

电力系统短期负荷预测

2017011048 电 75 牛腾腾

目录

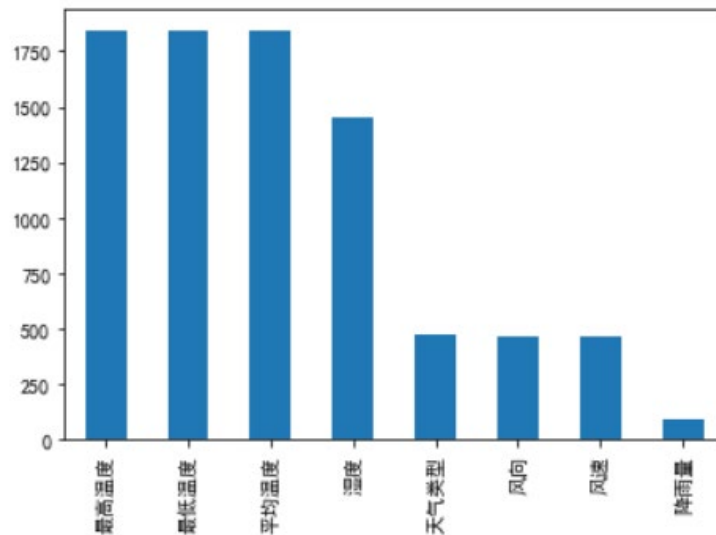
1. 数据预处理	2
1.1. 对气象数据的处理.....	2
1.1.1. 处理缺失值	2
1.1.2. 处理异常值	3
1.2. 对负荷数据的处理.....	3
1.2.1. 处理缺失值	3
1.2.2. 处理异常数据.....	3
1.3. 对时间类型数据的处理	4
1.4. 数据归一化	4
1.5. 最终结果	4
2. 模型搭建	4
2.1. 输入输出构造.....	4
2.2. 模型搭建	5
3. 训练与验证	5
3.1. 验证集划分	5
3.2. 超参数选择	5
3.3. 训练结果	5
4. 结果分析	8
4.1. 低精度数据分析	8
4.2. 改进空间	9
5. 总结	10

1. 数据预处理

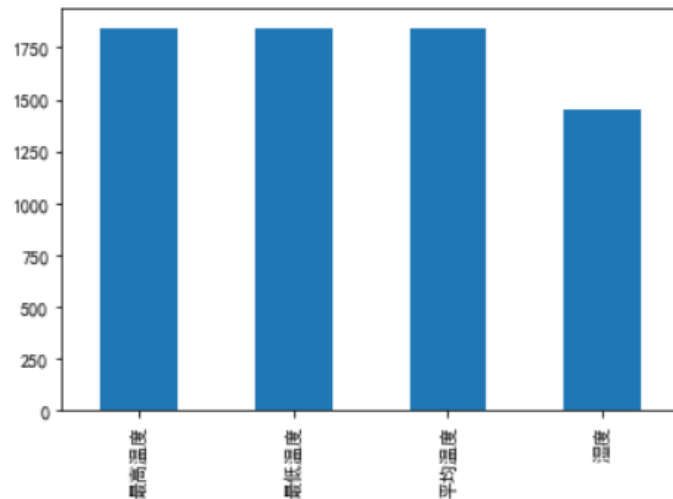
1.1. 对气象数据的处理

1.1.1. 处理缺失值

首先查看气象数据各字段的缺失情况



可以看到，天气类型、风向、风速、降雨量四个字段数据缺失严重，直接剔除剔除之后各字段数据缺失情况如下：



经过更加细致的检查，这四个字段数据缺失情况如下表所示：

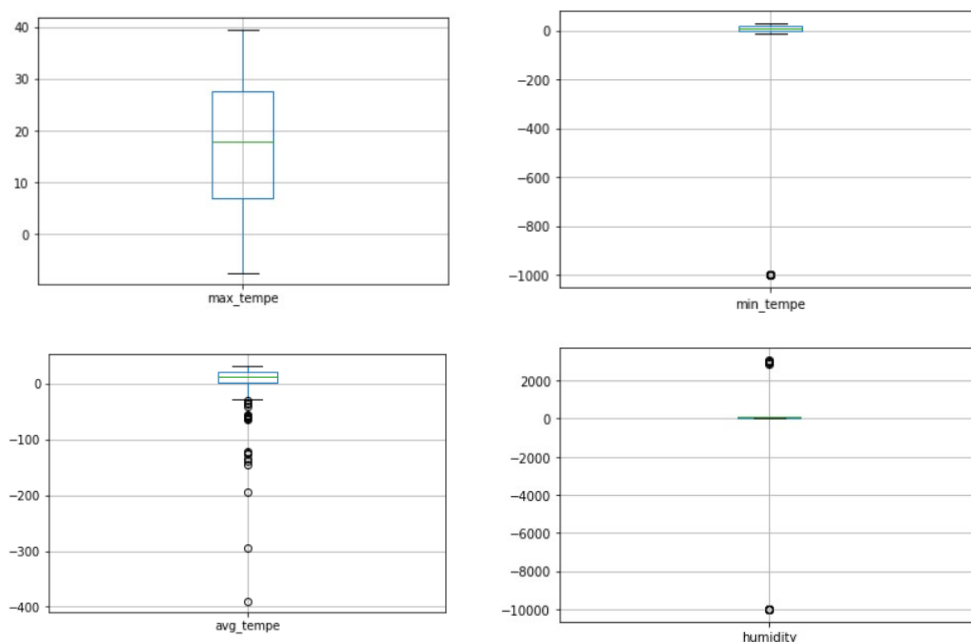
字段名	数据缺失数
平均温度	249
最高温度	249
最低温度	249

湿度	411
----	-----

查看数据发现，这些缺失值大多连片存在，比如 2003. 5. 10-2003. 12. 31 时间段内，这四个字段数据值全部缺失。因此这里对缺失值的处理方法是：使用相邻年份相同日期的数据填充缺失数据，比如用 2004. 5. 10-2004. 12. 31 的数据值填充 2003. 5. 10-2003. 12. 31 的缺失值。

1. 1. 2. 处理异常值

这里使用箱线图观察气象数据的异常情况：



可以看到，最低气温、平均气温、湿度三个字段有明显的异常数据，这里使用各字段的中位数替换异常数据

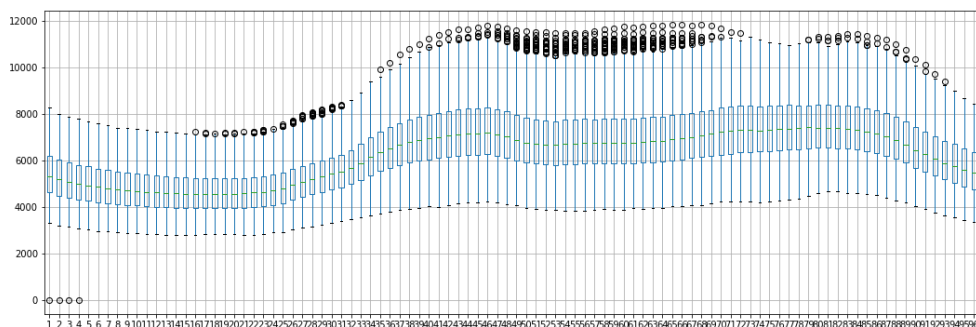
1. 2. 对负荷数据的处理

1. 2. 1. 处理缺失值

经过检查，符合数据仅有局部零星几个缺失数据，由于负荷数据在相邻的几天内，固定时间段的波动较小，这里处理缺失值的方法是：使用前一个非缺失值来填充缺失值。

1. 2. 2. 处理异常数据

查看负荷数据的箱线图



可以看到前 4 个字段出现 0 异常值，使用中位数进行填充

1.3. 对时间类型数据的处理

原始数据时间数据的格式为 yyyy/mm/dd，我们需要从中提取出每一天的星期类型、是否节假日，并进行 one-hot 编码。

1.4. 数据归一化

这里使用过 min-max 归一化，0-1 归一化，但在之后的训练测试过程中，发现模型的鲁棒性很差。经过多次测试后，使用原始数据 / 基值的方法进行归一化。具体而言，各字段的基值为：

字段	基值
负荷数据	7000MW
气温数据	20°C
湿度数据	100RH%

1.5. 最终结果

经过上述的数据处理过程后，最终得到了一张 $1982 * 109$ 的表，1982 指天数，109 指特征数，具体而言：

特征类型	特征数目
第 T 天的负荷数据	96
第 T 天的天气数据	4
第 T 天的星期类型（one-hot 编码）	7
第 T 天是否节假日（one-hot 编码）	2

2. 模型搭建

2.1. 输入输出构造

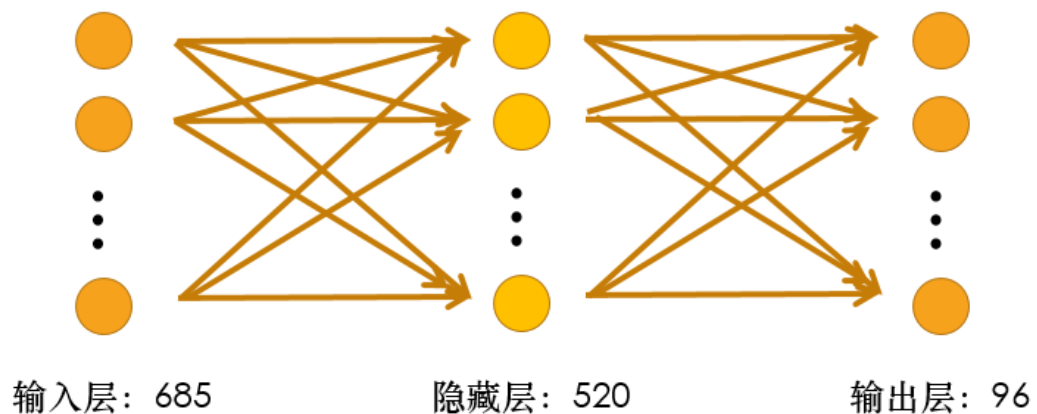
使用预处理后的数据构造输入输出，使用前七天的负荷与当天的气象、时间数据预测当天的负荷，每一条数据的格式如下：

输入	输出
历史负荷数据 第 T-7、T-6、T-5...T-1 天的负荷数据	第 T 天的 96 点负荷数据 (96 * 1)

(96 * 7)	
第 T 天的气象数据 (4 * 1)	
第 T 天的星期类型 (7 * 1)	
第 T 天是否节假日 (2 * 1)	

2.2. 模型搭建

搭建如下的 MLP 模型



3. 训练与验证

3.1. 验证集划分

数据集中总共有 1982 条数据，这里选择 2008.1.1-2008.6.4 的数据作为验证集（即后 155 条数据）

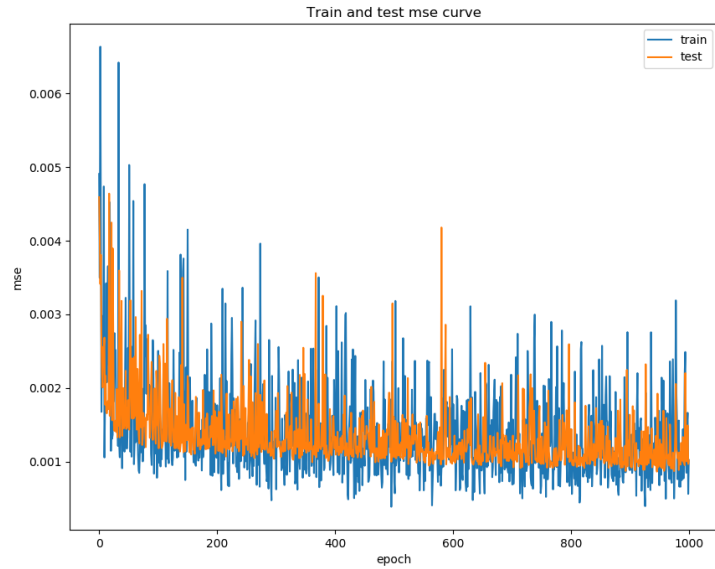
3.2. 超参数选择

经过不断的训练&验证，最终的超参数选择如下：

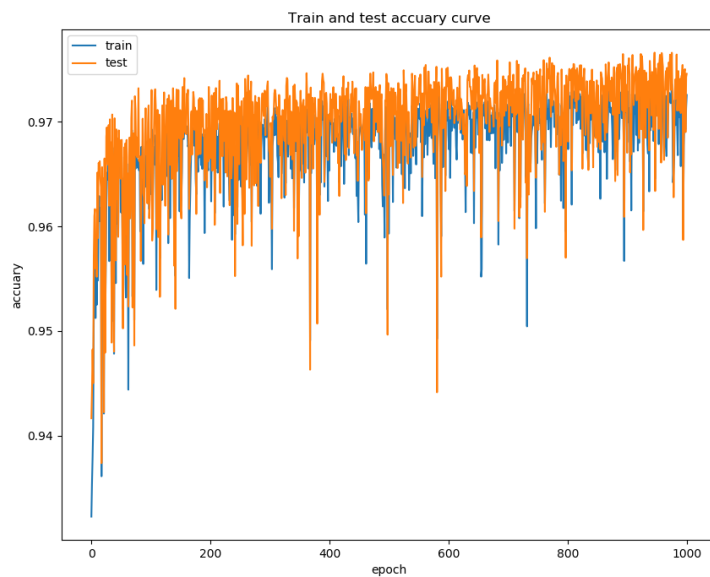
optimizer	Adam
learning_rate	1e-3
batch_size	32

3.3. 训练结果

训练集和验证集的 mse 曲线



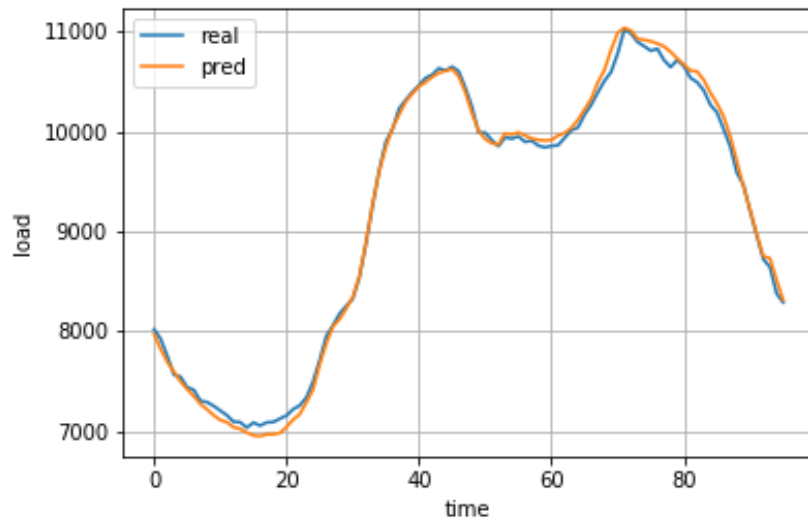
训练集和验证集的精度曲线



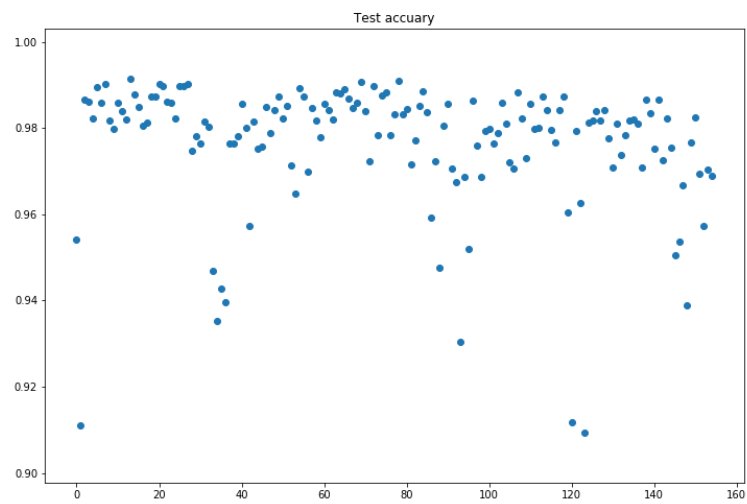
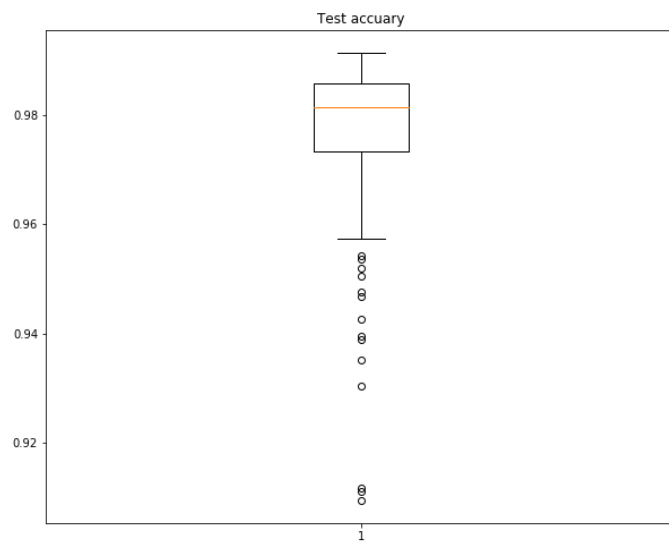
在第 942 次迭代中，获得了最优模型，模型效果如下：

	验证集
mse	0.000865
折合为负荷 mse	205.9296MW
平均精度	97.66%

验证集精度最高的一天：



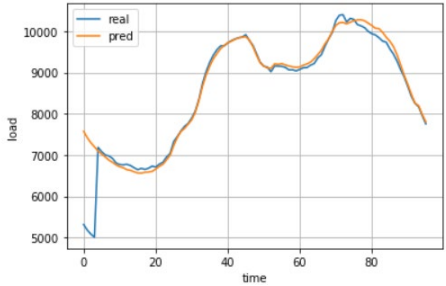
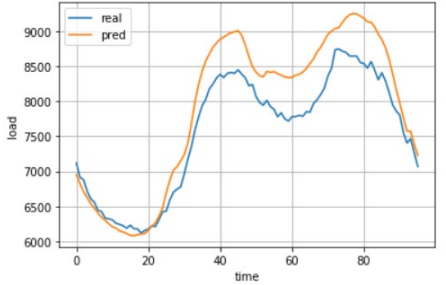
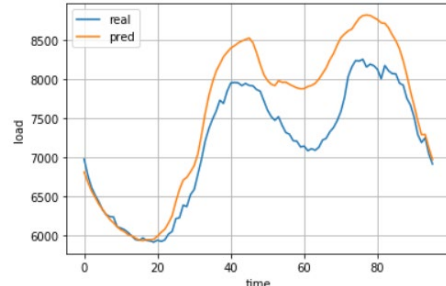
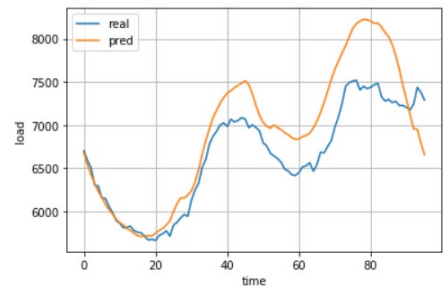
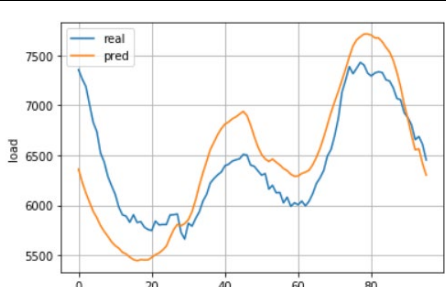
验证集预测精度箱线图与散点图

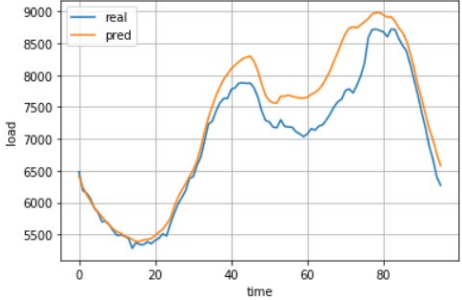
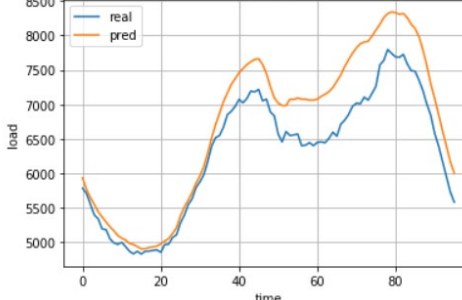
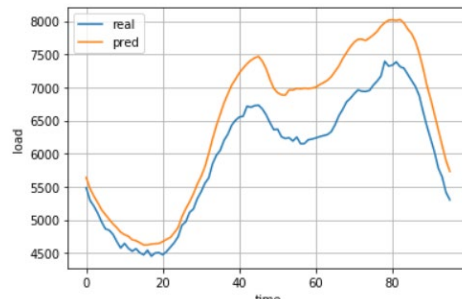
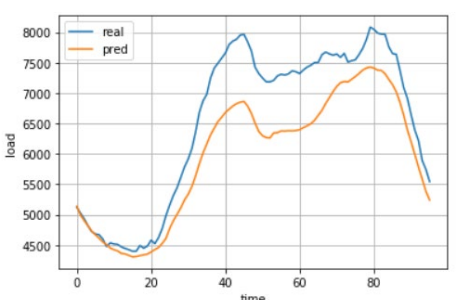
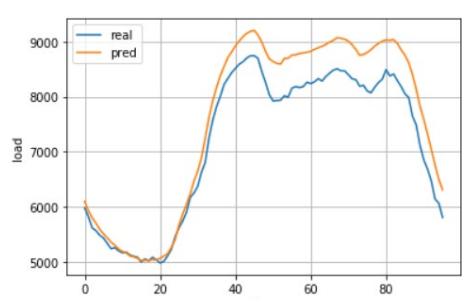


4. 结果分析

4.1. 低精度数据分析

从上面的精度散点图可以观察到有一些精度明显小于整体水平的点，这里把精度小于 95% 的点单独列出来进行分析，分析结果如下：

日期	精度	原因	预测值-真实值曲线
2008-01-03	91.1%	异常数据	
2008-02-04	94.7%	春节	
2008-02-05	93.5%	春节	
2008-02-06	94.3%	春节	
2008-02-07	94.0%	春节	

2008-03-30	94.8%	负荷数据与临近几天相比，突然降低	
2008-04-04	93.0%	清明节	
2008-05-01	91.2%	劳动节	
2008-05-04	91.0%	劳动节 (假期结束后，负荷水平突然增高)	
2008-05-29	93.9%	负荷数据与临近几天相比，突然降低	

可以看出，模型对节假日以及其他负荷水平突然变化的情况，预测精度不够高。

4.2. 改进空间

根据上面的低精度天数据的分析，提出下面两点改进空间

1. 对节假日单独建模，单独预测
2. 使用其他更适合于时间序列分析的模型比如 LSTM、GRU 等

5. 总结

本次实验搭建了一个 MLP 模型，其输入为前七天的负荷数据与待预测日的气象、时间特征数据，输出为待预测日的负荷数据。经过训练，模型在验证集共 155 条数据的平均预测精度为 97.66%，其中有 10 条数据预测精度低于 95%，原因主要是节假日负荷水平突变。