

基于 K-Means 和 CNN 的用户短期电力负荷预测^①



吕志星, 张 虢, 王沈征, 王 一, 程思瑾, 秦承龙

(国网山东省电力公司 泰安供电公司, 泰安 271000)
通讯作者: 吕志星, E-mail: zhixinglvgrid@163.com

摘 要: 随着电力物联网的不断发展, 用户级电力负荷预测在电力需求侧管理中呈现出日益重要的作用. 为了提高用户级电力负荷预测的性能, 本文提出一种基于 K-means 聚类与卷积神经网络特征提取的短期电力负荷预测模型. 首先, 利用 K-means 将用户聚为两类: 对于日相关性强的用户, 将相邻时刻和日周期的历史负荷作为输入, 采用 CNN 模型提取特征进行预测; 对于日相关性弱的用户, 仅将相邻时刻的历史负荷输入到 CNN 模型进行预测. 为了验证所提出算法的性能, 我们在实际的用户负荷数据上做了实验, 并与随机森林、支持向量回归机进行对比, 结果表明本文所构建模型的预测平均绝对百分误差降低了 20% 以上.

关键词: 聚类; 相关性分析; 卷积神经网络 (CNN); 短期电力负荷预测

引用格式: 吕志星, 张虢, 王沈征, 王一, 程思瑾, 秦承龙. 基于 K-Means 和 CNN 的用户短期电力负荷预测. 计算机系统应用, 2020, 29(3): 161-166. <http://www.c-s-a.org.cn/1003-3254/7287.html>

Hybrid Method for Short-Term Load Forecasting Based on K-Means and Convolutional Neural Network

LYU Zhi-Xing, ZHANG Xiao, WANG Shen-Zheng, WANG Yi, CHENG Si-Jin, QIN Cheng-Long
(State Grid Tai'an Power Supply Company, Tai'an 271000, China)

Abstract: With the development of smart grid technology, short-term power load forecasting becomes more and more important. To improve the accuracy of short-term electric load forecasting for individual users, this study proposes a load forecasting model, which is based on K-means and Convolutional Neural Network (CNN). Firstly, K-means is applied to group users into two categories. For users with strong daily correlation, the historical loads of adjacent time points and same time points in adjacent days are taken as input to the CNN model to extract abstract features for prediction. For users with weak daily correlation, the historical loads of the adjacent time points are utilized as features. To assess the performance of the proposed method, we conducted comparison experiments on real data with random forest and support vector regression. The experimental results show that the MAPE of the proposed approach is reduced by 20%.

Key words: clustering; correlation analysis; Convolutional Neural Network (CNN); short-term load forecasting

1 引言

智能电网的发展推动了用户侧智能电表的普及, 同时各种监控系统的大规模部署使得电网公司能够获

得多尺度、全方位的用户用电信息^[1], 这些大量、高分辨率的用户负荷数据不但可以用于刻画用户用电习惯, 还可以对用户的用电负荷进行预测, 因此, 挖掘用户负

① 基金项目: 国网公司科技项目 (SGSDTA00YXJS1800585)

Foundation item: State Grid Science and Technology Program of China (SGSDTA00YXJS1800585)

收稿时间: 2019-07-27; 修改时间: 2019-08-23; 采用时间: 2019-08-27; csa 在线出版时间: 2020-02-28

荷数据对于电网系统的调度优化、精细化经营管理以及服务于市场用户等方面具有重要价值^[2]。

长期以来,负荷预测是电力系统的研究热点之一。按照时间期限分类,电力负荷预测可以划分为长期负荷预测(10年及以上)、中期负荷预测(5年左右)和短期负荷预测。其中,长期负荷预测主要用于电力系统未来规划,短期负荷预测主要是为了合理安排发电机组的启停、减少旋转备用容量,从而降低发电成本、保障电网的安全稳定性。根据电力负荷数据时序性和非线性特点,短期负荷预测模型一般可以分为两类,一类是时间序列方法,通常是将电力负荷看成一种时间序列的集合,根据电力负荷的历史数据、相关影响因素构建预测模型,从而实现预测未来负荷值。常用的方法包括:傅里叶展开法^[3],指数平滑模型^[4],自回归积分滑动平均模型^[5]。这些传统的时间序列方法对历史数据随时间的平稳性要求比较高,在一定程度上过于强调对历史数据的拟合,导致预测精度不高。另一类是以机器学习为代表的人工智能方法,主要包括:统计机器学习方法中的随机森林算法(Random Forest, RF),支持向量机,以及深度学习中的长短期记忆网络(Long Short Term Memory, LSTM)等。邹云峰等^[6]将数据类进行K-means聚类后分别进行线性回归,在线损率预测上取得较好结果,但由于其采用线性回归,对非线性数据建模能力有限。黄晗等^[7]提出了基于随机森林的方法,针对历史负荷数据和天气信息对某地区进行预测。文献^[8]将小波变化与极限学习机组合,提出了一种集成学习预测模型。以上方法相较于传统的预测方法有了很大的提升,但难以兼顾到时序数据时间相关性。文献^[9,10]利用了LSTM的时间序列性,可以更好地预测非稳定状态的负荷数据。虽然这些方法可以在系统级负荷预测上取得很好的效果,但由于单个用户用电的波动性和随机性较大,系统级负荷预测方法在单个用户的负荷预测上的预测性能较差,无法满足电网精益化管理的需求。

当前短期负荷预测多以电网系统级负荷为研究对象,而对于用户级对象的研究工作较少。文献^[11]研究了不同内核的支持向量回归(Support Vector Regression, SVR)模型对单个家庭用电负荷的预测效果。文献^[12]针对呈指数级增长的电力用户数据提出一种改进的并行化的基于随机森林的负荷预测方法。文献^[13]考虑到了用户用电行为的变化,提出了扩展卡尔

曼滤波及核函数极限学习机相结合的方法对用户侧短期负荷进行预测。文献^[14]提出了一种基于决策树和小波去噪相结合的预测方法,对不同用电模式的用户进行个性化负荷预测。这些模型考虑了用户用电负荷影响因素,对不同用户的用电特性进行分析,但忽略了时间序列相关性对单个用户负荷预测的影响。

从相关工作来看,对于用户电力负荷预测而言,需解决两方面问题:一是如何有效提取与预测目标相关的抽象特征,以便预测模型学习其中规律;二是不同用户负荷特点差异较大,单一预测模型不能对特征差别较大的用户进行负荷预测,建立组合预测模型以提升预测精度是很有必要的。因此,本文从用户负荷数据特性出发,利用K-means聚类得到不同用户类别,针对每类用户的负荷数据时间相关性,搭建不同的负荷预测模型。为了能够提取与负荷相关的抽象特征,本文利用不同时间序列信息构建不同张量作为卷积神经网络的输入,构建出深层卷积网络来预测单个用户的负荷。最后,本文结合某地23个用户一年的真实用电负荷数据对本文提出的方法进行验证。

2 基于K-means聚类算法的用户分类

图1为随机选择的5个用户的某周负荷曲线图,不同颜色的曲线表示不同用户的周负荷变化情况。其中,横轴表示时刻点,每15分钟采样一次负荷数据,纵轴表示负荷数据。通过观察曲线发现,对于单个用户而言,由于生产和经营一般具有较强的周期性和规律性,其日负荷数据也具有规律性,具体可以表现为以下2个方面:如无重大突发情况,其短时间内负荷曲线波动较小(相邻时刻相关);不同日的相同时刻的负荷波动差异不大(日相关)。为此,首先我们利用K-means聚类算法来研究用户用电行为模式。

K-means算法是一种无监督学习方法,算法通过设定好的中心点对靠近它们的点进行归类,通过迭代的方式,逐次更新各聚类中心的值,直到得到最好的聚类结果。对处理大数据集而言,K-means聚类算法保持了较高的可伸缩性和高效性。算法描述如下:

(1) 选择合适的 k 值作为聚类的初始中心点。

(2) 计算其他的点到这 k 个中心点的距离,离哪个中心点的距离最小就被划分到哪个类。距离的具体计算公式如下:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

(3) 通过求平均值的方法更新该类的中心。

(4) 按步骤 (2)、(3) 迭代更新后, 直至各类中心点的值均不再变化, 迭代结束。

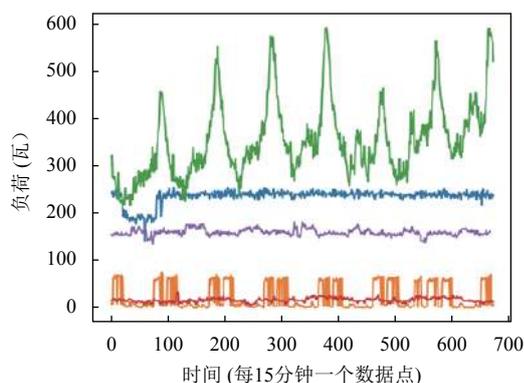


图1 某5个用户一周的负荷曲线图

本文采用 K-means 算法对真实数据集中的 23 个用户的历史数据进行聚类分析, 然后按照其时间序列的不同的相关性, 将不同类别的用户负荷序列转换为张量后输入至不同的卷积神经网络进行抽象特征提取. 聚类得到的结果如表 1 所示. 其中, A 类表示日相关性和相邻时刻相关性均较为明显的用户, B 类表示仅相邻时刻相关性较强的用户。

表 1 用户聚类结果

聚类结果	用户名称	聚类结果	用户名称
A	user1	B	user13
A	user2	B	user14
A	user3	B	user15
A	user4	A	user16
A	user5	B	user17
A	user6	B	user18
B	user7	A	user19
A	user8	B	user20
B	user9	B	user21
A	user10	B	user22
A	user11	B	user23
A	user12		

3 基于 CNN 特征提取的负荷预测模型

3.1 卷积神经网络

在深度学习的诸多算法中, 卷积神经网络 (Convolutional Neural Networks, CNN) 因其特殊的网络结构, 在计算机视觉领域应用广泛. 然而, 在电力负荷预测方面

的应用较少. CNN 的典型网络结构由输入层、卷积层、池化层、全连接层和输出层组成^[15]. 其中, 卷积层是 CNN 的核心, 包括卷积核、卷积层参数和激励函数 3 个重要部分. 卷积层内部包含多个卷积核, 组成卷积核的每个元素都对应一个权重系数和一个偏差量, 类似于前馈神经网络的一个神经元. CNN 的优势在于能够通过较少的权重参数训练提取得到特征^[16].

本文采用 CNN 网络来提取用户历史负荷的特征. 对于 A 类用户, CNN 模型用相邻的前 16 个负荷数据点和待预测点前 16 天的相同时刻的数据点分别构造成 4×4 的矩阵形式, 将这两个矩阵组成的张量作为 CNN 模型的输入, 如图 2 所示.

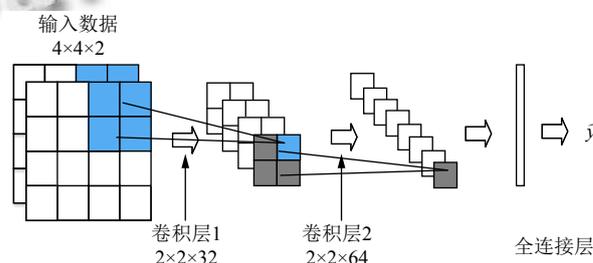


图 2 A 类用户卷积神经网络结构图

对于 B 类用户, 模型仅以相邻的前 16 个负荷数据点构建为 $4 \times 4 \times 1$ 的网络输入, 如图 3 所示. CNN 模型有两个卷积层, 卷积核分别为 $2 \times 2 \times 32$ 和 $2 \times 2 \times 64$.

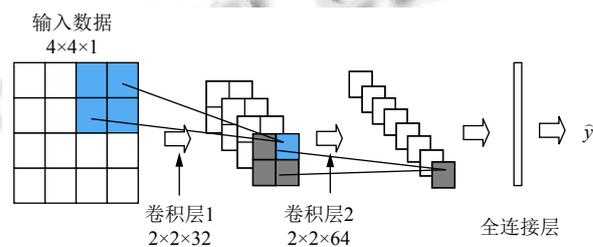


图 3 B 类用户卷积神经网络结构图

考虑到输入数据的维度不高, 因此本文在卷积层后直接接了两个全连接层, 而未使用池化层进行下采样, 保证了数据的完整性. 模型使用 ReLU 作为激活函数, 选择均方误差 (Mean Square Error, MSE) 作为损失函数.

3.2 算法流程

将历史负荷数据经过数据预处理得到用户用电负荷曲线, 通过 K-means 聚类将用户分为日相关性强和相邻时刻相关性强的两类用户. 针对不同用户负荷特

性,构建不同张量输入的 CNN 模型,具体处理流程如图 4 所示。

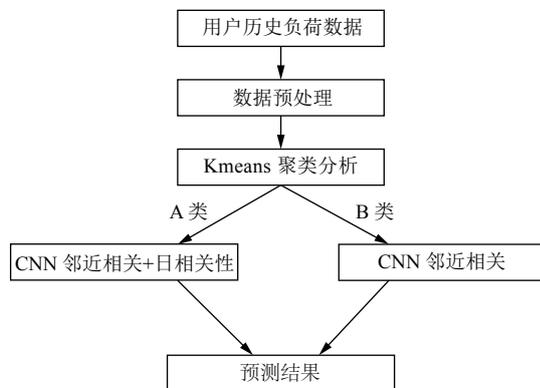


图 4 基于 K-means 和 CNN 的用户短期电力负荷预测流程

4 实验及结果分析

4.1 数据说明与数据预处理

本文所用数据集为某市 23 个用户 2015 年全年的负荷数据,每 15 分钟采集一个样本点,每天可获得 96 个负荷点数据。通过观察数据,发现原始采集数据存在缺失、数据为负或远大于该用户其他时刻点的负荷数据等失真情况。因此在用于模型训练和测试前,本文首先对采集到的负荷数据进行异常处理。对于缺失或失真的数据,本文利用该数据的前 15 个相邻数据点的平均值对其进行填充或修正,得到完整的负荷数据序列。同时,将所有的负荷数据通过归一化映射到[0,1]之间,公式如下:

$$x'_i = \frac{x_i - x_{\max}}{x_{\max} - x_{\min}} \quad (2)$$

其中, x_i 为原始负荷数据, x'_i 为归一化后的负荷数据, x_{\max} 为数据集样本中的最大负荷, x_{\min} 为数据集样本中的最小负荷。

4.2 评价指标

为了有效评价模型的准确度以及与其他方法进行对比,本文选取平均绝对误差 (Mean Absolute Percentage Error, MAPE) 作为评价指标,计算公式如下:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\left| \frac{\hat{y}_i - y_i}{y_i} \right| * 100\% \right) \quad (3)$$

其中, n 表示预测点的个数, y_i 表示第 i 个预测点的负荷真实值, \hat{y}_i 表示第 i 个预测点的负荷预测值,即 MAPE 值越小,预测准确率越高。

4.3 预测结果分析

采用本文提出的 CNN 算法对经过 K-means 算法分类的电力用户进行负荷预测,训练集与测试集划分比例为 8:2。训练集采用的是 2015 年 1 月 1 日至 2015 年 10 月 15 日的数据,共有 28 128 条记录;测试集采用的是 2015 年 10 月 15 日至 2015 年 12 月 31 日的数据,共有 7007 条记录。

(1) 将本文模型与随机森林和支持向量回归机进行比较,结果见表 2。

表 2 不同电力负荷预测方法 MAPE 比较 (%)

用户名称	RF	SVR	本文模型
user1	3.73	3.34	2.72
user2	89.95	354.25	67.40
user3	6.56	7.36	6.47
user4	21.52	27.10	22.18
user5	192.54	143.27	18.21
user6	51.35	123.89	57.48
user7	72.83	232.70	67.83
user8	311.50	116.82	104.16
user9	26.75	61.84	25.53
user10	8.20	16.41	7.79
user11	13.04	19.92	11.37
user12	43.74	83.95	29.94
user13	6.02	10.22	5.77
user14	17.38	197.96	18.42
user15	7.14	10.83	6.93
user16	18.92	409.27	14.87
user17	28.00	38.00	15.31
user18	43.89	64.04	41.87
user19	10.72	11.00	10.29
user20	24.22	22.7	12.33
user21	93.61	259	70.25
user22	64.18	60.95	25.66
user23	43.23	328.80	68.22
平均值	52.13	113.20	30.91

本文模型对绝大多数用户的预测效果都优于现有算法,平均 MAPE 为 30.91%,相对于现有最好的 RF 方法 MAPE 降低了 20%,说明本文模型可有效提高短期负荷的预测准确度。其中 user5 的准确度相较于其他模型提高得最多,因为 user5 的数据波动很大,通过归一化可以增加数据的稳定性,提高预测的准确度。由表 2 可见,不同用户的 MAPE 差距很大,这是因为不同用户的用电行为习惯差异很大。

图 5 为 user1 和 user8 在 2015 年 1 月 1 日至 2015 年 2 月 28 日的负荷曲线。图 5 中, user1 历史负荷曲线较平稳,其用电模式较规律,因此可得到较准确的预测结果,而 user8 在同段时间的用电行为模式难以捕捉,无

法达到较好的预测效果. 这说明了针对不同用户采用不同预测模型的有效性和重要性.

(2) 为了验证聚类方法的有效性, 将本文模型与仅输入相邻时刻相关的 CNN 模型、输入相邻时刻序列和日相关特征序列的 CNN 模型进行比较, 结果如表 3 所示.

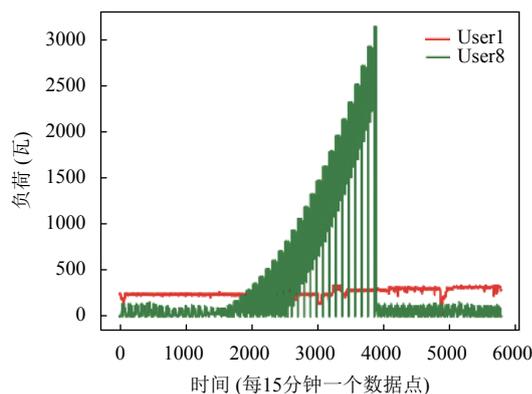


图5 user1 和 user8 2015 年 1 月 1 日至 2015 年 2 月 28 日负荷曲线

表 3 聚类效果 *MAPE* 分析 (%)

用户名称	CNN 邻近相关	CNN 邻近相关+日相关	本文模型
user1	3.09	2.72	2.72
user2	82.45	67.40	67.40
user3	6.74	6.47	6.47
user4	21.54	22.18	22.18
user5	19.03	18.21	18.21
user6	52.91	57.48	57.48
user7	67.83	62.93	67.83
user8	239.28	104.16	104.16
user9	25.53	27.44	25.53
user10	7.83	7.79	7.79
user11	13.63	11.37	11.37
user12	43.84	29.94	29.94
user13	5.77	5.79	5.77
user14	18.42	18.1	18.42
user15	6.93	7.21	6.93
user16	12.41	14.87	14.87
user17	15.31	17.73	15.31
user18	41.87	42.4	41.87
user19	10.69	10.29	10.29
user20	12.33	26.68	12.33
user21	70.25	83.73	70.25
user22	25.66	77.56	25.66
user23	68.22	98.48	68.22
平均值	37.89	35.69	30.91

从表 3 不难看出, 本文模型的 *MAPE* 比简单将数据放入 CNN 模型的 *MAPE* 降低了 5%, 证明了应用聚

类算法可更好区别不同用户的特性, 有助于提高预测性能.

5 结论

针对用户负荷差异明显、波动性大的问题, 本文利用 K-means 算法将用户用电历史负荷曲线进行聚类, 针对不同类别的用户利用 CNN 构造不同的特征提取网络结构, 训练建立预测模型. 实验结果表明本文模型均比 RF、SVR 及未经聚类的 CNN 模型的预测效果优秀, 预测 *MAPE* 降低了 21.22%. 由于当前预测模型仅参考了用户的历史负荷作为模型输入, 实际上用户负荷与温度、湿度及用户所在行业经济形势等多种因素相关, 后续将考虑筛选更多相关因素输入模型中, 进一步提高模型的预测准确度.

参考文献

- 王继业, 季知祥, 史梦洁, 等. 智能配用电大数据需求分析与应用研究. 中国电机工程学报, 2015, 35(8): 1829-1836. [doi: 10.13334/j.0258-8013.pcsee.2015.08.001]
- 马天男, 王超, 彭丽霖, 等. 多源异构大数据下综合能源系统用户用能行为预测分析研究. 智慧电力, 2018, 46(10): 86-95. [doi: 10.3969/j.issn.1673-7598.2018.10.014]
- Beiraghi M, Ranjbar AM. Discrete Fourier transform based approach to forecast monthly peak load. Proceedings of 2011 Asia-Pacific Power and Energy Engineering Conference. Wuhan, China. 2011. 1-5. [doi: 10.1109/APPEEC.2011.5748585]
- 史文博, 王健, 沈添福, 等. 基于改进倍比平滑法的节假日短期负荷预测方法. 电力建设, 2014, 35(10): 21-25. [doi: 10.3969/j.issn.1000-7229.2014.10.005]
- Hor CL, Watson SJ, Majithia S. Daily load forecasting and maximum demand estimation using ARIMA and GARCH. Proceedings of 2006 International Conference on Probabilistic Methods Applied to Power Systems. Stockholm, Sweden. 2006. 1-6. [doi: 10.1109/PMAPS.2006.360237]
- 邹云峰, 梅飞, 李悦, 等. 基于数据挖掘技术的台区合理线损预测模型研究. 电力需求侧管理, 2015, 17(4): 25-29. [doi: 10.3969/j.issn.1009-1831.2015.04.006]
- 黄晗, 孙堃, 刘达. 基于随机森林的电力系统小时负荷预测研究. 智慧电力, 2018, 46(5): 8-14. [doi: 10.3969/j.issn.1673-7598.2018.05.002]
- Li S, Goel L, Wang P. An ensemble approach for short-term

- load forecasting by extreme learning machine. *Applied Energy*, 2016, 170: 22–29. [doi: [10.1016/j.apenergy.2016.02.114](https://doi.org/10.1016/j.apenergy.2016.02.114)]
- 9 Muzaffar S, Afshari A. Short-term load forecasts using LSTM networks. *Energy Procedia*, 2019, 158: 2922–2927. [doi: [10.1016/j.egypro.2019.01.952](https://doi.org/10.1016/j.egypro.2019.01.952)]
- 10 Zheng J, Xu CC, Zhang Z, *et al.* Electric load forecasting in smart grids using Long-Short-Term-Memory based recurrent neural network. *Proceedings of the 2017 51st Annual Conference on Information Sciences and Systems*. Baltimore, MD, USA. 2017. 1–6. [doi: [10.1109/CISS.2017.7926112](https://doi.org/10.1109/CISS.2017.7926112)]
- 11 Aurangzeb K. Short term power load forecasting using machine learning models for energy management in a smart community. *Proceedings of 2019 International Conference on Computer and Information Sciences*. Sakaka, Saudi Arabia. 2019. 1–6. [doi: [10.1109/ICCISci.2019.8716475](https://doi.org/10.1109/ICCISci.2019.8716475)]
- 12 王德文, 孙志伟. 电力用户侧大数据分析并行负荷预测. *中国电机工程学报*, 2015, 35(3): 527–537. [doi: [10.13334/j.0258-8013.pcsee.2015.03.004](https://doi.org/10.13334/j.0258-8013.pcsee.2015.03.004)]
- 13 汤庆峰, 刘念, 张建华, 等. 基于 EMD-KELM-EKF 与参数优选的用户侧微电网短期负荷预测方法. *电网技术*, 2014, 38(10): 2691–2699. [doi: [10.13335/j.1000-3673.pst.2014.10.012](https://doi.org/10.13335/j.1000-3673.pst.2014.10.012)]
- 14 罗敏, 程将南, 王毅, 等. 基于小波去噪和决策树的个性化大用户负荷预测. *南方电网技术*, 2016, 10(10): 37–42. [doi: [10.13648/j.cnki.issn1674-0629.2016.10.006](https://doi.org/10.13648/j.cnki.issn1674-0629.2016.10.006)]
- 15 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. *计算机学报*, 2017, 40(6): 1229–1251. [doi: [10.11897/SP.J.1016.2017.01229](https://doi.org/10.11897/SP.J.1016.2017.01229)]
- 16 Huang CJ, Kuo PH. A deep CNN-LSTM model for particulate matter (PM_{2.5}) forecasting in smart cities. *Sensors*, 2018, 18(7): 2220. [doi: [10.3390/s18072220](https://doi.org/10.3390/s18072220)]