

# 基于 Hyperband-LSTM 模型的股票价格预测研究

陈健 刘伟基\*

(上海师范大学 商学院, 上海 200030)

**摘要：**本文提出将 Hyperband 算法与 LSTM 神经网络算法相结合，建立一个适用于股票价格预测的模型——Hyperband-LSTM 模型，以解决目前在股票预测任务上因 LSTM 神经网络重要超参数设定的不确定性和繁杂性而引起的预测精度不高和预测效率低下的问题。同时，以预测沪深 300 指数收盘价为例进行实证研究，引入 Bayes-LSTM 和 LSTM 模型作为对比，实证结果表明 Hyperband-LSTM 模型比 Bayes-LSTM 和 LSTM 模型具备更高的预测精度，并且 Hyperband 优化比 Bayes 优化节约近一半的时间成本，即 Hyperband-LSTM 模型表现出更高的预测效率。此外，本文通过 DM 检验证实 Hyperband-LSTM 模型的预测效果显著优于其它模型。最后，本文使用上证 50 指数和深证 100 指数进行重复实验，实验结果与沪深 300 指数的实验结果保持一致，表明 Hyperband-LSTM 模型具备较强的稳健性。

**关键词：**Hyperband 算法；LSTM 神经网络；股票价格预测；DM 检验

**中图分类号：**F830.91      **文献标识码：**A      **文章编号：**

---

[投稿日期]：2021-01-04；[修回日期]：2022-02-10

[作者简介]陈健，女，1972.-，经济学博士，上海师范大学商学院副教授，硕士生导师；研究领域：金融时间序列分析；电子邮箱：chenjian@shnu.edu.cn；联系电话：021-64322339；邮政编码：200030；通讯地址：上海市徐汇区桂林路 100 号。

通讯作者：刘伟基\*（通讯作者），男，1996.3-，经济统计学专业硕士研究生，上海师范大学商学院；研究领域：金融时间序列分析；电子邮箱：vickylau@163.com；联系电话：13172907221；邮政编码：200030；通讯地址：上海市徐汇区桂林路 100 号。感谢匿名评审专家和编辑部的宝贵意见，文责自负。

# **Research on Stock Price Prediction Based on Hyperband-LSTM**

## **Model**

Jian chen, Weiji Liu

(School of Finance and Business, Shanghai Normal University, Shanghai 200030, China)

**Abstract:** This paper proposes to combine the Hyperband algorithm with the LSTM neural network algorithm to establish a model suitable for stock price prediction, the Hyperband-LSTM model, in order to solve the problems of low prediction accuracy and low prediction efficiency caused by the uncertainty and complexity of the important Hyperparameter of the LSTM neural network. At the same time, taking the prediction of the closing price of the CSI 300 Index as an example, the Bayes LSTM and LSTM models are introduced as a comparison. The empirical results show that the Hyperband-LSTM model has higher prediction accuracy than Bayes-LSTM and LSTM models, and the Hyperband optimization saves nearly half of the time cost than Bayes optimization, that is, the Hyperband-LSTM model shows higher prediction efficiency. In addition, this paper confirms that the prediction effect of Hyperband-LSTM model is significantly better than other models through DM test. Finally, this paper uses the SSE 50 Index and the SZSE 100 Index to carry out repeated experiments. The experimental results are consistent with those of the CSI 300 Index, indicating that the Hyperband-LSTM model has strong robustness.

**Key Words:** Hyperband Algorithm; LSTM Neural Network; Stock Prediction; DM Test

## 一、引言

股票资产投资因其高收益性，受到了大量投资者的青睐。但是股票资产的高收益往往伴随着高风险，一种规避风险获得超额收益的方法是通过预测股票价格来进行投资组合的构建。由于机器学习模型能捕获股票序列中的非线性关系，被广泛应用于股票价格预测。但相对于传统的预测模型，机器学习虽然能提高预测精度，但在模型建立的过程中需要耗费更多的时间。由于投资组合的构建往往是需要对几百甚至几千只股票同时进行建模预测而确定，如李斌等（2019）<sup>[1]</sup>、Leippold et al.（2021）<sup>错误!未找到引用源。</sup>在建立投资组合时所使用的股票池包含了上千只股票，那么时间成本将会成倍数地增加，李斌的研究中为了加快实验进程使用了40台计算机进行集群实现。因此，现有机器学习模型存在的一个较大问题是建模成本较高，如何构建一个同时满足高精度和低成本股票预测模型成为重要的研究方向。在当前的股票预测方法研究中，大部分的研究都聚焦于精度提升，鲜有学者关注股票序列的预测成本。本文研究的落脚点是在保证预测精度的前提下，尽可能地降低股票序列预测的实现成本，力求使模型更加切合实际投资需求。尽管本文的研究重点是对预测方法的探索，暂不涉及投资组合的选择研究，但本文的研究目的是试图为建立股票投资组合提供一种能在保证预测精度的同时降低预测成本的方法。接下来，本文将通过梳理前人在股票预测上的研究成果，选择一种适合股票预测任务的基础模型，然后对这个模型进行优化以实现提高股票价格预测的效率。

国内外相关工作进行股票价格的预测研究，使用的方法大致可以分为传统的时间序列模型、传统机器学习算法以及使用神经网络算法(黄超斌和程希明，2021)<sup>[3]</sup>。传统的时间序列模型主要包括GARCH模型和ARMA模型及其衍生模型，这些模型的一个共性是其仅能挖掘数据的线性关系而忽略了非线性关系。因此，在存在复杂非线性关系的股票序列预测的应用上，传统的时间序列模型存在着较大的局限性。而机器学习的模型具备识别非线性关系的能力，近些年来越来越多的学者运用机器学习的方法对股票价格进行预测，并且证实了大部分的机器学习模型的预测效果优于传统的计量模型，如张贵生和张信东（2016）<sup>[4]</sup>，提出了SVM-GARCH预测模型，通过股指预测实验发现其预测效果在各个方面均优于传统的ARMA-GARCH模型；潘水洋（2019）<sup>[5]</sup>研究发现机器学习的非线性定价模型在样本外预测精度、分组多空策略业绩表现均优于传统线性定价模型。传统的机器学习模型虽然能识别出股票序列的非线性关系，预测精度高于传统的时间序列模型，但由于其浅层结构导致非线性

关系的识别能力仍然较弱，而且缺乏一种挖掘股票特征的能力。在神经网络算法中衍生出来的深度学习算法，有着深层的网络设计，具备更优的非线性识别能力和抽象化的特征提取能力，在包括股市预测在内的任务上表现出优于线性和传统机器学习模型的性能（Jiang, 2021）<sup>[6]</sup>。已有大量研究表明，基于深度学习算法的股票价格预测模型的效果优于传统机器学习模型。如黎镞等（2018）<sup>[7]</sup>发现使用GRU深度神经网络预测股票收盘价，GRU深度神经网络预测的相对误差均值要低于SVR模型和BP模型预测的相对误差均值。Fischer and Krauss（2018）<sup>[8]</sup>的研究发现在金融时间序列的预测效果上LSTM深度神经网络优于随机森林(RAF)、逻辑回归分类器(LOG)。欧阳红兵等（2020）<sup>[9]</sup>研究发现LSTM深度神经网络在股票序列预测任务上的预测精度要优于传统机器学习模型和传统的时间序列模型。

从相关文献的梳理中可以发现，目前深度学习算法在股票预测的研究中能取得比传统时间序列模型和传统机器学习模型更好的效果。虽然深度学习算法在预测精度上是高于其它模型，但是由于其复杂性——存在着许多需要训练而得到的权重，因此相比其它模型的拟合成本是有所增加的，如Borovkova and Tsiamas（2019）<sup>[10]</sup>发现：尽管深度学习能表现出最优的预测效果，但其计算成本要远远高于拉索回归和岭逻辑回归。相对于模型复杂性而产生的成本，深度学习算法的模型建立成本更多的是来源于灵活性——存在许多需要根据具体任务设定的超参数。超参数的选择对于机器学习方法的性能至关重要，因为它们控制了模型的复杂性（Gu et al., 2022）<sup>[11]</sup>。在大多数情况下，几乎没有关于如何“调优”超参数以优化样本外性能的理论指导（Gu et al., 2022）<sup>[11]</sup>。因此实验者需要通过重复实验去寻找一组适合的超参数组合来保证模型的预测性能，或者是利用设定好的程序实现超参数的自动搜寻，这个过程耗费了实验人员的大部分精力和计算机的算力，占据了整个实验的绝大部分时间。在实际应用中，深度学习模型存在着非常高维度的超参数，如LSTM（Long Short-Term Memory）模型需要设定每一层的神经元数量、神经元舍弃率和激活函数等的超参数，当引入层数变化时，超参数组合的范围将是巨大的。综上所述，目前深度学习算法存在因灵活性和复杂性而导致的建模成本高，以及因超参数的不确定性而引起的模型预测效果不佳的问题，因为股票序列的多样性和不稳定性，这些问题在股票序列预测应用中表现得尤为突出。针对深度学习算法在股票序列中现存的问题，本文选用目前在股票价格预测中效果较好的LSTM神经网络，在LSTM神经网络的基础上引入由Li et al.（2018）<sup>[12]</sup>提出的Hyperband算法，构造出一个能有效解决现存问题的模型——Hyperband-LSTM模型，此模型能够快速有效地自适应识别LSTM神经网络各类超参数，解决现有股票价格预测的预测成本高、预测效率低的问题，

实现在股票价格的预测上以更低成本达到更高预测精度的目标。

## 二、模型原理

本部分首先对LSTM神经网络的原理做一个简单的分析，通过阐述LSTM神经网络的原理来表明建立LSTM神经网络对股票价格序列预测研究的合理性和可行性，以及引入Hyperband算法的必要性。然后对Hyperband-LSTM模型进行说明，以此来表明Hyperband算法对LSTM神经网络预测精度提升和预测效率提高的合理性和可行性。

### (一) LSTM 神经网络原理

为了能更简明地描述神经网络的基本原理，本文以一个简单的双层神经网络为例。如图1，第一层（隐藏层）神经网络接受模型的输入值，经过神经元的运算  $g_i^1(X) = W_i^1 X + b_i^1$  将原输入值  $X = \{x_1, x_2, \dots, x_n\}$  进行了  $i$  次的抽象特征提取。接着  $g_i^1$  经过非线性激活函数  $f^1(\cdot)$  得出此层的输出  $y_i^1$ ，其中  $y_i^1 = f^1(g_i^1)$ ，最终得出经非线性变换的  $i$  个新特征向量作为输出层的输入。

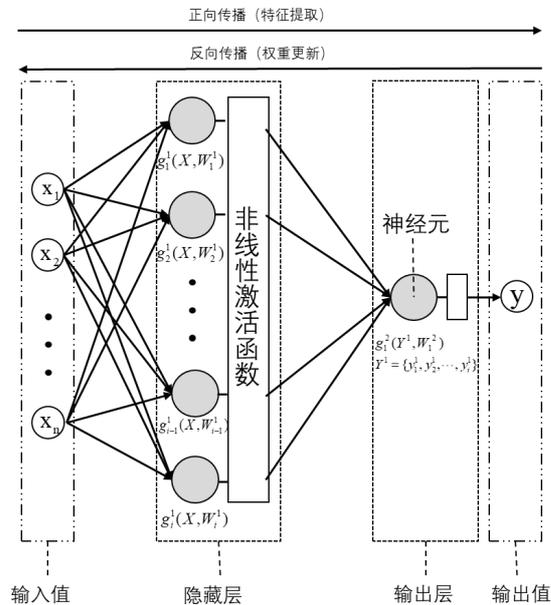


图1 双层神经网络

然后以  $Y^1 = \{y_1^1, y_2^1, \dots, y_i^1\}$  作为第二层（输出层）的输入，进行相似的运算最终得出网络的输出  $y$ 。这个过程称为神经网络的正向传播，其作用是将原始输入进行非线性的特征抽象提取，以达到拟合非线性数据的目的。最后，深度神经网络以预测误差为依据，以梯度下降为方向更新权重  $W = \{W_1^1, W_2^1, \dots, W_i^1; W_1^2\}$ ，权重的具体更新大小和方向依具体的优化器而定，此过程称为深度神经网络的反向传播，目的是得到适合的权重使得网络能提取到有效的特征，最终让模型具备一定的预测能力。

由深度神经网络的原理可以清晰地了解到：神经网络具备提取抽象化特征和拟合非线性

数据的能力。因此，对于存在非线性特性的股票价格预测使用神经网络是合理的。但是诸如股票价格等的时间序列存在着序列相关，而普通的神经网络无法识别出时间序列的这种特性，而RNN(Recurrent Neural Network)模型能够很好地识别出时间序列的序列相关，与MLP(Multilayer Perceptron)模型、CNN(Convolutional Neural Networks)模型等深度学习模型不同的是：其能关注到不同时期输入值与目标值的关联性，即能关注到时间序列中的序列相关。可是由于RNN模型存在对长期信息遗忘的问题，在长期存在序列相关的时间序列中，RNN模型容易把较早的信息遗忘掉。而大部分的时间序列预测目标不仅仅与历史信息有着短期联系，而是与历史信息有着较长的关联，如股票序列，人们并不会只关心过去一两期的信息，而是着眼于序列的长期走势。LSTM模型是RNN模型的一个变种模型，是在RNN模型中加入一个特殊的门控单元，如图2，这个门控单元包括遗忘门、输入门和输出门，其中遗忘门决定需要舍弃的旧信息、输入门决定新信息的保留率、输出门通过整合处理过的旧信息和新信息得出当前输出，使得LSTM模型能够解决RNN模型在长序列中的早期信息遗忘问题。

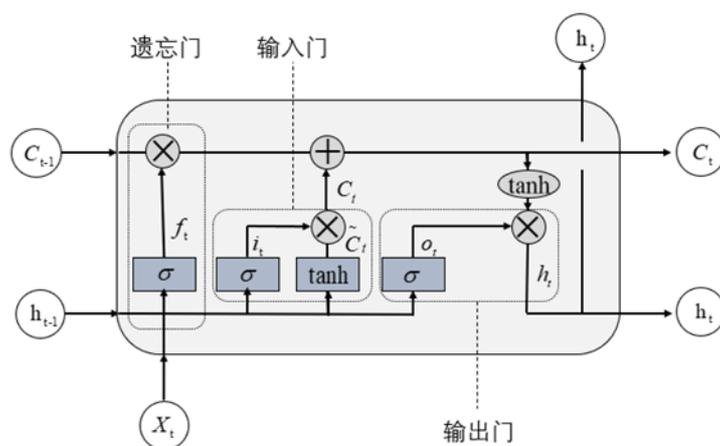


图2 LSTM单元

遗忘门： $X_t$ 为 $t$ 时刻的特征输入， $C_{t-1}$ 和 $h_{t-1}$ 分别为上一门控单元的细胞状态和隐藏层状态，同时也作为 $t$ 时刻的门控单元的输入，也就是当前的门控单元的输入包括 $X_t$ 、 $C_{t-1}$ 和 $h_{t-1}$ 。首先 $X_t$ 和 $h_{t-1}$ 进行如公式(1)的运算，输出 $f_t$ 将决定 $C_{t-1}$ 的舍弃率。

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (1)$$

输入门： $X_t$ 和 $h_{t-1}$ 进行如公式(2)的运算，这次运算的输出 $i_t$ 将决定新信息的保留率，

同时  $X_t$  和  $h_{t-1}$  进行如公式 (3) 的运算，此次运算的输出  $\tilde{C}_t$  为当前门控单元新的输入带来信息。

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, X_t] + b_c) \quad (3)$$

输出门：接着通过公式 (4) 计算出当前门控单元的细胞状态  $C_t$ ，最后  $X_t$  和  $h_{t-1}$  进行如公式 (5) 的运算，得出输出  $o_t$ ，如公式 (6)，将经  $\tanh$  函数变换的  $C_t$  相乘得出 LSTM 单元的输出  $h_t$ 。

$$C_t = f_t * C_{t-1} + i_t \otimes \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, X_t]) \quad (5)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (6)$$

其中， $\tanh(\bullet)$  表示  $\tanh$  函数变换， $[\bullet]$  表示矩阵的连接， $\otimes$  表示元素乘法， $W_f$ 、 $W_i$ 、 $W_c$ 、 $W_o$ 、 $b_f$ 、 $b_i$  和  $b_c$  都是模型权重，通过模型训练求得。

由上述 LSTM 神经网络的原理可知，LSTM 神经网络使用一种特殊的门控单元代替了普通神经网络的神经元，使得 LSTM 神经网络可以很好地识别出时间序列的长期序列相关性，因此使用 LSTM 神经网络来研究股票价格的预测问题是合理的。可是，对于 LSTM 神经网络中需要提前设定的超参数，如层数、神经元数量和激活函数的选择等，在实验开始前是难以准确判断的，因此需要一个行而有效的方法帮助实验中识别这些超参数，才能提高实验的效率和实验结果的准确度。

## (二) Hyperband-LSTM 模型原理

由上文 LSTM 神经网络的原理，可知与 LSTM 神经网络的预测精度直接相关的是各隐藏层神经元的权重和激活函数，因为权重和激活函数决定着模型的输入因子以何种抽象形式和以何种非线性形式转化为 LSTM 神经网络隐藏层的输出，即决定了隐藏层的非线性识别和特征提取的能力。权重的生成取决于 LSTM 模型中所设定的超参数，如权重的数量取决于神经元的数量和隐藏层层数、权重的更新大小和方向取决于优化器，而激活函数本身就是需要提前设定的一种超参数，因此 LSTM 神经网络的超参数在直接或间接地决定着模型的预测精度。

因此本文以LSTM神经网络为基础模型，引入Hyperband算法为LSTM神经网络快速有效地配置超参数组合，提高实验结果的准确性和实验的效率。

Hyperband算法的主要贡献在于实现了超参数组合数量  $n$  和每组超参数所能获得的资源量  $B/n$  之间的有效权衡，其中  $B$  为超参数优化过程中所消耗的总资源量。本文将LSTM神经网络嵌套进Hyperband算法中，实现流程如图3所示。给定每个参数组合所能得到的最大资源  $R$ ，参数组合保留率  $1/\eta$ 。初始化Hyperband算法的两个主要参数：

$s_{\max} = \lfloor \log_{\eta}(R) \rfloor$ ， $B = (s_{\max} + 1)R$ ，其中  $\lfloor \cdot \rfloor$  表示向下取整， $s_{\max}$  决定Hyperband算法

的外循环次数， $(s_{\max} + 1)$  为Hyperband算法的外循环次数。在给定了初始参数后，得出第一轮外循环的超参数组合数  $n$  和每组超参

数组合实际能分配到的资源量  $r$ ；接着初始化内循环——连续减半算法（Successive Halving Algorithm）的参数  $i$ ，进而确定内循环的第一轮内循环所要迭代的超参数组合数量  $n_0$  和每组超参数组合实际能分配到的资源量  $r_0$ ；内循环所要做的核心工作是将所获取的参数组合镶嵌到LSTM神经网络中使用训练集和验证集分别进行模型的训练和验证；在全部超参数组合的训练与验证工作结束后，通过比较不同超参数组合所构建模型产生的验证集误差值进行超参数组合的筛选，保留其中最后的  $1/\eta$  的超参数组合；接着进行下一次内循环，直达  $i = s$ ，结束内循环开始第二轮外循环；当  $s = 0$  时，结束最后一轮外循环，得到一组适用于LSTM神经网络的超参数组合，使得该模型能尽可能地发挥其最大的预测能力。

Hyperband的核心思想在于资源的合理配置，通过有差别地对待每一组超参数组合，给予对资源有高需求的超参数组合足够的资源，给予对资源依赖较低的超参数组合少量的资源，实现在给定的总资源数上完成更多超参数组合的验证。Hyperband优化算法通过合理的资源配置，能大大降低LSTM神经网络的优化成本，同时以更丰富的超参数组合规避了陷入局部

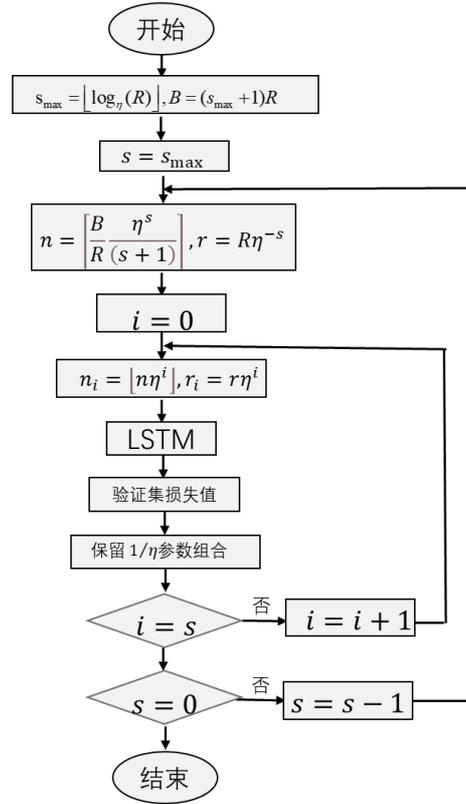


图3 Hyperband-LSTM原理

最优解的问题。因此，Hyperband-LSTM模型能自适应地、有效率地为LSTM神经网络配置一组最优的超参数组合，解决了手动设置超参数的繁杂性和不确定性，并且能识别每一组超参数组合对资源的需求程度，进行有效的资源优化配置，实现了以更低的时间成本对更大的超参数空间进行搜索，解决了现有超参数优化方法的优化效率不足问题，最后通过超参数的优化实现了预测精度的进一步提升。

### 三、实证研究

#### (一) 指标选取及数据说明

本文数据来源于万得金融数据库，包含了2020年8月至2021年8月的沪深300指数5分钟高频数据，样本总数为12720个。预测目标为沪深300指数每五分钟的收盘价，如表1所示，指数的预测因子由通过滞后处理的22个股票基本指标和技术指标构成。

表1 预测因子

符号	名称	符号	名称
<i>open</i>	开盘价	<i>CCI</i>	顺势指标
<i>high</i>	最高价	<i>OSC</i>	震荡量指标
<i>low</i>	最低价	<i>ROC</i>	变动率指标
<i>val</i>	成交额	<i>W&amp;R</i>	威廉指标
<i>vol</i>	成交量	<i>EMV</i>	简易波动指标
<i>close</i>	收盘价	<i>WVAD</i>	威廉变异离散量
<i>BBI</i>	多空指标	<i>CR</i>	中间意愿指标
<i>DDI</i>	方向标准离差指数	<i>PSY</i>	心理线指标
<i>MACD</i>	异同移动平均线	<i>VR</i>	成交量变异率
<i>MTM</i>	动力指标	<i>BB</i>	布林极限
<i>TRIX</i>	三重指数平滑平均线	<i>BOLL</i>	布林线指标

本文将数据集划分为训练集、验证集和测试集三个数据集，分别把三个数据集的数据经归一化处理变换为 $[0,1]$ 范围内的数据，训练集数据用于确定LSTM神经网络的权重组合，验证集数据用于确定LSTM神经网络的超参数组合，测试集数据用于评价模型的预测效果。基于对数据的充分利用的原则，本文将验证集的第一个样本的预测因子从训练集中获取，因此表2的三个数据集样本的起始时间和结束时间有一定的交集，但是模型训练时并不涉及有信息超前性，即模型仍然是基于历史信息对未来信息做出预测。本文使用的预测方法为滚动预测，滚动预测过程中设定使用过去24期的指标对未来1期的收盘价进行预测。

表2 数据集划分

数据集	样本数	起始时间	结束时间
训练集	7656	2020/8/03 9:35	2021/3/30 11:30
验证集	2568	2021/3/30 9:35	2021/6/18 11:30
测试集	2544	2021/6/18 9:35	2021/8/31 15:00

股票指数数据中的各个指标的量级差距过大，会导致模型在识别指标时，对不同量级的指标的关注度不同，量级较大的指标可能会引起模型的过度关注，而量级较小的指标甚至会被模型忽略掉。为了解决这一问题，需要对各数据集的数据进行归一化处理，消除数据的量纲和将数据缩小至同一量级。同时，归一化处理可以减轻模型的训练负担，使模型能更快地寻找到最优解。本文使用归一化方法为线性归一化，归一化过程如公式（7）。

$$y_{scale} = \frac{y - \min(Y)}{\max(Y) - \min(Y)} \quad (7)$$

其中， $Y$  表示训练集的全集， $\min(\cdot)$  为取最小值函数， $\max(\cdot)$  为取最大值函数， $y$  为原始值， $y_{scale}$  为归一化的结果值。

需要强调的是，由于本文假定测试集是未来的数据，因此在测试集中进行归一化时是使用训练集中的最大值和最小值。同时，为了使得模型在样本外的预测更具稳健性，验证集的归一化过程同样使用训练集中的最大值和最小值。

进行归一化后，数值转化为 $[0,1]$ 区间的数值，在预测结果呈现时需要把归一化后的数据进行反归一化处理得到符合原数据集量级的数据，反归一化过程如公式（8）。

$$y' = y_{pre}(\max(Y) - \min(Y)) + \min(Y) \quad (8)$$

其中， $y_{pre}$  为模型输出的预测值， $y'$  为反归一化的结果。

## （二）参数说明

如图4所示，本文LSTM神经网络的架构包括LSTM层、Dropout层、Flatten层和Dense层。LSTM层的作用是使得神经网络具备长期记忆性，同时考虑到可能存在的过拟合问题，本文在每一层的LSTM层引入Dropout层，Dropout层的作用是按给定的概率随机关闭一定比例的神经元，通过减小模型内部训练的权重数量减轻过拟合现象。然后在最后的Dropout层后加入Flatten层，Flatten层的作用是将LSTM的多维输出降低为一维输出用于最后一层Dense层的输入，最后Dense层将模型的预测结果输出。

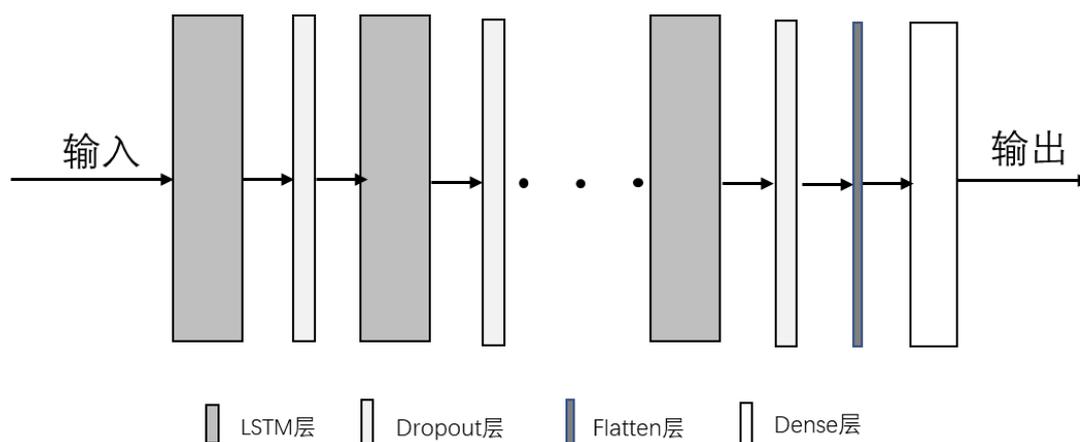


图4 LSTM神经网络的架构

由神经网络的万有逼近定理,可知在深度神经网络中只要存在一层的隐藏层,并且该隐藏层的神经元足够的多,所构造的深度神经网络就有能力以任意的精度去逼近任何形式的函数。但是,在现实的实验中,权衡地使用适当的层数和神经元可以使得预测任务更有效率(奥雷利安·杰龙, 2021) [13]。目前在股票价格预测应用中的研究文章中, LSTM神经网络神经元层数设为2层和3层时效率最高(胡聿文, 2021) [14]。而在LSTM层上的神经元数量的设定上并没有一致的认识,因此本文尽可能地扩大神经元数量的搜索范围,避免神经元的不足而影响模型预测性能。

深层神经网络的非线性关系识别能力主要体现在隐藏层的特征抽象提取和激活函数的非线性变换中,因此神经网络中的激活函数设定在非线性拟合过程中起到了不可或缺的作用。在LSTM层中,本文选用默认的Tanh激活函数;在Dense层激活函数的选择中,拟选择Tanh, Relu, Softplus和Elu函数,四种激活函数都能实现非线性映射,但其转换过程不尽相同,所能达到的效果也因具体数据而异。神经网络的训练过程大致可归纳为:随机初始化权重,使用训练集数据进行第一次训练;然后把训练效果反馈给模型,依梯度下降原则更新权重,进行下一次训练,此过程称作反向传播;不停重复此过程,直到模型能达到稳定的训练效果,最终确定模型权重。权重的更新方向和大小是基于LSTM神经网络设定的优化器决定的,目前较为成熟的优化器包括SGD、Adam、AdaGrad、RMSprop,但是并无相关理论指导在未知的数据中使用哪种优化器能得到较好的效果。因此,在实验之前是无法预知哪种激活函数和优化器适合当前数据的,因此激活函数和优化器的设定需要Hyperband算法搜索确定。

根据深度学习的基本原理,可以知道LSTM神经网络是一种基于数据驱动的预测模型,在实现不同的任务中需要灵活地变动相应的超参数才能最大化地激发深度神经网络的预测潜能,因此本文参考前人的经验和相关原理确定超参数搜索空间如表3所示,在Hyperband

算法中设定  $\eta = 1/3$ 、 $R = 50$ ,即内循环每次会保留三分之一的超参数组合，每组超参数最多能获得50次迭代。

表3 超参数搜索空间

	范围	步长
LSTM层数	[2,5]	1
神经元数	[16,256]	16
舍弃率	[0.1,0.6]	0.1
激活函数	Relu、tanh、softplus、elu	--
优化器	SGD、Adam、AdaGrad、RMSprop	--

### (三) 评价指标

本文引入5种评价指标对模型的预测效果进行评估，分别为均方误差（MSE）、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、皮尔逊相关系数(PCCs)和模型建立时间(Time)。MSE、MAE是衡量预测值与真实值之间的绝对误差，而平均绝对百分比误差表示预测值与实际值之间的相对误差（Ayala et al., 2021）<sup>[15]</sup>。模型建立时间是指在同样的实验环境（计算机硬件和软件）下LSTM神经网络超参数优化过程和权重迭代训练过程所耗费的总时长，是以时间角度上体现的模型建立成本。MSE、MAE、MAPE和Time越小表示模型的效果越佳。PCCs用于度量预测值和真实值之间的相关性，PCCs 的值越接近1，预测值与真实值的拟合度越高（党建武和从筱卿，2021）<sup>[16]</sup>。均方根误差、平均绝对误差、平均绝对百分比误差和皮尔逊相关系数的计算分别如公式（9）、公式（10）、公式（11）和公式（12）所示。

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \hat{y}_i - y_i \right| \quad (10)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (11)$$

$$PCCs = \frac{\sum_{i=1}^n (y_i - \frac{1}{n} \sum_{i=1}^n y_i)(\hat{y}_i - \frac{1}{n} \sum_{i=1}^n \hat{y}_i)}{\sqrt{\sum_{i=1}^n (y_i - \frac{1}{n} \sum_{i=1}^n y_i)^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \frac{1}{n} \sum_{i=1}^n \hat{y}_i)^2}} \quad (12)$$

其中  $\hat{y}_i$  为预测值， $y_i$  为真实值， $n$  为样本数。

#### (四) 评价指标检验

此外，仅从上述评价指标来比较两种模型的预测精度差异是缺乏统计依据的，因此本文参考Lee and Kim(2020)<sup>[16]</sup>、Baek and Kim(2018)<sup>[18]</sup>的做法，通过DM检验(Diebold Mariano test)验证两个模型的评价指标差异是否符合统计学上异同的意义。本文在以下内容中简要地阐述了DM检验的原理。

存在两模型的预测误差序列分别为：

$$\begin{aligned} E_1 &= \{f(y_1^1, \hat{y}_1^1), f(y_2^1, \hat{y}_2^1), \dots, f(y_T^1, \hat{y}_T^1)\} \\ E_2 &= \{f(y_1^2, \hat{y}_1^2), f(y_2^2, \hat{y}_2^2), \dots, f(y_T^2, \hat{y}_T^2)\} \end{aligned} \quad (13)$$

其中， $f(\bullet)$  为误差函数， $y$  和  $\hat{y}$  分别表示真实值和预测值，上标表示模型，下标表示序列长度。

求得误差序列的差值序列为：

$$D = E1 - E2 = \{d_1, d_2, \dots, d_T\} \quad (14)$$

其中， $d_T = f\left(y_T^1, \hat{y}_T^1\right) - f\left(y_T^2, \hat{y}_T^2\right)$ 。

求得序列  $D$  的均值  $\bar{D}$  和标准差  $\sigma(D)$

$$\bar{D} = \frac{\sum_{i=1}^T d_i}{T}, \quad \sigma(D) = \sqrt{\frac{\sum_{i=1}^T (d_i - \bar{D})^2}{T-1}} \quad (15)$$

可得DM统计量：

$$DM = \frac{\bar{D}}{\sigma(D)} \quad (16)$$

给出以下原假设和备择假设：

$H_0: \bar{D} = 0$ ，即两种模型的预测效果是相等的

$H_1: \bar{D} \neq 0$ ，即两种模型的预测效果是存在差异的

DM 检验假设统计量符合 t 分布，最后通过查询对应分布表得到检验的置信度，根据置

信度判断是否拒绝原假设。

### (五) 实证结果

为了体现Hyperband-LSTM模型在股票价格预测上的有效性，本文引入了Bayes-LSTM模型和LSTM模型作为本文的对比模型，Bayes-LSTM模型是指使用贝叶斯优化算法优化的LSTM神经网络。基于上文所讨论的实验设定，得出如表4的实验结果，由于LSTM模型没有经过优化阶段，因此没有考虑时间成本维度的评价指标。

表4 沪深300指数实验结果

模型	数据集	MSE	MAE	MAPE	PCCs	Time
Hyperband-LSTM	验证集	82.92	6.858	0.0013	0.9968	37m32s
	测试集	107.6	7.607	0.0015	0.9975	
Bayes-LSTM	验证集	215.0	11.08	0.0022	0.9913	67m19s
	测试集	322.8	12.89	0.0026	0.9916	
LSTM	验证集	1092	30.50	0.0060	0.9934	--
	测试集	864.0	26.12	0.0052	0.9945	

从实验结果看，经过优化的模型大体上要优于原LSTM模型，说明优化算法对LSTM神经网络预测的提升是有帮助的。无论在样本内的验证集上，还是样本外的测试集上，Hyperband-LSTM模型所能达到的预测精度都要优于Bayes-LSTM模型和LSTM模型，并且Hyperband-LSTM模型大大减小了模型优化训练的时间成本，可以说明Hyperband模型在效率上是有提升的。而且，Hyperband-LSTM模型在测试集和验证集中的预测指标的差距要低于Bayes-LSTM模型，表明Hyperband-LSTM模型具有较优的泛化性能。图5展示了3种模型在测试集上的预测效果，可以直观地看出Hyperband-LSTM模型的预测效果要优于Bayes-LSTM模型和LSTM模型的。因此，从实证角度来看，Hyperband-LSTM模型在沪深300收盘价预测的任务上表现出更高的预测精度和更低的时间成本，从而证实了本文所提出的Hyperband-LSTM模型在股票价格预测应用上的有效性。

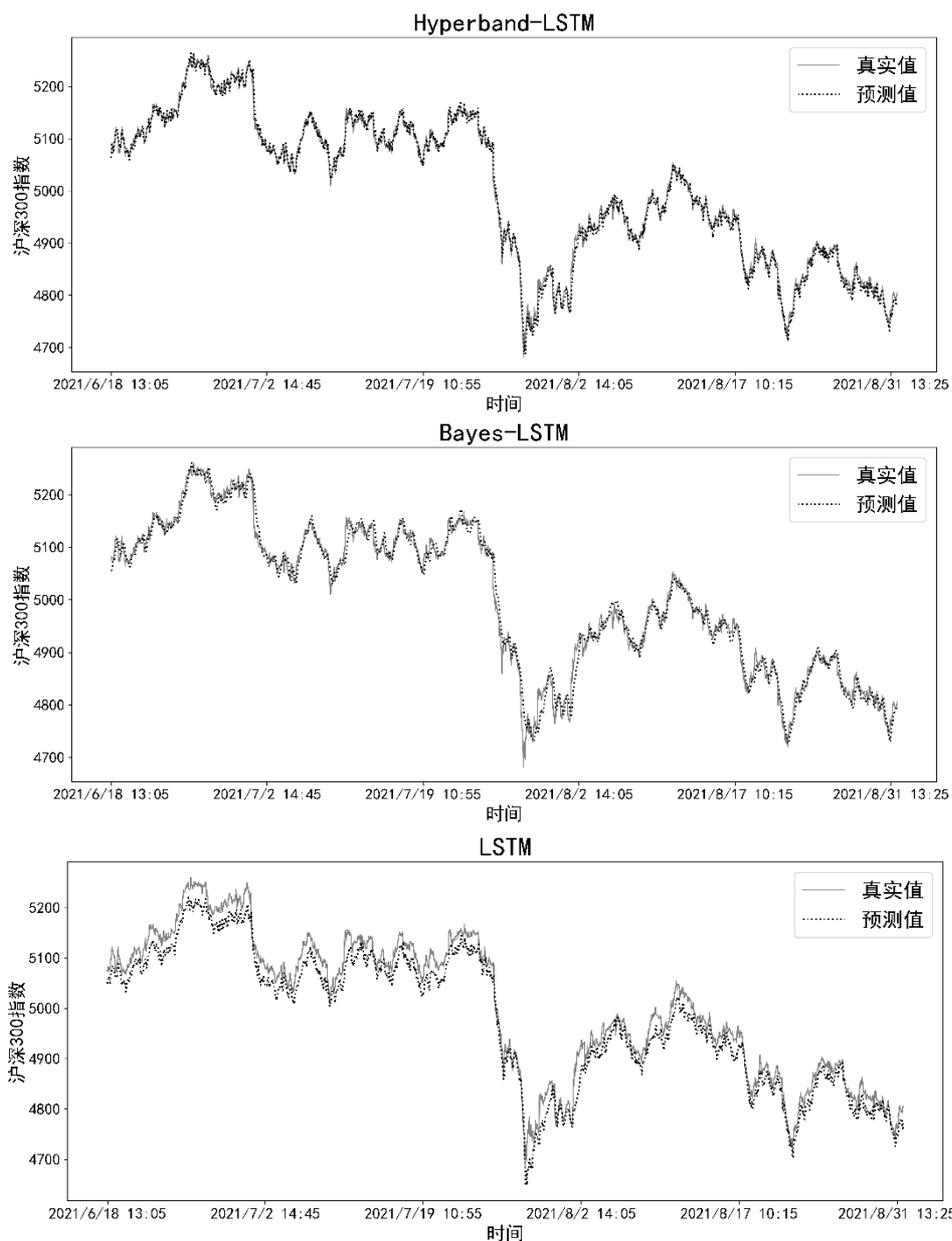


图5 预测效果图

本文使用上述3种模型在预测集中所得出的预测序列分别两两组合进行DM检验，表5为检验结果，由于篇幅问题，只展示了MAPE指标和MAE指标的检验结果。从检验结果看，两种评价指标在不同模型组合的检验值都是显著的，因此可以认为这三个模型的预测精度存在显著差异，并且Hyperband-LSTM模型与其它两种模型组合的检验统计量符号为负，表明Hyperband-LSTM模型相应的评价指标是显著小于其它两个模型，即Hyperband-LSTM模型的预测精度是显著高于其它两个模型，进一步表明了Hyperband-LSTM模型的预测能力是有所提升的。

表5 DM检验结果-沪深300

	MAPE	Hyperband-LSTM	Bayes-LSTM	LSTM
MAE				
Hyperband-LSTM	--	--	-25.01***	-68.19***
Bayes-LSTM	25.25***	--	--	-37.78***
LSTM	67.94***	38.36***	--	--

注: 表中结果为 DM 统计量, \*\*\*, \*\*, \* 分别表示 1%、5%、10%的显著性水平。

### (六) 稳健性检验

考虑到本文仅仅使用沪深300指数进行实验,并不能说明Hyperband-LSTM模型具有稳健性。因此,本文使用基于不同证券交易所上市的股票进行编制的和代表着中国两个证券市场动态和趋势的上证50指数和深证100指数进行重复实验,以检验本文模型的稳健性。两个样本的时间跨度、频率与沪深300指数样本相同,实证结果如表6所示,由实证结果观察可知本文所提出的Hyperband-LSTM模型在两个样本的预测精度、时间成本和预测泛化性都要优于Bayes-LSTM模型和LSTM模型,能与上文所进行的实验结果得出一致的结论,因此可证实本文所提出的Hyperband-LSTM模型在股票价格预测任务上具备较强的稳健性。

表6 实验结果-上证50、深证100

指数	模型	数据集	MSE	MAE	MAPE	PCCs	Time
上证50	Hyperband-LSTM	验证集	39.72	4.634	0.0013	0.9960	37m11s
		测试集	64.86	5.834	0.0018	0.9986	
	Bayes-LSTM	验证集	109.8	8.082	0.0023	0.9907	79m43s
		测试集	206.8	10.63	0.0032	0.9950	
	LSTM	验证集	470.1	20.00	0.0057	0.9934	--
		测试集	1869	28.77	0.0090	0.9731	
深证100	Hyperband-LSTM	验证集	257.4	11.76	0.0017	0.9976	37m44s
		测试集	342.9	13.39	0.0019	0.9963	
	Bayes-LSTM	验证集	689.0	19.47	0.0028	0.9936	91m40s
		测试集	994.0	22.64	0.0033	0.9888	
	LSTM	验证集	3882	58.00	0.0083	0.9942	--
		测试集	4147	59.15	0.0084	0.9921	

与沪深300指数实验目的一样,在上证50指数和深证100指数的实验中也进行了DM检验,

检验结果如表7所示。根据检验结果，两个指数在不同模型预测序列检验中的统计量都显著为负，检验结果和上文沪深300指数实验的结论一致，即Hyperband-LSTM模型的预测精度显著高于其它两种模型。因此，进一步验证了Hyperband-LSTM模型的稳健性。

表7 DM检验结果-上证50、深证100

指数	MAPE			
	MAE	Hyperband-LSTM	Bayes-LSTM	LSTM
上证50	Hyperband-LSTM	--	-24.93***	-36.86***
	Bayes-LSTM	25.47***	--	-27.33***
	LSTM	38.08***	27.73***	--
深证100	Hyperband-LSTM	--	-25.98***	-79.16***
	Bayes-LSTM	26.09***	--	-55.57***
	LSTM	78.98***	55.99***	--

注：表中结果为 DM 统计量，\*\*\*、\*\*、\* 分别表示 1%、5%、10%的显著性水平。

#### 四、结语

本文首先通过阅读前人的研究，发现在目前股票价格预测研究中能达到较好效果的一种方法是使用深度神经网络。在深度神经网络中，LSTM神经网络的预测效果总体上是优于其它深度神经网络的，因此本文选用LSTM深度神经网络进行研究。同时，本文也发现现有研究中的不足：搭建LSTM神经网络的成本太高；LSTM神经网络的预测潜能没有得到极大化的释放；表现为预测效率较低的状态。因此本文提出将Hyperband算法与LSTM神经网络模型相结合应用于股票价格预测中，弥补现有研究存在的不足。然后，本文简述LSTM神经网络和Hyperband算法的基本原理，从理论上阐明将两者结合起来的理论逻辑和将其应用于股票价格序列预测上的合理性和可行性。最后，本文通过一系列的实验证实Hyperband-LSTM模型在不同的评价指标上都是优于相应的对比模型，即Hyperband-LSTM模型在股票价格预测应用上是切实有效的。基于本文的研究，能够得出以下结论：Hyperband-LSTM模型能降低股票价格预测的实验成本；Hyperband-LSTM模型能提高股票价格预测的精确度；Hyperband-LSTM模型具有较强的稳健性，能应用于不同的股票价格预测任务。

## 参考文献

- [1] 李斌,邵新月,李玥阳. 机器学习驱动的基本面量化投资研究[J].中国工业经济,2019(08): 61-79.
- [2] Leippold M., Wang Q., and W. Zhou. Machine Learning in the Chinese Stock Market[J]. Journal of Financial Economics, 2021, 28(1):11-16.
- [3] 黄超斌,程希明. 基于 LSTM 神经网络的股票价格预测研究[J].北京信息科技大学学报(自然科学版),2021(01): 79-83.
- [4] 张贵生,张信东. 基于近邻互信息的 SVM-GARCH 股票价格预测模型研究[J].中国管理科学,2016(09): 11-20.
- [5] 潘水洋. 大数据、机器学习与资产定价[J].现代管理科学,2019(02): 6-8+33.
- [6] Jiang W.. Applications of Deep Learning in Stock Market Prediction: Recent Progress[J]. Expert Systems with Applications,2021, 184:115537.
- [7] 黎镭,陈蔼祥,李伟书,梁伟琪,杨思桐. GRU 递归神经网络对股票收盘价的预测研究[J].计算机与现代化,2018(11): 103-108.
- [8] Fischer T. and C. Krauss. Deep Learning with Long Short-term Memory Networks for Financial Market Predictions[J]. European Journal of Operational Research,2018, 270(2):654-669.
- [9] 欧阳红兵,黄亢,闫洪举. 基于 LSTM 神经网络的金融时间序列预测[J].中国管理科学,2020(04): 27-35.
- [10] Borovkova S. and I. Tsiamas. An Ensemble of LSTM Neural Networks for High-frequency Stock Market Classification[J]. Journal of Forecasting,2019,38(6):600-619.
- [11] Gu S., B. Kelly and D. Xiu. Empirical Asset Pricing via Machine Learning[J]. The Review of Financial Studies,2020, 33(5):2224-2273.
- [12] Li L., K. Jamieson, G. DeSalvo, A. Rostamizadeh and A. Talwalkar. Hyperband: A Novel Bandit-based Approach to Hyperparameter Optimization[J]. The Journal of Machine Learning Research,2018, 18(1):1-52.
- [13] [法]奥雷利安·杰龙. 机器学习实战基于 Scikit-Learn, Keras 和 TensorFlow [M].宋能辉和李娴译.北京:机械工业出版社, 2021.
- [14] 胡聿文. 基于优化 LSTM 模型的股票预测[J].计算机科学,2021 (S1): 151-157.
- [15] Ayala J., Noguera J.L.V., G.T. Miguel, G. Francisco and D. Federico. Technical Analysis

Strategy Optimization Using a Machine Learning Approach in Stock Market Indices[J]. Knowledge-Based Systems,2021, 225:107119.

[16] 党建武,从筱卿. 基于 CNN 和 GRU 的混合股指预测模型研究[J].计算机工程与应用,2021 (16): 167-174.

[17] Lee S.W. and H.Y. Kim. Stock Market Forecasting with Super-high Dimensional Time-series Data Using ConvLSTM, Trend Sampling, and Specialized Data Augmentation[J]. Expert Systems with Applications,2020, 161:113704.

[18] Baek Y. and H.Y. Kim. ModAugNet: A New Forecasting Framework for Stock Market Index Value with an Overfitting Prevention LSTM Module and a Prediction LSTM Module[J]. Expert Systems with Applications,2018, 113: 457-480.