

提交日期：2025 年 7 月 15 日中午 12 点之前（实际完成需至少 1 个月，视调试经验或有长短）。以链接中**报告pdf** 最终提交时间为准，晚于此时间视为没交。

提交链接（任意交一个即可）：<https://send2me.cn/5oE8thju/R-eDQGDeGdzoIg>
<https://send2me.cn/jbzcZCXI/SZ60PrPqAR6oUQ>

合作方式：**最多可 2 人组队**，请在报告上详细列明各位作者的贡献，并注明各自所属的研究领域。团队内部讨论为主，但这并不妨碍团队间的交流。如果从其他团队有借鉴之处（包括网上作者），必须在参考文献中明确标注，否则将被视为剽窃。

问题来源：随着智能家居与物联网技术的不断发展，家庭电力消耗监测与管理成为节能减排、降低用电成本和实现智能化生活的重要环节。通过对家庭电力消耗进行细致的数据采集和建模分析，不仅有助于居民了解自身用电行为，还能为电力公司合理调度和预测电力负荷、平衡供需提供技术支持。家庭电力消耗受多种因素影响，例如季节变换、节假日、家庭成员行为模式、用电设备种类及气象条件等，这使得准确预测具有较大挑战性和现实意义。

对家庭电力消耗进行多变量时间序列预测，可以帮助用户及时发现异常用电，合理安排用电时间，降低峰值负荷，从而节省电费和提升能源利用效率。同时，精确的用电预测对智能电网的动态调度、可再生能源接入与分布式能源管理等也有重要推动作用。典型数据集为UCI Machine Learning Repository 公开的“Individual household electric power consumption”数据集，可在<https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption> 获取，采集自法国一户家庭，记录时间跨度从2006年12月到2010年11月，数据粒度为每一分钟，包括全屋有功功率、无功功率、电流、电压、各路子表的能耗等多个变量。我们通常会以每天为单位对原始数据进行汇总，同时可融合天气等外部因素作为输入变量，构建多变量时间序列预测模型。数据以月为基础汇总数据，然后提取并添加相应的信息。天气信息可在下述网站获取：<https://www.data.gouv.fr/fr/datasets/donnees-climatologiques-de-base-mensuelles>。

数据各列含义如下：

global_active_power（全局有功功率）：家庭总的有功功率消耗，单位为千瓦（kW），表示实际消耗的电能。

global_reactive_power（全局无功功率）：家庭总的无功功率消耗，单位为千瓦（kW），表示储存在电路中并来回转换的能量。

voltage（电压）：电路中的平均电压值，单位为伏特（V）。

global_intensity（电流强度）：家庭总的平均电流强度，单位为安培（A）。

sub_metering_1（分表1能耗）：厨房区域的有功能量消耗，单位为瓦时（Wh）。

sub_metering_2（分表2能耗）：洗衣房区域的有功能量消耗，单位为瓦时（Wh）。

Sub_metering_3（分表3能耗）：气候控制系统的有功能量消耗，单位为瓦时（Wh）。

RR：月累计降水高度，单位为毫米的十分之一（即记录值需除以10）

NBJRR1 / NBJRR5 / NBJRR10：当月日降水 $\geq 1/5/10$ mm的天数

NJBROU：当月雾出现的天数

提示：可以根据其他三个变量，计算出第四个变量： $\text{sub_metering_remainder} = (\text{global_active_power} * 1000 / 60) - (\text{sub_metering_1} + \text{sub_metering_2} + \text{sub_metering_3})$

预测问题：根据最近的电力消耗情况，接下来的预期电力消耗是多少？即要求对接下来每一天的**总有功功率**进行预测。

预测任务：根据所提供的数据对未来**总有功功率**进行预测。基于过去90天的数据曲线来预测未来90天（短期预测）和365天（长期预测）两种长度的变化曲线（需要分别训练，即长期预测的模型参数不能用于短期预测）。按照方法分为三部分。

前两部分为基础题，第三部分为开放题，各占总分的三分之一：

1. 使用 LSTM 模型进行预测；
2. 使用 Transformer 模型进行预测；
3. 使用自己提出的改进模型进行预测，结构不限，例如可以结合卷积层提取局部特征后接 Transformer 编码以改进长期依赖建模能力。此部分以原理的新颖程度为首要评价标准，性能为次要评价标准。

训练与测试：数据集主要分为 train 和 test 两部分（具体见文件“train.csv”和“tes.csv”）。请使用两种评价标准进行测试，即均方误差（MSE）与平均绝对误差（MAE）。至少进行五轮实验，并对结果取平均值，同时提供标准差（std）以评估结果的稳定性。

提示：数据处理方面应该按照以下处理

global_active_power、global_reactive_power、sub_metering_1、sub_metering_2 按天取总和
voltage、global_intensity 按天取平均

RR、NBJRR1 / NBJRR5 / NBJRR10、NJBROU 取当天的任意一个数据

提交方式：实验报告应由以下四部分构成：1. 问题介绍、2. 模型（可以包含少量伪代码）、3. 结果与分析、4. 讨论。同时，需要提交代码（必须给出 Github 链接）。结果需以截图形式贴在报告中，并绘制出电量（power）预测与真实值（Ground Truth）曲线的对比图。请注意对三种方法进行比较。如果自行提出的方法虽然新颖但性能不佳，但原因分析有力，同样可以获得较高的分数。务必注明参考文献，否则将视为抄袭，每抄袭一处将扣除 33 分。允许使用 ChatGPT 、DEEPSEEK 一类工具撰写报告，但仅限于撰写部分，并注明，必要的参考文献仍然不可或缺。

注意：所提供的数据中可能会存在数据项缺失的情况，真实数据有缺失是正常现象，但这并不对我们的预测任务产生影响。尤其注意数据基本时间是以分钟划分，需要自行处理。此外，一个 Sample 的大小为（input+output），如果对时间序列数据处理方面有所疑问的同学，可以和上一届参加过机器学习课程的同学进行交流，如果对窗口大小和步长等概念不太清楚的同学可以参考以下博客。

[1]https://blog.csdn.net/qg_47885795/article/details/143462299

[2] https://blog.csdn.net/weixin_39653948/article/details/105431099

[3]https://datac.blog.csdn.net/article/details/105928752?fromshare=blogdetail&sharetype=blogdetail&sharerId=105928752&sharerefer=PC&sharesource=weixin_44709585&sharefrom=from link