

基于 XGBOOST-DNN 的中期电力负荷预测^①



杨 洋^{1,2}, 谷震浩²

¹(中国科学院大学, 北京 100049)

²(中国科学院 沈阳计算技术研究所, 沈阳 110168)

通讯作者: 杨 洋, E-mail: yangyang184@mails.ucas.ac.cn

摘 要: 精准的负荷预测是电力工作者重要的工作之一, 而负荷预测以预测周期的不同, 一般可以划分为短期电力负荷预测与中长期电力负荷预测. 其中中长期电力负荷预测相较于短期电力负荷预测而言, 该领域缺乏大量前沿工作者的探索. 因此本文提出一种可应用于中期电力负荷预测领域且基于 XGBoost-DNN 的算法. 该算法将树模型和深度神经网络相结合, 并将短期电力负荷预测引入到了中期电力负荷预测的工作中, 基于树模型自身特点, 将数据特征加工成高阶的交叉特征, 同时结合原有数据利用深度神经网络可学习到丰富的特征信息. 这里是以 2017 全球能源预测竞赛的数据进行算法分析, 其中实验表明, 在中期电力负荷预测领域, 该方法提出的 XGBoost-DNN 模型相较于 DNN, LSTM 而言, 其具备更加精准的准确性.

关键词: 负荷预测; 中期负荷预测; 多目标回归; 深度神经网络; XGBoost

引用格式: 杨洋, 谷震浩. 基于 XGBOOST-DNN 的中期电力负荷预测. 计算机系统应用, 2021, 30(9): 186-191. <http://www.c-s-a.org.cn/1003-3254/8098.html>

Mid-Term Power Load Forecasting Based on XGBoost-DNN

YANG Yang^{1,2}, GU Zhen-Hao²

¹(University of Chinese Academy of Sciences, Beijing 100049, China)

²(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

Abstract: Accurate load forecasting is one of the important tasks for power workers, and power load forecasting can be generally divided into short-term forecasting and medium- and long-term forecasting depending on the forecasting period. Compared with short-term power load forecasting, medium- and long-term forecasting is little explored by cutting-edge workers. Therefore, this study proposes an XGBoost-DNN-based algorithm that can be applied to mid-term power load forecasting. The algorithm combines the tree model with the deep neural network and introduces short-term forecasting into mid-term forecasting. According to the characteristics of the tree model, the data features are processed into high-order cross features, and in combination with the original data, the deep neural network is used to learn rich feature information. Algorithm analysis with the data of the 2017 Global Energy Forecasting Competition shows that in mid-term power load forecasting, the XGBoost-DNN model proposed by this method is more accurate than DNN and LSTM.

Key words: load forecast; mid-term load forecast; multi-objective regression; deep neural network; XGBoost

电力负荷预测在电力系统调度、维护和容量规划等方面起着至关重要的作用, 对指导电网规划建设 and 改善电力系统资源的优化配置具有重要意义. 实际的

电力负荷预测分为短期、中期和长期预测. 短期负荷预测 (Short Term Load Forecasting, STLF) 与未来几小时到几天的负荷预测相关, 中期负荷预测 (Medium

^① 收稿时间: 2020-12-15; 修改时间: 2021-01-11, 2021-01-27; 采用时间: 2021-02-02; csa 在线出版时间: 2021-09-02

Term Load Forecasting, MTLF) 则处理以几周到几个月为目标的预测, 长期负荷预测 (Long Term Load Forecasting, LTLF) 则是处理一年到几年的负荷预测. LTLF 可以协助规划新的电力系统; MTLF 有助于系统维护、购买能源和定价计划; 而 STLf 在配电和负荷调度中起着关键作用. MTLF 是一项及具有挑战性的任务, 因为它需要在较长的时间内完成及长期指导和预测的功能.

在前沿工作者的研究当中, 短期电力负荷预测的算法种类较多, 丰富多样. 如 Chen 等^[1] 在 2018 年提出了一种引入深度残差网络结构的短期电力负荷模型; 孔祥玉等^[2] 在 2019 年提出了一种基于模态分解与特征相关分析的短期负荷预测方法; Del Carmen Ruiz-Abellón 等^[3] 在 2018 年验证了集成学习方法随机森林在短期电力负荷的有效性. 肖勇等^[4] 在 2020 年提出了基于多尺度跳跃深度长短期记忆网络的短期负荷预测方法. 在中长期的电力负荷预测的研究工作中, 傅靖等^[5] 在 2020 年提出了基于基因表达式编程的中长期电力负荷预测算法, 并通过仿真实验验证了其算法的精确性; 王军^[6] 在 2018 年提出了基于深度神经网络的中期电力负荷预测, 在中期的负荷数据分析处理和神经网络模块中给出了具体的处理方案; 陈家慧^[7] 在 2019 年提出了灰色理论与深度神经网络在电力系统中长期负荷预测中的应用方案.

在本文中, 我们提出了基于 XGBoost 和 DNN 模型相结合的 XGBoost-DNN 算法, 最大程度地利用除负荷时序特征以外的特征信息, 并引入短期电力负荷的预测形成的交叉特征到最终的中期电力负荷模型中.

1 电力负荷预测的影响因素

电力负荷预测的数据主要是以时序形式存储, 其影响因素非常复杂, 因此需要观察不同时间段的天气、温度、湿度、节假日、经济状况来对数据进行不同角度的分析. 这里我们主要采用的数据源于全球能源大赛 GEF-Com2017, 其中的更多数据细节可以参考文献 [8] 进行查阅. 该数据提供了横跨美国新英格兰的 8 个区域的小时级电力负荷数据, 还提供了包括干球温度和露点温度在内的每小时天气数据. 当前数据的分析主要取 2003 年 3 月到 2017 年 12 月时期所有区域总和的负荷数据, 温度是 8 个地区气象站的加权值, 根据文献 [8] 所解释, 每个负荷区对应的气象站是

根据其于负荷区地理位置的接近程度来表示的“最佳选择”, 并不代表负荷区的“实际天气”, 而仅代表所列气象站, 其权重数据可具体参见 GEF-Com2017 数据集.

受于数据所限, 本文的负荷数据影响因素的分析将在干球温度和露点温度的数据上进行. 如图 1 和图 2 所示, 我们可以获取到从 2003 年到 2017 年在以小时为粒度, 温度和真实需求负荷之间的散点图, 我们从中进行了非线性函数的拟合, 最终确定温度和负荷之间的非线性关系.

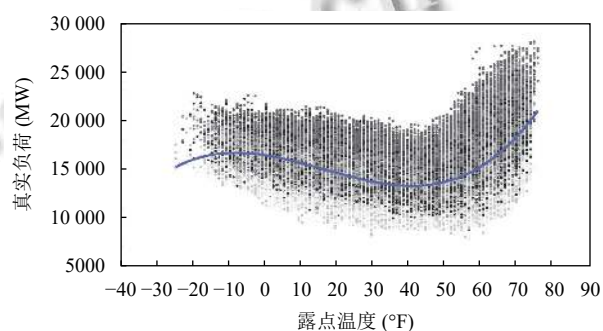


图 1 露点温度与负荷数据的散点图

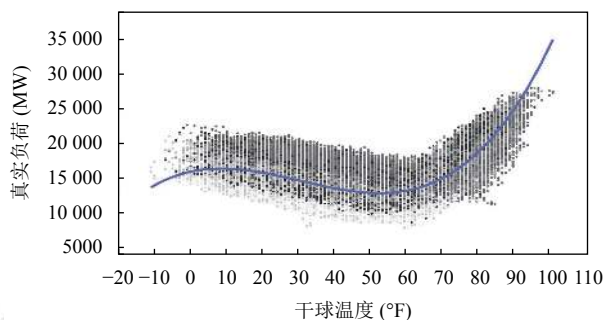


图 2 干球温度与负荷数据的散点图

数据在数据分析过程中, 我们不得不去验证数据温度的真实性, 即其温度数据是否符合我们的基础认知 (呈现线性关系). 如图 3 所示, 我们将所有的干球和露点数据进行了散点图的可视化, 并进行了线性函数的拟合. 可以发现, 干球温度和露点温度小时粒度的区分度非常明显, 并且线性拟合的函数满足于我们的基础认知的数据特性.

除以上干球和露点温度的影响因素外, 其负荷数据本身便存在一定的周期性变化, 如图 4 所示, 可以看出数据在 2003 年到 2017 年之间, 每一年的负荷数据都呈现了 3 个峰头和 3 个峰谷的周期性变化. 因此, 与周期性变化具有强关联性的时间特征是需要重点挖掘的数据特征, 比如季节、月份、上下旬、工作日、周末等.

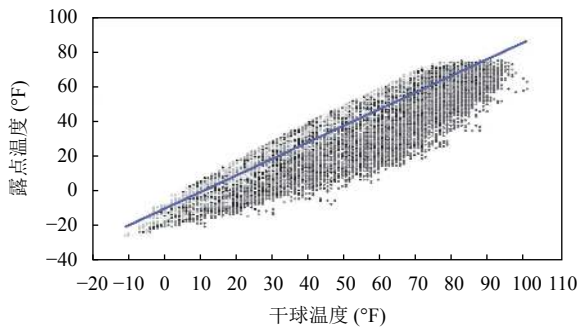


图3 露点温度与干球温度数据的散点图

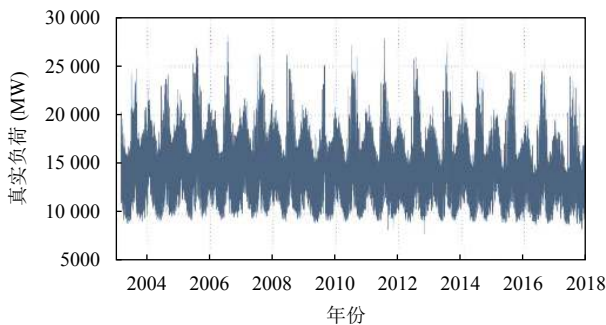


图4 负荷数据的周期性变化图

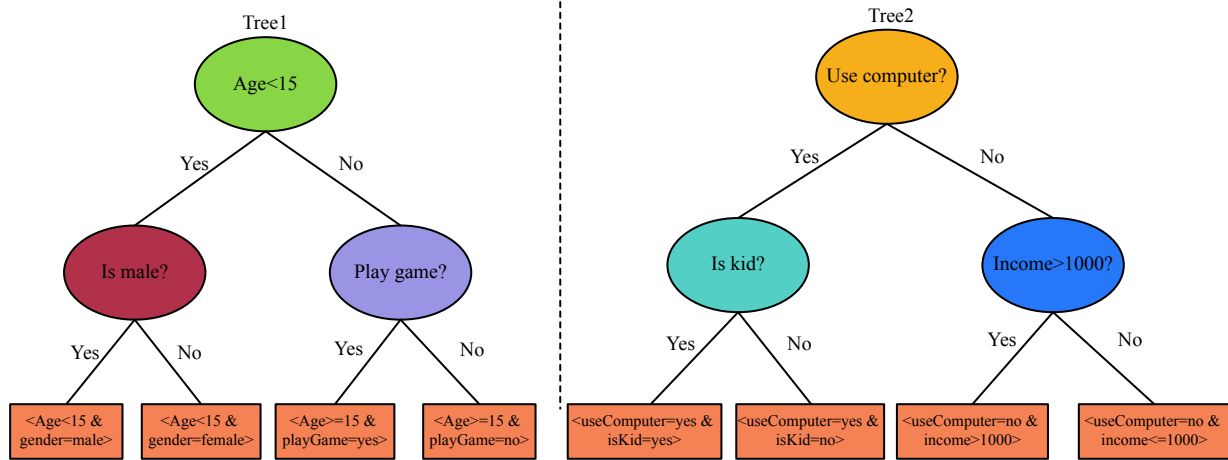


图5 树模型的特征交叉

在树模型的演化历史中,集成学习中的树模型又被分为了以随机森林 (random forest) 模型为代表的 bagging 思想和梯度提升树 (Gradient Boosting Decision Tree, GBDT) 为代表的 Boosting 思想,两者都是使用弱学习器的集成学习方式对需求任务进行学习,其中主要的不同是,前者是分为多个子模型对放回抽样的数据进行多次决策,最终预测结果依赖于全部子模型的投票结果,后者是将多个子模型进行堆叠,后一个子模

2 XGBoost-DNN 模型算法

本文所提出的 XGBoost-DNN 模型算法主要是由 XGBoost 和 DNN 两个模块,前者作为特征处理器,利用短期负荷预测对原始数据进行用来学习特征到目标的映射关系,后者将学习的映射关系 (叶子节点) 作为特征接入到 DNN 模型中去学习中期电力负荷的时序目标,由此完成中期电力负荷的预测任务。

2.1 XGBoost

树模型具备一般的 if-then 结构,简单的决策树模型会将输入数据集的特征空间进行划分,每个数据样本最终会被分在特征空间的子空间中,也就是每个数据样本会落在决策树上的某一个叶子节点上,而该叶子节点得益于 if-then 结构的原因就具备了一个天然的特征交叉效果,如图 5 所示,在 Tree1 中,如果样本 A 落在第一个叶子节点上,那么可以获得 <age < 15 & gender = male> 的二阶交叉特征.因此,通过树模型我们可以很自然地获取到高阶的交叉特征,基于这类特性,我们选择树模型可以将时序特征和类别特征加工成丰富的交叉特征。

型在前一个子模型的基础上学习其剩余的残差,即一个样本不再只落在一棵树的叶子节点上.如图 5 所示,当一个样本 A 通过 Boosting 的树模型,并且落在 Tree1 的第一个叶子节点和 Tree2 的第一个叶子节点,那么我们可以得到 <Age < 15 & gender = male> 和 <useComputer = yes & isKid = yes> 作为样本 A 的高阶特征表达。

XGBoost (eXtreme Gradient Boosting) 算法是由华盛顿大学陈天奇博士基于 GBDT 和随机森林提出的高

效和广泛使用的集成学习方法,既可以应用于分类任务中,也可以应用在回归问题中。相比 GBDT 和随机森林, XGBoost 不仅具备更加良好的精确性,还添加了正则化项增强了模型的泛化效果,同时支持并行化特征选择使其具备了更加高效的学习方式去提升模型的训练速度。损失函数的计算公式如下:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) + \Omega(f_i)$$

其中, $\hat{y}_i^{(t)}$ 是第 t 次迭代中第 i 个样本的预测结果, $f_i(x_i)$ 是第当前迭代中第 i 个样本的预测函数, $\Omega(f_i)$ 是正则项。

基于此,树模型本身的交叉特性和 XGBoost 的高精确性是我们选择它的主要因素。借助 XGBoost 模型对中长期电力负荷预测任务进行预训练,将中长期电力负荷预测的特征数据接入到 XGBoost 模型中,预测 Label 由多目标转为单目标,即取最近的时序负荷作为目标 Label,预测任务是短期负荷预测任务。该操作主要考虑到两个因素,其一是树模型本身不适用于做多目标回归任务;其二是 XGBoost 在 XGBoost-DNN 的模型结构里,承担的主要任务是提供高阶的交叉特征,而选择最近时序的负荷数据作为目标 Label 有助于 XGBoost 模型在学习过程中能更好地捕捉到以短期时序为目标的特征结构信息,让 DNN 模型学习到更丰富的特征空间。

2.2 DNN

深度神经网络 (Deep Neural Networks, DNN) 是基于误差反向传播算法的前向神经网络,具有较强的非线性拟合和学习能力,一般包含 3 层网络:输入层、隐含层和输出层。每一层都是由许多神经元组成,相同层的神经元之间没有任何连接,相邻层则是完全互相连接,其中隐含层大多情况下不止一层,而其中每个隐层的每个神经元都是非线性的激活函数,如 ReLU、Sigmoid 等。一般, DNN 在工业界中是由机器学习转到深度学习演进过程中必不可少的模型,其强大的非线性拟合和学习能力是得到业界共识的。在中期电力负荷的领域里, DNN 在多目标的回归任务中是学习时序关系必不可少的重要模型。它可以很好地捕捉到负荷数据到目标的非线性映射关系,这是 XGBoost 等树模型所不能替代的。

2.3 XGBoost-DNN 算法

XGBoost-DNN 模型是由 XGBoost 和 DNN 所构

成的组合模型,其模型结构如图 6 所示。在中期电力负荷预测的场景下一般是使用单个模块来实现特征信息的提取,如我们常使用 LSTM 提取时序特征信息,使用 DNN 用来学习特征到目标的非线性映射关系。目前在中长期电力负荷领域下,使用 DNN 的单模块模型在学习过程中更偏向于学习特征映射到多目标的时序关系,对特征本身的特征捕捉能力较少。这里 XGBoost 作为子模块之一,引入短期电力负荷的 Label 作为单目标回归任务的学习,相对中期电力负荷预测的多目标回归任务,单目标模型因为精度和损失函数问题会更多关注数据特征的深度挖掘,不会因为多目标的 Label 之间存在时序关系而导致模型更关注时序关系而忽略了特征本身。因此利用 XGBoost 学习到的从数据到目标的交叉特征,就是一份经过深度挖掘后的丰富特征信息,利用这份预训练学习到的交叉特征结合原数据通过 DNN 的特性去学习中期电力负荷的时序关系,可在组合模型中降低 DNN 单模型的学习阈值,并实现将中期电力负荷预测模型无法关注到的特征信息得以关注,以此来更适用于该场景下的预测任务。其算法流程如下详述:

1) XGBoost 获取交叉特征:输入层会将数据分为 5 折输入到 XGBoost 模型中进行一次短期负荷预测,然后通过交叉预测的方式对每个样本进行预测,进而获取到每一个样本在 XGBoost 中的所有叶子节点。例如,假设模型共 3 棵子树,并且 Tree1、Tree2、Tree3 分别对应叶子节点数目为 3, 2, 3, 样本数据分别落在第 2, 1, 3 的叶子节点上,其样本数据获得的交叉特征当为 [0, 1, 0, 1, 0, 0, 0, 1] 的特征向量。

2) 特征工程:这里主要采用的是针对时间和负荷数据做了大量的特征处理,比如,在时间数据上,年、月、日、季度、工作日、节假日、小时、月份上下旬是我们所处理后的时间特征;在负荷数据上,我们生产了多阶差分数据(上一个时间点的负荷与当前时间点的负荷差值)作为偏移值,这里只做了过去 72 小时的差分特征。同时,根据时间,为当前时间点产出过去 30 天的小时级负荷数据,同时引入前一年同期的前后 15 天小时级负荷数据。最后,将类别特征做独热编码的处理,并将全部产出特征和 XGBoost 所得到的交叉特征进行拼接,即可得到全部转换后的特征数据。

3) DNN 训练: DNN 模块这里主要采样了 3 层的结构, 激活函数采样 ReLU, 每层神经元数目根据塔式

变化进行调整, 并需保证最后一层输出层是 720 个神经元, 以此预测未来 30 天时间里的负荷数据.

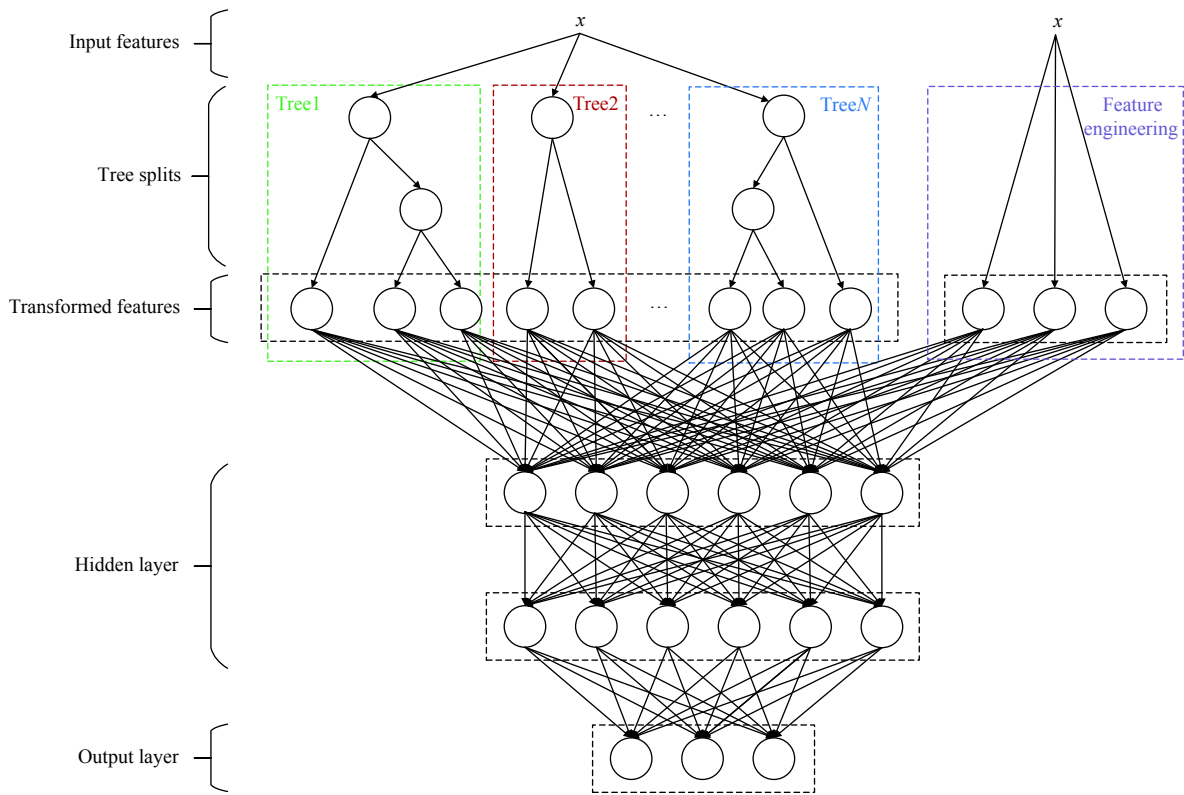


图 6 XGBoost-DNN 模型结构

3 实验分析

如前面所说, 实验所采用数据源自于全球能源大赛 GEF-Com2017, 主要是从 2013 年 3 月到 2017 年的 12 月的负荷和温度数据.

3.1 实验环境设置

本文中的实验环境是 Docker 容器, 镜像是 Tensor-Flow-GPU-1.4.0 版本, 对应 Keras-2.25 版本, 所用 Python 版本是 3.6.8. 除此之外, 其他配置为 6 核 CPU, 24 GB 内存, 一块 Tesla P40 GPU, 显存 23 GB.

3.2 实验评估指标

为评估预测结果的性能, 设置均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE) 来去评估实验预测结果的性能.

$$y_{RMSE} = \sqrt{\frac{1}{n * m} \sum_{i=1}^n \sum_{j=1}^m (X_{true}(ij) - X_{pred}(ij))^2}$$

$$y_{MAPE} = \frac{1}{n * m} \sum_{i=1}^n \sum_{j=1}^m \left| \frac{X_{true}(ij) - X_{pred}(ij)}{X_{true}(ij)} \right|$$

其中, n 是样本数目, m 是预测未来时间点的个数, $x_{true}(ij)$ 是第 i 个样本在第 j 个时间点的真实负荷数据, $x_{pred}(ij)$ 是第 i 个样本在第 j 个时间点的预测负荷数据.

3.3 实验结果分析

本次实验采用了 Grid Search 对 XGBoost 的参数进行了调试, 并通过控制变量法逐步对 DNN 模型的参数进行调优, 在固定网络层数的情况下, 测试不同的超参数对模型整体的影响. 如图 7 所示, 可以发现 XGBoost-DNN 的模型算法预测的某个样本未来 30 天里的负荷曲线图, 较之其他模型算法有较好的表现, 更接近于真实的预测结果.

各个算法的评测结果见表 1. 从表 1 中可知, XGBoost-DNN 的算法在所有的实验算法中拥有较好的指标提升, 可见这种架构模型的性能较好.

4 结论与展望

本文将 XGBoost 引入到 DNN 进行中长期电力负荷预测之中, 并将短期电力负荷和长期电力负荷相结合. 在

实验过程中,采用了全球能源大赛 GEF-Com2017 的数据,分别对每一个时间点进行特征工程和模型训练的工

作,并将 XGBoost-DNN 算法和其他模型进行了对比,验证了该模型架构的准确性,也证实了多特征交叉的有效性。

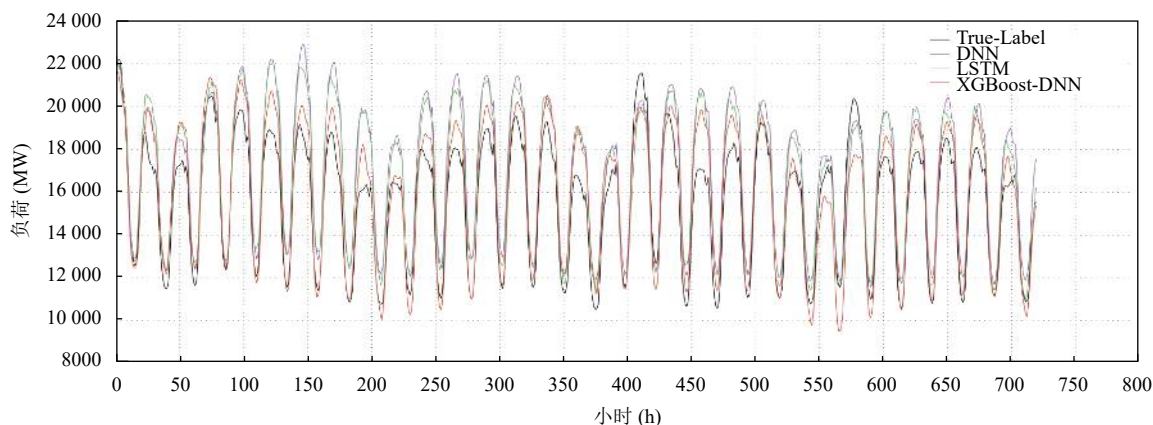


图7 各类算法在某一样本上对未来30天的预测结果

表1 各类算法实验结果

| 模型算法 | RMSE | MAPE |
|-------------|------------------|---------------|
| DNN | 1252.8518 | 0.0593 |
| LSTM | 1280.3640 | 0.0616 |
| XGBoost-DNN | 1119.8158 | 0.0542 |

参考文献

- 1 Chen KJ, Chen KL, Wang Q, *et al.* Short-term load forecasting with deep residual networks. *IEEE Transactions on Smart Grid*, 2019, 10(4): 3943–3952. [doi: [10.1109/TSG.2018.2844307](https://doi.org/10.1109/TSG.2018.2844307)]
- 2 孔祥玉, 李闯, 郑锋, 等. 基于经验模态分解与特征相关分析的短期负荷预测方法. *电力系统自动化*, 2019, 43(5): 46–52.
- 3 Del Carmen Ruiz-Abellón M, Gabaldón A, Guillamón A. Load forecasting for a campus university using ensemble

methods based on regression trees. *Energies*, 2018, 11(8): 2038. [doi: [10.3390/en11082038](https://doi.org/10.3390/en11082038)]

- 4 肖勇, 郑楷洪, 郑镇境, 等. 基于多尺度跳跃深度长短期记忆网络的短期多变量负荷预测. *计算机应用*, 2021, 41(1): 231–236.
- 5 傅靖, 王栋, 白阳, 等. 基于基因表达式编程的中长期电力负荷预测挖掘. *电力信息与通信技术*, 2020, 18(5): 7–12.
- 6 王军. 基于深度神经网络的中期电力负荷预测. *重庆工商大学学报(自然科学版)*, 2018, 35(6): 17–21.
- 7 陈家慧. 灰色理论与神经网络在电力系统中长期负荷预测中的应用 [硕士学位论文]. 苏州: 苏州大学, 2019.
- 8 Hong T, Xie JR, Black J. Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. *International Journal of Forecasting*, 2019, 35(4): 1389–1399. [doi: [10.1016/j.ijforecast.2019.02.006](https://doi.org/10.1016/j.ijforecast.2019.02.006)]