

doi: 10.3969/j.issn.0490-6756.2020.01.004

# 指数趋势预测的 BP-LSTM 模型

孙存浩, 胡兵, 邹雨轩

(四川大学数学学院, 成都 610064)

**摘要:** 本文根据股指、股价等数据的时序特征将人工神经网络(ANN)与深度学习中的循环神经网络(RNN)引入股指预测, 基于 BP 神经网络模型与长短期记忆(LSTM)神经网络模型构建了 BP-LSTM 模型. 基于上证指数, 本文进行了数值实验. 结果表明 BP-LSTM 预测模型的准确率相比传统机器学习模型有明显提升, 与普通 LSTM 模型相比也有较大提升.

**关键词:** BP 神经网络; 长短期记忆神经网络; 上证指数趋势预测

**中图分类号:** O29 **文献标识码:** A **文章编号:** 0490-6756(2020)01-0027-05

## A BP-LSTM trend forecast model for stock index

SUN Cun-Hao, HU Bing, ZOU Yu-Xuan

(School of Mathematics, Sichuan University, Chengdu 610064, China)

**Abstract:** In this paper, according to the time series characteristics of financial data such as stock index and stock price, and introducing the Artificial Neural Network (ANN) and Recurrent Neural Network (RNN) in deep learning to stock index prediction, we build the BP-LSTM model based on Back Propagation (BP) neural network model and Long Short-Term Memory (LSTM) neural network model. Numerical analysis shows that the accuracy of BP-LSTM prediction model is significantly higher than the traditional machine learning models and the LSTM model.

**Keywords:** Back propagation neural network; Long short-term memory neural network; Shanghai composite index trend forecast

## 1 引言

随着计算机计算能力的提升, 机器学习及神经网络在越来越多领域得到应用.

神经网络主要包含 BP(back propagation)神经网络, 卷积神经网络(convolutional neural networks, CNN)和循环神经网络(recurrent neural network, RNN)等. 其中, BP 神经网络已被应用于搜索引擎、推荐系统优化等方面, CNN 被应用于图像识别与目标检测, 而 RNN 在处理自然语言数据等时序信息时也有良好的表现.

近年来, 神经网络在金融领域的运用逐渐增加. 越来越多的研究者开始使用神经网络进行金融数据处理与预测. 谭朵朵<sup>[1]</sup>使用 BP 神经网络进行了指数期权定价, 说明 BP 神经网络相比传统 BS (black-scholes)公式有更好的适用性, 但是没有考虑期权价格数据的时序性. 陈佳等<sup>[2-4]</sup>针对标的价格的时序特征使用 LSTM 模型进行了价格预测, 说明 LSTM 模型在进行金融时序预测时相比于传统方法具有更高的预测精确度, 但其仅使用历史价格作为主要输入特征, 特征维度单一, 不能全面地反映市场状态. 文宇<sup>[5]</sup>提出了 CNN-LSTM 模型,

收稿日期: 2019-04-09

基金项目: 国家自然科学基金(11401407)

作者简介: 孙存浩(1994-), 男, 四川成都人, 硕士生, 主要研究方向为计算数学. E-mail: sch\_math@163.com

通讯作者: 胡兵. E-mail: hubingscu@scu.edu.cn

对金融二级市场数据进行分析,根据标的的历史价格和成交量信息扩展了更多维度的特征,并用 CNN 模型进行了空间上的特征提取,但是 CNN 模型在处理全局性的特征时具有一定的局限性。

考虑到股指数据高时序依赖、多组合特征的特点,本文针对股指数据的时序特征提出了一种基于 BP 神经网络与长短期记忆(long short-term memory, LSTM)神经网络的 BP-LSTM 模型。其中, BP 神经网络主要负责对日内全局特征进行特征提取与特征组合,将输入数据自适应地组合成特征传入 LSTM 神经网络;LSTM 神经网络提取特征中的时序特征,在处理具有强时序趋势的数据方面有显著的效果;该模型最后通过一层带有 sigmoid 激活函数的 BP 神经网络将输出结果非线性地映射到(0,1)区间内,由此对次日股票指数涨跌情况进行预测。实证分析的结果表明, BP-LSTM 模型比逻辑回归、梯度提升树<sup>[6]</sup>等传统预测方法及 LSTM 模型等神经网络方法具有更高预测准确率。

## 2 预备知识

### 2.1 人工神经元与 BP 神经网络

下图是一个从生物神经元中得到启发被设计出的 M-P 神经元<sup>[7]</sup>的示意图。

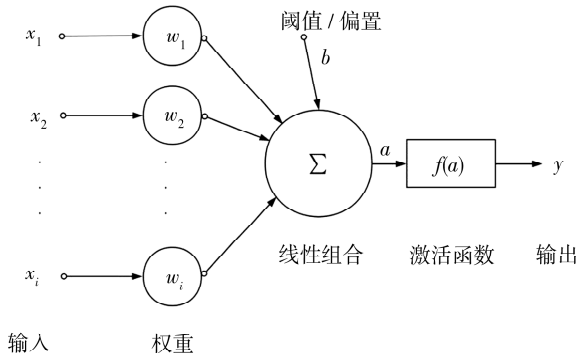


图 1 M-P 神经元  
Fig. 1 M-P neural cell

神经元正向传播的过程为

$$y = f(a) = f\left(\sum_{i=1}^n \omega_i x_i - b\right) \quad (1)$$

神经元接收  $n$  个输入信号(这些输入信号带有一定的权重进行传递),并将接收加权后的总输入与该神经元的阈值进行对比,通过激活函数处理产生输出,把多个这样的神经元按照一定的层次结构连接起来就形成了神经网络。神经网络理论上可以以任

意精度自适应非线性逼近连续函数。1989 年, Hecht-Nielsen<sup>[8]</sup>证明了单隐藏层的 BP 神经网络可以逼近闭区间内的任意连续函数,即万能逼近定理。

严格意义下的激活函数  $f$  应为单位阶跃函数,它将我们的输入映为“0”或“1”。然而,由于阶跃函数非连续,我们常用 Sigmoid 函数作为激活函数,它将实数域内的输入映射到 (0,1) 区间内。

在训练时,神经网络先输入数据进行正向预测,再结合预测结果与实际结果求误差并进行逆传递,通过不断地以梯度下降的方式调节神经元中的权重与阈值,来提高预测的准确率。

常见的 BP 神经网络是形如图 2 所示的层级结构,每层神经元都与下一层的神经元完全互连,神经元之间不存在同层、跨层连接的情况。在图 2 中,输入层第  $i$  个神经元与隐层第  $h$  个神经元间的权重与隐层第  $h$  个神经元的阈值分别用  $v_{ih}$  与  $\gamma_h$  表示,隐层第  $h$  个神经元与输出层第  $j$  个神经元间的权重与输出层第  $j$  个神经元的阈值分别用  $\omega_{hj}$  与  $\theta_j$  表示,神经网络的“学习”过程就是根据输入的训练数据调整神经元之间的权重以及阈值。

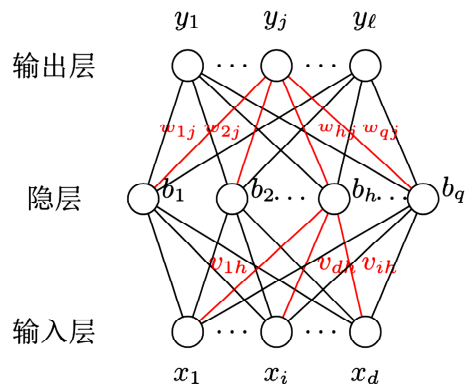


图 2 单隐层 BP 神经网络

Fig. 2 Back propagation neural network with one hidden layer

### 2.2 RNN 与 LSTM 神经网络

RNN<sup>[9-10]</sup>是一种特殊的神经网络,它的算法考虑了数据中存在的时序信息,具有时间维度,因而常用于对被对具有时序规律的数据进行建模。一个常见的 RNN 结构如图 3 所示:每个 RNN cell 会做定向循环把某些信息传到下一个时间的隐藏层中,从而保留数据之间的时序关系,有效处理时序数据。

然而,RNN 在进行循环的过程中,随着循环次数的增大,反向传播时往往会出现梯度极大(梯度

爆炸)或者梯度极小(梯度消失)的情况,因而无法处理过长的时序信息,否则会忘记较长期的信息。

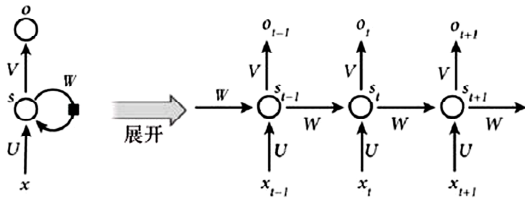


图 3 RNN 网络结构和展开结构

Fig. 3 RNN network structure and expansion

在这种情况下,人们基于 RNN 的设计提出了 LSTM 神经网络. 相比于传统的 RNN, LSTM 能够处理更长的信息序列。

LSTM 神经网络是 RNN 的一种变体,最早由 Hochreiter 与 Schmidhuber 在 1997 年提出<sup>[11]</sup>,设计初衷是解决神经网络中的长期依赖问题. LSTM 神经网络针对很多时序中间隔和延迟较长的问题(例如语音识别、智能问答、机器翻译、视频分类及情感分析等)都有较好的表现,此外它还可用于服务器响应时间预测<sup>[12]</sup>. LSTM 神经网络的基本结构与 RNN 的结构类似,如图 4 所示。

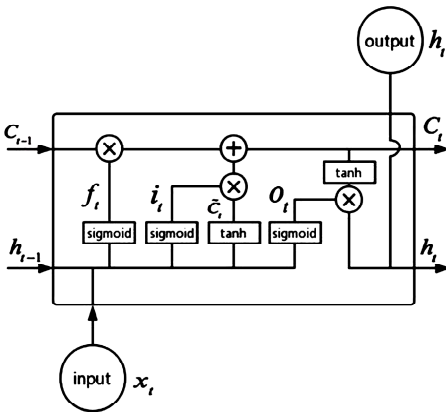


图 4 LSTM 网络结构

Fig. 4 Structure of LSTM network

### 3 BP-LSTM 模型

我们将数据与特征传入一个 BP 神经网络,通过神经网络进行特征重组,然后将重组后的特征输出到 LSTM 网络中来提取数据在时间序列上的特征,最终通过有一个带 Sigmoid 激活函数的神经元进行分类结果输出,这就是 BP-LSTM 模型,见图 5。

#### 3.1 模型构建

3.1.1 数据收集与特征选择 数据的收集与特征的选择对一个模型的表现至关重要. 合适的特征处

理不仅可以提高模型的准确率,还可以加快收敛速度. 根据金融相关知识,我们将数据特征分为两类: (i) 上证指数的交易数据,主要是指数的开、高、低、收与交易量等指标; (ii) 相关的技术指标,基于统计学的研究,以数据来论证股票的趋向、买点卖点,较好地量化市场状态. 我们所用指标主要是基于历史数据走势计算出的 KDJ、MACD、BIAS、AMA、MOM 以及部分根据个人经验设计的指标,共 25 维,按照时间顺序有序传入模型进行预测。

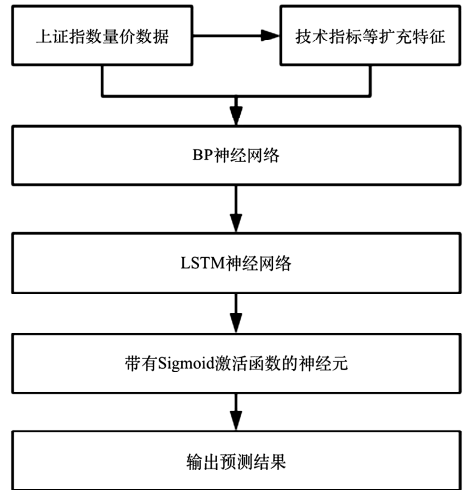


图 5 BP-LSTM 模型结构示意图

Fig. 5 Structure diagram of BP-LSTM model

对收集的数据及数值特征还进行了如下的归一化处理. 考虑到部分特征非负的特点,利用特征数据的最大值与最小值,将特征的值缩放到  $[0, 1]$  区间:

$$x_{\text{norm}}^{(i)} = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (2)$$

需要注意的是,此处归一化使用的最大值与最小值均为训练集中数据. 在测试集上进行测试时,也根据训练集数据进行归一化,有效防止数据泄露问题(即在当前时间可以获得未来的数据). 此类归一化作用不仅可以增强模型在小方差数据集上的稳定性,还可以维持稀疏特征中为 0 的样本. 实验结果也显示带有预先数据归一化的模型预测表现较不带预先归一化的模型更好。

3.1.2 构建用于特征处理的全连接神经网络 全连接神经网络可以全局接收传入模型的  $n$  维的数据与特征,通过单层含有  $m$  个神经元的全连接神经网络作用输出  $m$  维数据,以此来进行特征提取与特征降维(当  $m < n$  时). 这样的设计不仅使模型自动提取重要特征、进行特征融合,还在保证特征信息损失最少的前提下控制传入下一层长短期记

忆神经网络的特征维度,有效控制特征的复杂度,对缩短训练时间、加快模型收敛速度与提升模型预测效果有显著作用.接着,对全连接神经网络的  $m$  维输出做批归一化处理<sup>[13]</sup>可以进一步提升模型的收敛速度.

**3.1.3 构建处理时序数据的长短期记忆网络** 长短期记忆网络将全连接神经网络的输出按日期顺序传入长短期记忆网络中进行训练,有效处理时间序列问题与循环神经网络不同,长短期记忆网络可以学习更长的时间序列数据.我们分别对 7 天、15 天的历史数据集进行了建模预测,证明更长的数据周期对长短期记忆网络模型的预测准确度有较明显的提升.我们主要使用的是带 64 个 cell 的长短期记忆网络模型,即经过神经网络处理会输出 64 维的输出特征.

**3.1.4 构建用于数据输出的神经元** 数据输出由单个神经元负责.它接收长短期记忆网络输出的 64 维特征.将其与自带权重及阈值进行线性组合,通过 Sigmoid 激活函数将结果映射到(0,1)区间内即为对下一日上证指数涨跌的预测结果.

**3.2 模型训练与预测**

本次实验所用数据为 1991 年 3 月 4 日至 2018 年 12 月 27 日上证指数日量数据(数据来源: Tushare).其中,前 80%交易日的数据作为训练集与验证集,后 20%交易日的数据作为测试集,以保证测试数据在天级别的连续性.经统计,训练集与测试集中正、反例分布均匀,没有数据倾斜情况.我们选用的 Adam 优化器<sup>[14]</sup>比传统的基于梯度下降的 SGD 优化器具有更快的收敛速度,并且可以根据数据分布自适应调整学习率.模型选用的损失函数为:

$$L(y, y') = - \sum_{i=1}^n [y'_i \log y_i + (1 - y'_i) \log (1 - y_i)] \tag{3}$$

为防止过拟合问题,除了在长短期记忆网络的输出层加入 Dropout 层<sup>[15]</sup>之外,我们还在训练过程中加入了 Early Stopping<sup>[16]</sup>处理,使模型仅保留在验证集上最佳表现的参数.

**3.3 模型分析与评估**

图 6 为 BP-LSTM 网络的模型架构.

在特征工程环节我们选取 25 个维度的技术指标反映市场状态,并用有 25 个隐藏层的 BP 神经网络对特征进行自适应组合,然后将原始输入特征进行组合后传入 LSTM 神经网络进行时序信息学习. LSTM 神经网络作为 RNN 的变体对时序问题

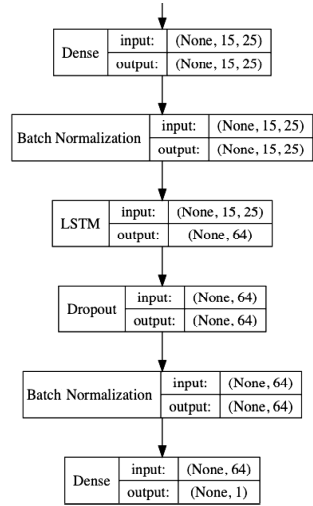


图 6 BP-LSTM 模型架构图

Fig. 6 Architecture diagram of BP-LSTM model

同样有好的学习能力,且多个门的设计有效解决了梯度消失与梯度爆炸的问题,对学习长时间序列数据有更好的作用.最后,我们通过单个含有 Sigmoid 激活函数的神经元对数据进行非线性输出. BP-LSTM 模型的训练同样通过反向传播完成,模型中的每个参数都可以通过反向传播计算梯度进行更新.为了防止训练停止在局部最小值点,我们使用 Adam 优化器代替 SGD.这也是一个效果很好的优化算法.

模型评估使用的指标是预测的准确率与 F1 值,计算方式如表 1 所示.

表 1 评估指标

Tab. 1 Evaluation index

指标名称	计算方式
准确率	$\frac{TP+TN}{TP+FP+TN+FN}$
精确率 (precision)	$\frac{TP}{TP+FP}$
召回率 (recall)	$\frac{TP}{TP+FN}$
F1 值	$\frac{2 * precision * recall}{precision + recall}$

其中 TP、TN、FP 和 FN 可以通过图 7 的混淆矩阵得到.

		True class	
		p	n
Hypothesis output	Y	TP (Ture Positives)	FP (False Positives)
	N	FN (False Negatives)	TN (True Negatives)
Column counts:		Pc	Nc

图 7 混淆矩阵

Fig. 7 Confusion matrix

### 3.4 数值实验

实验主要分为三类,分别为:在相同的数据集上构建传统的逻辑回归模型和梯度决策树模型,普通 LSTM 模型及带有初始数据归一化(表中以 processing 标示)的 BP-LSTM 模型.各模型预测准确率与 F1 值如表 2 所示.

表 2 不同模型的预测结果

Tab.2 Prediction results of different models

模型	准确率	F1 值
逻辑回归	51.72%	49.5%
梯度提升树	51.50%	44.82%
7day LSTM	53.89%	63%
15day LSTM	54.01%	65%
7 day processing BP-LSTM	83.8%	85.3%
15 day processing BP-LSTM	90.43%	91.41%

对比各项指标,基于较长期数据的 BP-LSTM 模型在上证指数趋势预测方面具有更高的准确率.实验表明 BP-LSTM 模型是一种比较有效的方法.

## 4 总 结

本文提出了基于 BP 神经网络与 LSTM 神经网络的 BP-LSTM 模型,模型相比于传统的机器学习模型及普通的 LSTM 神经网络具有更高的预测准确率.当然,鉴于本文所使用的数据及特征全部基于上证指数历史价量数据,缺乏更多维度的特征(如市场情绪、宏观决策等),在后续研究中我们还可以收集互联网上股民留言数据,使用 LSTM 网络进行市场情绪分析,然后加入本文的特征数据集中,以得到准确率更高的结果.

### 参考文献:

- [1] 谭朵朵. 基于 BP 神经网络的 S&P500 指数期权定价[J]. 统计与信息论坛, 2008, 11: 40.
- [2] 陈佳, 刘冬雪, 武大硕. 基于特征选取与 LSTM 模型的股指预测方法研究[J]. 计算机工程与应用, 2019, 55: 108.

- [3] 任君, 王建华, 王传美, 等. 基于正则化 LSTM 模型的股票指数预测[J]. 计算机应用与软件, 2018, 35: 44.
- [4] 龙奥明, 毕秀春, 张曙光. 基于 LSTM 神经网络的黑色金属期货套利策略模型[J]. 中国科学技术大学学报, 2018, 48: 125.
- [5] 文字. 基于 CNN-LSTM 网络分析金融二级市场数据[J]. 电子设计工程, 2018, 26: 75.
- [6] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [7] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity [J]. Bull Math Biophys, 1943, 5: 115.
- [8] Hecht-Nielsen R. Theory of the backpropagation neural network [M]. New York: Academic Press, 1992: 65.
- [9] Jordan M I. Serial order: a parallel distributed processing approach [M]. Amsterdam; North-Holland, 1997.
- [10] Elman J L. Finding structure in time [J]. Cognitive Sci, 1990, 14: 179.
- [11] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Comput, 1997, 9: 1735.
- [12] 刘承启, 林振荣, 黄文海. 基于 LSTM 的 WEB 服务响应时间大数据预测方法[J]. 四川大学学报: 自然科学版, 2019, 56: 71.
- [13] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [J]. ArXiv: 1502.03167, 2015.
- [14] Kingma D P, Ba J. Adam: a method for stochastic optimization [J]. ArXiv: 1412.6980, 2014.
- [15] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. ArXiv: 1207.0580, 2012.
- [16] Prechelt L. Automatic early stopping using cross validation: quantifying the criteria [J]. Neural Networks, 1998, 11: 761.

### 引用本文格式:

中文: 孙存浩, 胡兵, 邹雨轩. 指数趋势预测的 BP-LSTM 模型[J]. 四川大学学报: 自然科学版, 2020, 57: 27.

英文: Sun C H, Hu B, Zou Y X. A BP-LSTM trend forecast model for stock index [J]. J Sichuan Univ: Nat Sci Ed, 2020, 57: 27.