

## 考虑多因素的深度学习融合方法实现负荷预测

徐先峰, 陈雨露, 王研, 王世鑫  
(长安大学 电子与控制工程学院 陕西 西安 710064)

**摘要:** 综合考虑电力负荷预测受多因素影响特性, 提出一种深度学习融合方法即 K 近邻和长短期记忆网络融合模型(KNN-LSTM)实现电力负荷预测。为了充分挖掘影响电力负荷预测的关键因素, 引入了 KNN 算法衡量各因素与负荷数据之间的欧式距离以遴选最邻近的  $K$  个负荷影响因素, 基于选取的负荷因素数据构建监督学习数据集, 引入 LSTM 模型实现了电力负荷高精度预测。结果表明, KNN-LSTM 方法可以有效地提取关键负荷因素并获得良好的预测性能。

**关键词:** 负荷预测; 深度学习; 多因素; K 近邻; 长短期记忆网络

DOI: 10.3969/j.issn.1000-3886.2020.05.019

[中图分类号] TM711 [文献标志码] A [文章编号] 1000-3886(2020)05-0061-03

## Load Forecasting in the Deep Learning Fusion Method Considering Multiple Factors

Xu Xianfeng, Chen Yulu, Wang Yan, Wang Shixin  
(College of Electronics and Control Engineering, Chang'an University, Xi'an Shaanxi 710064, China)

**Abstract:** Considering that power load forecasting is affected by many factors, a deep learning fusion method, namely K-nearest neighbor and long short term memory (KNN-LSTM), was proposed to realize power load forecasting. In order to fully mine key factors affecting power load forecasting, KNN algorithm was introduced to measure the Euclidean distance between each factor and load data to select the nearest  $K$  load-influencing factors. Based on the selected load factor data, the supervisory learning data set was constructed, and an LSTM model was introduced to achieve high-precision power load forecasting. The results showed that the proposed KNN-LSTM method could effectively extract key load factors and achieve good prediction performance.

**Keywords:** load forecasting; deep learning; multi-factor; K-nearest neighbor (KNN); long short term memory (LSTM)

## 0 引言

智能电网是实现未来可持续战略的一个关键组成部分, 而高精度的电力负荷预测是智能电网有效运行的前提, 准确预测电力负荷有利于降低发电成本, 节能降耗, 保证电网安全稳定运行<sup>[1]</sup>。因此, 电力负荷预测技术<sup>[2-4]</sup>近年来成为热门的研究领域。

如今, 深度学习<sup>[5]</sup>是人工智能领域中最热门的机器学习方法之一, 在图像处理领域、语音识别领域和电力负荷预测领域等有着广泛的应用。文献[6]提出了长短期记忆神经网络负荷预测模型, 仅仅针对实际负荷和电价数据, 结果表明模型具有更高的预测精度。文献[7]提出了一种基于数据挖掘的 SVM 优化负荷预测方法, 所提模型具有良好的准确性, 但仅将负荷数据与气象因素融合在一起, 并未提及有关其他方面相关负荷的影响。文献[8]采用均衡 KNN 算法进行短期电力负荷预测以进行场景归类, 结果表明该算法具有优异的并行预测性能, 但其并未涉及电价因素且没有考虑因素之间的影响程度大小。

负荷信息是一种高维时间序列数据, 不仅受气候因素影响, 电力负荷量和电价等都会不同程度地作用于负荷信息。而现有算法只是针对单一因素或者特定因素研究电力负荷预测技术, 设计多负荷因素高精度预测的方法较为匮乏。为了充分挖掘负荷多影响因素以提高预测精度, 本文利用欧式距离适合处理随机序列相似度评价问题的优点, 引入 KNN 算法衡量各因素与负荷数

据之间的欧式距离以遴选最邻近的  $K$  个负荷影响因素。面向电力负荷原始实测数据集, 引入 LSTM 模型实现了电力负荷高精度预测。

## 1 基于 KNN-LSTM 的预测模型

本文结合 K 近邻算法(KNN)和长短期记忆网络(LSTM)两种算法的特点, 提出了一种考虑多因素的深度学习融合方法即 KNN-LSTM 电力

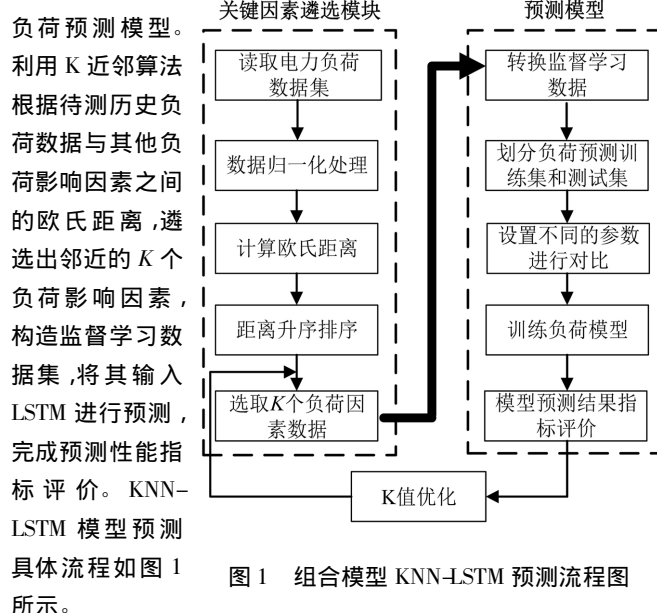


图1 组合模型 KNN-LSTM 预测流程图

定稿日期: 2020-02-18

### 1.1 基于 KNN 的关键因素遴选模块

K 近邻是一种监督学习方法,面对由历史负荷数据及其他多因素数据构成的测试样本,选用历史负荷数据作为参照样本,基于距离度量找出训练集中与参照样本最靠近的  $K$  个样本,然后综合参照样本及这  $K$  个“邻居”信息来进行预测。鉴于欧氏距离考虑到了微小时间段内不同时间序列的相似性程度,适合处理时间序列的相似度问题,因而本算法选用欧氏距离作为距离度量方法。基于 KNN 的关键因素遴选步骤如下:

步骤 1: 读取负荷数据,并进行归一化处理。

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

式中:  $x$  为电力负荷数据集中每个因素数据。

步骤 2: 计算欧式距离。计算数据集中待测负荷因素  $x$  与其他每个因素  $y_i$  之间的欧式距离  $d(x, y_i)$ 。

$$d(x, y_i) = \sqrt{\sum_{k=1}^i [x(k) - y_i(k)]^2} \quad (2)$$

式中:  $x(k)$  为待测负荷因素第  $k$  个负荷数据;  $y_i(k)$  为第  $i$  个因素的第  $k$  个负荷数据。

步骤 3: 根据步骤 2 的结果对原负荷因素数据按照递增关系进行排列。

步骤 4: 选取前  $k$  个负荷因素构造监督学习数据集。

### 1.2 LSTM 电力负荷预测模型

LSTM 是一种改进的循环神经网络,可以解决循环神经网络的梯度消失与梯度膨胀问题。本文利用 LSTM 模型适用于处理时间序列间隔这一特点,将 KNN 遴选的负荷数据构造的监督学习数据集输入 LSTM 进行电力负荷预测。计算节点如图 2 所示。

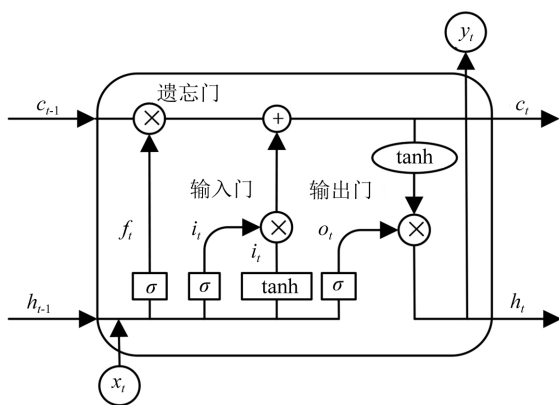


图 2 LSTM 结构图

当输入前一次的负荷信息状态从左至右穿过网络,它将先经过遗忘门,选择性地通过部分数据,输入门则通过 sigmoid 函数确定保留负荷信息的哪部分,用 tanh 函数为这些负荷信息生成一个向量  $i_t$ ,用来更新负荷信息状态。当前负荷信息状态  $c_t$  不经过任何变换直接输出,同时  $c_{t-1}$  也会经过 tanh 函数处理和输出门的过滤,作为负荷信息状态输出(即当前时间节点的输出  $y_t$ )。

$$i_t = \sigma(W_{xi}^T \cdot x_t + W_{hi}^T \cdot h_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf}^T \cdot x_t + W_{hf}^T \cdot h_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(W_{xo}^T \cdot x_t + W_{ho}^T \cdot h_{t-1} + b_o) \quad (5)$$

$$g_t = \sigma(W_{xg}^T \cdot x_t + W_{hg}^T \cdot h_{t-1} + b_g) \quad (6)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \quad (7)$$

$$c_t = h_t = o_t \otimes \tanh(c_t) \quad (8)$$

式中:  $W_{xi}$ 、 $W_{xf}$ 、 $W_{xo}$ 、 $W_{xg}$  为与当前输入的负荷数据  $x_t$  的连接权重;  $\sigma$  为 sigmoid 函数;  $W_{hi}$ 、 $W_{hf}$ 、 $W_{ho}$ 、 $W_{hg}$  为与前一时刻输出的负荷数据  $h_{t-1}$  的连接权重;  $b_i$ 、 $b_f$ 、 $b_o$ 、 $b_g$  为偏移项。

## 2 KNN-LSTM 模型仿真分析

### 2.1 数据集分析

为验证本文算法的有效性,利用新英格兰某地区电力负荷数据集进行试验分析。数据集包括 2017 年共 365 天的负荷数据,另外还包括日前竞标负荷、真实负荷、电价以及温度和湿度等 13 个负荷影响因素,数据每隔一小时采集一次,共 8 760 个数据样本。本文选取真实负荷为待测负荷因素,其余因素为辅助影响因素。在试验时,把数据集按照 9:1 的比例分为训练集和测试集。

### 2.2 LSTM 参数设置

本次试验中,LSTM 预测模型结合 KNN 算法遴选的其他负荷影响因素,根据前一小时历史负荷数据预测下一小时。为保证预测模型的性能,本文分别针对隐藏层神经元数量和学习率的设置进行试验仿真以确定最佳超参数。在损失函数为均方误差损失函数(MSE)情况下,试验设置了不同的隐藏层层数和学习率。为了评价预测效果,本文使用均方根误差(RMSE)和平均绝对误差(MAE)两个指标。试验对比结果如表 1 所示。

表 1 不同参数的结果对比

损失函数	隐藏层数量/层	学习率	RMSE	MAE
MSE	30	0.001	19.81	18.03
		0.010	11.38	10.16
	50	0.001	11.08	9.84
		0.010	3.48	2.81
	75	0.001	26.88	26.19
		0.010	15.87	15.10

观察表 1 可知,当隐藏层神经元数量取 50 时,学习率为 0.010 时预测结果最好。因此,本试验 LSTM 模型选择学习率为 0.010 的 MSE 作为损失函数。

## 3 试验结果分析

经过试验仿真,真实负荷预测曲线如图 3 所示,实线表示真实负荷值,虚线表示预测负荷值。此时,KNN 算法选出的相关负荷影响因素个数为 4,即  $K=4$  时,预测效果最佳。从图 3 可以看出,预测值与真实值具有非常高的拟合度。

### 3.1 对比试验

为了验证使用方法的有效性,本试验计算了模型的 RMSE 和 MAE,选取了 KNN、LSTM、SVM 与本文提出的算法进行对比, RMSE 和 MAE 平均降低了 25.38 和 17.62,预测结果如表 2 所示。KNN-LSTM 的误差更小,预测性能更好。

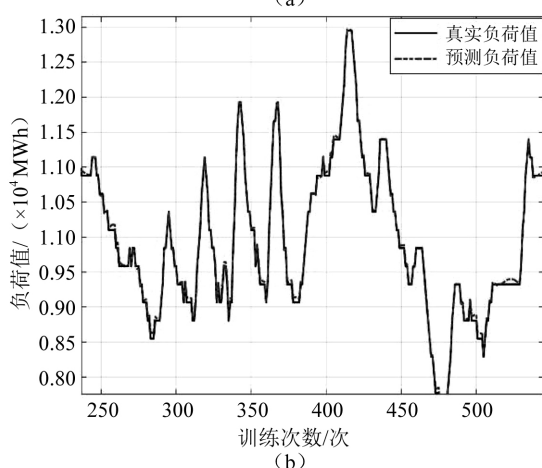
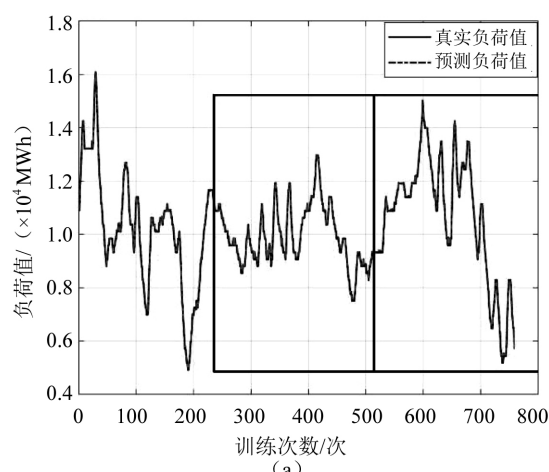


图 3 KNN-LSTM 模型预测曲线

表 2 不同模型的结果对比

模型结构	RMSE	MAE
KNN	35.70	22.75
LSTM	7.31	4.01
SVM	33.14	26.11
KNN-LSTM	3.48	2.81

### 3.2 K 值结果分析

K 值的大小直接关系着预测精度,故选取合适的 K 值是本实验很重要的一个环节。经过多次模型训练,得出 K 取 4 时,即真实负荷、日前竞标负荷、温度和湿度和模型拟合效果最佳。结合新英格兰所处的地理位置(每年降水量和降雪量较大)以及电价发展,筛选的负荷影响因素符合试验数据的分析。为了验证算法结果,计算了 K 从 1 到 14 的 RMSE 和 MAE,其变化曲线如图 4 所示。另外需要指明的是,本文采用的 K 值遴选算法同样适用于其他领域的多因素预测。

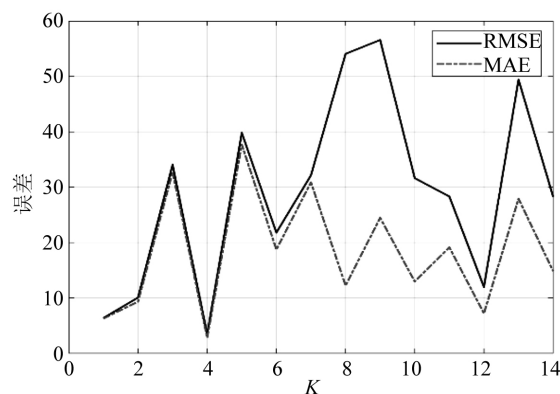


图 4 误差变化曲线

## 4 结束语

为了实现多因素的高精度负荷预测,本文建立了一种考虑多因素深度学习融合方法即 KNN-LSTM 电力负荷预测模型。为了评估提出的模型,分别与 KNN、LSTM、SVM 在同一数据集上进行了比较分析。结果表明,与传统预测模型相比,所提出的 KNN-LSTM 模型是一种效果更优的负荷预测方法,算法遴选出了包括历史负荷数据的 4 个负荷影响因素, RMSE 降低到了 3.48, MAE 降低到了 2.81,表明算法可以为多因素电力负荷预测提供必要的依据。

### 参考文献:

- [1] MASERA M, BOMPARD E F, PROFUMO F, et al. Smart (Electricity) grids for smart cities: assessing roles and societal impacts [J]. Proceedings of the IEEE, 2018, 106(4): 613-625.
- [2] 王茜,李皓然,王新娜,等. 基于混沌类电磁算法优化支持向量机的短期负荷预测[J/OL]. 计算技术与自动化, 2019, 38(4): 15-18.
- [3] 任成国,肖良良,简献忠,等. EMD-LSTM 算法在短期电力负荷预测中的应用[J]. 电力科学与工程, 2019, 35(8): 12-16.
- [4] 彭文,王金睿,尹山青. 电力市场中基于 Attention-LSTM 的短期负荷预测模型[J]. 电网技术, 2019, 43(5): 1745-1751.
- [5] LE CUN Y, BENGIO Y, HINTON G. Deep learning. Nature, 2015, 521: 436-444.
- [6] 李鹏,何帅,韩鹏飞,等. 基于长短期记忆的实时电价条件下智能电网短期负荷预测[J]. 电网技术, 2018, 42(12): 4045-4052.
- [7] XI Y, WU J, SHI C, et al. Research on refined load forecasting method based on data mining [D]. Singapore: Advances in Green Energy Systems and Smart Grid, 2018, 925: 3-13.
- [8] 林芳,林焱,吕宪龙,等. 基于均衡 KNN 算法的电力负荷短期并行预测[J]. 中国电力, 2018, 51(10): 88-94; 102.

【作者简介】徐先峰(1982—)男,山东泰安人,副教授,博士,研究方向为信号处理、深度学习理论及应用、智能电网等。陈雨露(1996—)女,陕西人,硕士生,研究方向为电力负荷预测。王研(1996—)女,陕西人,硕士生,研究方向为机械故障诊断。

保护动物就是保护我们的同类。  
动物是人类亲密的朋友,人类是动物依赖的伙伴。