

■ 长短时记忆算法(LSTM)介绍

LSTM 是一种改进后的循环神经网络，可以解决 RNN 无法处理长距离依赖的问题，引入了一种计算单元作为记忆单元，取代了网络隐藏层中的传统人工神经元。利用这些记忆单元，网络能够有效地将记忆关联起来，并能及时远程输入，从而适应随时间动态掌握数据结构，较好的解决了时间序列的预测问题。

■ 长短时记忆算法(LSTM)原理

神经网络的训练就是根据目标提取特征设计好神经网络的结构后，输入数据对模型进行计算。首先对网络结构中的各权重和偏置项设置一个初始值，选择好一个激活函数，设置一个适当小的正数作为学习率，然后把训练数据输入模型通过激活函数进行正向计算，得到输出层的输出值，然后使用误差反向传播算法，将输出值、实际值代入到误差方程中，求出各层神经节点的误差项，再根据误差项计算平方误差关于权重和偏置的偏导数，然后根据梯度下降算法求出代价函数和它的梯度，最后用梯度对权重和偏置项的值进行更新，这就完成模型的一次训练。

■ 长短时记忆算法(LSTM)应用

LSTM 在语音识别、图片描述、自然语言处理等许多领域中得到较好的应用，但在金融领域中，还处于探索阶段，随着神经网络理论本身以及相关技术的不断发展，神经网络的应用将更加深入。

■ 风险提示

机器学习量化策略的结果是对历史经验的总结，存在失效的可能。

崇盛棠 高级分析师

投资咨询编号：Z0010966

从业资格编号：F0257938

邮箱：chongst@shhxqh.com

田景山 数据分析师

从业资格编号：F3045626

邮箱：tianjs@shhxqh.com

施杨 分析师助理

从业资格编号：F3057470

邮箱：shiyang@shhxqh.com

关于我们：

客服电话：400-186-8822

公司地址：上海市黄浦区福州路 666 号
华鑫海欣大厦 21、22 楼

公司官网：<http://www.shhxqh.com>



内容目录:

1、神经网络	3
1.1 神经网络原理	3
1.2 神经网络结构	3
1.3 神经元工作数学表达	4
2、激活函数	4
2.1 激活函数	4
2.1.1 激活函数的作用	4
2.1.2 三种常用激活函数	4
2.1.2.1 Sigmoid 函数	4
2.1.2.2 Tanh 函数	5
2.1.2.3 ReLU 函数	6
3、输出过程	6
4、神经网络优化	7
4.1 代价函数介绍	7
4.1.1 代价函数原理	7
4.1.2 代价函数数学表达	7
4.2 误差反向传播法 (BP 法)	8
4.2.1 梯度下降法	8
4.2.2 误差反向传播法	8
4.3 小结	9
5、LSTM 模型对期货价格预测的实证分析(5 分钟、1 小时、1 日数据)	9
5.1 数据预处理	9
5.1.1 数据归一化	9
5.1.2 时间序列转换成监督学习	10
5.1.3 数据集划分	10
5.2 数据来源与特征项	10
5.2.1 单变量长短时记忆神经网络模型 LSTM	10
5.2.1.1 长短时记忆网络	10
5.2.1.2 Adam 优化器	11
5.2.1.3 训练结果	11
5.2.2 多变量长短时记忆神经网络模型 LSTM	13
6、交易策略	15
6.1 策略描述	15
6.2 策略实现	15

图表目录：

图一：神经网络工作图-----	3
图二：Sigmoid 函数图像-----	4
图三：神经元工作图-----	5
图四：Tanh 函数图像-----	5
图五：ReLU 函数图像-----	6
图六：神经网络平方误差图-----	7
图七：误差反向传播法的流程图-----	9
图八：LSTM 单变量模型训练损失率比较-----	11
图九：LSTM 单变量模型 5 分钟数据预测结果-----	12
图十：LSTM 单变量模型小时数据预测结果-----	12
图十一：LSTM 单变量模型日数据预测结果-----	13
图十二：LSTM 多变量训练损失率比较-----	13
图十三：LSTM 多变量模型 5 分钟数据预测结果-----	14
图十四：LSTM 多变量模型小时数据预测结果-----	14
图十五：LSTM 多变量模型日数据预测结果-----	15
图十六：LSTM 价格预测交易策略净值-----	16

当前机器学习的研究方向主要分为两类：一类是传统机器学习的研究，该类研究主要是研究学习机制，注重探索通过算法来模拟人的思维和学习机制，如：决策树、随机森林、人工神经网络、贝叶斯学习等。一类是大数据环境下机器学习的研究，大数据的价值体现主要集中在数据的转向以及数据的信息处理能力。大数据时代的到来，对数据的转换，数据的处理、数据的存储提供了更好的技术支持，产业升级和新产业诞生形成了一股强大的推动力量。

本报告对机器学习中的一类算法---神经网络中的长短时记忆模型（LSTM）进行探索。LSTM 是循环神经网络(RNN)的一种，改进后的 LSTM 较好的解决了时间序列的预测问题，因此也成为当下最主流的 RNN 算法，同时在语音识别、图片描述、自然语言处理等许多领域中得到较好的应用。

1、神经网络

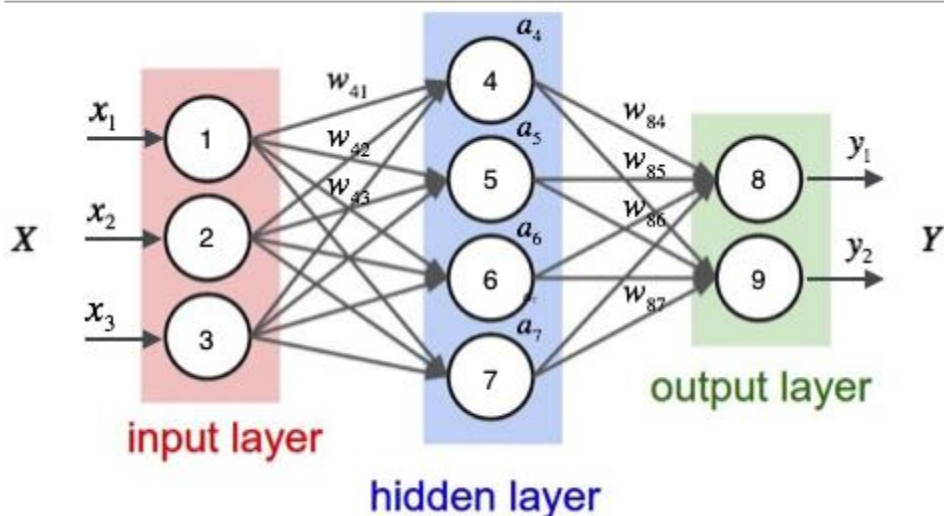
1.1 神经网络原理

神经网络的原理是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，依靠系统的复杂程度，通过调整内部大量节点之间相互连接的关系，从而达到处理信息的目的。

1.2 神经网络结构

典型的神经网络结构如下图所示，构成这个网络结构的层分别为：输入层、隐藏层、输出层。

图一：神经网络工作图



资料来源：华鑫期货研究所

输入层负责给予神经网络的信息；隐藏层对外部来说不可见，负责特征提取；输出层显示神经网络计算出的结果。

1.3 神经元工作数学表达

神经元的多个输入 $x_1, x_2, x_3, \dots, x_n$ 整理为数学表达式：

$$y=a(z), \quad z=w_1x_1+ w_2x_2+\dots+ w_nx_n+b$$

其中， w 为权重， b 为偏置， n 为输入的个数， $a(z)$ 为激活函数。

2. 激活函数

2.1 激活函数

激活函数（Activation Function）在神经网络的神经元上运行的函数，负责将神经元的输入映射到输出端。常用的激活函数有 Sigmoid 函数、Tanh 函数和 ReLU 函数。

2.1.1 激活函数的作用

激活函数是用来加入非线性因素，提高神经网络对模型的表达能力，解决线性模型所不能解决的问题。对于神经网络模型去学习、理解复杂和非线性的函数来说具有十分重要的作用。

2.1.2 三种常用激活函数

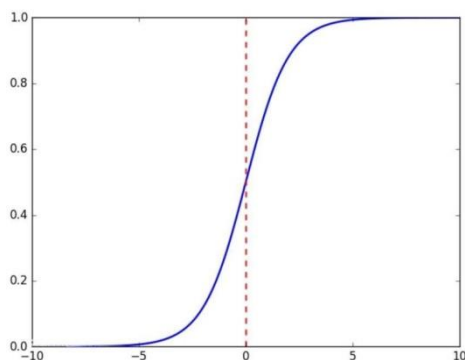
2.1.2.1 Sigmoid 函数

Sigmoid 函数是一个在生物学中常见的 S 型函数，也称为 S 型生长曲线。由于连续、光滑可导的两种性质使得 Sigmoid 函数很容易处理，因此它常被用作神经网络的阈值函数，将变量映射到 0,1 之间，公式如下：

$$f(x) = \frac{1}{1 + e^{-x}}$$

函数图像如下：

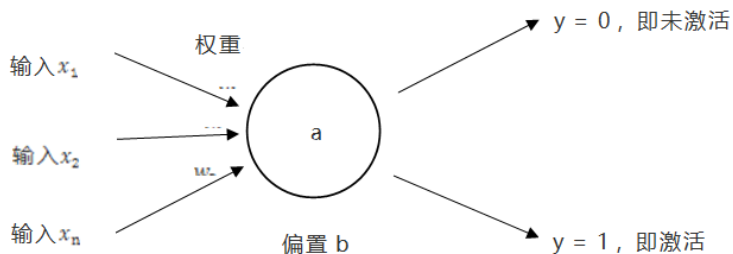
图二：Sigmoid 函数图像



资料来源：华鑫期货研究所

例：我们将神经元的示意图抽象化之后，对于输出信号，我们也对其生物上的限制进行一般化。根据激活与否，生物学上的神经元的输出 y 分布取值 1 和 0（激活与否用 1 和 0 表示）：

图三：神经元工作图



资料来源：华鑫期货研究所

这时表示激活与否的式子如下：

$$y = a(z), \quad z = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$

这里的函数 a 是建模者定义的激活函数，即 Sigmoid 函数， $x_1, x_2, x_3, \dots, x_n$ 是模型允许的任何数值， y 是函数 a 能取到的任意数值，其输出值是大于 0 小于 1 的任意值。

2.1.2.2 Tanh 函数

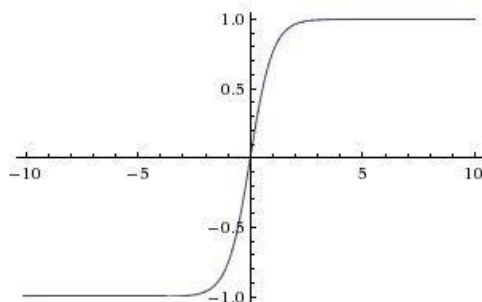
Tanh 是双曲函数中的一个，Tanh()为双曲正切。在数学中，双曲正切“Tanh”是由基本双曲函数双曲正弦和双曲余弦推导而来。

公式如下：

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

函数图像如下：

图四：Tanh 函数图像



资料来源：华鑫期货研究所

相较而言，在隐藏层，tanh 函数要优于 sigmoid 函数，可以认为

是 sigmoid 的平移版本，优势在于其取值范围介于-1 ~ 1 之间，数据的平均值为 0，而不像 sigmoid 为 0.5，有类似数据中心化的效果。但在输出层，sigmoid 也许会优于 tanh 函数，原因在于你希望输出结果的概率落在 0 ~ 1 之间，比如二元分类，sigmoid 可作为输出层的激活函数。

2.1.2.3 ReLU 函数

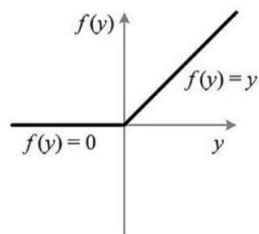
Relu 激活函数 (The Rectified Linear Unit)，用于隐层神经元输出。

公式如下：

$$f(x) = \max(0, x)$$

函数图像如下：

图五： ReLU 函数图像



$$ReLU(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

资料来源：华鑫期货研究所

实际应用中，特别是深层网络在“训练”时，tanh 和 sigmoid 会在端值趋于饱和，造成训练速度减慢，故深层网络的激活函数默认大多采用 relu 函数，浅层网络可以采用 sigmoid 和 tanh 函数。

3.输出过程

假定图一的神经网络中，激活函数采用 sigmoid，由输入节点 1、2、3 的值，有：

$$\begin{aligned} a_4 &= \text{sigmoid}(\vec{w}^T \cdot \vec{x}) \\ &= \text{sigmoid}(w_{41}x_1 + w_{42}x_2 + w_{43}x_3 + w_{4b}) \end{aligned}$$

输出节点 y1 的表达式为：

$$y_1 = \text{sigmoid}(\vec{w}^T \cdot \vec{a})$$

$$= \text{sigmoid}(w_{84}a_4 + w_{85}a_5 + w_{86}a_6 + w_{87}a_7 + w_{88})$$

同理，我们可以算出 a_5 、 a_6 、 a_7 、 y_2 的值，这就完成了神经网络输出值的计算。

4.神经网络优化

4.1 代价函数介绍

向神经网络提供学习数据，并确定符合学习数据的权重和偏置，这个过程称为学习或训练，在数学上一般称为最优化，最优化的目标函数是代价函数，也可以称为损失函数、目的函数、误差函数等。

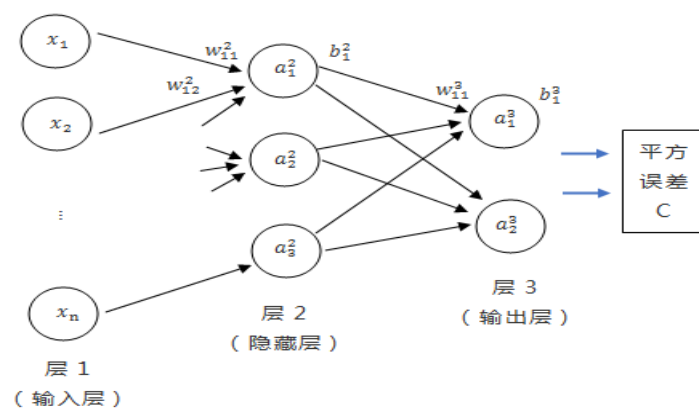
4.1.1 代价函数原理

神经网络确定权重和偏置的参数优化原理与回归分析相同，求使得表示误差总和 CT 的代价函数达到最小，即代价函数 CT 为误差的平方和， $CT=C1+C2+...Cn$ ，其中 $C1$ 、 $C2$ 、 $...Cn$ 是输出层的变量与该变量对应的正解的平方误差 C 。

4.1.2 代价函数数学表达

代价函数 CT 最小值条件的方程为 $dCT/dx=0$ ， $dCT/dy=0$ ， $dCT/dz=0$

图六：神经网络平方误差图



资料来源：华鑫期货研究所

神经网络算出的预测值用输出层的神经元的输出变量 a_1 、 a_2 来表示。设这些输出变量对应的正解为 t_1 、 t_2 。于是，预测值与正解的

平方误差 C 可以如下表示：

$$C = 1/2 * \{(t1 - a1)^2 + (t2 - a2)^2\}$$

在现实应用中，很多函数都可以作为代价函数，神经网络的世界中有名的代价函数是交叉熵，不过不论使用哪种代价函数，其思想和方法都是一样的。

4.2 误差反向传播法（BP 法）

虽然神经网络和回归分析确定模型的原理相同，但神经网络中的参数数量巨大，而且使用的函数（激活函数）并不是一次性的，使得代价函数变得较为复杂，所以神经网络相比回归分析需要更强大的数学工具---梯度下降法和误差反向传播法。

4.2.1 梯度下降法

最小化的方法中最有名的就是梯度下降法，它的理论思想就是沿着梯度 $(dCT/dx, dCT/dy, dCT/dz)$ 的方向一步一步地移动最终到达最小值，简洁数学表达式为

$$(\Delta C1, \Delta C2 \dots \Delta Cn) = -N * \nabla CT \quad (\text{哈密顿算子})$$

其中， CT 为误差总和， ∇CT 为它的梯度， N 为正的微小的常数（步长），在神经网络中也称学习率，是个只能通过反复试验来寻找恰当的值。

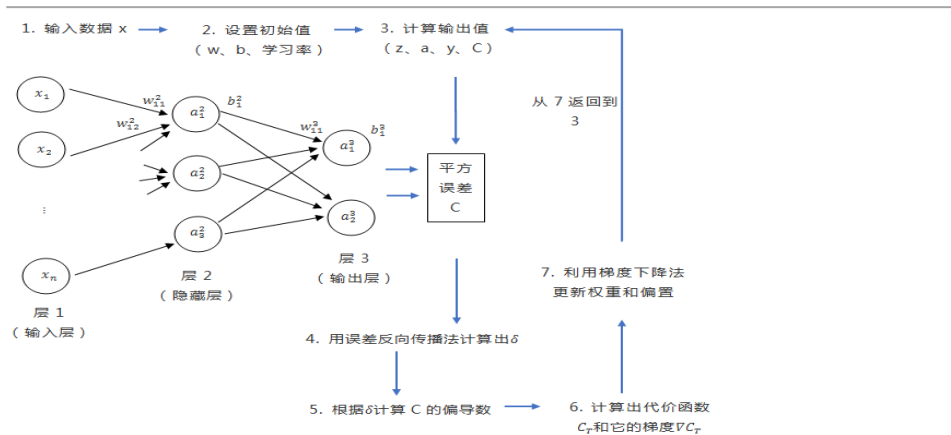
注：在神经网络中，由于计算求导过于复杂，因此不能直接使用梯度下降法，而是需要使用误差反向传播法。

4.2.2 误差反向传播法

误差反向传播法的工作原理就是以梯度下降的算法为基础，建立神经元误差 δ （ δ 为平方误差 C 关于权重 w 和偏置 b 的偏导数）的递推关系式，并通过这些递推关系式来回避复杂的导数计算。

神经元误差 δ 的数学方程： $\delta = dC/dz$ ； $dC/dw = \delta * a^{(-1)}$ ， $dC/db = \delta$ ；

图七：误差反向传播法的流程图



资料来源：华鑫期货研究所

第 3 步的处理称为前向传播，第 4 和 5 步的处理称为反向传播，误差反向传播法是将这两者组合起来的计算方法，即第 6 步是由第 3 至 5 步的全部数据相加而成。

4.3 小结

神经网络的训练就是根据目标提取特征设计好神经网络的结构后，输入数据对模型进行训练。通俗地讲，首先对网络结构中的各权重和偏置项设置一个初始值，选择好一个激活函数，设置一个适当小的正数作为学习率，然后把训练数据输入模型通过激活函数进行正向计算，得到输出层的输出值，然后使用误差反向传播算法，将输出值、实际值代入误差方程中，求出各层神经节点的误差项，再根据误差项计算平方误差关于权重和偏置的偏导数，然后根据梯度下降算法求出代价函数和它的梯度，最后用梯度对权重和偏置项的值进行更新，这就完成模型的一次训练。

5、LSTM 模型对期货价格预测的实证分析(5 分钟、1 小时、1 日数据)

5.1 数据预处理

5.1.1 数据归一化

由于使用的数据中，价格和成交量、持仓量的量纲不一样，数据特征不在一个维度，不具有可比较性。因此，在使用数据训练模型之前，首先我们需要对数据进行归一化处理，采用 MinMaxScaler 特征值进行线性变换缩放到[0, 1]区间，用以提升模型的精度和收敛的速度。

5.1.2 时间序列转换成监督学习

对于时间序列问题，我们通过使用最后一个 $t-1$ 时刻的值预测 t 时刻的值。监督学习问题由输入模式 (X) 和输出模式 (y) 组成，使得算法可以学习如何从输入模式预测输出模式。具体的操作就是用 Pandas 的 `shift()` 函数将整体的时间数据向后滑动一格，和原始数据拼接，就形成了监督的数据。

5.1.3 数据集划分

把数据集划分为三个部分：训练集、验证集和测试集，因为研究的是时间序列，按照时间对数据集进行划分，我们将前 70% 的数据作为训练集，接下来 15% 的数据作为验证集，剩下的 15% 作为测试集。

5.2 数据来源与特征项

采用螺纹钢期货主力连续合约行情数据，时间周期从 2016 年 7 月 1 日至 2019 年 8 月 30 日。单变量模型中，使用收盘价预测收盘价，多变量模型使用场内行情的开盘价、收盘价、最高价、最低价、成交量、持仓量，对收盘价进行预测。

5.2.1 单变量长短时记忆神经网络模型 LSTM

5.2.1.1 长短时记忆网络

长短时记忆网络(LSTM)是一种改进后的循环神经网络，可以解决 RNN 无法处理长距离依赖的问题，引入了一种计算单元作为记忆单元，取代了网络隐藏层中的传统人工神经元。利用这些记忆单元，网络能够有效地将记忆关联起来，并能及时远程输入，从而适应随时间动态掌握数据结构，具有较高的预测能力。在 t 时刻，LSTM

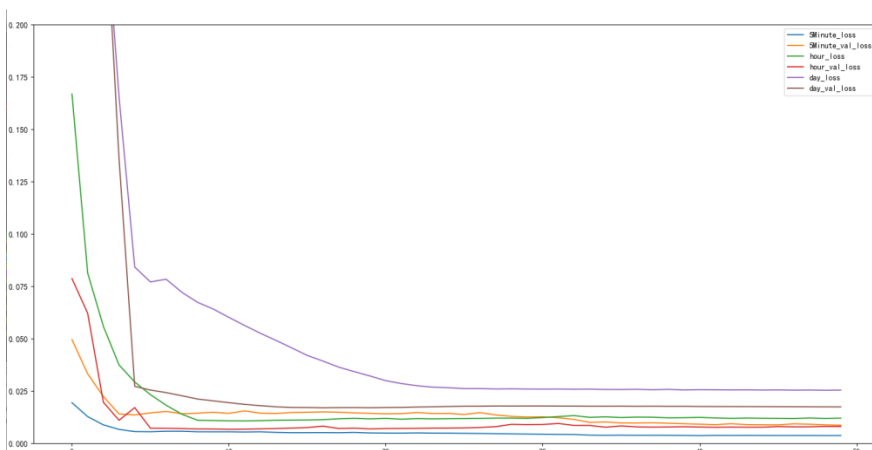
的输入有三个：当前时刻网络的输入 X_t 、上一时刻 LSTM 的输出 \hat{y}_{t-1} 、以及上一时刻的单元状态 C_{t-1} 。LSTM 的输出有两个：当前时刻网络输出值 \hat{y}_t 、当前时刻的单元状态 C_t 。

5.2.1.2 Adam 优化器

Adam 优化器，是一种计算每个参数的自适应学习速率的方法，考虑了过去的平方梯度的指数衰减平均值和过去梯度的指数衰减平均值。

5.2.1.3 训练结果

图八：LSTM 单变量模型训练损失率比较



资料来源：华鑫期货研究所

从上图看模型训练的损失率，5 分钟数据的好于小时数据，小时数据好于日数据。

性能报告：

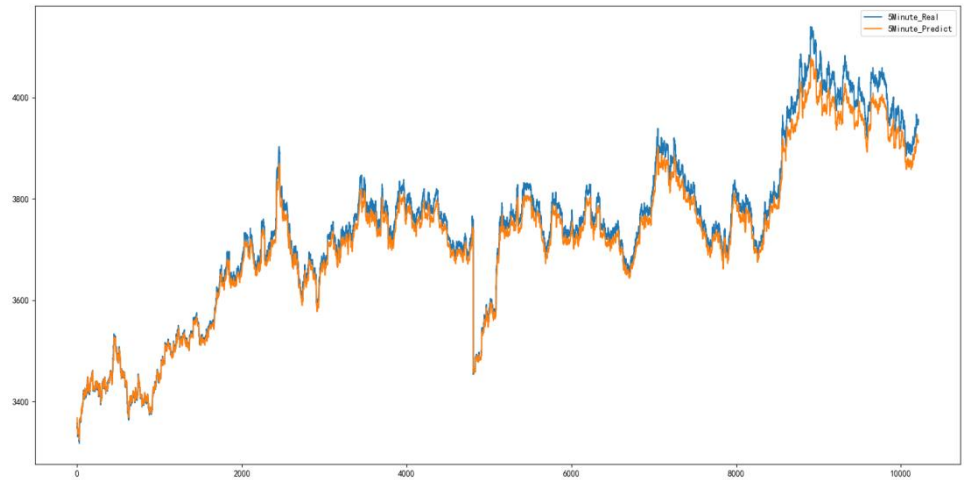
单变量 LSTM 不同频率模型训练比较

周期	训练损失率	验证损失率	模型训练耗时(秒)	RMSE
5 分钟	0.0037	0.0096	48.672	24.731
1 小时	0.0117	0.0082	6.183	24.212
1 日	0.026	0.0177	2.005	52.401

预测

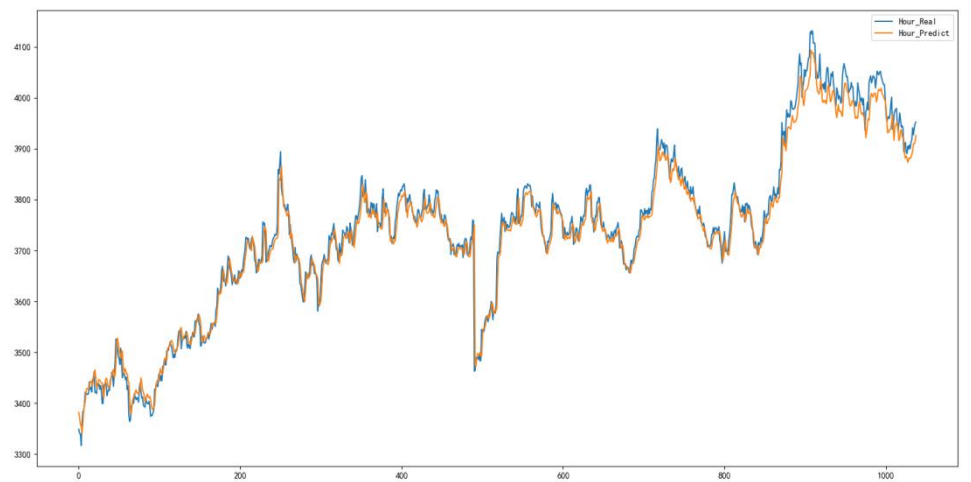
模型训练完成后，采用当天收盘价作为输入，对第二天收盘价进行预测，用滚动推进分析方法，得到收盘价预测值，与实际的收盘价绘制曲线如下：

图九：LSTM 单变量模型 5 分钟数据预测结果



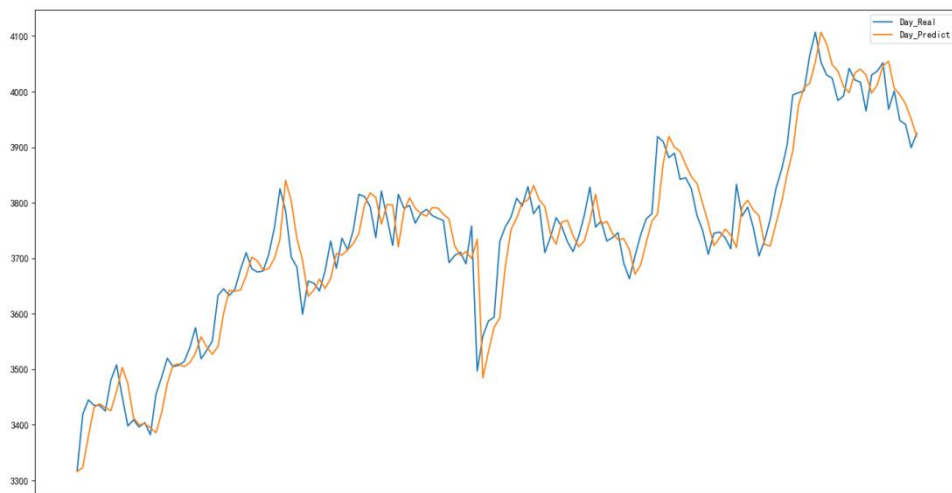
资料来源：华鑫期货研究所

图十：LSTM 单变量模型小时数据预测结果



资料来源：华鑫期货研究所

图十一：LSTM 单变量模型日数据预测结果



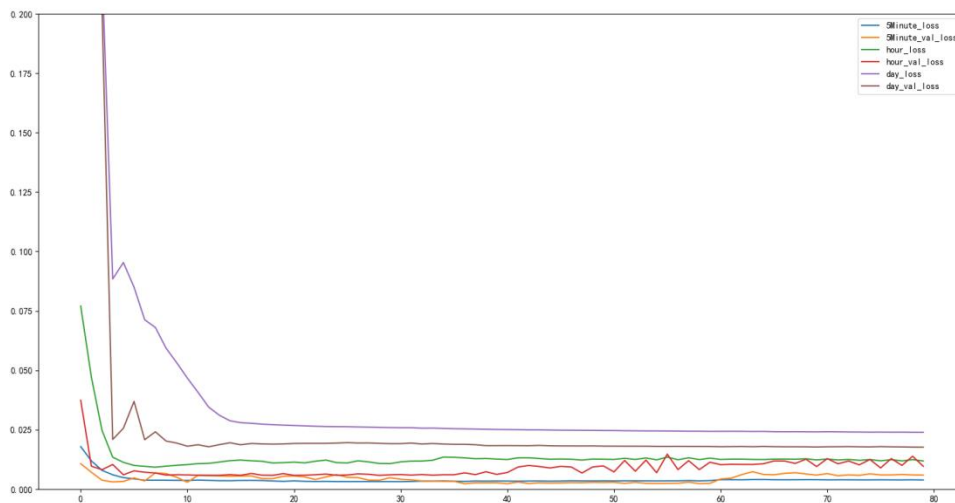
资料来源：华鑫期货研究所

5.2.2 多变量长短时记忆神经网络模型 LSTM

使用开盘价、收盘价、最高价、最低价，成交量、持仓量来预测收盘价，单步预测，即用 $t-1$ 时刻的数据预测 t 时刻的收盘价。网络结构定义及模型训练和单变量模型相同。

训练结果和预测：

图十二：LSTM 多变量训练损失率比较



资料来源：华鑫期货研究所

性能报告：

多变量 LSTM 不同频率模型训练比较

周期	训练损失率	验证损失率	模型训练耗时(秒)	RMSE
5 分钟	0.0040	0.0061	83.218	30.504
1 小时	0.0118	0.0095	17.764	23.744
1 日	0.0239	0.0176	5.029	45.100

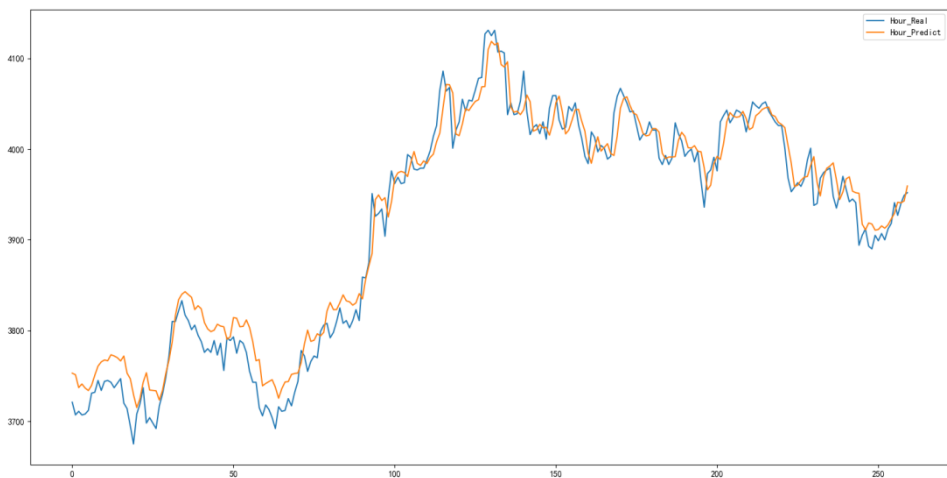
预测图

图十三：LSTM 多变量模型 5 分钟数据预测结果



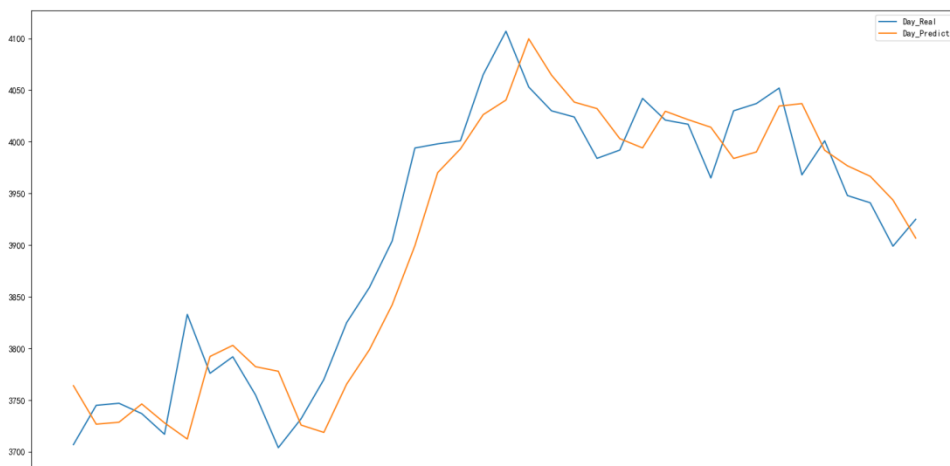
资料来源：华鑫期货研究所

图十四：LSTM 多变量模型小时数据预测结果



资料来源：华鑫期货研究所

图十五：LSTM 多变量模型日数据预测结果



资料来源：华鑫期货研究所

由以上实证分析可得结论：多变量模型要比单变量模型提取的特征项要多，拟合跟预测效果更好。从数据的采样频率角度看，5分钟数据的训练效果要好于1小时数据，1小时数据的训练效果要好于1日数据。

6、交易策略

6.1 策略描述

采用海龟交易法则的思想，用预测的收盘价和开盘价进行比较。如果下一开盘价高于预测价，则以开盘价买入开仓，收盘价卖出平仓；如果下一开盘价比预测价低，则以开盘价卖出开仓，收盘价买入平仓。

6.2 策略实现

选择测试中时间维度居中的小时数据，训练时间范围2016年7月1日至2019年7月8日，对2019年7月9日至2019年8月30日的小时收盘价进行预测，根据预测结果计算策略的收益，净值曲线如下：

图十六：LSTM 价格预测交易策略净值



资料来源：华鑫期货研究所

对近期两个月的数据进行了交易回测，累计收益 8%，最大回撤约 5%，夏普率为 2.33。

该回测结果未对神经网络结构、超参数等做任何优化调整，也未纳入手续费、交易滑点、资金管理、风险管理等因素。

■ 分析师声明

崇盛棠声明，本人具有中国期货业协会授予的期货从业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。

■ 本公司具备期货投资咨询业务资格的说明

华鑫期货有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得期货投资咨询业务许可。本公司及其投资咨询人员可以为期货投资人或客户提供期货投资分析、预测或者建议等直接或间接的有偿咨询服务。发布期货研究报告，是期货投资咨询业务的一种基本形式，本公司可以对期货及期货相关产品的价值、市场走势或者相关影响因素进行分析，形成投资评级等投资分析意见，制作期货研究报告，并向本公司的客户发布。

■ 免责声明

本研究报告由华鑫期货有限公司撰写，报告中的信息均来源于已公开的资料，我公司尽可能保证可靠、准确和完整，但并不保证报告所述信息的准确性、完整性和时效性，也不保证我公司做出的建议不会发生任何变更。在任何情况下，我公司不就本报告中的内容对任何投资做出任何形式的担保，本报告不能作为道义的、责任的和法律的依据或者凭证，对于本报告中提供信息所导致的任何直接或间接投资盈亏不承担任何责任。

本报告的版权归华鑫期货有限公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用发布，需注明出处为“华鑫期货有限公司”，且不得对本报告进行有悖原意的引用、删节和修改。华鑫期货有限公司对于本免责声明条款具有修改权和最终解释权。

华鑫期货有限公司 研究所

公司地址：上海市黄浦区福州路 666 号华鑫海欣大厦 21、22 楼

公司官网：<http://www.shhxqh.com>

咨询电话：400-186-8822

