

基于聚类 LSTM 深度学习模型的主动配电网电能质量预测^①

翁国庆^② 龚阳光 舒俊鹏 黄飞腾

(浙江工业大学信息工程学院 杭州 310023)

摘 要 针对较长时间跨度上电能质量(PQ)数据的时序性和非线性特点,提出一种基于 K-means 聚类和长短期记忆(LSTM)网络的主动配电网(ADN)电能质量预测方法。在构建 LSTM 深度学习模型的基础上,将大量的电能质量历史数据、环境因素及负荷数据以多维向量的形式进行 K-means 聚类,并针对每一类数据集分别使用 LSTM 模型进行网络的训练和性能评估,然后利用完成训练和评估的聚类 LSTM 网络模型进行主动配电网电能质量稳态指标项的预测。最后,通过 IEEE-13 节点含分布式电源的主动配电网仿真算例,分析验证了所提聚类 LSTM 网络法比时间序列预测法、反向传播(BP)神经网络法和标准 LSTM 网络法具有更优的预测性能。

关键词 电能质量(PQ)预测;深度学习;长短期记忆网络(LSTM);K-means 聚类;主动配电网(ADN)

0 引 言

随着分布式电源(DG)渗透率的不断提升、智能计量和信息通信技术的快速发展,配电网由被动控制过渡到主动控制成为趋势,主动配电网(active distribution network, ADN)已成为未来智能配电网最重要的发展模式之一^[1,2]。但是,由于 DG 的广泛接入以及 ADN 运行特性灵活多变,ADN 必将面对更加严峻的电能质量(power quality, PQ)问题^[3]。结合 ADN 自身“主动”特征,高水平的主动控制成为其电能质量解决方案的最重要内容^[4]。而实现高性能的电能质量态势预测、评估和预警,是能够有效进行电能质量主动控制的前提。

目前,针对电能质量预测问题国内外专家已开展了积极的探索,但系统深入的研究和取得的共识性成果尚少。文献[5]提出了基于线性回归法(LR)、随机时间序列法(RTA)和灰色模型(GM)的组合预测模型,改善了单一方法的预测精度,但提高

了权重确定及建模的难度。文献[6]结合自回归移动平均(ARIMA)模型与反向传播(back propagation, BP)神经网络进行电能质量指标项预测,利用二者各自在非平稳序列和高维非线性问题处理上的良好性能,但其对时序数据时间相关性的忽略使其在中长期预测时准确性明显降低。文献[7]利用离散傅里叶分解与时间序列自回归法(AR)相结合进行预测,因其去除了一部分频域分量,因此存在预测结果整体性缺失的缺陷。文献[8]提出了一种基于动态时间规整(DTW)及 Monte Carlo 算法的预测法,该方法引入聚类思路有效提高了预测性能,但 Monte Carlo 算法引入的随机性将导致电能质量预测结果具有一定的不确定性。文献[9]通过量化电能质量指标与温度、储能电池状态等因素的关系,提出一种预测效果较好的随机森林模型(RF)预测法,但由于其需考虑具体线路、设备参数及运行信息,仅适用于某些特定场合。上述文献所提方法各具特点,但在未来具有 DG 高渗透率的主动配电网背景下,面对系统电能质量多重影响因素和电能质量各指标项数

① 国家自然科学基金(51777193)和浙江省自然科学基金(LY17E070005)资助项目。
② 男,1977 年生,博士,副教授;研究方向:智能电网,电能质量智能监控;E-mail: wggq@zjut.edu.cn
(收稿日期:2019-07-13)

据之间的高维、非线性关联特性,如何更好地实现较长时间区段内的电能质量指标项预测性能仍具有非常大的挑战性。

近年来,随着人工智能和机器学习的兴起,包括卷积神经网络(CNN)、循环神经网络(RNN)以及深度置信网络(DBN)等深度学习模型在预测领域的各个方面得到了蓬勃发展和应用^[10],特别是长短时记忆网络(long short-term memory, LSTM)模型在负荷预测等时间序列预测方向开始扮演越来越重要的角色。文献[11]提出了一种基于 LSTM 神经网络的智能电网电力负荷预测方法,利用 LSTM 提取负荷数据在较长时间跨度上的依赖性,实现了不错的预测效果。文献[12,13]提出了一种基于模态分解及长短时记忆网络的负荷预测方法,首先对历史负荷数据进行模态分解,接着对每个模态分量建立 LSTM 模型,最后结合模型输出重构预测结果。文献[14]提出了基于 CNN-LSTM 混合神经网络模型的负荷预测方法,首先使用卷积神经网络对由负荷影响因素构造的连续特征图进行特征向量提取,接着利用提取的特征向量建立 LSTM 模型,最后进行负荷预测。LSTM 深度学习独特的网络特性及其强大的记忆功能,使其可以很好地记忆并充分考虑较长时间跨度内海量多维时序数据的时间相关性,具有良好的时间序列预测功能。

本文提出了基于改进 K-means 聚类和 LSTM 网络深度学习模型的主动配电网电能质量稳态指标预测方法。首先,提出基于聚类 LSTM 模型预测的系统框架;接着,对该框架中关键的功能模块实现原理进行详细分析,包括基于轮廓系数法的最优类别确定、基于 K-means 聚类方法的电能质量关联数据类别划分、基于随时间反向传播(BPTT)算法和适应性动量估计(Adam)梯度优化算法的 LSTM 深度学习网络模型构建;然后,以伪代码形式设计展示了所提基于聚类 LSTM 电能质量预测模型的训练学习和性能评估实施算法;最后,通过 IEEE-13 节点含分布式电源的主动配电网仿真算例,分析验证了所提主动配电网电能质量稳态指标预测方法的有效性和先进性。

1 聚类 LSTM 预测模型的系统框架

如图 1 所示,基于 K-means 聚类及 LSTM 深度学习网络模型的含 DG 主动配电网电能质量预测模型的系统框架,主要包括 PQ 数据获取、PQ 数据聚类、LSTM 模型训练与测试、PQ 数据预测 4 个功能模块。

各个功能模块主要任务如下。

PQ 数据获取模块。在含 DG 的主动配电网中合适位置,布置环境变量监测装置、负荷监测仪、电能质量监测仪等智能仪表,获取较长时间跨度内系统光照强度、温度等环境因素数据、负荷数据,以及对应的电能质量常用稳态指标项数据,并以相同时间标记为依据进行关联保存,作为预测模型的训练和性能评估数据源。

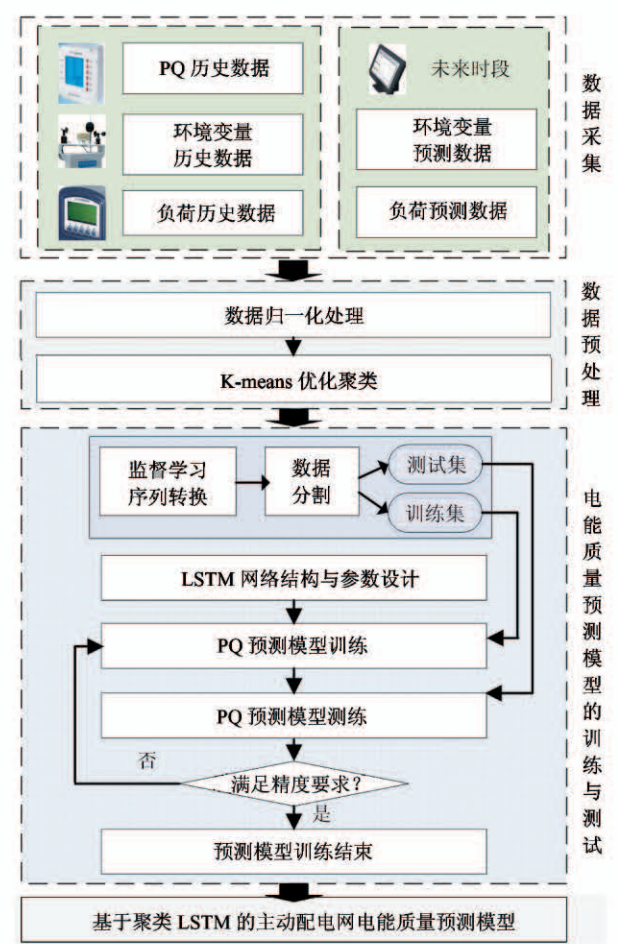


图 1 基于聚类 LSTM 模型的 PQ 预测系统框架

PQ 数据聚类。对于已经获得的 PQ 关联历史数据,在进行归一化预处理后,首先使用轮廓系数法确定其最优分类数,然后使用 K-means 聚类方法进行 PQ 关联数据的分类。

LSTM 模型训练与测试。将聚类得到的每一类 PQ 关联数据转换为监督学习序列,并将其按比例进行训练集与测试集的划分;构建 LSTM 深度学习网络模型,首先利用训练集中包含的输入和输出数据完成 LSTM 网络训练,然后利用测试集中包含的输入和输出数据完成 LSTM 网络模型的预测性能评估,直至确定各类 LSTM 预测模型。

PQ 数据预测。获取未来某时段目标电网的环境因素预测数据和负荷预测数据,并实施聚类确定其所属类别;以其作为输入数据,并调用已完成训练的相应类别的 LSTM 网络模型进行预测,模型输出即为待求的目标电网电能质量稳态指标项的预测数据。

2 K-means 聚类及 LSTM 深度学习算法

2.1 K-means 聚类及改进

由于电能质量各项指标数据在不同时间段的欧式空间内普遍呈现出相互靠近或相互远离的现象,而数据挖掘中的 K-means 硬聚类算法是以距离作为相似度的评价指标,因此,在实施电能质量预测前对其历史数据集进行 K-means 聚类使其中特性相互接近的数据归聚为一类^[15],有利于提高系统电能质量的预测精度。

常规 K-means 聚类法中对于聚类数 K 值是根据观察进行确定,具有一定的不确定性。针对这一缺陷,本文采用轮廓系数法进行最佳聚类数 K_{opt} 值的选取确定。

对于数据集中的某个样本点 P_i , 定义其轮廓系数:

$$S_i = \frac{c - a}{\max(a, c)} \quad (1)$$

其中, a 为凝聚度,表征 P_i 与同簇其他样本的平均距离; c 为分离度,表征 P_i 与最近簇中所有样本的平均距离。

最近簇的定义为

$$D_i = \arg \min_{D_k} \frac{1}{n} \sum_{q \in D_k} |q - P_i|^2 \quad (2)$$

其中, q 是簇 D_k 中的某个样本, n 为 D_k 中的样本个数。

求出所有样本的轮廓系数后,即可获得其平均轮廓系数 $\bar{S}$, 且 $\bar{S} \in [-1, +1]$ 。簇内样本的距离越近,且簇间样本的距离越远,则 $\bar{S}$ 越大,聚类效果越好。

2.2 长短时记忆神经网络

LSTM 神经网络是一种基于改进时间循环神经网络(RNN)的深度学习算法。由于记忆时间序列的长短期依赖信息是 LSTM 网络的默认行为,因此其特别适用于处理和预测时间序列中的间隔和延迟事件^[16,17]。

LSTM 神经网络由一个个内部结构相同的 LSTM 细胞单元链接而成,如图 2 所示。在整个 LSTM 网络结构中, LSTM 单元细胞状态 $\{\dots, C_{t-1}, C_t, C_{t+1}, \dots\}$ 最为关键。图 2 中上方从左贯穿到右的水平线像传送带一样,将信息从上一个细胞单元传送至下一个细胞单元,且和细胞单元中其他部分仅有很少的线性关联。LSTM 细胞单元内部结构中,网络依靠一些“门”结构让信息有选择性地影响循环神经网络中每个时刻的状态^[18]。“门”是一种能够使信息选择性通过的结构,由一个输出值在 $[0, 1]$ 区间 Sigmoid 函数和点乘操作符组成,每个 LSTM 细胞单元均包含遗忘门、输入门、输出门,其各自的核心构造、作用可由式(3)~(8)表征^[19]。

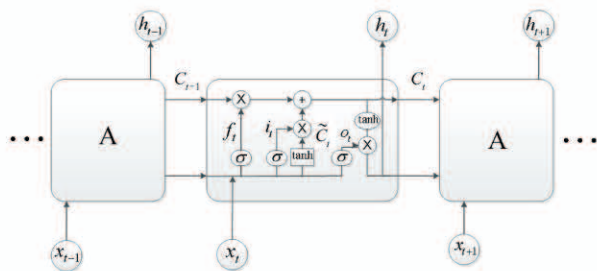


图 2 LSTM 细胞单元内部结构

遗忘门(forget gate)本质是以上一个单元输出 h_{t-1} 和本单元输入 x_t 为输入的 Sigmoid 函数,其为 C_{t-1} 中的每一项产生一个在 $[0, 1]$ 内的值,用以控制上一单元状态被遗忘的程度。图 2 中, f_t 为 Sigmoid

函数的输出序列:

$$f_t = \sigma(\mathbf{W}_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

式中, σ 为 Sigmoid 激活函数, $\mathbf{W}_f$ 为权重系数矩阵, b_f 为偏置项。

输入门(input gate)用于和一个双曲正切 tanh 函数配合控制哪些新信息可被加入。tanh 激活函数产生一个新的候选向量 $\tilde{C}_t$, 输入门为 $\tilde{C}_t$ 中每一项产生一个在 $[0, 1]$ 内的值, 控制被加入新信息的多少, 并使本单元的状态得到更新:

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (5)$$

$$\tilde{C}_t = \tanh(\mathbf{W}_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

输出门(output gate)用来控制当前的单元状态有多少被过滤掉。先将单元状态激活, 并为其中每一项产生一个在 $[0, 1]$ 内的值, 控制单元状态被过滤的程度:

$$o_t = \sigma(\mathbf{W}_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (8)$$

式中, $i_t, \tilde{C}_t, o_t$ 的形式与 f_t 相同, 但其各自的权重系数矩阵 $\{\mathbf{W}_i, \mathbf{W}_c, \mathbf{W}_o\}$ 及偏置项 $\{b_i, b_c, b_o\}$ 与 f_t 完全不同; Sigmoid 和 tanh 函数分别用以将输入连续实值“压缩”到 $[0, 1]$ 和 $[-1, 1]$ 之间一定范围, 其各自的函数曲线如图 3 所示。

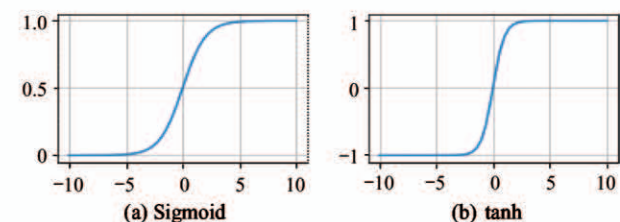


图3 Sigmoid 与 tanh 函数曲线

2.3 BPTT 与 Adam 算法

LSTM 深度学习网络模型训练过程采用随时间反向传播(BPTT)算法^[20], 其主要步骤如下。

步骤1 前向计算。按照前向计算方法计算式(3)~(8)中 LSTM 细胞的输出值。

步骤2 误差反向传播。反向计算每个 LSTM 细胞的误差项, 其值不仅与输出层有关, 而且与前后时间序列点的隐藏层有关。

步骤3 梯度计算。根据相应的误差项, 计算

每个权重的梯度。

步骤4 权值更新。根据相应误差项以及每个权重的梯度, 应用基于梯度的优化算法更新权值。

在步骤4中, 采用适应性动量估计(Adam)梯度优化算法, 其优势在于适合解决含大规模数据和参数以及包含很高噪声或稀疏度的问题。

3 基于聚类 LSTM 的 ADN 电能质量预测

3.1 数据预处理及聚类

3.1.1 数据归一化

使用多变量时间序列数据集进行含 DG 主动配电网的电能质量预测时, 需要考虑到影响电能质量的各个变量具有不同的量纲, 并且各个变量在各自的量纲内数值差别较大, 另外, 还需考虑 LSTM 深度学习模型中非线性激活函数的输入输出范围。因此, 为避免 LSTM 模型中神经元陷入饱和状态, 同时为保证各项变量能够平等地作用于电能质量的变化预测, 需要将各项变量以及电能质量指标项均进行归一化处理。

对温度、光照、用电负荷变量项以及电能质量指标项均进行归一化处理, 利用式(9)将其归算至 $[0, 1]$ 之间。与之对应, 经 LSTM 模型得到的电能质量指标项预测数据也为归一化数据, 为获得具有实际物理意义的电能质量数据, 需利用式(10) 对其进行反归一化处理。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

$$x = x' \times (x_{\max} - x_{\min}) + x_{\min} \quad (10)$$

式中, x, x' 分别表示归一化前后的指标项数值, $x_{\max}, x_{\min}$ 选取各指标项在历史数据集内的最大与最小限值。

3.1.2 电能质量数据聚类

为实现更佳的电能质量预测性能, 在进行 LSTM 预测网络的训练、评估以及真正实施基于 LSTM 模型的电能质量预测过程前, 本文针对包含目标电网环境因素以及负荷因素的电能质量数据, 将每个时刻对应的环境因素、负荷因素、以及电能质量指标数据作为一条聚类信息记录, 其结构为形如 [温度, 光照, 用电负荷, 电能质量指标] 的 4 维值向

量。然后,设置潜在可能的类别数 K_p ,采用 2.1 节中的轮廓系数法对按时间顺序排列的所有值向量进行轮廓系数计算,以此确定不同类别数下所有值向量的平均轮廓系数,并据此获得系统最佳的类别数 K_{opt} 。利用式(1)和(2),即可算出当分别取系统中所有聚类样本数据的潜在类别数 K_p 在 $[2,8]$ 之间时,其各自的平均轮廓系数 $\bar{S}$,如图 4 所示。

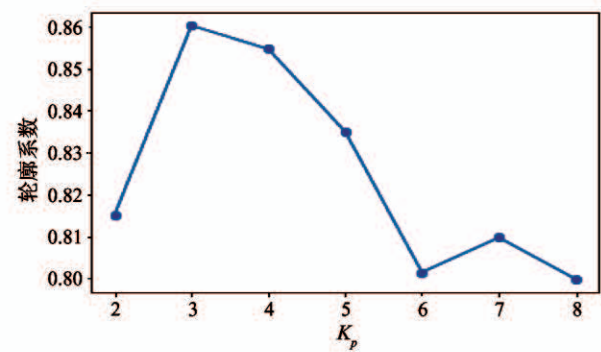


图 4 聚类数与轮廓系数关系

由图 4 可见,当类别数 K_p 取值为 3 时,其对应的平均轮廓系数 $\bar{S}$ 值最大,所以最佳类别数 K_{opt} 应为 3,该值表明了针对系统中所有聚类样本数据的聚类效果最好。然后,调用 K-means 算法,设置其聚类参数 K 为 3,并选定 K 个初始聚类中心按步骤进行数据集聚类,最后将完成聚类过程后得到的每一类子数据集分别单独保存。

3.1.3 数据分割

对于聚类处理后各类别的历史数据(包含各影响因素数据以及电能质量指标项数据),为获得可实现电能质量预测功能的 LSTM 深度学习模型,首先需要基于大量具体的输入变量历史数据和输出电能质量指标历史数据进行学习训练以获得 LSTM 网络模型的内部参数确定。其后,为评价所获得 LSTM 模型的电能质量预测性能,同样需要数量可观的具体历史数据进行测试、评估。因此,在选定用于确定电能质量 LSTM 预测模型的历史数据集后,需要对其进行数据分割。基于一般性原则,本文将采用的电能质量历史数据集按照时间顺序划分为训练集与测试集,其量值分别约占历史数据总集的 70% 与 30%。这样,既可保证 LSTM 预测模型可以充分学习历史数据集中各输入变量和输出指标项的

关联性和规律性,又能充分保障所获得预测模型性能评估的有效性。

3.2 基于聚类 LSTM 模型的电能质量预测算法实现

根据系统电能质量训练集中已完成归一化及聚类处理的输入变量和电能质量指标历史数据,可进行目标主动配电网电能质量预测的 LSTM 深度学习模型的网络训练。聚类 LSTM 预测网络模型的整体结构如图 5 所示,包含输入层、隐藏层和输出层。其中 LSTM 隐藏层中,每一个细胞单元的内部详细结构如图 2 所示。

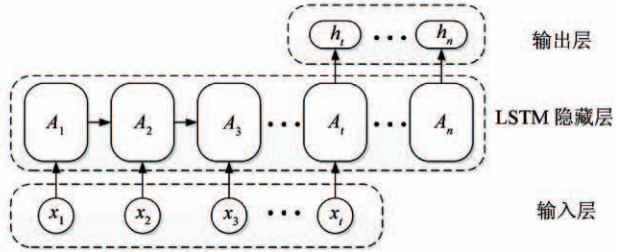


图 5 LSTM 网络 PQ 预测模型整体结构

与此对应,如表 1 中所示算法伪代码,展示了使用聚类 LSTM 深度学习模型进行含 DG 主动配电网的电能质量预测的整体实施过程。

表 1 基于聚类 LSTM 模型的 ADN 电能质量预测实施算法

PQ 预测算法实现过程	
1	读取目标电网 PQ 历史数据
2	调用 Clustering() 实现 K-means 聚类划分
1	For $i=0$ to K by 1 :do
2	读取第 i 类 PQ 数据
3	调用函数 series_to_supervised();do
	将原始时间序列转换为监督学习数据
4	跳出函数 series_to_supervised()
5	读取第 i 类监督学习数据
6	将第 i 类 PQ 数据分割为训练集和测试集
7	读取训练集监督学习数据
8	调用 Sequential();do
	将模块 Sequential() 赋予预测模型
9	将模块 LSTM() 添加到模型中
10	将模块 Dense(1) 添加到模型中
11	使用 compile (loss = 'mae', optimizer = 'adam') 模块编译模型
12	调用函数 fit();do
	拟合网络,训练 LSTM 模型
13	End

表 1 实施算法伪代码中, `series _ to _ supervised()` 是监督学习中包装好的一个集成函数, 作用在于处理时间序列数据, 可将原来的单一时间步长序列转换为可调整预测步长的时间序列; `Sequential()` 是深度学习 Keras 框架中的一个类, 实质为一个序贯模型, 其中 Keras 框架是一个开源的高度模块化的深度学习神经网络库; `LSTM()` 模块是一个包含若干个 LSTM 细胞单元的预测基础模型, 其优点在于可以将算法计算过程整合到一个模块中, 便于调参, 其次, 直接调用该模型, 能大程度上避免程序的复杂性; `Dense()` 模块代表 LSTM 网络中的全连接隐藏层, 可以根据输出维度要求直接调节括号内参数; `compile()` 模块用于编译模型, 可以将源程序解释为指令来交由 CPU 执行, 其参数 ‘mae’ 用于确定损失函数值; 参数 ‘adam’ 代表所选用的优化算法, 用于实现梯度优化, 可使梯度朝着期望的方向下降; `fit()` 为拟合函数, 其作用在于寻求训练集中影响因素与目标之间的表征关系。

3.3 聚类 LSTM 预测模型的性能评估

为合理评估经训练所确定聚类 LSTM 模型的预测性能, 一方面需要进一步借助相同类别中经数据分割后另存的测试集数据进行模型预测性能评估, 其实施过程如表 2 中伪代码所示; 另一方面需要考虑模型算法的复杂度, 其从预测结果分析以外反映了算法的性能。对于复杂度分析, 可以从模型算法的时间复杂度和空间复杂度入手, 时间复杂度可由算法的执行时间反映, 空间复杂度可由算法占用计算机内存方面反映。

表 2 聚类 LSTM 模型的预测性能评估实施算法

预测性能评估实现过程
1 读取测试集中的影响因素输入数据
2 调用函数 <code>predict()</code> : do 实现基于 LSTM 模型的 PQ 指标项预测
3 读取测试集中对应的电能质量指标项历史数据
4 For $j=1$ to m by 1: do
5 调用函数 <code>rmse()</code> : do 计算第 j 项 PQ 指标预测数值的均方根误差
6 调用函数 <code>mae()</code> : do 计算第 j 项 PQ 指标预测数值的平均绝对误差
7 End

与表 1 所示的聚类 LSTM 预测模型训练过程相对应, 对于经聚类划分的第 i 类别 PQ 历史数据子集, 读其测试集中环境、负荷等影响因素数据作为 LSTM 预测模型的输入信息, 然后通过调用 `predict()` 函数进行基于已完成网络训练的 LSTM 深度学习模型的电能质量指标项数值预测; 对于 LSTM 模型输出的任意 j 项电能质量指标数据, 均可将其与测试集中按时间标记对应的电能质量指标历史数据进行比对, 通过调用 `rmse()` 函数和 `mae()` 函数计算其均方根误差和平均绝对误差, 与此同时, 进一步通过对模型算法的复杂度分析, 以实现聚类 LSTM 预测模型的性能评估。

性能评估算法实现中, 为综合表征所得聚类 LSTM 预测模型的预测结果绝对误差性能以及预测值误差的实际情况, 选用适用于评价回归算法的均方根误差 (RMSE) 和均方误差 (MSE) 指标项, 其计算公式分别为

$$\varepsilon_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^N [y_i - \hat{y}_i]^2} \tag{11}$$

$$\varepsilon_{MSE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \tag{12}$$

式中, y_i 和 $\hat{y}_i$ 分别为基于测试集数据所获得的电能质量指标项的历史测量值与 LSTM 模型预测值; N 为用于测试评估的指标项数据的个数综合, i 为其数据编号。

4 算例分析

4.1 算例模型及数据获取

为更好地体现所提 LSTM 预测模型在系统输入变量及电能质量指标值具有较大波动特性下的各种网络运行场景中的有效性, 本文以 IEEE 13 节点配电网拓扑结构为系统基本网络架构, 基于 Matlab/Simulink 软件搭建分析算例的系统仿真模型, 系统额定电压等级和额定容量分别设定为 10.5 kV 和 500 kVA, 如图 6 所示。系统中, 为模拟含 DG 主动配电网网络特征, 在模型 13 节点处并网连接一个 10% 系统额定容量的光伏 DG 模型, 在节点 2 处布置系统电能质量在线监测点, 实时监测网络各项电

能质量数据。

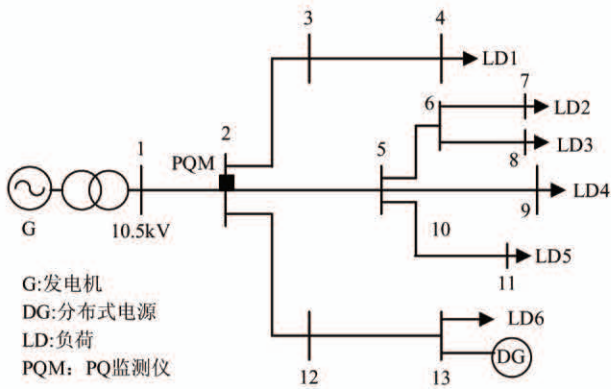


图6 含DG主动配电网结构

为获得有效、丰富的多场景下的系统输入变量和输出电能质量指标项历史数据,如图7所示,算例仿真模型中的光伏阵列模块可以通过设置接入的温度和光照参数的变化曲线,模拟外部环境因素变化以调整光伏DG的实时输出功率。仿真时,为更好体现数据真实性,以作者所在单位搭建的微电网实验平台环境监测装置所采集记录的2018年7、8、9三个月的现场光照强度和温度数据作为模型中光伏阵列的输入参数值。其中,8月份前3星期(共21天)的光照强度和温度参数的变化曲线如图8所示。同时,将模型中的LD1~LD6设置为不同类型、

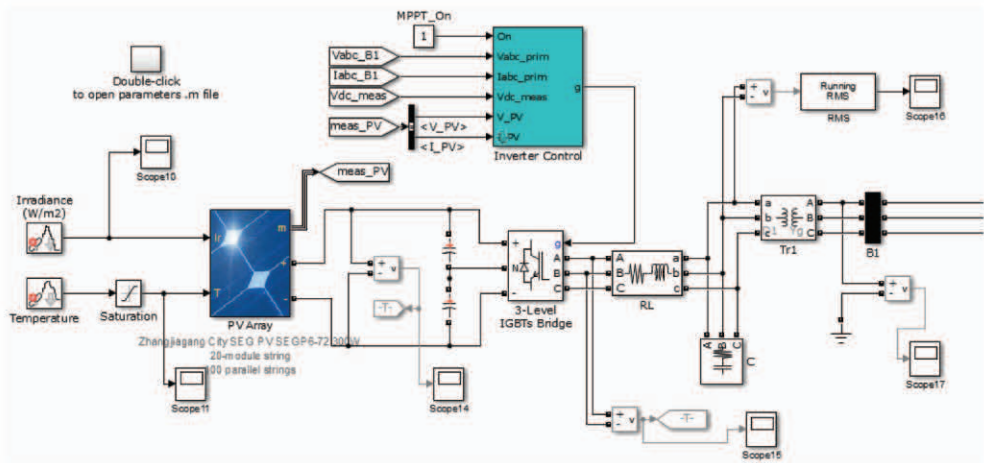


图7 光伏发电模块模型

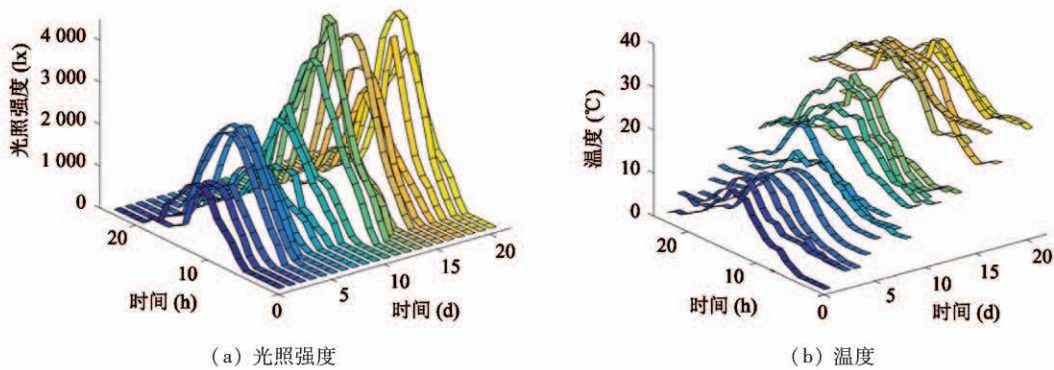


图8 21 d 内光照强度和温度分时变化情况(8月份)

不同特性的负载,并且所有负载单元均可根据设置的负荷曲线进行定时投切变换以体现网络结构和运行场景的多变性。

据此,共得3个月92组不同环境因素和系统负荷条件下的仿真数据。运行仿真模型,并获取每一

组内一天24 h 内的24组稳态电能质量各指标项数据。按照3.1小节中的数据分割准则,选取7、8两个月份的环境、负荷输入变量及对应电能质量稳态指标项数据作为LSTM预测模型的训练集,9月份的数据作为测试集。

4.2 对比预测模型选择

为验证所提基于聚类 LSTM 电能质量预测模型的优越性,选取差分自回归移动平均(ARIMA)、BP 神经网络、无聚类 LSTM 等预测模型进行对比分析。上述各个模型的基本特性和关键设置如下。

(1)ARIMA:统计模型中最常见的一种用以进行时间序列预测的模型,模型整定涉及 3 个参数(p 代表采用时序数据本身的滞后数, d 代表时序数据需要进行几阶差分化后才能稳定, q 代表采用的预测误差滞后数)。本文采用的优化对比模型为 ARIMA(5,1,0)。

(2)BP 神经网络:作为典型的前馈型神经网络,由输入层、隐含层和输出层构成,其主要特点是信号前向传输、误差反方向传播,若误差未达预期即通过误差反馈机制不断调整网络权重和阈值,使得网络平均误差和最小。本文采用的优化对比模型其动量因子设置为 0.9,输入层、隐藏层及输出层神经元个数分别设置为 4、18、1。另外,考虑到学习率过大虽然可在开始阶段加快收敛速度,但临近最佳点时易产生动荡致使无法收敛,故学习率选为较小值 0.01。

(3)无聚类 LSTM:即单独采用 LSTM 深度学习模型,但数据预处理阶段,不进行按照欧式距离原则进行的聚类划分,仅以完整的数据集统一进行电能

质量预测。

4.3 预测结果及对比分析

聚类 LSTM 预测模型输出可包括各类电能质量稳态指标项,算例仅选取其中最常用的“电压偏差”项作为所提预测算法性能的展示和对比项。基于 2018 年 9 月连续 30 d 的环境、负荷输入变量数据集,依据表 2 所示的性能评估实施算法,可进行该时段电压偏差指标项的预测,并根据预测结果计算分析其均方根误差及平均绝对误差(MAE)。同时,可与 4.2 节所提经参数调优的 ARIMA 模型、BP 神经网络模型以及无聚类 LSTM 网络模型的预测结果进行性能对比。

基于同样训练数据和评估数据条件下,ARIMA、BP 神经网络、无聚类 LSTM 和聚类 LSTM 这 4 种不同预测模型各自预测结果的 RMSE、MAE 对比结果如表 3 所示。其中,根据图 4 所示 K-means 聚类优化结果,可将本算例中电能质量历史数据集按最优类别数 K_{opt} 取值划分 3 类,表 3 中第 i ($i=1,2,3$) 类 LSTM 预测模型采用“类 i + LSTM”表示。此外,各种预测模型在其每一次训练中均存在随机因素,由此将导致其预测输出结果的不稳定性。为尽可能减少该现象对各模型性能评估的影响,表 3 中对每种预测模型均进行了 3 次测试并采用其平均值的应对措施。

表 3 各模型预测结果的 RMSE、MAE 对比

实验	ARIMA		BP 网络		LSTM 网络		类 1 + LSTM		类 2 + LSTM		类 3 + LSTM	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	8.38	6.75	9.73	7.85	6.54	5.34	6.08	4.86	5.58	4.40	5.75	4.68
2	8.39	6.76	9.51	7.61	6.64	5.41	6.09	4.87	5.62	4.41	5.68	4.65
3	8.34	6.73	9.59	7.48	6.68	5.44	6.08	4.86	5.57	4.40	5.67	4.64
平均	8.37	6.75	9.61	7.66	6.62	5.40	6.08	4.86	5.59	4.40	5.70	4.65

从表 3 可看出,对于含多输入变量的主动配电网电能质量预测,上述各模型的预测性能对比可简要描述为:(1)BP 神经网络的预测性能最差,其主要原因是在模型训练过程中,一方面因其本身缺乏记忆单元难以捕捉较长时间跨度上的信息,另一方面由于多影响因素输入使得模型确立变得复杂;(2)ARIMA 模型仅用时间作为变量因素,避免了多

影响因素变化导致的复杂性,其预测表现稍好于 BP 网络,但是如若参数选取不当,该模型的训练复杂度和时间将会明显增大;(3)标准 LSTM 网络模型,由于其可以综合考虑多变量影响因素以及 PQ 数据在时间跨度上的相关性,其预测性能明显优于 BP 网络模型和 ARIMA 模型;(4)聚类 LSTM 网络模型,由于其是在标准 LSTM 模型基础上进一步考虑了不同

类别数据之间的差异性,实现按最优聚类结果的分
类网络训练,其预测结果自然呈现出最佳的性能评
估结果。

在 30 d 预测数据集中,随机选取其中 2 天共
48 h 的预测结果数据进行对比,其实际值(即模型仿
真数据值)与各种预测模型的预测输出值对比以及
预测结果相对误差对比分别如图 9、图 10 所示。由
图可见,上述 4 种预测模型对于未来目标时段内的
电压偏差指标项呈现出性能差异的预测功能:BP 模
型和 ARIMA 模型的预测值曲线与实际值曲线大体

上满足趋势走向的一致性,但其拟合曲线波动性较
严重,整体偏离值相对较大,在部分时间节点偏差显
著;标准 LSTM 模型一定程度上改善了 BP 模型和
ARIMA 模型的上述缺陷,但对于变化趋势线中的峰
值点跟随性仍然欠佳,在部分时段内预测误差相对
较大;基于 K-means 聚类及 LSTM 网络的预测模型,
其对于电压偏差指标项的趋势预测最贴近实际值,
预测时段内全域无显著偏差点,且对于变化趋势线
中的峰值点跟随性良好。

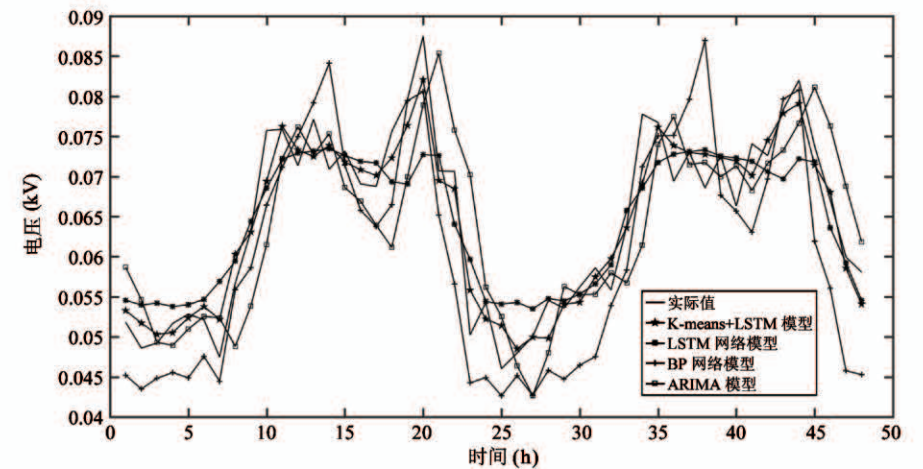


图 9 各个模型预测曲线与实际曲线对比

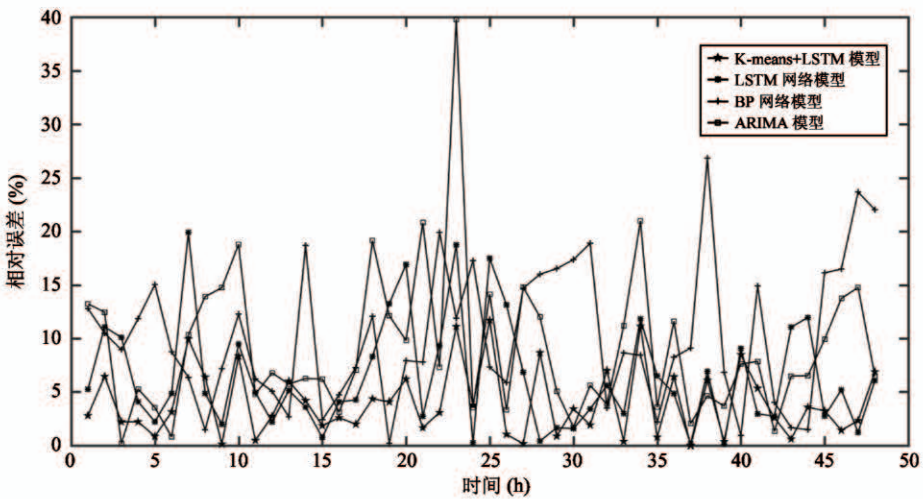


图 10 各模型预测结果相对误差对比

对上述预测结果进一步分析对比,各种预测模
型的预测结果在不同相对误差范围内的百分占比分
布对比如表 4 所示。对比分析可以看出,在这随机
抽取的 2 天共 48 h 内,本文所提聚类 LSTM 方法的

预测相对误差波动最小,其值全落于 15% 以内,且
近 70% 预测结果的相对误差处于 5% 以内,显示其
对于电压波动指标项预测效果最佳。对比分析其他
3 种预测模型,其预测结果相对误差波动峰值均超

过 25%,并且其相对误差分布在 5% 以内的相占比均不足 50% 概率,显示其预测效果与聚类 LSTM 预测模型均具有较大差距。

表 4 各模型预测结果相对误差分布对比

相对误差	ARIMA 模型	BP 模型	LSTM 模型	聚类 LSTM 模型
0 ~ 5%	29%	23%	46%	67%
5% ~ 10%	33%	33%	17%	27%
10% ~ 15%	27%	17%	19%	6%
15% ~ 20%	4%	21%	10%	0
20% ~ 25%	4%	4%	6%	0
> 25%	3%	2%	2%	0

经测试分析,表 4 中各模型的执行复杂度情况如表 5 所示。

表 5 各模型的复杂度对比

复杂度	ARIMA 模型	BP 模型	LSTM 模型	聚类 LSTM 模型
Tc(s)	246.3	11.7	7.4	5.0
Sc(MB)	71.9	166.3	162.0	165.4

表 5 中,Tc 代表模型执行的时间复杂度,Sc 代表模型执行的空间复杂度。可以看到,聚类 LSTM 模型的 Tc 值最小,其 Sc 值与 BP 模型和 LSTM 模型相近;与之对应,ARIMA 模型的 Tc 值最大,Sc 值最小。经分析可知,在当前计算机存储能力越来越强、高速化以及并行化已成为其主要特征背景下,对于测试模型对应的内存占用,计算机处理较难具有显著的区别度。因此,可主要通过时间复杂度对比,验证参与对比的 4 种不同预测模型中,所提聚类 LSTM 模型的效率最高、性能最佳。

5 结 论

本文针对未来含 DG 主动配电网中对于电能质量态势感知能力日益提高的需求,提出了基于 K-means 聚类及 LSTM 深度学习模型的电能质量预测方法。通过改进型 K-means 方法对系统电能质量历史数据进行基于欧式距离的聚类划分,然后针对各

类别数据分别建立并训练其对应的 LSTM 深度学习网络预测模型。算例分析选择其中“电压波动”指标项的预测结果表明,所提聚类 LSTM 预测模型优势显著。

LSTM 深度学习网络特性及其独特的“门”结构,可以记忆并充分考虑较长时间跨度内海量多维时序数据的时间相关性,面对含 DG 主动配电网中复杂的电能质量影响因素信息和变化特性,具有比传统预测模型更出色的态势感知和预测性能。

LSTM 深度学习网络具有可从多维信息中深度挖掘并拟合其复杂非线性关系的特性,非常适用于含多类型 DG 主动配电网电能质量预测这种典型多输入(各种环境变量、负荷变量等)、多输出(多电能质量指标项)预测应用场景,有效保障预测性能。

采用改进 K-means 聚类法可将系统中海量、多维的电能质量时序关联数据按照欧式距离原则进行最优聚类划分,使得特性相近数据归为同一类并按不同类别分别进行 LSTM 网络模型的训练和预测,可实现比标准 LSTM 模型更佳的电能质量预测性能。

后继工作中,为提高所提预测模型的普适性和预测精度,需进一步考虑高维复杂输入对于 K-means 聚类以及 LSTM 建模的影响。一方面,考虑到未来主动配电网中接入的 DG 类型多样性,可增加预测模型的输入变量类型,包括风速、风向、气压等环境变量以及 DG 控制方式等;另一方面,可考虑按季节、节假日、天气类型等不同场景进行更加合理的聚类划分。充分结合 K-means 聚类和 LSTM 网络模型两者优势,通过研究构建更加针对性、轻量化的预测模型,提高含 DG 主动配电网在各种场景下电能质量预测的实时性及准确性。

参考文献

[1] 赵波,王财胜,周金辉,等. 主动配电网现状与未来发展[J]. 电力系统自动化,2014, 38(18):125-135

[2] Hao J, Wang C, Zhao B, et al. A review of active management for distribution networks: current status and future development trends[J]. *Electric Power Components and Systems*, 2014, 42(3): 280-293

[3] Weng G, Huang F, Tang Y, et al. Fault-tolerant location of transient voltage disturbance source for DG integrated

- smart grid[J]. *Electric Power Systems Research*, 2017, 144: 13-22
- [4] 杨帆. 主动配电网中的电能质量解决方案——访江苏省电力公司电力科学研究院电网技术中心主任袁晓冬. *电力电气*, 2015, 34(23): 6-9
- [5] 丁泽俊, 刘平, 欧阳森, 等. 电能质量预测与预警机制及其应用[J]. *电力系统及其自动化学报*, 2015, 27(10): 87-92
- [6] 苏卫卫, 马素霞, 齐林海. 基于 ARIMA 和神经网络的电能质量稳态指标预测[J]. *计算机技术与发展*, 2014, 24(3): 163-167
- [7] 崔晓飞, 马智远, 许中, 等. 基于离散傅里叶分解的电能质量不平衡度指标 AR 预测方法[J]. *现代电力*, 2013, 30(6): 38-42
- [8] Bai J J, Gu W, Yuan X D, et al. Power quality prediction, early warning, and control for points of common coupling with wind farms[J]. *Energies*, 2015, 8(9): 1-18
- [9] Vantuch T, Mišák S, Jeřowicz T, et al. The power quality forecasting model for off-grid system supported by multiobjective optimization[J]. *IEEE Transactions on Industrial Electronics*, 2017, 64(12): 9507-9516
- [10] 王金甲, 陈浩, 刘青玉. 大数据下的深度学习研究[J]. *高技术通讯*, 2017, 27(1): 27-37
- [11] Zheng J, Xu C C, Zhang Z, et al. Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network[C] // 2017 51st Annual Conference on Information Sciences and Systems (CISS), Baltimore, USA, 2017: 1-6
- [12] Bedi J, Toshniwal D. Empirical mode decomposition based deep learning for electricity demand forecasting[J]. *IEEE Access*, 2018, 6: 49144-49156
- [13] 胡欣球, 马立新. VMD-LSTM 算法在短期负荷预测中的应用[J]. *电力科学与工程*, 2018, 34(6): 9-13
- [14] 陆继翔, 张琪培, 杨志宏, 等. 基于 CNN-LSTM 混合神经网络模型的短期负荷预测方法[J]. *电力系统自动化*, 2019, 43(8): 131-137
- [15] 白雪峰, 蒋国栋. 基于改进 K-means 聚类算法的负荷建模及应用[J]. *电力自动化设备*, 2010, 30(7): 80-83
- [16] 陈亮, 王震, 王刚. 深度学习框架下 LSTM 网络在短期电力负荷预测中的应用[J]. *电力信息与通信技术*, 2017, 15(5): 8-11
- [17] 李彬, 彭曙蓉, 彭君哲, 等. 基于深度学习分位数回归模型的风电功率概率密度预测[J]. *电力自动化设备*, 2018, 38(9): 15-20
- [18] 朱乔木, 李弘毅, 王子琪, 等. 基于长短期记忆网络的风电场发电功率超短期预测[J]. *电网技术*, 2017, 41(12): 3797-3802
- [19] Liu Y, Dong S T, Lu M M, et al. LSTM based reserve prediction for bank outlets[J]. *Tsinghua Science and Technology*, 2019, 24(1): 77-85
- [20] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE Transactions on Neural Networks*, 1994, 5(2): 157-166

Power quality prediction of active distribution network based on clustering LSTM deep learning model

Weng Guoqing, Gong Yangguang, Shu Junpeng, Huang Feiteng

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

Abstract

Aiming at the time series and nonlinear characteristics of power quality (PQ) data over a long time span, an active distribution network (ADN) power quality prediction method based on K-means clustering and long-term short-term memory (LSTM) network is proposed. Based on the LSTM deep learning model, a large number of power quality historical data, environmental factor data, and load data are clustered by K-means in the form of multi-dimensional vectors, next using the LSTM deep learning models to train and evaluate for each category of data, and then using the clustered LSTM network model of the completed training and evaluating to predict the data of the steady-state indicators of the power quality of active distribution network in the future. Finally, through the IEEE-13 node active distribution network simulation example with distributed power supply, the simulation results show that the proposed clustering LSTM network method has more significant prediction performance than the time series prediction method, back propagation (BP) neural network method and standard LSTM network method.

Key words: power quality (PQ) prediction, deep learning, long-term and short-term memory network (LSTM), K-means clustering, active distribution network (ADN)