**一、摘要**

本次项目主要使用LSTM模型进行短期电力负荷预测，并得到了比全连接神经网络更好的效果。

**二、项目简介：**

较为精准的电力负荷预测对电力现货交易，电网的安全运行，电力的保障等有重要的意义。预测方法主要分为统计学预测模型和智能预测模型。如神经网络、时间序列法、回归分析、支持向量机、模糊预测法等。近些年来，深度学习在电力系统负荷预测方面应用已经十分广泛，较为成熟。不少是结合了多种方法进行负荷预测。

本次项目是对未来24小时的电力负荷进行预测。

**三、数据处理：**

**1.数据观察**

(1) 数据集构成：21年1月至23年2月地区负荷采样数据（间隔15min）；气象数据(种类包括：天气状态、温度、湿度、风速、风力等级、风向角、压力、可见度、降水、光照强度)，数据粒度与负荷匹配。

(2) 获取原始电力负荷数据集后，根据了解到的知识以及对数据的观察发现，电力负荷在日，周和年上的变化有很强的规律性，周一和周日的负荷会相对低一些。

(3) 在法定节假日时电力负荷先下降后上升。在元宵节负荷基本没有什么变化，并且元宵节不是法定节假日，所以在这个数据集里不将元宵节作为节假日处理，这里holiday为1的都是国家法定节假日。

(4) 月份的第几天和负荷没啥关系。

(5) 随着年份的增加，电力负荷整体上呈上升趋势，并且年初电力负荷都经历了一次较大的下降，不知道是什么原因。

**2.处理**

（1）将时间特征表示为当前一天内的分钟数、周几(周一用1表示，周二用2表示...)、是否为法定节假日（是用1表示，调休用0.5表示，其他用0表示）、年份（2021用1表示，2022用2表示，2023用3表示）。

（2）2022年和2023年元宵节附近10来天电力负荷出现了很明显的下降，2021年在农历腊月十几电力负荷出现了明显的下降，持续时间和强度与2022和2023年年初类似，但却不是在元宵节期间，不知道为什么会明显下降。

（3）一年中不同时间负荷的变化受所在地区气候的影响比较大，可以用气象信息包含，不再加入月份特征。

（4）突发情况可由人工调节。

（5）电力负荷的最大值没发现异常，最小值发现了异常。异常值用其附近负荷的平均值代替。

（6）将object类型数据两边的引号去掉，然后转为数值类型，其值形如“‘0’”。

（7）气象数据未发现异常值。

（8）风速和风力等级重复了，并且这个数据集里风力等级只有0，1，2，3，17这5种取值，所以删去风力等级(wind\_level)列。

（9）将处理后的数据保存为new\_data.csv。

（10）使用决策树训练，得到的特征重要性作为筛选特征的依据。最终保留temperature, humidity, pressure, visibility, holiday, minute, week, year 这8个特征。

（11）2021年、2022年、2023年年初电力负荷均出现了明显的先下降后上升过程，但原因不明，需要和电力公司沟通，了解原因。因此在模型训练和测试时选择将这段时间的数据忽略。

（12）划分训练数据和测试数据。2021年、2022年前11个月和2023年的数据作为训练，2022年12月的数据作为测试。如将2023年的数据作为测试，2021和2022年的数据训练的话模型对年份特征的学习不够好，希望的情况是有过去3年以上的数据用作训练，最近一年的数据用作测试。

（13）对训练数据和测试数据使用训练数据的最大值和最小值进行归一化。

**四、模型训练：**

使用了决策树、全连接神经网络、LSTM三种模型。

使用方差（MSE）作为损失函数。

分别尝试了不同的模型结构和特征构造。

训练中发现全连接神经网络训练时损失下降较慢、波动较大，并且训练多轮后在训练集上的损失仍在较高的水平。LSTM训练时损失下降较快、波动较小，并且训练多轮后在训练集上的损失达到了一个很低的程度。

全连接神经网络训练一轮比LSTM要快许多。

**五、模型评价：**

1. APE（绝对百分比误差）：|y预测 - y真实| ÷ y真实 × 100%。评价指标：MAPE(所有APE的平均值)。 每次训练后若模型在测试集上的MAPE减小，则保存这次训练的模型，若连续多次训练在测试集上的MAPE没有小于之前的模型，则将之前的模型作为最终的模型。

2. 决策树的MAPE为9.274728886615331%，其只使用了预测时刻的数据用于训练，未加入过去时刻的数据进行训练。

3. 全连接神经网络性能最好的是MAPE为2.963206226104105% (2.4212825111301317%)，括号里的是去掉测试集12月后两天的误差较大的异常样本后MAPE。因为观察12月30日的数据发现，这天为周五，气象数据未发现异常，且不是节假日，12月31日才是元旦假期，但负荷相较前一天出现了明显的下降，推测这是导致误差较大的原因。

训练样本的数量为66629，测试样本的数量为2785。每个样本包括特征和标签，其中特征为前24小时数据（包括负荷值和特征）和未来24小时的数据(不包括负荷值)，标签为未来24小时的负荷值。生成的样本起始时刻依次相差15分钟。

使用前24小时和未来24小时的数据预测未来24小时的负荷。

设置训练数据的批量大小为128。

模型输入层为1632个神经元，有一个隐藏层，神经元个数为1000，输出层为96个神经元，对应未来24小时的96个时刻（间隔15分钟）。

训练过程中的学习率依次为0.01，0.001，0.0001, 0.00001。

4. LSTM性能最好的为LSTM-11和LSTM-19。

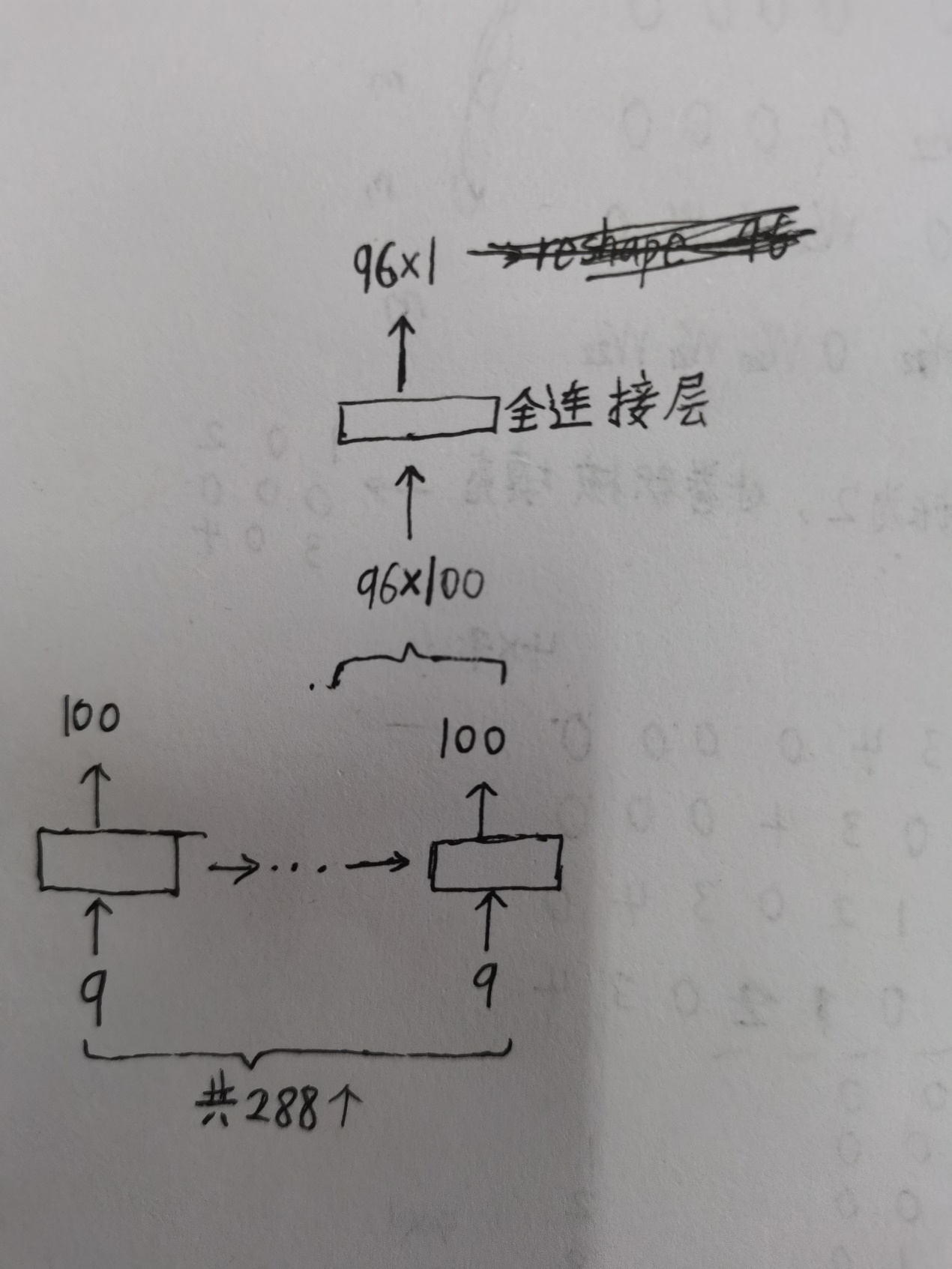
（1）LSTM-11的MAPE为2.4541228872315464%（1.851189349256456%）。

训练样本的数量为66149，测试样本的数量为2689。每个样本包括特征和标签，其中特征为前192个时刻和后96个时刻的气象数据和负荷值（后96个时刻的负荷值用1代替，只是占个位置，保证输入序列的特征数一致，没有实际含义），标签为后96个时刻的负荷值。生成的样本起始时刻依次相差15分钟。

使用前48小时和未来24小时的数据预测未来24小时的负荷。

设置批量大小为96。

模型结构如下图所示



先使用学习率为0.01训练, 训练集上的损失降低到一定程度时使用学习率为0.001训练。

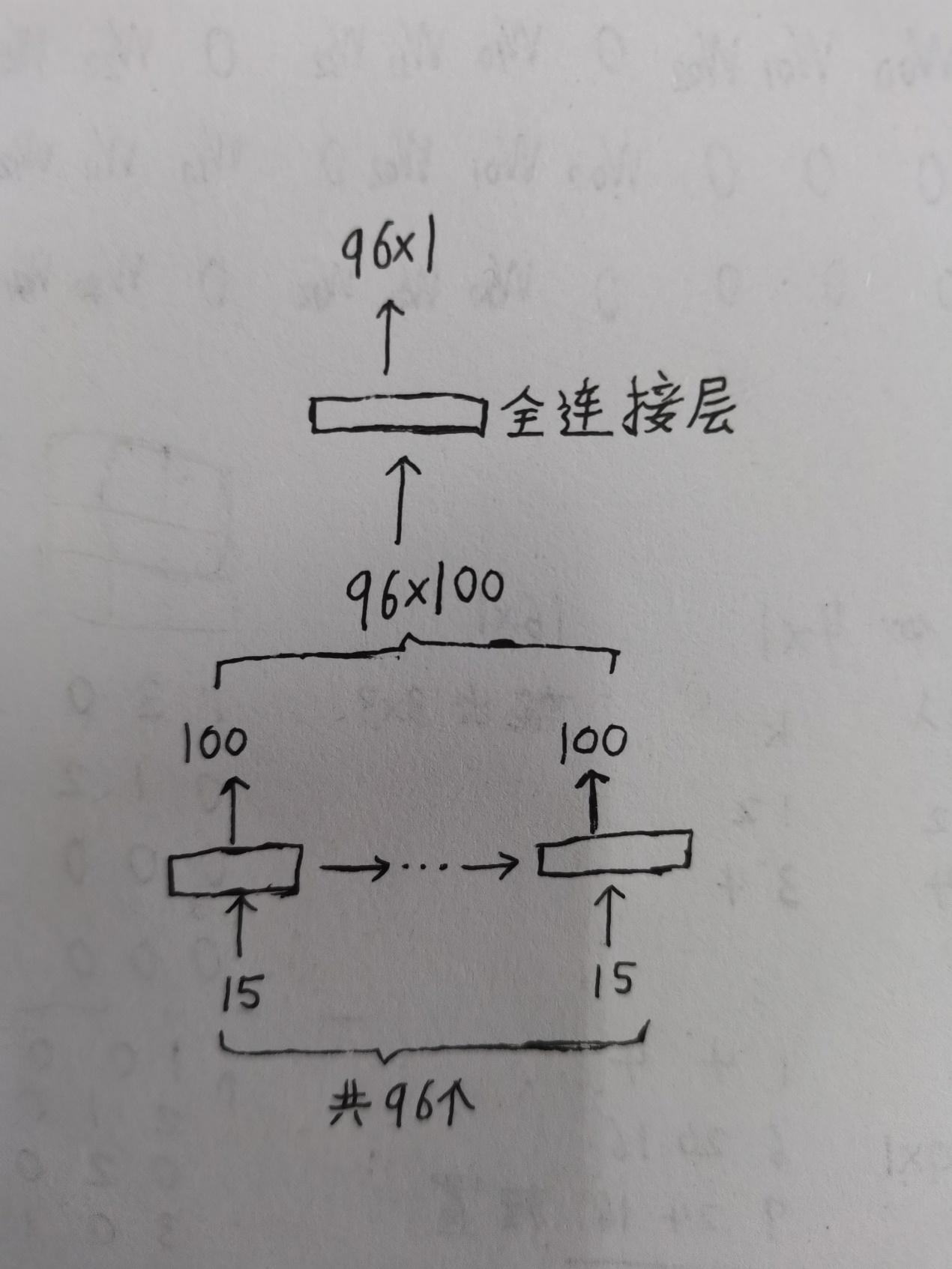
（2）LSTM-19的MAPE为2.3963830265355237%（1.4876579731935635%）。

训练样本的数量为63749，测试样本的数量为2209。每个样本包括特征和标签，其中特征有96个序列，每个序列为前7\*24小时同一个时刻的负荷和未来24小时同一时刻的数据，时刻从0:00到23:45，标签为未来96个时刻的负荷值。生成的样本起始时刻依次相差15分钟。

使用前7\*24小时和未来24小时的数据预测未来24小时的负荷

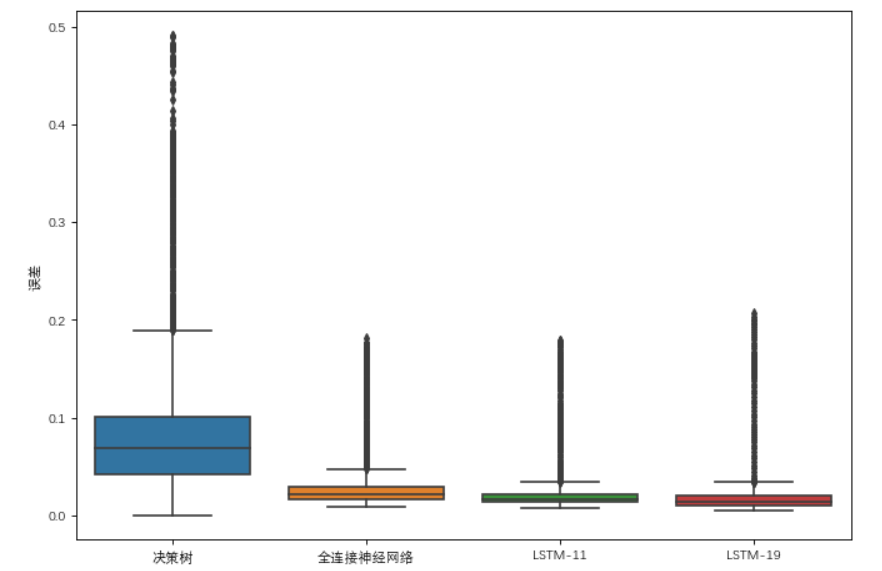
设置批量大小为128

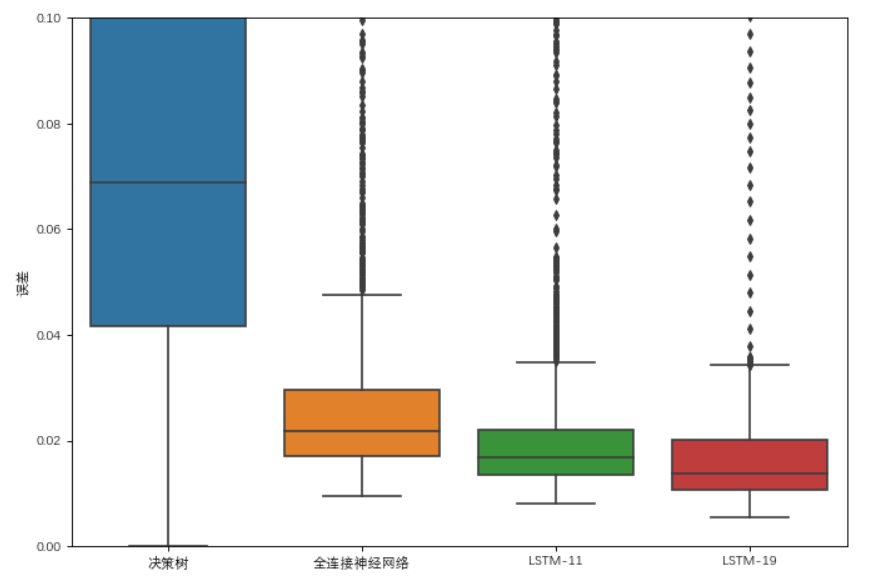
模型结构如图所示：



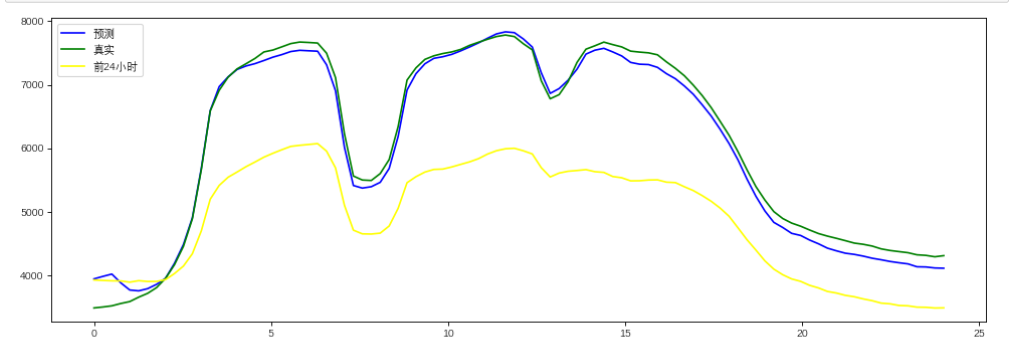
训练时先使用了学习率0.01，之后为0.001，最后为0.0001。

5. 测试样本MAPE的箱线图（纵轴为未乘以100%的MAPE）。





6. 对某个测试样本的预测



**六、不足：**

一些模型训练次数不够多。

特征和模型结构还有很多组合没有实验。

数据处理不够细致，某些分类特征可考虑embedding。

受未来天气预报准确性的影响。

对异常数据的处理还不充分，需要获得更多的相关信息，也可考虑增加异常情况列。

未实验其他的神经网络模型。

学习率之后可使用动态学习率，不再人工调节。

**七、结论：**

这次项目主要通过LSTM模型进行了短期负荷预测，得到了有一定的参考价值的效果。未来可以尝试更多的模型进行负荷预测，看是不是可以取得更好的效果，同时对特征的处理尝试进一步的改进。

**八、参考资料：**

[【负荷预测】电力负荷预测方法概述 - 知乎 (zhihu.com)](https://zhuanlan.zhihu.com/p/43279308)

[[干货]深入浅出LSTM及其Python代码实现 (zhihu.com)](https://www.zhihu.com/tardis/zm/art/104475016?source_id=1005)

[基于 PyTorch + LSTM 进行时间序列预测（附完整源码） - 知乎 (zhihu.com)](https://zhuanlan.zhihu.com/p/617126384)

[基于集成LSTM模型的数据驱动需求预测 (epet-info.com)](https://epet-info.com/html/2022/6/210826327.html)

[基于LSTM循环神经网络的短期电力负荷预测 (epet-info.com)](https://epet-info.com/html/2021/1/181224757.html)

[一种基于LSTM神经网络的短期用电负荷预测方法 (dwjs.com.cn)](http://e-press.dwjs.com.cn/dlxx/periodical/html/2017-15-09-19.html#b12--0-0-0-1710221715-1)