

基于预训练 GRU-LightGBM 的电力负荷预测^①



张 晓¹, 丁云峰², 王 刚³

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 系统与软件事业部, 沈阳 110168)

³(国网辽宁省电力有限公司 调度控制中心, 沈阳 110004)

通讯作者: 张 晓, E-mail: 1033610689@qq.com

摘 要: 本文研究工作围绕电网中的核心环节电负荷预测展开. 在总结分析之前研究学者的研究成果基础上, 提出了基于预训练 GRU 与 LightGBM 相结合的方法. 该方法首先使用电负荷数据训练一个特征提取网络 GRU, 然后使用该网络进行时序特征的提取, 并将提取到的时序特征与非时序特征使用 LightGBM 进行电负荷的预测. 本方法的创新点在于提出了预训练网络来扩充特征, 充分融合时序特征及非时序特征. 并且考虑到电网的地区差异性, 在整体训练过程中将 GRU 网络参数进行了适应性微调. 保证提取到的时序特征是符合当前地域特点的. 通过仿真实验最终发现该方法在各项指标上取得了 2% 的提升.

关键词: GRU 网络; 电负荷; 时序特征; 预训练网络

引用格式: 张晓, 丁云峰, 王刚. 基于预训练 GRU-LightGBM 的电力负荷预测. 计算机系统应用, 2021, 30(8):288-292. <http://www.c-s-a.org.cn/1003-3254/8053.html>

Electric Load Forecasting Based on Pre-Training GRU-LightGBM

ZHANG Xiao¹, DING Yun-Feng², WANG Gang³

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

²(System and Software Division, Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

³(Dispatch Control Center, State Grid Liaoning Electric Power Co. Ltd., Shenyang 110004, China)

Abstract: This study focuses on the electric load forecasting of the core link in the power grid. On the basis of summarizing and analyzing the research results of previous researchers, a method based on the combination of pre-training GRU and LightGBM is proposed. This method first uses electrical load data to train a feature extraction network GRU, then uses the network to extract timing features, and uses LightGBM to predict the electrical load of the extracted timing features and non-sequential features. The innovation of this method is to propose a pre-training network to expand the features and fully integrate the timing features and non-timing features. And taking into account the regional differences of the power grid, the GRU network parameters were adaptively fine-tuned during the overall training process. Ensure that the extracted time series features are consistent with the current regional characteristics. Finally, it is found through simulation experiments that this method has achieved a 2% improvement in various indicators.

Key words: GRU-network; electric load; timing characteristics; pre-train network

社会经济的发展离不开电网的可靠性保障. 电负荷是电网运行状态的重要表征, 因此电力负荷的预测是配

电网规划的关键部分. 在电网运行状态的评估以及电网故障的预警上都有重要意义, 减少电网故障的发生对提

① 收稿时间: 2020-11-20; 修改时间: 2020-12-21; 采用时间: 2021-01-08; csa 在线出版时间: 2021-07-31

高电力系统的社会效益和经济效益有重大意义. 随着电网智能化的推进, 电网中数据的智能化、信息化程度不断提高, 这为电力系统负荷预测提供了海量数据样本, 也为达成提高负荷预测准确性的目标奠定了基础.

短期电力负荷是对某一区域未来几个小时的用电负荷进行预测. 目前, 在海内外研究人员的探索、尝试下, 取得了很大的进展. 当下, 短期电力负荷预测的方法主要有传统的统计学方法和基于机器学习的预测方法^[1-3]. 其中统计方法包括时序预测法^[4]、多元线性回归 (MLR)^[5,6]等. 统计学方法取得了较好的成果, 但是这些方法需要对电网领域的专业知识有比较深的理解才可以有效的建模并给出结果, 使用起来学习成本较高. 而近年来日渐崛起的人工智能方法降低了应用门槛, 并且取得了更好的效果. 其中陈振宇等^[7]提出结合 LSTM 及 XGBOOST 方法的电负荷预测, 该方法的创新点在于两个模型的预测结果融合方法. 但是在预测中两个模型单独起作用, 实践表明单独的模型无法取得了比较好的结果. 马煜等^[8]提出使用 CNN 及 SVR 结合的方法, 其中 CNN 网络主要用于特征提取. 但是该提取方法, 打乱了本身的时序特征. 综上分析, 人工智能方法崛起后, 大家的主要研究方向集中在特征提取方法和是否单一模型工作上. 本文提出的方法也是在充分分析了前辈的工作成果以及电网负荷的变化特点后提出的.

电负荷影响因素有很多, 诸如历史负荷数据, 气象数据、电力设备使用时间以及峰谷电价等因素. 文献^[9]通过整理发现, 这些因素主要可以分为时序因素和非

时序因素, 因此本文提出了一种同时对时序特征和非时序特征建模的电负荷预测方法——基于预训练 GRU-LightGBM 的电负荷预测方法. 双向循环 GRU 网络作为时序特征提取器, 同时考虑历史时期和未来时期负荷影响因素对预测负荷的影响, 充分提取负荷特征数据之间有效潜在关系. 将提取到的时序特征与影响电力负荷的非时序特征 (例如区域电价) 相结合, 放入到对回归预测问题有着不错效果且运算速度快的 LightGBM 模型中, 进行模型训练与预测.

算例测试结果表明, 本文的组合模型由于针对时序特征与非时序特征进行了归纳整理、综合预测, 能够降低单一模型在极端场景下的误差, 相比于其他组合模型, 也具有更加稳定的预测效果.

1 算法总体结构设计

前文提出本文综合考虑电力负荷的实际特点我们提出了 GRU 与 LightGBM 相结合的方法. 算法的总体架构图如图 1 所示.

如图 1 所示, 整个算法结构分为 3 个阶段. 图中阶段 1 表示时序特征提取, 通过双向 GRU 网络提取电网时序特征中蕴含的时序依赖语义信息, 该阶段分别输出语义的前向依赖和反向依赖信息, 并将两个方向的信息进行组合. 本文提出了两种组合方案拼接和取平均值. 阶段 2 表示特征融合, 主要包括两部分特征, 一部分是阶段 1 双向循环 GRU 作为特征提取器提取的时序信息, 另一部分是非时序特征; 阶段 3 表示电负荷预测阶段. 下面对 3 个阶段进行详细介绍.

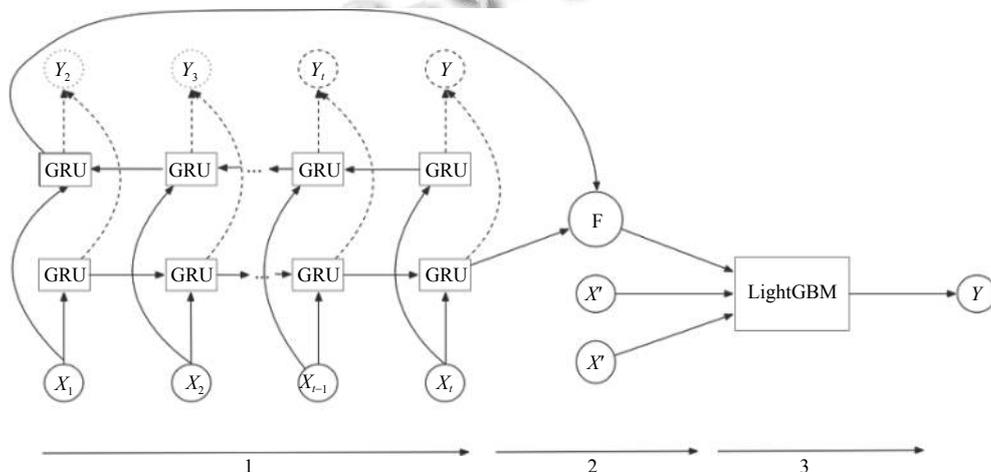


图 1 预训练 GRU-LightGBM 的电负荷预测方法整体架构图

1.1 GRU 网络设计

GRU (Gate Recurrent Unit) 是循环神经网络 (Recurrent Neural Network, RNN) 的一种. 与 LSTM (Long-Short Term Memory) 近似, 是为了缓解长时序下的梯度消失或者梯度爆炸而设计的一种方案. 在众多研究中 LSTM 表现要略胜 GRU, 但是 LSTM 参数多, 训练起来耗时更长. 并且本文 GRU 网络用于提取序列特征, 对实际预测效果依赖性弱. 相比之下, 时效性对整个系统更加重要. 因此我们选用了消耗时间少并且效果也不差的 GRU 作为主要结构.

图 2 是 GRU 结构示意图. 每一个 GRU 结构中, 有一个当前时刻的输入 x^t , 及上一个节点传递下来的隐状态 h^{t-1} , h^{t-1} 中蕴含了之前各个时刻的依赖信息. 结合 x^t 和 h^{t-1} , GRU 会得到当前隐藏节点的输出 y^t 和传递给下一个节点的隐状态 h^t .

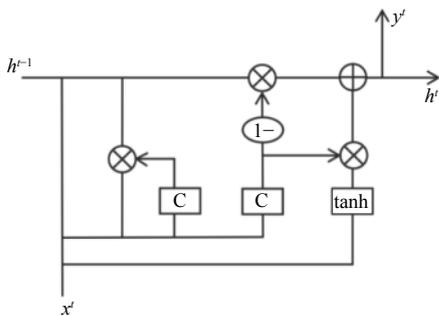


图 2 GRU 结构示意图

每个 GRU 中有重置门 r^t 和更新门 z^t , 如式 (1) 和式 (2) 所示. 两个门的计算逻辑都是拼接当前时刻的输入 x^t , 以及上一时刻的隐状态 h^{t-1} , 并通过 Sigmoid 函数将输出控制在 $[0, 1]$ 之间. 当输出趋向 0 时表示抑制, 当趋向 1 时表示激活, 这两种状态对应表示成门的开闭. 重置门和更新门的构建如下:

$$r^t = \sigma(W_r \cdot [h^{t-1}, x^t]) \tag{1}$$

$$z^t = \sigma(W_z \cdot [h^{t-1}, x^t]) \tag{2}$$

得到门控信号后首先使用重置门来重置信息 $r^t \cdot h^{t-1}$, 然后将重置后信息通过一个 tanh 函数将数据缩放到 $[-1, 1]$ 的范围内得到 \tilde{h}^t . 如式 (3), \tilde{h}^t 中包含当前时刻要加入的信息, 相当于记忆了当前时刻的状态.

$$\tilde{h}^t = \tanh(W_h \cdot [r^t \cdot h^{t-1}, x^t]) \tag{3}$$

最后一个阶段输出最终的隐藏信息. 这一步的操

作就是忘记传递下来的某些维度信息, 并加入当前节点输入的某些维度信息. 并根据这些隐藏信息输出当前时刻的输出 y^t , 如下:

$$h^t = (1 - z^t) \cdot h^{t-1} + z^t \cdot \tilde{h}^t \tag{4}$$

$$y^t = \sigma(W_o \cdot h^t) \tag{5}$$

本文的 GRU 网络是时间步长为 6 的双向 GRU 网络, 正向 GRU 网络主要提取正向时间 (相当于历史时期) 步间的依赖信息. 反向 GRU 网络则主要提取反向时间 (相当于未来时期) 步的依赖信息, 这样保证提取到更加全面的特征信息.

1.2 组合特征

阶段 2 主要融合双向 GRU 网络的隐藏单元特征. 其中在 GRU 网络的训练阶段, 我们尝试了拼接和取均值的两种融合方式. 经过训练对比后发现拼接要比取均值效果更好, 具体的训练细节以及原因分析在后面章节进行详细说明. 最后我们给出如式 (6) 所示的方案:

$$F = [h_t^1; h_t^2] \tag{6}$$

其中, $[\]$ 表示向量拼接操作, 举例来说 $h_t^1 = |x|_{r*c}$, $h_t^2 = |x|_{r*e}$, 则 $F = |x|_{r*(c+e)}$.

F 输出了时序特征向量. 将时序特征与离散特征共同作为最终预测模型的特征输入, 此时的时序特征在数据形态上与离散特征相同.

1.3 预测算法

本文核心算法使用 LightGBM, LightGBM 训练效率高、低内存使用、准确率更高、并且支持并行学习等优势, 非常适合处理特征维度大, 数据量大的电力数据.

LightGBM 是 GBDT 的一个变种. 基于 boosting 增强策略的加法模型, 训练的时候采用前向分布算法进行贪婪学习, 每次迭代都学习一颗 CART 树来拟合前 $t-1$ 棵树的预测结果与训练样本真实值的残差. LightGBM 提出了基于梯度的单侧采样 GOSS 算法. GOSS 算法的算法思想是去掉很大一部分梯度很小的数据, 只使用剩下的去估计信息增益. 由于梯度大的数据在计算信息增益的时候更重要, 所以 GOSS 在小很多的数据上仍然可以取得相当准确的估计值. 具体做法是保存梯度绝对值较大的 a 的数据, 然后对剩下的 $(1-a)$ 的数据, 再随机选取 b 的量. 最终使用 $a+(1-a)b$ 的数据量参与下一轮训练. 但是这样会引起分布变化, 所以对随机选择的那部分样本要乘以 $(1-a)/b$.

1.4 模型评价标准

模型选用均方根误差 (Root Mean Square Error, *RMSE*), 均方根误差亦称标准误差, 是观测值与真值偏差的平方和与观测次数 *N* 比值的平方根, 是用来衡量观测值同真值之间的偏差. *RMSE* 的公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2} \quad (7)$$

其中, *N* 为样本数, *y_i* 为第 *i* 条样本的真实值, *p_i* 为第 *i* 条样本的预测值.

2 模型的训练方案

根据第 1 节对模型结构的阐述, 本文模型属于多个子模型的嫁接. 一次训练整个模型比较困难, 加上 RNN 网络本身收敛难度大因此整个网络一起训练比较困难. 所以在模型训练上本文使用 fine-tuning 的思想. 首先训练双向 GRU 网络, 训练后的双向 GRU 网络作为时序特征提取器. 本文算法基于 Tensorflow 框架编写完成, 选用了内置 ADAM 优化算法作为 GRU 网络的优化器. 并在输出层中添加 DROPOUT 层, 比率选择 0.5. 在实际测试中 dropout 层加入在预测准度和运行时间上都有所优化.

特征提取器训练完成后组合 LightGBM 算法, 进行模型的整体训练. 在模型的整体训练阶段, 将特征提取器每个 GRU 单元的输出结构移除然后模型整体参

与训练. 在训练阶段使用反向传播, 更新 LightGBM 的参数及双向 GRU 网络的参数.

2.1 特征提取网络预训练

首先训练双向 GRU 网络, 通过分析电负荷在一天中变化发现<这个地方为贴那张负荷变化图, 然后解释一下>每 4 个小时为一个变化趋势, 所以使用 4 个 GRU 时间步. *t* = 1 时的输入, 预测 *t* = 2 刻的电负荷, 以此类推最后一个时刻 *t* 的输入预测 *t* + 1 时刻的电负荷.

特征提取双向 GRU 网络的训练效果如图 3 所示.

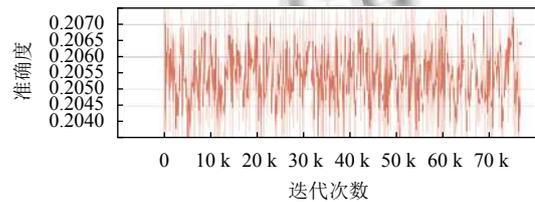


图 3 F 函数使用拼接方案的训练过程图

2.2 模型整体训练

模型整体训练, 首先将双向 GRU 网络的 *y* 输出删掉, 如图 4 所示. 预训练的特征提取网络最后一个状态的只输出对应的隐状态. 之所以这样做是因为双向 GRU 网络只负责时序特征的提取, 并且在整个网络训练阶段参与训练. 如果保留原 GRU 网络的 *y* 输出, 在反向传播时 GRU 网络也会根据自身 *y* 的输出进行参数的更新, 这与从 lightGBM 产生的梯度信息产生重合, 导致更新量过多, 最终导致模型效果不佳.

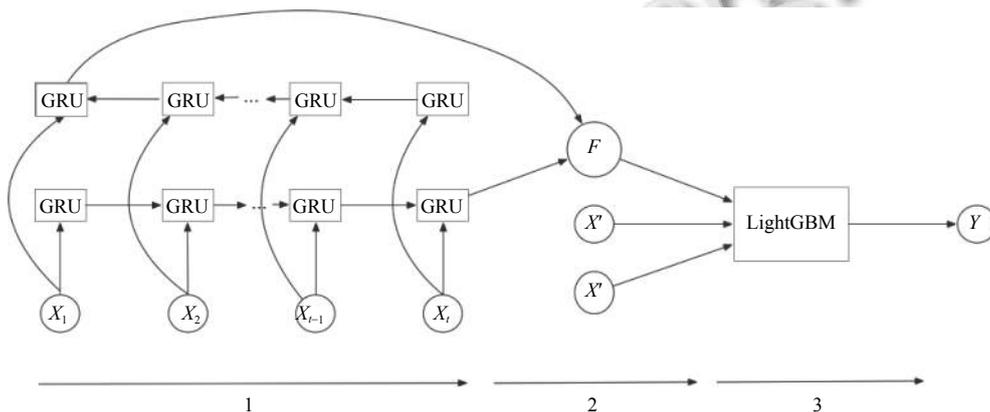


图 4 模型整体训练

模型从前向后一次经过预训练的 GRU 网络, 输出双向的隐状态. 然后拼接隐状态输入 LightGBM 中最终由 LightGBM 输出预测的电负荷值. 然后训练过程中, 直接使用梯度下降法依次更新 LightGBM 和 GRU

的内部参数. 在超参数的调整上只调整 LightGBM 的参数, 不对预训练的模型以及超参数做任何修改. 我们调整的 LightGBM 参数有: (1) 学习率. 学习率对树模型的调参非常重要, 最终选择了比较大的学习率 0.1.

(2) 树的深度和叶子结点数目. 使用网格搜索的方法最终得到 7 和 80 两个值. 其他参数我们使用了默认值.

本文在训练中对 F 函数的两种方案进行了分别测试. 其中使用均值方案训练过程如图 5 所示. 观察图 5 发现在迭代 7 万次后整个模型还没有收敛状态, 并且 $RMSE$ 值较高. 而拼接方案的效果则比较好, 已经收敛并且 $RMSE$ 值较低. F 函数使用拼接方案的训练过程如图 6 所示.

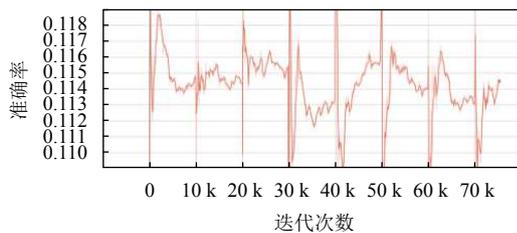


图 5 F 函数使用均值方案的训练过程图

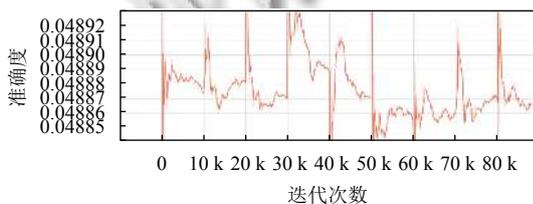


图 6 双向 GRU 网络的训练效果

3 实验仿真

3.1 数据集准备

本文的数据集来自辽宁省某市区的 2013 至 2019 年的电力数据. 数据选取的时序特征主要包括: 上一小时电力负荷、温度、湿度、风力、降雨量等, 非时序特征为: 区域电价、是否为节假日、是否发生重大社会事件、社会事件影响程度等级、电力设备使用年限等.

我们首先从数据集中使用 2013~2014 年一年的数据训练双向 GRU 网络. 然后使用后面的数据训练整个算法模型. 前文提到我们进行模型的整体训练时, 也要更新 GRU 网络的参数的另一个原因是我们训练的 GRU 网络所使用的数据时间比较早. 电负荷环境会随着社会的进步发生变化, 直接使用较早时间训练的预训练网络可能对特征的提取产生偏差. 为了缓解这种偏差考虑对 GRU 网络的参数进行更新.

3.2 不同模型对比

我们主要对比了前文提到的 LSTM、XGBoost 以

及 CVR 与 CNN 结合的方法. 如表 1 所示.

4 结论与展望

通过实验结果表明, 预训练 GRU-LightGBM 模型与其他模型相比在负荷预测的精度上有很大提高. 本文方法即考虑历史负荷数据、温度等时序特征对电力负荷的影响, 也考虑到社会、特殊事件等非时序特征对短期电力负荷预测的影响. 总的来说, 针对时序特征进行预训练再与非时序特征相结合放入预测模型训练的方法是切实有效的.

表 1 不同模型之间不同指标结果对比

方法	$RMSE$	$MAPE$	MAE
LSTM	0.0555	0.0499	5.1284
XGBoost	0.0602	0.0552	8.2356
GRU+CNN	0.0623	0.0521	8.1643
GRU+LightGBM	0.0488	0.0312	3.1451

参考文献

- 胡松峰, 彭显刚. 电网短期负荷预测方法综述. 武汉 (南方九省) 电工理论学会第 22 届学术年会、河南省电工技术学会年会论文集. 焦作, 中国. 2010.
- 夏博, 杨超, 李冲. 电力系统短期负荷预测方法研究综述. 电力大数据, 2018, 21(7): 22-28.
- Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting, 2016, 32(3): 914-938. [doi: 10.1016/j.ijforecast.2015.11.011]
- 艾欣, 周志宇, 魏妍萍, 等. 基于自回归积分滑动平均模型的可转移负荷竞价策略. 电力系统自动化, 2017, 41(20): 26-31, 104. [doi: 10.7500/AEPS20170119009]
- Fang JL, Xing Y, Fu Y, et al. Rural power system load forecast based on principal component analysis. Journal of Northeast Agricultural University, 2015, 22(2): 67-72.
- 王勇, 黄国兴, 彭道刚. 带反馈的多元线性回归法在电力负荷预测中的应用. 计算机应用与软件, 2008, 25(1): 82-84. [doi: 10.3969/j.issn.1000-386X.2008.01.031]
- 陈振宇, 刘金波, 李晨, 等. 基于 LSTM 与 XGBoost 组合模型的超短期电力负荷预测. 电网技术, 2020, 44(2): 614-620.
- 马煜, 黄哲洙, 钟丽波, 等. 基于 CNN 和 SVR 相结合的电力负荷预测分析. 东北电力技术, 2020, 41(2): 37-41. [doi: 10.3969/j.issn.1004-7913.2020.02.012]
- Wang SP, Sun J, Xu ZB. HyperAdam: A learnable task-adaptive Adam for network training. Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence. Honolulu, HI, USA. 2019. 5297-5304.