

用电量预测模型

罗磊 2018-07-04

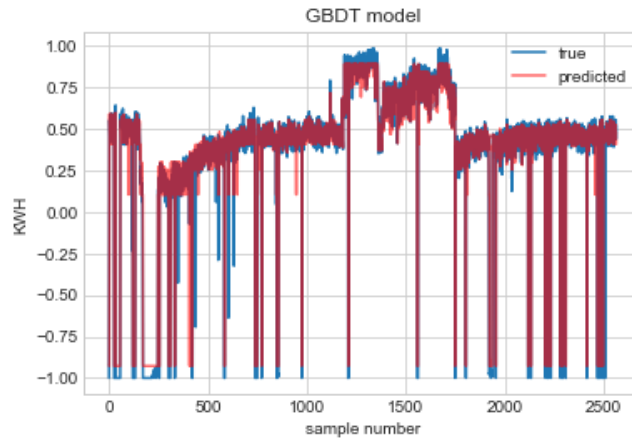
1. GBDT 模型

将数据样本按照 7: 3 的比例分割为训练集和测试集，使用上一时刻的数据来预测下一时刻的数据，使用 GBDT 模型对样本训练集进行训练，然后在测试集上进行测试。GBDT 模型的参数设置为：

表格 1. GBDT 模型参数设置

参数名称	参数值
n_estimators	300
learning_rate	0.01
min_samples_split	100
min_samples_leaf	200
max_depth	4
subsample	1
random_state	10
loss	'lad'

结果如下：



求得平均绝对百分偏差为 $MAPE = 0.1611$ 。

2. XGBoost 模型

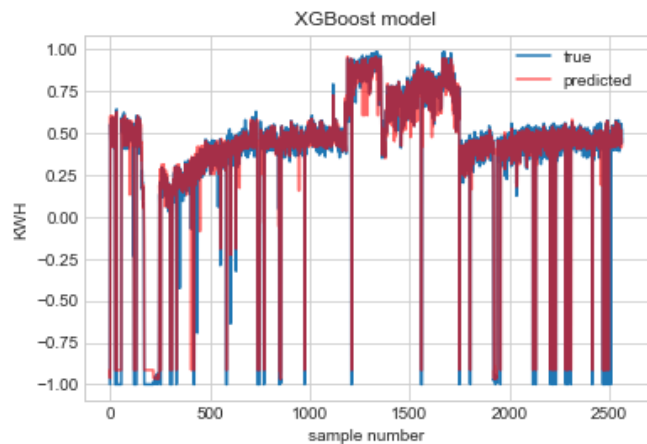
同 GBDT 模型一样，XGBoost 模型也使用上一时刻的数据来预测下一时刻的数据，参数设置为：

表格 2. XGBoost 模型参数设置

参数名称	参数值
n_estimators	100
learning_rate	0.1
seed	0

loss	'mse'
------	-------

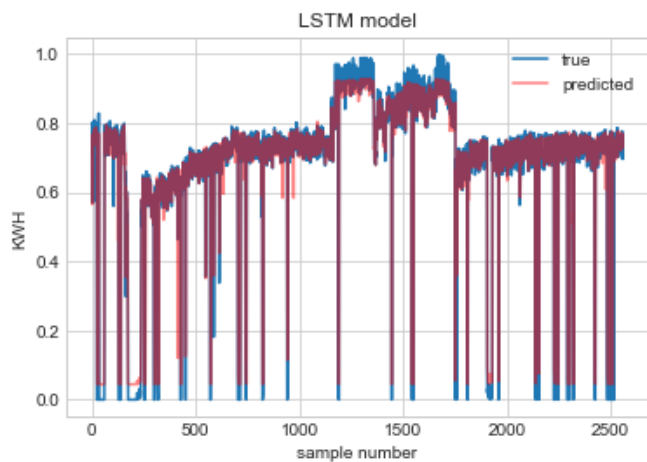
结果如下：



求得平均绝对百分偏差为 $MAPE = 0.1609$ 。

3. LSTM 模型

使用 Keras 框架搭建该 LSTM 模型，首先进行数据归一化，然后按照 7: 3 的比例，构造训练集和测试集。设置预测序列长度 $seq_len = 10$ ，即使用前 10 位用电量数据去预测接下来时刻的用电量数据。模型迭代次数 $epochs = 1000$ ，使用批式随机梯度下降法，设置 $batch_size = 3000$ 。对测试集预测结果如下：



求得平均绝对百分偏差为 $MAPE = 4.8434$ 。

4. K 折检验

下面均使用 K 折检验判断 XGBoost 模型和 LSTM 的准确度情况，其中参数 $k-fold = 10$ 。得到下表：

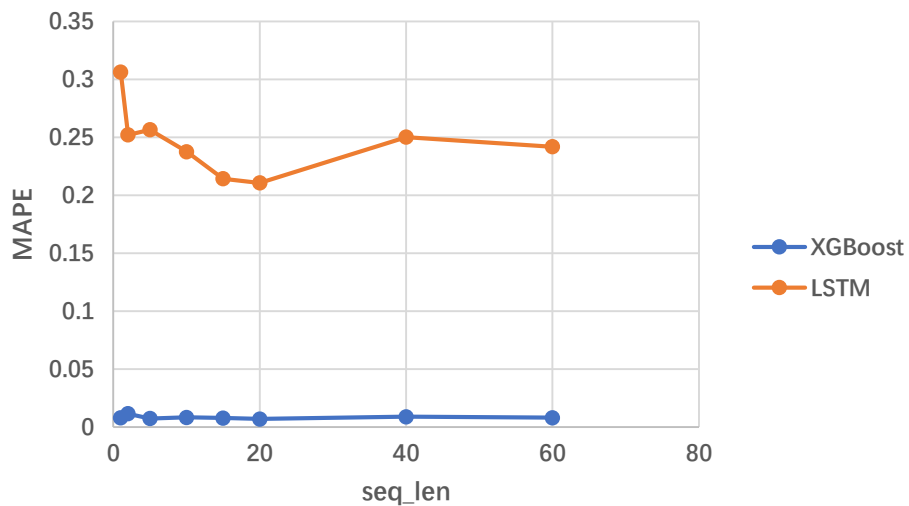
表格 3. XGBoost 和 LSTM 两模型进行 K 折检验得到的 MAPE 结果比较

XGBoost		LSTM	
0	0.007934	0	0.2237
1	0.005522	1	0.3739

2	0.005753	2	0.1490
3	0.006681	3	0.1564
4	0.006625	4	0.1755
5	0.007439	5	0.1742
6	0.006131	6	0.4220
7	0.013443	7	0.1591
8	0.008696	8	0.1532
9	0.004889	9	0.1750
mean	0.007312	mean	0.2162

从两模型的 MAPE 来看，XGBoost 略胜一筹，更适合解决这个问题，不过还需要后续调参来进一步确认。

5. 考察参数 seq_len 对模型 MAPE 的影响



从上图可以看出，随着 seq_len 的增大，LSTM 的预测准确度呈现先增加后降低的趋势，但在该程序中的表现不如 XGBoost（未调参）。