
1 问题介绍

1.1 问题背景

随着智能电表和物联网设备的普及，家庭级别的用电数据被以时间序列的形式大量采集，准确预测家庭的用电趋势不仅有助于提升电力系统的运行效率，优化能源分配，还能为居民提供个性化的用电建议，从而降低能耗、节约成本。然而，家庭用电行为受到诸多因素影响，如季节变化、居民作息、家电使用习惯等，呈现出明显的周期性、波动性与非线性特征，因此对建模方法提出了较高要求。近年来兴起的深度学习方法，特别是基于循环神经网络（如 LSTM）和自注意力机制（如 Transformer）的模型，为高精度的时间序列预测提供了新的解决思路。围绕这一问题，构建科学有效的预测模型，不仅对推进家庭能源智能化管理具有应用价值，也为研究复杂时间序列建模提供了典型案例。

1.2 问题重述

本研究旨在利用过去连续 90 天的多变量数据，分别预测未来 90 天（短期预测）和未来 365 天（长期预测）的全局有功功率变化情况。

数据集各列含义如表 1 所示：

表 1 数据集说明

字段	含义	单位
Global_activate_power	全局有功功率：家庭总的有功功率消耗，表示实际消耗的电能。	千瓦 (kW)
Global_reactive_power	全局无功功率：家庭总的无功功率消耗，表示储存在电路中并来回转换的能量。	千瓦 (kW)
Voltage	电压：电路中的平均电压值。	伏特 (V)
Global_intensity	电流强度：家庭总的平均电流强度。	安培 (A)
Sub_metering_1	分表 1 能耗：厨房区域的有功能量消耗。	瓦时 (Wh)
Sub_metering_2	分表 2 能耗：洗衣房区域的有功能量消耗。	瓦时 (Wh)
Sub_metering_3	分表 3 能耗：气候控制系统的有功能量消耗。	瓦时 (Wh)
RR	月累计降水高度。	1/10 mm
NBJRR1	当月中日降水 $\geq 1\text{mm}$ 的天数	天
NBJRR5	当月中日降水 $\geq 5\text{mm}$ 的天数	天
NBJRR10	当月中日降水 $\geq 10\text{mm}$ 的天数	天
NBJRROU	当月中出现雾的天数	天

2 模型介绍

本研究设计了三种深度学习模型：LSTM、Transformer 和 FeSA-LSTM，分别用于多变量时间序列的短期与长期预测任务。

2.1 LSTM

如图 1 所示，LSTM 模型采用双层双向 LSTM 架构，单元数分别为 128 和 64，输出层为全连接线性层。

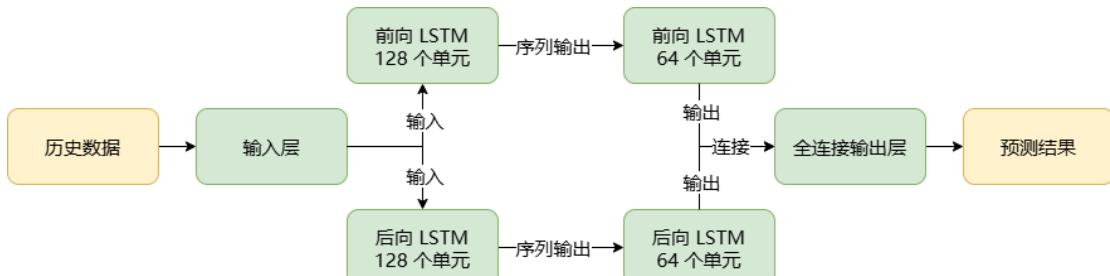


图 1 LSTM 模型结构

2.2 Transformer

Transformer 模型对原始结构进行了简化，如图 2 所示，维度 256，首先对输入进行可学习的位置向量编码，残差连接了 8 个注意力头组成的注意力组件和 512 维的 GELU 函数激活的两层前馈神经网络组件，每个组件都有层归一化，最后是进行全局平均池化，输出层为全连接线性层。

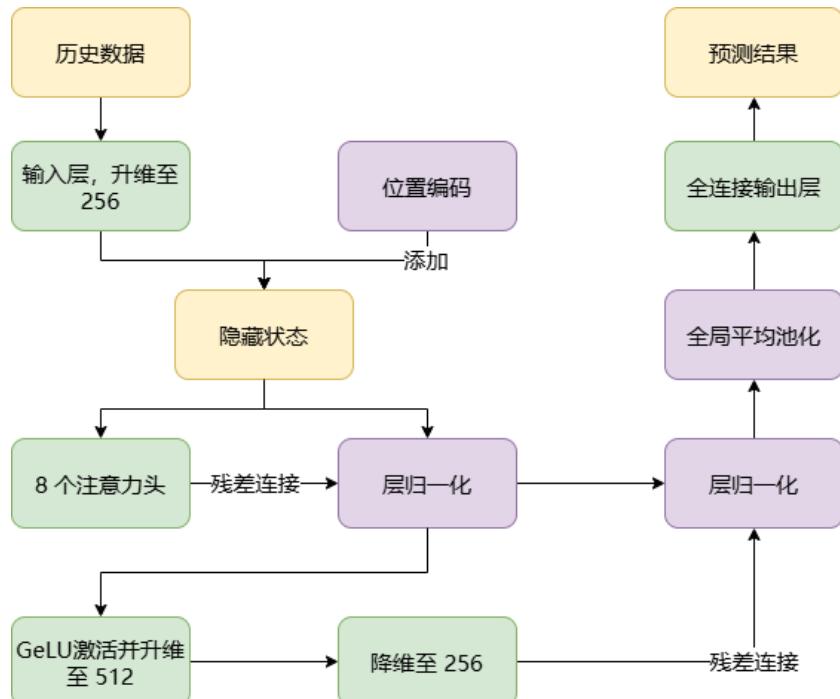


图 2 Transformer 模型结构

2. 3 FeSA-LSTM

Feature-Specific Attention LSTM (FeSA-LSTM) 是本研究提出的改进模型，旨在解决多变量 Transformer 注意力混合特征的弊端，整体结构如图 3 所示。该模型为每个特征构建独立注意力头，结合位置编码实现特征级建模，然后接入 LSTM 提取时序信息，融合局部与全局的特征表达能力。

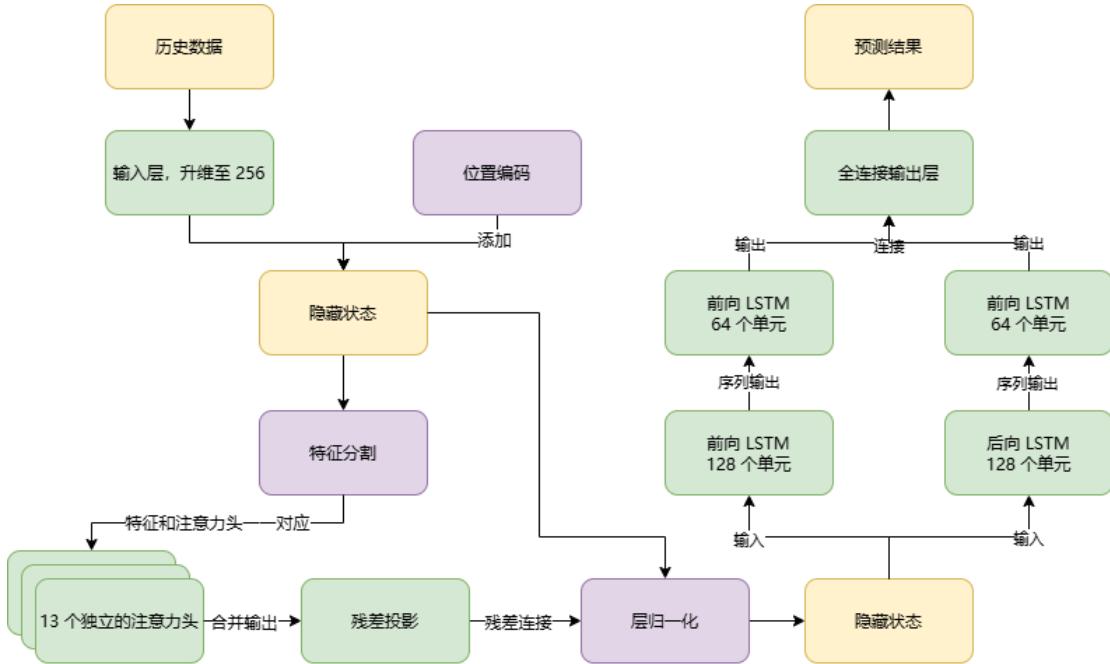


图 3 FeSA-LSTM 模型结构

FeSA-LSTM 的 Transformer 部分（260 维，是特征数的整数倍）首先对输入进行可学习的位置向量编码，接着实行特征分割，13 个注意力头各自关注自己的特征，然后将合并的残差投影到原始维度，最后进行层归一化和残差连接。

FeSA-LSTM 的 LSTM 部分采用双层双向 LSTM 架构，单元数分别为 128 和 64，输出层为全连接线性层。

3 结果与分析

3. 1 数据预处理

在训练模型之间，首先要对数据集进行预处理：

1. 数据清洗：将异常值“?”替换为 NaN；对于缺失值，使用前一天的数据进行填充，直到数据中再无 NaN。
2. 特征添加：计算家庭的其他电力消耗 $\text{Sub_metering_remainder} = (\text{Global_active_power} * 1000 / 60) - (\text{Sub_metering_1} + \text{Sub_metering_2} + \dots)$

Sub_metering_3), 该变量也作为特征。

3. 单位变换: RR 的单位是 1/10mm, 因此 $RR=RR/10$ 。
4. 数据聚合: 数据集的时间单位是分钟, 需要按天聚合。
 - (1) Global_active_power、Global_reactive_power、Sub_metering_1、Sub_metering_2、Sub_metering_3、Sub_metering_remainder 按天取总和;
 - (2) Voltage、Global_intensity 按天取平均;
 - (3) RR、NBJRR1、NBJRR5、NBJRR10、NJBROU 取当天的第一个数据。

3. 2 特征缩放

为提升深度学习模型训练的稳定性与预测精度, 本研究对所有输入特征进行了缩放处理。不同的特征具有不同的数值分布与量纲特性, 因此采用统一的缩放方式可能会导致模型偏向高方差或大尺度的特征。为此, 针对每一类特征的统计特征和分布形态, 选择最合适的方式, 以增强模型对不同特征的表达能力。

1. 如图 4 所示, Global_active_power、Global_reactive_power、Voltage、Global_intensity、Sub_metering_remainder 这些特征的分布连续平稳, 且量纲相对接近, 适合用标准归一化保持其自然波动幅度, 有助于加快模型收敛速度。

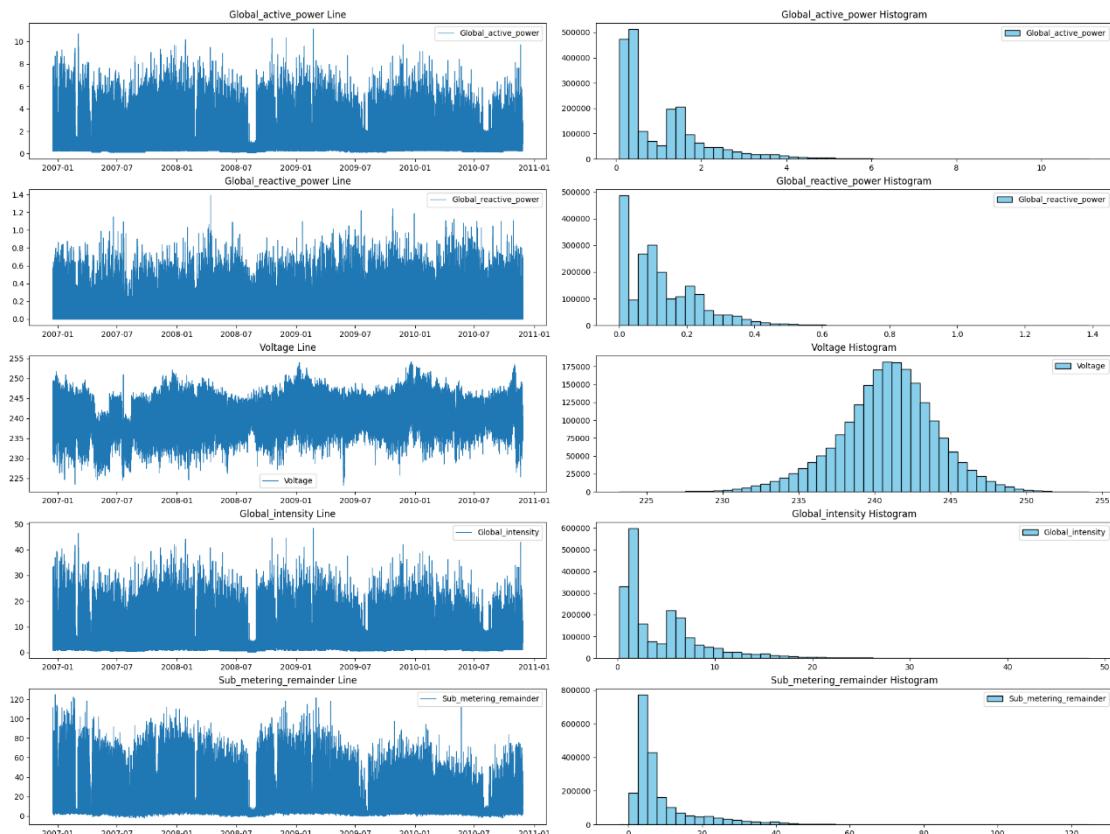


图 4 Global_active_power、Global_reactive_power、Voltage、Global_intensity、Sub_metering_remainder 特征分布情况

2. 如图 5 所示, Sub_metering_1、Sub_metering_2、Sub_metering_3 这些特征偏态分布明显、具有“稀疏激活”的特性, 先进行对数变换以缓解偏态, 再进行标准归一化, 减少异常值对训练结果的干扰。

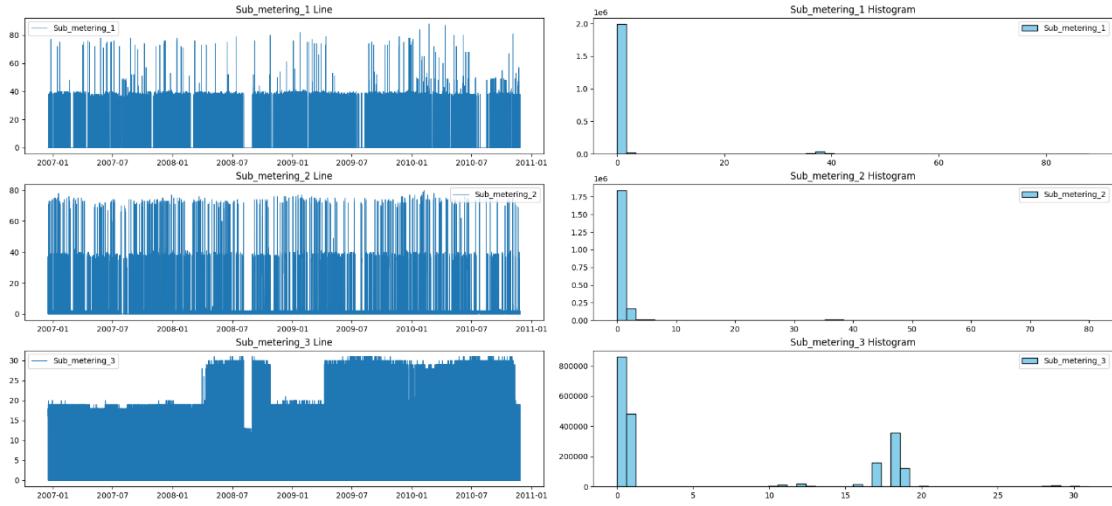


图 5 Sub_metering_1、Sub_metering_2、Sub_metering_3 特征分布情况

3. 如图 6 所示, RR、NBJRR1、NBJRR5、NBJRR10、NBJBROU 都属于气象类特征, 数值范围固定且取值离散, 采用最小最大归一化, 提升模型对气象因素的敏感度。

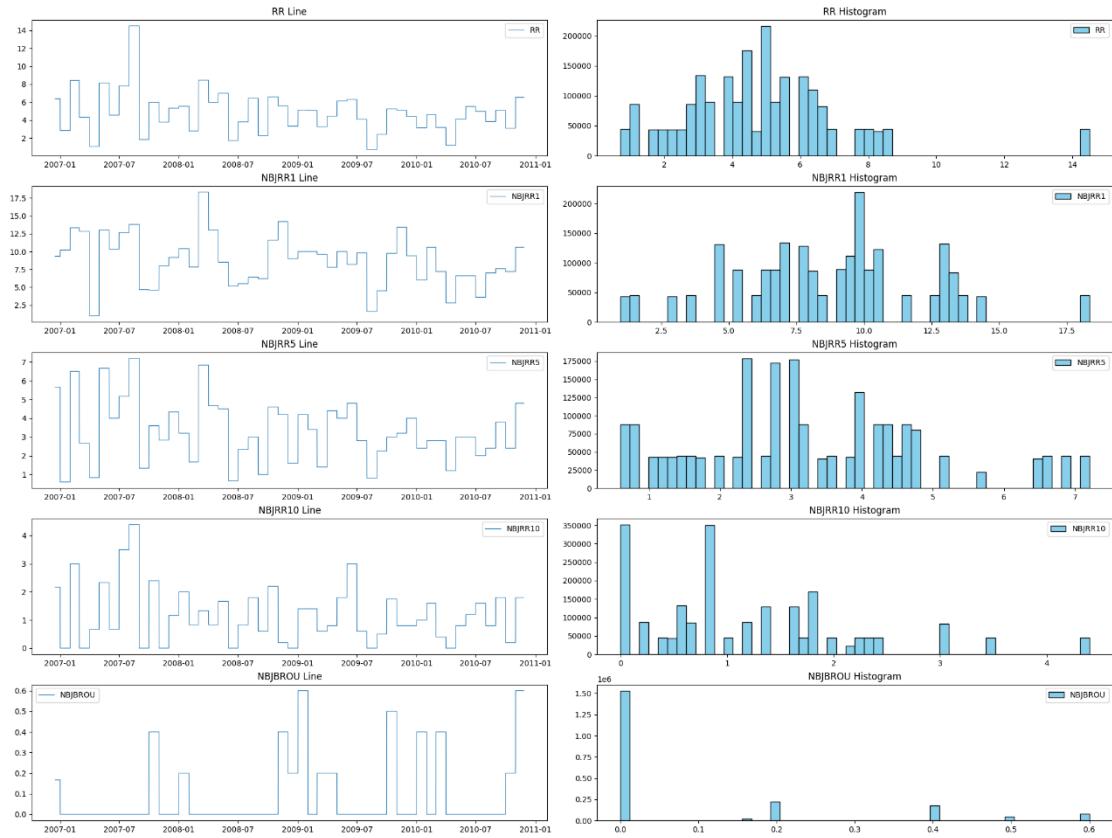


图 6 RR、NBJRR1、NBJRR5、NBJRR10、NBJBROU 特征分布情况

3.3 实验结果

本实验的短期预测和长期预测各自需要 3 个模型，模型参数独立。每个模型进行了 5 次独立的实验，计算每次实验的预测值和真实值之间的 MSE（均方误差）和 MAE（平均绝对误差），得到 5 次实验的平均 MSE 和平均 MAE，并计算 MSE STD（标准差）和 MAE STD 来证明评估结果的稳定性。

每个实验的 epochs 为 100，学习率为 0.001，优化器为 Adam。滑动窗口的步长为 1，预测方式为滚动预测，预测取值方式为指数移动平均， α 为 0.1。

实验结果如表 2，均保留两位小数：

表 2 实验结果

Task	Model	评估指标			
		AVG MSE	MSE STD	AVG MAE	MAE STD
90d	LSTM	186878.16	12436.26	339.18	11.11
	Transformer	170609.99	11622.64	318.47	13.91
	FeSA-LSTM	178208.70	10063.04	325.22	9.96
365d	LSTM	161190.73	3466.38	306.89	4.15
	Transformer	173851.17	6712.79	327.16	8.55
	FeSA-LSTM	167953.66	6066.18	314.96	6.89

最终绘制各个模型的平均预测值和真实值的对比图，短期预测结果如图 7 所示，长期预测结果如图 8 所示。

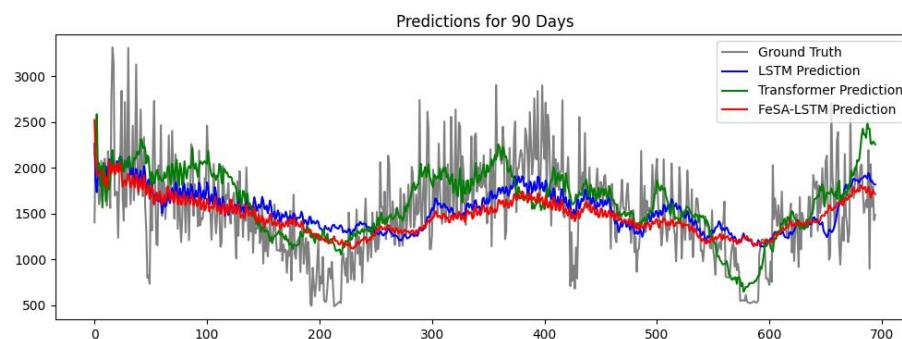


图 7 短期预测结果对比

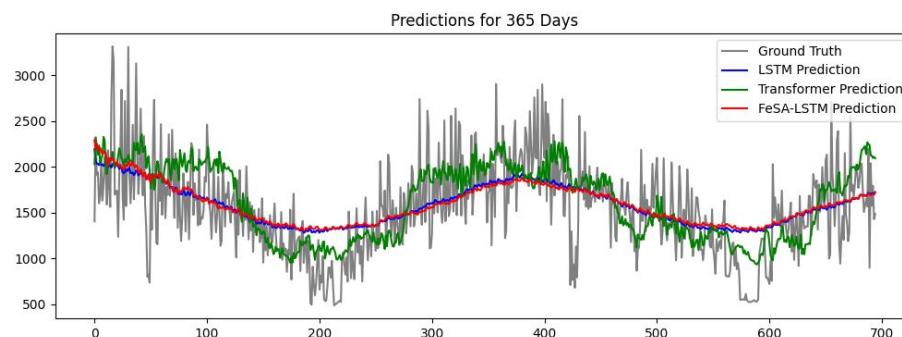


图 8 长期预测结果对比

4 讨论

4. 1 短期预测

从表中数据来看，整体性能 Transformer>FeSA-LSTM>LSTM，表明 Transformer 的短期拟合能力最强，可以捕捉到更多的特征变化；LSTM 对不同特征的交互表达较弱，略显劣势；FeSA-LSTM 在结合了特征独立注意力机制后，其拟合能力也超过了 LSTM。

从图中数据来看，Transformer 的整体走向最接近真实值，FeSA-LSTM 最为平滑，LSTM 偏差较大。

4. 2 长期预测

从表中数据来看，整体性能 LSTM>FeSA-LSTM>Transformer，表明 LSTM 的长期拟合能力最强，由于序列记忆使得其泛化性更强；FeSA-LSTM 由于其 LSTM 结构，预测能力保持稳定；由于采取滚动预测，Transformer 的注意力机制无法发挥其优势。

从图中数据来看，Transformer 的波动性最强，FeSA-LSTM 和 LSTM 非常平滑，拟合较为接近。

4. 3 模型稳定性

LSTM 在短期或长期窗口中能积累时间上下文信息，对波动趋势建模稳定；Transformer 对输入序列长度和全局注意力高度敏感，容易受到长序列噪声影响；FeSA-LSTM 通过特征划分缓解了注意力不集中问题，提升了短期稳定性。

参考文献

数据预处理：EAI2. 时间序列预测 11：用电量预测 01 数据分析与建模[EB/OL].(2024-08-17)[2025-07-01]. <https://datac.blog.csdn.net/article/details/105397315>.

问题背景章节和特征缩放章节由 ChatGPT 进行优化，所有模型结构图使用 draw.io 绘制。