模型的预测结果进行对比。实验表明，相同数据集与评价指标下，BLSTM 模型性能最优，其平均绝对误差较 BP、CNN、LSTM 模型分别降低了 66.12%、78.75%、27.19%。本研究为中央空调系统节能优化与智能调控提供了方法参考。

[关键词] 中央空调；负荷预测；AK-means 聚类算法；BLSTM；BES 算法

[中图分类号] TP273 [文献标识码] A

建筑能耗约占总能耗的 33%[1]，其中商业建筑能耗所占比重逐年增长。中央空调系统是商业建筑中的主要能耗来源，约占商业建筑总能耗的 40%[2]。降低中央空调系统能耗尤为重要，而准确的冷负荷预测是降低中央空调系统能耗的最有效途径之一。

根据现有的研究和建模方法，构建冷负荷预测模型可分为机理建模和数据驱动建模[3]。机理建模依赖于已知的物理、化学和生物等科学原理来描述系统行为和过程的建模方法。构建机理模型需要精确的物理参数，但在实际生产过程中这些参数难以获得，导致模型结果与实际情况有差异。数据驱动建模是利用数据来建立模型，从大量数据中挖掘出模式、规律和关联性，实现对系统行为的描述、预测和控制。相对于机理建模，数据驱动建模可以更快地构建模型，随时根据新数据更新模型，实现实时预测和决策。谢文举等[4]提出一种基于粒子群、卷积神经网络和长短期记忆网络的混合预测模型，该模型对网络的超参数进行寻优，提高拟合精度。陈馨凝等[5]提出一种基于改进磷虾群算法优化 ESN 神经网络的负荷预测模型，对网络层与层之间的连接权值进行优化，提高了预测结果的准确性。MaoYun 等[6]提出了一种基于非线性混沌映射的哈里斯鹰优化全 Elman 神经网络的建筑冷负荷预测模型，解决了模型结构参数调整复杂等问题。He Ning 等[7]构建一种基于时间卷积网络的中央空调系统预测模型，降低了过拟合风险。严良文等[8]提出一种基于支持向量回归机的中央空调负荷预测模型，并与 BP 神经网络负荷预测模型进行对比，证明支持向量回归机的预测精度更高。王蕾等[9]提出一种粒子群算法改进 BP 神经网络参数的预测模型，并引入了控制误差反馈参数来提高模型精度。上述方法多侧重于单一模型的优化或复杂参数调整，对中央空调负荷数据的动态时变特性与运行模式多样性的综合考量尚显不足。

本文提出一种基于 AK-means 聚类算法和 BLSTM 的中央空调负荷预测模型，即利用 AK-means(adaptive K-means)聚类算法预处理数据，并结合 BES 优化 LSTM 参数，构建 BLSTM 预测模型，提升预测精度与稳健性。白头鹰搜索算法[10](bald eagle search optimization algorithm，BES)是一种新兴的元启发式算法，具有计算精度高、参数少、收敛速度快等优点。本文引入 BES 算法优化长短期记忆网络(long short-term memory，LSTM)模

[收稿日期] 2024－06－18
[基金项目] 湖北省技术创新专项重大项目(2023BEB038)；湖北省教育厅科研规划项目(B2021358)
[第一作者] 丁 璐(1999－)，女，陕西宝鸡人，湖北工业大学硕士研究生，研究方向为智能算法改进。
[通信作者] 丁 坦(1978－)，男，湖北武汉人，工学博士，湖北工业大学讲师，研究方向为复杂系统优化控制与建模。

型,建立了白头鹰搜索算法-长短期记忆网络(BLSTM)的中央空调负荷预测模型,解决人工调参盲目选择参数的问题。

1　AK-means 聚类算法

K-means 聚类算法[11]是一种基于划分的经典聚类算法。其将给定的数据集分割成 K 个簇,使得同一簇中的数据点在特征空间中接近,而不同簇之间的数据点远离,用欧式距离作为衡量数据对象间相似度的指标,相似度与数据对象间的距离成反比,且相似度越大距离越小。但 K-means 聚类算法在运行时需先人为选择簇的数量 K,故容易受到主观判断的影响。固定的 K 值无法根据数据的内在结构进行调整,导致聚类结果的质量不稳定。因此,本文提出一种基于"肘部法"动态调整 K 值的 K-means 聚类算法(adaptive K-means 聚类算法)。AK-means 聚类算法能够根据数据自动选择最合适的簇数,避免人为设定 K 值的主观性,提高算法对异常值的鲁棒性。

本文采用误差平方和(SSE)作为聚类效果的评判指标。基于"肘部法"动态调整 K 值步骤如下:

步骤 1:设定候选 K 值范围,依次进行 K-means 聚类;

步骤 2:对范围内每个 K 值执行聚类,计算对应的误差平方和(SSE);

步骤 3:建立 K 值与 SSE 的对应关系曲线,识别 SSE 下降速率由快变慢的拐点,计算曲率最大值,确定最优聚类数。

AK-means 聚类算法具体步骤如下:

步骤 1:设定最大可能簇数和最大迭代次数。

步骤 2:使用肘部法动态选择 k 个聚类中心 $C_i(i \leqslant 1 \leqslant k)$。

步骤 3:计算其余数据对象与聚类中心 C_i 的欧氏距离,找出离目标数据对象最近的聚类中心 C_i,并将数据对象分配到聚类中心 C_i 所对应的簇中。

空间中数据对象与聚类中心间欧氏距离

$$d(X,C_i)=\sqrt{\sum_{j=1}^{m}(X_j-C_{i,j})^2} \tag{1}$$

其中,X 为数据对象;C_i 为第 i 个聚类中心;m 为数据对象的特征数;X_j,$C_{i,j}$ 为 X 和 C_i 的第 j 个属性值。

步骤 4:根据当前分配的簇,计算每个簇中数据对象的平均值作为新的聚类中心。

步骤 5:重复步骤 2 和步骤 3,进行下一次迭代,直到聚类中心不再变化或达到最大迭代次数时停止。

2　白头鹰搜索算法—长短期记忆网络(BLSTM)模型

2.1　白头鹰搜索算法

2.1.1　选择空间　第一阶段,白头鹰选择猎物最多的地方进行搜索,其方程如下:

$$P_{i,\text{new}}=P_{\text{best}}+\alpha\cdot\gamma(P_{\text{mean}}-P_i) \tag{2}$$

式中:$P_{i,\text{new}}$ 为第 i 只白头鹰的更新位置;P_{best} 为当前最佳位置;P_{mean} 为白头鹰个体的平均分布位置;$\alpha\in(1.5,2)$ 是控制白头鹰搜索位置的值;γ 为取值范围在 0～1 之间的随机数。

2.1.2　搜索空间　第二阶段,白头鹰选择空间之后,在选定的空间进行"螺旋式"搜索,找到最优的目标,其方程如下:

$$P_{i,\text{new}}=P_i+x(i)\times(P_i-P_{\text{mean}})+y(i)\times(P_i-P_{i+1}) \tag{3}$$

式中,$x(i)$,$y(i)$ 代表极坐标中白头鹰所在的位置,取值为 $(-1,1)$。

2.1.3　俯冲　第三阶段,白头鹰找到目标猎物之后,从最佳位置俯冲至目标猎物位置,其他白头鹰个体也同时向目标猎物俯冲,其方程如下:

$$P_{i,\text{new}}=\text{rand}\times P_{\text{best}}+x_1(i)\times(P_i-c_1\times P_{\text{mean}})+y_1(i)\times(P_i-c_2\times P_{\text{best}}) \tag{4}$$

c_1,c_2 可以使白头鹰向最佳位置移动速度加快,c_1,$c_2\in(0,1)$。

2.2　LSTM 模型

长短期记忆网络(LSTM)[12]是一种时间循环神经网络,将信息储存在每个记忆细胞中,不同隐藏层的记忆细胞之间通过少量线性交互形成一条传送带,实现信息的流动,并引入一种"门"的结构,用来新增或删除记忆细胞中的信息,控制信息的流动。LSTM 实现了遗忘门、输入门和输出门的计算,用来保护和控制细胞的状态,其结构如图 1 所示。

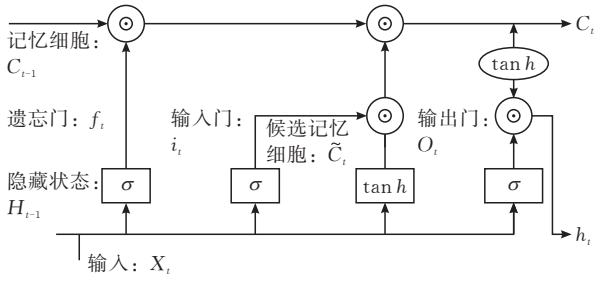


图 1　LSTM 结构

1)遗忘门　遗忘门负责决定保留上一时刻单元状态到当前时刻单元状态的哪些信息,其方程如下:

$$f_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_f) \tag{5}$$

式中:W_f,b_f 分别为遗忘门的权重和偏置;f_t 为遗忘门的函数值。

2)输入门　输入门负责决定保留多少当前时刻的输入到当前时刻的单元状态,其方程如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{7}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{8}$$

式中:W_i,b_i 分别为输入门的权重和偏置;i_t 为输入门的函数值;W_C,b_C 分别为内部状态的权重和偏置。

3)输出门　输出门负责决定当前时刻的单元状态有多少输出,其方程如下:

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

$$h_t = o_t \times \tanh(C_t) \tag{10}$$

式中:W_o,b_o 分别为输出门的权重和偏置;o_t 为输出门的函数值。

2.3　BES 算法优化 LSTM 模型(BLSTM 模型)

LSTM 模型的参数由人工调整,调参过程效率不高且缺乏系统化和自动化的调参机制,导致模型性能不能达到最优。因此,本文提出一种 BES 算法优化 LSTM 模型的 BLSTM 预测模型。该模型通过自动寻优机制,优化学习率及隐藏层神经元数量的关键超参数,BLSTM 预测模型的自动寻找最优参数代替了手动调参,减少了工作量,降低了人为因素的影响,提高了模型的精度和效率。其过程如图 2 所示。

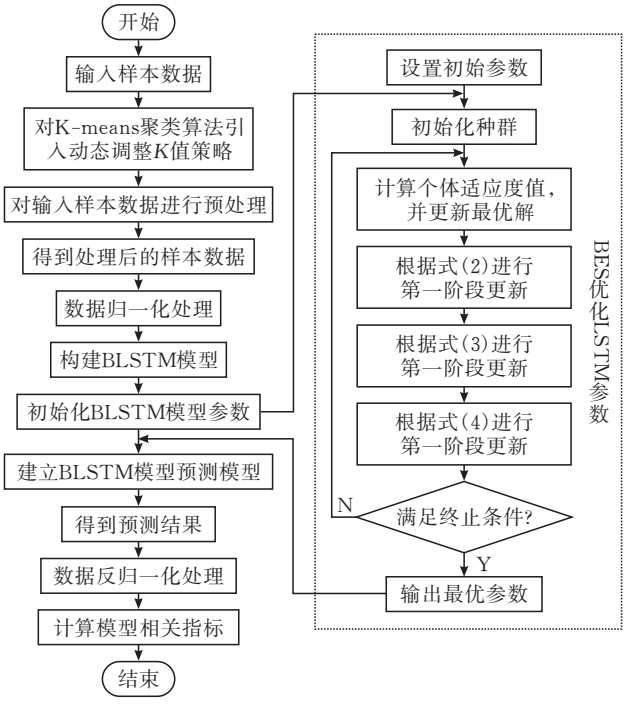


图 2　基于 AK-means 聚类算法和 BLSTM 模型的预测流程

2.4　基于 AK-means 聚类算法和 BLSTM 的模型预测

基于 AK-means 聚类算法和 BLSTM 的中央空调负荷预测,首先是利用 AK-means 聚类算法对输入数据进行数据处理,提高数据集质量,进一步提高预测效果,然后利用 BES 算法优化 LSTM 的参数,提高模型的预测精度,其流程如图 2 所示。基于 AK-means 聚类算法和 BLSTM 模型预测步骤如下:

步骤 1:对 K-means 聚类算法引入动态调整 K 值策略,得到改进后的 AK-means 聚类算法。

步骤 2:利用 AK-means 聚类算法对输入数据进行预处理。

步骤 3:将处理后的数据进行归一化。

步骤 4:设置初始参数,初始化种群。

步骤 5:计算个体适应度函数值,并更新最优解。

步骤 6:根据当前的最优解,利用式(2)、(3)、(4)更新所有个体的位置和速度。

步骤 7:判断是否满足终止条件,若不满足,则返回步骤 5;若满足,则输出最优解。

步骤 8:利用最优解优化 LSTM 网络参数,利用训练集数据训练网络。

步骤 9:利用训练后的网络对测试集数据进行负荷预测。

步骤 10:对数据进行反归一化处理。

步骤 11:计算模型相关指标。

2.5　评价指标

本文选取 RE(相对误差)、MAE(平均绝对误差)、MBE(平均偏差误差)、RMSE(均方根误差)四个指标对预测模型精度进行评价。RE 是衡量测量值与真实值之间的误差,相对于真实值的大小,可以直观地表示测量值与真实值之间的相对差异。MAE 是衡量一组预测值误差的平均大小,MAE 值越小,表示预测值与真实值之间的平均绝对误差越小,模型的预测精度越高;MBE 是衡量预测值的平均偏差,MBE 值接近零表示模型的预测值在总体上没有系统性偏差;RMSE 是衡量预测值与实际值之间差异平方的平均值的平方根,RMSE 值越小,模型的预测精度越高。它们的计算公式如下:

$$\mathrm{RE} = \frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\% \tag{11}$$

$$\mathrm{VMAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{12}$$

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{14}$$

式中:n 为样本个数;y_i 为真实值;$\hat{y}_i$ 为预测值。

3　实验分析

3.1　实验数据

本文实验数据来源于 2017 年“泰迪杯”全国数据挖掘挑战赛提供的某商业建筑中央空调系统监测数据集，采样间隔为 1 min，共包含 18 293 条有效原始时序数据。为构建适用于稳态负荷预测的高质量建模数据集，对全部原始数据进行了系统化预处理：针对缺失值采用线性插值法填补；计算每个15 min 时间窗口冷负荷的标准差，仅保留标准差低于该窗口内冷负荷平均值 3%（即 0.03 倍的平均值）的窗口数据，以此筛选稳态工况样本，并将每个入选时段内的分钟级数据按特征取平均，得到一个代表性的“稳态工况样本点”，共得到 315 个稳态样本。还引入本文提出的基于“肘部法”动态确定 K 值的自适应 K-means 聚类算法（AK-means），对全部 315 个样本的六维输入特征空间进行聚类分析，识别并剔除了 11 个异常样本，数据质量显著提升，轮廓系数由 0.22 增至 0.55。为使模型训练与算法对比实验更具普适性，采用分层随机抽样方法（抽样置信度 95%），从该高质量样本池中选出 100 个具有代表性的样本点，构成本文的核心实验数据集。每个样本包含 7 个物理特征（相对湿度、干球温度、系统热平衡、湿球温度、冷却水温差、总耗电量）作为模型输入特征，系统冷负荷作为模型预测的输出。其 100 个输出样本点如图 3 所示。

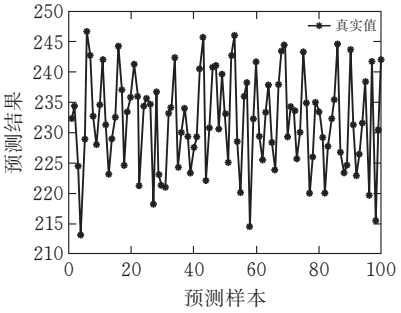
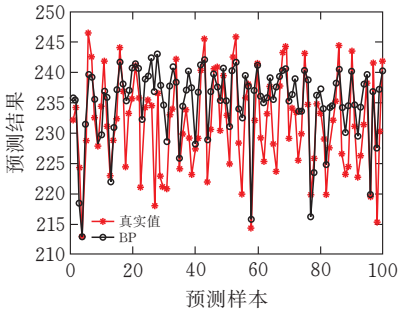


图 3　系统冷负荷输出样本数据分布

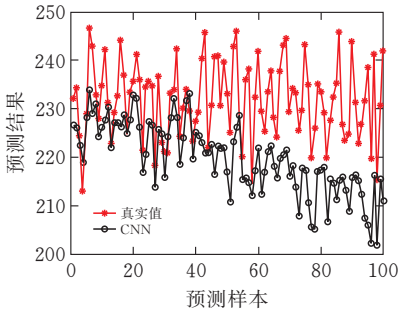
3.2　实验结果

为了验证本文所提出的基于 AK-means 聚类算法的数据处理和 BLSTM 模型的有效性，采用同一数据集，利用 BP、CNN、标准 LSTM 以及 BLSTM 等 4 种网络分别对中央空调的负荷进行预测。其中，BP、CNN 与 LSTM 模型均采用其常用默认参数，BLSTM 模型则通过 BES 算法实现了参数的自动优化。将各模型的预测值与真实值进行对比。预测结果对比如图 4 所示，预测值对比如表 1 所示。

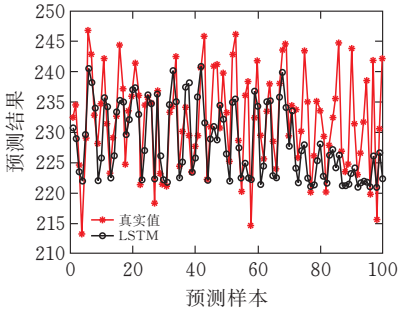
如图 4a 所示，BP 神经网络的预测曲线与真实值曲线偏离较大，尤其在样本点 X 至 Y 区间，表明



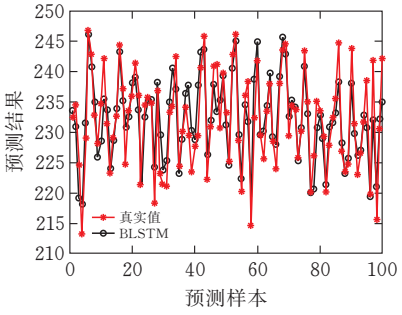
(a)真实值 vsBP



(b)真实值 vsCNN



(c)真实值 vsLSTM



(d)真实值 vsBLSTM

图 4　预测结果对比

其预测精度有限。图 4b 是真实值与 CNN 网络预测值对比。后期预测值与真实值的偏差很大，网络拟合精度差。图 4c 是真实值与 LSTM 网络预测值对比，与前两种网络相比，LSTM 网络具有较高的准确度，在 LSTM 网络的基础上改进，可得到更好的预测模型。图 4d 是真实值与参数优化后 BLSTM 网络预测值对比。通过对比图 4c 和 d 可得，LSTM 和 BLSTM 两个网络都能较为正确地预测出真实值的变化趋势，但 LSTM 网络后期预测值与真实值偏差较大。从表 1 数据可得，BP、CNN、

LSTM、BLSTM 网络预测值与真实值的误差分别为 1.8～20.9 kW、1.7～26.6 kW、1～9.6 kW、0.1～3.2 kW。如表 1 所示，BLSTM 模型的预测值与真实值的绝对误差范围(0.1～3.2 kW)明显小于 BP、CNN 和 LSTM 模型，初步表明其具有更高的预测精度。

表 1 部分预测值对比 kW

样本点	真实值	BP	CNN	LSTM	BLSTM	BP 与真实值绝对误差	CNN 与真实值绝对误差	LSTM 与真实值绝对误差	BLSTM 与真实值绝对误差
1	232.2	230.4	230.5	227.9	232.6	1.8	1.7	4.3	0.4
2	246.6	258.3	238.9	238.8	245.5	11.7	7.7	7.8	1.1
3	223.1	230.4	218.9	228.6	225.4	7.3	4.2	5.5	2.3
4	228.9	220.1	227.1	229.9	228.8	8.8	1.8	1	0.1
5	244.2	261.1	231.3	234.6	241.7	16.9	12.9	9.6	2.5
6	235.9	220.2	232.2	229.4	233.9	15.7	3.7	6.5	2
7	239.5	260.4	221.9	234.8	237.3	20.9	17.6	4.7	2.2
8	235.8	247.6	216.1	229.3	232.7	11.8	19.7	6.5	3.1
9	244.4	260.9	217.8	240.6	241.2	16.5	26.6	3.8	3.2
10	232.2	244.4	212.9	226.6	229.7	12.2	19.3	5.6	2.5

为避免预测模型的偶然性，对不同的网络模型分别进行多次实验，得到其平均值。将其平均值进行对比，分析每个预测模型的结果。表 2 为评估不同模型预测精度的相关指标平均值对比。根据表 2 列出的 RE、MAE、MBE 和 RMSE 指标对比可知，与 BP、CNN 模型相比，LSTM 模型的 RE 分别比 BP、CNN 网络降低 2.2%、4.69%；MAE 降低 53.46%、70.82%；MBE 降低 72%、83.8%；RMSE 降低 49.55%、70.68%。与前两种模型相比，LSTM 模型的拟合能力明显优于其他两个模型。但 LSTM 模型采用 BES 算法对其网络参数进行优化后，得到优化后的 BLSTM 模型，与 LSTM 模型相比，RE 降低 0.6%，MAE 降低 29.16%，MBE 降低 63.02%，RMSE 降低 29.7%。因此，经 BES 算法优化后的 BLSTM 模型，其各项误差指标均显著低于标准 LSTM 模型，预测精度得到进一步提升。图 5 为各个模型相对误差的对比。

表 2 相关指标 %

相关指标	BP	CNN	LSTM	BLSTM
RE	4.15	6.64	1.95	1.35
MAE	9.8	15.63	4.56	3.23
MBE	8.5	−14.75	−2.38	−0.88
RMSE	11.28	19.41	5.69	4

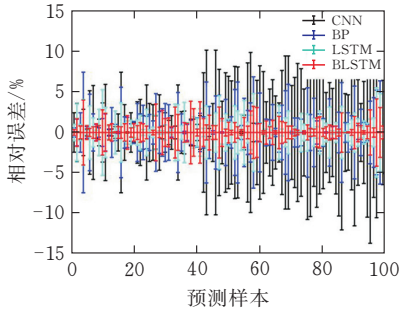


图 5 各模型相对误差对比

从图 5 可见，BLSTM 模型与其他预测模型相比，其相对误差明显较小。结合图 6 可得，BLSTM 模型预测值与真实值的误差均小于 5%。因此，BLSTM 模型的预测精度更高。

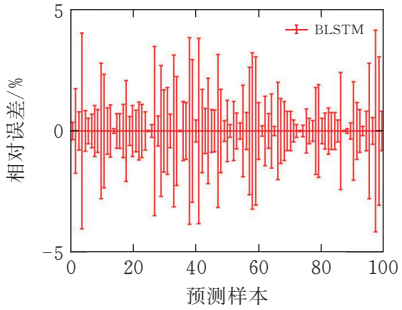


图 6 BLSTM 相对误差

4 结论

1)针对传统 K-means 聚类需要人为指定簇数、结果易受主观影响的问题，提出了基于动态调整 K 值的 AK-means 聚类算法。该算法在模型输入阶段对数据进行预处理，可有效识别并剔除异常值，提升数据质量，为后续高精度预测提供可靠数据支撑。

2)为减少人工调参的工作量与主观性，引入 BES 算法对 LSTM 模型的超参数进行自动优化，构建了 BLSTM 模型。该方法实现了模型参数的智能寻优，显著提高了预测精度与效率。

3)将优化后的 BLSTM 模型与 BP、CNN、LSTM 模型进行对比实验，BLSTM 模型在 RMSE、MAE 等多个误差指标上均表现更优，验证了本文预测模型具有更高的精度与实用性。

[参 考 文 献]

[1] ZHOU N, LIN J. The reality and future scenarios of commercial building energy consumption in China[J].

Energy & Buildings,2008,40(12):2121-2127.
[2] LEI L, WEI C, BING W, et al. A building energy consumption prediction model based on rough set theory and deep learning algorithms[J]. Energy & Buildings,2021,240.
[3] 宋文杰,林欣怡,张恒,等.基于傅里叶变换的多联机空调负荷提取及预测方法[J].暖通空调,2024,54(05):53-57.
[4] 谢文举,薛贵军,白宇.基于 PSO-CNN-LSTM 的短期热负荷预测模型[J].计算机仿真,2024,41(04):102-107.
[5] 陈馨凝,刘辉,刘雨薇,等.基于改进磷虾群算法优化 ESN 的短期负荷预测[J].湖北工业大学学报,2020,35(01):38-42.
[6] MAO Y, YU JQ, ZHANG N, et al. A hybrid model of commercial building cooling load prediction based on the improved NCHHO-FENN algorithm[J]. Journal of Building Engineering,2023,78.
[7] HE N, ZHANG L, QIAN C, et al. Short-term cooling load prediction for central air conditioning systems with small sample based on permutation entropy and temporal convolutional network[J]. Energy & Buildings,2024,310114115.
[8] 严良文,李韶鹏,韩涛,等.基于支持向量回归机的中央空调负荷预测模型[J].计量与测试技术,2016,43(06):35-38.
[9] 王蕾,张九根,李腾,等.基于改进 BP 神经网络的中央空调冷负荷预测研究[J].计算机测量与控制,2014,22(06):1690-1692.
[10] ALSATTAR H A, ZAIDAN A A, ZAIDAN B B. Novel metaheuristic bald eagle search optimization algorithm[J]. Artificial Intelligence Review: An International Science and Engineering Journal,2020,53(08):2237-2264.
[11] 刘明群,何鑫,覃日升,等.基于改进 K-means 聚类 k 值选择算法的配网电压数据异常检测[J].电力科学与技术学报,2022,37(06):91-99.
[12] 王晨,李又轩,吴其琦,等.基于 SVM-STL-LSTM 的区域短期电力负荷预测研究[J].水电能源科学,2024,42(04):215-218.

Research on Central Air Conditioning Load Prediction Based on AK-means Clustering Algorithm and BLSTM

DING Lu[1,3], QI Dezhong[1,2,3], TIAN Chen[1,3], DING Tan[4]

(1 *Agricultural Machinery Engineering Research and Design Institute of Hubei Univ. of Tech., Wuhan* 430068, *China*;
2 *College of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an* 2710182, *China*;
3 *Hubei Agricultural Machinery Equipment Intelligent Engineering Technology Research Center, Wuhan* 430068, *China*;
4 *Mechanical and Electrical College of Wuchang Shouyi University, Wuhan* 430068, *China*)

Abstract: Accurate cooling load forecasting is one of the most effective ways to reduce energy consumption in central air conditioning systems. To improve the accuracy of cooling load forecasting and reduce energy consumption, a central air conditioning load forecasting model based on the AK-means clustering algorithm and BLSTM is proposed. First, the K-means clustering algorithm is improved by introducing a dynamic adjustment strategy for the K value, resulting in the AK-means clustering algorithm. This algorithm is used to preprocess the input sample data, enhancing data quality. Second, the BES algorithm is employed to optimize the network parameters of the LSTM model, constructing the BLSTM load forecasting model. Finally, the BLSTM model is compared with three other models, and various performance metrics are calculated. Experimental results show that, compared to BP, CNN, and LSTM models, the average absolute error of the BLSTM model is reduced by 66.12%、78.75%and 27.19%, respectively, demonstrating higher forecasting accuracy.

Keywords: load prediction;BLSTM; BES algorithm,;central air conditioning

[责任编校:张　众]