About me:

联系方式：81363132@qq.com

以下为：系统设计依据的原理及实验结果分析部分

第二章 时间序列与神经网络

本章重点论述有关股票价格预测的基础——时间序列与神经网络。

时间序列（或称动态数列）是指将同一统计指标的数值按其发生的时间先后顺序排列而成的数列。时间序列分析是根据已有的时间序列数据，通过曲线拟合和参数估计来建立数学模型的理论和方法。[1]经济数据中大多数以时间序列的形式给出。

神经网络（Neural Network）20 世纪 80 年代以来人工智能领域兴起的研究热点，它是一种模仿动物神经网络行为特征，进行分布式并行信息处理的算法数学模型。神经网络有着优秀的非线性拟合能力和学习能力。所以它十分适合处理股票数据这类金融数据[2]。

2.1 时间序列

时间序列有着以下几个重要的特性，这些特性都在一定程度上帮助我们去发现和解决研究课题中的一些重要的问题。这些与时间序列相关的特性如下：

(1) 时间序列的趋势性

(2) 时间序列的随机波动性

(3) 时间序列的平稳性

时间序列的趋势性

时间序列都有着确定的或者随机的趋势。对于股票价格构成的时间序列而言，它的趋势本身是十分难以把握的。不同的趋势对我们在进行时间序列的预测和分析中的影响也不一样，因为我们预测的结果很大程度上是要符合其整体的趋势的。果时间序列时间序列的趋势是时间序列的宏观描述，下面将给出三幅与股票价格有关的图来具体阐述这一特性。

图 2-1 是平安银行 (000001.SZ)从 2015 年 1 月 5 日至 2019 年 4 月 30 日里共计 1053 天的每日最高价的形成的时间序列图。平安银行的股票价格始终在某一价格之间上下波动，这也暗示着这支股票价格的整体趋势是一种在定值间上下浮动的趋势。

图 2-2 是贵州茅台（600519.SH）从 2015 年 1 月 5 日至 2019 年 4 月 30 日里共计 1053 天的每日最高价的构成的时间序列图。不难看出贵州茅台的价格在四年间从 2015 年年初的 200 元涨到现在的 1000 元左右。它也暗示着这只股票的价格的整体趋势是一种增长的趋势。

图 2-3 是将两支股票的历史数据放在同一坐标图中展示的结果。从第三幅时间序列图的对比结果中可以更好的帮助我们看清时间序列的趋势。尤其在我们仔细观察图 2-3 后会发现相较与贵州茅台的股票价格而言平安银行的就显得异常平滑。在许多关于股票价格的预测文章中都会提及承认历史能够重演（股票价格的波动是有周期的）这一前提，这也是对时间序列关于趋势的一种肯定。只是这种趋势终归是历史的，我们可以用它来预测未来，但预测未必是准确的。因为时间序列还有另一个重要的特性——随机波动。

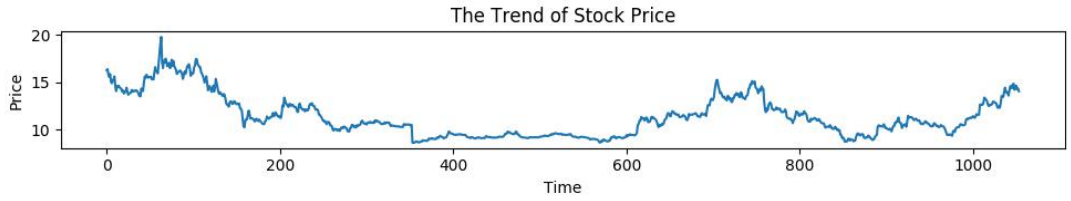

图 2-1 平安银行历史股价

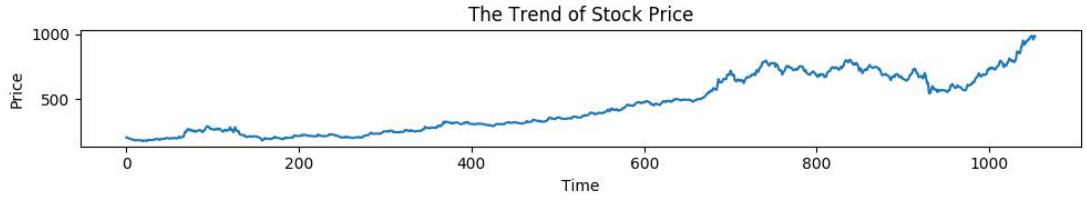

图 2-2 贵州茅台历史股价

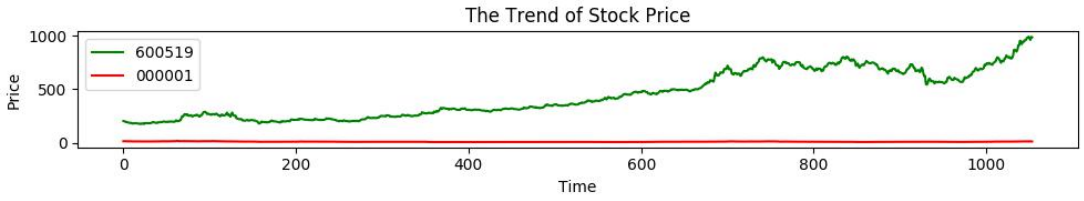

图 2-3 平安银行与贵州茅台的历史股价

时间序列的随机波动性

时间序列的随机波动是在时间序列预测中最大的随机变量。往往有很多因素会导致时间序列的随机波动，这些随机事件是极其难以预测的或者说是无法预测的。这时我们就不得不对时间序列进行处理，就像对信号去噪一样，通过一定的方法剔除时间序列里那些特别异常的影响（比如 2019 年 5 月，美国总统特朗普的关于中国关税的推特引发了中国股票市场不小的风波，沪深两市多只股票跌停）。至于为什么要进行这样的处理很容易理解，其一：这些特殊的数据不一定会周期行出现（特朗普也许不会每年五一劳动节之后都会发这样的一条推特）、其二：这些异常的数据对时间序列的整体稳定性是十分不友好的，他会影响时间

序列的整体趋势，从而影响我们对未来数据的预测。因此处理好这类特殊情况下的数据是十分有必要的。

尽管时间序列的随机波动性有着许多负面和不确定的影响，但正是这些随机波动性因素的存在，使得时间序列看起来更加有预测的必要性和研究的价值与意义。毕竟在这些随机波动中有一部分是很有价值的，如金融危机的定期来临，熟悉资本市场经济发展规律的人都知道资本主义会周期性产生金融危机。

在将时间序列的随机波动因素降低以后（类似信号处理中的去噪），我们就能看到时间序列的另一特性——平稳性。

时间序列的平稳性

时间序列的平稳性是时间序列分析的基础，只有基于平稳的时间序列的预测才是有效的。时间序列的平稳性是指时间序列的统计特征不随时间的推移而发生变化。直观的来说，就是时间序列无明显的上生或下降趋势，各观测值围绕某一固定值上下波动。

图 2-4 和图 2-5 分别是平安银行和贵州茅台的每日收益（这里的每日收益是今日的股票最高价减去前一日的最高价），从由股票的每日收益所构成的时间序列图的表现中可以看出，这种时间序列就是十分平稳的。它们都具有明显的平稳性。

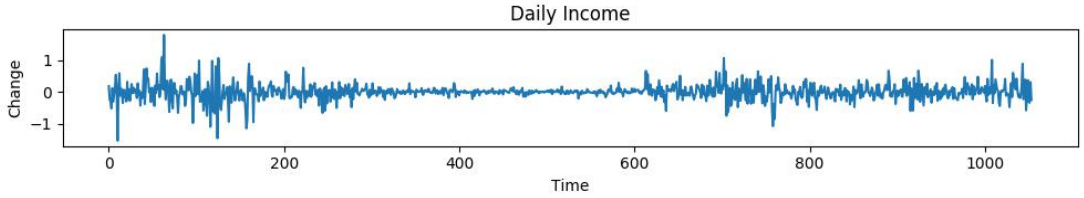

图 2-4 平安银行每日收益图

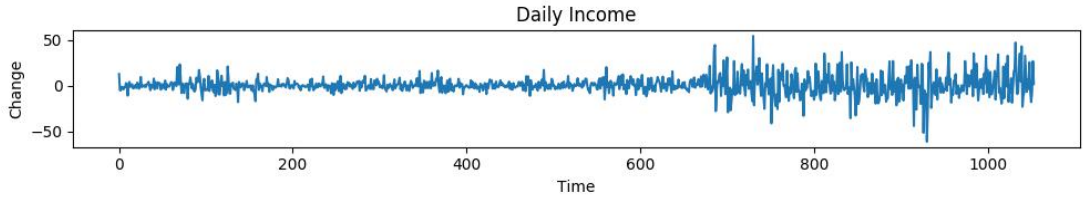

图 2-5 贵州茅台每日收益

两图中的观测值都围绕着某个固定的值上下波动，通过计算每日收益的平均值就可以估算出两只股票的每日收益组成的时间序列中所围绕的那个固定值的大小了。我们可以看到从 2015 年 1 月 5 日至 2019 年 4 月 30 日间平安银行的收益在 0.0034 上下[-2,2]的范围里波动。而贵州茅台的收益是在[0.7924]上下波动的。

表 2-1 两只股票的每日收益平均值

股票名称	平安银行	贵州茅台
每日收益平均值(元)	0.0034	0.7924

2.1.1 时间序列与股价数据

股票价格本身其实就是一种特殊的时间序列，为了更好的说明关于股票价格预测中遇到的一些问题和解决这些问题的方法，时间序列是不得不提的。就目前的研究而言，时间序列结合股票价格预测的问题中大致可以分为两类：回归问题（预测具体的价格和数据）与分类问题（预测价格的上涨或者下跌）。

为了更好的体现本次课题研究的重点与股票价格预测的实用性，预测未来一段时间内股票价格的最高价是最为合适不过的了。因为就投资者的目的而言，在得知股票的最高价格时抛售股票才能获得最大的收益，显然对比股票数据的其他几个特征（如收盘价、开盘价等）就显得更加有意义了。那么明天的股价是涨还是跌就显的不那么重要了。因此本课题的推进将着重针对股票最高价进行价格预测，希望通过预测最高价的只来为投资者带来最大化的收益。

时间序列分析与股价预测中有一些很经典的基于数理统计的应用，其中之一就是在时间序列分析中一个很简单的模型：移动平均模型。移动平均本质上就是利用时间序列的局部平稳性来进行预测的。移动平均方法进行时间序列预测的重点就是：将短时间的前一段历史数据的平均值视为预测结果。这一方法里短期内的异常值和短期波动相对来说是可控的——只要调整一下历史数据的个数（某一时间序列数据的个数）就能进行控制，从而避免那些波动较大的数据。

我们可以尝试着用这种方式来预测一下每日的收盘价格。表 2-2 给出了平安银行股票价格数据中的一部分数据（2019 年 4 月 18 日至 2019 年 4 月 29 日间的收盘价的 7 条交易记录）及部分处理数据（涨跌额、前五日涨跌额均值、预测价），其中涨跌额为当日收盘价减去前一日收盘价；前五日涨跌额均值为前五日涨跌额的平均值；预测价格为当日收盘价加前五日涨跌额均值。

表 2-2 平安银行七条数据及处理数据

序列编号	收盘价	涨跌额	前五日涨跌额均值	预测价
0 (20190430)	14.1	0.31	-0.188	13.912
1 (20190429)	13.79	-0.34	-0.042	13.748
2 (20190426)	14.13	-0.31	0.018	14.148
3 (20190425)	14.44	0.37	-0.102	14.338
4 (20190424)	14.07	-0.08	0.092	14.162
5 (20190423)	14.15	-0.58	0.262	14.412
6 (20190422)	14.73	0.39	0.16	14.89

我们可以通过图表来更加直观的观测预测结果，图 2-6 就是表 2-2 中的预测价格与真实价格的比较。从短期的预测结果来看，效果显得有点不足。图 2-7 是 150 条数据的预测结果与真实结果的比较。图 2-8 是 500 条数据的预测结果与真实结果的比较。

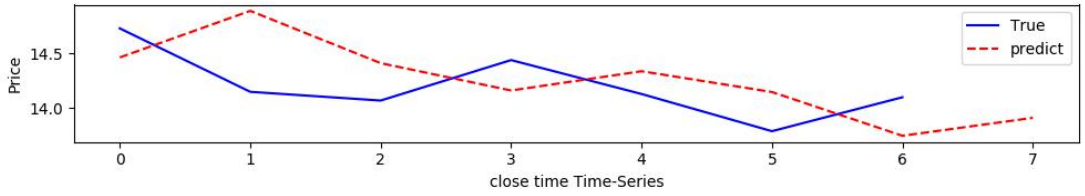

图 2-6 七条数据预测价格与真实价格对比

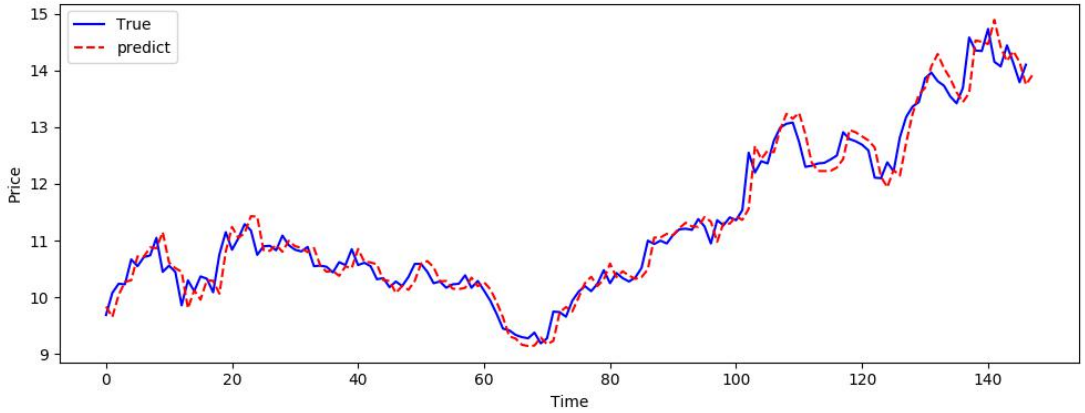

图 2-7 一百五十条数据预测价格与真实价格对比

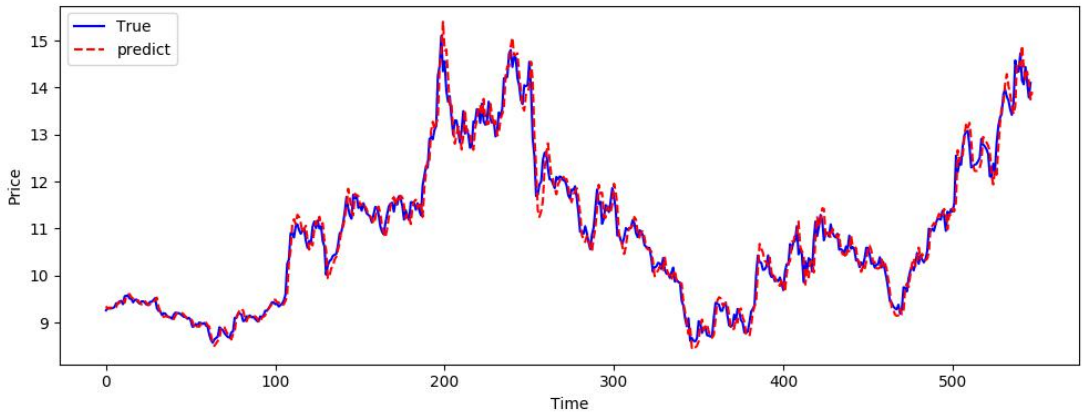

图 2-8 五百条数据预测价格与真实价格对比

仔细对比一下，不难发现，在数据量大的情况下观察，预测曲线与真实曲线的拟合程度也更加贴合。显然大部分情况下通过移动平均这种预测方法来进行时间序列的预测时，结果还是出乎意料的令人满意。

2.2 神经网络模型

人工神经网络是由人工建立的、以有向图为拓扑结构的动态系统，它通过对连续或断续的输入作为状态响应而进行信息处理。综合来源、特点和各种解释，神经网络可简单地表述为:人工神经网络是一种旨在模仿人脑结构及其功能的信

息处理系统。[3]

随着人工智能技术的不断运用与发展，许多神经网络模型也都被应用于股票预测的研究之中。基于目前研究热门的神经网络模型中，RNN（循环神经网络）是用于分析与预测时间序列中被提及最多的神经网络模型。RNN 是一类带有反馈回路的神经网络模型，很适合处理时间序列数据。

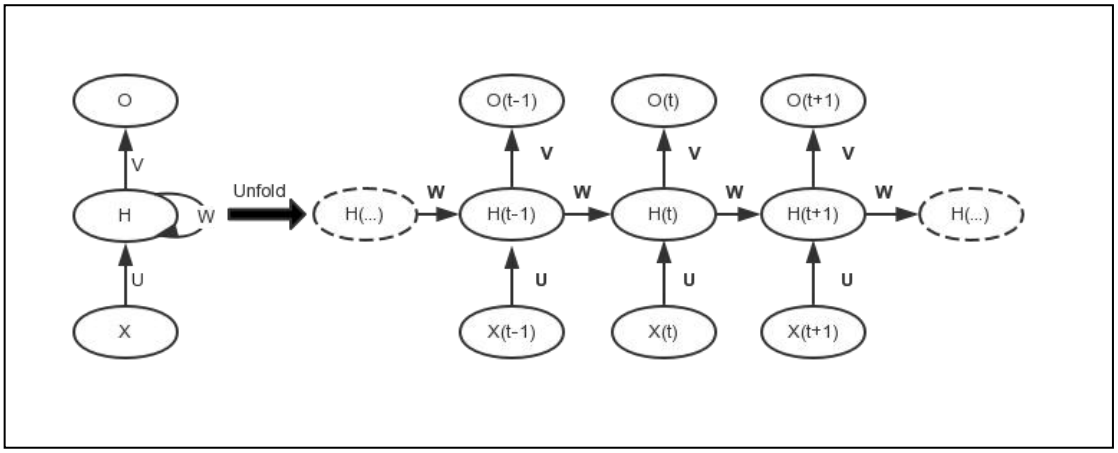

图 2-9 RNN 网络结构

图 2-9 是 RNN 模型按时间展开的图，在图中 X(t)是时间序列数据对应 t 时刻的数据样本的输入，H(t)表示的是隐藏层的节点向量，O(t)是训练样本序列的输出。U 表示输入层到隐藏层的权值矩阵，V 表示隐藏层到输出层的权值矩阵，W 表示反馈环路上的权值矩阵，这三个矩阵就是模型的线性关系参数。隐藏层 h 的值不仅取决于当前时刻的输入也取决与上一时刻的输出和（U，W，V）三个矩阵在网络中共享都反映了模型的"循环反馈"的思想。

循环网络的前向传播算法（对于 t 时刻）：

$$h^{(t)} = f(Ux^{(t)} + Wh^{(t-1)} + b)$$

式（2-1）

其中为激活函数，一般来说都会选择 tanh 函数，b 为偏置。则 t 时刻的输出：

$$o^{(t)} = Vh^{(t)} + c$$

式（2-2）

最终模型的预测输出为：

$$y^{(t)} = o^{(t)}$$

式（2-3）

RNN 本身也存在一定的局限性，因为随着时间序列长度（时间数据记录的条数）的增加，RNN 容易出现梯度消失的问题，“忘记”之前的状态信息。我们在对时间序列进行预测的时候，有着这样的一个既定前提——“历史可以重演”，因此所有的历史数据都是有用的。为了解决 RNN 梯度消失这一不足之处，有人引入了 LSTM（长短期记忆网络）和 ESN（回声状态网络）。

LSTM 通过刻意的设计来学习长期依赖信息并记住长期的信息。ESN 作为一种新型的人工神经网络，在时间序列预测上也有着不俗的表现，它本质上一个具有松散连接隐层的递归神经网络，从而巧妙的规避了 RNN 带来的梯度消失问题。

2.2.1 LSTM（长短期记忆网络）

LSTM（Long Short-Term Memory）即长短期记忆网络，LSTM 本身也是一种循环神经网络。但 LSTM 对 RNN 进行了改进，LSTM 与 RNN 最大的不同就是重复模块。如图 2-10 中所示，RNN 算法的重复模块中原本只有一层简单的结构（如 tanh 层），对比图 2-11 的 LSTM 网络结构，会发现 LSTM 有一个精心设计的判断信息是否有用的“门”结构——它是一种让信息选择式通过的方式，每一个 LSTM 细胞状态都有这样的三个门来保护和控制。

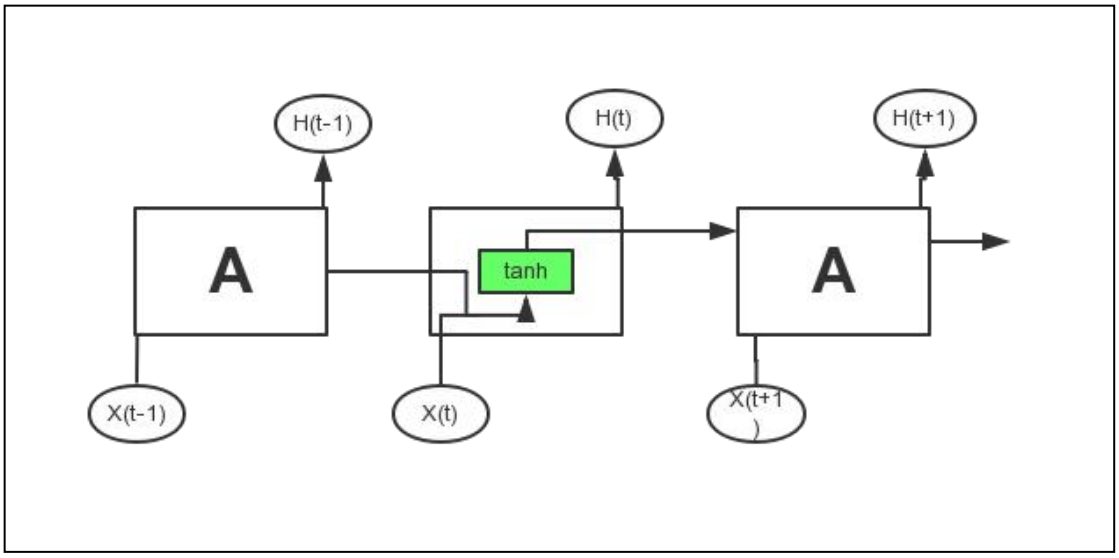

图 2-10 标准 RNN 重复模块中包含 tanh 层示意图

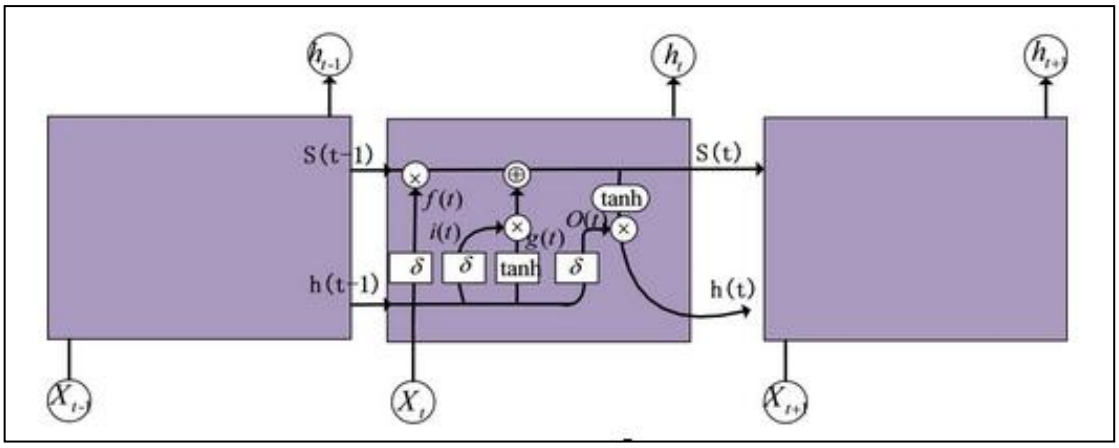

图 2-11 LSTM 重复模块中的四个层

LSTM 的神经元中有三个"门"，分别是：输入门、输出门、遗忘门。下面介绍各个控制们的计算原理：

首先计算输入门i_t的值和在 t 时刻输入细胞的候选状态值C_t，公式如下：

$$i_t = d(\mathrm{W}_i * (X_t, h_{t-1}) + b_i)$$

式（2-4）

$$C_t = \tanh(W_c * (X_t, h_{t-1}) + b_c)$$

式（2-5）

其次，计算在 t 时刻遗忘门的激活值f_t，公式如下：

$$f_t = d(W_f * (X_t, h_{t-1}) + b_f)$$

式（2-6）

由以上两步的计算，就可以计算出 t 时刻的细胞状态更新值c_t，公式如下：

$$c_t = i_t * C_t + f_t * c_{t-1}$$

式（2-7）

在计算得到细胞状态更新值后就可以计算输出门的值，其计算公式如下：

$$o_t = \sigma(W_o * (X_t, h_{t-1}) + b_o)$$

式（2-8）

$$h_t = o_t * \tanh(c_t)$$

式（2-9）

通过以上计算，LSTM 就可以有效的利用输入来使其具有长时期的记忆功能。

2.2.2 ESN（回声状态网络）

回声状态网络(Echo State Network)是 Jaeger 于 2001 年提出一种新型递归神经网络，ESN 通过随机部署大规模系数连接的神经元构成网络隐藏层——储备池。储备池中的神经元具有三个特点：储备池中神经元的连接状态随机；神经元的数目相对较多；储备池中神经元连接权重固定，不需要使用梯度下降进行权重更新。

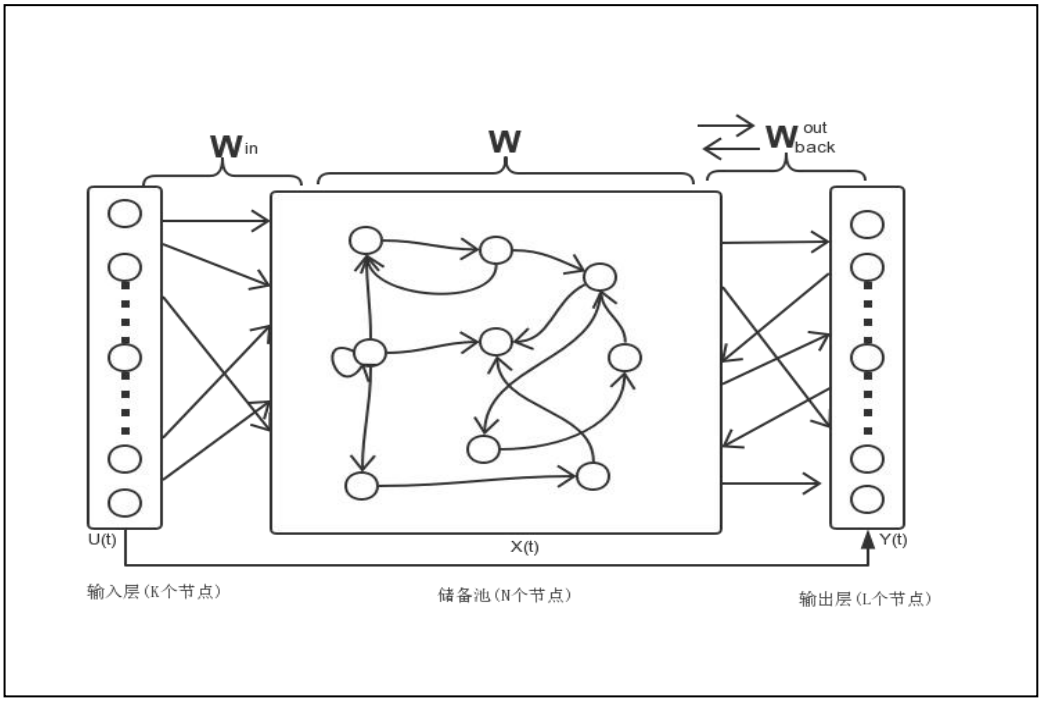

图 2-12 ESN 网络结构

ESN 的基本思想就是由储备池生成一个随输入不断变化的复杂动态空间，当这个状态空间足够复杂时，就可以利用这些内部状态，线性地组合出所需要的输出。ESN 网络结构有 3 层：输入层、隐含层和输出层。输入层节点数 K，储备池节点数 N，输出层节点数 L。

储备池内有许多稀疏神经元，这些些神经元暗含了系统的允许状态，具有短期记忆功能。输入层与储备池间的连接权值（W_{in}）和储备池内各个神经元间的连接权值（W）是随机初始化的，在网络训练的过程中不需要调整。ESN 网络训练过程，就是训练储备池（W_{out}）与输出层间连接权值（W_{back}）的过程，简单的线性回归可以完成对网络的训练过程。相关的连接矩阵和状态方程及有关计算如下：

当输入$u_{(t)}$时，储备池状态更新方程如下：（f 为激活函数 tanh）

$$x_{(t+1)} = f(W_{in} * u_{(t+1)} + W_{back} * x_{(t)}) \quad \text{式 (2-10)}$$

ESN 输出状态方程如下：（F 为输出层激活函数）

$$y_{(t+1)} = F[W_{out} * (u_{(t+1)}; x_{(t+1)})] \quad \text{式 (2-11)}$$

在 ESN 储备池的主要调节参数有谱半径$S_R = \max(|\lambda_1|, |\lambda_2| \ldots |\lambda_N|)$，其中的$\lambda_1, \lambda_2 \ldots \lambda_N$是 W 的特征值。$S_R < 1$ 时网络整体稳定。 式 (2-12)

2.3 股票价格预测

神经网络有许多应用，其中之一就是预测时间序列。一个很好的时间序列就是股票价格。股票价格预测一直都是一个十分具有研究意义的话题，尤其是随着近年来神经网络的兴起。神经网络作为模拟人类大脑突触结构进行信息处理的数学模型，它有着于区别数理统计等方法在股票价格预测上的独特优势。神经网络模型可以在训练的过程中根据预测误差不断的自动调整模型参数，从而去逼近理想的输出，具有很强的自适应能力，因此它适合股票价格预测这类复杂非线性问题。

然而，预测股票价格依然是十分困难的。根据我们前面提到过的时间序列的随机性可知，还有许多因素的影响是我们无法预知的。我们只能通过合理的方法去预测逼近那个真实的价格，想要做到真正的预测还是需要许多的努力的。

第三章 实验及结果分析

本章的主要内容是通过已经实现的神经网络来进行股票价格预测模型的设计与研究。结合上一章对时间序列和神经网络的介绍，本章主要将展开对具体的神经网络进行实验的结果和对结果的分析。

实验中涉及的一些主要步骤有：数据采集、数据的预处理、实验环境搭建及实验、神经网络调参和实验结果记录及分析。实验中用到的神经网络框架为TensorFlow——一个采用数据流图，用于数值计算的开源 Python 软件库。数据来源于 Tushare——Python 的一个开源财经数据接口。

实验中的神经网络模型主要有：单变量 LSTM（时间序列的输入是一维的）、多变量 LSTM（时间序列输入为多维）、ESN。对不同的神经网络模型进行调参进行记录，最后给出相应的评价并选择合适的神经网络模型。

3.1 数据来源及预处理

数据来源

本课题中所涉及的股票，主要来源于上海和深圳市场的沪深 300（CSI300）中 300 支股票中的部分。因为 CSI300 指数所涉及的股票反映的是流动性强和规模大的代表性股票，所以这类股票数据无论从市场的代表性还是股票数据本身都显得更好。所有数据均的来源方式都是通过 Tushare 获取并保存于本地服务器的。通过 Tushare 这类免费的财经数据接口，可以极大的减轻数据获取方面的工作量，如下图 3-1 所展示的是通过 Tushare 获取的关于平安银行（000001.SZ）的部分数据字段。

	trade_date	open	low	close	pre_close	change	high
0	20190430	13.99	13.59	13.85	14.1	-0.25	14.05
1	20190429	13.9	13.86	14.1	13.79	0.31	14.33
2	20190426	14.08	13.7	13.79	14.13	-0.34	14.25

图 3-1 平安银行 000001.SZ)部分字段部分示意

相关字段说明：trade_date(交易日期)、open(开盘价)、low(最低价)、close(收盘价)、pre_close(昨日收盘价)、change(涨跌额)、high(最高价)。这些字段是在股票价格预测的实验中常被提及的字段，而且也是股票交易中常被用来描述股票的具体字段。

其中获取的数据为各只股票 2015 年 1 月 5 日至 2019 年 4 月 30 日里共计 1053 天的历史数据，每天的历史数据都包含着以上几个数据字段的具体值。因为

Tushare 提供的数据相对完整不存在缺失数据，因此不需要做进一步的清洗和缺失值填充。但在获取到数据之后，后续还是需要对数据进行预处理的。我们通过 Tushare 获得的数据是以 CSV 文件的形式保存在本地的，因此处理数据的前提是将文件中的数据加载到 pandas（开源 Python 数据处理相关的函数模版）的数据框架中，然后在进行数据的预处理。下面将具体阐述：如何进行数据预处理以及进行预处理的原因。

数据预处理

在我们加载相关数据之后，将对数据进行预处理。这里主要是对数据进行标准化预处理，对数据进行标准化处理的目的主要是将预处理的数据限制在一定的范围之中，缩小过于偏离的数据值对时间序列导致的不良影响。这样做的主要原因是：当我们输入的时间序列分布并不标准或者变化幅度（标准差）过大时，就会减慢神经网络的学习和收敛速度，也会阻碍网络的学习效率。因此这里我们对时间序列数据进行标准化是十分必要的，采用的数据标准化方法为：z-score 标准化方法。其数学表达如下：

对序列 $x_1, x_2 \ldots x_n$ 进行标准化变换的：

$$y_i = \frac{x_i - \bar{x}}{s} \quad (\text{这里} \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}) \qquad \text{式（3-1）}$$

这种方法基于原始数据的均值（mean）x 吧和标准差（standard deviation）s 进行数据的标准化。将对应的 x 标准化为 y。结合前一章节中对时间序列平稳性的介绍，通过 z-score 标准化方法处理过的数据将更加平稳。

下图 3-2 是对平安银行（000001.SZ）中最高价（high）组成的时间序列数据进行标准化后的结果与原始数据的对比。我们可以从图中看出标准化并没有改变时间序列的趋势，但标准化后的数据范围从原来的[10,20]变为了[0,5]，使得时间序列数据整体上更加紧凑。

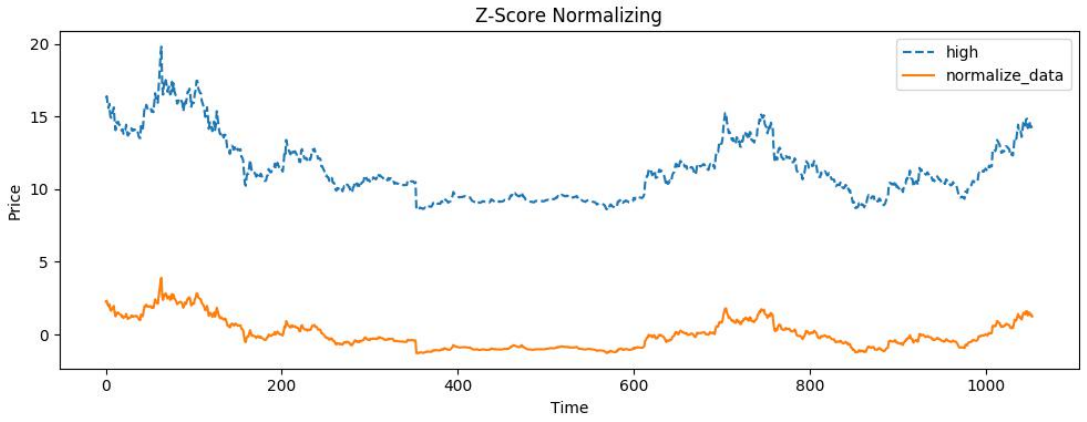

图 3-2 平安银行(000001.SZ)中最高价(high)标准化处理

其中平安银行（000001.SZ）最高价（high）构成的时间序列中，相关原始数

据的均值和标准差如表 3-1 所示。

表 3-1 最高价的均值和标准差

均值	11.37
标准差	2.16

在对数据进行处理完之后就可以运用神经网络模型来对股票价格进行预测了。当然多变量 LSTM 的数据标准化方法的思路是一样的，只不过需要处理的变量更多了。在我们数据进行完标准化后，就可以将标准化后的数据输入到神经网络中进行训练和预测了。最后当我们得到预测结果后，将结果逆标准化后就能得到我们所需要的预测结果了。可以这样做的目的，主要是因为标准化并不会影响原始的数据间的特征。

3.2 LSTM 预测股价

在通过 LSTM 进行股票价格预测的实验中，考虑到影响股票价格因素的多样性，因此针对数据的输入进行了具体的划分：单变量 LSTM（时间序列预测输入为一维）、多变量 LSTM（时间序列输入为多维）然后分开进行模型的训练与预测。通过具体的实验来进行预测和对实验结果进行分析。

3.2.1 单变量 LSTM

这一节主要介绍通过单一变量构成的时间序列来训练 LSTM 然后进行预测，所谓的单变量 LSTM 具体是指输入的时间序列数据为一维。结合 2.1.1 时间序列与股价预测这一章节中我们对股票最高价意义的说明，我们在单变量 LSTM 实验中主要就是对最高价这一变量进行预测，例如平安银行（000001.SZ）六个股票数据字段中的最高价（high）构成的时间序列来进行训练和预测的。以下我们将通过具体的步骤来进行详细介绍：

模型训练

这一部分主要是关于单变量 LSTM 进行时间序列预测股票最高价中的模型训练部分相关的内容。有关神经网络训练模型并进行预测的原理大致如图 3-3 所示，将完整的时间序列按时间步长划分这种划分类似于滑动窗口。

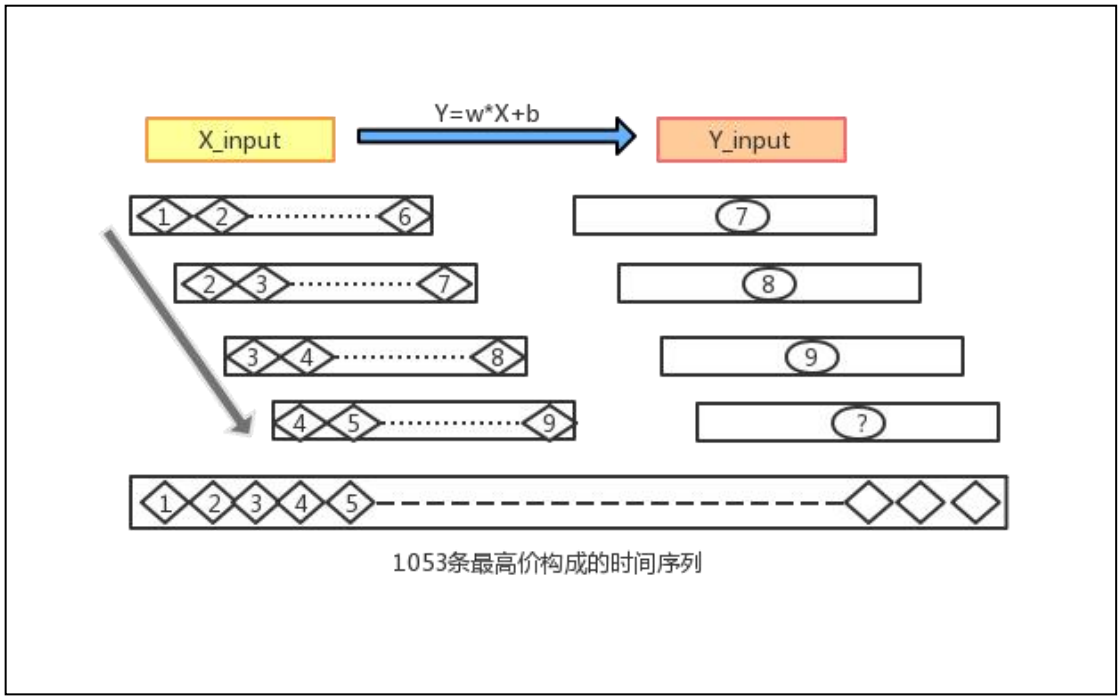

图 3-3 LSTM 预测的原理

图 3-3 中形象的说明了单变量 LSTM 进行时间序列预测时的原理。在经过标准化处理后的时间序列按一定的时间步长进行划分，然后就是类似于线性回归的训练过程那样，通过神经网络训练的主要目的就是求出 Y=w*X+b 中的 w（权重）和 b（偏置）并保存。当然这里的 w 和 b 并不是简单的线性回归中的数字。

结合 2.2.1 LSTM 中对 LSTM 的介绍，TensorFlow 便是以此为依据 LSTM 进行封装的。下图 3-4 是关于具体的神经网络模型的结构，以及通过 LSTM 进行训练和预测的 TensorFlow 计算图。TensorFlow 的计算图反映的就是相关代码在运行过程中的数据流向图。

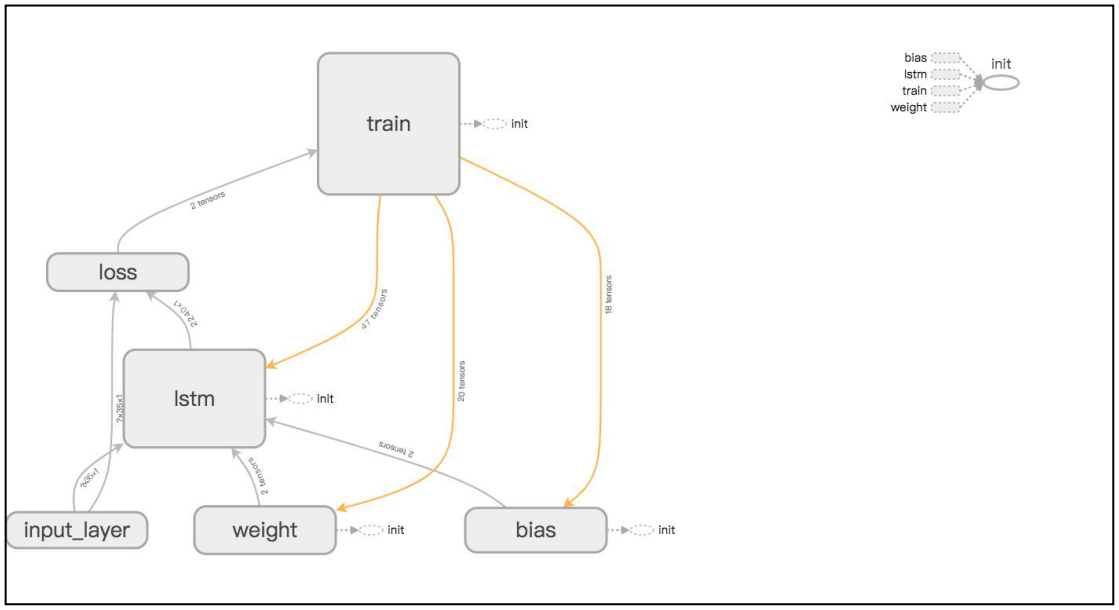

图 3-4 单变量 LSTM 的 TensorFlow 计算图

结果评价

在进行结果分析前，我们预测了收盘价（close）最后七天的结果与图 2-6 用传统移动平均预测的结果进行对比。图 3-5 上半部份的图片为通过移动平均进行预测的结果，图 3-5 下半部份的图片为通过 LSTM 神经网络训练后的模型进行预测得到的预测结果。从预测结果的对比可以看出通过 LSTM 神经网络预测的结果和历史数据 (真实数据) 更加贴合、拟合的效果更好，充分说明了 LSTM 预测的结果是好于传统的通过数理统计来进行预测的。

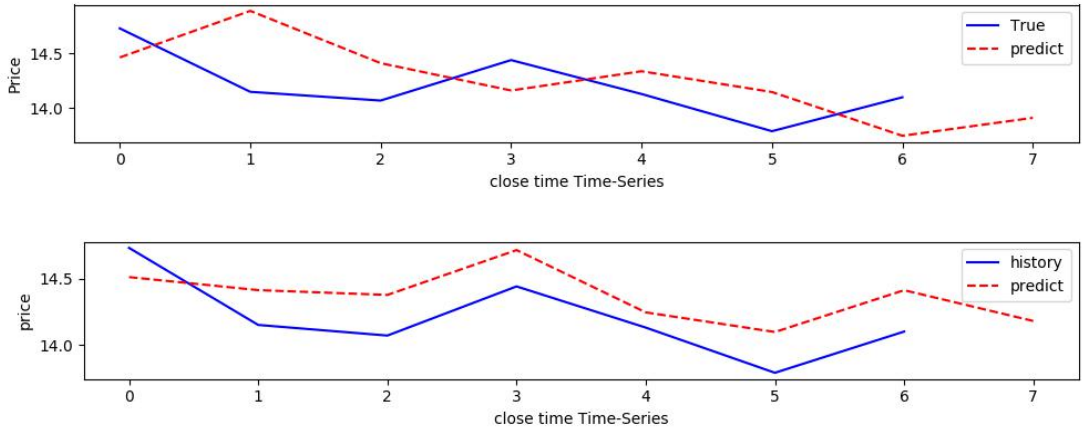

图 3-5 收盘价用移动平均(上)和 LSTM(下)预测结果对比

结果分析部分主要是对预测的结果进行分析，并记录一些超参数调参的过程和对实验结果的影响，在最后得出最优的预测结果时参数的具体值。这里我们还需额外花费一些篇幅来介绍一下关于评判训练后模型结果好坏的衡量指标 MAE（Mean Absolute Error）值——平均绝对误差值,后面几个模型都将以此为评价标准。MAE 的具体计算方法如下：

$$\mathrm{MAE}(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1} |y_i - \hat{y}_i| \qquad \text{式 (3-2)}$$

其中，y_i 为真实（历史）数据 $\hat{y}_i$ 为预测数据，MAE 的值越小表明预测的数据与原始数据之间的绝对误差越小，说明预测的结果也就越好。

参数调整

在单变量 LSTM 预测股票最高价中，我们需要调整的参数主要有：时间步长、每一批次训练样本数、训练次数。按照传统的控制变量法，衡量不同参数对 LSTM

神经网络的影响。如时间步长的影响中唯一变动的参数就只有时间步长，保持其它因素不变，并记录相对应的 MAE。

（1）时间步长的影响

表中的取余值为：时间序列的长度对时间步长进行模运算的结果。此时对应固定值的的变量：每一批次训练样本数为 64、训练次数为 100。

表 3-2 时间步长对 MAE 值的影响

时间步长	3	4	5	6	7	8
MAE 值	0.1616	0.1595	0.1627	0.1599	0.1819	0.1601
取余值	1	2	4	4	4	6

通过观察实验记录的数据可以发现当时间步长为偶数时对应的 MAE 值相对于时间步长为奇数时要小。而实验中时间序列的长度为 1054，而且通过对比取余值会发现取余值越小的数字，说明用于训练的数据样本越多被抛弃的数据越小，预测的效果也就更好。

（2）每一批次训练样本数的影响

此时的时间步长为：4、训练次数为：100。

表 3-3 批次大小对 MAE 值影响

批次大小	4	16	32	64	96
MAE 值	0.2013	0.1729	0.1610	0.1600	0.1616

此外批次大小对训练的损失率影响也很大，下图 3-6 中左侧为批次大小为 32 时的损失率图像，右侧为批次大小为 64 时的损失率图像。

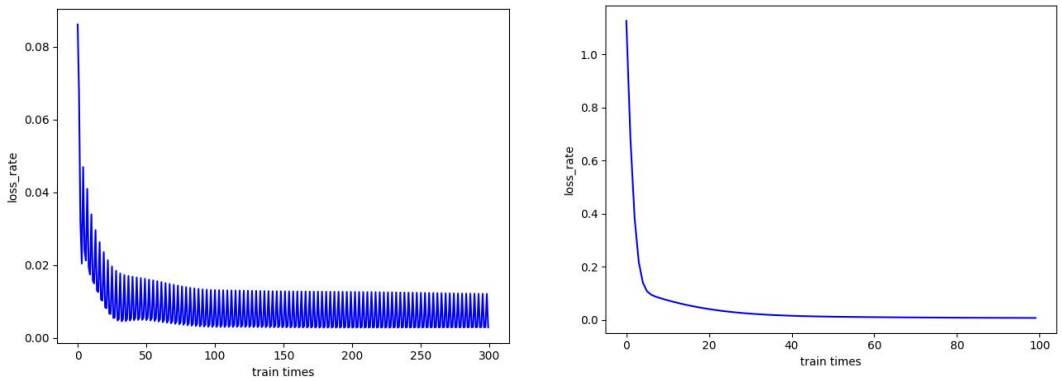

图 3-6 批次大小为 32（左）和 64（右）对于的损失率图像

其实就对应的训练批次大小而言，批次大一些的收敛得快，需要训练的次数相对也少一些，准确率上升得也比较稳定。

（3）训练次数的影响

此时的时间步长为：4、此时的每一批次训练样本数为：64。

表 3-4 训练次数对 MAE 值的影响

训练次数	100	500	1000
MAE 值	0.1636	0.1635	0.1580

从表中的数据可以看出训练次数对 MAE 值的影响。训练的次数越多预测的结果也越好但同时其费时也越久。

最后综合以上 3 个参数的调整记录，可以得出相对较好的结果应设置参数：时间步长：4、批次大小：64、训练次数：1000。下图 3-7 就是历史数据与预测数据拟合的结果，可以看到预测的数据与历史数据几乎重合。而且此时的 MAE 值是 0.158，可以认为预测值与真实值的偏差为±0.158。平安银行的股票价格大致为 14 元，那么偏差大致为 1.12%，准确率为 98.88%。

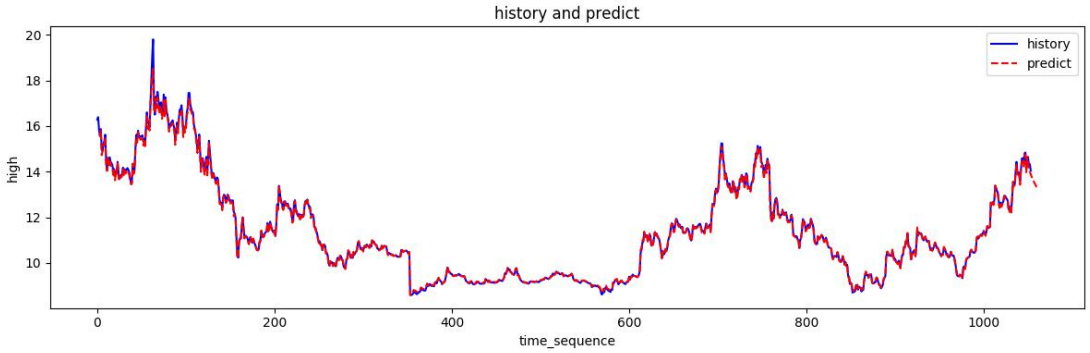

图 3-7 单变量 LSTM 的历史数据与预测数据

3.2.2 多变量 LSTM

这一节主要介绍通过多变量构成的时间序列来进行训练和预测，所谓的多变量 LSTM 具体是指输入的时间序列的数据是多维的。例如平安银行（000001.SZ）六个股票数据字段中的除最高价（high）外的 open(开盘价)、low(最低价)、close(收盘价)、pre_close(昨日收盘价)、change(涨跌额)五个变量构成的时间序列来进行训练和预测。

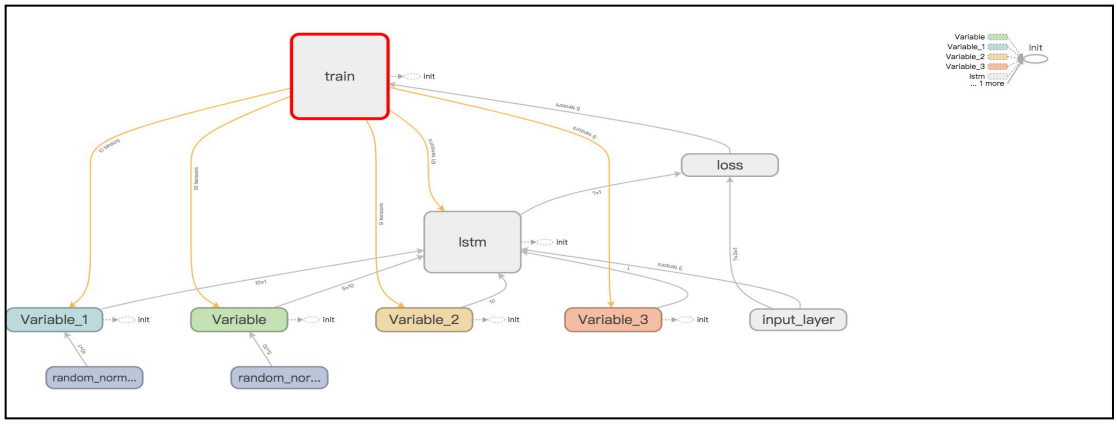

图 3-8 多变量 LSTM 的 TensorFlow 计算图

有关预测的原理和图 3-3 中的预测原理大致相似，只不过多变量 LSTM 的输入时间序列的维度为五维不再是一维的，同时有关的 TensorFlow 计算图也不一样如上图 3-8。多变量 LSTM 输入数据的预处理也是用的 z-score 标准化方法，需要说明的是这里是对五个变量都进行了标准化。

参数调整

在多变量 LSTM 预测股票最高价中，我们需要调整的参数主要有：时间步长、每一批次训练样本数、训练次数。依旧采用传统的控制变量法来进行实验，衡量不同参数对 LSTM 神经网络的影响。

（1）时间步长的影响

此时的每一批次训练样本数为：64、训练次数为：500。

表 3-5 时间步长对 MAE 值的影响

时间步长	3	4	5	6	7	8
MAE 值	0.0695	0.0758	0.1196	0.0705	0.0803	0.0767

对比单变量 LSTM 的实验结果会发现多变量 LSTM 的 MAE 值整体更小。其实就 MAE 值而言，不同的时间步长之间对应的 MAE 值变化其实已经很小了。

（2）每一批次训练样本数的影响

此时的时间步长为：3、训练次数为：500。

表 3-6 批次大小对 MAE 值影响

批次大小	16	32	64	96
MAE 值	0.0763	0.0972	0.0695	0.0792

同样在多变量 LSTM 的实验中 MAE 值都是整体偏小的。

（3）训练次数的影响

此时的时间步长为：3、此时的每一批次训练样本数为：64。

表 3-7 训练次数对 MAE 值的影响

训练次数	100	500	1000
MAE 值	3.1491	0.0695	0.0774

从表中的数据可以看出训练次数对 MAE 值的影响。从结果可以看出对于多变量 LSTM 中训练次数很少时，影响是很大的。

最后综合以上 3 个参数的调整记录，可以得出相对较好的结果应设置参数：时间步长：3、批次大小：64、训练次数：500。下图 3-9 就是历史数据与预测数据拟合的结果，可以看到预测的数据与历史数据几乎重合。而且此时的 MAE 值是 0.0695，可以认为预测值与真实值的偏差为±0.0695。平安银行的股票价格大致为 14 元，那么偏差大致为 0.49%，准确率为 99.51%。

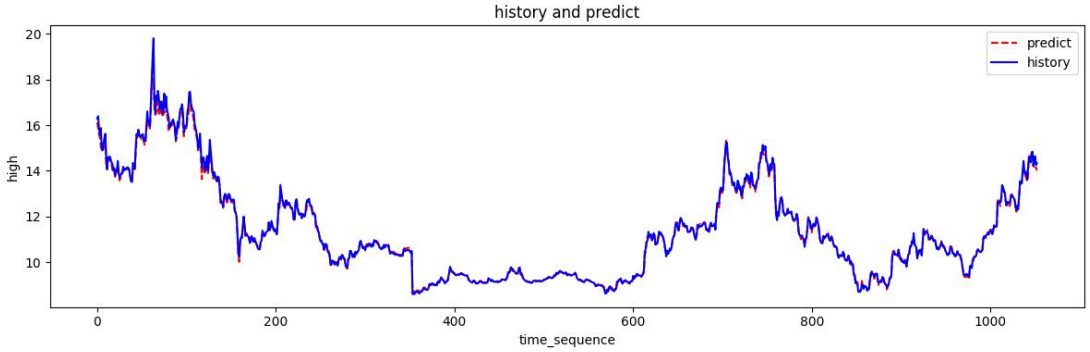

图 3-9 多变量 LSTM 历史数据与预测数据

3.3 ESN 预测股价

这一节主要介绍 ESN（回声状态网络），作为另一种不同类型的神经网络。ESN 区别于其他神经的主要原因在于：ESN 在不同神经元间的随机连接，它不是层之间整齐连接的，因此 ESN 的训练方式也不一样。我们只需要训练网络的输出权重，这就加快了神经网络的训练速度，提供了更好的预测。ESN 的训练速度快，不存在分支，易于实现。

模型训练

在使用 ESN 预测未来股票价格最高价的实验中，ESN 模型训练和预测的原理大致如下图 3-10 ESN 预测原理所示。我们进行预测时使用前面的 997 个数据点来预测未来两天的数据。如此，滑动的窗口长为 999，最后 2 个数据作为验证。

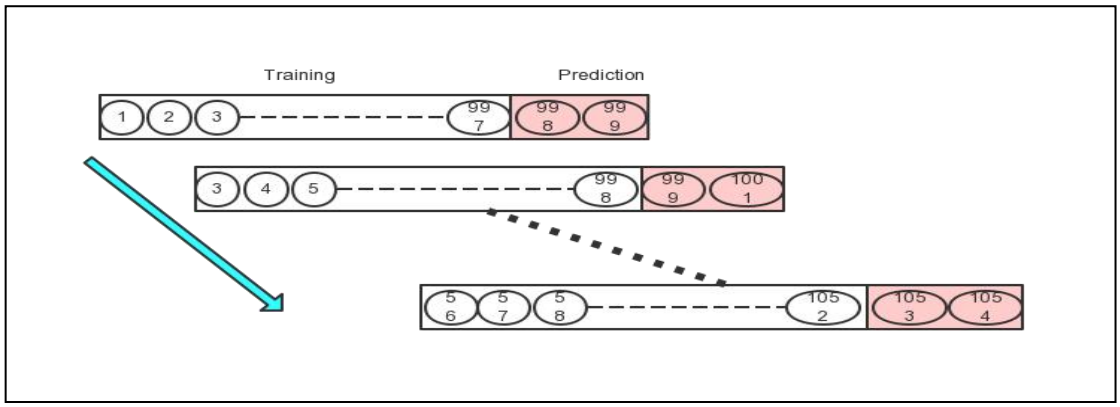

图 3-10 ESN 预测原理

ESN 模型预测结果的衡量我们还是采用前面介绍过的 MAE(平均绝对误差)，关于 ESN 相关的参数主要有：谱半径、噪声和窗口长度。噪声的目的主要是为了防止过拟合。

参数调整

有关参数说明：R：谱半径、N：噪声、M：MAE 值。

(1) 当窗口长度为 1000 时：

表 3-8 窗口长度为 1000 时各参数组合对应的 MEA 值

R	0.9			1			1.1		
N	0.001	0.004	0.006	0.001	0.004	0.006	0.001	0.004	0.006
M	0.392	0.403	0.409	0.391	0.405	0.409	0.394	0.407	0.413

(2) 当窗口长度为 1050 时：

表 3-9 窗口长度为 1050 时各参数组合对应的 MEA 值

R	0.9			1			1.1		
N	0.001	0.004	0.006	0.001	0.004	0.006	0.001	0.004	0.006
M	0.221	0.249	0.256	0.224	0.244	0.251	0.226	0.238	0.242

从实验结果的数据中可以看出，窗口长度对 MAE 值的影响较大。而谱半径和噪声对预测结果的影响就相对小了很多。

最后综合以上 3 个参数的调整记录，可以得出相对较好的结果应设置参数：

窗口长度：1050、噪声：0.001、谱半径：0.9。下图 3-11 就是历史数据与预测数据拟合的结果，可以看到预测的数据与历史数据几乎重合。而且此时的 MAE 值是 0.221，可以认为预测值与真实值的偏差为±0.221。平安银行的股票价格大致为 14 元，那么偏差大致为 1.57%，准确率为 98.43%。回声状态网络分析时间序列的能力和金融预测的数据是高度非线性的，因此 ESN 可以称得上是一个不错的工具的。

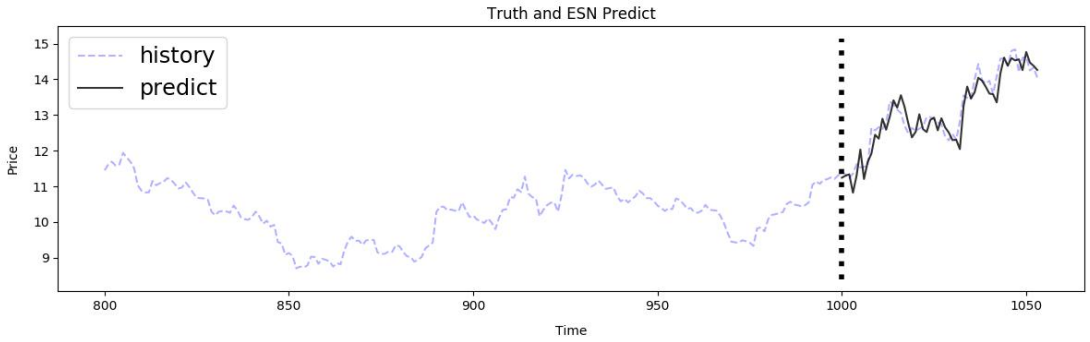

图 3-11 ESN 历史数据与预测数据

实验总结

最后我们在获得对应的网络模型参数的最优化结构也考虑验证了时间花销问题，分别单独的对三个神经网络的训练和预测进行计时。ESN 训练预测一次花费的时间为：26s，时间花费相对较小但 MAE 值较大；单变量 LSTM 训练预测一次花费的时间为：45s，时间花费相对适中 MAE 值适中；多变量 LSTM 训练预测一次花费的时间为：49s，时间花费较大但 MAE 值最小。

表 3-10 不同神经网络的耗时表

神经网络模型	单变量 LSTM	多变量 LSTM	ESN
耗时（s）	45	49	26
MAE 值	0.158	0.0695	0.221

因此就以上结论，当我们在设计股票价格预测模型时，如果要实现分钟级别的预测应该采用 ESN 模型，如果对时间要求不是十分严苛的如日级别的预测可

以采用多变量 LSTM 模型，对于小时级别的预测则可以采用多变量 LSTM 模型。