基于LSTM-RNN的股票预测研究

刘茁，1500011438，物理学院

胡承开，1800932142

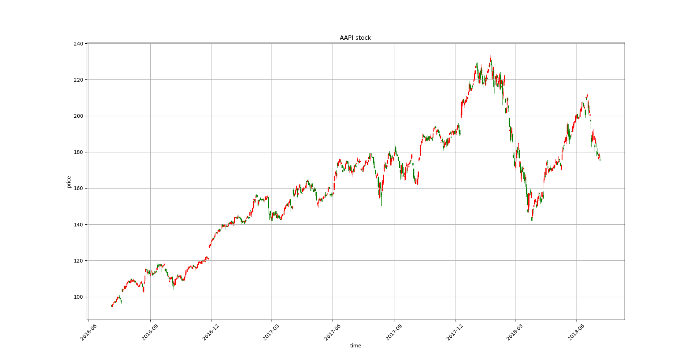
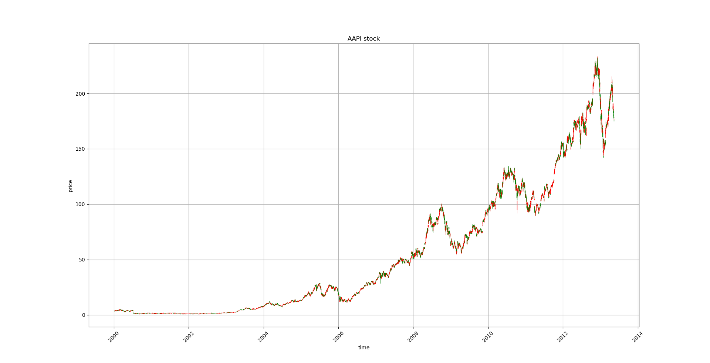
人们一直以来都期望可以预测股票未来价值，尽管这非常困难。这种困难源于股票的非平稳行为且没有任何明确的形式。因此，最好通过分析金融股票数据进行预测。为了处理大数据集，当前普遍使用滑动平均方法。然而，利用小波变换代替移动平均去噪股票信号，可以使金融数据更平滑，更准确地分解。LSTM(Long Short-Term Memory)是长短期记忆网络，是一种时间递归神经网络，适合于处理和预测时间序列中间隔和延迟相对较长的重要事件。

LSTM已经在科技领域有了多种应用。通过这些方法的实施，我们可以得到更准确的股票预测，进而增加利润。

本次研究包括以下6步：

**1、获取股票价格数据**

使用雅虎财经API获取，以苹果公司2000年1月1日-2019年1月1日的数据为例。共4884个样本，每个样本包括日期、开盘价、最低价、最高价、收盘价、已调整收盘价和成交量。



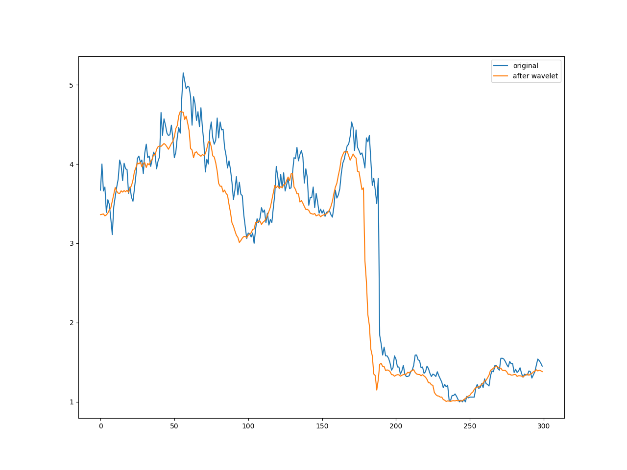
**2、小波变换去除数据中的噪声**

任何给定的股票都与高波动性、小样本量、非平稳和非线性行为以及大量高噪声变量有关，这些都使得其分析变得困难。高噪声数据导致股票历史交易价格与未来价格之间存在不完全信息缺口。目前，当涉及到处理和消除金融数据的时候，行业标准是使用所谓的滑动平均值。虽然滑动平均值已被证明是处理数据的一种可行方法，但我们提出了一种通过小波去噪程序进行替换的方法，这将产生改进的结果。利用小波变换代替滑动平均值去噪股票信号，可以使金融数据更平滑，更准确地分解。

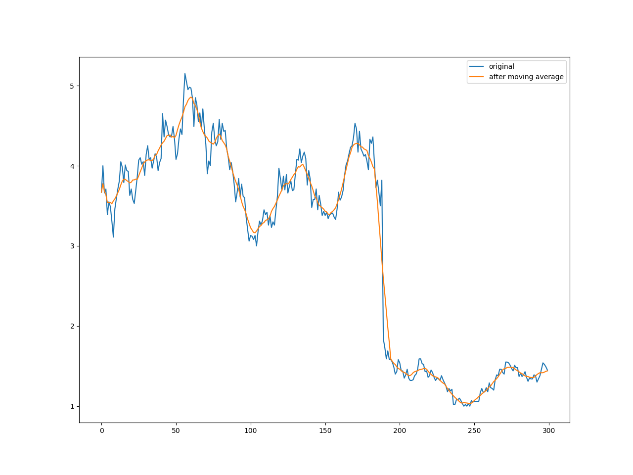
小波变换是一种数据分析技术，其中数据集被转换为小波域，并然后分成低频和高频部分。认为大部分数据的噪声都隐藏在小波域的高频部分。通过去噪程序，可以隔离和消除几乎每只股票中存在的许多不希望出现的噪声，所有这些噪声都有助于股票的波动性，但不一定表示股票的特征。

下面的例子展示了小波变换和滑动平均对数据集的去除一个标准差以外的数据结果以及使用新的系数生成平滑后的数据结果。我们可以观察到，滑动平均去噪后的股票与小波去噪后的股票相比更平滑，但滑动平均去噪后的股票过度平滑。小波去噪生成的数据误差明显较小，提高了预测精度。

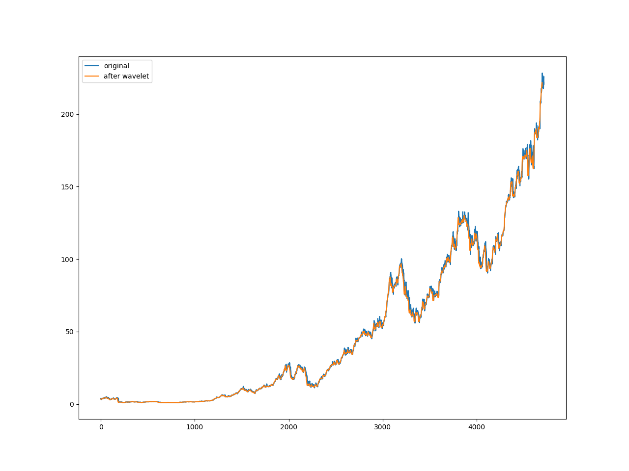
小波变换对数据集的去除一个标准差以外的数据结果：



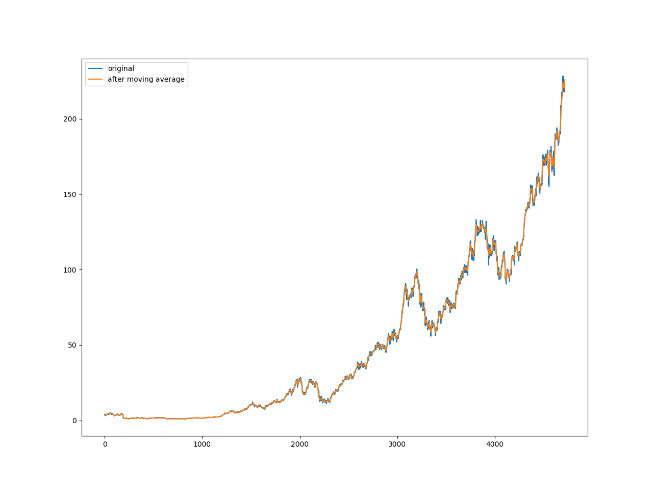
滑动平均对数据集的去除一个标准差以外的数据结果：



小波变换的使用新的系数生成平滑后的数据结果



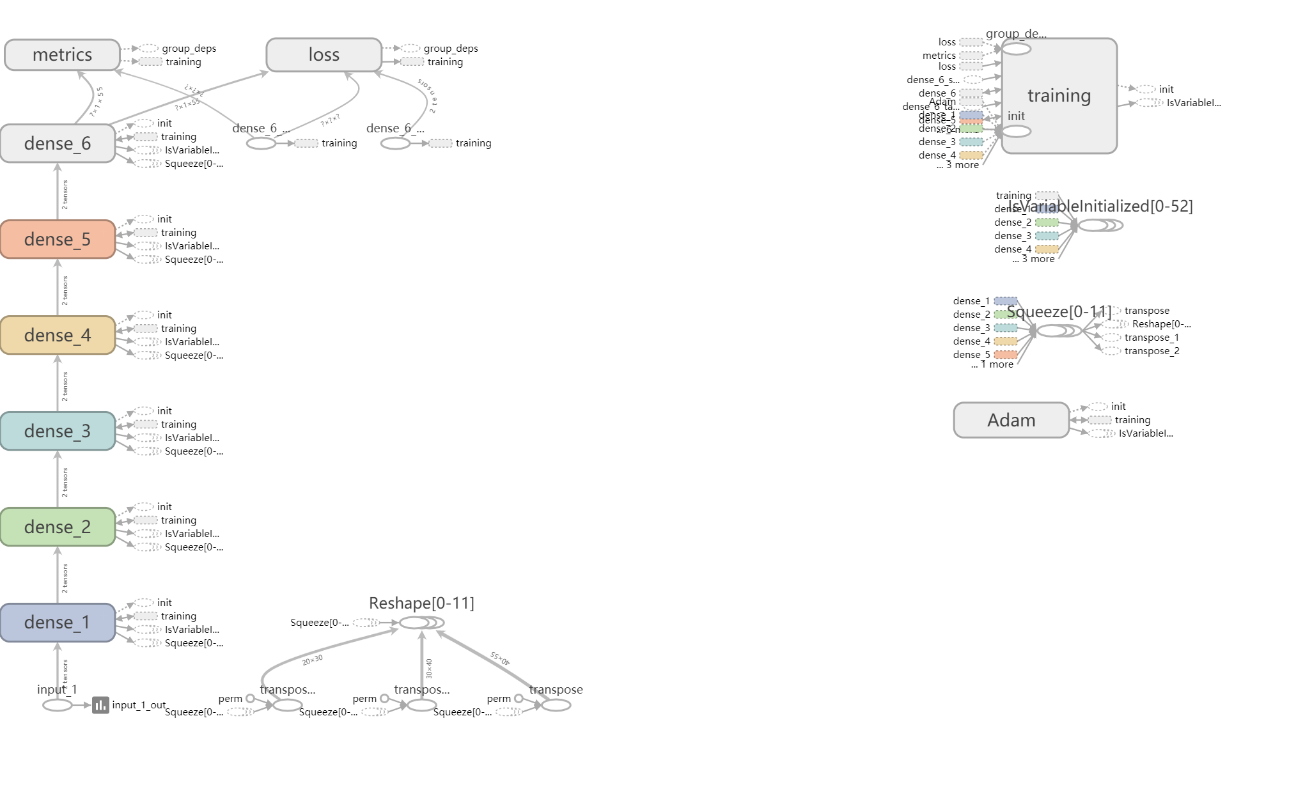
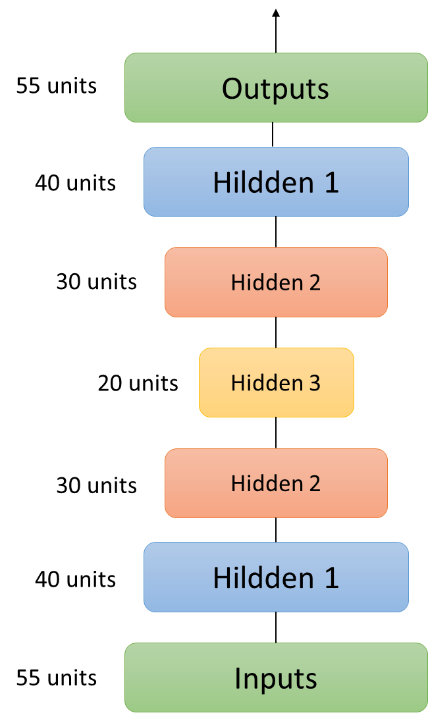
滑动平均的使用新的系数生成平滑后的数据结果：



最终我们用数据的前1/4作为自编码器的训练数据，后3/4的前80%为LSTM训练数据，后20%为测试集，使用相邻之间股价差异的对数值作为训练和测试数据。

**3、用栈式自编码器来提取特征**

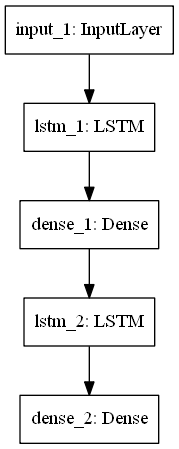
通常在机器学习中，往往需要利用领域知识来构建特征提取的方法，也可以使用受限玻尔兹曼机来自动提取特征。此项目采用的是栈式自编码器，由多个自编码器（自编码器是一种能够通过无监督学习，学到输入数据高效表示的人工神经网络。）堆叠而成，主要是考虑到栈式自编码器的可解释性更好。栈式自编码器把数据先从高维压缩到低维、然后再扩展到高维，这个过程的压缩部分就是很有效的一种特征提取的方法。此项目采用的栈式自编码器由3个自编码器级联构成。



**4、用这些特征来训练LSTM模型**

LSTM是RNN的特殊类型，RNN由于梯度消失的原因只能有短期记忆，LSTM由于各种结构的cell和gate的存在将短期记忆与长期记忆结合起来，可以有效的捕捉相隔很远的两个node之间的关系，从而在时间序列的预测问题上效果非常好。

模型结构：

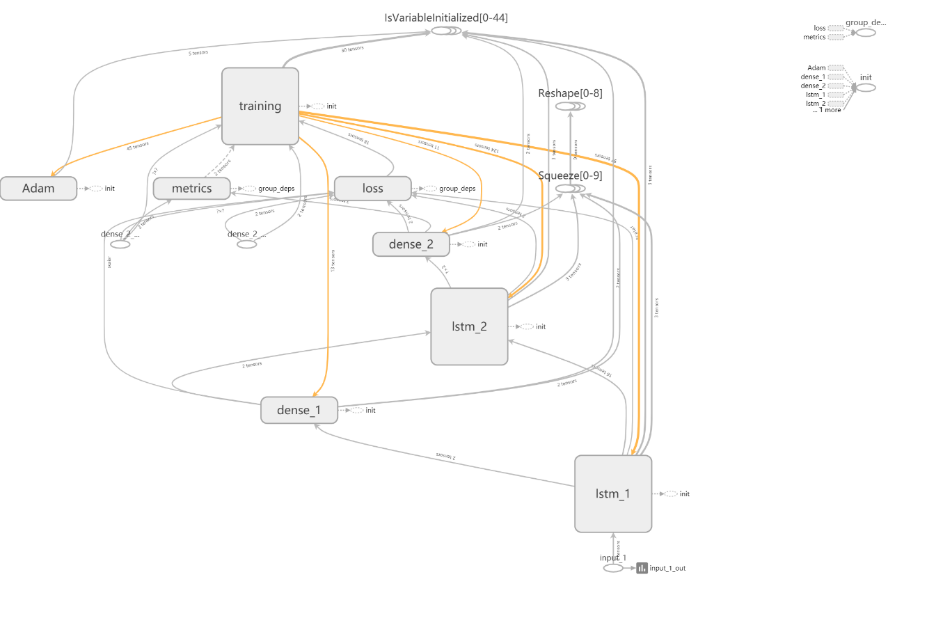
输入：20维特征值

第一层LSTM输出5维，dense输出5维，第二层LSTM输出2维，dense输出最终第二日收盘价差值的对数值

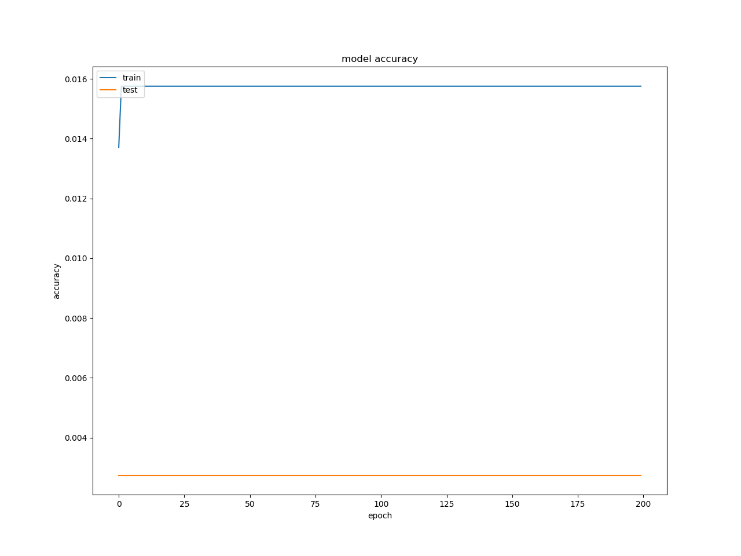
每层采用L2正则化和drop\_out

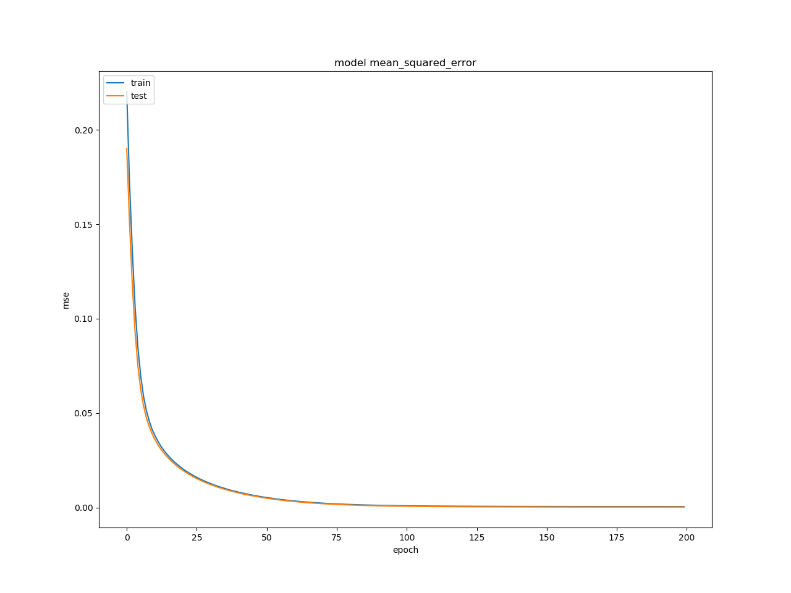
优化方法为adam，动态调整学习率和梯度

采用mean\_square\_loss作为损失函数

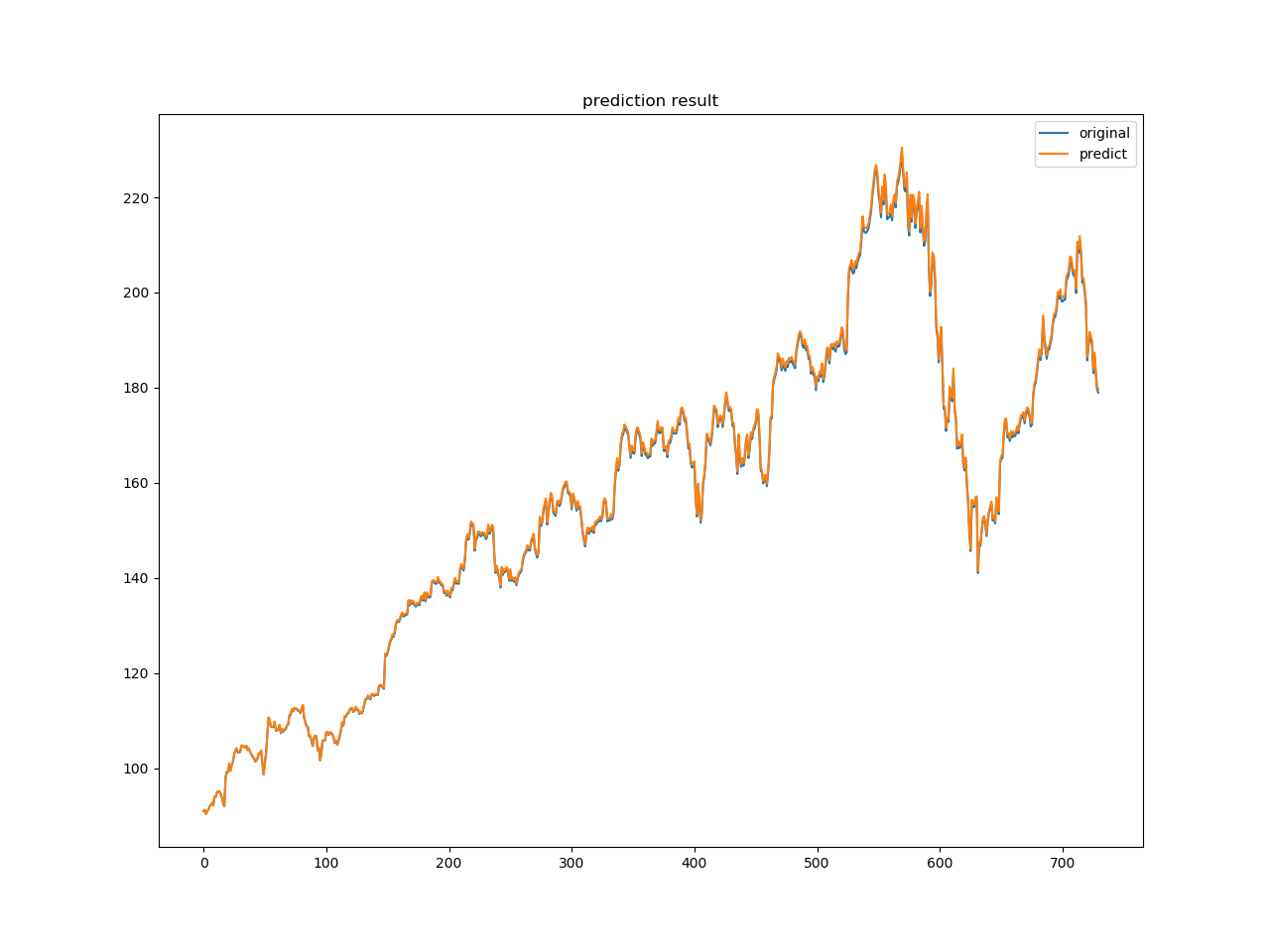
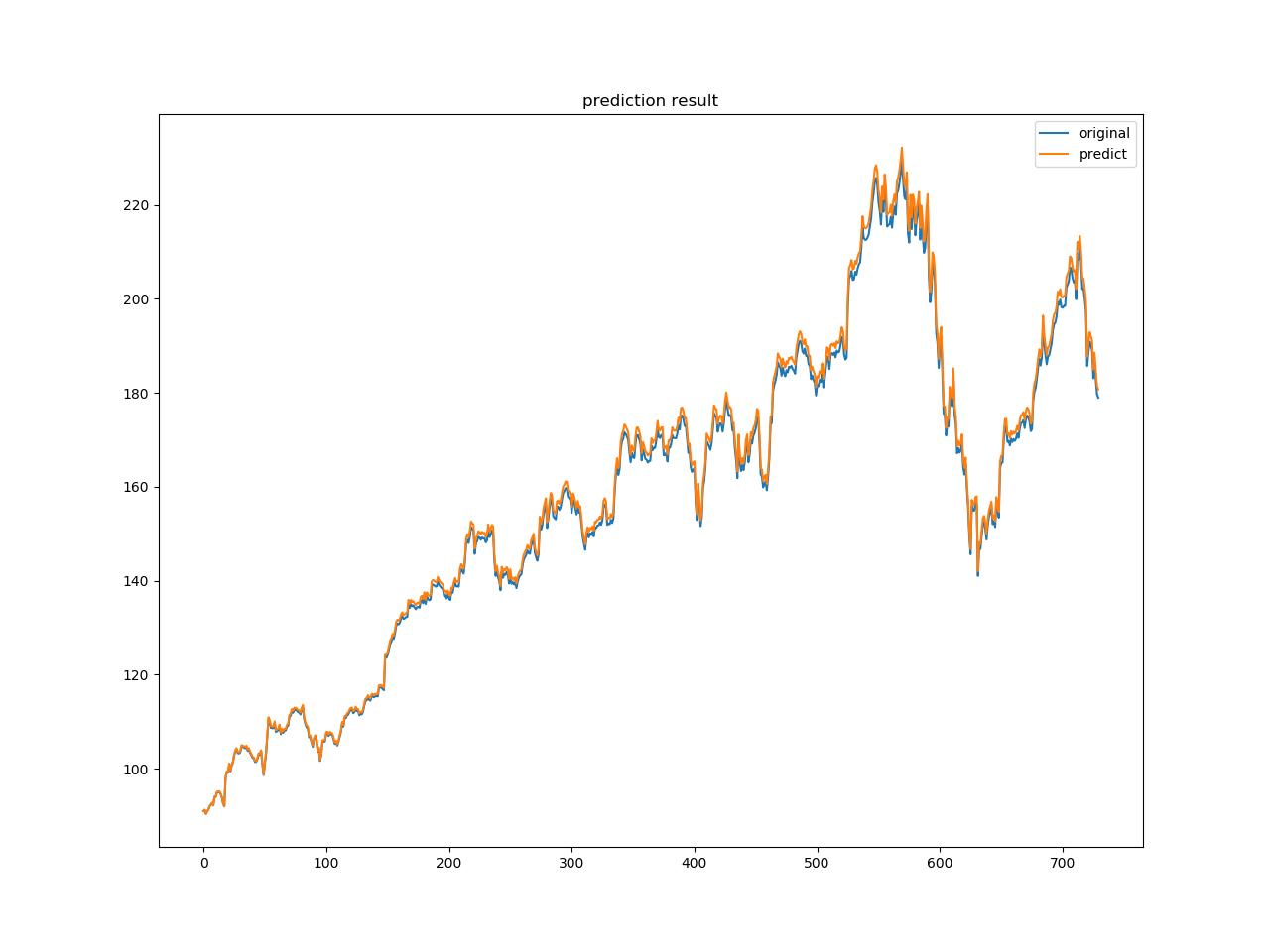
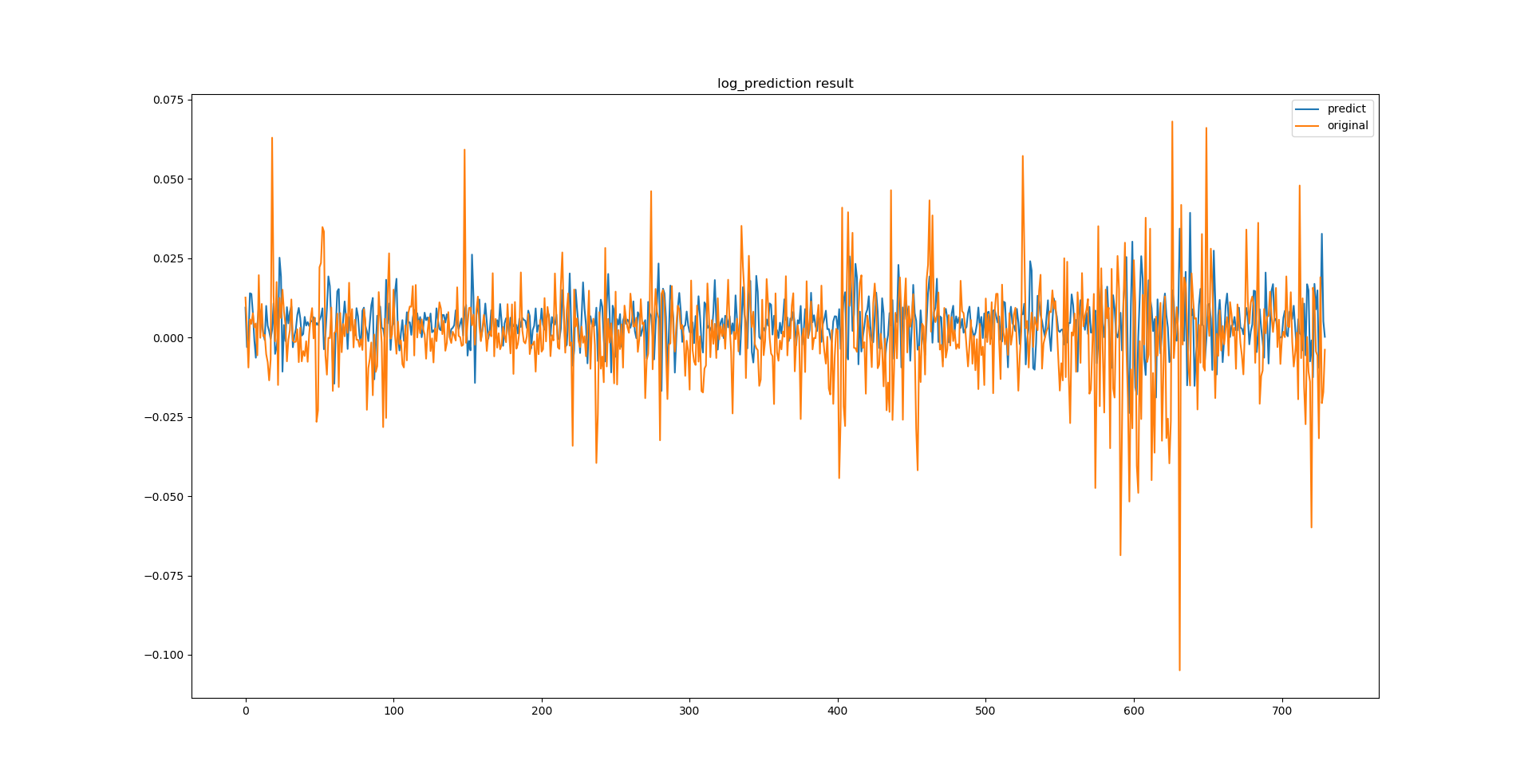


**5、训练**

****

****

**6、结果**

****